Proceedings of

Measuring Behavior 2012

8th International Conference on Methods and Techniques in Behavioral Research

Utrecht, The Netherlands
August 28-31, 2012

Editors
Andrew Spink
Fabrizio Grieco
Olga Krips
Leanne Loijens
Lucas Noldus
Patrick Zimmerman

Noldus
Information Technology

2012

Bibliographical reference

Proceedings of Measuring Behavior 2012, 8th International Conference on Methods and Techniques in Behavioral Research (Utrecht, The Netherlands, August 28-31, 2012)

Edited by A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, and P.H. Zimmerman

ISBN 978-90-74821-87-2

Copyright

P.O. Box 268

6700 AG Wageningen

The Netherlands

www.noldus.com

Printed in the Netherlands

Table of Contents

Keynote Lectures

Special sessions

Measuring Driver and Pilot Behaviour

From Classic to Automated HomeCage Phenotyping

Measuring Engagement: Affective and Social Cues in Interactive Art and Media

Tests for Mild Cognitive Impairment

The Role of Behavior Measurement in Persuasive Settings

Social Behavior and Communication - From Mice to Primates

Electrophysiological Correlates Of Behaviour

The Development Of A Diverse Battery Of Behavioral Tasks Using Touchscreen Equipped Operant Boxes for the Study Of Cognition In The Rodent

Technical Support for Analysis of Human Error in Task Performance

Measuring Behavior in a Game Context

Progress in Assessing Animal Welfare in Relation to New Legislation: Opportunities for Behavioural Researchers

Fish in Behaviour Research: Unique Tools with a Great Promise!

General sessions

Advanced Statistical and Other Analysis of Behavioral Data

Measurement of Human Behavior and Interaction in Natural Contexts

Measurement of Animal Behavior in Natural and Semi-natural Contexts

Automation and Analysis of Rodent Tests in Mazes

Use of Sensors and Other Techniques to Automate the Measurement of Behaviors

Measuring Human-Computer Interaction

Posters

Workshops

Welcome to *Measuring Behavior 2012*

We are delighted to welcome you all to the 8th *Measuring Behavior* conference. Building on the format that has emerged from previous meetings, this year we have also endeavored to bring together an international audience engaged in the development and validation of methods for recording and understanding behavior and actions in their broadest sense. We thank all of you for bringing your best and most exciting science.

This year sees the return of *Measuring Behavior* to its birth place, Utrecht. It was here, that two visionaries, Berry Spruijt and Lucas Noldus, realized their ambition to bring together the scientific community interested in the objective recording and interpretation of behavior. Both remain actively involved in the further development of this field and play significant roles in this meeting, and the scientists present here today owe them a great deal of foresight and vision. Utrecht University with its international reputation in behavior and veterinary science was the ideal breeding ground for the conference. The measuring of behavior has always been an important topic in its research, and is currently reinforced in the University's strategic themes Life Sciences and Youth & Identity.

Scientifically, this year's program encapsulates the recent conceptual progress in the development of translational tools highly relevant for medical research, combined with technical advances in terms of hard- and software for improved recording and as a response to the ever increasing challenges of data analyses. Encompassed are Special Sessions dedicated to selected topics, Workshops, User Meetings, Tutorials, Demonstrations, General sessions and Poster presentations representing the global endeavors to record, quantify and understand behavior. Highly recommendable are also the Keynote lectures delivered by three outstanding and highly respected speakers. Finally, we trust you will enjoy the scientific tours in the Utrecht region and the social events, into which a lot of hard work has gone by the Local Organizing Committee.

We hope this program caters for many of your interests and we look forward to seeing and hearing your contributions and trust it will become a productive, exciting and memorable conference.

Andrew Spink (Noldus Information Technology, The Netherlands)

Gernot Riedel (University of Aberdeen, United Kingdom)

Remco Veltkamp (Utrecht University, The Netherlands)

Berry Spruijt (Utrecht University, The Netherlands)

Conference Chairs

The *Measuring Behavior* Conferences

Measuring Behavior is a unique conference about methods and techniques in behavioral research. While most conferences focus on a specific domain, *Measuring Behavior* creates bridges between disciplines by bringing together people who may otherwise be unlikely to meet each other. At a *Measuring Behavior* meeting, you find yourself among ethologists, behavioral ecologists, neuroscientists, experimental psychologists, human factors researchers, movement scientists, robotics engineers, software designers, electronic engineers, human-computer interaction specialists… to name but a few. Experience tells us that the focus on methodological and technical themes can lead to a very productive cross-fertilization between research fields. Crossing the boundaries between disciplines and species (from astronauts to zebras) can be extremely inspiring.

Measuring Behavior started in 1996 as a workshop in the framework of a European research project “Automatic Recording and Analysis of Behavior”, aimed at sharing the results of our project with colleagues from abroad. Organized by Noldus Information Technology and hosted by Utrecht University, *Measuring Behavior ’96* attracted over 150 participants. From that modest beginning, the conference has grown to a significant international event with several hundred delegates from thirty plus countries. We have also grown in terms of the scientific quality of the conference, with selection of papers now being determined by a process of independent peer-review by many hundreds of reviewers. The scientific program committee is very grateful for all that work that many of you reading this have contributed towards. In 2012, we return to our ‘birthplace’ and, appropriately, our first conference chair, Prof. Berry Spruijt, is honorary chair this year.

Noldus Information Technology serves as conference organizer and main sponsor. For a small company like ours, the conference is a major investment. We gladly do this, because we believe that the focused attention on behavior research methods and techniques will eventually lead to a higher demand for our solutions. To prevent commercial bias, however, the scientific program is put together under auspices of an independent Scientific Program Committee, consisting of international experts from a broad variety of disciplines (see the Scientific Program Committee on page 520) and many different research groups and companies have contributed and participated in this series of conferences.

Over the years, the conference has been hosted by a variety of universities:

Year	**City**	**Conference chair**
1996	Utrecht	Berry Spruijt
1998	Groningen	Jaap Koolhaas
2000	Nijmegen	Alexander Cools
2002	Amsterdam	Gerrit van der Veer
2005	Wageningen	Louise Vet
2008	Maastricht	Harry Steinbusch
2010	Eindhoven	Boris de Ruyter
2012	Utrecht	Remco Veltkamp & Gernot Riedel

In the scientific program, which is well balanced between human and animal research, you can find a variety of formats for presentation, interaction and exchange of information. In the past years we have seen that the special sessions (with speakers invited by session chairs) have become more prominent, and also the Demonstration Showcase has become more popular.

Measuring Behavior is a scientific conference, so special attention is paid to publication of the work presented at the meeting. The format of papers in the Proceedings is always a matter of debate, due to the different conventions of the many disciplines represented at the conference. After trying a variety of formats over the years, we have settled on extended abstracts (1-2 pages for posters and 2-4 pages for oral presentations), and that seems to be the best compromise between a text with enough content for both lasting value and being possible to review and short enough so that the effort for the writers is not excessive. An important feature of the conference Proceedings is that they are all available as open access from measuringbehavior.org. As usual, we will be having post-conference publications as special editions of selected journals, and we hope to be able to announce which journals are selected during the conference itself.

Now you find yourself at the 8th *Measuring Behavior* conference. The organizers have done their best to prepare an optimal mix of scientific, technical, and social ingredients. We hope that you will find *Measuring Behavior 2012* a rewarding and stimulating experience and wish you a pleasant stay in Utrecht.

Lucas P.J.J. Noldus

Managing Director, Noldus Information Technology bv

l.noldus@noldus.nl

Andrew Spink

Chair of Scientific Program Committee, Measuring Behavior 2012

Andrew@measuringbehavior.org

Measuring Behavior 2012: Highlights of the scientific program

The eighth edition of *Measuring Behavior* covers an even wider diversity of scientific topics than its predecessors. Within that diversity, some clear trends are visible. Tracking technologies are becoming more important. A few years ago, video tracking dominated the field, and although advances are being made in automatic behavior detection from video images, the tracking itself now uses a variety of different techniques as (for instance) GPS signals become more accurate and use lighter sensors. The rule remains that there is not much point collecting behavioral data if you cannot analyze and interpret it, and there are papers on analyzing both spatial data and other sorts of behavioral data. Indeed, technology and science constantly influence each other.

Another trend over the past few years is for physiological and other sensors to be used more and more in combination with traditional behavioral measures. New technologies enable less intrusive and more accurate measurements, and as those measurements have become more common, so the interpretation of them has become more reliable. As in many areas in the field, this is one domain where insights in controlled conditions studying animal behavior have spilled over into both studies of human behavior, and studies in less confined contexts. As well as physiological data, signals in terms of auditory (and other) communication, the measurement of engagement, interactions with touch screens and capturing human behavior interacting with computers and other machines has really moved into the mainstream. Human factor research has become such an important domain that a number of different sessions are dedicated to it, including control of cars and planes, interaction of (and engagement with) games, and measurement and analysis of errors in peoples' interaction with software.

The above list of domains indicates that the proportion of behavioral measurement carried out in purely artificial environments is increasingly small. Most of the psychological advances presented at *Measuring Behavior* are experiments carried out in realistic environments, and especially in the context of animal research, the developments of home cages has allowed for the assessment of a much richer and natural range of behaviors (ethogram) than was present a few years ago. The importance of the possibility of a subject displaying a natural behavioral repertoire has become increasingly clear to researchers, and with it the necessity for good welfare conditions (in addition to the ethical imperative). *Measuring Behavior* has increased the welfare requirements for submitted papers this year and there is also a special session devoted to this subject.

Animal research is sometimes carried out because the animals' behavior is interesting in itself, and at other times because it may act as a model for human behaviour or disease. In a few cases that can best be achieved by developing new animal models (see the example of zebra fish as they have a better developed visual than olfactory cortex, and are in this respect a better model for human cognition than the traditional rodent models) and in some other cases by further development of existing models to critically examine their translational value.

Technical and methodological development critically relies on ‘hands-on’ experience, that is, the knowledge of what each subject is capable of and how training procedures can be implemented so that data are gathered unambiguously reporting on the research questions to be answered. This know-how comes to a great extent from MSc or PhD students. This year, we sought an even stronger incorporation of this group of researchers, and so we have reduced the student registration fees, increased the number of student travel grants and introduced a prize for the best student presentations.

Measuring Behavior 2012 uses the same presentation formats that have been built up over the past seven conferences. In 2012, most presentations will take place in special sessions (previously called symposia and special interest groups), in which the session chair has invited speakers. By contract, the general sessions (full papers) and posters are submitted as individual abstracts and clustered together by the conference organization. Tutorials are longer presentations teaching existing techniques, and demonstrations showcase new prototypes of software and hardware relevant to our research. The conference program is completed by four types of events outside the scientific program; user meetings, scientific tours, a commercial exhibition and social events.

In 2012, *Measuring Behavior* returns to the city and university where the conference started 16 years ago. Utrecht University is the oldest university in the country (founded 1636) and in the Shanghai ranking comes out as the best in the country. It has an outstanding record in behavioral research and is a very fitting location for us to return to.

Andrew Spink

Chair of Scientific Program Committee, Measuring Behavior 2012
Andrew@measuringbehavior.org

Gernot Riedel

Conference Chair, Measuring Behavior 2012
G.Riedel@abdn.ac.uk

Remco Veltkamp

Conference Chair, Measuring Behavior 2012
R.C.Veltkamp@uu.nl

Keynote lecture

Professor Yadin Dudai

Weizmann Institute of Science, Rehovot, Israel; New York University

About the speaker

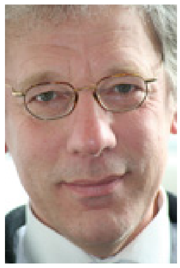
Yadin Dudai is the Sela Professor in Neurobiology at the Weizmann Institute of Science, Rehovot, Israel, the Albert and Balanche Willner Family Global Distinguished Professor of Neuroscience, New York University. He studied biochemistry and genetics, with supplements in modern history, at the Hebrew University in Jerusalem, received his Ph.D. in Biophysics from the Weizmann Institute of Science, and conducted his postdoctoral training at the California Institute of Technology, Pasadena, where he had been on the team that pioneered neurogenetic analysis of memory. Over the years he has been a visiting Professor at many academic and research institutions in the US and Europe. With research interests that include brain and behavioral mechanisms of learning and memory, as well as their interaction with culture, Prof. Dudai has over 190 professional publications in the field of brain and memory, including key books in the discipline. He has been awarded numerous honors, is a member of professional bodies in the fields of science, education, and science-society interactions, and serves on the boards of scientific journals in the neural- and cognitive sciences. Prof. Dudai also has professional experience in administration and R&D planning. He has held multiple posts in public and academic life, including Advisor on Science Policy in the Prime Minister's office in Jerusalem, member of the Granting and Planning Committee of the Israeli Council for Higher Education, member of the Israeli and International Pugawsh group, chair of the life-sciences teaching program and of the postdoctoral programs at the Feinberg Graduate School of the Weizmann Institute of Science, Dean of the Faculty of Biology and Chair of the Department of Neurobiology and of the Brain Research Centers at the Weizmann Institute. Prof. Dudai also serves as the scientific director of the Israeli Center for Research Excellence (I-CORE) in the cognitive sciences.

The Neurobehaviorist Paradox: When Knowing More is Less

The stunning developments in molecular and cellular methodologies bring behavioral analyses of so called “model organisms”, ranging from worms to rodents, into the toolbox of every self-respecting biology laboratory. Yet in most cases, a dissonance is apparent between the depth of analyses of the molecular mechanisms on the one hand and that of the behavior on the other. To this date it is not rare to encounter reliance on a single behavioral measure as if it is a litmus test, evading the complexity underlying the behavior in real-life. Paradoxically, however, some major findings linking molecules to behavior found their way into the cannon of the scientific literature because the lack of detailed knowledge about the behavioral complexity allowed the investigators to reach their conclusions. Had the behavior of model organisms been known at the same resolution as that of the experimenter, the authors would have been much more reluctant to link identified molecular mechanisms to distinct attributes of normal and pathological behavior. I will discuss the tension between knowing too little and knowing too much about behavior, and illustrate how in the neuroscience, one might wish to adhere to that level of resolution of behavior that is just useful enough to allow productivity without sacrificing realism on the one hand but boldness on the other.

Keynote lecture

Professor Berry M. Spruijt



Delta Phenomics, Utrecht, the Netherlands and Department of Biology, Utrecht University, Utrecht, the Netherlands

About the speaker

Prof. Dr. B.M. Spruijt is holds a Chair in Ethology and Welfare, which involves research into the sensitivity of reward systems and positive emotions. This research is aimed at the assessment of welfare from the perspective of the animal. An important premis is that for the early detection or prevention of welfare problems one should measure the animal's appraisal of its environment rather than wait until post hoc pathological symptoms of chronic stress have been developed. The need for welfare measurement in, for instance, livestock husbandry practices at a large scale, and the need for objective measurements inspired an already long existing interest in automation and valorisation of ethological procedures. Furthermore, also in behavioural neuroscience, behavioural pharmacology and toxicology an urgent need for less laborious and more reliable and reproducible methodology exists, being more animal friendly at the same time. For rodents, a system (PhenoTyper®) has been developed that prevents the influence of confounding factors such as the presence of human observers, transport before the test, etc. (Spruijt and de Visser, 2006; de Visser et al., 2007, 2006). It also allows the distinct assessment of reactions to novelty, anxiety and cognitive performance as the system is equipped with various programmable (aversive and rewarding) stimuli to activate various behavioural systems in the home cage.

The collaboration with Noldus Information Technology (Wageningen, the Netherlands) resulted in the founding of a company Delta Phenomics, offering contract research services for pharma, food companies and other research institutes.

Back to the future II - Validation of paradigms of the past and technology of the future

The first open field study for rodents was conducted 1934; the issue of what is exactly measured in such an open field test has been addressed in the first review written in 1973[1]. In a recent review Haller & Alicki [2] mention that the open field is still a frequently used test nowadays, despite innovative alternatives. The open field is used for exploratory behaviour, habituation, anxiety depression, schizophrenia, but also for more pharmacological questions. The ethogram used has always been a topic of debate. Technological innovations in molecular biology, biochemistry electrophysiology and histochemistry have enormously facilitated the level of resolution of independent variables, the acceleration of production of results and yielding new read out parameters and new insights. Now, the question emerges: Are there comparable benefits from technological innovations for behavioural science in terms of new parameters, results and insights as well? What has changed in the course of 80 years?

The introduction of the Skinner-box with its increasingly automated electro-mechanical devices is indistinguishable connected with the behaviouristic approach and has induced a discipline on its own already a long time ago. According to a review of Haller and Alicki2 all innovations are mainly used by the innovators themselves. In fact, the oldest tests are still most frequently used, almost in the same way for decades. Has it just been a mere improvement in efficiency and costs replacing the human hand and eye by a device? This is more than a rhetoric question.

The topic of this paper highlights the necessity to validate novel technologies in behavioural sciences, an attempt that is hampered by the fact that the classical behavioural tests are poorly validated.

The driving force behind the use of well-known tests for studying the proximate causation of behaviour is the scientific discipline of comparative or translational research which uses animal models for human health and medical care. Of all 2.637.720 behavioural studies (including economics etc.) 775.662 are on human behaviour, 123.726 are on rat behaviour and 59.320 are on mouse behaviour. So by far, animal behaviour means ’rodent’ behaviour studied as model for man in neuroscience and preclinical research.

Oversimplified ethological set ups are characteristic for laboratory tests: next to Skinner boxes, various mazes, open field, light dark boxes, conditioned fear shock box, defensive behaviour paradigms, attention tasks, all are classical tests. While the Skinner box is limited in its underlying assumptions resulting in conceptual constraints, ignorance of genetically determined natural behaviours and focus on a few artificial read out parameters (nose pokes, lever presses etc.), most other tests suffer from biased human observations and short test durations. Often, the focus is on behaviour of individuals, mostly of the male sex. We may assume that humans may be skilled in interpreting primate behaviour, but the nocturnal rodent who is strongly relying on olfactory senses, touch and also sound – this is a different story. Furthermore, humans can hardly resist interpreting behaviour while observing. For example, when the distance between two animals declines, the human observer easily scores ‘approach behaviour” as if he/she can recognize the intention of the animal.

The subjective influence of the human observer and the large inter-observer and inter-lab differences are substantial. The ongoing debate on the fuzzy and hardly reproducible results culminated in papers in the nineties and early 2000 when mouse mutants became available in large numbers3.

Automated methods could solve the reliability issue. But then scientists rightfully demand validation of these methods and techniques as there is no a priori ‘intuitive’ knowledge of a specific behaviour defined by a computer algorithm.

To understand why some tests have been successfully automated and others not one has to distinguish two types of research question using these tests:

1. A test for addressing a straightforward question where a few behavioural read-out parameters are used to assess dose-response relationship. Example: amphetamine induces locomotor activity. Distance moved is then used as a litmus paper indicative for the effective dose. This approach is applied in behavioural and safety pharmacology.

2. The same parameters in the simple test are used for the functional interpretation of a treatment. Let us say the same test is used for the efficacy of diazepam in anxiety. Now low activity (immobility or freezing) is assumed to be indicative of anxiety and the drug acts anxiolytic by increasing locomotor activity.

Tests used to answer straightforward questions are not necessarily validated for functional interpretations. Thus, the validity of a specific set up depends on the scientific question and the interpretation of behaviour depends on the scientific context. Nonetheless, we see whole batteries of tests (mainly simple tests previously used for rats) being used for describing the effects of a novel drug or the phenotype of unknown mouse mutants in functional terms.

Given the broad array and the huge number of studies conducted over time using those classical tests of open field and elevated plus maze, the need for validation seems to have faded away against the background of a familiar and extensive mixed set of data. The interpretation of immobility in case of an anxiolytic drug action and low activity in case of a stimulant may be acceptable when the treatment effects are well known.

The question remains: How can behaviour such as immobility be used for a biological relevant functional interpretation of behaviour in terms of anxiety when an unknown drug or mutant mouse is studied?

The validation of innovative technology is difficult to design in a convincing way because of unclear validation criteria. Let us consider two aspects:

First, the aim is to get rid of human (mainly visual) observations. However, there is that desired link with the previous intuitive way of scoring and interpretation. We want a new technology to score immobility the same way as human observers do, even when inter-observer reliability is low. We have to accept that if new technology has to match previous methods, little will be gained. For straight forward questions such as: amphetamine induces locomotion, previous results will be confirmed.

Second, novel technologies allow extending the time of observation and ethograms. Novel drugs or mutations may have all kind of effects on behaviours which have not been measured before, complicating the previously assumed and simplified specificity.

The caveat of a more automated comprehensive approach, detecting a wider spectrum of behaviours over a longer period of time is that it yields a complex set of data. Those numbers representing movements, velocity, (changes in) contour of animals etc. do not have a direct intuitive representation in terms of behaviours as one is familiar with by using classical studies. Thus, at first sight the data set might be apparently without any meaning. This may be business as usual in other scientific disciplines where math is more generally used and accepted, but cumbersome for behavioural scientists, who like to rely on their own way of observing animal behaviour which leads us back to the first item.

How to get out of this vicious circle?

It requires special tools for exploring the data. Descriptive studies require special data mining tools. Hypothesis driven studies have a priori defined parameters and those which appeared to have changed without a priori prediction need to be explained. Since the number of parameters may be very large, an appropriate analysis and interpretation is intriguing and challenging. Behaviour as human observers see it, always elicits associations in terms of meaning or differences. Scepticism and doubt emerge about the meaning of those newly generated abstract numbers. One is tempted to fall back on familiar and 'simple' tasks.

The user – not the observer in case of automated observations – has to rely on numbers, not knowing exactly how these numbers are generated. Therefore, the need for proper validation is justified. But this requires an appropriate test paradigm which allows the detection of the parameters generated by the technology. It is the tradition and the apparent need for simplicity which have restricted the validation and use of automation and technological innovation.

Now, automation of the measurement of parameters in classical tests focusing on individual animals in a relatively simple environment lent itself for automation and it is there where technology found its entrance.

The advantage is mainly a matter of saving time, more precision etc. but hardly any new parameters. Distance moved and time spent in certain areas are probably the most commonly used parameters.

For the progress of behavioural sciences in methods and paradigms we need the two available kinds of technical innovations:

1. Statistical and methodological procedures to go beyond the mere counting of frequencies and durations.

2. Data collection instruments by use of sensors connected to a computer.

ad1) That behaviour is more than a sum of frequencies and/or duration had been noticed before: behaviour is a stream of events organized in time and space: temporal organization or sequential organization of behaviour as is explained in various ethological text books. Both time and space had been neglected in the most frequently used tests. The elegant studies of Golani and Benjamini [4] show that exploration is a gradually extending patterning of behaviours in the available space, which has to be bigger than anything regularly used in laboratories. The analysis of the temporal organization by using transition matrices or t-patterns had also been used decades ago, but did not evolve into a general tool. The rapid access to computer power eased the use of these methods but even that did not yield a breakthrough of such methods. However, mainstream users sticking to classical methods slow down the progress in this field.

ad 2) Infra-red beam devices are a simple way of monitoring activity; a kind of grid and the interruption of a beam could allocate the animal to certain area in an open field-like apparatus. Later on, video- images were digitized and then fed into a computer. These methods are activity based thus, quantifying a change in position of an animal. Another set of features was added later. The animal was now not reduced to a single point such as for instance the centre of gravity, but the nose tail points of the animal and the contour of its periphery was used to include more form related characteristics. The dynamic changes of the contour allowed the distinction of some behaviours such as grooming, rearing exploration, immobility etc. The need for high throughput phenotyping tried to take advantage of automated devices. However, well known classical tests were often just combined in a battery of tests and therefore counteracting the initial approach.

The full potential of using technology which could bring the whole discipline of behavioural science to a higher level requires a paradigm shift as the animals have to be provoked to display the whole potential of behaviours. Though, technology now allows a shift in paradigms as the laborious and complex human observations can be avoided and new parameters can be added, it requires validation of the paradigm and the technology. This is one of the reasons for the lack of progress in using new paradigms in behavioural neuroscience.

Pharmacological validation asks for the use of reference drugs. Thus, if a new system with the capacity to measure anxiety related behaviour is tested, the antagonistic effect of e.g., diazepam has to be demonstrated first. In contrast, effects of diazepam are based on the limited technology one wants to improve. A new technology might show that anxiety is more than immobility and that diazepam has more effects than counteracting immobility. It is not easy to break through this circular reasoning, unless it's accepted that diazepam has more effects than previously assumed and that it is not the ultimate reference drug. For testing anxiety, consensus on what anxiety is under certain conditions is required. Therefore, we may have to return to and upgrade ethological concepts.

Advanced analyses have been developed and novel measuring techniques have been introduced. There is a huge opportunity for improvement in the main stream of behavioural sciences. The strength of automated methods is that one has to explicitly formulate the behaviours of interest before they can be translated into a computer algorithm. A machine has the ability to perform with sustained attention and, thus, the factor time can now really be involved in the stream of behaviours. Rhythmicity can be revealed and behaviour of nocturnal animals can easily be monitored during the dark phase of the circadian rhythm. Surprisingly and to our disappointment, extending the duration of testing is not really extensively used yet.

Another issue is the reproducibility. Whereas inter-observer reliability is low, a machine scores the same stream of events almost identically to the advantage of reliability.

We have to demonstrate that new methods of observation and analysis tools catalyse the use of new paradigms yielding more and biologically relevant information, stimulating cross fertilization between innovative technology and expertise in behavioural sciences, an acceleration of its development is the near future. Studies will be conducted under biologically-relevant conditions. This will yield animal models with enhanced translational value in preclinical research, new main effects and unwanted side effects of drugs can be detected within the same paradigm. If we accept the limitations of existing paradigms and accept that behaviour is more complicated than previously assumed then technology may help to explore and understand behaviour in depth and stimulate for instance drug discovery to contribute to human health.

References

1. Kulkarni, S.K., Dandiya, P.C. (1973). Effects of antidepressant agents on open field behaviour in rats. *Psychopharmacologia* **33**, 333-338.

2. Haller, J., Alicki, M. (2012). Current animal models of anxiety, anxiety disorders, and anxiolytic drugs. *Curr. Opin. Psychiatry*. **25**, 59-64.

3. Wahlsten, D., Bachmanov, A., Finn, D.A., Crabbe, J.C. (2006). Stability of inbred mouse strain differences in behavior and brain size between laboratories and across decades. *Proc. Natl. Acad. Sci.* U.S.A. **103**, 16364-16369.

4. Golani, I., Benjamini, Y., Eilam, D. (1993). Stopping behavior: constraints on exploration in rats (Rattus norvegicus). *Behavioural Brain Research* **53**(1-2), 21-33.

Keynote lecture

Professor Maja Pantic

Intelligent Behaviour Understanding Group (iBUG), Imperial College London, London, UK

About the speaker

Maja Pantic received M.S. and PhD degrees in computer science from Delft University of Technology, the Netherlands, in 1997 and 2001. From 2001 to 2005, she was an Assistant and then an Associate professor at Delft University of Technology, Computer Science Department. In April 2006, she joined the Imperial College London, Department of Computing, UK, and continued working on machine analysis of human non-verbal behaviour and its applications to Human-Computer Interaction (HCI). In October 2010, she became Professor of Affective & Behavioural Computing and the leader of the Intelligent Behaviour Understanding Group (iBUG) at Imperial College London. From November 2006, she also holds an appointment as the Professor of Affective & Behavioural Computing at the University of Twente, Computer Science Department, the Netherlands. In 2002, for her research on Facial Information for Advanced Interface (FIFAI), she received Innovational Research Award of Dutch Research Council as one of the 7 best young scientists in exact sciences in the Netherlands. In 2007, for her research on Machine Analysis of Human Naturalistic Behavior (MAHNOB), she received a European Research Council Starting Grant (ERC StG) as one of 2% best junior scientists in any research field in Europe. She is also the Scientific Director of the large European project on Social Signal Processing. In 2011, Prof. Pantic receivedthe BCS Roger Needham Award, awarded annually to a UK based researcher for a distinguished research contribution in computer science within ten years of their PhD.

She is the Editor in Chief of the Image and Vision Computing Journal (IVCJ/ IMAVIS), Associate Editor of the IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics (IEEE TSMC-B), Associate Editor of the IEEE Transactions on Pattern Analysis and Machine Intelligenve (IEEE TPAMI), and a member of the Steering Committee of the IEEE Transactions on Affective Computing. She is a Fellow of the IEEE.

Prof. Pantic is one of the world's leading experts in the research on machine understanding of human behavior including vision-based detection, tracking, and analysis of human behavioral cues like facial expressions and body gestures, and multimodal analysis of human behaviors like laughter, social signals, and affective states. She is also one of the pioneers in design and development of fully automatic, affect-sensitive human-centered anticipatory interfaces, built for humans based on human models. She has published more than 150 technical papers in the areas of machine analysis of facial expressions and emotions, machine analysis of human body gestures, and human-computer interaction. Her work is widely cited and has more than 25 popular press coverage (including New Scientist, BBC Radio, and NL TV 1 and 3). See also: http://ibug.doc.ic.ac.uk/~maja/

Machine Analysis of Facial Behaviour

Facial behaviour is our preeminent means to communicating affective and social signals. There is evidence now that patterns of facial behaviour can also be used to identify people. This talk discusses a number of components of human facial behavior, how they can be automatically sensed and analysed by computer, what is the past research in the field conducted by the iBUG group at Imperial College London, and how far we are from enabling computers to understand human facial behaviour.

Special Sessions

On Road Observational Survey of Seat Belt Use Among Young Drivers in Qatar

Khaled Shaaban

Department of Civil Engineering and Architectural Engineering, Qatar University, Doha, Qatar. kshaaban@qu.edu.qa

Abstract

Traffic accidents are one of the main cause of death in modern societies. Next to circulatory diseases and cancer, road accidents are probably the third major cause of death in the developed world [1]. Traffic accidents kill 1.2 million people every year and injure or disable as many as 50 million more. They are the second leading cause of death globally among young people aged 5 to 29 years and the third leading cause of death among people aged 30 to 44 years [2]. In 2010, road traffic accidents were the cause of 228 people deaths and 568 major injuries in Qatar. In October 2007, a new traffic law was implemented in the State of Qatar in order to reduce the growing number of accidents and reckless driving in Qatar. The new law forced seat belt usage for all drivers and front-seat passengers. The objective of the paper was to investigate the seat belt use among university students drivers in Qatar after the implementation of the new traffic law using observational survey. Two higher educational facilities in Qatar were selected to conduct the research. Four students pursuing an engineering degree were trained in general traffic data collection methods and procedures and specifically on how to observe the vehicles approaching and how to collect data in sheets prepared in a specific coded way to gather the information needed. The trained observers stationed at four different entrances of the studied universities. The results of this study suggest that a significant percentage of university students do not wear their seat belts and is a clear indication that driving habits in Qatar have not improved since the implementation of the 2007 traffic law.

Introduction

Traffic accidents, one of the most important challenges of modern societies, are the third leading cause of death in most countries. There is no doubt that traffic accidents cause social and economic problems and leave a direct impact on people. With respect to economic problems, traffic accidents constitute a big burden on the society as a result of loss of life, injury and disability and increase in the amount of insurance and compensations. Although, in recent years, significant developments have been made in road safety in Qatar, and national traffic low was implemented, traffic statistics indicate that there is increasing in accidents in Qatar. In October 2007, a new traffic law was implemented in the State of Qatar. This law contained a package of rules and regulations in traffic system in addition to the awareness and preventive campaigns. The new law forced seat belt usage for all drivers and front-seat passengers. However, young drivers often do not wear seat belts. After more than four years of the implementation of the traffic law, it was necessary to investigate the seat belt use among young drivers. Better understanding of the behavior of young drivers is important since it will shed more light on developing plans or strategies to improve the traffic safety of this vulnerable group in society.

Studies show that seat belts can save lives and significantly decrease motorists' injury severity in crashes [1,8]. In 2008, 25,351 passenger vehicle occupants were killed in traffic accidents in the USA. Among these fatalities, more than 55 percent of the occupants were unrestrained [3]. The National Highway Traffic Safety Administration (NHTSA) estimates that an 80 percent safety belt use rate saves more than 15,000 lives per year and an overall societal cost of 50 billion dollars in the country each year [4]. NHTSA estimated that 13,250 lives were saved in 2008 due to the use of safety belts [5] among passenger vehicle occupants over age 4. Finally, according to the NHTSA, seat belts are approximately 50 percent effective at preventing fatalities in car crashes and save about 15,000 lives each year in the U.S. [6].

In order to reduce the growing number of accidents and reckless driving in Qatar, the government has introduced new traffic laws stipulating severe penalties in October 2007. This law contained a package of rules and regulations in traffic system in addition to the awareness and preventive campaigns. The new law forced seat belt usage for all drivers and front-seat passengers. Children under age of 10 have also been banned from sitting in the front seat of the car. However, young drivers often do not wear seat belts. Young drivers and students are known to be a problem age group for road traffic safety in several countries. Motor vehicle crashes are the leading cause of death for US teenagers, accounting for 40% of fatalities. In a study in Colorado, a higher proportion of young drivers were found to be involved in over-speeding, reckless driving, charged with traffic violations, and safety belt non-use [7]. Young drivers in New Zealand contribute disproportionately to traffic accident injuries and deaths. In 1999, the 15-24 year age group accounted for 25% of road deaths and 31% of reported injuries [8]. The objective of the paper was to determine the seat belt use among university students drivers using observational survey.

Survey Design and Implementation

Two higher educational facilities in Qatar; Qatar University (QU) and College of the North Atlantic (CNA) were selected to conduct the research. These two high educational facilities were chosen due to the high number of students (approximately 8,200 and 4,600 students respectively) compared to other higher educational facilities such as Texas A&M University at Qatar (approximately 390 students) and Carnegie Mellon University in Qatar (approximately 320 students).

Four students pursuing an engineering degree were trained in general traffic data collection methods and procedures and specifically on how to observe the vehicles approaching and how to collect data in sheets prepared in a specific coded way to gather the information needed. Each student received a day-long training explaining the procedure and practicing on how to conduct the field data collection. Prior to conducting the actual survey, a pilot survey was administered. The four field observers were used to conduct the survey at one site. During the pilot survey, it was difficult to observe if drivers are wearing seat belt or not in some of vehicles (approximately 20%) due to the vehicles speeding. Therefore, it was decided to choose new locations where vehicles are forced to slow down. During the final survey, the trained observers stationed at four different entrances of the studied universities. The four locations were selected carefully where vehicles had to slow down at the speed humps located at the different entrances. This strategy allowed for easy capturing of all data needed from all vehicles approaching as shown in Figure 1.

The data was collected for 701 student vehicles between 7:00 am and 9:00 am, at the beginning of the school day and between 14:00 pm and 15:00 pm at the end of the school day. At the studied locations, the observers team collected the data while the approaching vehicles slowed down to enter the gate of the university. The surveys objective was to collect accurate and reliable data. They made the following observations: 1) whether the vehicles is a passenger car or SUV; 2) whether the driver is wearing a seat belt or not; 3) whether front-back seat passenger is wearing seat belts or not; 4) whether front-back seat passenger is wearing seat belts or not; 5) gender; 6) student or adult; and race. Observations were made during daylight hours to ensure clear visibility. It should be noted that gender and race for the driver, front-seat passenger, and back-seat passengers were identified as part of the survey. The driver and passenger races were categorized as Arab-Qatari, Arab-Non Qatari, Asian, and American/European. The data collected were recorded manually on paper survey forms to

Figure 1. Inspection place showing visibility of seat belt usage.

avoid any issues related to the visibility of the screen of electronic data collection equipment under sunlight conditions. According the observes, the manual method was effective, accurate, and allowed them to verify and easily fix any issue at the time of data entry.

Analysis of Data

After conducting the survey, data was transferred from the field sheets to a main Excel spreadsheet by two team members and verified for accuracy by the other two team members. The verification for accuracy was achieved by comparing the results obtained by comparing the survey forms against the Excel spreadsheet. Three separate worksheets were established: 1) data dictionary; 2) survey data; and 3) survey data analyzed.

Conclusions

After collecting and analyzing the survey data, it was concluded that the method used in the collection was effective and adequate. The results of this survey provide a snapshot of seat belt use in Qatar among young drivers. Findings of this survey indicated that nearly 57.3% of the young drivers were not wearing seat belts. The number of male and female drivers in the random sample were 467 and 234 respectively. The paired t test, a parametric test, was conducted to identify if there were significant differences between male and female drivers. Using the two-tail t test, the null hypotheses was rejected at the 5% significance level, as the p value was 0.000. The results show that male drivers have a lower seat belt use rate with 71.5% of male drivers found not wearing seat belts compared to 29.1% of female drivers not wearing seat belts. In addition, it was found that Qatari students have a lower seat belt use rate than other non-Qatari categories. 76.1% of Qatari drivers were found not wearing seat belts compared to 32.9 non-Qatari not wearing seat belts ($p = 0.000$). Finally, vehicle type seems to show a significant gap between unbelted and belted students. Students in SUV's have a higher non-seat belt use rate (65.9%) than those in other vehicle types (46.6%).

The results of this study suggest that a significant percentage of university students do not wear their seat belts and is a clear indication that driving habits in Qatar have not improved since the implementation of the 2007 traffic law. There are no simple remedies for drivers not wearing seat belts, however, there are a variety of countermeasures. These countermeasures include children education, driver education and awareness, legislation and enforcement. Not using the safety belt is a behavioral issue, so educational programs targeting the change of the all driver's behavior will also lead to an increase in the safety use rate in the driving population. Previous research indicated that programs designed to encourage seat belt use in young children can increase seat belt use by the children and their parents, particularly in low income neighborhoods [9]. In addition, implementation of more restrict regulations will be necessary. Another study showed that an increase in fine level from $25 to $60 was associated with a three to four percent increase in observed seat belt use [10].

Acknowledgements

This research was made possible by a grant from the Qatar National Research Fund under its Undergraduate Research Experience Program award number UREP 11-109-3-022. Its contents are solely the responsibility of the author and do not necessarily represent the official views of the Qatar National Research Fund.

References

1. Abdel-Aty, M. (2003). Analysis of driver injury severity levels at multiple locations using ordered probit models. *Journal of Safety Research* **34**, 597-603.

2. O'Flaherty, C.A. (1997). *Transport Planning and Traffic Engineering*. Arnold Publications, London.

3. NHTSA's National Center for Statistics and Analysis (2009). *Traffic Safety Facts – 2008 Traffic Safety Annual Assessment – Highlights*. U.S. Department of Transportation, NHTSA, DOT HS 811 172.

4. The National Initiative for Increasing Safety Belt Use Buckle Up America Campaign (2005). Eight Report to Congress, Sixth Report to the President, U.S. Department of Transportation, NHTSA.

5. NHTSA's National Center for Statistics and Analysis (2008). *Traffic Safety Facts, 2008 Data – Occupant Protection*, U.S. Department of Transportation, NHTSA, DOT HS 811 160.

6. NHTSA's 2004 Motor Vehicle Occupant Protection Facts (2006). Washington DC: National Highway Safety Administration, U.S. Department of Transportation, DOT HS 810 654.

7. Gonzales M.M., Dickinson, L.M., DiGuiseppi, C., Lowenstein, S.R. (2005). Student Drivers: A Study of Fatal Motor Vehicle Crashes involving 16-Year-Old Drivers, *Ann. Emerg. Med.* **45**(2), 155-156.

8. New Zealand Land Transport Safety Authority (2000). *Motor Accidents in New Zealand*.

9. Hazinski, M.F., Eddy, V.A., Morris, J.A. (1995). Children's traffic safety program: influence of early elementary school safety education on family seat belt use. *J. Trauma Injury Infect. Crit. Care* **39**, 1063–1068.

10. NHTSA's Strategies to Increase Seat Belt Use (2010). *An Analysis of Levels of Fines and the Type of Law*. Washington DC: National Highway Safety Administration, U.S. Department of Transportation, DOT HS 8110 413.

Studying Driver's Lane Changing Behavior Under Heavy Traffic Volumes

Khaled Shaaban

Department of Civil Engineering and Architectural Engineering, Qatar University, Doha, Qatar. kshaaban@qu.edu.qa

Abstract

The goal of this paper was to develop a method to study the driver's lane changing behavior occurring on arterial streets under heavy traffic volumes. This work focus on studying roadway segments with short spacing between signalized intersections suffering from delay problems due to the lane changing behavior. The heavy delay problems result from the through vehicles changing lanes to turn right at the downstream intersection conflicting with side street vehicles entering from an upstream intersection and changing lanes to turn left or go through at the downstream intersection. The method utilized video cameras and road tubes to collect the data. It was found that this method was effective, accurate, and allowed for collecting accurate information regarding the driver's lane changing behavior.

Introduction

Two sites were selected for the analysis. These two sites suffer from a delay problem due to the lane changing behavior. The first site was on State Road 421 between the I-95 Off-Ramp and Airport Road in Port Orange, Florida and the second site was on State Road 50 between State Road 408 Off-Ramp and Bonneville Drive in Orlando, Florida. The two sites exist at the exit ramp of a diamond interchanges where the side street vehicles enter the arterial street through a free right turn lane. These two sites have the following criteria: relatively short spacing between two signalized intersections that are running in coordination; moderate to heavy road volumes; and no driveways or median openings between the two signalized intersections. The arterial segment had two through lanes. The downstream intersection had a left turn lane and a right turn lane. Figure 1 shows the studied movements occurring on the two arterial segments. These movements caused heavy delay along the two segments as shown in Figure 2.

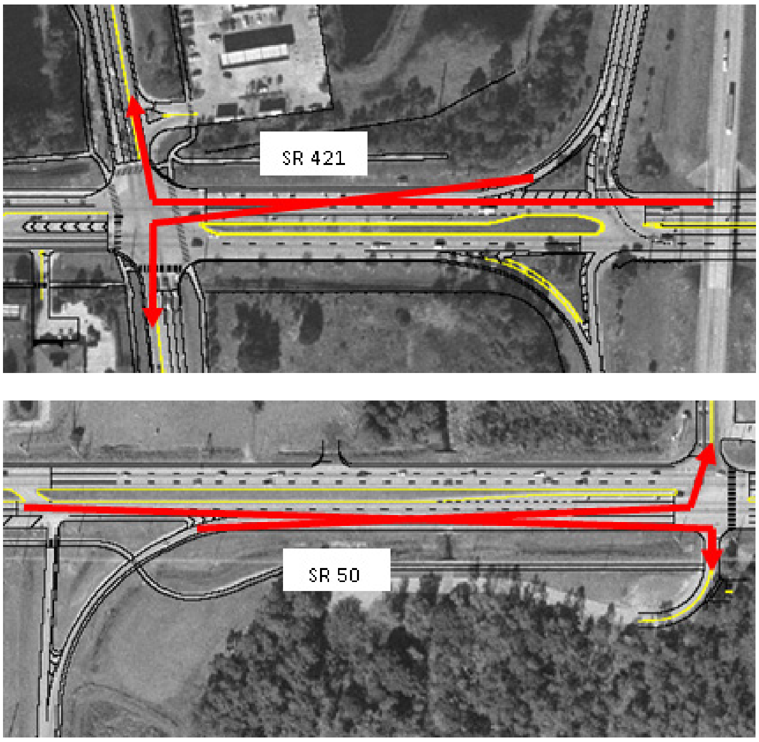


Figure 1. Aerial maps showing the studied movements at the two studied sites.

Figure 2. Breakdown conditions on the two arterials.

Data Collection

Video cameras were used to collect the data. The cameras were used for two purposes. First, the cameras were used to record the driver’s behavior. Second, the cameras were used to obtain volume counts and turning percentages along the arterials. To be able to achieve these two goals, the cameras were positioned on a high position (the I-95 bridge and the SR 408 bridge) to cover the studied segments (see Figure 3). The lane changing area was defined as the area between the end of the gore area at the first intersection to the stop bar at the second intersection. The cameras were zoomed in to capture the movement of each vehicle within the lane changing section. In order to determine the location where the vehicle performed the lane changing, road tubes were placed at a 100 feet spacing starting at the gore area as shown in Figure 3. The tubes acted as distance meters. In addition to the video, aerial photographs and detailed sketches of the two sites were obtained. These sketches included the geometry of each site including the number of lanes, channelization, auxiliary lanes, and the distance between the two signalized intersections. At each site, eight hours of data were collected on a normal weekday using the video recording equipment. The time periods were selected so that two hours in the morning period (7:00 a.m. to 9:00 p.m.), two hours in the midday period (11:00 a.m. to 1:00 p.m.), and four hours in the evening period (2:00 p.m. to 6:00 p.m.) were observed.

The reduction of the field data involved observing the videotapes of each site. The videotapes were used to observe the lane changing movements and also to obtain accurate counts and turning percentages along the arterial. This method was used since it was hard to observe the lane changing and to count the vehicles in real time at high volumes. Accuracy in video data is due mainly to the fact that the viewer is able to view the videotape more than one time. Therefore, the viewer can concentrate on a single movement and then when finished rewind the tape and observe a different movement. Data reduction sheets were created for each site so

Figure 3. Video camera positioned on the bridge and road tubes placed every 100 feet along the arterial.

that the lane changing distance and the origin-destination patterns of individual vehicles could be recorded. The lane changing distance is defined as the distance from the gore area to the location where the vehicle crossed to the desired lane. Videos were then watched in slow motion to verify the lane changing distance, the origin-destination information, and the number of lane changes required to complete the movement. The origin-destination volumes, the lane changing distances, and number of lane changes were recorded in one-minute increments.

Analysis of Data

It was also found that there are five types of lane changing movements occurred. Type 1 and Type 2 were originated from the mainline and attempted to reach the right turn lane at the second signalized intersection. Type 1 vehicles had to perform one lane change in order to complete the desired lane changing maneuver. Type 2 vehicles had to perform two lane changes in order to complete the desired maneuver (change one lane to the second through lane then a second lane change to the right turn lane). Types 3, 4, and 5 were lane changing movements originated from the side street free right turn lane to go through or turn left at the second signalized intersection. Type 3 vehicles had to perform one lane change in order to complete the desired lane changing maneuver (move to the through lane). Type 4 vehicles had to perform two lane changes in order to complete the desired lane changing maneuver (change one lane to the first through lane then a second lane change to move to the second through lane). Type 5 vehicles had to perform three lane changes in order to complete the desired lane changing maneuver; the first land change to move to the first through lane, the second lane change to move to the second through lane, and the third lane change to the left turn lane.

It was found that 64% of the lane changing movements were originated from the side street and 34% was originated from the main street. The majority of lane changing volume occurred between Type 1 (35%) and Type 3 (40%), which accounted for 75 % of the total lane changing volume. Type 2 was the lowest lane changing volume (1%), which indicated that most vehicles that wanted to perform the lane changing movement from the main street preferred to change lanes to be in the outside through lane before entering the lane changing area to minimize the number of lane changes to only one lane change. The percentage of Type 4 was 13%, which indicated that some of the vehicles preferred to change two lanes to be in the inside through lane on the main street. This is probably due to the impression that the inside through lane will be faster than the outside through lane due to less distraction after the intersection. Type 5 (11%) is mainly based on the number of vehicles that had to perform a left turn at the second intersection.

The average lane changing distance for the 4,443 vehicles tracked for each type of lane changing were calculated. The average lane changing distance (D1) is the average of the lane changing distances required to perform the first lane change measured from the end of the gore area (applicable to all types). The average lane changing distance (D2) is the average of the lane changing distances required to perform the second lane change measured from the end of D1 (applicable only to types 2, 4 and 5). The average lane changing distance (D3) is the average lane changing distances required to perform the third lane change measured from the end of D2 (applicable only to type 5).

It was found that Type 5 has the minimum value of D1. These vehicles had to perform three lane changes and they had to start the lane changing movement as soon as they enter from the side street to the main street. D1 for Type 1 was also low because some of the vehicles in this type started the lane changing movement before the end of the gore area (driving on the gore area striping). The maximum value of D1 was for Type 3 where vehicles had to perform only one lane change. D1 for Type 2 and Type 4 were very close (157 feet and 143 feet respectively). These two types had to perform the same number of lane changes (two) in order to complete the desired lane changing maneuver.

A comparison of D1, D2, and D3 was done for the two studied sites. The main difference between the two sites was the distance between the end of the gore area to the stop line at the second intersection (LG). LG for the first site was 532 feet and for the second site was 730 feet. It was found that D1, D2, and D3 decreased dramatically when LG decreased which indicates the great effect of the distance between the two intersections on the average lane changing distance for the different lane changing types.

Conclusions

This paper has examined the different lane changing movements occurring between two close-spaced intersections for two sites in Florida. The two sites have a heavy right turn volume entering from the side street and close-spaced intersections. The paper has also studied the breakdown conditions occurring on the two arterial segments and caused by the lane changing movements. It was found that the breakdown conditions occur in two cases. The first case occurred when the main street through volume was heavy with moving queues observed extending onto the first intersection. In this case, vehicles entering from the side street could not find adequate gaps on the main street and had to reach a complete stop waiting for a gap on the main street. In the second case, the left turning volume at the second intersection was heavy and blocking the whole left turn lane. Although the main street volumes were moderate and adequate gaps were available, vehicles entering from the side street and willing to perform a left turn at the second intersection had to stop blocking the free right turn lane and waiting for the left turning vehicles to clear.

The analysis also revealed that the lane changing distances were also affected by the distance between the two intersections. As the spacing between the two intersections increased, the lane changing distances for all movements increased. By increasing the distance between the two intersections, drivers will have more space and time to adjust and to perform the lane changing movement. In addition, the lane changing distances within the same site were affected by the number of lanes changed. If a driver wants to change three lanes, he/she will perform the first lane change at a much shorter distance than a vehicle that wants to change only one lane. Based on the analysis introduced in this paper, it was concluded that lane changing movements on arterial streets can cause major delay problems streets unless adequate spacing between intersections is provided. It was found that the data collection method utilized was effective, accurate, and allowed for collecting accurate information regarding the driver’s lane changing behavior.

Selection of a Measurement Battery for Human Behaviour Assessment in Serious Games in the Aviation Domain

G.D.R. Zon

Training, Simulation and Operator Performance department, National Aerospace Laboratory – NLR, Amsterdam, The Netherlands. zon@nlr.nl

The challenge

Is measurement of operator performance behaviour in serious games different from measurement in other environments? A serious game can be seen as an environment in which subjects are allowed and encouraged to play and experiment. The aim is that, by doing so, the subjects learn in an environment that changes, which is safe, and which elicits to play again. As such the serious game can be seen as a mixture of a training environment, a simulation and a game. For each of these domains different ways of measuring behaviour of the human operator exist. The question is to what extent the measurement methodologies that are common for those different domains are also applicable to the measurement of behaviour in a serious game.

Human behaviour assessment in simulator experiments

Normally when assessing operator performance or behaviour in a simulator experiment, the experiment compares a baseline with one or more experimental conditions. From these conditions one or more hypotheses are drawn and compared. These hypotheses may comprise aspects like improved situational awareness, decrease of mental workload, better operator performance, increased trust or operator acceptance, etc. In order to measure those, a number of tools exist. From that toolset a measurement battery for that particular experiment is selected. This is approach is applicable in all kinds of simulation environments ranging from civil or military cockpit to air traffic control environments [3, 7, 8].

Methodological Triangulation

The method "methodological triangulation" is also known as "converging evidence". The method can be summarised by stating that when behaviour of an operator is assessed at least one measure per main category needs to be applied. There are three main categories (see **Error! Reference source not found.**), namely:

- Bio-behavioural or (objective) data (for example psychophysiological or eye tracking data);
- Performance data (for example simulator output like reaction(time) or errors made by the operator);
- Perception or subjective data (for example questionnaires, interviews or workshop outcomes).

The reason why these three main categories are used is that each of these provides information about what has happened during the experiments. However they do that each from their own perspective. A more complete overview of what has truly happened develops when data from the different main categories are compared.

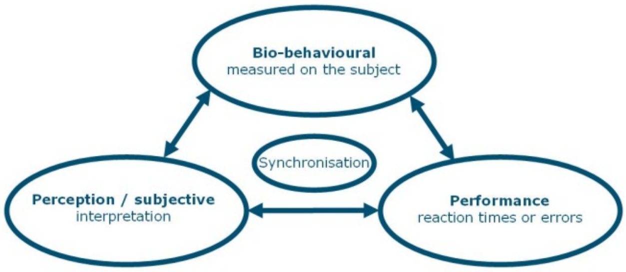


Figure 1. The three main categories that need to be compared and synchronised for methodological triangulation.

Relevant information may be missed when just one or two of the main categories are used. Hence, after participation in an experiment a subject may say that s/he has worked hard and performed well, but are both indeed true? The only way to be certain of that, is by checking other sources of information. Methodological triangulation underlines that: the whole is more than the sum of its individual components.

Serious games

Serious games have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement [1]. They are games that aim to teach the players competencies that are important and relevant for their professional development. Serious games are often used in parallel with other learning tools and environments, such as lectures, e-learning forums, and simulators. They enable a student to 'play' with course material and to see for him/herself how actions that s/he takes may work out without resulting in the consequences that may result from experimenting in real life. They range from very detailed, for part-task training, to very global, to provide students with a good understanding of the coursework. Serious games can even be used as an 'umbrella' for a complete course.

The advantage of using serious games over other learning tools is that when serious games are well-developed, they enhance students' motivation [4] for example because a game is fun to play as well. This increased motivation can lead to students spending more time on the training-task and therefore to better results [6]. In fact the game elicits them to play, and therefore learn, over and over again. A serious game is considered well-developed when the correct balance between entertainment and education was found. Therefore, the development of a serious game is a flexible process with many interactions between developers and potential users to ensure the correct balance. In other words, a game that is both entertaining and ensures real impact (i.e. transfer of learning back into the work-place).

Case studies

Two serious games are currently under development. One within the ADAHR[1] [5] project. This project focuses at the foreseen increase in level of automation in future air traffic control and airport environments. Different roles, or professions, that were identified might be influenced by increased levels of automation in 2035 and 2050. Currently games are being developed in which these new concept can be played with, and tested. There will be two hardware based games in which simulation in realistic environments plays an important role. For a number of relevant aspects that can not be simulated in those environments paper based games will be developed as well. The aim of these games is to acquire better insight in the effect that these higher levels of automation may have on human operators.

The other game is part of the MASCA[2] [2] project. MASCA supports management of airlines and airports in the process of being better able to quickly and swiftly adapt to change. This change can either be induced by authorities but also by market developments or customer requirements. MASCA will create a board, or paper based, game. The game will support the so called collaborative decision making at airports. The players will learn that collaboration between different stakeholders at the airport supports better and faster handing of aircraft. Therefore it will teach them to adjust their behaviour accordingly in the future.

---

[1] ADAHR = Assessment of Degree of Automation on Human Roles (http://www.adahr.eu/ADAHR/). This project is sponsored by the EU and EUROCONTROL in the context of SESAR. The project partners in are: National Aerospace Laboratory of the Netherlands (NLR), Ingenieria de Sistemas para la Defensa de España (Isdefe), Centro de Referencia de Investigación, Desarrollo e Innovación ATM in Spain (CRIDA), Deutschen Zentrums für Luft- und Raumfahrt (DLR).

[2] MASCA = MAnaging System Change in Aviation (http://www.masca-project.eu/). This project is sponsored by the EU. The project partners are: National Aerospace Laboratory of the Netherlands (NLR), Trinity College Dublin (TCD), Kungliga Tekniska Högskolan (KTH), Scandinavian Airlines System (SAS), Swedavia, Aeroporto d'Abruzzo – Pescara, Thales, KITE.

At the moment that this abstract is written the games and as such also the methodology for measurement of the behaviour of the human operators who will be playing the game is under development. Therefore the results and conclusions can not be provided here.

During the presentation

Since a serious game can be seen as kind of hybrid between simulation, game and training, aspects like: operator state (situational awareness, workload, stress, trust, acceptance), operator performance, training competencies that are mastered, feedback, but also relevance for the domain and to what extend the game is fun and elicits to play it more often may need to be assessed. The tools and technologies that were considered, and the set that was eventually selected for this measurement will be presented at Measuring Behavior 2012.

References

1. Abt, C. (1970). *Serious Games*. New York: The Viking Press.
2. Corrigan, S. and Zon, G,D.R. (2011). Managing System Change in Aviation: A New Approach in Human Factors. *Symposium on Human Factors for Future Aviation 2011*. Amsterdam Schiphol.
3. Dijk, H. van, Merwe, K. van de, and Zon, G.D.R. (2011): A Coherent Impression of the Pilots' Situation Awareness: Studying Relevant Human Factors Tools, *The International Journal of Aviation Psychology* **21**:4, 343-356.
4. Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming* **33**(4), 441-467.
5. Koolloos, M.F.J. and Zon, G.D.R. (2011). The Effect of a Highly Automated Environment on Human Behaviour. Plans and first results of the WP-E project "ADAHR". *Proceedings of SESAR Innovation Days 2011*, Toulouse.
6. Mautone, T., Spiker, V.A., Karp, M.R., and Conkey, C. (2010). Using Games to Accelerate Aircrew Cognitive Training. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).*
7. Zon, G.D.R. and Dijk, H. van, (2009). A selection of Human Factors tools: Measuring HCI aspects of flight deck technologies. *Proceedings of HCI International 2009*. San Diego.
8. Zon, G.D.R., Waard, D. de, and Heffelaar, T. (2008). Human Factors measurement- and analysis tools for cockpit evaluation and pilot behavior. *Proceedings of Measuring Behavior 2008*, Maastricht.

Managing Driver Workload Using Continuous Driver Workload Assessment

Michel Alders[1], Jasper Michiel van Hemert[1], Jasper Pauwelussen[2], Tobias Heffelaar[3], Riender Happee[4], Joop Pauwelussen[5]

[1]*Tom Tom, Amsterdam, The Netherlands. Michel.Alders@tomtom.com, JasperMichiel.vanHemert@tomtom.com*

[2]*TNO, Soesterberg, The Netherlands. Jasper.Pauwelussen@tno.nl,*

[3]*Noldus Information Technology, Wageningen, The Netherlands. T.Heffelaar@noldus.nl*

[4]*Delft University of Technology, Delft, The Netherlands. R.Happee@tudelft.nl*

[5]*HAN University of Applied Technology, Nijmegen, The Netherlands. Joop.Pauwelussen@han.nl*

The last decades, the concern for the impact of our mobility system on safety, environment and congestion, has resulted in many innovations in vehicles and infrastructure. It is absolutely clear that the driver plays a crucial role in this. Pushing the throttle, not taking sufficient distance, taking risks on the road by aggressive driving behaviour are all factors being controlled by the driver. For that reason, engineers have tried to find innovative solutions by giving better support to the driver. Regarding safety, still over 80 % of all accidents are caused by human error, being primarily a result of poor recognition and inappropriate decision making. It is therefore remarkable that the type of driver and the driver state are hardly taken into account in these systems. We may distinguish between young drivers (responsible for 20 – 30 % of all driver deaths), elderly drivers (one out of five being older than 65 in Japan), truck drivers (combining logistic and driving tasks), etc. The driver state refers to mental workload experience (individual judgment of driving performance under possibly critical conditions), driver fatigue, alertness, drowsiness, driving skills, etc. It seems obvious that better (more effective) driver support may be obtained by adding driver state information to vehicle state data and information about traffic and road conditions, as input to the support system. This is referred to as DSE (Driver State Estimation) where we will emphasize on workload estimation.

The project ADVICE will take up the challenge to derive improved driver support systems, by taking driver workload into account. In more general terms (research question), ADVICE will explore the added value of workload estimation to interpret the actual driver ability to recognise traffic conditions and to make decisions, in order to contribute to a more effective driver support. The Workload is estimated from the context (such as various road characteristics, close by departure/destination etc.), from direct or indirect driver behaviour, e.g. the steering intensity and frequency, the observed use of throttle and brakes in relationship to traffic conditions, but also from physiological information (e.g. heart rate) and parameters describing the gain and delay of the driver response on changes in traffic and road conditions. The required workload estimation algorithm on the basis of driver observation can also be used to validate driver support systems in general with reference to workload, i.e. as part of a first generation standardized validation tool. Finally, by being able to derive workload data under varying traffic conditions, we can add information to a geographical map, to indicate locations with high workload expectation. These correspond to potential critical spots in the road network.

ADVICE will result in the following deliverables:

- A **Driver Support System** (a ‘proof of principle’ for a personalised user interface, on the basis of the monitored state (emphasis on workload, furthermore alertness, skills, …) of the driver. Extended with geographical information on high workload locations.
- A **Driver State Validator**, a first generation standardized research tool to validate Driver Support Systems in instrumented vehicle (field-) tests and naturalistic driving studies.

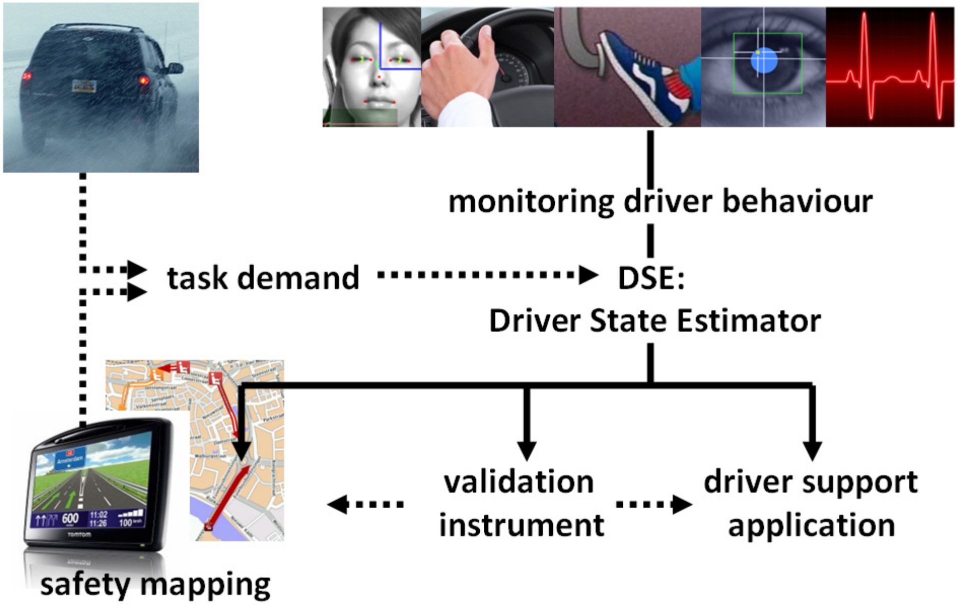


Figure 1. ADVICE system layout.

See figure 1 for a schematic layout of the proposed system. A consortium of organisations, with partners which are well recognised authorities in their specific field, has been formed to execute this project: TomTom, Noldus Information Technology, TNO, Delft University of Technology and the HAN University of Applied Sciences (project- and research management). In addition, specific associated SME partners have joined the ADVICE team, bringing in valuable knowledge and tools related to driver observation and data acquisition. Within the consortium, the industrial partners will focus on producing feasible and practical results, whereas TNO, HAN and Delft University will support this work through their in-depth knowledge, identification algorithms, and validation through experiments in-door (driving simulation), out-door (instrumented vehicle, under naturalistic conditions), software-in-the-loop and hardware-in-the-loop. ADVICE will use results of related previous projects such as the DRIVOBS project (methods for driver observation in car simulators) and the Dutch SPITS project, with a focus on affordable and open solutions for an Intelligent Transport System.

The presentation will focus on the need to adapt in vehicle information systems to driver state, and on foreseen affordable methods to estimate driver state.

Can We Discriminate Safe and Unsafe Visual Scanning in Multitask Driving Conditions?

P.M. van Leeuwen, J.C.F. de Winter, R. Happee

Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft, The Netherlands.
P.M.vanLeeuwen@tudelft.nl, J.C.F.deWinter@tudelft.nl, R.Happee@tudelft.nl

Introduction

Traffic accidents are a major public health concern with 1.2 million fatalities occurring every year and with millions more individuals getting injured [1]. More than 90% of the accidents are, at least partially, caused by the driver [2]. Figure 1 shows the overrepresentation of younger and older drivers among these crashes. The overrepresentation of younger drivers [1] can be explained by an increased willingness to take risk, poor anticipation of hazards, and insufficiently learned lateral and longitudinal vehicle control [3]. Figure 1 also shows the overrepresentation of elderly drivers which is caused by decline of cognitive and physical abilities [4] such as visual impairment (e.g., glaucoma).

Generic predictors of crash risk exist (e.g., age, gender). However, detailed knowledge about how drivers control their vehicle and combine various subtasks related to driving does not exist. Combining the visual scanning, vehicle control and decision making tasks makes the driving task complex. Drivers not only control the vehicle but also anticipate oncoming events (e.g., hazards, traffic) and combine the driving task with other tasks (navigating, cell phone use). Driving is predominantly a visual task [6] and individual differences in visual scanning behavior are found as a function of increasing driving experience [7], visual impairment [8] and environmental complexities [9]. For example, novice drivers have less visual attention to latent hazards compared to experienced drivers [10] and show visual scanning strategies that rely less on top down mechanisms of visual attention [11].

In our research, we aim to distinguish safe from unsafe drivers based on their visual scanning behavior. In this paper we will demonstrate two examples of visual scanning behaviors: (a) when using an in-vehicle system and (b) when performing a highway driving task, both in a driving simulator. Using these results the differences of drivers visual scanning in (multitask) driving and the applicability of non-intrusive eye tracker hardware in driver behavior research will be demonstrated.

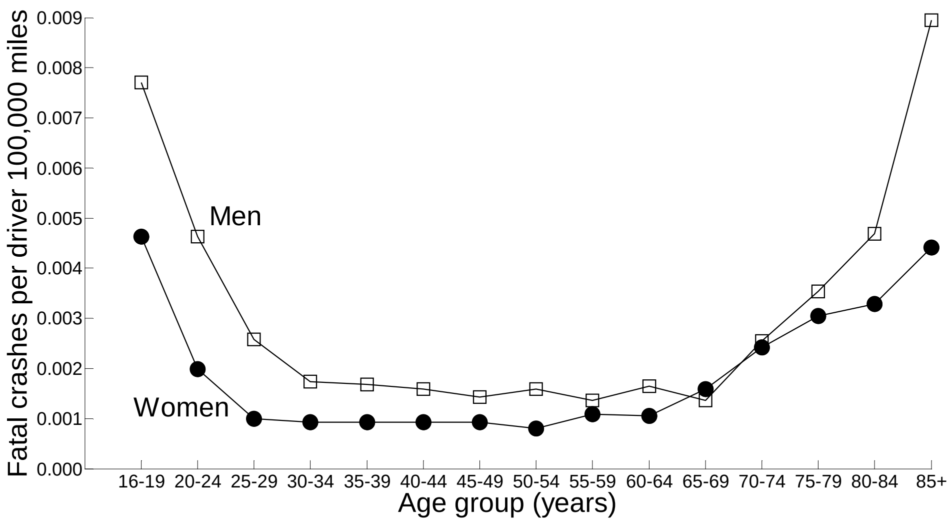


Figure 1. Fatal accident involvement per driver 100,000 miles as a function of age [5].

Methods

The driving simulator used in these studies is the Green Dino [12] driving simulator which is used for initial driver training in the Netherlands. This medium fidelity, fixed-base driving simulator provides a 180 degree horizontal field of view. A high field of view increases perceptual fidelity and is assumed to result in more realistic scanning behavior. The simulator controls were based on controls from a real car and steering feel was passively calibrated with respect to on-road vehicles [13]. This driving simulator has previously been used for research of training and assessment of student drivers [14]. Head motion and gaze direction was measured in both studies with a remote mounted eye-tracker system using infrared illumination. Cameras were mounted outside of the visual scenery in the driving simulator and calibration took place for each individual participant.

Results

The visual scanning of novice drivers using a concurrent lane position feedback system [15] is shown in figure 2. This feedback system allowed learner drivers to improve their lane keeping performance by using the feedback on their momentary lateral position error presented on their vehicle dashboard. The figure shows how drivers directed their gaze at the feedback area for longer periods of time.

In figure 3 the difference in visual scanning patterns are shown for two inexperienced drivers during a traffic-free highway driving task. One driver showed a small variance in horizontal fixations, with most fixations aimed at the roadway. A second driver showed large horizontal variance in fixations and shorter fixation times indicative of increased visual scanning of the roadway and its surroundings.

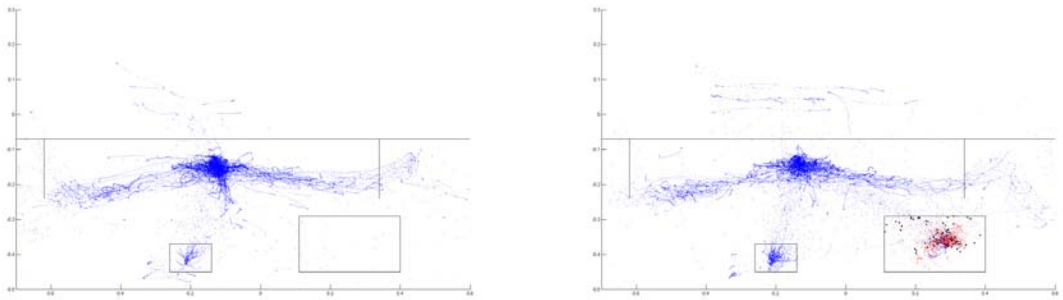


Figure 2. Raw gaze data for a rural road driving task. Gaze is mainly directed to the road ahead and the swirls left and right indicate looking into corners. Gaze pattern of a novice driver showing extensive use of an in-vehicle system (right).

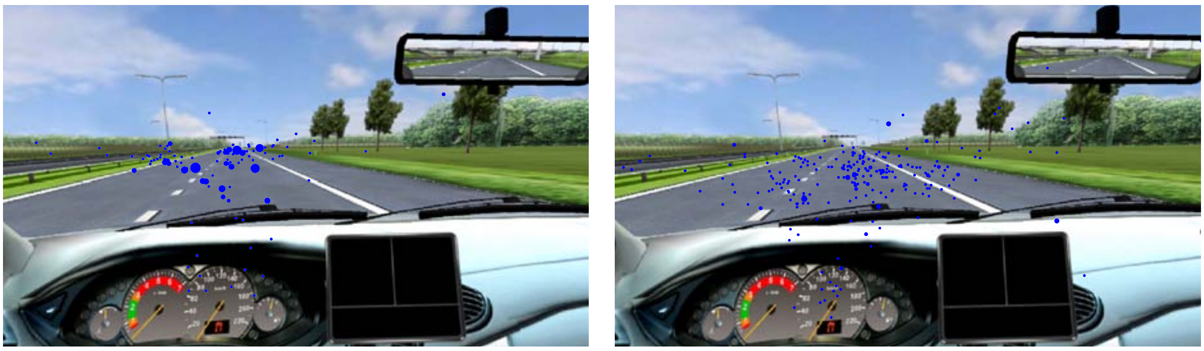


Figure 3. Visualization of fixations during a highway driving task, fixation duration is indicated by the dot size. Fixation pattern of two inexperienced drivers showing differences in visual scanning.

Discussion

Using driving simulators and high-end eye-tracking hardware, we have shown to be able to discriminate the differences in visual attention while using an in-vehicle system and the different fixation patterns of drivers performing the same task, showing the variability between drivers. In future work we will relate driving simulator measures of both the vehicle (e.g., lane position) and driver performance (e.g., steering activity) with gaze direction based metrics (e.g., fixation patterns) and measures of driver workload (e.g., heart rate variability). With these relations we aim to distinguish safe drivers from unsafe drivers and apply this knowledge in driver support systems, driver training and assessment.

Acknowledgements

This work was funded by the Dutch Ministry of Economic affairs through the High Tech Automotive Systems (HTAS) program, grant no. HTASI09004–E!5395 of the Driver Observation in Car Simulators project. The research of Joost de Winter is supported by the Dutch Technology Foundation (Stichting voor de Technische Wetenschappen, STW), the Applied Science Division of the Netherlands Organization for Scientific Research (Nederlandse Organisatie voor Wetenschappelijk Onderzoek, NWO) and the Technology Program of the Ministry of Economic Affairs.

References

1. Williams, A. F. (2006). Young driver risk factors: Successful and unsuccessful approaches for dealing with them and an agenda for the future. *Injury Prevention* **12,** i4-i8.
2. Smiley, A. & Brookhuis, K. A. (1987). Alcohol, drugs and traffic safety. In J. A. Rothengatter & R. A. de Bruin (eds) *Road users and traffic safety* (Assen: Van Gorcum), 83-105.
3. Lee, J. D. (2007). Technology and teen drivers. *Journal of Safety Research* **38**, 203-213.
4. Stelmach, G. E., *&* Nahom, A. (*1992*). Cognitive-motor abilities of the elderly driver. *Human Factors* **34**, 53-65.
5. Janke, M. K., Masten, S. V., McKenzie, D. M., Gebers, M. A., & Kelsey, S.L. (2003). Teen and senior drivers (Report No. 194). Sacramento, CA: California Department of Motor Vehicles
6. Sivak, M. (*1996*). The information that drivers use: is it indeed 90% visual? *Perception* **25**, 1081-1089
7. Underwood, G. (2007). Visual attention and the transition from novice to advanced driver. *Ergonomics* **50**, 1235-1249
8. Medeiros, F. A., Weinreb, R. N., Boer, E. R. & Rosen, P. N. (in press). Driving simulation as a performance-based test of visual impairment in glaucoma. *Glaucoma*.
9. Crundall, D. E. & Underwood, G. (1998). Effects of experience and processing demands on visual information acquisition in drivers. *Ergonomics* **41**, 448-458
10. Pradhan, A. K., Hammel, K. R., DeRamus, R., Pollatsek, A., Noyce, D. A., & Fisher, D. L. (2005). The use of eye movements to evaluate the effects of driver age on risk perception in an advanced driving simulator. *Human Factors* **47**, 840-852.
11. Vlakveld, W. P. (2011). Hazard anticipation of young novice drivers : assessing and enhancing the capabilities of young novice drivers to anticipate latent hazards in road and traffic situations (Doctoral dissertation). University of Groningen, the Netherlands.
12. Green Dino. (2012). Retrieved March 27, 2012, from http://www.greendino.nl/.

13. Katzourakis, D., De Winter, J. C. F., De Groot, S., & Happee, R. (in press). Driving simulator parameterization using double-lane change steering metrics as recorded on five modern cars. *Simulation Modelling Practice and Theory*.

14. De Winter, J. C. F., De Groot, S., Mulder, M., Wieringa, P. A., Dankelman, J., & Mulder, J. A. (2009). Relationships between driving simulator performance and driving test results. *Ergonomics* ***52***, 137-153

15. Van Leeuwen, P. M., De Groot, S, Happee, R, De Winter, J. C. F. (2011). Effects of concurrent continuous visual feedback on learning the lane keeping task. *Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 482-488. Retrieved from:

http://drivingassessment.uiowa.edu/sites/default/files/DA2011/Papers/070_vanLeeuwendeGroot.pdf

Identifying Driver Behaviour in Steering: Effects of Preview Distance

H.J. Damveld[1], R. Happee[2]

[1]*Faculty of Aerospace Engineering, Delft University of Technology, Delft, the Netherlands.*
H.J.Damveld@tudelft.nl

[2]*Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft, the Netherlands.*
R.Happee@tudelft.nl

Operator identification methods have been used extensively to identify pilot models while controlling aircraft dynamics [1,4]. The current state of the art allows us to simultaneously identify the separate contributions of the pilot's visual system, the vestibular system and the neuromuscular system [1,2,3]. System identification was also used successfully to observe changes in the estimated pilot model parameters due to changes in the simulator motion cueing algorithm [5].

In this paper, the first steps are taken to bring identification in driver steering tasks up to the same level as identification in aircraft piloting tasks. A study was performed to provide measurement data, to allow identification of the driving behaviour in a curve negotiation task. The study aimed to separate two essential driver control components, being:

1. preview (feedforward) control using visual information regarding the upcoming road curvature, and
2. compensatory (feedback) control to minimise lateral position and heading error.

Figure 1 shows a model adapted from [6] which shows two compensatory control loops, responding to a heading error (ψ) and a lateral error (y). Also a feedforward loop is shown, responding directly to the commanded heading ψ_c.

In a fixed base driving simulator, drivers were instructed to drive as they would normally do over a winding road. Two separate stimuli, commonly called forcing functions, were simultaneously applied being 1) the commanded road heading ψ_c, and 2) a steering wheel disturbance δ_{sd}, representing disturbances due to wind and

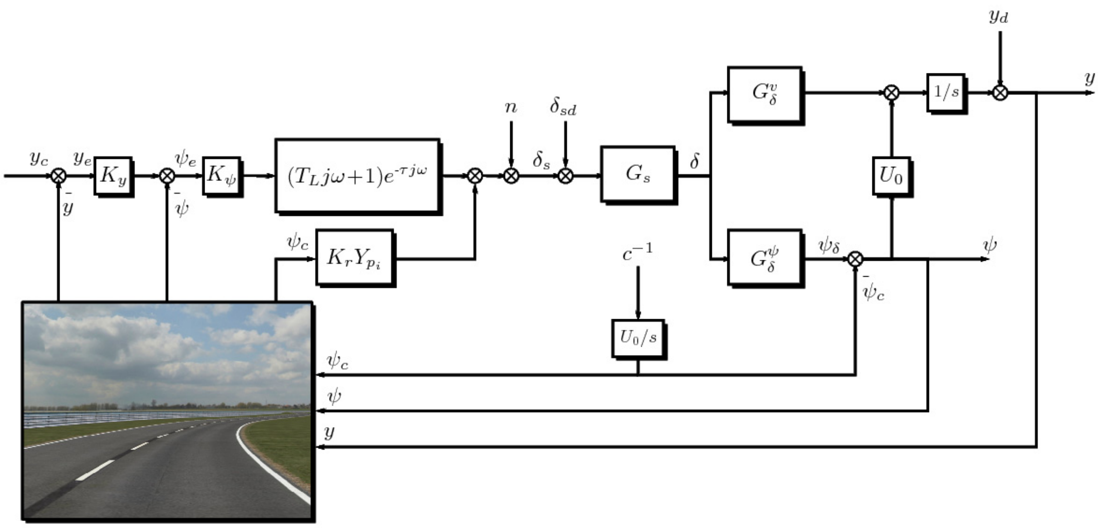


Figure 1. Driver-Vehicle Model adapted from [6].

Proceedings of Measuring Behavior 2012 (Utrecht, The Netherlands, August 28-31, 2012)
Eds. A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, and P.H. Zimmerman

100m preview

15m preview

5m preview

Figure 2. Three different preview distances.

road irregularities. The application of the forcing functions allows the simultaneous identification of preview and compensatory driver control loops. The forcing functions were defined in the frequency domain and ranged between 0.02 and 1 Hz. To further investigate the use of preview information, the distance of the path preview that was visible to the driver was varied. Figure 2 illustrates the three different preview conditions.

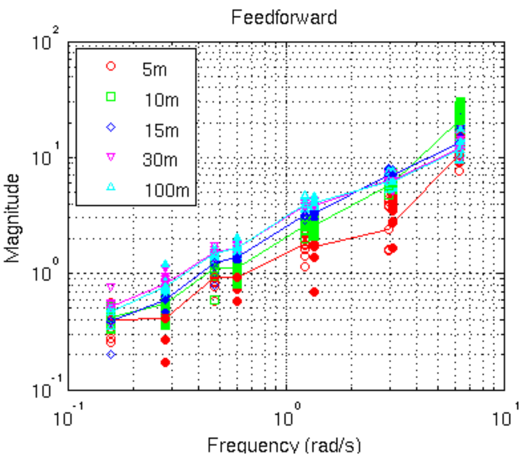
Figure 3 shows the resulting driver steering action magnitude relative to the road curvature. The feedforward system shows an increasing gain with higher frequencies, and this behaviour almost perfectly compensates for the vehicle characteristics which have a reduced gain for higher frequencies. The feedback system shows a more complex behaviour resulting from combined position and heading compensation.

Results with varying preview distance show that drivers rely more on preview with increasing preview distance and that compensatory behaviour is reduced. Although not all results were statistically significant, several measures show that the highest performance was not reached at the maximum preview distance used in this study (100m) but at a shorter distance of 15m. This indicates that with preview above a certain point, drivers no longer minimize lateral error, but use the additional preview to obtain a smooth path.

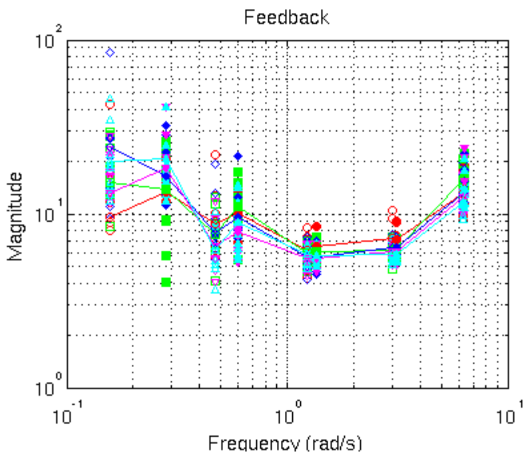
The presentation will further illustrate methods and results as an example of driver (and pilot) model identification.

References

1. Damveld, H.J. (2009). “A Cybernetic Approach to Assess the Longitudinal Handling Qualities of



Feedforward system



Feedback system

Figure 3. Identified driving behavior. The open symbols represent the commanded road heading, the close symbols represent the steering wheel disturbance.

Aeroelastic Aircraft," *Ph.D. dissertation*, Delft University of Technology, the Netherlands.

2. Damveld, H.J., Abbink, D.A., Mulder, M., Mulder, M., Van Paassen, M.M., Van der Helm, F.C.T., Hosman, R.J.A.W. (2009). "Measuring the Contribution of the Neuromuscular System During a Pitch Control Task," in *Proceedings of the AIAA Modeling and Simulation Technologies Conference*, Chicago, Illinois, Aug. 10-13, 2009, no. AIAA-2009-5824. American Institute of Aeronautics and Astronautics.

3. Damveld, H.J., Abbink, D.A., Mulder, M., Mulder, M., Van Paassen, M.M., Van der Helm, F.C.T., Hosman, R.J.A.W. (2010). "Identification of the Feedback Component of the Neuromuscular System in a Pitch Control Task," in *Proceedings of the AIAA Guidance, Navigation, and Control Conference* 2 - 5 August 2010, Toronto, Ontario Canada, M. Silvestro, Ed., no. AIAA 2010-7915, American Institute of Aeronautics and Astronautics. American Institute of Aeronautics and Astronautics, Aug. 2010, 1–22.

4. Pool, D.M., Zaal, P.M.T., Damveld, H.J., Van Paassen, M.M., Mulder, M. (2009). "Pilot Equalization in Manual Control of Aircraft Dynamics," in Proceedings of the 2009 *IEEE International Conference on Systems, Man, and Cybernetics San Antonio*, TX, USA - October 2009, Institute of Electrical and Electronics Engineers. Institute of Electrical and Electronics Engineers, 2480–2485.

5. Zaal, P.M.T., Pool, D.M., Van Paassen, M.M., Mulder, M. (2011). "Comparing Multimodal Pilot Pitch Control Behavior Between Simulated and Real Flight," in *Proceedings of the AIAA Modeling and Simulation Technologies Conference*, Portland, Oregon, Aug. 8-11.

6. Allen, W.R., McRuer, D.T., Hogge, J.R., Humes, R.W., Stein, A.C., O'hanlon, J.F., Hennessy, R.T., Kelley, G.R., and Bailey, G.V. (1977). "Drivers' Visibility Requirements For Roadway Delineation, Volume I: Effects of Contrast and Configuration on Driver Performance and Behavior," Systems Technology, Inc. FHWA-RD-77-165.

Advantages and Disadvantages of Driving Simulators: A Discussion

J.C.F. de Winter, P.M. van Leeuwen, R. Happee

Department of BioMechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft, The Netherlands. j.c.f.dewinter@tudelft.nl

Introduction

For half a century at least, each new generation of computer chip, appearing every two years or so, has provided twice as many transistors per unit cost (as predicted by Moore's law [1]). Extrapolating the accelerating pace of technological potential, logic dictates that computational devices will be tiny and powerful by the early 2020s, giving rise to virtual-reality applications such as displays built into our eyeglasses [2]. By the 2030s, going to a website could mean entering a totally realistic and compelling virtual environment facilitated by miniature computers that interact with brain cells, and people will probably spend most of their time in virtual reality [3]. Although these prospects may sound farfetched, we can already see that virtual-reality-based applications are increasingly used in such common tasks as driving. We use driving simulators for assessment, rehabilitation, learner driver training, race-car driver training, research into safety, and for at-home entertainment and in amusement halls.

Advantages of driving simulators

Driving simulators offer various advantages compared to real vehicles, including:

1. *Controllability, reproducibility, and standardization.* Behavior of virtual traffic, weather conditions, and the road layout can be manipulated (offline or in real-time) as a function of the training needs or research aims. Purpose-developed scenarios enable trainees to practice a large number of dedicated maneuvers per time unit. Wassink et al. [4] describe software architecture for generating dynamic scenarios in a driving simulator. With the aim of maximizing the effectiveness of the training, the authors apply a metaphor from the 1998 movie The Truman Show: everything surrounding the learner driver responds to the driver's behavior. Using simulators, participants in different physical locations can drive under the exact same conditions. This is beneficial for creating standardized driving tests and reproducible research results. In contrast, the real traffic environment is largely random.

2. *Ease of data collection.* A driving simulator can measure performance accurately and efficiently. With a real vehicle, it is far more cumbersome to obtain complete, synchronized, and accurate measurement data. It is a fundamental challenge to get an accurate recording of where a real vehicle actually is in the world. For example, in one study using an instrumented vehicle and a driving simulator, it was impossible to determine the distance between the vehicle and a stop line on the road, while in the simulator this information was readily available [5]. Measurement of lateral position is challenging as well, as this requires visible lane markers while weather conditions, reflection, and shades may affect the quality of the measurement [6]. Santos et al. [7] found that lateral position measurements of the instrumented vehicle were of marginal quality while this information was accurate in the simulator, leading the authors to conclude that "problems with field studies in an instrumented vehicle have been confirmed" (p. 145). Because of the measurement capabilities of simulators, new types of behavior analyses come within reach, such as trigonometric analysis of time-to-line crossing [8] or object detection and hazard perception research using eye-tracking [9].

3. *Possibility of encountering dangerous driving conditions without being physically at risk.* Simulators can be used to prepare trainees to handle unpredictable or safety-critical tasks that may be inappropriate to practice on the road, such as collision avoidance or risky driving [10]. In addition, simulators make it possible to study hazard anticipation and perception by exposing drivers to dangerous driving tasks, which is an ethically challenging endeavor in real vehicles [9]. Flach et al. [11] stated that simulators

"offer an opportunity to learn from mistakes in a forgiving environment" (p. 134). Allen et al. [12] made a similar case: "Motor vehicle crashes are significantly higher among young drivers during the first year of licensure, and crash risks decline with increased experience. […] This produces an interesting dilemma about how to provide young drivers with driving experience without significantly increasing their crash risk. Driving simulation may be the solution to this dilemma."

4. *Novel opportunity for feedback and instruction.* Simulators offer the opportunity for feedback and instruction that is not easily achieved in real vehicles. For example, it is possible to freeze, reset, or replay a scenario [13]. Feedback and instructions can also be delivered in other modalities besides speech, such as visual overlays to highlight critical features in the environment.

Disadvantages of driving simulators

However, simulators have several known disadvantages and challenges, including:

1. *Limited physical, perceptual, and behavioral fidelity.* Low-fidelity simulators may evoke unrealistic driving behavior and therefore produce invalid research outcomes. Simulator fidelity is known to affect user opinion. Participants may become demotivated by a limited-fidelity simulator and prefer a real vehicle instead (or a more costly high-fidelity simulator for that matter). Interestingly, while safety is often cited as an advantage of driving simulation (see above), sometimes this same feature is interpreted as a disadvantage. For example, Käppler [14] pointed out that real danger and the real consequences of actions do not occur in a driving simulator, giving rise to a false sense of safety, responsibility, or competence. Simply investing resources to increase fidelity is not necessarily a desirable solution, as it adds to the complexity of the device and might hamper experimental control. In some cases, deliberate deviations from reality yield valid results [15][16]. Evans [17] provided an interesting thought experiment, arguing against a blind focus on high-fidelity driving simulation: "Consider a make-believe simulator consisting of an actual car, but with the remarkable property that after it crashes a reset button instantly cancels all damage to people and equipment. What experiments could be performed on such make-believe equipment that would increase our basic knowledge about driving? The answers provide an upper limit on what might be done using improved simulators" (p. 190).

2. *Shortage of research demonstrating validity of simulation.* A growing body of evidence indicates that driving-simulator measures are predictive for on-the-road driving performance [18]-[23]. However, only a few studies have investigated whether skills learned in a driving simulator transfer to the road (see [24]-[26] for a few exceptions). Note that in the field of aviation, studies on the transfer of training are far more common [27], but even in aviation critical questions remain unanswered, for example whether a motion base provides added value for the effectiveness of flight training [28].

3. *Simulator discomfort, especially in older people or under demanding driving conditions.* Simulator sickness symptoms may undermine training effectiveness and negatively affect the usability of simulators. This is a serious concern, but fortunately, useful technological and procedural guidelines are available to alleviate it [29]. Research shows that simulator sickness is less of a problem for young drivers [30]. Experience shows that limiting the horizontal field of view, avoiding sharp curves or stops during driving, and using short sessions (≤10 min) with sufficient rest breaks improves or even eliminates simulator sickness.

Conclusion

Because of the increasing potential of computer technology, we foresee increasing use of driving simulation in areas such as driver assessment, driver training, research, and entertainment. Low-cost virtual-reality applications will come within the reach of many organizations. However, several research questions may need to be answered before ubiquitous driving simulation becomes feasible, particularly questions related to simulator fidelity, predictive validity of driving simulators, simulator-to-reality transfer of learning, and simulator discomfort.

References

1. Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electron* **38**, 114-117.

2. Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. Viking, New York.

3. Kurzweil, R. (2004). Foreword. In *Virtual humans: A build-it-yourself kit, complete with software and step-by-step instructions of book* (pp. xi-xiii). New York: Amacom.

4. Wassink, I., Van Dijk, B., Zwiers, J., Nijholt, A., Kuipers, J., Brugman, A. (2006). In The Truman Show: Generating dynamic scenarios in a driving simulator. *IEEE Intelligent Systems* **21**, 28-32.

5. Godley, S. T., Triggs, T.J., Fildes, B.N. (2002). Driving simulator validation for speed research. *Accident Analysis and Prevention* **34**, 589-600.

6. Roskam, A.J., Brookhuis, K.A., De Waard, D., Carsten, O.M.J., Read, L., Jamson, S., … Victor, T. (2002). HASTE Deliverable 1: Development of experimental protocol. Retrieved from <http://ec.europa.eu/transport/roadsafety_library/publications/haste_deliverable_1_v1_1.pdf> Accessed 31 May 2012.

7. Santos, J., Merat, N., Mouta, S., Brookhuis, K., De Waard, D. (2005). The interaction between driving and in-vehicle information systems: Comparison of results from laboratory, simulator and real-world studies. *Transportation Research Part F* **8**, 135-146.

8. Van Winsum, W., Brookhuis, K. A., De Waard, D. (2000). A comparison of different ways to approximate time-to-line crossing (TLC) during car driving. *Accident Analysis and Prevention* **32**, 47-56.

9. Underwood, G., Crundall, D., Chapman, P. (2011). Driving simulator validation with hazard perception. *Transportation Research Part F* **14**, 435-446.

10. Hoeschen, A., Verwey, W., Bekiaris, E., Knoll, C., Widlroither, H., De Waard, D., et al. (2001). TRAINER: Inventory of driver training needs and major gaps in the relevant training procedures (Deliverable No 2.1). Brussels, Belgium: European Commission. Retrieved from <http://ec.europa.eu/transport/roadsafety_library/publications/trainer_deliverable_2_1.pdf> Accessed 31 May 2012.

11. Flach, J.M., Dekker, S., Stappers, P.J. (2008). Playing twenty questions with nature (the surprise version): Reflections on the dynamics of experience. *Theoretical Issues in Ergonomics Science* **9**, 125-154.

12. Allen, R.W., Park, G.D., Cook, M.L., Fiorentino, D. (2007). The effect of driving-simulator fidelity on training effectiveness. *Proceedings of the Driving Simulation Conference North America*, Iowa City, IA.

13. Vlakveld, W.P. (2005). The use of simulators in basic driver training. HUMANIST Workshop on the application of new technologies to driver training, Brno, Czech Republic. Retrieved from http://www.esafetysupport.org/download/research_and_development/HUMANISTA_13Use.pdf

14. Käppler, W.D. (1993). Views on the role of simulation in driver training. *Proceedings of the 12th European Annual Conference on Human Decision Making and Manual Control*, Kassel, Germany, 5.12-5.17.

15. Lee, J.D. (2004). Simulator fidelity: How low can you go? *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society*, Santa Monica, CA.

16. Reed, M.P., Green, P.A. (1999). Comparison of driving performance on-road and in a low-cost simulator using a concurrent telephone dialling task. *Ergonomics* **42**, 1015-1037.

17. Evans, L. (2004). Traffic safety. Bloomfield Hills, MI: Science Serving Society.

18. Bédard, M.B., Parkkari, M., Weaver, B., Riendeau, J., Dahlquist, M. (2010). Assessment of driving performance using a simulator protocol: Validity and reproducibility. *The American Journal of Occupational Therapy* **64**, 336-340.

19. De Winter, J.C.F., De Groot, S., Mulder, M., Wieringa, P.A., Dankelman, J., Mulder, J.A. (2009). Relationships between driving simulator performance and driving test results. *Ergonomics* **52**, 137-153.

20. Lee, H-C., Lee, A.H., Cameron, D., Li-Tsang, C. (2003). Using a driving simulator to identify older drivers at inflated risk of motor vehicle crashes. *Journal of Safety Research* **34**, 453-459.

21. Lee, H-C., Cameron, D., Lee, A.H. (2003). Assessing the driving performance of older adult drivers: on-road versus simulated driving. *Accident Analysis and Prevention* **35**, 793-803.

22. Lew, H.L., Poole, J.H., Lee, E.H., Jaffe, D.L., Huang, H-C., Brodd, E. (2005). Predictive validity of driving-simulator assessments following traumatic brain injury: A preliminary study. *Brain Injury* **19**, 177-188.

23. Shechtman, O., Classen, S., Awadzi, K., Mann, W. (2009). Comparison of driving errors between on-the-road and simulated driving assessment: a validation study. *Traffic Injury Prevention* **10**, 379-385.

24. Strayer, D. L., Drews, F. A. (2003). Simulator training improves driver efficiency: Transfer from the simulator to the real world. *Proceedings of the second international driving symposium on human factors in driver assessment, training and vehicle design*, Park City, UT, 190-193.

25. Strayer, D.L., Drews, F.A., & Burns, S. (2005). The development and evaluation of a high-fidelity simulator training program for snowplow operators. *Proceedings of the third international driving symposium on human factors in driver assessment, training and vehicle design*, Rockport, ME, 464-470.

26. Uhr, M. B. F., Felix, D., Williams, B. J., Krueger, H. (2003). Transfer of training in an advanced driving simulator: Comparison between real world environment and simulation in a manoeuvring driving task. In *Proceedings of the driving simulation conference North America*, Dearborn, MI.

27. Jacobs, J.W., Prince, C., Hays, R.T., Salas, E. (1990). *A meta-analysis of the flight simulator training research* (TR-89-006). Orlando, FL: Naval Training Systems Center.

28. De Winter, J.C.F., Dodou, D., Mulder, M. (2012). Training effectiveness of whole body flight simulator motion: A comprehensive meta-analysis. *The International Journal of Aviation Psychology* **22**, 164-183.

29. Kolasinski, E. M. (1995). *Simulator sickness in virtual environments* (Technical Report 1027). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

30. Brooks, J.O., Goodenough, R.R., Crisler, M.C., Klein, N.D., Alley, R.L. *et al*. (2010). Simulator sickness during driving simulation studies. *Accident Analysis and Prevention* **42**, 788-796.

Driver and Pilot Identification and Model Parameter Estimation; Modelling the Visual, Vestibular, and Neuromuscular Control Loops Describing Driver and Pilot Behaviour

R. Happee[1], H.J. Damveld[2], D.A. Abbink[1], M.M. van Paassen[2], M. Mulder[2]

[1]*Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft, The Netherlands.*
R.Happee@tudelft.nl; D.A.Abbink@tudelft.nl

[2]*Faculty of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands.*
H.J.Damveld@tudelft.nl, M.M.vanPaassen@tudelft.nl, M.Mulder@tudelft.nl

Summary

Driver and pilot modelling has been successfully applied to unravel driver and pilot behaviour, and to design systems optimally matching driver and pilot capacities. Modelling has been applied in particular for continuous manual control tasks such as steering where drivers apply continuous steering actions to control vehicle heading and lateral position. In such models the visuomotor loop describes steering wheel rotation as a function of visual information regarding the road and the vehicle state. Some pilot models include a vestibular component which describes how vestibularly perceived motion contributes to pilot behaviour. Interfaces such as steering wheels in cars provide drivers with haptic (force) feedback regarding the vehicle state, and to optimally design such interfaces neuromuscular models have been developed describing how operators use reflexes to control limb position, force or stiffness.

This paper describes state of the art driver and pilot models and shows how the visual, vestibular, and neuromuscular control loops have been tested and modelled. In particular it describes how multiple stimuli have been applied simultaneously to identify operator behaviour in relatively short experiments eliciting steady state behaviour. Human control behaviour is shown to adapt effectively and systematically to the dynamics of the system being controlled as well as to task instructions and applied stimuli.

Methods

Driver and pilot (or operator) model identification enables the estimation of operator model parameters based on measured operator responses to certain stimuli. Operator model identification has been successfully applied to capture human control behaviour in continuous closed-loop tasks such as aircraft control, car following and steering. Human control actions could be described as (linear, time-invariant) function of task-related stimuli and perceivable responses of the controlled vehicle. Such models describe how operators dynamically control the acceleration, velocity and position of vehicles to follow a desired path and to correct for possible disturbances. Complex operator models have been derived simply fitting available data and/or hypothesizing plausible control loops [9]. This paper focusses on operator models which can be uniquely identified from dedicated experiments on individual drivers. Here unique identification refers to the possibility to uniquely derive a set of operator parameters with good experimental reproducibility [16]. Unique identification makes operator model identification a sensitive design tool. For proposed vehicles and user interfaces, operator behaviour can be identified using driving or flight simulators. Comparing operator model parameters for different vehicle designs, we can measure whether and how operators adapt their control behaviour. Effects of proposed vehicle modifications have shown to provide valuable insight in operator behaviour, and have shown significant and relevant changes in cases where traditional performance measures reflecting task accuracy were not significantly changed as a result of effective adaptation by the operator

Operator models have been identified using data collected in real vehicles and in simulators (see Figure 1). To uniquely identify the various control loops dedicated stimuli have been designed aiming to create realistic task conditions as well as enabling unique identification. In general such stimuli are designed to be random appearing such that operator behaviour is elicited for an unknown task rather than testing “pre-programmed” behaviour.

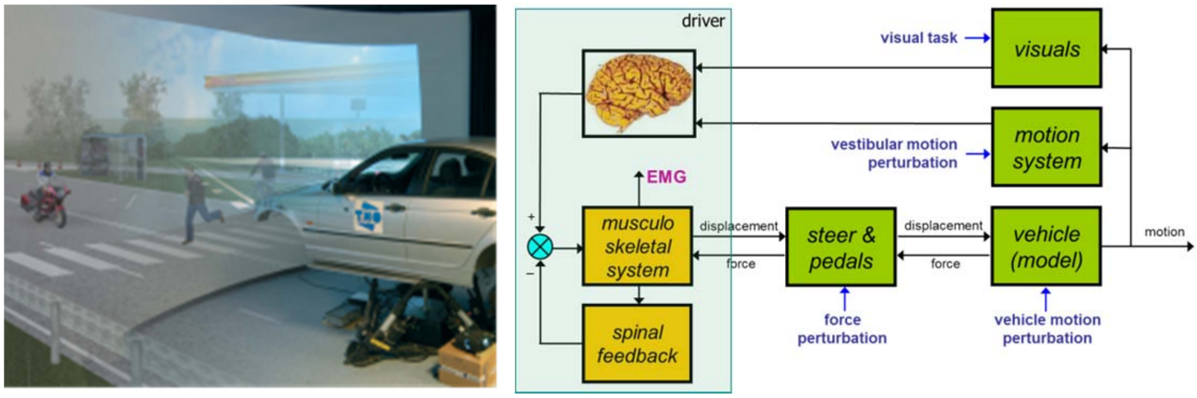


Figure 1. Driving Simulator; a human operator interacts with a virtual vehicle in a virtual driving environment. The driver receives visual task stimuli and various perturbations and driver actions and vehicle responses are recorded.

Visual stimuli primarily represent task related information, and can include desired states, actual states and errors between desired and actual states. Simple visual stimuli such as desired heading, speed, and position can be presented on dashboard or cockpit displays. Complex visual stimuli such as road geometry and other vehicles are naturally available in real vehicle testing, and can be presented virtually in simulators. Obviously in mathematical models such complex stimuli require derivation of simple variables assumed to be perceived by the operator from the available visual information field. Such variables represent perceived ego-vehicle motion: e.g. longitudinal and lateral position, heading and their derivatives in time, observed road geometry, and (relative) position, velocity and heading of other vehicles. Visual stimuli are primarily applied to identify the operators visuomotor control behaviour.

Vehicle motion stimuli can represent perturbations of the vehicle motion for instance related to wind or other physical disturbances. Vehicle motion stimuli can be used to jointly identify the visuomotor and vestibular control loops. As illustrated in Figure 1 a vehicle motion perturbation will be perceived visually as well vestibularly, and hence such stimuli are not suitable to separate these two components.

Vestibular motion stimuli can be applied in simulators only. Where in actual vehicles the visual and vestibular information regarding vehicle motion are mechanically linked, in a simulator additional perturbations can be applied on the simulator motion without affecting the visually perceived vehicle motion. In principle this enables separate identification of the vestibular control loop, but humans easily notice a mismatch between vestibular and visual information, and discard apparently incongruent vestibular information.

Neuromuscular stimuli can include disturbance forces applied to steer, pedals and other interfaces. Measuring the resulting displacement and muscular activity (EMG) we can model the human stiffness or compliance, and estimate neuromuscular feedback delays and gains reflecting position control using muscle spindles and force control using Golgi tendon organs [10]. Neuromuscular identification is valuable in particular for the development of haptic (force) feedback systems supporting the driver with guiding forces, where identification showed a systematic adaptation of neuromuscular control when haptic feedback is provided [1,3].

Results

Operator model identification methods have been used extensively to identify pilot models while controlling aircraft dynamics [5,13]. Neuromuscular control models were already identified in the 90ties [17,18]. The current state of the art allows us to simultaneously identify the separate contributions of the pilot’s visual system, the vestibular system and the neuromuscular system [5,6,7]. System identification was also used successfully to observe changes in the estimated pilot model parameters due to changes in the motion cueing algorithm [19,20].

Neuromuscular feedback of the ankle joint while controlling a gas pedal was identified and used to design a prototype haptic gas pedal [1,4,11,12] that has led to the marketing of the Distance Control Assist (DCA) system available in the Nissan Infiniti in Japan and USA. With the haptic gas pedal, subjects were shown to adapt their neuromuscular feedback strategy controlling force rather than position. In these studies, a visual control loop was identified using the perceivable distance and relative velocity towards the lead vehicle. This visual feedback model for car following, was also used to evaluate a visual driver support system, and showed significant beneficial effects of a distance and acceleration display on the driver parameters [14,15].

Neuromuscular feedback of the upper extremities in car steering has been identified [2] and the use of visual information in car steering will be illustrated in a separate presentation at measuring behaviour [8].

References

1. Abbink, D.A., Mulder, M. (2010). Neuromuscular Analysis as a Guideline in designing Shared Control. *Advances in Haptics.* pp. 499-516.

2. Abbink, D.A, Mulder, M., van der Helm F.C.T., Mulder, M., Boer, E.R. (2011). Measuring Neuromuscular Control Dynamics During Car Following With Continuous Haptic Feedback, *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics.*

3. Abbink, D.A., Mulder, M., van Paassen M.M. (2011). Measurements of Muscle Use during Steering Wheel Manipulation, in: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Anchorage, Alaska, USA, pp. 1652-1657, IEEE, 2011.

4. Abbink, D.A., Mulder, M. (2009). Exploring the Dimensions of Haptic Feedback Support in Manual Control. *J. Comput. Inf. Sci. Eng.* **9**(1), 011006 (9 pages). DOI:10.1115/1.3072902.

5. Abbink, D.A. (2006). *Neuromuscular analysis of haptic gas pedal feedback during car following*. PhD dissertation, Delft University of Technology, the Netherlands.

6. Damveld, H.J. (2009). *A Cybernetic Approach to Assess the Longitudinal Handling Qualities of Aeroelastic Aircraft*, Ph.D. dissertation, Delft University of Technology, the Netherlands.

7. Damveld HJ, Abbink DA, Mulder M, Mulder M, van Paassen MM, van der Helm FCT, Hosman RJAW. Measuring the Contribution of the Neuromuscular System During a Pitch Control Task. In *Proceedings of the AIAA Modeling and Simulation Technologies Conference*, Chicago, Illinois, Aug. 10-13, 2009, no. AIAA-2009-5824. American Institute of Aeronautics and Astronautics, Aug. 2009.

8. Damveld HJ, Abbink DA, Mulder M, Mulder M, van Paassen MM, van der Helm FCT, Hosman RJAW. Identification of the Feedback Component of the Neuromuscular System in a Pitch Control Task. In *Proceedings of the AIAA Guidance, Navigation, and Control Conference* 2 - 5 August 2010, Toronto, Ontario Canada, M. Silvestro, Ed., no. AIAA 2010-7915, American Institute of Aeronautics and Astronautics. American Institute of Aeronautics and Astronautics, Aug. 2010, 1–22.

9. Damveld, H.J., Happee, R. (2012). Identifying Driver Models in Steering: Effects of Preview Distance. *Measuring Behaviour, Utrecht, The Netherlands, August 2012.*

10. Katzourakis, D., Droogendijk, C., Abbink, D., Happee, R., Holweg E. (2010). Driver Model with Visual and Neuromuscular Feedback for Objective Assessment of Automotive Steering Systems. *AVEC2010, International Symposium on Advanced Vehicle Control*, August 2010.

11. Mugge, W., Abbink, D.A., van der Helm, F.C.T, DeWald, J. (2009) A rigorous model of reflex function indicates that position and force reflexes are flexibly tuned to position and force tasks. *Experimental Brain Research* **200**, 325-334.

12. Mulder, M., Abbink, D.A., van Paassen, M.M. (2011). Design of a Haptic Gas Pedal for Active Car-Following Support. *IEEE Transactions on Intelligent Transportation Systems* **12**(1), 268-279.

13. Mulder, M., Pauwelussen, J.A., van Paassen, M.M., Mulder, M., Abbink, D.A. (2010). Active Deceleration Support in Car Following. *IEEE Transactions on Systems, Man, and Cybernetics—Part A* **40**(6), 1271-1284.

14. Pool, D.M., Zaal, P.M.T., Damveld, H.J., van Paassen, M.M., Mulder, M. (2009). Pilot Equalization in Manual Control of Aircraft Dynamics," in Proceedings of the 2009 *IEEE International Conference on Systems, Man, and Cybernetics San Antonio*, TX, USA - October 2009, Institute of Electrical and Electronics Engineers, 2480–2485.

15. Saffarian, M., Happee, R. (2011). Supporting Drivers in Car Following: A Step towards Cooperative Driving. *The intelligent vehicles symposium (IV'11)*, Baden-Baden, Germany, June 2011.

16. Saffarian, M., Winter, J.C.F. de, Happee, R. (2012). Enhancing driver car following performance through distance and acceleration display. *In Press. IEEE Transactions on Systems Man & Cybernetics, Part A.*

17. Steen, J., Damveld, H.J., Happee, R., van Paassen, M.M., Mulder, M. (2011). A Review of Visual Driver Models for System Identification Purposes. *IEEE SMC Conference 2011*.

18. van Paassen MM. (1995). A Model of the Arm's Neuromuscular System for Manual Control. *Proc. IFAC Analysis, Design and Evaluation of Man-Machine Systems*, Cambridge, USA 1995.

19. van Paassen, M.M. *Biophysics in Aircraft Control—a Model of the Neuromuscular System of the Pilot's Arm*, Ph.D. Dissertation, Faculty of Aerospace Engineering, Delft Univ. of Technology, Delft, The Netherlands, June 1994.

20. Zaal, P.M.T., Pool, D.M., de Bruin, J., Mulder, M., van Paassen, M.M. Use of Pitch and Heave Motion Cues in a Pitch Control Task. *Journal of Guidance, Control, and Dynamics* **32**(2), 366-377, DOI: 10.2514/1.39953.

21. Zaal, P.M.T., Pool, D.M., van Paassen, M.M., Mulder, M., Comparing Multimodal Pilot Pitch Control Behavior Between Simulated and Real Flight. In *Proceedings of the AIAA Modeling and Simulation Technologies Conference*, Portland, Oregon, Aug. 8-11, 2011.

These references list research as published by TU Delft. Work from other groups has been cited extensively in the above papers.

Can We Trust Driver Behaviour Assessment?
Examples from Research in Simulators and in the Field

Katja Kircher, Christer Ahlström

The Swedish National Road and Transport Research Institute (VTI), Linköping, Sweden

Introduction

It is very common to compare mean values of driving performance indicators (PI) like mean speed, the standard deviation of lateral position, time headway, mean glance duration, and many more, in order to investigate possible differences between different treatment groups. Just like all the PIs mentioned here, most of them describe aspects of the control level of driving behaviour according to Michon's control hierarchy [1]. When means differ significantly between treatment groups, this is often interpreted in relation to traffic safety gains or losses.

In this paper we are going to discuss possible pitfalls with the use and interpretation of such PIs, based on examples from different studies. Finally, we suggest a number of possible solutions to avoid some of the issues discussed here.

Individual differences

As people may perform differently under similar conditions, it is common to compare the mean values of groups to find out whether the groups differ systematically in their performance. How the individuals differ within the groups, and what the possible reasons for those differences are, is often neglected. This is illustrated with an example from a study conducted in two different driving simulators (Figure 1). In addition to a baseline drive the participants had to drive on a relatively curvy rural road while either performing a mostly visual, a solely cognitive, or a mostly haptic task. They did this once under car-following and once under free driving conditions. The additional tasks were presented on three different levels of difficulty. The participants were instructed to complete those tasks as best they could, while still having in mind that they were driving, that is, the main focus was directed at the additional task [2].

Several aspects can be noticed by taking a closer look at **Error! Reference source not found.**, which shows the time in seconds a driver spent with at least two wheels in the oncoming lane. The total time driven per condition was 30 s. The thin lines stem from individual drivers. Especially for the visual condition it is noticeable how much the values vary between participants. What is more interesting, they do not only vary in a quantitative manner, with one participant getting into the oncoming traffic more than the other, but also in a qualitative manner. Except for the visual condition, most participants do not leave their lane at all, or for less than 2 s, as can be deducted from the median values. Some other participants, however, drive with at least two wheels in the oncoming lane for a third of the time or more. This may be interpreted as some participants losing control, or consciously accepting a drift into the oncoming lane, while others manage to keep the vehicle within the lane boundaries. With those data and interpretations as background, it does not seem meaningful to analyse mean values for PIs like this one. Instead it might be much more educative to investigate the underlying factors that might lead to the quantitatively different individual results.

Systematic differences

Especially when comparing the results from different studies, systematic differences between those studies can influence the results in unforeseen ways. These differences can for example be coupled to participant selection, the test bed in use, or the way the PI was computed.

It has already been indicated by Greenberg et al. [3] that motion cueing may have a strong impact on lateral driving PIs when disturbances, like the execution of additional tasks, are present. A number of further studies

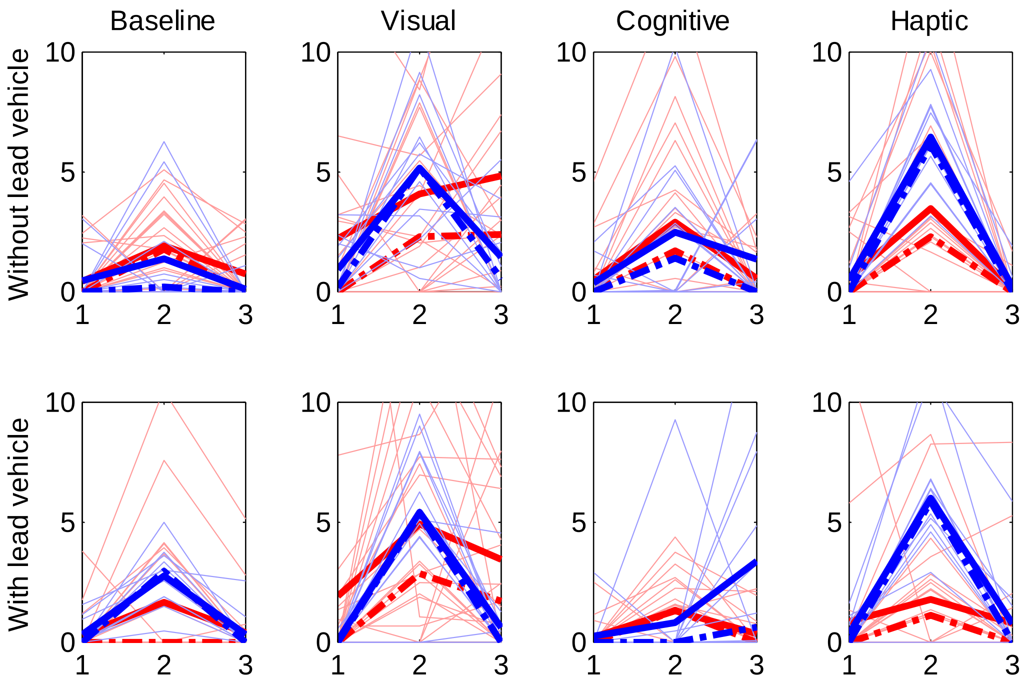


Figure 1. Individual, mean (solid lines) and median (dashed lines) lane departure time results (in s) for driving without (upper row) or with (lower row) a lead vehicle under baseline conditions, or with a visual, cognitive or haptic additional task (from left to right) in a high fidelity moving base simulator (red) or a static simulator (blue).

corroborate the fact that the type of motion cueing influences results [4,5,6]. Therefore, when comparing results between different simulators it has to be kept in mind that the measured Pis are not necessarily directly comparable.

Participant recruitment is another factor that may influence results. It is not uncommon that universities recruit participants amongst students, as this is convenient. In the study mentioned above, the participants who drove in the moving base simulator were recruited from the general public, whereas the participants for the fixed base simulator were recruited from engineers who were employed at the company running the study. The participants had also been asked to perform the additional tasks while standing still, that is, in this condition the additional task was the only task performed, so simulator differences were eliminated. For all three modalities tested, the engineers clearly and significantly outperformed the members of the general public. This difference in performance between groups in the additional task may have had a systematic impact on the capability to perform the driving task, such that studies with different types of participants may lead to qualitatively or quantitatively differing results.

Furthermore, a PI used in different studies under the same name can be computed in different ways. To illustrate this, we focus on the window size that is used to compute the PI “standard deviation of steering wheel angle”, an indication of how variably the driver turns the steering wheel over a certain period of time. The data used in the example stem from a field study, in which the drivers used an experimental car as their own during the period of a month [7] . With the help of the distraction detection algorithm AttenD, which uses eye tracking data as input, the drivers were classified as either distracted or attentive for certain time windows [8,9]. Depending on the time window used to compute the “standard deviation of steering wheel angle” there was either no significant difference or there was a significant difference between distracted and attentive drivers (**Error! Reference source not found.**). Relating back to what was discussed above, the graph containing the standard deviation of the PI in the right part of **Error! Reference source not found.** illustrates once more how wide the behavioural range is within each group, in spite of the significant differences in mean values for some of the window sizes.

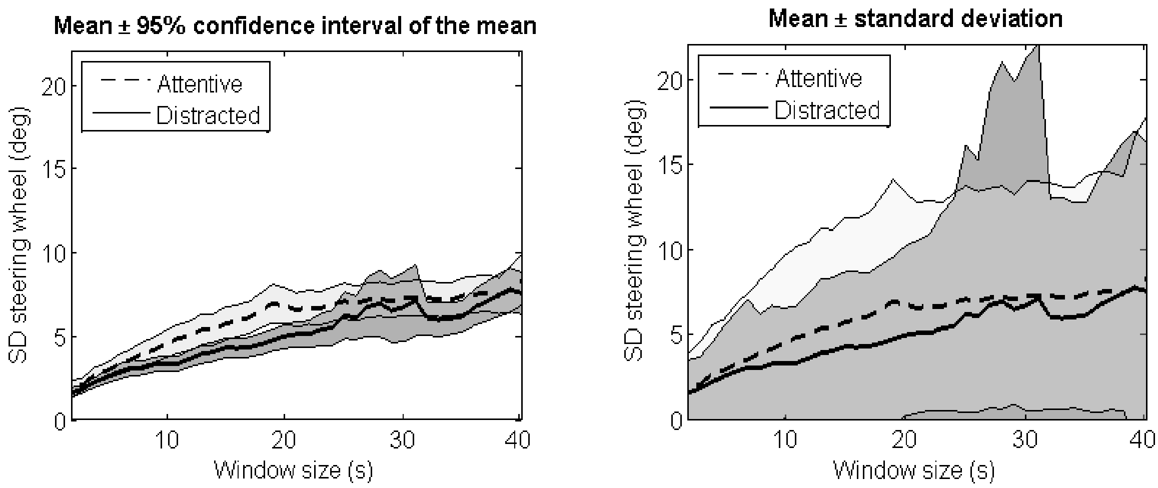


Figure 2. The PI "standard deviation of steering wheel angle" computed for data logged in the field for window sizes between 1 s and 40 s. In the graph on the left the mean with its 95 % confidence interval and in the graph on the left the mean with one standard deviation is shown.

What do performance indicators really indicate?

When it is found that on average the standard deviation of lateral position (SDLP), that is, how much a vehicle sways around its medium lateral position in the lane, increased with, say, 12 cm, what does this tell us? Obviously, the trajectory of the vehicle is somewhat less straight. Can this be linked directly to a safety hazard, though? We argue that more information is necessary before such an interpretation can be made. We would need to know, for example, whether the vehicle actually leaves the lane, whether other traffic is present, whether the other traffic comes dangerously close due to the increased swaying, whether the driver is aware of the increased swaying, whether he accepts it in the situation at hand, and so on.

Any difference in means can become significant when enough participants are used, such that the size of the measured effect plays an important role. In addition to the purely statistical effect size, which tells us how much the mean values differ in relation to the standard deviation within the population, the "practical effect size" should be considered as well. As mentioned in the example above, which practical implications does an increased SDLP have for safety, road construction, etc.? On a wide and straight road, with an average lateral position in the middle of the lane, the 12 cm increase may not have any safety implications at all. On a narrow road, however, this could mean the difference between crossing over into the oncoming lane or not. Then again, a driver may use the oncoming lane on purpose, when cutting the curves on a rural road. Obviously, the driver would not do so when oncoming traffic was present, such that an increased SDLP as a consequence of intentional behaviour does not need to be correlated to changes in the level of traffic safety at all.

Possible solutions?

It is very difficult to link this kind of PIs to actual crashes. A large-scale attempt will be made during the SHRP II programme, with 2000 instrumented vehicles and an even higher number of drivers. However, even studies of that scale will not be able to answer all the questions of what changes in driving behaviour related PIs really mean. Therefore we suggest a number of pragmatic measures to enhance the safety related validity of PIs used in studies.

Instead of mainly using PIs that describe the control level of driving behaviour, it is suggested to develop indicators that describe the tactical level. These are expected to be easier to understand and interpret. It is also assumed that the tactical level is affected by external factors sooner than the control level is, which would make such PIs more sensitive. A drawback is the fact that tactical PIs may have to be adapted to the present situation much more, which makes them harder to standardise for comparisons between studies.

PIs should have a link to violations or other behaviour that already has been linked to an increased risk in traffic. Of course, PIs can also be used to estimate efficiency or other aspects of the traffic system, but then a link to those concepts should be documented.

In order to incorporate the qualitative/quantitative aspect of behavioural differences it may be meaningful not only to evaluate the difference in mean values between groups, but to focus on the percentage of drivers who are able to perform a task within given boundaries, for example, to send an sms within a certain amount of time while remaining within one's lane and within a certain speed interval.

References

1. Michon, J. A. (1985). A Critical View of Driver Behavior Models. What Do we Know, What Should we Do? In L. Evans & R. Schwing (Eds.), *Human Behavior and Traffic Safety*. New York: Plenum Press.

2. Kircher, K., Ahlstrom, C., Rydström, A., Ljung Aust, M., Ricknäs, D., Almgren, S., et al. (2012, in press). *Secondary task workload test bench - 2TB. Final report*. Linköping, Sweden: The Swedish National Road and Transport Research Institute.

3. Greenberg, J., Artz, B., & Cathey, L. (2003). *The effect of lateral motion cues during simulated driving*. Paper presented at the DSC North America.

4. Correia Grácio, B. J., Wentik, M., Valente Pais, A. R., van Paassen, M. M., & Mulder, M. (2011). Driver behavior comparisons between static and dynamic simulation for advanced driving maneuvers. *Presence: Teleoperators and Virtual Environments* **20**(2), 143-161.

5. Feenstra, P., van der Horst, R., Correia Grácio, B. J., & Wentik, M. (2010). Effect of simulator motion cueing on steering control performance. Driving simulator study. *Transportation Research Record: Journal of the Transportation Research Board* **2185**, 48-54.

6. Valente Pais, A. R., Wentik, M., Van Paassen, M. M., & Mulder, M. (2009). Comparison of thee motion cueing algorithms for curve driving in an urban environment. *Presence: Teleoperators and Virtual Environments* **18**(3), 200-221.

7. Kircher, K., Kircher, A., & Claezon, F. (2009). *Distraction and drowsiness. A field study* (VTI Report No. 638A). Linköping: VTI.

8. Kircher, K., & Ahlstrom, C. (2009). Issues related to the driver distraction detection algorithm AttenD. *First International Conference on Driver Distraction and Inattention*.

9. Kircher, K., & Ahlstrom, C. (2010). Predicting visual distraction using driving performance data. *Association for the Advancement of Automotive Medicine Annual Scientific Conference*, Las Vegas, NV, USA.

Home-cage Automated Cognitive Phenotyping in Mice

Valter Tucci

Department of Neuroscience and Brain Technologies, Istituto Italiano di Tecnologia, Genova, Italy. valter.tucci@iit.it

Mouse mutant lines constitute promising translational models to investigate mental illness and to develop new therapeutic approaches. As advances in mouse genetics technology allow to engineer mouse models carrying complex genetic variants, there are now several large-scale projects, world-wide, to create collections of mouse mutants based on relevant phenotypic data [1-5].

However, the vast repertoire of mouse models is not currently paralleled by a correspondingly rich set of phenotyping methods. A “phenotype” can be defined as any functional or anatomical expression of the organism’s genotype and its interactions with the environment. Phenotypic assessment is required to build any link between genes and mental functions, but classical research on mental health using rodent models has mainly relied on coarse assessments (such as rudimentary behavioral analyses), which have limited translational value.

Moreover, while most biological sciences are progressing towards a holistic approach to investigate the complexity of organisms (i.e., “systems biology” approach), mouse functional genomics has embraced a more reductionist strategy that focuses on intermediate phenotypes (endophenotypes) [6]. The investigation of single endophenotypes has been a powerful approach in mouse genetics but, in many cases, it was not followed by a better understanding of human disease susceptibility [7, 8]. The reason for this poor translational result is that the genetic analysis of mouse phenotyping does not necessarily rely on an increase of isolated behavioral analyses. Instead, the success of large-scale genetic investigations by using laboratory animals very much depends on the ability: (i) to integrate meaningful and multi-dimensional comparable traits; (ii) to monitor time-variable phenotypes; (iii) to develop novel technologies that detect and analyze phenotypes in home-cage environment. In addition, phenotypic data must become easily accessible for the whole scientific community [9] to promote the continuous integration with new scientific information.

Thus there is an urgent need for new investments in basic and integrated robust phenotyping technologies [10].

Here I will present a series of behavioural/cognitive measures in mice that were derived with automated behavioural protocols. Mice were subjected to a long-term investigation in their home-cage environment. A novel operant wall was installed in regular mouse cages. The wall is equipped with two lateral hoppers which are connected to pellet dispensers and one central hopper that is used by the mouse to self-initiate the trial. All animals were trained on a timing task that required them to judge when to switch from one feeder to another for reward maximization [11]. In a given trial, reward could occur at one of the two feeding locations with some predetermined probability and after different corresponding latencies. The optimal performance in this task requires animals to have an accurate representation of time intervals (i.e., short and long intervals), the endogenous uncertainty about these intervals, and the exogenous uncertainty (i.e., probability of reward occurring at two different locations). We have also introduced long and short probe trials with different probabilities.

After a standard training, all mice were subjected to three experimental phases starting with ratio 1:2 (3 s vs. 6 s). In each phase we introduced probe (not rewarded) trials, with different probabilities, to test the endogenous uncertainty according to risk exogenous assessment. During the first experimental phase (Exp. 1) short probe (S_p) and long probe (L_p) trials were introduced with the same conditional probabilities, $P(S_p|S) = P(L_p|L) = 0.2$. In the second experimental phase (Exp. 2) $P(S_p|S) = 0.5$ and $P(L_p|L) = 0.2$, and vice versa for the third one. Furthermore, the analysis of the probe trials along the experimental phases allows to understand and quantify the mice perception of the changes in the exogenous assessment.

The analysis of timing behaviors across different experiments and across 24 hours has shown significant circadian effects on some behavioral measures. Moreover, we present evidence that daily variations in behavioral performance correlate with circadian-sleep processes. In particular, when the pressure of sleep is thought to increase, the behavioral performance decreases.

Comprehensive phenotyping of thousands of mutant mouse strains (i.e., across mouse genetics consortia around the world) will provide an unprecedented volume of data in the future. Current bioinformatic tools, such as *ontologies* [12-15], are designed to provide a repository for multimodal data but are not well-suited for integrated analyses between genome and detailed and dynamic phenome information. In this project we will investigate, by means of our multi-dimensional experiments, novel computational strategies to investigate the link between genotypes and phenotypes by considering new factors, i.e., time-changing variables.

Through the symposium we will discuss the advantages and the current challenges of realizing fully automated systems for behavioral screens in mice.

References

1. Brown, S. (2011). Straight talk with... Steve Brown. Interview by Hannah Waters. *Nat Med* **17**(11), 1332.
2. Brown, S.D., P. Chambon, and M.H. de Angelis, (2005). EMPReSS: standardized phenotype screens for functional annotation of the mouse genome. *Nat Genet*. **37**(11), 1155.
3. Brown, S.D. et al. (2009). The functional annotation of mammalian genomes: the challenge of phenotyping. *Annu Rev Genet* **43**, 305-333.
4. Morgan, H. et al. (2010). EuroPhenome: a repository for high-throughput mouse phenotyping data. *Nucleic Acids Res* **38**(Database issue), D577-85.
5. Mallon, A.M., A. Blake, and J.M. Hancock. (2008). EuroPhenome and EMPReSS: online mouse phenotyping resource. *Nucleic Acids Res* **36**(Database issue), D715-D718.
6. Kas, M.J. and J.M. Van Ree. (2004). Dissecting complex behaviours in the post-genomic era. *Trends Neurosci* **27**(7), 366-369.
7. Flint, J. and M.R. Munafo. (2007). The endophenotype concept in psychiatric genetics. *Psychol Med* **37**(2), 163-80.
8. Munafo, M.R. and J. Flint. (2004). Meta-analysis of genetic association studies. *Trends Genet* **20**(9), 439-44.
9. Schofield, P.N., et al. (2009), Post-publication sharing of data and tools. *Nature* **461**(7261), 171-173.
10. Schofield, P.N. et al. (2010). Research funding. Sustaining the data and bioresource commons. *Science* **330**(6004), 592-593.
11. Balci, F., D. Freestone, and C.R. Gallistel. (2009). Risk assessment in man and mouse. *Proc Natl Acad Sci USA* **106**(7), 2459-2463.
12. Schofield, P.N., R. Hoehndorf, and G.V. Gkoutos. (2012). Mouse genetic and phenotypic resources for human genetics. *Hum Mutat*.
13. Chen, C.K., et al. (2012). MouseFinder: Candidate disease genes from mouse phenotype data. *Hum Mutat*.
14. Wimalaratne, S.M. et al. (2012). An infrastructure for ontology-based information systems in biomedicine: RICORDO case study. *Bioinformatics* **28**(3), 448-450.
15. Schofield, P.N. et al. (2011). New approaches to the representation and analysis of phenotype knowledge in human diseases and their animal models. *Brief Funct Genomics* **10**(5), 258-265.

Comparison of Home-Cage Activity Systems Using Transgenic Mouse Lines and Pharmacological Interventions

L. Robinson, G. Riedel

School of Medical Science, College of Life Science and Medicine, University of Aberdeen, Aberdeen, Scotland, UK.
l.robinson@abdn.ac.uk

Introduction

A number of different Home-Cage activity systems have been developed in order to assess the behaviour of mice. Automated home-cage activity systems enable the establishment of continuous and long-term monitoring of locomotor activity, feeding behaviour and circadian rhythmicity in rodents [1-4]. They allow a handling free assessment of drug and behavioural effects in a longitudinal design.

Aim

The main aim of this present study was to evaluate the performance of transgenic and pharmacologically treated mice in two variants of home-cage activity systems, allowing us to perform a comparison of the behavioural setups.

Methods

Two variants of home-cage activity observation systems were used in this study: 1) the PhenoMaster (TSE Systems, Bad Homburg, Germany) and 2) the PhenoTyper (Noldus IT, Wageningen, The Netherlands). In both tests, animals were singly housed and subjected to a 12 hour light/dark cycle (lights on @ 7am, temperature 23 ± 2°C, relative humidity of 40 – 60 %) with free access to food and water. All experiments being performed in accordance with Home Office regulations. Mice were placed into the cages and given 2 days of habituation before assessment of circadian/ultradian activity over continuous day-night cycles.

1. The PhenoTyper is a video-based observation system for long-term continuous assessment of behavioural activity in mice via a built-in digital infrared sensitive video camera and infrared light sources in a top unit of each cage. The body-centered recording of the animal (EthoVision 3.1) was sampled at 12.2 Hz. Data was stored online and calculated for parameters including: i) total locomotion during day and night cycles; ii) time spent in the food zone (area in front of the feeder) and iii) time spent in the water zone (area in front of the water bottle).
2. The PhenoMaster system consists of a test cage (42 x 26.5 x15 cm) positioned in a metal frame containing regular infrared beams in both X and Z coordinates at a distance of 3 cm and at a rate of 100Hz. These enable the continuous recording of exploratory activity over long time periods. Cage lids were fitted with two weight transducers supporting a feeder and a water bottle allowing an automated measurement of food and water intake. All data was stored online and parameters calculated included: i) daily food consumption; ii) daily water consumption and iii) total distance moved in day and night cycles.

Two different studies were employed: i) Evaluation of the cannabinoid antagonist AM251 as a possible treatment for obesity and ii) Behavioural phenotyping of a mouse model of Rett Syndrome.

i) C57BL6 mice (25-30g) were used in this study to assess the effect of the CB1 antagonist AM251 on body weight and feeding behaviour. Mice were matched for body weight prior to an intraperitoneal injection of either AM251 (10 mg/kg) or vehicle (Tween 80) 1-2 hours before the start of the dark phase of testing. In addition to the parameters measured by the activity cages the body weight of each mouse and weight of food hopper (food intake) and water bottle (water intake) were recorded.

ii) Female mice in which the endogenous Mecp2 gene was silenced by insertion of a targeted *lox stop* cassette ($Mecp2^{Stop}$) were crossed with males hemizygous for the CreESR transgene [5]. For the purpose of this study only heterozygous $Mecp2^{Stop/+}$ females without the cre-ER transgene and wild-type (WT) littermates were used and were aged 10 – 12 months at the start of testing. Mice were assessed in the PhenoMaster and PhenoTyper for a period of 5 days in order to evaluate locomotor activity, circadian activity and food/water intake.

Results

i) Acute treatment with AM251 in both the PhenoMaster and PhenoTyper induced a suppression of food intake and a reduction in body weight compared to vehicle treated mice. AM251 treated animals spent less time in the food zone of the PhenoTyper and displayed a significant reduction in food consumption in the PhenoMaster. In addition to AM251 induced effects on feeding behaviour a reduction in locomotor activity was also evident in the activity cages although this was only apparent for the initial few hours of the dark phase.

ii) $Mecp2^{Stop/+}$ mice displayed a significant reduction in overall locomotor activity compared to WT's in both the PhenoMaster and PhenoTyper. The differences in activity were particularly evident for the nocturnal dark phases of testing with little to no difference apparent during the light phases. In both home cage activity systems the $Mecp2^{Stop/+}$ mice displayed normal circadian rhythm with both genotypes displaying elevated locomotor activity during the dark hours and were less active during the light phase. In addition to differences in locomotion a significant alteration in feeding behaviour was also observed with $Mecp2^{Stop/+}$ mice presenting with an increased body weight compared to WT's, along with enhanced food consumption in the PhenoMaster and time spent in food zone in the PhenoTyper.

Conclusions

These findings suggest that the home cage activity observation systems ' PhenoTyper' and 'PhenoMaster' are sensitive and effective at determining drug induced changes in feeding behaviour and in the validation of the locomotor phenotype evident with a mouse model of Rett syndrome. Replication of behavioural observations in two different recording environments supports the quality of the equipment and the reliability of the phenotype.

References

1. De Visser, L., van den Bos, R., Kuurman, W.W., Kas, M.J.H., Spruijt, B.M. (2005). Novel approach to the behavioural characterization of inbred mice: automated home cage observations. *Genes, Brain and Behav* 1-9.

2. Riedel, G., Fadda, P., McKillop-Smith, S., Pertwee, R.G., Platt, B., Robinson, L. (2009). Synthetic and plant-derived cannabinoid receptor antagonists show hypophagic properties in fasted and non-fasted mice. *Br. J. Pharmacol.* **156**, 1154-1166.

3. Theander-Carrillo, C., Wiedmer, P., Cettour-Rose, P., Nogueiras, R., Perez-Tilve, D., Pfluger, P., Castaneda, T.R., Muzzin, P., Schürmann, A., Szanto, I., Tschöp, M.H. (2006). Rohner-Jeanrenaud F. Ghrelin action in the brain controls adipocyte metabolism. *J Clin Invest.* **116**(7), 1983-1993.

4. Edelsbrunner, M.E., Painsipp, E., Herzog, H., Holzer, P. (2009). Evidence from knockout mice for distinct implications of neuropeptide-Y Y2 and Y4 receptors in the circadian control of locomotion, exploration, water and food intake. *Neuropeptides* **43**(6), 491-497.

5. Guy, J., Gan, J., Selfridge, J., Cobb, S., Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. *Science* **315**, 1143–1147.

A SWOT Analysis on Automating "Measuring Behaviour"

B.M. Spruijt[1], R.C. de Heer[2]

Delta Phenomics, Utrecht, The Netherlands
Department of Biology, Utrecht University, Utrecht, The Netherlands

[1]*berry.spruijt@deltaphenomics.com*, [2]*Raymond.deHeer@deltaphenomics.com*

Introduction

In the past decades technological innovations in many scientific fields, such as molecular biology, biochemistry electrophysiology and histochemistry have facilitated enormously the degree of resolution of independent variables, in addition to accelerating the production of results and read-out parameters. Consequently, this has resulted in new scientific insights. Similarly, a lot of technological innovations have been introduced in studies on (animal) behaviour. The question emerges, however, whether behavioural science has benefited accordingly in terms of generating new parameters, results and insights. We will address this question by applying a SWOT-analysis to discuss the strengths, weaknesses, opportunities and threats of technological innovations that have culminated in the development of automated (high-throughput) home cage systems.

Strengths

Several software packages that enable automating behavioural experiments are commercially available. Among the largest manufacturers are TSE, CleverSys, ANY-maze, Biobserve and Noldus IT. These manufacturers have in common that they are involved in the development of both software and hardware to automate behavioural experiments. Most of them provide their own high throughput home cage system, such as e.g. PhenoMaster (TSE), HomeCage Plus (Biobserve) and PhenoTyper (Noldus IT). Although these systems and/or software each have their own advantages and disadvantages, they have several strong points in common:

- They facilitate the automation of almost all so-called standard behavioural tests, like Open Field, Elevated Plus Maze, Radial Maze, Morris Water Maze, etc.

- The time needed to carry out an experiment is largely reduced.

- They offer more accurate measurements compared to human observations, since a computer is not that easily distracted and/or fatigued.

- They yield a higher reproducibility, because a computer will "observe" the same stream of events in an identical manner.

When it comes to the home cage systems, another strong point of automating behavioural measurements is that the factor time can really be involved, since a computerized system is able to observe 24 hours a day, 7 days a week. Another advantage is that with automated home cage observations the handling and transport of animals from the home cage to the test apparatus becomes redundant, which results in a smaller influence of variation caused by human intervention.

Weaknesses

The focus on automating behavioural experiments is mainly on increasing efficacy and therefore saving time. Hardly any new parameters have been developed. Furthermore, little effort has been made to prolong the time an animal is observed. An automated elevated plus maze test is still run for only 10 minutes rather than measuring for a longer period. Other weaknesses we would like to point out are:

- Instead of observing the animals directly, the experimenter is transformed into a user of a computer application. He or she has to rely on numbers without exactly knowing how they were produced.

- Increased dependence on computer hard- and software. If the computer and/or software malfunction, more and more users will not know how to carry out the experiment.

- The data obtained with a computer can still be “noisy” data that needs to be filtered. Large mistakes, like a disturbance of the experimental setting, will probably be detected, but many smaller mistakes are most likely not. Even a computer can make mistakes by tracking/following something else than the animal.

- Manufacturers of soft- and hardware tend to predetermine how a certain experiment should be carried out, whereas it could be the other way around.

Opportunities

By taking innovation a step further, it will be possible to test animals using a much more ethological approach. For example, by combining transponder technology with video-tracking it will be possible to test a single animal even when it is group-housed, while at the same time determine what other animals in the group are doing at that particular time point. Other opportunities are:

- Enlarging the cage (instead of reducing it) to allow for an integrative measurement of behaviour, that can easily be acquired using dedicated computer software.

- Making use of cleverly designed analytical tools, like the ones created by Golani and colleagues (Golani, Benjamini et al. 1993; Drai, Benjamini et al. 2000; Drai, Kafkafi et al. 2001).

- Collecting vast amounts of data, that facilitates the development of new variables and endpoints.

Threats

The long history of classical methods has yielded an intuitive consensus, be it only superficial, that it is not necessary to describe the used ethograms in detail and that just a simple reference to a previous paper is sufficient. Everybody assumes that behaviours such as freezing, immobility etc. seen in an open field are similar to what one is accustomed to see. This is only based on a kind of historically developed intuition. But when automated methods are introduced, scientists demand validation due to the absence of this intuition as there is no *a priory* conception of computer-used algorithms for a specific behaviour.

The caveat of a more comprehensive approach and a wider spectrum of behavioural elements over a longer period of time is that it yields a complex set of data, especially when technical devices are used which collect data at a high frequency rate. The numbers representing movements, velocity, contour of animals etc. do not have a direct familiar representation in terms of behaviours as one has been familiarized with through the use classical studies. Thus, the data set is at first sight without any meaning. This may be business as usually in other disciplines where math is more generally used and accepted. This is cumbersome however for behavioural scientists who rely on their own way of observing animal behaviour.

Moreover, the analysis requires special tools for exploring the data. When the study is descriptive in nature, special data mining tools are required. When the study is hypothesis driven, the parameters *a priori* defined and those which appeared to have changed without an *a priori* prediction need to be tested. Since the number of parameters may be large, the appropriate analysis and interpretation can be challenging and difficult.

The problem is that behaviour, when observed by humans, elicits associations in terms of meaning or differences. Without human observations, scepticism and doubt emerge about the meaning of abstract numbers. In that case, one is tempted to fall back on familiar and easier behavioural tasks.

So far, technological innovations have given us the opportunity to increase the efficiency by doing more experiments in the same time. However, they have not yielded an increase in the number of variables/endpoints when the data is analysed. If the scientific community can make a distinction between the validation of a paradigm and technological validation, data analysis could benefit from innovation as well. Here, we will present data to illustrate this.

References

1. Drai, D., Benjamini, Y., *et al*. (2000). Statistical discrimination of natural modes of motion in rat exploratory behavior. *Journal of Neuroscience Methods* **96**(2), 119-131.

2. Drai, D., Kafkafi, N. *et al*. (2001). Rats and mice share common ethologically relevant parameters of exploratory behavior. *Behavioural Brain Research* **125**(1-2), 133-140.

3. Golani, I., Benjamini, Y. *et al*. (1993). Stopping behavior: constraints on exploration in rats (Rattus norvegicus). *Behavioural Brain Research* **53**(1-2), 21-33.

Mouse Phenotyping in the IntelliCage: From Spontaneous Behavior to Cognitive Function

D.P. Wolfer[1,2], V. Voikar[3], E. Vannoni[1], G. Colacicco[1], H.P. Lipp[1]

[1]*Institute of Anatomy, University of Zurich, Zurich, Switzerland. dpwolfer@anatom.uzh.ch*

[2]*Institute of Human Movement Sciences, ETH Zurich, Zurich, Switzerland. dwolfer@ethz.ch*

[3]*Neuroscience Center, University of Helsinki, Helsinki, Finland. vootele.voikar@helsinki.fi*

Introduction

Increasingly sophisticated and efficient methods of mutagenesis produce an exploding number of genetically modified mouse lines, many of which are inducible or pharmacologically controllable. Additional models are produced using viral vectors and RNA interference techniques. Because many of these mouse models address questions related to nervous system function and disease, there is a large demand for comprehensive behavioral phenotyping of genetically modified mouse lines. During the past 25 years our laboratory has been involved in the behavioral phenotyping of >11,000 mice and >130 different mutations. Traditional behavioral tests for mice are inefficient, labor intensive, and often difficult to interpret because they were originally developed and validated for rats. Traditional tests typically involve social isolation, sensory deprivation, exposure to unfamiliar apparatus, and repeated handling by humans. The resulting stress responses introduce artifacts and reduce test reliability. A new approach with the potential to eliminate most of these problems is to automate behavioral experiments and to run them in the home cage of the animals. Home cage behavior can be monitored continuously over extended periods of time, whereas observation times of traditional tests are typically short, often too short. Several systems permitting to test mouse behavior in the home cage have appeared on the market in recent years. The IntelliCage (IC), used by my lab since 5 years, is peculiar in the sense that it permits to phenotype mice in a social home cage setting. As shown in Figure 1, IC is a large home cage designed for 10-16 mice with four computer-controlled learning corners that provide space for one mouse at a time and contain two drinking bottles each. Water access can be blocked by motorized doors that open in response to nosepokes. Because the mice carry RFID tags, all corner visits, nosepokes and licking events can be attributed to individual mice. Each corner can present light stimuli, deliver air puffs as punishment, and is fully programmable as an operant conditioning chamber.

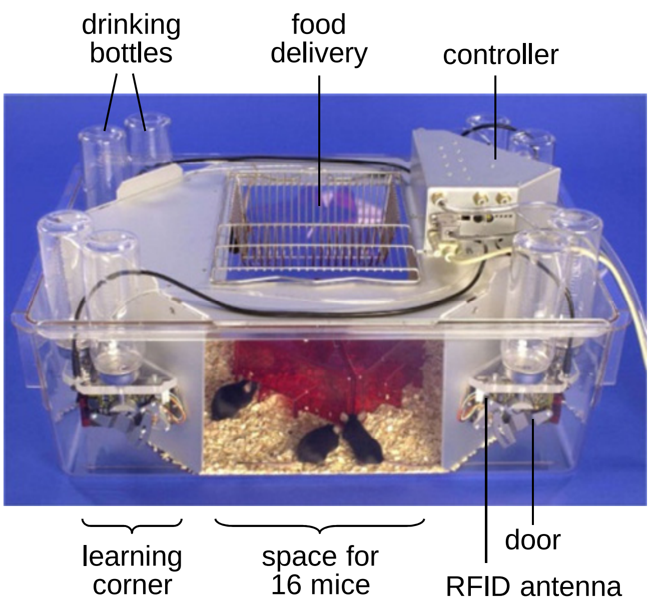


Figure 1. The IntelliCage apparatus.

Spontaneous behavior

While many specialized protocols have been developed for IntelliCage to test learning and memory, attention, impulsivity and emotional responses (see below), all mice begin testing with about one week of *free adaptation*, during which all water bottles are freely accessible at all times. Spontaneous corner visits, nosepoking patterns and licking activity are monitored 24/7 during this phase. Since the IC system was last presented at the Measuring Behavior conference in 2008, we have collected data on 50 behavioral parameters of >800 mice. These parameters include duration, content, spatial distribution and temporal patterning of visits, nosepokes, as wells as licking episodes. Subsequent factor analysis of these parameters extracted 12 orthogonal factors accounting for 81% of total variance. Comparison of factor scores of C57BL/6, DBA/2, BALB/c and 129S2 mice revealed a unique profile for each strain. Analysis of mice with hippocampal, prefrontal and striatal lesions also yielded unique profiles for each condition. Monitoring of mutant mice with known deficits in hippocampus-dependent tests produced profiles very similar to those of hippocampal lesions [1, 4]. On the other hand, this analysis revealed strikingly similar changes of diurnal activity distribution in the 5xFAD mouse which over-expresses β-amyloid precursor protein (βAPP) and a βAPP hypomorphic mutant [6]. Thus, already the monitoring of spontaneous behavior during a week of free adaptation in IC permits high throughput prescreening of mutant mice. On the other hand, our data indicate that tight control of genetic background remains essential also if behavioral testing occurs in the home cage. In order to familiarize the mice with the operant properties of IC used in most learning tasks, free adaptation is followed by *nosepoke adaptation*, during which doors are closed but can be opened at any times once per corner visits by performing a nosepoke. This is followed by a *drinking session adaptation* protocol in which the mice learn to adapt their activity to a regime of scheduled drinking sessions. This form of learning seems highly and specifically sensitive to hippocampal lesions.

Cognitive function

In order to asses spatial learning and memory, we have developed a suite of tasks that are based on a drinking session protocol but make the additional restriction, that for each individual mouse only one of the four corners delivers water reward in response to nosepokes. The rule which the mice must learn in order to predict where water will be available next changes between tasks and becomes increasingly difficult. The simplest implementation is a *corner preference* task in which each mouse can obtain water in the same corner during a week. So far, this task was learned successfully by all tested mouse models, indicating that this is a simple control procedure (sort of an IC equivalent of the cue navigation protocol in the water-maze) useful to verify that the mice possess the sensory and motor abilities needed to perform the other tasks in this suite. Next, the mice proceed to a *reversal learning* phase in which the reward location is moved to the opposite corner, and finally to a *serial reversal protocol*, in which the location of reward changes between drinking sessions in a random fashion. A multi-lab study has demonstrated that these protocols deliver consistent results independently of lab environment [3]. Finally, the mice learn two protocols in which the location of reward changes between individual visits. In the *chaining task*, the animals must learn to visit corners in a clock- or anti-clock-wise sequence. The *patrolling task*, is similar but more difficult to learn because the reference that determines location of reward is not the most recently visited corner, but the one in which reward was last obtained. In both tasks, acquisition of the rule is followed by a reversal phase in which the direction is reversed. Patrolling reversal appears to be extremely difficult for mice to learn, even intact C57BL/6 learn extremely slowly. Mice take about 2 months to proceed through all the tasks of this suite. In our validation studies, we have found that mice with bilateral complete hippocampal (but not prefrontal or dorso-lateral striatal) lesions are strongly impaired in reversal, serial reversal as well in as chaining, patrolling and other forms of sequence learning. Since the last presentation of IC at the Measuring Behavior conference in 2008, this suite of tasks has revealed strong impairments in mutant mouse models with deficient function or connectivity of the hippocampus, for example in mice with a null mutation of alphaPix/Arhgef6, a gene associated with X-linked intellectual disability [4], as well as in a mouse model of Alzheimer's disease [1]. Massive deficits were also revealed in a βAPP hypomorphic mutant due to motor deficits cannot be tested in many of the conventional spatial learning tasks [6]. The suite of spatial tasks is complemented by a *corner avoidance* protocol in which the animals learn to avoid one of the four corners in which nosepokes are punished by air puffs during a training period of 24h hours. In order to test long-term memory, the animals can be reintroduced into IC for a probe trial after varying retention intervals outside IC. If left for several days in the IC, the mice show *extinction learning*, that is gradually resume drinking in the no longer punished corner. Performance in this test is sensitive to hippocampal manipulations as well [1, 5]. Together with other tests, this paradigm has revealed that mice lacking Ras-GRF1 are selectively impaired in fear conditioning but have intact spatial memory [2]. More recent developments whose validation is still in progress include an IC version of the *conditioned taste aversion* and *social transmission of food preferences* tests.

Adaptive decision making and inhibitory response control are important aspects of cognition that are often neglected in mouse phenotyping because their testing in conventional operant setups is extremely time consuming and resource intensive. Therefore, we have recently developed a simple *reaction time task* for IntelliCage. At any time and in any corner a mouse can start a trial by entering the corner and making a first nosepoke. After a variable delay, stimulus lights will go on and the mouse is allowed to drink at the door where the first poke was made. Principal readouts are the number of premature responses as a measure of motor impulsivity and the latency of correct responses (reaction time). In a strain comparison this test recapitulated the results obtained in conventional operant reaction time tasks, such as the five-choice serial reaction time task. Mice with lesions of the medial habenula show increased impulsivity in this IC task as well as in the traditional 5-choice serial reaction time task. In order to assess how well mice can control their response to reward, we have also developed a *delay discounting task* for IC. In all corners the mice are given a choice between a bottle with plain water and one with 0.5% saccharin solution. In a training phase, the mice are allowed to learn about the existence of this choice and to establish their (usually exclusive or almost exclusive) preference for the saccharin solution. In the following test phase, access to saccharin is increasingly delayed at a rate of 0.5s per day. As a consequence of this imposed delay, the mice will eventually abandon their preference for the saccharin solution. Reduced delay tolerance in this test can provide evidence for increased cognitive impulsivity, whereas excessive delay tolerance in conjunction with saccharin overconsumption and compulsive nosepoking at the saccharin door are indicators of a poorly controlled response to reward. The latter observation was made when we tested mice with lesions of the medial habenula in this test, as well as in a transgenic mouse line over-expressing erythropoietin selectively in the CNS.

Thus, the fully automated IC system permits efficient prescreening of socially housed mutant mice by recording profiles of spontaneous individual behavior, as well as assessment of specific aspects of cognitive function using specialized operant tasks.

Acknowledgements

This research has been supported by: University of Zurich, ETH Zurich, Zurich Center for Integrative Human Physiology (ZIHP), Swiss NF NCCR Neural Plasticity and Repair, EC FP6 INTELLIMAZE, EC FP7 EUROSPIN. Procedures with animals were approved and their use monitored by the Veterinary Office of the Canton of Zurich, Switzerland. DPW and HPL are members of the Zurich Center for Neuroscience (ZNZ).

References

1. Codita, A., Gumucio, A., Lannfelt, L., Gellerfors, P., Winblad, B., Mohammed, A.H., Nilsson, L.N. (2010) Impaired behavior of female tg-ArcSwe APP mice in the IntelliCage: A longitudinal study. *Behav. Brain Res.* **215**, 83-94.

2. d'Isa, R., Clapcote, S.J., Voikar, V., Wolfer, D.P., Giese, K.P., Brambilla, R., Fasano, S. (2011) Mice lacking Ras-GRF1 show contextual fear conditioning but not spatial memory impairments: convergent evidence from two independently generated mouse mutant lines. *Front. Behav. Neurosci.* **5**, 78.

3. Krackow, S., Vannoni, E., Codita, A., Mohammed, A.H., Cirulli, F., Branchi, I., Alleva, E., Reichelt, A., Willuweit, A., Voikar, V., Colacicco, G., Wolfer, D.P., Buschmann, J.U., Safi, K., Lipp, H.P. (2010) Consistent behavioral phenotype differences between inbred mouse strains in the IntelliCage. *Genes Brain Behav.* **9**, 722-731.

4. Ramakers, G.J., Wolfer, D., Rosenberger, G., Kuchenbecker, K., Kreienkamp, H.J., Prange-Kiel, J., Rune, G., Richter, K., Langnaese, K., Masneuf, S., Bosl, M.R., Fischer, K.D., Krugers, H.J., Lipp, H.P., van Galen, E., Kutsche, K. (2012) Dysregulation of Rho GTPases in the alphaPix/Arhgef6 mouse model of X-linked intellectual disability is paralleled by impaired structural and synaptic plasticity and cognitive deficits. *Hum. Mol. Genet.* **21**, 268-286.

5. Voikar, V., Colacicco, G., Gruber, O., Vannoni, E., Lipp, H.P., Wolfer, D.P. (2010) Conditioned response suppression in the IntelliCage: assessment of mouse strain differences and effects of hippocampal and striatal lesions on acquisition and retention of memory. *Behav. Brain Res.* **213**, 304-312.

6. Weyer, S.W., Klevanski, M., Delekate, A., Voikar, V., Aydin, D., Hick, M., Filippov, M., Drost, N., Schaller, K.L., Saar, M., Vogt, M.A., Gass, P., Samanta, A., Jaschke, A., Korte, M., Wolfer, D.P., Caldwell, J.H., Muller, U.C. (2011) APP and APLP2 are essential at PNS and CNS synapses for transmission, spatial learning and LTP. *EMBO J* **30**, 2266-2280.

Combining Classical and Automated Neurophenotyping in Mice and Rats

S. von Hörsten, A.-C. Plank, Y. Urbach, F. Canneva

Dept. Experimental Therapy, Franz-Penzoldt-Center, Universitätsklinikum Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany. Stephan.v.hoersten@ze.uni-erlangen.de

Introduction

Translational research requires the establishment of comprehensive neurobehavioral screening systems, dedicated to fill the gap between post genomic generation of state-of-the-art animal models (i.e. transgenic rodents) on the one hand and their added value for really predictive experimental preclinical therapy on the other hand. Due to these developments in the field, neuroscientists are frequently challenged by the task of detecting discrete behavioral differences in rodents. Systematic, comprehensive phenotyping covers these needs and represents a central part of the process. Here we provide an overview on theoretical issues related to classical comprehensive neurobehavioral phenotyping and propose specific procedures as well as concepts for integrating automated, intra-home-cage technologies. The latter allows continuous screening of various behavioral and physiological dimensions on an ethological basis, which represents important added value to the comprehensiveness of the screening approach.

Approach and Methods

Comprehensive phenotyping of rodents is a process of defining characteristics during their ontogeny. During the past 20-30 years rather standardized approaches evolved using combinations of classical behavioral assays spanning all different behavioral domains. Specific guidelines for the appropriate conduction of such classical phenotyping work have been proposed. However, more recently, due to the consideration of certain short-comings in classical approaches (non-ethological based, stress-confounded, non-repeatable under the same test-construct), intra-home-cage automated phenotyping technology was developed, partly validated, and is now up to be integrated/associated with comprehensive classical phenotyping approaches.

Here we will connect the present status of guide-lines for classical neurobehavioral phenotyping and focus on open questions and issues brought up by integrating automated phenotyping into large comprehensive screens. Among other reasons, most of those problems are derived from the novel quality of multidimensional parallel acquisition of behavioral outputs within the automated systems. These issues span areas such as data mining, analysis of multiple parallel measures in longitudinal intraindividual experimental designs, provide new chances for the identification and interpretation of novel complex multidimensional combined output variables and – last but not least – require feasibility and validation of cognitive testing within home-cage-environment on an ethological based non-touch testing scenario.

To deal with these presumptions, transgenic Huntington's disease (tgHD) rats were repeatedly screened within modular, intra-home-cage phenotyping systems measuring spontaneous activity, feeding, temperature, metabolic performance, cognition by operant procedures, social, and emotional parameters. This investigator-independent (ethological) approach was further validated and compared with classical behavioral assays (social interaction test, prepulse-inhibition, accelerod, two-way active avoidance) in the same animals. Apart from outlier detection, multivariate analysis was used to explore correlation patterns of variables in each genotype as well as for finding similarities and dissimilarities between genotypes and the key variables that discriminate between genotypes. Along with circadian changes in energy metabolism, automated phenotyping revealed higher specific motor activities in tgHD rats, with spontaneous free rearing correlating with individual performance in the accelerod test. Principle component analysis revealed a separation by genotype in juvenile tgHD rats that differed from adult animals, being further resolved by Partial Least Squares Discriminant Analysis detecting "temperature" (juvenile) and "rearing" (adult) as phenotypic key variables in the tgHD model. These studies illustrate that automated-intra-home-cage-phenotyping in combination with multivariate analysis is capable of characterizing a complex phenotype by detecting novel physiobehavioral markers at similar sensitivity and presumably better standardization using fewer resources.

Conclusion

We propose a broadening of the guidelines for comprehensive phenotyping in order to cope with issues derived from automated phenotyping and provide proof-of-concept for operant learning paradigms with the home-cage using intra-home-cage automated technology. Multivariate statistics were successfully applied to identify components with the parallel multivariate data sets that correlate with classical behavioral responses.

SPECIAL SESSION

Measuring Engagement: Affective and Social Cues in Interactive Media

Chair: Anton Nijholt[1], Alessandro Vinciarelli[2]

[1]*Human Media Interaction, University of Twente, Enschede, The Netherlands. anijholt@cs.utwente.nl*

[2]*Department of Computing Science, University of Glasgow, Scotland, UK. vincia@dcs.gla.ac.uk*

Introduction

In game or entertainment environments the 'user' may take part in events that require bodily interaction with sensor-equipped environments. Embedded motion-capture and gyroscopic devices capture movements. Thanks to Nintendo's remote WII, motion-controlled games are now widespread. Cameras, microphones, pressure sensors, and proximity sensors have been added. Thanks to Microsoft's Xbox Kinect whole body interaction games have become popular. Apart from entertainment, such sensor equipped and game oriented environments can be designed to improve health conditions, sports performance, or (therapeutic) physical rehabilitation.

More sensors are becoming available, allowing a game to obtain more information about a player, in particular his or her bodily and emotional conditions. In addition, these sensors allow more input modalities for interaction with an environment [1]. There are many examples of advanced games where posture, gestures, body movements, facial expressions and brain activity are among the modalities that are used for control. Control can be direct, but it can also be mediated (for example through a balance board, a tangible or a wearable). Observations of the face and body can be used in different forms, depending on whether the user may take the initiative to control the interaction or whether the application takes the initiative to adapt itself to the user. Therefore we can distinguish:

- Control: the user consciously produces facial expressions, head movements or body gestures to control a game. For example, to navigate, move game characters, or change appearances.
- Adaptation: the gamer's spontaneous facial expressions and body poses are interpreted and used to adapt the game. For example, this information may lead to changes in the difficulty level, the appearance of the game environment, the interaction modalities, or the narrative.

The game environment is not necessarily at home. It can be a training and simulation field or a less controlled setting such as an urban area, where bodily movements and activities are detected by mobile devices and webcam networks.

Engagement

One of the key issues in (video) games and in interactive entertainment and art is engagement. Research questions are: how can we design and predict engagement? How can we adapt a game to its users and audience to increase or decrease engagement? Again, when we talk about a 'game', we are talking about environments that provide entertaining interaction opportunities to a 'user'. Automatically obtained engagement information allows adaptation of a game or game environment by the game controller. That may include physical and virtual appearance, availability and characteristics of interaction modalities, feedback by actuators, game narrative, and game strategy. It should be mentioned that above we used the term 'engagement' in an informal way. Numerous terms and definitions have been introduced to describe aspects or features of engagement, or the other way around, to define engagement as an aspect of user experience concepts such as presence, levels of immersion or flow [2].

The questions mentioned above need to be investigated. Knowledge about factors that affect engagement can be used in game design and game evaluation. Although there is theorizing around concepts such as flow, presence, and immersion, there is a considerable gap with applied knowledge coming from design, performing arts, narrative and drama theory. Let alone the ‘gut feeling’ of a game designer who has played games for thousands of hours himself and has access to the experiences of hundreds of game testers that have spent hundreds of hours playing a particular game before it gets its final form and is made publicly and commercially available.

Nevertheless, detecting different levels of engagement during game playing can make it possible to make decisions about increasing or decreasing a level of difficulty, to provide in-game explanations, to adapt the storyline, or to adapt the behavior of game characters. It should be noted that results on engagement modeling in sensor-equipped game, entertainment and interactive art environments do not necessarily translate to other applications where, for instance, task efficiency is an important issue. In the game and entertainment situations we address, immersion, presence, and flow need to be considered when investigating user experience, user satisfaction, and user enjoyment. As mentioned by Berthouze et al. [3], there are lots of theories that attempt to define and model aspects of engagement but, “most theories of engagement have focused purely on its mental aspects”.

Special Session “Measuring Engagement”

The aim of this special session at Measuring Behavior 2012 is to look at engagement and ways to measure engagement in situations where users are not glued to their chair and keyboard, that is, in sensor-equipped environments that are able to perceive nonverbal interaction behavior. And, moreover, we focus on activities that are not necessarily aimed at performing a particular task in the most efficient way. Rather the focus is on enjoyment during performance and satisfaction during and after finishing the interaction.

Real-time information providing sensors allow us to detect cues from the behavior of one or more participants in the environment from which different levels of engagement during performance can be concluded. And, also, different kinds of feedback and adaptation of the game environment can be decided. Social and affective behavioral cues [4], to be detected from nonverbal behavior, need to be recognized and interpreted in order to be used as input to our algorithms that provide us with information about engagement, immersion, and flow. All these issues will be covered by the six contributions for this special session at Measuring Behavior 2012.

Acknowledgements

This research has been supported by the European Community’s 7th Framework Programme (FP7/2007–2013) under grant agreement no. 231287 (SSPNet) and the CaSA (Computers as Social Actors) project of the EIT ICT Labs (EIT-ICT-12124).

SPECIAL SESSION CONTENTS (sorted by paper ID)

Video-Based Multi-person Human Motion Capturing

Nico van der Aa, Lucas Noldus (Noldus Information Technology, The Netherlands) and Remco Veltkamp (Utrecht University, The Netherlands)

Making Ambient Spaces into Playgrounds

Dennis Reidsma, Daniel Tetteroo and Anton Nijholt (Human Media Interaction / Creative Technology, University of Twente, The Netherlands)

Building Corpora of Bodily Expressions of Affect

Marco Pasch and Monica Landoni (University of Lugano, Switzerland)

Measuring Fun and Enjoyment of Children in a Museum: Evaluating the Smileyometer

Frans van der Sluis, Betsy van Dijk and Bert Perloy (University of Twente, The Netherlands)

On Making Engagement Tangible

Egon L. van den Broek (TNO; University of Twente, Radboud University Medical Center Nijmegen, The Netherlands)

What Can Body Movement Tell Us About Players' Engagement?

Nadia Bianchi-Berthouze (University College London, UK)

References

1. Pantic, M., Nijholt, A. Pentland, A., Huang, T. (2008). Human-Centred Intelligent Human-Computer Interaction (HCI²): How far are we from attaining it? *J. Autonomous and Adaptive Communications Systems* **1**(2), 168-187.

2. Nijholt, A., Pasch, M., van Dijk, E.M.A.G., Reidsma, D., Heylen, D.K.J. (2011). Observations on Experience and Flow in Movement-Based Interaction. Chapter 9 in: *Whole Body Interaction*. D. England (Ed.), Human–Computer Interaction Series. Springer-Verlag, London, 101-119.

3. Bianchi-Berthouze, N., Kim, W.W., Darshak, P.I. (2007). Does body movement engage you more in digital game play? And Why? *Affective Computing and Intelligent Interaction (ACII 2007)*, LNCS 4738, Lisbon, 102-113.

4. Vinciarelli, A., Pantic, M., Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. *Image and Vision Computing* **27**(12), 1743-1759.

Video-Based Multi-person Human Motion Capturing

Nico van der Aa[1], Lucas Noldus[1], Remco Veltkamp[2]

[1]*Noldus Information Technology, Wageningen, The Netherlands. n.vanderaa@noldus.nl; l.noldus.noldus.nl*

[2]*Utrecht University, Utrecht, The Netherlands. R.C.Veltkamp@uu.nl*

Abstract

Observing people using standard cameras provides a non-intrusive way to capture human motion for further study of human behavior in a scene. The system presented uses multiple cameras to detect subjects, track them over time and gives 3D insights in how a subject moves around. These cues are important for the interpretation of human behavior in response to its environment, to interaction with other people or with interactive systems.

Introduction

To capture people's behavior in specific studies is complex. To illustrate, in the Restaurant of the Future 1, researchers want to study people's behavior with respect to food selection and food consumption. Knowing how people choose their food, how other people influence this choice, and if they like what they eat are some of the research questions. The Restaurant of the Future is a field lab equipped with 23 pan-zoom-tilt cameras to ensure non-intrusive observations for food selection and consumer behavior studies. Most studies use The Observer® XT 2: an event logging software for the collection, analysis, and presentation of observational data. Because of the diversity in study setups, full automation is nearly impossible. Therefore, a software module is created based on computer vision algorithms called the Video Analysis and Recognition Toolbox (VidART) to help researchers to focus on the important parts. The tedious manual annotations from video streams are partly reduced by using vision-based algorithms to detect and track people from video streams. To illustrate this, imagine you want to know how often your product is chosen. By tracking the subjects' position automatically and knowing your product's position, the system can select only those frames for further analysis, where people actually approach your product.

Video Analysis and Recognition Toolbox (VidART)

Video analysis in the Restaurant of the Future is challenging due to complex lighting conditions (big windows, different illumination possibilities, large shadows, etc.), the unconstrained subject's appearance (wide clothes, reflective jewelry, hair fashion, etc.) and the presence of static objects in the scene. VidART consists of four modules: (1) background subtraction and shadow detection; (2) contour tracking; (3) voxel reconstruction; and (4) people position tracking.

Background subtraction techniques, like the Mixture of Gaussians 3, are used to detect people or objects in a camera view. The key idea is that any object present in the current frame, but absent in the background model, is identified as foreground. This background model is computed from one or more previous frames captured by the same camera. A subject's silhouette provides essential information about its shape, and other features are easily derived from it like its appearance. It forms the basis for further analysis like tracking and pose estimation. Although background subtraction gives these silhouettes, it requires static cameras and it is sensitive to illumination changes and shadows. Shadows can be detected by assuming that shadows do not alter the texture of the object or the color chromaticity. However, shadows will also be present inside the silhouettes, since the light source's position is in general not the same as that of the camera. Removing such shadows will cause undesirable holes in the silhouette, leading to severe problems in the voxel reconstruction.

An alternative to image segmentation using background subtraction can be found in segmentation based on evolving a so-called level set function. Two appearance models are used: one for the background appearance and one for the foreground appearance. These models, each consisting of a histogram computed from an initial segmentation, are used to drive the segmentation towards the boundary of the desired object. It has the same goal as background subtraction, but it removes the disadvantages. Unfortunately, it is very slow in general. Since we are using video streams, the segmentation result from the previous frame can be used to speed up the process. Our level set tracker based on 4 goes even one step further. It distinguishes rigid body motion from deformations. The rigid body tracking assumes that the shape does not change and locates the position of this rigid body in the next frame. This is represented by a warp of a bounding box, keeping the segmentation inside fixed. The deformation phase finds an optimal segmentation in the bounding box by using the level set segmentation. Although computationally expensive, it is reduced to a small region-of-interest, making the overall method fast. Because of the distinction between rigid body movement and shape deformations some additional correction might be necessary, since deformations can also include rigid body movements. Therefore, an additional drift correction is computed to keep the segmentation in the center of the bounding box. Figure 1 shows the steps of the level set based tracker applied to track a person's head.

Once the silhouettes are obtained for each camera view, we can project the 2D foreground silhouettes to the 3D world. The technique used is voxel reconstruction 5, which keeps a registration of which pixel in each camera view belongs to which voxel (the 3D variant of a pixel) in the 3D world. These pixel-to-voxel correspondences can only be obtained if the cameras are calibrated. In other words, we have to know how a 3D point is captured by the camera and projected to the 2D image. A common way to have the cameras calibrated is to use a calibration object, like a checkerboard, of which the size and dimensions are known 6. Voxel reconstruction labels a voxel as "visible", if all corresponding pixels in the camera views are indicating a foreground object. As a result, a 3D reconstruction of the person's or object's silhouette is available. In Figure 2 an example is given.

To track a person, the silhouettes from either background subtraction or level set based tracking are used to estimate the position of each person in the scene. Initially, a histogram is created for each person to capture his/her appearance. In 7 the concept of vertical reference lines is explained, which are the 2D correspondences of selected 3D lines perpendicular to the ground floor projected on the image planes. For each vertical reference line, the pixel count is registered for each person. For each of the pixels on this vertical reference line that are classified as being foreground pixels, the probability is computed that this pixel belongs to person *k*. The pixel counts for the person with the highest probability. The vertical reference line with the highest pixel count for person *k* is taken as his/her position in the 2D view. Once these lines are found in all cameras views, they are projected back to 3D. To handle occlusions, like when a person is in front of another person in a certain view, the concept of Best Visibility View is used 7. This concept uses the principle that the back-projection of the lines only is done using those cameras that see the person best. Figure 3 shows an example where each person is best visible in two cameras. This makes the tracking of people more robust.

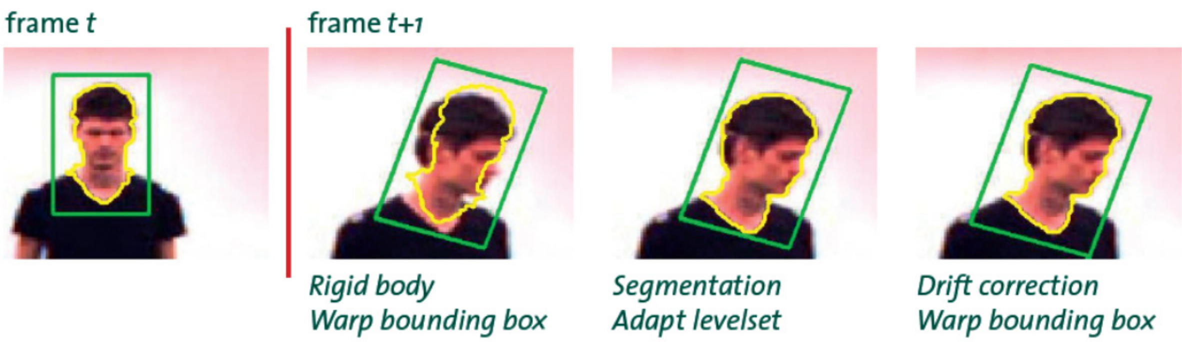


Figure 1. Principle of level set based contour tracking: the bounding box (green) including the initial segmentation (yellow) are given in time step *t*. In the next time step *t*+1, the rigid body tracking, the re-segmentation and the drift correction step are given.

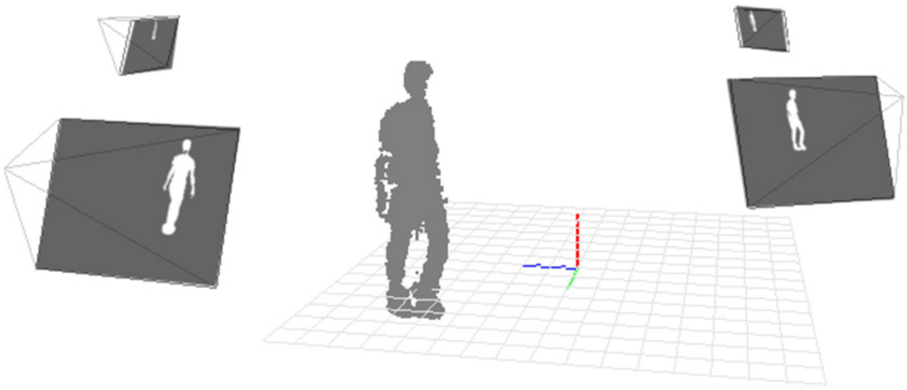


Figure 2. Example of how foreground detection of multiple cameras are combined to obtain a voxel reconstruction of the foreground objects.

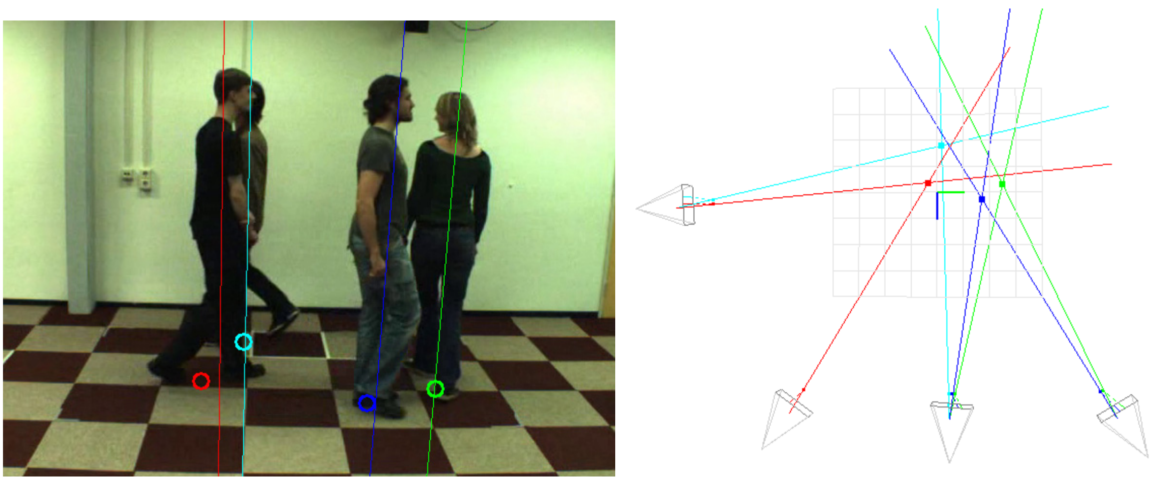


Figure 3. Principle of people tracking with vertical reference lines (*left*) and a top view of the ground floor position using the best visibility view (*right*).

Validation

The software modules have to be tested on accuracy and speed. Since the Restaurant of the Future is a challenging environment, we created the Utrecht Multi-Person Motion (UMPM) benchmark 8 to provide a benchmark for articulated human motion capturing for multi-person motion and interaction in a more controlled environment. UMPM exists of scenarios including human-human and human-object interaction, where each scenario is recorded with up to 4 subjects in the scene. For each scenario, UMPM includes synchronized video sequences for the four cameras used with 644 x 484 resolution at 50 fps. The scenarios include subjects that (1) walk, jog and run in an arbitrary way among each other, (2) walk along a circle or triangle of a predetermined size, (3) walk around while one of them sits or hangs on a chair, (4) sit, lie, hang or stand on a table or walk around it, and (5) grab objects from a table. These scenarios include individual actions, but the number of subjects moving around in the restricted area cause inter-person occlusions. We also include two scenarios with interaction between the subjects: (6) a conversation with natural gestures and (7) the subjects throw or pass a ball to each other while walking around. Since the UMPM benchmark is created to capture human motion, 3D ground truth information is provided, knowing marker positions captured by a motion capture system. The UMPM benchmark including the videos, background images, calibration data and ground truth 3D information, is available for research purposes on www.projects.science.uu.nl/umpm/. In the figures shown in this paper the UMPM benchmark is used to illustrate the principles of the toolbox.

Conclusions and future work

The VidART software ensures people tracking in 3D from video streams, which will be an important tool for automatic detection of people’s behavior. Although it is currently tested on the UMPM benchmark, which is a controlled indoor environment, the toolbox will be extended to be applicable in more advanced sceneries like the Restaurant of the Future in the near future.

Acknowledgement

This research has been supported by the GATE project, funded by the Netherlands Organization for Scientific Research (NWO).

References

1. Wageningen UR (2011). Restaurant of the Future, <http://www.restaurantvandetoekomst.wur.nl/UK/>, Accessed 27 March 2012.

2. The Observer XT software, < http://www.noldus.com/observer>, Accessed 27 March 2012.

3. Stauffer, C., Grimson, W.E.L. (1999). Adaptive background mixture models for real-time tracking, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 1999*, 246-252.

4. Bibby, C. Reid, I. (2008). Robust Real-Time Visual Tracking using Pixel-Wise Posteriors, *Proceedings of the 10th European Conference on Computer Vision, Marseille, France, October 2008,* Part II, 831-844.

5. Kehl, R., Bray, M., Gool, L. van (2005). Full body tracking from multiple views using stochastic sampling, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition 2005*, 129-136.

6. Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(11): 1330-1334.

7. Luo, X., Tan, R.T., Veltkamp, R.C. (2011). Multi-person Tracking Based on Vertical Reference Lines and Dynamic Visibility Analysis. *IEEE International Conference on Image Processing (ICIP 2011).*

8. Aa, N.P. van der, Luo, X., Giezeman, G.J., Tan, R.T., Veltkamp, R.C. (2011), Utrecht Multi-Person Motion (UMPM) benchmark: a multi-person dataset with synchronized video and motion capture data for evaluation of articulated human motion and interaction, *Proceedings of the Workshop on Human Interaction in Computer Vision (HICV), in conjunction with ICCV 2011*.

Making Ambient Spaces into Playgrounds

Dennis Reidsma, Daniel Tetteroo, Anton Nijholt

Human Media Interaction / Creative Technology, University of Twente, Enschede, The Netherlands.
d.reidsma@utwente.nl

Introduction

Ambient spaces are characterized by sensors, which allow detailed observation of visitors and inhabitants, and by displays and actuators that allow the system to react to the observed behavior. Starting from a general ambient space, we are working towards an interactive environment that invites and seduces people to interaction with each other and with the environment. We aim at “play” rather than at “game”; the visitor has no obligation to interact with the system, nor is there any specific goal to reach or task to complete.

Background and motivation

The development of our interactive playground (see Figure 1) is inspired by children’s play, combining the rich interaction possibilities of computer games with the physical and open-ended aspects of traditional playground play [1,2]. The “open-ended play” afforded by such a playground may benefit children in many aspects of their development [3, 4].

Offering specific toys, play objects, or mechanisms on a playground does not necessarily mean they will be used as intended. Children appropriate the things that are present in playgrounds, such as swings and seesaws, and re-purpose them for their own play and games. The boundaries between play and game are blurred, the two types of activity merge seamlessly. While they play a game, playful activity occurs within, besides and around the specific rules and goals. “Dramatic performance in play” (bodily acts, a special jump or shout, etc.) occurs, and “in fact it often becomes the most important activity of the play in itself, and it becomes the one admired and imitated by others.” [5]. In the pilot tests with the first prototype of our interactive playground, for example, one pair of children decided to designate the playground to be an environment for expressive dancing. On the other hand, during open play, *ad hoc* rules are continually being defined, adapted and discarded, and spontaneous “gamelets” emerge continually. Many such temporary rule introductions were observed during the pilot tests. For example, one of the children shouted at some point: “We must all run away from the yellow shapes that are

Figure 1. Children playing with the HMI Interactive Playground

following you!” (interestingly, in the thematic concept underlying this playground, the shapes that are following you were actually envisioned to be “your shapes”, and not hostile at all). At another point, a child said: “come on everybody, we must all get off the play area for a moment, come on…”. This little “rule” was introduced with an ulterior motive: this child was the first person to run onto the field again and claim one of the colored shapes that was previously “caught” by one of the others.

Measuring and influencing social and affective behavior in the interactive playground

Intelligent interactive playgrounds offer possibilities beyond those of a seesaw or a swing, which just sit there, waiting to be used in play. An intelligent playground is more like another person, responding to what you do, but also proactively contributing things to the play that occurs. By well-placed interventions in the interactive play, an intelligent playground can influence many social and physical properties of the play that emerges [6]. To this end, the playground needs to observe and interpret the social and affective behavior of the players, and determine appropriate (re)actions to change the interaction between visitors [7].

For example, the amount of nonverbal coordination (“motion synchrony”) between visitors may be measured to yield insight into their social relation [8], but might also be manipulated through coordinated interaction between the system and the individual visitors, in order to change their relation [9]. The playground may contain sensors for the detection of engagement and group forming, and when a person is observed to be uninvolved in the play, the interactive elements of the playground may explicitly link this person to the others in an attempt to draw them into the play. The system may respond to motion and sound in such a way that the children are encouraged to jump around wildly and scream at the top of their voices, or, conversely, to be very quiet and careful. Through designing such interactive interventions, we aim at increasing the engagement of people with the system and with each other.

Pilot evaluation of the first prototype

The first prototype of the playground was evaluated in a series of user tests with 19 children, in groups of 2-4, playing for approximately 30 minutes inside the playground. The first 10 minutes they were allowed to play without any instruction at all. After that, they were asked to tell about the interaction patterns that they had discovered in the system, and they received a 5 minute explanation of some of the patterns that they had not yet discovered. The goal of these tests was twofold. Firstly, we wanted to know whether the children would display the mix of play and game typical of children’s activity on a playground, and whether we would observe them “telling stories” about the possible interaction patterns and changing these stories throughout the session. As can partially be seen from the examples above, this was the case. Secondly, we attempted to introduce a set of interaction patterns in the system that specifically targeted the social dimension of *cooperation* and *collaboration*, hoping to increase the presence of either of these aspects in the play of the children. However, the amount data that we collected was too small to draw clear conclusions about that.

Future work

We are currently involved in developing an annotation scheme and an inventory of relevant dimensions of play behavior, concerning both physical and social dimensions. This annotation scheme is based upon recordings made of the evaluations of the first prototype playground and upon observations of children’s play at a number of primary school playgrounds. This should lead to the development of algorithms for automatic detection of relevant social and affective behavior, and the development of a series of new interactive playground prototypes. Future experiments will focus more on the elicitation and influencing of various types of social and physical behavior in the children’s play.

References

1. Tetteroo, D., Reidsma, D., van Dijk, B., and Nijholt, A. (2011). Design of an interactive playground based on traditional children's play. *Proc. of INTETAIN ’1.*, LNICST 78, Springer Verlag, 129-138. In print.

2. J. Sturm, T. Bekker, B. Groenendaal, R. Wesselink, and B. Eggen (2008). Key issues for the successful design of an intelligent, interactive playground. *Proc. of IDC '08.* ACM, New York, NY, USA, 258-265.

3. Y. Rogers and S. Price (2004). Extending and augmenting scientific enquiry through pervasive learning environments. *Children, Youth and Environments* **14**, 67-83.

4. T. Bekker, J. Sturm, R. Wesselink, B. Groenendaal, and B. Eggen (2008). Interactive play objects and the effects of open-ended play on social interaction and fun. *Proc. of Advances in Computer Entertainment Technology '08.* ACM, New York, USA, 389-392.

5. Lund, H. H. and Jessen, C. (2005). Playware - Intelligent technology for children’s play, *Technical Report TR-2005-1*. Maersk McKinney Moller Institute for Production Technology, University of Southern Denmark.

6. Lindley, S., Le Couteur, J., and Bianchi-Berthouze, N. (2008). Stirring up Experience through Movement in Game Play: Effects on Engagement and Social Behaviour. *Proc. Of CHI ’08.* ACM, New York, USA, 511-514.

7. Zancanaro, M., Oliviero, S., Tomasini, D., and Pianesi, F. (2011). A socially aware persuasive system for supporting conversations at the museum café. *Proc. of the 16th international conference on Intelligent user interfaces (IUI '11).* ACM, New York, NY, USA, 395-398.

8. F. Ramseyer and W. Tschacher (2008). Synchrony in dyadic psychotherapy sessions. *Simultaneity: Temporal Structures and Observer Perspectives.* World Scientific, Singapore, 329-347.

9. Tomida, T., Ishihara, A., Ueki, A., Tomari, Y., Fukushima, K., and Inakage, M. (2007). MiXer: the communication entertainment content by using "entrainment phenomenon" and "bio-feedback". *Proc. Of Advances in Computer Entertainment Technology ‘07.* ACM, New York, USA, 286-287.

Building Corpora of Bodily Expressions of Affect

Marco Pasch, Monica Landoni

Faculty of Informatics, University of Lugano, Lugano, Switzerland. {marco.pasch, monica.landoni}@usi.ch

Abstract

Studies aimed at measuring users' engagement with interactive products mostly rely on collecting self-report data. However, there is an increasing body of research into systems able to assess the affective state of a user from objective data such as body movements. To train such systems we need corpora of motion capture data showing people in emotionally charged situations. Here, we report on our ongoing research into building corpora of movement data. In particular, we focus on two crucial areas in the building process. Firstly, challenges related to recording corpora "in-the-wild", i.e. in ecologically valid situations. Secondly, challenges related to obtaining ground truth labels, i.e. ensuring that individual sequences of movements are labeled correctly with affective states that can be seen in them.

Introduction

Studies aimed at measuring users' engagement with interactive products mostly rely on collecting self-report data 1. However, people can only report on what they are aware of. Self-report also requires that people can give an accurate description of what they think about a product. Children, for example, often have difficulties in reflecting and communicating their thoughts, beliefs, and wishes. Another potential bias in self-report data is that people not necessarily give us truthful answers. They may leave out items they feel embarrassed about or tell lies instead. Often these lies are innocent when people simply tell us what they believe we want to hear instead of their actual opinion.

An alternative to subjective self-report methods are those based on objective measurements. In recent years, the field of affective computing has come up with a number of approaches that provide more objective data. Typically, these approaches detect one or several modes of affective feedback. Intentionally or not, we continuously express affect. This can happen verbally, through the tone of our voice and our choice of words, as well as non-verbally, through our facial expressions, posture, gestures, amongst others. Physiological data like heart rate or skin conductance have also been used to infer the affective state of the user. When we talk to other people we intuitively pick up on cues to their affective states. As social beings this comes natural to us. However, having computers automatically recognize a person's affective state is still one of the big challenges in affective computing.

Our research aims at assessing engagement from bodily expressions of affect, in a way enabling a system to read the body language of users. This could be beneficial in a number of settings. It could provide researchers and practitioners evaluating interactive products with an initial assessment of how a user felt during interaction, which they can use as a starting point for conducting post-interaction interviews. It could also be an important step towards intelligent systems, which can adapt to their users 1. For instance, if a video game is able to detect the user is frustrated, it can adapt the gameplay and ensures to keep the user in the flow zone 3. Same holds for tutoring systems that ensure the learner is in a positive affective state.

The system we are working on is based on the Microsoft Kinect sensor. The Kinect sensor is a camera-based movement sensor, emitting an infrared grid and aggregating the reflected light into a skeletal representation of the body. Using the Kinect SDK we developed our own motion capture system which records the body movements of up to two players, represented in the form of skeletons. It is fairly robust in terms of ability to function in changing (indoor) environments and totally non-intrusive for the user.

To build systems able to assess affective states from body movements we must know which features within a movement carry cues to a person's affective state. To identify these features we need a corpus of movement data, i.e. must record movement data and label it with the according affective state. Existing corpora of affective behavior usually consist of acted data, which means they often show exaggerated movements done by people who were instructed to depict a particular affective state in a lab environment. However, for affective computing to advance it is important to move from acted data to data obtained from people actually feeling the expressed affective states in ecologically valid environments 4.

Apart from recording “in the wild”, we also must ensure that the data is labeled as accurate as possible.

Observing movement and affective behavior

We conducted initial observations of children playing active video games in a primary school in southern Switzerland. Observing children in a school is a perfect scenario for us, as the school is a very natural environment for its students. After establishing a contact, we conducted observations in the setting of the school's after-school program as well as physical education classes.

The after-school program is a playful setting, where children can spend their time doing homework or play games under supervision until their parents come to pick them up. The physical education class is a formal setting and comes with the great advantage that a central learning objective of this class is for children to learn how to deal with positive and negative emotions, which makes it an ideal scenario for our study.

In both settings we introduced a Nintendo Wii console and observed how children play the Wii Sports games, a set of games simulating tennis, bowling, baseball, and boxing. We also videotaped some of the observation sessions of the physical education classes and reviewed the videos with the teacher. This revealed that we can observe three classes of movements: task-related movements, affective expressions, and social behavior.

Task-related movements are all movements that have to do with playing the game. In the context of the Nintendo Wii this means mostly swings of the arms and flicks from the wrists. Obviously, these movements do also carry affective information. We can observe from the way a child executes, for example, the throw of a bowling ball whether he/she is joyous because the last throw was a strike, or whether he/she was just embarrassed by classmates prior to the throw.

Affective expressions are movements that have no utilitarian function and that only serve the communication of affective states. An example for this is a child throwing his/her hands in the air after scoring a strike in bowling. Another example is a child walking spiritlessly away from the console with hanging shoulders after loosing a game. There is obviously a social component present in that the expression of affect has a social function, such as eliciting sympathy when feeling sad. In fact, most affective expressions happen in a social context. When there is no recipient present, we feel less need to communicate our affective states. Our third class of movements is social behavior. Here, movements have no utilitarian function and are also not primarily communications of affect. We see for instance one girl hugging another girl in a gesture of empathy or a boy putting an arm around another boy to express their friendship.

Recording movement data “in-the-wild”

An important requirement for this is ensuring that the data is obtained from a naturalistic setting. This means that the data comes from people expressing real, i.e., actually felt, affect rather than simply portraying affective states. It also means that we capture the data in an environment that is familiar to people from whom we obtain it. Ideally, it also means they are doing things they are familiar with. In essence, we try to reduce a potential novelty effect and a feeling of insecurity due to alien circumstances as much as possible.

We are currently piloting a study aimed at capturing movement data for which we return to the aforementioned school. Apart from the school being a familiar environment, this setting comes with the advantage that the children there are already used to our presence. Similar to our previous observations, we want the children to engage in playing active video games as the Nintendo Wii. From our observations we already know that the Wii games are known to most of the children. So also the use case is familiar for them.

Children will be playing in couples with Wii Sports games of their choice, playing in a competitive fashion and asked a few questions at the end of a session. Here we want to know how they feel during playing so that we can match their subjective experience against direct observations and the teacher's analysis of the whole experience. In addition to recording movements with the Kinect based recognition system we will video tape the session for later analysis.

Having the teacher on board and enthusiastic to take part in the study is invaluable for us. Not only his professional knowledge on student behavior, but also the fact he has known most of the students since they enrolled in the school has already provided us with many insights during our initial observational study. One of the main learning goals of the physical education syllabus is for students to learn to identify and control their emotions. So our study does not interfere with the class' activities, it may actually provide the teacher with new insights and aid achieving the learning goals.

Labeling data

In order to be able to use the motion capture data for training and testing an affect recognition system, we need to label individual sequences with what affective state can be seen in them. In affective computing, this process is known as ground truth labeling.

As the labels are the benchmark that a system is tested against, it is an important step for the creation of affect recognition systems. Usually, labels are obtained from letting observers rate the sequences and for instance by majority vote assign a label to each sequence. Our own labeling procedure will involve two researchers assigning labels and discussing those with the expert (teacher) in order to validate these. We will develop an appropriate coding protocol to be followed by all human judges involved.

However, we argue that this common practice of ground truth labeling can introduce biases. Often judges are recruited from colleagues of the researchers. This can introduce a population bias into the ratings and makes a generalization to how people in general rate expressions of affect difficult.

Because of this we want to explore how emerging methods such as crowdsourcing can help in improving ground truth labeling. Crowdsourcing here means letting a large population of observers rate the sequences online. This way we can include a bigger population from a wider background into the rating process. However, this also means letting go of a controlled environment where we have no control over the circumstances in which the observers rate the sequences. Still, we believe that exploring different rating schemes is an undertaking worthwhile. Ideally we can find more robust rating schemes. For the very least, we hope this can inform researchers in developing future corpora.

Conclusion

We described our ongoing research into building an automatic affect recognition system based on body movements. In particular, we discussed the necessity of corpora of annotated movement data that shows affective behavior of humans in naturalistic settings for training such systems as well as requirements towards building them. Once we have recording and labeled the data, we plan to make it available to other researchers. Existing corpora of affective behavior usually consist of acted data, which means they often show exaggerated movements done by people who were instructed to depict a particular affective state in a lab environment. However, for affective computing to advance it is important to move from acted data to data obtained from people actually feeling the expressed affective states in ecologically valid environments. We believe our corpus can help other researchers to test their systems with more valid data than what is available today.

References

1. Bargas-Avila, J. A., Hornbæk, K. (2011). Old wine in new bottles or novel challenges? A critical analysis of empirical studies of user experience. *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI'11)*, ACM Press, pp. 2689–2698.

2. Pantic, M., Nijholt, A., Pentland, A., Huang, T. (2008). Human-Centred Intelligent Human-Computer Interaction (HCI²): how far are we from attaining it? *Int. J. of Autonomous and Adaptive Communicative Systems* **1**(2), 168-187.

3. Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM* **50**(4), 31–34.

4. Picard, R.W. (2011). Measuring Affect in the Wild. *Proceedings of the 4th Int. Conf. on Affective Computing and Intelligent Interaction* (ACII'11), Springer, p. 3.

Measuring Fun and Enjoyment of Children in a Museum: Evaluating the Smileyometer

F. van der Sluis, E.M.A.G. van Dijk, L.M. Perloy

Human Media Interaction, University of Twente, Enschede, The Netherlands. f.vandersluis@utwente.nl

Introduction

Measuring fun and enjoyment with children is not trivial. Subjective measures are known to suffer from an experimenter effect and often lack detail in their answering. The experimenter effect refers to answering in accordance with the expectations of the experimenter rather than reflecting the opinion of the subject. The lack of detail in answering is due to the oft-found finding (which we will illustrate in this paper as well) that children tend to select the extremes of a scale: either very high or very low, leaving out more fine-grained possibilities. Markopoulos et al. (2008) [1] treat related issues 'satisficing' and 'suggestibility' as important issues to deal with when using survey methods with children. Suggestibility refers to the influence of the experimenter that is generally expected to be more important for children than for adults because of age differences. A respondent that is satisficing gives more or less a superficial answer that appears reasonable without carefully considering the question and the answer options. With children satisficing can easily occur when they find the questions difficult to understand.

With increasing age, children develop increasing skills and ability which explain their different answering tendencies as compared to adults. Piaget noted in his pioneering work the development of thinking skills around the age of 11, where children learned to use abstract concepts and reflect upon them. This includes self-regulation, a skill consisting of self-motivation, attention control, self-monitoring, and self-evaluation [2]. For a subjective measure of fun and enjoyment these skills are pivotal: the ability to monitor oneself and identify these abstract concepts is requisite.

Several attempts have been made to counter the described effects. For example, objective measurements allow (indirect) indicators of fun such as total time spent, total interactions performed, or video recordings. Regarding fun and enjoyment, subjective measures are common and easily applied. Hanna et al. [3] compared the usefulness of several rating methods. Special subjective instruments have been developed for children, such as the Smileyometer or the Visual Analogue Scale (VAS). The Smileyometer (See **Error! Reference source not found.**) uses images of smileys to make the items on the scale more recognizable for the participant. It was designed with the help of children, leading for it to have a non-neutral face at the centre of the scale [4]. In this way, the scale was expected to be more recognisable to children, leading to more fine-grained answers; i.e., less extreme answers. A VAS uses a question format with pictorial representations that children use to identify their feelings or opinions. According to Read et al. [5], VAS scales can be very useful for younger children but when used to elicit opinions about software or hardware products younger children almost always indicated the highest score on the scale. Hence, in this paper we will focus on a popular subjective method for measuring fun with children, the Smileyometer.

The Smileyometer will be reviewed based on two studies we performed in Museon, a science museum in The Hague, The Netherlands. The museum is an educational museum for a broad audience and the main themes are the human and his role in the society, nature, culture and science and technique. Its goals are education, study

Figure 1. The Smileyometer.

and enjoyment. With over 150.000 visitors each year, of which the majority are children, it is a popular and modern location for fun and learning. As a test location, the museum assures a natural environment, intensifying the need for measurement devices which are reliable and valid in less constraint, non-laboratory settings as well.

Two studies will be reported, both evaluating the enjoyment of children during a quest through the museum. This quest started at a multi-touch table with the selection of topics of interest from the permanent exhibition of Museon.

This selection was used to generate the electronic quest that children answered in the exhibition area. Finally, the quest ended at the multi-touch table with an end game. For both studies an informed consent was given by the parents. The difference between the studies lies in the interfaces used at the multi-touch table, being a direct change of the beginning and the end of the quest and an indirect change on the middle part: the quest changes according to the interactions of the first part. Study One used a prototype whereas Study Two used an evolved version of the system. The focus here, however, will not be on the interfaces but on the Smileyometer, in particular for its experimenter effect and coarse answering.

Study One: Prototype

We used the Smileyometer at the end of the quest through the museum to measure reported enjoyment of the three main parts of the quest: the beginning (initial selection of topics), middle (quest through the permanent exhibition), and end (final game at the multi-touch table). **Error! Reference source not found.** shows the Smileyometer. For each of the three parts of the quest the children used the Smileyometer to answer the question “How much fun was it to do that part?”

Table 1 shows the means and standard deviations (SD) of the first study in which 36 children took part. The answers on the Smileyometer are re-coded to 1 (for awful) until 5 (for brilliant). The results indicate the Smileyometer does not alleviate the mentioned problems: the answers were still almost always very high. Partly this can be explained by the experience actually being fun. Looking at the difference between age groups gives further insights in the causes. Table 1 also shows the descriptive statistics for the younger children (7 to 9) and the older children (10 to 15) separately. Two-tailed t-tests confirmed that for the begin and the middle part, the scores measured with the Smileyometer were significantly higher for children between 7 and 9 compared to children between 10 and 15 (begin part: $t(34)=2.345$, $p= 0.025$, middle part: $t(34)=.0367$, $p=0.004$). However, the final part showed no significant difference. We found significant correlations between the scores on the Smileyometer and age groups for the begin part (Spearman’s rho = 0.371, $p=0.026$) and for the middle part (Spearman’s rho = 0.425, $p=0.010$). Younger children scored higher. The scores of the final part were not correlated with age. In the begin and middle part the highest score was selected more often by younger children than older children (begin part 58.3% vs. 16.7%, middle part 83.3% vs. 45.8%), no difference was found for the final part (both groups 41.7%). The data indicate younger children have a higher tendency to select the highest ratings, but do select none-extreme scores as well.

Table 1. Enjoyment in a museum quest as measured by the Smileyometer - Descriptive statistics of study 1.

	All			Age 7-9			Age 10-15		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Begin	36	4.11	0.75	12	4.50	0.67	24	3.92	0.72
Middle	36	4.42	0.77	12	4.83	0.39	24	4.21	0.83
End	36	4.11	0.98	12	3.92	1.24	24	4.21	0.83

Table 2. Enjoyment in a museum quest as measured by the Smileyometer - Descriptive statistics of study 2.

	All			Grade 7			Grade 8		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Begin	26	3.77	0.77	16	4.00	0.82	10	3.40	0.52
Middle	48	3.27	1.13	27	3.78	1.05	21	2.62	0.87
End	48	3.73	0.98	27	3.74	0.90	21	3.71	1.10

Study Two: Matured Version

A study similar to Study One was performed at the museum, using a later version of the multi-touch table interfaces at the beginning and at the end of the quest, including improvements based on the evaluation of Study One. This study employed a between-subjects design comparing a personalized quest created by the improved interface versus a normal, fixed, quest without any personalization nor multi-touch interaction at the beginning.

Table 2 shows the descriptive statistics for the Smileyometer of the second study with 48 children. Notice that the number of children that played the initial selection game is roughly half of the total number of participants because of the experimental setup of this study. These results are not extremely high any more. This can partly be explained by the ages of the children. All children were 10-12 years old and were part of two classes (grade 7 and grade 8 in the Dutch school system) of a primary school in The Hague.

We studied age differences again by comparing the results between the class from grade 7 (10-11 years old) and the class from grade 8 (11-12 years old). Table 2 also shows the descriptive statistics per grade level. Even though the difference in age is only one year, again for the begin and the middle part, the scores measured with the Smileyometer were significantly higher for children from grade 7 than for children from grade 8. (results two-tailed t-tests begin part: $t(24)=2.071$, $p= 0.049$; middle part: $t(46)=4.090$, $p<0.001$). Again, no significant difference was found for the final part. We found a strong correlation between the score on the Smileyometer and grade level for the middle part (Spearman’s rho = -0.512, $p<0.0005$) and a marginally significant correlation for the begin part (Spearman’s rho = -0.375, $p=0.059$). The scores of the final part were not correlated with grade level. Selection of the highest score in the begin and middle part was also very different between children from grade 7 and 8. For the begin part the highest score was chosen by 5 children, all from grade 7 (31.2% vs.0%) In the middle part 8 children, all from grade 7, chose the highest score (29.6% vs. 0%). Again no difference between grade 7 and 8 was found for the final part (22.2% vs. 28.6%). Remarkable here is that the majority of the children who chose the highest score (9 out of 12) were from the group who did not play the initial game and hence played with the multi-touch table for the first time during the end game. The data indicate that even this age difference of one year shows differences in the tendency to select the highest ratings, but at the age of 10-11 years many children do select none-extreme scores as well. There even seems to be a turning point at the age of 11-12 years (just before Dutch children go to secondary school). These children were much more critical in their reported enjoyment.

Conclusion

We reviewed the Smileyometer as used in two field studies, allowing us to compare two different age groups twice: in Study One ages 7-9 against 10-15 and in Study Two ages 10-11 versus 11-12. The results indicate that the Smileyometer, as customized subjective measurement, did not solve the mentioned problems: the reported fun remained high. Both studies indicated age was a factor in the reported fun, such that younger children give higher ratings. The reasons for this effect cannot be proven from the studies, but do point to the problems mentioned in the Introduction: a tendency towards extreme answers and an experimenter effect.

An alternative explanation for the relation between age and reported fun is that younger children simply had more fun. There is no evidence from observation and discussions between the experimenters and the participants to further corroborate this theory, making this an unlikely yet not impossible counter-explanation.

Although the Smileyometer was not able to solve the problems inherent to subjective questionnaires for children, it did partly alleviate the problems. A possible method to further alleviate the problems is by extending the measurement scale from five to, for example, seven items as is often used with semantic differentials. At least for adults, subjective scales tend to become more precise when consisting of more items. For example, this holds even for up to ten items [6].

Summarizing, using pictures was shown to give a reliable scale, and in particular for older children, give fine-grained results. For younger children, however, some problems still remained. Consequently, when using the Smileyometer or similar subjective scales, age should be included in the analyses as a covariant.

Acknowledgements

This work was part of the PuppyIR project, which is supported by a grant of the 7th Framework ICT Programme (FP7-ICT-2007-3) of the European Union.

References

1. Markopoulos, P., Read, J.C., MacFarlane, S., Höysniemi, J. (2008). *Evaluating children's interactive products.* Elsevier Inc., Amsterdam.

2. Ormrod, J.E. (2004). Human Learning. Pearson Education Inc., Englewood Cliffs, NJ, USA.

3. Hanna, L., Neapolitan, D., Risden, K. (2004). Evaluating computer game concepts with children. *Proceedings of Conference on Interaction Design and Children,* 49 - 56.

4. Read, J.C., MacFarlane, S., Casey, C. (2002). Endurability, engagement and expectations: Measuring children's fun. *Proceedings of Conference on Interaction Design and Children*, 189-198.

5. Read, J.C., Gregory, P., MacFarlane, S., McManus, B., Gray, P., Patel, R. (2002). An investigation of participatory design with children - informant, balanced and facilitated design. *Proceedings of Conference on Interaction Design and Children,* 53-64.

6. Van den Broek, E. L., Van der Sluis, F., and Dijkstra, T. (2012). Modeling an acoustic profile for ubiquitous stress assessment. Personal and Ubiquitous Computing **15**(6) (in press).

On Making Engagement Tangible

Egon L. van den Broek[1,2,3]

[1]*Media and Network Services (MNS), Technical Sciences, TNO, Delft, The Netherlands*

[2]*Human Media Interaction (HMI), Faculty of EEMCS, University of Twente, Enschede, The Netherlands*

[3]*Karakter University Center, Radboud University Medical Center (UMC) Nijmegen, Nijmegen, The Netherlands*

vandenbroek@acm.org

Abstract

In this article the complexity of the construct engagement and three theories on this topic are discussed. Csikszentmihalyi's theory of flow is taken as starting point for the measurement of engagement. The measurement of each of its eight aspects is discussed, including its pros and cons. Regrettably, no overarching computational model is available. This article ends with a concise discussion.

Introduction

In Human-Computer Interaction (HCI) and related (multidisciplinary) fields, engagement is often approached as what one could baptize "*a closed gray box*", surrounding an application. The box is mostly closed because the application is often studied in isolation; that is, solely the application and its characteristics are taken into account not all processes surrounding it (e.g., its development and marketing). The box is gray because concepts such as engagement, involvement, and flow are considered as fuzzy concepts, hard to pinpoint. With this article, I aim to satisfy two aims: i) open up the closed box and ii) bring the box from gray to white.

In this article, I will focus on making engagement tangible. In the next section, the concept engagement will be discussed, including three theoretical frameworks. Subsequently, I will discuss how engagement can be measured, using one of the three theoretical frameworks. However, since this is a complex endeavour, I will also present a coarse but pragmatic approach on measuring engagement. I will end this article with a brief discussion.

The concept engagement

Until recently, there was no consensus present on the concept engagement throughout scientific literature. This is why I baptized the box to be gray instead of white. However, in 2008 this already changed with O'Brien and Toms' work with which they answered the question: What is User Engagement? Additionally, they provided a conceptual framework and a definition for user engagement, with a bias towards its relation with technology. In this article, I adopt their definition of engagement, which is: "*a quality of user experience characterized by attributes of challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control.*" [1] On the one hand, this definition captures most crucial aspects of today's dominant theoretical framework in a concise manner. On the other hand, this definition requires substantial explanation to enable (true) understanding. However, the concept engagement can also be linked to various (other) theoretical frameworks. Due to reasons of brevity, we will refrain from providing an exhaustive review and mention three.

In 1985, Deci and Ryan introduced their Self-Determination Theory (SDT) [2]. Their theory consists of five mini theories, namely: i) cognitive evaluation theory (i.e., on intrinsic motivation); ii) organismic integration theory (i.e., on various forms of extrinsic motivation); iii) causality orientations theory (i.e., on orientation toward environments and regulation of behavior); iv) basic psychological needs theory (i.e., psychological needs and their relations to psychological health and well-being); and v) goal contents theory (i.e., emerging from the distinctions between intrinsic and extrinsic goals and their impact on motivation and wellness). Together these theories aim to assess people's motivation and/or personality functioning. Although SDT has a high construct and ecological validity, it is fragile with respect to content validity. That is, what signals (e.g., biosignals, questionnaires, and system usage) represent what aspects of what mini-theory? Therefore, I will not adopt the SDT as theoretical framework.

In 1989, Davis [3] introduced his Technology Acceptance Model (TAM), which he derived from the theory of reasoned action and related to (information) technology. TAM poses that people's intention to become engaged with a system is determined by its perceived usefulness and perceived ease of use. Hereby, TAM assumes that people are free to act after (s)he has formed an intention to it. However, in practice people's acts are limited by various sources, amongst which limited ability, time, environmental or organizational limits, and unconscious habits. The latter issue limits the ecological validity of TAM significantly. Therefore, I have refrained from using TAM as theoretical foundation in this article.

In the same period as the previous two theories were invented, Csikszentmihalyi defined his theory on flow [4]. An optimal engagement is closely related to, or the same as, optimal User eXperience (UX) or flow [4]. When involved in products (e.g., media and games), the underlying goal is to move towards a flow. According to Csikszentmihalyi [4], this requires: i) a concrete task to complete, ii) the ability to concentrate, iii) clear goals, iv) immediate feedback, v) deep but effortless involvement, vi) a sense of control, vii) sense of self disappears but increases unnoticeable during flow, and viii) the internal clock to be influenced. This decomposition illustrates the complexity of the construct flow, which is even further increased by its relation with constructs such as motivation, passion, presence, and engagement, which all hint towards the same phenomenon (i.e., the feeling of optimal UX). Nevertheless, Csikszentmihalyi [4] provided aspects that each by itself can be measured.

Measuring engagement: The eight aspects of flow

The eight aspects of flow (or engagement) as Csikszentmihalyi [4] identified would span up an eight dimensional space. However, although each individual dimension is appealing and (in principle) could be assessed via objective measurement, an integral model of engagement, in particular, in real world settings is beyond reach. Nevertheless, I pose that the theory of Csikszentmihalyi already provides the means to bring the box from gray to white. Therefore, in the remaining article, I will try to open the box and bring if from lab to life [5].

Three of the eight aspects of engagement can be measured rather straightforward, namely: a concrete task, a clear goal, and immediate feedback. Although the level of abstraction people can handle varies considerably among them, in principle, a concrete task to complete can always be defined (e.g., play a game). The goals related to the task at hand can be made as explicit as needed (e.g., reach level *x* of game *y*). Feedback on people's behavior (i.e., on cognitive, affective, or physical level) can be given on several levels (e.g., direct and/or indirect and conscious and/or unconscious), using all possible modalities (e.g., tactile, auditory, and visual). The remaining five aspects of engagement are harder to pin point.

The ability to concentrate is known to have a high variance both within and between people. As Csikszentmihalyi [2] states "*To pursue mental operations to any depth, a person has to learn to concentrate attention. Without focus, consciousness is in a state of chaos.*" (p. 26) So, the ability of concentration can be assessed using tests founded on mental operations (e.g., calculation), with or without distractors. However, as such, it is hard to capture the ability to concentrate in real world practice, as this is not a well-developed research area [6]. Recently, several attempts have been published of measuring the level of involvement of people in real life situations. For example, in 2009 Sohn [7] investigated the impact of magazine and television social comparison processes on people's body perception. For this study, he developed and, subsequently, validated a new scale, based on the Affect, Reason, and Involvement (ARI) model, measuring involvement in the context of body image. One year later, Van den Ende, Hoonhout, and Meesters [8] developed a 25 items questionnaire to measure people's involvement with audio/video content. However, note that these and other questionnaires most likely will heavily depend on the task and context at hand.

Sense of control has been measured in various contexts. For example, in 2002 Dudek, Merecz, and Makowska [9] have done so in the context of occupation. They introduced their Sense of Personal Control at Work (SPCW) questionnaire. In the same year, Jackson and Eklund introduced their Dispositional Flow Scale-2 (DFS-2) for sports, which has been used in various other contexts as well (e.g., gaming) [10]. So, sense of control is approached from various angles and can be assessed using either validated questionnaires or tailored questionnaires to the topic at hand, which can be founded on the existing theoretical frameworks and one or multiple of the related questionnaires. Sense of self can be conveniently assessed using the Sense of Self Scale (SOSS). In addition, several other questionnaires that assess related constructs can be valuable as well; for example, the Self-Concept Clarity Scale (SCCS), the Self-Monitoring Scale (SMS), and the Extended Measure of Ego Identity Status 2 (EOM-EIS II) [11]. However, these questionnaires do not anticipate on a quickly changing level of sense of self, as Csikszentmihalyi [2] describes. So, for the purpose of measuring engagement such questionnaires do have their limitations.

Roughly one century ago, the notion of human's internal clock was noticed. Nevertheless, the theoretical foundation on the internal clock mechanism is still a topic of debate as current models cannot describe all possible states. Further, research on this issue is limited to lab studies and has not been brought to real life [12]. However, even despite these limitations, significant differences both between people and within people (over time) have been observed [12]. Taken together, this makes the internal clock mechanism very hard to operationalize and, hence, to be used to measure engagement in practice.

Discussion

This article sketched the complexity of the construct engagement. Subsequently, it discussed three of the most influential theories on this topic. Next, Csikszentmihalyi's theory of flow [2] was taken as starting point for the operationalization of the measurement of engagement. The measurement of each of the eight aspects of the theory of flow [2] has been discussed, including its pros and cons. Its most important drawback is that although each of the aspects can measured by itself no overarching computational framework is available or presented here. Most likely, such a framework would require a significant additional effort in basic as well as in applied research.

As was indicated in the previous section, founding the measurement of engagement on Csikszentmihalyi's eight aspects is at least labor-intensive if not impossible in real life practice (cf. [5]). In real life contexts, a pragmatic approach would be of high value, even if it can only bring us a coarse approximation of the level of engagement. Such an approximation can be derived from many more angles. Let me mention one: Virtual Reality (VR). In VR settings one distinguishes between immersion and presence, which can be complemented by aspects of UX (e.g., involvement, interest, and emotion) [13]. Although constructs such as UX do not map one-on-one on engagement, such related constructs should be considered as well as the methods explored to measure them.

If anything, studying the construct engagement requires that an interdisciplinary stance is taken, including insights from psychology [4], physiology, information and computer sciences [1], media, organization theory, and marketing. Moreover, both theory and experience from practice have to be embraced and blended, which is rare in practice (cf. [8]). Further, one should realize that engagement is of importance on all possible levels and all possible situations; for example, the engagement of subjects in scientific studies or student with their lectures [6], of consumers with products [7], of (knowledge) workers with their occupation [9], and of athletes with their sport [10]. As such, engagement is omnipresent and central in our lives.

In sum, on the one hand, nowadays engagement is acknowledged for its key importance in daily practice. On the other hand, engagement still has to become more than yet one more buzz word. It has to be defined and operationalized properly. A definition and theoretical framework was adopted. Moreover, it has been outlined how engagement can be measured, including its pros and cons. As such this brief article can perhaps provide a useful springboard for further research.

Acknowledgments

This publication was supported by the Dutch national program COMMIT (projects P4 Virtual worlds for well-being and P7 SWELL). Further, the author gratefully acknowledges the special session chairs for their invitation and the three anonymous reviewers for their to the point and constructive remarks.

References

1. O'Brien, H.L., Toms, E.G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology* **59**(6), 938-955.
2. Deci, E., Ryan, R. Self-Determination Theory: An approach to human motivation and personality. http://www.selfdeterminationtheory.org/. Accessed on 13 July 2012.
3. Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly* **13**(3), 319-340.
4. Csikszentmihalyi, M. (1997). *Finding flow: The psychology of engagement with everyday life*. New York, NY, USA: BasicBooks / HarperCollins Publishers, Inc.
5. Broek, E.L. van den, Janssen, J.H., Westerink, J.H.D.M. (2009). Guidelines for Affective Signal Processing (ASP): From lab to life. In J.F. Cohn, A. Nijholt, M. Pantic (Eds.), *Proceedings of the IEEE International Conference on Affective Computing & Intelligent Interaction* (Amsterdam, The Netherlands, September 10-12), Volume I, 704-709.
6. Alapin, I., Fichten, C.S., Libman, E., Creti, L., Bailes, S., Wright, J. (2000). How is good and poor sleep in older adults and college students related to daytime sleepiness, fatigue, and ability to concentrate? *Journal of Psychosomatic Research* **49**(5), 381-390.
7. Sohn, S.H. (2009). Body image: Impacts of media channels on men's and women's social comparison process, and testing of involvement measurement. *Atlantic Journal of Communication* **17**(1), 19-35.
8. Ende, N. van den, Hoonhout, J., Meesters, L. (2010). Measuring involvement with audio/video content. In T. McEwan, L.M. McKinnon (Eds.), *People and Computer XXIV - Games are a serious business: Proceedings of HCI 2010, The 24th BCS Interaction Specialist Group Conference* (Dundee, UK, September 6-10), 319-327.
9. Dudek, B., Merecz, D., Makowska, Z. (2002). Theoretical assumptions and psychometric characteristics of the sense of control at work questionnaire. *International Journal of Occupational Medicine and Environmental Health* **15**(1), 29-36.
10. Jackson, S.A., THomas, P.R., Marsh, H.W., Smethurst, C.J. (2001). Relationships between flow, self-concept, psychological skills, and performance. *Journal of Applied Sport Psychology* **13**(2), 129-153.
11. Ickes, W., Park, A., Johnson, A. (*in press*). Linking identity status to strength of sense of self: Theory and validation. *Self and Identity*.
12. Hancock, P.A., Szalma, J.L., Oron-Gilad, T. (2005). Time, emotion, and the limits to human information processing. In D. McBride, D. Schmorrow (Eds.), *Quantifying Human Information Processing, Chapter 5*, p. 157-175. Oxford, UK: Lexington Books / Rowman & Littlefield Publishers, Inc.
13. Slater, M. (2003). A note on presence terminology. *Presence-Connect* **3**, 3.

What Can Body Movement Tell Us About Players' Engagement?

N. Bianchi-Berthouze

UCLIC, University College London, London, UK. n.berthouze@ucl.ac.uk

Introduction

With the advent of full-body technology, we are witnessing a growing interest in the role played by the body in shaping and also measuring experiences [1]. There are three main reasons for this. First, the emergence of multimodal technology has brought our body back to the centre of the interactive experience. Second, the availability of increasingly cheaper and unobtrusive motion capture sensors allows for a more feasible analysis of continuous behaviour. Third, an increasing understanding, from various research fields, of the fact that body movement and posture affect the way we perceive and evaluate both ourselves and the environment (e.g. [2, 3]), raises new interesting and challenging questions regarding measuring engagement.

According to the pragmatist philosopher John Dewey and to embodied interaction research [4, 5, 6], experience is not predefined by the design of technology. Rather, it emerges through the interaction between user and technology within a context where the body is a vehicle for engagement. This has led HCI and game researchers to investigate more thoroughly the role of body movement in player experience [7-10]. In [7], the author proposes a taxonomy of movements that is important to player experience and lets the players appropriate the game, i.e. actively construct meanings of their own game experience. In this paper, we use this taxonomy and the studies reported in [7, 12, 13] to suggest body movement measures to gain insight into the player's engagement.

Body movement as a measure of commitment to gameplay

According to the taxonomy proposed in [7], there are two types of body movements that are related to the performance of the game: task-control movements and task-facilitating movements. Task-control movements are the movements set by a game designer for the control of the game. Task-facilitating movements are movements that players use to facilitate the control of the game (i.e. distributed cognition over body resources available [14, 15]). Since coordinating available resources requires knowledge about the activity (i.e. the game), these types of body movements are generally used only by experienced players [14].

What should we measure of these two categories of movements? In [12] and [13], the authors found that players appropriated the controllers, and hence the body movements to control the game, according to the motivations that led them to play the game. When they were led by wanting a hard-fun type of game (i.e. challenging their skills to win the game), players reported using their body movements to exploit optimally the functionalities offered by the controller to gain points. Hence, a first measure of engagement should provide a measure of the player's effort to optimize the control of the game. Four consecutive phases could be envisaged as part of this optimization process: exploration, consolidation, mastering and performing.

During the exploration phase, the player is engaged in understanding what the control movements are and how their body can accomplish them. A body-pattern-entropy measure could be used to detect these explorative phases. This measure could be computed for different body parts or groups of body parts. Investigating the profile of these measures over time could provide insights into the effort and progress of the players. It could be expected that, rather than decreasing over time, their profile may be characterized by multiple local minima (each corresponding to the learning of a particular body movement pattern), until certain body patterns are acquired. If a player remains in a local minimum for too long and his/her task-control movements are still far from optimal, boredom or frustration may appear and the player may disengage. When a set of coordinated task-control movements has been learnt, i.e. their distance from their optimal execution is below a certain threshold, their consolidation phase may start. In this phase, we may expect that the measure of body pattern entropy may increase only on the body parts involved in the type of patterns to consolidate. A measure of distance between the executed movement and the optimal target movement, together with a measure, over time, of the tuning exploration space, could be used to capture the player's effort towards achieving his/her goal.

The phase of mastering may see the emergence of task-related movements (e.g. involvement of other body parts) to further optimize the learnt control patterns. Differently from the explorative phases, where various body parts may be involved without really contributing to the game control, the task-related body movements should be aligned (e.g. in synchrony, or sequentially organized) with the task-control body movements. For example, in [7], the rhythm of foot tapping (task-related movement) is synchronized with the pressing of the guitar buttons (task-control movement) in the Guitar Hero music game [7]. Hence, measures of relatedness between task-control body movements and emerging body movements may be used to detect the players' engagement in appropriating the game and creating their own body movement strategies. Finally, during the performing phase, the player is using the acquired body patterns to win the game. We should expect low entropy at this stage and a variety of task-control movements being exploited.

Whilst we have presented these four phases as consecutive and fully independent phases, we may in fact expect that they overlap as the learning of each different movement pattern may not start at the same time. Also, as the players get better, they may take on new challenges and hence a new learning process may start.

Creation and steering of the interactive game experience.

An important aspect of the player experience is the way the player continuously builds and assigns meaning to the interaction within the game. The studies reported in [7] show that, through proprioceptive feedback, body movement plays a key role in facilitating a transition from a pure hard-fun engagement, where the players' motivations are mainly to win and challenge themselves, to an experience that grounds its pleasure in taking up the role-play that the game offers. Particularly important in facilitating this shift are role-play body movements.

When easy fun and emotional experience motivated play, players fully engaged with and gained pleasure from their body movements, even when these interfered with scoring points [7, 16]. Hence, optimal control of the game may not always be what players look forward to; what players may be working towards is to achieve enjoyable role-play movements. Automatic recognition of role-play body movements may be necessary to measure automatically such types of engagement. The increasing availability of data captured using new movement sensing technologies (e.g. CHALEARN Gesture Challenge[1]) is providing the space and the resources for such capabilities to be developed and we are assisting with a steady progress in this direction [17-22].

Another source of information on the engagement of the players is provided by their affective and social body expressions. Various efforts have been made to create software that can automatically recognize such expressions [11], with increasing interest in naturalistic expressions. For example, [23, 24] have investigated this question in the whole-body game context with promising results. The recognition of these expressions not only informs us of how the player feels but also about the player's motivations for playing the game. In fact, hard fun and easy fun are generally characterized by a different set of emotions [7]. Furthermore, it may also be interesting to explore when the affective expressions may appear to be acted (i.e. part of the role-play). This may shed light on when how the players are actively building the social and affective parts of their experience. Finally, recognizing the players' affective body expressions can also provide information about their action tendency, i.e. their readiness to act upon the events [25].

A problem related to the detection of affective expression is its dependence on the type of action that is performed (e.g. a backhand vs. forehand in a tennis game). Studies on acted body expressions (e.g. [11, 26]) have shown that certain features for affective body expression discrimination appear to be constant over different types of action. However, the fact that these results are based on acted expressions may raise the concern that this could be true only for stereotypical expressions. It would be interesting to explore computational models that can optimize the solution of the affective discrimination problem by automatically exploiting information about

[1] http://www.kaggle.com/c/GestureChallenge. Downloaded on 1 April 2012.

the context (i.e., type of action, identity of the person, etc.). These kinds of approaches have shown promising results (e.g. [27]) in the recognition of emotions by using computational models that co-learn tasks that are orthogonal to emotion recognition (e.g. learning to recognize person identity) by exploiting the knowledge that these tasks are based on quasi-orthogonal (i.e. separate) features. This could be the case with action and emotion. In fact, results by Atkinson et al. [28] have highlighted that, whereas recognition of the type of action being performed (e.g. walking vs. playing tennis) is based on local body joint information, recognition of the affective content of an expression is based on overall measures of body configuration and dynamics.

Conclusions

The emerging full-body technology and the increasingly cheaper motion capture systems offer interesting possibilities for the design and evaluation of entertainment technology. Body movement is a very important source of information about player experience. It offers the means to measure how the player is appropriating the game and at the same time it is a window on the player's affective experience. The field is emerging and providing interesting features for measuring player experience. We propose here to investigate changes in body movement entropy as a way to gain insight on how the player is appropriating the game control. Four different phases were discussed where the profile of such measure may differ: exploration, consolidation, mastering and performing. Furthermore, affective and social body expressions may be used to measure not only how the players feels but also to understand if the players is actively building a hard-fun type of experience or if s/he is looking forward to an easy-fun type instead. When easy-fun is the main motivation of the player, the body patterns to be measured are also (and possibly mainly) the role-play ones rather than only the ones to optimally control the game.

References

1. Muller F., Bianchi-Berthouze, N. (2010) Evaluating Exertion Games Experiences from Investigating Movement Based. *Human-Computer Interaction Series*, Part 4, Springer, 187-207.
2. Niedenthal, P.M., Barsalou, L.W., Winkielman, P., Krauth-Gruber, S., Ric, F. (2005). Embodiment in Attitudes, Social Perception, and Emotion. *Personality and Social Psychology Review* **9**(3), 184–211.
3. Carney, D.R., Cuddy, A.J.C., Yap, A.J. (2010). Power posing: Brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological Science* **21**(10), 1363-1368
4. Petersen, M.G. (2004). Aesthetics of Interaction – A Pragmatist Perspective*. Proceedings Workshop on Aesthetic Approaches to Human-Computer Interaction*, 31- 32.
5. Petersen, M.G., Iversen, O.S., Krogh, P.G. (2004). Aesthetic Interaction – A Pragmatist's Aesthetics of Interactive Systems. *Proceedings on Designing Interactive Systems*
6. Dourish, P. (2001). *Where the Action Is*. MIT Press, Cambridge, MA, USA.
7. Bianchi-Berthouze, N. (in press). Understanding the role of body movement in player engagement. *Human-Computer Interaction,* (in press).

   http://web4.cs.ucl.ac.uk/uclic/people/n.berthouze/BerthouzeHCI12.pdf
8. Isbister, K., Schwekendiek, U., Frye, J. (2011). Wriggle: An Exploration of Emotional and Social Effects of Movement. *Proceeding of SIGCHI conference on Human factors in computing systems,* 1885-1890.
9. Melzer, A., Derks, I., Heydekorn, J., Steffgen, G. (2010). Click or Strike: Realistic versus Standard Game Controls in Violent Video games and Their Effects on Agression, In H.S Yang, R. Malaka, J. Hoshino & J.H. Han (Eds) *Entertainment Computing, LNCS 6243,* Springer: Berlin, 171-182.
10. Isbister, K. and Dimauro, C. (2011). Waggling the form baton: Analyzing body-movement-based design patterns in Nintendo Wii games, toward innovation of new possibilities for social and emotional experience. In *Whole Body Interaction*. D. England, Ed. Springer, London.

11. Kleinsmith, A., and Bianchi-Berthouze, N. (in press). Affective Body Expression Perception and Recognition: A Survey. *IEEE Transaction on Affective Computing*.

12. Pasch, M., Bianchi-Berthouze, N., van Dijk, B., & Nijholt, A. (2009). Movement-based Sports Video Games: Investigating Motivation and Gaming Experience. *Entertainment Computing* **9**(2), 169-180.

13. Nijhar, J., Bianchi-Berthouze, N., Boguslawski, G. (in press). Does Movement Recognition Precision affect the Player Experience in Exertion Games? *International Conference on Intelligent Technologies for interactive entertainment*.

14. Kirsh, D., Maglio, P. (1994). On distinguish epistemic from pragmatic action. *Cog. Sci.* **18**, 513-549.

15. Maglio, P., Wenger, M.J., Copeland, A.M. (2008). Evidence for the role of self-priming in epistemic action: Expertise and the effective use of memory, *Acta Psychologica* **127**(1), 72-88.

16. Shaker, N., Asteriadis, S., Yannakakis, G.N, Karpouzis, K. (2011). A Game-based Corpus for Analysing the Interplay between Game Context and Player Experience. *Affective Computing and Intelligent Interaction Conference; Emotion in Games Workshop*, Springer.

17. Park, J.Y., Yi, J.H. (2006). Gesture recognition based interactive boxing game. *International Journal of Information Technology* **12**(7), 36–44.

18. Park, J.Y. Lee, Y.C., Jo, B.S. Lee, C.W. (2011). Virtual playing ground interface using upper-body gesture recognition. *Workshop on Frontiers of Computer Vision*, pp. 1-5.

19. Reidsma, D., Welbergen, H., van Poppe, R., Bos, P., Nijholt, A.: Towards bi-directional dancing interaction. Harper, R., Rauterberg, M., Combetto, M. (eds.) *ICEC 2006, Lecture Notes in Computer Science*, LNCS 4161, Springer, 1-12.

20. Bleiweiss, A., Eshar, D., Kutliroff, G., Lerner, A., Oshrat, Y., Yanai, Y. (2010). Enhanced interactive gaming by blending full-body tracking and gesture animation. *ACM SIGGRAPH ASIA.*

21. Roccetti, M., Marfia, G., Semeraro, A. (2011). A Fast and Robust Gesture Recognition System for Exhibit Gaming Scenarios. *International Workshop on Distributed Simulation and Online Gaming.*

22. Roccetti, M., Marfia, G., Semeraro, A. (2012). Playing into the Wild: A Gesture-based Interface for Gaming in Public Spaces. *Journal of Visual Communication and Image Representation*, **23**(3), 426-440.

23. Kleinsmith, A., Bianchi-Berthouze, N., & Steed, A. (2011). Automatic Recognition of Non-Acted Affective Postures. *IEEE Transactions on Systems, Man, and Cybernetics, Part B* **41**(4), 1027-1038.

24. Savva, N., Scarinzi, A., & Bianchi-Berthouze, N. (2012). Continuous recognition of player's affective body expression as dynamic quality of aesthetic experience, *IEEE Transactions on Computational Intelligence and AI in Games*.

25. de Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society* **364**(3), 3475-3484.

26. Roether, C., Omlor, L., Christensen, A., Giese, M.A. (2009). Critical features for the perception of emotion from gait, *Journal of Vision* **8**(6,15), 1-32.

27. Romera Paredes, B., Argyriou, A., Bianchi-Berthouze, N., Pontil, M. (2012). Exploiting Unrelated Tasks in Multi-Task Learning. *Proceedings of International conference on Artificial Intelligence and Statistics*.

28. Atkinson, A.P. (2009). Impaired recognition of emotions from body movements is associated with elevated motion coherence thresholds in autism spectrum disorders. *Neuropsychologia*, **47**(13), 3023-3029

29. D'Mello, S., Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. *Cognition & Emotion* **25**(7), 1299-1308.

The Radial Arm Maze (RAM) for the Evaluation of Working and Reference Memory Deficits in the Diurnal Rodent *Octodon degus*

E. Tarragon[1], L. Lopez[1], F. Ros-Bernal[2], J.E. Yuste[1], V. Ortiz-Cullera[1], E. Martin[3], E. Schenker[4], F. Aujard[5], R. Bordet[6], J.C. Richardson[7], M.T. Herrero[1]

[1]*Clinical & Experimental Neuroscience (NiCE) and Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), School of Medicine, University of Murcia, Spain. ertarr@um.es*

[2]*University Jaume I of Castellon;* [3]*University of Albacete;* [4]*Institut de Recherche Servier*

[5]*Centre National de la Recherche Scientifique, UMR;* [6] *University Lille Nord de France*

[7] *GlaxoSmithKline R&D China, U.K. Group*

Spatial memory is highly relevant in biology because it is related with both individual and species survival. Among behavioural tests, one of the most suitable devices for measuring spatial learning and memory is the radial arm maze (RAM) [1,2]. Briefly, RAM consists of eight horizontal arms (57x11 cm) placed radially around a central platform above the floor. Automated doors (20 cm high) are located at the entrance of each arm (Fig. 1). Experimental subjects are placed on a central platform from which they have to collect hidden baits placed at the end of the arms.

The standard version of the RAM animals are habituated to the environment, placed on the central platform and allowed to explore the maze for 15 min per day. Reinforcers (or baits) are scattered on the arms. On the last day of habituation (day 3), the amount of reinforcer is reduced to half, and the session ends when all eight arms have been visited. Following habituation, the animals are trained one session per day for eight consecutive days. One piece of reinforcer is placed at the end of each arm in a well that hides the food from sight, and the animal is allowed to freely explore the maze. Each session lasts until (a) all eight arms have been entered (consider enter an arm when the whole body, except the tail, is inside the arm), (b) 10 min passed since the start of the test, or (c) 2 min passed since the animal’s last arm entrance [3,4]. Arm entries are recorded for later analysis. To prevent odour cues, the maze must be wiped clean between animals. The variables commonly used for the analysis of the performance are (a) the number of errors in each session (entering an arm that has been visited previously counted as an error) and the total number of errors across eight sessions, (b) the number of correct choices in the first eight arm entries of each session, (c) the location of the first error in each session, (d) the number of adjacent arm entries in each session, (e) the time taken to visit each arm (total time to complete the session divided by the total number of arm entries), and (f) the number of sessions to reach the criterion of one error or less, averaged over four consecutive days of training [3-5]

There is extensive evidence that attending to the visual cues located outside the apparatus is one of the elements that subjects use to avoid re-entering the different arms in RAM. Furthermore, at least in rats, a correct

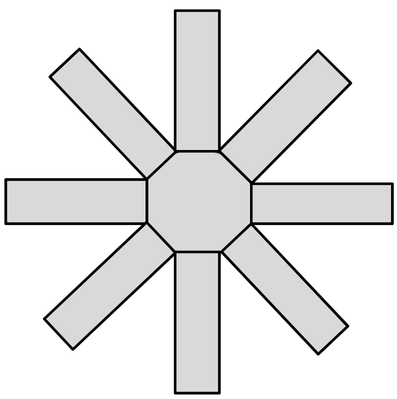


Figure 1. Schematic diagram of the eight-arm radial maze.

performance of the task seems to depend primarily on extra-maze cues [6]. On contrary, olfactory cues were rather related with the improvement of choice accuracy in reduced-visibility situations [7]. Avoiding revisits has been directly related to the amount of surrounding environment cues available, as well as to the viewing duration [8] Moreover, when extra-maze cues predicted the reinforced arms, rats performed almost perfectly on the RAM, whereas following intra-maze cues (such as odour) push performance outcome not better than chance [9]. In this sense, it seems that external cues apparently control choice behaviour when they are easily accessible [10]. Under these circumstances, each visit to an arm may be regarded as a go-no-go discrimination based on extra-maze cues [10]. Finally, it was assumed that when extra-maze cues are limited a representation of spatial locations rather than intra-maze cues might be used when navigating in the radial maze [11].

Modifications of the initial procedure permitted the distinction of spatial working memory errors (double entries into baited places) *versus* spatial reference memory errors (entering never baited arms) [12]. This version of the RAM aims to test working and reference memory at the same time. In their version of the task, only four maze arms are baited [13]. The same maze arms are baited each day and, across sessions, the rats learn to ignore the remaining four arms, which never contain reward. This is the reference memory component of the task, and entry into a never-baited arm is considered a reference memory error. Within a training session, re-entry into one of the four baited arms would be considered a working memory error.

Traditionally, the most extensively studied species in the RAM is the rat. However, a number of experiments have recently demonstrated that another rodent, *Octodon degu* (*O. degu*), is also a valid model for testing reference and working memory in this paradigm.

The experiments in our laboratory show that O. degu are able to learn the task following the traditional protocol for reference memory evaluation, as well as the combined protocol for assessing both working and reference memory (unpublished data). Moreover, animals show a clear learning curve in several variables registered, such as latency to the reinforced arms, total time to complete the essay, the number of errors committed within session, and the number of errors at the end of the training. Furthermore, it has also proved to be sensitive to the temporary impairment in memory caused by sleep deprivation. In this sense, this model results not only valid as a memory testing in normal states, but also in discriminating possible deficits in memory following different protocols for memory impairment.

Conclusions

RAM is a consolidated paradigm for the evaluation of memory. Despite the fact that the most extended model used in this kind of experiments is the rat, RAM has also demonstrated its validity across other animal species (including humans). Over the years, different versions and variations of the RAM have developed, all of them proved as solid as the traditional regarding memory assessment. Accordingly, animal models have also augmented. O. degu is a rodent that has been lately discovered as a valid experimental animal to test memory impairment associated with dementia. In this sense, evaluating the performance of this recent model using RAM results very useful to understand multiple cognitive and behavioural components of memory testing. Moreover, the validity demonstrated with this paradigm opens numerous possibilities within the field of memory and learning studies, especially those regarding cognitive impairment, which in last term will contribute to a better knowledge of these processes.

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement No 115009, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA's in kind contribution (www.imi.europa.eu).

References

1. Olton, D.S., Samuelson, R.J. (1976). Remembrance of placed passed: spatial memory in rats. *Journal of Experimental Psychology: Animal Behavior Proceedings* **2**, 97-116.

2. Olton, D.S. (1978). Characteristics of spatial memory. In: Hulse S.H., Fowler H, Honig W.K. (Eds.), *Cognitive processes in animal behavior*, Hillsdale, NJ: Lawrence Erlbaum Associates, 341-73.

3. Liu, P., Bilkey, D.K. (2002). The effects of NMDA lesions centered on the postrhinal cortex on spatial memory tasks in the rat. *Behavioral Neuroscience* **116**, 860-73.

4. Liu, P., Rushaidhi, M., Collie, N.D., Leong, M.T., Zhang, H. (2008). Behavioral effects of intracerebroventricular microinfusion of agmatine in adult rats. *Behavioral Neuroscience* **122,** 557-569.

5. Liu, P., Bilkey, D.K. (1999). The effect of excitotoxic lesions centered on the perirhinal cortex in two versions of the radial arm maze task. *Behavioral Neuroscience* **113,** 672-682.

6. Suzuki, S.W., Augerinos, G., Black, B.A. (1980). Stimulus control of spatial behavior on the eight-arm maze in rats. *Learning and Motivation* **11**, 1-18.

7. Lavenex, P., Schenk, F. (1995). Integration of olfactory information in a spatial representation enabling accurate arm choice in the radial arm maze. *Learning and Motivation* **2**, 299-319.

8. Mazmanian, D.S., Roberts, W.A. (1983). Spatial memory in rats under restricted viewing conditions. *Learning and Motivation* **14**, 123-139.

9. Olton, D.S, Collison, C. (1979). Intramaze cues and `odor trials' fail to direct choice behavior on an elevated maze. *Animal Learning and Behavior* **7**, 221-223.

10. Babb, S.J., Crystal, J.D. (2003). Spatial navigation on the radial maze with trial-unique intramaze cues and restricted extramaze cues. *Behavioral Processes* **64**, 103-111.

11. Brown, M.F. (1992). Does a cognitive map guide choices in the radial-arm maze? *Journal of Experimental Psychology: Animal Behavioral Processes* **18**, 56-66.

12. Jarrard, L.E. (1978). Selective hippocampal lesions: differential effects on performance by rats of a spatial task with preoperative versus postoperative training. *Journal of Comparative and Physiological Psychology* **92**, 1119-1127.

13. Jarrard, L.E. (1983). Selective hippocampal lesions and behavior: effects of kainic acid lesions on performance of place and cue tasks. *Behavioral Neuroscience* **97,** 873-889.

Evaluating MCI in AD Patients and the Effect of Symptomatic Drug Treatment

J. Deguil[1], R. Bordet[2]

[1,2]*Department of Medical Pharmacology, University of Lille 2, Lille, France.* [2]*regis.bordet@univ-lille2.fr*

Alzheimer's disease is a progressive brain disorder that causes a gradual and irreversible decline in memory and cognitive abilities. Until today the pharmacological therapy of AD is still limited to symptomatic temporary improvement or stabilization of cognitive performance and the reduction of neuropsychiatric symptoms of the disease. Five drugs are currently marketed for the treatment of AD including four cholinesterase inhibitors — tacrine, donepezil, galantamine, and rivastigmine — and one glutamate antagonist (memantine). However, owing to the extensive and multifocal nature of neurodegeneration in AD, the effects of transmitter modulators are modest. In recent years, a new therapeutic approach (disease modifying approach) has emerged. Unlike treatment that target symptoms, disease modifying therapies interact on the natural course of the disease by interrupting early pathologic events thus preventing underlying pathophysiologic processes. Although very promising, to date none disease modifying therapies have been clearly shown to be efficacious. In this context, the development of new drugs with symptomatic effect remains necessary.

The clinical development of drugs in Alzheimer's disease has been confronted with challenging methodological difficulties. Taking into account the financial stakes involved taking drug candidates to the phase III stage of development, and the risk of investing time and resources fruitlessly in the evaluation of poor candidate drugs, the crucial decision remains whether to proceed from phase II to phase III (Go/Nogo). The aim of phase II studies is to select a molecule likely to be effective in phase III, but also to eliminate candidate-drugs with an inadequate effect. No consensus currently exists on the best possible design of Phase II studies to inform the Go/Nogo decision optimally. At present time, neuropsychological-based tools are the most established and approved method of assessing outcomes in AD pharmacotherapy in part because they are widely available and do not require technological instrumentation. Because of the difficulties in demonstrating the efficacy of a candidate-drug using only clinical and cognitive tools, development of new assessment tools have become more important over the past few years [1].

Neuropsychological assessment

The Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) represent the most used cognitive scale to evaluate disturbances of memory, language, praxis, attention and other cognitive abilities which are often referred as the core symptoms of AD. Several systematic reviews reported significant improvement in ADAS Cog score in MCI or AD subjects treated with donepezil, reflecting beneficial effects on cognitive status. In contrast, in healthy volunteers (HVT), donepezil induces a slight improvement in the retention of training on complex aviation tasks, verbal memory for semantically processed words and might improve long term visual memory. Moreover, two studies reported transient negative effects on episodic memory and no improvement in the Canbridge Neuropsychological Test Automated Battery (CANTAB), a computer-based cognitive assessment system consisting of a battery of neuropsychological tests.

EEG recording

It was recently reported that the electroencephalogram (EEG) could be a technique to predict the evolution of the Alzheimer disease [2]. **Resting-state EEG** is a suitable technique for monitoring the effects of pharmacological interventions because it is a highly reproducible condition that does not depend on task difficulty and subjects' anxiety and cooperation. Data available in the literature indicated that long- and short-term treatment with donepezil reduced significantly the deterioration of EEG spectral activity and REM sleep EEG disturbance observed in AD patients correlated with cognitive improvement rate on the ADAS-cog. In a small study, it has been demonstrated the ability of a **qEEG variance** combined with a delayed recognition memory task to measure accurately treatment effects on patients with AD. In healthy elderly subjects, a single dose of memantine has been shown to compensate **diurnal vigilance** fluctuation measured by EEG recording.

Another EEG component, the P300 is of special interest as it is related to brain functions such as cognition and attention, which are severely impaired in patients with dementia. In AD patients performing auditory and visual oddball tasks, significant P300 latency changes were observed already during the first month of donepezil administration and were significantly correlated with various neuropsychologic tests scores changes. It has been suggested an effect of donepezil more evident, in the advanced stage of AD. EEG studies seems to confirm the validity of neurophysiological measurements as an additional instrument to evaluate the pharmacologic response to drug in patients affected by cognitive impairment since the effect of donepezil was rapidly and consistently evident. However there is a lack of data on healthy people.

Measurement of imaging markers

Neuroimaging techniques, which can reliably and noninvasively assess of neuroanatomy, chemistry, physiology and pathology, also hold promise as biomarkers. Neuroimaging is based on two different methods: MRI with different acquisition sequences; metabolic imaging by PET scan [3].

The clinical trials, using **structural MRI technique** to quantify the effect of drugs on brain structure volume changes, have confirmed that hippocampus was the most sensitive structure to change in AD patients exposed to donepezil or memantine treatment. These anatomical modifications have been shown to relate closely to neuropathologic and clinical data. Only one study failed to show alteration of hippocampal volume after 2 years of donepezil treatment in AD patients. Other anatomical structures such as white matter and whole brain, have also been studied but did not show any volume changes after donepezil therapy. These findings support the feasibility of using hippocampal atrophy detection by MRI as outcome measures in dementia treatment trials. However, shorter-term effect of these drugs has not yet been investigated in healthy subjects and AD patients.

Functional MRI is uniquely suited for evaluation of cognitive- enhancing agents. In 2 cognitive paradigms of visual memory, donepezil has been reported to produce activation in the ventrolateral prefrontal cortex and in the fusiform gyrus in patients with MCI or AD. In another study, donepezil has been shown to reverse the deficit of activation in fronto parietal region during a working memory task in patients with MCI. Furthermore, in a context of sleep deprivation-induced episodic memory deficit, donepezil enhanced activation of cerebral regions involved in attention and memory encoding processing during a semantic judgement task. Lastly, it has been observed during an auditory attention control task, a decreased of the prefrontal cortex and anterior cingulate cortex activation after memantine administration. Despite the limitations inherent to methodological problems and small sample size of these studies, results point to specific cortical substrates underlying the actions of donepezil and memantine, which can be tested in future studies. It appears as a sensitive imaging marker for very short-term therapy.

The **(18)F-fluorodeoxyglucose (FDG-PET)** technique was used in clinical trial to detect biochemical changes in tissue that precede anatomical changes. A study conducted in resting conditions have demonstrated that treatment with donepezil in AD patients may slow the decline in functional brain activity in right parietal lobe, left temporal lobe and right and left frontal lobes. In addition, it has been showed that the metabolic changes induced by donepezil, during a passive audio-visual stimulation, were limited to the right hippocampus and the left prefrontal cortex and independent of effects on cognitive performance. Finally, in patients treated with memantine for 52 weeks, it has been reported that glucose metabolism in all brain areas studied was preserved longer.

In conclusion, the physiological endpoints mentioned above might constitute a set of surrogate markers of cognitive-enhancing drug response. More sensitive neuropsychological tests or physiological endpoints like imaging and electrophysiological techniques could be relevant as biomarker of pharmacological intervention and could improve the selection of more efficient drug during the development of new compounds.

References

1. Irving, E.A, Upton, N. (2007). Symptomatic treatment of Alzheimer's disease: Identification of biomarkers to aid translation from bench to bedside. *Biomarkers Med* **1**, 93-110.

2. Adamis, D., Sahu, S., Treloar, A. (2005). The utility of EEG in dementia: a clinical perspective. *Int. J. Geriatr. Psychiatry* **20**, 1038-1045.

3. Goekoop, R., Scheltens, P., Barkhof, F., Rombouts, S.A. (2006). Cholinergic challenge in Alzheimer patients and mild cognitive impairment differentially affects hippocampal activation – a pharmacological fMRI study. *Brain* **129**, 141-157.

The Circular Platform Task for Evaluation of Mci in the Grey Mouse Lemur (*Microcebus murinus*), a Non-human Primate Model

A. Rhaman[1], F. Pifferi[1], S. Languille[1], J.L. Picq[2], F. Aujard[1]

[1]*CNRS UMR 7179, MNHN, Brunoy, France.*
[2]*Laboratoire de psychopathologie et de neuropsychologie, EA, 2027, Université Paris 8, St-Denis, France.*
palash20may@yahoo.com

The grey mouse lemur, a non-human primate model of aging

Mouse lemur (Microcebus murinus) is a nocturnal prosimian primate originating from Madagascar with a life expectancy of 8-10 years. The grey mouse lemur presents specific characteristics that make it a good primate model to study and evaluate behavioral and cognitive parameters. In particular they have a small size and weight (80 to 120 g) with an omnivorous dietary habits, and the possibility to assess their behavioral and cognitive performances with specific tasks which have been developed and adapted in our laboratory [1,2]. Some mouse lemurs spontaneously develop cognitive impairments after five years or older [3]. These cognitive impairments have been related in particular with age-related cerebral atrophy [1]. Among the cognitive task available to evaluate mouse lemurs cognitive performances, we use the circular platform task that evaluate spatial reference memory (hippocampus-dependent), and consists in a less stressful version of the Morris water maze.

The circular platform task

The circular platform task apparatus was an adaptation of the device described by Barnes [4] for mouse lemurs. It consisted of a white circular platform (diameter, 100 cm) with 12 equally spaced circular holes (each 5 cm in diameter) located 3 cm from the perimeter. The platform could be rotated. The maze platform was placed 60 cm above the floor, and a cardboard nest box (10 cm × 10 cm × 20 cm) could be inserted and removed beneath each hole and served as a refuge (goal box). To prevent the mouse lemur from escaping, the platform was entirely surrounded with a white wall 25 cm high and covered with a transparent Plexiglas® ceiling that permitted the mouse lemurs to see the extra-maze visual cues. The apparatus was surrounded by a black curtain hung from a square metallic frame (length of the side, 120 cm) located 110 cm above the floor. The roof of the frame was a one-way mirror to allow observation. Attached beneath the one-way mirror and along perimeter of the maze (about 50 cm above the platform), there are 24 evenly spaced 2-W lights, illuminating the maze. The maze center was also illuminated with a 60-W light. Between the one-way mirror and the upper edge of the wall, various objects were attached along the inner surface of the curtain to serve as visual cues. The starting box was an open-ended dark cylinder positioned in the center of the platform. Transparent radial Plexiglas partitions (25 cm high x 20 cm long) were placed between the holes to prevent the strategy used by some mouse lemurs to go directly to the periphery of the platform and then walk along the wall and inspect each hole one by one. Consequently, animals had to return to the center of the platform after each hole inspection.

Animals were given one session of habituation and training (day 1) and one session of testing (day 2). Each session included four trials, each of which began with placement of the animal inside the starting box. After 30 seconds, the box was lifted to release the animal. For the lemurs, the objective was to reach the goal box positioned beneath one of the 12 holes, kept constant relative to the cues for all trials. When the animal entered the goal box, the trial was stopped, and the animal was allowed to remain in its nest box for 2 minutes. After each trial, the platform was cleaned and randomly rotated on its central axis to avoid the use of intra-maze cues, although the position of the goal box was kept constant relative to the cues.

On day 1, trials 1 and 2 consisted of placing the animal in a four-walled chamber containing only the opened goal compartment (one-choice test). For trials 3 and 4, the platform comprised six evenly spaced open compartments (six-choices test). These two trials permitted the animal to explore the maze, observe the visual cues, and further learn the position of the goal box. On day 2 (testing day), 12 compartments were opened during the four trials. Performance was assessed based on the time required for the animal to reach the correct exit (expressed in sec) and the number of errors (wrong quadrant visit) prior to reaching the goal box. Data were expressed as the mean time to reach the correct exit and mean number of errors during the 4 trials of day 2.

Results

The mean number of errors was 1.8 for the young adult group and 6.3 for the older group (see Figure 1). The difference between young and aged was significant (unpaired t-test, p=0.043 ; t=2.216 ; df=14). However, a large individual variability emerged within the older group, with some of the older animals performing as well as younger ones, whereas other older animals were severely impaired. A significant correlation between age and performance was identified ($r = 0.67$, $p < 0.01$).

Conclusions

The circular platform test seems to be adapted to observe significant age-related cognitive impairments in grey mouse lemurs.

Ethics statement

All experiments were performed in accordance with the Principles of Laboratory Animal Care (National Institutes of Health publication 86-23, revised 1985) and the European Communities Council Directive (86/609/EEC). The Research was conducted under the authorization n° 91–305 from the “Direction Départementale des Services Vétérinaires de l'Essonne” and the Internal Review Board of the UMR 7179. All the experiments were done under personal license (authorization number 91–460, issued 5 June, 2009) delivered by the Ministry of Education and Science.

References

1. Picq, J.L. (2007). Aging affects executive functions and memory in mouse lemur primates. *Exp Gerontol.* **42**, 223-232.
2. Dal-Pan, A. Pifferi, F. Marchal, J. Picq, J.L. Aujard, F. on behalf of RESTRIKAL Consortium. (2011). Cognitive Performances Are Selectively Enhanced during Chronic Caloric Restriction or Resveratrol Supplementation in a Primate. *PLoS One* **6**, 16581.
3. Languille, S., Blanc, S., Blin, O., Canale, C.I., Dal-Pan, A., Devau, G., Dhenain, M., Dorieux, O., Epelbaum, J., Gomez, D., Hardy, I., Henry, P.Y., Irving, E.A., Marchal, J., Mestre-Francés, N., Perret, M., Picq, J.L., Pifferi, F., Rahman, A., Schenker, E., Terrien, J., Théry, M., Verdier, J.M., Aujard, F. (2012). The grey mouse lemur: a non-human primate model for ageing studies. *Ageing Res Rev.* **11**, 150-62
4. Barnes, C.A. (1979). Memory deficits associated with senescence: a neurophysiological and behavioral study in the rat. *J Comp Physiol Psychol.* **93**, 74-104.

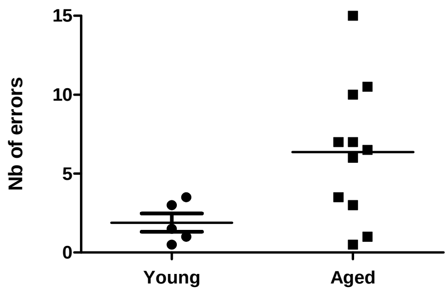


Figure 1. Performance. expressed as the number of errors before finding the correct exit of young and aged adult mouse lemurs in the circular platform task.

The Use of Touchscreens as a New Tool in Mouse MCI Profiling

S.L. Dix[1], S. Billa[1], D. Delotterie[2], C. Dorner-Ciossek[2], J. Gartlon[3], T. Jacobs[4], C.A. Jones[4], L. Lerdrup[5], A. Marti[2], J.C. Talpos[4]

[1]*Eli Lilly, Lilly Research Centre, Windlesham, UK. (Dix_Sophie@Lilly.com, Billa_sophie@network.lilly.com*

[2]*Boehringer Ingelheim Pharma, Dept. of CNS Diseases Research, Biberach, Germany, david.delotterie@boehringer-ingelheim.com, cornelia.dorner-ciossek@boehringer-ingelheim.com, anelise.marti@boehringer-ingelheim.com*

[3]*Eisai Ltd., European Knowledge Centre, Hatfield, UK, Jane_Gartlon@eisai.net*

[4]*Janssen Research & Development, Beerse, Belgium. tjacobs1@its.jnj.com, cjones34@its.jnj.com, jtalpos@its.jnj.com*

[5]*Synaptic transmission, H. Lundbeck, Valby, Denmark, lmat@lundbeck.com*

Introduction

Operant chambers equipped with touch-sensitive screens ('Touchscreens') are becoming an increasingly popular technique for the development of novel paradigms for use in cognitive testing [1]. Here, we describe techniques and pilot data generated by members of the PharmaCog touchscreen working group within the Innovative Medicines Initiative. PharmaCog is dedicated to establishing imaging, biochemical, electrophysiological and behavioural translatable biomarkers for use in research into Alzheimer's disease (AD) and mild cognitive impairment (MCI). The touchscreen working group, comprising members from 5 pharmaceutical companies, is tasked with harmonising and validating touchscreen protocols for use in mice. The first large cross-site study aims to assess the reproducibility of pharmacological effects using donepezil, memantine and scopolamine on visual discrimination (VD) performance in trained young c57/bl6 mice. We will then compare the VD acquisition performance of three transgenic mouse models of amyloid pathology in Alzheimer's disease: PDAPP (APP_{717}); Tg TASTPM (double mutant APP_{swe} x $PS1_{M146V}$); Tg APP-PS2-Tau (triple mutant APP_{swe} x $PS2_{N141I}$ x Tau_{P301L}), and age-matched wild-type C57BL/6J control mice. We are in the process of comparing protocols to assess paired associate learning in these mice.

Visual Discrimination

All experiments were performed in compliance with the European Commission Directive 86/609/EEC and had approval by the local ethical review committee at each laboratory. In a visual discrimination task, the rodent is presented with pairs of stimuli [2]. The same pair is used throughput the study and one of the stimuli is deemed the correct response and is given an appetitive reward (food pellet or sucrose - dependent on the hardware). If the animal makes an incorrect response there is a timeout period in which the house lights are turned off and no reward is given. If the animal fails to respond within x sec then it is counted as an omission, and the next trial begins. Prior to VD training the animals are trained to respond to a white box on the screen using a 4-stage procedure. Pilot strain-comparison data showed that DBA/2J and 129S mice were slower to learn to respond to the screen than CD1 and c57/bl6 mice. However the DBA/2J mice were subsequently faster to acquire the VD, then the 3 other strains.

The cross site study will start in April and be conducted within three laboratories. Male c57/bl6 mice will be trained to criterion in the VD task. They will then be treated with donepezil, memantine and scopolamine (serially) to assess the reproducibility of data across sites and the impact of AD relevant pharmacology on performance in trained animals. To this aim, we attempted to harmonise the protocols wherever possible. We agreed on independent variables such as the inter-trial interval (10s), use of correction trials (to be included), length before omissions (30s), time-out duration (10 s), trial number (30) and session duration (45 min) by consensus. Stimuli will be presented in two apertures cut in a black mask covering the touchscreen sized 5 x 7.5 cm, and change locations from trial to trial in a counter-balanced fashion We performed pilot studies to determine the appropriate stimuli to use for further studies. Initial data from two sites showed strong preferences

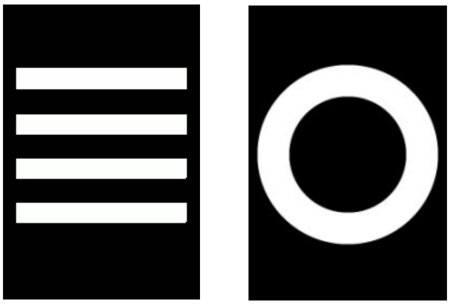


Figure 1. Final stimuli used in the visual discrimination studies.

A+	C	
A+		B
C	B+	
	B+	A
	A	C+
B		C+

Figure 2. The 6 possible permeations of stimuli using three stimuli (A, B and C) presented in 2 of 3 locations within a PAL task. '+' indicates the correct response to gain an appetitive reward.

for certain images, but that performance was matched when 'lines' and 'circle' were used (see Figure 1). These findings were confirmed within two additional laboratories and selected as the stimuli to be used.

We will also profile the three transgenic lines selected for use within PharmaCog using the harmonised VD protocol. Preliminary data has shown that aged male TASTPM (15 – 18m) were unimpaired in the VD acquisition. Indeed the mice showed enhanced learning compared to age-matched wild type c57/bl6 mice, and were comparable in their performance to young (6m) male c57/bl6 mice. Using a different amyloid transgenic line, tg2576, aged female mice (13m) were again shown to have enhanced performance relative to wild type control animals. This study also examined the pretraining regimen used to train animals to respond to the screen. The results suggested that lengthy pretraining protocols introduced variance into the subsequent VD acquisition data.

Paired Associates Learning

The second task that the team is exploring is paired associates learning (PAL). Poor PAL performance in MCI patients has been shown to be predictive of AD pathology [5]. In this task, a subject must remember not only the visual stimulus that he has been presented with, but also where it was located ('what' and 'where'). A rodent version of the task has been developed using touchscreens in both in rats [3] and mice [4]. In this paradigm an animal is presented with, typically, 2 stimuli in 2 of 3 apertures within a touchscreen mask. An animal learns that it must respond to. stimulus A, e.g., when it is in location 1 but not when it appears in locations 2 or 3 (Figure 2). Hence the animal must learn to associate a visual stimulus with a specific location ('what' and 'where'). It is also possible to increase or decrease the possible combinations of stimuli and location by increasing the number of apertures (i.e. possible locations) and / or changing the number of stimuli that are presented animal is presented with. Studies are ongoing to optimise the number of stimuli and locations to detect deficits in AD transgenic models. PDAPP mice show acquisition deficits in a simple version of the task in which 2 stimuli are presented within 4 locations (4 permutations) compared to age-matched wild type littermate control animals.

Discussion

Touchscreens are a relatively new technology that confers significant advantages over traditional lever-equipped operant chambers due to the range of stimuli that can be presented to the animal. Furthermore, there is translational potential to have analogous paradigms across species including humans. This large collaboration demonstrates the clear advantages inherent in the consortium approach. This multisite series of experiments will allow for novel protocols to be further developed, harmonized and validated rapidly with the aim of identifying translatable paradigms that are sensitive to AD pathology in transgenic mice and hence can be used in drug development and research into MCI and AD.

VD is simple task and we did not anticipate a learning deficit in these mice. If the selective deficit in PAL in the amyloid transgenic animals is found to be robust, this would provide strong validation for the use of this paradigm in AD research. It would demonstrate that amyloid pathology can produce a selective deficit in PAL acquisition rather than a non-specific deficit in touchscreen-based learning *per se* which could potentially reflect a non-cognitive phenotype such as visual deficit. The final stage of this project will be to assess drugs currently used for the symptomatic treatment of AD (donepezil and memantine) and amyloid lowering compounds against performance deficits in mouse models of Alzheimer's disease in touchscreen tasks.

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement No 115009, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA's in kind contribution (www.imi.europa.eu).

References

1. Bussey, T.J., Holmes A., Lyon L., Mar A.C., McAllister K.A.L., Nithianantharajah J., Oomen C.A., Saksida L. M. (2012). New translational assays for preclinical modelling of cognition in schizophrenia: The touchscreen testing method for mice and rats. *Neuropharmacology* **62**, 1191-1203.

2. Talpos J.C., Fletcher, A.C., Circelli C., Tricklebank, M.D., Dix S. L. (2012). The pharmacological sensitivity of a touchscreen-based visual discrimination task in the rat using simple and perceptually challenging stimuli. *Psychophrmacology,* DOI 10.1007/s00213-011-2590-z.

3. Talpos J.C., Winters B.D., Dias R., Saksida L.M., Bussey T.J. (2009). A novel touchscreen-automated paired-associate learning task sensitive to pharmacological manipulation of the hippocampus: a translational rodent model of cognitive impairments in neurodegenerative disease. *Psychopharmacology (Berl)* **205**, 157-168.

4. Bartko, S.J., Vendrell I., Saksida L. M., Bussey T.J., (2011). A computer-automated touchscreen paired-associates learning (PAL) task for mice: impairments following administration of scopolamine or dicyclomine and improvements following donepezil. *Psychopharmacology (Berl)* **214**, 537-548.

5. Swainson R., Hodges J.R., Galton C.J., Semple J., Michael A., Dunn B.D., Iddon J.L., Robbins T.W., Sahakian B.KJ. (2001) Early detection and differential diagnosis of Alzheimer's disease and depression with neuropsychological tasks. *Demnt. Deriatr. Cogn. Disord.* **12**, 265-280.

New Insight for the Study of Mild Cognitive Impairment: The Novel Object Recognition Task and the Single Day Morris Water Maze in Total Sleep Deprived Rats

M. Pellitteri*, V. Colavito, E. Moscardo, F. Schio, B. Salvetti, M. Bentivoglio, P.F. Fabene, G. Bertini

Department of Neurological, Neuropsychological, Morphological and Motor Sciences, University of Verona, Italy.
**michele.pellitteri@univr.it*

Introduction

Mild cognitive impairment (MCI) is a brain disorder in which cognitive abilities are mildly impaired. Subjects with MCI are of major clinical importance because they have an increased risk of developing Alzheimer-type dementia. Appropriate animal models are necessary in order to understand the pathogenic mechanisms of MCI and develop drugs for its treatment. Although MCI is associated with a variety of symptoms, memory deficits are a dominant feature of the syndrome and is essential to select appropriately difficult behavioural tasks in order to point out the deficits can be associated with neuropathological alterations. We are all aware of the detrimental effects of sleep loss. Preventing or disrupting sleep after the acquisition of new information have been repeatedly shown to impair the recollection of memories. Few studies have investigated, however, the role of sleep prior to the learning experience in affecting subjects' ability to acquire new information. We examined the effect of 6h of total sleep deprivation (TSD) by gentle handling immediately before the learning process, in two different mnemonic tests: the novel object recognition task (NOR) and the single-day Morris-water-maze task (1d MWM).

Materials and Methods

We assessed recognition memory in Sprague-Dawley rats (4 months old) after 6hrs of total sleep deprivation by gentle-handling performed at 2 different periods of the resting phase, respectively during the first 6 hours (0-6h) and the last 6 hours (7-12h) after lights on. Both time points were tested in each of the subjects. Tests were carried out 3 weeks apart. Immediately after sleep deprivation, during the acquisition phase, animals were allowed to familiarize with 2 identical objects for 10 minutes in PhenoTyper cages (Noldus Information Technology, Wageningen, The Netherlands). To test recognition memory, after a 1h delay, animals were presented with 2 objects (recognition phase): one was identical to the previously presented objects (Familiar object), the other one was never seen before (Novel object). During the delay time animals were placed in their own home cages. Automated tracking was performed using the EthoVision® XT software (Noldus Information Technology, Wageningen, The Netherlands).

The distance moved by each animal in the cage (cm) and the duration (s) of when the nose of the animals was tracked within a 2 cm area around the objects were recorded. Recognition memory was inferred by comparing the time spent exploring the novel object vs. the familiar one, during the recognition session. Object type and position of the novel object were randomized.

Furthermore, we focused our studies evaluating the effects of TSD in a spatial learning task. A modified experimental classic protocol of the Morris water maze was adopted. A circular pool with a diameter of 130 cm and a height of 90 cm, filled with 20±1 °C water to a depth of 60 cm, was placed in a completely dark room. The maze was divided into four equal quadrants and release points were designed at quadrant as NW, SW and SE. A hidden circular platform (11 cm in diameter), was located in the centre of the NE quadrant, submerged 1.5 cm beneath the surface of the water. In order to increase the colour contrast between the water and the animal, a nontoxic paint was used to get the water darker. Fixed, extra maze visual cues were present at various locations around the maze. The task requires rats to swim to the hidden platform guided by distal cues. Immediately after 6h hours of total sleep deprivation (0-6h), rats were subjected to three blocks of 3 trials each, with an inter-trial time of 1h. Three different starting positions were equally distributed around the perimeter of maze and randomized within each block. After reaching the platform, rats were allowed to remain there for 30 s until the start of the next trial. The animals were given a maximum of 120 s to find the platform; if they failed to find the platform in this time, they were placed by the experimenter on the platform and allowed to stay there for 30 s. After completion of each block of trials, the animals returned to their home cage. 1h after the end of the 3 blocks, an extra 120 sec trial (probe trial) was performed. Platform was removed and time spent in each quadrant was recorded. A camera was mounted above the center of the maze and automated tracking was performed using EthoVision® XT software. For all the trials, time (s) spent to find the platform (latency) and total travelled distance were measured. Moreover the time spent in each quadrants was recorded during the probe trial.

Results and Discussion

Regarding the NOR task, we found that locomotor activity was not affected by SD protocols. On the other hand, sleep deprivation impaired object recognition when it occurred during the first 6 hours of the light phase. When TSD was postponed to the last 6 hours of the light cycle, it had no effect on cognitive performance. Control rats behaved similarly both when the test was administered in the light and in the dark phase.

In the spatial learning task (1d MWM), learning occurred in both animal groups; animals during the probe trial spend majority of the time swimming within the quadrant where the platform was previously placed. The mostly interesting point is that total sleep deprived animals didn't learn the position of the hidden platform as fast as the controls animals.

These findings show that preventing the initiation of sleep at the end of a normal period of activity (i.e., when homeostatic sleep pressure is high) impairs subsequent learning. 1) TSD has been proved to be a valid challenge in order to induce acutely MCI in rodents; 2) we demonstrated that both NOR and single-day MWM can detect TSD-induced MCI; 3) Pro e cons of the two different techniques are weighted. Therefore these behavioural tests allow to study the role of drugs during the different stage of the memory shaping.

Ethical statement

The experiments received authorisation from the Italian Ministry of Health, and are conducted following the principles of the NIH Guide for the Use and Care of Laboratory Animals, and the European Community Council (86/609/ EEC) directive. Rats are housed and used according to current European Community rules. Experiments on rats are approved by the research committees from the University of Verona and Italian National Institute of Health

References

1. Ennaceur, A., Delacour, J. (1988). A new one-trial test for neurobiological studies of memory in rats. 1: Behavioural data. *Behavioral Brain Research* **31**, 47-59.

2. Morris, R.G. (1984) Developments of a water-maze procedure for studying spatial learning in the rat. *J Neurosci Methods* **11**, 47-60.

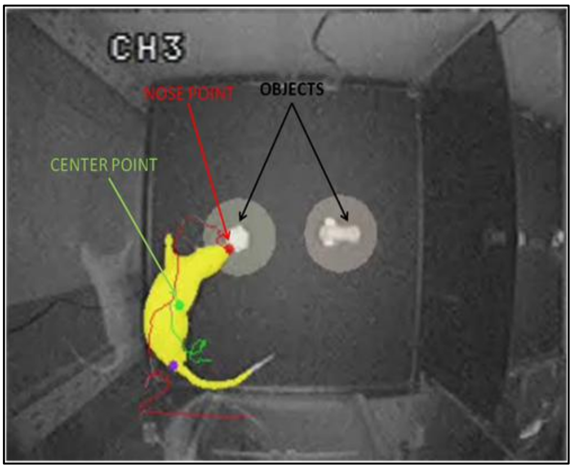


Figure 1. Automatic tracking of a rat during the recognition phase in the NOR test. The time spent by the animal with the nose within the virtual circular area drawn around the objects was counted as exploration time.

Whole Body Vibration and Spatial Learning: c-Fos and ChAT as Neuronal Correlates of Cognitive Improvements

E.A. Van der Zee[1], G. Riedel[2], F. Postema[1], B.J. Venema[1], J.N. Keijser[1]

[1]*Dept of Molecular Neurobiology, University of Groningen, Groningen, The Netherlands. e.a.van.der.zee@rug.nl*

[2]*Inst. Med. Sci., University of Aberdeen, UK*

It has been well established that physical activity (exercise) by way of running wheel or treadmill activity improves motor and cognitive performance in (sedentary) mice [1,2,3]. This type of exercise, however, cannot be performed well by aging mice or mice with motor deficiencies. Notably the amount of voluntary exercise of old mice in running wheels decreases largely (up to 90% as compared to young mice). Here, we aimed to find an alternative type of exercise and examined whether whole body vibration (WBV) as a form of passive exercise can improve motor and cognitive performance. WBV stimulates the entire body, including brain, via controlled vibrations as used in human power plates. A scale model was build for mice, by which different frequencies and g-forces could be applied to mice [4]. The mice were placed in a cage which was directly connected to the vibrating platform of the vibrator located in the middle of the cage. Mice received WBV for 5 weeks, 10 minutes per day, week days only. Pseudo-WBV was performed by placing the mice in the cage in the absence of the actual vibrations. WBV continued during the days of testing, but was always performed after the behavioral testing was done at the end of the day in order to prevent acute effects of WBV on behavior. No signs of distress or anxiety were observed during the 10 minutes of WBV treatment. In NMRI mice, three frequencies (20, 30 and 45 Hz) and two g-forces (0,2 and 1,9 g) were compared for the capacity to improve motor performance, using rotarod, hanging wire and balance beam as tests. Best motor performance was found at 30 Hz and 1,9g [4]. Thereafter, these settings were used as standard, and the balance beam test the standard test for motor performance. Next, C57BL6 or CD1 mice (3 or 24 months of age) received these mild vibrations. WBV improved sensory-motor abilities as measured in the balance beam test, irrespective of age and strain. Learning and reversal learning in a Y maze was used as a spatial learning task. Results in C57Bl6 mice showed a significant improvement in Y-maze learning (rate of acquisition), but no improvement in Y-maze reversal learning [3]. A direct comparison of WBV, running wheel and treadmill activity revealed that the cognitive improvement in the Y maze by WBV, although somewhat less, was of a similar degree as induced by the active forms of exercise. WBV in CD1 mice failed to show significant improvement in the Y maze, in contrast to the improvement of motor performance. The improvements in motor performance due to WBV developed at a slower in old than young subjects. To examine the neurophysiological brain changes induced by WBV treatment, c-Fos and ChAT immunoreactivity was measured in experimentally naïve young C57Bl6 or CD1 mice after 1, 3, and 5 weeks of WBV, and one day after 5 weeks of WBV. Pseudo-WBV treatment served as control. The results showed that hippocampal c-Fos expression gradually and strongly increased in C57Bl6 mice after 1, 3 or 5 weeks, but only marginally in CD1 mice. c-Fos expression was back to baseline values one day after the last WBV session. The pattern of c-Fos expression in the brain, with areas reacting in varying degree from highly activated to no change, suggests that sensory stimulation via the whiskers is a primary pathway by which WBV stimulates the brain [5]. Strong increases in c-Fos staining were observed in hippocampal CA1 and CA3 regions and the prefrontal cortex, among others. In contrast to c-Fos, no significant WBV-induced changes were found for Arc or Egr1/zif268. Immunoreactivity for choline-acetyltransferase (ChAT), the acetylcholine synthesizing enzyme, was measured in the nucleus basalis, medial septum, hippocampus, neocortex and amygdala (C57Bl6 mice only). No WBV-induced changes were found in the nucleus basalis, whereas significant increases were found in the medial septum. The WBV-induced increase in the medial septum was lost one day after the last WBV session, but remained present in hippocampus, neocortex and amygdale. The ChAT data suggests increased cholinergic responsiveness due to WBV, like described for physical exercise [6]. Significant changes in the hippocampal ChAT staining, probably contributing to cognitive improvements, developed slightly slower than the improvements in motor performance. This suggests that WBV is slightly more effective in improving sensory-motor abilities than spatial learning abilities. Taken together, these findings indicate that WBV as a form of passive exercise is suitable for improving cognitive performance in young and old subjects. It also suggests that c-Fos and ChAT are WBV-specific neuronal correlates serving a ‘priming’ of the septohippocampal circuit for improved acquisition of spatial learning. A next step.we made was to study the effect of WBV on cognitive performance in humans. Recent studies at our University within a consortium of Life Sciences, Movement Sciences and Neuropsychology revealed that WBV significantly improves cognitive performance (Stroop test). Taken together, our data clearly shows the potential of WBV as an intervention to improve motor and cognitive performance in subjects not able (or willing) to perform active forms of exercise. This is of particular interest to aging human subjects in need of an alternative, non-aerobic type of exercise.

References

1. Van Praag, H., Christie, B.R., Sejnowski, T.I., Gage, F.H. (1999) Running enhances neurogenesis, learning, and long-term potentiation in mice. *Proc. Natl. Acad. Sci. USA* **96**,13427-13431.

2. Fisher, B.E., Petzinger, G.M., Nixon, K., Hogg, E,, Bremmer, S., Meshul, C.K., Jakowec, M.W. (2004) Exercise-induced behavioral recovery and neuroplasticity in the 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine-lesioned mouse basal ganglia. *J. Neurosci. Res* **77**, 378-390.

3. Van der Borght, K., Havekes, R., Bos, T., Eggen, B.J., Van der Zee, E.A. (2007) Exercise improves memory acquisition and retrieval in the Y-maze task: relationship with hippocampal neurogenesis. *Behav. Neurosci.* **121**, 324-334.

4. Timmer, M., Van der Zee, E.A., Riedel, G. (2006) Whole body vibration and behaviour: investigation of the role of various neurotransmitters. *FENS*, A 089.31.

5. Van der Zee, E.A., Riedel, G., Rutgers, E.H., de Vries, C., Postema, F., Venema, B.J., Keijser, J.N., Luiten, P.G.M. (2010) Enhanced neuronal activity in selective brain regions of mice induced by whole body vibration. *FENS*, 024.49.

6. Fordyce, D.E., Farrar, R.P. (1991) Enhancement of spatial learning in F344 rats by physical activity and related learning-associated alterations in hippocampal and cortical cholinergic functioning. *Behav. Brain Res.* **46**,123-133.

SPECIAL SESSION

The Role of Behavior Measurement in Persuasive Settings

Chair: Robbert Jan Beun

Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands.
r.j.beun@uu.nl

Abstract

The goal of this special session is to investigate the role of sensor and behavior measurement in persuasive systems, in particular systems that are aimed at adherence to exercises and sustainable behavior change of humans, such as virtual coaching and training systems.

Persuasive settings can be characterized as communicative contexts in which it is common ground that one participant intentionally tries to change the behavior of another participant. Persuasive settings are dominant in health, education and commerce: a doctor tries to convince a patient to take the prescribed medication, a packet of cigarettes ‘warns’ a smoker for the lethal effects of smoking, and a salesperson tries to sell a particular product for a particular price. Persuasion often incorporates elements of resistance and contradiction: resistance, because without support of the persuader the individual will not show the desired change; contradiction, because despite the knowledge that a particular behavior supports an individual’s well-being, something withholds the individual to perform the desired behavior in practice. Patients, for instance, may find exercises too strenuous or may simply forget the medication, smokers may lack motivation to quit smoking and the salesman’s price may be too high. Persuasion is, therefore, a precarious process where timing, modality and content of messages should be carefully chosen to avoid the individual’s withdraw and a premature stop of the communication process.

Currently, an interest emerges in developing automated systems that provide behavior changing support to patients and consumers without human interference. In these systems, methods from so-called persuasive technology are applied to implement effective communication strategies that support self-care, adherence to prescribed exercises and sustainable behavior change. These systems provide for anonymous and frequent monitoring, feedback and counseling that would otherwise be impossible. Modern mobile technology and web applications enable the delivery of advice and interventions in the appropriate form and modality, at the appropriate time, location and device; wearable sensors enable continuous measuring of relevant momentary information. In persuasive settings, tailoring of information plays a central role. Superfluous interruptions, erroneous messages and irrelevant information have to be avoided as much as possible. As a result, these systems should become familiar with a blend of characteristics of the individual and its environment: the current activity, daily behavior, ability to perform an exercise, physical circumstances, and so on. In other words, a prerequisite for the implementation of adequate persuasive strategies is a strong awareness of the characteristics and activities of the individual. Basically, there are two channels that provide the system of this information: a sensory channel that enables automated measuring of behavior, biosignals and environmental data, and a symbolic channel for the exchange of intentionally produced messages such as natural language.

Automated measuring of behavior by sensor information may have some important advantages over symbolic messages. First, data can be collected in a more objective and reliable way than when intentionally entered by human individuals (e.g. stress related or smoking behavior); second, the individuals need not to focus on the measurement process and can, therefore, avoid the tedium of tracking their own performance (e.g. a step-counter) and do not have to worry about the timing of the measurement (e.g. in cases of emergency); third, it enables individuals to become aware of otherwise unobservable behavior (e.g. sleep activity); fourth, frequent feedback of a device often provides motivation to perform the desired activity. As a result, the process may give the persuaded individual, the persuasive system and third parties relevant information for diagnosis, understanding and exercise performance and so substantially improve the quality of the interaction.

This special session is intended to bring together a group of researchers in the field of automated coaching and other behavior changing applications. Questions concentrate on, but will not be restricted to, sensor types and behavioral measurements in well-known persuasive settings, the reliability of behavioral measurement, the influence of the behavioral information on the interaction (in particular the timing, modality and content of messages), and the integration of symbolic input and sensor data. In this session, various persuasive applications will be discussed, ranging from automated sleep therapy to sustainability and activity coaching.

SPECIAL SESSION CONTENTS (sorted by paper ID)

Inter-usability and the Presentation of Multi-modal Feedback for Physical Activity and Diabetic Type II Patients

Randy Klaassen, Rieks op den Akker

University of Twente, Enschede, The Netherlands. R.klaassen@utwente.nl, h.j.a.opdenakker@utwente.nl

Context

In the EU Artemis project Smarcos we developed a personal behaviour change support system that supports users in attaining a healthy lifestyle. The coach sends timely, context-aware feedback about daily activities through a range of interconnected devices. The system is designed for two targeted user groups: Diabetics Type 2 patients and office workers. The Diabetics group gets feedback about their medication intake, activity level, glucose level and food intake. The office workers receive feedback about physical activity and food intake. The feedback to the users of the system consists of personalized reminders, advice and tips, or learn more about why healthy behaviour is important.

The prototype system combines two existing (single-device) behaviour change support systems. The multi-device service system we present here extends the single device feedback systems bringing them on one integrated service platform. It allows multiple contact points with users in various contexts: at home, at work, on the road and the system is able to gather more context information of the user. Interventions with several contact points are expected to be more effective in stimulating change in health behaviour than those that use a single contact points [5]. The input devices in the prototype are a smart pill dispenser and a 3D accelerometer. The pill dispenser is connected to a server via the GSM network. Every time the user opens the pill dispenser it sends a message to the server. The accelerometer is measuring the amount of physical activity all day long. The user must connect the accelerometer to his computer to upload the data to the server. The output devices of the prototype system are smartphones (Android and iOS), desktop/laptop computers and (smart) TV's. Smartphones can act as an input and output device. Smartphones can gather information about location of the user, can ask for self-report and are able to receive feedback messages from the system.

Feedback

The input devices provide the system with context information of the user about medication use, physical activity and (semantic) location. The system knows for all users their medication moments, the number of pills and the time when they should take the pills. The system also knows the personal and daily activity goal of all the users. Smartphones provide the system with location information and self-report. The user can connect his devices to the coaching system so the system knows what kind of input and output devices the user can use.

The system receives and evaluates updates on the user context and compares the new data with the conditions of the coaching rules. When one of the rules is fired, the system should send feedback to the user. When a message should be presented four aspects are important; the selection of the device, the timing of the message, the content of the message and the modality in which the message is presented to the user.

Device selection. The system should select the best available device on which it can present the feedback. Based on the available devices and the location of the user the best device can be selected. If the user is at home, the television is maybe the best available device. When the user is on the road the only available device can be the mobile phone of the user.

Content. The content of the feedback can differ based on the context information of the user. Feedback messages can be a regular (daily or weekly) report about physical activity, calories, or medication intake. Or it can be a reminder, warning, an advice, an assessment, or maybe a message that asks for specific information from the user. Variation of feedback messages is important in order to prevent annoyance. Bickmore et al. reports: "One surprising finding from the interviews was that, even though the dialogue scripts had been authored to provide significant variability in each days' interaction, most participants found the conversations repetitive at some point during the month. This repetitiveness annoyed subjects, and a few subjects even indicated that it negatively impacted their motivation to exercise..." [2]. It is also important not to send too many messages to the user. It can be possible that some coaching rules can fire several times a day; in this case the system should send only a maximum number of messages.

Timing. When the system decides that it is time to send a feedback message to the user, it should decide when to send the message. The timing of the message is dependent on the context of the user and the content of the message. Warnings about medication intake should be send as quick as possible, but a presentation of the weekly report of physical activity can wait until the user is at home (and the television is on, so the system can present the overview on a large screen).

Modality of presentation. Feedback can be presented using different modalities. The possible formats in which the system can present feedback to the user are a spoken animated graphical user, simple text messages, a graph showing an overview, text that refers to a picture or sound or a tune that signals the user that he has forgotten to take his pills. These modalities are available on desktop, mobile phones and television.

User centered design

The coaching system as described in this paper is designed in a user centered design process. In order to provide feedback at the right time and using the right modality it is necessary to take into account the context of the user. Feedback models define the interactions between system and the user. To capture the requirements for the system, two context-mapping studies have been conducted. Context mapping is a procedure for conducting in-depth research with users. The objective of this procedure is to gain knowledge about the use of products from implied and implicit information that is provided by prospective users during an intensive process [6]. Through active participation by users and stakeholders during the design process, it informs and inspires design teams and ensures a good fit between the design and the use of a product. These qualitative studies consisted of a diary booklet with exercises and a semi-structured interview. Over a period five days, users were probed to describe their daily activities and reflect on their behaviours relevant for the system based on small assignments. The study provides insights in the actual behaviour of the users, awareness of the importance of healthy behaviour and their barriers to a desired behaviour. The requirements collected with this approach were further refined by a role-playing exercise during a workshop.

When the initial requirements were defined two online questionnaires were provided to find out in which situations users would like to receive what kind of feedback on which device. The questionnaires were also used to assess the clarity of the icons that were used in the interface, and to get an idea of how the participants perceived the presented concept.

During a six-week user evaluation of a physical activity coaching application two alternatives were compared for providing digital coaching to users of a physical activity promotion service. Participants received personalized feedback on their physical activity levels for a period of six weeks. Feedback was provided weekly either by e-mail or through an embodied conversational agent (ECA). User's perception of the digital coaching was assessed by means of validated questionnaires after three weeks and at the end of the study. Results show significantly higher attractiveness, intelligence and perceived quality of coaching for the ECA.

The results of the previous studies were incorporated into the first prototype of the coaching system. New, long term user evaluations are planned to investigate the overall user experience of the coaching program and the effectiveness of the coaching program when feedback is presented using different presentation modalities and output devices. The effectiveness of the coaching system will be measures by logging the actual behaviour of the participants (the level of physical activity, the medication intake and the progress toward the personal goal of the user, therapy adherence and retention rate). The overall user experience [1,4] and quality of coaching [3] will be measured using validated questionnaire before, halfway and at the end of the user evaluation.

References

1. Bartneck, C., Kulić, D., Croft, E., Zoghbi, S. (2009). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* **1**, 71-81.

2. Bickmore, T.W., Picard, R.W. (2005). Establishing and maintaining long-term human-computer relationships. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction* **12**, 293-327.

3. Cote, J., Yardley, J., Hay, J., Sedgwick, W., and Baker, J. (1999). An exploratory examination of the coaching behavior scale for sport. *AVANTE* **5**, 82-92.

4. Hassenzahl, M. (2005) The thing and I: understanding the relationship between user and product. *Human-Computer Interaction Series* **3**, 31-42.

5. Noar, S., Benac, C., and Harris, M. (2007). Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. *Psychological Bulletin* **133**, page 673-693.

6. Sleeswijk Visser, F., Stappers, P.J., van der Lugt, R. (2005). Contextmapping: experiences from practice. *CoDesign: International Journal of CoCreation in Design and the Arts* **1**(2), 119-149.

It's LiFe!: A Monitoring- and Feedback Tool to Stimulate Physical Activity, Embedded in Primary Care

S. van der Weegen[1], R. Verwey[1], M.D. Spreeuwenberg[1], H.J. Tange[2], T. van der Weijden[2], L.P. de Witte[1]

[1]Department of Health Services Research, CAPHRI school for primary care and public health, Maastricht University, Maastricht, The Netherlands. S.vanderweegen@maastrichtuniversity.nl

[2]Department of General Practice, CAPHRI school for primary care and public health, Maastricht University, Maastricht, The Netherlands

Introduction

Physical activity improves the long-term prognosis and quality of life of people with a chronic disease like COPD or type 2 diabetes. Therefore, care standards state that stimulating physical activity is a central element in the treatment of those patients in primary care [1,2]. Although there are many initiatives to increase the level of physical activity, few patients manage to be sufficiently active. This may be because care standards also state that the practice nurses and/or GPs see these patients only one to four times a year. Consequently, the level of success regarding this element depends for the greater part on the degree in which patients succeed to execute their self-management role. Therefore, primary care providers should involve patients in self-management decisions and seek together with the patient for lifestyle interventions that fit with the motivation, needs and capabilities of the patient [3,4]. In addition to the support of the care provider, persuasive technology might play an important role in accomplishing the behaviour change, adherence to the new behaviour and the self-management role. A big advantage of persuasion by technology over human persuasion is that technology is more persistent and can go where humans cannot [5]. Through this, personalised feedback can be given on individual performance in relation to goal setting. Self-monitoring of physical activity using a pedometer or an accelerometer has been identified as an effective approach towards behaviour change [6,7]. If embedded in primary care, the combination of human persuasion and persuasive technology could complement and reinforce each other, especially when the two are aligned carefully. In the current project It's LiFe! an innovative monitoring- and feedback tool is being developed and tested, which will be embedded in a Self-management Support Program that is being developed together with this technology. As a prerequisite for useful technology and a successful intervention that meets the requirements and preferences of the end-users, it is important to involve the end users in the design process at an early stage.

In the present study, potential end-users were interviewed in order to answer the following research questions: 1. How the tool could be designed in order to be attractive, easy to use and suitable to wear on a daily basis; 2. Which feedback patients need to optimally support them in their self-management; 3. How this feedback should be presented; 4. How the use of the monitoring and feedback tool could be integrated in a Self-management Support Program (SSP) that is based on the principles of patient involvement, shared decision making and current insights into disease management and the chronic care model. This paper describes the designed intervention and the user-centred design process, which has led to this intervention.

Methods

A user-centred design method was applied to develop the monitoring- and feedback tool and the SSP. The procedure of this method is depicted in Figure 1. Defining the user requirements (phase A) was an iterative process starting with user and context identification (stage 1). Subsequently, the conceptual idea was developed in collaboration with several experts and business partners. The conceptual idea was described in a use case from the perspective of a patient with COPD, a patient with diabetes and a practice nurse (stage 2). A use case is a description of steps or actions between a user and a software system [8]. In stage 3 user requirements were elicited in seven open interviews with patients and semi-structured interviews with seven patients and fifteen health care professionals (general practitioners, physiotherapists, nurse specialists in diabetic and pulmonary care and practice nurses). Patients were asked where they wanted to wear the activity monitor and what other

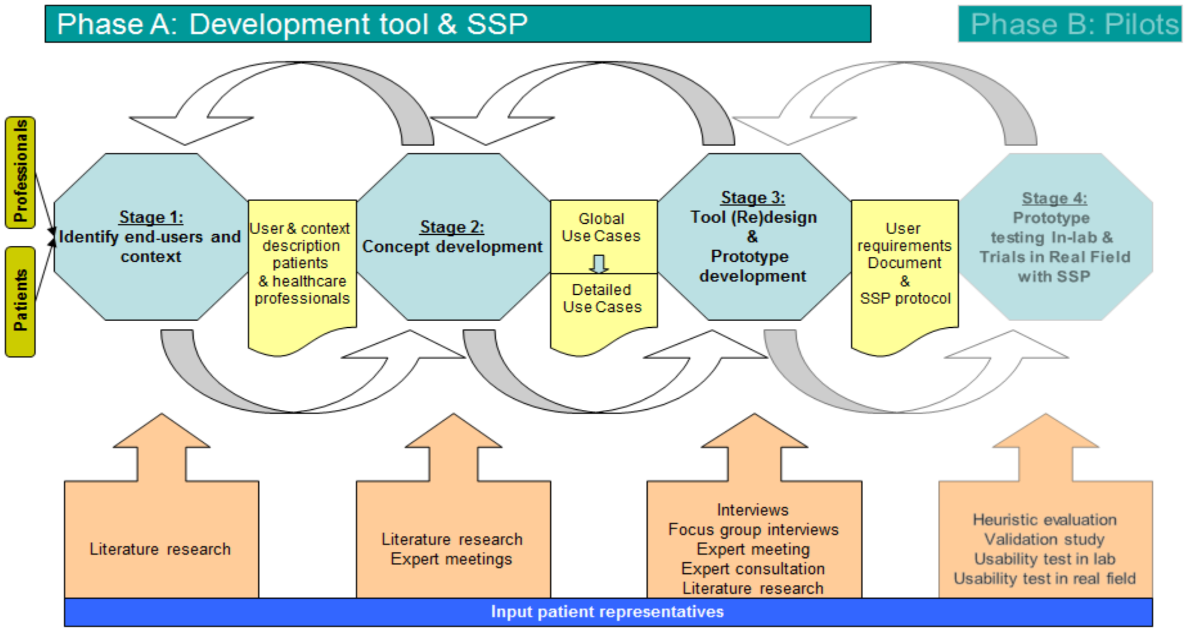


Figure 1. Methodological model.

requirements they had for the activity monitor. Also, on what device they want to see their activity data, in what unit and format and how the care professional could be involved. Care professionals were asked which data they needed to support the patients in their self-management and in what format they wanted to receive this. In the patient group, two focus group discussions were used to complement and confirm the gathered data. In the care professionals group experts from different fields were asked to comment on the care program developed. Patients were included until data saturation was reached. During the whole process two patient representatives were present at most research meetings to represent their patient group and there was continuous collaboration with the engineering team. The technological development took place in collaboration with two companies: Maastricht Instruments Ltd and Sananet Care Ltd. The research team of Maastricht University conducted the research and 'fed' the engineering team with information about user requirements for the tool, so that gradually the elements of the technology evolved.

Results

The development process has led to an innovative monitoring- and feedback tool and a SSP, which meets the requirements of the end-users. The requirements from the end-users where determining factors in the design of the devices for activity monitoring and feedback, the feedback strategy and the number of consultations with the practice nurse.

Tool design

The It’s LiFe! tool consists of three different elements:

1. a 3D accelerometer worn on the hip;
2. an application (app) on a Smartphone;
3. a web application and server called ‘It’s LiFe! Online’ (powered by Sananet).

The transmission of data from the accelerometer towards the app is via a Bluetooth connection. Every 15 minutes the Smartphone will connect with the accelerometer, provided that the accelerometer is within a 5 meter distance of the Smartphone. The transmission of data from the Smartphone towards the It’s LiFe! server is via an internet connection.

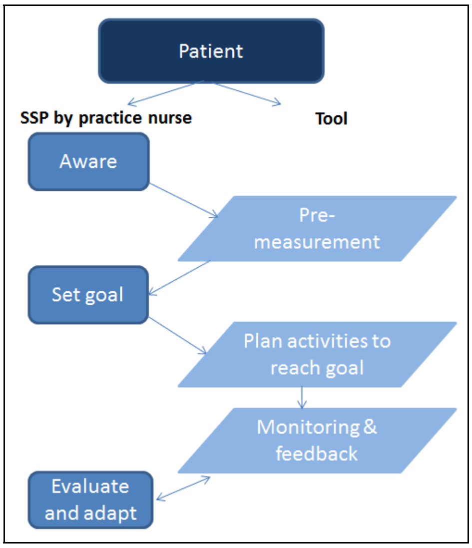


Figure 2. The intervention; Tool integrated in the Self-management Support Program.

Embedding of the tool in the Self-management Support Program

The whole intervention is an embedding of the tool in the SSP (see Figure 2).

To influence patients' motivation for behaviour change, the intervention starts by increasing patients' awareness about the risks of inactivity in a consultation with the practice nurse. Practice nurses indicated that they want to do this in a positive way. Next, the patients are given insight into their own activity pattern by a two-weeks pre-measurement. In the pre-measurement, the activity level of the patient is measured each day and dialogue sessions are sent to reveal which activities the patient does, likes, can do, with whom and which barriers he/she has to overcome. After two weeks, an activity profile is made available on the server which the practice nurse can discuss with the patient in order to set an appropriate goal in minutes a day. The patients and the care professionals indicated that they preferred a pre-measurement and to set a goal in collaboration with each other to make sure that the goal is realistic, challenging, and tailored to the individual preferences and abilities of the patient.

The goal will be set in the server, which is connected to the app. The patient receives a dialogue session in which he/she can plan activities to reach the goal. Planning this goal in more detail such as when, with whom and where you will be active, will narrow the gap between intention and behaviour and makes it more likely that the patient will reach his or her personal goal. If the plan is set, the monitoring and feedback period starts.

Manner and content of the feedback

In the monitoring and feedback period the patient gets three types of feedback:

1. Simple statistics about the amount of physical activity in relation to an activity goal in minutes per day divided in moderate-intense and intense activities (real-time);
2. Motivational messages based on activity results;
3. Responses of the practice nurse based on activity results during consultations.

On the Smartphone, the patient can see information about his or her activity level such as the intensity during a day, a week and a month. The activity monitor does not have a display since the patients indicated that a screen on the monitor would be too small and not easily accessible. Based on the activity, data messages will be sent. There are several types of messages, such as tips, encouragement, positive trend, rewards, barriers, facilitators and suggestions to adjust the goals/target value. Users will receive such messages when they reach or do not reach their goal after 3, 5 and 14 days. In some cases the goals have to be reached for 100% and other rules are based on 80%. Participants will also get messages on positive trends. All messages are written in a positive way. After two to three months, the patient will have a consultation with the practice nurse again to evaluate the results and discuss barriers and facilitators. After the consultation, the monitoring and feedback period can be extended. In between consultations the practice nurse can see the activity results as well and is free to choose whether to react on them or not.

Discussion and future work

According to Fogg, an intervention to change people's behaviour should not only focus on ability and motivation, but should also include a trigger to change [5]. This project provides motivation and ability through self-monitoring of behaviour and the Self-management Support Program that is unfolded by the practice nurse. Supplementary, it provides a trigger by delivering feedback on physical activity on a timely base and in an actionable format, namely related to concrete, personal goals.

Having followed a user-centred design, it is expected that the usability and acceptability of the tool and the SSP are increased. The usability will be tested in a heuristic evaluation and by patients in a lab environment. In addition, the tool and the SSP will be evaluated by looking at technical performance, user experience and acceptance in a pilot in two general practices with 10 patients each. The tool and the SSP will be adjusted based on the findings of the pilot. The effects of the tool embedded in the SSP, will be evaluated in a Randomized Controlled Trial. This will be a trial with three branches, each with 80 patients from eight different general practices: one group will receive "care as usual", one group will receive only the care described in the SSP, and one group will receive the complete intervention with both the SSP and the tool. Primary outcome measure will be physical activity, measured with an accelerometer different from the one developed in this project. Secondary outcome measures will be self-efficacy and quality of life, both measured with questionnaires. This project focuses on patients with COPD or diabetes who are treated in primary care, but in case of proven effectiveness, patients with other chronic conditions could also use the tool and the model. Furthermore, the activity monitor may be extended by measuring other parameters.

References

1. Long Alliantie Nederland. (2010). Zorgstandaard COPD. *Amersfoort: Long Alliantie Nederland* .

2. Nederlandse Diabetes F. (2007). NDF Zorgstandaard : transparantie en kwaliteit van diabeteszorg voor mensen met diabetes type 2. Amersfoort: Nederlandse Diabetes Federatie.

3. Bodenheimer, T., Lorig, K., Holman, H., Grumbach, K. (2002). Patient self-management of chronic disease in primary care. *Jama [Research Support, Non-U.S. Gov't]* **288**, 2469-2475.

4. CBO. (2012). "Self-management, working together towards patient centred care". Zelfmanagement, samen werken aan zorg die bij de patient past. [cited 2012 28-3-2012].

5. Fogg, B.J. (2003). *Persuasive technology : using computers to change what we think and do*. Amsterdam; Boston: Morgan Kaufmann Publishers.

6. van Achterberg, T, Huisman-de Waal, G.G., Ketelaar, N.A., Oostendorp, R.A., Jacobs, J.E., Wollersheim, H.C. (2011). How to promote healthy behaviours in patients? An overview of evidence for behaviour change techniques. *Health Promot Int. [Review]*. **26,** 148-162.

7. Tudor-Locke, C., Lutes, L. (2009). Why do pedometers work?: a reflection upon the factors related to successfully increasing physical activity. *Sports Med.* **39,** 981-993.

8. Jacobson, I. (1992). Object-oriented software engineering: a use case driven approach. [New York]; Wokingham, Eng., Reading, Mass. ACM Press, Addison-Wesley.

Unobtrusive Sleep Monitoring

Reinder Haakma[1], Robbert Jan Beun[2]

[1]*Philips Research, Eindhoven, The Netherlands. reinder.haakma@philips.com*

[2]*Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands. r.j.beun@uu.nl*

Summary

In this paper we will briefly discuss the role of unobtrusive sleep monitoring in behavioral treatments of sleep deprivation and sleep disorders. Lack of sleep is an issue for a large number of people. Sleep quality metrics can help these people in getting an objective view upon the quality of their sleep and discover trends in that quality over time. The golden standard for diagnosis sleep disorder, polysomnography, is not suitable for this purpose because it is obtrusive and uncomfortable. Actigraphy is a better choice, but it is limited in the richness and the accuracy of its sleep metrics. In a persuasive setting, sleep hygiene and CBT-I approaches could benefit not only from unobtrusive sleep monitoring techniques, but also from monitoring of adherence to sleep hygiene rules and therapy compliance for CBT-I.

Sleep and sleep quality

Sleep is a universal phenomenon. All people and all animals sleep. In former times, sleep was highly associated with the night, but nowadays the association is weakened. In contrast to fires and candles, electrically-generated artificial light enables people to be active during the night. Sleep deprivation is not without risk, however, and adequate sleep is critical for optimal day-time functioning [1,2]. Insufficient sleep may cause neurobehavioral impairments, including lapses of attention, reduced working memory and depressed mood. Chronic sleep deprivation is associated with poor health. There is evidence that poor sleep has adverse endocrine, metabolic and inflammatory effects. The underlying causes of sleep deprivation are diverse. It may be caused by sleep disorders, but also medical conditions, occupational or social demands, and environmental conditions can hamper the quality of sleep.

Many people suffer from a lack of sleep. An outcome of ‘Het Grote Slaaponderzoek 2008’ [3] was that one out of five people in the Netherlands sleep poorly. Fifteen percent report not sleeping well multiple nights per week, while 6% indicate they almost never experience a good night sleep. In the USA, the National Sleep Foundation found in the 2009 ‘Sleep in America poll’ [4] that 26% of the respondents experienced a good night sleep only a few times a week, while 24% reported experiencing a good night sleep only a few days a month or even less.

In sleep science, sleep quality is a neglected area [5]. The term has not been rigorously defined and there is only a vague understanding of how objective measures relate to the subjective sleep quality experience. There is evidence that the need for sleep differs between individuals [1,2] and that there is variation between people in neurobehavioral responses to sleep restriction. Nevertheless, in both clinical and sleep research settings sleep metrics are applied that serve as a sleep quality indicator. Examples of metrics are sleep onset latency, total sleep time, number of awakenings, amount of REM-sleep and amount of deep sleep.

Sleep monitoring

Sleep monitoring could help people to understand the quality of their sleep, their sleep behavior and the relation between the two. Since sleep state is characterized by limited consciousness, people may not be aware of the characteristics of their sleep. In sleep studies, it is not uncommon to find discrepancies between objective sleep metrics and subjective sleep reports e.g. acquired via sleep questionnaires or sleep diaries [5].

In sleep laboratories, polysomnography (PSG) is typically used for diagnosing sleep disorders. PSG is a diagnostic test during which a number of physiological variables are measured and recorded during sleep. Trained medical staff places sensor leads on the patient in order to record a large number of parameters,

Proceedings of Measuring Behavior 2012 (Utrecht, The Netherlands, August 28-31, 2012)
Eds. A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, and P.H. Zimmerman

including brain electrical activity, eye and jaw muscle movement, leg muscle movement, airflow, respiratory effort (chest and abdominal excursion), ECG (electrocardiogram), and oxygen saturation. Information is gathered from all leads and fed into a computer and output as a series of waveform traces. This output enables the lab technicians to visualize the various waveforms and score the sleep of the patient. In this way the hypnogram is derived that reflects the stages and cycles of the sleep through the night. From it, sleep quality indicators are derived, such as the ones stated above.

For polysomnography, the sensors are mounted on or glued to the head and the body, making it a very obtrusive and uncomfortable way of assessing sleep. Therefore, this technique is not suitable for longitudinal, ambulatory monitoring. Home-based, longer-term monitoring of sleep can better be done using a technique called actigraphy [6]. An actigraph is a wrist-worn device that records the wearer's movements by means of an accelerometer. Actigraphy identifies sleep onset after a period of continuous inactivity that rarely occurs while people are awake. In a similar way, waking up is detected. In comparison to PSG, actigraphy is more comfortable to wear and more easy to install. However, it is limited to sleep-wake detection and is also less accurate because body movements are an indirect measure of sleep. In particular sleep-onset detection is less precise with actigraphy.

Sleep monitoring in persuasive settings

Persuasive settings in the domain of sleep deprivation can typically be found in the treatment of insomnia. Insomnia may be characterized as a persistent difficulty initiating and/or maintaining sleep [7]. Today it is widely accepted that cognitive and behavioral factors play an important role in the condition of insomnia, making cognitive behavior therapy for insomnia (CBT-I) a generally accepted, non-pharmacological treatment [8] producing sustainable positive changes in this condition. Basically, CBT-I and other behavior related treatments offer a number of exercise types whose effectiveness is proven for sleep therapy – i.e. sleep restriction, stimulus control, relaxation, and sleep hygiene [9]. Sleep restriction involves curtailing the time spent in bed to stabilize the sleep pattern and then lengthening sleep time as sleep efficiency improves. Stimulus control is aimed at the coachee's re-association of the bed and the bedroom with sleep and to re-establish a consistent sleep-wake schedule. Relaxation training involves methods aimed at reducing somatic tension. Sleep hygiene education aims to make the person aware of behavioral practices and environmental factors that may promote or hamper sleep; examples are the advice to keep a regular schedule, to avoid daytime napping and to refrain from caffeine, alcohol and nicotine before bedtime [10].

In general, behavior treatments such as CBT-I require a great deal of effort and extensive self-discipline of the insomniac. People may enthusiastically start a particular self-help sleep therapy and discover that sizing down their time spent in bed or getting out of bed in the middle of a cold and dark night requires a great deal of effort. Consequently, sleep quality deteriorates, people feel worse, get into a downward motivation spiral and the therapy is terminated prematurely. People may find the exercises too strenuous (e.g. sleep restriction), they do not believe that it contributes to a solution of the problem or simply forget to perform the exercise (e.g. a relaxation exercise).

In order to motivate the individual to adhere to the exercises, a combination of persuasive strategies may be applied. Some of these strategies pertain to the adaptation of the exercises to the individual's characteristics, such as ability and preferences. Other strategies refer to the ease of the activities that have to be performed during the therapy. For instance, people have to fill in a sleep diary to gain insight of the sleep behavior during the therapy. Not only are sleep diaries notoriously unreliable with respect to reporting sleep characteristics, they also put an extra burden on the individual of reporting and processing various types of sleep relevant information.

In this context, sleep monitoring may have important advantages. First, because individuals are never able to intentionally indicate that they are asleep, information can be obtained where direct access would otherwise be impossible (e.g. the individual's sleep stages). Second, insomniacs do not need to focus on the measurement process and do not have to worry about the timing of the measurement; as a result, they can avoid the tedium of tracking their own performance. Third, data can be collected in a more objective way than when entered by the individual. Fourth, sleep monitoring may smoothly be integrated with other behavioral measurements to offer (semi-) automated and individually tailored intervention programs. Consequently, sleep monitoring enables individuals to become aware of relevant unobservable behavior, provide sleep indicators to track sleep quality over time and feedback upon their sleep-hygienic behaviors, and may release the individual from a range of therapy related activities.

From the previous it should not be concluded that sleep monitoring is beyond doubt. Integrity of the data processing and reliable sensing are important prerequisites for acceptation of a sleep monitoring system. Also, intrusive measurements should be avoided wherever possible in a first step therapy. Consequently, a trade-off should be found between reliability, intrusiveness and integrity of the information. Since CBT-I and sleep hygiene requires longer-term, home-based monitoring, unobtrusiveness and comfort are important requirement for the assessment techniques, in combination with a sufficient level of accuracy to observe trends in quality of sleep.

Conclusion

We outlined the role of unobtrusive sleep monitoring in behavioral treatments of insomnia that aim to improve people's sleep by promoting more effective cognitions and behaviors around sleep. In a persuasive setting, sleep hygiene and CBT-I approaches can benefit from sleep monitoring techniques, not only by providing objective sleep information but also from monitoring of adherence to sleep hygiene rules and therapy compliance for CBT-I.

References

1. Kryger, M.H., Roth, T., Dement, W.C. (2005). *Principles and practice of sleep medicine*. Elsevier/Saunders.

2. Banks, S., Dinges, D.F. (2007). Behavioral and physiological consequences of sleep restriction. *J. Clin. Sleep Med.* **3**(5), 519-528.

3. GezondheidsNet.nl, Het Grote Slaaponderzoek 2008. http://www.gezondheidsnet.nl/geest/artikelen/1040/het-grote-slaaponderzoek-2008 Accessed 3 April 2012.

4. National Sleep Foundation, Sleep in America poll http://www.sleepfoundation.org/category/article-type/sleep-america-polls Accessed 3 April 2012.

5. Krystal, A.D., Edinger, J.D. (2008). Measuring sleep quality. *Sleep Medicine* **9** Suppl. 1, S10-7.

6. Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., Pollak, C.P. (2003) The role of actigraphy in the study of sleep and circadian rhythms. *Sleep* **26**(3), 342-392.

7. Morin, C.M. (1993). *Insomnia: Psychological assessment and management.* New York: Guilford Press.

8. Bain, K.T. (2006). Management of chronic insomnia in elderly persons. *Am. J. Geriatr. Pharmacother.* **4**(2), 168-192.

9. Ancoli-Israel, S., Ayalon, L. (2006). Diagnosis and treatment of sleep disorders in older adults. *Am. J. Geriatr. Psychiatry* **14**(2), 95-103.

10. Stepanski, E.J., Wyatt, J.K. (2003). Use of sleep hygiene in the treatment of insomnia. *Sleep Med. Rev*. **7**(3), 215-225.

Unobtrusively Measuring Stress and Workload of Knowledge Workers

S.J. Koldijk[1,2], M.A. Neerincx[2], W. Kraaij[1,2]

[1]*ICIS, Radboud University, Nijmegen, The Netherlands.* [2]*TNO, Soesterberg, The Netherlands.*

{saskia.koldijk, mark.neerincx, wessel.kraaij}@tno.nl

Imagine a typical working day of a knowledge worker, i.e. someone who is predominantly concerned with interpreting and generating information. Bob gets into the office at 9, starts up his computer, takes a look at his mails and calendar and plans what things he has to do this day. Then he starts working on one of the important tasks that have to be completed this week. When an e-mail comes in, he quickly reads it. As it is not relevant to him, he continues his task. When a colleague drops by, Bob looks up and talks to him, which results in some more to-do's for the day. As they are quite urgent, Bob decides to do them right away. After completion he switches back to his previous task and continues his working day. At 5 'o clock Bob notices that he has not completed all planned tasks yet and he feels somewhat stressed. He starts wondering where the time went and whether he should work overtime again to finish up.

Bob and many other knowledge workers experience working days like this, with important tasks to complete, while interruptions cause task switches and unplanned things have to be handled. These people often experience stress building up, which in the worst case results in burn-out [1, 2]. To prevent this from happening, knowledge workers should become more aware of what makes them feel stressed and how they can handle or avoid that.

Our goal

In our research we want to investigate how well-being at work can be improved by means of computer tools. We particularly focus on the unobtrusive measurement of workload and stress in order to give knowledge workers feedback and support in their way of working, preventing burn-out in an early stage. In this extended abstract we describe how the context and user state of an individual knowledge worker will be captured, interpreted and used in a coaching tool aimed at changing behaviour, resulting in improved well-being at work.

Well-being at work

In order to design an effective coaching tool we need to know what is of influence on well-being at work. Based on a literature study we identified four categories of determinants of well-being at work: the work itself, the working conditions, private circumstances and personal factors [3]. First of all, the type of work has influence on the well-being of employees. It is important that the work is diverse and the workload is not too high and not too low. In addition, high autonomy and feedback on work performance are beneficial to well-being at work. By contrast, interruptions and responsibilities outside one's control reduce well-being.

Not only the work itself but also factors relating to work conditions play a role. Interpersonal relationships are important, causing a stable social environment for the employee, which can help him or her deal with challenging situations. In addition, factors like career opportunities, good payment or job security are also important determinants of well-being at work. Furthermore, physical work conditions and psychological comfort also play a role.

Besides work conditions, also the private circumstances of employees can have an influence on their well-being. Again, social support plays an important role, as it can act as a buffer against problems at work. In general, it is important that there is a good balance between work and life. Problems at home can have a negative effect on the well-being of employees.

The final important category that influences well-being comprises personal factors. Two people in the same situation can experience different well-being. Factors of influence are age, sex, education, social and intellectual skills. Moreover factors like expectations, emotional stability and sensitivity to suffer from stress are of influence on well-being.

In our research we want to develop a computer tool that can help people to better cope with the negative determinants of well-being. Preferably it should be possible to capture relevant factors by sensing interactions with a computer. Therefore, we focus on the factors of the work itself and the working conditions. We aim to automatically recognize the task and content users are involved in, as well as their social and physical context. Moreover, we intend to automatically infer the user state, in terms of cognitive load, valence and arousal in order to get an indication of the stress level experienced. These aspects can then be coached upon.

Regarding the other determinants, the category of personal factors should certainly also be kept in mind. Well-being is subjective, and so individual differences should be considered by personalizing the tool. Moreover, the private circumstances of the user are important to consider, as they may have a great influence on well-being. Nevertheless measuring these factors will only be possible by asking the employee.

Automatic recognition of context and user state

For providing effective support to knowledge workers it is crucial to understand their current situation. In our system we define four relevant forms of contexts (see Figure 1). With task context we mean the task the knowledge worker is currently working on (e.g. writing a report, making a presentation). Content context refers to the project or content the person is currently involved in. With social context we mean recent contacts with colleagues. Finally, physical context includes aspects as location, movement and posture.

We intend to capture these contexts automatically, without user effort. The inference of context will be formulated as a pattern recognition problem. From various sensors, specific features will be extracted which are provided to a classifier to assign a label. In a pilot study data from several knowledge workers will be collected during their working day and annotated with context labels. The annotated data will be used to train and test various pattern recognition approaches.

Important for automatic context recognition is that it requires little user effort. The sensors should preferably be unobtrusive, available in a normal office setting or otherwise relatively cheap. Moreover, for real-time support, the system should be able to quickly come from sensor data to context labels. It is important that the automatic recognition of context is reliable and robust, so that the users gain trust in the system. Finally, the privacy of the user has to be taken in mind when personal data is accessed.

Figure 1 depicts a first setup for linking different sensors to the described contexts and user states. A first selection of sensors has been made based on costs, availability, obtrusiveness and expected usefulness. The first set of sensors is located at the computer or mobile phone and is therefore readily available for most knowledge workers. Information from keyboard, mouse and applications can be used to infer what task the person is performing [4]. With information about the content that the user accessed, the content context can be determined. Analysing phone calls made and e-mail conversations can yield information about the social context of the user. Moreover, several sensors can be used to infer aspects of the user's physical context. GPS and accelerometers can be used to infer the position and level of activity of the user. With visual information from a camera or Kinect the body posture can be recognized and audio input can be used to recognize the level of noise in the environment.

Combining information from different contexts can give insights in the mental state of the users. The cognitive task load can be determined based on three dimensions: The duration of a task, the difficulty of the task and the amount of switching between tasks [5]. The information of the tasks performed and the contents worked on can thus give an indication of the experienced workload, which is a first indication of the stress experienced during the day. Furthermore, the social context and physical context can influence the stress level experienced.

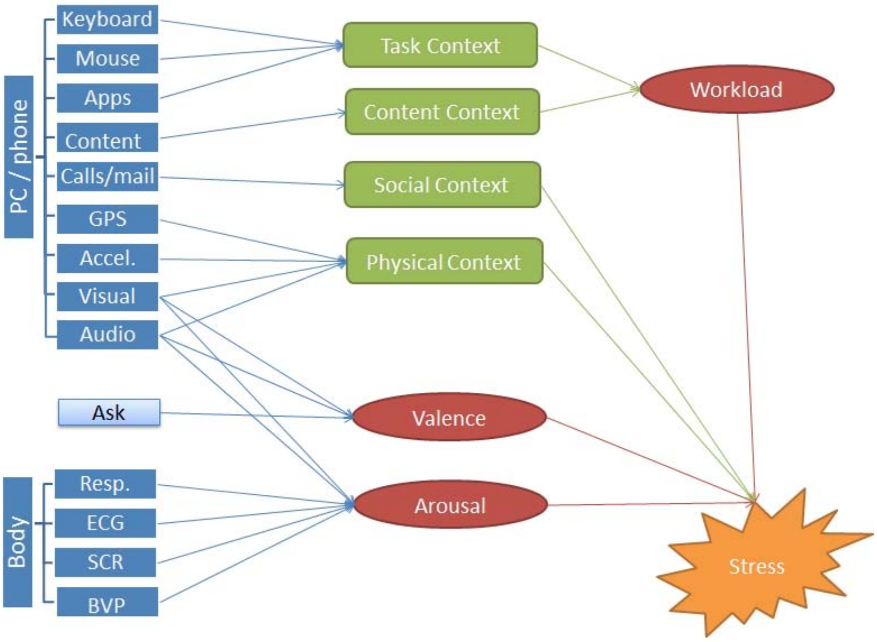


Figure 1. First setup for linking different sensors to the described contexts and user states.

In addition, information from facial expressions or speech can be used for inferring the emotional state of the user, in terms of valence and arousal [6]. For determining valence, also asking the user can yield valuable information (e.g. do you like or dislike the work you are performing now). Besides computer sensors, wearable sensors to measure body signals can be used. The respiration of the user, his heart signal (ECG), skin response (SCR) and blood volume pulse (BVP) can be used to determine the level of arousal [7]. The recognized arousal and valence can then additionally be used to estimate the level of stress that the knowledge worker is experiencing. Our hypothesis is that a high workload in combination with high arousal and negative valence causes a negative feeling of stress.

Based upon these inferred contexts and user states, an e-coaching tool could support knowledge workers. For example, more awareness of performed activities could be created by providing an overview of active tasks and projects, as well as information about durations of tasks and the amount of task switching. Most informative is probably how the current behaviour relates to usual behaviour of the user. Slow or sudden changes or anomalies could be detected and users could be warned when, for example, the task load gets too high or the stress level is building up. By linking the workload and stress curves to the tasks performed, knowledge workers could gain insights into what has caused their feelings of stress. They could then learn to avoid stress by planning the week differently or taking more short breaks to relax.

Personalization

The general model described in the previous section will have to be adapted to the specific user, due to personal characteristics. The same task can for example have a lower or higher workload for a knowledge worker depending on his level of experience and expertise. Therefore a user model will be created for each user, which holds relevant information of personal factors, models of typical behaviour and how low level information links to the experienced well-being of this user. The tool should be person adaptive (applying personalization/ tailoring), as well as situation adaptive (taking the user’s current context into account), to provide effective support and to gain trust of the user, so that the tool is used efficiently and over an extended period of time.

Discussion

Our contribution is focussed on providing knowledge workers efficient support. A state of the art study we performed revealed that although several applications with this aim exist on the market, applications which make use of sensors and smart interpretation of data to base support upon are rare. Typically the applications require much user effort and lack personalisation and adaptation to the user's context. In our research we want to overcome these problems by automatically inferring information about the context and state of the user. Unobtrusive and relatively cheap sensors will be used for estimating the workload and stress of knowledge workers. We will perform most research in real-world office settings and test our reasoning techniques on realistic data. In this way, our solution will be affordable and applicable for many knowledge workers, opposed to related work on workload and stress recognition, in which often rather obtrusive or expensive sensors are used and tested in controlled experiments. Finally, often the success of applications on the market is questionable, so we will perform a field study to evaluate the effect of our tool on well-being at work.

Research agenda

This work is part of the SWELL project[1], in which smart reasoning systems for well-being at work and at home will be developed. Important aspects of this project include the development of a stable and flexible framework for combining data from various sensors, and performing reasoning upon this data. Automatic inference of various context and user aspects will be enabled, as described here. Furthermore, possibilities to provide information support to knowledge workers by, for example delivering relevant information just in time, will be investigated. Policies for guaranteeing the privacy of the user will be developed. Finally, appropriate methods for achieving long lasting behavioural change and improved well-being will be applied and evaluated in field studies.

Acknowledgements

This publication was supported by the Dutch national program COMMIT (project P7 SWELL).

References

1. Michie, S. (2002). Causes and management of stress at work. *Occup. Environ. Med.* **59**, 67-72.

2. Kalimo, R. (1999). Knowledge jobs - how to manage without burnout? *Scand. J. Work Environ. Health* **25**(6), 605-609.

3. Vos, P., van Staalduinen, M., Koldijk, S. (2010). Studie naar werkvreugde en determinanten: Literatuurstudie en reflectie van experts. *TNO rapport*.

4. Koldijk, S., van Staalduinen, M., Neerincx, M., Kraaij, W. (2012). Real-time task recognition based on knowledge workers' computer activities. *Proceedings of ECCE 2012* (Edinburgh, Scotland, 28-31 August 2012).

5. Neerincx, M. A., Kennedie, S., Grootjen, M., & Grootjen, F. (2009). Modeling the cognitive task load and performance of naval operators. In *Foundations of augmented cognition: Neuroergonomics and operational neuroscience* **5638** ,60-269.

6. Caridakis, G., Malatesta, L., Kessous, L., Amir, N., Paouzaiou, A., Karpouzis, K. (2006). Modeling naturalistic affective states via facial and vocal expression recognition. *Int. Conf. on Multimodal Interfaces*, 146-154.

7. A. Haag, S. Goronzy, P. Schaich, J. Williams (2004). Emotion recognition using bio-sensors: first step towards an automatic system. *Affective Dialogue Systems, Tutorial and Research Workshop* (Kloster Irsee, Germany, 14-16 June 2004).

[1] http://www.commit-nl.nl/projects/swell-smart-reasoning-systems-for-well-being-at-work-and-at-home

A Context-Aware Adaptive Feedback Agent for Activity Monitoring and Coaching

H. op den Akker[1,2], V.M. Jones[2], L.S. Moualed[1,3], H.J. Hermens[1,2]

[1]*Roessingh Research and Development, Cluster Telemedicine, Enschede, The Netherlands. h.opdenakker@rrd.nl*

[2]*University of Twente, Telemedicine Group, Enschede, The Netherlands*

[3]*University of Twente, Biomedical Engineering Group, Enschede, The Netherlands*

Abstract

A focus in treatment of chronic diseases is optimizing levels of physical activity. At Roessingh Research and Development, a system was developed, consisting of a Smartphone and an activity sensor, that can measure a patient's daily activity behavior and provide motivational feedback messages. We are currently looking into ways of increasing the effectiveness of motivational messages that aim to stimulate sustainable behavioral change, by adapting its timing and content to individual patients in their current context of use.

Introduction

One of the focus areas in the treatment of chronic diseases is to optimize levels of physical activity. We are investigating how to support this by objectively measuring a patient's daily activity patterns and providing regular feedback on the patient's performance via a Smartphone. The positive effects of regular feedback on behavior change are supported by numerous scientific studies. For example Kreuter et al. [1] showed the positive effect of regular feedback on four different kinds of health related behavioral change including increase of physical activity. Our own research has shown the positive effects of giving feedback to Chronic Low Back Pain patients throughout the day [2], while similar effects have been shown for patients with COPD and Chronic Fatigue Syndrome [3].

The system that was used in our activity monitoring and feedback research is depicted in Figure 1 below. On the left is the ProMove-3D physical activity sensor, developed by Inertia Technology; in the center is a screenshot of the Smartphone interface showing a graph of daily activity, and on the right is a screenshot showing a motivational advice (feedback) message: "*Take a nice walk!*".

These motivational text messages, which will be the focus of the rest of this article, are currently used in a relatively ad hoc manner. The Smartphone can be set to generate a motivational message every hour (or every *x* minutes), and at that time, the message to be presented to the user is picked randomly from a pre-defined list of

Figure 1. The Activity Monitoring and Feedback system from Roessingh Research and Development. Depicted is the ProMove-3D activity sensor, and two screenshots of the Smartphone interface showing a graph of daily activity and a motivational message.

around 25 messages. This message is then shown to the patient by displaying the text on the screen. As described in [4], we distinguish between three aspects of feedback: *timing*, *content*, and *representation* (when to give feedback, what to tell the user, and how to convey this to the user) which we are focusing on in our current research. The next section will shortly describe our work regarding feedback timing and content. The representation of feedback is an aspect left for future work.

Method: Improving Feedback Timing and Content

Although the positive effect of giving regular feedback messages has already been shown, compliance to individual feedback messages is relatively low (59% on average over measurements from 86 different patients). Compliance here is defined as the immediate response to a given feedback message. If the patient is told to become more active (i.e. "*Please go for a walk.*"), and the patient does become more active in the 30 minute interval after the feedback message, as measured by the activity sensor, we define this as being compliant to the message, and vice versa. This immediate behavior change after feedback messages (compliance) can be computed in real time on the Smartphone. We use this direct feedback loop to the system to enable the system to learn to adapt to the patient in terms of when to give feedback and how to phrase the feedback message. The idea is that every individual patient has its own preferences in terms of this feedback timing and content, and we should be able to increase compliance to individual messages, by adapting these aspects to individuals. The way in which this real-time personalization is realized in the feedback system is explained for the feedback timing and the feedback content below.

In order to find an optimal timing for feedback messages, we have analyzed the messages generated in previous studies. We have a dataset of 2,374 feedback messages for which compliance is known. As mentioned earlier, only 59% of these messages where complied to, so the goal of this work was to find out why some messages where complied to and others were not. For this we have created a classifier that is able to predict compliance to feedback messages based on a set of features, consisting of context information (e.g. weather, time of day) and features related to the history of use of the system (e.g. how many messages where already received today). We were able to predict feedback compliance with **86% accuracy** on average over all patients. Compared to an average baseline score of 60%, this constituted a 64% increase in accuracy over baseline, proving the viability of the approach. The work in [5] describes the details of this work.

For the personalization of feedback content to individual patients, we took a slightly different approach. We exploit the same principles by using the direct measurement of behavior (compliance calculation) after a feedback message to adapt the system to individuals. But for the content generation system we chose not to take a classical machine learning approach. We designed an ontology of feedback messages in which motivational messages are organized by the type of activity they prescribe. For example, at the top of the Ontology, messages are divided into "Outside", "Inside" and "Generic" categories (see Figure 2). These are three categories that relate to the content of the given advice: the "Outside" tree contains all advices that tell the patient to do something outside, the "Inside" tree contains everything that can be done inside, while the "Generic" tree contains advices, like "Be more active!". We then developed an algorithm that can traverse the ontology, and at each level can select an appropriate path based on the user's context and the user's responses to previous messages. For example, choosing a message from the "Outside" section of the Ontology is only appropriate if the weather is nice. By looking at the patient's context (weather at current location), the algorithm can decide to prune that part of the Ontology. Then, when a selection has to be made between Inside and Generic messages, the algorithm will favor the node to which the patient previously responded best. Similarly the Ontology can include other nodes that can be pruned based on contextual information. For example, a node "Bike", containing advices that tell someone to go for a bike ride, can be pruned if the user has indicated that he or she does not own a bike. The Ontology based message selection algorithm and method is described in more detail in [6].

Issues and Future Work

The feedback timing and content approaches are in an advanced state, both being implemented and currently undergoing trials on the Smartphone. A current pressing issue is the problem of *cold start*. Whenever a new

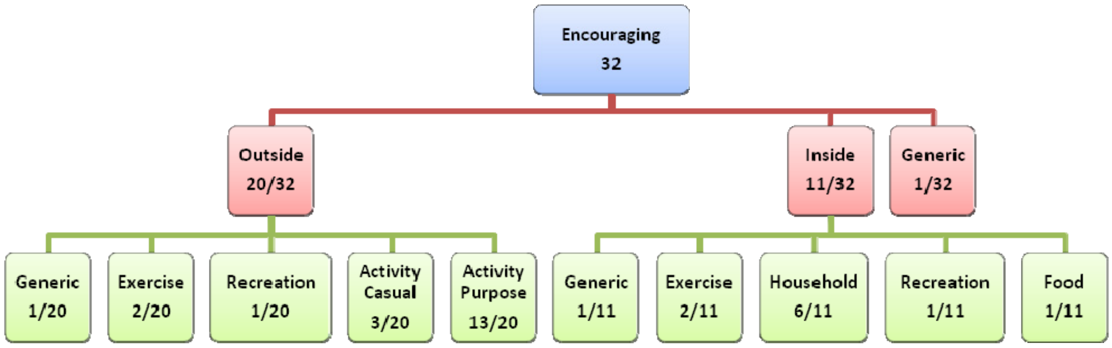


Figure 2. Message Ontology with high level categories for motivational feedback messages. The numbers in each node give information on the distribution of leave nodes (actual advice messages) under each node. E.g., in this example Ontology, there are 32 leave nodes: 20 under “Outside”, 11 under “Inside” and 1 under “Generic”. Of all the subcategories under “Outside”, there are 20 leave nodes, 3 of which are in the “Activity Casual” category, etc. These fractions are the prior probabilities of selection for each node, if no user data is available.

patient starts using the system, there is no data available to start the personalization algorithms. For the feedback timing module, we aimed to solve this problem by creating a cold start classifier. This classifier is trained on data from all of our previous patients and is used to find the right timing for the first 25 motivational messages (after which the system will train an individual classifier and switch to using this). This approach works, but the classification accuracy of this cold start classifier is only 67% (compared to 86% for the individual classifiers). On the one hand this shows that data from different patients is really different, which is a strong motivator for doing personalization in the first place. On the other hand, the cold start issue remains unsolved for our current application.

We are now looking into ways of improving our system’s cold start phase, specifically by looking into “behavior change personality clustering” of patients, or, clustering patients who respond similarly to motivational messages. This way we could compare a new patient with patients from earlier studies and hopefully increase our predictive power. This approach is well known in the field of recommender systems, and a similar approach in the field of eHealth was taken by Cortellese et al. [7], but to the best of our knowledge never implemented into a real-time system.

References

1. Kreuter, M.W., Strecher, V.J. (1996). Do tailored behavior change messages enhance the effectiveness of health risk appraisal? Results from a randomized trial. *Health Education Research* **11**(1), 97-105.

2. Van Weering, M.G.H., Vollenbroek-Hutten, M.M.R., Tönis, T.M., Hermens, H.J. (2009). Daily physical activities in chronic lower back pain patients assessed with accelerometry. *European Journal of Pain* **13**(6), 649-654.

3. Evering, R.M.H., Tönis, T.M., Vollenbroek-Hutten, M.M.R. (2011). Deviations in daily physical activity patterns in patients with the chronic fatigue syndrome: a case control study. *Journal of Psychosomatic Research* **71**(3), 129-135.

4. op den Akker, H., Moualed, L.S., Jones, V.M., Hermens, H.J. (2011). A Self-Learning Personalized Feedback Agent for Motivating Physical Activity. *Proceedings of ISABEL 2011: the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*. Barcelona, Spain.

5. op den Akker, H., Jones, V.M., Hermens, H.J. (2010). Predicting Feedback Compliance in a Teletreatment Application. *Proceedings of ISABEL 2010: the 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies*. Rome, Italy.

6. Wieringa, W., op den Akker, H., Jones, V.M., op den Akker, R., Hermens, H.J. (2011). Ontology-Based Generation of Dynamic Feedback on Physical Activity. *Proceedings of the 13th Conference on Artificial Intelligence in Medicine (AIME)*, Bled, Slovenia, 55-59.

7. Cortellese, F., Nalin, M., Morandi, A., Sanna, A., Grasso, F. (2009). Static and Dynamic Population Clustering in Personality Diagnosis for Personalized eHealth Services. *Proceedings of the 3rd International Workshop on Personalisation for e-Health, AIME2009*, 1-9.

SPECIAL SESSION

Social Behavior and Communication – From Mice to Primates

Chair: Markus Wöhr[1], Marcel van Gaalen[2]

[1]*Behavioral Neuroscience, Experimental and Physiological Psychology, Philipps-University of Marburg, Marburg, Germany. woehrm@staff.uni-marburg.de*

[2]*Neuroscience Research, CNS Pharmacology , Abbott GmbH & Co. KG, Ludwigshafen, Germany. marcel.vangaalen@abbott.com*

Abstract

Social interactions and communication are among the most complex forms of behaviors. Analyzing these behaviors remains a time consuming challenge which requires practice, skills and patience. However, new techniques became available to speed up analysis and standardize methods. This is particularly relevant as social behaviors are a very prominent part of the behavioral repertoire in many species. In humans, a variety of social behaviors are disturbed in neurological and psychiatric disorders such as schizophrenia, Alzheimer's disease and autism. The speakers of this symposium will present methods that allow reliable assessment of social behavior and communication – from mice to primates. They will demonstrate improved techniques for analysis, but also give examples of behaviors that can so far only be analyzed by human observations.

Marcel van Gaalen will give examples of measurement of dominant submissive behavior and various forms of aggression in mice and rats. He will show that an increased understanding of the pharmacology of these behaviors is relevant for drug discovery for CNS disorders. In addition, he will give examples of experimental compounds that may have potential for treating social interaction and communication disruptions in CNS disorders.

Louk Vanderschuren will show that social play is the most characteristic component of the social repertoire of young mammals, which is of great importance for the development of physical, cognitive and social capacities. Social play is highly rewarding and an incentive for maze learning, lever pressing and place conditioning in rats and primates. It is modulated through neurotransmitters implicated in the motivational and hedonic properties of food and drug rewards, such as endogenous opioids, dopamine and endocannabinoids.

Michael Lukas will describe how behavioral assays like the social recognition/discrimination and the social preference paradigms can be used to study elementary social abilities that are essential for functioning social communication such as social preference and individual social memory. Such social behaviors are regulated by highly conserved neuropeptides, like vasopressin, oxytocin, and the newly described neuropeptide S. Michael Lukas will present examples showing how social recognition/discrimination and the social preference can be combined with central pharmacological manipulations and in vivo microdialysis to investigate the involvement of relevant neuropeptides.

Markus Wöhr will show how measuring ultrasonic vocalizations and scent marks can be applied to study communication deficits in mouse models of autism. Autism is a severe neurodevelopmental disorder characterized by three core symptoms: 1) abnormal reciprocal social interactions, 2) deficits in social communication, and 3) repetitive behaviors. While reliable behavioral assays for abnormal reciprocal social interactions and repetitive behaviors are available since several years, relevant behavioral assays in the field of communication became available only very recently. Ultrasonic vocalizations and scent marks are particularly interesting as they may help to detect acoustic and olfactory communication deficits, respectively. Markus Wöhr will describe under which conditions ultrasonic vocalizations and scent marks can be recorded and what equipment is needed.

Dominik Seffer will give examples how playback experiments can be used to study social behavior and ultrasonic communication in rats. Different types of rat ultrasonic vocalizations serve distinct communicative functions, eliciting opposite behavioral and neural responses. Dominik Seffer will present a recently developed behavioral paradigm to study social approach behavior elicited by appetitive high-frequency ultrasonic vocalizations and will provide experimental data showing that the paradigm is sensitive for detecting differences in social experience during early life.

Hans Slabbekoorn will address how standardized behavioral tests on temporary captive animals can provide an index of personality. He will argue that such a quantification of personality traits can help in the interpretation of avian playback results in natural territories.

Finally, **Kurt Hammerschmidt** will present a description of the vocal repertoire of nonhuman primates, which consists of a limited number of call types that vary substantially within these categories. One can find all variation from highly graded to more or less distinct vocal repertoires, and it may be unavoidable that call variants at category boundaries are difficult to categorize. This must not be a problem for animals which grew up in the same community, because the efficiency of categorical perception allows the receiver to respond correctly even in graded signaling systems. However, to understand the evolution of acoustic communication and to compare studies in bioacoustics it is necessary to have comparable units, or to know how different categorization levels influence the outcome of an acoustic analysis. Kurt Hammerschmidt will discuss different ways to establish a reliable categorization.

SPECIAL SESSION CONTENTS (sorted by paper ID)

Studying the Neurobehavioral Mechanisms of Social Behavior in Adolescent Rats

Louk .J.M.J. Vanderschuren, E.J.M. Achterberg (Utrecht University, The Netherlands), P.J.J. Baarendse, R. Damsteegt, L.W.M. Van Kerkhof (UMC Utrecht, The Netherlands), V. Trezza (University “Roma Tre”, Italy).

Ultrasonic Communication in Mouse Models of Autism

Markus Wöhr (Philipps-University of Marburg, Germany)

Ultrasonic Communication in Rats: Insights from Playback Studies

Dominik Seffer, Rainer K.W. Schwarting, Markus Wöhr (Philipps-University of Marburg, Germany)

Categorizing Vocal Repertoires of Nonhuman Primates

Kurt Hammerschmidt, Philip Wadewitz (German Primate Center Göttingen, Germany)

Central Neuropeptides Social Recognition, Social Preference and Social Fear in Rodents

Michael Lukas, Iulia Toth, Inga D. Neumann (University of Regensburg, Germany)

Individual Differences in F0 Imitation: Causes and Effects

Marie Postma, Eric Postma (Tilburg University, The Netherlands)

Measuring Behavioural Changes to Assess Anthropogenic Noise Impact on Singing Birds

Hans Slabbekoorn (Leiden University, The Netherlands)

Measuring Social Behavior in Drug Discovery

Marcel van Gaalen, Thomas Appl, Anton Bespalov (Abbott, Ludwigshafen, Germany)

Studying the Neurobehavioral Mechanisms of Social Behavior in Adolescent Rats

L.J.M.J. Vanderschuren[1,2], E.J.M. Achterberg[1], P.J.J. Baarendse[2], R. Damsteegt[2], L.W.M. Van Kerkhof[2], V. Trezza[3]

[1]*Department of Animals in Science and Society, Division of Behavioural Neuroscience, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands. l.j.m.j.vanderschuren@uu.nl*

[2]*Rudolf Magnus Institute of Neuroscience, Department of Neuroscience and Pharmacology, University Medical Center Utrecht, Utrecht, The Netherlands*

[3]*Department of Biology, University "Roma Tre", Rome, Italy.*

Social play behavior

In between weaning and puberty, the young of all mammalian species, including humans, display a characteristic form of social interaction known as social play behaviour or rough-and-tumble play. This form of social behaviour is highly rewarding and essential for the development of social and cognitive skills [2,5,6,7,9,19]. Our research focuses on elucidating the neural and behavioral underpinnings of social play behaviour in adolescent rats, and its role in behavioural development. To that aim, we have developed methods to study the performance of social play behavior, the rewarding properties of social play behavior and the motivational properties of social play.

Assessment of different behavioral components of social play behavior

The performance of social play behavior is studied in dyadic encounters. Rats (25-40 days old), that are matched for sex (we usually use males) and weight are used. To preclude the influence of existing dominance hierarchies on social play, cage mates are never tested against each other. To enhance the motivation for social behavior, rats are briefly socially isolated before the experiment. Isolation for up to 24 hr causes an orderly increase in social play behavior during the test [4,18,20]. To enable the ready detection of both increases and decreases in social play behavior, an isolation time is used that evokes a half-maximal increase in social play behavior, i.e. 3.5 hr [4].

In rats, a bout of social play behavior starts with one rat soliciting ('pouncing') another animal, by attempting to nose or rub the nape of its neck. The animal that is pounced upon can respond in different ways: if the animal fully rotates to its dorsal surface, 'pinning' is the result. Pinning is regarded as the most characteristic posture in social play in rats. As rats mature into adulthood, the structure of social play changes. Most prominently, the response to pouncing with full rotation to the dorsal surface (resulting in pinning) occurs less and is replaced by evasions and partial rotations [6]. The following behaviors are relevant measures of social play in rats [6,9]:

Pouncing: Nuzzling the nape of the conspecific's neck with the tip of the snout followed by a rubbing movement.

Evasion: Upon solicitation, the recipient animal avoids contact with the nape by leaping, running, or turning away from the partner.

Partial Rotation: Upon contact of the nape, the recipient animal begins to rotate along its longitudinal axis, but then stops and keeps one or both hind feet firmly planted on the ground.

Pinning: Upon contact of the nape, the recipient animal fully rotates around the longitudinal axis of its body, ending in a supine position with the other subject standing over it.

Boxing/Wrestling: Rearing in an upright position towards the other subject combined with rapidly pushing, pawing, and grabbing at each other, or wrapping around the other subject.

Following/Chasing: Moving or running forward in the direction of or pursuing the other subject, who moves away.

In addition to these measures, social exploration (sniffing, licking or grooming any part of the body of the test partner, including the anogenital area) is assessed as a measure of general social interest, which may not necessarily be associated with playful social behavior. Also, non-social exploratory behavior is measured, in order to detect any general behavioral impairments as a result of neural or pharmacological manipulations.

To understand the behavioral structure of social play in more detail, we have also established setups to study specifically the rewarding and motivational properties of social play. The rewarding properties of social play behavior are assessed using a place conditioning setup [1,15]. In place conditioning experiments, an apparatus is used that consists of two chambers that can be clearly distinguished by the animals, for example on the basis of different visual, tactile and olfactory cues. Preferably, the animals should have no innate preference for one of the two chambers before conditioning starts. During conditioning, the animals are allowed to engage in social play behavior in one chamber and are placed alone in the other chamber, in alternating sessions. In this way, the cues in one chamber gain meaning to the animal because they get paired with something behaviorally relevant, i.e. social play. During testing, the animals can freely move around the two chambers and the time they spend in either one is recorded. If they spend more time in the chamber that has been associated with social play, than this is said to evoke conditioned place preference, which is interpreted as social play having positive subjective ('pleasurable') effects. In order for social play behavior to evoke conditioned place preference, our parametric analysis has indicated that social isolation during conditioning and testing (causing the animals to be able to engage in social play only during conditioning sessions) is optimal. We have found that 8 conditioning sessions, lasting 30 min each, induces robust, social play-induced conditioned place preference [11].

At this moment, we are also developing an operant conditioning setup, in which animals have to lever press in order to gain the opportunity to engage in a short episode of social play. Social isolation during training sessions causes animals to readily acquire the task. Once trained, the animals perform in a stable fashion under both fixed ratio (where a fixed number of lever presses throughout the test session is required to receive the reward) and progressive ratio schedules of reinforcement (where an increasing number of lever presses after each reward is required).

The neuropharmacology of social play behavior

Our recent work has focused on the neurobiology of social play behavior. These studies have revealed important roles for cannabinoid, opioid, dopaminergic, noradrenergic and serotonergic neurotransmission in social play [3,8,12-14,20, for reviews see 9,17]. This is in keeping with the rewarding properties of social play, as these neurotransmitter systems have been widely implicated in the positive subjective properties of food, sex and drugs. In-depth analysis of the underlying neural substrates has identified the nucleus accumbens as a site of action for the stimulating effects of mu-opioid receptor agonists on social play behavior [10]. Furthermore, analysis of immediate early gene expression patterns after social play revealed activation of limbic corticostriatal pathways during social play. Based on these data, pharmacological inactivation studies have subsequently confirmed the importance of the prelimbic cortex, medial orbitofrontal cortex and nucleus accumbens core for social play [16].

Conclusions

Our studies have shown that the performance of social play, as well as its rewarding and motivational properties can be readily assessed in adolescent rats. Interacting opioid, cannabinoid and dopaminergic systems within the corticolimbic circuits underlying incentive motivation and reward modulate social play behaviour.

References

1. Bardo, M.T. and Bevins, R.A. (2000). Conditioned place preference: what does it add to our preclinical understanding of drug reward? *Psychopharmacology* **153**, 31-43.

2. Fagen, R. (1981). *Animal play behavior*, Oxford University Press, Oxford.

3. Homberg, J.R., Schiepers, O.J.G., Schoffelmeer, A.N.M., Cuppen, E., Vanderschuren, L.J.M.J. (2007). Acute and constitutive increases in central serotonin levels reduce social play behaviour in peri-adolescent rats. *Psychopharmacology* **195**, 175-182.

4. Niesink, R.J.M. and Van Ree, J.M. (1989). Involvement of opioid and dopaminergic systems in isolation- induced pinning and social grooming of young rats. *Neuropharmacology* **28**, 411-418.

5. Panksepp, J., Siviy, S.M., Normansell, L. (1984). The psychobiology of play: theoretical and methodological perspectives. *Neurosci. Biobehav. Rev.* **8**, 465-492.

6. Pellis, S.M. and Pellis, V.C. (2009). *The playful brain: venturing to the limits of neuroscience*, Oneworld Publications, Oxford.

7. Špinka, M., Newberry, R.C., Bekoff, M. (2001). Mammalian play: training for the unexpected. *Q. Rev. Biol.* **76**, 141-168.

8. Trezza, V., Baarendse, P.J.J., Vanderschuren, L.J.M.J. (2009). Prosocial effects of nicotine and ethanol in adolescent rats through partially dissociable neurobehavioral mechanisms. *Neuropsychopharmacol.* **34**, 2560-2573.

9. Trezza, V., Baarendse, P.J.J., Vanderschuren, L.J.M.J. (2010). The pleasures of play: pharmacological insights into social reward mechanisms. *Trends Pharmacol. Sci.* **31**, 463-469.

10. Trezza, V., Damsteegt, R., Achterberg, E.J.M., Vanderschuren, L.J.M.J. (2011). Nucleus accumbens m-opioid receptors mediate social reward. *J. Neurosci.* **31**, 6362-6370.

11. Trezza, V., Damsteegt, R., Vanderschuren, L.J.M.J. (2009). Conditioned place preference induced by social play behavior: parametrics, extinction, reinstatement and disruption by methylphenidate. *Eur. Neuropsychopharmacol.* **19**, 659-669.

12. Trezza, V. and Vanderschuren, L.J.M.J. (2008a). Bidirectional cannabinoid modulation of social behavior in adolescent rats. *Psychopharmacology* **197**, 217-227.

13. Trezza, V. and Vanderschuren, L.J.M.J. (2008b). Cannabinoid and opioid modulation of social play behavior in adolescent rats: Differential behavioral mechanisms. *Eur. Neuropsychopharmacol.* **18**, 519-530.

14. Trezza, V. and Vanderschuren, L.J.M.J. (2009). Divergent effects of anandamide transporter inhibitors with different target selectivity on social play behavior in adolescent rats. *J. Pharmacol. Exp. Ther.* **328**, 343-350.

15. Tzschentke, T.M. (1998). Measuring reward with the conditioned place preference paradigm: A comprehensive review of drug effects, recent progress and new issues. *Prog. Neurobiol.* **56**, 613-672.

16. Van Kerkhof, L.W.M., Damsteegt, R., Trezza, V., Gao, P., Voorn, P., Vanderschuren, L.J.M.J. (2011). Corticostriatal systems activated during social play behavior in adolescent rats. *Soc. Neurosci. Abstr.* **943**, 20.

17. Vanderschuren, L.J.M.J. (2010). How the brain makes play fun. *Am. J. Play* **2**, 315-337.

18. Vanderschuren, L.J.M.J., Niesink, R.J.M., Spruijt, B.M., Van Ree, J.M. (1995). Effects of morphine on different aspects of social play in juvenile rats. *Psychopharmacology* **117**, 225-231.

19. Vanderschuren, L.J.M.J., Niesink, R.J.M., Van Ree, J.M. (1997). The neurobiology of social play behavior in rats. *Neurosci. Biobehav. Rev.* **21**, 309-326.

20. Vanderschuren, L.J.M.J., Trezza, V., Griffioen-Roose, S., Schiepers, O.J.G., Van Leeuwen, N., De Vries, T.J., Schoffelmeer, A.N.M. (2008). Methylphenidate disrupts social play behavior in adolescent rats. *Neuropsychopharmacology* **33**, 2946-2956.

Ultrasonic Communication in Mouse Models of Autism

Markus Wöhr

Behavioral Neuroscience, Experimental and Physiological Psychology, Philipps-University of Marburg, Marburg, Germany. markus.woehr@staff.uni-marburg.de

Abstract

Autism is a complex neurodevelopmental disorder characterized by 1) aberrant reciprocal social interactions, 2) deficits in social communication, and 3) repetitive, stereotyped patterns of behaviors, along with narrow restricted interests. Designing mouse behavioral assays with face validity to the three core symptoms of autism in humans is a daunting challenge for behavioral neuroscientists. Mice produce vocalizations in the ultrasonic range, i.e. clearly above the human frequency threshold and therefore not audible to humans. Such ultrasonic vocalizations serve as situation-dependent affective signals and convey important communicative information**.** Measuring the emission of ultrasonic vocalizations and the behavioral responses to ultrasonic vocalizations in playback experiments provides therefore an unique tool to study communication in mice. This is of high relevance for mouse models of autism, since delayed language and poor communication skills are fundamental to the diagnosis of autism.

Introduction

Autism is a complex neurodevelopmental disorder characterized by 1) aberrant reciprocal social interactions, 2) deficits in social communication, and 3) repetitive, stereotyped patterns of behaviors, along with narrow restricted interests (DSM-IV-TR, 2000). While the causes of autism remain unknown, a strong genetic component in the etiology of autism is indicated by extraordinarily high [1-3]. Genome-wide and pathway-based association studies led to the identification of more than hundred autism candidate genes [4-6]. Considerable efforts are made to better understand the genetic causes of autism. The generation of genetically modified mice is a powerful tool to study the physiological functions of such candidate genes. Mouse models with homologous mutations, e.g. Shank3 knockout mice, have revealed behavioral phenotypes relevant to the three autism core symptoms [7,8].

Designing mouse behavioral assays with face validity to the three core symptoms of autism in humans is a daunting challenge for behavioral neuroscientists. While reliable tests are available to assess the first core symptom, social deficits, and third core symptom, repetitive behavior, in mice, it is particularly difficult to measure communication deficits in mice [9].

Mice produce vocalizations in the ultrasonic range, i.e. clearly above the human frequency threshold and therefore not audible to humans. Such ultrasonic vocalizations serve as situation-dependent affective signals [10,11]. Currently, three types of ultrasonic vocalizations are known in mice. Their occurrence is dependent on the animal's age. 1) Isolation-induced ultrasonic vocalizations are emitted by mouse pups during the first two weeks of life when separated from mother and littermates [12]. Isolation-induced ultrasonic vocalizations were the first purely ultrasonic signals reported to be produced by mice. They were discovered by Zippelius and Schleidt in 1956. Zippelius and Schleidt suggested that isolation-induced ultrasonic vocalizations serve communicative purposes, since they have observed that mothers leave the nest to retrieve vocalizing pups scattered outside the nest, whereas no retrieval behavior was seen in response to pups that have been anesthetized and hence did not emit isolation-induced ultrasonic vocalizations. Later, their communicative function was confirmed by playback experiments [13-15]. 2) Interaction-induced ultrasonic vocalizations occur during social investigation in juvenile mice aged three to four weeks [16]. Because their occurrence is positively associated with social investigation behavior, it was suggested that interaction-induced ultrasonic vocalizations help to maintain social contact [16]. 3) Female-induced ultrasonic vocalizations are uttered by adult male mice when exposed to females [17] or female urine [18]. As shown in devocalization and playback studies, female-induced ultrasonic vocalizations serve an important communicative function as well, namely to attract females [19,20].

Measuring the emission of ultrasonic vocalizations and the behavioral responses to ultrasonic vocalizations in playback experiments provides therefore an unique tool to study communication in mice. This is of high relevance for mouse models of autism, since delayed language and poor communication skills are fundamental to the diagnosis of autism.

Methods

Isolation-induced ultrasonic vocalizations in pups: For the induction of isolation-induced ultrasonic vocalizations, pups are separated from mother and littermates under room temperature for 10 minutes during the first two weeks of life. Pups are removed individually from the nest at random and gently placed into an isolation box made of plastic (an example is shown in Figure 1; [21]) or an isolation container made of glass, containing clean bedding material [22,23]. Emission of ultrasonic vocalizations is monitored by an UltraSoundGate Condenser Microphone CM 16 (Avisoft Bioacoustics, Berlin, Germany) placed in the roof of the sound attenuating box. The microphone is connected via an UltraSoundGate 116 USB audio device (Avisoft Bioacoustics) to a personal computer, where acoustic data are recorded with a sampling rate of 250,000 Hz in 16 bit format by Avisoft RECORDER (version 2.97; Avisoft Bioacoustics). The microphone that is used for recording is sensitive to frequencies of 15-180 kHz with a flat frequency response (±6 dB) between 25-140 kHz.

Interaction-induced ultrasonic vocalizations in juveniles: For the induction of interaction-induced ultrasonic vocalizations pairs of mice are allowed to socially interact after a short period of social deprivation. For acoustic recording, the same equipment can be used as for isolation-induced ultrasonic vocalizations (Avisoft Bioacoustics).

Female-induced ultrasonic vocalizations in adult males: For the induction of female-induced ultrasonic vocalizations adult male mice are exposed to a drop of female urine. To elicit high rates of ultrasonic vocalizations female urine needs to be fresh [24]. Also, male mice with female experience emit more ultrasonic

Figure 1. **Setup for measuring isolation-induced ultrasonic vocalizations in rodent pups.** The roof and one wall are made of transparent plastic to allow video observation during the test. The isolation box is placed in a sound attenuating isolation cubicle (51x71x51 cm; Coulbourn Instruments, Allentown, USA) equipped with 2 white-light LED spots (63 lux, Conrad Electronic GmbH, Hirschau, Germany) and a black/white CCD camera (Conrad Electronic GmbH) connected to a DVD recorder (DVR-3100 S, Pioneer, Willich, Germany).

vocalizations than male mice without such an experience [24]. For acoustic recording, the same equipment can be used as for isolation-induced ultrasonic vocalizations (Avisoft Bioacoustics).

Acoustical analysis: For acoustical analysis, recordings are transferred to Avisoft SASLab Pro (version 4.50; Avisoft Bioacoustics) and a fast Fourier transform is conducted (512 FFT length, 100 % frame, Hamming window and 75 % time window overlap [21]). Correspondingly, the spectrograms are produced at 488 Hz of frequency resolution and 0.512 ms of time resolution. Call detection is provided by an automatic threshold-based algorithm (amplitude threshold: -40 dB) and a hold-time mechanism (hold time: 10 ms). A high-pass filter of 20 kHz is used to reduce background noise outside the relevant frequency band to 0 dB. The accuracy of call detection by the software is verified manually by an experienced user. When necessary, missed calls are marked by hand to be included in the automatic parameter analysis. Total number of ultrasonic vocalizations is calculated for the entire session and in 60 s time bins, to visualize the time course of the ultrasonic vocalization response. Additional parameters include peak frequency and peak amplitude, i.e. loudness, which are derived from the average spectrum of the entire call, are determined automatically. Peak amplitude is defined as the point with the highest energy within the spectrum. Peak frequency is defined as the frequency at the location of the peak amplitude within the spectrum. In addition, the extent of frequency modulation, i.e. the difference between the lowest and the highest peak frequency within each call, is measured automatically. Temporal parameters include latency to start calling, total calling time, and call duration.

Results

Results obtained provide an in-depth characterization of auditory communication in mice throughout life. Exemplary spectrograms are shown (see Figure 2).

Discussion

BTBR T+tf/J mice are currently among the most commonly used mouse models of autism [9]. In a series of experiments it was shown that all three types of ultrasonic vocalizations are affected in this mouse model. As

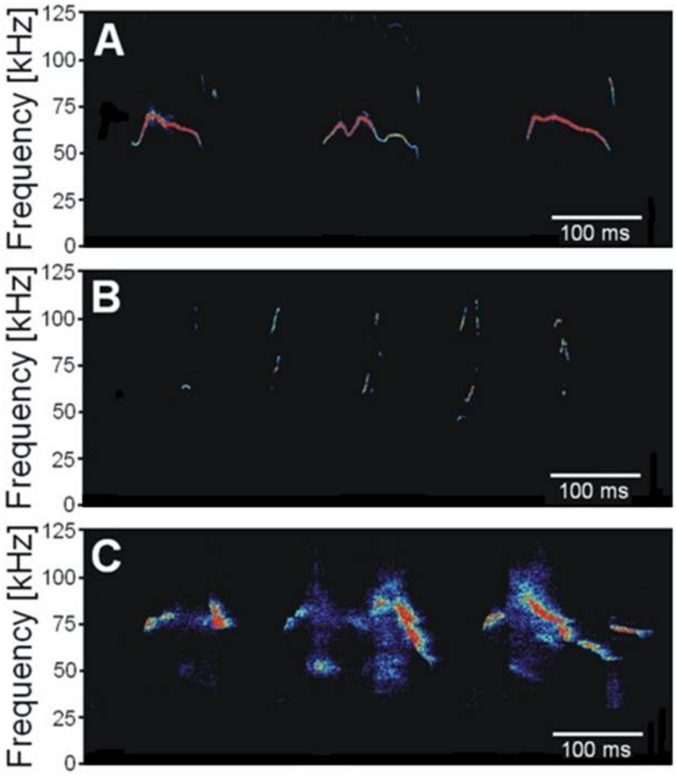


Figure 2. **Exemplary spectrograms of mouse ultrasonic vocalizations.** A) Isolation-induced ultrasonic vocalizations; B) Interaction-induced ultrasonic vocalizations; C) Female-induced ultrasonic vocalizations.

pups, BTBR T+tf/J mice display an untypical repertoire of isolation-induced ultrasonic vocalizations [25]. As juveniles, reduced social interaction behavior is paralleled by a reduction in interaction-induced ultrasonic vocalizations [26]. As adults, male BTBR T+tf/J mice displayed minimal ultrasonic vocalization responses to female urine obtained from both BTBR T+tf/J and C57BL/6J, a standard mouse strain showing high levels of social behavior [27]. The lack of female-induced ultrasonic vocalizations was associated with untypically low scent marking, a measure of olfactory communication [27]. This indicates that the BTBR T+tf/J mouse model of autism incorporates phenotypes relevant to the second diagnostic symptom of autism, communication deficits, along with its strong behavioral phenotypes relevant to the first and third diagnostic symptoms, impairments in social interactions and high levels of repetitive behavior [24].

Candidate genes for autism include genes encoding for the SHANK family of synaptic scaffolding proteins. Mutations in *SHANK3* and deletions containing *SHANK3* have been detected in several autistic individuals [28]. Despite behavioral phenotypes with relevance to the first and third diagnostic symptoms of autism, ultrasonic communication behavior was found to be only mildly impaired in Shank3 knockout mice [7,8]. Shank1 is a related scaffolding protein within the postsynaptic density, promoting the morphological and functional maturation of synapses. When tested for isolation-induced USVs as pups, Shank1 null mutants emitted fewer USVs as compared to wild-type littermates; and as adults in response to female urine, the USV production by Shank1 null mutant males was characterized by an unusual time pattern [23]. In addition, Shank1 null mutant males displayed lower levels of scent marking in proximity to the female urine spot [23].

Together, these data support the relevance of ultrasonic vocalizations for mouse models of neuropsychiatric disorders characterized by social deficits, such as autism. Importantly, the automated assessment of ultrasonic vocalizations at different time points during mouse development will allow high-throughput testing of social communicative behavior in mice.

References

1. Folstein, S., Rutter, M. (1977). Genetic influences and infantile autism. *Nature* **265**, 726-728.
2. Le Couteur, A., Bailey, A., Goode, S., Pickles, A., Robertson, S., Gottesman, I., Rutter, M. (1996). A broader phenotype of autism: the clinical spectrum in twins. *Journal of Child Psychology and Psychiatry* **37**, 785-801.
3. Polleux, F., Lauder, J.M. (2004). Toward a developmental neurobiology of autism. *Mental Retardation and Developmental Disabilities Research Reviews* **10**, 303-317.
4. Abrahams, B.S., Geschwind, D.H. (2008). Advances in autism genetics: on the threshold of a new neurobiology.*Nature Reviews Genetics* **9**, 341-355.
5. Betancur, C. (2011). Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. *Brain Research* **1380**, 42-77.
6. Persico, A.M., Bourgeron, T. (2006). Searching for ways out of the autism maze: genetic, epigenetic and environmental clues. *Trends in Neurosciences* **29**, 349-358.
7. Peça, J., Feliciano, C., Ting, J.T., Wang, W., Wells, M.F., Venkatraman, T.N., Lascola, C.D., Fu, Z., Feng, G. (2011). Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. *Nature* **472**, 437-442.
8. Yang, M., Bozdagi, O., Scattoni, M.L., Wöhr, M., Roullet, F.I., Katz, A.M., Abrams, D.N., Kalikhman, D., Simon, H., Woldeyohannes, L., Zhang, J.Y., Harris, M.J., Saxena, R., Silverman, J.L., Buxbaum, J.D., Crawley, J.N. (in press). Reduced excitatory neurotransmission and mild autism-relevant phenotypes in adolescent Shank3 null mutant mice. *The Journal of Neuroscience*.
9. Silverman, J.L., Yang, M., Lord, C., Crawley, J.N. (2010). Behavioural phenotyping assays for mouse models of autism. *Nature Reviews Neuroscience* **11**, 490-502.

10. Portfors, C.V. (2007). Types and functions of ultrasonic vocalizations in laboratory rats and mice. *Journal of the American Association for Laboratory Animal Science* **46**, 28-34.

11. Wöhr, M., Schwarting, R.K.W. (2010). Rodent ultrasonic communication and its relevance for models of neuropsychiatric disorders. *e-Neuroforum*, **4.10**, 71-80.

12. Zippelius, H.M., Schleidt, W.M. (1956). Ultraschall-Laute bei jungen Mäusen. *Naturwissenschaften* **43**, 502.

13. Ehret, G. (1987). Left hemisphere advantage in the mouse brain for recognizing ultrasonic communication calls. *Nature* **325**, 249-251.

14. Sewell, G.D. (1970). Ultrasonic communication in rodents. *Nature* **227**, 410.

15. Wöhr, M., Dahlhoff, M., Wolf, E., Holsboer, F., Schwarting, R.K.W., Wotjak, C.T. (2008). Effects of genetic background, gender, and early environmental factors on isolation-induced ultrasonic calling in mouse pups: An embryo-transfer study. *Behavior Genetics* **38**, 579-595.

16. Panksepp, J.B., Jochman, K.A., Kim, J.U., Koy, J.J., Wilson, E.D., Chen, Q., Wilson, C.R. & Lahvis, G.P. (2007). Affiliative behavior, ultrasonic communication and social reward are influenced by genetic variation in adolescent mice. *PLoS One* **2**, e351.

17. Sewell, G.D. (1967). Ultrasound in adult rodents. *Nature* **215**, 512.

18. Whitney, G., Aapern, M., Dizinno, G., Horowitz, G. (1974). Female odors evoke ultrasounds from male mice. *Animal Learning & Behavior* **2**, 13-18.

19. Hammerschmidt, K., Radyushkin, K., Ehrenreich, H., Fischer, J. (2009). Female mice respond to male ultrasonic 'songs' with approach behaviour. *Biology Letters* **5**, 589-592.

20. Pomerantz, S.M., Nunez, A.A., Bean, N.J. (1983). Female behavior is affected by male ultrasonic vocalizations in house mice. *Physiology & Behavior* **31**, 91-96.

21. Wöhr, M., Schwarting, R.K.W. (2008). Maternal care, isolation-induced infant ultrasonic calling, and their relations to adult anxiety-related behavior in the rat. *Behavioral Neuroscience* **122**, 310-330.

22. Kurz, A., Wöhr, M., Walter, M., Bonin, M., Auburger, G., Gispert, S., Schwarting, R.K.W. (2010). Alpha-synuclein deficiency affects brain Foxp1 expression and ultrasonic vocalization. *Neuroscience* **17**, 287-298.

23. Wöhr, M., Roullet, F.I., Hung, A.Y., Sheng, M., Crawley, J.N. (2011). Communication impairments in mice lacking Shank1: Reduced levels of ultrasonic vocalizations and scent marking behavior. *PLoS One* **6**, e20631.

24. Roullet, F.I., Wöhr, M., Crawley, J.N. (2011). Female urine-induced male mice ultrasonic vocalizations, but not scent-marking, is modulated by social experience. *Behavioural Brain Research* **216**, 19-28.

25. Scattoni, M.L., Gandhy, S.U., Ricceri, L., Crawely, J.N. (2008). Unusual repertoire of vocalizations in the BTBR T+tf/J mouse model of autism. *PLoS One* **3**, e3067.

26. Scattoni, M.L., Ricceri, L., Crawley, J.N. (2011). Unusual repertoire of vocalizations in adult BTBR T+tf/J mice during three types of social encounters. *Genes Brain & Behavior* **10**, 44-56.

27. Wöhr, M., Roullet, F.I., Crawley, J.N. (2011). Reduced scent marking and ultrasonic vocalizations in the BTBR T+tf/J mouse model of autism. *Genes Brain & Behavior* **10**, 35-43.

28. Durand, C.M., Betancur, C., Boeckers, T.M., Bockmann, J., Chaste, P., Fauchereau, F., Nygren, G., Rastam, M., Gillberg, I.C., Anckarsäter, H., Sponheim, E., Goubran-Botros, H., Delorme, R., Chabane, N., Mouren-Simeoni, M.C., de Mas, P., Bieth, E., Rogé, B., Héron, D., Burglen, L., Gillberg, C., Leboyer, M., Bourgeron, T. (2007). Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nature Genetics* **39**, 25-27.

Ultrasonic Communication in Rats: Insights from Playback Studies

Dominik Seffer, Rainer K.W. Schwarting, Markus Wöhr

Behavioral Neuroscience, Experimental and Physiological Psychology, Philipps-University of Marburg, Marburg, Germany. seffer@staff.uni-marburg.de

Abstract

Mice and rats produce vocalizations in the ultrasonic range, i.e. clearly above the human frequency threshold and therefore not audible to humans. Such ultrasonic vocalizations (USVs) serve as situation-dependent affective signals and convey important communicative functions. Here, we tested whether different acoustic stimuli, including aversive 22-kHz USVs and appetitive 50-kHz USVs, elicit distinct behavioral response patterns that can be differentiated by the automated video tracking system EthoVision. Results show that playback of acoustic stimuli elicited distinct patterns of behavioral activity. While playback of aversive 22-kHz USVs were followed by behavioral inhibition, appetitive 50-kHz USVs induced an increase in locomotor activity, which is directed towards stimulus source. The present findings further support the conclusion that aversive 22-kHz USVs serve an alarm function to warn conspecifics about external danger and that 50-kHz USVs serve a social affiliative function as social contact calls. The distinct behavioral response patterns elicited can be measured by using the automated video tracking system EthoVision. This will allow high-throughput testing of social communicative behavior in rats.

Introduction

Mice and rats produce vocalizations in the ultrasonic range, i.e. clearly above the human frequency threshold and therefore not audible to humans. Such ultrasonic vocalizations (USVs) serve as situation-dependent affective signals [1; 2]. Rodent USVs have attracted considerable interest as it was suggested that they can serve as a direct measure of affect, i.e. of the animal's subjective emotional state. Furthermore, as rodent USVs convey important communicative information, measuring them offers the opportunity to study communication in rodent models of neuropsychiatric disorders characterized by social and communication deficits such as autism and schizophrenia. In the rat, three USV types are known: 1) 40-kHz USVs, which are emitted by pups when isolated from mother and littermates; 2) 22-kHz USVs, which are emitted by juvenile and adult rats in aversive situations such as fear-conditioning or social defeat; and 3) 50-kHz USVs, which are emitted by juvenile and adult rats in appetitive situations such as rough-and-tumble play, mating or when tickled playfully. In the mouse, similar call types exist with the exception of 22-kHz USVs, which are not present in mice.

In aversive situations, such as fear learning, 22-kHz USV production was found to be strongly correlated with freezing behavior, the most commonly used fear measure in rodents, and was shown to be highly dependent on the situation's aversiveness. Rats exposed to higher shock intensities emitted more aversive USVs than rats exposed to lower shock intensities during fear learning and testing, i.e. in absence of shock application when exposed to a formerly neutral stimulus that has been repeatedly paired with shocks [3]. In addition to situational factors, individual differences in disposition for anxiety-related behavior play an important role. Rats characterized as anxious based on their anxiety-related behavior on the elevated plus maze emitted more aversive USVs during fear learning than less anxious rats [4]. In contrast, in appetitive situations, such as rough-and-tumble play, 50-kHz USV production could serve as indicators of high welfare. When tickled playfully by an experienced human experimenter, appetitive USV production was found to be highly correlated with the level of newborn cells in the hippocampus, a brain area important for learning and memory as well as emotion and stress regulation. Rats displaying a depressive-like phenotype with low rates of appetitive USVs were characterized by low levels of newborn cells [5].

Behavioral, neuronal and pharmacological studies provide evidence that USVs serve communicative purposes. It is known for a long time that 40-kHz USVs elicit maternal search and retrieval behavior. More recently, however, it was demonstrated that also 22-kHz USVs and 50-kHz USVs induce call-specific behavioral responses in the receiver [6]. While 22-kHz USVs induce freezing behavior, indicating an alarm function, 50-kHz USVs induce social approach behavior, supporting the notion that they serve as social contact calls. The opposite behavioral responses are paralleled by distinct patterns of brain activation [7]. While 22-kHz USVs induce activation in amygdala and periaqueductal gray, 50-kHz USVs are followed by activation in the nucleus accumbens. As indicated by subsequent pharmacological studies [8], social approach behavior in response to 50-kHz USVs depends on an intact opioid system.

Here, we tested whether different acoustic stimuli, including aversive 22-kHz USVs and appetitive 50-kHz USVs, elicit distinct behavioral response patterns that can be differentiated by an automated video tracking system (EthoVision, Noldus, Wageningen, The Netherlands).

Methods

Animals and housing: Male Wistar rats (HsdCpb:WU, Harlan-Winkelmann, Borchen, Germany) served as subjects. Animals were housed in groups on Tapvei peeled aspen bedding (indulab ag, Gams, Switzerland) in a Macrolon type IV cage (size: 378x217x180 mm, plus high stainless steel covers). Lab chow (Altromin, Lage, Germany) and water (0.0004% HClsolution) were available ad libitum. Animals were housed in an animal room with a 12:12 h light/dark cycle (lights on 7–19 h) where the environmental temperature was maintained between 20–25° Celsius. Prior to testing, all animals were handled for 3 days in a standardized way (5 min each day).

All experimental procedures were approved by the local government ethical committee (Regierungspräsidium Giessen, Germany; MR 20/35, Nr. 50/2010).

Experimental setting: Testing was performed on a radial maze of gray plastic with 8 arms (9.8x40.5 cm) extending radially from a central platform (diameter: 24 cm), which was elevated 52 cm above the floor (see Figure 1 & 2; for details see: [9]). Acoustic stimuli were presented through an ultrasonic speaker (ScanSpeak, Avisoft Bioacoustics, Berlin, Germany; see Figure 3) using an external sound card with a sampling rate of 192 kHz (Fire Wire Audio Capture FA-101, Edirol, London, UK) and a portable ultrasonic power amplifier with a frequency range of 1–125 kHz (Avisoft Bioacoustics). The loudspeaker had a frequency range of 1–120 kHz with a relatively flat frequency response (± 12 dB) between 15–80 kHz. It was placed 20 cm away from the end of one arm at a height of 52 cm above the floor. Testing was performed under red light (approximately 11 lux in the center of the maze and between 9 and 12 lux in the arms) in a testing room with no other rats present. All behavioral tests were conducted between 9–17 h. Prior to each test, behavioral equipment was cleaned using a 0.1 % acetic acid solution followed by drying.

Behavior was monitored by a video camera (Panasonic WV-BP 330/GE, Hamburg, Germany) from about 150 cm above the maze, which fed into DVD recorder (DVR-3100 S, Pioneer, Willich, Germany). Playback of acoustic stimuli were monitored by two UltraSoundGate Condenser Microphones (CM 16; Avisoft Bioacoustics; see Figure 3) placed 20 cm away from the maze at a height of 55 cm above the floor. One out of these two was placed next to the loudspeaker, i.e. in front of the three proximal arms, whereas the other one was placed vis-à-vis in front of the three distal arms. These microphones were sensitive to frequencies of 15-180 kHz with a flat frequency response (± 6 dB) between 25–140 kHz, and were connected via an Avisoft UltraSoundGate 416 USB Audio device (Avisoft Bioacoustics) to a personal computer, where acoustic data were displayed in real time by Avisoft RECORDER (version 2.7; Avisoft Bioacoustics), and were recorded with a sampling rate of 214,285 Hz in 16 bit format.

Acoustic stimuli: The following three acoustic stimuli were used: 1) 22-kHz USVs, 2) 50-kHz USVs, and background noise (see Figure 4). All stimuli were presented for 1 min with a sampling rate of 192 kHz in 16 bit format. USVs were presented at about 69 dB (measured from a distance of 40 cm), and noise was presented with about 50 dB, which corresponds to the background noise during playback of the other stimuli. The 22-kHz USVs had been recorded from a male Wistar rat after applications of foot-shocks (for setting and recording see: [3]). The 50-kHz USVs had been recorded from a male Wistar rat during exploration of a cage containing scents from a cage mate (for setting and recording see: [10]). The presentation was composed of a sequence of 3.5 s, which was repeated for 1 min, i.e. 17 times, to assure the presentation of a high number of frequency-modulated calls

Figure 1. Setup for playback experiments and red light device. Elevated eight arm radial maze equipped with two ultrasonic microphones and two loudspeakers (Avisoft Bioacoustics; Berlin Germany). The rat's behavioral responses during playback of ultrasonic stimuli are recorded with a video camera positioned above the elevated eight arm radial maze. Testing is performed under dim red light.

within a relatively short period of time. Since the two acoustic stimuli presented contained background noise, i.e. sounds, which occur when a rat is exploring an arena with bedding, background noise without calls was presented to control for its possible effects.

Experimental procedure: A given animal was placed onto the central platform of the radial maze, facing the arm opposite to the loudspeaker. After an initial phase of 15 min where no acoustic stimuli were presented (habituation), the rat was exposed to three presentations of acoustic stimuli for 1 min, each followed by an inter-stimulus-interval of 10 min.

Behavioral and acoustic analysis: Behavioral analysis was performed using an automated video tracking system (EthoVision, Noldus Information Technology, Wageningen, The Netherlands). Previously, the following parameters were obtained: 1) total distance travelled (cm), 2) number of arm entries into the three proximal or distal arms (n) and 3) the time spent thereon (6; 8). Here, the track profile was acquired. For the automated

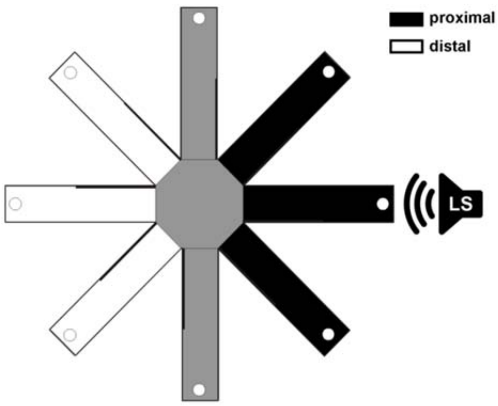


Figure 2. Schematic illustration of the elevated radial eight arm maze. Arms close to the loudspeaker are denoted as proximal, central arms as neutral and opposite arms as distal.

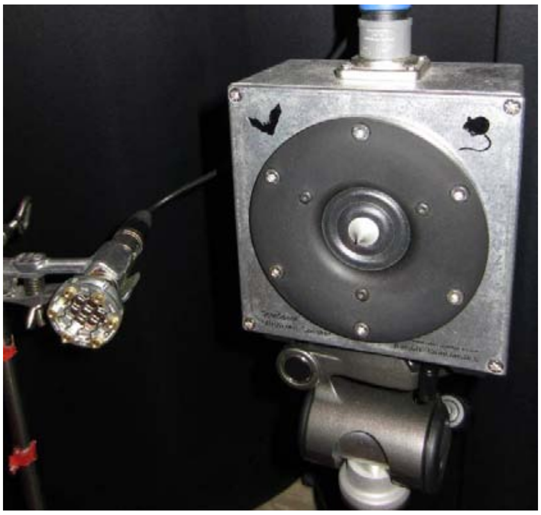


Figure 3. Ultrasonic loudspeaker for playback presentations and microphone for ultrasonic recording.

analysis, input filters were activated to avoid an over-estimation of locomotor activity due to head-movements: a minimal distance moved of 8 cm was used for the total distance travelled, whereas a minimal distance moved of 3 cm was used for the arm entries.

For acoustical analysis, recordings were transferred to SASLab Pro (version 4.38; Avisoft Bioacoustics) and a fast Fourier transform was conducted (512 FFT-length, 100 % frame, Hamming window and 75 % time window overlap). Correspondingly, the spectrograms were produced at 488 Hz of frequency resolution and 0.512 ms of time resolution.

Results

Playback of acoustic stimuli elicited distinct patterns of behavioral activity (see Figure 5). While playback of aversive 22-kHz USVs were followed by behavioral inhibition, appetitive 50-kHz USVs induced an increase in locomotor activity, which is directed towards stimulus source. Playback of noise led to an undirected increase in behavioral activity. The results show that different acoustic stimuli, including aversive 22-kHz USVs and appetitive 50-kHz USVs, elicit distinct behavioral response patterns that can be differentiated by an automated video tracking system (EthoVision).

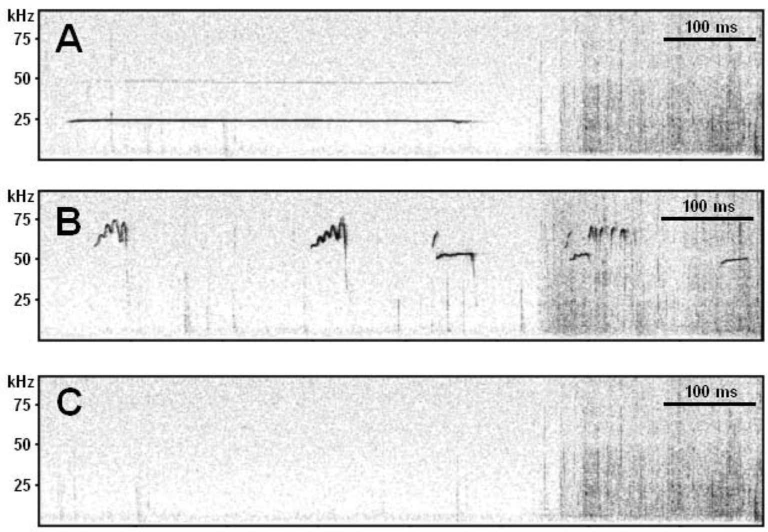


Figure 4. Exemplary spectrograms of the presented acoustic stimuli. A) 22-kHz USVs, B) 50-kHz USVs, and C) background noise.

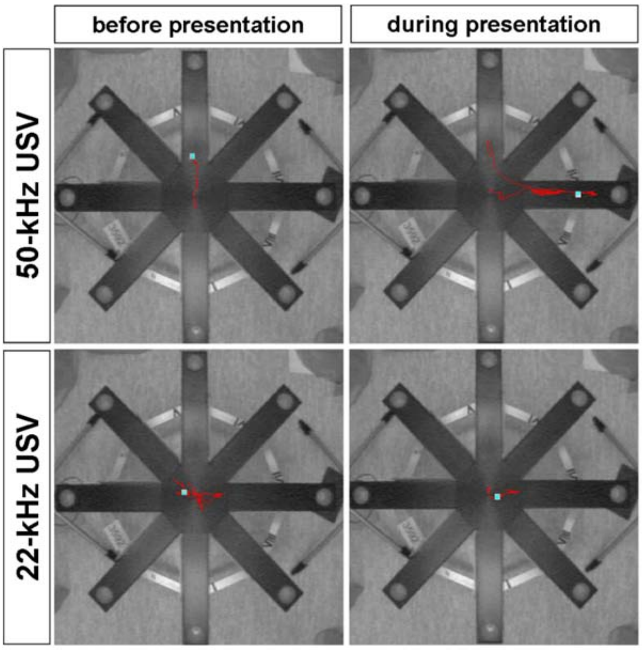


Figure 5. Exemplary patterns of behavioral activity measured by using EthoVision.

Discussion

In line with previous findings [6; 7; 8], it was found that aversive 22-kHz USVs and appetitive 50-kHz USVs induce call-specific behavioral responses in the receiver. The present findings further support the conclusion that aversive 22-kHz USVs serve an alarm function to warn conspecifics about external danger and that 50-kHz USVs serve a social affiliative function as social contact calls. Importantly, the present findings demonstrate that the behavioral response patterns elicited by aversive 22-kHz USVs and appetitive 50-kHz USVs can be measured by using an automated video tracking system (EthoVision). This will allow high-throughput testing of social communicative behavior in rats.

Social communicative deficits occur in a number of neuropsychiatric disorders, among them neurodevelopmental disorders such as autism and schizophrenia. Autism is characterized by aberrant reciprocal social interactions, deficits in social communication, and repetitive, stereotyped patterns of behaviors (DSM-IV-TR, 2000); and there are overlapping symptoms between autism and schizophrenia, particularly the negative symptoms such as poverty of speech and social withdrawal (DSM-IV-TR, 2000). To study underlying alterations in genes, neurotransmitter systems, neuronal plasticity, etc., the development of new rodent behavioral assays with face validity to social and communication deficits seen in neurodevelopmental disorders is crucial.

Since rats are social animals and rough-and-tumble play during adolescence has an important role for social development, separation from conspecifics during this phase is known to impair social behavior. Post-weaning social isolation in rats is a widely used animal model to induce behavioral effects and developmental changes in the nervous system related to symptoms in psychiatric disorders like schizophrenia [11; 12]. It is therefore of great interest to see whether the newly developed rat social communication assay is sensitive enough to detect deficits in social communication induced by juvenile social isolation.

References

1. Portfors, C.V. (2007). Types and functions of ultrasonic vocalizations in laboratory rats and mice. *Journal of the American Association for Laboratory Animal Science* **46**, 28-34.

2. Wöhr, M., Schwarting, R.K.W. (2010). Rodent ultrasonic communication and its relevance for models of neuropsychiatric disorders. *e-Neuroforum*, **4.10**, 71-80.

3. Wöhr, M., Borta, A., Schwarting, R.K.W. (2005). Overt behavior and ultrasonic vocalization in a fear conditioning paradigm - a dose-response study in the rat. *Neurobiology of Learning and Memory* **84**, 228-240.

4. Borta, A., Wöhr, M. & Schwarting, R.K.W. (2006). Rat ultrasonic vocalization in aversively motivated situations and the role of individual differences in anxiety-related behavior. *Behavioural Brain Research*, **166**, 271-280.

5. Wöhr, M., Kehl, M., Borta, A., Schänzer, A., Schwarting, R.K.W. & Höglinger, G.U. (2009). New insights into the relationship of neurogenesis and affect: Tickling induces hippocampal cell proliferation in rats emitting appetitive 50-kHz ultrasonic vocalizations. *Neuroscience*, **163**, 1024-1030.

6. Wöhr, M., Schwarting, R.K.W. (2007). Ultrasonic communication in rats: can playback of 50-kHz calls induce approach behavior? *PLoS ONE* **2**, e1365.

7. Sadananda, M., Wöhr, M., Schwarting R.K.W. (2008). Playback of 22-kHz and 50-kHz ultrasonic vocalizations induces differential c-fos expression in rat brain. *Neuroscience Letters* **435**, 17-23.

8. Wöhr, M., Schwarting, R.K.W. (2009). Ultrasonic communication in rats: effects of morphine and naloxone on vocal and behavioral responses to playback of 50-kHz vocalizations. *Pharmacology Biochemistry and Behavior* **94**, 285-295.

9. Görisch, J., Schwarting, R.K.W. (2006). Wistar rats with high versus low rearing activity differ in radial maze performance. *Neurobiology of Learning and Memory* **86**, 175-187.

10. Wöhr, M., Houx, B., Schwarting, R.K.W., Spruijt, B. (2008). Effects of experience and context on 50-kHz vocalizations in rats. *Physiology and Behavior* **93**, 766-776.

11. Fone, K., Porkess, V. (2008). Behavioral and neurochemical effects of post-weaning social isolation in rodents – Relevance to developmental neuropsychiatric disorders. *Neuroscience and Biobehavioral Reviews* **32**, 1087–1102.

12. Lapiz, M.D., Fulford, A., Muchimapura, S., Mason, R., Parker, T., Marsden, C.A. (2003). Influence of postweaning social isolation in the rat on brain development, conditioned behavior, and neurotransmission. *Neuroscience & Behavioral Physiology* **33**, 13-29.

Categorizing Vocal Repertoires of Nonhuman Primates

Kurt Hammerschmidt, Philip Wadewitz

Cognitive Ethology Lab, German Primate Center, Göttingen, Germany.
kurt.hammerschmidt@cog-ethol.de

Abstract

The vocal repertoire of animals consists typically of a limited number of call types which vary substantially within these categories. One can find all variation from highly graded to more or less distinct vocal repertoires, and it may be unavoidable that call variants at category boundaries are difficult to categorize. This must not be a disadvantage for the communication of animals which grew up in the same community, because the efficiency of categorical perception allows the receiver to respond correctly even in highly graded signaling systems. However, to understand the evolution of acoustic communication and to compare studies in bioacoustics it is necessary to have comparable units, or to know how different categorization levels influence the outcome of an acoustic analysis.

In bioacoustics it is not uncommon to use the visual cues of spectrograms to establish vocal repertoires. In some cases the visual approach is support by a parametric description of the Fourier transformed wave signals. In other cases the classification of calls is guided according to their proposed function, grouping them to ‘alarm’, ‘contact’ or ‘food calls’. All these procedure have the disadvantages that they rely on prior assumptions. In cases of distinct vocal repertoires a time-consuming statistical analysis seems to be dispensable. However, vocal repertoires of most mammals consist of call types which vary substantially within their categories and it is difficult to identify reliable boundaries without any statistical approach. Therefore, it seems to be important to have more objective procedures to establish vocal repertoires.

In a case study of Barbary macaques’ vocal repertoire Hammerschmidt and Fischer [1] have tested a k-means cluster approach to establish vocal categories. As k-means cluster procedures do not directly present the optimal cluster solution, they calculated two indices to assess the power of the different cluster solutions: First, they calculated $Eta_{(k)}$ which reflects the relative reduction in variance by a given cluster solution in relation to the overall variation of the unpartitioned data set ('Zero-model'). With a growing number of clusters Eta increases successively. Therefore, it is mainly of interest to identify peaks in the course of Eta, which indicate promising cluster solutions. This can be done by a coefficient (pre) which describes the relative improvement in the reduction of variance compared to the previous solution ($pre_{(k)}$=1-(sum of $squares_{(k)}$/ sum of $squares_{(k-1)}$).

The most difficult part conducting such cluster approaches is the identification of relevant acoustic variables. A high number of variables make it difficult for cluster algorithms to find appropriate cluster centers. In addition, acoustic parameters are often highly correlated which can shift the result in the direction of these parameters. A principle component analysis (PCA) can be a useful strategy to reduce the number of acoustic parameters. In addition the factors of a PCA are uncorrelated and can be used instead of the original acoustic variables. However, cluster analyses based on factor loadings have a tendency to find a lower number of clusters. Therefore, we decided to estimate the best cluster solutions for all possible combination of characteristic acoustic variables.

In a last approach we focused on question whether classical cluster algorithms based on Euclidean distance measures are really the appropriate tool to describe vocal repertoires of nonhuman primates. As mentioned at the beginning the categorical perception of a receiver must not work in Euclidean fashion. To achieve this goal we replaced the Euclidean distance measure by a log-likelihood distance measure. This probability based distance has the additional advantage to be able to handle both continuous and categorical variables. We used a two-step cluster procedure and took the Schwarz-Bayesian criterion (BIC) as measures of the best cluster solution. In the talk we will present data of vocal repertoires of Barbary macaques and Baboons as examples of graded and discrete vocal repertoires and compare the results of the different approaches.

References

1. Hammerschmidt, K., Fischer, J. (1998). The vocal repertoire of Barbary macaques: A quantitative analysis of a graded signal system. *Ethology* **104**, 203-216.

Central Neuropeptides Social Recognition, Social Preference and Social Fear in Rodents

M. Lukas, I. Toth, I.D. Neumann

Department of Behavioral and Molecular Neurobiology, University of Regensburg, Regensburg, Germany. michael.lukas@ur.de

Abstract

Highly conserved brain neuropeptides, such as oxytocin (OT), vasopressin (AVP) and the newly described neuropeptide S, have the potential to regulate elementary social abilities, i.e. social approach and social preference, social fear as well as individual social memory. These social skills are essential for functioning social communication among conspecifics. To provide evidence for neuropeptide-dependent social abilities we established and further developed, respectively, various behavioral assays like the social recognition/social discrimination, the social preference and the social fear conditioning paradigms. The involvement of neuropeptides in these behaviors is investigated using central pharmacological manipulation and in vivo microdialysis.

Introduction

Besides of the general interest in social behavior regulation by neuropeptides, there is substantial interest in the regulation of abnormal sociability accompanying various psychopathologies including depression, bipolar disease, autism or anxiety disorders. To reveal the neurobiological underpinnings and to further understand the involvement of OT and AVP also from a translational perspective in human sociability, appropriate animal models are essential.

Here, we will describe our efforts to establish and employ rodent models of social behaviors resembling human behaviors including the social recognition (social discrimination paradigm), a test for social approach, as well as an experimental setup to test for social fear using a social fear conditioning paradigm in rats and mice where appropriate.

Methods and Paradigms

Social discrimination. The ability of adult male rats and mice to discriminate between a previously encountered (same) and a novel con-specific was tested in the social discrimination test according to Engelmann et al. [1] with some important modifications [2]. The social stimulus, i.e. a male juvenile, was introduced into the home cage of the experimental male rat or mouse for 4 min (acquisition period). After a defined inter-exposure interval (consolidation period), the same stimulus was re-introduced along with a novel conspecific for 4 min (retrieval period). The percentage of time investigating the same and the novel stimulus animal during the retrieval period was calculated. A lower investigation time directed towards the same versus the novel stimulus was interpreted as social recognition and memory.

Social preference. As recently described [3], adult male rats or mice were placed in a novel arena allowing a 30- sec habituation period. Then, a small empty wire-mesh cage (object stimulus) was placed at one side wall of the arena for 4 min (rats) or 2.5 min (mice). The empty cage was then exchanged by an identical cage containing an unknown male con-specific (social stimulus) for additional 4 min (rats) or 2.5 min (mice). Before each trial, the arena was cleaned with water containing a low concentration of detergent. In rats, all tests took place during the active phase starting one hour after lights off using an unknown male rat as social stimulus. Mice were tested in the early light phase (one hour after lights on) using an unknown male mouse as social stimulus.

Social fear conditioning. As described in great detail recently [4], social fear can be characterized using a novel social fear conditioning paradigm for mice. For social fear c*onditioning (day 1),* mice were placed in the conditioning chamber and, after a 30-s adaptation period, an empty wire mesh cage was placed as a non-social stimulus near one of the short walls. The mice were allowed to investigate the non-social stimulus for 3 min, before it was replaced by an identical cage containing an unfamiliar male mouse. Unconditioned mice were allowed to investigate the social stimulus for 3 min. Conditioned mice were given a 1-s electric foot shock (0.7 mA, pulsed current) each time they investigated the social stimulus, defined by direct contact with the mouse. The time mice spent investigating the non-social stimulus served as a pre-conditioning measure of non-social anxiety. *Extinction (day 2 or 15).* To investigate whether conditioned mice displayed social fear and whether this fear could be extinguished, social investigation was assessed in the home cage 1 or 15 days after social fear conditioning for short-term and long-term social fear, respectively. *Extinction recall (day 3 or 16).* To investigate whether repeated exposure to social stimuli during extinction leads to a complete reversal of social fear, social investigation was assessed in the home cage 1 day after extinction. Extinction recall consisted of exposing the mice to six unfamiliar social stimuli for 3 min, with a 3-min inter-exposure interval.

Pharmacological manipulation. In order to manipulate the respective neuropeptide system, neuropeptide synthetic agonists or antagonists were infused either icv or directly into a brain target region via previously implanted guide cannulas (2 mm above the right lateral ventricle or bilaterally above the region of interest). Drugs were infused mostly 10 to 20 min before behavioral testing.

Animals received either an icv or a bilateral infusion into the region of interest of either synthetic OT/AVP (exogenous stimulation; rats: 0.01-0.1 µg), a selective receptor antagonist (blockade of endogenous release; rats: 0.1-0.75 µg, mice: 2-20µg), or vehicle via an infusion cannula.

Intracerebral microdialysis. For measurement of intracerebral oxytocin or vasopressin release during ongoing behavioral testing in the social discrimination paradigm we implanted a microdialysis probe stereotaxically in the respective target region (lateral septum). Two days later, five consecutive 30-min dialysates (3.3 µl/min, Ringer, pH 7.4) were collected during the social discrimination procedure: dialysates 1 and 2 were taken under baseline (undisturbed) conditions, dialysate 3 during the social memory acquisition period starting with the 4-min exposure to the first 3-week-old rat, and dialysates 4 and 5 were taken thereafter. Microdialysates were collected and vasopressin/oxytocin content was quantified using a highly sensitive radioimmunoassay.

Involvement of central neuropeptides in social behaviors

We provide further evidence for the a prominent role of OT and AVP in basic social behaviors, i.e. social approach and social recognition in male rats and mice, that are prerequisites for diverse social behaviors like aggression, sexual behavior, and pair bonding. Furthermore, we present a new model for measuring conditioned social fear in mice.

Naturally, male rats and mice demonstrate a preference for social stimuli over non-social stimuli. We demonstrated that the social preference behavior in both rats and mice is dependent on the endogenous OT system. Thereby, we provided first evidence for the involvement of endogenous OT in social approach behavior and affiliation towards a same-sex social stimulus in rodents, which is thereby not linked to reproductive behavior. We also demonstrated that social stressors, like acute social defeat, significantly impair social preference. Social defeat caused social avoidance in male rats when tested in the social preference paradigm, and synthetic OT was able to reverse such social deficits in rats, but could not further elevate social preference behavior in normal, undefeated rats.

Concerning the involvement of OT and AVP in rodent social cognition, we demonstrated that brain OT is required for the consolidation of social memory in male rats and mice. We demonstrate that OT receptor activation in the lateral septum is needed for social memory, whereas OT receptor activation in the medial amygdala is important for social memory for adult females only. This suggests a key role for the medial amygdala in mediating social memory depending on the biological relevance of the social stimulus.

Endogenous AVP release into the lateral septum of male rats was measured, accompanying social memory acquisition during social discrimination. This not only supports the involvement of septal AVP in male social recognition, but also provides evidence that septal AVP release can also be triggered during a non-aggressive encounter.

We also measured AVP release within the lateral septum of postnatally stressed male rats (maternal separation) during social recognition. We found a blunted AVP release in response to the initial social stimulus during social recognition in these postnatally stressed rats. Control experiments, using non-social stimulation, proved that this blunted release was specifically related to the social stimulation. Finally, the impairment in social recognition was causally linked to the blunted AVP release during social memory acquisition, as administration of synthetic AVP enhanced social recognition abilities of stressed rats to the level of unstressed rats. These findings strongly imply a link between a deficient central AVP signaling and social cognition deficits in rats.

Neuropeptide S tested in the social discrimination and social preference paradigms had no effects on social behavior.

Using the social fear conditioning paradigm we can show that by administering electric foot shocks during the investigation of a con-specific (social fear conditioning), social investigation can be severely reduced for both short- and long-term periods. In contrast, other behaviors such as fear of novelty, general anxiety, depressive-like behavior, or locomotion were found unchanged, which is indicative of specific social fear as a result of social fear conditioning. Both diazepam and paroxetine were able to reverse social fear thus validating the model.

Thus, our models of social preference/social avoidance, and of social fear conditioning provide unique animal models for increasing our understanding of normal and dysfunctional social behaviours, in particular of social fear as element of social anxiety disorders and for developing and testing novel treatment options.

References

1. Engelmann, M., Wotjak, C.T., Landgraf, R. (1995). Social discrimination procedure: an alternative method to investigate juvenile recognition abilities in rats. *Physiol. Behav.* **58**(2), 315-321.

2. Lukas, M., Bredewold, R., Landgraf, R., Neumann, I.D., Veenema, A.H. (2011). Early life stress impairs social recognition due to a blunted response of vasopressin release within the septum of adult male rats. *Psychoneuroendocrinology* **36**(6), 843-853.

3. Lukas, M., Toth, I., Reber, S.O., Slattery, D.A., Veenema, A.H., Neumann, I.D. (2011). The Neuropeptide Oxytocin Facilitates Pro-Social Behavior and Prevents Social Avoidance in Rats and Mice. *Neuropsychopharmacology* **36**, 2159-2168.

4. Toth, I., Neumann, I.D., Slattery, D.A. (2012). Social Fear Conditioning: A Novel and Specific Animal Model to Study Social Anxiety Disorder. *Neuropsychopharmacology* doi: 10.1038/npp.2011.329.

Individual Differences in F0 Imitation: Causes and Effects

M. Postma, E. Postma

TiCC, Tilburg University, Tilburg, The Netherlands
m.nilsenova@tilburguniversity.edu; e.o.postma@tilburguniversity.edu

Introduction

Phonetic imitation plays an important role in human interaction in that it reflects the closeness of the social bond between two individuals. Past studies have indicated the importance of the region between 50 Hz to 300 Hz (the fundamental frequency (F0) region) which is the most important source of information regarding emotions, stands and attitudes in the voice [1,2]. The same region also provides acoustic information for imitation exploited in promoting social convergence and status accommodation [3-8] and expressing ingroup-outgroup bias [9]. Interestingly, there appear to be large individual differences in speakers' ability to imitate F0. Using the shadowing task paradigm, originally introduced by [10], a recent study [11] found a considerable amount of variation in F0 accommodation, with some subjects actually diverging from the F0 of the model talker.

We hypothesized that the individual differences in speakers' ability to imitate F0 may at least partly be due to their neurocognitive ability to extract information about pitch from the speech signal. In particular, due to neuroanatomical differences found in the lateral Heschl's gyrus (the 'pitch processing center'), some listeners show an auditory perception bias for the sound as a whole (fundamental listeners), while others (spectral listeners) focus on its harmonic constituents [12,13]. The auditory perception bias has been almost exclusively analyzed in the context of musical training, but the results of individual studies indicate that it may also affect linguistic performance [14,15]. This study is the first attempt to explore the role of perception bias in imitation and thus its possible impact on social convergence.

Experimental study

Participants. Eighty-eight Dutch native speakers (67 female) between the age of 17-25y (M=20.48, SD=2.12) participated in the experiment for course credit. None of them reported any hearing difficulties. Fourteen of the participants were left-handed; about a half of the experimental group described their musical proficiency as low to average, the other half assessed their proficiency as high to professional.

Measuring auditory perception bias. Participants' auditory perception bias was determined with a variation of the psychoacoustic perceptual test described in [16], [17] and [18]. For the perceptual test, we constructed 36 pairs of complex harmonic tones, all 160 ms long, that consisted of 2-4 harmonics, with the same harmonic composition as employed by [17]. Participants were asked to categorize 18 perceptually ambiguous stimuli consisting of two complex tones, A and B, that were composed of a number of upper harmonic tones with the same highest harmonic but different levels of virtual fundamental pitch (derived from the harmonics as the best fit) and spectral pitch (based on the lowest harmonic). For example, the sequence A = (800 Hz + 1000 Hz) and B = (667 Hz + 1000 Hz) would be perceived as rising by a fundamental listener who would derive the missing F0 to be 200 Hz for tone A and 333 Hz for tone B; it would, however, be perceived as falling by a spectral listener who would focus on the lowest harmonics, i.e., 800 Hz and 667 Hz. The other 18 stimuli served as control trials in that their interpretation is unambiguous but helps to determine a participant's level of attention to the task. Listeners were instructed to categorize each experimental stimulus (tone pair) as either 'rising' or 'falling', depending on their perception of the sequence. Based on their answers, we calculated their individual 'Coefficient of Sound Perception Preference' (D_p) using the formula $D_p = (F\text{-}Sp)/(F+Sp)$, where F is the number of virtual fundamental classifications and Sp the number of spectral classifications. We calculated the 'Listener Attention Coefficient' (D_A) as the proportion of correctly categorized unambiguous stimuli. In order to test the validity of the perceptual test, we repeated the measurement approximately one month later under the same conditions with a subset of the participant set (N=64). The Shapiro-Wilks test of normality revealed that the coefficient D_p was not normally distributed in our experimental group: the majority of our participants performed as fundamental listeners (Mean_D_p=.397, SD=.406). A comparison of the first and the second measurement showed that even without feedback, repeated exposure to the ambiguous stimuli results in a shift towards fundamental auditory bias (Spearman's rho = .69, $p < .0001$). In order to explore the possibility that the difference between the first and the second measurement of participants' sound perception preference was due to the level of attention devoted to the task, we compared the absolute difference between the first and the second D_p to the attention coefficient D_A. The correlation between the attention coefficient D_A in the first measurement and the $| D_{p1} - D_{p2} |$ was significant (Spearman's rho = -.35, $p < .01$), indicating that poor attention to the task in general may have been the reason for the observed shift in D_p. In the subsequent analysis relating speakers' perceptual bias to their ability to imitate F0 in a shadowing task, we used the value of D_p collected during the first measurement (i.e., in the same session as the shadowing task). We excluded the data from participants whose attention coefficient D_A was lower than .9, i.e. those who classified two or more of the unambiguous stimuli incorrectly; this results in a sample of N=41.

Shadowing task. The material used during the shadowing task consisted of 16 sentences (8 declaratives and 8 interrogatives) that were presented four times in different orders to each participant. During the first and the fourth presentation, the sentences were presented one-by-one on a computer screen and the participant was instructed to read them in a neutral manner. During the second and the third presentation, the material was presented in an auditory form through a Sennheiser HMD26-600 headset. The participant was asked to repeat the sentences as precisely as possible. For model talkers, we used four different Dutch speakers (two male and two female) who were selected from a set of ten candidates on the grounds of speech clarity and lack of regional accent. The model talkers pre-recorded the 16 sentences in a soundproof booth with a Sennheiser HMD26-600 mic headset. Per participant, we used the recordings of a single model talker in order to increase exposure to the model talker's pitch. The participants were randomly divided between two experimental conditions; half of the participants heard full speech recordings while the other half heard recordings that were filtered with a 300-Hz high-pass Butterworth filter implemented in the Signal Processing Toolbox in Matlab. We calculated the 'Degree of F0 Imitation' by subtracting the absolute difference between the second and third F0 measurement (where the participant was shadowing) from the absolute difference between the model talker's F0 and the participant's F0 in the first measurement (baseline). Thus, a positive value of the 'Degree of F0 Imitation' indicates that the participant adapted to the model talker's F0, a negative value means the participant diverged and 0 represents no measurable change in mean F0. Given the considerable size of the speech corpus that was collected (5632 utterances in total), for the purpose of the analysis reported here, we focused solely on the initial voiced segment in one of the interrogative sentences ('Kan ik bij een vertraging mijn geld terug krijgen?' - "Can I get my money back in case of delays?"). The segment was extracted using a semi-automatic method and the pitch was determined with the autocorrelation method in Matlab.

The F0 measurements were analyzed with multiple regression with the between-subject condition F0 Filter (full speech signal vs. signal with frequencies above 300 Hz) and the participant's Coefficient D_p as predictors and the Degree of F0 Imitation as the dependent variable. As shown in Table 1 and 2, respectively, there was a main and interaction effect of the two predictors on the Degree of F0 Imitation, both for the second measurement (the first time the participants were shadowing the model talker's speech), $F(2,38)=3.443$, $p<.05$, as well as for the

Table 1. Summary of Multiple Regression Analysis for the Degree of F0 Imitation in the second measurement (N=41).

Variable	B	SE(B)	Beta	t	Sig.(p)
F0 Filter	50.71	16.20	.591	3.130	.003
Coefficient D_p	51.36	18.31	.443	2.805	.008
F0 Filter * Coefficient ∂_p	-81.08	36.62	-.400	-2.214	.033

Note. r^2=0.252

Table 2. Summary of Multiple Regression Analysis for the Degree of F0 Imitation in the third measurement (N=41).

Variable	B	SE(B)	Beta	t	Sig.(p)
F0 Filter	55.85	19.55	.558	2.857	.007
Coefficient D_p	43.96	22.10	.325	1.989	.054
F0 Filter * Coefficient ∂_p	-61.10	44.19	-.259	-1.383	.175

Note. r^2=0.199

third measurement (the second shadowing), $F(3,37)=3.065$, $p < .05$. This result shows that more fundamental listeners are better in F0 imitation than less fundamental (spectral) listeners, especially in conditions where the F0 information is missing and needs to be derived from the speech signal.

Concluding Discussion

In the study reported here, we used the standard psychoacoustic perceptual test with missing fundamental frequencies to determine a speaker's listener bias. In general, a listener's ability to perceive the missing F0 plays an important role in sound perception in that it helps to track prosodic contours in speech even when they are masked by noise or not transmitted [19], as in phone speech where the region up to 300 Hz is missing. The outcome of the task is somewhat dependent on the composition of the stimuli with the missing fundamental (the number of harmonics favors either the fundamental - for 4 harmonics, or the spectral - for 2 harmonics - interpretation; see [20]). Given that of the 18 ambiguous trials, 12 consisted of 3 and 4 harmonics, the task as such may have favored fundamental perception bias, which might explain the skewed distribution of the coefficient ∂_p in our participant group. A future study may therefore involve a more balanced composition of the stimuli, possibly also varying their length [21]. A comparison of the coefficient ∂_p measured at two different moments within a one-month period shows that the coefficient is not stable. In particular, we observed a slight shift towards a more pronounced fundamental perception, which correlated with the participant's attention to the task. Interestingly, a similar training-independent increase in the salience of the virtual fundamental pitch has been earlier reported by [22], who ascribed it to learning-induced long-term plasticity reflecting the biological relevance of pitch sensation [23]. We subsequently collected speech data in the classical shadowing task with two conditions, one with a full speech signal and one with high-pass filtered speech above 300 Hz. In both conditions, speakers with fundamental listener bias adapted more to the F0 of the model talker; as might be expected, the effect was more pronounced in the high-pass filtered condition. This result suggests advantages for fundamental listeners in communicative situations where F0 imitation is used as a behavioral cue. Future research needs to determine to what extent auditory perception bias may be influenced by training and whether it affects other social processes that rely on parsing of the prosodic information.

References

1. Juslin, P.N., Laukka, P. (2003). Communication of emotions in vocal expression and musical performance: different channels, same code? *Psychological Bul.*, **129**, 770-814.
2. Ververidis, D., Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech Com.* **48**, 1162-1181.
3. Gregory, S.W. Jr. (1983). A quantitative analysis of temporal symmetry in microsocial relations. *American Sociological Review* **48**, 129-135.
4. Gregory, S.W. Jr., Hoyt, B.R. (1982). Conversation partner mutual adaptation as demonstrated by Fourier series analysis. *J. Psycholing. Res.* **11**, 35-46.

5. Gregory, S.W. Jr., Webster, S.W., Huang, G. (1993). Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language and Communication* **13**, 195-217.
6. Gregory, S.W. Jr., Webster, S.W. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *JPSP* **70**, 1231-1240.
7. Gregory, S.W. Jr., Dagan, K., Webster, S.W. (1997). Evaluating the relation of vocal accommodation in conversation partners' fundamental frequencies to perceptions of communication quality. *J. Nonverb. Beh*. **21**, 23-43.
8. Gregory, S.W. Jr., Gallagher, T.J. (2002). Spectral analysis of candidates' nonverbal vocal communication: Predicting U.S. presidential election outcomes. *Social Psychological Quarterly* **65**, 298-308.
9. Babel, M. (2009). *Phonetic and social selectivity in speech accommodation.* Ph.D. diss., University of California, Berkeley.
10. Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review* **105**, 251-279.
11. Babel, M., Bulatov, D. (2011). The role of fundamental frequency in phonetic accommodation. *Language and Speech*, 1-17.
12. Rousseau, L., Peretz, I., Liégeois-Chauvel, C., Demany, L., Semal, C., Larue, S. (1996). Spectral and virtual pitch perception of complex tones: An opposite hemispheric lateralization? *Brain and Cognition* **30**, 303-308.
13. Schneider, P., Wengenroth, M. (2009). The neural basis of individual holistic and spectral sound perception. *Contemporary Music Review* **28**, 315-328.
14. Wong, P.C.M., Perrachione, T.K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics* **28**, 565-585.
15. Wong, P.C.M., Warrier, C.M., Penhune, V.B., Roy, A.K., Sadehh, A., Parrish, T.B., Zatorre, R.J. (2008). Volume of left Heschl's gyrus and linguistic pitch learning. *Cerebral Cortex* **18**, 828-836.
16. Smoorenburg, G.F. (1970). Pitch perception of two-frequency stimuli. *J. Acoust. Soc. Am.* **48**, 924-942.
17. Laguitton, V., Demany, L., Semal, C., Liégeois-Chauvel, C. (1998). Pitch perception: A difference between right- and left-handed listeners. *Neuropsychologia* **36**, 201-207.
18. Schneider, P., Sluming, V., Roberts, N., Bleeck, S., Rupp, A. (2005). Structural, functional and perceptual differences in Heschl's gyrus and musical instrument preference. *Ann. N.Y. Acad. Sci*. **1060**, 387-394.
19. Seither-Preisler, A., Johnson, L., Krumbholz, K., Nobbe, A., Patterson, R.D., Seither, S., Lütkenhöner, B. (2007). Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and non-musicians. *J. Exp. Psychol.: Hum. Percept. Perform*. **33**, 743-751.
20. Houtsma, A.J.M. (1979). Musical pitch of two-tone complexes and predictions by modern pitch theories. J. *Acoust. Soc. Am*. **66**, 87-99.
21. Beerends, J.G. (1989). *Pitches of simultaneous complex tones*. Unpublished Ph.D. diss, University of Eindhoven.
22. Seither-Preisler, A., Johnson, L., Preisler, E., Seither, S., Lütkenhöner, B. (2009). The perception of dual-aspect tone sequences changes with stimulus exposure. *Brain Research Journal* **2**, 125-148.
23. Schneider, P., Sluming, V. Roberts, N., Scherg, M., Goebel, R., Specht, H.J., Dosch, H.G., Bleeck, S., Stippich, C., Rupp, A. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nature Neuroscience* **8**, 1241-1247.

Measuring Behavioural Changes to Assess Anthropogenic Noise Impact on Singing Birds

H. Slabbekoorn

Institute of Biology, Leiden University, Leiden, The Netherlands. H.W.Slabbekoorn@Biology.LeidenUniv.nl

Sound aspects of avian singing

Many animals use sounds for communication. Calls and songs often serve functions that have direct fitness consequences, for example in warning for predators, defending critical resources, and attracting mates. The tight link with fitness makes such sounds suitable for studying the impact of natural selection on vocal variation and the consequences for evolutionary change in acoustic signals. Birds are well-known and well-studied with respect to acoustic behaviour. Most species have a repertoire of calls but the majority of studies have addressed the typically most prominent vocal behaviour of singing. The two most important functions of singing in birds concern defending a territory for food, nesting and hiding places and attracting a social partner for mating. Visual cues often also play an important role in avian communication, such as during display of colours and plumage patterns or during courtship rituals and song flights. However, sounds, in concert with visual cues or by themselves, are a prime communication channel for the majority of all birds.

The importance of acoustics in animal communication in general and for birds in particular is closely related to features that are more or less specific to sound. For example, sound can travel over relatively large distances, does not rely on the presence of light, and is relatively little affected by obstacles such as leaves and branches. Furthermore, sound is multidirectional both from the perspective of the sender as from the receiver of a signal. Acoustic signals radiate away in all directions with only relative amplitude being affected by the facing direction of the singing bird. Similarly, receivers hear acoustic signals coming in from all directions and do not need to face the source direction to detect its presence. Although sounds have many benefits for communicating animals, there are also several environmental factors causing problems. Attenuation and degradation of signals during transmission though the environment and interference of other sounds can lead to selection on the acoustic design of signals. Especially the latter impact of ambient noise on production and perception of acoustic signals has received a lot of recent attention due to the global rise of anthropogenic noise.

Noise impact factors

Natural habitats are full of natural sounds from abiotic sources, such as wind and water flow, and biotic sources, such as calling frogs and singing birds. However, human activities elevate these ambient noise levels artificially. The so-called anthropogenic noise from traffic, industry, and recreational activities concerns a global pollution factor which is still expanding in time and space. The artificially high noise levels are typically of relatively low frequency and can cause many different problems such as physical damage in cases of extreme overexposure, physiological stress, and auditory masking, which are all factors that an animal is more or less passively undergoing (see Figure 1). Furthermore, anthropogenic noise may lead to spatial deterrence, behavioural interruption, and signal modifications, which are all factors that involve some sort of active response from the animal. The three passive factors are typically positively related: if one is getting worse the others are likely to follow the same pattern. An increase in any of the three passive factors also increases the probability that any of the three active factors occurs. Vice versa the opposite may be the case: the active factors have the potential for providing relief on the passive factors by reducing the level and duration of exposure.

Many negative effects of increased noise levels along highways, in cities, or at industrial sites, have been confirmed for birds [1]. It has for example been shown that anthropogenic noise causes passive effects such as increased heart beat in birds (physiological stress), reduced mating success (auditory masking), and reduced reproductive success (through physiological stress and/or auditory masking). There are also reports on birds that start singing at more quiet times of day and/or stop singing after bursts of noise or with rising noise levels

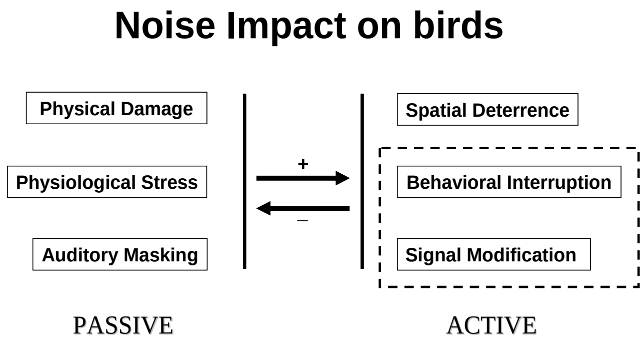


Figure 1. Schematic overview of six dominant factors that may play a role in assessing the impact of anthropogenic noise on bird welfare and fitness. The arrows indicate the relationship between factors. The 'passive' factors on the left reflect consequences of undergoing a particular exposure, which typically also increase the probability of occurrence of the 'active' factors on the right. The 'active' factors, in contrast, involve behavioural decisions by the individual that are likely to reduce the severity of the 'passive' factors. The purpose of this subdivision into categories is just to emphasize the complexity of noise impact assessments as, although the overall representation is generally true, the relative importance of factors and the magnitude and even direction of the relationships will vary per exposure level and duration, per species, and with physiological and environmental conditions. The dashed line block demarcates the focus of the reported experimental noise exposure studies.

(behavioural interruption). Especially common are reports on birds singing louder or higher in frequency, or with more concentrated sound energy bands in more noisy conditions (signal modification). Patterns of noise level dependent song frequency use have been found both between and within populations. Great tits (*Parus major*) in cities across Europe, for example, were found to sing with consistently higher minimum frequency on average compared to birds in nearby forest sites [2]. Individual great tits in and around Leiden were also found to have a repertoire of song types with higher minimum frequencies if they had a relatively noisy territory compared to nearby birds that were singing further away from traffic.

Mechanistic explanations

These patterns of noise-dependent frequency use in birdsong raised the question whether the association had emerged at an evolutionary, ontogenetic, or more immediate temporal scale [3]. Habitat-dependent selection could drive micro-evolutionary changes and explain song differences between populations in relatively noisy and more quiet habitat, but the urban habitat and the rise in traffic noise is relatively recent and is also less likely to be an explanation for variation between individuals in territories that differ in noise level but are only a few hundred meters apart. Reduced audibility at noisy sites of low-frequency relative to high-frequency songs could result in a noise-dependent bias when young birds are building up a song repertoire from memorizing what they have heard in their surroundings. This possible mechanism could account for patterns between and within populations, but would not allow for more rapid frequency modifications in response to current fluctuations in noise level. Such ability of relatively immediate flexibility was revealed in great tits for the first time using experimental exposure tests in the field.

Singing great tits were exposed to traffic-type 'city noise' to test for changes in frequency use before, during, and after the period of artificially increased ambient noise level [4]. It turned out that this species was not shifting upward in frequency by singing a specific song type at higher frequencies, but that it used a noise-dependent way of switching between song types. Individual birds that were exposed to noise continued to repeat song types for longer when these song types had a minimum frequency that was relatively high. If they happened to sing a song type with a minimum frequency that was relatively low, there was more masking and they switched more quickly. These tendencies provide an explanation for the correlations between song frequency and noise levels through a noise-dependent assortment: low song frequencies with quiet conditions and high song frequencies with noisy conditions. Interestingly, this means that the spectral adjustment to noise is actually achieved in great tits by temporal adjustment (singing particular song types for shorter or longer periods).

However, although immediate flexibility can explain all patterns of noise dependent frequency use between and within populations, a contribution of micro-evolutionary or ontogenetic changes can not be excluded yet.

Methodological issues of experimental exposure

Collecting song recordings

Noise-dependent frequency use may contribute to survival ability of birds in environments heavily exposed to anthropogenic noise [5]. It is therefore important to collect more song recordings and study more species that have a different singing style than the great tit to investigate whether they can also achieve spectral adjustment and in what way. Comparative data will eventually provide insight into the species requirements for behavioural flexibility under noisy conditions and has the potential to help in predicting and mitigating noise impact in the context of conservation. Planning and conducting new studies starts with checking previous investigations to sort out, for example, how the singing styles of potential target species compare to e.g. the song type repertoire style of singing of the great tit [4] or the syllable repertoire style of singing of the chiffchaff (*Phylloscopus collybita*) [6].

It is important to get enough song recordings with and without anthropogenic noise, or with low and high levels of ambient noise, that are not affected by any other acoustic disturbance. Songs or calls of other animal species overlapping the song recordings of the target individual have the potential to impede accurate measurements. Prominent activity of other individuals of the same species is also better avoided as in that case social stimulation may cause dramatic fluctuations in song characteristics that may overrule noise-dependent vocal changes. Ten clear song recordings or more in each noise level category should do for most types of analyses. It is good to realize in this context that bird species singing with so-called *eventual variety* will repeat one song or syllable type for a number of times before switching to another. Bird species singing with so-called *immediate variety* will sing another variant every other song they sing. And then there are also species that are single song type singers which have only variety in the spectral and temporal details of the one song variant they continuously sing. This type of variety among species has the potential to affect the sample distribution and can play a critical role in understanding the mechanism underlying noise-dependent song characteristics, as revealed for the great tit.

Experimental exposure conditions

The target of experimental exposure experiments in the field is to induce a rise in traffic-like ‘city noise’ at the position of a singing bird while all other factors remain the same as much as possible. Keeping recording periods for songs under different noise conditions, before, during, and possibly also after experimental exposure, as close as possible together in time will guarantee the optimal extent of condition similarity. Acoustic data collection on free-ranging animals always requires sensible action by the investigators. Approaching a singing bird off-trail, setting up equipment that is a-typical to be there, pointing at the target individual with a long and pointy dark microphone, and staring at it repeatedly by one or more investigators for the duration of the trial, will all make the probability higher that the bird will stop singing (see Figure 2). It may also fly away to another perch to resume singing at a distance from the speaker planned to use for the experimental noise exposure. However, careful operation and selection of optimal distances for the recording and playback equipment, which will depend on season, specific target species, and local habituation to people, will usually allow getting sufficient recordings.

It is important to consider the implications of stimulus choice for the noise exposure file to be played back to the target bird. Different approaches are possible to avoid the problem of pseudo-replication when using just a single recording of traffic noise as the overlapping noise stimulus file. One can apply a replicated set of traffic noise recordings [7] or generate a completely artificial sound file with an average traffic-type noise spectrum [4]. Note that alternative masking approaches have the potential to get answers to more fundamental research questions. One can apply for example a noise stimulus file which is below and non-overlapping in frequency with the song frequency range of the target species [8] or which is similar to ‘city noise’ but spectrally reversed; putting most energy at higher frequencies and masking the upper instead of the lower part of birdsong more [4].

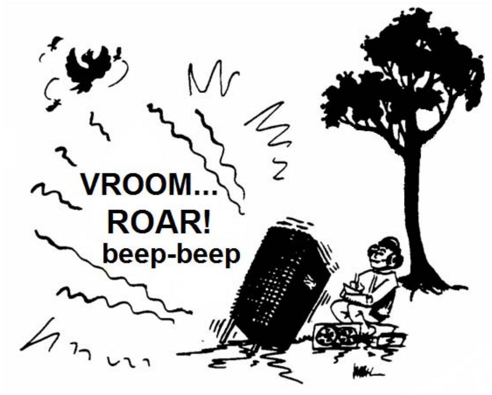


Figure 2. Cartoon by Mart Ottenheim showing an example of how experimental noise exposure can interrupt natural singing behaviour of a free-ranging bird.

It is also important to assess the background noise levels and the achieved rise in noise levels after the trial is over at the song perch used by the focal bird during exposure. Note that the noise level broadcasted through the speaker will decline with 6 dB per doubling of distance. This means, when you playback a noise file at 90 dB(A) as measured at one meter in front of the speaker, that the noise exposure level will range between 72 and 66 dB(A) at a distance between 8 and 16 meters in the direction at which you have pointed the speaker. Also note that the noise spectrum at the bird will be determined by the natural spectrum of the background noise together with the artificial spectrum, which is dependent on the noise stimulus file, the broadcasting properties of the speaker, and site-specific attenuation and reverberation. As the noise coming from the speaker will also affect the recording quality of the singing bird, it is best to use a highly directional microphone pointed at the bird and away from the speaker. An optimal recorder position is at a side of the speaker and pointed parallel to the broadcasting direction. Furthermore, getting as close as possible to the bird and as far as possible away from the speaker, positioning some natural objects in between microphone and speaker if possible, will guarantee the best recording quality for measurements.

Sonographic measurements

A variety of measures can be taken from sonographic representations of the recorded songs, which will depend partly on the target bird species and on the acoustic software used. Interesting temporal parameters include the duration of song bouts, individual songs, syllables, elements, or intervals between particular song units. Note that the start of a sound is typically a very accurate and reliable measure, while the assessment of the end of a sound is typically affected to some extent by reverberations or echoes. The associated measurement error increases in relative importance with decreasing duration of the song units and is only a problem when reverberative qualities of the environment co-vary with noise conditions. Interesting spectral parameters include: the minimum and maximum frequency, which are typically assessed by marking the lowest and highest visible trace on the sonogram attributed to the bird’s song; the so-called peak frequency, which is the frequency at any particular moment in the song with the highest amplitude, and the dominant frequency, which is the frequency with the highest level of accumulated energy over the duration of the target song unit. Note that this terminology is not always consistent across publications and across software packages. So, always check the details of acoustic definitions. Some software packages, such as Avisoft Bioacoustics (http://www.avisoft.com/), have the option to extract additional and potentially interesting measures on the energy distribution within the song frequency range by providing spectral demarcation values for energy quartiles below which 25%, 50%, and 75% of all energy is concentrated [8].

Measuring spectral changes in song recordings from quiet and noisy periods has the potential for noise-related artefacts. These artefacts can yield measurement biases away or in the direction of the expected changes in noise-dependent vocal production. This is especially true for minimum and maximum frequency and less true for peak and dominant frequency. The extent and direction of this potentially problematic measurement error depends on the software program and the extent to which measurements rely on cursor placements on the sonographic display. Note that more digital assessments are typically more objective but also less accurate than manual assessments. Our visual interpretation of sonographic traces of signals and noise are usually best to rely on. A double-blind procedure with the measurements being taken by somebody unaware of the presence or absence of treatment is difficult as the noise level is always audible and at least visible to some extent. The best way to go is to assess the potential measurement error by conducting a playback-rerecording session with known signal properties and controlled variation in ambient noise level at the position of the microphone [6]. Note that also the energy distribution values, as measured in for example Avisoft Bioacoustics, are potentially affected by noise conditions, as they accumulate energy from birdsong as well as ambient noise, and they are therefore only feasible with high signal-to-noise ratios. An advantage is, however, that in this case the potential measurement artefact due to experimental noise exposure is typically in the opposite direction of the expected change in song frequency use.

Conclusions

Vocal communication often plays a critical role in the life of birds while noisy human activities are deteriorating acoustic conditions for signalling in and around cities, along highways, and in the vicinity of airports. Spectral song adjustments seem a general ability among bird species that persist in such noisy areas, although the relief of the negative impact from masking noise is not straightforward due to compromises on signal value related to the adjustment [9]. Nevertheless, the use of higher frequencies reduces the extent of masking by anthropogenic noise, which usually has a bias to relatively low frequencies. Experimental exposure provides a tool to investigate the underlying mechanism and an opportunity to compare different species with different singing styles and different coping abilities. Methodological concerns involve aspects of both the data collection in the field as the data processing in the laboratory. However, recognition of potential problems and applying adequate solutions provide perfect opportunities for future studies by which we can gain fundamental understanding about the ecology and evolution of signal plasticity. At the same time, this information may become of applied value in a world that is becoming ever more noisy due to human activities.

References

1. Kight, C.R., Swaddle, J.P. (2011). How and why environmental noise impacts animals: an integrative, mechanistic review. *Ecology Letters* **14**, 1052-1061.
2. Slabbekoorn, H., den Boer-Visser, A. (2006). Cities change the songs of birds. *Current Biology* **16**, 2326-2331.
3. Patricelli, G.L., Blickley, J.L. (2006). Avian communication in urban noise: causes and consequences of vocal adjustment. *Auk* **123**, 639-649.
4. Halfwerk, W., Slabbekoorn, H. (2009). A behavioural mechanism explaining noise-dependent pitch shift in urban birdsong. *Animal behaviour* **78**, 1301-1307.
5. Slabbekoorn, H., Peet, M. (2003). Birds sing at a higher pitch in urban noise. *Nature* **424**, 267.
6. Verzijden, M.N., Ripmeester, E.A.P., Ohms, V.R., Snelderwaard, P., Slabbekoorn H. (2010). Immediate spectral flexibility in singing chiffchaffs during experimental exposure to highway noise. *Journal of Experimental Biology* **15**, 2575-2581.
7. Gross, K., Pasinelli, G., Kunc, H.P. (2010). Behavioral plasticity allows short-term adjustment to a novel environment. *American Naturalist* **176**, 456-464.
8. Hanna, D., Blouin-Demers, G., Wilson, D.R., Mennill, D.J. (2011). Anthropogenic noise affects song structure in red-winged blackbirds (*Agelaius phoeniceus*). *Journal of Experimental Biology* **214**, 3549-3556.
9. Halfwerk, W., Bot, S., Buikx, J., van der Velde, M., Komdeur, J., ten Cate, C., Slabbekoorn, H. (2011). Low-frequency songs lose their potency in noisy urban conditions. *Proceedings of the National Academy of Sciences USA* **108**, 14549-14554.

Measuring Social Behavior in Drug Discovery

Marcel van Gaalen[1], Thomas Appl[2], Anton Bespalov[3]

Neuroscience Research, GPRD, Abbott, Ludwigshafen, Germany
[1]marcel.vangaalen@abbott.com, [2]tom_appl@yahoo.de, [3]anton.bespalov@abbott.com

Abstract

Disturbed social interaction is seen in many psychiatric and neurological disorders. The majority of animal models used in the pharmaceutical industry, however, do not allow assessment of social behaviors. This may due to the beliefs that measuring social behavior is too labor intensive, lacks robustness, or is too complicated to measure. Here, we describe two models of social domination: Dominance Submissive Behavior and Social Isolation-Induced Aggression. These models can be useful to estimate efficacy for compounds that may be beneficial for treatment of depression, mania and disorders with symptoms of enhanced aggressive behavior such as Alzheimer's disease and schizophrenia. We demonstrate, by showing data of training sessions and effect imipramine on established submissive behavior, that Dominance Submissive Behavior can be studied in a simplified setting with automatic color recognition video tracking. In addition, we show that the level of aggressive behavior in mice can be easily assessed with simple "real time" latency measurement. However, the example of effect of buspirone indicates that detailed analysis of drug-induced effects can be partially automated, but require manual inspection of behavior of video analysis played back on reduced speed. Finally, training data from both models indicate that selection of animals that show reliable dominant/submissive or aggressive behavior is essential for drug testing.

Introduction

Disruption of social behaviors is symptoms that occur in many psychiatric and neurological disorders. Yet, the use of animal models allowing assessment of social behavior in drug discovery is limited. Here, we evaluate two models that may allow estimation of efficacy profile of compounds in preclinical drug discovery. Dominant submissive behavior models tax hierarchy formed through social interactions and can be based on naturally competition for various resources [1]. Dominant submissive behavior has been described for rodents in special designed arenas. Dominant behavior has been shown to be attenuated by antimanic drugs, whereas submissive behavior can be reduced by antidepressant drugs [2]. Here, we evaluate effects of the antidepressant drug imipramine in standard behavioral equipment (open field) with video tracking analysis approach.

Aggressive behavior has been described in disorders such as schizophrenia and Alzheimer's disease [3,4]. These symptoms are distressing for the patients, but also for caregivers and family members. For Alzheimer's disease, aggression is often the catalyst for institutional care [4]. Various animal models have been described. Here, we evaluate the development of aggression social isolated animals and compare various analysis methods to detect the effects of 5-HT_{1A} receptor partial agonist buspirone in aggressive mice.

Methods

Animals

Male Han Wistar rats (Janvier, France, weighing between 250-300g) housed in groups of four were used for Dominant Submissive Behavior. Male Swiss OF1 (residents, purchased form Charles River, Germany) and NMRI (intruders, purchased from Janvier, France) 6 weeks of age at arrival were used for social isolation-induced aggression. Residents were single housed for 4-6 weeks before the start of an experiment. Intruders were housed in groups of five. Animals were kept under standard laboratory conditions (21±1 °C air temperature and 55±15% humidity), with free access to food (except for Dominant Submissive Behavior) and water. Experimental procedures were approved by Abbott's Animal Welfare Officer, and were performed in accordance with the recommendations and policies of the U.S. National Institutes of Health "Principles of laboratory animal care" (1996 edition), and performed in an AALAAC certified institute.

Dominant Submissive Behavior

Training: Animals were tested in pairs (same composition throughout the experiment) derived from different home cages, once a day, five days a week. A test session lasted for 5 min and started by placing the pairs in an open field arena (60 x 60 cm) with a centrally inserted beaker glass (50 ml, 6 cm diameter in a hole in the floor plate) with condensed milk (Cat milk, Attika, < 0.4% lactose). Animals were color marked on their head and neck. During the session, a video recording was made of the central area of the arena. The time the color mark covered the beaker glass was analyzed by video tracking (EthoVision XT, Noldus Information Technology, Wageningen, The Netherlands) and taken to indicate time at the milk beaker. In order to make the analysis less time consuming, the daily recorded videos were merged to a single video. Analysis of the video in 5 min bins allowed analysis of single pairs. After each test, animals received laboratory chow ad libitum for only 1 hour, but had free access to chow during the weekends.

Twenty pairs (from 40) showed robust and stable dominant submissive behavior and were selected for drug treatment. Submissive animals received imipramine (10 mg/kg) or vehicle once per day and examined for consecutive 3 weeks.

Social Isolation-induced aggression.

Resident mice were tested during a training sessions. A training session started by placing an intruder in the home cage of the resident, and lasted for 5 min or shorted when the intruder was attacked and the intruder showed submissive behavior. Resident mice were selected for aggression after 3-4 training sessions (one session per day), and animals with short attack latency were selected for drug treatment studies. A drug testing session started by placing the resident in a test arena (open field 28 x 37 cm). During the first 5 min, locomotor activity was analyzed using EthoVision XT. A naïve intruder was introduced for the next 5 min of the session, which was recorded on video. Analysis of behavior (latency to attack, and number of bites) was performed manually with play back at half speed. Behavior as also manually analyzed and categorized in exploratory behavior, social interaction and inactivity using The Observer XT (Noldus Information Technology). Buspirone was tested at 0.03, 0.1 and 0.3 mg/kg.

Results

Stable and robust dominant submissive behavior was observed in approximately 50% of the tested pairs. Dominant animals continued to spend more time at beaker after vehicle treated animals, whereas submissive behavior who received imipramine did not differ from the animal that showed increased time at the beaker before treatment. We demonstrated that dominant submissive behavior can be measured in standard behavioral equipment (open field). Furthermore, merging video files reduced time spent for analysis.

Aggression levels, in terms of latency to attack varied strongly between residents. Training data over consecutive days indicated that animals that show initial low attack latencies, strongly decrease attack latencies over time, whereas animals that show long attack latencies do not show decreased latencies over time. Approximately 50% of the animals showed consistently very low attack latencies and were selected for drug testing. Video tracking revealed that locomotor activity was dose dependently reduced by buspirone during the first 5 min before the introduction of the intruder. Buspirone increased latencies to attack, although a large variation of the data was seen for this parameter. Number of bites was less variable, and reduced by the highest dose of buspirone. Buspirone tended to increase the time spent inactive, but had no effect on social interaction time or exploration time.

Discussion

Here, we demonstrated that dominant submissive behavior can be studied in standard behavioral equipment (open field). Video recording and analysis allows behavior testing in multiple arenas simultaneous. In order to enhance sensitivity of the recording, only the center of the open field, including the beaker glass was recorded. This enhances the ratio background to target (colored spot on the animals), and allows more flexibility in light settings. This is particularly of interest when several arenas are in use, placed next to each other. Video recording results in large data files. In order to allow analysis over night, we merged the videos and rendered and tracked the videos overnight. Video analysis with software that allows using a "play list" would be beneficial and further reduce analysis time.

The data indicate that dominant submissive behavior is established only approximately 50% of the pairs tested. A selection of animals is essential for drug testing. As an alternative, animals could be tested in larger groups (3-4 rats) from the same holding cage, to estimate hierarchy in the home cage. As such, submissive and dominant animals could be identified in each group (ongoing evaluation). Color identification of 3-4 animals is, however, technically more demanding.

Simple measurement of attack latency during training sessions is adequate to select highly aggressive animals for drug testing. However, analysis of the buspirone sessions indicated that latency to attack has high variability and the number of bites is a more robust read out. Analysis of the number bites requires the videos to be played at reduced speed. Video tracking data indicate that buspirone reduces locomotor activity, whereas manual analysis of social interaction, exploratory behavior and inactivity does not reveal an effect of buspirone. The analysis have been performed without intruder (video tracking) and with intruder (manual analysis), which may explain the difference. Alternatively, video tracking may be more sensitive. Nevertheless, number of bites was reduced by 60% by the highest dose buspirone, leaving social interaction unaffected.

Dominant Submissive Behavior and Social Isolation-Induced aggression can be reliably measured with standardized laboratory equipment (Open field). Read out are stable and robust, however, a selection of animals is necessary. Automatic analysis of the behavior increases speed of analysis, but can not be applied for all parameters.

References

1. Blanchard, D.C., Spencer, R.L., Weiss, S.M., Blanchard, R.J., McEwen, B., Sakai, R.R. (1995). Visible burrow system as a model of chronic social stress: behavioral and neuroendocrine correlates. *Psychoneuroendocrinology* **20**, 117-134.

2. Malatynska, E., Knapp, R.J. (2005). Dominant-submissive behavior as models of mania and depression. *Neurosci Biobehav Rev* **29**, 715-737.

3. Frogley, C., Taylor, D., Dickens, G., Picchioni, M. (2012). A systematic review of the evidence of clozapine's anti-aggressive effects. *Int. J. Neuropsychopharmacol.* 1-21.

4. Ballard, C., Howard, R. (2006). Neuroleptic drugs in dementia: benefits and harm. *Nat. Rev. Neurosci.* **7**, 492-500.

SPECIAL SESSION

Electrophysiological Correlates of Behaviour

Chair: Bettina Platt[1], Karsten Wicke[2]

[1]*School of Medical Sciences, College of Life Science and Medicine, University of Aberdeen, Scotland, UK*

[2]*Neuroscience Research, GPRD, Abbott GmbH & Co KG, Ludwigshafen, Germany*

Abstract

This symposium explores novel ways to monitor and analyse physiological brain activity in awake rodents and humans, for basic research purposes that addresses key aspects in cognitive and sleep research, as well as medically oriented applications with respect to the pharmacology and translational value of electrophysiological activity, and the use of biosignals for improved human-machine interfaces and a web-based platform for diagnosis of CNS conditions in humans.

Specifically, the first presentation by Nadine Becker and Matt Jones (University of Bristol) explores how spatial information acquired in a T-maze task is encoded in firing patterns of cortical and hippocampal areas, monitored in rats via chronically implanted electrode arrays, and discusses how area- and task-component-specific firing patterns can be monitored and deciphered, and what this implies for our understanding of cognitive processes in specific brain areas.

The second presentation by Valter Tucci (Genoa, Italy) investigates how genetic variables determine biological clocks and how information is time-stamped in the mouse brain. He will discuss the interplay between different time-keeping mechanisms on the basis of mouse models with mutations of clock genes, and how these affect the ability to time intervals at different timescales, related to sleep, circadian rhythms and an experimental timing task. Behavioural analysis was combined with long-term electrophysiological measures in the home-cage and uncovered close relations between timing behaviors and sleep physiology.

The third presentation by Lianne Robinson and Gernot Riedel (Aberdeen, UK), embarks on pharmacological aspects of electrophysiological and behavioural profiles recorded in chronically implanted rodents. The talk will explore how sleep-wake cycles, and vigilance stage-specific EEG parameters, are affected by cannabinoids in mice and rats, and how this can be exploited for cognitive and therapeutic research.

Continuing with this line of research, Karsten Wicke (Abbott, Germany) will present results from rat sleep EEG recordings that represent a back-translation from clinical human data to preclinical experiments and are used as a predictive translational model for the clinical potential of new pharmacological mechanisms. The antidepressants potential of drugs affecting the glutamatergic system will be discussed.

The remaining two presentations venture further into human and clinically oriented applications. Felix Putze and Tanja Schultz (Karlsruhe, Germany) will present developments in the area of 'cognitive technical systems', which utilises a range of biosignals emitted from the human body such as EEG and muscle activity. These signals are measured and interpreted by machines and offer an inside perspective on human mental activities, intentions, or needs and thus complements the traditional way of observing humans from the outside. Current and future applications for this technology will be discussed.

Finally, Björn Crüts (BrainMarker, The Netherlands) will introduce an internet-based platform for qEEG-based diagnosis of neurological and neurodegenerative disorders. He will focus particularly on examples regarding EEG markers for depression and respective treatments, to explain how their database and algorithms aid diagnosis and therapeutic applications in hospitals.

Together, these presentations span a wide range of technical and methodical approaches and their applications, and thus provide an exciting overview of a currently booming are of research and development.

SPECIAL SESSION CONTENTS (sorted by paper ID)

Sleep, Circadian Rhythms and Interval Timing

Valter Tucci (Istituto Italiano di Tecnologia, Genova)

Modulation of Sleep-Wake Cycles in Mice and Rats with Cannabinoids

Lianne Robinson, Andrea Plano, Anushka Goonawardena, Bettina Platt and Gernot Riedel (University of Aberdeen)

Cognitive Technical Systems

Felix Putze and Tanja Schultz (Karlsruhe Institute of Technology)

Neural Correlates of a Spatial Learning Task in Parietal Cortex, Prefrontal Cortex and Hippocampus

Nadine Becker and Matt Jones (University of Bristol)

NMDA Receptor Antagonists Induce Antidepressant-Like Sleep Changes: A Translational Model from Rats to Humans?

Karsten Wicke and G. Gross (Abbott, Germany)

Quantitative EEG for the Diagnosis of Disease States

Björn Crüts (BrainMarker, The Netherlands) and Pascal Römkens (Atrium Medical Centre, The Netherlands)

Sleep, Circadian Rhythms and Interval Timing

Valter Tucci

Department of Neuroscience and Brain Technologies, Istituto Italiano di Tecnologia, Genova, Italy. valter.tucci@iit.it

Timing is a fundamental property of many biological systems, however, a systems biology approach to timing is still missing. Indeed, timing can be investigated at genetic/epigenetic, cellular, neuronal and behavioral level [1].

The temporal interplay between genetic and epigenetic mechanisms determines behavioral, physiological and molecular phenotypes across the organism. The understanding of how genetic sequences translate in complex phenotypes is a major challenge in current functional genomics. This difficulty is due to the fact that modules of genes co-express [2] and a particular phenotype results from the precise temporal dynamics (i.e., daily oscillations) of such modules.

Single-cell organisms adapted to the environment by entraining to external stimuli (caused by the earth's rotation) in order to set the time of their metabolic processes and sleep-wake rhythms evolved from this rest-activity cycle. The success in discovering molecular loops that set the circadian clock has favored the diffusion of clock-like models in other fields of investigation. For instance, the presence of a pacemaker-like mechanism has been hypothesized in timed biological processes across many time-scales (microseconds, seconds, hours). Thus, lots of effort in research has been dedicated to develop comprehensive clock-like computational models across different timescales.

Timing is also fundamental in cognitive processes although the mechanism by which the brain codes for timing is still unknown [3-5]. In mice, as in many other species, conditioning behaviors are subjected to temporal determinants and a brief (seconds to minutes) time interval (namely interval timing) itself may embody the proper information to be learned.

However, as the mechanisms to time information across different timescales vary, we investigate timing at different timescales and at systems level [6]. An integrated investigation of timing mechanisms represents a powerful instrument to understand strategies that evolution has developed to timestamp information in daily life [1]. Here we will present evidence that support the use of mice in understanding the biology of timing properties in cognitive, circadian and sleep homeostatic processes. In particular, by studying mouse models in which specific genetic and epigenetic variables were manipulated, we identified a series of abnormal timing and sleep phenotypes.

We investigated cognitive timing in circadian and genomic imprinting mutant mouse lines. We have paralleled the behavioral analysis with long-term electrophysiological measures in home-cage. An analysis of the sleep-wake patterns of mice reveals a close relation between timing behaviors, circadian and sleep physiology.

We will show evidence that interval timing and temporal uncertainty [7] are compromised in a circadian mouse model. Circadian mutants showed asynchrony daily activity compared to littermate controls. Moreover, the performance during the dark phase of the light/dark cycle was significantly worse in the mutants compared to controls. Interestingly, we found that during the dark period (12 hours, from 7 p.m. to 7 a.m.) the control mice made significantly less mistakes respect to the mutants in every experimental phase. The control mice had a better performance compared to the mutants and they also showed a circadian trend in the error rate that is disrupted in the mice carrying the mutation in both heterozygosity and homozygosity.

A reduction of the spread of the timing peak of the behavioral responses in control mice during probes (not rewarded trials) occurred when the probability of probes was high, which testify a reduction of temporal uncertainty. On the contrary, the spread of responses was high and not significantly different in both heterozygous and homozygous mutant mice at each change of the probability of probes. This suggests that under different risk assessments the temporal uncertainty of mutants does not decrease. The reduction of the uncertainty, due to different risk assessments, produced also a change in the strategy. When the control mice become confident in a low probability condition they decide to anticipate the switch to the long location, while when the probability to receive the reward in the long location is lower, they persist more in the short location, postponing the switch.

Moreover, we will present evidence for a role of an imprinted gene, *Gnas* [8], in sleep and sleep-related timing functions. Sleep homeostasis refers to a process in which the propensity to sleep increases and decreases as wakefulness and sleep, respectively, progress [1]. Importantly, sleep modulates cognitive performance. Recent advances in functional studies of sleep states strongly suggest that non-REM sleep facilitates memory consolidation via modulation of synaptic plasticity through highly synchronized slow cortical activity, while REM sleep sustains consolidation by means of hippocampus-generated rhythms [9]. However, a direct measurement of synaptic plasticity and sleep and EEG activity is lacking. We have used a unique *in vivo* setup that allows the simultaneous and synchronized monitoring of EEG and behavior (by a home-cage 24-h monitoring system) to study sleep-cognition interactions. REM and NREM sleep are two distinct stages of the sleeping brain which are involved in the modulation of metabolic, physiological and cognitive processes. Despite these advances in functional understanding of REM and NREM sleep states, genetic and epigenetic mechanisms of sleep-dependent plasticity and memory processing are not currently well understood.

Several clinical and experimental evidence pointed out so far that imprinting genes are important in the regulation of sleep. However, the fundamental question whether genomic imprinting exerts a direct role in sleep, has remained elusive up to now. To address this crucial question we have started, a few years ago, to study sleep physiology and cognition in mouse models with an altered parent-of-origin expression profile. In this work we show that REM and NREM brain states are differentially modulated by the maternally expressed imprinted gene, *Gnas*. In particular, NREM-dependent physiologic and cognitive functions are enhanced while REM and REM-linked functions are inhibited. This is the first demonstration that we are aware of a specific effect of genomic imprinting on sleep states and their associated effects on cognition.

Interestingly, we observe a trade-off across different behavioral measures. For example, an increased precision in timing estimation is negatively correlated with the performance in behavioral fear responses.

References

1. Tucci, V. (2011). Sleep, circadian rhythms, and interval timing: evolutionary strategies to time information. *Front Integr Neurosci* **5**, 92.

2. Litvin, O. et al. (2009). Modularity and interactions in the genetics of gene expression. *Proc Natl Acad Sci USA* **106**(16), 6441-6446.

3. Buonomano, D.V. (2003). Timing of neural responses in cortical organotypic slices. *Proc Natl Acad Sci USA* **100**(8), 4897-4902.

4. Buonomano, D.V. (2007). The biology of time across different scales. *Nat Chem Biol* **3**(10), 594-597.

5. Buonomano, D.V. and U.R. Karmarkar. (2002). How do we tell time? *Neuroscientist* **8**(1), 42-51.

6. Buonomano, D.V. and R. Laje (2010). Population clocks: motor timing with neural dynamics. *Trends Cogn Sci* **14**(12), 520-527.

7. Balci, F., D. Freestone, and C.R. Gallistel. (2009). Risk assessment in man and mouse. *Proc Natl Acad Sci USA* **106**(7), 2459-2463.

8. Williamson, C.M. et al. (2004). A cis-acting control region is required exclusively for the tissue-specific imprinting of Gnas. *Nat Genet* **36**(8), 894-899.

9. Diekelmann, S. and J. Born. (2010). The memory function of sleep. *Nat Rev Neurosci* **11**(2), 114-126.

Modulation of Sleep-Wake Cycles in Mice and Rats with Cannabinoids

L. Robinson, A. Plano, A. Goonawardena, B. Platt, G. Riedel

School of Medical Sciences, College of Life Science and Medicine, University of Aberdeen, Aberdeen, Scotland, UK.
l.robinson@abdn.ac.uk

Introduction

The main psychoactive ingredient of marijuana Δ-[9] tetrahydrocannabinol (Δ-[9] THC) and endocannabinoids are involved in the modulation of the sleep-wake cycle. Systemic administration of the CB_1 antagonist/inverse agonist SR141716A dose-dependently induced an increase in wakefulness and a reduction in slow wave sleep (SWS) and rapid eye movement (REM) sleep [1]. In keeping with this observation, intracerebroventricular [2] or direct administration of anandamide into the peduncle pontine tegmental nucleus and medial preoptic area [3], two regions involved in the regulation of REM and non-rapid eye movement (NREM) sleep, increased NREM and REM sleep and reduced wakefulness. This effect was reversed with SR141716A [3, 4] indicating that the effects on sleep are mediated by activation of CB_1 receptors. However, the exact mechanisms by which cannabinoids modulate sleep-wake cycle are not fully understood.

Aim

The aim of the present study was to further our understanding of the endocannabinoid system in the modulation of sleep-wake cycle in freely moving rats and mice using a novel microchip-based data logging system.

Methods

C57BL6 mice (30-40g) (Harlan, UK) and Lister Hooded rats (250-300g) (Harlan, UK) were used in this study with all experiments performed under UK Home Office regulations and in accordance with the Federation of European Laboratory Animal Science Associations (FELASA) guidelines.

During surgery animals were anaesthetised with isoflurane and placed in a stereotaxic frame. Gold screw electrodes were positioned above the medial pre-frontal cortex and parietal cortex overlaying the left and right hippocampus. Further electrodes were located in the parietal and occipital areas as reference and ground. The pins were then assembled and secured to the skull surface using dental cement and tissue glue. Post-surgery, animals were allowed 7 days recovery prior to the start of recordings.

Mice received an acute intraperitoneal (ip.) injection of either the CB1 agonist WIN55,212-2 (WIN-2), CB1 antagonist AM251, the neutral CB1 antagonist ABD459 or vehicle (Triethylene glycol and phosphate buffered saline). Whereas, rats were injected with either i) WIN-2; ii) AM251; iii) WIN-2 + AM251 or iiii) vehicle. EEG recordings were performed immediately following drug administration with a recording duration of 6 hours. For recording purposes a wireless data Neurologger (New Behavior, Zurich, Switzerland) was used to monitor EEG activity. The Neurologger was attached to the head stage of the animal and allowed recording from up to 4 channels at a sampling rate of 200Hz (high pass filter: 0.1Hz, low pass filter:70Hz). A built-in accelerometer registered the animals movements throughout the experiment. Recordings were downloaded offline to a PC for analysis. Data were transformed using Matlab before being imported to SleepSign for the analysis of vigilant staging (NREM, REM, wake) and EEG spectral power [5].

Results

In mice, a significant reduction in REM sleep was evident following all cannabinoid treatments, combined with shorter REM episodes relative to vehicle controls. The cannabinoids however induced differential effects on wakefulness and NREM sleep; administration of WIN-2 resulted in a marked decrease in wakefulness while

AM251 induced an increase (see Figure 1). The CB_1 agonist also significantly increased NREM sleep whereas no change was observed with either of the antagonists.

In rats, WIN-2 also increased NREM sleep and the length of NREM episodes. Administration of AM251 failed to reverse the WIN-2 effect. However, AM251 did block the overall reduction in EEG spectral power observed during NREM and wakefulness after WIN-2 exposure. Treatment with AM251 alone had no effect on sleep stages but did increase the overall spectral power during NREM sleep.

Conclusions

All cannabinoids disrupted normal sleep patterns to some degree. The findings with WIN-2 and AM251 in this study corroborate previous studies reporting an increase in NREM and a reduction in wakefulness following exposure to anandamide or Δ^9-THC [2,3,6] and increased wakefulness after treatment with SR141716A [1]. Overall the results are consistent with the endocannabinoid system being involved in modulating sleep and support the possibility of using cannabinoid agonists, neutral antagonists/inverse agonists for the treatment of sleep disorders.

References

1. Santucci, V, Storme, JJ, Soubrie, P, Le, FG. (1996). Arousal-enhancing properties of the CB1 cannabinoid receptor antagonist SR 141716A in rats as assessed by electroencephalographic spectral and sleep-waking cycle analysis. *Life Sci*, **58**, L103-L110.

2. Murillo-Rodriguez, E, Sanchez-Alavez, M, Navarro, L, Martinez-Gonzalez, D, Drucker-Colin, R, Prospero-Garcia, O. (1998). Anandamide modulates sleep and memory in rats. *Brain Res*, **812**, 270-274.

3. Murillo-Rodriguez, E, Cabeza, R, Mendez-Diaz, M, Navarro, L, Prospero-Garcia, O. (2001). Anandamide-induced sleep is blocked by SR141716A, a CB1 receptor antagonist and by U73122, a phospholipase C inhibitor. *Neuroreport*, **12**, 2131-2136.

4. Murillo-Rodriguez, E, Blanco-Centurion, C, Sanchez, C, Piomelli, D, Shiromani, PJ. (2003). Anandamide enhances extracellular levels of adenosine and induces sleep: an in vivo microdialysis study. *Sleep*, **26**, 943-947.

5. Jyoti, A, Plano, A, Riedel, G, Platt, B. (2010). EEG, Activity, and Sleep Architecture in a Transgenic AbetaPPSWE/PSEN1A246E Alzheimer's Disease Mouse. *J Alzheimers Dis*ease. **22**(3): 873-887.

6. Buonamici, M, Young, GA, Khazan, N. (1982). Effects of acute delta 9-THC administration on EEG and EEG power spectra in the rat. *Neuropharmacology*, **21**, 825-829.

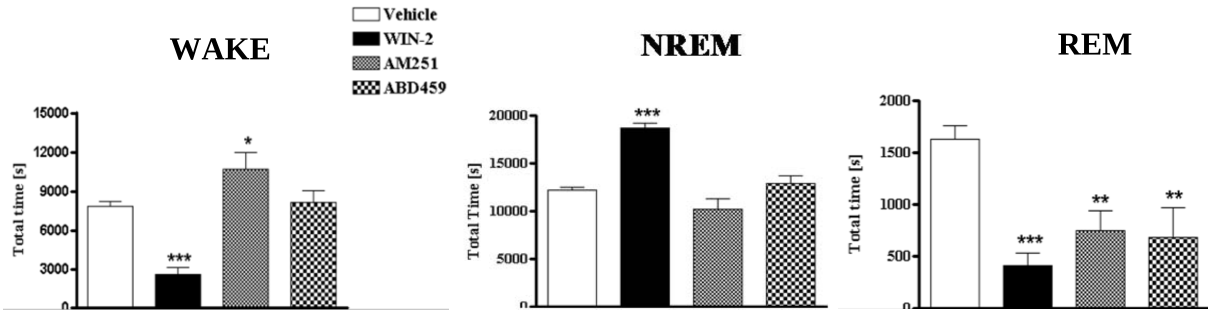


Figure 1. Effect of cannabinoids on time spent in WAKE, NREM and REM stages in mice. Mean + SEM. Asterisks denote reliable differences to Vehicle group (*= p<0.05; **= p<0.01; ***= p<0.001).

Cognitive Technical Systems

Felix Putze, Tanja Schultz

Cognitive Systems Lab, Karlsruhe Institute of Technology, Karlsruhe, Germany.
{felix.putze, tanja.schultz}@kit.edu

Electrophysiological measurements like Electroencephalography (EEG) and Electromyography (EMG) enable us to measure human behavior by capturing neural and muscular activity. While traditional applications of those techniques were developed for medical and psychological purposes, since a few years they are also applied in the design of innovative human-machine interfaces. They allow a very direct measurement of user behavior and can also capture internal mental states which influence current and future behavior and which are not directly visible by other means. This presentation will show a number of state-of-the-art EEG- and EMG-based applications developed at the Cognitive Systems Lab.

EEG-based Brain-Computer-Interfaces (BCIs) are often categorized into two groups: Active BCIs are interfaces which allow direct and voluntary control of a system using natural or artificial “mental commands”. In contrast, passive BCIs constantly measure the user’s natural cognitive processes and react to them in an appropriate way without explicit user intervention. State-of-the-art examples of active BCIs are systems for spatial control. They are based on the paradigm of motion imagery, i.e. the imagination of movement of arm or leg. Imagined motion triggers characteristic activation patterns in the motor cortex which can be automatically detected. Traditional systems can recognize a small set of discrete motions. We present our current BCI for spatial control which includes speed trajectory reconstruction of continuous hand movement of complex natural movements.

A major application of passive BCIs is the recognition of mental workload. Using machine learning techniques, a system learns the most relevant frequency bands to discriminate low workload and high workload conditions. As the workload level influences behavior and performance of the user, the system adapts its own interaction strategy to make optimal use of the remaining cognitive resources of the user. We evaluate this system on a large corpus of over 150 participants in a driving scenario and achieved an accuracy of 82% for the person-independent workload recognition and an accuracy of more than 90% for person-dependent but session-independent recognition in another study. We also present results of a study which shows that implementing a workload adaptive interaction system improves both effectiveness and efficiency compared to a workload-oblivious system. Another application of passive BCIs deals with tracking the execution of cognitive processes. To optimally support the user during a complex task, it is useful to know which cognitive processing stages (perception, memory retrieval, attention shift …) are currently active. One example of such a system is an error-BCI which detects Error Related Potentials (ERPs) in the EEG signal. Those potentials occur when the user becomes aware of a mistake during interface operation or notices erroneous behavior of the system. Detecting ERPs allows the system to proactively recover from errors in the workflow. In this presentation, we show a gesture recognition system that uses this technique to automatically detect its recognition errors from single trials and proactively recovers from error states.

Electrophysiological signals are often polluted with biological and technical artifacts. A major challenge of modern BCI systems which are employed in uncontrolled environments and deal with complex tasks is the removal of artifacts and the identification of task-relevant brain activity. Techniques for signal decomposition and blind source separation like Independent Component Analysis are employed to isolate different influences on the signal and to separate relevant from irrelevant contributions. This way, we cannot only remove the influence of typical artifacts like eye blinks but also extract the most relevant components related to the task.

When dealing with tasks involving or allowing motor execution for control, EMG-based user interfaces are often an alternative to BCIs. An example are Silent Speech Interfaces which use facial surface EMG to record facial activity of the user while speaking. Using state-of-the-art speech recognition technology, these signals can be evaluated to transcribe the user’s speech input to the system even if it is uttered without producing audible sounds. Such system is useful in noisy environments, when silence is required (e.g. during meetings) or when the user cannot produce audible speech after a laryngectomy.

Neural Correlates of a Spatial Learning Task in Parietal Cortex, Prefrontal Cortex and Hippocampus

N. Becker, M.W. Jones

School of Physiology & Pharmacology, University of Bristol, Bristol, United Kingdom.
nadine.becker@bristol.ac.uk

Learning and execution of complex behaviours require interactions between networks of neurons that encode, store and share information across numerous brain regions. Systems level analyses of how the brain processes information are therefore central to understanding how behaviour is generated. Spatial learning in rodents constitutes a useful model in which to correlate neurophysiology with cognitive behaviours. We studied neuronal activity recording multiple individual neurons and local field potentials from chronically implanted tetrode arrays in the posterior parietal cortex (PPC), the medial prefrontal cortex (mPFC) and the hippocampus (HPC) of rats while they were trained in a maze task that involved decision making, working memory and rule learning based on spatial navigation. To date, recording single neuron firing from multiple brain areas in behaving animals is an unparalleled approach to shed light on the neuronal mechanisms underlying behaviour with millisecond resolution [1].

A key feature of hippocampal activity is the firing of place cells, HPC principal neurons encoding the spatial location of an animal. Like the HPC, the PPC is involved in spatial processing, with PPC neurons selectively responding to navigational features that are also represented in the HPC, such as directed body movements and spatial novelty [2]. Interestingly, there is no direct anatomical connection between PPC and HPC, yet the simultaneous processing of spatial information renders their coordination highly likely. In recent years, evidence accumulated that PPC neurons are concurrently controlled by a range of spatial and non-spatial features, involving temporal sequence and intention [3, 4, 5]. Like the mPFC, the PPC is part of the association cortex, which is thought to be the key site for integrating sensory and internal information to form decisions. PPC and mPFC receive a wide range of sensory afferents [6] and neurons in both structures have been found to adaptively respond to current cognitive demands [7, 8]. In addition, lesions of either brain region lead to impaired spatial learning [9, 10]. We hence expect spatial information to be coordinated between PPC, mPFC and HPC. A mechanism of how neuronal activity can be coordinated within and across brain regions has been linked to the second hippocampal hallmark, the theta rhythm, a 4-12 Hz oscillation arising during spatial exploration and other complex behaviours requiring working memory and decision making. Recent studies suggest that phase-locking of neuronal firing to hippocampal theta frequency underlies the functional connectivity between brain regions involved in solving complex spatial tasks [11, 12]. Furthermore, LFP coherence across distant brain regions has been suggested as a measure of synchronised neuronal activity, indicating transient coordination [11, 13]. We set out to assess how spatial information is differentially encoded in PPC, mPFC and HPC and how information processing is coordinated between these brain regions.

To this end, we chronically implanted rats with microelectrode arrays consisting of 20 individually adjustable tetrodes distributed over PPC, mPFC and HPC. In each brain region, we simultaneously recorded spiking activity of multiple individual neurons and local field potentials (LFP) as a measure of neuronal population activity. Recordings were made over up to seven weeks from each rat while they were trained on two subsequent sets of rules on an end-to-end T-maze. For each set of rules, rats were trained to perform 85% correct trials per day. The maze tasks involve distinct working memory and decision-making (choice) episodes and control (guided) episodes, which are spatially segregated. Same clock time-stamping was used for the neurophysiological recordings at a rate of 2 kHz for LFP and 32 kHz for action potential recordings and video-tracking of the animal's position via head-mounted diodes at a frame rate of 25 Hz. This experimental approach enabled us to assess coordination of neuronal activity on the single cell and population level within and across brain regions in relation to distinct task episodes, spatial location and training stages. Specifically, we analysed the firing patterns of each neuron using linearised firing rate maps with a spatial bin size of 11 cm, shuffle-corrected cross-correlations of spike trains of ± 0.5 seconds and 5 ms bins, and multi-taper Fourier analysis with a frequency resolution of 3 Hz and a temporal resolution of 0.5 seconds to calculate LFP coherence. We compared these parameters in choice versus guided trials to assess the impact of cognitive load and decision. Correct versus error trials were compared to determine the neurophysiological and temporal origin of error. For all trial types, left hand and right hand trials were analysed separately to examine the spatial component of solving the task. Finally, all analyses were performed at different training stages to check for neurophysiological correlates of learning.

As shown previously, HPC and PPC are required for spatial navigation, while HPC and mPFC are critically involved in working memory and are coherently active during decision making. Accordingly, we expected all three brain regions to be activated differentially during task episodes and learning stages that selectively recruit working memory, decision-making and rule acquisition. We demonstrate that spatial features were crucially involved in determining firing patterns in all three areas, but trajectory encoding was highly complex and not limited to spatial features in the cortical areas. While hippocampal firing was almost entirely controlled by location and running direction, the neuronal code in the cortical areas was determined by a combination of factors such as location, start-to-goal sequence, turn sequence and correct versus error trials. PPC and mPFC neurons hence exhibited much more complex firing patterns than place cells, nevertheless the patterns were highly reproducible on a trial-by-trial basis. Analysis of LFP coherence and single neuron activity across brain regions revealed task episode-specific coordination of neuronal activity in a training-dependent manner. In summary, we present an experimental approach that is able to shed light on the fundamental nature of information coding, sharing and integration in the brain.

All procedures were conducted in accordance with the UK Animals (Scientific Procedures) Act, 1986 and with the approval of the University of Bristol Ethics Committee.

References

1. Miller, E.K., Wilson, M.A. (2008). All my circuits: Using multiple electrodes to understand functioning neural networks. *Neuron* **60**(3), 483-488.
2. McNaughton, B.L., Mizumori, S.J., Barnes, C.A., Leonard, B.J., Marquis, M., Green, E.J. (1994). Cortical representation of motion during unrestrained spatial navigation in the rat. *Cereb. Cortex* **4**(1), 27-39.
3. Nitz, D.A. (2006). Tracking route progression in the posterior parietal cortex. *Neuron* **49**(5), 747-56.
4. Whitlock, J.R., Pfuhl, G., Dagslott, N., Moser, M.B., Moser, E.I. (2012). Functional Split between Parietal and Entorhinal Cortices in the Rat. *Neuron* **73**(4), 789-802.
5. Harvey, C.D., Coen, P., Tank, D.W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature*. doi: 10.1038/nature10918. [Epub ahead of print]
6. Miller, E.K., Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167-202.
7. Freedman, D.J., Assad, J.A. (2006). Experience-dependent representation of visual categories in parietal cortex. *Nature* **443**(7107), 85-88.
8. Stoet, G., Snyder, L.H. (2004). Single neurons in posterior parietal cortex of monkeys encode cognitive set. *Neuron* **42**(6), 1003-1012.
9. Kolb, B., Buhrmann, K., McDonald, R., Sutherland, R.J. (1994). Dissociation of the medial prefrontal, posterior parietal, and posterior temporal cortex for spatial navigation and recognition memory in the rat. *Cereb. Cortex* **4**(6), 664-680.
10. Mizuhara, H., Yamaguchi, Y. (2007). Human cortical circuits for central executive function emerge by theta phase synchronization. *Neuroimage* **36**(1), 232-244.
11. Jones, M.W., Wilson, M.A. (2005). Theta rhythms coordinate hippocampal-prefrontal interactions in a spatial memory task. *PLoS Biol*. **3**(12), e402.

12. Benchenane, K., Peyrache, A., Khamassi, M., Tierney, P.L., Gioanni, Y., Battaglia, F.P., Wiener, S.I. (2010). Coherent theta oscillations and reorganization of spike timing in the hippocampal- prefrontal network upon learning. *Neuron* **66**(6), 921-936.
13. Rodriguez, E., George, N., Lachaux J.P., Martinerie, J., Renault, B., Varela, F.J. (1999). Perception's shadow: long-distance synchronization of human brain activity. *Nature* **397**(6718), 430-433.

NMDA Receptor Antagonists Induce Antidepressant-like Sleep Changes: A Translational Model from Rats to Humans?

K.M. Wicke, G. Gross

Neuroscience Research, GPRD, Abbott GmbH & Co KG, Ludwigshafen, Germany.
karsten.wicke@abbott.com, gerhard.gross@abbott.com

Polysomnographic recordings enable the analysis of sleep behaviors and sleep pattern. Various classes of drugs exert common effects on sleep pattern: the vast majority of antidepressant drugs, for example, reduces Rapid Eye Movement (REM) sleep, whereas drugs that increase the cholinergic tone have the opposite effect. These observations have been used to translate effects of such medications across various species, from murines, cats, and dogs to healthy human volunteers and patients.

The anesthetic drug ketamine has recently been demonstrated to exhibit antidepressant-like effects in treatment–resistant major depressive disorder (e.g. [1,2]) and bipolar depression [3]. Antidepressant-like ketamine effects in animal models of antidepressant efficacy seem to correspond to these observations in humans (e.g. [4,5]). In this context we investigated whether antidepressant-like effects can also been observed in rat polysomnographic studies, as such studies are considered to be robust predictors of antidepressant effects in humans.

All experiments were conducted in accordance with the recommendations and policies of the National Institutes of Health Guide for the Care and Use of Laboratory Animals (1996 edition) and all applicable German laws and were approved by the respective Ethic Committee and Animal Welfare Offices. They were conducted in facilities with full Association for Assessment and Accreditation of Laboratory Animal Care accreditation.

Test compounds were administered to male Fisher rats chronically implanted with supracortical EEG electrodes. After dosing animals were transferred into individual cages inside a sound shielded box. Animals were randomly divided into 2 groups and tested in a crossover study design. At the first recording day they received either test compound or vehicle (sc), subsequently sleep EEG was recorded for 8h. After 7 days EEG was recorded a second time after receiving the alternative treatment (e.g. day 1 = test compound, day 8 = vehicle, or vice versa). All EEG recordings and analysis were performed with Sagura-Polygraph V2.0 (Dr. Cüppers Computer and Medizintechnik, Kelkheim, Germany). Analysis of sleep stages was performed in an automated manner and subsequently controlled by an experienced experimenter. For further polysomnographic analysis the following stages and patterns were differentiated: wake time, sleep I time (mild sleep), sleep II time (deep sleep including e.g. sleep spindles), REM time, number of REM episodes, latency to sleep and latency to REM.

Ketamine (3 - 30 mg/kg s.c.) significantly affected REM sleep parameters in an antidepressant-like manner (see Figure 1) similar to commonly used antidepressants (inhibitors of serotonin reuptake, tricyclics, and inhibitors of the monoamine oxidase). The number of REM episodes, and the total REM time were significantly decreased; likewise, latency to first REM was increased. These effects are mechanism-related as they were reproduced with the NMDA antagonist MK-801.

The above results clearly demonstrate an antidepressant-like profile of ketamine in rat polysomnographic studies consistent with efficacious doses found in other antidepressant models [6]. These results are in line with the clinical antidepressant effect of ketamine therapy.

NEWMEDS - the work leading to these results has received funding from the Innovative Medicines Initiative Joint Undertaking (IMI) under Grant Agreement N° 115008.

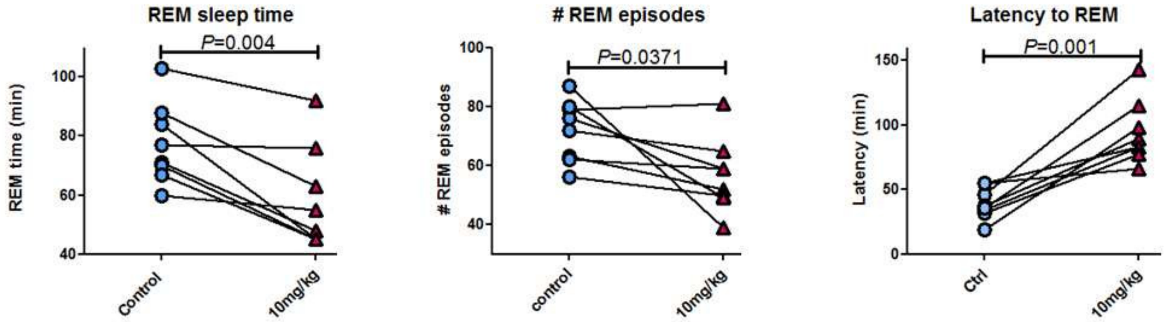


Figure 1. Ketamine (here at 10 mg/kg, ip) significantly affects REM sleep parameters in an antidepressant-like manner. The number of i) REM episodes and ii) the total REM time were significantly decreased; iii) latency to first REM was increased. (paired t-test, see P-values in figures)

References

1. Berman, R.M., Cappiello, A., Anand, A., Oren, D.A., Heninger, G.R., Charney, D.S., Krystal, J.H. (2000). Antidepressant effects of ketamine in depressed patients. *Biological Psychiatry* **47**, 351–4.

2. Ibrahim, L., Diazgranados, N., Franco-Chaves, J., Brutsche, N., Henter, I.D., Kronstein, P., Moaddel, R., Wainer, I., Luckenbaugh, D.A., Manji, H.K., Zarate, C.A., Jr. (2012). Course of Improvement in Depressive Symptoms to a Single Intravenous Infusion of Ketamine vs Add-on Riluzole: Results from a 4-Week, Double-Blind, Placebo-Controlled Study. *Neuropsychopharmacology* doi: 10.1038/npp.2011.338. [Epub ahead of print]

3. Diazgranados, N., Ibrahim, L., Brutsche, N.E., Newberg, A., Kronstein, P., Khalife, S., Kammerer, W.A., Quezado, Z., Luckenbaugh, D.A., Salvadore, G., hado-Vieira, R., Manji, H.K., Zarate, C.A.,Jr. (2010). A randomized add-on trial of an N-methyl-D-aspartate antagonist in treatment-resistant bipolar depression. *Archives of General Psychiatry* **67**, 793-802.

4. Garcia, L.S., Comim, C.M., Valvassori, S.S., Reus, G.Z., Andreazza, A.C., Stertz, L., Fries, G.R., Gavioli, E.C., Kapczinski, F., Quevedo, J. (2008). Chronic Administration of Ketamine Elicits Antidepressant-Like Effects in Rats without Affecting Hippocampal Brain-Derived Neurotrophic Factor Protein Levels. *Basic & Clinical Pharmacology & Toxicology* **103**, 502-506.

5. Pittenger, C., Sanacora, G., Krystal, J.H. (2007). The NMDA receptor as a therapeutic target in major depressive disorder. *CNS & Neurological Disorders Drug Targets* **6**, 101–15.

6. Maeng, S., Zarate Jr,C.A., Du, J., Schloesser, R.J., McCammon, J., Chen, G., Manji, H.K. (2008). Cellular Mechanisms Underlying the Antidepressant Effects of Ketamine: Role of [alpha]-Amino-3-Hydroxy-5-Methylisoxazole-4-Propionic Acid Receptors. *Biological Psychiatry* **63**, 349-52.

Quantitative EEG for the Diagnosis of Disease States

Björn Crüts[1], Pascal Römkens[2]

[1]*Brainmarker BV, Gulpen, The Netherlands. info@brainmarker.com*

[2]*Dept. of Medical Psychology, Atrium Medical Centre, The Netherlands. p.romkens@atriummc.nl*

An increasing number of patients with psychopathologies is being treated with pharmacological interventions. Currently the decisions whether or not to start pharmacological treatment and which drug is the most appropriate are not only based on evidence from clinical studies, but also rely heavily on personal experience of the clinician and trial-and-error processes. The trial-and-error method is essential, since outcomes from clinical trials cannot entirely predict treatment outcomes in a highly variable patient population as seen in the hospitals and practices. Knowledge generated by these trial-and-error processes have therefore to be captured and fed back into the clinical system, a procedure that is only minimally implemented in current day-to-day care.

In a consortium of clinics and practices in The Netherlands, existing behavioral measurements (such as questionnaires and neuropsychological tests) are combined with new measurements (brain activity) in order to objectify treatment choice and outcomes, thereby capturing the trial-and-error processes in the day-to-day care. The addition of brain measurements (in the form of quantitative EEG) are based on a growing body of evidence that shows that quantification of brain activity can evaluate effects of pharmacological interventions in a much more sensitive way. Above that, changes in bodily functions due to drug use can be detected much faster using brain measurements. Therefore, it is expected that the addition of brain measurements will ultimately result in a more reliable choice of drug treatment and faster and more reliable outcome monitoring in patients who use these drugs. We tested these hypotheses by means of a pilot study in which we added quantitative EEG in the routinely psychological testing of adult patients with depression and anxiety disorder of the Psychiatry Department and the Medical Psychology Department of the Atrium Medical Centre. We included 125 patients who filled out common symptoms- and personality questionnaires, before and after treatment. We found significant correlations between the use of certain drugs and brain activity. We also found typical brain activation patterns for depression and anxiety disorder. Some of these results are consistent with existing publications in the field of psychology and pharmaco-EEG, other results are not yet described in scientific literature in a consistent way. Especially these novel patterns in patients with depression seem to predict treatment outcomes of antidepressants.

Based on the results of this initial project, a larger project in two hospitals is now being planned, where brain and behavior measurements are implemented by means of a quality management system. The QM system will monitor and evaluate the following items and questions:

A) the predictability of the combination of brain and behavioral measurements for the effects of pharmacological interventions in patients with psychopathologies: can these measurements guide/quantify the choice of drugs in these patients?

B) the added value of brain measurements for the fast detection of treatment effects: can brain measurements evaluate drug interventions faster and more reliable than existing behavioral measurements, so that changes in the type and dose of drugs can be made faster and more reliable?

C) quantifying the added value per test: which tests are necessary for getting the required outcomes, and which tests do not provide any additional value and could potentially be eliminated from the measuring and monitoring process?

D) the added value of the combination of brain and behavioral measurements as a quality management system in clinical practice: can standardized measurement procedures improve patient satisfaction while reducing costs of pharmacological interventions?

This second project will include more than 1500 patients, ranging from adult patients with depression and anxiety disorder to children with development disorders (ADHD). We will present 1) results from the first study, and 2) consequences from these results for the quality management system in the second, larger study.

Keywords. mental health, pharmaco-EEG, quality management system, practice based evidence.

SPECIAL SESSION

The Development of a Diverse Battery of Behavioral Tasks Using Touchscreen Equipped Operant Boxes for the Study of Cognition in the Rodent

Chair: John Talpos

Janssen Research and Development, Beerse, Belgium. jtalpos@its.jnj.com

The use of touchscreen equipped operant boxes for the study of cognition in rodents has grown in popularity since the mid-1990s, when the first publications using the method appeared. The combination of numerous stimuli types and the large number of response rules that may be implemented when using a touchscreen equipped operant box affords the user the flexibility of a maze-based hand testing approach, with the consistency and throughput associated with operant testing. Accordingly, touchscreen technology is currently being used by a growing number of academic and industrial research groups. These groups are reinventing existing approaches, and creating the next generation of behavioral tasks to study cognition in the rodent.

This session will provide a general overview of the touchscreen approach, and demonstrate how it is making a unique contribution to the study of basic cognitive function and central nervous system disorders. Four areas will be discussed in additional detail. First is use of the touchscreen approach to study spatial "cognition" and neurogenesis using the trial unique non-match to location (TUNL) and spatial reversal learning paradigms, as well as a test of spatial reversal learning. Moreover this session will bridge the species boundary by also discussing how spatial cognition is studied in non-human primates using the touchscreen-based self-ordered spatial search paradigm. Secondly, because one of the most frequently used tasks within the touchscreen apparatus is the "visual discrimination", the utility of visual discriminations in studying animal models of schizophrenia will be discussed, as will the cross-site validation of pharmacological manipulations that is occurring within the NEWMEDS academic-industrial collaboration as part of the IMI (www.newmeds-europe.com). Next (3), the use of the touchscreen approach in studying cognitive flexibility will be considered with a focus on reversal learning and other novel developments. Finally (4), an overview of a touchscreen based paired associates learning (PAL) paradigm will be presented with a focus on its utility in the study of schizophrenia. This presentation will discuss the influence of specific brain regions on behavior in PAL, pharmacological sensitivity, as well steps that have been taken towards developing an acute model of schizophrenia for drug discovery. Through this diverse combination of topics and speakers, we hope to provide a firsthand practical account of how this technology is being applied in academic and industrial settings to enhance research methods for the study of cognition and novel treatments for disease.

The session will close with a dedicated question and answer period where attendees will have the chance to ask speakers questions of a practical and theoretical nature regarding their research using the touchscreen approach, and to allow a more thorough discussion of ongoing touchscreen activities with various IMI consortia (www.imi.europa.eu).

SPECIAL SESSION CONTENTS (sorted by paper ID)

Pharmacological Manipulation of a Rodent Paired Associates Learning (PAL) Paradigm, and Other Tasks for Use in Disease Research

John Talpos (Janssen Research and Development, Belgium)

How Can a Touch-Screen Based Visual Discrimination Help to Better Characterize Rodent Models of Schizophrenia?

Laetitia Fellini (Janssen Research and Development, Belgium)

Assessment of Behavioural Flexibility and Executive Function Using Novel Touch Screen Paradigms

Adam Mar, J. Alsiö, A. Haddenhorst, C.U. Wallis, L.M. Saksida, T.J. Bussey, T.W. Robbins (University of Cambridge, UK) and A. Trecker (Heinrich-Heine Universitaet Duesseldorf, Germany)

Perspectives on the Non-Human Primate Touch-Screen Self Ordered Spatial Search Paradigm

Jane Sutcliffe, Daniel Hutcheson

Neuroscience & CNS Research, Maccine Pte Ltd, Singapore.
jane.sutcliffe@maccine.com, daniel.hutcheson@maccine.com

Abstract

The self-ordered spatial search (SOSS) working memory task implemented as the computerised touch-screen task is a highly translatable platform providing a good comparison of human and non-human primate (NHP) cognitive performance measures. This communication describes how the manipulation of integral task components provides the ability to alter the dynamic range of range of performance suitable for the experimental approach being used. Accordingly, in the case of disruptors such as scopolamine for modelling cognitive decline and some aspects of brain disease states high baseline performances may be set to allow demonstration of disruption and reversal with potential new therapeutics. However, the SOSS task may be manipulated to obtain a large dynamic window of baseline performance optimised to detect the pure pro-cognitive impact of novel compounds.

Introduction

Rodents are an invaluable species for the evaluation of cognitive function during early drug development. However the use of non-human primates affords the ability to investigate complex behavioural processes that are more closely aligned with human function and disease with utility especially during the later stages of pharmaceutical drug lead optimization. Data derived from the NHP provides an invaluable, translatable link between data generated in rodents and in the human. Similarities in anatomy, neuronal circuitry, neurochemistry, functional and cognitive abilities exhibited between NHPs and humans provide an arguably more translatable approach to investigate higher cortical-mediated behaviours in particular, and allows for direct comparisons of cognitive function to be made using the same endpoints across species. The CAmbridge Neuropsychological Test Automated Battery (CANTAB) is a well established system for assessing cognitive function in patients across several neurodegenerative and psychiatric indications and for assessing responses to therapeutic agents. Touch-screen based cognitive tasks have been developed for use in NHP research from the early 1980's [1] and from conception the NHP-CANTAB tasks were adapted from the available clinical versions [2]. The SOSS task within the CANTAB battery shares obvious roots with the classic Hamilton search task [3-4], builds on the research of Mishkin [5-6] and is analogous to the rodent radial arm maze.

Methodology

In the clinic, the SOSS task is demonstrated as sensitive to frontal lobe damage [7] [8], disrupted in Schizophrenic patients [9] and in idiopathic Parkinson's disease [10] even when symptoms are mild [11]. Furthermore, Lange et al., 1992 [12] have demonstrated a distinct decline in SOSS performance in L-DOPA medicated Parkinsonian patients when "off" medication. The objective in the clinical version of the CANTAB spatial working memory task is to collect 'blue tokens' which are hidden beneath the boxes presented on the touch-screen and to then use the found tokens to fill an empty column present at the side of the screen. The subject must search by touching one box at a time to find the hidden token, on any one trial only one token is hidden. Once found, the next token will then become available in different location, the key feature of the task is that once a blue token has been discovered in one box location, that box location will not be used again to hide another token within that trial. Subjects are required to thus remember where they have found the previous token and not to return to that location. Non-verbal strategies have to be employed to train an NHP on the SOSS task and during initial training the NHP is gradually introduced to the concept of the task via the introduction of two coloured stimuli boxes. On touch selection of the stimuli the box may change colour or 'blink' to indicate that

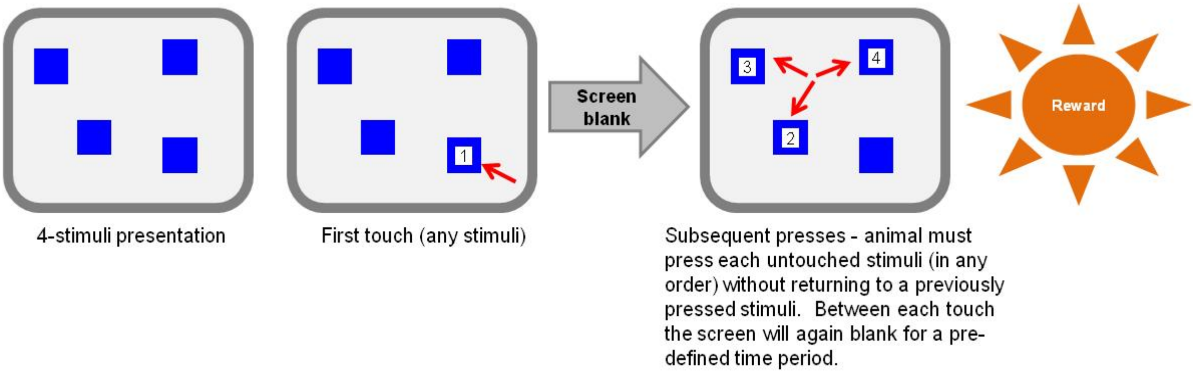


Figure 1. Schematic representation of the SOSS paradigm (4-stimulus trial).

the alternate stimuli must be sequentially selected (touched) in order to obtain the reward. As an individual's performance improves, the difficulty of the task is increased in a step wise manner: stimuli remain on screen, (no colour change or blinking) and the screen blanking is lengthened until terminal conditions are obtained. Importantly, during training care is taken to ensure rule comprehension and also that motivation to complete the task remains high. The training protocol for the cynomolgus macaque is adapted from the available rhesus macaque [13] and marmoset [14] literature. In this case up to six small coloured boxes (stimulus) may be displayed on the touch-screen in up to sixteen different screen locations. A set of stimuli are presented together at the start of each individual trial. The subject must touch each presented stimulus box without returning to a stimulus already selected for each trials set of stimuli presented in order to receive a reinforcer fruit food pellet (Figure 1). The trial is completed when the animal has either touched all boxes without repetition (correct), touched a box that had previously been selected in that trial (error), or failed to touch a box within 30 seconds of stimulus presentation (omission). After an inter-trial interval of 3 or more seconds, a new trial is presented with randomised stimuli of each selected difficulty level in newly allocated screen positions.

The SOSS task is very amenable to manipulation of the dynamic range of performance appropriate to demonstrate either the pro-cognitive impact of novel compounds or to conduct disruptor studies (e.g. scopolamine, ketamine). Disruptor induced reductions in performance must ensure easy level performance remains similar to control levels (i.e. providing an unaltered measure of global behaviour), but with working memory deficits apparent at the higher difficulty levels. This may be attained by starting with a higher baseline performance level which can be achieved by overtraining the subjects, using fewer stimuli, or minimising the inter-response screen blanking. Due to the flexible manipulations the SOSS task allows one to run studies at both high performance levels (for disruptor studies) and in separate sessions to create a large window of performance ranges (as low as 10-20% correct and stepwise up to >80%) in order to detect pure pro-cognitive changes in the same group of subjects.

Case Study

Scopolamine has been shown to disrupt cognitive performance on a wide range of memory tasks in numerous species including rats [15-17], nonhuman primates [18-20] and humans [21-25]. This effect is believed to be mediated through antagonism of the muscarinic subtype of the cholinergic receptor, and thus, the amnestic action produced in animals by scopolamine has been widely used pre-clinically as a model to characterize the potential cognition-enhancing influence of compounds. The SOSS task has previously shown sensitivity to scopolamine insults [26] in the young male rhesus macaque (~4 years) resulting in the proportion of correctly completed trials in the SOSS task being reduced with increasing trial difficulty (increasing the number of boxes in a trial) and by increasing dose levels of scopolamine. Importantly, performance on the 'easy' level condition (2-stimuli) was preserved with increasing dose levels of scopolamine whereas performance on the 3- and 4-stimulus was significantly impaired. The impact of scopolamine on the SOSS task has been investigated in our laboratory using the middle-aged (~15 years) and mature (~6 year) female cynomolgus macaque with similar effects observed when compared to the rhesus literature. During scopolamine titration studies extreme care must be taken to obtain an appropriate, individual, cognitive deficit without adversely affecting global behaviour. When using scopolamine to evaluate its actions on the central nervous system, particularly regarding memory, pretesting intervals of 30–45 min should be enforced since shorter pretest intervals may result in dominant peripheral anti-muscarinic side effects which will hinder task performance. In all cases the observed scopolamine deficit was replicated 2 or 3 times per individual prior to initiating reversal pharmacology. A robust, partial reversal of the scopolmaine induced cognitive deficit is observed with Donepezil pretreatment, confirming the clinical experience that cognitive improvements in Alzheimers disease after cholinesterase inhibitor administration exerts variable, often limited, responses in patient populations.

Conclusion

The NHP SOSS task is a highly translatable and sensitive neuropsychological paradigm which enables the pre-clinical researcher to model certain aspects of neurological disease states associated with cognitive decline. While a rodent touch-screen model of the SOSS task does not yet exist, its development would further aid translation across species.

Ethical Statement

All experimental procedures were conducted in an AAALAC accredited facility and approved by the Institutional Animal Care and Use Committee of Maccine Pte Ltd.

References

1. Gaffan, D. et al. (1984). Effects of fornix transection upon associative memory in monkeys: Role of the hippocampus in learned action. *The Quarterly Journal of Experimental Psychology Section B* **36(3)**, 173-221.
2. Roberts, A.C. and Sahakain, B.J. (1993). Comparable tests of cognitive function in monkey and man. *Behavioural Neuroscience: A Practical Approach*. (eds.) Rickwood, D & Hames, B.D, Oxford University Press, 164 - 184.
3. Hamilton, G.V. (1911). A study of trial and error reactions in mammals. *Journal of Animal Behavior* **1**, 33-66.
4. Levin, E.D. and Bowman, R.E. (1986). Scopolamine effects on Hamilton search task performance in monkeys. *Pharmacology Biochemistry and Behavior* **24(4)**, 819-821.
5. Mishkin, M. (1978). Memory in monkeys severely impaired by combined but not by separate removal of amygdala and hippocampus. *Nature* **273(5660)**, p. 297-298.
6. Mishkin, M. (1982). A Memory System in the Monkey. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **298(1089)**, 85-95.
7. Owen, A.M. et al. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia* **28,** 1021-1034.
8. Owen, A.M. et al. (1995). Visuo-spatial short-term recognition memory and learning after temporal lobe excisions, frontal lobe excisions or amygdalo-hippocampectomy in man. *Neuropsychologia* **33**, 1-24.
9. Piskulic, D. et al. (2007). Behavioural studies of spatial working memory dysfunction in schizophrenia: A quantitative literature review. *Psychiatry Research* **150,** 111-121.
10. Owen, A.M., et al. (1993). Visuospatial memory deficits at different stages of Parkinson's disease. *Neuropsychologia* **31,** 627-644.
11. Owen, A.M. et al. (1997). Spatial and non-spatial working memory at different stages of Parkinson's disease. *Neuropsychologia* **35**, 519-532.

12. Lange, K.W. et al. (1992). L-dopa withdrawal in Parkinson's disease selectively impairs cognitive performance in tests sensitive to frontal lobe dysfunction. *Psychopharmacology (Berl)*, **107,** 394-404.

13. Weed, M.R. et al. (1999). Performance norms for a rhesus monkey neuropsychological testing battery: acquisition and long-term performance. *Brain Res Cogn Brain Res* **8,** 185-201.

14. Roberts, A.C. (1996). Comparison of cognitive function in human and non-human primates. *Brain Res Cogn Brain Res* **3,** 319-27.

15. Fleck, C. et al. (2010). Investigation into the in vivo effects of five novel tacrine/ferulic acid and beta-carboline derivatives on scopolamine-induced cognitive impairment in rats using radial maze paradigm. *Arzneimittelforschung* **60,** 299-306.

16. Li, H.B. et al. (1996). NMDA-but not AMPA-receptor antagonists augment scopolamine-induced spatial cognitive deficit of rats in a radial maze task. *Brain Res* **725**, 268-71.

17. Rispoli, V. et al. (2004). Choline pivaloyl esters improve in rats cognitive and memory performances impaired by scopolamine treatment or lesions of the nucleus basalis of Meynert. *Neurosci Lett* **356**, 199-202.

18. Bartus, R.T. and Johnson, H.R. (1976). Short-term memory in the rhesus monkey: disruption from the anti-cholinergic scopolamine. *Pharmacol Biochem Behav* **5**, 39-46.

19. Buccafusco, J.J. (2009). *The Revival of Scopolamine Reversal for the Assessment of Cognition-Enhancing Drugs*. 2nd ed. Methods of Behavior Analysis in Neuroscience. Boca Raton (FL): CRC Press.

20. Terry, A.V. Jr., Buccafusco, J.J. and Jackson, W.J. (1993). Scopolamine reversal of nicotine enhanced delayed matching-to-sample performance in monkeys. *Pharmacol Biochem Behav* **45**, 925-9.

21. Robbins, T.W. et al. (1997). Effects of scopolamine on delayed-matching-to-sample and paired associates tests of visual memory and learning in human subjects: comparison with diazepam and implications for dementia. *Psychopharmacology (Berl)* **134**, 95-106.

22. Preston, G.C. et al. (1988). The scopolamine model of dementia: determination of central cholinomimetic effects of physostigmine on cognition and biochemical markers in man. *J Psychopharmacol* **2**, 67-79.

23. Broks, P. et al. (1988). Modelling dementia: effects of scopolamine on memory and attention. *Neuropsychologia* **26**, 685-700.

24. Drachman, D.A. (1977). Memory and cognitive function in man: does the cholinergic system have a specific role? *Neurology* **27**, 783-90.

25. Drachman, D.A. and Leavitt, J. (1974). Human memory and the cholinergic system. A relationship to aging? *Arch Neurol* **30**, 113-21.

26. Taffe, M.A., Weed, M.R. and Gold, L.H. (1999). Scopolamine alters rhesus monkey performance on a novel neuropsychological test battery. *Brain Res Cogn Brain Res* **8**, 203-12.

Pharmacological Manipulation of a Rodent Paired Associates Learning (Pal) Paradigm, and other Tasks for Use in Disease Research

J.C. Talpos

Janssen Research and Development, Beerse, Belgium.
jtalpos@its.jnj.com

Ethical Statement

All experiments described in this abstract have been approved by the local Ethical Committee at Janssen Pharmaceutica**.**

Introduction

Paired Associated learning (PAL) as part of the Cambridge Neurological Test Battery (CANTAB) is a task of great potential for disease research. The initial interest in this task was sparked by studies of patients with Alzheimer's disease (AD) or mild cognitive impairment (MCI), often viewed as a prodromal form of AD. AD patients were highly impaired on PAL, more importantly, so were a sub-population of MCI patients. Subsequent studies would reveal that this "impaired" sub-population of MCI patients showed a much higher conversion rate to AD than the rest of the cohort. Through the use of PAL and related tests, researchers were able to predict who would, and who would not convert to AD from an at-risk population several years earlier than with other available clinical tests (1). While the link between PAL and AD was being established, another line of research began to associate PAL with schizophrenia. Specifically patients suffering from first episode psychosis were found to be impaired in PAL. Furthermore performance on PAL was found to be more predictive of global cognitive function in schizophrenia than the extra dimensional / intra dimensional shift (ED/ID) task, often considered a gold standard for research into cognitive impairments in schizophrenia (2). As such this one task appears to have great value in researching cognitive impairments associated with AD and schizophrenia, and a pre-clinical model of this assay would be of great utility.

Although PAL exists in many forms, we are most concerned with modeling the CANTAB version. In this form PAL is a task of visuo-spatial learning. Participants are shown a series of images in distinct locations upon a computer screen (4-8 locations depending on the difficulty). One at a time, these same images are displayed in the middle of the screen and the participant is asked to touch where they were previously seen. In this way subjects are required to pair a specific object with a specific spatial location. Remembering the object or the location alone is not enough to solve the task; rather the participant must use both modalities.

The CANTAB version of PAL has been modeled in the rodent by using touch-sensitive computer monitors in an attempt to capture the visual and spatial elements of the human task. In the rodent version of the task the screen is divided into three distinct locations, each location is associated with a specific image. (Talpos et al., 2009) On each trial two images are shown, one in the correct location to earn a reward, and one in an incorrect location (see Figure 1, adapted from 4). In this way an animal can only earn a reward by making the connection between a specific object and a specific location. In order to learn this task rats or mice must be trained for many sessions. The rodent version of the task is not quickly acquired, nor can trial-unique stimuli be used as is the case in the human version of the task. However this rodent variant does require the binding of spatial and visual features in a similar way as CANTAB PAL. While this rodent PAL task has many differences from its human counterpart, it also has important similarities.

The neurobiological underpinnings of PAL in rodents are still being explored. However it has been shown that when rats have learned PAL the task is sensitive to manipulations of the hippocampus, whereas a highly similar control task is not. However the primary impetus for development of this task was specifically to aid the drug discovery process. In order to do this the pharmacological under-pinning of the task must be explored.

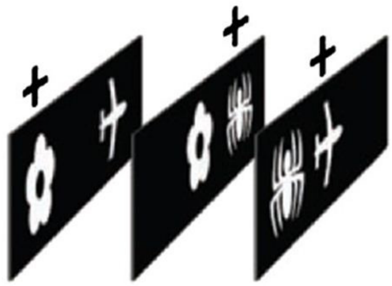


Figure 1. Scheme of the CANTAB version of PAL for rodents.

While acquisition of PAL could be used to discover potentially cognitive enhancing compounds, at the moment the task is learned far too slowly to make this practical. As such animals that have achieved a steady state will likely need to be used. However in PAL, animals routinely achieve a steady state of behavior at about 80% correct, leaving little room to detect a statistical improvement. Furthermore it is unclear if it would be possible to detect enhancement in normal animals performing at a near optimal level. To circumvent these challenges, we focus on frequently used pharmacological manipulations in an attempt to develop a model of disease and further understand the neuro-pharmacological basis of this task.

We initially focused on 4 compounds, PCP, Ketamine, amphetamine, and LSD, representing compounds used to model schizophrenia, and also known to disrupt learning, memory, or perception.

Material and Methods

Male Lister-hooded rats were used for this work. Rats were maintained on a food restricted diet keeping them at about 85% of free feed body weight.

The chambers used were modified med-associates operant boxes running commercially available software (K-Limbic, by Conclusive Solutions). Critical modification of these boxes included the addition of a “shelf”/ counter weighted flap. The purpose of this is to slightly slow the animals’ response, increasing the likelihood that they will attend to the displayed stimuli. Experience dictates that this modification is extremely important for successful task acquisition. A removable mask was used to divide the chamber into distinct regions.

Pre-training: Rats were trained with a procedure very similar to that used by Talpos et al. (2009). Briefly, rats were first habituated to the chamber. This was done but putting animals in the test chambers with small amounts of a reward pellet/peanut butter mixture placed upon the “flap”, “mask”, and screen. This was done to encourage animals to explore the screen. Animals were left in the chambers for approximately 30mins a day until they had eaten all of the peanut butter from the screen—typically one day. Next animals were trained to associate a tone with a food reward. A trial would start with the delivery of a food pellet, the sounding of the reward tone, and the activation of the magazine light. Once the pellet was collected the reward light would go out and a 30sec delay would start. Once 30 sec had passed another pellet would be delivered in combination with activation of the reward light and tone. This continued for 60 trials or 60mins. An animal was considered trained on this stage once they managed to complete 60 trials within 60mins (typically two days). Animals would then progress to screen touch training. A session would start with the delivery of a pellet. Collecting the pellet would cause the magazine light to deactivate and the screen to illuminate. A response at any portion of the screen would start the reward sequence (de-activation of the screen, delivery of a pellet, activation of the magazine light, and a short tone). Once the pellet was collected the reward light would go out and the inter trial interval (ITI) would begin. At the end of the ITI the magazine light would again be activated, and a poke at the light would start the next trial. A session was considered complete when an animal could complete 60 trials in 45mins. An animal was considered trained when they could successfully complete a session (typically 2 or 3 days). Successful animals were advanced to the “one location” stage, where only a single region of the screen was activated/rewarded, but in all other regards this was identical to the previous stage. Typically animals only needed one day to master this and were then progressed to the one random location. Again, this was the same as the previous stage, however now the location moved between trials. This was included to help avoid any potential side biases that may have developed during the pre-training. Most animals would complete 60 trials within 45mins on the first day, and were therefore ready to start task training.

Task Training: A trial began with activation of the reward magazine and delivery of a free pellet. Once this pellet is collected the magazine light is deactivated and two images are displayed upon the screen. A response at the correct location activates the reward sequence. Once the reward is collected the ITI begins, and at the end of the ITI the reward magazine is again illuminated. A response at the magazine will cause it to turn off, and begin the next trial. However if the incorrect stimuli is selected then the images are removed from the screen and the house light turns off (punishment period) for 10 seconds. Once 10 seconds have passed the ITI begins, and then the trial proceeds as normal except that the animal will experience the same trial until a correct response is made (correction trial). However these correction trials are not counted towards the total of trials completed. The reason for using the correction trials is to prevent the development of a side bias during the training of this long and difficult task. Animals are typically trained for 72 trials or 45mins, whichever occurs first. Training on this task is very long, requiring about 3-4 months to reach an asymptotical performance level of approximately 80% correct. Once animal behavior has stabilized animals are considered ready for compound testing.

Results

The NMDA antagonists, PCP and Ketamine, had partially distinct profiles in this test. Ketamine (up to 10 mg/kg) did not induce an impairment in accuracy. However the highest dose was associated with an increase in response latency and suppression of behavior. PCP (1.5 mg/kg) on the other hand did impair accuracy, but this effect was modest and also occurred with some side effects. A larger deficit was observed at 2.0 mg/kg, but this was also associated with a more robust side effect profile. Perhaps not surprisingly, LSD did not disrupt accuracy, although abnormalities were seen on secondary measures. Finally, amphetamine induced a substantial decrease in accuracy (0.5 mg/kg) without changes in secondary measures. While a larger, albeit non-selective impairment was also observed at 0.75 mg/kg. Of the compounds tested only amphetamine appears to possess the dynamic range and selectivity to serve as pharmacological challenge model (Figure 2). Moreover previous work has indicated these concentrations of PCP have little or no effect on performance of a touch-screen based visual discrimination, a related, but less challenging assay (Talpos et al., 2011).

Additional research has been performed with the amphetamine impairment model indicating that the effect is extremely robust (significant impairments at 0.5 mg/kg in over 15 instances). Our initial validation efforts have focused on reversing this deficit with antipsychotics. Studies indicate that clozapine, haloperidol, and risperidone are all capable of at least partially reversing this deficit; and the profiles of these compounds within this model are partially distinct.

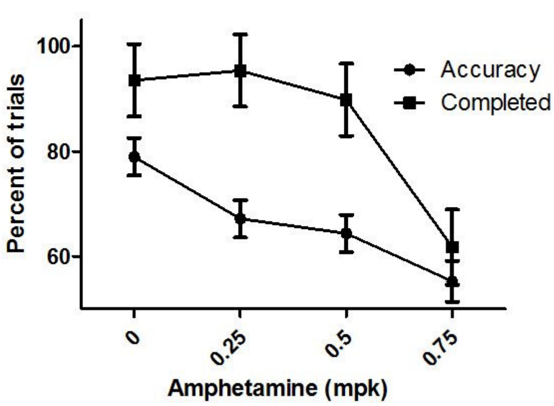


Figure 2. The effects of amphetamine on performance.

These current data indicate that PAL in the rat is exquisitely sensitive to dopaminergic hyperfunction induced by a relatively low dose of amphetamine. Moreover this deficit can be reversed by numerous antipsychotic compounds that have D2 receptor antagonism as their primary mode of action. Considering the lack of overt side effects, we do not think this is a model of hyperactivity or motoric impulsivity, although this cannot be ruled out. However additional work will be required to determine the origin of this deficit and to determine if this model of cognitive impairment can add value to the drug discovery process.

References

1. Blackwell, A.D., Sahakian, B.J., Vesey, R., Semple, J.M., Robbins, T.W., Hodges, J.R. (2004). Detecting dementia: novel neuropsychological markers of preclinical Alzheimer's disease. *Dement Geriatr Cogn Disord* **17**,42-48.

2. Barnett, J.H., Robbins, T.W., Leeson, V.C., Sahakian, B.J., Joyce, E.M., Blackwell, A.D. (2010). Assessing cognitive function in clinical trials of schizophrenia. *Neurosci Biobehav Rev* **34**, 1161-1177.

3. Talpos, J.C., Winters, B.D., Dias, R., Saksida, L.M., Bussey, T.J. (2009). A novel touchscreen-automated paired-associate learning (PAL) task sensitive to pharmacological manipulation of the hippocampus: a translational rodent model of cognitive impairments in neurodegenerative disease. *Psychopharmacology (Berl)* **205**, 157-168.

4. Clelland, C.D., Choi, M., Romberg, C., Clemenson, G.D., Jr., Fragniere, A., Tyers, P., Jessberger, S, Saksida, L.M., Barker, R.A., Gage, F.H., Bussey, T.J. (2009). A functional role for adult hippocampal neurogenesis in spatial pattern separation. *Science* **325**, 210-213

5. Talpos, J.C., Fletcher, A.C., Circelli, C., Tricklebank, M.D., Dix, S.L. (2011). The pharmacological sensitivity of a touchscreen-based visual discrimination task in the rat using simple and perceptually challenging stimuli. *Psychopharmacology (Berl).*

The Touchscreen Cognitive Testing Method for Mice and Rats

Tim Bussey

Department of Experimental Psychology and The MRC and Welcome Trust Behavioural and Clinical Neurosciences Institute, University of Cambridge, UK. tjb1000@cam.ac.uk

Overview

What are the characteristics of the ideal rodent cognitive testing method? In my talk I will suggest the following, which I put forward in a recent publication in a special issue of *Neuropharmacology* (Bussey *et al*., 2011).

"i. It should be automated

Notwithstanding some unique arguments to the contrary (Crabbe and Morris, 2004), most researchers agree that the same type of automation that has paid such dividends in areas like molecular biology is desirable for cognitive testing as well. Certainly the cognitive testing of clinical populations is becoming increasingly automated (e.g., the Cambridge Neuropsychological Test Automated Battery (CANTAB), CogState, Mindstreams), and so automation may help to achieve another desirable: effective translation (discussed in more detail below). But automation of cognitive testing in animal models confers advantages of its own; perhaps the most obvious to researchers 'on the ground' is a reduction in labour: while one experimenter is testing a single mouse in a maze, an experimenter armed with automated apparatus can test many mice simultaneously. Post-docs working in our lab, for example, routinely test 20 animals at a time, adding up to between 1 and 2 hundred a day. This realization negates the argument that some automated tests take more sessions to generate a result, as such tests would have to be 20 times slower before the equivalent non-automated testing methods start to make any time-economical sense. But it is the scientific arguments for automation that are perhaps the most compelling. With an automated test the computer can present many trials to an animal without having to interfere with the animal during testing. Such interference can have a huge effect, not just by introducing variability into the way the task is run, but also by introducing variability and potential confounds relating to the way the experimenter handles the animals, or even the sight or smell of the experimenter (Wahlsten et al., 2003). Finally, because automated tasks are controlled by a computer, parameters such as delays and stimulus presentation are identical on every trial, for every animal, and the measures the computer gets back from the animal, for example latencies, can be accurate to the millisecond. All this adds up to a method that is much more likely to be reliable than non-automated methods.

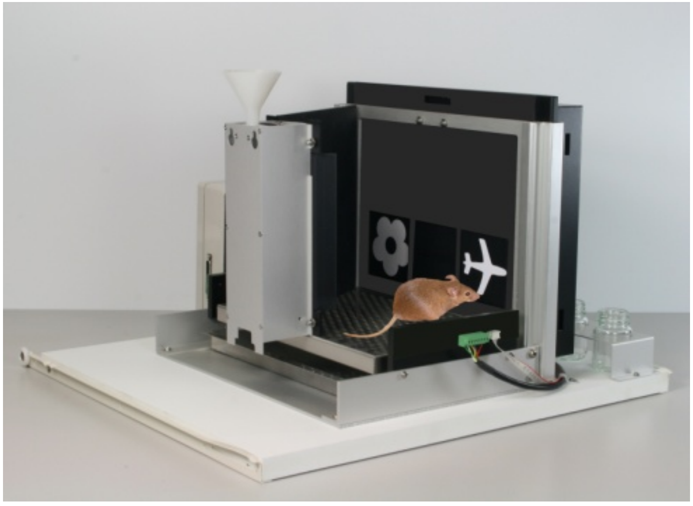


Figure 1. Touchscreen testing setup.

ii. It should be non-aversive and low-stress

Although some animal models, including those of schizophrenia, utilize stress as a manipulation, unnecessary stress as a feature of a cognitive test is best avoided. Everyone knows that stress can interfere with our experimental manipulations, and can have strong modulatory effects on cognition (Joels and Baram, 2009). Unwanted stress can therefore be a major impediment to the accurate assessment of cognition. Yet some of the most popular methods for cognitive testing involve cold water or electric shock. Such unwanted stressors are best avoided if possible, and so we would suggest that the ideal cognitive testing method should if possible be appetitive (reward-based) rather than aversive. Of course the use of appetitive reinforcers is not always problem-free either; some perturbations in experimental animals can affect animals' motivation to perform vigorously to obtain reward. However appetitive reinforcers themselves generally have a less potentially confounding effect on the organism than stressors.

iii. It should assess multiple cognitive domains

Currently we often assess an animal's memory in, say, a pool of water, its ability to extinguish a learned association using, say, foot shock, and its attentional capacities using, say, an operant chamber. Then we try to compare across these tests. Obviously there are difficulties with this approach. Ideally we would be able to investigate all of these aspects of cognition under the same conditions -- in the same apparatus, using the same types of stimuli, and the same rewards and responses. We can then employ a 'battery' of tests to establish a cognitive profile of an animal, in conditions under which the tests can be compared in a less confounded manner.

iv. It should be translational

The goal of all of this is of course translation, from mouse to clinic. Currently the tests we use in animals are usually nothing like the tests we use in clinical populations; some of the test mentioned above provide good examples of this disconnect. In order to translate effectively our findings from animals to humans, the tasks ideally would be as similar as possible. Such 'face validity' does not guarantee construct or predictive validity – further experimentation is needed to do that – but it makes these types of validity, and effective translation, much more likely than when the tests appear on the face of them to have little in common."

In my talk I will describe one method – the touchscreen testing method -- which I and others have been working on for about 20 years now, which we think satisfies this list of desirables and could prove to be very valuable in the cognitive testing of rodent models. I describe that method generally, before describing some of the tests that are currently implemented and providing some examples of how these tests can be used. These tests can be used as part of a 'flexible battery' approach for the cognitive profiling of rodent models.

References

1. Bussey, T.J., Holmes, A., Lyon, L., Mar, A.C., McAllister, K.A.L., Nithianantharajah, J., Oomen, C.A. & Saksida, L.M. (2011) New translational assays for preclinical modelling of cognition in schizophrenia: the touchscreen testing method for mice and rats. *Neuropharmacology,* **62**, 1191-1203.

2. Crabbe, J.C., Morris, R.G. (2004). Festina lente: late-night thoughts on high-throughput screening of mouse behavior. *Nat Neurosci* **7**, 1175-1179.

3. Joels, M., Baram, T.Z. (2009). The neuro-symphony of stress. *Nat Rev Neurosci* **10**, 459-466.

4. Wahlsten, D., Metten, P., Phillips, T.J., Boehm, S. L., 2nd, Burkhart-Kasch, S., Dorow, J., Doerksen, S., Downing, C., Fogarty, J., Rodd-Henricks, K., Hen, R., McKinnon, C.S., Merrill, C.M., Nolte, C., Schalomon, M., Schlumbohm, J.P., Sibert, J.R., Wenger, C.D., Dudek, B. C., Crabbe, J.C. (2003). Different data from different labs: lessons from studies of gene-environment interaction. *J Neurobiol* **54**, 283-311.

How Can a Touch-Screen Based Visual Discrimination Help to Better Characterize Rodent Models of Schizophrenia?

L. Fellini

Janssen Research & Development, Beerse, Belgium. LFellini@its.jnj.com

Ethical statement

All animals were treated in accordance with the European Ethics Committee (decree 86/609/CEE), the Animal Welfare Act (7 USC 2131) and the Guidelines for the Care and Use of Mammals in Neuroscience and Behavioral Research (National Research Council 2003). The study protocol was approved by the local animal experimental ethical committee at Janssen Research & Development (Beerse, Belgium).

Introduction

Existing in many forms, two of the most commonly used cognition paradigms in the rodent are rule acquisition and reversal learning (e.g. swim mazes, operant tasks, T-mazes). Despite the wide spread use of these simple assays, interpretation is often not straightforward and important controls are commonly over looked. For example, when acquiring a simple discrimination (a response to stimulus “A” is rewarded, whereas a response to stimulus “B” is not), necessary for many forms of operant learning, the animal must learn several concepts. Rodents must (1) learn to distinguish the stimuli, (2) learn that only one of the stimuli is associated with reward, (3) and that earning a reward is dependent upon their own behavior. Accordingly an impairment in task acquisition could be caused by a failure in any one of these processes or other non-specific effects, like alterations in motoric abilities. Reversal learning has similar, but unique, difficulties in interpretation. In rodents, reversal learning paradigms can take days, and are often split into three distinct phases. Early reversal is the first stage of the process, occurring immediately after the switch in rule contingencies associated with reward. The rodent’s behavior is capable of earning a reward, but its current behavior is not; it must learn the new contingency. Until this occurs, the animal will consistently perform below chance. As it comes to understand this change, the animal will begin to experiment with other responses to see if they earn a reward (middle reversal), and their accuracy will begin to increase towards chance level performance. Finally, as the animal learns that it can still earn a reward, it must learn which stimuli is now associated with a reward, a stage called late reversal. Depending upon the stage reversal learning does require different neurobiological and psychological processes (for example see Bussey et al., 1997, Chudasama and Robbins, 2003), some of which may overlap with acquisition learning. Considering the numerous processes involved in acquisition of a discrimination, and the overlap between reversal learning and acquisition of a discrimination, it is important to understand the psychological process impaired, not just the stage impaired (i.e. visuo-perceptual changes will impair acquisition, but this does not mean the animal is impaired at learning).

To this end, we used touch-screen technology to develop the Double Visual Discrimination task (DVD). Using modified operant chambers animals are trained to nose-poke at “images” displayed upon a touch-sensitive computer monitor to earn a reward. The “rules” dictating this reward can be modified to fit the experimental question at hand and should help to dissociate the individual elements of rule learning, stimulus discrimination, recall and reversal learning, in one procedure. In the DVD procedure animals are first trained to perform a visual discrimination (learn that a response at image “A” results in a reward whereas a response at image “B” does not). When this has been acquired a second pair of images is introduced, however some trials are still given with pair 1. In this way animals must learn a second discrimination while still performing the first. Finally, once the second pair has been learned, the reward outcomes of pair 1 are reversed (see figure 1 for example baseline data). In this way the DVD procedure should be able to dissociate the cognitive processes affected by a manipulation, rather than just the gross behavior (recall, acquisition, reversal learning). Although this new paradigm can be applied to numerous disease models, in this instance we have selected acute PCP challenge. Often used as a model of schizophrenia, the NMDA receptor antagonist is also known to disrupt learning and other cognitive functions. While evidence suggests that PCP given acutely can influence acquisition of a task, reversal learning, and baseline task performance, published studies do not use cohesive methodologies. As such, the comparative nature of these deficits, the concentration at which they occur, and the associated side effect profiles remain largely undescribed in a unified testing environment. Accordingly, we have performed dose response studies at each stage of the DVD procedure to elucidate the nature of PCP impairments on several cognitive processes using a unified testing procedure.

Materials and methods

Animals: Adult male Lister-Hooded rats (purchased from Harlan, The Netherlands) were used in this study (230- 250 g) at the beginning of the training). Access to food was restricted so that rats were maintained at 80-85% of their free-feeding body weight.

Apparatus: Testing was conducted in Med-associate chambers (internal dimensions: around L 33 cm x W 40.5 cm x H 29 cm) equipped with a touch screen (40 x 25 cm). A metal flap (36.5 x 5 cm) was located in front of the touch-screen, around 5 cm above the grid floor. A black metal insert was added on the touch-screen (1 cm from it) to create 4 response windows (8.2 x 15 cm). The wall located opposite to the touch-screen presents the food magazine (equipped with an infrared nose pokes detectors and a light), the loud-speaker and the house light. The testing apparatus was located in a sound-attenuating chamber (H: 60cm x W: 74cm x L: 60cm) ventilated by a fan. Stimuli and data recordings were controlled using K-Limbic software.

Pre-training: The experiment began with a short pre-training procedure during which the rats learn to vigorously respond to the touch-screen. This started with a tone training procedure where rats were taught to associate a “tone” with a delivery of a food reward. This was followed by a procedure where any response at the screen was rewarded while it was illuminated. When rats where reliably doing this they were advanced to the final stage of pre-training where only a portion of the screen would be illuminated and only responses at the illuminated portion were rewarded. A response at non-illuminated areas of the screen had no consequence.

The DVD task: The task itself takes place in three stages: (1) Simple Visual Discrimination, (2) acquisition of a second discrimination while still performing the first discrimination, and finally (3) the reversal of one of the previously learned pairs. During acquisition of the first stage a response at one image always results in a reward, whereas a response at the other image would lead to a short timeout (10sec) during which the house light was deactivated. Once the timeout period elapsed, the animal would attempt the same trial again, i.e. a correction trial. During this stage, training sessions were made of 48 trials or lasted 30 min whichever occurred first. Once animals had mastered this task and reached criterion (i.e. accuracy higher than 85% and standard deviation of less than 10 for at least three successive days) they were moved to the second experimental stage, the double visual discrimination. During this step, they were required to learn a second discrimination with new images, while also being presented with the original image pair. Both sets were presented but on separate trials and were randomly mixed across trials. Finally, after the second discrimination had also been mastered, the rats experienced a reversal where the contingencies associated with the first learned pair were reversed. Training stages with two sets of stimuli consisted of a total of 96 trials or lasted 60 minutes whichever occurred first. For a graphical depiction of this procedure, see figure 1.

Stimuli: In this paradigm, two pairs of stimuli were used: Flower/Bomb and Spider/Plane, each stimulus being ~5.5cm X 5.5cm size and highly similar to those used by Bussey et al. (1). The pair of stimuli presented during each experimental stage were counterbalanced between animals. For each pair of stimuli, only one stimulus was rewarded (S+) and the S+ of each pair being counterbalanced between rats. Importantly, the location of the S+ and S- were randomized over the four windows between trials. 4 windows were used to prevent the animal from seeing the two possible stimuli arrangements as a “scene” and responding using an “if ‘A’ go right, if ‘B’ left”, and to instead perform the task as a two-way discrimination.

The comparison between the acquisition of the first pair and acquisition of the second pair may allow the dissociation between procedural and stimuli learning, versus stimuli learning alone. Moreover, during acquisition of the second pair, the original pair also serves as a control for changes to “recall” or non-specific effects that may disrupt cognition. Finally, the comparison between acquisition phases and reversal learning will help determine if any deficits observed are specific to reversal learning, or if they affect learning in general.

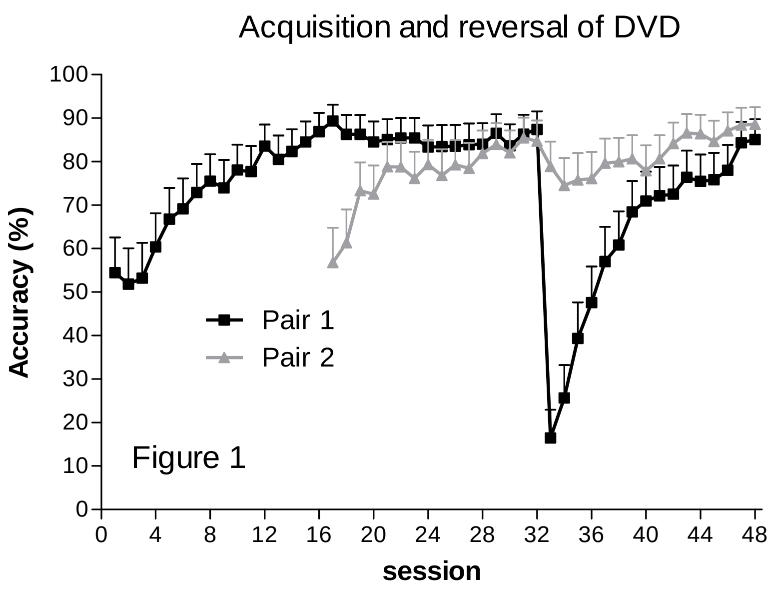
Treatment: Rats received either subcutaneous injections of phencyclidine (PCP group) or vehicle (vehicle group) 30 min before each training session. Solutions were administrated at 1ml/kg (s.c.).

Behavioral analysis: Animals' performance was recorded and collected automatically using K-Limbic software. For each training session the following measures were analyzed: the percentage of completed trials (number of trials completed /maximum number of trial*100; the mean accuracy (percentage of correct response relative to the number of completed trials), the mean response latency (i.e. time to select the correct/incorrect choice once the two stimuli are displayed on the screen) and the mean collection latency (i.e. time to collect the food reward in the food magazine once it has been delivered). Importantly, correction trials were not taken in consideration in the analysis of these parameters. They were however used to calculate correction trial accuracy, i.e. the number of correction trials required for an animal to get a correct response after it responded incorrectly to a trial.

Statistical analysis: Accuracy was analyzed using a non-linear regression model and complementary analysis was performed using a two-way ANOVA with treatment as the between group factor and training sessions as the within group factor. Dunnett post-hoc comparisons were made where appropriate, having the vehicle group as a control group. All other parameters were analyzed using two-way ANOVA as described above.

Results

Doses of PCP used in behavioral studies can often cause marked hyper-locomotion and impairments in motoric functions. However in these studies very low doses of PCP (0.25-1 mg/kg) were used in hopes that any effects observed would be the result of changes in "cognition" as opposed to non-specific side effects influencing the fundamentals of task performance. During acquisition 1, PCP caused dose-dependent disruptions in learning, with a modest effect seen at 0.5 mg/kg and a more robust impairment at 1.0 mg/kg. Importantly, no signs of motoric impairments were observed. In fact, animals had a tendency to respond faster when given PCP, likely a result of PCP's stimulant properties. On the second acquisition, significant impairments were seen at 0.5 mg/kg with minimal effects on the previously learned pair. These data suggest that PCP can disrupt image acquisition independently of rule learning. Finally, PCP also impaired reversal learning, while having minimal effects on the unchanged stimuli. These data indicate that while PCP may disrupt learning about stimuli, it is also capable of disrupting adaptation to new reward contingencies since the stimuli themselves do not change during reversal learning.



Conclusions

Using this procedure, we have demonstrated that PCP disrupts numerous aspects of cognition within a unified paradigm. Moreover, these impairments occur at doses lower than those often used in behavioral studies, and occur with minimal side effects. While PCP may disrupt reversal learning, it also impairs acquisition which should also be involved in reversal learning. As such a PCP challenge appears to impair task acquisition in general, as opposed to being specific to reversal learning, although more research is necessary to confirm this. Importantly, these effects on reversal learning and acquisition occurred with minimal effect on recall of a second pair of images suggesting that the effects seen are not the result of a global disruption.

Despite its popularity as a research tool, acute PCP treatment as a pharmacological tool to study cognition or as a model of psychosis has not been comprehensively described. Using the DVD procedure, we have described how low doses of PCP effect acquisition, reversal learning and task performance within a unified procedure. These data highlight the utility of the DVD procedure in allowing the simultaneous measurement of multiple cognitive processes within a single animal. The DVD procedure can be used to gain confidence in the selectivity of a specific change in cognition. Furthermore, DVD can also be applied to other behavioral models, and may be particularly effective at helping to describe the effects of chronic impairment models, be it sub-chronic PCP, early life stress, lesions, or transgenic animals, through the use of multiple acquisition and reversal phases. When combined with other tasks currently available for use within touch-screen equipped operant boxes, the DVD procedure should make an important contribution to the characterization of disease models and accurate interpretation of data.

Acknowledgements

The work leading to these results has received funding from the Innovative Medicines Initiative Joint Undertaking (IMI JU) under Grant Agreement N° 115008.

References

1. Bussey, T.J., Muir, J.L., Everitt, B.J., Robbins, T.W. (1997). Triple dissociation of anterior cingulate, posterior cingulate, and medial frontal cortices on visual discrimination tasks using a touchscreen testing procedure for the rat. *Behav Neurosci* **111**, 920-936.

2. Bussey, T.J., Padain, T.L., Skillings, E.A., Winters, B.D., Morton, A.J., Saksida, L.M. (2008). The touchscreen cognitive testing method for rodents: how to get the best out of your rat. *Learn Mem* **15**, 516-523.

3. Chudasama, Y., Robbins, T.W. (2003). Dissociable contributions of the orbitofrontal and infralimbic cortex to pavlovian autoshaping and discrimination reversal learning: further evidence for the functional heterogeneity of the rodent frontal cortex. *J Neurosci* **23**, 8771-8780.

Assessment of Behavioural Flexibility and Executive Function Using Novel Touch Screen Paradigms

A.C. Mar[1], J. Alsiö[1], A. Haddenhorst[1], C.U. Wallis[1], A. Trecker[2], L.M. Saksida[1], T.J. Bussey[1], T.W. Robbins[1]

[1]*Department of Experimental Psychology, Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, UK. am682@cam.ac.uk*

[2]*Heinrich-Heine Universitaet Duesseldorf, Duesseldorf, Germany*

The Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) initiative has recently identified a NIMH consensus cognitive battery comprising seven major domains for use in clinical trials assessing cognitive impairment associated with schizophrenia (CIAS). Major cognitive domains identified include Attention/Vigilance (assessed using variants of the continuous performance test), Visual Learning and Memory (e.g., visuospatial memory tests), Working Memory (e.g., spatial span or delayed response tasks), Speed of Processing (e.g., trail-making) and Reasoning and Problem-Solving. There is a great need for a biologically and statistically-validated consensus cognitive test battery for rodents which can reliably translate to these CIAS clinical endpoints. Here we report on the development of three novel touchscreen-based tasks for the rodent – a 3-Stimulus Discrimination and Reversal test, a Continuous Performance test, and a Self-Ordered Working Memory test – that tap into several of the CIAS domains, and have responses requirements and outcome measures that are highly related or identical to those used in clinical assessment.

3-Stimulus Discrimination and Reversal Learning

Recent evidence suggests that schizophrenia patients are unimpaired in simple rule learning but appear inflexible when rules change, possibly reflecting difficulties in using negative feedback information and in switching attention [1]. Reversal learning is thus a promising target for translational studies because it is simple, specific, clinically-relevant, and putatively targets orbitofrontal dysfunction.

We have recently developed a 3-stimulus discrimination and reversal learning paradigm as a tool for measuring rule learning and flexibility for visual stimuli in rats. The basic task uses 3 stimuli randomly presented across 3 screen locations on the touchscreen. Data will be presented on two task variants in which, in addition to the S+ and S- included in standard 2-choice reversal learning paradigms, the third stimulus is included as an additional S-, or as a neutral distracting stimulus (So). The 3-stimulus paradigm was devised in effort to obtain 1) better discernment of stimulus perseveration versus avoidance through analysis of choices of the unchanged third stimulus (S- or So), 2) lower reinforcement of spatial biases (e.g., reinforced 33% versus 50%) 3) less possibility of solving the task through configural or other forms of learning (triplets vs pairs). The relative merits and difficulties of the 3-stimulus paradigm will be discussed in relation to comparable data acquired using a standard 2-stimulus procedure and with respect to its ability to detect effects within animal models of schizophrenia and in response to pharmacological agents.

Continuous Performance Test

The continuous performance test (CPT) is a paradigm comprising several procedural variants, some of which have proven reliable indicators for the presence of, and susceptibility to, schizophrenia [2,3]. In its basic form, the CPT involves rapid, externally-paced (e.g., not under subject control) presentation of a large number of consecutive stimuli, to which subjects are instructed to respond only upon detection of a designated target stimulus or sequence. Performance on the CPT is traditionally assessed by analyzing numbers of correct responses ("hits" or, inversely, omissions or "misses") and incorrect responses ("commission errors" or "false alarms"). Signal detection analysis can also be used to combine these traditional measures and extract d' (ability to discriminate signal from background) and B (bias toward responding). The CPT is generally considered primarily as a measure of vigilance or sustained attention, with performance variables typically worsening as the session proceeds. However, CPT performance is also dependent on task parameters and may be designed to be more or less sensitive to elements of sustained attention, selective attention, impulsivity or executive control.

We report here on our development of an analogue of the CPT using the rat touchscreen. We trained rats to attend to a central location on the screen and presented stimuli every 3-7s (vITI) continuously throughout the task (the only pauses occurring after a correct response). Stimuli are presented for 1.5s stimulus duration and rats have 2s to respond to the stimulus after stimulus onset (limited hold). Initial probabilities are 50:50 for target (in example X) and nontarget (O,H,I) stimuli, where is never possible to have more than 4 targets or nontargets in a row (except correction trials). The stimulus presentation and response requirements were thus highly similar to the human version of the CPT. The basic version of the CPT task was acquired within 20 sessions. Task validation procedures confirmed detrimental effects on d' sensitivity measures using variable event rate, reduced stimulus duration and reduced stimulus contrast. Moreover, rats showed performance decrements toward the latter trials of the session suggestive of taxing vigilance processes. The results will be further discussed in relation to data collected on animal models of schizophrenia and in response to putative cognitive enhancing agents.

Self-Ordered Working Memory Task

Working memory dysfunction is a key aspect of cognitive impairment in schizophrenia where impairments have been reliably documented in schizophrenic patients, as well as back-translated to studies of primate and rodent models of schizophrenia. Evidence exists linking dopaminergic systems that may underlie schizophrenia-like working memory impairments in rhesus monkeys, with similar results found in rodents [4]. Moreover, visuospatial working memory (SWM) has been identified as a putative cognitive endophenotype of schizophrenia [5].

We have developed a self-ordered working memory task for rats in the touchscreen environment. In the basic version of the task, rats were trained on each trial to touch 3 spatially distinct squares on the touchscreen, receiving rewards for each new square touched. If a square was revisited within a trial, a negative feedback image was displayed and houselight illuminated. A trial was completed after all 3 squares had been touched once. A variable inter trial interval (25-45s) separated each trial. Rats were given up to 60 trials in a 60 min session. Rats were observed to progressively improved in the task, making fewer errors per trial than chance, and completing more than 42% perfect trials per session (chance performance = 22.2%). Probe tests to examine sequencing behaviour revealed that, despite some evidence for sequencing, rats continued to perform considerably better than chance when sequencing was discouraged (different forced starting squares). This suggests that the task may be reflective of spatial working memory. Further behavioural and pharmacological probe tests will be presented to illuminate putative underlying cognitive processes.

In summary, we have recently developed 3 novel cognitive tests in the rodent touchscreen apparatus (3-Stimulus Reversal; rodent CPT and rodent SOWM) that are highly similar to corresponding clinical paradigms (CANTAB visual discrimination; X-CPT and CANTAB spatial working memory tests) in terms of the stimuli, responses and outcome measures. Evidence suggests that these tasks can lead to reliable levels of performance (3-stimulus reversal has already been replicated across laboratories) that are sensitive to parametric behavioural manipulations that alter cognitive load. These tests are currently being validated in terms of their underlying neural substrates (e.g., medial and/or orbital prefrontal cortex lesions) and with respect to their sensitivity to detection of cognitive enhancing agents. These novel touchscreen tasks are highly promising for use in flexible translational battery with relevance to the examination of cognitive processes in schizophrenia.

Acknowledgements

The research leading to these results has received support from the Innovative Medicine Initiative Joint Undertaking under grant agreement n° 115008 of which resources are composed of EFPIA in-kind contribution and financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013).

References

1. Leeson, V.C., Robbins, T.W., Matheson, E., Hutton, S.B., Ron, M.A., Barnes, T.R., Joyce, E.M. (2009). Discrimination learning, reversal, and set-shifting in first-episode schizophrenia: stability over six years and specific associations with medication type and disorganization syndrome. *Biol. Psychiatry* **66**, 586-593.

2. Cornblatt, B.A., Keilp, J.G. (1994). Impaired attention, genetics, and the pathophysiology of schizophrenia. *Schizophr. Bull.* **20**, 31-46.

3. Chen, W.J., Faraone, S.V. (2000). Sustained attention deficits as markers of genetic susceptibility to schizophrenia. *Am. J. Med. Genet.* **97**, 52-57.

4. Floresco, S.B., Phillips, A.G. (2001). Delay-dependent modulation of memory retrieval by infusion of a dopamine D1 agonist into the rat medial prefrontal cortex. *Behav. Neurosci.* **115**, 934-939.

5. Saperstein, A.M., Fuller, R.L., Avila, M.T., Adami, H., McMahon, R.P., Thaker, G.K., Gold, J.M. (2006). Spatial working memory as a cognitive endophenotype of schizophrenia: assessing risk for pathophysiological dysfunction. *Schizophr. Bull.* **32**, 498-506.

SPECIAL SESSION

Technical Support for Analysis of Human Error in Task Performance

Chair: Ronald Poppe[1], Tobias Heffelaar[2], Jordi Bieger[3]

[1]*Human Media Interaction, University of Twente, Enschede, The Netherlands. R.W.Poppe@utwente.nl*

[2]*Noldus Information Technology, Wageningen, The Netherlands.*

[3]*VicarVision, Amsterdam, The Netherlands*

Safety, efficiency and worker well-being often depend on correct task performance. Yet, the structured or even repetitive nature of these tasks can lead to a worker's loss of attention, fatigue, stress or other mental states that negatively affect the performance of the work. Such states are more likely to cause the introduction of human errors, with associated risk.

By bringing sensor technology into the work place, human behavior can be analyzed and interpreted automatically. Cameras can observe human posture, head movement and gestures. From these, the task at hand can be analyzed. Eye trackers can determine the worker's focus of attention. Microphones can pick up subtleties in the worker's voice. From these measurements, deviations from normal task flow or worker's alertness can be identified at an early stage, and might trigger alarms or start check or safety procedures.

As the sensing and interpretation technologies are currently at the point of maturing, there is a growing interest to apply these techniques at the workplace. This requires the development of integrated solutions of automatic human error assessment and prevention that combine sensor technology and robust interpretation with knowledge of the task and the user. These solutions are starting to find their place in automation industry, call centers and assembly lines. Also, desktop-related working environments can benefit from such solutions. In addition, there is a growing demand for these techniques to be applied in training and simulation (e.g. driving simulators), and in consumer products (e.g. ATMs).

In this special session, we aim at bringing together an audience from both academia and industry. We highlight recent advances from a technical point of view, and requirements from an industry perspective. In our discussion, we focus on current challenges and collaborations and ways to address these.

SPECIAL SESSION CONTENTS (sorted by paper ID)

Eliciting Control Errors and Measuring Error Correlates

Michael Lindenthal (University of Applied Sciences Münster, Germany)

Watching People Making Errors: Vision Architectures for Monitoring Task Performance

Marten den Uyl (VicarVision, The Netherlands)

Measuring Electrodermal Activity of Both Individuals With Severe Mental Disabilities and Their Caretakers During Episodes of Challenging Behavior

Matthijs Noordzij[1], Patrick Scholten[2], Marleen Laroy-Noordzij[2]

[1] *University of Twente, Enschede, The Netherlands. m.l.noordzij@utwente.nl*

[2] *De Twentse Zorgcentra, Enschede, The Netherlands. patrick.scolten@detwentsezorgcentra.nl; marleen.laroy@detwentsezorgcentra.nl*

Abstract

Here, we present evidence for the feasibility of measuring changes in electrodermal activity in both individuals with severe mental disabilities and their caretakers during challenging (aggressive) behavior. In addition, we will provide guidelines on how to realize these measurements while minimizing distress for individuals with severe mental disabilities. These measurements will both provide insight into the physiology of intense, real life emotions and a better understanding of emotional fluctuations from those who cannot readily express feelings in a verbal manner.

Author keywords. Ambulatory measurement, Skin Conductance, Aggression, Ideographic studies, Safety, Mentally Handicapped Persons; Institutionalized Persons, Emotions, Arousal.

ACM classification keywords. J.3 Life and medical sciences: Health, Medical information systems, J.4 Social and behavioral sciences: Psychology.

Introduction

The inspection of health care [1] estimates that in The Netherlands about 20% of the 30.000 people with an intellectual disability (ID) who live in a residential setting show "severe behavioral problems" in which aggression and self-injurious behavior (SIB) are the most predominant. This challenging behavior (CB) is a serious problem in the daily care of people with a mental disability. At present, caretakers try to anticipate CB by relying on their experience on what mix of visual and/ or verbal cues from the client (in a particular context) might be indicative of upcoming aggression. Caretakers often report being surprised by CB. An important marker for upcoming aggression might be found in changes in their physiology (e.g. heart rate (variability), electrodermal activity (EDA)). Of course, changes in these markers are not readily perceived by caretakers. Interestingly, these physiological changes are relatively easy to measure, and these types of measurements are becoming extremely "wearable" and non-intrusive. Because of these developments it is now seems feasible to continuously measure changes in the physiology of clients in (almost) any situation. Caretakers express a great need for a better understanding of the emotional fluctuations of their clients. The ambulatory measurement of EDA now finally gives the opportunity to provide caretakers with additional, objectively measured information regarding the inner states of their clients.

Another important factor is that actions of direct care staff are thought to be antecedents as well as consequences of a large proportion of CB such as aggression [2]. Their actions may increase or decrease the likelihood of future problem behaviors. A low rate of adequate staff responding and even counterproductive responding have been reported [3]. The possibility of the occurrence of CB can be perceived as a stressor in working with clients who have a history of showing CB. This direct stressor can produce a psychological experience, a feeling of frustration or arousal within the caretaker without the actual presence of CB. This heightened level of arousal in the caretaker can further negatively influence the task performance through attention narrowing [4]. Care takers can display the tendency to restrict the range of attention and to ignore surrounding information sources in the context of high (physiological) arousal. Therefore, we measured physiological changes in both clients and caretakers.

Measuring electrodermal activity

For well over 100 years researchers have been measuring electrodermal signals. This has resulted in quite strict standardization regarding both terminology and measuring methods [5]. Typically EDA is recorded as skin conductance (SC, in µSiemens) by applying a direct current (with two silver electrodes) to the skin (exosomatic method). Central to this measure is the electrodermal response (EDR). The EDR constitutes a sharp rise in the SC value, followed by a slower drop in conductance. For example, a sudden loud burst of noise will result an EDR 1-2 seconds later, and this is easily visible in the raw data signal. In general, changes in SC are closely linked to activity of the sympathetic part of the autonomic nervous system (ANS) (i.e. "the fight or flight" system). Therefore, researchers and practitioners have taken EDA measurements as further operationalization for constructs such as attention, stress, anxiety, workload, pain, and arousal.

Developments in electronic devices have made it possible to measure EDA outside the laboratory, in ambulatory fashion. Very recently a shift in measurement location and the application of dry electrodes has further broadened the scope of EDA measurements to true observational field studies [6]. Our sensor, the Affectiva Q sensor, is a wrist worn, wireless sensor (electrodes are placed on the ventral side of the arm) that can easily measure EDA for 24 hours with a sampling frequency of 32 Hz. This measurement location does not adhere to current standards [5]. On the one hand, it has been shown to produce comparable results as the advised locations [7]. On the other hand, the number of sweat glands differ hugely across the body, and it seems to be the case that sensitivity for detecting EDR's can be considered lower at the wrist than at other locations (e.g. the palm of the hand), especially for stimuli that are only weakly arousing [8].

EDA during challenging behavior

For nine (fixed) couples of clients and caretakers we measured EDA (with the Q sensor) during 24 sessions of approximately two to three hours, which were all recorded on video. EDA measurements were successful in most sessions (many with CB) for both clients and caretakers in clearly showing fluctuations in EDA level and displaying the prototypical EDR's, albeit with far higher amplitudes than what is typically reported in lab studies [5]. This data, and the systematic comparison between physiological changes and behavior, will be reported elsewhere.

Figure 1 shows the data of a client from one session in which both CB (aggressive behavior, 1A) and strongly, positive emotional behavior (while playing music, 1B) were displayed. In the details on the right of this Figure the (superimposed) EDR's are clearly visible. Standard ways of classifying the reactivity of the sympathetic part of the ANS consist of computing the number of EDR's per minute and the mean amplitude of these responses. In addition, the whole level of EDA can rise (i.e. the space under the EDA line), which can be clearly seen on several occasions on the left side of Figure 1. It is important to note that the reactivity of the sympathetic ANS can rise dramatically during both negative and positive emotional episodes. As such the EDA is not particularly sensitive to the valence of an emotion, and caution should be applied when EDA is said to be used to uniquely measure a particular negative experience (such as pain, see [9]).

Guidelines

To measure EDA and heart rate variability (not further discussed here) both the client and the caretaker need to wear a Q sensor wrist band and a Polar chest band and receiver watch. In the interaction with individuals with a severe mental disability (client), who are known to show challenging behavior, every added question or demand to the normal daily routine can be seen as a stressor and therefore can be viewed as an act to cause challenging behavior. For the caretaker this added task can also bring stress to the relationship. Putting on and taking off these materials can be viewed as added tasks that can disrupt the life of the client and the relationship between the client and the caretaker. Below we specify our methods and guidelines to minimize these levels of distress.

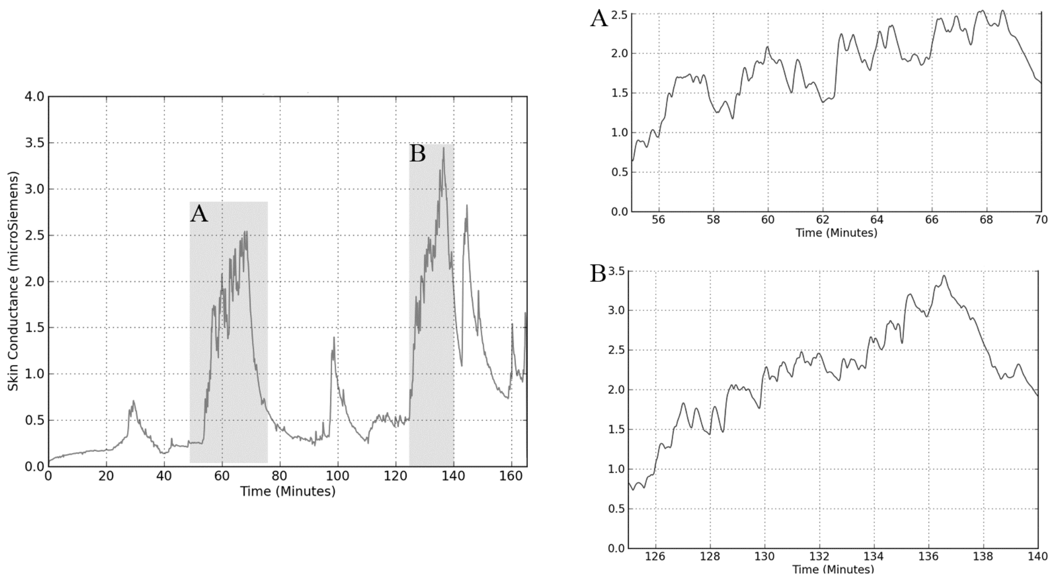


Figure 1. An example of one session in which the client displayed strongly negative challenging behavior (marked with an A in the overview figure and in the detail on the right), and strongly positive behavior associated with making music (marked with a B). Time (in minutes) is on the x-axes, and SC (in µSiemens) is on the y-axis. Data were low-pass filtered (2nd order Butterworth filter with a cut-off frequency of .002 Hz.).

Measurement materials are presented and removed by the caretaker. In this way the client does not need to build a new demanding relationship with unknown researchers. Caretakers are individually trained by the researchers in executing a measurement material applying (MMA) protocol during a two hour training session. In this training session the MMA protocol is personalized for the specific client-caretaker couple. For each caretaker an individual MMA protocol instruction card was made.

Central to the MMA protocol is maximizing the feeling of influence on the situation for the client. By making the situation predictable (in announcement and repeatability) and maximizing the level of participation the client can follow, recognize and understand the actions needed to put on and take off the measurement materials. Personalization is accomplished in two ways.

1. *Announcing the start and finish of the task to the client at his own determined level of communication*. For every client the level of communication is determined on the basis of the ComVoor [10]. A client can function on sensation, representation or presentation level. On the level of sensation, predictability is hard to accomplish. At best, recognition and familiarity by can be reached by repetition. Time is taken to let the client get used to the feeling of actual materials on their skin. Executing the MMA protocol in a consistent manner is needed and practiced by the caretaker in the training. At the level of representation, predictability in a short window of time can be reached. The client is presented with the blue box or the picture of the blue box in which the materials are stored just before putting on or taking off. With every time the blue box is introduced, the actual introduction of the measurement materials will be a bit later in time. Through this principle of backward chaining the client can form a link between presentation of the blue box and the action of putting on or taking off the measurement materials. At the level of presentation the picture of the blue box is part of the timeline that is shown to predict the upcoming day or part of the day. Covering the picture with a red cross indicates the end of the measurements.

2. *Determining the clients optimal level of participation in applying the materials*. We added caretaker help from least to most. “Least to most help” is defined as a continuum from verbal instruction, modeling (“showing how” on yourself or a demonstration doll), applying materials together, to making the client aware of total takeover by the caretaker. The caretaker is instructed to wait five seconds before adding the next level of help.

After each session the caretaker was interviewed by the researches to determine the level of participation and the level of intrusiveness of the materials employed. The level of participation by the clients was surprisingly high. In five cases verbal instruction was combined with modeling and physical assistance at the start of the research. In four cases the caretaker totally took over. During the course of the research for all but one client the level of participation increased.

During the training additional attention was paid to the several ways clients with a severe mental disability could show resistance [11]. For each client possible behaviors that could indicate resistance were determined. If a possible form of resistance was noted the researcher organized a multi-disciplinary meeting in which the behavior was evaluated by his/ her physician, legal representative and treatment coordinator. Depending on this multi-level interpretation of the behavior, participation in the research was continued or not. During the 24 x 9 session this ‘resistance protocol’ was only set in motion once. The behavior was interpreted as not specifically related to resisting to wearing the measurements materials. Participation in the research was continued.

Conclusions

EDA measurements are now possible under the most challenging circumstances. Here, we show how it is possible to measure EDA during CB of severely mentally disabled people and their caretakers. One of the future possibilities of an application based on this information is to assist caretakers in the recognition of CB by means of a technological aid that is based on online variations of the clients physiology. This signal could perhaps simply help caretakers to direct focused attention to this client and establish whether action is required to prevent escalation and thereby increase the safety and health of both the client and the caretaker. In addition, caretakers could also be informed about changes in their own physiology and possibly shortly refrain from interaction with the client, or actively engage in relaxation techniques, when their arousal seems to high. These measurements will both provide insight into the physiology of intense, real life emotions (see also [12] for similar attempts with a psychiatric patient) and the outlook of getting a better understanding of emotional fluctuations of those who cannot readily express feelings in a verbal manner [6].

Ethical statement

This study was approved by the local Medical Ethical Committee, MST hospital, Enschede (METC no. P11-27 NL 37314.044.11, approved on 06-09-2011). This study was also registered in the Netherlands Trial Register (Trial Code 3043).

Acknowledgements

This study has been supported by “De Twentse Zorgcentra”.

References

1. IGZ (2005). Inspectie voor gezondheidszorg: Complexe gedragsproblematiek bij mensen met een ernstige verstandelijke handicap vereist bundeling van specialistische expertise. Den Haag, Nederland.

2. Embregts, P. J. C. M., Didden, R., Huitink, C. and Schreuder, N. (2009). Contextual variables affecting aggressive behaviour in individuals with mild to borderline intellectual disabilities who live in a residential facility. *Journal of Intellectual Disability Research* **53**, 255-264.

3. Embregts, P. J. C. M. (2002). Effect of resident and direct-care staff training on responding during social interactions. *Research in Developmental Disabilities* **23**, 353-366.

4. Warm, J. S., Parasuraman, R. and Matthews, G. (2008). Vigilance requires hard mental work and is stressful. Human Factors: *The Journal of the Human Factors and Ergonomics Society* **50**, 433-441.

5. Boucsein, W. (2012). *Electrodermal Activity*. Springer, New York, NY, USA.

6. Picard, R. W. (2009). Future affective technology for autism and emotion communication. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**, 3575-3584.

7. Poh, M.-Z., Swenson, N. C. and Picard, R. W. A. (2010). Wearable Sensor for Unobtrusive, Long-Term Assessment of Electrodermal Activity. *IEEE Transactions on Biomedical Engineering* **57**, 1243-1252.

8. Van Dooren, M., de Vries, J. J. G. and Janssen, J. H. (2012). Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & Behavior* **106**, 298-304.

9. Eriksson, M., Storm, H., Fremming, A. and Schollin, J. (2008). Skin conductance compared to a combined behavioural and physiological pain measure in newborn infants. *Acta Pædiatrica* **97**, 27-30.

10. Noens, I., Van Berckelaer-Onnes, I., Verpoorten, R. and Van Duijn, G. (2006). The ComFor: an instrument for the indication of augmentative communication in people with autism and intellectual disability. *Journal of Intellectual Disability Research* **50**, 621-632.

11. NVAZ (1999). Gedragscode voor artsen bij beoordeling van verzet bij mensen met een verstandelijke handicap. *Tijdschrift voor de Vereniging van Artsen in de Zorg voor Mensen met een Verstandelijke Handicap* **17**, 1-3.

12. Kuijpers, E., Nijman, H., Bongers, I. M. B., Lubberding, M. and Ouwerkerk, M. (2012). Can mobile skin conductance assessments be helpful in signalling imminent inpatient aggression? *Acta Neuropsychiatrica* **24**, 56-59.

Generic Tool for Online Classification of Physical and Mental Workload

Christoph Hintermüller[1], Günther Edlinger, Christoph Guger

Guger Technologies OG, Schiedlberg, Austria, [1]hintermueller@gtec.at

Introduction

Physiological signals like the heart rate or the galvanic skin response are used to assess the mental [1] and physical stress of a subject during specific situations and tasks. Similar experiments are used to investigate and recognize emotions [2] and related emotional states of the subject in response to images [3]. Recent developments in the field of human machine [4, 5] and robot interaction require the online measurement and classification of physiological signals and features with respect to emotions and low level intentions in real-time. In the following sections the g.PhysioObserver system and its use for online real-time studies will be presented. The results of an initial test experiment on the detection of the mental and physical loads occurring during four different exercises and their discrimination based on corresponding physiological features and signals will be shown.

Methods

For the evaluation and online classification the g.PhysioObserver™ (Guger Technologies OG, Austria) system, shown in Figure 1, is used, which represents a MATLAB™/Simulink™ (Mathworks Inc, USA) based rapid prototyping system for the simple and thereby flexible design and implementation of study paradigms. The platform allows to record biosignals like the ECG, EEG, respiration (Resp) and galvanic skin response (GSR) signal extract many different physiological parameters and features thereof. It is possible to form application specific feature vectors and classify them with respect to the mental and physical load associated with a specific

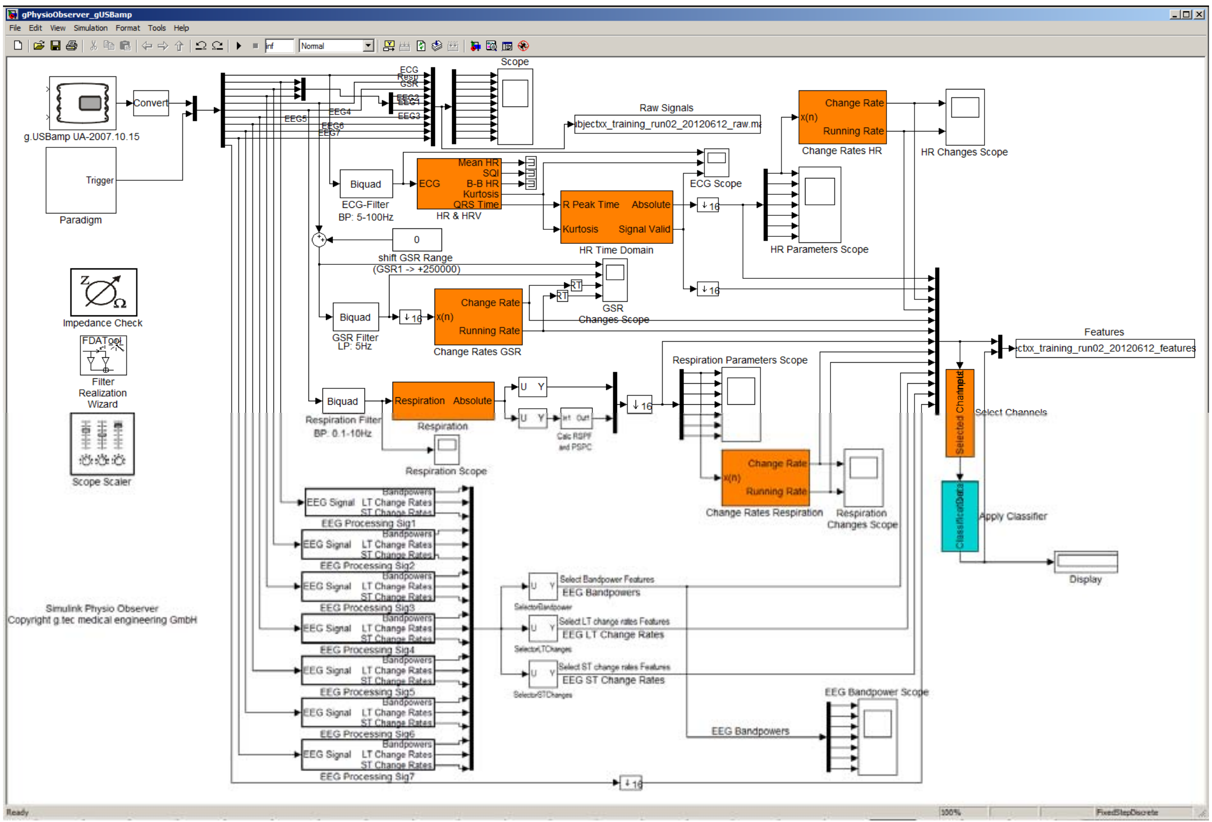


Figure 1. The g.PhysioObserver system allows to assess and evaluate different mental and physical sates of a subject online in real-time based on physiological signals and features extracted thereof.

exercise, mental or emotional state.

The system currently extracts from the ECG the Heart Rate, Heart Rate Variability, Root Mean Square of Sample Differences (RMSSD) or the RR interval, the percentage of consecutive RR intervals which vary by more than 50ms (pNN50) and the HR/RRMSD ratio. From the Resp signal it extracts the respiration rate (RespR), the duration and depth of the inspiration and expiration phases and a signal corresponding to the beaks during inspiration and expiration. The latter is indicated by a saddle point in the Resp signal exposing a zero value in its first and second derivative. The so called respiration stop signal is converted into a signal corresponding to the average number of breaks per minute and the number of breaks per respiration cycle. For any of the signals and Features it is possible to compute long term and short term changes rates. The long term rate indicates the overall change of the signals compared to an initial baseline segment at the beginning of the experiment. The short-term rate captures relative changes between the current evaluation period and its preceding period. Further it is possible to include measures like the power in the alpha-band, beta-band and theta band from up to 7 EEG channels.

An experiment consists of at least two phases, a training run and an online real-time classification and feedback run. During the training run all data required for computing a classifier are collected. This classifier is used during the online classification run to identify the current mental or physical state of the subject or to identify the exercise executed by the subject. Both runs can fully be controlled through two text files defining the corresponding paradigm. The first file defines the exercises the user has to execute along with the list of tasks to be accomplished for each exercise. Alternatively an external HTML page may be specified defining and controlling the tasks of the experiment. The second file determines the total number of trials and assigns each of them to one single exercise. The duration of one single trial, the amount of time the subject should be engaged in each exercise and the duration of breaks between distinct exercises can be freely defined through experimental parameters. The extracted feature signals are sampled at 16 Hz and stored independently from the raw biosignals which are acquired with a sampling frequency of 256 Hz. Thereby it is possible to keep the amount of data stored for each experiment at reasonably low levels, without losing any relevant information.

Experimental setup

The described experiment tries to identify the physical and mental load of the subject based on the ECG, GSR, Resp and 7 EEG signals only. The ECG signal is based on the bipolar derivation between the right shoulder and the Wilson lead V5. In the discussed preliminary result a thermistor based air flow sensor (SleepSens, S.L.P. Inc , USA) was used to record the respiration signal. Alternatively a piezzo based respiration belt sensor can be used. The EEG signals were recorded from the FCz, Cz, P3, Pz, P4, PO7 and PO8 positions of the international 10-20 system. For acquiring the GSR single two metal electrodes of the g.GSRsensor2™ were mounted on the medial phalanx of the subjects left index and middle finger.

For both, the recording of the EEG signals and the ECG signals active electrodes connected to a g.GAMMAbox were used to reduce undesired artefacts and to improve signal quality. The common ground electrode was placed on the AFz position of the international 10-20 system and the common reference electrode was attached to the left ear lobe. The initial experiments consisted of four distinct exercises of 5 Minutes duration each. In the first exercise (REST) the subject had to sit still and relax. During the second exercise (D2-Test) the subject had to perform a D2-Attention test. For this test the subject has to mark all symbols which show a d character annotated by exactly 2 short lines. During the remaining two exercises, SLEEP and SPORT, the subject had to lay down and sleep and bend the knees respectively. Each of the exercises was split into 5 consecutive trials of one minute duration.

Results

The data recorded during the training run was used to compute a Linear Discriminant Analysis (LDA) based classifier for discriminating the REST, exercise, from D2-TEST, SLEEP and SPORT. The feature signals used to generate the LDA classifier were selected by applying the distinction sensitive learning vector quantization

(DSLVQ) feature weighting method to a set of 192 trials recorded for 4 individual subjects. The resulting set of features included the minimum, mean and maximum HR, the inhalation depth and stops per cycle the long- and short-term change rates of GSR the power in the α-band of the EEG signals recorded from the positions P3, Pz and P4 and the power in the b-band of the positions P4, PO7and PO8. The resulting classifier was used in an online real-time evaluation run to identify, which exercise the subject was performing. In this run the order and duration of each exercise to be accomplished predefined by the paradigm equal to the training run.

Figure 2 depicts the classification errors estimated by the LDA for each time point within all trials of one specific class. The plots indicate the classification error for all four different exercises, (REST class 1, D2-Test class 2, SLEEP class 3 and SPORT class 4) individually and the average overall error achieved. In this first experiment an error of 0% could be achieved at any time point of each of the exercises. The resulting classifier was than tested during the online validation run. In this run each sample consisting of an individual set of values for the selected features was analyzed and assigned to one of the exercise. In case the probability that a specific sample belonged to the class selected by the classifier was below 95 % it was assigned to the virtual zero class instead. Figure 3 shows the average classification error (blue curve) over all error achieved for all four exercises. This classification error was computed by comparing the class assignment of each individual sample recorded during the online evaluation run with the class expected for the trial the sample was recorded for. In the depicted experiment an average classification error of 5 % was achieved through out all experiments. In addition to the classification error the rate of false positive classifications which was approximately 5 % is shown.

Discussion

The relatively high rate of false positive classifications compared to the average classification error may be explained by the low number of only 5 trials per class used to compute the classifier. Thereby it was necessary to use the same data for computing and testing the classifier which may be over fitted to the training data. The results obtained during the initial tests show that the presented system allows to extract a large variety of physiological features and use them to discriminate the mental and physical loads of user defined exercises.

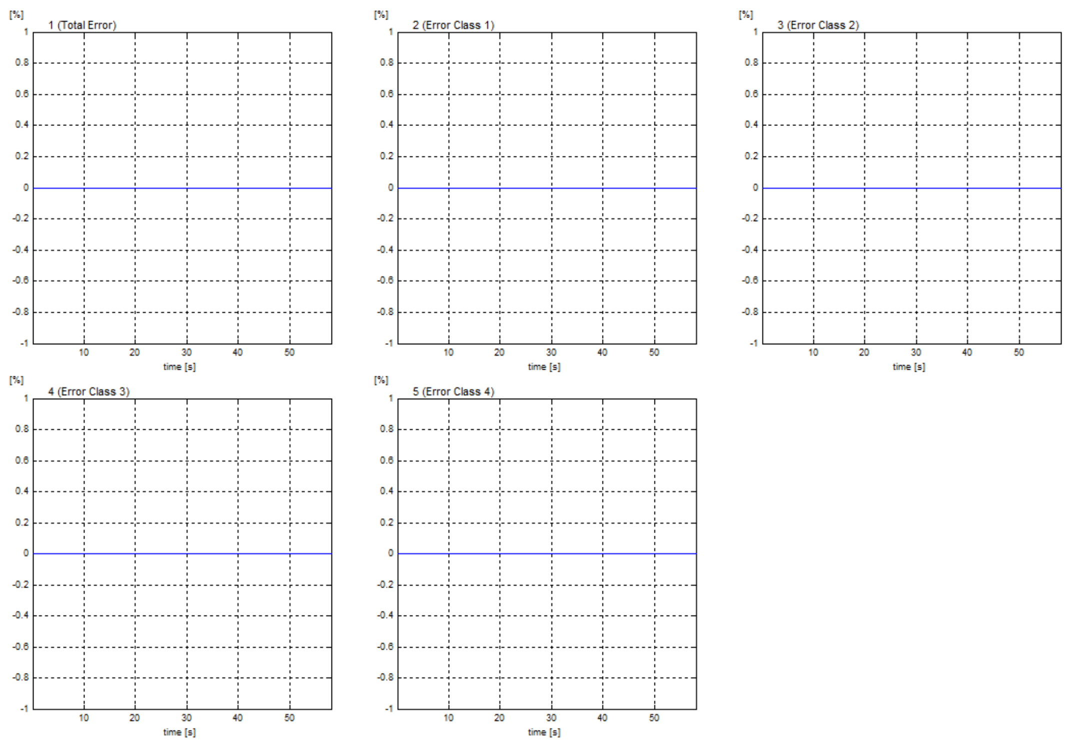


Figure 2. Classification error for discriminating the four exercises Rest, D2-Test, Sleep and Sports. The break between the exercises was set to 90s.

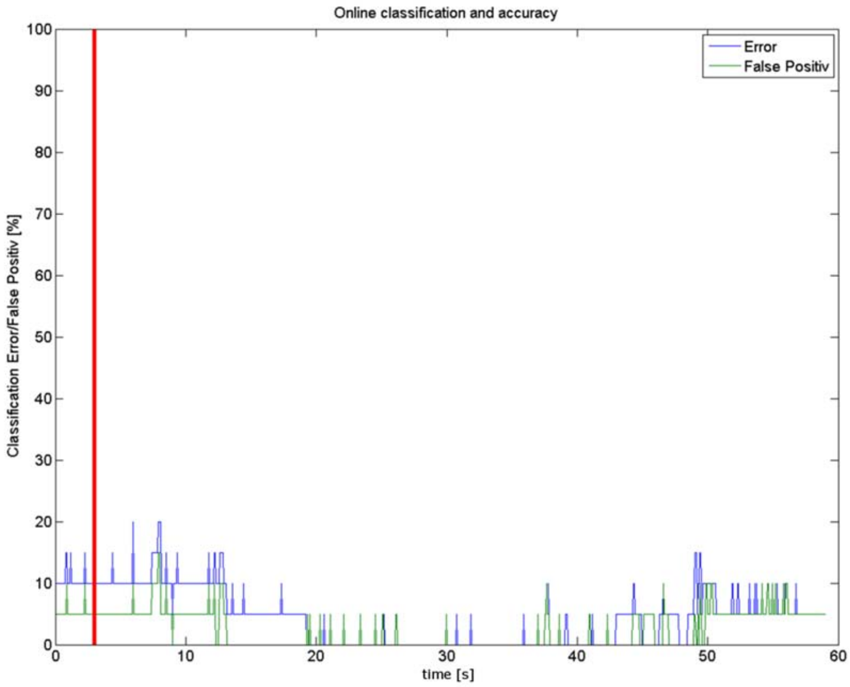


Figure 3. Online classification error and false positive rate achieved by the subject during the feedback session.

Acknowledgements

This research is supported by the European Union FP7 Integrated Project VERE (No. 257695).

References

1. Insulander, P. Vallin, H. (2005). Gender differences in electrophysiologic effects of mental stress and autonomic tone inhibition: a study in health individuals. *J. Cardiovasc. Electrophysiology* **16**(1), 59-63.

2. Scherer, K. (2007). What are emotions? And how can they be measured. *Social Science Information* **44**(4), 695-720.

3. Park B., Jang, E., Kim, S., Huh, C., Sohn, J. (2012). Identification of Optimal Emotion Classifier with Feature Selections Using Physiological Signals, *AHCI 2012*, 224-229.

4. Wong, M. (2006). Emotion Assessment in Evaluation of Affective Interfaces, *Neuron* **65**(3), 293.

5. Pantic M., Sebe, N., Cohn, J.F., Huang, T. (2005). Affective multimodal human-computer interaction. In *Proceedings of the 13th annual ACM international conference on Multimedia*, 669-676.

The Neural Origins and Applications of Human Error Processing

Tsvetomira Tsoneva, Gary Garcia-Molina

Philips Research, Eindhoven, The Netherlands
tsvetomira.tsoneva@philips.com, gary.garcia@philips.com

Introduction

The human brain has developed complex neural and cognitive mechanisms that deal with errors. Error awareness and processing is mediated by the supervisory cognitive control system. This system is concerned with the effortful activation and allocation of cognitive resources in the selection and processing of task-relevant information for the purposes of maximizing performance on tasks involving high difficulty, complexity, response conflict, or novelty [1]. Two components of control are distinguished: the *regulative control* is responsible for activation and implementation of control related processes; whereas the *evaluative control* deals with monitoring for the presence of errors and/or response conflicts and if needed with taking compensatory actions. Evaluative control is associated with the anterior cingulate cortex (ACC), which is also responsible for regulating emotional responses, while the regulative control is associated with various sites in the prefrontal cortex (PFC).

Objective

The objectives of this talk are: 1) to familiarize the audience with the brain mechanisms dealing with awareness of erroneous responses, 2) to explore options for measuring physiological correlates of such brain mechanisms, and 3) to present practical applications that can benefit from these measurements.

Brain responses to errors

Brain activity produces electromagnetic signals which result from the synaptic current flowing between the neurons in the brain cortex. These signals can be measured non-invasively along the scalp using an appropriate sensing technology, such as the electroencephalogram (EEG). EEG can be very useful in studying the functional role and the interaction between the different systems involved in various mental states and processes. *Event-related potentials* (ERP) are patterned voltage changes embedded in the ongoing EEG that reflect a process in response to a particular event.

The brain responses to errors can be also quantified using this technique. The electrophysiological reflection of the evaluative cognitive control function is called Error-related potential (ErrP). Depending of the source and the brain's awareness of the error different types of ErrPs can be distinguished. ErrP emerging shortly after an error made by a human subject, usually in a choice reaction task, requiring a quick response to a stimulus, is called error-related negativity (ERN) or response ErrP. Another type of ErrP, the feedback ErrP, appears shortly after feedback indicating an erroneous response from the subject, in e.g. a reinforcement learning task. Whereas these neural correlates of error awareness are manifested after errors committed by the subjects themselves, ErrPs are also present after an observation of an error, for example committed by the interface the subject is interacting with. These are known as interaction ErrPs.

Possible applications

Understanding the origins of the different types of error potentials, their characteristic features and the factors affecting them would allow for their application in various contexts including: ensuring higher levels of safety, performance, and user experience enhancement. Interaction ErrPs are already used in brain-computer interface (BCI) systems to adapt the pattern recognition algorithms and increase the classification accuracy. Other systems relying on automated pattern recognition, such as speech recognition, face detection, etc., can also benefit from similar approach. Detecting and preventing human errors when time-critical decision making is required, for professionals such as pilots, air-traffic controllers, surgeons, or drivers, can significantly improve safety. Error recognition can also help in optimal behavior learning by decreasing the likelihood of repeating erroneous decisions in the same context (biofeedback).

References

1. Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., Cohen, J.D. (2001). Conflict monitoring and cognitive control. *Psychol Rev*. **108**, 624-652.

Dealing with False Alarms in Camera Surveillance

F.L. Kooi, W.D. Ledegang

TNO Behavioural & Social Sciences, Soesterberg. The Netherlands
frank.kooi@tno.nl, wietse.ledegang@tno.nl

With the advancement of automated detection, CCTV operators run the risk of being swamped by false alarms. In order to avoid distraction, repeated false alarms may cause operators to start to ignore all alarms, knowing that each time it will most likely be false. In anticipation we have developed a camera surveillance interface with the purpose to handle substantial false alarm rates. The interface is designed to reduce task switches and minimize operator workload when false-alarms occur regularly. This challenge requires a set of audio-visual design measures. The 'false-alarm robust' interface may be viewed as supplementary to the other measures of human behavior described in the adjoining presentations. The prototype interface contains three approaches to help deal with false-alarms.

Sources of false-alarms

Besides originating in faulty hardware, false-alarms may have a variety of origins such as:

- Software false-alarms: acoustic event (sound) detection.
- Camera based false-alarms: license plate recognition, a person crossing a virtual 'trip wire', face recognition.
- Human false-alarms: store and bank alarms (e.g. robbery).

In this article we are interested in most false-alarms that do not originate in faulty hardware. The number and importance of each False-alarm varies with the information source. Few but important false alarms are expected to be generated by persons, such as a bank employee accidentally pressing a button which is linked directly to the police or to a surveillance centre. Automated alarms are more likely to cause errors. Acoustic event detection and face recognition are examples of technologies which are particularly prone to false alarms [1].

Priority alarms

The most important alarms we categorize as 'priority 1'. These alarms should not be missed and require a high attention value (i.e. a high conspicuity) leading to a quick response. Priority 1 alarms require a 2nd, 'back-up' operator if the primary operator does not respond within a specified amount of time. This approach should

Figure 1. Conventional Video wall with numerous cluttered video images, ideal to hamper situation awareness and vigilance (http://www.videonext.com).

Figure 2. The interface (Left) centered on a map display showing color-coded camera viewing cones. The map display (Right) increases spatial awareness of the camera locations, orientations, and zoom. The color coding helps to match the image above to the viewing cone. Located at the CIV (Centrum voor Innovatie & Veiligheid) of the city of Utrecht.

guarantee a timely response to the most important alarms, irrespective why the alarm is missed. Priority 1 alarms will at least include bank (robbery) alarms. The other alarms will be given lower classifications, depending on the required response time.

Spatial awareness

Camera surveillance of public spaces typically makes use of a ‘video wall’ with numerous video screens (see Figure 1). The spatial awareness of the operator is limited by the ability to integrate the individual images into a consistent whole. Our prototype interface contains a street map with color-coded field-of-view cones as an attempt to combine the top-down map view with the camera perspective [2,3,4]. The viewing cones are particularly useful in combination with pan-tilt-zoom cameras. Each camera cone is uniquely color-coded corresponding to the edge of the camera image to help switch eye fixation between the ‘video wall’ and the map display (see Figure 2). The colors are chosen to be easily identifiable from the corner of the eye.

Sound Surveillance

The interface further contains ‘Sound surveillance’, designed to draw attention to (sound) events that - most likely - occur outside the camera field-of-view. Each event that surpasses a threshold value on an automatic ‘suspiciousness’ scale will briefly be played to the operator [1,5]. The location of the microphone recording the suspicious sound is simultaneously shown on the map display by smoothly expanding circles (not shown). The operator himself decides whether to pay closer attention to the event, saving precious time in case of obvious false-alarms. In an experimental study we examined ‘sound events’ interspersed with false alarms, simulating 3 real events for every 17 false-alarms, i.e. a false-alarm rate of 85%.

Conclusion

Interfaces may be expanded to handle alarms, real as well as false. To maintain vigilance, priority is to minimize the effort needed to check on alarms, thereby minimally distracting from the main (surveillance) task. Sound and vision are by and large - but not entirely - complementary senses. Sound does not, for example, require eye fixations. The same holds true for alarms that are easily visible from the periphery [6,7]. Smooth motion and color are particularly useful to attract attention but not distract [8]. Lastly, coordination among co-workers is a must to not miss the most important alarms such as a robbery alarm.

References

1. Van Hengel P., Huisman M., Appell J.E. (2009). Sounds like trouble. *Human Factors - Security and Safety, D. de Waard, J. Godthelp, F. Kooi, and K. Brookhuis, Eds. Shaker Publishing* (Maastricht, the Netherlands, July 2009), 369-375.

2. Kooi F.L. & Zeeders R. (2009). Improving Situational Awareness in Camera Surveillance by Combining Top-view Maps with Camera Images. *Human Factors, Security and Safety. D. de Waard, J. Godthelp, F. Kooi, and K. Brookhuis, Eds. Shaker Publishing.* (Maastricht, the Netherlands, July 2009), 355-368.

3. Endsley, M.R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors*, **37**(1), 65-84.

4. Endsley, M.R. (2004). Situation awareness: Progress and directions. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory, measurement and application*. Aldershot, UK: Ashgate Publishing, 317-341.

5. Truong K.P. & van Leeuwen D.A. (2007). Automatic discrimination between laughter and speech. *Speech Communication* **49**, 144-158.

6. Kooi, F.L., Toet, A., Tripathy, S.P., Levi, D.M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision* **8**(2), 255-279.

7. Toet, A., Kooi, F.L., Bijl, P., Valeton, J.M. (1998). Visual conspicuity determines human target acquisition performance *Optical Engineering*, **37**(7), 1969-1975.

8. Kooi F.L., Mosch M. (2006). Peripheral motion displays: tapping the potential of the visual periphery. *Proceedings of the Human Factors and Ergonomics Society (HFES) 50th annual meeting* (San Fransisco October 2006), 1604-1608.

Eliciting Control Errors and Measuring Error Correlates

M. Lindenthal

Center for Medical Engineering, Usability and Ergonomics, University of Applied Sciences Münster, Steinfurt, Germany. michael.lindenthal@fh-muenster.de

Introducing the PATRIA2 Project

In professions such as the medical sector, as well as in power plants, chemical plants and in transportation, people are often busy with work that can be risky. In many cases the fate or well-being of individuals depends on the people who are responsible performing their tasks correctly. A safe performance is crucial, but the structured or even repetitive nature of many tasks can lead to a worker's loss of attention, stress or other mental states which negatively affect performance by increasing the likelihood of human error. Errors may arise from haste, carelessness, fatigue, or deviating from prescribed procedures, and they often lead to avoidable adverse events (AAE).

The occurrence of AAEs can be reduced by monitoring the actions of workers by automatic systems. This means detecting deviations in task performance by observing it with cameras, extracting performance parameters, comparing the parameters to normative models for the correct implementation of the task, and initiating a corrective action when deviations are detected, either correcting the deviation or alerting the worker. It also means predicting the risk that an error might happen by detecting internal states of the worker that are correlated with error-proneness.

The goal of PATRIA2 – the project this presentation serves to introduce – is an integrated hardware/software solution to automatically assess behaviour and actions of a single operator interacting with a computer while engaged in some routine activity. The solution derives a parameter indicating the likelihood of an erroneous action in a specific task related setting. To this end, the PATRIA2 tool that is currently being developed non-invasively records eye gaze, head direction, gestures, facial expression, and user-input (keystrokes and mouse clicks). It relates operator's actions as indicated by user input with other properties of the operator's behaviour. In this way, the tool serves to mitigate risks posed by sub-standard task performance and human error.

Not every kind of error and associated risk lends itself to technical mitigation though. PATRIA2 concentrates therefore on slips and lapses (errors of attention or memory) rather than on rule-based or knowledge-based mistakes (i.e. the execution of a wrong plan of action), which require re-training and other organizational solutions. Measuring such errors can be difficult, owing to the low probability of their occurrence. Finding a way of improving the assessment of errors has been a crucial part of the project.

Analysing Error Proneness through Differentiation of Error Components

In our experimental scenario, participants are required to operate a computer-simulated medical device, with different user interface versions, while simultaneously working on a secondary task that consumes attention and thus increases the likelihood of making errors. Each use error is subdivided into two components according to the subtasks of action: execution and evaluation. Execution errors and evaluation errors, as components of use errors, have a higher probability than the use error as a whole and are therefore easier to measure. Under normal circumstances, evaluation errors can only be discovered if an execution error was made beforehand. To detect evaluation errors independently of execution errors, we simulate execution errors (e.g., entering a wrong number) and analyse the user's reaction in order to detect if they are followed by an evaluation error (also called a control error: failing to check if the input was done correctly).

Our studies use two different user interface concepts of a syringe pump, the pump being simulated on a computer screen. Using a simulation makes experimental manipulation, as well as data collection, easier. During the experiment, each participant is required to programme the pump a certain number of times, to operate at various different flow rates, as if they would give different drugs to various patients. The task is complicated by a secondary task that puts a strain on working memory capacity, by time pressure and a turbulent environment – circumstances that often appear, in principle, in the real application of medical and other devices (albeit in a different form), thus simulating a key feature of the final use cases in a controlled environment. The simulator introduces faked execution errors by replacing the user's input with a different number on a selection of trials. Depending on experimental conditions, subjects fail to detect up to 30% of these faked execution errors (about 10% on average), thus committing a control error.

This scenario can easily be adapted to the investigation of a related error class, namely mode errors – doing something that would be correct if the device was actually in the mode one thinks it is in. In this case, one would use or simulate a device with two or more modes and covertly change the mode in the course of the experiment. The scenario is therefore suitable for at least two relevant error categories: failing to control one's own input or failing to control the system state before or while making an input. These kinds of errors are among those that could possibly be compensated for by an automated behaviour observation system.

Conclusion

A method has been developed for studying the occurrence of control errors following an action that could be compromised by an execution error. This method introduces faked execution errors to enhance the likelihood that control errors can be detected. All keystrokes and mouse clicks are recorded together with the test participants' eye-tracking data, head movements, facial expressions and gestures. Data from these different sources can then be used to develop or test hypotheses about indicators of risks and critical situations. Applying such knowledge is a crucial part of the PATRIA2 project that aims at developing an automatic monitoring system to reduce human-error induced risks in critical, high responsibility tasks, either through training or real-time warnings.

Watching People Making Errors: Vision Architectures for Monitoring Task Performance

Marten den Uyl

VicarVision, Amsterdam, The Netherlands. denuyl@smr.nl

Abstract

People at work sometimes make errors, which can be of grave importance in some circumstances –e.g. medical, traffic- and bring nuisance and economic costs in other areas, e.g. order picking for web shops. It would be of value then, to have systems that can observe task performance and signal when –possibly- errors are being made, and thus may help to save lives and money. The Patria2 project (EURegio 2010) develops a system for observing task performance in medical and other settings, using primarily camera systems, and evaluates to what extent such a system could help in reducing avoidable errors by signaling increased risk of error. This presentation focusses on the challenges and problems of achieving adequate monitoring of task performance through the integration of results, in real time, from a number of vision systems under development at VicarVision.

A human supervisor could see increased risk of error in the way a subject performs a task from observing two kinds of deviations from standard or low risk behavior: 1) deviation by subject from standard or prescribed task procedures; and 2) deviation from optimal affective and motivational work attitude, e.g. stress, frustration or boredom may have negative effects on task performance. To see whether the subject deviates from standard procedures may require detailed knowledge of these procedures for the task at hand, as well as the ability to perceive specific details of task performance, hand and arm movements, direction of gaze. To perceive mental states and emotional responses in a subject may require even more attention to details of facial actions and body movements. And while performing a task, a subject may move outside the limited scope of a single camera. Thus, multiple camera systems will be required just to register task performance in sufficient detail. And for task monitoring to be relevant in most tasks, it will be necessary that the monitoring system responds in real time to an increased risk of error observed.

The Patria2 system involves three vision subsystems. FaceReader, a system for analyzing emotional expressions on the face that is available as a commercial product for behavioral research since almost 5 years www.noldus.com/human-behavior-research/products/facereader. GazeTracker, for determining direction of gaze from standard camera images, e.g. a webcam, and BodyReader, for the analysis of body pose en motion from video, are both under development for some years, but have not yet achieved maturity as a product. A typical set up for the Patria2 system uses six cameras.

The presentation will give an overview of the vision systems comprising the Patria2 architecture, status, performance and challenges. A number of special issues will be discussed:

- Alternative architectures for processing and integrating six or more camera streams in real time;
- Use of stereo and mono camera configurations;
- Integration of standard infrared eye tracking systems;
- Use of active depth sensors (Kinect);
- Problems of measuring mental states from video;
- Problems in identifying the precise steps in task performance.

Structural Knowledge Assessment: Change in Cognitive Structure Due to Playing a Serious Game

P. Wouters

Department of Information and Computing Sciences, Utrecht University, Utrecht, Netherlands.
P.J.M.Wouters@uu.nl

Introduction

Verbal assessment (e.g., knowledge tests, transfer tests) is often used to determine the effectiveness of complex learning environments such as serious games (computer games used for learning and instruction). However, current thinking in cognitive science emphasizes the importance of knowledge organization in learning complex skills [1]. An adequate organization of knowledge in knowledge structures warrants the capability to integrate new information with existing knowledge making the new information meaningful. We propose that structural assessment can be used to measure the quality of knowledge structures.

Structural assessment

The underlying assumption of the structural assessment approach is that we organize our knowledge in knowledge structures containing the important concepts of a domain and the relations among those concepts [2]. Although several methods exist to measure knowledge structures (e.g., ordered-tree techniques, cluster analysis), we have chosen the Pathfinder approach. Two important reasons for this choice were: (1) it does not force a hierarchical structure on the data but identifies meaningful links between concepts, (2) it is a rather straightforward method with predefined concepts, requiring less introspection and cognitive effort from the assessee. Three steps can be distinguished to implement a structural approach to measure and interpret knowledge structures.

1. Knowledge elicitation

Rating the relation between concept pairs is applied in cognitive psychology and is assumed to capture the underlying knowledge organization. The selected concepts are randomly combined and presented in n(n-1)/2 pairs (n is the number of concepts) to the learner who has to rate these on a 5, 7 or 9 point Likert scale. The pair ratings result in a matrix of proximity values between concepts indicating how closely the concepts are related (see Figure 1, left).

	A	**B**	**C**	**D**	**E**
A	0	1	3	2	3
B	1	0	1	4	6
C	3	1	0	5	5
D	2	4	5	0	4
E	3	6	5	4	0

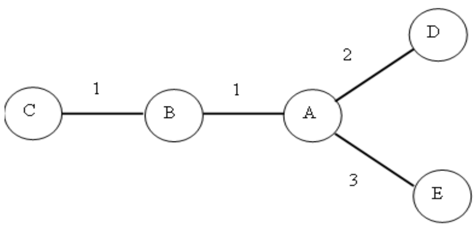


Figure 1. Proximity data (left) and their representation in a PFNet (right).

Pairs A-B and B-C have a distance of 1 and are directly connected because no shorter, indirect path exists (there is no lower link weight sum than 1). A-D has a distance of 2. A direct link connects the concept pair A-D because no shorter path can be found. As an example, consider the potential shorter path in which A-D is connected via A-B and B-D. A-B has a link weighed sum of 1, B-D one of 4 which makes 5 which is not a shorter path. The pairs A-E and A-C have distances of 3. The procedure creates a direct link for the pair A-E, but not for A-C. In the latter case a shorter indirect path from A to C exists via B.

2. Knowledge representation

The proximity values are difficult to interpret, therefore the elicited knowledge should be represented. The Pathfinder procedure uses a graph-theoretic distance technique (it uses the distance measured by the minimum number of links connecting two nodes in a graph) to represent the proximity matrix in a graphical network structure (a PFNET) [3]. Pathfinder links all concepts and assigns a weigh to each link (based on the ratings). Next, the Pathfinder algorithm removes direct links if there exists a shorter, indirect path that connects both concepts (see Figure 1).

3. Knowledge evaluation

The third step concerns the evaluation of the knowledge representation with a referent knowledge structure (e.g., instructor, expert). For the selection of an appropriate referent, issues such as the agreement on the core concepts in a domain and the number of instructors/experts used to generate the referent structure should be taken into account. In this paper we will use: *similarity* (the number of common links between both networks divided by the total number of links), *coherence* (an indirect measure of similarity by correlating the ratings given for each item in a pair with all of the other concepts) and *graph-theoretic indices of the focal node* (analyze and compare the central concepts).

Methods

Participants and materials

We used the game Code Red: Triage in which advanced (n = 10, M_{age} = 41) and novice players (n = 9, M_{age} = 31) learn in 20 minutes the basic procedure to classify victims according to their injuries. Before and after the game we administered a verbal test (10 mpc items on procedural and declarative knowledge) and structural assessment (based on 13 concepts, 78 pairs, 3 experts were used for the referent structure, 9-point Likert scale).

Results

The verbal test shows that both groups performed better after the game, but the differences before the game between advanced learners and novices had disappeared after the game. A closer look reveals that novices benefit most with regard to procedural knowledge, whereas advanced learners benefit most with regard to declarative knowledge.

The similarity of novices with the referent increased due to the game. This was not the case for advanced learners. The initial larger similarity with the referent for advanced learners -in comparison to novices- disappeared after the game. The coherence scores for novices did not differ from the advanced learners neither before nor after the game. For novices the decrease in coherence after the game was weakly significant, while the decrease was significant for the advanced learners.

Table 1. Results of verbal and structural assessment.

	Novices (N = 9)		Advanced learners (N = 10)	
	Before game	After game	Before game	After game
Knowledge test				
All items	4.00 (1.22)	6.89 (1.76)	7.70 (1.16)	8.30 (1.16)
Declarative items	2.89 (.93)	3.56 (.53)	3.90 (.74)	4.30 (.67)
Procedural items	1.11 (.78)	3.33 (1.50)	3.80 (.92)	4.00 (1.15)
Structural assessment				
Similarity scores	.11 (.07)	.16 (.07)	.24 (.13)	.21 (.08)
Coherence scores	.50 (.12)	.31 (.24)	.49 (.24)	.30 (.20)

	mobility	crt-nail	crt-head	pulse	weather	triage tape	classify	length	breathing	airway	stabilize	circulation	darkness	
mobility		2	4	2	3	3	1	4	2	2	2	3	5	2.8
crt-nail	2		2	1	1	3	1	4	2	2	2	1	3	2.0
crt-head	4	2		2	1	3	3	4	3	2	3	1	1	2.4
pulse	2	1	2		2	1	2	3	1	1	2	1	3	1.8
weather	3	1	1	2		3	2	4	2	2	3	1	2	2.2
triage tape	3	3	3	1	3		2	1	1	2	3	2	4	2.3
classify	1	1	3	2	2	2		3	1	1	1	2	4	1.9
length	4	4	4	3	4	1	3		2	4	4	4	5	3.5
breathing	2	2	3	1	2	1	1	2		2	3	1	3	1.9
airway	2	2	2	1	2	2	1	4	2		1	1	3	1.9
stabilize	2	2	3	2	3	3	1	4	3	1		2	4	2.5
circulation	3	1	1	1	1	2	2	4	1	1	2		2	1.8
darkness	5	3	1	3	2	4	4	5	3	3	4	2		3.3
	5	4	4	3	4	4	4	5	3	4	4	4	5	

median node

center node

Figure 2. Path distance matrix for the averaged expert PFNet.

We computed the three focal nodes: *highest degree node* (the node(s) with the greatest number of links), *median node* (the node(s) with the smallest average distance to all other nodes) and *center node* (the node(s) with the smallest maximum distance to any other node) for all PFnets (see Figure 2 for an example). Based on these nodes we defined the central concepts in the expert's PFNet. Advanced learners, but not novices, show a converging of central concepts with experts after the game.

Conclusion and discussion

The results suggest that structural assessment measures an individual's understanding of a domain at least differently from verbal assessment. While verbal assessment may provide a more nuanced picture regarding declarative and procedural knowledge, structural assessment may add an in-depth understanding of the concepts that are regarded important in a domain. The pattern of results yields at least three questions that should be discussed. First, we failed to find an increase in similarity between the advanced learners and the expert referent after the game. It is possible that there is no 'ideal' referent structure, but that each level of expertise requires a different referent structure (cf. [2], [4]). The second question concerns the high number of links in the PFNets. A possible explanation is that participants have used many extreme relatedness ratings (so many 1s and 9s) or rated many pairs as highly related (8s and 9s). For example, most novices show an increase in extreme relatedness ratings after the training which may have caused many tied values and Pathfinder includes all links when there are tied distances. The third question pertains to the decrease in coherence. It is possible that the lower coherence can be partly attributed to the fatigue and decreasing concentration during the second rating. An alternative explanation may be that the knowledge organization became more differentiated in connected clusters. For example, a cluster around physical body characteristics versus a cluster around environmental, external characteristics. Finally, we propose four guidelines to effectively use structural assessment in serious games:

1. Determine whether the domain allows a referent based structural assessment with similarity (agreement on important concepts in a domain) or that a referent free assessment with coherence is more appropriate.
2. When a referent based structural assessment is chosen, consider carefully which type of referent is most suitable. Especially when different levels of expertise are involved.
3. The concepts should be unambiguous and their number should not be too high (13-20 concepts). Consider to use a subset of a domain when a domain involves more than 20 concepts.
4. Consider the analysis of the graphical knowledge representations to obtain in-depth information about the quality of the knowledge structures.

References

1. Day, E. A., Arthur Jr., W., Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology* **86**, 1022–1033.

2. Kraiger, K., Ford, J. K., Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology* **78**, 311-328.

3. Schvaneveldt, R. W., Durso, F. T., Dearholt, D. W. (1985). *Pathfinder: Scaling with network structures* (Memorandum in Computer and Cognitive Science, MCCS-85-9, Computing Research Laboratory). Las Cruses: New Mexico State University.

4. Acton, W. H., Johnson, P. J., Goldsmith, T. E. (1994). Structural knowledge assessment: Comparison of referent structures. *Journal of Educational Psychology* **86**, 303-311.

Player-Centric Game Design: Adding UX Laddering to the Method Toolbox for Player Experience Measurement

Bieke Zaman[1], Vero Vanden Abeele[2]

[1]*IBBT-CUO, KU Leuven, Leuven, Belgium. bieke.zaman@soc.kuleuven.be*

[2]*GROUP T-Leuven Engineering College, Leuven, Belgium. vero@groept.be*

Abstract

In this paper, UX Laddering is introduced with its underlying theoretical perspective, Means-End Chain theory, and its methodological approach to measure game player experiences. Firstly, UX Laddering is positioned along the spectrum of game experience research methods. Then, a Means-End Chain theory inspired Game eXperience Model is presented. Next, the four-step approach of the UX Laddering methodology is briefly described, including

1) product interaction, 2) preference ranking, 3) Lenient Laddering and 4) a qualitative and quantitative analysis. This paper ends by explaining the applicability of the outcome of UX Laddering, the Hierarchical Value Map, as a communication tool towards game designers to support them in taking informed decisions during the game development process, from the analysis to game design and evaluation phase.

Introduction

A thorough understanding of game experiences requires insight into the complex interplay between three factors, including the 1) players, 2) characteristics of the game system, and the 3) game context [1]. Game experience can be measured on each of these dimensions. Studies that aim for an understanding of the game play *context* rely mostly on ethnographically inspired methods. The methods employed in studies on the assessment of the game *system* used to focus on bug testing, but recently also have incorporated more event-logging user data [e.g. gameplay metrics] [1]. The assessment methods of the idiosyncratic game experience of the *player* are most diverse. To illustrate, these cover amongst others objective, quantitative psychophysiological measurements [2], subjective surveys such as the Game Experience Questionnaire [3], as well as usability-oriented methods such as the RITE method [4] and more qualitative formative user testing methods like Deep Gameplay or the Initial Experience Playtest [5]. Many of the methods mentioned above have been introduced to support the design and development of games. Over the last years, game designers and researchers have been inspired by user-centered design practices in their search for appropriate methods to include players in a more player-centric game design process [5]. Typically, these measures are employed rather late in the game design cycle when most decisions that shape the gameplay have already been defined by the team of game designers. Yet, this is unfortunate because many design decisions can benefit from insights from user research from an early design stage onwards. Game design constitutes a second order design [6], meaning that a designer can only design the rules of the game and never the direct player experience. It is the challenge of finding the right theoretical perspective and methodological approach to account for the motivations and experiences of the players throughout the whole game design and development process. As an answer to this challenge, UX Laddering and its underlying theoretical framework – Means End Chain theory- are introduced to the game research community.

Games as a means to an end

Means-End Chain (MEC) theory has been developed in the domain of consumer research to provide an understanding of how people's needs and desires can be fulfilled through the interaction with products. In essence, Means-End Chain theory focuses on the linkages (*chains*) between product attributes (the *means*) on the one hand, and consequences of product use and personal values (the *ends*) on the other hand. The underlying premise is that people perceive certain product *attributes* as most likely to have certain desired *consequences*, which in turn seem beneficial to their individual *values* [7, 8]. Recently, it has been recognized that MEC theory can support research studies in other domains than consumer research because of its comprehensive perspective

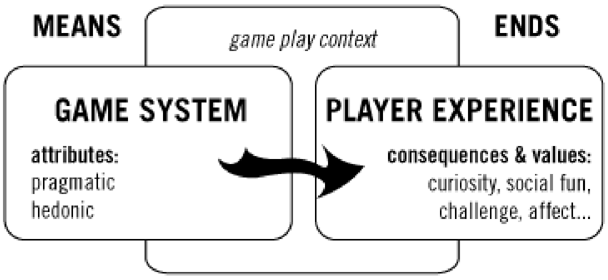


Figure 1. A Means-End Chain theory inspired Game eXperience Model.

on how to understand and analyse product experiences. More particularly, it has been argued that the user experience construct lends itself especially well to analysis in the context of MEC theory [9–11]. MEC theory's underlying premise points out that people prefer products, not because of the product's characteristics, but rather because of what these attributes can do for them in a particular context; the consequences of using a product unfold in the interaction between user, product attributes and context. This premise has also been agreed upon among user experience researchers. Moreover, it is also this premise that draws the parallels with game design as a second-order design problem. The design of a game system can never be a goal in itself, but rather that it is a means to provoking meaningful play as well as fulfilling player needs and values in a specific game play context.

Based on the insights from the game/user experience concept and the MEC theory we have summarized the potential of the MEC perspective for measuring player experiences, see Figure 1. In this MEC inspired Game eXperience Model, it is highlighted that in a particular game play context, both the pragmatic (e.g. usability-related) and hedonic (e.g. relating to aspects as flow or fantasy) game system attributes should be considered as a means to engage players in a meaningful game experience.

UX Laddering

The MEC theory provides us with several methods to reveal and understand product experiences and preferences. Laddering is one of these methods and foresees in a specific in-depth one-on-one interview protocol aimed at eliciting the respondents' attributes-consequences-values ladders with respect to a (limited set of) product(s). Laddering also refers to a particular data analysis protocol [12]. Recently, the potential of Laddering for Human-Computer Interaction (HCI) researchers has been put forward, which resulted in 2009 in the definition of UX Laddering [13]. UX Laddering is characterized by a four-step approach involving 1) product interaction, 2) preference ranking, 3) Lenient Laddering and 4) a qualitative and quantitative data analysis approach. UX Laddering always places the product under interest in a comparative, meaningful choice context, in which the experience with one or more alternative products or product classes is considered. As experiences unfold in the interaction between the individual and the system, and this in a particular context, people can only be questioned when they are able to reflect upon actual product interaction, and this preferably in a meaningful setting. Hence, UX Laddering's first step foresees in a short product interaction. Then, in the second step, the attribute elicitation phase follows by means of preference ranking. When the respondents are asked to rank two or more products, they will focus more on visible and concrete differences between the products than when the task is to elaborate freely on important product features [14]. The third step of UX Laddering consists in the actual interview in which the respondents *are ask*ed to reflect on those criteria that are important in distinguishing the products and in explaining what the experience with these products mean to them. The respondent is then repeatedly asked why-questions as many times as needed to reveal all relevant and salient criteria. It is this why-probing that typifies the Laddering interview approach during which the interviewer continuously asks for higher ordered, more abstract reasons and thus strives to climb up the ladder of attributes, consequences and (if possible and/or relevant) values [8, 15].

As for the fourth and last step, the UX Laddering data analysis, both a qualitative and a quantitative phase can be distinguished. It begins with the analysis of the interview transcripts according to the standards of qualitative

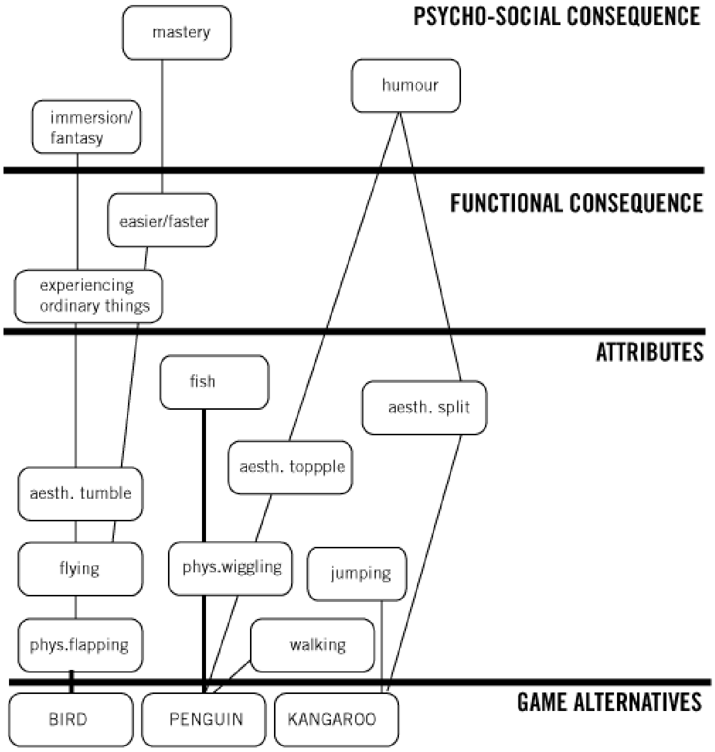


Figure 2. Example of a Hierarchical Value Map of a UX Laddering study in which children's dominant motivations of playing three types of cuddly toy interaction games are presented, to be read from bottom to top [16].

research and content analysis. The, the quantitative data analysis follows in order to reveal those dominant motivations that do no longer represent individual ladders but means-end chains at the aggregate level. This aggregation process results in the identification of the dominant relations at group level whereby marginal, individual links that are not repeatedly mentioned are ignored, and this all visually presented in a Hierarchical Value Map (HVM). Figure 2 shows an example of a HVM [16]. Such a HVM is the ultimate goal of UX Laddering as it provides an overview of dominant perceptions and motivations for using a certain product [15].

The Hierarchical Value Map as a communication and design support tool

The end result of UX Laddering foresees in a visual summary, the HVM, that highlights the most important results at a single glance, but at the same time includes sufficiently openness to account for the rich, qualitative insights that were gained during the in-depth interviews. Using the HVM as both a communication and design support tool is promising because the best communication strategy among researchers and designers, is one that enables both types of experts to talk in a similar jargon [17] and in which the results are concisely presented but at the same time do not lose all their richness [18]. The HVM accounts for these aspects and also facilitates a communication and decision strategy in which the right priorities are set based on the research conclusions while also leaving sufficient openness for design interpretation and discussion [18].

For instance, referring back to example in Figure 2, the HVM explains why certain children preferred one cuddly toy interaction game to the others. More particularly, it revealed that the bird prototype was preferred because of its physical interaction (flapping the wings) that led to the psycho-social consequence of immersion or mastery whereas the penguin (wiggling) and kangaroo (jumping) alternatives were preferred for benefits related to humour. The UX Laddering study showed the specific meanings these interactions had and the kind of experiences these could evoke among their players. Yet, these insights do not immediately prescribe new design aspects. Although there is still room for discussion, the design team will be able to take more informed decisions as they have insights into how and which concrete elements of their game relate to which player benefits [16].

Conclusion

This paper introduced UX Laddering and its underlying theoretical framework as a comprehensive approach to understand player experience whereby the results can be used to support the game design process. The UX Laddering approach reveals the link between players' motivations and game play benefits on the one hand and specific game design elements on the other hand. Its data gathering is based on interviews and the data analysis makes a transition from qualitative to quantitative analyses in order to reveal the dominant patters of player experiences. UX Laddering can be added to the game experience toolbox of methods that inform the design process of a game from the early development phases onwards. Its underlying theoretical framework, Means-End Chain theory, provides us with an array of concepts to describe the emerging game experience. When the interaction between the game system and the individual player is reflected upon, the UX Laddering method gives game researchers and designers the methodological support to understand it.

References

1. Nacke, L.E. (2009). *Affective Ludology: scientific measurement of user experience in interactive entertainment*, PhD thesis, Blekinge Institute of Technology, Karlskrona, Sweden.
2. Mandryk, R.L. (2008). Physiological Measures for Game Evaluation. In *Game Usability: Advice from the Experts for Advancing the Player Experience*. Isbister, K., and Shaffer, N., Eds., Morgan Kaufmann.
3. IJsselsteijn, W., Poels, K., de Kort, Y. (2008). The Game Experience Questionnaire: Development of a self-report measure to assess player experiences of digital games. *FUGA Deliverable D3.3*.
4. Medlock, M., Wixon, D., Terrano, M., Romero, R.L., Fulton, B. (2002). Using the rite method to improve products: a definition and a case study. In *UPA conference*, Orlando, Florida, USA, published online http://www.microsoft.com/download/en/details.aspx?id=20940, accessed 28 March 2012.
5. Pagulayan, R.J., Keeker, K., Wixon, D., Romero, R.L., Fuller, T. (2003). User-centered design in games. In *The Human-Computer Interaction Handbook. Fundamentals, Evolving Technologies and Emerging Applications*. Jacko, J. A., and Sears, A., Eds., 883–906, L. Erlbaum Associates Inc., Hillsdale, NJ, USA.
6. Salen, K., Zimmerman, E. (2003). *Rules of Play: Game Design Fundamentals*. Cambridge, MIT.
7. Kardes, F., Cronley, M., and Cline, T. (2010). *Consumer Behavior*. 1st ed., South-Western College Pub.
8. Olson, J. C., Reynolds, T. (2001). Understanding consumer Decision Making. In *Understanding Consumer Decision-Making: The Means-end Approach to Marketing and Advertising Strategy*. Reynolds, T. J., and Olson, J. C., Eds., 3–20, Lawrence Elbaum Associates, Mahwah N.J.; London.
9. Cockton, G. (2004). Value-centred HCI. In *NordiCHI '04*, 149–160, Tampere, Finland.
10. Friedman, B. (1996).Value-sensitive design. *Interactions **3***, 16-23.
11. Hassenzahl, M. (2004). The thing and I: understanding the relationship between user and product. In *Funology: from usability to enjoyment*. Blythe, M., Overbeeke, K., Monk, A., and Wright, P., Eds., 31–41, Kluwer Academic Publishers, Norwell, MA, USA.
12. Grunert, K., Bechlarsen, T. (2005). Explaining choice option attractiveness by beliefs elicited by the laddering method, *Journal of Economic Psychology **26***, 223-241.
13. Vanden Abeele, V., Zaman, B. (2009). Laddering the User Experience! In *UXEM09 workshop*, Published online http://soc.kuleuven.be/com/mediac/cuo/?p=publication&id=165, accessed 28 March 2012.
14. Costa, A.I., Dekker, M., Jongen, W.M. (2004). An overview of means-end theory: potential application in consumer-oriented food product design, *Trends in Food Science & Technology **15***, 403-415.

15. Reynolds, T.J., Gutman, J. (1988). Laddering theory, method, analysis, and interpretation, *Journal of Advertising Research* ***28***, 11-31.

16. Vanden Abeele, V., Zaman, B., and De Grooff, D. (2012). User eXperience Laddering with Preschoolers: Unveiling Attributes and Benefits of Cuddly Toy Interfaces, *Personal and Ubiquitous Computing,* ***16****(4):451-465*. Zaman, B., Vanden Abeele, V., Markopoulos, P., and Marshall, P., Eds.

17. Rogers, Y. (2004). New theoretical approaches for HCI, *Annual review of information science and technology* ***38***: 87–143.

18. Visser, F. S., Lugt, R. V. D., and Stappers, P. J. (2007). Sharing User Experiences in the Product Innovation Process: Participatory Design Needs Participatory Communication, *Creativity and Innovation Management* ***16***: 35–45.

Effects of Playing a Serious Game: A Comparison of Different Cognitive and Affective Measures

Erik D. van der Spek

Department of Industrial Design, Eindhoven University of Technology, Eindhoven, Netherlands.
e.d.vanderspek@tue.nl

Introduction

Through a number of experiments we tried to determine general game design rules that improved the efficacy of a certain serious game. The game was called *Code Red Triage*, a total conversion modification of Half-Life 2, and was purposefully made in order to systematically empirically test the aforementioned game design rules, which in turn were based on better aligning the game with the player's cognitive system.

In Code Red Triage, a player would take on the role of a medical first responder, who has to learn how to perform a mobility (sieve) triage, or categorizing the many victims of a mass casualty event according to urgency of needed medical attention. The player started at the central hall of a train station where he (or she) received a message that a terrorist bomb strike had taken place on one of the subway platforms. The player then had to find the way to the subway platform via a number of winding corridors. Arriving at the platform, the player could triage victims by walking up to them and entering the triage menu; here, the player could choose a number of triage actions and assign one of four triage categories. From the feedback provided after performing the triage actions, as well as after categorizing the victim, the player had to deduce, iteratively through nineteen victim cases, what the correct procedure of the triage was. The feedback was presented in the form of short text snippets and a game score which detailed how well the player had performed the triage procedure on the current victim.

A total of six studies were performed with the game: a pilot study, a media comparison study where the efficacy of learning with the game was compared to learning with a conventional PowerPoint slideshow, and four value added experiments where the effect of one or more interventions were compared with a control group. The experiments were based on Mayer's cognitive theory of multimedia learning [1], which states that one goes through three cognitive processes in order to learn something: the selection of relevant over irrelevant information, the organization of the relevant information into coherent knowledge structures, and the integration of these knowledge structures into prior knowledge. The four experiments were: (1) the introduction of auditory and visual attention cues to ameliorate the selection of relevant information; (2) a 2-way comparison between just-in-time and just-in-case option complexity versus a massed or a spaced approach to victim complexity, to see which approach led to improved organization of information; (3) the introduction of an adaptation engine where, if the player scored well on a certain victim, similar victims were deleted, to see whether this led to more efficient information organization; and (4) the introduction of surprising events to stimulate better knowledge integration. A complete description of the experiments, as well as the results, can be seen in [2]. Here instead, we will focus on the used measurement instruments and how the participants' scores correlate, as an indication of their validity.

Game experience

Any experiment on learning with games may be confounded by a participant's prior game experience for a number of reasons. For instance, a novice to games and 3d games especially, may need to afford considerably more mental effort to spatially navigate an avatar through a game world than a person that has ample experience in this, which in turn could lead to less cognitive capacity that may be used for learning. Additionally, experienced gamers will be more accustomed to ways in which a game conveys important information to the learner.

One of the problems we faced was how to measure prior game experience. A conventional way would be to ask how many hours a participant spends playing games per week; however there are three problems with this measurement. Firstly, a person may have spent ample time playing games in the past, but has been too busy to play games lately, in which case he would be incorrectly categorized as an inexperienced gamer. Secondly, it is unclear what type of game experience it measures; a person may play many hours of solitaire on an office computer but still be unable to spatially navigate a 3d environment. Lastly someone may not play games, but like to read about them extensively.

We had purposefully made the way from the train station to the subway platform labyrinthine, in order to measure the time it took a participant to reach the platform, under the presupposition that experienced gamers would have less problems with spatially navigating the virtual environment, and therefore arrive quicker. However, in a pilot study we found this to be completely uncorrelated with time spent gaming [$r(19) = 0.13$, n.s.]. By surveying the paths taken and asking the participants afterwards, we discovered that experienced gamers were more inclined to wander off and explore the game world, whereas inexperienced gamers were more inclined to follow the signs towards the subway platform. In the end we therefore settled with a self-report measure that asked whether the player a) rarely played games, b) sometimes played games, or c) considered themselves a gamer. This was highly correlated to time spent gaming in the four experiments [$r(41) = 0.85$, $r(56) = 0.69$, $r(28) = 0.67$, $r(41) = 0.70$, respectively, all p's < 0.001] and, for the first (attention cueing) experiment was able to discern a significant effect of prior game experience on the ability to understand and correctly use the visual cues.

Learning

Arguably the most important part of measuring the efficacy of serious games for learning is measuring how much is learned. Three measures for learning were used: an in-game score, a pen-and-paper knowledge test and a structural knowledge assessment.

In-game score. The in-game score was a measure for how well the player performed the triage procedure on a victim, ranging from 1 to 100 for any victim, to a total of 1900 for all 19 victims. Aligning the thing to be learned with the core game mechanics leads to better learning [3]; but at the same time it also makes for an excellent stealth assessment [4], where the player's learning progress can be continuously monitored without him (or her getting) disengaged from the game. A possible problem with using the in-game score as a measure for learning could however be that choosing the wrong option will lead to a lower score, but provides the player with the opportunity to learn from his mistake, possibly even more so than a person that chose the right answer and doesn't reflect on his actions. The performance in the game may therefore not be an accurate depiction of actual learning after the game is finished.

Pen-and-paper knowledge test. We consequently also had the participants perform a knowledge test before and after the game. The questions were in the form of 'given a victim where you just measured x, then the next step in the primary triage procedure would be?'. Eight of the questions were verbal only, eight additional questions also contained a screenshot of a victim as it would appear in the game. The pictorial versions of the questions were added because, in research done by [5], it was found that people who played a game remembered visual information better than text. In the four value-added experiments the participant had to choose one of four possible answers per question, in the comparison experiment with a PowerPoint presentation, each question had one of six possible answers and an additional retention test was performed a week after the intervention.

Structural knowledge assessment. While pen-and-paper knowledge tests are a good way to measure the learning of declarative knowledge and how well a participant can replicate this knowledge afterwards, it is ill-suited to measure deeper learning, or how well the learner stores knowledge structurally. We therefore used another instrument, the structural knowledge assessment, specifically to measure deeper learning. For this, the participant had to rate pairs of concepts of the triage procedure on their degree of relatedness. From these relatedness ratings, a computer program (PCKNOT) uses the Pathfinder algorithm to create graphs, where concepts that are closely related to each other only have one or a few links separating them, whereas concepts that are unrelated appear further away in the graph [6]. These graphs can subsequently be compared with those of an expert leading to a similarity score [7]. This score is then a measure of how well a participant's knowledge structure resembles that of an expert and an indication of transfer of training [6].

Correlations and conclusions. Due to space constraints, we will only report the interesting correlations between different measurement types. For the four value-added experiments they are printed in Table 1. In addition, in the study comparing the game to a PowerPoint, both the verbal knowledge test and the pictorial knowledge test were correlated with the in-game score after playing the game [$r(48) = 0.90$, $p < 0.01$; $r(48) = 0.60$, $p < 0.01$ resp.], as well as for the delayed knowledge tests [$r(47) = 0.78$, $p < 0.01$; $r(47) = 0.75$, $p < 0.01$ resp.]. The structural knowledge assessment after playing the game was correlated with both the verbal knowledge test and the pictorial knowledge test after the game [$r(47) = 0.30$, $p < 0.05$; $r(47) = 0.30$, $p < 0.05$], but not with the in-game score.

Table 1. Correlations between the different learning metrics over the four different experiments. SKA = Structural Knowledge Assessment, * = p < 0.05, ** = p < 0.01

	Score – Verbal knowledge test	Score – Pictorial knowledge test	SKA – Verbal knowledge test	SKA – Pictorial knowledge test	SKA - Score
Experiment 1	0.573**	0.744**	0.364*	n.s.	n.s.
Experiment 2	0.291*	0.505**	n.s.	0.363**	0.293*
Experiment 3	0.546**	n.s.	n.s.	n.s.	n.s.
Experiment 4	0.409**	n.s.	0.378**	n.s.	n.s.

Given the results mentioned above and in Table 1, there doesn't seem to be a measure that is entirely redundant. The verbal knowledge test is always correlated with the in-game score, which corroborates the notion that the in-game score is a valuable measure for learning in the game. Interestingly, the pictorial knowledge test seems to more closely measure what was learned in the game in the first two experiments, but then is uncorrelated to the in-game score in the latter two experiments. The reason for this ostensible switch is unclear. In addition, Table 2 shows the measures that revealed a significant effect of the experimental condition.

In Table 2 one can see that the structural knowledge assessment has been able to pick up effects of an experimental condition that a pen-and-paper knowledge test was unable to discover, meaning that game design decisions sometimes activate the deeper processing capabilities of a player without affecting more superficial learning, or even the in-game score. Given the previous considerations, we therefore advise the usage of all three different measures for learning.

Table 2. Measures that showed a significant effect of condition, * = significant effect but unimportant due to differences in the way the condition was set up

	Verbal knowledge test	Pictorial knowledge test	Struc. Knowledge Assessment	Score
Experiment 1	n.s.	n.s.	Sig.	Sig.
Experiment 2	n.s.	n.s.	n.s.	Sig.*
Experiment 3	Sig.	Sig.	Sig.	Sig.*
Experiment 4	n.s.	n.s.	Sig.	n.s.

Affect

We lastly want to briefly discuss the affective measure(s) that we used. Primarily we wanted to test whether the game design interventions did not negatively influence the engagement of the game, as some cognitive guidance techniques could take the player out of the experience. We therefore needed to measure both the enjoyment/engagement of the player as well as their feelings of presence. For this we used the engagement subscale of the ITC Sense Of Presence Inventory (SOPI) [8], as this seemed to give a good mix of the two while not overburdening the participant with questionnaires. However the Cronbach alpha's were poor (< 0.60) for two of the four experiments and in between eight experimental interventions, only one saw a small significant improvement in engagement. We offer three main reasons for this: One, small cognition-based game design interventions do not impact engagement. Two, the scale is rated on a five point Likert scale and participants rarely used the extremes, making any possible effect very subdued. Two, in letting the participant play only one condition of the serious game, it is unclear what the user's reference point was to indicate how engaged he was. Was it another form of instruction, in which case the game could have been appraised favorably, or was it a (entertainment) game, in which case our comparably dull serious game may have suffered. We therefore propose that researchers use 7 point Likert-scales as well as give the player a clear referent to measure the serious game (condition) against.

References

1. Mayer, R.E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning*, 31-49. New York, NY: Cambridge University Press.

2. Van der Spek, E.D. (2011). *Experiments in serious game design: a cognitive approach*. Doctoral dissertation.

3. Habgood, M.P.J., Ainsworth, S.E. (2011). Motivating children to learn effectively: exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences* **20**(2), 169-206.

4. Shute, V.J., Ventura, M., Bauer, M.I., Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects*, 295-321. New York, NY: Routledge.

5. Belanich, J., Sibley, D.E., Orvis, K.L. (2004). *Instructional characteristics and motivational features of a PC-based game* (Research Report No. 1822). Alexandria, VA: U.S. Army Research Institute.

6. Wouters, P., Van der Spek, E.D., Van Oostendorp, H. (2011). Measuring learning in serious games: a case study with structural assessment. *Educational Technology Research and Development* **59**(6), 741-763.

7. Thompson, L.A., Gomez, R.L., Schvaneveldt, R.W. (2000). The salience of temporal cues in the developing structure of event knowledge. *The American Journal of Psychology* **113**(4), 591-619.

8. Lessiter, J., Freeman, J., Keogh, E., Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-sense of presence inventory. *Presence: Teleoperators and Virtual Environments* **10**(3), 282-297.

Assessing the Personality Trait Compliance in a Game Context

Christof van Nimwegen[1], Herre van Oostendorp[1], Alec Serlie[2], Joost Modderman[2]

[1]*Department of Information and Computing Sciences, Faculty of Science, Utrecht University, Utrecht, The Netherlands. C.vanNimwegen@uu.nl*

[2] *GITP Research, Rotterdam, The Netherlands. A.Serlie@gitp.nl*

Introduction

Serious games for contexts *not primarily aimed at entertainment* are a well-established phenomenon. Well-designed serious games captivate players for a specific purpose, and in spite of the often non-entertainment purpose they provide engaging environments. Players are provoked to endeavor in higher-order cognitive skills such as handling of information, decision making and solving problems. Serious games are widely implemented in areas as diverse as healthcare, the military, and persuasive technology or applications aimed at attitude change [1]. There is also the rising trend to use serious games in recruitment contexts, e.g. to acquaint people with certain professional areas and enthuse them, of which America's Army [2] is a famous example. Using serious games for *assessment* in a recruitment procedure, is one step further, which is gaining popularity as a screening means next to standard interviews. Playing a game with the goal of assessment is about uncovering personal traits, skills and capabilities (or, of course, deficiencies) that are present in persons. Personality tests have been in use for a long time already e.g. for employee selection, team building or training. They can be paper tests, interviews, or computerized versions of these, but there are a few disadvantages, one of them being the fact that it is sometimes too easy to give socially desirable answers. Furthermore, a self-report questionnaire does not provide information about the players' decision process. Our idea is that accommodating assessment in a game (although not purely meant to entertain) provides a richer, less repetitive and predictable, and more immersive context. Players might act more natural and intuitive compared to holding a pen and paper, and thus contribute to the validity of the measurement. Serious games can uncover individual personality traits of people in a structural manner, uncolored by the chemistry between interviewer and (future) employee.

The Compliance Game

This paper describes the process of developing a first version of "The Compliance Game" and its evaluation, followed by a second version and its evaluation. An additional aim of the project, using the game as a test bed, is to develop a notation and visualization method called "GameDNA" (<u>Game</u> <u>D</u>iscourse <u>N</u>otation and <u>A</u>nalysis). It is player-centric and accommodates mental but also other actions that players face in the game, along with the discourse between system and player [3]. In serious games where balancing between mental decisions and effort from the player, and reactions to the system is the issue, this becomes important. With this tool people from different disciplines (as in this research project) can clearly see what the player has to do, how the flow of information is shaped and how the prospected measurable features are balanced and correctly implemented.

There were three partners in the project. GITP is an international Human Resource Management consultancy firm whose goal is to enhance organizational effectiveness and performance by assessing people. In this project, they wanted a game developed to measure "Compliance": the tendency of an individual to agree with an instruction from another person, even if he/she does not believe it, only to satisfy the other(s) or to avoid conflicts. Serious games have been used in the context of compliance, such as "Straight Shooter!" by [4], a real-time 3D e-learning game aimed at compliance. However, this involved training, while our aim was assessment, uncovering which attitudes are already present in people *before* they possibly enter a company. Compliance is a hot topic today in the aftermath of financial crises in banking and financial services. There is a need for better assessment, as the rules are getting stricter and policies are changing. The technical development was done by Ranj, a serious game studio that is endeavoring to serious games for assessment, such as their award winning "Houthoff Buruma The Game" [5]. The third partner is the Department of Information & Computing Sciences (ICS) at Utrecht University and is involved in various gaming projects such as GATE [6].

The Compliance Game version 1

Content

In the Compliance game, the player acts in a setting in which Compliance and its associated behavior is an everyday occurrence: the sales department of an international company that manufactures smart connected devices (e.g. computer pads). The player acts in a story, makes decisions and expresses opinions. In the scenes in the game assignments must be carried out (or not), managerial decisions must be taken, suggestions to act immoral (but profitable for the company) must be acted upon.

The game "storyline" was made as engaging as possible, and was constructed such (especially the dialogues) that players will believe that the game characters reactions are a specific reaction to the player's own actions. This is not the case in reality, the game is sequential in the sense that all players have to pass the same milestones, although there is of course freedom in actions/answers during the game. The measurements (see next paragraph) are quantitatively analyzed afterwards, therefore it should be safeguarded that all players are confronted with the same events to act on. The building stones of the game are:

- Animated realistic characters (Figure 2) which are moving and talking (pictures of humans shown as moving stills, audio fully streamed). They talk to the player and ask questions.
- Text shown, e.g. the actions the player can choose from.
- A Smartphone (Figure 1). Text messages, emails and voice message arrive here, and can be read/listened to.
- Actions: grab phone with the mouse, open an item and close it, and choose actions and make decisions by clicking on one of them.

Measurements in the game

What is measured is the construct "Compliance" and 4 constructs of "Machiavellianism", one's propensity to distrust others, engage in amoral manipulation, seek control over others, and seek status, for which scales already exist in reliable reference tests [7]. 5 constructs are measured: Compliance, and of Machiavellianism: "Distrust in others", "Amorality", "Desire for control" and "Desire for status". The actions in the story that players choose, correspond to a 4-point Likert-scale of which the answers range on the spectrum on one of these 5 constructs. A typical situation in the game is that the player is confronted with the dilemma whether to accept assistance from someone else or not (there might be a hidden agenda behind it), involving considerations of whether and how to delegate. The player has to indicate the degree to which he/she would act compliant on a certain proposition (Figure 3).

What this game should accomplish is to measure the same as the reference tests do, i.e. correlate highly. If

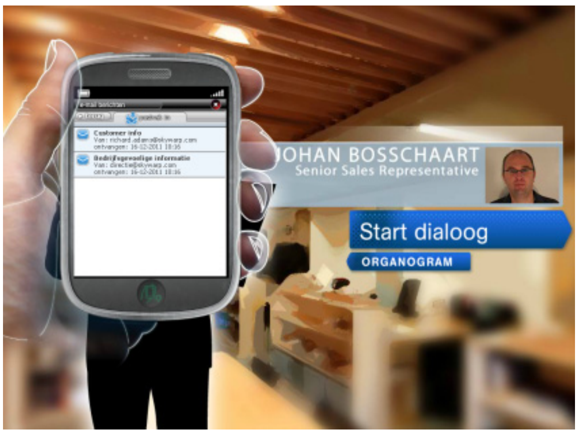


Figure 1. Using the smartphone in the game.

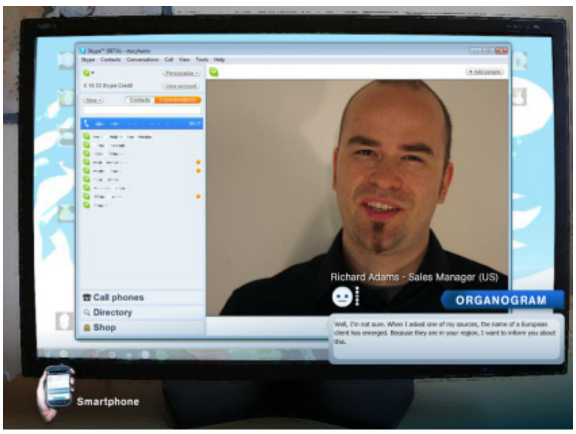


Figure 2. A talking and moving character in the game.

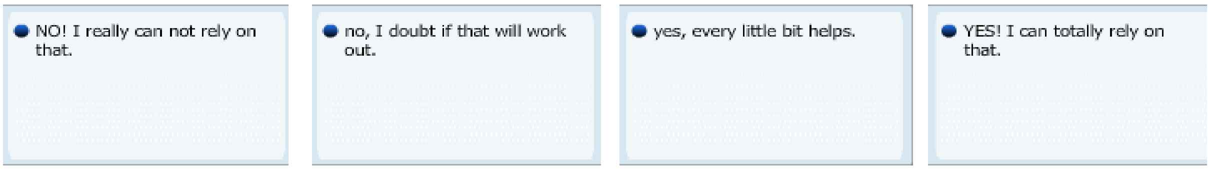


Figure 3. Answers in the game players can choose from ranging on a (in this case "distrust") scale

similar scores on the 5 constructs are yielded, one can say that the game can be used in a valid way to measure such personality traits. The first game version consisted of 5 scenes. Each scene is a part of the story and accommodates 3 different constructs in the storyline on which the player has to act. In each scene, there are 3 actions pertaining to the constructs involved, 2 actions during dialogue, and 1 action at the end. When the game is played, there are 3 measurements on each construct, which are subsequently averaged to a score indicating the position of a player on that construct.

Evaluation of game version 1

The first game version was played by 72 test participants from diverse age groups, backgrounds and professions. After playing the game, they also filled out the "classical", validated reference test (online) consisting of 36 questions, aimed at measuring the same constructs. This in total took about 45 minutes. The aim in this phase of the project was to have a game that yields the similar scores as the reference tests. Although at first glance similar *average* scores on the constructs of the game and reference tests were found, the results were not yet convincing. On none of the constructs, correlations between game score and reference test score reached significance. Analysis showed that the problem probably was the reliability; the internal consistency (Cronbach's alpha) of the 3 items of the 5 constructs was very low, respectively Compliance .30, Distrust .35, Amorality -.32, Control .15 and Status .14. In other words, the separate items of each construct did not quite measure the same thing. Qualitative questionnaires that were administered in addition revealed that in general, the test participants claimed to like to play this game, at times they felt "immersed" and found it more interesting (and at times more difficult) than filling out questionnaires.

The Compliance Game version 2

Development

After the informative, but not sufficiently satisfying results of the evaluations of the first game version, careful analysis took place. Using GameDNA, the game and the empirical results were carefully dissected and we scrutinized the places where inconsistent answers took place. The measurements of each construct were revised or replaced, and the amount of measurements was extended to 4, in order to obtain more reliable means that are based on more scores. The game was shortened a bit, answers that had the risk of ambiguity were refined, unnecessary interaction between system and player removed and more details were changed, including some graphics and audio.

Evaluation

Now there were 64 participants from diverse age groups, backgrounds and professions. The preliminary results are slightly more promising. Amorality in the game and amorality by the reference test had a significant correlation, $r(63) = .53$, $p < .01$, and Cronbach's alpha of this construct (4 items) measured in the game was in the right direction at .42. Compliance in the game and compliance by the reference test also had a significant correlation, $r(63) = .25$, $p < .05$., but surprisingly, Cronbach's alpha of the 4 items of the construct was `0`. The correlations between game scores and reference test scores on "Distrust", "Desire for control" and "Desire for status" still did not reach significance yet. Cronbach's alpha of these constructs were low; .22, .12 and .18 respectively. Summarizing, on 2 of the 5 constructs similar scores were obtained both in the game and in the reference test.

Conclusion and Discussion

Stakeholders from different disciplines cooperated on our assessment game measuring Compliance by embedding psychological constructs in a serious game. We have seen that is very difficult to consistently replicate personality trait constructs as the one we envision. The first game version did not result in convincing correlations between game scores and reference test scores. But since all moves and actions of the players were logged, the sessions were highly informative because by using GameDNA [3] we could see where it went wrong in the game. Unlike when measuring personality traits with standard non-game reference tests, the player acts in a story. The questions to be answered are not standalone; story elements of a scene can yield a situated judgment of one kind on a construct, while in another scene the judgment on a similar question is different. Careful iteration, using GameDNA does on the one hand dissect the game events and actions, and on the other hand it provides a helicopter view of the discourse and information flow in the game story. This resulted in a second, more balanced and elegant game version. With the second game version on 2 of the 5 constructs similar scores were obtained both in the game and in the reference test. For yet another iteration, it is useful to scrutinize on “Distrust”, “Desire for control” and “Desire for status”. An iterative approach as described here might prove imperative in the development and validation of serious games of the type we envision. The multidisciplinary nature of the project, cooperating with domain experts in assessment and game developers, calls for one shared language to shape the narrative and map game events and information flow. For this purpose, GameDNA will be further developed and applied in the course of future game development.

References

1. Susi, T., Johannesson, M., Backlund, P. (2007). Serious games - an overview. Technical Report HS-IKI -TR-07-001. University of Skövde.

2. America’s Army© (2002). http://www.americasarmy.com. Accessed 20 February 2012.

3. Nimwegen, van C., Oostendorp, van H., Modderman, J., Bas, M. (2011). A test case for GameDNA: conceptualizing a serious game to measure personality traits. Proceedings of the *16th International conference on Computer Games: AI, Animation, Mobile, Interactive Multimedia, Educational & Serious Games* (CGAMES 2011 USA), IEEE digital library.

4. Prensky, M. (2004). *Digital Game-Based Learning*. New York: McGraw-Hill

5. Houthoff Buruma. The Game©. http://www.experiencehouthoff.com. Accessed 2 January 2011.

6. GATE. http://gate.gameresearch.nl/. Accessed 15 December 2010.

7. Dahling, J.J., Whitaker, B.D., Levy, P.E. (2009). The development and validation of a new Machiavellianism scale. *Journal of Management* **35**, 219-257.

SPECIAL SESSION

Progress in Assessing Animal Welfare in Relation to New Legislation: Opportunities for Behavioural Researchers

Chair: Penny Hawkins

Research Animals Department, RSPCA, Southwater, UK. phawkins@rspca.org.uk

Abstract

The need for effective assessment and retrospective review of laboratory animal welfare, including recognition of pain, suffering or distress, is recognized in many laws and guidelines that regulate animal use. For example, the US Institute for Laboratory Animal Research (ILAR) *Guide* emphasizes the importance of post-approval monitoring that includes regular review of adverse effects, and the new European Union (EU) Directive that regulates laboratory animal care and use requires retrospective assessment and reporting of the level of suffering experienced by animals.

Accurately recognizing, recording, analyzing and reporting animal behaviors are all important in order to conduct a proper retrospective assessment of the level and nature of suffering, and behavioral researchers are clearly in an especially good position regarding the ability to achieve this. They have also contributed greatly to our knowledge of behaviors associated with negative states, such as discomfort, pain, anxiety and distress – and positive wellbeing – in a wide range of species, as well as to the development of accessible techniques for observing and monitoring animals. Increased dialogue between behavioral researchers and those working in other fields would therefore be helpful in improving the standard of assessments of laboratory animal behavior. The outcome would be not only better animal welfare, but also better science, as it is now very widely recognised that avoidable suffering can lead to experimental confounds, for example due to physiological responses to stress.

This symposium will explore how researchers using and developing behavioral recognition and monitoring techniques are contributing towards improving the understanding and interpretation of animal behavior, particularly relating to the assessment of animal welfare and suffering.

SPECIAL SESSION CONTENTS (sorted by paper ID)

Monitoring Burrowing and Nest Building Behavior as Species-specific Indicators of Animal Wellbeing

Paulin Jirkof

Institute of Laboratory Animal Science and University Hospital Zurich, University of Zurich, Zurich, Switzerland.
paulin.jirkof@ltk.uzh.ch

Impairment and pain in laboratory rodents

Legislative bodies and the general public both demand that suffering is minimized when animals are used for scientific purposes, which is important for both ethical and scientific reasons. In addition to refined experimental methods and good husbandry practice, the recognition of reduced wellbeing is a prerequisite to assure the welfare of laboratory animals. It is essential to be able to recognise and assess pain reliably so that it can be treated effectively, e.g. following painful experimental procedures or when animals are suffering from disease. Although the mouse is the most widely used laboratory animal species, there has been a lack of non-invasive methods to assess moderate, lasting pain.

As behavior can be observed easily in a non-invasive manner, and can provide meaningful indicators of animal wellbeing, behavioral changes have been proposed as a useful and simple means to assess welfare in small laboratory species [1, 2, 3, 4].

We conducted a series of experiments using two highly motivated, species-specific home cage behaviors, burrowing and nest building, as sensitive indicators of mild to moderate post-operative pain. These behaviors could also be used to test the efficacy of analgesic treatment. Mice of the commonly used C57BL/6J strain underwent a minor, one-sided laparotomy under inhalation anesthesia, with or without pain treatment (the non steroidal anti-inflammatory drug Carprofen). Control animals received anesthesia/analgesia only. We recognise that it is not usual practice to withhold perioperative analgesia, but the decision was made that it would be justified in this case because the aim was to achieve the benefit of better pain recognition for laboratory mice. All animal housing and experimental protocols were approved by the Cantonal Veterinary Department, Zurich, Switzerland.

Burrowing performance

This study investigated the potential use of burrowing performance as a measure of mild to moderate post-operative pain in laboratory mice. The influence of minor surgery on burrowing was analysed in C57BL/6J mice of both sexes, using a modified rodent burrowing test within the animal's home cage. A standard, opaque plastic water bottle filled with food pellets identical to those of the animal's normal diet was provided for burrowing, and an additional empty bottle was provided to serve as a shelter for the animal (see Figure 1). We measured the latency to the displacement of food pellets (burrowing) from the burrowing apparatus, using infrared sensitive cameras to record animals for 24 hours in the absence of a human observer.

Almost all (98%) healthy mice burrowed (mean latency 1.3 h). After minor surgery with anesthesia but no pain treatment, latency of burrowing was significantly prolonged (mean Δ latency 10 h). Analgesic treatment with Carprofen decreased the latency of burrowing after surgery (mean Δ latency 5.5 h) to the level found in mice that had only been anesthetized (mean Δ latency 5.4 h) or had received anesthesia and analgesia (mean Δ latency 4.6 h). Analgesia during surgery was associated with a significantly earlier onset of burrowing compared to surgery without pain treatment.

Nest complexity scoring

In a previous study we demonstrated a general correlation of post-operative pain and nest building performance in laboratory mice [2]. However, a standardised protocol for the assessment of pain using nest building

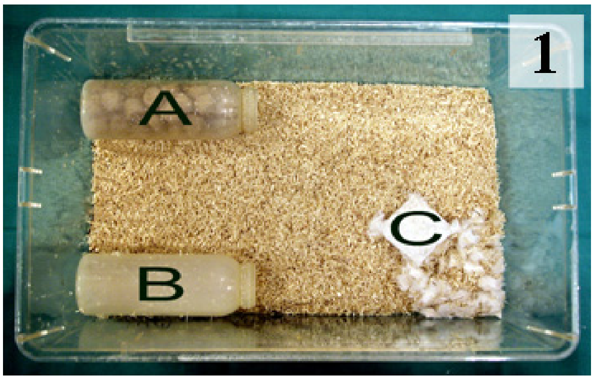


Figure 1. Experimental setup. A: Burrowing test apparatus, B: shelter, and C: nesting material (nestlet) in home cage [3].

Figure 2. Example of complex nest: nest score 5; made from commercially available nestlet (Indulab).

performance has so far not been developed. Here we used a nest complexity rating scale to standardise nest scoring for the assessment of post-operative pain (score 0 = no nest building activity to score 5 = complex nest; see Figure 2) in female C57BL/6J mice. As mice tend to destroy and rebuild their nests in a circadian rhythm, successful assessment of nest building performance depends on the time when the nest is observed. Therefore we analysed the normal 24h nest building rhythm in healthy mice to determine a suitable time to score nest complexity.

Healthy mice spend 4.3% of their daily time budget on nest building and there are two peak times for this behavior; (i) in the beginning and middle of the light phase and (ii) in the second half of the dark phase. For pain assessment we choose a scoring time nine hours after the experiments, at the end of the light phase following the second nest building activity peak. Healthy mice had mean nest scores of 2-4 in baseline measurements. After experiments a significant decrease of nest complexity was seen with mean nest score 0 after surgery without pain treatment, 0.7 after surgery with pain treatment and 1.7 in the anesthesia only group.

Conclusion

Analyzing changes in species-specific behaviors has been suggested as a promising approach to assess both pain severity and the efficiency of pain management regimens [e.g. 3, 5]. The burrowing test and nest complexity scoring were both shown to be reliable and non-invasive tools for the assessment of post-operative impairment of general condition and pain in laboratory mice, and can be applied easily within the normal laboratory routine. However, detailed evidence for the efficacy of specific pain treatments will need further investigations.

References

1. Deacon, R. (2012). Assessing burrowing, nest construction, and hoarding in mice. *Journal of visualized Experiments* **59**.

2. Arras, M., Rettich, A., Cinelli, P., Käsermann, H.P., Bürki, K. (2007). Assessment of post-laparotomy pain in laboratory mice by telemetric recording of heart rate and heart rate variability. *BMC Veterinary Research* **3**, 16.

3. Jirkof, P., Cesarovic, N., Rettich, A., Nicholls, F., Seifert, B., Arras, M. (2010). Burrowing behaviour as an indicator of post-laparotomy pain in mice. *Frontiers in behavioral neuroscience* **4**, 165.

4. Van Loo, P.L.P., Kuin, N., Sommer, R. (2007). Impact of 'living apart together' on postoperative recovery of mice compared with social and individual housing. *Laboratory Animals* **41**, 441-455.

5. Langford, D.J., Bailey, A.L., Chanda, M.L. (2010). Coding of facial expressions of pain in the laboratory mouse. *Nature Methods* **7**, 447-452.

The Assessment of Pain Using Facial Expressions in Laboratory Rodents

Matt Leach

Comparative Biology Centre, Medical School, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK. Matthew.Leach@newcastle.ac.uk

Background

Pain in animals is of considerable public concern, particularly where animals are used in biomedical research. Pain can compromise not only animal welfare, but also the validity of scientific results. In order to alleviate pain, we need to be able to assess its severity and duration effectively [1]. If we cannot do this then we are unable to provide adequate pain management or to develop more effective and humane endpoints.

Behavioural indicators are increasingly being considered as effective cage-side indices of pain in many laboratory animal species, including rats [2], mice [3] and rabbits [4]. However, they have limitations:

- Behaviour may offer only a direct measure of an animal's response to the sensory afferent barrage (nociceptive input) rather than the emotional consequences of pain ('how pain makes animals feel') [5].
- Although behavioural indices are considered to offer an accurate and reliable assessment of pain in many species [6], there is still room for improvement.
- A recent study, investigating how we observe rabbits, showed that observers focus predominately on the face rather than on body areas where behavioural indices of pain are observed [7], reducing the effectiveness of behavioural indicators.
- The identification, validation and recording of behavioural indicators of pain is time consuming, which has led to pain behaviours only being identified for a very limited range of procedures in a small number of laboratory animal species.

The recent development of the Mouse Grimace Scale (MGS: [8]) and Rat Grimace Scale (RGS: [9]), which use facial expressions to assess pain, may overcome these difficulties. These studies demonstrate that mice and rats undergoing routine rodent nociceptive tests exhibit characteristic changes in facial expressions. Preliminary data from Langford et al. [8] raises the possibility that facial expression could indicate the affective component of pain in animals as it does in humans. Lesioning of the rostral anterior insula (implicated in the affective component of pain in humans) prevented changes in facial expression but not abdominal writhing (behavioural marker of nociception) in mice. The authors found both the MGS and RGS to be very accurate (72-97%) consistent and reliable (Interclass correlation: 0.9 respectively) both between and within observers. The assessment of pain using facial expressions should be less time consuming to apply than full behavioural scoring, allowing effective indicators of pain to be rapidly identified for a greater range of procedures. All of the indicators are located in one small area (i.e. the face), so exploiting the human tendency to focus on animal faces when assessing pain.

Assessment of facial expressions

In order to assess pain using facial expressions, images of mice and rats taken before and after potentially painful procedures are scored using the MGS and RGS respectively. The MGS is composed of 5 facial action units; orbital tightening, cheek bulge, nose bulge, ear position and whisker position (please see Langford et al. [8] for further details). The RGS is composed of 4 facial action units; orbital tightening, nose/cheek flattening, ear position and whisker position (please see Sotocinal et al. [9] for further details). Each facial action unit is scored on a 0 to 2 scale (0 = not present, 1 = moderately present, 2 = obviously present) by treatment-blind observers using mouse or rat grimace scale pictograms. A cumulative "grimace" score is then calculated for each image by simply adding the scores for each facial action unit that comprises the grimace scale.

Although the work of Langford et al. [8] and Sotocinal et al. [9] has demonstrated that mouse and rat facial expressions change in response to routinely used acute and chronic nociceptive tests, there has been little attempt to assess how facial expressions change in response to pain following routine surgical or other painful procedures. In addition, no attempt has been made to correlate changes in facial expression with the behavioural indices that we currently consider the most relevant to assessing post-operative and other post-procedure pain.

Assessment of post-surgical pain

We therefore have carried out a number of studies in rodents to determine whether routine surgical procedures are also associated with changes in facial expressions, and whether facial expressions are as effective for assessing post-surgical pain and analgesic efficacy as the established behavioural indices in mice and rats.

Study 1 aimed to determine if routine vasectomy in CD1 mice (n=18) induced changes in the MGS, and whether these could be used to effectively assess post-surgical pain and the efficacy of routinely used analgesia (20ml/kg saline [*sc*], 20mg/kg meloxicam [*sc*], 5mg/kg Bupivicaine [*li*]). In this study pain was assessed using MGS and validated behavioural indicators of pain [3]. The results showed that vasectomy did induce significant changes in the MGS, and that both the MGS and the scoring of pain behaviours identified clear differences between the pre- and post-surgery periods and between the animals receiving analgesia or saline post-operatively. Both these assessments exhibited a high positive correlation with each other. The MGS demonstrated high accuracy and reliability within and between observers.

Study 2 aimed to determine if routine laporatomy in Wistar (n=16) and Lister Hooded (n=16) rats induced changes in the RGS, and whether these could be used to effectively assess post-surgical pain and the efficacy of routinely used analgesia (0.4ml/kg saline [*sc*], 2mg/kg meloxicam [*sc*]). In this study pain was assessed using RGS and validated behavioural indicators of pain [2]. The results showed that laporatomy did induce significant changes in the RGS, and that both the RGS and the scoring of pain behaviours identified clear differences between the pre- and post-surgery periods and between the animals receiving analgesia or saline post-operatively. Both these assessments exhibited a high positive correlation with each other. The RGS demonstrated high accuracy and reliability within and between observers.

In order to reduce any unnecessary suffering in the saline treated groups (no analgesia), a rescue analgesia protocol was in place, where these animals received analgesia immediately after the data was collected (~ 1h post-surgery) to ensure that they went untreated for short a period as possible. The studies were conducted under licence from the UK Home Office and the research programme and protocols were reviewed and approved by the University of Newcastle’s local ethical review process.

Conclusions

The results of these studies suggest that the assessment of facial expressions offers a means of assessing post-surgical pain in rodents that is as effective as behavioural pain assessment. The assessment of pain using facial expressions was easy and rapid to carry out, required minimal training (less than 5 minutes), and provided a reliable and accurate means of assessing pain following vasectomy and laporatomy in mice and rats respectively. However, further research needs to be carried out into the effectiveness of facial expressions for assessing pain following different procedures and in different species.

References

1. Hawkins, P. (2002). Recognising and assessing pain, suffering and distress in laboratory animals: a survey of current practice in the UK with recommendations. *Laboratory Animals* **36,** 378–395.

2. Roughan, J.V., Flecknell, P.A. (2001). Behavioural effects of laparotomy and analgesics effects of ketoprofen and carprofen in rats. *Pain* **90,** 65-74.

3. Wright-Williams, S.L., Courade, J.P., Richardson, C.A., Roughan, J.V., Flecknell, P.A. (2007). Effects of vasectomy surgery and meloxicam treatment on faecal corticosterone levels and behaviour in two strains of laboratory mouse. *Pain* **130,** 108-118.

4. Leach, M.C., Allweiler, S., Richardson, C.A., Roughan, J.V., Narbe, R., et al. (2009). Behavioural effects of ovariohysterectomy and oral administration of meloxicam in laboratory housed rabbits. *Research in Veterinary Science* **87,** 336-347.

5. Flecknell, P.A. (2011). Affective state and quality of life in mice. *Pain* **152,** 963-964

6. Roughan, J.V., Flecknell, P.A. (2006). Training in behaviour-based post-operative pain scoring in rats - An evaluation based on improved recognition of analgesic requirements. *Applied Animal Behaviour Science* **96,** 327-342.

7. Leach, M.C., Coulter, C.A., Richardson, C.A., Flecknell, P.A. (2011). Are we looking in the wrong place? Implications for behavioural-based pain assessment in rabbits (*Oryctolagus cuniculi*) and beyond? *PloS ONE* **6,** 1-9.

8. Langford, D.J., Bailey, A.L., Chanda, M.L., Clarke, S.E., Drummond. T,E., et al. (2010). Coding of facial expressions of pain in the laboratory mouse. *Nature Methods* **7,** 447-449.

9. Sotocinal, S.G., Sorge, R.E., Zaloum, A., Tuttle, A.H., Marton, L.J., et al. (2011). The rat grimace scale: a partially automated method for quantifying pain in the laboratory rat via facial expressions. *Molecular Pain* **7,** 55.

Recognising and Assessing Positive Welfare: Developing Positive Indicators for Use in Welfare Assessment

Wanda McCormick

Moulton College, West Street, Moulton, Northamptonshire, UK. Wanda.mccormick@moulton.ac.uk

The use of positive welfare indicators: a proactive approach

Welfare assessment methods have developed greatly over recent decades with regards to both behavioural and physiological indicators. However, whether welfare is being assessed on the basis of biological function, affective state or ‘naturalness’, the emphases have remained heavily on identifying those key indicators of poor welfare or welfare concerns that require more immediate action, often via legislative change. The reactive approach in identifying welfare concerns has been the underpinning force behind legislative changes to mould animal management systems into more acceptable arrangements but as consumer interest in food sourcing grows, it is becoming increasingly apparent that acceptable is no longer enough and positive welfare indicators offer a future avenue for developing improved, ‘higher welfare’ systems in a more proactive manner [1]. The benefits of employing positive welfare indicators as assessment tools offer potential avenues of development for all animal sectors, not just within agriculture. Positive welfare indicators can fall into 3 categories: indications of contentment / pleasure, luxury behaviours, and behaviours that support ability to cope with challenge.

Indications of contentment / pleasure

Good welfare is not just an absence of negative experiences but also an ability to experience positive affective states. The ability of non-human animals to experience emotions in the same manner as humans remains a controversial topic, however, recent advances in neuroscience have been used to evidence the existence of positive affective states in animals with regards to behaviours such as positive anticipation [2].

Luxury behaviours

Luxury behaviours are often deemed to be key positive welfare indicators due to them being the first behaviours to be lost during challenging situations. As such, any occurrences have been assigned the ‘good welfare’ indicator label – if the animal is performing these behaviours then the situation must be of a sufficient welfare level to enable them to. These luxury behaviours have often been considered separate from those ‘fundamental’ behavioural needs [3] so commonly used to determine minimum requirements in welfare legislation. There is, however, a risk that by focusing on the more urgent ‘demand’ behaviours, a situation may never be achieved that is high enough for the luxury behaviours to occur. As the luxury behaviours may be indicative of less transparent long-term rewards, their absence could still result in negative or less favourable consequences for the animal involved. The identification of key luxury behaviours for different species could therefore become essential tools in the assessment of good rather than acceptable welfare, and behaviours such as play, allogrooming and certain vocalisations could become increasingly important [4-7]. In the absence of clear indicator luxury behaviours, it may be enough to assess behavioural diversity as a measure of welfare (e.g. using the Shannon-Weaver diversity index) by identifying those systems where a greater spread across the behavioural repertoire is encouraged, a method increasingly employed in zoo collections (McCormick & Melfi, unpublished) [8]. The use of behavioural diversity as an assessment tool can also help to eliminate the risk of over-recording luxury behaviours that are acting as self-rewarding mechanisms in stressful situations.

Behaviours that support the ability to cope with challenge.

It may not be enough to simply identify the presence of behaviours indicative of positive emotions, but focus may also be required on identifying the existence of systems that allow animals to cope. Management systems do not need to be absent of any potential stressors – in fact this could be argued to result in a lack of stimulation – but should enable animals to cope with these acute events in a way that prevents a chronic stress situation from occurring. As such, the study and assessment of positive indicators of welfare should also include those behaviours / situations that allow coping. Studies in rats have identified that negative affective states are associated with faster startle responses and reduced anticipatory behaviours [1,9] but investigations into the situations that link positive affective states to coping are limited.

Previous social studies in domestic cattle have focused on the negative perspectives, i.e. dominance and aggression, but no work had been conducted on the potential positive benefits of social bonding. A large, dynamically managed, dairy cattle herd was studied to establish the existence of preferred partners according to proximity scorings. During these initial stages, it was established that heifers (52%, n=34) were significantly more likely to have one or more preferred partners compared to cows (32%, n=12) ($X^2 = 8.210$, df=1, p=0.004) with relation to the probability of being associated with a particular individual – although it was suggested that this could have indicated that experience of repeated regroupings in cows had caused a dissolution of previously formed bonds.

Following stage 1, commercial regrouping practices continued to affect the possible existence of preferred partnerships. During stage 2, six focal animals, previously identified as having a preferred partners (PP) that they were still housed alongside were separated off from the rest of the group for 30 minutes in a holding pen, either with their preferred partner or with a random individual from within the group. Three separation conditions were studied: 1. Focal animal (A) separated with their PP (B); 2. Focal animal (A) separated with a random individual (C) whilst B is released back to the home group; 3. Animal B separated with a random individual C whilst focal animal A is released back into the group. The heart rates of focal animal A during stage 1 and stage 2, and animal B during stage 3, were monitored using a Polar (Protrainer 5 Equine RS800) heart rate monitor, adapted for use in cattle, at 15 sec intervals. All protocols were approved by a Moulton College ethical committee. Heart rates differed significantly between stages (P=0.001) with focal animals (A) having lower heart rates when separated with their preferred partner (80.2 BPM) compared to being separated with a random individual (82.6 BPM). Interestingly, the partners' heart rates when with a random individual (79.9 BPM) were lower than the focals' in either situation. Although the actual difference in heart rates was small, the distinction was clear and this opens up further areas requiring research with regards to threshold levels beyond which chronic stress becomes physiologically damaging. There is also potential that the presence of any familiar cow could provide some degree of support but that the distinction lay within the level of familiarity.

From these results it can be suggested that some individuals seem to benefit from having their preferred partner present during a potentially stressful situation. It could also be suggested that some individuals are more sensitive to stressful situations compared to others. When developing / assessing dairy management systems, those that support the formation of social bonding and subsequently manage animals in a way that these bonds can be maintained may in turn provide a higher welfare system by improving the coping ability of individuals.

Conclusions

Much emphasis has been placed on identifying indicators of poor welfare but the effective assessment of good welfare is an area gaining momentum both within academia and industry [10]. A wide range of behavioural methods exist that can be utilised as good welfare indicators in all animal management systems but the subtle nature of these indicators requires rigorous validation and development to create systems that can easily be employed by users on site.

Acknowledgements

This presentation is sponsored by Noldus Information Technology. However, the content of the presentation is independent of Noldus.

References

1. Boissy, A., Manteuffel, G., Jensen, M.B., Oppermann, Moe. R., Spruijt, B., Keeling, L.J., Winckler, C., Forkman, B., Dimitrov, I., Langbein, J., Bakken, M., Veissier, I., Aubert, A. (2007), Assessment of positive emotions in animals to improve their welfare. *Physiology & Behaviour* **92**, 375-397.

2. Spruijt, B.M., van den Bos, R., Pijlman, F.T. (2001). A concept of welfare based on reward evaluating mechanisms in the brain: anticipatory behavior as an indicator for the state of reward systems. *Applied Animal Behaviour Science* **72**, 145-71.

3. Jensen, P., Toates, F.M. (1993). Who needs "behavioral needs"? Motivational aspects of the needs of animals. *Applied Animal Behaviour Science* **37**, 161-181.

4. Donaldson, T.M., Newbwerry, R.C., Spinka, M., Cloutier, S. (2002). Effects of play experience on play behavior of piglets after weaning. *Applied Animal Behaviour Science* **79**, 221-231.

5. Thornton, P.H., Waterman-Pearson, A.E. (2002). Behavioral responses to castration in lambs. *Animal Welfare* **11**, 203-212.

6. O'Connell, J., Giller, P.S., Meaney, W. (1989). A comparison of dairy cattle behaviour patterns at pasture and during confinement. *Irish Journal of Agricultural Research* **28**, 65-72.

7. Schön, P.C., Puppe, B., Manteuffel, G. (2004). Automated recording of stress vocalization as a tool to document impaired welfare in pigs. *Animal Welfare* **13**, 105-110.

8. Clark, F.E., Melfi, V.A. (2011). Environmental enrichment for a mixed-species nocturnal mammal exhibit. *Zoo Biology* **29**, 1-17.

9. von Frijtag, J.C., Reijmers, L.G., van der Harst, J.E., Leus, I.E., van den Bos, R., Spruijt, B.M. (2000). Defeat followed by individual housing results in long-term impaired reward-and cognition-related behaviors in rats. *Behavior Brain Research* **117**, 137-146.

10. Ohl, F., van der Staay, F.J. (2012). Animal welfare: At the interface between science and society. *The Veterinary Journal* **192**, 13-17.

Measuring Behavioural Changes to Assess Anthropogenic Noise Impact in Adult Zebrafish (*Danio rerio*)

H. Slabbekoorn

Institute of Biology, Leiden University, Leiden, The Netherlands. H.W.Slabbekoorn@Biology.LeidenUniv.nl

Sounds underwater

Underwater habitats are full of natural sounds from abiotic sources, such as wind and water flow, and biotic sources, such as chorusing fish and snapping shrimps. However, human activities elevate these ambient noise levels artificially. The so-called anthropogenic noise from traffic, industry, and recreational activities concerns an only recently recognized pollution factor underwater which is expanding in time and space [1]. The artificially high noise levels are typically of relatively low frequency and can cause many different problems such as physical damage in cases of extreme overexposure, physiological stress, and auditory masking, which are all factors that an animal is more or less passively undergoing (see Figure 1). Furthermore, anthropogenic noise may lead to spatial deterrence, behavioural interruption, and signal modifications, which are all factors that involve some sort of active response from the animal. The three passive factors are typically positively related: if one is getting worse the others are likely to follow the same pattern. An increase in any of the three passive factors also increases the probability that any of the three active factors occurs. Vice versa the opposite may be the case: the active factors have the potential to provide relief on the passive factors by reducing the level and duration of exposure.

Many fish species generate sounds to be heard by conspecifics for communication about for example competition for resources and attraction of social or mating partners. Probably even more common is the use of hearing abilities to find prey or to detect predators. Furthermore, the geographic distribution of habitat-specific abiotic and biotic sound sources provides cues to fish for orientation and localization of specific areas for migrating, feeding, hiding, or spawning [2]. The majority of these biologically relevant sounds are relatively low in frequency, within the range of hearing sensitivity for most fish species, but also overlapping dramatically with the typical spectrum of anthropogenic noise. Therefore, the widespread occurrence of artificially elevated noise levels due to human activities has the potential to mask these biological sounds and affect the behaviour of many acoustically dependent fish. Besides masking, a large impact may also result from interruption of natural

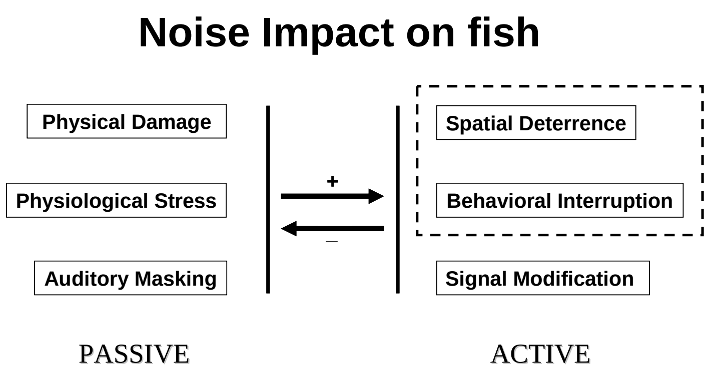


Figure 1. Schematic overview of six dominant factors that may play a role in assessing the impact of anthropogenic noise on fish welfare and fitness. The arrows indicate the relationship between factors. The ‘passive’ factors on the left reflect consequences of undergoing a particular exposure, which typically also increase the probability of occurrence of the ‘active’ factors on the right. The ‘active’ factors, in contrast, involve behavioural decisions by the individual that are likely to reduce the severity of the ‘passive’ factors. The purpose of this subdivision into categories is just to emphasize the complexity of noise impact assessments as, although the overall representation is generally true, the relative importance of factors and the magnitude and even direction of the relationships will vary per exposure level and duration, per species, and with physiological and environmental conditions. The dashed line block demarcates the focus of the reported zebrafish studies.

behaviours due to startle responses, attentional disruption, and lower efficiency of acoustically guided behaviours. Such impact is depending on the nature of fluctuations in the noise background and the nature of noise events. Although there are still only few data papers, all of these noise effects may negatively affect growth, reproductive success, and survival or cause fish to escape noisy areas.

Behaviour and stress

Behavioural changes related to artificial noise exposure may have detrimental effects by themselves but may also be associated with stress: physiological changes that yield energetic costs and have negative effects on growth, reproduction, or survival. Several studies have shown an impact of extreme anthropogenic noise exposure on, for example, increased heart rate and elevated cortisol levels [3]. However, although common and more moderate exposures that lead to behavioural changes will be accompanied by underlying physiological changes, these should not all necessarily be regarded as stress. There are general physiological changes related to the energetic demands and the state of mind required for making and executing adequate behavioural decisions to external stimuli. Different kinds of stimuli may for example raise interest, arousal, fear, or anxiety [4]. The associated physiological changes are an integral part of the natural regulatory capacity of healthy fish and are not necessarily different between stimuli or states that we would intuitively regard as positive or negative. Nevertheless, especially exposure to unpredictable and uncontrollable noise events and biologically relevant sounds that are critical for survival and become unpredictable and uncontrollable due to masking noise may yield physiological changes that deserve the label of stress [5].

Zebrafish as a high-throughput model organism

The zebrafish (*Danio rerio*) is emerging as a new model organism and rapidly gains popularity in a variety of disciplines [6]. They earn this popularity by practical, financial, and ethical advantages over other vertebrates, such as rodents and primates. They are relatively easy and cheap in maintenance and reproduce readily in captivity. Importantly, zebrafish also share many physiological similarities with terrestrial vertebrates. They are highly similar and comparable in the structure and function of neuro-chemical and behavioural systems as well as in the general organization of their stress-regulating systems. They have proven to be suitable for laboratory studies on behavioural assessments of anxiety, while technological advances now allow high-throughput drug screening and discovery.

One of the most commonly used paradigms to test behavioural changes in adult zebrafish is the novel environment test [7]. A single or a small group of individual fish is transferred to a new tank in which the behavioural response to the novel environment (and novel social conditions) can be recorded on video and monitored manually or automatically. Typical behavioural assessments include: startle responses and erratic swimming movements, swimming speed, accumulated swimming distance, group cohesion, freezing bouts, delay to first entry of upper half or upper third of the fish tank, and number of entries and total time spent in that same area relatively close to the surface. The impact of a particular treatment is evaluated by whether it yields an incline or a decline in the anxiety-related responses. Stronger anxiety is for example related to faster swimming, stronger group cohesion, and staying away from the surface and closer to the bottom for longer [8].

Noise impact assessment in zebrafish

Although it does not apply to all noise exposure related to human activities, there is certainly potential for unpredictable and uncontrollable sound events that could yield physiological stress. The occurrence of anthropogenic noise in the natural environment is characterized by variety in time at various scales. There are more or less continuous noises from vessels, pumping systems, windmill farms, and gas extraction platforms. There are also repetitive sounds from pile driving, sonar use, and seismic surveys, and there are very brief but loud noise events related to explosions. Furthermore, there are many noise types that are in between these temporal extremes, such as sounds generated by dredging, water scooters, boats changing gear, and general construction activities in or close to the water. Very few noise impact studies have addressed such temporal variety, although it likely plays a critical role in the potentially negative effects of noise on fish. Neo et al. [9]

Figure 2. Picture of the two adjacent fish tanks connected by a pvc-pipe swim tunnel for the second experiment (see main text) of Neo et al. (unpublished). Both fish tanks are acoustically insulated from each other and their surroundings of the office building by air, four rubber shock pads each, and a layer of Styrofoam between the fish tanks and two separate trolley tables on rubber wheels. Water flows into the right tank through the tunnel to the left tank and out of the system on the left, but the circulation is stopped during noise exposure trials. There is an underwater speaker on the end of each fish tank opposite the tunnel for passage. Both speakers were used in alternating sequence: one speaker broadcasting artificial noise in one trial and the other in the next trial to avoid side effects.

investigated the disruptive effects of continuous noise as well as noise pulses on the behaviour of captive zebrafish and also compared the effects of slow and fast pulse rates, and predictable and unpredictable pulse intervals.

Neo et al. [9] assessed the impact of temporal variety on fish behaviour in two different experiments. In the first experiment, they used moderate exposure levels of about 112 dB re 1 μPa in a single tank without acoustic escape possibility. In the second experiment, they used higher exposure levels up to 140 dB re 1 μPa in a double-tank system with acoustic escape possibility. The first experiment revealed that noise pulses of moderate noise level could already alter behaviour of a small group of five zebrafish. Noise exposure through an in-air external speaker generated a relatively homogenous underwater field of sound pressure conditions, which resulted in changes for several behavioural measures such as swimming speed, group cohesion, and tendency to move up to the surface. Furthermore, the impact turned out to vary significantly among exposure regimes. The second experiment revealed that noise conditions can be made distinct in two adjacent fish tanks connected by a pvc-pipe swim tunnel of 35.0 cm in length and 12.5 cm in diameter (see Figure 2). Noise exposure through one of two underwater speakers generated sound pressure gradients in the left or right fish tank, while leaving the sound levels in the other close to baseline levels. Frequent tunnel passages by six individual zebrafish allowed testing of spatial avoidance of the noisy fish tank. The experiments were only carried out after ethical evaluation and approval by the Animal Experiments Committee of Leiden University (DEC# 10069).

Implications and extrapolations

Variable impact of different temporal patterns of experimental noise exposure on captive fish may have implications for noise impact assessments in natural environments, although one has to be cautious with extrapolations. I believe there is for example value in the relative differences of impact on anxiety-like behaviour for different noise exposure regimes. The variation in e.g. startle threshold levels, duration required for habituation, or effects on spatial avoidance, as assessed in captivity in response to different temporal patterns of artificial noise exposure, may also all apply to natural conditions. However, studies comparing such behavioural measures during experimental noise exposure in captivity and in natural water bodies are needed to confirm this statement. Absolute values for certain thresholds from inside noise exposure testing are almost certainly not likely to be very useful outside. This is due to the fact that inside and outside conditions are inherently very different, both with respect to the acoustics and with respect to the behavioural and physiological state of the fish.

Noise impact conditions in the wild differ from those in captivity as the participation in natural activities such as migration, feeding, or spawning, may dramatically alter perceptual threshold levels as well as the physical impact. The same is true for the continuous possibility of a predator attack in the wild compared to the complete safety in the laboratory fish tank. Furthermore, differences in absolute values and even in relative effects of different temporal patterns may vary by species, age, sex, and season. Neo et al. [9] focussed for example on sound pressure levels without assessing the exposure characteristic in the fish tank in terms of the particle motion component of sound. This may be right for adult zebrafish, which are sound pressure sensitive through the pressure-to-motion transduction by the so-called Weberian ossicles, which are connecting the swim bladder to the inner ear. However, juvenile zebrafish and many other species without these specialized adaptations for hearing are more motion sensitive and will experience noise conditions differently. Particle motion is more difficult to assess than sound pressure and also a more complex feature of sound close to the source and in the confined area of a fish tank.

In addition to the possibilities to extrapolate indoor insights to the reality of the outside world, it is also important to realize that for many fish the reality is inside. Many fish are kept in captivity by hobbyists at home, by zoo keepers in public aquaria, by professional breeders in aquaculture, and by scientists in research facilities. Typically, good health conditions are key to the purpose of keeping fish by these very divergent groups of people. They are often interested in the beauty of colour, body forms, and behavioural display, optimal growth and reproduction, or repeatable patterns of semi-natural behaviours or stereotypic responses to artificial stimuli. Nevertheless, noise conditions in fish tanks are customarily not consciously controlled and can be very high (e.g. through pumping systems) and unpredictable (e.g. due to knocking visitors). Habituation to continuous presence or habituation to repeated and predictable exposure can bring some relief, but still we are currently most often ignorant of the potentially detrimental effects of anthropogenic noise on the well-being of fish and on reaching the specific targets of the fish keepers in terms of the body condition and behavioural state of their fish.

In conclusion, there is a huge potential for fundamental studies on noise impact in fish with lots of opportunities for applied value. Fish tank studies in the laboratory are currently most accessible and they benefit significantly from the advantages of having the zebrafish as an ideal model organism and from the technological advances of measuring and processing behaviour automatically [6]. However, in parallel to the indoor studies we also need to go outside. We need to learn more about the role of sound in natural activities of fish and investigate the potentially negative effects of anthropogenic noise [1]. We can gain important insights through observational studies at noisy locations as well as through experimental exposure studies at locations that are still relatively unaffected acoustically by human activities. I hope future studies will be able to integrate studies from the laboratory and from the field to exploit the best of both worlds: the possibilities to assess effects of anthropogenic noise on fish in great detail and in large numbers in captivity with the ecological relevance and acoustic reality of the natural world.

References

1. Slabbekoorn, H., Bouton, N., van Opzeeland, I., Coers, A., ten Cate, C., Popper, A.N. (2010). A noisy spring: the impact of globally rising underwater sound levels on fish. *Trends in Ecology and Evolution* **25**, 419-427.

2. Fay, R.R. (2009). Soundscapes and the sense of hearing of fishes. *Integrative Zoology* **4**, 26-32.

3. Wysocki, L.E., Dittami, J.P., Ladich, F. (2006). Ship noise and cortisol secretion in European freshwater fishes. *Biological Conservation* **128**, 501-508.

4. Luca, R.M., Gerlai, R. (2012). In search of optimal fear inducing stimuli: Differential behavioral responses to computer animated images in zebrafish. *Behavioural Brain Research* **226**, 66-76.

5. Koolhaas, J.M., Bartolomucci, A., Buwalda, B., de Boer, S.F., Flügg, G., Korte, S.M., Meerlo, P. Murison, R., Olivier, B., Palanza, P., Richter-Levin, G. Sgoifo, A., Steimer, T., Stiedl, O., van Dijk, G., Wöhr, M, Fuchs, E. (2011). Stress revisited: A critical evaluation of the stress concept. *Neuroscience and Biobehavioral Reviews* **35**, 1291-1301.

6. Champagne, D.L., Hoefnagels, C.C.M., de Kloet, R.M.E., Richardson, M.K. (2010). Translating rodent behavioral repertoire to zebrafish (*Danio rerio*): Relevance for stress research. *Behavioural Brain Research* **214**, 332-342.

7. Stewart, A., Gaikwad, S., Kyzar, E., Green, J., Roth, A., Kalueff, A.V. (2012). Modeling anxiety using adult zebrafish: A conceptual review. *Neuropharmacology* **62**, 135-143.

8. Egan, R.J., Bergner, C.L., Hart, P.C., Cachat, J.M., Canavello, P.R., Elegante, M.F., Elkhayat, S.I., Bartels, B.K., Tien, A.K., Tien, D.H., Mohnot, S., Beeson, E., Glasgow, E., Amri, H., Zukowska, Z., Kalueff, A.V. (2009). Understanding behavioral and physiological phenotypes of stress and anxiety in zebrafish. *Behavioural Brain Research* **205**, 38-44.

9. Neo, Y.Y., Parie, P., Bakker, F., Snelderwaard, P.' Slabbekoorn, H. (unpublished). Artificial noise exposure affects zebrafish behaviour but not their spatial distribution.

Automated Assessment of Animal Health and Wellbeing

J.E. van der Harst[1], B.M. Spruijt[2]

[1,2]Delta Phenomics, Utrecht, The Netherlands; Department of Biology, Utrecht University, Utrecht, The Netherlands; SenseWell consortium[i]

[1]johanneke.vanderharst@deltaphenomics.com

Health and wellbeing are issues of major importance in both animal husbandry and laboratory animal science. Animal health and (mental) wellbeing have suffered dramatically with increases in the size of farms and the number of animals kept per farm. Public support for intensive livestock farming has decreased as a result, and so innovation with respect to more welfare-friendly husbandry is urgently needed. Similarly, the use of animals for research purposes is a major concern for many people, and objective and transparent welfare monitoring of laboratory animals could help to reassure the public that suffering is being effectively recognised and alleviated. It is also important to monitor and improve laboratory animal welfare, since scientific data will be more relevant if the physical and mental condition of the subjects can be improved.

Animal welfare is receiving growing attention, in the form of European Union research initiatives (e.g. www.welfarequality.net) and EU legislation that requires reliable and objective monitoring of animal welfare on a large scale (e.g. Directive 2010/63/EU[ii] that regulates laboratory animal care and use). Objective tools to assess animal wellbeing and health are essential in order to help meet societal demands for better animal welfare in both food production and research.

Effective early detection of animal welfare problems can require automated observation of behaviour, including vocalisations and physiological parameters, if the number of animals in a production farm or animal research facility is too large in relation to the number of staff, and available experts on animal behaviour.

A consortium of scientific and technical partners is currently working on the 'SenseWell' project, which aims to develop an automated system for the early detection of symptoms of disease and chronic stress, and also of positive indicators of wellbeing. This way, animal health and wellbeing can be assessed 24/7, which supports the responsible management of large numbers of animals. This system will consist of an intelligent sensor network that is able to measure behaviour, vocalizations and physiology of freely moving subjects. It is based on recent advances in sensor technology, image processing, data fusion, acoustical signal analysis, pattern recognition and software architecture design. The outcome of this project will be a platform technology that is suitable for both behavioural research and welfare monitoring, which could make significant contributions to both animal welfare and the quality of the science.

In the SenseWell project, rats are used as a model for laboratory animals in general as well as for farmed animals such as pigs. Animals were housed in a large home cage environment (PhenoTyper® 9000, 90x90cm), which was developed in line with the PhenoTyper concept (Noldus Information Technology, Wageningen, The Netherlands), in which animals can be monitored for periods of several hours, for several days or weeks, in a familiar environment without handling or other disturbance. This large, home cage monitoring environment was designed specifically to further increase the scientific relevance of home-cage monitoring from an ethological perspective. That is, animals can express more dynamic movement patterns and therefore effects on distance and velocity of movement can be larger between groups / subjects of investigation. Furthermore, in a larger automated home cage, animals can be tested under social conditions, with more animals in a cage, which is more applicable to real-life situations in both laboratory animal housing and farming. Importantly, by the increased surface of the monitoring environment, dynamics in social interactions and structures can be amplified, in the sense that approach and avoidance between individuals can be expressed within a broader range.

The large home cage monitoring environment (PT-9000) contains bedding, 1 or 2 large shelters, food and water, and is equipped with a camera and infrared lighting. Video images are analysed in real-time using tracking software (EthoVision® XT, Noldus Information Technology). The PT-9000 is further equipped with microphones to measure (ultrasonic) vocalisations (SonoTrack®, Metris BV) and can be further extended with telemetry receivers for wireless monitoring of physiological parameters such as heart rate and body temperature (PhysioLinq®, Telemetronics Biomedical BV). The latter system is specifically suitable for simultaneous measurements of a group of animals since individual signals can be identified. The main goal of the SenseWell project is to integrate all 3 systems and their data streams to create a 24/7 animal monitoring platform with reference-values for future support of early detection of welfare problems including automatic notifications. The results can be translated to larger animals and find their way in the design of new automated monitoring tools and advanced housing systems for farm animals.

By using different animal models and protocols, evoking specific responses such as stress, play and reward-anticipation, and simultaneous monitoring of behaviour, (ultrasonic) vocalisations, heart rate and body temperature, the aim is to construct a reference library of automatically generated (sets of combined) parameters that can be used as indicators of both positive and negative animal health and wellbeing.

Recent results and a current update on the status of the above described project will be presented.

Ethical statement

Experiments were performed according to the legal requirements of The Netherlands concerning research on laboratory animals (Wod/Dutch Animal Welfare Act) and have been approved by an Animal Ethics Committee ('Lely-DEC'). All procedures complied with the regulations controlling animal experimentation within the EU (European Communities Council Directive 86/609/EEC).

---

[i] The SenseWell consortium consists of the following academic and commercial partners: Noldus Information Technology BV (L.P.J.J. Noldus); Metris BV (R.J.A. Bulthuis); Telemetronics Biomedical BV (G.J. van Essen); BioMetris, Wageningen University (G. van der Heijden); Faculty of Veterinary Medicine, Utrecht University (J.A. Stegeman); EEMCS, TU Delft (P.J. French); Delta Phenomics BV (J.E. van der Harst); Faculty of Science, Utrecht University (B.M. Spruijt).

[ii] http://ec.europa.eu/environment/chemicals/lab_animals/home_en.htm

SPECIAL SESSION

Fish in Behaviour Research: Unique Tools with a Great Promise!

Chair: Robert Gerlai

Department of Psychology, University of Toronto Mississauga
Ontario, Canada. robert_gerlai@yahoo.com

Abstract

Zebrafish are becoming increasingly popular in behavioural brain research as this species offers good translational relevance and has several genetic and biological techniques developed for it. Other species of fish may also have utility in behavioural research as their analysis can not only shed light on interesting species-specific characteristics but also allow the investigator to examine complex behavioural phenomena in a simpler, evolutionarily older "design", and thus make conclusions about how our own behaviour may have evolved. The current symposium will have five speakers. Dr. Gerlach will talk about olfactory imprinting in coral reef fish and zebrafish and will discuss both the ecological and neurobiological aspect of her research. Dr. Agrillo will review findings on numerical abilities in fish presenting data obtained with the use of behavioural paradigms developed for fish species. These paradigms include spontaneous choice tasks as well as operant conditioning methods. Ms. Buske will describe how social behavior develops in zebrafish and how dynamic changes in shoaling may be quantified using video-tracking systems. Dr. Cretton will present his novel hardware and software solutions as to how to quantify larval zebrafish behavior including anxiety-like responses. Dr. Merlin will present his research on modeling and analyzing attention deficit hyperactivity disorders using zebrafish. Last, Dr. Gerlai will examine social affiliation, and how one can induce group cohesion using computer animated stimuli. He will also present data on the neurobiological mechanisms of group cohesion in zebrafish.

SPECIAL SESSION CONTENTS (sorted by paper ID)

Social Affiliation in Zebrafish: From Synthetic Images to Biological Mechanisms

Robert Gerlai

University of Toronto Missisauga, Department of Psychology, Mississauga, Ontario, Canada.
robert_gerlai@yahoo.com

The zebrafish is an increasingly popular species in behavioural neuroscience

Zebrafish are small freshwater fish that inhabit the tropical waters (small lakes, rivers and streams) of Southern and Eastern portions of Asia, e.g. India and Nepal. This species has been becoming increasingly popular among behavioural neuroscientists for many reasons (e.g. [4, 15]). First, due to the past four decades of extensive developmental biology research conducted with this species, powerful genetic tools have been developed for it that now may be utilized by several other disciplines of biology, including brain and behaviour research. Second, the zebrafish appears to strike an optimal compromise between system complexity (it is a typical vertebrate with basic neuroanatomy, neurophysiology, neurochemistry, and biochemistry similar to those of higher order vertebrates including mammals) and practical simplicity (it is small, easy and cheap to maintain in the laboratory, and it is highly prolific) [2]. Third, due to these above features and because numerous genes identified in zebrafish have been found to have nucleotide sequences highly homologous to mammalian counterparts, the zebrafish is believed to be an excellent translational tool, i.e. it may be utilized for modeling and analysing human diseases [5].

The zebrafish is a highly social species

Both in nature and in the laboratory zebrafish form tight groups, called shoals [7-10]. A shoal is defined as an aggregate of individuals in which group members remain in close proximity to each other. In nature, this behaviour is believed to serve several adaptive functions, including protection from predators, increased foraging efficiency and increased mating success. In the laboratory, this behaviour is also prevalent and may be quantified both in freely moving groups of fish [7, 8] as well as in situations where a single subject responds to the presence of social stimuli [6]. The prevalent and robustly inducible shoaling behaviour in zebrafish may serve as an excellent tool with which one can investigate the mechanisms of vertebrate social behaviour. For such investigation to be successful, however, the stimuli that trigger this affiliative response must be well known and their method of delivery must be standardized.

The zebrafish is a diurnal species that has excellent vision.

Zebrafish, similarly to our own species and unlike rodents, are diurnal and are active during the day and sleep at night. Although zebrafish use several modalities for perceiving stimuli in their environment, e.g. lateral line, olfaction and auditory perception, these fish also have excellent vision and respond to visual cues. The zebrafish visual system and its development and physiology have been well characterized and we know, for example, that this species has four different cones (tetrachromatic) to detect specific wavelength ranges of light, i.e. their colour vision is excellent. The practical aspect of good visual perception abilities of zebrafish is that the experimenter may be able to use visual cues to affect and manipulate the behaviour of these fish and, given that our own species also uses vision as one of the primary mode of perception, simple consumer grade "visual devices", such as cameras and TV screens are available for a reasonable cost for such research.

Visual stimuli to induce social responses

We have started the analysis of visual stimuli and explored if they induce social responses. We first investigated how zebrafish respond to live fish in a test where the zebrafish and the stimulus fish were allowed to swim freely [11]. We found that zebrafish prefer staying close to conspecifics and do not shoal with heterospecific fish irrespective of whether these heterospecific fish are social (shoaling) or non-social (not group forming) species. We also found that zebrafish shoal with (i.e. prefer staying close to) conspecifics even if the colour of the conspecific differed from their own (we used a pigment mutant “gold” variant of zebrafish to investigate this question). Subsequently, we examined if zebrafish may be sensitive to particular features of conspecifics and started modifying such features using computer animated images [11]. The results suggested that zebrafish respond to moving images of conspecifics just as they would respond to real live fish. The results also revealed that zebrafish are not very sensitive to the pattern of the conspecifics, i.e. in a binary choice test they preferred images that did not have the wild type horizontal stripe pattern equally to those that showed the normal wild type horizontal stripe pattern. Furthermore, the test subjects also showed equal preference to the wild type pattern and images that exhibited vertical stripes. Also, zebrafish responded normally to images that were coloured red. Interestingly, experimental zebrafish preferred yellow coloured images to the wild type colour (blue, green and yellow stripes). Last, zebrafish showed equal preference towards a shortened (and “fatter”) image and the wild type body proportions, but did show a robust avoidance of the opposite, a lengthened and narrower fish image. These results suggest that computerized animated images may be utilized to induce social behaviour and that the decomposition of what zebrafish may regard as species specific stimuli representing their conspecifics will be possible.

Quantification of social responses

In this presentation I focus on quantification of shoaling behaviour, i.e. social affiliative responses. These responses may be induced and quantified in principally two different contexts. One, the behaviour of freely moving fish may be measured [8] or two, the behaviour of a single subject to social stimuli may be recorded [6]. In the former case, one can track all subjects of the freely moving shoal and measure numerous parameters that define shoal cohesion. For example, the average inter-individual distance measures the distance between a particular fish from all of its shoal members and calculates the average (mean) of these distances. Similarly, the variance of inter-individual distances may also be calculated for each shoal member. These measures are dependent upon the shoal size, i.e. how many members form the shoal, but they are highly informative as they take all distances between all possible pairs of fish in the shoal into account. Another measure often used in the quantification of group forming behaviour in fish, birds and other species that form aggregates, is the nearest neighbour distance. This measure is independent of shoal size, which is an advantage if one wants to compare social behaviour across a range of shoal sizes. However, it does not take into account all distances within the shoal and thus it may be less informative compared to the inter-individual distance measure. The second context in which one may quantify shoaling responses is with a single subject. In this case, a social stimulus, live stimulus fish or computer animated social stimuli, may be presented to the experimental subject adjacent to its test tank and, for example, the subject’s distance from the stimulus is quantified [6].

The sight of conspecifics is rewarding

Using visual stimuli of conspecifics one may be able to trigger robust social responses [6]. These responses do not need to be trained and appear spontaneously in response to the social stimulus. It appears therefore that zebrafish are innately motivated to perform these responses. Briefly, social stimuli may be rewarding [3]. Indeed, this is what we found in a learning paradigm in which zebrafish were required to associate a visual cue (a red cue card) with the presence of conspecifics (in this case live stimulus fish) [1]. Zebrafish were found to be able to acquire the CS (cue card) US (stimulus fish) association within 20 trials and showed robust preference (conditioned response) to the CS alone when it was presented to them at a probe trial, a results that confirmed that the US (the conspecific stimulus fish) was a strong reinforcer. Given that this task was a typical appetitive conditioning task, we conclude that the sight of conspecifics is rewarding in zebrafish. A similar conclusion could be drawn from subsequent learning paradigms in which the mechanisms of associative learning were investigated with zebrafish [12-14]. In these studies, we found that the presence of conspecifics could provide sufficient drive (motivation) to learn both associative and spatial tasks and that learning performance in these tasks was NMDA-Receptor dependent, as a selective NMDA-R antagonist, MK801 could disrupt learning and memory at concentrations that did not affect motor function and/or vision.

The dopaminergic system, reward, and the sight of conspecifics

The dopaminergic system is known to mediate reward, among other functions, in mammals. Dopamine receptors (D1-R, D2-R D3-R and D4-R) have been identified in zebrafish and the nucleotide sequence of the corresponding genes has been found highly homologous to that of mammalian counterparts. We have started the mechanistic analysis of the whether the dopaminergic system is involved in social behaviour in zebrafish. First, we investigated whether presentation of conspecific images may induce changes in the levels of dopamine and DOPAC (a metabolite of dopamine). We also analyze potential changes in other neurochemicals. This analysis is still on-going but so far we have found unequivocal evidence showing that dopaminergic responses are enhanced by the sight of conspecifics, and in a manner that correlates with the length of the exposure to this visual stimulus. We also know that 6the stimulus must have certain features resembling the conspecifics as scrambled images or the testing procedure without conspecific images do not induce the dopamine responses. Interestingly, the serotoninergic system does not appear to be engaged by the presentation of social stimuli. Subsequently, we examined if disruption of the dopaminergic system may affect responses to social stimuli. We utilized a D1-R antagonist (SCH23390). D1-R is the most abundantly expressed post-synaptic dopamine receptor subtype in zebrafish. We found the antagonist to significantly impair social behavioural responses (to increase the distance of the treated experimental zebrafish from the stimulus screen on which animated images of conspecifics were shown) without affecting motor function and/or perception. Taken together the above results suggest that dopamine is likely to mediate the rewarding aspect of shoaling.

Concluding remarks

Zebrafish behavioural neuroscience is in its infancy. The number of zebrafish behavioural tests is orders of magnitude smaller than what is available for laboratory rodents, the rat and the house mouse. Nevertheless, the past few years have seen an exponential increase in the number of zebrafish behavioural studies. Our knowledge of zebrafish behaviour is rapidly increasing and our tools with which we can investigate this species from a behavioural perspective are continuously improving. Given the translational relevance of this fish and its practical advantages, I believe zebrafish will become one of the primary model organisms of biomedical research. The above studies represent just one small subset of this field. But I hope they demonstrate the utility of how one may be able to employ zebrafish in the analysis of such complex functions of the brain, as for example, social behaviour.

Acknowledgements

I would like to thank Dr. Noam Miller, Margarette Sison, Christina Saverino, Lina Al-Imari, Muhammed Saif, Christine Buske, Dr. Diptendu Chatterjee. Supported by NIH/NIAAA (USA) and NSERC (Canada).

References

1. Al-Imari, L., Gerlai, R. (2008). Conspecifics as reward in associative learning tasks for zebrafish (*Danio rerio*) *Behavioural Brain Research* **189**, 216-219.

2. Chatterjee, D., Gerlai, R. (2009). High Precision Liquid Chromatography Analysis of Dopaminergic and Serotoninergic Responses to Acute Alcohol Exposure in Zebrafish. *Behavioural Brain Research* **200**, 208-213.

3. Gerlai, R. (2011) Associative Learning in Zebrafish (*Danio rerio*). *Methods in Cell Bioology* **101**, 249-270.

4. Gerlai, R. (2011). A small fish with a big future: Zebrafish in Behavioral Neuroscience. *Reviews in the Neurosciences* **22**, 3-4.

5. Gerlai, R. (2010). High-throughput Behavioral Screens: the First Step towards Finding Genes Involved in Vertebrate Brain Function Using Zebrafish. *Molecules* **15**, 2609-2622.

6. Gerlai, R., Chatterjee, D., Pereira, T., Sawashima, T,, Krishnannair, R. (2009). Acute and Chronic alcohol dose: Population differences in behavior and neurochemistry of zebrafish. *Genes, Brain and Behavior* **8**, 586-599.

7. Miller, N., Gerlai, R. (2011a). Redefining membership in animal groups. *Behavior Research Methods* **43**, 364-371.

8. Miller, N., Gerlai, R. (2011b). Shoaling in Zebrafish: What We Don't Know. *Reviews in the Neurosciences* **22**, 17-25

9. Miller, N., Gerlai, R. (2008). Oscillations in Shoal Cohesion in Zebrafish (*Danio rerio*) *Behavioural Brain Research* **193**, 148-151.

10. Miller, N., Gerlai, R. (2007). Quantification of Shoaling Behaviour in Zebrafish (*Danio rerio*). *Behavioural Brain Research* **184**, 157-166.

11. Saverino, C., Gerlai, R. (2008). The social zebrafish: Behavioral responses to conspecific, heterospecific, and computer animated fish. *Behavioural Brain Research* **191**, 77-87.

12. Sison, M., Gerlai, R. (2011a). Associative learning performance is impaired in zebrafish (*Danio rerio*) by the NMDA-R antagonist MK-801. *Neurobiology of Learning and Memory* **96**, 230-237.

13. Sison, M., Gerlai, R. (2011b). Behavioral performance altering effects of MK-801 in zebrafish (*Danio rerio*). *Behavioural Brain Research* **220**, 331-337.

14. Sison, M., Gerlai, R. (2010). Associative learning in zebrafish (*Danio rerio*) in the plus maze. *Behavioural Brain Research* **207**, 99-104.

15. Sison, M., Cawker, J., Buske, C., Gerlai, R. (2006). Fishing for genes of vertebrate behavior: Zebra fish as an upcoming model system. *Laboratory Animal* **35**, 33-39.

Diving Deeper into Zebrafish Development of Social Behavior: Analyzing High Resolution Data

Christine Buske, Robert Gerlai

University of Toronto, Department of Cell & Systems Biology, Toronto, Ontario, Canada
chr.buske@mail.utoronto.ca

Zebrafish as a behavioral neuroscience model of disease

The mechanisms and development of vertebrate social behavior are not fully understood. Numerous human clinical conditions exist in which abnormal social behavior is the core symptom, among these are alcoholism and fetal alcohol syndrome. Animal models may facilitate the understanding of the mechanisms of social behavior and the mechanisms underlying abnormal social behavior. Zebrafish are known to shoal, i.e. aggregate in groups. We have just completed a study that described, for the first time, the ontogenesis of shoaling, i.e. revealed significant increases in shoaling behavior with developmental stage in zebrafish [1].

The effect of embryonic alcohol exposure on social behavior development

In the current study, we aim to investigate how embryonic alcohol exposure affects the ontogenesis of shoaling. Although a large body of literature exists on the effects of fetal alcohol exposure, many studies have been completed using unrealistically high alcohol concentrations. To make our investigation more easily extrapolated to realistic clinical circumstances, we employed a modest alcohol dosing regimen of low doses and a short exposure time (2 hours). A previous study has demonstrated this dosing regiment to be sufficient to cause significant reduction in social preference (reduced response to images of zebrafish) when adult fish were tested [2]. The current study looks at this phenomenon from a group perspective. Zebrafish are a highly social species and studying them in groups, versus the individual, opens up the possibility to better understand complex behaviors for this species. As not sufficient information exists yet, and the only previous study looking in depth at group behavior was recently completed, we approach this in an explorative manner: starting from individual plots for each subject, for short time periods of 1-5minutes, we aim to visually observe potential patterns emerging from positional data. This serves as a starting point to investigate the possibility of modeling the observed social behavior. Previous studies from our research group has shown an oscillating pattern in shoal cohesion of zebrafish [3], here we attempt to identify additional patterns that can be observed from group interactions. Collective behavior is becoming a topic of interest and can be a direct link to better understanding how the brain works. Research suggests that animals act on environmental data using probabilistic estimation as a means of making behavioral decisions [4]. Work performed by Perez-Escudero and Plavieja (2011) supports patterns of choice of *Gasterosteus aculeatus* correspond very well to probabilistic estimation using the social information. The link found between these patterns of behavior and mathematical models can help in future experimental design using multi-subject testing. In particular, group cohesion or aggregation behavioral patterns might be influenced by pharmaceutical insults (such as embryonic alcohol exposure) and can provide us with an alternate or richer resource to test subtle effects of different compounds on the (developing) brain through behavioral research.

Quantification of social behavior

Zebrafish are a highly social species and aggregate in groups [1]. Previous research has shown that the sight of conspecifics is also rewarding to zebrafish [5]. Here we chose to observe zebrafish in groups of ten, mimicking a realistic shoaling situation (vs. exposure to either automated images of zebrafish or the sight of live conspecifics separated by a glass barrier.

Positional data were generated for each subject in a ten-member group during the testing period, at a frequency of 30 times per second. Shoaling trials were repeated regularly throughout development from 7 to 90 days post fertilization. From this data, it is possible to extract the distance between all members of the group, speed for each subject and the average for the group, as well as other parameters that can be deduced from positional data. For example, thigmotaxis of an individual or the group as a whole, the establishment of a home base, or potential dominance patterns. Distances between shoal members have been previously used as a measure of shoal cohesion, but this project aims to go beyond this and explore how high resolution positional data for each member can be used to gather more information on group dynamics in zebrafish social behavior, throughout development. A first step towards understanding behavioral changes throughout development, and the effect of embryonic alcohol exposure on these behaviors, is to use shoal cohesion as a measure (inter individual distances over time). The second step is to model high resolution positional data to determine if throughout development or as a result of a pharmaceutical insult (such as embryonic alcohol exposure), changes can be observed. For example, [3] found that zebrafish groups show an oscillating pattern in their shoaling behavior. One of our working questions is whether this changes over time, or due to embryonic alcohol exposure. Measuring this behavior can provide greater insights in subtle behavioral changes perhaps not captures purely with average inter individual distances.

The open field task

Ten member groups are observed while shoaling in an open field. The open field task is an established and frequently used tool in rodent studies, but still quite new in zebrafish research. While exploring patterns in social interactions between shoal members can be very informative, we also investigate exploratory patterns of a single fish in the open field task. Relative to other models, such as rodents, little is known with regards to the behavioral repertoire of zebrafish. Preliminary results show a difference in how zebrafish explore the open field compared to rodents. Our data suggests that zebrafish heavily explore the open field within the first 5 minutes of a 30 minute trial. The remaining time, the fish increasingly spends more time closer to the boundaries of the arena, and thus is showing stronger thigmotactic behavior. This is the opposite of what is observed in rodent behavior [6]. A noteworthy observation that can potentially be exploited in behavioral paradigms.

Behavioral patterns and data mining for these studies is performed using R: A language and environment for statistical computing.

Neurochemical analysis

To complement behavioral data obtained as described above, we have aimed to identify some neurochemical targets of interest. Dopamine and serotonin are implicated in several conditions with aberrant behavior as a component (such as depression and anxiety disorders), we use these targets to assess the effects of embryonic alcohol exposure on the brain. Our studies show that one time exposure of moderate to low doses of alcohol permanently disrupts normal dopamine and serotonin levels. Levels for both neurochemicals, as well as their metabolites DOPAC and 5-HIAA are significantly depressed in whole brain samples of 40, 70, and 102 day old zebrafish. These observations lay the foundation to further investigate the effects of embryonic alcohol exposure on the brain, over time.

Conclusions

Zebrafish are an increasingly popular model organism for human diseases. As zebrafish mature as a model in behavioral neuroscience, a better understanding of this species' behavioral repertoire is increasingly more important for future research. We aim not only to establish some new behavioral paradigms, but also investigate the long term effects of one time, low dose, alcohol exposure on the development of social behavior and neurochemistry in zebrafish.

References

1. Buske, C., Gerlai, R. (2010). Shoaling develops with age in Zebrafish (*Danio rerio*). *Prog Neuropsychopharmacol Biol. Psychiatry*. doi:10.1016/j.pnpbp.2010.09.003.

2. Fernandes, Y., Gerlai, R. (2009). Long-term behavioral changes in response to early developmental exposure to ethanol in zebrafish. *Alcohol Clin. Exp. Res.* **33**, 601-609: doi:10.1111/j.1530-0277.2008.00874.x.

3. Miller, N.Y., Gerlai, R. (2008). Oscillations in shoal cohesion in zebrafish (*Danio rerio*). *Behavioural Brain Research* **193**, 148-151. doi:10.1016/j.bbr.2008.05.004.

4. Pérez-Escudero, A, de Polavieja, G.G. (2011). Collective animal behavior from Bayesian estimation and probability matching. *PLoS Comput. Biol.* **7**, e1002282. doi:10.1371/journal.pcbi.1002282.

5. Al-Imari, L., Gerlai, R. (2008). Sight of conspecifics as reward in associative learning in zebrafish (Danio rerio). *Behavioural Brain Research* **189**, 216-219. doi:10.1016/j.bbr.2007.12.007.

6. Blaser, R.E., Chadwick, L., McGinnis, G.C. (2010). Behavioral measures of anxiety in zebrafish (*Danio rerio*). *Behavioural Brain Research* **208**, 56-62. doi:10.1016/j.bbr.2009.11.009.

Zebrafish Assays to Measure ADHD Endophenotypes

M. Lange[1], W.H.J. Norton[1], M. Chaminade[1], P. Vernier[2], K.P. Lesch[3], L. Bally-Cuif[1]

[1]*Zebrafish Neurogenetics, Neurobiology & Development, INAF, CNRS UPR3294, Gif-Sur-Yvette, France. lange@inaf.cnrs-gif.fr*

[2]*"Development and Evolution of Neurotransmission", Neurobiology and Development, CNRS UPR 3294, Gif-Sur-Yvette, France*

[3]*Department of Psychology, Biological Psychology, Clinical Psychology, and Psychotherapy, University of Würzburg, Würzburg, Germany*

Attention Deficit Hyperactivity Disorder: Introduction

Attention-deficit hyperactivity disorder (ADHD) is a developmental disorder characterized by hyperactivity, impulsivity and inattention. It currently affects around 5% of children worldwide and frequently has long-term consequences. However, although there is a significant genetic component to ADHD, relatively few risk genes have been identified and characterized [1-3]. The drug treatments available for the disease are poor with variable efficacy and significant side effects. Furthermore the etiology of ADHD is poorly understood and there are relatively few animal models for the disease. Recently, Arcos-Burgos and colleagues [4] reported evidence for a risk haplotype in the gene encoding Latrophilin 3 (*LPHN3*), revealed by a linkage scan in a genetic isolate population and subsequent fine mapping. LPHN3 is a potential G protein-coupled receptor with the propensity to moderate cell-cell interaction (adhesion-GPCR). However the physiological function of LPHN3 is not well understood and endogenous ligands have yet to be identified.

ADHD endophenotypes

It is therefore essential to use animal models in order to investigate the genetics basis which leads to disease pathology. Unfortunately however, the complicated genetic basis of ADHD makes it difficult to fully recreate them in animal models. One way to simplify this problem is to measure endophenotypes, biological markers that correspond to a disease-gene's activity [5]. Endophenotypes only model a few of the symptoms of a psychiatric disorder, they should also be connected to the symptoms of the disease, perhaps reflecting some of the underlying changes in neurobiology [6]. The creation of animal model for at least some endophenotypes of ADHD will allow in-vivo analysis and subsequently a better understanding of disease neurobiology.

Modeling ADHD in zebrafish

In recent years, the zebrafish (*Danio rerio*) was established as a valid animal model for probing the genetics and developmental bases of behavior, based on the combination of a well-characterized ontogeny, availability of numerous tools for genetic modification and a battery of tests for behavioral characterization [7-9]. By six days, larval fish swim continuously, search for food and are able to escape from predators, thus demonstrating a range of innate behaviors. Furthermore, the formation and function of neurotransmitter signaling pathways is conserved between zebrafish and other vertebrates, highlighting the usefulness of studies conducted in fish.

To validate LPHN3 as a potential ADHD risk factor, we analyzed the function of zebrafish Lphn3 in larvae. We used morpholino injection to achieve a transient reduction of Lphn3 activity. We next characterized the larval behavioral phenotype of lphn3 morphants.

Quantification of the *lphn3* morphant endophenotypes: hyperactivity and motor impulsivity

We focus our attention on two behavioral endophenotypes of ADHD: hyperactivity and motor impulsivity. Locomotion can be easily measured in larval fish. We use automated videotracking software to measure fry zebrafish. Automated software allows us to record and measure simultaneously multiple parameters. We can also readily modify the parameters of the assays (area size, detection threshold, integration time…). We first measured larval behavior at 6 days and found an increase in the distance swum by the morphants in a 5 mm experiment, as revealed by an increase in the mean swimming speed (the resting time was unchanged). Furthermore, this hyperactivity can be rescued by application of the pharmacological treatments for ADHD, Methylphenidate and Atomoxetine. ADHD patients exhibit constant hyperactivity [10]. To determine whether *lphn3* morphant larvae were constantly hyperactive, we plotted the distance swum every three seconds during a 90 second experiment for several larvae. We found that the morphants displayed remarkably stable locomotion over time. ADHD-associated impulsivity can be subdivided into both motor and cognitive components [11]. We developed a method to analysis the motor impulsivity in 6dpf larvae by plotting the distance swum in short time windows for individual animals. We observed sharper acceleration peaks in the morphants' curves, confirming a motor impulsivity endophenotype. Together these results suggested that *lphn3.1* morphants show 2 main behavioral ADHD-like endophenotypes: a stable hyperactivity, and motor impulsivity.

Neurobiological characterization of the morphants

We currently aim to connect potential neuroanatomical defects with the observed behavioral endophenotypes. We focused on monoaminergic systems because of their known link with both ADHD and locomotor control [12]. While the serotonergic (5-HT) and noradrenergic systems remained unaffected, we found that *lphn3* morphant larvae display a disorganization of dopaminergic (DA) neurons. In parallel, immunohistochemical, in situ hybridization and high pressure liquid chromatography (HPLC) studies of other neurotransmitter systems, (including NA, 5-HT, GABA and glutamate) failed to reveal obvious defects in these other neurotransmitter signals following loss of *lphn3.1* function.

Standardization of the ADHD-like endophenotypes assays in popular zebrafish strains

As zebrafish neuroscience research becomes more and more popular, the definition of standard thresholds for measurable behaviors (such as sleep/wake activity, locomotion, predation, phototaxis, learning, etc.) and standardization of tests/techniques become crucial. Using the methods developed to study the *lphn3* morphant, we characterized the locomotion activity/impulsivity throughout zebrafish life (6dpf/1month/3months). The zebrafish community uses several wild type strain of *D. rerio,* among which we chose AB, WIK, TU and EK strains to conduct a behavioral comparative study. We also decided to add to this study the transparent Casper line [13], which has become more popular with the emergence of optogenetics and live-imaging studies. We developed and conducted high-throughput analyses of the distance travelled, speed and resting time for 6dpf/1month/3months zebrafish with automated sorting of the data. We also created routine assays to record the stable activity and the motor impulsivity described above. We obtained the locomotion threshold profile for each strain according to their intrinsic polymorphisms. All this information will help to choose the best strain for a given experiment depending on the locomotion parameters to be investigated. Moreover the methods developed provide a robust and precise *modus-operandi* to study ADHD-like endophenotypes (as locomotion) in zebrafish.

Summary and conclusion

In the first part we provided evidence implicating Lphn3 in the control of locomotor activity and DA development. *lphn3* morphants have the potential to become a new animal model for the locomotor endophenotypes associated with ADHD and will allow us to expand our knowledge of Lphn3 function. However, there is still a clear need to expand the number of endophenotypes that can be measured, in particular to include those that quantify cognitive impulsivity and attention. We also adapted the assays that measure ADHD-like endophenotypes to obtain a standardized *modus-operandi* to study the locomotion behavior in zebrafish at different stages. Moreover we used those automated solutions to describe the locomotion parameters of popular zebrafish strains. Finally the development of validated protocols to measure different ADHD-like endophenotypes in zebrafish suggest that zebrafish has the potential to become one of the primary model organisms for translational research of psychiatric disorders.

Acknowledgements

We are grateful to all LBC lab members for their support and would like to thank Michael Rebagliati for helpful discussion and comments. Work in L. B-C's laboratory is supported by the Neuroscience School of Paris (ENP), the ANR (Chair of Excellence ANR-08-CEXC-001-01), the FRM (program DPR 20081214424**)**, the PIME program, the Schlumberger Association (grant DLS/GP/LB090305) and the EU projects NeuroXsys (grant agreement FP7-2007-2013- no 223262) and ZF-Health (grant agreement HEALTH-F4-2010-242048).

References

1. Lesch, K. (2009). Attention-deficit/hyperactivity disorder (ADHD). in (ed.) L. F. Ed., *Encyclopedia of Molecular Mechanisms of Disease*, Springer: Heidelberg, New York, 188-191.

2. Polanczyk, G., Rohde, L.A. (2007). Epidemiology of attention-deficit/hyperactivity disorder across the lifespan, *Curr Opin Psychiatry* **20**, 386-392.

3. Swanson, J.M., Sunohara, G.A., Kennedy, J.L. (1998). Association of the dopamine receptor D4 (DRD4) gene with a refined phenotype of attention deficit hyperactivity disorder (ADHD): a family-based approach, *Mol Psychiatry* **3**, 38-41.

4. Arcos-Burgos, M., Jain, M., Acosta, M.T. (2010). A common variant of the latrophilin 3 gene, LPHN3, confers susceptibility to ADHD and predicts effectiveness of stimulant medication, *Mol Psychiatry* **15**, 1053-1066.

5. Gottesman, I.I., Gould, T.D. (2003). The endophenotype concept in psychiatry: etymology and strategic intentions, *Am J Psychiatry* **160**, 636-645.

6. Castellanos, F.X., Tannock, R. (2002). Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes, *Nat Rev Neurosci* **3**, 617-628.

7. Fero, K., Yokogawa, T., Burgess, H.A. (2010). The behavioural repertoire of larval zebrafish.

8. Lieschke, G.J., Currie, P.D. (2007). Animal models of human disease: zebrafish swim into view, *Nat Rev Genet* **8**, 353-367.

9. Norton, W., Bally-Cuif, L. (2010). Adult zebrafish as a model organism for behavioural genetics, *BMC Neurosci* **11**, 90.

10. Taylor, E. (1998). Clinical foundations of hyperactivity research, *Behav Brain Res* **94**, 11-24.

11. Zhuang, X., Oosting, R.S., Jones, S.R. (2001). Hyperactivity and impaired response habituation in hyperdopaminergic mice, *Proc Natl Acad Sci USA* **98**, 1982-1987.

12. Gizer, I.R., Ficks, C., Waldman, I.D. (2009). Candidate gene studies of ADHD: a meta-analytic review, *Hum Genet* **126**, 51-90.

13. White, R.M., Sessa, A., Burke, C. (2008). Transparent adult zebrafish as a tool for in vivo transplantation analysis, *Cell Stem Cell* **2**, 183-189.

Automated Analyses of Behavior in Zebrafish Larvae

Sean D. Pelkowski[1], Holly A. Richendrfer[1], Ruth M. Colwill[2], Robbert Creton[1*]

[1]*Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, Rhode Island, USA.*
[*]*Robbert_Creton@brown.edu*

[2]*Department of Cognitive Linguistic and Psychological Sciences, Brown University, Providence, Rhode Island, USA*

Early brain development may be influenced by numerous genetic and environmental factors with long-lasting effects on brain function and behavior. Identification of these factors is facilitated by behavioral studies in animal model systems. However, large-scale screening in whole organisms remains challenging. We developed a novel high-throughput imaging system capable of analyzing complex behaviors in zebrafish larvae. The three-camera system can image twelve multi-well plates simultaneously and is unique in its ability to provide local visual stimuli in the wells of a multi-well plate. The acquired images are converted into a series of coordinates, which characterize the location and orientation of the larvae. The imaging techniques were tested by measuring avoidance behaviors of zebrafish larvae in response to visual stimuli. The system effectively quantified larval avoidance responses and revealed an increased edge preference in response to a 'bouncing ball' stimulus. We further demonstrated that this edge preference, or thigmotaxis, can be interpreted as a measure of anxiety; larvae exposed to valium showed a reduced edge preference, while larvae exposed to caffeine showed an increased edge preference. We are currently using the developed imaging system to screen for behavioral defects in response to modulators of calcium signaling and organophosphate pesticides. The imaging system and assays for measuring avoidance behavior may also be used to screen for a variety of other genetic and environmental factors that cause developmental brain disorders and for novel drugs that could prevent or treat these disorders.

Fish as a Model to Study Non-verbal Numerical Abilities

Christian Agrillo, Maria Elena Miletto Petrazzini, Laura Piffer, Marco Dadda, Angelo Bisazza

Department of General Psychology*, University of Padova, Italy. christian.agrillo@unipd.it*

Numerical abilities are widespread among vertebrates

Abilities such as recording the number of events, enumerating items in a set, or comparing two different sets of objects can be adaptive in a number of ecological contexts. Lyon [1], for instance, reported a spontaneous use of numerical information (egg recognition and counting) in a natural context as a strategy to reduce the costs of conspecific brood parasitism in American coots. McComb and co-workers [2] using playback experiments found that wild lions based the decision whether or not to attack a group of intruders on a comparison of the number of roaring intruders they had heard and the number and composition of their own group. Extensive laboratory research carried out on apes and monkeys has revealed the existence of non-verbal systems of numerical representation that non-human primates apparently share both with infants and with human adults tested in comparable conditions [3]. In the last five years, an increasing number of studies has focused their attention also on numerical abilities of species more distantly related to humans, such as fish, in order to broaden our knowledge on the evolutionary origin of number processing.

Spontaneous quantity discrimination in poeciliid fish

There is substantial evidence that, in social situations, individual fish in unfamiliar environments tend to join other conspecific and, if choosing between two shoals, they exhibit a preference for the larger group [4-8]. This spontaneous preference for joining the larger shoal is commonly adopted to study fish ability to discriminate between quantities. In these tests the experimental apparatus is usually composed of three adjacent tanks. The central one, the 'subject tank', houses the test fish. At the two ends two 'stimulus tanks' (in which two shoals are inserted) face the subject tank. Subject is inserted in the subject tank and his/her behaviour is recorded for 10-15 minutes. Shoal preference is calculated as the time spent by the subject near the glass facing either of the stimulus tank.

Both mosquito-fish (*Gambusia holbrooki*) and guppies (*Poecilia reticulata*) proved able to discriminate between shoals differing by one unit up to 4 (1 vs. 2, 2 vs. 3 and 3 vs. 4) showing apparently the same effort (that is, fish accuracy was not affected by numerical ratio in the small number range). At the same time, fish could discriminate larger shoals provided that numerical ratio was at least 1:2 (i.e. 4 vs. 8). Performance above 4 units showed ratio-dependence and the capacity to discriminate between two quantities became increasingly accurate as the ratio between them increased [9]. Similar results have been reported in angel-fish (*Pterophyllum scalare*) [6-7]. As a reference, in a recent study [5] a group of undergraduate students was required to estimate the same numerical ratio presented to poeciliid fish. Participants had to estimate the larger of two groups of dots while prevented from verbal counting. Interestingly, humans and fish showed almost identical performance patterns for small and large quantities, suggesting that the evolutionary emergence of our quantity abilities may be more ancient than we have previously thought.

Use of number by poeciliid fish

Much debate has arisen over the exact mechanisms enabling animals to make such a discrimination. Since stimulus numerousness co-varies with non-numerical extent, such as the total area occupied by objects, the sum of their contour, their density and luminance, organisms can provide quantity judgments without necessarily being capable of numerical representation [10-11]. For instance, in a shoal choice test fish may select the larger shoal by using the overall space occupied by the groups instead of numbers. The assessment of numerical capacities in animals requires careful controls to exclude non-numerical cues being used in place of number.

One experimental strategy employed to exclude the use of continuous cues in spontaneous preference tests is the sequential presentation of items within each set, so that subjects can never have a global view of the entire contents of the sets. In order to choose the preferred or the reinforced set in these experiments animals have, therefore, to attend to each item and to build a representation of the contents of the set on the basis of the items that come sequentially into view. Then they have to repeat the process for the second set and, finally, to compare the two representations. For example, in Hauser and colleagues' study [12] rhesus monkeys observed experimenters that placed pieces of apple, one at a time, into each of two opaque containers. Monkeys had no opportunity to see the matched groups before selecting one of them and thus had to mentally add the number of food items inserted in each container.

We have recently adapted the item-by-item procedure in a shoal choice test using mosquito-fish as model [13]. The apparatus for shoal choice was modified by confining each stimulus fish in an adjacent and separate compartment. In this way subjects could choose between one large and one small group of companions but they could only see one fish at a time, thus preventing the possibility that they could use non numerical attributes of the shoal, such as the cumulative area occupied by fish, to estimate the larger set. Mosquito-fish proved to successfully discriminate the larger group of social companions also in this test, suggesting a spontaneous number representation.

To investigate whether animal species can process discrete (numerical) discrimination, extensive training procedures are also reported in literature. In a recent series of experiments we adopted one of these procedures consisting of training the subject to discriminate between sets containing different numbers of geometric figures. Mosquito-fish were placed in an unfamiliar tank and trained to discriminate between two doors in order to re-join their social group. Doors were associated with a pair of stimuli consisting of groups of figures differing in numerosity. These figures were controlled for non-numerical variables, therefore fish could solve the task by attending numerical information only. Fish proved be able to use numerical information both in the small (2 vs. 3) and in the large (4 vs. 8) number range [14-15]. Interestingly, their accuracy was not affected by the total set size: mosquito-fish can discriminate 4 vs. 8 as well as 100 vs. 200. On the contrary, numerical ratio affected the performance, and discriminating a 1:2 numerical ratio was easier than a 2:3 or 3:4 ratio. As reference we tested adult humans presenting the same stimuli: again, the performance of humans largely overlapped that of fish. This further supports the idea of similar non-verbal numerical systems shared among vertebrates.

Conclusions and future directions

Recent studies have demonstrated that numerical abilities not only predate verbal language but also have a very ancient evolutionary origin. Fish are able to spontaneously discriminate between quantities and use numerical information when continuous variables are controlled for. To date, two main procedures have been reported in literature: spontaneous shoal choice test and training procedure using social reward. Only recently we have developed a new training procedure [16] using food reward. Subjects are singly housed in rectangular tanks. At intervals, two stimuli (groups of figures differing in numerosities) are introduced at opposite ends of the tank and food is delivered near the stimulus to be reinforced. Time spent near positive stimulus in probe trials is taken as a measure of discrimination performance. To validate the method, we replicated two published studies that used operant conditioning to investigate the mechanisms of numerical discrimination in mosquito-fish: our data indicate a complete overlap of the results obtained using the two different methods. The novel procedure, however, proved to be less time-consuming and showed less limitations than operant conditioning. In this sense, in the next future we might investigate the full range of numerical abilities in poeciliid fish.

More generally, the new paradigm may also be well suited for automation. Stimulus delivery on a computer screen could be synchronized with automated tracking of the fish movements using one of the available programs, which could also serve for the automated measurement and analysis of visual choice. This might provide a system for high-throughput conditioning of fish in a manner similar to approaches already used with rodents and offer a powerful tool in many fish studies involving learning.

References

1. Lyon, B.E. (2003) Egg recognition and counting reduce costs of avian conspecific brood parasitism. *Nature* **422**, 495-499.
2. McComb, K., Packer, C., Pusey, A. (1994) Roaring and numerical assessment in contests between groups of female lions, *Panthera leo*. *Animal Behaviour* **47**, 379–387.
3. Cantlon, J.F., Brannon, E.M. (2006) Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science* **17**, 401–406.
4. Agrillo, C., Dadda, M. (2007) Discrimination of the larger shoal in the poeciliid fish *Girardinus falcatus*. *Ethology, Ecology and Evolution* **19**,145–157.
5. Agrillo, C., Piffer, L., Bisazza, A., Butterworth B (2012). Evidence for two numerical systems that are similar in humans and guppies. *PLoS ONE* **7**(2), e31923.
6. Gòmez-Laplaza, L.M., Gerlai, R. (2011a) Can angelfish (*Pterophyllum scalare*) count? Discrimination between different shoal sizes follows Weber's law. *Animal Cognition* **14**(1),1–9.
7. Gòmez-Laplaza, L.M., Gerlai, R. (2011b) Spontaneous discrimination of small quantities: shoaling preferences in angelfish (*Pterophyllum scalare*). *Animal Cognition* **14**(4),565–574.
8. Buckingham, J.N., Wong, B.B.M., Rosenthal, G.G. (2007) Shoaling decisions in female swordtails: how do fish gauge group size? *Behaviour* **144**,1333–1346.
9. Agrillo, C., Dadda, M., Serena, G., Bisazza. A (2008) Do fish count? Spontaneous discrimination of quantity in female mosquitofish. *Animal Cognition* **11**(3),495–503.
10. Feigenson, L., Carey, S., Spelke, E.S. (2002) Infants' discrimination of number vs. continuous extent. *Cognitive Psychology* **44**, 33–66.
11. Pisa, P.E,, Agrillo, C. (2009) Quantity discrimination in felines: A preliminary investigation of the domestic cat (Felis silvestris catus). *Journal of Ethology* **27**, 289–293.
12. Hauser, M.D., Carey, S., Hauser, L.B. (2000) Spontaneous number representation in semi-free-ranging rhesus monkeys. *Proceeding of the Royal Society of London B* **267**,829–833.
13. Dadda, M., Piffer, L., Agrillo, C., Bisazza, A. (2009) Spontaneous number representation in mosquitofish. *Cognition* **112**,343–348.
14. Agrillo C., Dadda, M., Serena, G., Bisazza, A. (2009) Use of number by fish. *PLoS ONE* **4**(3), e4786.
15. Agrillo, C., Piffer, L., Bisazza, A. (2010) Large number discrimination by mosquitofish. *PLoS ONE* **5**(12), e15232.
16. Agrillo, C., Miletto Petrazzini, M.E., Piffer, L., Dadda, M., Bisazza, A. (2012). A new training procedure for studying discrimination learning in fish. *Behavioural Brain Research*, online first.

Olfactory Signals Involved in Kin Recognition in Zebrafish

Gabriele Gerlach[1], Cornelia Hinz[2]

Carl von Ossietzky University, Oldenburg, Germany
[1]gabriele.gerlach@uni-oldenburg.de, [2]Cornelia.Hinz@uni-oldenburg.de

Kin recognition

In many species - invertebrates and vertebrates - social interaction differs according to the genetic relatedness of individuals. Individuals that are able to assess the genetic relatedness of conspecifics can preferentially allocate resources towards related individuals and avoid inbreeding.

Several mechanisms have been proposed by which individuals may discriminate kin. One way to identify possible relatives is to treat any conspecific that shares a particular location or degree of familiarity as kin. Kin recognition requires more stringent criteria when proximity and familiarity with conspecifics are not sufficiently reliable to detect true genetic relatedness. In a more specific method of kin recognition known as phenotype matching, an individual learns a template of its own phenotype [1] and/or that of its familiar kin [2], and later compares the phenotypes of unfamiliar animals with this template [3]. Such phenotype matching depends on a consistent correlation between phenotypic and genotypic similarity, so that detectable traits are more alike among close relatives than among more distantly related or unrelated individuals [4].

Olfactory imprinting and kin recognition in zebrafish

We have addressed the question of how animals acquire the ability to recognize kin by studying the development of olfactory kin preference in zebrafish (*Danio rerio*). We exposed zebrafish eggs directly after fertilization to different conditions: We either kept a single larva isolated in a small beaker or exposed larvae to olfactory and visual cues of their siblings at different days of their development. When 10 to 15 days old, we tested larvae for their olfactory preference of kin in a two channel choice flume with a steady driven flow. Each channel of the flume contained holding water of either kin or non-kin. We recorded the position of the fish in one or the other water flow every 10 s during two 3-min periods separated by a 1-min transition period to switch water sources as a control for possible (non-olfactory) side bias of the fish. Our results show that larvae could only recognize and preferred kin when having been exposed to kin during a 24-hour time window on day six post fertilization (dpf) [5]. Exposing larvae to kin at any other day than at 6 dpf does not induce any recognition of kin later in life; isolated larvae did also not express any kin recognition later in life [6,7].

Surprisingly, larvae do not imprint on the cues of non-kin during this sensitive phase for imprinting, indicating that a predisposition for the olfactory cues might exist or that larvae might use additional cues [5].

Our results on kin recognition in juvenile and adult zebrafish suggest that juveniles prefer to shoal with relatives while adults might use the same kin recognition mechanism to avoid mating with close relatives. Indeed, in the same choice flume adult females preferred olfactory cues of foreign males over those of brothers [7].

Identification of olfactory signals triggering kin recognition

Extreme genetic diversity of genes of the major histocompatibility complex (MHC) provides the underlying mechanism for kin recognition [8]. The MHC is a cluster of genes that has been intensively studied for its importance in immune reactions and immune recognition [9], and shows striking similarity among many vertebrates. Variation in MHC genes contributes also to unique individual odors (odortypes). MHC/peptide complexes expressed at cell surfaces are thought to be shed from the cell surface and their fragments appear in the urine and other body fluids. In recent studies we could identify peptide ligands of MHC receptors as signals triggering imprinting and kin recognition [10]. When adding mixtures of nine amino acid long peptides we could evoke later kin recognition in zebrafish with specific MHC class II alleles.

Benefits of associating with kin

We examined the potential benefits of kin preference by comparing growth rate, shoaling, and aggressive behavior in juvenile zebrafish housed in groups of either familiar kin or unfamiliar non-kin [11]. Over an observation period of 5 days animals grew 33 % more in kin groups; however, neither shoaling nor the frequency of aggressive interactions was different in groups of related versus unrelated individuals. Shoaling behavior increased with increasing observation time and increasing age, while aggressive behavior remained the same. We conclude that associating with kin probably creates a less stressful environment that allows for higher growth rates, which can lead to higher direct fitness based on increased survival and earlier reproduction. Kin recognition leading to kin-structured groups may therefore be under positive selection

References

1. Mateo, M. Johnston, E. (2000). Kin recognition and the 'armpit effect': evidence of self-referent phenotype matching. *Proceeding of the Royal Society of London B* **267**, 695-700.

2. Sherman, P.W., Reeve, H.K., Pfennig, D.W. (1997). Recognition systems. Pages 69-96 *in* J. R. Krebs and N. B. Davies, editors. *Behavioral ecology: an evolutionary approach*. Blackwell Scientific, Oxford UK.

3. Tang-Martinez, Z. (2001). The mechanisms of kin discrimination and the evolution of kin recognition in vertebrates: A critical re-evaluation. *Behavioural Processes* **53**, 21-40.

4. Holmes, W., Sherman, P. (1983). Kin recognition in animals. *The American Scientist* **71**, 46-55.

5. Gerlach, G., Hodgins-Davis, A., Avolio, C., Schunter, C. (2008). Kin recognition in zebrafish: A 24-hour window for olfactory imprinting. *Proceedings of the Royal Society, London B* **275**, 2165-2170.

6. Mann, K.D., Turnell, E.R., Atema, J., Gerlach, G. (2003). Kin recognition in juvenile zebrafish (*Danio rerio*) based on olfactory cues. The Biological Bulletin **205**, 224-225.

7. Gerlach, G., Lysiak, N. (2006). Kin recognition and inbreeding avoidance in zebrafish, *Danio rerio*, is based on phenotype matching. *Anim. Behav.* **71**, 1371-1377.

8. Hepper, P.G., Cleland, J. (1999). Developmental aspects of kin recognition. *Genetica* **104**, 199-205.

9. Klein, J. (1986). Natural history of the major histocompatibility complex and human evolution. *Trends in Genetics* **6**, 7-11.

10. Hinz, C., Namekawa, I., Behrmann-Godel, J., Oppelt, C., Dittmar, J., Jäschke, A., Friedrich, R.W., Gerlach, G. (submitted). A peptide makes you like your kin: MHC peptide ligands can evoke olfactory imprinting on kin in zebrafish.

11. Gerlach, G., Hodgins-Davis, A., MacDonald, B., Hannah, R.C. (2007). Benefits of kin association: related and familiar zebrafish larvae (*Danio rerio*) show improved growth. *Behavioral Ecology and Sociobiology* **61**, 1765-1770.

General Sessions

Assigning and Combining Probabilities in Single-Case Studies

R. Manolov[1], A. Solanas[2]

Department of Behavioral Sciences Methods, University of Barcelona, Barcelona, Spain

[1]*rrumenov13@ub.edu*, [2]*antonio.solanas@ub.edu*

Abstract

There is currently a considerable diversity of quantitative measures available for summarizing the results in single-case studies (e.g., nonoverlap indices, regression coefficients or R-squared values). Given that the interpretation of some of them is difficult due to the lack of established benchmarks, the current paper proposes an approach for obtaining further numerical evidence on the importance of the results, complementing substantive criteria, visual inspection and the summary indices themselves. This additional evidence consists in obtaining the likelihood of the outcome in case the intervention was ineffective. The probability is expressed in terms of p values, which can then be used to integrate the results of several studies; an integration which is problematic when different metrics are used across primary studies and raw data are not available. Two methods for combining probabilities in the context of single-case studies are pointed out – one based on a weighted average and the other on the binomial test.

Introduction

Currently there is a great variety of techniques proposed for quantifying the magnitude of effect in single-case data. However, not all of these procedures are accompanied by unquestionable interpretative benchmarks in order to judge the relevance of the results obtained. Even in the case of regression-based procedures which yield R-squared values the use of Cohen's guidelines in single-case studies has been put in doubt due to greater effects usually found [1]. Apart from the interpretation of individual studies' results, another important question yet to be answered is how to integrate the results of several studies conducted on the same topic. As regards integration meta-analysis is the option of choice, but combined significance may be useful when effect sizes are not reported or there is not an established effect size measure for some data analytic procedure [2]. Furthermore, the great proliferation of analytical techniques expressed in different metrics and the lack of consensus on which technique to use in order to summarize the results hinders carrying out meta-analyses [3] and opens the possibility for combining probabilities. Note, however, that when raw data are available the researcher can compute the metric of choice regardless of the original primary indicator and combine the results afterwards. The present study focuses on two complementary topics: a) the additional assessment of intervention effectiveness in an individual single-case study – this is achieved estimating statistical significance after constructing the relevant sampling distribution and b) the quantitative integration of several single-case studies using different indices for quantifying the magnitude of effect.

Additional evidence for effectiveness in individual studies

Rationale

Single-case study results should be analyzed both visually and numerically, while also using substantive criteria of what effect is relevant in the specific behavioral context. However, it would also be useful for the researcher to have a statistical criterion complementing the substantive one. Given the lack (or inadequacy) of benchmarks for most indices, we propose a method for further assessing the relevance of an effect size. This additional evidence is based on obtaining the statistical significance associated with the index computed. p values have already been used in single-case designs via randomization tests [4], but in the proposal made here the sampling distribution is not constructed after permuting the data or the points of change in phase. The basic idea is that each effect size index, standardized or unstandardized, has its sampling distribution and its expected value in the conditions of no intervention effect, that is, when the null hypothesis is true. The value actually obtained (i.e., the "outcome") can be located in the sampling distribution in order to know whether a value as large as or larger than the outcome is likely to be obtained only by chance in absence of effect.

The maximal reference approach (MRA)

The current proposal is related to simulation modeling analysis, SMA [5], in which the outcome is located in a sampling distribution based on generated samples extracted from a population where the following parameters are specified: phase lengths equal to the ones of the original sample, no intervention effect, phase serial dependence equal to the autocorrelation estimated in the original sample, normal disturbance. However, at least two assumptions are being made with the SMA: a) data are normal, which may be questionable, and b) the autocorrelation is estimated precisely, which is problematic when few measurements are available. In case these assumptions are not met, the sampling distribution constructed may not be appropriate. In order to reduce the uncertainty around these unknown data features and gain confidence on the validity of the sampling distribution as a reference, it may be necessary to construct not one but several sampling distributions, provided that the typical values of the primary indicators are expected to vary according to series/phase lengths, the data generation processes, the degrees of serial dependence, the random variable distributions, etc. Our proposal is to follow a conservative approach, the maximal reference approach (MRA), in which the index values associated with several key *p* values (e.g., .90, .80, ..., .20, .10, .05) are identified for several conditions. If the *p* values are tabulated, then a researcher can compare the outcome to the reference values from the table in order to know in what range of probability is such an outcome expected at random. For instance, suppose that a researcher obtains an R-squared value of .60 after carrying out a regression analysis, whereas another researcher obtains a value of 95% using a nonoverlap index. Are the effects large? In which study is the intervention effect stronger? We consider that the MRA, when used jointly with visual analysis and substantive criteria, may aid professionals interested in these questions.

Quantitative integration of single-case studies using different metrics

Combining weighted vs. unweighted probabilities.

One of the possible sources of invalidity when integrating results is weighting equally studies with different sample sizes, but other factors affecting the reliability and validity of the results should also be considered [6]. Weighting gives more information and has proven to be more powerful when combining studies with different sample sizes [7]. Additionally, weighting is already inherent to the meta-analytical combination of effect sizes.

Methods for combining probabilities.

Of the diversity of existing approaches for combining probabilities only two will be highlighted here, given that they are especially relevant for the approach presented in the previous section. One of the methods for combining probabilities that can be used is similar to a proposal consisting in testing the statistical significance of the mean of the *p* values of the studies to be integrated [8]. However, as we underline the need of weighting the *p* values, the statistical test presented by Edgington [8] is not applicable. As regards the exact weighting procedure, a minimal requirement would be to use series length (the single-case equivalent of sample size) as a weight. However, it should be considered that the ideal weight is the inverse of the error variance, that is, the Fisher information [7]. Therefore, one option would be to use the inverse of the variance of the summary index, given that the amount of dispersion of the index values about its mathematical expectancy is closely related to the reliability of the results. It has to be highlighted that this approach is equivalent to the one usually followed in the meta-analytical integration of studies' findings. In single-case designs, using the index's variance is impeded by the fact that different studies use different metrics. Given that variances are not directly comparable, we propose using the coefficient of variation (CV). In order to follow the MRA, for each outcome the researcher should assign a probability and a CV, both which need to be made available, for instance, in a tabular format. Given that the exact *p* value associated with the outcome is not known in the MRA, Edgington's [9] conservative solution can be followed, taking the upper limit of the probability given by a table as the probability itself (e.g., if the information available is that $p < .05$, then $p = .05$ should be the value used when integrating studies). The weights obtained via the CV could then be used to compute a weighted mean of the *p* values which can be compared to a predefined reference value.

The other method for combining probabilities potentially used for single-case studies is the binomial test, which has been deemed both quick and simple [10]. The main advantage is that it allows using studies which report only information on whether the *p* value was above or below .05 (i.e., it is possible to follow the MRA). The test consists in comparing each *p* value associated with the outcome to the nominal level *α* predetermined by the researcher. All outcome *p* values lower than *α* are counted as "successes" (versus "failures" in case of greater values) in terms commonly used when working with the binomial distribution [10]. After that the probability of obtaining as many successes, denoted by *s*, is referred to a binomial distribution with *n* (the number of trials), which is equal to the number of independent individual studies, and *π* (the probability of success in each trial), which is equal to *α*. The probability of obtaining *s* or more successes can be calculated directly from $\sum_{x=s}^{n}\binom{n}{x}\alpha^{x}\left(1-\alpha\right)^{n-x}$. For instance, suppose that a researcher wants to integrate the results of ten studies using the same intervention and similar participants, but not necessarily the same effect size indicator. Further suppose that the *α* selected is .05 and that three out of the ten studies are assigned $p < .05$ via the MRA. In this case, the probability of obtaining such a result only by chance is .0115, which indicates that, when considered together, the studies point at a strong intervention, although such an interpretation is always conditional on the professional's criteria, client's perceptions, etc.

Discussion

Strengths and limitations of the additional evidence for effectiveness in individual studies

The additional evidence has been presented in terms of *p* values, that is, the probability of obtaining as large as or larger outcome only by chance. Nonetheless, this evidence can also be expressed as the proportion of outcomes lower than or equal to the outcome, an idea similar to the one underlying percentiles. In any case, given the limitations of *p* values, these should not be used as the sole indicator, but rather as a complement to practitioner's experience, visual analysis, and the index quantifying behavioral change. The MRA is based on the idea of creating several plausible scenarios which allows researchers to make solidly supported decision. In terms of the efforts required from the researcher, when using the MRA it is only necessary to make a comparison to an already available maximal reference for the index and phase lengths used. However, given that the greatest *p* value of all conditions is used, this approach is rather conservative, which can be thought of as loss of power. The loss of power is attenuated for those cases in which there is evidence on the data features in a specific field and there is a narrower set of scenarios that need to be represented by the sampling distributions. Moreover, the probability assigned will be closer to the actual one – and the approach will be less conservative – in case there is a greater amount of reference values, for instance, for *p* values of .90, .80, …, .10, .05, and .01. Finally, it is not possible to rule out the possibility of publication bias – researchers may feel more inclined to submit for publication results for which the *p* value associated with the outcome is of specific magnitude. Nonetheless, journal editors endorse full reporting and maximal information around the effect size computed and they are not expected to reject an article only on the basis of one indicator, the p value, which is not the main focus of the findings.

Strengths and limitations of the quantitative integration of single-case studies using different metrics

The MRA allows combining studies whose results are summarized using different indices, which is important, given that the lack of a common effect size in single-case designs. The present paper focuses specifically on two ways of combining individual studies' *p* values. The weighted means approach has the drawback of the impossibility to test the composite *p* statistically. It only allows a comparison with a reference *p* determined by the researcher prior to carrying out the combination of individual studies' results. The binomial test is based on counting the amount of studies in which the *p* values is below a predefined nominal α and seems the most natural complement of the MRA. Another advantage of this method is that its logic is relatively simple to understand and its use is straightforward. The binomial test allows specifying any nominal α, given that it uses this reference as the probability of a positive result in for each study to integrate. Additionally, although the test is performed on *p* values, the original information expressed in terms of the magnitude of effect indices commonly used in single-case research is not lost. In case the idea subjacent to this approach for obtaining additional evidence and integrating it quantitatively is accepted by applied researchers, a logical step would be to obtain the reference values for the most frequently used procedures and for the variety of data patterns described here.

References

1. Parker, R.I., Brossart, D.F., Vannest, K. J., Long, J.R., Garcia De-Alba, R., Baugh, F.G., Sullivan, J.R. (2005). Effect sizes in single case research: How large is large? *School Psychology Review* **34**, 116-132.

2. Becker, B.J. (1987). Applying tests of combined significance in meta-analysis. *Psychological Bulletin* **102**, 164-171.

3. Beretvas, S.N., Chung, H. (2008). A review of meta-analyses of single-subject experimental designs: Methodological issues and practice. *Evidence-Based Communication Assessment and Intervention* **2**, 129-141.

4. Edgington, E.S., Onghena, P. (2007). *Randomization tests* (4th ed.). London: Chapman & Hall/CRC.

5. Borckardt, J.J., Nash, M.R., Murphy, M.D., Moore, M., Shaw, D., O'Neil, P. (2008). Clinical practice as natural laboratory for psychotherapy research: A guide to case-based time-series analysis. *American Psychologist* **63**, 77-95.

6. Noble, J.H. (2006). Meta-analysis: Methods, strengths, weaknesses, and political uses. *Journal of Laboratory and Clinical Medicine* **147**, 7-20.

7. Whitlock, M.C. (2005). Combining probability from independent tests: The weighted Z-method is superior to Fisher's approach. *Journal of Evolutionary Biology* **18**, 1368-1373.

8. Edgington, E.S. (1972b). A normal curve method for combining probability values from independent experiments. *Journal of Psychology* **82**, 85-89.

9. Edgington, E.S. (1972a). An additive method for combining probability values from independent experiments. *Journal of Psychology* **80**, 351-363.

10. Darlington, R.B., Hayes, A.F. (2000). Combining independent p values: Extensions of the Stouffer and binomial methods. *Psychological Methods* **5**, 496-515.

Data Fusion by Kernel Combination for Behavioral Data

D. Fekas[1], P. Zimmerman[2], C. J. F. ter Braak[3]

[1,3]Biometris¸ Wageningen University, Wageningen, The Netherlands. [1]d.fekas@noldus.nl

[2]Noldus Information Technology bv, Wageningen, The Netherlands

Introduction

Technological advances have radically changed the collection of behavioral data in animals. Modern tools complement, or even replace human observations in some labs. Also, behavior manifests itself in many different and complex ways, so behavioral scientists have developed a strategy of subjecting the animals to multiple observation systems, for longer periods of time, in order to get as much information as possible. Methodologically, the sample size (i.e. the number of animals) should be balanced from a statistical and an animal welfare point of view. Efficient statistical methods are therefore required to combine and analyze data from different types of experiments, while using a minimum number of animals. In this article, we present a generic framework that enables such a combination of diverse behavioral data.

Data description

We used behavioral observations of 4 strains (i.e. classes) of mice; a control group and three mutants which were designed to model Parkinson's disease (alpha-synuclein, synphilin-1 and double transgenic). There were 12 animals per strain and the raw data consisted of both video tracking data in a home cage (PhenoTyper®) and gait data from a gait analysis system (CatWalk® XT).

Video tracking within the home cage, which is performed by EthoVision® XT, produces large amounts of data, as the animals are tracked at a sample rate of 15 per second, during period of one week. There are predefined zones inside the cage and, also, automated cognitive tests take place at specific times. The output parameters are of a spatio-temporal nature: the 2D-body position in the cage, velocity, angular movement, relative position (within the cage) and shape are calculated on each frame.

The traditional way to analyze such longitudinal data is to split them in time-windows (bins) and then summarize (using mean values and coefficients of variation) all the calculated parameters per bin. We did this for both 12 hour long bins, corresponding to the light/dark cycle in the home cage, and for 1 hour long bins, taking only the first hour of the dark period for each day. However useful, the shortcoming of this approach is that if the bin is too long then you average over a very long period and hence lose information, while if too short, synchronization problems might appear; different animals might be engaged in different activities in each bin. Therefore, one could argue that comparisons across animals are not meaningful for short bins.

In this paper, we introduce an additional approach of representing behavior through strings. We define a set of behaviors of interest (alphabet) according to specific rules and then we produce strings or "words" that correspond to sequences of these behaviors in specific time periods. Consequently, we try to overcome the problem of misalignment in time by defining an appropriate similarity measure that detects similarities also between non-contiguous strings. This second approach results also in longitudinal, but very different type of

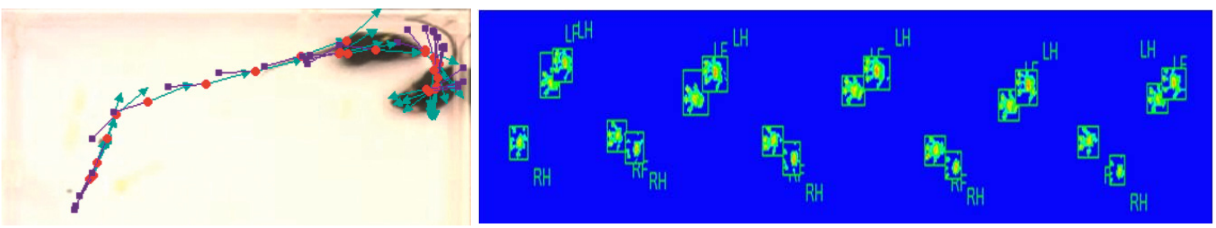


Figure 1. Snapshots of automated animal tracking (left) and gate analysis (right).

data; effectively a set of strings for each animal, whereas the time bin data sets are numeric. In addition, the string comparison method enables us to use the zone transitions (within the home cage) in the same classification context.

The gait analysis system on the other hand, produces numeric, but not particularly longitudinal data. CatWalk XT generates a large number of parameters related to a) individual paw prints (e.g., size of the print, mean intensity, stand, swing), b) spatial relationship between paws (e.g., base of support, print positions) and c) temporal relationships between paws (e.g., phase dispersion). All parameters are calculated for each analyzed run, i.e., a crossing from one side of the glass plate to the other on the CatWalk system. Runs can be acquired at different time points throughout the animal's life.

In total, from the video tracking data and gait analysis data we derived a total of 7 data sets, each measuring different aspects of the recorded behavior:

1. Bin12 - Using 12-hour long bins from the home cage, where we use means and coefficients of variation for each parameter as our input variables.
2. Bin1 - As above, but only using data from the 1st hour of the dark period each day.
3. String12 - 12-hour long string representations (from the home cage).
4. String1 - 1st-hour string representations (from the home cage).
5. Zone12 - Zone transition string (12-hour bin).
6. Zone1 - Zone transition string (1st-hour bin).
7. CatWalk - Gait analysis data.

We then tried to combine all this information sources in a single analysis.

Methods

In this research we tried to create a good classifier; a way to predict the strain of an animal, given the behavioral data from all the available information sources. That requires a model to be trained to find out what are the characteristic parameters for each strain. Therefore, even if it seems counter-intuitive at first glance to try to predict something we already know (the strain of each mouse), the process of doing so reveals a lot concerning the differences between groups.

A kernel is a similarity measure; when given a set of input variables and their corresponding classes, we can represent the relations between them using the kernel trick [1, 2]. For data fusion, we used a kernel combination approach [3, 4, 5, 6, 7]. The advantage of this method is that we can create one composite kernel for all the data, thus building one single classifier. The alternative would be to build a classifier for each separate data set, but then we would have to optimally combine them.

In short, the method works as follows. Consider a classification problem with S different feature spaces, from which we have D_s-dimensional predictor variables x_n^s, where $s = 1,\ldots, S$ and the corresponding target variables $t_n \in 1,\ldots, C$ for $n = 1,\ldots, N$, with C the number of classes and N the number of observations. Our approach lies in embedding the information from the different feature spaces into Hilbert spaces using the kernel trick, and then combining these into one composite kernel space:

$$K^{\beta\Theta}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{s=1}^{S} \beta_s K^{s\theta_s}(\mathbf{x}_i^s, \mathbf{x}_j^s),$$

where β is an S x 1 vector of weights and Θ is an $S \times D^s$ matrix containing the kernel parameters θ_s of the base kernels. Each of the base kernels K^s is constructed based on our prior knowledge about the data source and its desired similarity metric. Unknowns are estimated as in [8].

Results

The classification performance can be measured by the error rate (i.e. the percentage of misclassified samples) using cross-validation [9]. Part of the data is used to train the classifier which we then use to predict the strain for the "left out" data.

In Figure 2 below, one can see the improvement in performance of the combined data (all) compared to the classification achieved by the individual data sets. It needs to be mentioned that this is a four-class problem which explains the big variation in the estimates. Wild types are much easier distinguished from the transgenic groups than the transgenic groups amongst each other.

Conclusions

In this article we provide a way to combine behavioral data from different tests in a classification framework and we present the improved classification performance using this approach. The data sets used in this experiment are used only to present the new possibilities. The main advantage of this approach is its flexibility. In essence, one can combine any kind of data from the same subjects in the same analysis. At the same time, we propose a novel behavior representation and analysis which renders it possible to compare behaviors that are not necessarily synchronized.

Ethical statement

The aforementioned data resulted from an experiment that was approved by the Ethical Committee of Utrecht University.

Acknowledgements

Dimitris Fekas is funded by the European Commission in the framework of Neuromodel, a Marie Curie Initial Training Network (ITN).

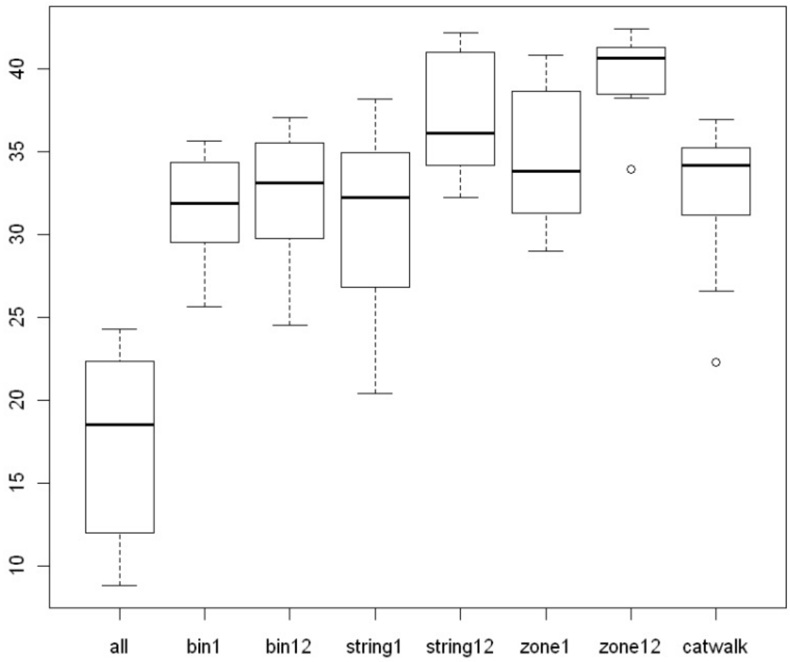


Figure 2. Error rate for the 4 class problem. The estimate of the combined data (all) is much lower than those obtained by each individual data set.

References

1. Shawe-Taylor, J., Cristianini, N., (2004). *Kernel methods for pattern analysis*. Cambridge University Press, 2004.

2. Scholkopf, B., Smola, A.. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press.

3. Damoulas, T., Girolami, M. (2009). Combining feature spaces for classification. *Pattern Recognition* **42**(11), 2671-2683.

4. Lanckriet, G., Cristianini, N., Bartlett, P., Ghaoui, L., Jordan, M. (2004). Learning the kernel matrix with semidefinite programming. *The Journal of Machine Learning Research* **5**, 27-72.

5. Ong, C. Smola, A. Williamson, B. (2005). Learning the kernel with hyperkernels. *Journal of Machine Learning Research* **6**, 1045-1071.

6. Lewis, D. Jebara, T. Noble, W. (2006). Support vector machine learning from heterogeneous data: an empirical analysis using protein sequence and structure. *Bioinformatics* **22**(22), 2753-2760.

7. Lee, W. Verzakov, S. Duin, R. (2007). Kernel combination versus classifier combination. *Multiple Classifier Systems (Proc. 7th Int. Workshop, MCS 2007, Prague, May 23-25, 2007), no. 4472*: 22-31.

8. Damoulas, T. Girolami, M. (2009). Pattern recognition with a bayesian kernel combination machine. *Pattern Recognition Letters* **30**(1), 46-54.

9. Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*, *Series B* **36**, 111-147.

A Markov Transition Score for Characterizing Interactive Behavior of Two Animals and its Application to Genetic Background Analysis of Social Behavior of Mouse

Toshiya Arakawa[1], Aki Takahashi[2], Akira Tanave[3], Satoshi Kakihara[4], Shingo Kimura[5], Hiroki Sugimoto[6]
Toshihiko Shiroishi[7], Kazuya Tomihara[8], Tsuyoshi Koide[9], Takashi Tsuchiya[10]

[1]Department of Statistical Science, The Graduate University for Advanced Studies (SOKENDAI), Tokyo, Japan.
[2,3,9]Mouse Genomics Resource Laboratory, National Institute of Genetics (NIG), Shizuoka, Japan.
[2,3,9]Department of Genetics, SOKENDAI, Shizuoka, Japan. [4,10]National Graduate Institute for Policy Studies, Tokyo, Japan.
[5]NEC Software Kyushu, Ltd., Fukuoka, Japan.
[6]Division of Genetic Therapeutics, Center for Molecular Medicine, Jichi Medical University, Tochigi, Japan.
[7]Mammalian Genetics Laboratory, NIG, Shizuoka, Japan.
[8]Faculty of Law, Economics and Humanities, Kagoshima University, Kagoshima, Japan.
[10]The Institute of Statistical Mathematics, Tokyo, Japan.

[1]arakawa@ism.ac.jp, [2]aktakaha@lab.nig.ac.jp, [3]atanabe@lab.nig.ac.jp, [4]skakihara@gmail.com
[5]shingo.k-m-r.0715@voice.ocn.ne.jp, [6]hisugimo@lab.nig.ac.jp, [7]tshirois@lab.nig.ac.jp
[8]tomihara@leh.kagoshima-u.ac.jp, [9]tkoide@lab.nig.ac.jp, [10]tsuchiya@grips.ac.jp

1. Introduction

Characterizing interactive behavior of two animals is an important and interesting subject of research in social behavior study of animals. In this paper, we propose a score called Markov transition score to characterize interactive behavior of two animals, and apply this score to analyze social behavior of mouse.

Takahashi et al. [1][2] conducted a genetic study of mice to understand a mechanism by which individuals develop differences in social behavior. The study was based on (C57BL/6J(B6)-based) inter-subspecific consomic strains established from MSM/Ms (MSM) and C57BL/6J. Here, consomic strain is a strain where one chromosome of C57BL/6J is replaced with the same chromosome of MSM/Ms. Comparing social interactive behaviors of each consomic strain with C57BL/6J and MSM/Ms, we may well speculate the role of each chromosome in determining social interacting nature of mouse. In their study, two littermates with the same consomic strain, sex, aged 10 weeks were introduced into a novel new open field for 10 minutes and their behavior was recorded through a video tracking system (Image SI; O'Hara & Co., Tokyo, Japan; see Figure 1), which was based on a public domain image processing and analysis program, NIH Image, from the National Institute of Health (USA). The location of two mice is recorded every 1/3 second so that a video tracking data consists of 1800 frames. As shown in Table 1, a total of 530 video data was collected, where approximately 10 video data are available for each consomic strain and sex. In Takahashi et al. [1][2], a genetic mapping was conducted based on total duration time and frequency of contact of each consomic strain and sex.

We will analyze the same set of data from a different viewpoint utilizing a two-state Markov model[3]. We consider that the pair is in either "indifference" state or "interactive" state, and that the state sequence obeys to a Markov process. The model has two parameters, i.e., the transition probabilities from "indifference" to "interactive", and from "interactive" to "indifference." Therefore, we can represent each pair or each consomic strains as a point on a 2D plane. Looking the behavior of the consomic mice pairs through this model, we will demonstrate that a mixture nature of consomic strains are "quantitatively" clearly seen with the scattering plots of B6, MSM and consomic strains on the 2D plane. In the analysis, we need to identify the state of the pair at each tracking time point. This is an intelligent task which ideally is to be done by a human expert. In this study, we employed a hidden Markov model to automate this task. This paper is organized as follows. We introduce the method of application of Markov model and hidden Markov model in Section 2. In Section 3, we demonstrate the analyzed result. Section 4 is conclusion.

Ethical statement : In this study, All procedures were conducted in accordance with the NIG (National Institution of Genetics) guidelines and approved by the Institutional Committee for Animal Care and Use.

2. Markov Transition Score

In this section, we explain the two-state Markov model. The location of the two mice is recorded every 1/3 second in the duration of 10 minutes and therefore we have a total of 1800 tracking time points for each video data. We consider that a mice pair is either in one of the two states, namely, “indifference” and “interactive” at each tracking time point. We mean by the state “interactive” that the two mice are interacting each other like sniffing, genital grooming, following, etc., whereas by the state “indifference” that the two mice pay little attention to each other. We consider that, when moving from one tracking time point to the next, the transition between the two states occurs according to the Markov process described as below.

Let p_{00}, p_{01}, p_{10} and p_{11} be, respectively, the transition probability from “indifference” to “indifference”, from “indifference” to “interactive”, “interactive” to “indifference” and “interactive” to “interactive”. There are just two independent parameters in this model. We choose p_{01} and p_{10} as independent parameters. Let (μ_0, μ_1) be the stationary probability of this Markov process. It is easy to see that

$$\mu_0 = \frac{p_{10}}{p_{01}+p_{10}}, \quad \mu_1 = \frac{p_{01}}{p_{01}+p_{10}}$$

This means that all Markov models with the same stationary probability lie on a ray emanating from the origin in (p_{01}, p_{10}) plane. Then, what is the difference between two Markov models with the same stationary probability? They are different in the number of switching between the two states. The two models exist on the same ray on (p_{01}, p_{10}) plane. Intuitively, switching occurs more often in the model farther from the origin, though the two models have the same stationary probability. The probability of occurrence of switching per step (a unit tracking interval) is given by the following formula:

$$\mu_0 p_{01} + \mu_1 p_{10} = \frac{2 p_{01} p_{10}}{p_{01}+p_{10}}$$

Estimation of Markov transition probability from the tracking data of a pair is done as follows.
(a) Mark the state of the pair at every time step.
(b) Count the number of transitions n_{00}: the number of transitions of “indifference” → “indifference”, n_{01}: “indifference” → “interactive”, n_{10} : “interactive” → “indifference” and n_{01} : “interactive” → “interactive”
(c) Estimate p_{00}, p_{01}, p_{10}, p_{11} according to the following formulas:

$$p_{00} = \frac{n_{10}}{n_{01}+n_{10}}, \quad p_{01} = \frac{n_{01}}{n_{01}+n_{10}}, \quad p_{10} = \frac{n_{10}}{n_{01}+n_{10}}, \quad p_{11} = \frac{n_{11}}{n_{01}+n_{10}}$$

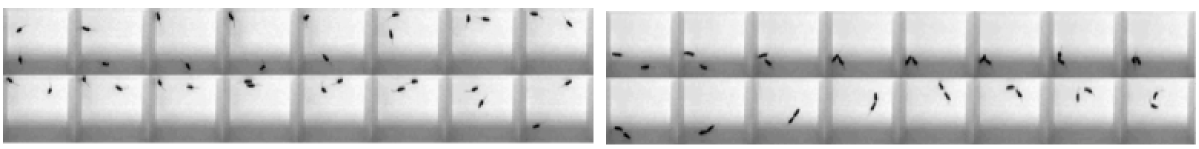


Figure 1. Sequential snapshots of a mouse pair (left: indifference, right: interactive).

Table 1.

(Consomic) Mice	B6	Chr 1	Chr 2C	Chr 2T	Chr 3	Chr 4	Chr 6C	Chr 6T	Chr 7	Chr 8	Chr 9	Chr 11
Male	14	13	10	11	11	11	10	12	13	11	11	10
Female	11	11	11	11	11	11	11	11	11	10	10	10
Total	28	24	21	22	22	22	21	23	24	21	21	20
(Consomic) Mice	**Chr 12C**	**Chr 13**	**Chr 14**	**Chr 15**	**Chr 16**	**Chr 17**	**Chr 18**	**Chr 19**	**Chr XC**	**Chr XT**	**Chr Y**	**MSM**
Male	11	11	11	11	16	11	6	12	10	10	11	10
Female	11	10	10	12	12	11	3	10	10	10	0	10
Total	22	21	21	23	28	22	9	22	20	20	11	20

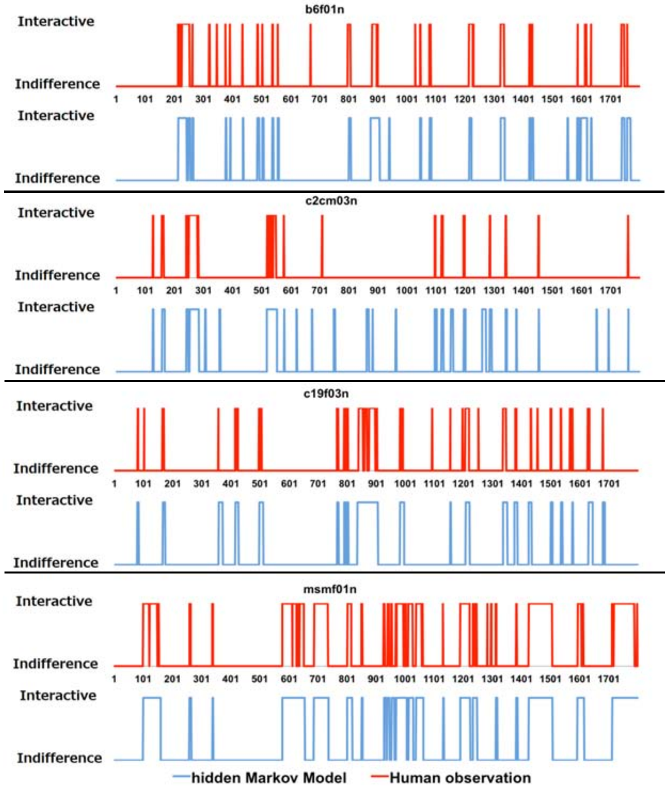


Figure 2. Comparison of results by time-series.

Though this procedure appears simple, marking of the two states watching the recorded video of a mice pair is a highly intelligent task, and ideally this should be done by a human expert. However, it is difficult to do so if we have so many video data to analyze. We developed a system which automates this task utilizing the hidden Markov model (HMM)[4][5][6][7]. We omit the details of HMM, but show the time series of human judgement and the judgement of HMM for 4 pairs (Figure 2). It is seen that HMM judgment agree fairly well with human observation. In c2cm03n, HMM tends to take more interactive states than human observation. We checked the video and confirmed that the judgment by HMM is reasonable. A total of 530 mice pair data are available, and we calculated p_{01} and p_{10} for each pair with HMM. About ten data are available for each consomic strain and sex, and we define the *Markov transition score (MTS) of a consomic strain and sex as the average of* (p_{01}, p_{10}) *over all pairs of that consomic strain and sex.* MTS of each consomic strain and sex is shown in Figure 3.

3. Results

From Figure 3, we see that MTS of male consomic strains are either (i) located around B6, or (ii) located along in the downward-sloping straight line segment connecting B6 and MSM, except for C6c. From this figure, we observe the following:

1. The genetic mixture nature of consomic strain is indeed reflected “quantitatively” in Markov transition probability plane. This is the most interesting and important observation we made through this analysis.
2. The chromosomes associated with the consomic strains belonging to (i) seem to have little influence on interactive behavior, while the second one do have strong effects. It is interesting that that one replacement of chromosome changes so drastically interactive behavior of mice.
3. The scattering plot also suggests that the Markov probability from “interactive” to “indifference” is determined uniquely when Markov probability from “indifference” to “interactive” (from “interactive” to “indifference”) was determined.

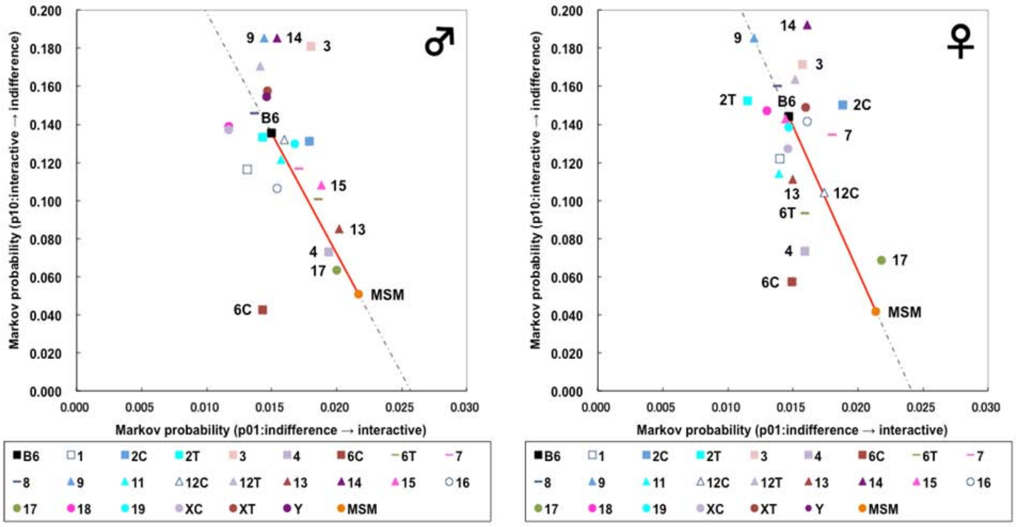


Figure 3. Markov transition score.

4. MTS of C6c is located apart from the straight line connecting B6 and MSM. This position of C6c makes a good contrast with other chromosomes in the following meaning. Basically, we observe “higher” Markov transition probability of “indifference” to “interactive” when a consomic strain has “lower” Markov transition probability of “interactive” to “indifference”. The strain C6c violates this rule: transition probability of “indifference” to “interactive” remains almost the same in spite of that transition probability of “interactive” to “indifference” is lowered. This suggests a special function of chromosome C6c in exposition of interactive behavior of mouse, and definitely deserves further study.

4. Conclusion

In this study we proposed the Markov transition score to capture interactive behavior of two animals and applied the score to analyze the role of each chromosome on social behavior of mouse. We think this score can be useful in many situations including human interaction behavior.

References

1. Takahashi, A., Nishi, A., Ishii, A., Shiroishi, T., Koide, T. (2008) Systematic analysis of emotionality in consomic mouse strains established from C57BL/6J and wild-derived MSM/Ms. *Genes, Brain and Behavior* **7**, 849–858.
2. Takahashi, A., Tomihara, K., Shiroishi, T., Koide, T. (2009). Genetic Mapping of Social Interaction Behavior in B6/MSM Consomic Mouse Strains. *Behav. Genet.* **40**, 366–376.
3. Arakawa, T. et al. (2012). Automatic Segmentation of Mice States Using Hidden Markov Model and Characterization of Mouse Strain with 2-state Markov Model. *Proceedings of the Institute of Statistical Mathematics* **60**(1) (Japanese, in press).
4. Huang, X.D. Ariki, Y., Jack, M.A. (1990). *Hidden Markov Models for Speech Recognition*. Edinburgh University Press, Edinburgh.
5. Eddy, S.R. (1996). Hidden Markov Models. *Current Opinion in Structural Biology* **6**, 361-365.
6. Eddy, S.R. (2004). What is a hidden Markov model? *Nature Biotechnology*, **22**(10), 1315-1316.
7. Rabiner, L.R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. IEEE*, **Vol. 77**(2), 257-228.

Speech Inversion with Acoustic Classification

C. Neufeld

Department of Speech-Language Pathology, Oral Dynamics Lab, University of Toronto, Toronto, Canada.
christopher.neufeld@utoronto.ca

Introduction

Inferring the shape of the vocal tract from speech acoustics is a complex machine learning problem that has been explored by many researchers over the past decades using various methodologies [1-3]. This problem is central for several domains of research. It has been argued that something akin to a `motor simulation' is an important component of human speech perception, and therefore speech inversion may be an intrinsic component of humans' ability to parse speech [4]. Modeling such behavior accurately is an important part of this line of research in human perception. It has also been demonstrated that the inclusion of articulatory data improves automatic speech recognition, and improvements in speech inversion may translate into improvements in automatic speech recognition [5]. This paper presents a novel method for improving speech-inversion with Mixture Density Networks (MDNs) [6] by classifying acoustic space into distinct aerodynamically defined classes using a Hidden Markov Model (HMM) [7], and training a separate speech inverter for each acoustic class.

Aerodynamic modes and acoustic nonlinearities

There are two types of nonlinearities which complicate the project of speech inversion. The first is related to the concept of a phase transition, where the gross characteristics of a system's output change abruptly as some control variable changes continuously. We can easily identify four basic aerodynamic phases in speech: laminar flow through the mouth during vowels, liquids and glides (here symbolized as V), laminar airflow through the nose (N), turbulent airflow during aspiration, fricatives and affricates (H), and zero airflow as during the production of a stop consonant (C). Transition between these phases may be abrupt: for instance, an infinitesimally small increase in subglottal air pressure may result in an abrupt change from laminar to turbulent airflow exiting the vocal tract. The second type of nonlinearity can be observed within phases, where large changes in the shape of the vocal tract may result in negligible spectral changes, or conversely, small changes in the shape of the vocal tract may result in significant changes [8]. These characteristics of speech acoustics greatly complicate the process of speech inversion. Here a novel approach to this problem is presented by training a set of specialist speech inverters which only estimate the vocal tract shape for some specific acoustic class. In this model, inversion consists of two steps. For some acoustic speech data, a classifier identifies which portions of the speech signal belong to which acoustic class [C, V, N, H]. Then, the trajectories of speech articulators are estimated from acoustics by specialist inverters: e.g. all those portions of the acoustic signal classified as C are inverted by a C-inverter, all those portions classified as V are inverted by a V-inverter, etc. By classifying acoustic space and training different inverters for the four classes identified here, each specialized inverter only has to learn within-phase nonlinearities, while a more general speech inverter must learn both within- and across-phase nonlinearities. This approach allows the optimization of the acoustic feature set from which articulation is derived for each acoustic class. It is found that for some acoustic classes, acoustic features derived with a relatively large window allow for the best reconstruction of articulation, while for other classes, a relatively small window size provides the best acoustic features for speech-inversion.

Data

All data comes from the TORGO database [9] which consists of speech data from video sessions, as well as 3D Electro-magnetic Articulograph (EMA) sessions, from speakers with cerebral palsy or amyotrophic lateral sclerosis, and matched controls. Here only data from matched controls is used, giving three female and four male speakers. Subjects were recorded reading English text from a screen. Stimuli include non-words, short words, restricted sentences and unrestricted sentences. The acoustic data from the video sessions was used for training and testing the acoustic classifier. The acoustic and articulatory data from the EMA sessions was used to train, validate and test the speech inverters. Acoustic data was recorded with an array microphone at 22.1 kHz, downsampled to 16 kHz, and pre-emphasized. Every 5 ms a vector of acoustic features was extracted using Hamming windows with widths ranging from 10 to 200 ms. These features were 12th order Linear Predictive Coefficients, 12th order Mel-filter Cepstral Coefficients, root-mean-squared energy, and these features' 1st and 2nd time derivatives. Articulatory data was recorded with the Carsten's AG-500 Electro-Magnetic Articulograph (EMA) at a sampling rate of 200 Hz. Lateral and vertical displacement of sensors attached to the tongue tip, body and dorsum, upper and lower lip, and jaw (TT, TB, TD, UL, LL, JA) were corrected for head movement, and filtered with an 11-point Butterworth filter with a cut-off frequency of 6 Hz. All data was normalized within subjects, and transformed by a sigmoid function of the form $1/(1+e^{-x})$ to lie on the interval [0,1].

Model outline

The speech inversion model consists of an HMM, which classifies acoustic data, and a set of MDNs, each optimized to reconstruct the trajectory of a particular articulatory channel from each acoustic class. So, for instance, in reconstructing the trajectory of the lower lip during the production of the word cancer [kʰænsɹ], the HMM would ideally assign the hidden state sequence CHVNHV given the sequence of acoustic feature vectors extracted from the waveform of the word. Those portions of the acoustic signal classified as C would be inverted by an MDN trained and optimized exclusively to reconstruct the articulation of the lower lip during consonants, those portions of the acoustic signal classified as V would be inverted by an MDN trained and optimized to reconstruct articulation of the lower lip during vowels, and so forth. The construction of this model is detailed in the following sections. In outline, the data is divided into 5 sets. Audio data from video sessions is used to train the HMM, and is divided into a training set and test set. Audio and EMA data are used to construct the MDNs, and are divided into a training set, a validation set used to select the best MDN for each acoustic class, and a test set to compare the performance of this model against a baseline model. The baseline model is a set of MDNs which have been trained on the same acoustic/EMA dataset but without any acoustic classification: i.e. the baseline model is trained to reconstruct the articulation of particular EMA channels for all types of speech acoustics, rather than specific acoustic classes.

Classification

Audio speech data from the video sessions in the TORGO database was manually annotated by visual inspection of the spectrogram into four classes described above: C(onsonant), H (turbulence), N(asal) and V(owel). Four subjects (2M, 2F) contributed 100 utterances each. The remaining 3 subjects were excluded either because they did not participate in the video session, or there were significant problems with the audio recording. These

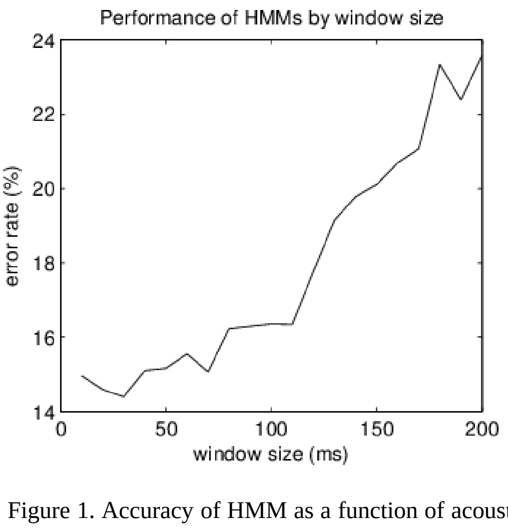


Figure 1. Accuracy of HMM as a function of acoustic window size. The best HMM is trained using acoustic features derived from a 30 ms window.

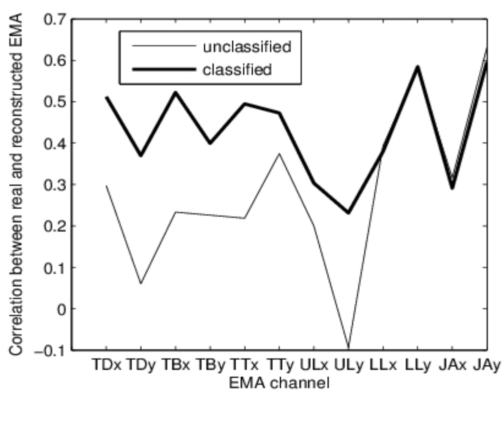


Figure 2. Correlation between real and reconstructed EMA by channel. For most channels, articulation is better reconstructed by specialist inverters operating over classified acoustic data (bold line) than by generic inverters operating over unclassified acoustic data (light

annotations were used to train a set of 4-state HMMs (each hidden state corresponds to one of the classes) with a 10-component mixture-of-Gaussians output (the acoustic features). These HMMs were trained using acoustic feature sets derived from different window sizes ranging from 10-200 ms. 90% of the data was used for training, and the remaining 10% of the data was used to test the accuracy of each HMM, and select the window size which produces the classifier with the highest accuracy. Figure 1 shows the error rate for each HMM as a function of window size. The best HMM is one where acoustic features are derived from a 30 ms window, with an overall error rate of 14.4%. Error rates for each class are [C=19.5% , V=14.4%, N=29.1%, H=10.1%].

Inversion

300 utterances each from all subjects were used for the inversion step. One subject's data was reserved for testing, and was not used for training or validation. 70% of the data from each remaining subject was used for training, 20% for validation and 10% for testing. The training data was used to construct sets of MDNs with 100 hidden units and 10 mixture components using acoustic features derived from windows ranging in size from 10 to 200 ms. For the unclassified data, the validation data was used to select the optimal acoustic window size for each EMA channel. For the classified data, validation data was used to select the optimal acoustic window size for each EMA channel *and* each acoustic class. The validation procedure for selecting the best speech inverter for the j^{th} acoustic class and k^{th} EMA channel: $cMDN_{j,k}$, is given here. The trajectory of the k^{th} EMA channel estimated by the MDN trained on acoustic features derived from the i^{th} window size for the j^{th} acoustic class is:

1. $EMA_{i,j,k} = [MDN_i(AC_j)]_k$

C is an i x j x k matrix storing the correlation coefficient between the true articulation of the k^{th} EMA channel underlying the acoustics of the j^{th} acoustic class, and the estimated EMA articulation derived from the MDN trained on the jth acoustic class and i^{th} window size

2. $C_{i,j,k} = corr(EMA_{j,k} , EMA_{i,j,k})$

The optimal MDN for the j^{th} acoustic class and k^{th} EMA channel is defined as that MDN for which the correlation between the true and estimated EMA is highest:

3. $cMDN_{j,k} = argmax_i(C_{j,k})$

Construction of the optimal set of MDNs for unclassified acoustics proceeds in exactly the same way, except without distinctions of acoustic class. The articulation of the k^{th} EMA channel estimated by the MDN trained on acoustic features derived using the i^{th} window size is:

4. $EMA_{i,k} = [MDN_i(AC)]_k$

C is an i x k matrix storing the correlation coefficient between the true articulation of the k^{th} EMA channel, and the estimated EMA articulation derived from the MDN derived from the i^{th} window size

5. $C_{i,k} = corr(EMA_k , EMA_{i,k})$

The best MDN for the k^{th} EMA channel is the MDN with the highest correlation between true and estimated EMA:

6. $uMDN_k = argmax_i(C_k)$

Results

Figure 2 shows the average correlation coefficient between the real and estimated articulations for each EMA channel, for classified and unclassified speech inverters. It can be seen that, for all but the lower lip and jaw channels, classified speech inverters perform better than their unclassified counterparts. A two-way ANOVA (TYPE x CHANNEL) revealed significant main effects for both TYPE [$F(1,4776) = 11.34$, $p<10^{-16}$] and CHANNEL [$F(11,4776) = 20.21$, $p<10^{-37}$], as well as a significant interaction [$F(11,4776) = 1.64$, $p<0.0005$].

Figure 3 shows the window size used to derive the acoustic features which are used for the best MDN (as determined by the validation step) for each EMA channel and acoustic class. It can be seen that, for the unclassified inverter, almost every channel has the same optimal window size, suggesting that this is an average. However, for the classified inverters, there is a very wide range of diversity, indicating that one of the means by which classified inversion achieves its superior performance is by being able to perform fine-grained optimization over specific acoustic classes and articulatory channels. For example, it can be seen that the position of articulators during turbulent airflow (class H) is generally best reconstructed using a very large window size, while for laminar oral airflow (class V) a medium-to-small window size is optimal. These results show that classifying acoustic speech data prior to inversion, and training specialist inverters to derive articulation for each acoustic class improves speech inversion. The variation of window sizes used for inverters across acoustic classes indicates that for every acoustic class, there are two levels of optimization to exploit: the window size which derives the acoustic features which are the input to the MDN, and the training of the MDN itself. This is a promising initial result, and further work is needed to explore whether these results can be replicated or improved upon with different classifier and inverter architectures.

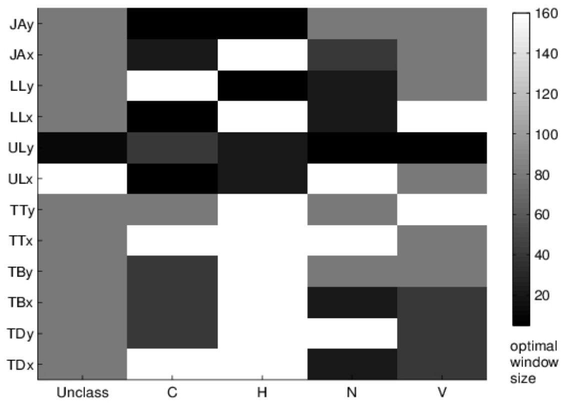


Figure 3. Window size used for each MDN by EMA channel and acoustic class. Most MDNs for unclassified data have the same window size, while the MDNs for classified data (columns C,V,N,H) employ acoustic features derived from a variety of window sizes.

References

1. Papcun, G., Hochberg, J., Thomas, T.R., Laroche, F., Zacks, J. and Levy, Simon. (1992). Inferring articulation and recognizing gestures from acoustics with a neural network trained on x-ray microbeam data. *Journal of the Acoustical Society of America* **92**(2), 18688-18700.

2. Hogden, J., Rubin, P., McDermott, E., Katagiri, S. & Goldstein, L. (2007). Inverting mappings from smooth paths through Rn to paths through Rm: A technique applied to recovering articulation from acoustics. *Speech Communication* **49**, 361-383.

3. Rudzicz, F. (2010). Adaptive kernel canonical correlation analysis for estimation of task dynamics from acoustics. *Proceedings of the 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP'10).

4. Liberman, A. M., Cooper, F. S., Shankweiler, D. P., Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review* **74**(6), 431-461.

5. Ghosdh, P.K., Narayan, S. (2011). Automatic speech recognition using articulatory features from subject-independent acoustic-to-articulatory inversion. *Journal of the Acoustical Society of America* **130**(4): EL251-EL257.

6. Bishop, C.M. (1994). *Mixture density networks*. Technical Report. Aston University, Birmingham. (unpubl.).

7. Rabiner, L.R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE* **77**(2), 257-286.

8. Stevens, K.N. 1989. On the quantal nature of speech. *Journal of Phonetics* **17**: 3-4.

9. Rudzicz, F., Namasivayam, A.K., Wolff, T. (2011). The TORGO database of acoustic and articulatory speech from speakers with dysarthria. *Language Resources and Evaluation.* DOI 10.1007/s10579-011-9145-0

Validating the Cross-Validation: A 3-Dimensional Model for Multiple Informant Data (3D-MMID)

J.A. Malik[1,2], H.M. Koot[2]

[1]*National Institute of Psychology, Quaid-i-Azam University, Islamabad, Pakistan, [1]ja.malik@yahoo.com*

[2]*Developmental Psychopathology, VU University, Amsterdam, The Netherlands. [2]j.m.koot@vu.nl*

Conceptual validation is always an issue of measurement in social and clinical sciences. Social scientists and clinicians are of the view that sometime single source is not enough to cater a concept [1] either due to potential biasness or partial involvement/contact with the target population [2]. Although almost all social science disciplines face the issue at some time, mainly research in developmental psychology & psychopathology meets the said challenge often [5, 6]. That is because many of the researched constructs/variables in question can be measured only by involvement of more than one source [2]. For instance, to measure adolescents' behavioral problems data may be required from home environment as well from school environment. Another variation might be that from the same environment (i.e., home) data is required from both parents on the same construct as studies shows that father and mother reports differ for a child's underlying state [11]. This issue is of the concern in a situation where subjective involvement is required rather than objective measurement of a physical status or medical condition [10]. For instance, exploration on a question of interest "how much stress is faced by an individual?" may require involvement of all stake holders in one's life (i.e., self-report, parent / partner / practitioner report, teacher / colleague / boss report).

A variety of solutions to the multi-informant data have been presented [2,10]. Some suggested the use of optimal source [4,13], and ignoring others which actually means there is no need to collect data from other sources. The decision that which source is optimal shall be made at the time of designing the project [10]. Another frequently used method is to use data from all sources separately [7,9], which most of the time increases confusion rather than converging the problem. As concluded in a review article using multi-informant, that there is low to medium level of correlations between informants (i.e., Teacher-Parents = .27; Self-Peer = .26 etc.) [2]. Correlations only for the same informants appeared above .50 (i.e., Parent-Parent = .59; Teacher-Teacher = .64 etc.) [2]. Another way to combine multi-informant data is pooling. Various methods of pooling have been introduced and algorithms have been developed to commence a solution for multi-informant data [4,10 ,12]. For instance the "And" "Or" rules, the "And" rule suggest that information is accepted only if presented in reports of all informants whereas "Or" rule suggest that an information is considered if any of the source reported it [10,13]. The "And" rule costs sensitivity and the "Or" rule costs specificity. Although depending on the nature of the research question any of the method can be a more preferable solution over the other yet this type of pooling offer solution only for dichotomous scales/variables.

Another method of pooling is averaging the scores [3], which can be used for continuous scales but has its' own pros and cons. This type of pooling is seriously affected with outliers; additionally it loses a lot of information. Let's consider an example: on a five point Likert scale an adolescent (A) is rated at 4 by his teacher and rated 2 by his parent, the resulting average score will be 3, on the other hand an individual (B) who is rated 3 by both parent and teacher also results on an average score of 3. None of the informant rated the two cases as same but using the average method, the researcher considers both cases as same which is not only loss of information, rather it also indicates that to the researcher none of the informant is reliable as the method researcher used is not representative of any or both of the informants.

Others presented more complex algorithms, for optimal use of multi-informant data [8,10]. Most of the solutions presented for composite score and to the extent of my knowledge, so for there is not a single method addressing the issue at the scale level. This study is an effort to develop a method for use of multi-informant data at the scale level hence retaining all the scale characteristics. The study is aimed to advance the pooling methods by overcoming some of its shortcomings. In terms of optimal informant, a method is presented which accounts for informant error and in terms of averaging the method is modified to preserve the distinct position of each individual.

The example being used here include parent rated and teacher rated problem behaviors. Both parent and teacher reported on a 5 point Likert type scale. The pooling was conducted on three steps and is reported as such.

Step 1

An agreement scale was computed based on the following matrix. The matrix is produced by cross-tabulating the two informants.

Agreement		Parents				
		1	2	3	4	5
Teacher	1	5	4	3	2	1
	2	4	5	4	3	2
	3	3	4	5 (B)	4	3
	4	2	3 (A)	4	5	4
	5	1	2	3	4	5

Although the example presents two informants, the SPSS syntax to compute an agreement scale can be extended to as many informants as included in the study. This step accomplishes the horizontal stretch of the concept/construct.

Step 2

At the second step, the researcher needs to decide whether he/she want to use one optimal source or both sources. In case of using one optimal source the data can be corrected for potential bias simply by multiplying the agreement score with the score of the informant. For instance, in parent report, conflict on item was reported by parents as 4, whereas adolescent reported conflict on the same item as 3. Using the cross-tabulation, the agreement score for the item emerges as 4. If the researcher now wants to use parent report, the parent reported intensity i.e., 4 shall be multiplied with the agreement score hence 4*4=16. To use adolescent perspective, adolescent reported intensity shall be multiplied with the agreement score that is 3*4 =12. This type of pooling corrects report of an optimal informant for the potential biasness. But in case, the use of multi-informant is based on the fact that one informant is not sufficient to cater the concept, then this type of pooling is not useful. In such situation, the second step would be to generate an average intensity measure simply by averaging two reports i.e., (i1+i2)/2. An average report measure of such a five points Likert scale looks as shown below.

Intensity		Parents				
		1	2	3	4	5
Teacher	1	1	1.5	2	2.5	3
	2	1.5	2	2.5	3	3.5
	3	2	2.5	3 (B)	3.5	4
	4	2.5	3 (A)	3.5	4	4.5
	5	3	3.5	4	4.5	5

This step accomplishes the vertical stretch of the concept/construct.

Step 3

To grab the three dimensional sight of concept/construct, a final step involves combining the two matrices. This step can simply be accomplished by multiplying the agreement scale with the average scale, resulting into a 3D-MMID (a 3 dimensional model for multi informant data).

Agreement * Intensity	5	6	6	5	3
	6	10	10	9	7
	6	10	15 (B)	14	12
	5	9 (A)	14	20	18
	3	7	12	18	25

Considering the case reported earlier: on a five point Likert scale an adolescent (A) is rated at 4 by his teacher and rated 2 by his parent, the resulting average score will be 3, on the other hand an individual (B) who is rated 3 by both parent and teacher also results on an average score of 3. None of the informant rated the two cases as same but using the average method, researcher considers both cases as same. Now look at the situation, the first individual gets a score of (3*3)=9, whereas the second individual gets a score of (5*3)=15, both the individual are reported different by both the raters and their score as computed by 3D-MMID also presents them as different individuals compare to the average method which reports the two cases similar. Though not the perfect solution, the resulting matrix solves most of the issues of the average method yet a conceptual analysis of the resulting 3D-MMID matrix raises other issues. Analyzing the first row of the resulting matrix, we can see that 2nad and 3rd cell of the row presents two cases as similar but they are not similar. That is they are rated similar by teacher but rated different by teachers. Another more severe problem emerges by comparing 1st and 5th cells of first row. The 1st cell is presenting a subject for whom both teacher and parents agree on the lowest probability of presence of a problem behavior. Comparatively, 5th cell present a case where though one source (parent) report lowest probability of presence of that problem behavior yet at least one source (teacher) reports the highest probability of presence of the same problem behavior. A conceptual analysis of the two situation suggest that the child facing the situation in 5th cell is more vulnerable to have the particular problem behavior than the child facing the situation presented in 1at cell, yet our 3D-MMID matrix suggest opposite results i.e., rating the child in 1st cell at 5 and rating the child in 5th cell at 3. To solve these issues agreement matrix was revised by placing a magnitude of 0.6 for lowest agreement and 1.0 for highest agreement. The resulting agreement matrix is presented below:

Agreement		Parents				
		1	2	3	4	5
Teacher	1	1.0	0.9	0.8	0.7	0.6
	2	0.9	1.0	0.9	0.8	0.7
	3	0.8	0.9	1.0 (B)	0.9	0.8
	4	0.7	0.8 (A)	0.9	1.0	0.9
	5	0.6	0.7	0.8	0.9	1.0

Step 3 the multiplication step was revised to generate new interaction matrix which resulted as:

Agreement * Intensity	1	1.35	1.6	1.75	1.8
	1.35	2	2.25	2.4	2.45
	1.6	2.25	3 (B)	3.15	3.2
	1.75	2.4 (A)	3.15	4	4.05
	1.8	2.45	3.2	4.05	5

The final 3D-MMID-R matrix resolved all the issues raised in the previous version of the matrix. As presented in matrix 5, not only that subject A and B are rated differently, values in different cells are also more representative of the respective position of subjects.

Example Scales

Two data sets are presented as example cases. Adolescents with type 1 diabetes and their parents rated Parent-Child conflicts on a 17 items, 5 point Likert type scale and a 29 items family responsibility questionnaire. Descriptives of the scales are presented for 7 types of computed scores. 1. Parents reports, 2. Adolescents report, 3. Average score, 4. Parents report corrected for adolescents perspective, 5. Adolescent report corrected for parents perspective, 6. 3D-MMID scores using the 3rd matrix, and 7. 3D-MMID-R scores using the 5th Matrix.

Parent-Child Conflict:

	N	Mean	S.E	S.D	Variance	Skew	S.E	Kurt	S.E	Alpha
Adolescent report	411	29.30	0.36	7.24	52.43	0.92	0.12	1.49	0.24	0.84
Parent reported	432	30.44	0.43	8.98	80.57	0.86	0.12	0.46	0.23	0.81
Average score	408	29.79	0.35	6.99	48.87	0.86	0.12	1.28	0.24	0.84
Parents corrected for adolescent	408	27.11	0.32	6.42	41.17	1.06	0.12	2.44	0.24	0.82
Adolescents corrected for parents	408	27.83	0.36	7.28	53.04	0.79	0.12	0.87	0.24	0.93
3D-MMID Conflict	408	125.75	1.38	27.87	776.64	1.43	0.12	4.36	0.24	0.79
3D-MMID-R Conflict	408	27.47	0.30	6.09	37.04	1.14	0.12	2.76	0.24	0.82

Family Responsibility:

	N	Mean	S.E	S.D	Variance	Skew	S.E	Kurt	S.E	Alpha
Adolescent report	386	71.70	0.96	18.77	352.24	0.27	0.12	-0.20	0.25	0.92
Parent reported	386	93.96	1.01	19.77	391.00	-0.13	0.12	-0.19	0.25	0.93
Average score	322	83.35	0.33	5.93	35.11	-0.26	0.14	1.05	0.27	0.73
Parents corrected for adolescent	322	61.09	0.75	13.37	178.75	-0.35	0.14	-0.19	0.27	0.73
Adolescents corrected for parents	322	46.89	0.72	12.86	165.28	0.33	0.14	-0.03	0.27	0.88
3D-MMID responsibility	322	206.30	2.48	44.44	1974.91	0.04	0.14	-0.39	0.27	0.68
3D-MMID-R responsibility	322	53.99	0.33	5.86	34.39	-0.24	0.14	1.02	0.27	0.72

Conclusion

The polling method presented as 3D-MMID-R captures individual differences very well and as well presents a reliable solution to multi-informant data and shall be preferred on average method.

References

1. Achenbach, T. M. (1993). Implications of multiaxial empirically based assessment for behavior therapy with children. *Behavior Therapy* **24**, 91-116.

2. Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. *Psychol Bull* **101**(2), 213-232.

3. Allen, J. P., Kuperminc, G., Philliber, S., & Herre, K. (1994). Programmatic prevention of adolescent problem behaviors: the role of autonomy, relatedness, and volunteer service in the Teen Outreach Program. [Comparative Study Research Support, Non-U.S. Gov't]. *Am J Community Psychol* **22**(5), 617-638.

4. Bird, H. R., Gould, M. S., & Staghezza, B. (1992). Aggregating data from multiple informants in child psychiatry epidemiological research. [Research Support, U.S. Gov't, P.H.S.]. *J Am Acad Child Adolesc Psychiatry* **31**(1), 78-85. doi: 10.1097/00004583-199201000-00012

5. Breton, J. J., Bergeron, L., Valla, J. P., Lepine, S., Houde, L., & Gaudet, N. (1995). Do children aged 9 through 11 years understand the DISC version 2.25 questions? *Journal of the American Academy of Child Adolescent Psychiatry* **34**(7), 946–956.

6. Edelbrock, C., Costello, A. J., Dulcan, M. K., Kalas, R., & Conover, N. C. (1985). Age differences in the reliability of the psychiatric interview of the child. *Child Development* **56**, 265–275.

7. Feldman, S. S., & Weinberger, D. A. (1994). Self-restraint as a mediator of family influences on boys' delinquent behavior: a longitudinal study. [Research Support, Non-U.S. Gov't]. *Child Dev* **65**(1), 195-211.

8. Fitzmaurice, G. M., Laird, N. M., Zahner, G. E., & Daskalakis, C. (1995). Bivariate logistic regression analysis of childhood psychopathology ratings using multiple informants. [Research Support, U.S. Gov't, P.H.S.]. *Am J Epidemiol* **142**(11), 1194-1203.

9. Gould, M. S., Fisher, P., Parides, M., Flory, M., & Shaffer, D. (1996). Psychosocial risk factors of child and adolescent completed suicide. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. *Arch Gen Psychiatry* **53**(12), 1155-1162.

10. Horton, N. J., Laird, N. M., & Zahner, G. E. P. (1999). Use of multiple informant data as a predictor in psychiatric epidemiology. *International Journal of Methods in Psychiatric Research* **8**(1), 6-18.

11. Jensen, P. S., Traylor, J., Xenakis, S. N., & Davis, H. (1988). Child psychopathology rating scales and interrater agreement: I. Parents' gender and psychiatric symptoms. *J Am Acad Child Adolesc Psychiatry* **27**(4), 442-450. doi: 10.1097/00004583-198807000-00012

12. Offord, D. R., Boyle, M. H., Racine, Y., Szatmari, P., Fleming, J. E., Sanford, M., & Lipman, E. L. (1996). Integrating assessment data from multiple informants. [Research Support, Non-U.S. Gov't]. *J Am Acad Child Adolesc Psychiatry* **35**(8), 1078-1085. doi: 10.1097/00004583-199608000-00019

13. Zahner, G. E., Pawelkiewicz, W., DeFrancesco, J. J., & Adnopoz, J. (1992). Children's mental health service needs and utilization patterns in an urban community: an epidemiological assessment. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. *J Am Acad Child Adolesc Psychiatry* **31**(5), 951-960. doi: 10.1097/00004583-199209000-00025

How to Describe the Process of the Establishment of a Social System Within a Wolf Pup Model Group Using Traditional Ethological Indexes and the Detection of Hidden Patterns

A.A. Yachmennikova[1], A.D. Poyarkov[1]

[1]*Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Moscow, Russian Federation.*
[1]*felis.melanes@gmail.com*

Abstract

The present study compares what can be shown in the process of wolf pup social system establishment with traditional ethological methods; (index of individuals association, index of synchronization of activity, distances between sleeping individuals, sociograms, and the variability of contact intensity) and also with the detection of hidden patterns, Theme (Noldus Information Technology) was used for this purpose. A simple social group of four growing pups which were kept in captivity and were used as a test model, the sample is 2 groups.

Keywords. Social system development, observational methodology, T-patterns, group behavior, traditional ethological methods, comparison of methods

Introduction

The first social influence in animal life is the maternal influence, but siblings are important also for developing the behavior of each other in ontogenesis. Therefore the purpose of investigating the reciprocal influence of the same age younglings is a matter of most importance.

Social environment is the most predictable factor on the one hand, though it is difficult to detect changes in it sometimes, but on the other hand, it is very important for the individuals survival for all of its' life. [3, 5, 12]. How can small pups organize their complex social system as elements? How can the observer measure this process? It is important to find the methodological way of describing and analyzing the ontogenesis of social connections in a group from all perspectives of the dynamics. Classical and traditional methods of measuring of the social connections between animals exist [2, 6, 8, 10, 11]. But they cannot illustrate the full picture, just shed light on some principal moments or describe different aspects with a unidirectional outlook. If we try to look at the social group as a complex functional system, where animals are elements of it we have to understand how uninterruptedly the behavior of every animal is organized in the flow of time and how it is connected with the behavior of all other animals within the group. Hidden patterns can show this type of organization; and their dynamics shows how it evolves over time. When the social structure changes with the animals and their relations development then other elements with other connections between them build the hidden patterns that can be defined. This methodology was successfully applied for the analysis of human communication in different ways [1, 13], we tried to test how it works in social animals living and developing within a group.

Aims

The aim is to study how to explain completely the process of the simple social system of four wolf pups development during their ontogenesis. The main purpose of the paper is to illustrate the changes in the social structure situation during establishment by using six categories of methods (traditional indexes and hidden patterns). Other purposes are to test how such an observational instrument can be applied for such measurement; to compare methods that we have used.

Methods

Two model groups each of four animals were investigated. We had two males and two females of the same age in each group. We received the animals from zoos at the age of twenty-five days old, after the imprinting phase of socialization; in one group we united animals from two different litters before the beginning of collecting behavioral data. We raised the animals in big enclosures, square of up to one hectare in size, which were just parts of a natural forest with a rich habitat. Animals were not handled, but were raised without parents by humans; they knew personally all observers and behaved themselves naturally but this kind of behavior they did not extrapolate toward all people generally. We collected data in three ways: (1) made 24 hour observations once a week, recorded information on a dictaphone every 1 minute for each animal – a method of time steps was used for this; (2) collected video data – each animals behavior was filmed during a period of 10 minutes with focal observation method; (3) made 6 hour observations describing all behavior and interactions, recorded information on to a dictaphone – continuous logging. Different types of data collecting used in different days. We collected 927 hours of Dictaphone recorded observations and 20 hours of video for group #1; 281 hours of dictaphone observations and 18 hours of video for group #2. The wolf pups were under observation for the age period of between thirty five days and four and a half months.

Instrument

During 24 hour observations we recorded types of activity. There were two activity categories: individual types of activity and social types. The first category included: sleeping, resting, moving, manipulation of objects, feeding, excreting, comfort activity, activity of orientation and tracking of observer (human). Second category: contact with observer, consolidation of pups, agonistic activity, friendness activity, following each other, avoidance of each other, play, stress reaction, and reaction to an unknown stimulus If two senses of behavioral actions were observed, we registered the activity type that had a more definitive motivational indication and was strongly shown in the morphology of behavior [9].

Data analysis

Traditional methods of social connection estimation

We used five types of indexes for estimating the level of social connection between animals. Three of them are often used for analyzing behavior in dyads; namely: index of activity synchronization; index of individuals' association and distances between individuals during sleeping time. For counting the index of activity synchronization for the one type of activity we used formula $\boldsymbol{I}_{syn\text{-}AB\text{-}actviv}= \mathrm{Sum}_{\boldsymbol{Syn}}/\ (\mathrm{Sum}_{\boldsymbol{A}}+\mathrm{Sum}_{\boldsymbol{B}})$. Here $\mathrm{Sum}_{\boldsymbol{Syn}}$ – quantity of time steps when individual's *A* и *B* had the same activity type; $\mathrm{Sum}_{\boldsymbol{A}}$ or $\mathrm{Sum}_{\boldsymbol{B}}$ – summary of time steps with this type of activity during 24 h for *A* or for *B* only. If *A* with *B* are fully synchronized, then index meaning should be 0,5. For counting the index of individuals' association we used $Ia=Nab/(Nab+Na+Nb)$ formula. Here *Nab* – summary of cases of observation, when *A* and *B* were together; *Na* – counted cases of observation, when *A* was without *B; Nb* – counted cases of *B* without *A.* Maximum meaning of index is 1 – animals are together every time, minimum meaning is 0 [11]. For analyzing distances during sleeping time, we allocated distances of 6 types: "close contact"; "touching"; "animals are close; but without touching" (less than a length of a body between them); "nearby" (from 1 to 5 lengths of a body); "far" (more than 5 lengths of a body); "too far" (out of a visibility). We counted all categories of the sleeping distances frequency and the dynamics of it. 282 sleeping periods were analyzed in all.

Two other types of estimation of social bonding changes were sociograms and distribution of partners during interactions. We pictured a system of interactions as classical sociograms through analyzing the video-files; the social ranking of the animal was recorded from it (α, β) as classical ethological works describes [4, 14]. Index of individuals rank counted as $I_{rank}=(Sum_{in}-Sum_{get})/(Sum_{in}+Sum_{get})$, where Sum_{in} – summary of all initiated contacts, Sum_{get} – summary of all contacts was received by the animal from all others. Sociograms were made for all positive, negative and neutral contacts. When the frequency of distribution of partners during interactions was counted, we took all 15 variations of animals' combinations: not only dyads, but trios and all-together group contact.

Hidden patterns analysis

Data of four pups 24 hour activity for four months ontogenesis period were analyzed using Theme v. 5 (NOLDUS); the retrieving patterns were allocated to data base Theme v. 6 (beta version). The algorithm was made for hidden T-pattern detection [7]. In one event, we used information about the author of activity, beginning or ending of activity, type of activity, partner for this activity. Example of one event: dunai,b,agon, dn (wolf Dunai begins his fight with wolf Nika). All patterns consist from such events, which are connected. We used p-level < 0.005 for pattern detection; minimal occurrence for pattern means 5.

A pattern can be written as a line: (((dunai,b,agon,dn_dunai,e,agon,dn) rada,b,play,ar)(amur,b,play,and amur,e,play,and)) or as a tree-scheme (Figure 1). The tree-scheme shows the events occurring within the pattern, listed in the order in which they occur in the time. The first event in the pattern appears at the top and the last at the bottom. The pattern diagram (the lines connecting the dots) shows the connection between events. This example pattern consists of two blocks of events (3&2). If read we have: male Dunai begins fight with female Nika (1) and ends it by himself (2); next event that was in the time after these is connected with these two – female Rada begins to play with male Amur (3); other two events that were in the time after previous three are connected with all described block – male Amur begins to play with other male Dunai and female Nika (4) and ends it by himself (5). This all has behavioral sense that the agonistic interaction of two pups can provoke the behavior of reconciliation, initiated by others consistently; such pattern was repeated more than 5 times in a 24h period.

So we can analyze in such away all the patterns as they appear in the animal's ontogenesis with relationships between individuals developing. For this, we have to make a table with line-written patterns in the order that they appear in ontogenesis, the example is shown in Table 1. It is simpler if you organize patterns with some characters. For example through SQL we can have all patterns with agonistic or playing activity only. When we analyze patterns from this table by sense in their order, we can have the total picture of the social situation in group developing and establishing. In this way patterns with agonistic, friend and playing behavior were analyzed. Then all results of all methods were used combined for a resulting picture of relations between animals establishment understanding.

Comparison of the methods which were used

We compared methods which we used to understand which method that behavior characteristics could describe in a better way. We tested the contribution of different methods with discriminant, cluster and factor analysis. With cluster and factor analysis, we compared results of the index of synchronization of activity; index of animals' association and quantity of inclusiveness of pairs of animals in the patterns structure. With discriminant analysis we compared quantity of inclusiveness of individual in all patterns; in friend, agonistic and playing patterns as author of activity; inclusiveness of individual in all patterns as recipient of activity (these all were dependent variables), which were discriminated by sex, rank status, individual, litter and age period of development (independent variables).

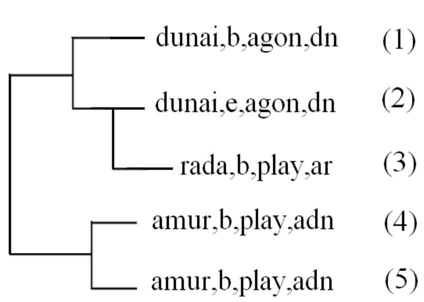


Figure 1. Tree-scheme of one hidden pattern type from 24.08.2008 for investigated wolf pup group #2, pattern was repeated 7 times on the day it was allocated. Rada, Dunai, Amur – names of wolf pups; b/e – begin or end of activity; agon/ play – agonistic/playing type of activity, dn/ar/adn – first letters of interacting wolves' names: dunai-nika/amur-rada/amur-dunai-nika.

Table 1. Example of table with hidden T-patterns where playing behavior is included. Results for wolf pup group #2. nika, amur, dunai, rada – pups names; b/e –beginning or ending of activity; rest, play, move, manip – types of resting, playing, moving or manipulation with objects activity; and, dr, all – combinations of first letters of interacting pups names – amur-dunai, dunai-rada; all (all 4 pups together).

Pups' age (days)	Pattern
37	(nika,e,rest,and ((amur,b,move,a_rada,e,rest,r) (dunai,b,play,dr_dunai,e,play,dr)))
50	(((nika,e,comf,n rada,b,agon,nr) amur,b,play,an) (amur,e,play,an dunai,e,manip,d))
50	((amur,b,orient,a nika,b,play,nr) (nika,e,play,nr(amur,e,manip,a_rada,e,move,r)))
58	((amur,b,manip,a rada,e,manip,r) ((dunai,b,play,ad (dunai,e,play,ad nika,b,play,dn)) (dunai,e,play,dn nika,e,play,dn)))
58	((rada,e,manip,r (nika,e,manip,n rada,e,play,ar)) (amur,b,mark,a_amur,b,rest,a))
68	(nika,e,manip,n (((nika,b,manip,n rada,b,play,dr) (dunai,b,play,dr rada,e,play,dr))(rada,b,comf,r rada,e,comf,r)))
76	(((nika,b,play,nr nika,e,play,nr)(amur,b,manip,a rada,b,manip,r)) (amur,e,manip,a ((dunai,b,manip,d dunai,e,manip,d) (amur,b,orient,a nika,e,orient,n))))
76	((rada,b,play,dr rada,e,play,dr) (((amur,b,move,a amur,e,move,a) amur,b,manip,a) (amur,e,manip,a (dunai,b,manip,d nika,b,manip,n))))
86	((amur,b,play,ad dunai,b,play,adr) ((dunai,e,play,adr (dunai,b,play,all rada,b,play,all)) (nika,b,play,all (amur,e,play,all dunai,e,play,all))))
86	(nika,b,play,all ((amur,b,play,all rada,e,play,all)(amur,e,play,all (dunai,b,play,adr dunai,e,play,adr))))
95	(((dunai,b,move,d nika,b,orient,n)(nika,e,orient,n rada,e,orient,r)) (nika,e,move,n (rada,b,play,r rada,e,play,r)))
95	((dunai,b,play,dr dunai,e,play,dr) ((amur,b,manip,a nika,b,manip,n) (dunai,b,move,d nika,e,manip,n)))
101	(rada,b,play,all ((amur,b,play,all amur,e,play,all) (rada,e,play,all (dunai,e,play,all rada,b,manip,r))))
110	((rada,b,play,nr (nika,b,play,nr nika,e,play,nr)) ((amur,b,play,all amur,e,play,all) dunai,b,play,and))
110	*(((dunai,b,agon,dn dunai,e,agon,dn) rada,b,play,ar) (amur,b,play,and amur,e,play,and))*
124	((dunai,b,move,d (dunai,e,move,d rada,e,move,r)) (nika,b,play,nr nika,e,play,nr))
135	(nika,b,play,all (amur,b,play,all amur,e,play,all))

Results

(1) Traditional methods can show the most important moments of social situation in group changes in the time scale during development. For an example period, when animals became a group, this event takes place in nonlinear way in the wolf pups aged 68-110 days. The Index of synchronization of activity can show that dependence between two animals exist, but cannot show the character of this dependence. Index of association of the individuals can show intensity in attachment in pairs of individuals, but it is oriented much toward the pups from one litter. Distances between sleeping individuals shows periods when animals were especially close and these dynamics have a difference of a period of 5-7 days delay with the dynamics of indexes described above. Using Sociograms we can imagine how the social system looks from the hierarchical point of view. (2) Hidden pattern analysis can show all nuances of process of social structure establishment. In the beginning when group is not structured and pups do not have an opinion about each other there is less agonistic activity in patterns and pups play as pairs (as patterns show); then ranking position of the α-male is allocated; after that it is the beginning of the social position of β-male appearance and ranking position of α-female allocation; after that all group units as a whole; than social position of β-male establishes and after social positions of β-female allocation – primary social group system becomes stabilized. (3) When comparing methods it is the indexes, which describe better variability of connections in pairs than patterns. Methods are not interchangeable, they describe situation differently, and they do not repeat the results of each other (factor and cluster analysis has shown this). With discriminant analysis we tested only pattern characteristics. Inclusiveness of the individual to all patterns as an author of activity and as a connected recipient better describes the individual's characteristics:

rank status, sex, litter. Intensity of inclusiveness of the individual in patterns with playing, agonistic and friendless activity better describes the periodization of development.

Conclusion

When we describe ontogenesis or the structure of group establishment we have to use different methods. But the full picture can be shown if combine all results of the different methods together, because different aspects are specified with them. Using T-pattern analysis we have opportunity to describe such processes like the social position of a female allocating this with a mathematical significance. This method seems to be very promising. Though it describes a social connection between animals in dyads poorly when compared to the use of classical indexes, it has worked out for the use of complex analysis, so it can show a picture of a real situation and the processes in reality. This type of analysis is independent and separate (as cluster analysis has shown), we cannot use it instead of classical ethological indexes, and we cannot reliably replace these old analyses with the more modern T-pattern analysis. We have different types of information resulting from them, but we can combine them. We can use the same patterns for analysis in different aspects to understand features of the wolf pups development or the features of individuals' behavior, or picturing their social system.

References

1. Blanchet, A., Batt, M., Trognon, A., Masse, L. (2005). Language and behavior patterns in a therapeutic interaction sequence. In *The hidden structure of interaction: from neurons to culture patterns* (Eds. Anolli, L., Duncan S. Jr., Magnusson, M.S., Riva G. – Amsterdam IOS Press), 124-139.
2. Crook, J.H. (1970). Social Behaviour in Birds and Mammals. In *Essays on the Social Ethology of Animals and Man*. (Ed. Crook, J.H. Department of Psychology; The University, Bristol, England), 492;
3. Goltsman, M.E. (1983). Social control of mammals behavior: audit of the domination concept. *Results of a science and technics. Zoology of the vertebrates* **12,** 71-131;
4. Hinde, R.A. (1970). *Animal behavior: a synthesis of ethology and comparative psychology* (New York, NY, US: McGraw-Hill).
5. Kruchenkova, E.P. (2002). *Principles of mother-cub relations at mammals*. PhD thesis (in Russian). Moscow St. Univ., 30.
6. Lenher, P.N. (1996). *Handbook of Ethological Methods*. Cambridge University Press, 672.
7. Magnusson, M.S. (2000). Discovering Hidden Time Patterns in Behavior: T_Patterns and Their Detection. *Behav. Res.Methods Instrum. Comput*. **32**(1), 93-110;
8. Nikolsky, A.A. (1992). *Ecological bioacoustics of mammals.* MSU, Moscow, 1-120.
9. Ovsjanikov, N.G. (1993). *Behavior and social organization of arctic fox*. MSU, Moscow, 183.
10. Panov, E.N. (1983). Methodological problems in communication and social behavior researching. *Results of a science and technics. Zoology of the vertebrates* **12**, 5-70.
11. Popov, S.V., Il'chenko, O.G. (2008). *Guidelines to Studies in Zoos: Guidance for Ethological Observations of Mammals in Zoos*. Moscow, Moskovskii zoopark, 165.
12. Popov, S.V., Tchabovsky, A.V. (2005). The concept of a sociality in researches of mammals. *Zoo Journ. Moscow. Nauka* **84**(1), 4-15.
13. Sakaguchi, K., Jonsson, G.K., Hasegawa, T. (2005). Initial interpersonal attraction between mixed sex dyad and movement synchrony. In *The hidden structure of interaction: from neurons to culture patterns* (Eds. Anolli, L., Duncan S.Jr., Magnusson, M.S., Riva G. – Amsterdam IOS Press), 108-119
14. Zimen, E. (1982). Wolf pack sociogram. In *Wolves in the world* (Ed. Harington FH. P.C. Paquet). New Jersey, Park Ridge, 282-323.

Sustainability Goes Change Talk: Can Motivational Interviewing Be Used to Increase Pro-Environmental behavior?

F. E. Klonek[1], S. Kauffeld[1]

[1]Institute of Psychology, Technical University Braunschweig, Braunschweig, Germany. f.klonek@tu-bs.de

Abstract

Motivational Interviewing (MI) is an interviewing style that has been used extensively in the field of addiction as a treatment intervention for clients that are either resistant to or ambivalent about change [11]. Since its origins in the field of addiction treatment, the use of MI has also been extended to health psychology, clinical psychology [6] and to a minor extent also coaching psychology [13]. This study explores the feasibility and efficacy of motivational interviewing in the field of ecological psychology. Specifically, we compared the effects of an MI with those of a Non-MI control interview on client change and sustain talk language about pro-environmental behavior. Interviewers in the intervention condition were trained in MI to talk with participants about their ecological behavior and to increase pro-environmental behavior. Seventy-one interviews were videotaped, and data was analyzed using a combination of two behavioral coding schemes: the German version of the motivational interviewing treatment integrity [5] and the motivational skill code for client language [10]. Results on client change talk show that clients in the MI condition uttered significantly more reasons for change and ability to change. It is suggested that MI may offer a method to increase pro-environmental behavior by means of increasing client change language.

Introduction

The negative impact humans have on the ecological environment constitutes a major problem to our society [6]. Specifically, human behavior has effects on global warming, air and water pollution, depletion of environmental and energy shortages. A challenging task for researchers and policy makers is to develop methods that can reduce the negative human impact on the environment. As most of the respective effects are rooted in human behavior, psychology can contribute to the solution of this problem by developing methods that increase pro-environmental behavior [15]. Pro-environmental behavior consists of all behavioral efforts made not to harm the environment or even to protect and safe environmental resources [15]. MI offers an innovative and promising approach for communicating the severity of environmental problems, increasing actions to conserve environmental resources and also reducing environmentally harmful behavior. MI originates in the treatment of difficult patients in drug therapy, especially the treatment of alcohol problems [11]. Its major aim is to increase a client's intrinsic motivation for change. Over the last two decades the method has increasingly been used in a variety of behavioral domains, for example for reducing risk behaviors, for the treatment of psychological problems [6] and also as a method to increase healthy behavior [cf., 8]. In the most recent meta-analysis on MI, the authors claim that they "have likely not yet found the limits of the types of problems (...) to which MI can be profitably applied" [8, p. 154] . A key skill of MI is to listen for and prompt change talk. Change talk (CT) is any client language that is directed towards changing a target behavior (in this case: pro-environmental behavior). This can include reasons, desires, needs, abilities or commitments by the client to demonstrate the specific target behavior. In contrast, sustain talk (ST) constitutes any client language against changing. This would include benefits of the current behavior or resistance to change. Previous research using MI in a drug population has demonstrated that CT and ST can predict outcomes of a target behavior, such as abstinent rates [1]. This line of reasoning suggests that MI can also be implemented as a method to increase CT about pro-environmental behavior. This is what the present study seeks to investigate. Specifically, we are interested in whether participants who talk about their ecological behavior with an interviewer trained in MI show higher levels of CT than participants who talk about their ecological behavior with an untrained interviewer.

Methods

Sample. Seventy-one participants (interviewees) received feedback on their environmental behavior. Interviewees were allocated to one of two different groups of interviewers: i) an interviewer that conducted feedback about ecological behavior in an MI style or ii) a control interviewer. All dyadic interactions were videotaped and analyzed by means of the German version of the motivational interviewing treatment integrity [MITI-d, 5] and the motivational skill code for client language [MISC 2.1, 10].

Intervention condition. MI has two aims: the first is to increase clients' intrinsic motivation for change; the second aim is to set goals for the new target behavior in collaboration with the client and develop measures for how to achieve them. The interviewing style is based on four principles: 1) expressing empathy, 2) rolling with resistance, 3) developing discrepancies, and 4) supporting self-efficacy. Within the MI style, specific techniques are used, such as asking evocative questions (e.g., "What is a good reason for you to act pro-environmentally?") that aim to prompt client CT. Also, reflecting client utterances is considered a key method of active listening skills.

> Reflection
>
> Client: "I don't want to pay 100 € for my electric bill!"
>
> Interviewer: "You see better way spending 100 €, and saving energy is an easy method for you to achieve that."

Training for interviewers in the intervention condition. Interviewers in the intervention group received training in MI. The interviewers were thirteen bachelor students of Psychology, one master student in Human Resource Development and one PhD Psychology student. The author of the present study, who is certified in MI, conducted the training of interviewers. Training took place from November 2011 until January 2012 and lasted for a total of about 21 hours. It contained exercises aimed at improving MI performance using an active, empathic listening style that minimizes confrontation. Additionally, interviewers took part in bi-weekly peer coaching sessions to increase their communication skills. Supervision was included at the end of training by means of tape recordings of these peer coachings. Control interviewers did not receive MI training. They were given the task to convince their conversational partner to increase their pro-environmental behavior.

Procedure. To exclude the recruitment of clients who are already motivated to talk about their environmental behavior, clients were kept blind about the topic of the interview until the interview started. Beforehand, they received a questionnaire about their ecological behavior. They were told that it was a questionnaire to determine their ecological footprint. Interviewers were given enough time to assess their respective client's ecological behavior prior the interview by means of this questionnaire. Interviewers in both groups were given a short written agenda that listed the topics which needed to be covered during the conversation. These topics were: 1) Setting the agenda, 2) Asking current environmental behavior, 3) Giving feedback about environmental behavior to clients, 4) Asking for measures for increasing pro-environmental behavior, 5) Planning measures and giving advice.

Instruments. Two instruments were combined to code the dyadic interaction: for analyzing the interviewer behavior, the German Motivational Interviewing Treatment integrity, MITI-d, was used [5]. This instrument is a reduced version of the MISC and has specifically been designed to code interviewer behavior only. It includes seven different codes that are intended to capture behavioral micro-skills in MI (see left side of Figure 1). In order to have a mutually exclusive and exhaustive coding scheme, an eighth category "Other behavior" was added. This allowed coding every utterance of the conversation. Client behavior was analyzed by means of a German version of the Motivational Interviewing Skill Code, MISC 2.1, for client language [10]. The MISC includes 16 different codes that can be differentiated by their valence: client utterances with a positive inclination towards change are termed "change talk", whereas utterances that have a negative inclination toward change are termed "sustain talk". Coding was performed using INTERACT software [9], see Figure 2.

Interviewer (MITI-d)	Client (MISC)
	
MI Adherent (M) a. Asking permission before giving advice or information b. Affirming the client c. Emphasizing the client's control, d. Supporting the client **Open question (o)** **Simple reflexion (e)** **Complex reflexion (k)**	Change Talk (+) / Sustain Talk (-) **Reasons (G/g)** "I should"; "I must"; arguments for and against change **Desire (W/w)** "I want to…"; "I'd like to…"; "I love to…" **Ability (F/f)** "I can"; "I am able to…"
MI Non-adherent (m) a. Giving advise without permission b. Confronting c. Directing the client by giving orders, commands or imperatives.	**Need (N/n)** "I need"; "I must" **Other (A/a)** Client movement towards or away from change that is not captured by the other categories
Neutral **Giving Information (I)** a. Providing feedback from assessment instruments b. Personal feedback about the client that is not already available c. Explaining ideas or concepts relevant to the intervention d. Educating about a topic **Closed Question (c)**	**Taking steps (S/s)** Concrete and specific steps towards or away from the target behavior **Commitment Language (V/v)** Agreements; intention to change; obligations **Follow Neutral (O)** No inclination towards or away from change

Figure 1. Coding Schemes: the left side shows MITI-d codes for interviewer behavior; the right side shows MISC codes for client codes. Keyboard codes for each code are given in brackets.

To date, a subset of 28 dyads has been rated by two raters who had received 30 hours of training in the MITI-d and MISC. Within this subset, 17 subjects belonged to the intervention group (MI-interview), and 11 subjects belonged to the control group. To assess inter-rater agreement, a convenience sample of three dyads was double-coded by both raters. Interact-time-event sequential data files were converted into Sequential Data interchange standard (SDIS) using ActSds [3]. Interrater agreement (Time-unit kappa and event alignment kappa) was determined using GSEQ [3]. Time-unit kappa was K=.77-.78 (80-82% agreement) and event alignment Kappa was K =.66 (71 % agreement).

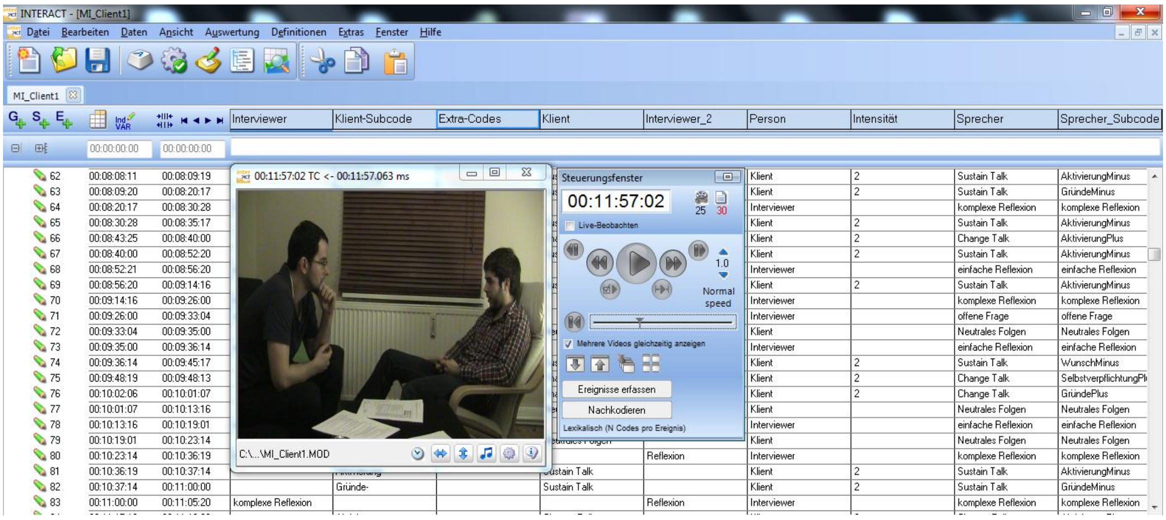


Figure 2. Coding of interviewer (left person) and client (right person) by means of Mangold Interact Software.

Preliminary Results

Interviewer language. In order to adjust for time differences in the interview length between both groups, we calculated rates for separate codes (i.e., frequency of a code per 60 minutes). Interviewers that received training in MI asked significantly more open questions and significantly fewer closed questions in comparison to interviewers in the control group. The MI group interviewers also showed significantly higher levels in active listening skills (rate of reflections) and less MI non-adherent behavior in comparison to the control group. We compared behavioral summary scores of the MI interviewer sample with indicators of good MI [12]. Using these benchmarks, the MI group can be classified within the beginning proficiency level (see Table 1).

Client language. Independent t-tests on client variables revealed that clients in the MI condition uttered significantly more reasons to change ($t(26) = 3.03$, $p<0.005$) and showed significantly more language indicating ability to change ($t(16) = 2.60$, $p<0.019$). Overall, clients in the MI group had more utterances of CT than of ST ($t(26) = 2.02$, $p= 0.054$) – although this was only nearly significant.

Table 1: MI quality benchmarks for competency and beginning proficiency: Means and standard deviations (in brackets) for the MI and the control group.

Behavioral measure	Expert level	Beginning proficiency	*Mean values for MI group*	*Mean values for control*
Percentage of open questions	70 %	50%	*52% (17.8)*	*23 % (9.6)*
Percentage of complex reflexions	50%	40%	*65% (14.8)*	*69% (19.8)*
Reflections to questions ratio	2:1	1:1	*1.2: 1*	*0.46:1*
Rate of reflections in 10 min.	>15	>10	*10.6 (2.2)*	*4.7 (2.2)*
Percentage of MI-adherent statements	100%	90%	*95% (9.2)*	*68% (24.7)*
Percentage of talk time	< 50%	< 60%	*42% (0.5)*	*46% (0.4)*

Discussion

Our results show that interviewers in the MI group demonstrated proficiency in MI as measured by the MITI-d, whereas interviewers in the control group perform below the MI quality performance threshold. More interestingly, clients in the MI condition had more reasons to change and showed higher ability to change – as captured by their natural language. These effects are in line with principles of MI that opt to increase clients' reasons to change and support their self-efficacy [11].

Preliminary Conclusions

Our study reveals that MI can be easily adapted to an ecological psychology framework. Further, the MI style effected client language in terms of increasing their reasons and ability to change. A previous study of MI in a drug population sample showed that reasons and ability to change can have a significant effect on outcome variables, such as abstinent rates [1]. Further analyses are needed to demonstrate whether this link can also be supported for environmental behavior. When the final coding is completed, sequential analysis [2] will be carried out to investigate sequential hypotheses from MI theory, and recurrence analysis will be used to detect specific interaction patterns [14].

References

1. Baer, J. S., Beadnell, B., Garett, S. B., Hartzler, B., Wells, E. A., & Peterson, P. L. (2008). Adolescent change language within a brief motivational intervention and substance use outcomes. *Psychology of Addictive Behaviors* **22**(4), 570–575. doi:10.1037/a0013022
2. Bakeman, R., & Quera, V. (2008). ActSds and OdfSds: Programs for converting INTERACT and The Observer data files into SDIS timed-event sequential data files. *Behavior Research Methods* **40**(3), 869–872. doi:10.3758/BRM.40.3.869
3. Bakeman, R., & Quera, V. (2011). *Sequential Analysis and Observational Methods for the Behavioral Sciences*. Cambridge, UK: Cambridge University Press.
4. Bakeman, R., Quera, V., & Gnisci, A. (2009). Observer agreement for timed-event sequential data: A comparison of time-based and event-based algorithms. *Behavior Research Methods* **41**(1), 137–147. doi:10.3758/BRM.41.1.137
5. Brück, R. K., Frick, K., Loessl, B., Kriston, L., Schondelmaier, S., Go, C., … (2009). Psychometric properties of the German version of the Motivational Interviewing Treatment Integrity Code. *Journal of Substance Abuse Treatment***,** **36**, 44–48. doi:10.1016/j.jsat.2008.04.004
6. Burke, B. L. (2011). What can motivational interviewing do for you? *Cognitive and Behavioral Practice*, (18), 74–81. doi:10.1016/j.cbpra.2009.08.004
7. Gardner, G. T., & Stern, P. C. (2002). *Environmental problems and human behavior* (2nd edition). Boston, MA: Pearson Custom Publishing.
8. Lundahl, B. W., Kunz, C., Brownell, C., Tollefson, D., & Burke, B. L. (2010). A meta-analysis of motivational interviewing: Twenty-five years of empirical studies. *Research on Social Work Practice***,** **20**(2), 137–160. doi:10.1177/1049731509347850
9. Mangold. INTERACT Quick Start Manual V2.4. Retrieved from www.mangold-international.com
10. Miller, W. R., Moyers, T. B., Ernst, D., & Amrhein, P. (2008). *Manual for the motivational interviewing skill code (MISC): Version 2.1*. Unpublished manual. Retrieved from http://casaa.unm.edu/download/misc.pdf
11. Miller, W. R., & Rollnick, S. (2002). *Motivational interviewing: Preparing people for change*. New York: Guilford Press.

12. Moyers, T. B., Martin, T., Manuel, J. K., Hendrickson, S. M., & Miller, W. R. (2005). Assessing competence in the use of motivational interviewing. *Journal of Substance Abuse Treatment* **28**(1), 19–26. doi:10.1016/j.jsat.2004.11.001
13. Passmore, J. (2007). Addressing deficit performance through coaching – using motivational interviewing for performance improvement at work. *International Coaching Psychology Review* **2**(2), 265–275.
14. Quera, V. (2008). RAP: A computer program for exploring similarities in behavior sequences using random projections. *Behavior Research Methods* **40**(1), 21–32. doi:10.3758/BRM.40.1.21
15. Steg, L., & Vlek, C. (2009). Encouraging pro-environmental behaviour: An integrative review and research agenda. *Journal of Environmental Psychology* **29**(3), 309–317. doi:10.1016/j.jenvp.2008.10.004

Functional Analysis of Challenging Behaviours in People with Severe Intellectual Disabilities Using The Observer XT 10.0 Software

C. Delgado[1], R. García[1,2], J.I. Navarro[1], E. Hinojo[2]

[1]*Department of Psychology, University of Cadiz, Spain.* [2]*AFANAS-JEREZ, Centro Vistahermosa, Spain.*

[1]*candida.delgado@uca.es*

Abstract

Challenging behaviours (self-injuries, physical or verbal aggression, and material destruction) are particularly prevalent in adults with intellectual disabilities and are one of the main challenges for researchers. Several studies show that such people are 3-to-5 times more at risk of behaviour problems than the general population. The development of functional analysis to assess behaviour has provided clinicians with effective methodology for the empirical assessment of the functions of a particular behaviour. Therefore, in order to undertake effective intervention, we should carry out a prior functional assessment of behaviour. The aim of this is to gather and synthesize information in order to define the behaviour, determine its motivation, and describe the environment associated with high and low rates of this behaviour. This paper presents a new system of assessment and the behavioural analysis of a person with a severe intellectual disability and challenging behaviour. This system uses a triangular design analysing the participant, environment and interaction. Assessment was carried out using recordings and *The Observer XT 10.0.* This software has several advantages over previous behavioural methods. The statistical analysis module allowed us to filter data, select information and visualize the context in which challenging behaviours occur, in order to discover functions, antecedents and consequences of those behaviours, as well as frequency, duration and intensity.

Introduction

There is a consensus that behaviour problem interventions are more effective when based on the results of a functional evaluation identifying the variables underlying the individual's challenging behaviour [7]. Therefore, in order to undertake effective intervention, we should carry out a prior functional assessment of the behaviour. The aim of this is to gather and synthesize information to define the behaviour, determine what underlies it, and describe the environmental context associated with high and low rates of this behaviour [6]. Then we can implement an intervention that changes the environment to increase appropriate behaviour and reduce challenging behaviour, teaches new skills, and guarantees the safety of all those involved [3, 4]. Epidemiological studies of self-injury and other challenging behaviours [1] have shown that the function of those behaviours can be identified in most cases. However, a study by Thompson *et al.* [8] which evaluated aggressive behaviour in a child with severe intellectual disabilities showed that the conclusions obtained by a functional analysis were inconclusive in those cases in which the self-injury, aggressive or destructive behaviour was supported by automatic reinforcement. This suggests that when challenging behaviour is supported by social responses, treatment usually involves breaking the reinforcement, extinction and positive reinforcement of adapted behaviours. However, when the reinforcement responsible for maintaining the challenging behaviour is automatic, it is more difficult to design an intervention because the specific reinforcer that is responsible for maintaining the behaviour is unknown or impossible to manipulate.

In this study we present a system of objective evaluation and the consequent behavioural analysis of a person with a severe intellectual disability. This system uses a triangular design which analyses the participant, environment and interaction [5]. Assessment was carried out using recordings and *The Observer XT version 10.0.* The aim of this process was to design an individual programme to reduce or eliminate the individual's challenging behaviour, replacing it with pro-social behaviour.

Methodology

Participant. A 40-year-old male who has been diagnosed with profound intellectual disability, autism and extremely challenging behaviour participated in this study. His behaviour, defined as violent and sometimes dangerous, includes: (1) *self-injury*, defined as behaviour in which he injures himself; (2) *physical aggression:* behaviour causing physical pain to others; (3) *repetitive or stereotyped behaviours:* unusual or bizarre behaviours; (4) *socially-offensive behaviours:* behaviours that offend others; (5) *disruptive behaviours:* those that interfere with the work of others. These behaviours have been present since childhood. The user has twice undergone a neurosurgical intervention (bilateral cryohypotalamotomy) to reduce problem behaviour and has been prescribed psychotropic drugs permanently. Following neurosurgical intervention, aberrant behaviours decreased slightly, although they continue to be present in his repertoire. We used various methods of behaviour modification to reduce or eliminate aberrant behaviour, including negative reinforcement, extinction, negative punishment, positive punishment, and time-out. However, even though his behaviour changed, challenging behaviours still remained in his repertoire, harming both himself and those around him.

Materials and procedure. First we gathered information about the participant through the review of his medical records, collecting data concerning his diagnosis, etiology, behavioural problems, medication, physical health and mental health, intervention undertaken so far and his personal history. We conducted an interview with his care staff and family, in order to understand why these challenging behaviours appear, whether there has been any variation in these behaviours, and what is currently being done to correct them. We also administered different evaluation tests *Diagnostic Assessment for the Severely Handicapped*, *Autism Diagnostic Interview-R* and *Inventory for Client and Agency Planning*). Finally, we performed behaviour observation by recording him in his natural environment in different contexts at the Day-Stay Unit. The family was informed and agreed to this study.

The participant‘s operative challenging behaviours were defined as follows: (1) *self-injury*: banging his head, belly, chest, leg, hand or arm, (2) *physical aggression*: kicking, head banging, hitting or punching, (3) *repetitive or stereotyped behaviours*: wandering or swinging, (4) *socially-offensive behaviours*: urinating in the work room, being naked in public, (5) *disruptive behaviours*: escaping from the room during working hours or throwing or destroying the material. Repetitive or stereotyped behaviours were defined by duration, in order to know the time involved in such behaviour. Self-injurious, physical aggression, socially-offensive and disruptive behaviours were defined as discrete behaviours and the frequency and intensity with which they occurred were recorded. We also defined *antecedents* (stimuli that explain the presence of challenging behaviours) and *consequences* (social response of the environment when challenging behaviours happen) individually, after the information provided by care staff and family. So, antecedents were classified as: (1) Not paying attention, defined as those situations in which no attention had been paid to the subject before the onset of challenging behaviour, (2) Interruption, defined as those situations in which someone interacted with the subject and prevented him from doing something he liked, (3) Transitions, understood as those times when a change was made from one activity to another, (4) A difficult task, defined as those circumstances in which the subject was asked to do an activity, (5) Receiving an order, conceptualized as the time when the subject was asked to collaborate and (6) Attempted communication, understood as those situations where the subject attempted to deliver a message to the staff member. Consequences were classified as: (1) Obtain self-stimulation: situations where the subject was involved in episodes of challenging behaviour without getting any social result, (2) Get attention: situations in which the subject was involved in challenging behaviours and immediately received attention from the direct care staff, (3) Get an object or an activity: situations where the subject was involved in episodes of challenging behaviours and immediately received an object or performed a desired activity, (4) Avoid orders: situations in which, following an order by the direct care staff, the subject was involved in challenging behaviours and avoided obeying the order, (5) Avoid activity: situations where, after a request for an activity to be done, the subject was involved challenging behaviours and managed to avoid doing it (6) Avoid people: situations in which the subject was involved in challenging behaviours in response to social interaction with someone.

We selected a continuous observation method because of the low frequency of some behavioural problems. Recordings each lasted two hours and covered the entire period in which the user was working at the Day-Stay

Unit. Data collection was carried out for 15 non-consecutive days, over a total period of a month and a half, for a total of 30 hours. Recordings were made in different contexts (classroom, dining and on entertainment trips). Finally, these materials were recorded manually and later analysed with The Observer XT 10.0, a tool for observing and analysing all types of behaviour.

Results and Discussion

Data obtained by The Observer XT provided extensive information about the frequency or duration of behaviours, and this allowed us to design a person-centred intervention, aimed at reducing or eliminating challenging behaviours and increasing the occurrence of appropriate behaviour. Figure 1 shows challenging behaviour frequency. The most predominant behaviour in the behavioural repertoire was self-injury (N=391). The most prevalent behaviours were: hitting his leg (=13.67), head banging (= 6.80) and hitting his hand (=3.00). In terms of physical aggression (N=83), the most frequent was hitting others (=4.07) (Table 1).

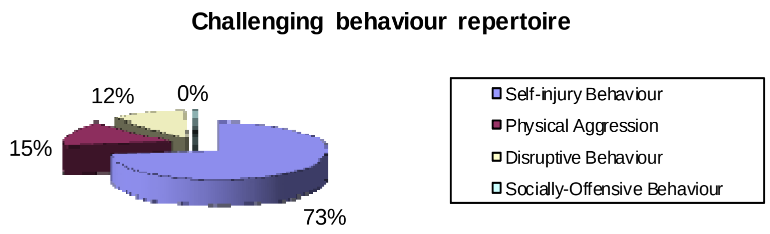


Figure 1. Frequency of challenging behaviours observed in 15 days.

	N	Min.	Max.	Mean	SD	Rate/min
Self-injury behaviours	**391**					
Hitting his leg	205	0	53	13.67	18.02	0.17
Banging his head	102	0	22	6.80	9.03	0.09
Hitting his hand	45	0	19	3.00	6.00	0.05
Hitting his arm	20	0	15	1.33	3.89	0.02
Hitting his belly	9	0	4	0.60	1.18	0.008
Hitting his face	6	0	3	0.40	1.06	0.005
Hitting his chest	4	0	3	0.27	0.80	0.003
Physical Aggression behaviours	**83**					
Hitting	61	0	20	4.07	5.81	0.05
Kicking	13	0	4	0.87	1.51	0.01
Head banging	5	0	3	0.33	0.90	0.004
Punching	4	0	1	0.07	0.26	0.002
Socially-Offensive behaviours	**2**					
Being naked in public	1	0	1	0.07	0.26	0
Urinating in classroom	1	0	1	0.07	0.26	0
Disruptive behaviours	**62**					
Escaping the work room	35	0	9	2.33	2.72	0.03
Throwing material	27	0	6	1.80	2.43	0.02

Table 1. Results obtained after analysis of the challenging behaviours observed.

The Observer XT allowed us to make a more precise behavioural analysis by considering the frequency of *antecedents* and *consequences*, thus identifying functionality for the observed behaviours. Most problem behaviours occur when the subject is not receiving any attention (F=50), he is interrupted (F=19), after receiving an instruction (F=7), he is doing a difficult activity (F = 6), he becomes frustrated because of his inability to communicate (F=1), or he changes activity (F=1) (Table 2). Therefore, challenging behaviours functions were: achieve self-stimulation (F=33), seek attention (F=18), obtain a desired object, undertake a desired activity

(F=9), avoid other people (F=9), avoid carrying out instructions (F=8), or avoid carrying out tasks (F=6). Figure 2 represents the frequency and duration of all challenging behaviours observed in the subject after The Observer XT. Each colour represents a different behaviour (e.g. "*banging his head*" is shown in blue), the frequency of each behaviour is represented by a vertical line, while the duration is represented by a bar whose length is determined by the time of the behaviour.

Table 2. Relationship between antecedents and consequences of the challenging behaviours observed.

	Self-injury							Physical aggression			Disruptive		Total
	Hitting his leg	Banging his head	Hitting his hand	Hitting his arm	Hitting his arm	Hitting his belly	Hitting his chest	Hitting	Punching	Kicking	Escaping	Throwing material	
Antecedents													
Not paying attention	9	6	13	3	2	2	2	2	2	1	7	1	50
Interruption	1	1	5	-	-	-	-	7	1	2	1	1	19
Transition	-	-	-	-	-	-	-	-	-	-	1	-	1
Difficult task	1	2	-	-	-	1	-	2	-	-	-	-	6
Receive an instruction	1	1	1	-	-	2	-	-	-	-	1	1	7
Communication attempt	-	-	-	-	-	-	-	1	-	-	-	-	1
Consequences													
Get self-stimulation	7	6	12	3	2	2	-	1	-	-	-	-	33
Get object/activity	1	1	3	-	-	-	-	2	-	-	2	-	9
Gain attention	1	-	1	-	-	1	1	2	2	1	8	1	18
Avoid instruction	1	-	2	-	-	2	-	2	-	-	-	1	8
Avoid activity	1	2	-	-	-	1	-	1	1	-	-	1	6
Avoid people	1	1	1	-	-	-	-	3	1	2	-	-	9

This software allowed us to describe challenging behaviours accurately, quantify human behaviour, and generate results through a rigorous statistical analysis letting us select each variable analysed. In summary, The Observer XT offers the possibility of visualizing the context in which challenging behaviours occur many times, in order to discover functionality of those behaviours, the antecedents and consequences that determine them, as well as their frequency, duration and intensity. At the same time, the statistical analysis module allows us filter data and select relevant information in order to discover the factors that may be involved in maintaining the participant's

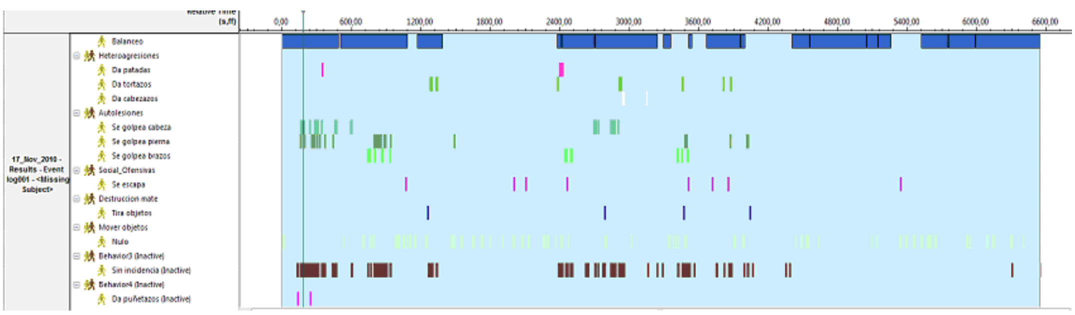


Figure 2. Frequency and duration of challenging behaviours observed in the subject in a day of observation.

challenging behaviours. This is very different from the other behaviour recording systems (e.g. *O'Neill's details of systematic observation*, *Aberrant Behaviour Checklist Scale*) used to perform a functional or topographical assessment. They do not offer the possibility of repeating the episode in which challenging behaviour occurred in order to perform a precise analysis of the antecedents and consequences explaining occurrence of the behaviour. This may cause confusion in the observer when defining the function of that behaviour. This may also affect the reliability and validity of the study, misrepresenting elements involved in such behaviour. In practice, the antecedent and consequent periods involved in narrative ABC records are rarely defined in a sufficiently precise way that either conditional or unconditional probabilities can be determined with any degree of confidence [2]. As a result, the information provided by narrative ABC charts can only have a very general value in determining basic response dimensions (e.g. the rate of occurrence of challenging behaviours) and the relationship between challenging behaviour and general setting variables (e.g. time, location). Thus, while the narrative descriptions of antecedent and consequent events may generate some tentative hypotheses for further investigation, they are far from providing a "convincing demonstration" of underlying behavioural processes. Although the software used clearly has advantages over previous methods of behavioural observation, we would like make some comments about our experience. The Observer XT had some limitations in terms of cost and time on data collection. Secondly, there are limitations in the supported video formats, which do not include HD-quality video. This makes the data entry process very slow, with more time being spent on converting video than on data entry.

References

1. Derby, K. M., Wacker, D. P., Sasso, G., Steege, M., Northup, J., Cigrand, K., & Asmus, J. (1992). Brief functional assessment techniques to evaluate aberrant behaviour in an outpatient clinic: A summary of 79 cases. *Journal of Applied Behavior Analysis* **25,** 713–721.

2. Emerson, E. (1995). *Challenging behaviour: Analysis and intervention in people with learning difficulties.* Britain: Cambridge.

3. Horner, R. & Carr, E. G. (1997). Behavioral supports of students with severe disabilities: Functional assessment and comprehensive intervention. *Journal of Special Education* **31** (1), 84-104.

4. Horner, R., Dunlap, G., Beasley, J., Fox, L., Lambara, L., Brown, F. & cols. (2005): Positive support for behavioural, mental health, communication and crisis needs. In K. Lakin & A. Turnbull (Eds.), *National goals and research for people with intellectual and developmental disabilities* (pp. 93-208). Washington, DC: American Association on Mental Retardation.

5. Pascual-Leone, J. (2007). Comments to the Practice Guidelines for diagnostic, treatment and related support services for persons with intellectual disabilities and problem behaviour.

6. Repp, A. C. & Horner, R. H. (1999). *Functional analysis of problem behaviour: From effective assessment to effective support.* Belmont, CA: Wadsworth.

7. Sigafoos, J., Arthur, M., & O'Reilly, M. (2003). Challenging behaviour and developmental disability. London: Whurr Publishers.

8. Thompson, R. H., Fischer, W. W., Piazza, C. C., & Kuhn, D. E. (1998). The evaluation and treatment of aggression maintained by attention and automatic reinforcement. *Journal of applied behaviour analyses* **31**, 103-116.

Crowd and Pedestrian Dynamics: Empirical Investigation and Simulation

S. Bandini[1], A. Gorrini[2], L. Manenti[1], G. Vizzari[1]

[1] *CSAI - Complex Systems and Artificial Intelligence Research Center, University of Milan-Bicocca, Milan, Italy. {bandini,manenti,vizzari}@disco.unimib.it*

[2] *Information Society Ph.D. Program, University of Milan-Bicocca, Milan, Italy. a.gorrini@campus.unimib.it*

Abstract

The investigation of crowd dynamics is a complex field of study, which involves different types of knowledge and skills. From the socio-psychological perspective the definition of crowd is still controversial. For this reason, we propose to analytically investigate this topic focusing on pedestrian dynamics in high-density situation. In this paper we propose a multi-disciplinary approach, which is based on observations and experimental procedures focused on proxemic behaviour of walking groups, and Multi-Agent Systems, which are useful technique to simulate what-if scenarios. The methodological approach that we propose can be synthetically represented as composed of research in vivo (observations), in vitro (experimental procedures) and in silicon (modeling and simulations).

Keywords. Crowd, Groups, Pedestrian Dynamics, Proxemics.

Introduction

The investigation of crowd dynamics is a complex field of study, which involves different types of knowledge and skills. Early interest in studying crowd behavior started from the pioneering study of Gustave Le Bon [1], who defined a crowd as a potential threat to society: as members of a crowd people display a loss of sense of self-awareness and an increase of violent behaviors. Far from this approach, the ESIM-Elaborated Social Identity Model [2] proposes a social-normative conception of collective behavior, based on the spontaneous transition from an individual to a common social identity [3]. From the socio-psychological perspective, the definition of crowd is still controversial, because of the lack of standard guidance for the empirical investigation of the phenomenon, and its variability among size and typology. For this reason, we propose to analytically investigate this topic focusing on pedestrian dynamics in high-density situations. The study of people movement dynamics, in relation with the levels of density and the physical characteristics of the environment, represents an innovative discipline that involves different knowledge areas and provides applicative results (e.g. a more efficient management of pedestrian circulation in public spaces). The multi-disciplinary approach that we propose is based on the integration between:

- Socio-Psychological perspectives about: groups[1], the basic elements which the crowd is composed of, and proxemics [4], chosen as an analytical indicator of spatial behavior dynamics within the crowd;
- Multi-Agent System simulation technique [5].

In analogy with territorial behavior in animals [6], proxemics [4] is defined as a type of nonverbal communication that conveys information about the nature of participants' relationship, by means of the dynamic regulation of interpersonal distances[2]. In situations of high density, the invasion of personal space [7] is strictly linked with crowding [8], a negative subjective state typically associated with: psycho-physiological responses of arousal and stress, cognitive performance decrements, and aggressive response.

---

[1] A group can be defined as two or more people who interact for a shared goal, perceiving a membership based on a shared social identity [3].

[2] The regulation of spatial distances (intimate, personal, social, and public distances) is influenced by age, gender, culture, and personality [7].

In Computer Science, a Multi-Agent System is defined as a system in which multiple entities, called agents, interact in a shared environment, aimed to achieve some individual or collective goals [9]. The Multy-Agent System are useful approach in modeling and simulation of pedestrian and crowd dynamics [5], by means of the representation of crowd as a system of heterogeneous agents.

We propose a methodological approach that includes studies already performed (1, 4), and on-going works (2, 3):

1. Observations of proxemic behavior of walking groups, focusing on: spatial arrangement (degree of alignment and cohesion), walking speed, level of density, group size and gender;

2. Experimental investigation on size and shape of pedestrian personal space

3. Experimental investigation on the capability of group proxemic behavior to reduce the effects of crowding

4. Design what-if scenarios about pedestrian proxemic behavior by means of a simulation tool.

Group Pedestrian Behavior

Group proxemic behavior reveals the psychological relationship among the group members [10]; in motion situation, it generates typical patterns, which allow communication and spatial cohesion among members during the movement [11], in relation to the level of density[3]. At low density, group members tend to walk side by side, forming a line perpendicular to the walking direction (line-abreast pattern); as the density increases, the linear walking formation turn into a V-like pattern, with the middle individual positioned slightly behind in comparison to the lateral individuals; in situation of high density, the spatial distribution of group members leads to a river-like pattern and lane formation, characterized by the presence of a leader that coordinates the group members to cross the space [12]. To further investigate proxemic behavior of walking groups in real social context, we performed the observation of the incoming pedestrian flow to the admission test of the Faculty of Psychology at the University of Milan-Bicocca in September 1, 2011 (about two thousand students attended the test). The observation was aimed at gathering empirical data, by means of a people counting activity supported by video footages of the event[4]. The survey was focused on: percentage of groups within pedestrian flow[5]; relationship between walking speed and group size, group spatial arrangement, gender of members; level of density and level of service; lane formation and queues. Data analysis shows that more than the 65% of the incoming flow was composed of groups (77% couples, 19% triples, 4% larger groups). Moreover, a one-way analysis of variance (ANOVA) showed that the size of groups affects walking speed in situation of medium density ($p < 0,05$). More in detail, the differences in walking speed between singles and couples, singles and triples, couple and triples, were confirmed by a T-test analysis ($p < 0,01$). No significant differences in walking speed were detected in reference to spatial arrangement and gender composition ($p > 0,05$).

Pedestrian Personal Space

Personal space is traditionally defined as a circular area surrounding individuals, whose size varies depending on subject head orientation and visual mechanisms [13]. Pedestrian personal space is assumed to be different from the one in static situation, because of the need to reserve additional margins in the front zone to avoid collision with obstacles, and visual mechanism of eye convergence [14]. For these reasons, the front zone is assumed to be larger than the one in static situation, and the lateral zones are assumed to be smaller than the ones in static situation. We propose to use the “stop-distance” procedure [13], designed between subjects (60 subjects): 2

[3] Proxemic behavior of walking group is related to group size, and the features of the members among gender and age [12].

[4] The use of people tracking tools although is a useful contribution in detecting spatial movements of pedestrian within the crowd, it is not enough calibrated to recognize proxemic indicators among group members.

[5] The identification of groups in the streaming of passers-by was assessed considering verbal and non-verbal communication indicators.

groups composed of 30 subjects for each modality of the independent variable “state of movement of the subject” (subject stationary or subject walking straight ahead); the dependent variable “size” of pedestrian personal space will be deduced from the “stop-distance” requested by the subject in the moment of the perceived discomfort related to spatial invasion. The procedure will be conducted in a laboratory setting, and it will employ digital video-recording equipment for data analysis.

Group Proxemic Behavior and Crowding

Group proxemic behavior reveals the psychological relationship among members and, in static situations, it can mitigate the effects of crowding, by producing spatial boundaries that shield group members from unwanted physical interactions [8]. Starting from these assumptions, typical spatial patterns of walking groups are assumed to be an adaptive stress-reducing behavior to crowding by maintaining group cohesion, increasing inter-group distances and rising intra-group capability to communicate. In relation to the modalities of the independent variables “belonging to a group” and “state of movement of the subject” (2x2), the experimental procedure will be designed between subjects (120 subjects), with 4 groups composed of 30 subjects for each modalities. Following the standard procedure for crowding measurement [8], the subjects will be taken into a room for 30 minutes, as a short-term exposure to high-density situation (4 persons for m^2); crowding will be deduced as dependent variable, and it will be measured by means of the level of the salivary cortisol, which is an accurate indicator of endocrine activity and crowding [8, 15].

Proxemic Behavior: Modeling and Simulation

In order to support the design of what-if scenarios related to pedestrian dynamics, we use a software simulation platform, based on Multi Agent System technique, developed by CSAI [16]. The tool represents the environment of the scenario, imported as CAD file, as discretized into squared cells of 40x40 cm (the typical space occupied by a pedestrian in high density situations), in which the groups of pedestrians are driven by behavioral and perception rules. The agents are able to perceive and avoid spatial elements that represent obstacles and other agents, and to move toward their targets. The agents are also able to recognize those that belong to their group, and to behave according to proxemic behavioral parameters about: group cohesion (stay closed its group members reaching their targets), invasion of personal space (preserve a certain distance from other pedestrians, which belong to a different group). These rules are based on several behavioral rules about proxemics, which have to be calibrated starting from existing and known empirical data, and by means of the empirical results achieved by observation and experimental procedures (e.g. the typical patterns of walking group, the size of pedestrian personal space and the spatial cohesion among group members to avoid crowding).

Discussion

The innovative approach that we propose consists of a first phase of modeling and calibration by using empirical data, and a second phase of performing simulations. The results collected by means of the observation already performed, and data, which will be achieved by means of the on-going experimental procedures, will be used to improve the simulation model. In particular, a more detailed definition of the perceptive capabilities and behavioral specifications of each agent will be performed, in relation to their interpretation of the other elements in the system (e.g., agents, objects, obstacles). Although there are some objections about the simplified level of correspondence between simulations and real phenomena, we propose to use simulations as a method to study complex system from a different layer of abstraction, and as a virtual experimental laboratory to study those scenarios which are difficult to be empirically investigated (such as evacuation processes of a crowd in case of emergency). Simulations produce a series of statistical data, that can be analyzed and interpreted, and that can lead to further empirical investigations. The methodological approach that we have proposed can be synthetically represented as a “virtuous cycle” composed of three main parts: research in vivo (observations), in vitro (experimental procedures) and in silicon (modeling and simulations). The results achieved by means of empirical investigations and simulations of what-if scenarios can be useful to support the design of applicative strategies related to a more efficient management of pedestrian circulation dynamics.

References

1. Le Bon, G. (1895, trans. 1947) *The Crowd: A Study of the Popular Mind*. Ernest Benn, London.

2. Reicher, S. (2001). *The psychology of crowd dynamics.* In *Blackwell handbook of social psychology: Group processes*, Wiley-Blackwell, 182-208.

3. Turner, J. (1982). Toward a cognitive definition of the group. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, 15-40.

4. Hall, E.T. (1966). *The Hidden Dimension*. Doubleday, New York.

5. Gilbert, G.N., & Troitzsch, K.G. (2005). *Simulation for the social scientist*. Open Univ Pr.

6. Hediger, H. (1961). The evolution of territorial behavior. In *Social life of early man*, Wenner-Gren Foundation for Anthropological Research, 34-57.

7. Aiello, J. R. (1987). *Human spatial behavior*. In *Handbook of environmental psychology*, Wiley, New York, 1, 389-505.

8. Baum, A., & Paulus, P. B. (1987). *Crowding*. In *Handbook of environmental psychology*, Wiley, New York, 1, 533-570.

9. Ferber, J. (1999). *Multi-agent systems: an introduction to distributed artificial intelligence* (Vol. 222). Addison-Wesley, Boston.

10. Knowles, E. S. (1973). Boundaries around group interaction: The effect of group size and member status on boundary permeability. *Journal of Personality and Social Psychology* **26**(3), 327.

11. Moussaïd, M., Perozo, N., Garnier, S., Helbing, D., & Theraulaz, G. (2010). The walking behaviour of pedestrian social groups and its impact on crowd dynamics. *PLoS One* **5**(4), e10047.

12. Costa, M. (2010). Interpersonal Distances in Group Walking. *Journal of Nonverbal Behavior* **34**(1), 15-26.

13. Hayduk, L. A. (1983). Personal space: Where we now stand. *Psychological bulletin* **94**(2), 293.

14. Gibson, J. J. (1954). The visual perception of objective motion and subjective movement. Psychological Review **61**(5), 304.

15. Evans, G. W., & Wener, R. E. (2007). Crowding and personal space invasion on the train: Please don't make me sit in the middle. *Journal of Environmental Psychology* **27**(1), 90–94.

16. Bandini, S., Rubagotti, F., Vizzari, G., Shimura, K. (2011). An agent model of pedestrian and group dynamics: Experiments on group cohesion. In Pirrone, R., Sorbello, F., eds.: AI*IA. Volume 6934 of Lecture Notes in Computer Science., Springer 104–116.

Measuring Situation Awareness of the Microneurosurgeons

Shahram Eivazi

School of Computing, University of Eastern Finland, Joensuu, Finland. Shahram.eivazi@uef.com

Microneurosurgery is performed using miniature movements of microinstruments. The conduct of any procedure consists of complex series of actions with microinstruments hold in the right and left hand. The operative field is illuminated and magnified by a neurosurgical operation microscope. The microscope can be linearly focused onto the operative plane where the tips of the micro-instruments are functioning. The microneurosurgery processes require concentration and any distraction prolongs the procedure and may lead to iatrogenic errors. In Europe usually in order to become a board certified neurosurgeon, trainee should conduct at least 800-1000 surgeries during the five years training (www.eans.org).

It is essential to understand behavioral processes involved in the microneurosurgery such as situation awareness. The situation awareness is a perception of the available information, events, resources, and environment within a given time and space. Humans have limited abilities to obtain and maintain situation awareness, as they need to carefully orchestrate the available resources. A failure to maintain situation awareness may lead to serious errors in human behavior [2]. For example, in the miconeurosurgery when micronsurosurgeon manually repositioning microscopes to obtain a new view on the brain or when modifying the parameters and settings of the device, the surgeon needs to remove hands from the site and this might reduce situation awareness of the microneurosurgeons.

The amount of existing evidence shows that the users’ behavior, including situation awareness, could be modeled where the accurate measure of user’s behavior exist. Measuring the situation awareness of microneurosurgeons in operating room is a key challenge for deeper understanding of microneurosurgeons’ behavior and it is one of the main motivations for this research project.

There are different techniques to gain access into the users’ behavior patterns. For example, recording think aloud or eye movement protocols. Both think aloud and eye tracking methods have been indicated as measurement techniques that often are used to show operator situation awareness during a defined task [4]. Since past decade the eye tracking methodology has become attractive both in everyday activities and controlled laboratory studies. Analysis of the relations between eye movements and users’ behavior has indeed proved fruitfully in different domains, such as reading comprehension, visual search, selective attention, and studies of visual working memory. As accuracy of the eye trackers increased, recording eye movements has become a popular method in the medical research. For example, radiology image perception, eye-controlled microscope for surgical applications, and finding differences between medical specialties.

We first aim to measure microneurosurgeons’ situation awareness using an eye tracker outside of the operating room and second we plan to integrate a binocular eye tracker within a neurosurgical microscope to record the eye movements of microneurosurgeons in the operating room. The measuring microneurosurgeons’ behavior help us to study surgeons’ interrupting factors, purposefulness movements, hazardous tendencies, decision making errors. The result of the captured and analyzed knowledge, will allow to more explicitly describing the processes involved in critical situations:

1. To find the various strategies which surgeons applying during surgery.
2. To find the primary factors in accidents attributed to the surgeon error.
3. To find the differences between surgeons (experts) and resident (novices).
4. To be used in training surgeons- to continuously show gaze orientation of experts in educational videos-.

5. To be used in intelligent proactive monitoring of on-going operations.

In an eye-tracking study of skill differences between surgeons and residents shows a marked difference in the way expert and novice microneurosurgeons attend to snap shots of a surgery video. In the Eivazi and Bednarik study [3] the finding related to the mean fixation duration was confirm the fact that the microneurosurgeons have to be able to focus for prolonged times at cretin points due to the limited size area of the surgery. While little has been done in the field of microneurosurgery in regard to measuring behavior, Eivazi and Bednarik [3] study motivated us to build a system that is able to record microsurgeons' eye movements in the operating room. The goal is to define the characteristics of the microneurosurgeons' behavior, and search for the underlying eye movement features. We believe that understanding microneurosurgeons' strategies in the operating room will help to develop an advance method to support surgeons and residents during the surgery.

References

1. Charlier, J., Sourdille, P., Behague, M. Buquet, C. (1991). Eye-controlled microscope for surgical applications. *Developments in ophthalmology* **22**, 154-158.

2. Eivazi, S. (2011). Situation awareness in neurosurgery: A user modeling approach. *User Modeling, Adaption and Personalization, volume 6787 of Lecture Notes in Computer Science*, chapter 39, pages 409-413. Springer Berlin / Heidelberg, Berlin, Heidelberg.

3. Eivazi, S., Bednarik, R. (2012). Gaze Behaviour of Expert and Novice Microneurosurgeons Differs During Observations of Tumor Removal Recordings. To appear in the ETRA. Santa Barbara, CA, USA.ACM.

4. Salmon, N. Stanton, G. Walker, Green D. (2006). Situation awareness measurement: A review of applicability for c4i environments. *Applied Ergonomics* **37**, 225–238.

Behavioral Dynamics (in Staff Meetings): What Patterns Lead To Success?

A.M.G.M. Hoogeboom[1], C.P.M. Wilderom[2]

Department of Business Administration, University of Twente, Enschede, The Netherlands

[1]a.m.g.m.hoogeboom@utwente.nl, [2]c.p.m.wilderom@utwente.nl

Analyzing the patterns of behavior in an organizational context is of paramount importance to the knowledge of effective corporate governance. This study aims to reveal new insights in the social-behavioral repertoire of members during (and around) staff meetings. In this research we will not only test the currently most known leadership model (the so called transformational model), we will also capture the interactive (video-taped and -analyzed) dynamics between leaders and their co-workers (i.e., followers) in and around regular staff and staff meetings (i.e., influence patterns). This proposed effectiveness study of the behavioral patterns during staff meetings builds on earlier field studies of actual behavior of leaders in staff meetings [1] (see appendix 1 for initial results on the leaders' behavioral repertoire). Here we extend the survey- and video-field research approach to capture the behavioral patterns of staff meetings. On these premises, a number of specific hypotheses will be developed, based on various theory, coming from leadership, social influence (i.e., Human Interaction Dynamics: HID), entrepreneurship/intrapreneurship, small-group and team research. These hypotheses will be examined with the following sources of information combined: (a) reliable behavioral coding of video-taped leader and followers in and around staff meetings; (b) reliable expert ratings of both individual leaders and of their team's effectiveness; (c) team member surveys for the assessment of more generic interpersonal relationships, and (d) self-reports on cognitive characteristics.

Research topic

Acknowledging that human actors form the 'spine' of all organizations, various subareas of Management research show a growing interest in the behavioral complexity of human work carried out in small and large ventures or organizations. A rich literature on leading group, departmental or teamwork is meeting to some extent that need. A behavioral research focus on the complexity of effective team processes during (and around) staff meetings would fill a large gap; the specific behavioral repertoire of both the leaders and followers and its underlying motivational, cognitive and social-interaction aspects may add to extant knowledge, especially when also drawing upon the Leadership field of scholarly action [2-5]. The extant research on effective team behavior has thus far largely ignored the dynamic interplay among team members and its subsequent effects on team or leader effectiveness and meeting success. Also the effects of these influence or interaction forces are still unknown, e.g., how does effective interaction stimulate intrapreneural behavior (i.e., in terms of creation of new-business venturing; innovativeness; self-renewal; proactiveness, see Antoncic & Hisrich, 2001) or other relevant functional behaviors (e.g., decision making or information exchange).

While a lot of scientific attention is devoted to teamwork in general (for reviews see e.g., [6]), our knowledge as regards to how teams effectively interact and respond in routine, time constraint events (i.e., regular staff meetings) is still unaddressed. Previous work suggests that patterns of interaction among team members are found to be consequential and highly generalizable antecedents of team effectiveness [7]. In this sense, Greer and Van Kleef [8] noted that team interaction styles are related to functional or dysfunctional outputs. In recognition hereof and partially driven by the increasing time devotement in meetings, we aim to answer the following general question: What functional behaviors affect/are related to different meeting goals? This study aims to reveal the multi-faceted nature of the various interaction events during regular staff meetings and subsequent outcomes. The study integrates research in: team behavior during team meetings (i.e., assembling objective and direct measures of actual leader and follower behavior in the field [9]); social-interaction processes (using Bales' and Borgatta's [10,11] interaction theory as building block); and team cognitive diversity as predictors of team effectiveness (i.e., individual characteristics, values, self-efficacy [12-19]).

Approach

A video-methodological approach, already fruitfully applied to leader-behavior during randomly selected, real-life or non-experimental staff meetings, is likely to bring more objective insights on effective dynamics and interaction patterns in regular staff meetings, especially if paired with the more customarily used methods in this area (such as the interview; survey; and expert ratings). The research design is built up accordingly to the underlying assumptions of triangulation. Video-coding behavior (and interaction through for example sequential analysis) will result in reliable systematically-video coded data (as defined and conceptualized by Bakeman & Gottman [20]), representing fine-grained behavior and interaction patterns.

Observing behavior. Shondrick and Lord [21] argued that understanding leadership behavior through perceptions of followers may result in categorizations of leaders which match available role models. Instead, many leadership authors, including for instance Hunter et al. [22] and Hindmarsh and Heath [23] have advocated the use of direct measurements of actual field behavior, other than through surveys.

The importance of investigating specific observed behaviors as a basis for exploring leader effectiveness is written about in the work of Uleman [24] and Yukl et al. [25]. Yukl [26] argues that specific behaviors would provide a solid base for theory generation of leader effectiveness. He noted a void in existing studies of leader behavior; analyzing specific behaviors contributes to a better understanding of leader effectiveness. When emphasis is directed toward more specific behavior, solid ground for the behavioral and interaction theories would be established [27].

Behavioral coding scheme. In order to clearly specify leadership behavior during daily work practices, a coding scheme was developed and pilot tested [27,9]. A solid base for this scheme was found in the work of Bales [10] and Borgatta [11]. Both Bales and Borgatta observed interaction processes between leaders and their followers in small group settings. Bales' Interaction Process Analysis (IPA) coding scheme and Borgatta's Interaction Process Scores (IPS) analysis distinguished between positive-social emotional behavior, neutral-task oriented behavior and remaining socio-emotional behavior. Their work and that of others led to a set of mutually exclusive behaviors to provide a scheme for the coding of a full range of a leader's behavioral repertoire. Another study, using an experimental approach towards measuring leader behavior, was Feyerherm's extension [28] of the work of Bales [10] and Borgatta [11] with several task- and social-oriented directly observable behaviors, which was taken into consideration as well. These three frameworks have in common that they (a) assessed directly observable *behavior* and (b) used their behavioral schemes to code leader behavior in a group context. In addition to the results of this experimentally oriented research within small groups, we also incorporated behaviors that had been assessed or covered by questionnaire-based measurement tools. The bulk of the leadership literature, though, especially for examining transformational versus transactional *style*, uses namely survey measures as the sole determinant of effective 'behavior.' A critical note to this approach is that it fails to capture actual, observable leadership behavior but mainly presents *follower perceptions* of a leaders style [29]. But we could not neglect the emergent behavioral pattern that so many leadership scholars referred to and studied (e.g., [30, 31, 25]). As another key reference point for comparison purposes we also included as much as possible into our own scheme the behavioral taxonomy of Yukl et al. [25]. The resulting integrative leader behavior coding scheme, used in this study, is listed in Table 1.

Data analysis

Regular staff meetings in a group-setting will be audio- and video-recorded, using special equipment for the *behavioral* measure of leader and follower behavior in a field setting. This set-up allows for a *measure of interaction processes between leader and followers and vice versa (i.e., or between members of the meeting),* as coded interdependently by several trained coders. A thoroughly developed and pilot-tested 15-page codebook is used. This codebook, consisting of 11 mutually exclusive behaviors, is based on the extensive experience and work of prominent behavioral scholars (such as [10,11,28,25]). They laid the foundation of our more-or-less objective measure of the interaction processes within a team context (see also [32]). The current behaviorally anchored coding scheme was extensively developed and pilot tested (e.g., [27,9]). Video-analyses will be

Table 1. Behaviors of highly vs. moderately effective leaders*.

Behavior	Highly effective leaders	Moderately effective leaders	
	n = 16	*n* = 13	*n* = 29
1. Showing disinterest	0.6%	0.5%	0.5%
2. Defending one's own position	0.4%	0.4%	0.4%
3. Providing negative feedback	0.7%	0.5%	0.6%
4. Directing	3.5%	3.3%	3.4%
5. Verifying	6.3%	6.6%	6.4%
6. Structuring the conversation	7.2%	7.5%	7.4%
7. Informing	16.3%	15.5%	15.9%
8. Visioning	10.3%	10.9%	10.6%
9. Intellectual stimulation	5.7%	5.8%	5.8%
10. Individualized consideration	10.1%	10.0%	10.1%
11. Listening	38.9%	39.0%	38.9%
Total	100%	100%	100%

* = no significant differences were found between the highly and moderately effective leaders (t-tests, p< .05).

conducted in the Leadership Lab at the University of Twente, utilizing a software program specifically developed for analyzing video-based behavioral data: The Observer XT [33]. Next to this, Theme (Noldus Information Technology, Wageningen, The Netherlands) will be used for unraveling the detailed time structure of behavior, using T-Analysis for the analysis of social interactions.

Besides video-filming the meetings, expert-raters will be selected on the basis of either being in a supervisory position vis-à-vis the leader or working in close proximity to the focal leaders. These subjective performance assessments will be complemented with more objective evaluation forms obtained from the organization, where possible. Next to the expert-raters both the leaders and followers are asked to provide several measures related to behavioral variables and group effectiveness indicators (i.e., values, MLQ, LMX, TMX, SI, interaction; leader, team and meeting effectiveness).

References

1. Hoogeboom, A.M.G.M., Wilderom, C.P.M., Nijhuis, J.H.E., Van Den Berg, P.T. (2011). Leader values, style and behavior during meetings: Triangulated evidence of effective leadership. Paper Presented at the OB division of the Academy of Management Meetings, Texas: San Antonio.

2. Burke, C.S., Stagl, K.C., Klein, C., Goodwin, G.F., Salas, E., Halpin, S.M. (2006). What type of leadership behaviors are functional in teams? A meta-analysis. *Leadership Quarterly* **17**, 288-307.

3. Glynn, M.A., Raffaelli, R. (2010). Uncovering mechanisms of theory development in an academic field: Lessons from leadership research. *The Academy of Management Annals* **4**(1), 359-401.

4. Ilgen, D.R., Hollenbeck, J.R., Johnson, M., Jundt, D. (2005). Teams in organizations: From Input-Process-Output models to IMOI models. *Annual Review of Psychology* **56**, 517-544.

5. Sonnentag, S., Volmer, J. (2009). Individual-level predictors of task-related teamwork processes: the role of expertise and self-efficacy in team meetings. *Group & Organization Management*, **34**(1), 37-66.

6. Kozlowski, S.J.W., Ilgen, D.R. (2006). Enhancing the effectiveness of work groups and teams, *Psychological Science in the Public Interest* **7**(3), 77-124.

7. Waller, M.J., Gupta, N., Giambatista, R.C. (2004). Effects of adaptive behaviors and shared mental models on control crew performance. *Management Science* **50**, 1534-1544.

8. Greer, L.L., Van Kleef, G.A. (2010). Equality versus differentiation: the effects of power dispersion on group interaction. *Journal of Applied Psychology* **95**(6), 1032-1044.

9. Van Der Weide, J.G. (2007). *An explorative video-study to the behavior of effective middle managers*. Unpublished doctoral dissertation, University of Tilburg, the Netherlands.

10. Bales, R.F. (1950). A set of categories for the analysis of small group interaction. *American Sociological Review* **15**(2), 257-263.

11. Borgatta, E.F. (1964). A note on the consistency of subject behavior in interaction process analysis. *Sociometry* **27**(2), 222-229.

12. Brown, M.E., Trevino, L.K. (2009). Leader–Follower values congruence: Are socialized charismatic leaders better able to achieve it? *Journal of Applied Psychology* **94**(2), 478-490.

13. Chen, G., Kanfer, R. (2007). Toward a system theory of motivated behavior in work groups. In B.M. Staw (Ed.), *Research in organizational behavior* **27**, 223-267. Greenwich, CT: JAI press.

14. Erdogan, B., Liden, R. C. (2002). Social exchanges in the workplace: A review of recent developments and future research directions in leader-member exchange theory. *Leadership*. In L. L. Neider, & C. A. Schriesheim (Eds.), Greenwich, CT: Information Age Press, 65-114.

15. Fu, P.P., Tsui, A.S., Liu, J., Lan L. (2010). Pursuit of whose happiness? Executive leaders' transformational behaviors and personal values. *Administrative Science Quarterly* **55**(2), 222-254.

16. Gibson, C.B. (2001). Me and us: Differential relationships among goal-setting training, efficacy and effectiveness at the individual and team level. *Journal of Organizational Behavior* **22**(7), 789-808.

17. Liao, H., Liu, D., Loi, R. (2010). Looking at both sides of the social exchange coin: A social cognitive perspective on the joint effects of relationship quality and differentiation on creativity. *Academy of Management Journal* **53**(5), 1090-1109.

18. Locke, E.A., Latham, G.P. (2002). Building a practically useful theory of goal setting and task motivation. *American Psychologist* **57**(9), 705-717.

19. Lord R.G., Brown, D.J. (2001). Leadership, values, and subordinate self-concepts. *Leadeship Quarterly* **12**(2), 133-152.

20. Bakeman, R., Gottman, J.M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). UK: Cambridge University Press.

21. Shondrick, S.J., Lord, R.G. (2010). Implicit leadership and followership theories: Dynamic structures for leadership perceptions, memory, and leader-follower processes. *International Review of Industrial and Organizational Psychology* **25,** 1-33.

22. Hunter, S.T., Bedell-Avers, K.E., Mumford, M.D. (2007). The typical leadership study: assumptions, implications, and potential remedies. *Leadership Quarterly* **18**(5), 435-446.

23. Hindmarsh, J., Heath, C. (2007). Video-based studies of work practice. *Sociology Compass* **1**(1), 156-173.

24. Uleman, J.S. (1991). Leadership ratings: Toward focusing more on specific behaviors. *Leadership Quarterly* **2**, 175-187.

25. Yukl, G., Lepsinger, R. (2004). *Flexible Leadership*. Jossey-Bass.

26. Yukl, G. (2008). *Leadership in organizations* (11th ed.). Upper Saddle River, NJ: Prentice Hall.

27. Gupta, K., Wilderom, C., Van Hillegersberg, J. (2009). Exploring the behavior of highly effective CIOs using video analysis. *AMCIS 2009 Proceedings*. Retrieved from http://aisel.aisnet.org/amcis2009/463.

28. Feyerherm, A.E. (1994). Leadership in collaboration: A longitudinal study of two interorganizational rule-making groups. *Leadership Quarterly* **5**(3/4), 253-270.

29. Brown, D.J., Keeping, L.M. (2005). Elaborating the construct of transformational leadership: The role of affect. *Leadership Quarterly* **16**(2), 245-272.

30. Avolio, B.J. Bass B.M. (1995). Individual consideration viewed at multiple levels of analysis: A multi-framework for examining the diffusion of transformational leadership. *Leadership Quarterly* **6**, 188-218.

31. Pearce, C.L., Sims, H.P.Jr., Cox, J.F., Ball, G., Schnell, E., Smith, K.A., Trevino, L. (2003). Transactors, transformers and beyond: A multi-method development of a theoretical typology of leadership. *Journal of Management Development* **22**(4), 273-307.

32. Nam, C.S., Lyons J.B., Hwang H., Kim, S. (2009). The process of team communication in multi-cultural contexts: An empirical study using Bales‟ interaction process analysis (IPA). *International Journal of Industrial Ergonomics* **39**, 771-782.

33. Noldus, L.P.J.J., Trienes, R.J.H., Hendriksen, A.H.M., Jansen, H., Jansen R.G. (2000). The observer video-pro: New software for the collection, management, and presentation of time-structured data from videotapes and digital media files. *Behavior Research Methods, Instruments and Computers* **32**, 197–206.

34. Uhl-Bien, M. Pillai, R. (2007). The romance of leadership and the social construction of followership. In: B. Shamir, R. Pillai, M. C. Bligh and M. Uhl-Bien, eds., *Follower-centered perspectives on leadership: A tribute to the memory of James R. Meindl*, Information Age Publishing, Greenwich, CT, 187–209.

Autism and Somantics: Capturing Behaviour in the Wild

D.J. Walker[1], W. Keay-Bright[2], D. Cobner[3]

[1]*Department of Applied Psychology, Cardiff Metropolitan University, Cardiff, UK. ldwalker@cardiffmet.ac.uk*

[2]*Cardiff School of Art and Design, Cardiff Metropolitan University, Cardiff, UK. 2wkbright@cardiffmet.ac.uk*

[3]*Cardiff School of Sport, Cardiff Metropolitan University, Cardiff, UK. 3dmcobnert@cardiffmet.ac.uk*

Abstract

This paper looks at the use of a novel software application created with the inclusive participation of the target population, young people with autistic spectrum disorders (ASD). Somantics has been used in a variety of settings, therapeutic and educational, as a method of eliciting non-formal, user-led interaction. This is in contrast with many information technology interventions that target specific behaviours via structured tasks. Moreover, they are often conducted in controlled environments. By investigating behavior 'in the wild' we offer a genuine alternative to traditional paradigms in understanding ASD. This paper highlights the eclectic methodologies (e.g. micro-ethnography, charettes, performance analysis) and tools that we have used to capture ASD behavior. It emphasizes the benefits of adopting a genuine multidisciplinary approach of individuals and processes (e.g. product designers, psychologists, sports scientists, touch therapists) when investigating real-world phenomena.

Keywords. Autism, information technology, mixed methods, ethnography, video observation, performance analysis, charrette.

Introduction

Around 130,000 children in the UK [1] are diagnosed with ASD. Factoring in the considerable impact that this has on families, it is estimated around 2 million people in the UK could be affected directly or indirectly by this condition. Although diagnosis is often difficult, clinicians look for impairments in social interaction and communication, along with problems in emotional regulation. Furthermore, narrow interests and restricted repetitive behaviours are present [2]. ASD is commonly accompanied with anxiety that impacts on communication for approximately 40-45% of young people [3]. The role, nature and treatment of anxiety in ASD has been studied over the last ten years, with some notable clinical [4] and research [5] examples. Children with ASD often find 'comfort' interacting with computers; as this provides them with an outlet for their restrictive behaviours and attention to detail without the need for social interaction in the real-world [6]. There exists a multitude of information technologies for training individuals with ASD in specific skills such as emotion recognition [7] and understanding the mental states of others [8]. Most of these packages encourage active, user-driven, learning and run on any home computer. However, the success of these interventions has been mixed, with some studies reporting that improvements within the computer environment often fail to generalize to real-world environments [9]. Thus, there is a paucity of evidence looking at the effects of technologies that encourage children with autism to engage in creative play through movement and touch [10].

Somantics

The idea was to create a suite of applications for the iPad that use tactile exploration and playfulness to enhance the communicative potential of the most impaired and misunderstood young people on the autism spectrum. Furthermore the potential for gestural communication was boosted by taking the interface beyond the iPad tablet into the environment using the gesture recognition functionality of the *Kinect 3D* camera. Somantics promotes the evolution of gestured-based communication through 3 distinct modes of interaction:

Figure 1. A child interacting with the iPad.

Figure 2. Group interaction using Kinect 3D camera.

Mode (1) *playful interaction*. Research indicates that one of the main barriers to communication for people with ASD is anxiety. We created applications that use repetitious and rhythmic, tactile interaction using the iPad touch sensitive screen, to afford relaxation. People on the spectrum are known to avoid social communication when they are confused by the nuances of face-to-face conversation, when they do not understand the purpose of communication and when sensory information in the environment is overwhelming [11].

Mode (2) *gesture based interaction.* Using the internal two-way iPad camera to capture facial and gestural expressions (see Figure 1.) we produced applications that enable young people to manipulate their captured images in order to create, highly individual 'artworks', effectively augmenting their images with their own graphical gestures. Research shows that people with autism have a poor sense of identity and low self esteem. Our applications will enable young people to explore identity, to determine aspects of self that are meaningful, and to share these with others.

Mode (3) *real-time gestural interaction*. What really makes Somantics novel in terms of gestural communication is the facility project these artworks into the environment, and use the *Kinect 3D* camera (see Figure 2.) to make the projections a multi-touch surface. The high visibility and repetitious mirroring of action into the projected space can reinforce feelings of self-control - an essential pre-requisite for communication. Furthermore, the opportunity to see that control reflected within the environment means that the child's available attention for communication is widened, shaped and shared through interaction with others.

Emergent evaluation methods and tools

The methods continue to emerge and use a mixed methods approach [12] yielding both quantitative and qualitative data. The main evaluation tool is video observation of behaviours. Our studies are thus situated broadly within a visual sensory ethnographical framework [13]. Data generated from more formal discrete planned sessions is assimilated into a more continuous ethnographic narrative, sustained through a combination of field notes, interviews, video and photographs. We have used more formal video analysis software from the psychological and performance analysis worlds, which facilitate systematic quantitative analyses using duration/time and event/activity based coding. They also allow qualitative analysis of video footage by allowing the coding of 'key moments', generating video clips that can be supplanted with textual annotation Indeed, our approach is consistent with our contemporaries in educational settings. For example Snell [14] has highlighted the benefits of combining systematic quantitative analysis with micro-ethnographic analysis of video data.

We thus see these software packages as key evaluation tools and an integral part of the methodological evidence base. Moreover, we envisage their use in both group and case based analyses of Somantics. For example, they can serve as a powerful longitudinal database of a child's behaviour over multiple sessions with Somantics. This archive allows systematic tracking of behaviours, with early sessions serving as quasi-baseline data to see the efficacy of targeted interventions. Quantitative analyses of behaviours have been *triangulated* with input (e.g.

interviews) from key stakeholders such as teachers, carers and parents. Thus we have strove to avoid the reductionism possible, by over-reliance on objective coding and statistical analysis of video data.

Ethnographical Techniques

Our approach to methodology, whilst theoretically informed has also been eclectic and pragmatic. We have borrowed tools and techniques from ethnography, without necessarily conducting a full-blown ethnography. For example, many of us have spent regular (often weekly) one hour sessions in *Special Needs Schools* using Somantics with low functioning autistics (LFA) whom we have become well acquainted with. Of equal importance is the trust we have gained from teachers and supporting carers ‘in the field’ and how we have integrated them as much as possible in the research process (e.g. interviews, commenting on video footage) to achieve full *triangulation* [15]. In terms of a general ethnographical observation approach we have probably straddled the continuum between *observer-as-participant* and *participant-as-observer* [16]. Thus in some situations members of the team have been known and recognized, but related more to the participants (particularly the LFA children) primarily as a researcher. Whereas in other situations, the researchers have seen themselves more fully integrated with the life of the group and are acknowledged not just as researchers but as ‘friends’.

Readers may wonder given the impoverished linguistic skills of many of the autistic children whether ethnographical techniques allow them to have a ‘voice’ and indeed whether this research format is appropriate. We don’t see this as a problem, believing that the nuanced bodily communication facilitated by Somantics serves in part as their voice and it is our challenge to capture and interpret this. Moreover, from its earliest inception in anthropology, ethnography has been concerned with non-literate populations. It also has a large knowledge base that has dealt with gesture and movement, e.g. Kinesics [17].

Charrettes and Performance Analysis

A charrette is a method of structuring thoughts from experts and users following intense periods of working collaboratively, often over multiple sessions [18]. They allow the quick generation of a design solution, but crucially allow the input of all stakeholders to be incorporated. At the invitation from an internationally renowned movement therapy centre, *The Touch Trust*, we trialled Somantics Apps in the context of a range of therapeutic activities. This has offered the chance to assess the feasibility of planned measurement tools and provide rapid feedback for both the design and evaluation processes. We collaborated with experts from sport performance analysis, who have worked with British rugby and netball teams, but have also used their skills in applications such as health and safety. They designed a discrete multi camera set-up, with real-time capture of behaviours using *non-participant observation*. We piloted the use of coding schemes informed by the SCERTS framework [19] and the feasibility of capturing behaviours in both real-time (e.g. discussing main performance patterns; coding specific behaviours using a predesigned template) and post video capture coding. We also experimented with *participant observation* via wireless capture and real-time coding of behaviours using iPads and wearable cameras. We saw genuine benefits of performance analysis expertise complimenting the psychological and design based expertise of the team.

Table 1 shows quantitative data from a charette from one child. They attended two sessions of approximately one hour length. We can quantify the amount of time that can be considered as positive and negative emotional regulation. We can also quantify unwanted behaviours (e.g. touching shirt, violating personal space of others) and we can look at aspects of transactional support through behaviours that require carer intervention. You can see from the below table that there was a reduction in the amount of negative emotional regulation from session 1 to session 2, that unwanted behaviours were also reduced (e.g. invading the personal space of a partner) and this is further corroborated with less need for carer intervention.

Table 1. Behaviours recorded from two charette sessions

	Session 1			Session 2		
	Number	**Time**	**%**	**Number**	**Time**	**%**
Total time		54.27			60.48	
Positive	10	21.45	40	3	48.02	80
Negative	10	32.42	60	3	12.46	20
Touch shirt	10			9		
Personal Space	18			11		
Carer intervention	7			5		

However, it was the qualitative data captured by the use of performance analysis that was arguably of most value, particularly in the early stages of trialling Somantics. For example, emergent behaviours that we hadn't expected to capture and the impact of practitioner reflection and review, were one of the most relevant and empowering factors of the performance analysis system. Indeed the lead author (coming primarily from a quantitative background) quickly realised that the statistics were only meaningful when viewed in the context of parent/practitioner/stakeholder participation. Thus the review process, facilitated by the performance analysis system, provided the most compelling testimonies we needed. Furthermore, it became a shared vocabulary that complimented other data.

Conclusions and further research

We have found the adoption of mixed methodologies ‘in the wild’ to be challenging in terms of measurement, but it has delivered a richer understanding of behavior. The research team has approached the design and evaluation of Somantics from an interdisciplinary perspective and this ethos has been more formally adopted at our newly created Centre for Applied Research in Arts and Design (CARIAD). In the coming months we will be adding to the functionality of Somantics, by including more sound and music based tools, given their known therapeutic benefits [20]. We will also be extending the research to domains such as stroke rehabilitation and dementia.

Note: in the above research full ethical approval was granted and informed consent was obtained from parents, teachers and carers. All children gave their assent. Permission to use still and moving images has been granted.

References

1. Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T; Meldrum, D. & Charman, T. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *Lancet* **368**, 210-215.

2. American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: Author.

3. White, S., Oswald, D., Ollendick, T., Scahill, L. (2009). Anxiety in children and adolescents with autism spectrum disorders. *Clinical Psychology Review* **29**, 216-229.

4. Howlin, P., Rutter, M. (1987). *Treatment of autistic children*. Chichester, UK: Wiley.

5. Muris, P., Steerneman, P., Merckelbach, H., Holdrinet, I., Meesters, C. (1998). Comorbid Anxiety Symptoms in Children with Persuasive Developmental Disorders. *Journal of Anxiety Disorders* **12**, 387-393.

6. Porayska-Pomsta, K., Frauenberger, C., Rajendran, T., Smith, H., Pain, R., Menzies, M., Foster, A., Alcorn, S., Wass, S., Bernadini, S., Avramides,K., Keay-Bright, W., Chen, J., Waller, K., Guldberg, K., Good, J.,

Lemon, O. (2012) Developing Technology for Autism Needs: An Interdisciplinary approach. *Personal and Ubiquitous Computing* **16**, 117-127.

7. Golan, O., Baron-Cohen, S. (2006). Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. *Development and Psychopathology* **18**, 591-617.

8. Grynszpan, O., Nadel, J., Martin, J.C., Simonin, P.B., Wang, Y., Gepner, D., Barillier, F.L., Constant, J. (2011). Self-Monitoring of Gaze in High Functioning Autism. *Journal of Autism and Developmental Disorders* [ePub ahead of print].

9. Parsons, S., Lewis, A., Ellins, J. (2009). The views and experiences of parents of children with autistic spectrum disorder about educational provision: comparisons with parents of children with other disabilities from an online survey. *European Journal of Special Needs Education* **24**, 37-58.

10. Keay-Bright, W. (2009). ReacTickles: playful interaction with information communication technologies. *International Journal of Arts and Technology* **2**, 133-151.

11. Bogdasina, O. (2003). *Sensory Perceptual Issues in Autism: Different Sensory Experiences – Different Perceptual Worlds.* London. Jessica Kingsley.

12. Bryman, A. (2006). *Social Research Methods* (2nd ed.). Oxford: OUP.

13. Pink, S. (2009). *Doing Sensory Ethnography.* London: Sage.

14. Snell, J. (2011): Interrogating video data: systematic quantitative analysis versus micro-ethnographic analysis. *International Journal of Social Research Methodology* **14**, 253-258.

15. Fetterman, D.M. (2010). *Ethnography Step by Step* (3rd ed.). Thousand Oaks, CA: Sage.

16. Angrosino, M. 2007). *Doing Ethnographic and Observational Research*. Thousand Oaks: Sage.

17. Bernard, H.R. (1988). *Research Methods in Cultural Anthropology*. Newbury Park, CA: Sage.

18. Wikipedia. http://en.wikipedia.org/wiki/Charrette**.** Accessed 27 March 2012.

19. Prizant, B.M., Wetherby, A.M., Rubin, E., Laurent, A.C., Rydell, P. (2006). *The SCERTS Model: A Comprehensive Educational Approach for Children with Autistic Spectrum Disorders*. Baltimore, MD: Paul H. Brookes Publishing.

20. Wigram, T, Gold, C. (2006) Research evidence and clinical applicability of music therapy for autism spectrum disorder'. *Child Care: Health and development* **32**, 535-542.

Observing Flow in Child/Music Machine Interaction

A.R. Addessi[1], L. Ferrari[2]

Department of Music and Performing Arts, University of Bologna, Bologna, Italy.

[1]annarita.addessi@unibo.it, [2]laura.ferrari10@unibo.it

Abstract

The aim of this extended abstract is to introduce a grid created with the software The Observer XT10.5 (Noldus Information Technology) to observe and analyze the state of Flow (Csikzsentmihalyi [6]) in 4 and 8 year old children when they interact with a musical machine for children music improvisation: the MIROR Impro. The MIROR Impro is one of the three components of the MIROR platform, which is an innovative adaptive system for music learning and teaching based on the "reflexive interaction" paradigm [1,3,4,10]. The theory of Flow has been used to explain the ability of the IRMS (Interactive Reflexive Music Systems) to imitate the music style of the user and their ability to enhance and maintain the attention of the user [11]. The results would support the hypothesis that the IRMS and the reflexive interaction may generate an experience of well-being and creativity. The Flow grid worked in an effective way and it was possible to indicate some aspects of the MIROR Impro to be improved. The research is carried out in the framework of the EU-ICT Project MIROR (Musical Interaction Relying On Reflexion).

Theoretical Background: the MIROR Impro and the theory of Flow

The MIROR Impro is an IRMS (Interactive Reflexive Music Systems), and it is characterized by the idea of letting the users manipulate virtual copies of themselves through designed machine-learning software. "The basic concept in the paradigm of reflexive interaction is to establish a dialogue between the user and the machine, in which the user tries to 'teach' the machine his/her musical language" p.19 [3]. According to Pachet [11,10] the reflexivity in interactive systems is characterized by: similarity or mirroring effect, agnosticism, scaffolding of complexity and seamlessness. The basic hypothesis of the MIROR project is that the "reflexive interaction" enhances musical learning and creative processes in children. This hypothesis is based on several interactional, perceptual and neurophysiology mechanisms generated in children during the interaction with an IRMS: as well as imitation, self-imitation, imitation recognition, turn-taking, repetition/variation processes and the experience of Flow is considered as one of the creative experience enhanced by the reflexive interaction [1]. Leman et Al. [9] indicate the theory of Flow as one of areas of expertise, along with the concept of "presence", which should be explored to study the human/machine interaction. The MIROR Impro, tested in the protocol n.1 inside the MIROR project, is an evolution of a previous IRMS: the Continuator. The studies on the Continuator/preschool children interaction underlined as the Continuator, and so the IRMS, could be defined as Flow machine [4]8]. According to Csikzsentmihalyi [6], the state of Flow is a condition of well-being, an optimal experience reached when a balance between the skills of the subject and the challenges is present. The Flow is characterized by the presence of high levels of a series of variables: focused attention, clear-cut feedback, clear goals, pleasure, control of situation, awareness merged, no worry of failure, self-consciousness disappeared, the change of the perception of time, pleasure. From the 80s some studies and researches applied the theory of Flow to different field as daily life, economy and sport, and in recent years to the music education, performance and composition [4,5,7,12].

Method

Participants: 48 children involved in the protocol, 24 children observed = 12 (4 year old), 12 (8 year old).

Setting: a room used as atelier in a kindergarten; the library of the primary school. The keyboard and the two loudspeakers were on a table. The child played the keyboard resting in front of it. The laptop was on a little table near the other one. The camera was resting on a tripod in front of the child and was visible by the child.

Equipment: MIROR-Impro prototype v. 2.5; a music synthesizer KORG X50; a notebook TOSHIBA - Techra (Windows 7, 64 bits); two amplifiers M-AUDIO AV30; a video camera SONY (recording in HD).

MIROR Impro: three different set-up of the MIROR Impro prototype were used: set up A “same”, based on the imitation of the input phrase; set up B “very different”, the answer of the system presents minor similarities with the user's input; set up “nothing”, the system was inactivated and there was no answer to the user.

Procedure: every child realized three sessions in three consecutive days. In each session, the operator asked the child to play four different games (tasks), as follows: Task 1 = the child played alone with the keyboard; Task 2 = the child played alone with the keyboard and the MIROR Impro Prototype; Task 3 = the child played the keyboard with a friend; Task 4 = the child played the keyboard and MIROR Impro Prototype with a friend. The children were invited to tell when he/she was tired and want to stop the activity and/or to change the task. The tasks were given in random order.

Independent variables: the use of MIROR-Impro (with/without the system); the set-up (“same”, and “very different”); the age of children (4 and 8 year old); the gender of children; the exposition to the system (3 sessions); presence/absence of the friend.

Experimental hypothesis: the basic hypothesis is that the reflexive interaction and the mechanism of repetition/variation implemented by the IRMS could enhance creativity and learning processes. Consequently the experimental hypothesis is that the Flow emotional state increases when children playing with the MIROR-Impro and with set-up “same”.

Data collected: video recordings, photos, MIROR Impro recordings, drawings made by the children, children profile filled by the teachers, questionnaires filled by the parents (provided only by the preschool teachers).

The Flow Grid

The grid has been created with The Observer XT 10.5 software (Noldus Information Technology, Wageningen, The Netherlands). The research by Csikzsentmihalyi measures the Flow using interviews and diaries in which subjects, in first person, describing situations in which there is the Flow (self-reported data). Instead, in the grid of Flow here described, the Flow is measured by the observation of the children's behaviour. For this reason, only the variables that could be described through an operational behaviour have been considered, as follows: focused attention, clear feedback, clear goals, control of the situation, pleasure. In some cases, to define better the operational behaviours of the variables the indicators of Custodero [7] were used. The basic idea of this grid was that the observer did not observe/register directly the Flow state, but rather the “variables” and the intensity of each variable. The Flow grid allows recording the presence/absence and the duration of each behaviour. Furthermore, the grid allows recording the level of intensity of each behaviour, by using the “Modifier” tool: 1 = low level of intensity; 2 = medium level of intensity; 3 = high level of intensity. The Observer XT calculates, through a specific data profile, the combinations of the different levels of the different behaviours: when the levels of all behaviours are recorded with high levels, the presence of the state of Flow is indicated as present. The data profile was based essentially on a series of consecutive “nests” over the behaviours. The Nesting function allows selecting among the data registered by the observers the combined presence of pre-defined level of the behaviours. Before starting the observation, each independent observer received a document named “Instructions for the observers”, within the definitions of the behaviours, the modifier and several “practical actions” to register the behaviour by the software. Reliability tests within the five independent observers have been realized before to start the registration and during the registration of the observation. The cases of disagreement were solved by collective discussions and observations. After the observation, the data collected have been analysed with The Observer XT and in particular by the behavioural analysis. The software calculates: total duration (total duration of the video observed), analyzed duration (the duration of the video observed with “dead moments”) duration of each behaviour, percentage (analyzed duration) of each behaviour. The results obtained with The Observer XT have been exported to statistical software (SPSS) to make MANOVA and t-student analysis.

Results

The percentage of state of Flow is higher, for all subjects, when the child/children played with the MIROR Impro (T2=Task 2; T4 = Task 4); supporting of this outcome, the lowest percentage of Flow has been recorded when the child played alone without partner and without the system (T1 = Task 1), see Figure 1.

Thus, we can affirm the MIROR Impro enhances the state of Flow in children. The presence of the Flow is always (considering the four tasks) higher in the sessions of 8 years rather the 4 years old. About this finding, we could assume the MIROR Impro is more suitable for older children rather the younger. In the sessions with the set up A "same" the percentage of Flow is always higher than the set up B "very different", see Figure 2 and in all 4 tasks. The difference between set-up "same" and set-up "very different" is significant (p=.004). This result support the experimental hypothesis that the Flow state increases when the system's reply is more similar to the input played by the children, that is when the system's reply is more "reflexive". In this case, it is therefore possible to say that the result support also the hypothesis that the reflexive interaction could enhance the flow state. Considering all the sessions, the trend of the Flow remains constant between the first and the second session, while it rises between the second and third session. The understanding of the rules of the system, the presence of high level of control of situation, the pleasure to play may make the children more self-confident and these may be important factors to enhance a state of well-being and the state of Flow.

Conclusion

The analyses carried out with The Observer XT confirm the experimental hypothesis: the results show that the Flow increases not only when children play with MIROR-Impro, but also when they play with the set-up "same", that is the more "reflexive" set-up used in the experiment. Furthermore, the Flow state is more evident when children play alone with the system. The creation of the coding scheme and of the data profile, the observation with The Observer XT has been a very demanding task. Nevertheless, it has been also very helpful because it allowed to reflect on and define better the behaviours (finding operational definitions), to observe and code many data that could be analysed in different ways (different data profiles), to control many variables, to make some basic statistical analysis. In the realization of the grid some difficulties raised especially in the definition of the data profile, during the observation (for example how to solve the problem of the presence of "dead moments" in the video), how to organize the results in the behavioural analysis and in the exportation of the data to other software for the statistical analysis.

Acknowledgments

This study was partially supported by the EU-ICT Project MIROR (Musical Interaction Relying On Reflexion, www.mirorproject.eu). The authors thank Fabrizio Grieco, trainer of Noldus Information Technology, for his helpful support during the construction of the Flow grid.

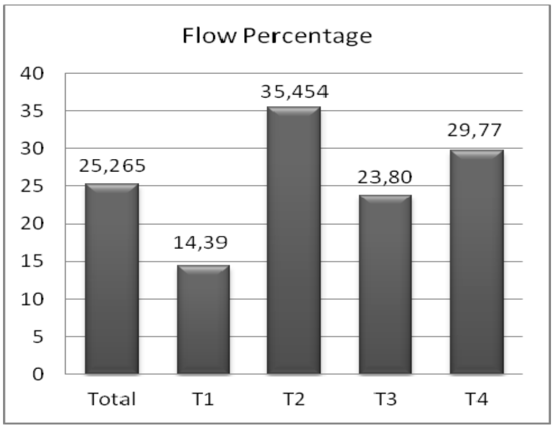
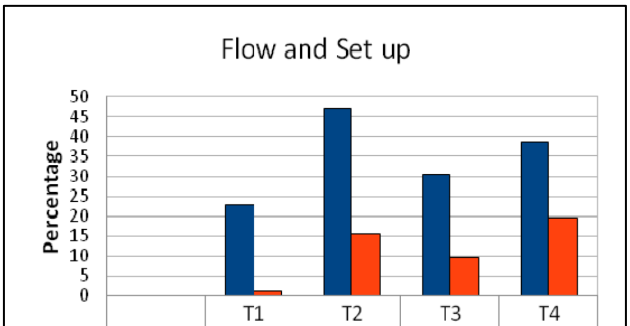


Figure 1. The percentage of the duration of the Flow, all subject in the four tasks.



		T1	T2	T3	T4
A	0	22,78481	47,11538	30,42254	38,55932
B	0	1,324503	15,57377	9,696969	19,60784

Figure 2. The percentage of the duration of Flow in all Subjects with the set up A (same) and B (very different).

References

1. Addessi, A.R. (2012). From Econ to the mirror neurons: Founding a systematic perspective on the reflexive interaction paradigm. In E. Cambouropoulos, et Al. (Ed.), *ICMPC-ESCOM2012 Proceedings*, 23-28 July 2012, Thessaloniki, Greece.

2. Addessi, A.R., Ferrari, L., Carugati, F. (2012). Observing and measuring the flow experience of children interacting with the MIROR Platform. In E. Cambouropoulos, et Al. (Ed.), *ICMPC-ESCOM2012 Proceedings*, 23-28 July 2012, Thessaloniki, Greece.

3. Addessi, A.R., Volpe, G. (2011). The MIROR Project. *Proceedings of the 6th Conference on Technology Enhanced Learning "Toward Ubiquitous Learning" (EC-TEL2011)*, Palermo (Italy), September 2011, Springer, 15-28.

4. Addessi, A.R., Pachet, F. (2006). Young children confronting the Continuator, an interactive reflective musical system. *Musicae Scientiae, Special Issue 2005-2006*, 13-39.

5. Byrne C., McDonalds, R., Carlton, L. (2003). Assessing creativity in musical composition: Flow as an assessment tool. *British Journal of Music Education*, **20**(3), 277-290.

6. Csikzsentmihalyi, M. (1997). *Finding flow the psychology of engagement with everyday life.* New York: Basic Book.

7. Custodero, L.A. (2005). Observable indicators of flow experience: A developmental perspective of musical engagement in young children from infancy to school age. *Music Education Research* **7**(2), 185-209.

8. Ferrari, L., Carlotti, S., Addessi, A.R., Pachet, F. (2004). Suonare con il Continuator è un'esperienza ottimale? In Biasutti, M. (Eds) *Proceedings of conference on Psychology and music education (PME04*), Padua 29/30 November, Padua, Italy: CLUEP Edition (available on CdRom).

9. Leman, M., Lesaffre, M., Nijs, L., Deweppe, A. (2010). User-oriented studies in embodied music cognition research, *Musicae Scientiae*, Special Issue, 203-224.

10. Pachet, F. (paper in preparation). MIROR: a software suite to study reflexive interactions in music improvisation and composition.

11. Pachet, F. (2004). On the design of Flow Machine. The future of learning. In M. Tokoro and L. Steels (Eds), *Learning zone of one's own*, IOS Press.

12. Sheridan, M., Byrne, C. (2002). Ebb and flow of assessment in music. *British Journal of Music Education* **19**(2), 135-143.

Measuring User Behavior in a Complex USAR Team Evaluation

N.J.J.M. Smets[1], M.A. Neerincx[1,2], R. Looije[1]

[1]TNO, Soesterberg, The Netherlands. Nanja.Smets;Mark.Neerincx;Rosemarijn.Looije@tno.nl; [2]Man-Machine Interaction Group, Delft University of Technology, Delft, The Netherlands

Introduction

Performing urban search and rescue (USAR) is a stressful and high demand task. The rescuers can use an extra pair of eyes and ears in the field, especially in places where they themselves cannot go. A ground robot can support rescuers in their task. To develop such a ground robot high fidelity and realism testing is needed [1]. High fidelity evaluations are difficult to set up, because the circumstances need to be complex and truthful. In this paper we performed a complex team experiment in the USAR domain. We measured a large number of variables to assess participants' performance, task load and emotion. This paper focuses on the question whether the analyses tools (Observer and FaceReader [2]) and measures gave an adequate indication of and insight in rescuers' operations for a robot with two levels of automation (no versus partial).

Methods

Participants. Ten firemen participated, three in a pilot and one participant canceled, so that 6 participants in total completed the evaluation (5 male and 1 female , average age of 42). The mean number of years the participants had a driver's license was 24. Three participants had experience with operating robots. And three participants played (first person) computer games.

Procedure. Participants first had to read a general instruction about the experiment and then fill out some questionnaires. After which the participant was trained to use the robot (see Figure 1). Training was first performed in line-of sight and later out of line-of-sight. When the participant was confident enough and the instructors were satisfied with their performance, they received instructions about the main evaluation. The main evaluation consisted of unit tasks (different tests of different level of abstraction) [3] and the tunnel scenario. The tunnel scenario consisted of a reconnaissance with a team: Unmanned Ground Vehicle (UGV) operator, Unmanned Aerial Vehicle (UAV) operator and the mission commander. The participant performed the role of UGV operator. The tunnel scenario took 30 – 40 minutes, the whole evaluation took 3 hours. The evaluation was between subjects, they either used the UGV without autonomy (tele-operation) or with autonomy (such as waypoint and speech navigation).

Task. The participants controlled the ground robot to perform reconnaissance after a tunnel accident. The robot was deployed to gather more information about the situation inside. The participants specifically had to look for cars in the tunnel, the lay out of the situation, victims (where, how many and where) and look for fire and dangerous substances, depicted by pictures of warning signs.

Figure 1. Left is a picture of the UGV, in the middle is a picture of the tunnel scenario on the right a participant is tele-operating the UGV.

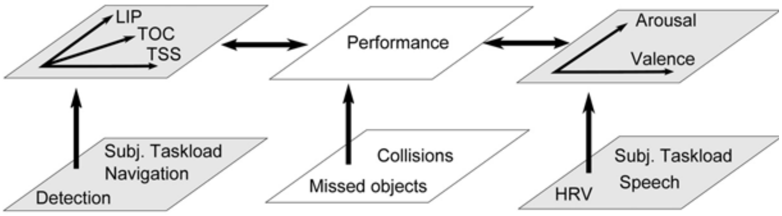


Figure 2. Relation of our measures and the task load (LIP Level of Information Processing; TOC Time Occupied; TSS, Task Set Switches), performance and emotion (arousal and valence) [reference].

Measures. We used the following measures, the relation between the measures and the cognitive task load, performance and emotional state are also depicted in Figure 2.

- Workload questionnaire. Every 2 minutes we asked the participant what his current workload was on a scale from 1 (none at all) to 5 (far too much). We also logged the reaction time to answer this question.
- Heart rate variability was measured with a belt around the chest.
- An observer rated the behavior of the participant. The behaviors that were rated are: communication with mission commander, situation report, tele-operating ground robot, operate robot by using speech commands, waypoint navigation (the last three behaviors are mutually exclusive). These behaviors were adapted from [4] to fit the available data. Novel is that also non mutually exclusive categories were used.
- Emotional state, the participant's face was recorded using a webcam and later analyzed using FaceReader.
- Performance, number of collisions: objective (observer) and subjective (by participant).

Results

First we looked at the behavior of the participants during the scenario. See Figure 3 for an example of how the data can be visualized in Observer. Figure 2 shows that this participant communicated a lot with the mission commander, tele-operated the UGV by using the compass rose, speech (in the beginning of the evaluation) and touched the operator control unit a few times (the screen showing the different camera and laser images from the robot [5]). These views give us a good overview of the behavior of the participants in time.

Behavior analyses for the rated behaviors is shown in Table 1. The first three participants only used tele-operation to control the robot, this means no data is available for speech and waypoint navigation. Remarkable is that participant four spent a lot of time navigating by speech and less time communicating with the MC or using the compass rose. The other participants in the automated condition (participants 4, 5 and 6) communicated more with the MC than the participants in the manual condition (1,2 and 3), this can be related to the usage of the situation report which was less for the participants in the automated condition. When the manual navigation is heavily used (participants 5 and 6), they also collided more.

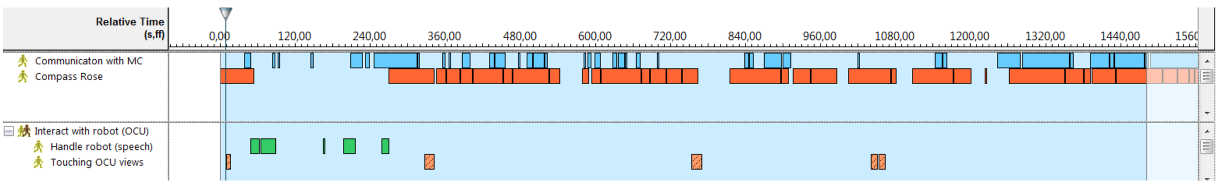


Figure 3. Screenshot of Observer with visualized rating data from the observers.

Table 1. The number and duration (seconds) of participants' behaviors (standardized data, range 0-1) and the number of UGV collisions (objective and subjective).

Partic.	Communication with MC	Compass Rose	Speech navigation	Waypoint navigation	Mapping/ situation report	Performance # of collisions
1	40, 0.52	41, 0.65	-	-	12, 0.23	10, 4
2	45, 0.33	50, 0.82	-	-	13, 0.12	7, 3
3	40, 0.17	53, 0.62	-	-	12, 0.15	5, 2
4	37, 0.45	17, 0.29	33, 0.31	3, 0.02	6, 0.07	6, 0
5	68, 0.51	87, 0.78	7, 0.08	3, 0.01	1, 0.01	19, Many
6	68, 0.50	64, 0.73	0, 0	10, 0.08	0, 0	17, 4

We also asked the participant how much effort certain tasks took (see Table 2). The tasks related to the automated conditions took little effort and using the compass rose took more effort. We successfully collected heart rate variability data. Because of the number of participants we could not test for significance, but rest data may differ from event data (such as finding a victim), see Figure 4. Due to the limited number of participants and the realistic scenario that caused events to be presented in a different order for each participant we were not able to link experienced workload to the heart rate data. The emotional state of the participants during the tunnel scenario is shown in Table 3. The face recognition did not work optimal because of the lighting conditions and the participant did not always face the camera, for instance when talking to the mission commander. Expressions of participant 6 were well-recognized because of her make up.

Conclusions

The detailed behaviors of the participants showed us what they were doing and for how long, thus giving us detailed understanding of human-robot teamwork and how certain behaviors interact with each other. The emotional data are interesting, but due to the nature of the task where the participant does not still in front of the camera (when talking to the MC) and the lighting conditions are not optimal, there were too much missing data. When the environmental conditions are better and with an eye tracker, we will probably get more useful data of the operator's emotion. With respect to heart variability data, the method of measuring should be improved. More extensive measuring in rest is needed, this way we expect a good baseline and better differences when events occur. The scenario needs to be improved, events should be more stressful to have a better effect on heart rate variability. For instance not only finding a victim, but also hearing or being able to interact with a victim.

Table 2. Effort on tasks ranging from 1 (no effort) to 5 (a lot of effort).

Task	Effort
Managing the speed and direction of the robot using the compass rose	3,3
Interacting with people (such as the mission commander) during the scenario	2,6
Mapping/ situation report	2.3
Managing the speed and direction of the robot using voice commands	2,0
Using waypoint navigation	2,0
Using voice commands to direct the robot	1,8

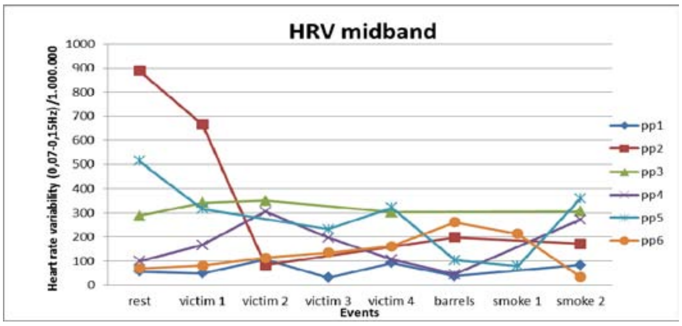
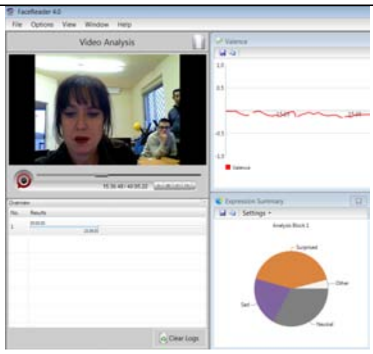


Figure 4. Heart rate variability of participants against the events in the scenario.

Table 3. Screenshot of FaceReader with mug shot of participant, valence chart with positive and negative dimensions and the momentary state in terms of Ekman's basic emotions (surprised, sad, neutral and other). The right part of table depicts the participant and the total number of times an emotion was recognized and how many seconds this emotion lasted in seconds during the tunnel scenario.



Subject	Happy	Disgusted	Sad	Angry	Surprised	Neutral
1	2 3.33s	5 7.33s	0	0	0	2 1.67s
2	0	2 3.80s	11 33.87s	8 10.80s	1 3.73s	31 89.21s
3	0	2 2.4s	0	0	0	0
4	1 1.33s	8 8.47s	1 2.00	0	1 1.53s	2 1.20s
5	0	0	0	0	0	1 2.07s
6	4 7.33s	0	42 77.41s	0	112 426.03	116 399.17s

Acknowledgements. This paper describes research done under the EU-FP7 ICT247870 NIFTi project. For more about NIFTi, please visit http://www.nifti.eu.We would like to thank the fire fighters of FDDo (Dortmund, Germany) and of SFO (Montelibretti, Italy) for their support.

References

1. Smets, N., Bradshaw, J., Diggelen van, J., Jonker, C., Neerincx, M., de Rijk, L., Senster, P., Sierhuis, M., ten Thije, J. (2010). Assessing human-agent teams for future space missions. *IEEE Intelligent systems* **25**(5), 46–53.

2. Noldus Information Technology. The Observer XT10, FaceReader 4.0. http://www.noldus.com/

3. Mioch, T., Smets, N.J.J.M., Neerincx, M.A. (submitted). Comparing Human-Robot Performances in Complex Situations by Means of Unit Task Tests. *Ro-Man. 21st IEEE International Symposium on Robot and Human Interactive Communication*, Paris, France, 9-13 September 2012.

4. Gómez, A.V. (2010). *Evolutionary design of human-robot interfaces for teaming humans and mobile robots in exploration missions*. Ph.D. dissertation, Universidad Politécnica de Madrid.

5. Larochelle, B. Kruijff, G. J. Smets, N. J. Mioch, T. and Groenewegen, P. (2011). Establishing human situation awareness using a multi-modal operator control unit in an urban search & rescue human-robot team. In *20th IEEE International Symposium on Robot and Human Interactive Communication. International Symposium on Robot and Human Interactive Communication* (RO-MAN-2011), Atlanta, GA, United States.

Unobtrusive Emotion Sensing in Everyday Life

Joyce Westerink[1], Anton van Boxtel[2], Geraldine Fitzpatrick[5], Wijnand Ijsselsteijn[3], Joris Janssen[1], Martin Ouwerkerk[1], Thérèse Overbeek[4], Petr Slovak[5,] Gert-Jan de Vries[1], Marjolein van der Zwaag[1]

[1]*Body Brain Behaviour Group, Philips Research, Eindhoven, The Netherlands, joyce.westerink@philips.com*
[2]*Department of Medical Psychology and Neuropsychology, Tilburg University, Tilburg, The Netherlands*
[3]*Human Technology Interaction, Eindhoven University of Technology, Eindhoven, The Netherlands*
[4]*Research & Development, Pluryn, Hoenderloo, The Netherlands*
[5]*Institute of Design and Assessment of technology, Vienna University of Technology, Vienna, Austria*

Decades of laboratory research in psychophysiology have shown us that our bodily reactions are connected to our emotions and to our mental state. Well-known examples are our heart rate reactions to demanding events, the change in our skin conductance in cases of various (positive and negative) types of arousal, and the way our breathing patterns reflect surprises and other excitement. These physiological parameters and many more thus can give us insight into how we feel and how we react to situations.

This opens up opportunities for applications in everyday life in various areas of society. For instance, becoming aware of our bodily reactions might also help us to realize how we feel about a certain situation. It might be a source of reflection for ourselves, so that we get grip on the events that comprise our busy lives. Or we could trust someone else to receive our physiological signal, maybe our romantic partner, with whom we want to feel close despite a physical distance. Maybe our physiological signal could also serve as an additional clue for professionals who provide care to us, a clue that helps them to treat us well and keep us comfortable. As a last possibility, our physiological signal could be interpreted by the devices we use, so that they can adapt their functioning to the mental state we are in or would like to be in.

However, several steps have to be taken to arrive at really working products. For instance, the physiological measurements have to be done in an unobtrusive way. Some people find it acceptable to carry a heart measurement device during running, but most of us would not think of wearing a similar device throughout each and every working day. Thus unobtrusiveness is an important feature in many of the applications mentioned above, and it fosters the research and development of miniaturized wearable sensors.

In addition, the measured data have to be interpreted correctly for each individual. Traditional psychophysiological research usually averages effects over people in order to get significant results. In real life applications, however, signal interpretations have to be one-shot, first-time right, which considerably complicates things. On the other hand, when doing intra-individual measurements, it is possible to look at smaller changes and to really get to know the user and his reactions, and thus increase the accuracy of the interpretation.

The need for unobtrusive devices and individual interpretations gives rise to a new branch of more applied research, of which I'll show a few examples: a wristband skin conductance sensor, and individual modeling of the relation between heart rate and emotions. I will also present and discuss some prototype applications we based on these skin conductance and heart rate signals: an affective music player that directs you mood and a live connectedness token for romantic partners.

Pine Weevil (*Hylobius abietis*) Feeding Pattern on Conifer Seedlings

Frauke Fedderwitz[1], Niklas Björklund, Velemir Ninkovic, Göran Nordlander

Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden.
[1]frauke.fedderwitz@slu.se

Abstract

The pine weevil (*Hylobius abietis*) is one of the most important forest pests in Europe, yet there is very little known about its detailed feeding behaviour. We study the temporal feeding pattern of individual pine weevils of both sexes for 24 hours with two treatments, intact and girdled seedlings. Properties of a meal, such as feeding duration, size and ingestion rate are of particular interest. The shortest interval considered to separate one feeding bout from another, the meal criterion, has never been published and it is only available for a few other insect species. Video recordings are analysed for feeding behaviour (e.g. duration of feeding activity, interval length between feeding activities, movements between and within feeding scars). We measured general activity patterns as there is insufficient knowledge on the daily behavioural patterns. We thereby got an in-depth view of the pine weevil feeding activity that would otherwise be difficult to assess.

Introduction

Herbivorous insects reduce the fitness of plants directly and indirectly by feeding despite inducing plant defence systems against herbivory [1]. Trees have especially low defence and tolerance as seedlings in comparison to other phases in plant ontogeny except for over-mature and senile plants [2].

The pine weevil (*Hylobius abietis*) is economically one of the most important forest pests in Europe [3]. Adults feed on the stem bark of conifer seedlings [4] and can cause seedling mortality of up to 90 % in the first three years [5]. Economic losses tend to be especially high in conifer forests that are managed by clear-cutting and replanting [3]. Despite the importance of behavioural components of insect feeding for understanding bases of host plant resistance [6-9], a lot of aspects of pine weevil feeding behaviour have not yet been studied. Studies on feeding behaviour in this species have mostly been concentrating on the total amount of consumed bark after a specific time period in relation to different aspects, for example, presence of other weevils, weevil size, microclimate, chemical composition of the bark or treatments with light and/or nitrogen [10-13] or as comparisons between different tree species [14-16].

In this ongoing project we are assessing the detailed pine weevil feeding behaviour and time budget on Norway spruce (*Picea abies*) seedlings during a 24-hour period. Generally feeding over longer durations, i.e. several days, often shows little variation in total intake of nutrition between individual insects, whereas the detailed feeding pattern of individuals varies greatly even under controlled environmental conditions [17]. The detailed feeding behaviour can be described by groups of feeding events, i.e. meals. These can be determined by the bout criterion, the shortest non-feeding interval considered to be between meals. The bout criterion is based on the distribution of non-feeding intervals since it can be expected that animals make a large number of short non-feeding intervals within a meal and relatively few longer intervals between meals [e.g. 18]. The bout criterion itself has an indirect importance for understanding feeding behaviour as it allows separating meals on an ethological meaningful basis. Thereby meal properties (size, duration and ingestion rate) and potentially other indicators can be related to initiation and termination of meals and thus may have important implications for insect-plant interactions. By understanding these properties it might be possible to influence the length of a meal. Shorter meals might reduce the risk of large feeding scars that can endanger the seedling's survival. For testing the effect of previous damage caused by larger feeding scars we are also measuring the effect of girdling. Girdling represents a type of pine weevil damage that is often observed on planted seedlings [19]. Furthermore wounding of plants can induce different effects of defence mechanisms as shown in experiments on several conifer species [20-22].

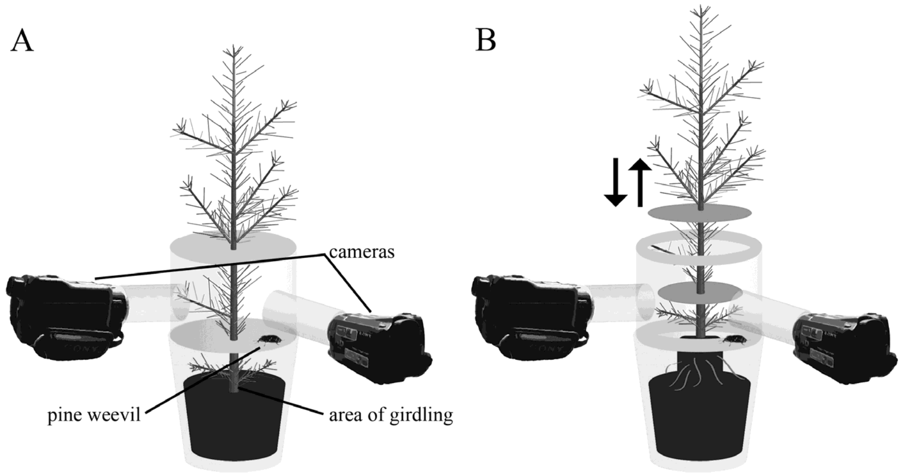


Figure 1. Experimental set-up during observation (A). A small water tube and a dark coloured shelter for weevils to hide under inside the cylinder are not shown. Changing of seedling after a 24 hour acclimatization period (B). The cylinder had a hole in the bottom and the lid which was closed by two plastic discs placed around the plant. Thereby it was possible to replace the plant with a fresh seedling with a minimum disturbance of the weevil that was inside the cylinder.

Materials and Methods

We video recorded the behaviour of 6 male and 6 female pine weevils on one-year old Norway spruce seedlings under 24 hour periods for assessing the bout criterion and time budget. Two cameras were placed at a 45 degree angle for high visibility of weevil activity regardless of the feeding position of the weevil. We limited the weevil's access to the middle part of the seedlings by placing a transparent plastic cylinder around the stem (see Figure 1A). This area can be assumed to represent a rather homogenous feeding source with regard to the distribution of resin ducts and growth factors. The experiment was conducted at room temperature under an artificial day-night cycle of 7 h light – 6 h darkness – 11 h light. During the day the light was provided by several light sources with a combined intensity of 59 μmol m^{-2} s^{-1} (Skye Instruments LTD, Sky 200 SKP). A light bulb at low intensity (0.1 μmol m^{-2} s^{-1}) as used to imitate a Swedish summer night with an additional red light (0.9 μmol $m^{-2}s^{-1}$) to increase the visibility during the night.

A single weevil was placed with a seedling in the cylinder 24 hours before filming so that the weevil had time to acclimatize to its new surroundings. Pre-test weevils were rarely feeding during this first day although they were eating before and afterwards when placed in Petri dishes with pieces of Scots pine (*Pinus sylvestris*). After the acclimatization phase the plant was replaced with a new seedling to avoid effects of possible feeding scars (see Figure 1B). The weevil was disturbed as little as possible. Video recording was started when the new seedling was installed and ended after 24 hours. The area eaten (mm^2) was measured on both plants and the location of feeding scars on the second seedlings. Each weevil was tested with two treatments: an intact seedling and a seedling that had been manually girdled under the first node 24 hours prior to filming to test effects of induced defence. These two treatments were started alternating between the weevils. Between the treatments the weevils were placed for two days in Petri dishes with Scots pine pieces.

Feeding behaviour was defined as manipulation of bark, needles and eggs with the mouth parts. It was not possible to distinguish periods of bark removal and ingestion. Consequently it cannot be completely asserted that the weevil was feeding when it showed feeding behaviour as defined above (from here on only feeding).

The duration of each feeding occurrence and the length of intervals between them were measured (The Observer XT 10, Noldus Information Technology, Wageningen, The Netherlands). In addition locomotion behaviours were noted. All behaviours were measured continuously. The length of non-feeding intervals was assigned to bins of the next full second. The distribution of the frequency of all non-feeding intervals was used to assess the bout criterion. The bout criterion was then used to determine average meal length and sizes as well as feeding rate. For the time budget the day was split up into one hour bins in which the additive time spent feeding or in locomotion was calculated. From these values relations to the dark/light cycle as well as between different activities and individuals were made. Additionally we measured the feeding scar distribution on the seedling and in which order they were made. We will thereby be able to assess if there is a general pattern in location of feeding on the plant. We also evaluated how often a pine weevil returned to an old scar in order to continue feeding there. Furthermore we noted direction in which they continue to feed after each non-feeding interval. That means we determined if the first bite after a non-feeding interval is extending the feeding scar upwards, downwards or horizontally. In connection with the meal determination we will also be able to see whether there is a link between meals and feeding scars, e.g. one meal is one feeding scar. This kind of information might indicate avoidance of locally induced defence systems or large resin ducts. The differences between treatments were assessed for all measurements.

Results and Discussion

So far only a part of the data has been analysed. This preliminary analysis indicates that the bout criterion lies close to 16 seconds and does not show a difference for the two different treatments. This duration is relatively short as the bout criterions established for other insects are in the range of minutes; examples include the Colorado potato beetle (*Leptinotarsa decemlineata*) with 4.8 minutes [8] and different grasshopper species such as *Taeniopoda eques* with 8 minutes [7] or *Locusta migratoria* with 2-6 minutes [9]. The short bout criterion could indicate a larger difference between intrameal and intermeal intervals in pine weevils. The longer non-feeding intervals recorded during this experiment range several hours and were probably mostly incomplete due to the cut off points before and after 24 hours.

Some weevils showed locomotion and feeding behaviour influenced by the dark/light rhythm whereas others expressed a similar activity pattern during both phases. In some cases there might even be an influence of the lighting on either behaviour but not on the other. Individuals expressing a daily pattern in locomotion showed more activity during the dark phase. Similar results have been shown previously [23]. For feeding activity Merivee and co-workers [23] found a significant pattern for male weevils. Our preliminary data points however towards an effect of individuals, meaning that there are both females and males that seem (not) to be influenced. In individuals showing a temporal pattern, feeding occurred mainly towards the end of the night or shortly afterwards. Other weevil species have been found to be predominantly nocturnal in relation to feeding behaviour [24; 25].

Based on these results it seems that the presented method is suitable for assessing the bout criterion and a detailed time budget of pine weevils. The detailed behaviour that is measured in this study is partly in the range of seconds and a lot of aspects need continuous observation. Therefore it would be very difficult to assess this type of data with a more automated or less precise method, e.g. instantaneous or one-zero sampling, at this stage. It should be possible to transfer the set up to other herbivorous insects for comparable studies.

References

1. Bernays, E.A. (1998). Evolution of feeding behavior in insect herbivores - Success seen as different ways to eat without being eaten. *Bioscience* **48**, 35-44.

2. Boege, K.,Marquis, R.J. (2005). Facing herbivory as you grow up: the ontogeny of resistance in plants. *Trends in Ecology & Evolution* **20**, 441-448.

3. Långström, B.,Day, K.R. (2004). Damage, control and managment of weevil pests, especially Hylobius abietis. In: *Bark and wood boring insects in living trees in Europe, a synthesis*. (Eds.) F. Lieutier, K.R. Day, A. Battisti, J.-C. Grégoire,H.F. Evans. Springer. Dordrecht. 415-444.

4. Day, K.R., Nordlander, G., Kenis, M.,Halldórson, G. (2004). General biology and life cycles of bark weevils. In: *Bark and wood boring insects in living trees in Europe, a synthesis*. (Eds.) F. Lieutier, K.R. Day, A. Battisti, J.-C. Grégoire,H.F. Evans. Springer. Dordrecht. 331-349.

5. Petersson, M.,Örlander, G. (2003). Effectiveness of combinations of shelterwood, scarification, and feeding barriers to reduce pine weevil damage. *Canadian Journal of Forest Research-Revue Canadienne de Recherche Forestiere* **33**, 64-73.

6. Jones, C.G., Hoggard, M.P.,Blum, M.S. (1981). Pattern and process in insect feeding-behavior: a quantitative-analysis of the mexican bean beetle, Epilachna varivestis. *Entomologia Experimentalis et Applicata* **30**, 254-264.

7. Bright, K.L., Bernays, E.A.,Moran, V.C. (1994). Foraging patterns and dietary mixing in the field by the generalist grasshopper Brachystola magna (Orthoptera, Acrididae). *Journal of Insect Behavior* **7**, 779-793.

8. Mitchell, B.K.,Low, R. (1994). The structure of feeding behavior in the Colorado potato beetle, Leptinotarsa decemlineata (Coleoptera: Chrysomelidae). *Journal of Insect Behavior* **7**, 707-729.

9. Simpson, S.J. (1982). Patterns in feeding: a behavioral analysis using Locusta migratoria nymphs. *Physiological Entomology* **7**, 325-336.

10. Bylund, H., Nordlander, G.,Nordenhem, H. (2004). Feeding and oviposition rates in the pine weevil Hylobius abietis (Coleoptera: Curculionidae). *Bulletin of Entomological Research* **94**, 307-317.

11. Wainhouse, D., Boswell, R.,Ashburner, R. (2004). Maturation feeding and reproductive development in adult pine weevil, Hylobius abietis (Coleoptera : Curculionidae). *Bulletin of Entomological Research* **94**, 81-87.

12. Wainhouse, D., Staley, J., Johnston, J.,Boswell, R. (2005). The effect of environmentally induced changes in the bark of young conifers on feeding behaviour and reproductive development of adult Hylobius abietis (Coleoptera : Curculionidae). *Bulletin of Entomological Research* **95**, 151-159.

13. Wainhouse, D., Staley, J.T., Jinks, R.,Morgan, G. (2009). Growth and defence in young pine and spruce and the expression of resistance to a stem-feeding weevil. *Oecologia* **158**, 641-650.

14. Manlove, J.D., Styles, J.,Leather, S.R. (1997). Feeding of the adults of the large pine weevil, Hylobius abietis (Coleoptera: Curculionidae). *European Journal of Entomology* **94**, 153-156.

15. Månsson, P.E.,Schlyter, F. (2004). Hylobius pine weevils adult host selection and antifeedants: feeding behaviour on host and non-host woody Scandinavian plants. *Agricultural and Forest Entomology* **6**, 165-171.

16. Carrillo-Gavilán, A., Moreira, X., Zas, R., Vilà, M., Sampedro, L. (2012). Early resistance of alien and native pines against two native generalist insect herbivores: no support for the natural enemy hypothesis. *Functional Ecology* **26**, 283-293.

17. Simpson, S.J. (1995). Regulation of a Meal: Chewing Insects. In: *Regulatory Mechanisms in Insect Feeding*. (Eds.) R.F. Chapman,G. de Boer. Chapman & Hall. New York. 137-156.

18. Simpson, S.J. (1990). The pattern of feeding. In: *Biology of grasshoppers*. (Eds.) R.F. Chapman,A. Joern. John Wiley and Sons, Inc. Somerset, New Jersey, USA. 73-104.

19. Lindgren, B.S., Nordlander, G.,Birgersson, G. (1996). Feeding deterrence of verbenone to the pine weevil, Hylobius abietis (L) (Col, Curculionidae). *Journal of Applied Entomology - Zeitschrift für angewandte Entomologie* **120**, 397-403.

20. Lewinsohn, E., Gijzen, M.,Croteau, R. (1991). Defense-mechanisms of conifers: Differences in constitutive and wound-induced monoterpene biosynthesis among species. *Plant Physiology* **96**, 44-49.

21. Moreira, X., Zas, R.,Sampedro, L. (2012). Quantitative comparison of chemical, biological and mechanical induction of secondary compounds in Pinus pinaster seedlings. *Trees-Structure and Function* **26**, 677-683.

22. Nault, J.R.,Alfaro, R.I. (2001). Changes in cortical and wood terpenes in Sitka spruce in response to wounding. *Canadian Journal of Forest Research-Revue Canadienne de Recherche Forestiere* **31**, 1561-1568.

23. Merivee, E., Sibul, I.,Luik, A. (1998). Diel dynamics of feeding and locomotor activities in the large pine weevil, Hylobius abietis. *Baltic Forestry* **4**, 59-62.

24. Barratt, B.I.P., Evans, A.A.,Ferguson, C.M. (1995). Circadian patterns of oviposition and feeding in Listronotus bonariensis (Kuschel) (Coleoptera: Curculionidae) in the laboratory. *Proceedings of the Forty Eighth New Zealand Plant Protection Conference* (Rotorua, 08-10 August 1995), 219-223.

25. Duan, J.J., Weber, D.C., Hirs, B.,Dorn, S. (1996). Spring behavioral patterns of the apple blossom weevil. *Entomologia Experimentalis et Applicata* **79**, 9-17.

The Cause of Stereotypic Behaviour in a Male Polar Bear (*Ursus Maritimus*)

P.W.F.H. Cremers[1], S.L. Geutjes[2]

[1,2] Wildlife Management, University of Applied Sciences Van Hall Larenstein, Leeuwarden, The Netherlands.
[1]peggy_cremers@hotmail.com, [2]sannegeutjes@hotmail.com

Abstract

This study was focused on finding the cause of stereotypic behaviour in a male polar bear (Ursus maritimus). The zoo wants to reduce the stereotypies present in their male polar bear. However, this can only be accomplished if the cause of the behaviour is identified and targeted [1]. In this study external factors from 3 different categories (i.e. husbandry, geography and environment) are investigated to find out which possible stressors affect the polar bear's behaviour. Video recordings of the polar bear's behaviour and the different external factors were processed with The Observer XT 7.0 software program and later analysed. Many different factors influence the male's behaviour, both positively and negatively. The polar bear's response to these external factors lead to believe his stereotypies have a coping origin, where the animal uses the rewarding endorphins, that are released due to his stereotypies, to cope with continuous stressful situations in his environment [2]. Further experimental research is recommended to investigate to what degree the different factors found in this study can be altered to create a more suitable environment for this polar bear.

Key words: Stereotypic behaviour, Ursus maritimus, polar bear, captivity.

Introduction

Polar bears have a long history of high popularity in zoo settings. However, many studies indicate that this wide-ranging species expresses a wide variety of abnormal repetitive behaviours in captivity [3], such as stereotypic walking, head swinging and repetitive swimming bouts [4, 5]. This is also the case in the 13 year old male polar bear in this study, which spends large portions of his day expressing stereotypic behaviour. Stereotypic behaviour is often a sign of a decreased welfare in an animal [2], because their choice to express natural behaviour has become limited [6]. The expression of natural species-specific behaviours is important for a zoo to maintain appropriate levels of animal welfare [7], create a satisfying visitor experience and also enable proper public education [8]. To enable positive visitor experiences and public education about polar bears and their natural species-specific behaviour and to increase the polar bear's welfare, the stereotypies present in this polar bear's behavioural repertoire need to be reduced. His long-term stereotypies must be treated, however, this can only be accomplished if the cause of the behaviour is identified and targeted [1]. This study is focused on identifying the cause of the stereotypic behaviour in this polar bear male by testing the effects of external factors on his behaviour.

Methods

In this study, different husbandry-, geography- and environment-related factors were investigated to see if one or more of these trigger the stereotypies of the polar bear. External factors that were scored in relation to husbandry were feeding regimes, keeper presence and its purpose, and enrichment effects. Geography-related factors covered geographical preferences within the exhibit and related substrate use, while environment-related factors focused on visitor numbers and employees, the behaviour of other polar bears (an adult female and her two cubs) at the zoo, construction work noises and the effects of time of day. The events that were scored from the female and cubs were defecating, vocalizing, the presence of food in their exhibit, being within sight of the male polar bear in the outside exhibits and being inside the night dens. These events were chosen because of their visual-, sound- and/or scent aspects, which could affect the male polar bear's behaviour. Through continuous recording and focal sampling the male polar bear's behaviour was observed to determine the extent of his stereotypic behaviour in relation to these external factors. A total of 116 observation sessions of approximately 30 minutes were conducted on set times throughout the day between 09.15 and 17.15 daily over a 24-day period. This resulted in 54 hours and 18 minutes of behavioural data. Four behaviour categories (i.e. ‘active’, ‘inactive’, ‘stereotypic’ and ‘in night den’) were used to describe the polar bear’s behavioural state. Also three short point behaviours were scored. These point behaviours can either be a stereotypy by itself (i.e. ‘head swing’), is believed to be related to stress (i.e. ‘yawn’) or short distractions (i.e. ‘variation’) from the male’s stereotypic pacing. The ‘variation’ point behaviour means that the polar bear was distracted for a few seconds but not abandoned his stereotypic pacing to investigate any further. All relevant polar bear behaviours were video recorded and all observed abnormalities and external factors were scored on an observation sheet at the time at which the event occurred. The Observer XT 7.0 computer program (Noldus Information Technology, Wageningen, The Netherlands) was used to score the frequency, duration and location of all state behaviours expressed by the polar bear, as well as the number of occurring point behaviours. Information about external factors possibly related to the male polar bear’s behaviour was also scored. The Observer XT’s visualization function behaviour scores were used to register which and to what extend behaviours and behavioural shifts occurred while external factors were present. After data were collected from the different external factors, a total of 354 data points were processed from 55 randomly chosen observation sessions to gather baseline data about behavioural shifts. These baseline data points present behavioural shifts between the ‘active’, ‘stereotypic’ and ‘in night den’ behaviours at moments where no external factors were occurring. Over 300 baseline data points were needed for statistical tests to detect differences between baseline situations and moments where external factors were occurring. The 55 chosen observation sessions were gathered from the total 116 observation sessions. These baseline data points were then compared to the data collected about the different external factors from this study.

Results

Data analysis shows that stereotypic behaviour covers a large part of the polar bear’s daily activity. Stereotypic behaviour was observed for 45.54% (±3.89) of the total data collection period. After stereotypic behaviour, the male was observed most in his night den (35.79% (±3.87)). It must be noted that the ‘in night den’ behaviour is an unknown behaviour because the observers could not see the polar bear in his night den. ‘Active’ behaviour occurred for 16.29% (± 2.34) of the time, while ‘inactive’ behaviour was only observed for 1.75% (±1.0) of the total observation time. The male only displays stereotypic behaviour on concrete exhibit floors, with a preference for two specific areas in his enclosure. New enrichment was provided on 7 occasions which resulted in 17 observed enrichment interaction moments covering only 3.12% of the total observation period. 3 interactions were with food-based enrichment (e.g. frozen fish), 3 with toy-based enrichment (e.g. tyres) and 11 other interactions were substrate-related (e.g. sand, shells or wood chips). No significant effects were found between the male polar bear’s behaviour and enrichment objects. However, 80% of the inactive behaviour was observed after keepers provided new substrate. A total of 292 head swings, 69 yawns and 144 variations were also observed. The male polar bear displayed significantly more stereotypic behaviour in mornings compared to afternoons ($F(3.104)=5.358$; $P=0.002$), while he spent more time in his night den in late afternoons ($F(3.112)=4.591$; $P=0.005$). The polar bear displayed significantly less stereotypic behaviour during observation sessions were he was fed ($F(2.88)=10.920$; $P=0.001$), while stereotypic behaviour was 30% higher on days where he was not fed at all. Keeper presence decreased stereotypic behaviour and increased shifting back and forth between active behaviour and retreating into the night den ($X2=237.190$; $df=8$; $P\leq 0.001$). The presence of another bear in the night den area also resulted in increased back and forth shifting between active behaviour and retreating into the night den ($X2=86.385$; $df=8$; $P\leq 0.001$). An increased employee count near the exhibit elicited an increase in the point behaviour ‘variation’, while traffic passing his exhibit, or noises over 70dB showed to cause an increase in stereotypic point behaviours ‘head swing’ and ‘variation’.

Conclusions

This study showed that many external factors affect the polar bear’s behaviour in both a positive and negative way. It was therefore impossible to point out one specific stressor that is causing his stereotypies. Motivational frustration described by Olsson et al. [2] as a cause of the bear’s stereotypic behaviour seemed to be a well-supported option at the start of this study. However, motivational frustration is usually focused on a specific behaviour that cannot be completely executed [9]. Due to the large number of factors affecting his behaviour positively and negatively, this motivational frustration now seems like a less suitable cause. Coping on the other hand now seems more plausible, because in such a coping mechanism an abnormal repetitive behaviour is expressed for the release of endorphins that an animal then uses to cope with continuous stressful situations [2]. Keeper presence for example resulted in shifting continuously between the night dens and his outside enclosure, making his behaviour seem more ‘restless’. This might suggest the male polar bear’s anticipation of food. To either support or discard the finding in this study, further experimental research is recommended to test possible relations between behaviour and the stimulation of positive factors such as increased feeding and regular provision of soft substrate. Also, the reduction/elimination of negative factors such as traffic and loud noises over 70dB might result in decreased stereotypies. It is also recommended to assess the polar bear’s current living conditions and maybe consider different housing and husbandry strategies in the near future. Recommendations on current husbandry conditions include increased feeding and more variation and regulation of new enrichment objects. For housing conditions, a larger exhibit would be preferred, or at least provide more soft substrates along with implementing more permanent enclosure attributes such as a viewing platform or more sloped edges leading to the water moats.

References

1. Vickery, S.S., Mason, G.J. (2003). Understanding stereotypies in captive bears: the first step towards treatment. *Animal Behaviour Research Group, Department of Zoology* (Oxford University. UK), 38-47.

2. Olsson, I.A.S., Würbel, H., Mench, J.A. (2011). Chapter 9. Behaviour. *Animal Welfare. 2nd Edition. Cabi Publishing*, 147-148.

3. Moore, D., Shepherdson, D. (2010). Behavior Management. IHC Polar Bears International. <http://www.polarbearsinternational.org/sites/default/files/ihc/IHC_BehaviorMgt.pdf> Accessed 28 September 2011.

4. Wechsler, B. (1991). Stereotypies in polar bears. *Zoo Biology* **10**, 177-188.

5. Wechsler, B. (1992). Stereotypies and attentiveness to novel stimuli: A test in polar bears. *Applied Animal Behavior Science* **33**, 381-388.

6. Clubb, R., Mason, G. (2003). Captivity effects on wide ranging carnivores. *Nature* **425**, 473-474.

7. Skibiel, A.L., Trevino, H.S., Naugher, K. (2007). Comparison of several types of enrichment for captive felids. *Zoo Biology* **26**, 371-381.

8. Kutska, D. (2009). Variation in Visitor Perceptions of a Polar bear's Enclosure Based on the Presence of Natural vs. Un-Natural Enrichment Items. *Zoo Biology* **28**, 292-306.

9. Mason, G., Rushen, J. (2006). Stereotypic Animal Behaviour. *Fundamentals and Applications to Welfare. 2nd Edition.* CAB International, 34.

Analysis of Sequences in Aggressive Interactions of Pigs for the Development of an Automatic Aggression Monitoring and Control System

M. Oczak[1,2], A.M. Costa[3], G. Ismailova[3], L. T. Sonoda[4], M. Fels[4], J. Hartung[4], M. Guarino[3], S. Viazzi[1], D. Berckmans[1], E. Vranken[1,2]

[1]*M3-BIORES: Measure, Model & Manage Bioresponses, Katholieke Universiteit Leuven, Leuven, Belgium;* [1]*stefano.viazzi@biw.kuleuven.be,* [1]*daniel.berckmans@biw.kuleuven.be,* [1]*erik.vranken@biw.kuleuven.be*

[2]*Fancom Research, Fancom, Panningen, The Netherlands;* [2]*moczak@fancom.com,* [2]*evranken@.fancom.com*

[3]*Department of Veterinary Science and Technologies for Food Safety, Faculty of Veterinary Science, Milan, Italy;* [3]*annamaria.costa@unimi.it,* [3]*gunel.ismayilova@unimi.it,* [3]*marcella.guarino@unimi.it*

[4]*Institute for Animal Hygiene, Animal Welfare and Farm Animal Behaviour University of Veterinary Medicine Hannover, Foundation, Hannover, Germany;* [4]*liliasonoda@hotmail.com,* [4]*michaela.fels@tiho-hannove.de,* [4]*joerg.hartung@tiho-hannover.de*

Introduction

Aggression is one of the most significant welfare problems when pigs are grouped in our modern production systems. When introduced to unfamiliar conspecifics, pigs naturally engage in aggressive interactions to determine their social hierarchy status [4]. Ordinarily, when the hierarchy is already established in a group, levels of aggression should be relatively low [15]. However, conditions of confined environment like limited space allowance [19], feeding systems promoting competition [13], barren environment [5], low fibre feed composition [16] and repeated changes in group composition [13] can result in persistent aggression. The most obvious negative impact of aggression is the skin lesions - which are often seen on the head, ears, shoulders, flanks and hindquarters [11, 14]. In terms of internal physiology, aggression results in activation of the sympathetic-adrenal-medullary axis and the hypothalamic-pituitary-adrenal axis, resulting in increased heart rate [12], increased plasma cortisol concentrations, increased plasma epinephrine and norepinephrine levels [13]. Besides the negative impact of aggression on health and welfare of pigs it results in lower productive performance, by increasing return to oestrus rate in adult sows [3,17] and by decreasing growth rates of growing pigs [18]. Although the problem of aggression was investigated by many researchers, there was no successful, long term, practical solution developed for lowering aggression level among pigs. There is a necessity for more basic studies in order to understand the phenomenon properly [10] and new solutions must be developed to the problem of aggression among pigs. Basic studies on pig aggression reveal that during a contest, animals successively sample bits of information concerning their relative fighting ability, by observing displays or otherwise gathering relevant information through the behaviour of the opponent. Therefore, when a sufficiently certain assessment cannot be achieved by means of a particular display, the animals will switch to another type of behaviour with a higher information return. This will cause a contest to proceed through a number of different phases, each with a typical behavioural characteristic. Each new phase will consist of more costly behaviour and the last phase will consist of direct sampling of actual fighting ability, through overt, dangerous fighting [10]. Therefore identification of patterns in gradual development of aggressive behaviour should allow for prediction of highly damaging aggressive behaviours expressed in the final phase of these interactions. In order to stop (prevent) these behaviour among pigs with automatic monitoring and control system (PLF –Precision Livestock Farming) we defined the following objective for this research: Identification of differences between aggressive sequences that lead to highly damaging biting behaviour and those that don’t lead to this behaviour.

Aggression monitoring and control by PLF technology

Automatic monitoring and control system of pig aggressive behaviour by utilization of Precision Livestock Farming (PLF) technology offers possibilities for permanent management of highly aggressive attacks. PLF is currently regarded as the heart of the engineering endeavour towards sustainability in (primary) food production. Its application allows making optimal use of knowledge and information from the animal in the monitoring and

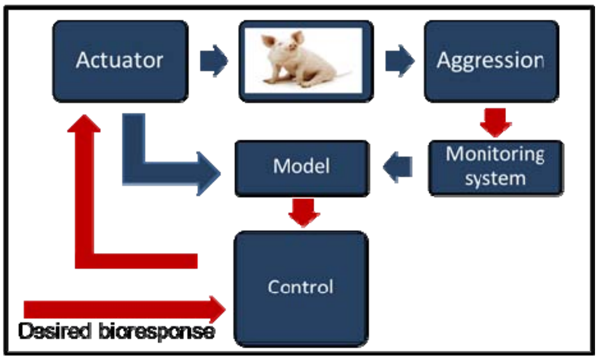


Figure 1. Aggression control by PLF.

control of processes [1]. Application of PLF technology as an automatic aggression monitoring and control system offers a new, unprecedented possibility to effectively lower aggression level among pigs. An integrated automatic monitoring and control system is one which collects information from a variety of sources, including sensors, databases and knowledge bases, processes the data and provides outputs, which may be recommendations to the producer, or direct process control actions [7]. The basic component of aggression monitoring and control system is an automatic monitor which by dynamic analysis of the individual interactions between animals will be able to detect aggressive attacks automatically. The second component of the system is an automatic control which by utilization of an actuator (i.e. sound, smell) will change pigs behaviour in a way that aggression level will be reduced (Figure 1). The strategy for monitoring and controlling of aggressive behaviour among pigs developed in accordance with PLF concept (Figure 1) comprises in using the aid of modern technology in order to develop online tools that can identify detailed animal behaviours in an automatic and continuous way, with an accuracy that allows detecting specific behavioural aspects. When the system automatically detects early signs that are used for predicting when the animal is entering in a particular status that leads to aggressive behaviour, it will use various triggers as actuators, such as light and sounds, to attract and redirect the animals' attention. An automatic reward mechanism, such as provision of food, will also be used to straighten animal responsiveness and to control their behaviour [2]. The most effective actuating mechanism will be chosen in a selection procedure comprising of series of practical experiments. The innovative approach is the development of a real time monitor that allows redirecting undesired livestock behaviour.

Materials and Methods

An observation was carried out at a commercial farm, in practical conditions, during 3 days after mixing on a group of 11 male pigs weighing on average 23 kg and kept in a pen of 4m x 2.5m. A total of 8 hours of video recordings were taken during the 3 days with a top-view camera in the pen and manually labeled afterwards. The video recordings were performed using a camera (Allied Vision Technologies®, model F080C) with 4.8 mm lens, placed above the pen in central position at a height of 2.3m, that permitted a top view image of the whole pen. Colour images were captured with a frame rate of 11 frames per second with a resolution of 1032 x 778 pixels. The videos were stored in a computer for later analysis. Each day a number of video recordings were registered (day 1: 2 h, day 2: 3 h, day 3: 3 h). Data obtained as a result of the visual analysis of the video recordings (labelling procedure) were carefully analyzed in order to identify sequences in pigs aggressive behaviour. An aggressive sequence was defined as a series of aggressive interactions following one after another, between at least two pigs. Successive aggressive behaviours had to be performed by the same pigs - within 30 seconds time period - to be labelled as part of the same sequence. Within aggressive sequences an initiating phase (nosing phase) was distinguished. The phase was defined as consisting of nose to nose, ear chewing and mounting behaviours [6,8,9,10]:

Nose to nose - the nose approaches the snout or head of the receiver
Ear chewing - chewing movements while the ear of a pen mate is in the mouth
Mounting - mount with front legs on the back of another pig while hind legs stay on the ground

Results and Conclusions

A total of 157 sequences of aggression were observed during our observations. Out of which 100 (63.7%) were classified as sequences that didn't lead to biting behaviour and 57 (36.3%) were classified as sequences that led to biting behaviour (Table 1). Duration of sequences that led to biting behaviour was longer (38.8 s, P<0.01) than of sequences that didn't lead to biting behaviour (18.1s) (Table 2). Initiating phase of sequences that led to biting behaviour was lasting for 3.3 s while of sequences that didn't lead to biting behaviour 1.9 s (Table 3).

We identified a difference between duration of aggressive sequences that led to biting behaviour and those that didn't lead to this behaviour. Duration of sequences that led to biting behaviour was longer (38.8 s, P<0.01) than of sequences that didn't lead to biting behaviour (18.1 s). The observed difference is an indication for the automatic monitoring and control system of pig aggression that aggressive sequences lasting longer are more probable leading to biting behaviour. Although not fully significantly, it was observed a trend (P=0.07) in longer duration of initiating phase between sequences that led (3.3 s) and didn't lead to biting behaviour (1.9 s). This finding might be utilized for the automatic, sensor based, early detection of severe aggression. Intervention into pig's behaviour with PLF technology in accordance with the strategy developed [2] could be applied in the early phase of aggressive interaction – during or just after the initiating phase of aggressive sequence.

Table 1. Number of aggressive sequences

Type of sequences	Number of sequences	
	N	Percentage
Sequences that didn't lead to biting	100,0	63,7%
Sequences that led to biting	57,0	36,3%

Table 2. Duration of aggressive sequences

Type of sequences	Duration of aggressive sequences	
	Mean (s)	SEM (s)
Sequences that didn't lead to biting	18.1 ^{A}	4.3
Sequences that led to biting	38.8 ^{B}	5.7

(A, B) values differ for P<0.01.

Table 3. Duration of initiating phase of aggressive sequences

Type of sequences	Duration of initiating phase	
	Mean (s)	SEM (s)
Sequences that didn't lead to biting	1.9 ^{a}	0.4
Sequences that led to biting	3.3 ^{b}	0.6

(a, b) values differ for P=0.07

References

1. Berckmans D. (2008). Precision livestock farming (PLF) - preface. *Computers and Electronics in Agriculture* **62**(1), 1-1.

2. Berckmans, D., Vranken, E., Hartung, J., Fels, M., Guarino, M., inventors, (2012). Automated monitoring and controlling of undesired livestock behaviour. Patent GB1202577.1.

3. Bokma S. (1990). Housing and management of dry sows in groups in practice: partly slatted systems. Proceedings of an International Symposium on Electronic Identification in Pig Production. RASE, Stoneleigh, UK, 37–45.

4. D'Eath R.B. and Turner S.P. (2009). The natural behaviour of the pig. J.N. Marchant-Forde (Ed.), *The Welfare of Pigs*, Springer Science + Business Media, B.V., Dordrecht, p. 13

5. Durrell J. L., Sneddon I. A., Beattie, V. E. (1997). Effects of enrichment and floor type on behaviour of cubicle housed sows. *Anim Welfare* **6**, 297–308.

6. Fraser D. (1974). The behaviour of growing pigs during experirnental social encounters. *J. Agric. Sci*. **82**, 14-163.

7. Frost A.R., Schofield C.P., Beaulah S.A., Mottram T.T., Lines J.A., Wathes C.M. (1997) A review of livestock monitoring and the need for integrated systems. *Computers and Electronics in Agriculture* **17**, 139-159.

8. Gonyou H. W. (2001). The social behaviour of pigs. Page 147–176 in *Social Behaviour in Farm Animals*. L. J. Keeling and H. W. Gonyou, ed. CABI Publ. Wallingford, U.K.

9. Jensen M. B. and Pedersen L. J. (2010). Effects of feeding level and access to rooting material on behaviour of growing pigs in situations with reduced feeding space and delayed feeding. *Appl. Anim. Behav. Sci.* **123**, 1-6.

10. Jensen P., Yngvesson J., (1998). Aggression between unacquainted pigs-Sequential assessment and effects of familiarity and weight. *Appl. Anim. Behav. Sci.* **58**, 49-61.

11. Kirkwood R., Zanella, A. J. (2005). *Infl uence of gestation housing on sow welfare and productivity*. NPB Final Report.

12. Marchant J. N., Mendl M. T., Rudd A. R., Broom D. M. (1995) The effect of agonistic interactions on the heart rate of group-housed sows. *Appl. Anim. Behav.* Sci. **46**, 49-56.

13. Marchant-Forde J. N, (2010). Social behaviour in swine and its impact on welfare. *Proceedings of the 21st IPVS Congress*, Vancouver, Canada, pp. 18-21.

14. McGlone J. J. (1985) A quantitative ethogram of aggressive and submissive behaviours in recently regrouped pigs. *J. Anim. Sci.* **61**, 559-565.

15. Mendl M., Zanella A.J. and Broom D.M. (1992). Physiological and reproductive correlates of behavioural strategies in female domestic pigs. *Anim. Behav.* **44**, 1107-1121.

16. Meunier-Salaun M. C., Edwards S. A., Robert, S. (2001). Effect of dietary fibre on the behaviour and health of the restricted fed sow. *Anim. Feed. Sci. Tech.* **90**, 53–69.

17. Simmins P.H. (1993). Reproductive performance of sows entering stable and dynamic groups after mating. *Anim. Prod*. **57**, 293–298.

18. Stookey J. M., and H. W. Gonyou. (1994). The effects of regrouping on behavioural and production parameters in finishing swine. *J. Anim. Sci*. **72**, 2804–2811.

19. Weng R.C., Edwards S.A., English P.R. (1998). Behaviour, social interactions and lesion scores of group-housed sows in relation to floor space allowance. *Appl. Anim. Behav. Sci.* **59**(4), 307-316.

High-Throughput Phenotyping of Plant Resistance to Insects

Karen J. Kloth[1,2,3], Manus P.M. Thoen[1,2,3], Harro J. Bouwmeester[2],

Maarten A. Jongsma[3], Marcel Dicke[1]

[1]*Laboratory of Entomology, Wageningen University, Wageningen, The Netherlands. karen.kloth@wur.nl*

[2]*Laboratory of Plant Physiology, Wageningen University, Wageningen, The Netherlands*

[3]*Business Unit Bioscience, Plant Research International, Wageningen University and Research Center, Wageningen, The Netherlands*

Abstract

Devastating insect pests and environment-malignant pesticides that are applied against them are still a major problem in agriculture and horticulture. In plant breeding and genomic research there is an increasing demand for efficient screening methods of plants. In this study we develop a high-throughput phenotyping system to screen plants for resistance to aphids by video-tracking of insect behaviour.

Introduction

Studying the behaviour of herbivorous insects is essential for the improvement of agricultural pest management. Devastating insect pests still cause a significant amount of yield loss and a reduced marketability of food and ornamental crops [1]. Environment-malignant and costly pesticides are widely applied to control pests and diseases. There is, however, an increasing urge to improve the sustainability in the food production chain. Host-plant resistance is one of the cornerstones of successful and environmentally benign pest management systems [1,2]. Its concept is to reduce the intrinsic plant susceptibility to insect damage and thereby minimizing the need to apply pesticides. To improve plant resistance, suitable cultivars need to be selected among hundreds of plant lines. Screening plants for resistance to insects is generally a costly exercise in terms of space, time and labour. Bioassays often require a lot of greenhouse space including outbreak prevention measurements, and will take several weeks, depending on the insect life cycle. Eventually such a screening is mainly a manual procedure, such as counting of offspring, monitoring the development time to adulthood or measuring insect weight [3,4]. Plant breeding companies therefore need to invest a significant amount of money into these activities. Also, plant researchers face the need for efficient large-scale screenings now the costs for genotyping have rapidly declined and next-generation sequencing has rendered a wealth of genomic information [5]. For genomic research the screening of numerous natural accessions of crop-related plants is often a necessity. Devastating pests and diseases only rarely occur in nature, which is due to the tremendous degree of natural variation in plant defense mechanisms [6,7]. Only a relatively small degree of such variation is contained in cultivated crop populations [8], but wild populations provide ample opportunities for discovering novel mechanisms responsible for resistance to insects. In order to identify genes involved in plant resistance to insects, genomic studies are merely interested in the mechanisms of resistance rather than solely the identification of resistant plants. An effective screening method should therefore be able to discriminate distinct plant mechanisms that reduce insect colonization and feeding (e.g. plant volatile cues, deterrent surface structures, or toxins in the internal plant tissue) [9] (Figure 1).

Video-tracking of minute animals, such as 2 mm sized aphids, is a challenge. It requires a different approach in the hard- and software setup compared with video-tracking of larger sized animals, particularly when insects and plants have indistinctive colours. Although the miniscule size is demanding for automated subject detection, video-tracking of insect behaviour has a lot to offer in terms of accuracy and high-throughput; assets that are hard to achieve simultaneously during manual observations.

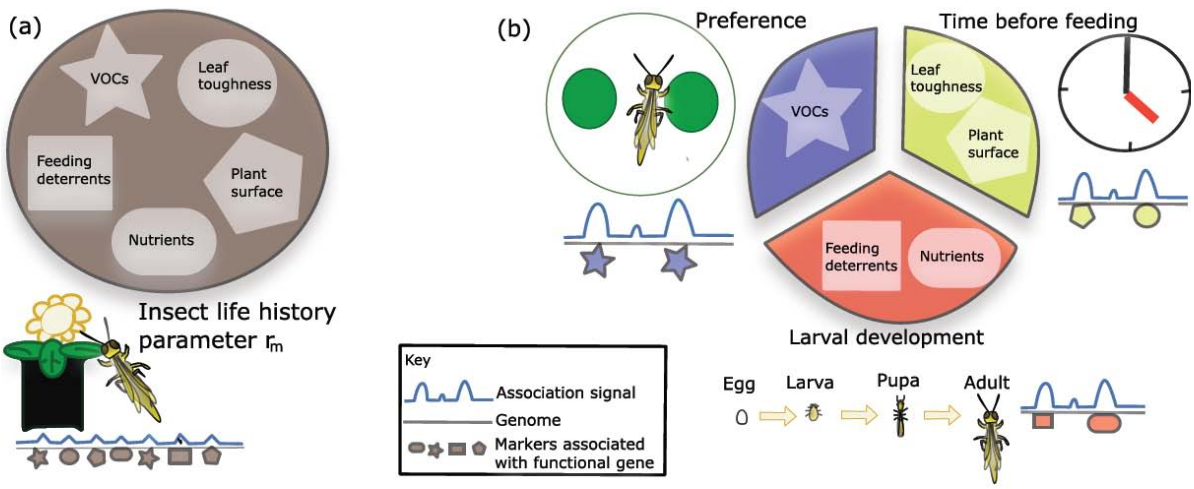


Figure 1. To find genes associated with resistance to insects, it is of particular importance to identify the mechanism(s) involved. This is illustrated in (a) where the life-history parameter rm (the intrinsic rate of population increase) of the insect shows which plant is resistant but seems to be associated with hundreds of genetic loci. A better approach to pinpoint the genes of importance, is to dissect the complex phenotype into component traits related to resistance mechanisms (b); e.g. insect preference (detection of repellent volatiles), time before the insect starts feeding (screening for the influence of leaf toughness and deterrent structures on the plant surface) and larval development (detection of, e.g. feeding deterrents, toxins and nutrient content) (figure adopted from [9]).

Aims

In this study we develop a high-throughput phenotyping system to screen for plant resistance to herbivorous insects. The system is based on video-tracking of insect behaviour and can be applied both for breeding purposes and genomic research with various insect and plant species. We study three components of insect behaviour: (1) host preference in a two-choice situation, (2) host acceptance (latency to the first feeding event and occurrence of dispersal events), and (3) feeding behaviour in general. We focus on the green peach aphid, *Myzus persicae*, a generalist phloem-feeding insect, that feeds on plants in more than 40 plant families and occurs virtually worldwide [10,11].

Methods and materials

The camera set up consists of arenas mounted on a backlight unit and two cameras with helicopter view for recording parallel sessions (Figure 2). The aphids are exposed to either a non-choice situation consisting of one leaf disc, or a two-choice situation consisting of two leaf discs of different plant types. We quantify two major behavioural components of aphid behaviour: (1) Probing behaviour (the activity when the aphid penetrates the plant tissue with its mouth parts in order to reach the phloem vessels) and (2) Non-feeding activities, such as the

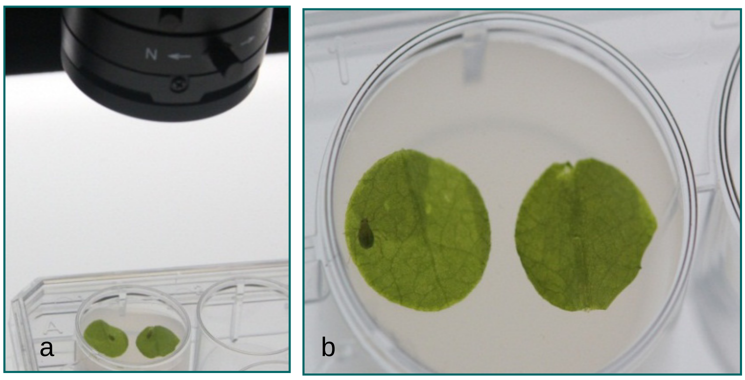


Figure 2. (a) Camera set-up. (b) Two-choice arena with a green peach aphid.

distance moved and the number of events when the aphid leaves the leaf disc. These parameters are registered over multiple time intervals, with multiple arenas observed simultaneously. We use the software EthoVision® XT for automated video tracking [12]. Data management and statistical analysis are performed with the program R [13]. Aphids exposed to resistant plants are expected to have a higher latency to the first feeding event and a higher frequency and duration of non-feeding activities compared with aphids exposed to susceptible plants.

With this video-tracking system we will screen 350 wild type plant accessions of mouse ear-cress, *Arabidopsis thaliana*. We will use the behavioural data of the aphids and the available genomic data of the Arabidopsis lines in a genome-wide association mapping study, in order to identify genes that are involved in resistance to aphids.

References

1. Schoonhoven, L.M., van Loon, J.J.A., Dicke, M. (2005). *Insect-plant biology* (second edition). Oxford University Press.
2. Panda, N. and Khush, G.S. (1995). *Host Plant Resistance to Insects* (CABI International).
3. Krips, O.E., Witul, A., Willems, P.E.L., Dicke, M. (1998). Intrinsic rate of population increase of the spider mite Tetranychus urticae on the ornamental crop gerbera: intraspecific variation in host plant and herbivore. *Entomologia Experimentalis et Applicata* **89**, 159-168.
4. Poelman, E.H. Galiart, R.J.F.H., Raaijmakers, C. E., van Loon, J. J. A., van Dam, N. M. (2008). Performance of specialist and generalist herbivores feeding on cabbage cultivars is not explained by glucosinolate profiles. *Entomologia Experimentalis et Applicata* **127**, 218-228.
5. Ingvarsson, P.K. and Street, N.R. (2011). Association genetics of complex traits in plants. *New Phytol.* **189**, 909-922.
6. Alonso-Blanco, C. and Koornneef, M. (2000). Naturally occurring variation in Arabidopsis: an underexploited resource for plant genetics. *Trends Plant Sci.* **5**, 22-29.
7. Anderson, J.T. and Mitchell-Olds, T. (2011). Ecological genetics and genomics of plant defences: evidence and approaches. *Funct. Ecol.* **25**, 312-324.
8. Gols, R., Bukovinszky, T., van Dam, N. M., Dicke, M., Bullock, J. M., Harvey, J. A. (2008). Performance of generalist and specialist herbivores and their endoparasitoids differs on cultivated and wild Brassica populations. *J. Chem. Ecol.* **34**, 132-143.
9. Kloth, K.J., Thoen, M.P.M., Bouwmeester, H.J., Jongsma, M.A., Dicke, M. (2012). Association mapping of plant resistance to insects. *Trends Plant Sci.* **17**(5), 311-319.
10. Blackman, R.L., Eastop, V.F. (2006). *Aphids on the World's Herbaceous Plants and Shrubs* (John Wiley & Sons, Ltd. & Natural History Museum, London, UK).
11. Van Emden, H.F., Harrington, R. (2007). *Aphids as crop pests* (CABI International).
12. Noldus, L.P.J.J., Spink, A.J., Tegelenbosch, R.A.J. (2002). Computerised video tracking, movement analysis and behaviour recognition in insects. *Comput. Electron. Agric.* **35**, 201-227.
13. Bates, D., Hornik, K. (2012). http://www.r-project.org/. Accessed 29 May 2012.

Temporal Patterns of Rodent Behavior in the Elevated Plus Maze Test

M. Casarrubea[1*], M.S. Magnusson[2], V. Roy[3], A. Arabo[3], F. Sorbera[1], A. Santangelo[1], G. Crescimanno[1]

*[1]Laboratory of Behavioral Physiology, Dept. of Experimental Biomedicine and Clinical Neurosciences, University of Palermo, Palermo, Italy. *maurizio.casarrubea@unipa.it*

[2]Human Behavior Laboratory, University of Iceland, Reykjavik, Iceland.

[3]Laboratoire de Psychologie et de Neurosciences de la Cognition et de l'Affectivité, Université de Rouen, Mont-Saint-Aignan, France.

Abstract

The aim of the present study was to evaluate, by means of a temporal pattern analysis, rat's behavior in the elevated plus maze test. A specific software called THEME has been used. On the basis of an ethogram encompassing 24 behavioral elements, results demonstrated that 14 components represented the 98% of the displayed activities. 145 different temporal patterns have been detected; length distribution of such patterns showed that three-, four-, and five-elements patterns were more represented than two-elements ones. Finally, a complex temporal pattern encompassing 8 behavioral elements has been discussed. Present article demonstrated the presence of complex temporal patterns characterizing the behavior of rats in the elevated plus maze test.

Keywords. t-pattern analysis, elevated plus maze, anxiety, rat

Introduction

Introduced by Handley and Mithani [7], the elevated plus maze (EPM) apparatus is a well-known and widely used model to assess anxiety-related behavior in rodents. The rationale underlying the utilization of the EPM is based on the assumption that naïve subjects, exposed to the apparatus, will respond to an approach/avoidance conflict between safe parts of the apparatus that are closed and protected, and aversive parts that are open, unprotected and more brightly lit [12]. Surprisingly, despite the large utilization of the apparatus and the large amount of important articles published, scanty data are available concerning the temporal organization of rat behavior in this experimental assay. Several questions still remain unsolved. For instance: do the behavioral events occurring on the EPM show a specific temporal arrangement? If so, would it be possible to identify specific and/or characteristic sequences of recurring behavioral activities? Last but not least, do such sequences, when/if manifest, encompass an underlying ethological meaning useful to better understand rodents' anxiety related behavior? The aim of this preliminary study was to shed light on above questions performing a detailed temporal analysis of rat behavior in the EPM by means of a specific software. This software, called THEME (PatternVision Ltd, Iceland; Noldus Information Technology, The Netherlands), utilizes a complex algorithm able to detect statistically significant relationships among the behavioral events in the course of time [10][11]. In recent years THEME has been used in different experimental approaches, for instance, to study route-tracing stereotypy in mice [1], behavior in neuro-psychiatric diseases [9], behavioral patterns associated with emesis [8] or, in our laboratories, to investigate exploration and anxiety-related behavior in rodents [4][5][6].

Method

Subjects and housing

Five 90 days old male Wistar rats, specific pathogen free, were used. Animals were born in the animal facility of the University of Rouen (France) and breeders originated from Janvier (Le Genest-St-Isle, France). Subjects were group-housed (3/4) in a room maintained at the constant temperature of 21 ± 2 °C, under a partially reversed light/dark cycle (light on: 12 noon – light off: 12 midnight). Food and water were freely available.

Experimental apparatus

The EPM (Intellibio, France) used in the present study was made of ivory Perspex, the arms were 50 cm long and 10 cm wide, and the apparatus was elevated at a height of 45 cm. The closed arms were surrounded by a 50 cm wall while open arms presented 0.5 cm edges in order to maximize open-arm entries [12]. The floor of the maze was covered with grey plastic. Environmental temperature was maintained equal to the temperature measured in the housing room. The testing room was illuminated with a dim white light that provided 100 lux for the open arms and 50 lux for the enclosed arms.

Procedure

Rats were transported from housing room to testing room inside their home-cages to minimize transfer effect. To avoid possible visual and/or olfactive influences, animals were allowed to acclimate for 30 minutes far from observational apparatus. Each subject, experimentally naïve, was placed in central platform of the EPM and allowed to freely explore for 5 min. After each observation, the EPM was cleaned with ethyl alcohol (10%) to remove scent cues left from the preceding subject. Experiments were recorded through a digital videocamera and video files stored in a personal computer for following analyses.

Ethogram and coding

The first step in a behavioral analysis is normally represented by the construction of a formal list, namely an ethogram, containing descriptions of behavioral elements. It is important to note that establishing an ethogram is always a quite critical aspect because an error (for instance, a behavioral element not described or, worst, misinterpreted) is able to negatively influence the following analysis. This is even more important if data will be studied by means of t-pattern analysis: an event can be quite uncommon, e.g. occurring only few times for each subject and/or not in all subjects, nonetheless the temporal relationships it establishes can be extremely important for the behavioral architecture. The ethogram used in the present study encompasses 24 behavioral elements and is presented in Table 1.

However, once video files have been collected and the ethogram is ready/available, the following step is normally represented by a coding process, that is, the utilization of specific software that allows the researcher to note the occurrences of all the behavioral elements carried out by the actor in the observed video file. The result of such a coding process is the so called *event log file*, that is, a sequence of behavioral events occurring at specific time points (namely, milliseconds, seconds or, even, video frames). In the present study all video files have been coded using The Observer (Noldus Information Technology, The Netherlands).

Data analysis

THEME software processes time stamped event log files searching for the particular but wide T-pattern class of hierarchical, sequential and significantly timed real-time patterns [10][11]. The search advances following a bottom-up process. In brief, being A, B, X three hypothetical events occurring in a given time window, the algorithm compares the distributions of each pair of the behavioral elements A and B searching for a time interval after occurrences of A such that, more occurrences of A are followed by at least one B within that window. In this case A and B are indicated as (A B) and form a t-pattern. After that, such first level t-patterns are marked and considered as potential A or B terms in higher patterns, e.g., ((A B) X), which are then also marked (added to the data) and may become an A or B part in still higher patterns, and so on up to any level of pattern complexity that may be found in the data. To avoid multiple detection of the same underlying patterns an evolution algorithm compares all detected patterns and selects only the most complete while deleting partial and redundant detections. Thus, more complex patterns may be created, step by step, following this bottom-up detection process. Before a t-pattern search is performed the software requires specific search parameters. In the present study t-pattern analysis was carried out by using the following search parameters: "*significance level*" (maximum accepted probability of any critical interval relationship to occur by chance) = 0.0001; "*lumping factor*" (forward and backward transition probability above which A and B of a t-pattern (A B) are lumped, that is, A and B are not considered separately but only as the (A B) pattern) = 0.90; "*minimum samples*" (minimum

Table 1. Ethogram of rat behavior in the elevated plus maze test. (*) = the behavioral element is considered protected (p-) when performed in the central platform or in a closed arm, unprotected (u-) when performed in an open arm; (**) = head dip is considered protected (p-) only in the central platform and unprotected (u-) in an open arm.

Behavioral element	Abbreviation	Description
Closed Arm Entry	CA-Ent	rat moves from the Central Platform to a Closed arm (all four paws in)
Open Arm Entry	OA-Ent	rat moves from the Central Platform to an Open Arm (all four paws in)
Closed Arm Return	CA-Ret	the rat puts only head and forepaws in the central platform, then rapidly re-enters in the closed arm
Closed Arm Walk	CA-Wa	rat walks in a Closed Arm
Open Arm Walk	OA-Wa	rat walks in an Open Arm
Central Platform Entry	CP-Ent	rat moves from an Open or Closed Arm to the Central Platform
Immobile Sniffing (*)	p-ISn; u-ISn	rat sniffs the surrounding area without walking activity
Corner Sniffing (*)	p-CSn; u-CSn	rat sniffs the entrance border of a Closed Arm
Stretched Attend Posture (*)	p-SAP; u-SAP	rat stretches its head and shoulders forward and then returns to the original position
Head Dip (**)	p-HDip; u-HDip	scanning over the sides of the maze towards the floor
Rearing (*)	p-Re; u-Re	rat maintains an erect posture
Defecation (*)	p-Def; u-Def	excrements are produced
Grooming (*)	p-Gr; u-Gr	rat licks/rubs its face and/or body
Paw Licking (*)	p-PL; u-PL	rat licks its paws
Immobility (*)	p-Imm; u-Imm	an immobile posture is maintained

percent of subjects in which a t-pattern must occur to be detected) = 100; "*minimum occurrences*" (minimum number of times a t-pattern must occur to be detected) = 5

Ethical statement

The experiment was conducted in accordance with the European Communities Council Directive 86/609/EEC concerning the protection of animals used for experimental scientific purposes.

Results

Coding process produced event log files containing a total of 1165 behavioral elements. Their percent distribution is presented in Figure 1. Concerning t-pattern analysis, 145 different t-patterns were detected. Figure 2 illustrates the distribution of such patterns on the basis of their length (that is, how many different t-patterns of two, three or *n* elements were detected). Figure 3 illustrates one of the most complex t-patterns detected, occurring twelve times and containing a sequence of eight behavioral events.

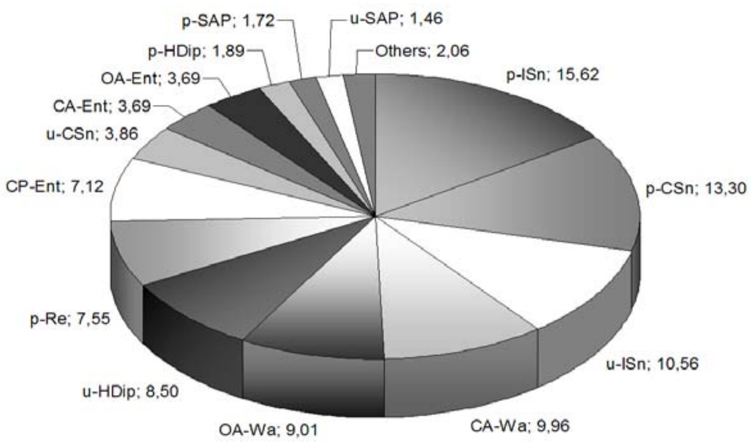


Figure 1. Per cent distribution of behavioral elements in the EPM. Others = p-Gr, u-Gr, CA-Ret, u-Def, p-Def, p-Imm, u-Imm, p-PL, u-PL, u-Re. Data obtained from the analysis of five subjects. See Table 1 for abbreviations.

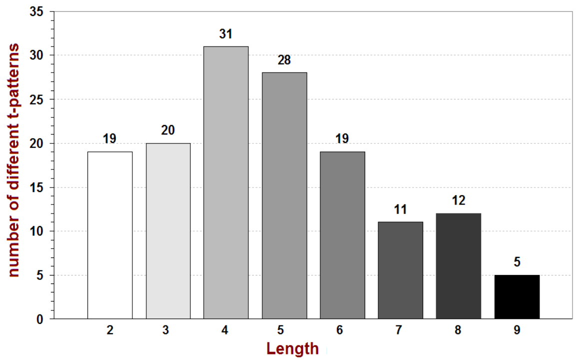


Figure 2. Number of different t-patterns on the basis of their length. X-axis= number of behavioral elements encompassed in the t-pattern's structure. Y-axis= number of t-patterns of different composition. Data obtained from the analysis of five subjects.

Discussion

Present research demonstrates the presence of temporal patterns in the behavior of Wistar rats observed in the elevated plus maze test. Pie chart shows that, among the 24 behavioral elements of the behavioral repertoire (see Table 1), 14 components represent the 98% of the behavior in the EPM (see Figure 1). On the contrary, groomings, immobilities and paw lickings are much more seldomly observed. Such behavioral elements share a common aspect: they require, to be performed, an immobile position. Hence it is possible to conclude that naïve rats in the EPM show a behavioral repertoire heavily oriented toward locomotion and exploration. Such a result appears to be quite noticeable if a comparison with different assays to study anxiety, such as the open-field and the hole-board, is carried out. Indeed both in open field and in hole-board groomings, immobilities and paw lickings are, by far, much more represented [2][3][4]. We hypothesize that such an important difference could be the result of the heavier impact of the EPM in terms of approach-avoidance conflict. T-patterns length distribution (see Figure 2) clearly shows that rodent's behavior in the EPM is well structured from a temporal point of view. Taking into consideration the peculiar bottom-up process carried out by THEME's search algorithm, it is quite common the detection of a large basis of two-elements t-patterns and a lower number of higher-order and more complex patterns [4][5][6] as the search proceeds. The condition in the EPM is very different since three-, four-, and five-elements t-patterns are more represented than two-elements ones. Such an evidence suggests that rat's behavior in EPM has more complex and structured temporal characteristics in comparison with open field and/or hole-board [4][5][6].

The t-pattern in Figure 3 well depicts rodent behavior in the EPM: the animal starts exploring by entering in an open arm (OA-Ent) and explores the unprotected zone (u-ISn, OA-WA); after that, the sniffing of central

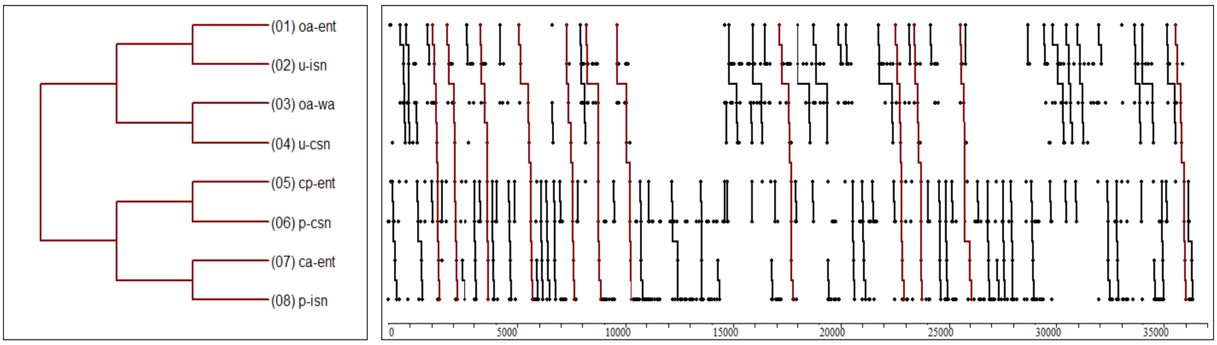


Figure 3. T-pattern of 8 behavioral events occurring 12 times. T-pattern tree structure is illustrated in the left box; numbers in brackets indicate the order of appearance of behavioral events. Onset and occurrences of the detected t-pattern along with incomplete sub-patterns are indicated in the right box. Data obtained from the analysis of five subjects. See Table 1 for abbreviations.

platform’s corners (u-CSn) is followed by the entrance in the central structure of the maze (CP-Ent); the t-pattern ends with the sniffing of a closed arm’s border (p-CSn), followed by the entry in the closed arm (CA-Ent) and, finally, a further sniffing activity (p-ISn). Interestingly, such a t-pattern (see Figure 3), as several others detected in the same subjects, is structurally organized on the basis of sub-patterns related with the three main EPM sections, that is, open arm, central platform and closed arm. Further studies are in progress in our laboratories to deepen and better clarify the temporal features of rodent behavior in the EPM.

References

1. Bonasera, S.J., Schenk, A.K., Luxenberg, E.J., Tecott, L.H. (2008) A novel method for automatic quantification of psychostimulant-evoked route-tracing stereotypy: application to Mus musculus. *Psychopharmacology (Berl)* **196**, 591-602

2. Casarrubea, M., Sorbera, F., & Crescimanno, G. (2009a). Structure of rat behavior in hole-board: I. Multivariate analysis of response to anxiety. *Physiology & Behavior* **96**, 174-179.

3. Casarrubea, M., Sorbera, F., & Crescimanno, G. (2009b). Structure of rat behavior in hole-board: II. Multivariate analysis of modifica-tions induced by diazepam. *Physiology & Behavior* **96**, 683-692.

4. Casarrubea, M., Sorbera, F., Crescimanno G. (2009c). Multivariate data handling in the study of rat behavior: an integrated approach. *Behavior Research Methods* **41**, 772-781.

5. Casarrubea, M., Sorbera, F., Magnusson, M., Crescimanno, G. (2010). Temporal patterns analysis of rat behavior in hole-board. *Behavioural Brain Research* **208**, 124-131.

6. Casarrubea, M., Sorbera, F., Magnusson, M.S., Crescimanno, G. (2011). T-pattern analysis of diazepam-induced modifications on the temporal organization of rat behavioral response to anxiety in hole-board. *Psychopharmacology (Berl.)* **215**, 177-189.

7. Handley, S.L., Mithani, S. (1984). Effects of alpha-adrenoreceptor agonists and antagonists in a maze exploration model of “fear”-motivated behavior. *Naunyn-Schmiedeberg’s Archives of Pharmacology* **327**, 1-5

8. Horn, C., Henry, S., Meyers, K., Magnusson, M.S. Behavioral patterns associated with chemotherapy-induced emesis: a potential signature for nausea in musk shrews. (2011). *Frontiers in Neuroscience* **5**, 1-17.

9. Kemp, A.S., Fillmore, P.T., Lenjavi, M.R., Lyon, M., Chicz-Demet, A., Touchette, P.E., et al. (2008) Temporal patterns of self-injurious behavior correlate with stress hormone levels in the developmentally disabled. *Psychiatry Research* **157**, 181-189.

10. Magnusson, M.S. (2000). Discovering hidden time patterns in behavior: t-patterns and their detection. *Behavior Research Methods Instruments and Computers* **32**, 93-110.

11. Magnusson, M.S. (2004). Repeated patterns in behavior and other biological phenomena. In: Oller DK, Griebel U, editors. *Evolution of Communication Systems: A Comparative Approach. Cambridge: The MIT Press*, 111-128

12. Roy, V., Chapillon, P., Jeljeli, M., Caston, J., Belzung, C. (2009). Free versus forced exposure to an elevated plus-maze: evidence for new behavioral interpretations during test and retest. *Psychopharmacology* **203**, 131-141.

A Computer-Based Application for Rapid Unbiased Classification of Swim Paths in the Morris Water Maze

Kurt R. Stover, Timothy P. O'Leary, Richard E. Brown

Department of Psychology and Neuroscience, Dalhousie University, Halifax, Canada. kurt.stover@dal.ca

Introduction

The Morris water maze (MWM) is a commonly used apparatus for measuring visuo-spatial reference or working memory in rodents. In the reference memory task in our lab, mice are trained to locate a submerged platform in a circular pool filled with opaque water by using distal extra-maze visual cues. Learning in the MWM is typically assessed with measures of performance including latency and distance travelled to reach the platform, as well as swim speed and thigmotaxis (the amount of time spent swimming near the walls of the pool). To assess memory, probe trials are used in which the platform are removed and the amount of time mice spend in the quadrant of the pool where the platform was located and the frequency that they swim across the location of the platform is measured. Visual ability and motivation to escape are assessed in training trials with a visible escape platform.

Although latency and swim distance are the most commonly used measures of learning in the MWM it is not possible to determine whether or not mice are using visuo-spatial cues based on these measures of performance [1]. Furthermore, latency and distance may not always have sufficient sensitivity to detect impairments in learning, because mice that use non-spatial strategies may not differ in performance from those that use visual-spatial strategies. As mice learn the location of the escape platform in the MWM they typically display a progression of search strategies with an increasing use of visuo-spatial cues [2,3]. On the first day of training, mice often begin with a random search strategy and then, as training continues, progress to non-spatial procedural strategies, such as swimming in a circle around the wall a fixed distance from the edge of the pool, then to strategies using extra-maze cues, such as scanning around the general area of the platform. As they learn to use the extra-maze cues, mice begin to swim from the release point directly to the platform. Mice may also swim in small circles, looping around the maze, engage in thigmotaxis, or float, which may be a result of non-cognitive factors including impaired motor abilities, stress, or hypothermia [4,5].

Latency and distance measures are not able to discriminate between search strategies in the MWM [6]. Therefore, using traditional measures of MWM performance without analyzing search strategy use may result in incorrect conclusions drawn about visuo-spatial ability. Although search path analysis is important for the assessment of learning in the MWM, it is rarely carried out. This may be due to the increased time and labor required for these analyses, a perceived lack of objectivity of search path analysis, or the potential for experimenter bias in search path analysis. While these limitations can be avoided using data-driven classification with relatively sophisticated multivariate statistical techniques, such as discriminate function analysis, this requires a large number of behavioural measures to be recorded [6], has not yet been validated across different laboratories, and may not be suitable for high-throughput behavioural experiments.

In order to efficiently categorize MWM swim paths we developed a simple computer program called SwimPath to display the image of a swim path and allow the user to indicate which search strategy was used. SwimPath can be used in conjunction with commercially available tracking systems. The SwimPath program ensures that the scorer is blind to the identity of the mouse and trial number, and allows responses to be recorded electronically and exported into a database. To assess the usefulness of this program in detecting strain differences, we analyzed the swim paths of male and female mice of five inbred strains which have different levels of visual acuity [7]. We also recorded latency and distance to the platform, swim speed, and thigmotaxis, and correlated these traditional measures of performance with search strategy use in the MWM. To determine the relationship between search strategy and memory we correlated search strategy use on day six of training with performance in the probe trial.

Methods

Mice

Seventy-four mice, with approximately equal numbers of males and females from each of five strains were used: A/J (5F, 4M) and BALB/cByJ (5F, 6M), which are albino; C3H/HeJ (4F, 7M) which have retinal degeneration (RD) [8]; BTBR T+ tf/tf (11F, 12M), which have normal vision but are used as a model of autism; and C57BL/6J (10F, 10M), which have normal vision. The mice ranged from 2 – 5 months of age. Mice were housed in same sex groups of two to four mice per age and had free access to food and water. The BTBR T+ tf/tf mice were provided by Dr. Valerie Bolivar (Wadsworth Center, New York State Department of Health), and all other mice were purchased from the Jackson laboratory (Bar Harbor, ME). The protocol for this experiment was approved by the Dalhousie University Committee on Animal Care.

Morris water maze procedure

Mice were tested in the MWM reference memory task using the procedure described by Wong and Brown [9]. The test lasted 8 days, which included 3 days of acquisition training (4 trials/day), 3 days of reversal training (4 trials/day) (where the platform is moved to the opposite side of the pool), a probe trial with no platform present, and a cued test with a visible flag added to the platform (4 trials). The latency and swim distance to the platform, swim speed, and thigmotaxis were calculated by the tracking program (Watermaze, Actimetrics Inc., Wilmette, IL) which also recorded the swim paths, the time spent in each quadrant and the number of annulus crossings in the probe trial. Strain differences for each behaviour were analyzed using a mixed design ANOVA for acquisition and reversal, and a between subjects ANOVA for the probe trial. Mice were pooled over sex for these analyses.

Swim path analysis

The swim paths of the mice were classified into nine different types, which were then combined into four strategies: spatial accurate, spatial inaccurate, procedural, and non-cognitive. Spatial accurate strategies consisted of (1) spatial direct (Figure 1A), a straight path leading directly to the escape platform, with no turns greater than 180º or (2) spatial indirect (Figure 1B), a relatively direct path to the location of the escape platform but with some deviation away from the most optimal path. Spatial inaccurate strategies were (1) focal-correct (Figure 1C), swimming towards the quadrant with the escape platform followed by a focused search confined to the area immediately surrounding the platform, and (2) focal-incorrect (Figure 1D), swimming directly to a quadrant of the pool without the escape platform then searching in this area with circling behaviour. Procedural strategies included (1) scanning (Figure 1E), when mice swam selectively in the center area of the pool, (2) looping (Figure 1F), a circular path at a similar distance from the wall that the escape platform was positioned, passing through at least two adjacent quadrants before reaching the escape platform,. The non-cognitive strategies included (1) thigmotaxis (Figure 1G), a path within a 5 cm zone next to the wall of the maze which may include sporadic movement away from the walls into the center of the maze, and (2) random (Figure 1H), when the mouse did not selectively swim in any area of the pool and contained straight movement or wide circular arcs through the center of the pool and (3) circling (Figure 1I), a tight circular path, or a short path with several tight circles. When a path contained features of more than one type they were categorized based on which type occupied the majority of the path. The SwimPath program (coded in Python 3.2, and can be run on any operating system capable of running Python) presented the images to the user in a random order and rotated the image randomly 0, 90, 180, or 270º. Each search strategy was assigned to a key, which was used to classify the images into one of the nine strategies.

The proportion of trials in which mice used each of the four search strategies was calculated for all training days (not the probe trial) and analyzed for strain differences using a Kruskal-Wallis one way analysis of variance by ranks. Post-hoc analyses were conducted using the Mann-Whitney U test. Search strategy use on the last day of training was correlated with measures of performance in the MWM on the last day of training and probe trial performance using the Pearson product-moment correlation coefficient.

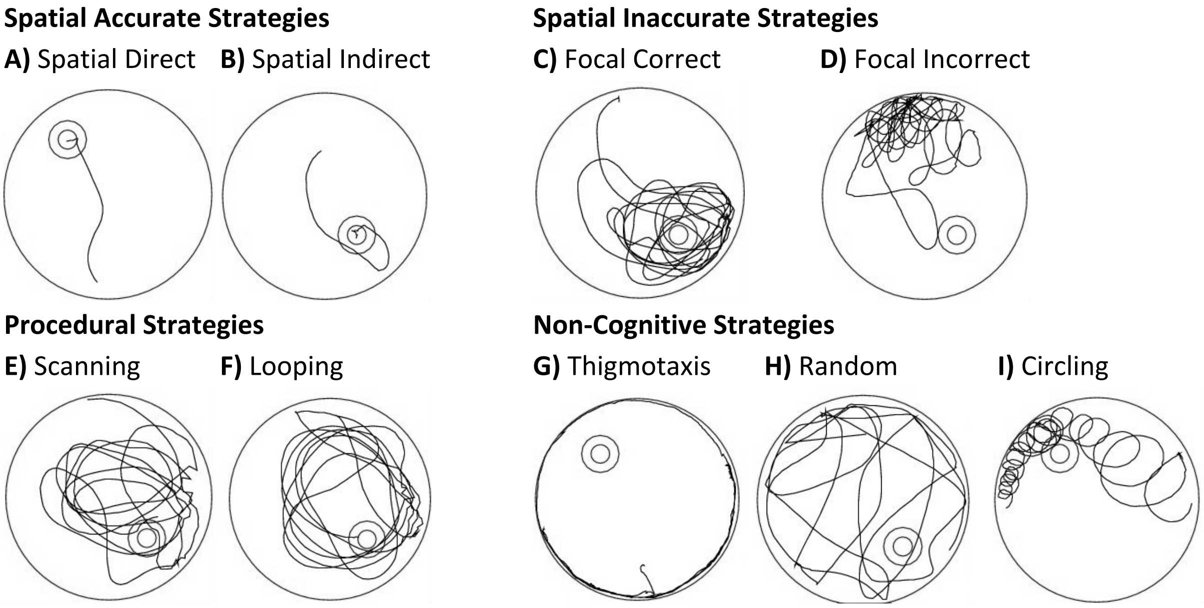


Figure 1. Example swim paths of the three nine types of search strategies which were classified into three categories: Spatial Accurate: A) Spatial Direct and B) Spatial Indirect, Spatial Inaccurate: C) Focal Correct and D) Focal Incorrect, Procedural: E) Scanning and F) Looping, and Non-Cognitive: G)Thigmotaxis, H) Random, and I) Circling.

Results

Tracking system measurements

Both pigmented and albino mice performed better than RD mice during acquisition, reversal, probe, and the visible platform trials. Within the albino mice, BALB/cByJ often performed better than A/J mice (Figure 2 A&B). In the probe trial memory test, C57 mice performed the best followed by BTBR, BALB/cByJ, C3H/HeJ, and A/J mice (Figure 2 C&D).

Swim path analysis

In general all mice began with non-cognitive strategies then shifted to procedural or spatial accurate strategies. Spatial inaccurate strategies were rarely used (Figure 3). There were significant strain differences in search strategy use on all days except spatial accurate and spatial inaccurate on day 1. By day six of training and on the visible platform trials, the C57, BTBR, and BALB/cByJ mice used the spatial accurate search strategies more than the A/J and C3H/HeJ mice (Figure 3A). BALB/cByJ mice used the spatial inaccurate strategies more than other mice during training and during the visible platform trials (Figure 3B). BTBR mice used the procedural strategies more than all other strains on all but day six of training and there were no differences between strains in the visible platform trials (Figure 3C). On day one all strains used the non-cognitive search strategies. The C57, BALB/cByJ, and BTBR mice quickly reduced their use of non-cognitive strategies over days, but A/J and C3H/HeJ did not (Figure 3D). In the visible platform trials the A/J and C3H/HeJ continued to use non-cognitive strategies.

Correlations between swim paths and tracking system measures of performance

On day 6 of training, use of the spatial accurate strategies was negatively correlated with latency ($r=-0.884$, $p<0.001$, $n=74$) and distance ($r=-0.858$, $p<0.001$, $n=74$) to the platform, and thigmotaxis ($r=-0.685$, $p<0.001$, $n=74$), and positively correlated with two measures of probe trial performance: percentage of time spent in the correct quadrant ($r=0.623$, $p<0.001$, $n=74$), and number of correct annulus crossings ($r=0.520$, $p<0.001$, $n=74$). Use of the spatial inaccurate strategies was not correlated with any measures of MWM performance. Use of the procedural strategies was positively correlated with distance to the platform ($r=0.312$, $p<0.01$, $n=74$) and swim speed ($r=0.328$, $p<0.01$, $n=74$). Use of the non-cognitive strategies was positively correlated with latency

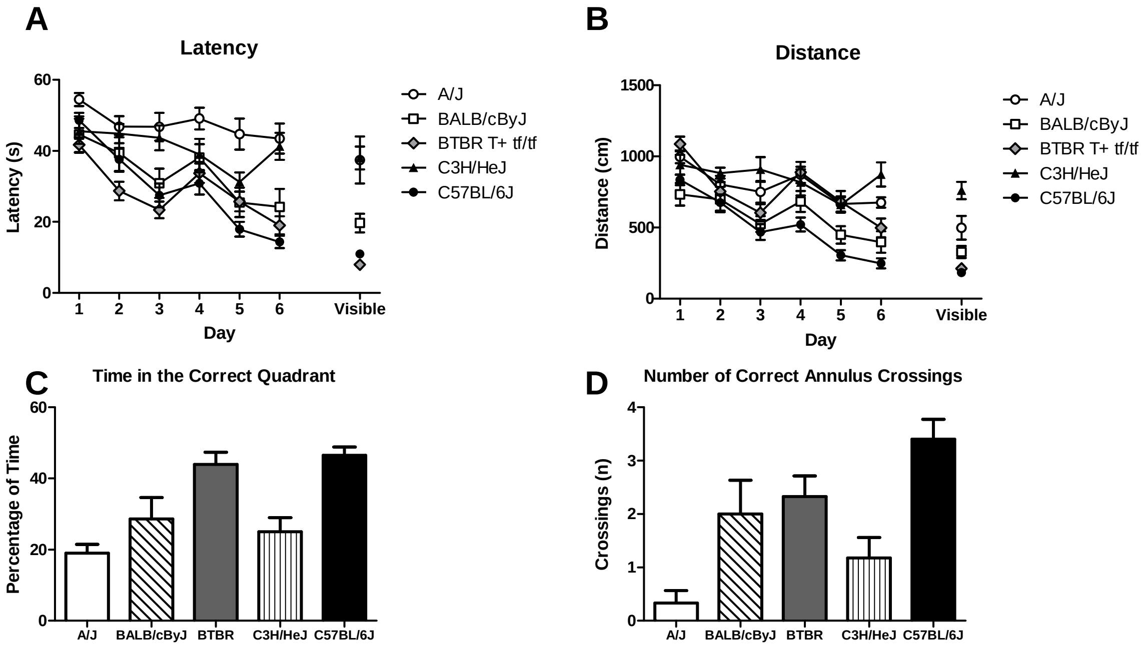


Figure 2. The (A) latency and (B) swim distance to the hidden platform over six acquisition days and the visible platform trials in the MWM. (C) The percentage of time in the correct quadrant and the (D) number of correct annulus crossings during the probe trial in the MWM. A/J and BALB/cByJ are albino, C3H/HeJ have retinal degeneration, BTBR T+ tf/tf have normal vision but are used as a model of autism, and C57BL/6J have normal vision. Error bars represent the standard error of the mean

($r=0.798$ $p<0.001$, $n=74$) and distance ($r=0.660$, $p<0.001$, $n=74$) to the platform, thigmotaxis ($r=0.862$, $p<0.001$, $n=74$), and negatively correlated with two measures of probe trial performance: percentage of time spent in the correct quadrant ($r=-0.519$, $p<0.001$, $n=74$), and number of correct annulus crossings ($r=-0.404$, $p<0.001$, $n=74$).

Discussion

These results demonstrate that search path classification using our SwimPath program is able to detect differences in learning in the MWM across strains with differing levels of visual ability [10]. Use of the spatial accurate search strategy by C57, BALB/cByJ, and BTBR mice, was correlated with shorter latencies and distance to the platform and better visuo-spatial memory scores in the probe trial. Use of the spatial inaccurate strategy was not correlated with any measures of performance in the MWM, but it was able to distinguish between the BALB/cByJ mice and all other strains, indicating that the search path analysis is measuring an aspect of MWM performance which is not captured by the performance measures typically used in the MWM. The higher use of the procedural strategies in the BTBR mice differentiates them from the BALB/cByJ mice, which have very similar levels of spatial accurate and non-cognitive search strategies. Mice using spatial accurate strategies had lower latencies and shorter distances to reach the escape platform during training and spent more time in the correct quadrant and had more platform crossings in the memory test than mice using non-cognitive strategies. These results demonstrate that increased use of the spatial accurate strategies is correlated with better performance and use of non-cognitive strategies with worse performance in the MWM. Overall we have demonstrated that this method of scoring search strategies is an effective measure of MWM performance which is able to detect differences between strains that traditional measures were unable to detect. The use of search strategy analysis is the only way to determine whether or not mice are actually using a visuo-spatial search strategy to learn in the MWM and this has implications for understanding strain differences in the neural basis of learning and memory in the MWM.

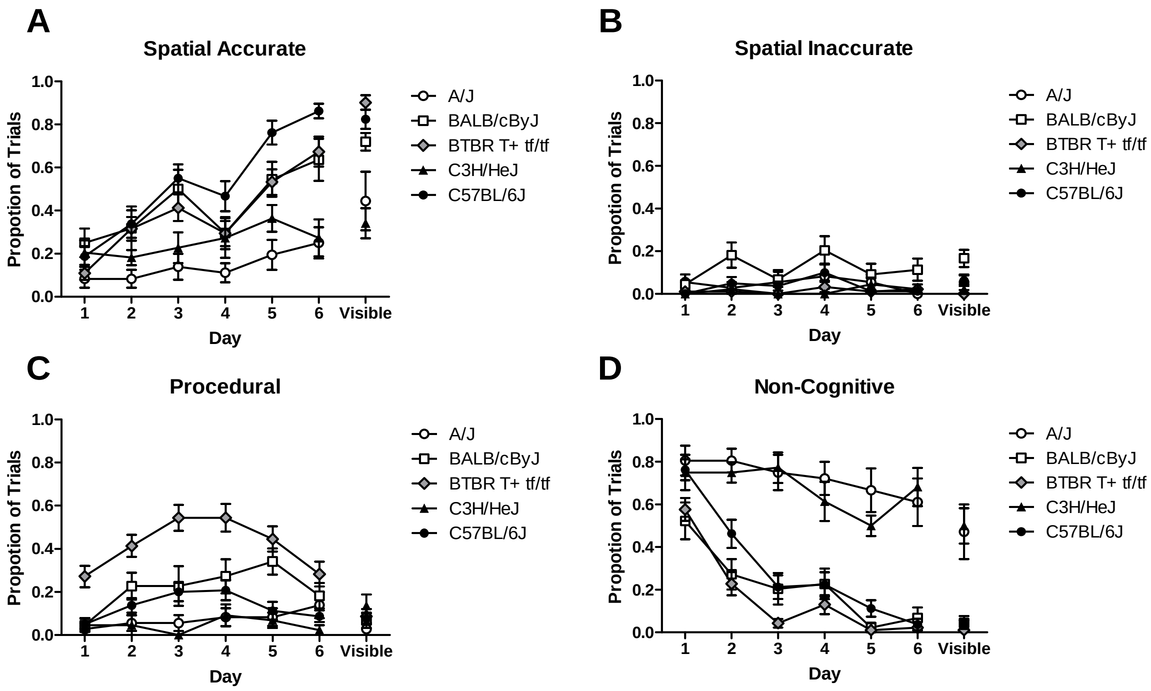


Figure 3. The proportion of usage of (A) spatial accurate, (B) spatial inaccurate, (C) procedural, and (D) non-cognitive search strategies in five strains of mice over six acquisition days and the visible platform trials in the Morris water maze. A/J and BALB/cByJ are albino, C3H/HeJ have retinal degeneration, BTBR T+ tf/tf have normal vision but are used as a model of autism, and C57BL/6J have normal vision. Error bars represent the standard error of the mean.

References

1. Cahill L., McGaugh J. L., and Weinberger N. M. (2001) The neurobiology of learning and memory: some reminders to remember. *TRENDS in Neurosciences* **24**, 578-581.
2. Janus, C. (2004). Search strategies used by APP transgenic mice during navigation in the Morris water maze. *Learning & Memory* **11**, 337-346.
3. Baldi, E., Lorenzini, C. A. and Burcherelli, C. (2003). Task solving by procedural strategies in the Morris water maze. *Physiology and Behaviour* **78**, 785-793.
4. Wolfer, D. P., Stagljar-Bozicevic, M., Errington, M. L. and Lipp, H-P. (1998). Spatial memory and learning in transgenic mice: Fact or artifact? *News in Physiological Science* **13**, 118-123.
5. Iivonen, H., Nurminen, L., Harri, M., Tanila, H. & Puoliväli, J. (2003). Hypothermia in mice tested in Morris water maze. *Behavioural Brain Research* **141**, 207-213.
6. Graziano, A., Petrosini, L. and Bartoletti, A. (2003). Automatic recognition of explorative strategies in the Morris water maze. *Journal of Neuroscience Methods* **130**, 33-44.
7. Wong, A.A., Brown, R.E. (2006). Visual detection, pattern discrimination, and visual acuity in 14 strains of mice. *Genes Brain and Behaviour* **5**, 389–403.
8. Sidman, R. L., Green, M. C. (1965). Retinal degeneration in the mouse: Location of the RD locus in linkage group XXVII. *Journal of Heredity* **56**, 23-29.
9. Wong, A.A., Brown, R.E. (2007). Age-related changes in visual acuity, learning and memory in C57BL/6J and DBA/2J mice. *Neurobiology of Aging* **28**, 1577-1593.
10. Brown R.E., Wong A.A. (2007). The influence of visual ability on learning and memory performance in 13 strains of mice. *Learning & Memory* **14**, 134-144.

Pressure-Sensor System for Sleeping-Posture Classification

Sabri Boughorbel, Fons Bruekers, Koen de Groot

Philips Research Laboratories, Eindhoven, The Netherlands.
{sabri.boughorbel, fons.bruekers, koen.de.groot}@philips.com

Abstract

In this paper we present our research work on the classification of sleeping postures using pressure sensing. A densely pressure-sensor grid placed under the matrass is first used for the posture classification task that will serve next as a reference. Pressure data obtained from 21 subjects for 4 sleeping postures (supine, prone, left and right side) has been analysed to determine the optimal location and the number of pressure sensors for an efficient sleeping-posture classification system. Obtained settings are used to build the optimal system and a second experiment including 11 subjects has been carried out with the new system. The obtained results are very promising and show the potential and the interest of the proposed method compared to existing systems.

Introduction

Bedridden patients are confined to bed because of illness or infirmity. This category of people requires additional care to avoid the formation of pressure ulcers due to the long sleep on the same posture. The pressure applied to the skin causes a deprivation of blood on the pressed skin area. If the pressure remains too high on the same spot for a couple of hours, skin tissue gets damaged and even may die. Since the discomfort of these pressure ulcers for the patients is enormous and the costs for treatment are high, the prevention of pressure ulcers is important. Prevention starts with using special materials for mattresses and sheets. In addition the patient is turned every 2 hours which requires the presence of nurses. In order to reduce nursing costs fewer nurses need to take cares of more patients at the same time by noting the time of turning. In the last years a number of lawsuits have happen against mistakes in the treatment of bedridden patient. Therefore the automatic monitoring of the sleeping postures can be very helpful to avoid mistakes by providing, e.g., alarms in case the posture did not change for a long period. It allows also logging of sleeping data and tracing other sleep related issues.

The topic of posture classification based on pressure data has been mainly researched for applications related to sitting [4, 5, 6] and sleeping. In [1] 18 sitting postures have been classified using PCA. In [2] **Error! Reference source not found.**8 sleeping postures have been classified from pressure data based on kurtosis and skewness estimation. The alternative approach is to use a camera for posture classification [3]. Many drawbacks can be expected using computer vision approaches such as changes in viewing angle, lighting condition, clothing and covers.

Posture Classification System Design

The goal of this research is to develop an unobtrusive system that measures sleeping postures based on pressure sensors. Therefore we have used pressure sensor mats placed under the mattress on top of an additional thin mattress. This choice avoids the direct contact of the subject with the pressure sensors and therefore prevents any possible safety issue related to the presence of the sensors close to the subject body. In our current research we focus on the optimization of the number, location and size of pressure sensors in order to have an efficient system in term of posture classification accuracy and costs. First a dense grid of pressure sensors (42x192 sensors) is used to determine the optimal location and size of the pressure sensors. Then the obtained settings are used to build the optimal system. The system works as follows: first a feature vector is extracted from each pressure sensor frame. A classifier is trained using features labelled with the sleeping posture. A new feature vector is then fed to classifier to obtain the estimated sleeping posture. Different types of classifiers such as linear, quadratic, nearest neighbourhood and SVM have been tested with a collected dataset of 21 subjects. Figure 1 depicts an example of the 4 most frequent sleeping-postures.

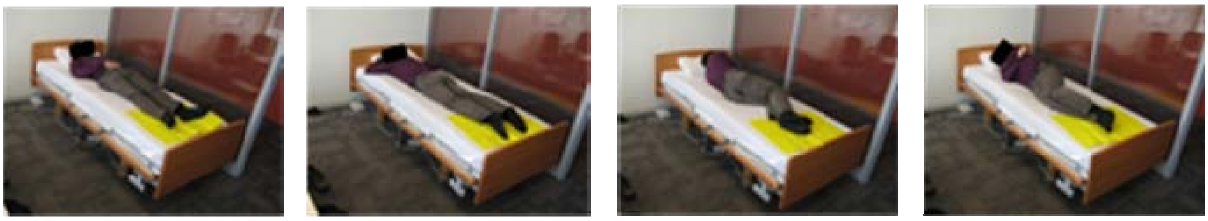


Figure 1. A subject during the recording of the 4 analysed postures: supine, prone, left and right side.

Experiments and Results

In the experiments four TekScan© pressure-sensor mats have been placed between a thin mattress, which is placed on the bed slats, and the mattress. A data frame that is recorded by the pressure system is a matrix of pressure values ranging from 0 to 255. The sampling rate was set to 1 frame every 3 seconds. In the first experiment the sleeping-posture data has been collected from 21 different subjects. Each subject was asked to lay-down for about 3 minutes on each of the four chosen posture, i.e., supine, prone, left-side and right-side and behave naturally for the chosen posture. For each subject and for each posture 240 frames are recorded to capture movement variability. Additionally a full night recordings for about 10 nights from 2 subjects were collected. It turned out that collected data is representative enough and similar to the full night recording in term of pressure values and patterns.

The feature extraction is as follows: firstly a coarse grid of bins is defined and secondly the pressure values are summed in each bin. Finally the resulting feature vector is normalized by the total pressure to account for weight differences between subjects. Different grids of bins have been tested and it turned out that 3x12 and 4x10 bins are good choices, see Figure 2. To deal with the differences of subject positions in the bed we compare the centre point of gravities and apply linear interpolation techniques.

Different classifiers have been tested [7] *for the* task from PRtools matlab toolbox. Table 1 shows the obtained error classification results for linear, quadratic, k-NN and SVM classifiers. Two training settings have been used. The first setting is subject-split where subjects included in the training are not included in testing. The second setting is measurement-split where postures of the same person can be in both training and testing. e.g., left posture used for training and supine posture of the same subject is used for testing. This difference in splitting did not show significant difference in terms of classification accuracy and showed that the classification does not require personalization to achieve high accuracy. This result shows that the system can be pre-trained in factory and therefore makes the deployment of the system easier and cheaper.

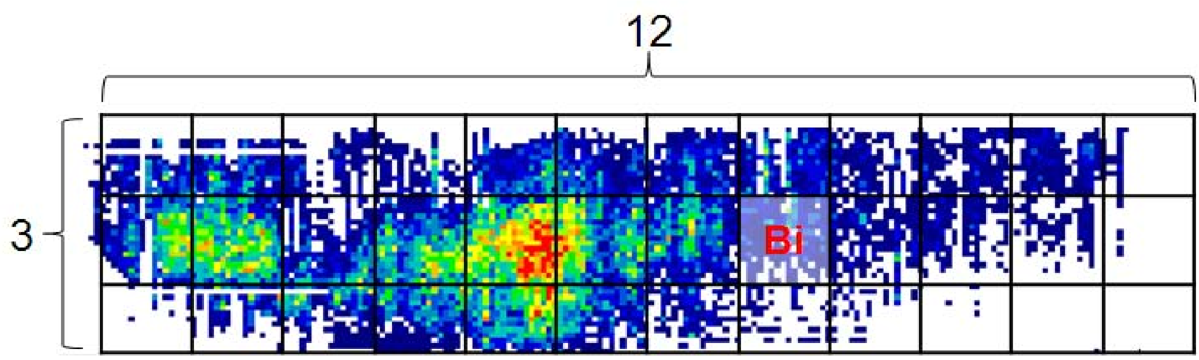


Figure 2. Pressure frame of left-side posture with the head on the left and the legs on the right. A grid of 3x12 is applied by summing the pressure in each cell.

Table 1. Mean and standard deviation of the sleep-posture classification error for two training conditions.

Classifier	Subject split	Measurement split
Linear	10.69%±6.8%	9.24%±3.02%
Quadratic	10.9%±8.1%	11.03%±2.66%
1-NN	6.66%±5.42%	5.31%±1.86%
SVM	2.91%±2.91%	2.90%±0.97%

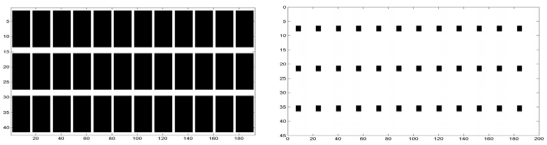


Figure 3. Pressure sensor size used in the first experiment (left) and second experiment (right).

In order to further reduce system requirements in terms of the number of pressure sensors, the size of the bins as in Figure 2 is reduced as illustrated in Figure 3. In this case the obtained feature vector of size 36 (3x12) is based on less pressure points since the overlaying mattress serves as a spatial low-pass filter, thus this bin-size reduction is allowed. To further optimize the sensor layout, different distances between sensors and different distances from the mattress border are tested.

The first experiment was based on the TekScan© pressure-sensor mats that are much narrower than the mattress. As a consequence a person may be positioned off the mat. To extent the width of the sensor area we have increased the number of rows from 3 to 4 and to end up with about the same number of bins we have decreased the number of columns from 12 to 10. As a consequence the length of the feature vector has increased from 36 to 40.

For each setting the feature vectors for all subjects are extracted and the train and testing cross-validation procedure is applied. Figure 4 shows the results. The left column shows the initial sensor configuration and middle column shows the final sensor configuration. The rows illustrate the distance offsets from the left border of the mattress. The right column shows the classification error as a function of distance between sensors. The optimal settings for a 4x10 sensor grid is a distance of 6 cm between sensors and a left offset of 20 cm from the left side (head side). This optimal settings lead to a classification error of 2.22% ± 2.7% (third row). The related confusion matrix is depicted in Table 2. Although the sensor size is reduced compared to the first experiment, position optimization allowed a good system performance.

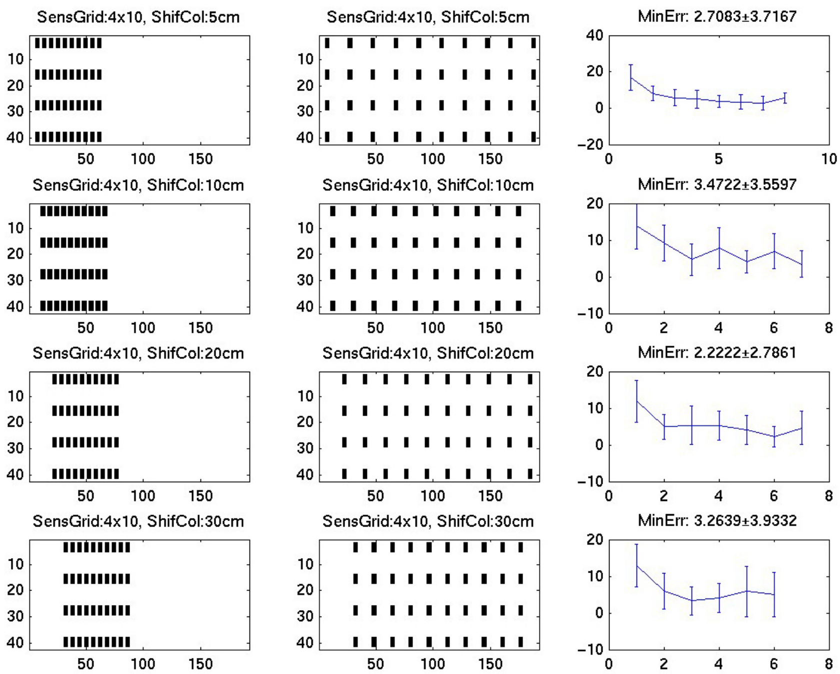


Figure 4. Sensor grid optimization by changing the distance between sensors.

Table 2. Confusion matrix obtained using the optimized hardware system (sensor grid of 4x10).

	Supine	**Prone**	**Left**	**Right**
Supine	97.22%	2.22%	0%	0.55%
Prone	0.55%	95.55%	3.88%	0%
Left	0%	0.55%	98.33%	1.11%
Right	0%	0%	0%	100%

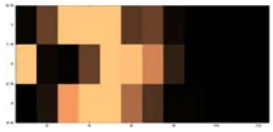


Figure 5. Optimal sensor location map. High intensity indicates that the sensor is important for posture classification.

It is not required to have the sensors lying on a rectangular grid and therefore it could be possible to further reduce the number of sensors needed. A forward feature selection algorithm is thus applied to identify the most important sensor location. Figure 5 shows the relevance of sensors that are most import for the sleep-posture classification. It can be seen clearly that the right part of sensor area (legs and feet) are much less important for posture classification.

Conclusions

In this paper we investigate the problem of sleep posture classification using pressure sensors placed under the mattress. The main application is the monitoring of bedridden subjects for pressure-ulcer prevention. The design steps of the system are described and the choice of sensor configuration is presented. The proposed system showed a promising performance (about 2% of classification error). As future work we plan to include temporal aspects to further improve the classification by avoiding abrupt change between postures.

References

1. Kamiya, K. et al. (2008). Sitting posture analysis by pressure sensors. *International Conference on pattern recognition* (ICPR).

2. Hasia, C.C. et al. (2009). Analysis and comparison of sleeping posture classification methods using pressure sensitive bed system. EMBC.

3. Nakajima, K. Matsumoto, Y., Tamura, T. (2000). A monitor for posture changes and respiration in bed using real time image sequence analysis. In: *Engineering in Medicine and Biology Society, 2000. Proceedings of the 22nd Annual International Conference of the IEEE.*

4. Hong Z. Tan et al. (2001). A Sensing Chair Using Pressure Distribution Sensors, IEEE/ASME Transactions on Mechatronics, **6**(3), 261-268.

5. Nishida Y. et al. (1999). Body Parts Positions and Posture Estimation System Based on Pressure Distribution Image, *IEEE International Conference on Robotics & Automation.*

6. Harada, T., Saito, A., Sato, T., Mor, T. (2000). Infant Behaviour Recognition System Based on Pressure Distribution Image. IEEE ICRA.

7. Duda R.O., Hart, P.E., Stork, D.G. (2001). *Pattern Classification* (2nd Edition). Wiley-Interscience, NY.

Evaluation of Alterations in Behaviour, Cognition, and Neuronal Excitability Induced by Administration of QTracker® 800 Quantum Dots

B. Salvetti[1], S. Becchi[2], F. Schio[1], E. Moscardo[1], G. Bertini[1], P.F. Fabene[1], F. Osculati[1]

[1] *Department of Neurological, Neuropsychological, Morphological and Movement Sciences, University of Verona, Verona, Italy. beatrice.salvetti@gmail.com*

[2] *Brain and Mind Research Institute, University of Sydney, Sydney, NSW, Australia.*

Abstract

Studies of Quantum Dots biodistribution have shown that these nanoparticles accumulate in various organs, including brain parenchyma, possibly causing functional alterations. To investigate its potential neurotoxicity, we administered QDs to young mice and studied their behaviour during the following weeks. Spontaneous locomotor behaviour in the home cage was not significantly altered in treated mice, the Rotarod and Grip Strength Meter tests were normal, whereas memory deficits were found with the Novel Object Recognition test up to three weeks after the injection. We also found a significant increase in the neuronal excitability of QDs-treated mice, following the administration of a Pentylenetetrazol challenge.

Keywords: Nanoparticles, Behaviour, Neurotoxicity, Novel Object Recognition.

Introduction

Quantum Dots are nanometre-sized particles generally composed of a heavy metal core made of cadmium-selenium (CdSe) or cadmium-telluride (CdTe) and a shell made of zinc-sulphide (ZnS) or zinc-selenide (ZnSe) to render QDs' bioavailability [1]. These semi-conductors, fluorescent nanocrystals are gaining increasing attention because of the large numbers of potential applications in biomedical research and neuroscience. However, the assessment of biocompatibility and biosafety of QDs is a critical issue for further applications as diagnostic and imaging tools on humans and still little is known about their potential toxicity. Previous studies in our laboratory have shown that QDs accumulate in the liver, spleen, and lungs. Moreover, we found that QDs can cross the blood brain barrier and can be observed by means of transmission electron microscopy in different cell types (both neurones and glia), both in the nucleus and in the cytosol.

We hypothesized that nanoparticles in the brain parenchyma may elicit an inflammatory response caused by an alteration in proinflammatory and signalling molecule expression, inducing fever and profound psychological and behavioural changes generally called ''sickness behaviour'' [2]. To such aim, we investigated possible alterations in basal locomotor behaviour, cognitive functions, and neuronal excitability in QDs treated mice.

Materials and Methods

BALB/c male mice received a single i.v. treatment (via tail vein), of 10µl/g 40pM QTracker® 800 QDs solution (Invitrogen). Animals were housed in PhenoTyper® chambers (45 cm W x 45 cm L x 60 cm H, Noldus Information Technology, Wageningen, The Netherlands) from the day of their arrival, under a reversed 12h light/12h dark cycle (lights off at 7 AM). Basal locomotor activity of the animals was recorded both before and after the treatment. Recognition memory was assessed using the NOR task [3] before the injection (baseline), and then 24 hours, 1 week, 2 weeks, and 3 weeks after the injection. Pentylenetetrazol (PTZ), a CNS stimulant, was injected 3 weeks after QDs injection.

Animal activity and behaviour during the NOR test was automatically recorded in the PhenoTyper®. Each chamber was divided into two portions with a white plastic divider, so that it accommodated two animals.

Automated tracking was performed using the EthoVision XT® software (Noldus Information Technology, Wageningen, The Netherlands).

The distance moved by each animal in the cage and its mean velocity (cm/s), mobility, and immobility durations were continuously recorded for 24 hours at different time points: before treatment (baseline), and 48 hours, 1 week, 2 weeks, and 3 weeks after treatment.

The NOR test consists of two phases, a sample trial (10 min duration) and a choice trial (5 min duration) with 60 min retention interval between the two (see Figure 1). During the sample trial two identical objects were placed in the PhenoTyper cage 10 cm away from the sidewalls and 7 cm from each other. The animal was left undisturbed for the delay period and then presented with a third copy of the previously seen objects and with a new one. To place and remove objects mice were confined at the end of the cage by the use of a mobile separator made of paper, which allows placing objects in the correct position without being seen by the animal. Ten different objects (in triplets) made of plastic, wood or glass that differed in terms of height, colour, shape and surface texture are used. The objects are fixed to the floor of the arena with tape to ensure that the mice could not displace them, they had no known ethological significance for the mice, and had never been paired with a reinforcer. Pilot studies ensured that mice of the BALB/c genetic background strain could discriminate between the objects, and there was no per se preference for one of these objects. The position and the identity of the novel object and familiar object were randomized between animals. The time spent exploring the objects was recorded. Exploration was defined as touching and sniffing the objects or directing the nose to them at a distance ≤ 2 cm. Discrimination ratio was used as an index of recognition memory, i.e. percentage of time spent exploring the novel object, relative to the total time spent exploring both objects $((t_{novel}/t_{novel}+t_{familiar})*100)$. Values higher than 50% suggest preference for the novel object.

In order to exclude a decreased object exploration due to impairments in locomotor functions animals were tested immediately after the NOR test for coordination and strength. Coordination and balance of the animals were measured with the Rotarod test (Ugo Basile, Comerio, Italy). Forelimb strength was assessed with the Grip Strength Meter test (Ugo Basile, Comerio, Italy).

To test further a possible alteration in neuronal excitability we injected PTZ i.p., a CNS stimulant with pro-convulsivant properties, 3 weeks after QDs injection. Latency from PTZ injection to 5 different convulsive stages was measured [4].

- Stages 1-3: hypoactivity and abdomen on bottom of the cage; twitches; mouse falls but recovers.
- Stage 4: falling and forelimbs parallel to body axis and stiff.
- Stage 5: major convulsion resulting in death.

Animals were sacrificed after a period of 30 minutes elapsed from stage 4. However, all animals reached stage 5 before this cut off.

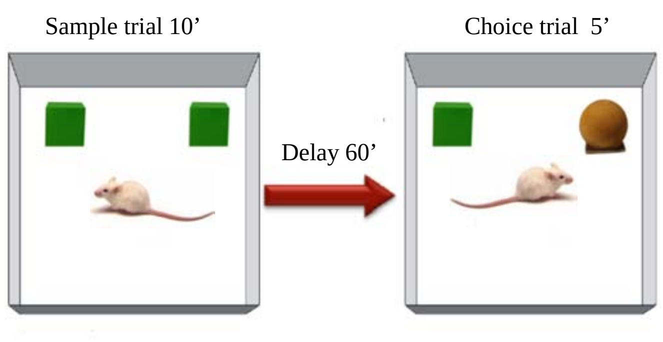


Figure 1. Schematic representation of novel object recognition test.

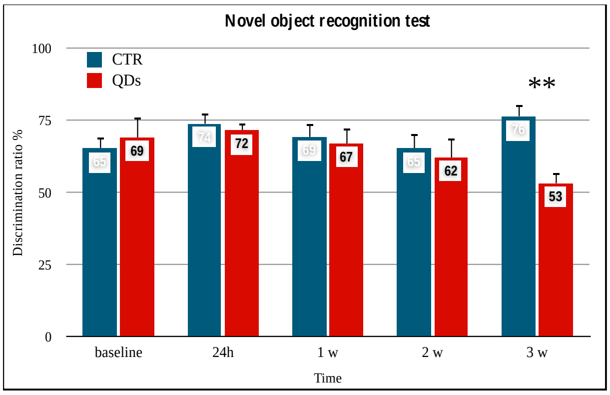


Figure 2 Novel Object Recognition results.

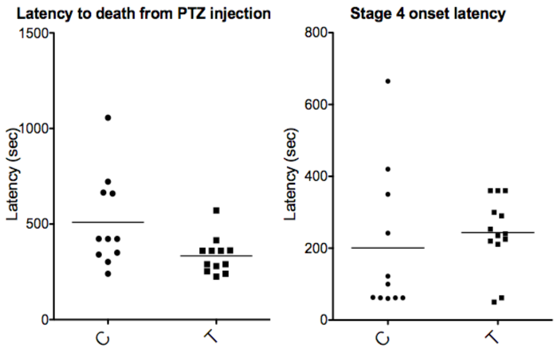


Figure 3. PTZ challenge results, latency from injection to death and stage 4.

Results and Discussion

Analysis of mice locomotor activity inside the chambers revealed no differences between QDs treated mice and controls in any of the considered parameters and time points. Rotarod and Grip Strength Meter tests revealed no differences between control and treated animals at any of the five time points. The NOR test revealed no differences between control and treated mice at baseline, 24 h, 1 week, 2 weeks but a difference was found in the discrimination ratio at 3 weeks after the treatment (t_{14}=4.283, $p<0.01$, see **Error! Reference source not found.**). In the PTZ challenge significant differences emerged between controls and treated mice in latency to death from injection (t_{21}=2.331; p=0.0298), and latency to stage 4 onset from the injection (t_{21}=2.508; p=0.0204, see **Error! Reference source not found.Error! Reference source not found.**). Data indicates that exposure to QDs induced a decrease in NOR performance over time, with an impairment at 3 weeks. Decrease in NOR performance might be caused by the presence of exogenous compound in the brain parenchyma, which can cause inflammatory response resulting in functional alterations. Data from the PTZ challenge show an increased duration of the seizures after the injection in control mice, that is, QDs treated mice died sooner than control ones after the treatment indicating an increased neuronal excitability possibly related to the presence of QDs in brain parenchyma.

Ethical statement

The experiments received authorisation from the Italian Ministry of Health, and are conducted following the principles of the NIH Guide for the Use and Care of Laboratory Animals, and the European Community Council (86/609/ EEC) directive. Mice are housed and used according to current European Community rules. Experiments on mice are approved by the research committees from the University of Verona and Italian National Institute of Health.

References

1. Akerman M.E., Chan W.C.W., Laakkonen P., Bhatia S.E., Ruoslathi E. (2002). Nanocrystal targeting in vivo. *PNAS* **99**(20), 12617-21.

2. Kent S., Bluthé R.M., Kelley K.W., Dantzer R. (1992). Sickness behaviour as a new target for drug development. *Trends in Pharmacological Sciences* **13**, 24-28.

3. Ennaceur A., Delacour J. (1988). A new one-trial test for neurobiological studies of memory in rats. 1: Behavioural data. *Behavioral Brain Research* **31**, 47-59.

4. Westmark C.J., Westmark P.R., Beard A.M., Hildebrandt S.M., Malter J.S. (2008). Seizure Susceptibility and Mortality in Mice that Over-Express Amyloid Precursor Protein. *International Journal of Clinical and Experimental Pathology* **1**(2), 157–168.

Monitoring Facial Expressions During the Mars-500 Isolation Experiment

R.D. Gorbunov[1], E.I. Barakova[1], R.M.C. Ahn[1], G.W.M. Rauterberg[1]

[1]*Departement of Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands*

There are many long-term missions that are performed by small crews in isolation. As examples we can mention missions performed on the international space station, polar research stations, submarines, oil platforms, meteorological stations and etc. Success of such missions strongly depends on emotional states of the crew members as well as on interpersonal relationships between them. Emotional problems of crew members and/or conflicts in a team can lead to a complete failure of a whole mission. The problem of the psychosocial health is also very important because the long-term isolated missions are usually accomplished by extreme physical conditions (such as weightlessness, low or high temperature, radiation, bad lighting and so on) which often have a strong and negative influence on the psychosocial atmosphere in the team. Moreover, these missions are also accomplished by negative psychological factors such as the isolation itself, spatial and social confinement, risks and so on. In this context it is very important to develop methods for monitoring and improvement of psychosocial atmosphere in isolated goal-oriented teams. In the present work we address the question of measuring of emotional states and interpersonal relations of small crews performing long-term missions in a strong isolation.

Among the methods for monitoring of psychosocial atmosphere we can distinguish the following three classes: usage of questionnaires, monitoring of nonverbal behavior (facial expressions [1,2], body movements [3,4], voice intonations [5]) and monitoring behavior of team member in cooperative computer games [6,7,8]. One of the advantages of methods based on questionnaires is that crew members can directly provide comprehensive information about their emotional states and interpersonal relations. However, people can be biased and not honest especially if they need to reveal their feelings and their vision of the conflicts that they are involved in. Moreover, a regular answering of extensive questionnaires can be very boring. In this context two other approaches becomes very useful. By nature people express their emotions through facial expressions and voice intonation which makes monitoring of nonverbal behavior very interesting. Interpersonal relationships are naturally expressed through interactions between people. In this context a use of computer games is very promising because they provide a way to formalize and simplify interactions between people as well as an easy way to measure behavior. More details about strength and shortcomings of different approaches as well as how they can be combined can be found in our earlier works [9,10] in which we described the design of our monitoring system used within the Mars-500 isolation experiment [11].

In the Mars-500 isolation experiment six participants have been isolated for 520 days to model a flight to Mars. Within this isolation experiment the six crew members biweekly participated in our experiments which were approximately half-an-hour long each. In the experiments the participants interacted with each other using software that we have provided. The interaction consisted of playing computer games and teaching each other. During the experimental sessions we made video records of the facial expressions of the participants. This work presents analysis of the obtained videos.

The extraction of the facial expressions from the video was done by FaceReader [12] software developed by Noldus Information Technology Company. The software automatically recognizes facial expressions by distinguishing six basic emotions plus natural. The accuracy of the facial expressions recognition was reported to be about 89%. More particularly, FaceReader classifies happy, angry, sad, surprised, scared, disgusted and neutral. The facial expressions recognition is based on Ekman and Friesen's theory of the Facial Action Coding System (FACS) that states that basic emotions correspond with facial models [13].

In total 33 experiments have been conducted. Every experiment provides 6 videos corresponding to different participants of the experiment. The average duration of the video is about 30 minutes.

To detect a long-term effect of the isolation on the emotional states of the crew members we have analyzed how different components of the facial expressions changes with time. The video data has been separated into two approximately equal parts: before the simulation of the landing on Mars and after that. The difference between the average components of the facial expression, corresponding to the two parts of the experiment, has been calculated. The statistical significance of the difference between the average values has been estimated and found to be low (ranging from 0.06 to 0.73 for different crew members).

To study the short-term effects we have studied a correlation between the facial expressions from the neighboring experiments (*i.e.* experiments separated by two weeks). To derive information about the interpersonal relations in the crew we have analyzed correlation between facial expressions of different crew members. The Pearson product-moment correlation coefficients between the components of the facial expression of different participant were found to be very low. However, the statistical significance of the values is very high. To study the relation between the facial expressions of different crew members in more detail, we have considered the joint distributions of the components of the facial expressions corresponding to two different crew members. Deviations of these distributions from the distributions of unrelated variables have been calculated and found to be statistically very significant. The relation between the emotional states of different crew members can be used as a measure of how emotionally bound two persons are. The form of the relation can be used not only do describe the strength of the bonding but also its type.

To get more information about the emotional states of the crew member within a given experiment we have analyzed not only the distribution of different facial expressions but also their time dynamics. In particular we analyzed how stable different emotional states (facial expressions are) and how frequently they changes.

To find the most condensed and complete description of the facial expressions we searched for independent degree of freedom in the components of the facial expressions. In order to get a better intuition of the emotional states of the crew member we used the FaceReader to extract a set of the most representative facial expressions. The representative expression can be used by people to interpret the emotional states of the crew members and then the real facial expression can be replaced (as an approximation) by the closest representative expression.

We have also presented general recommendations to researches and potential participants of similar experiments to help to prevent problems that we have encountered during our experiments.

References

1. Mase, K. (1991). Recognition of facial expressions from optical flow. *IEICE Transc.* **74**, 3474-3483.

2. Yacoob, Y., Davis, L. (1994). Computing spatiotemporal representations of human faces. *Proceedings of Computer Vision and Pattern Recognition 1994,* 70-75.

3. Barakova, E., Lourens, T. (2010). Expressing and interpreting emotional movements in social games with robots. *Personal and Ubiquitous Computing* **14**, 457-467.

4. Lourens, T., Barakova, E. (2009). Retrieving emotion from motion analysis: In a real time parallel framework for robots. *Proceedings of 16th International Conference on Neural Information Processing,* 430-438.

5. Oudeyer, P. (2003). The production and recognition of emotions in speech: features and algorithms. *International Journal of Human-Computer Studies* **59**, 157-183.

6. Rauterberg, M., Neerincx, M., Tuyls, K., van Loon., J. (2008). Entertainment computing in the orbit. *In Proceedings of New Frontiers for Entertainment Computing* **279**, 59-70.

7. Hennes, D., Tuyls, K., Neerincx, M., Rauterberg, M. (2009). Micro-scale social network analysis for ultra-long space flights. *In Proceedings of the IJCAI-09/SMC-IT-09/IWPSS-09 Workshop on Artificial Intelligence in Space.*

8. Gorbunov, R., Barakova, E., Ahn, R., Rauterberg, M. (2011). Monitoring interpersonal relationships through games with social dilemma. *Proceedings of International Conference on Evolutionary Computation Theory and Applications,* 5-12.

9. Voynarovskaya, N., Gorbunov, R., Barakova, E., Ahn, R., Rauterberg, M. (2010). Nonverbal behavior observation: collaborative gaming method for prediction of conflicts during long-term missions. *Proceedings of 9th International Conference on Entertainment Computing 2010*, 103-114.

10. Voynarovskaya, N., Gorbunov, R., Barakova, R., Rauterberg, M. (2010). Automatic mental health assistant: monitoring and measuring nonverbal behavior of the crew during long-term missions. *Proceedings of Measuring Behavior 2010, 7th International Conference on Methods and Techniques in Behavioral Research (Eindhoven, The Netherlands, August 24-27, 2010)*.

11. Russian Academy of Sciences' Institute of Biomedical Problems. The MARS-500 isolation experiment. http://mars500.imbp.ru/index.html. Accessed 28 June 2012.

12. VicarVision. FaceReader product page http://www.vicarvision.nl/FaceReader.html. Accessed 28 June 2012.

13. Ekman, P. and Friesen. (1977). *Manual for the Facial Action Coding System*. Consulting Psychologists Press, Palo Alto, CA, 1977.

Automatic Clustering of Conversational Patterns from Speech and Motion Data

S. Feese[1], B. Arnrich[1], G. Tröster[1], B. Meyer[2], K. Jonas[2]

[1]*Wearable Computing Lab., ETH Zurich, Zurich, Switzerland.* [1]*sfeese@ethz.ch*
[2]*Social and Business Psychology, University of Zurich, Zurich, Switzerland*

Abstract

Behavioral observation of teams is critical not only for team research, but also for leadership trainings. However, most current behavior observation methods used in team research often rely on manual video annotation which is time consuming and thus costly. In our work, we follow a sensor-based approach to automatically extract important pattern that capture conversational structure. Relying on automatic speaker detection from multiple microphones, we first employ a rule-based clustering to measure the proportions of typical conversational regimes taking place in a conversation and extend our method to learn more complex regimes from data using finite mixture models. Furthermore, we investigate the use of motion activity cues extracted from motion sensors. We apply our methods to a recent study on leadership in small groups and show how team meetings can be characterized.

Introduction

Conversation regimes characterize who speaks with whom in the course of a conversation. Typical pattern that can occur in a three person meeting are monolog, dialog and chat. In monologs, only one person speaks while other persons listen, whereas in dialogs two persons speak in an alternating manner. In the Chat regime all three persons participate in the part of the conversation and talk equally much. For a three person meeting all possible combinations are illustrated in the top line of Figure 2.

In this work we will focus on clustering such conversation patterns from audio recordings in a meeting scenario including three participants. We will first employ a rule-based clustering to illustrate the benefit of clustering. Second, we will use Gaussian Mixture Models to cluster conversational slices that are described by speech cues in order to capture more of the conversational dynamics. Additional to the speech modality, we apply the clustering method to motion data.

Related Work

Previous work on automatic analysis of social interactions in small groups dealt with automatic inference of conversation structure, analysis of social attention and the detection of personality traits and roles. A review on the on the topic can be found in [1]. These works mostly relied on computer vision based techniques and speech related cues such as speaking length, speaker turns and number of interruptions. Conversational patterns have been discovered using topic models in [2] and conversational scenes have been modeled with Dynamic Bayesian Networks in [3]. Wearable sensors in form of sociometric badges have been employed by Pentland and collaborators to measure “honest signals” in daily life [4]. In addition to speech activity cues as in [2], we investigate motion activity cues extracted from data measured with on-body motion sensors.

Methods

The complete processing chain is illustrated in Figure 1. Speaker diarization was performed by employing a threshold based approach. On a sliding window (length: 25~ms; step size: 10~ms) the signal energy was calculated for each group member. Speech was detected if the energy difference between the group members’ energy value and the mean value of the other group members was greater than an empirically set threshold. Speech activity segments shorter than 30~ms were then removed and segments of the same speaker within 1000~ms were merged.

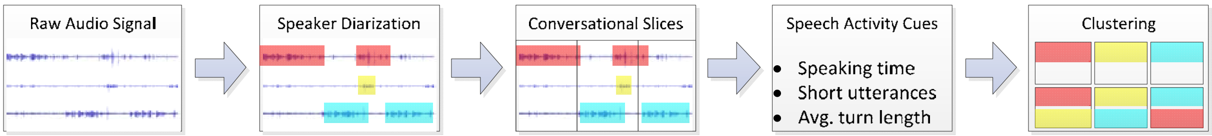


Figure 1. Processing steps.

To summarize the speech status of each group member over a two minute period conversational slices are extracted. Based on the speech activity cues proposed in [2], we calculate for each slice the following speech activity cues: total speaking time, number of short utterances as well as median and inter-quartile-range of the turn duration. We choose two minute slices because it resembles a rough minimum of a dialog length if one assumes that one speech act is about 30 seconds in length and that it requires a minimum of two alternating turns of each person.

For the rule-based clustering, we only use the total speaking time of each group member per conversational slice. We define seven clusters that correspond to the conversational regimes monolog, dialog and chat as follows: In a monolog one group member speaks more as twice as much as each other group member, in a dialog two group members speak more as twice as the third group member each. The chat regime captures all other cases.

In order to incorporate more of the discussion dynamics, we also learn Gaussian Mixture Models with k (k between 1 and 10) mixture components using for each group member the above mentioned speech activity cues as observed data. The emission probabilities are modeled as single 12 dimensional Gaussians with diagonal covariance and all model parameters are learned with the Expectation Maximization Algorithm (see [5] for more on mixture models).

One important parameter to estimate is the number of mixture components, which we assume to be equal to the number of clusters in the data. As we are interested in clusterings that are most reliable against noise in the data, we follow a resampling approach to evaluate cluster stability for a different number of components. We employ the cluster stability criterion presented in [6] which measures the similarity of clustering solutions that were obtained on random subsets of the data to the solution that was obtained on the complete data set.

The principle of clustering conversational slices can be generalized to other modalities. As an example, we show how the method can be used to discover conversational clusters from motion data. Analogous to above, each conversational slice can be represented through a set of motion activity cues which simply capture the activity of each body limb in terms of movement time and the number of short movements.

Data Set

We apply the clustering methods to a recent study on leadership behavior [7] in which groups of three participants were led by two different types of leaders. Fifty-five groups were asked to work on a hidden-profile decision making task to rank four fictive candidates with regard to their suitability for an open job position. Under the guidance of the group leader, the group had to discuss the suitability of each candidate and agree on a rank order which served as a measure of group performance. Groups with an individually considerate leader who encouraged all team members to contribute to the solution achieved a higher performance as opposed to groups that were lead authoritarian.

In the experiment each group member was equipped with a separate clip-on lapel microphone to record the speech of all group members (sampling frequency 44.1 kHz). Additionally, the upper body motion of each group member was captured with six inertial measurement units (XSens MTx) which were located on the lower arms, the back and the head (sampling frequency 32 Hz). In total, the data set contains 44 group discussions (16 groups were led authoritarian and 18 with individual consideration) with three participants each. Our data set totals to over 15 hours of discussion time.

Results and Discussion

In Figure 2 the results of the rule-based clustering solution are shown in terms of the mean speech distribution and the corresponding mean head orientation. The different types of conversation regimes are identified as expected. In case of monologs, one group member dominates in terms of speaking length. It appears natural that the speaker is looked at, which we can infer from the mean head orientation. In case of dialogs, speech is distributed across two speakers and considering the mean head orientation we can confirm that the dialog partners are orientated towards each other. To illustrate how the clustering solutions can discover some of the reasons why groups of individually considerate leaders achieved higher group performances, we present the cluster distributions across all groups. The bottom line of Figure 2 displays the mean and standard deviation of the relative amount of time that each cluster is present in a group discussion. In the experiment the leader was always person P1. As one can see, groups that were led by individually considerate leaders stayed longer in the regimes in which persons P2 and P3 speak with each other. This suggests that individually considerate leaders gave their followers more room to talk to each other which in turn led to more information being shared and thus higher group performance. The results of the model based clustering of speech activity cues are presented in the top of Figure 3. Not considering the trivial clustering solution of one cluster, we find from the stability analysis (top left) that models with four components are most stable in our data. The found clusters can be described by their corresponding cluster means which are illustrated in the top right of Figure 3 (different colors encode different group members). Cluster 1 represents conversational slices in which person P2 speaks mostly, while person P1 contributes with varying turn lengths. Person P3 speaks most in Cluster 2 and P1 acknowledges with short utterances. In Cluster 3 person P1 speaks most with long turns and person P2 and P3 acknowledge with short utterances. Finally, Cluster 4 represents conversational slices in which the group members talk with short utterances.

The bottom of Figure 3 presents the results of the cluster analysis in which we used motion activity cues to capture the movement of the lower arms, the back and the head during the conversational slices. The stability analysis suggests three clusters, which as we can see from the cluster means belong to slices in which one of the group member was active while the others did only shortly move their upper body limbs.

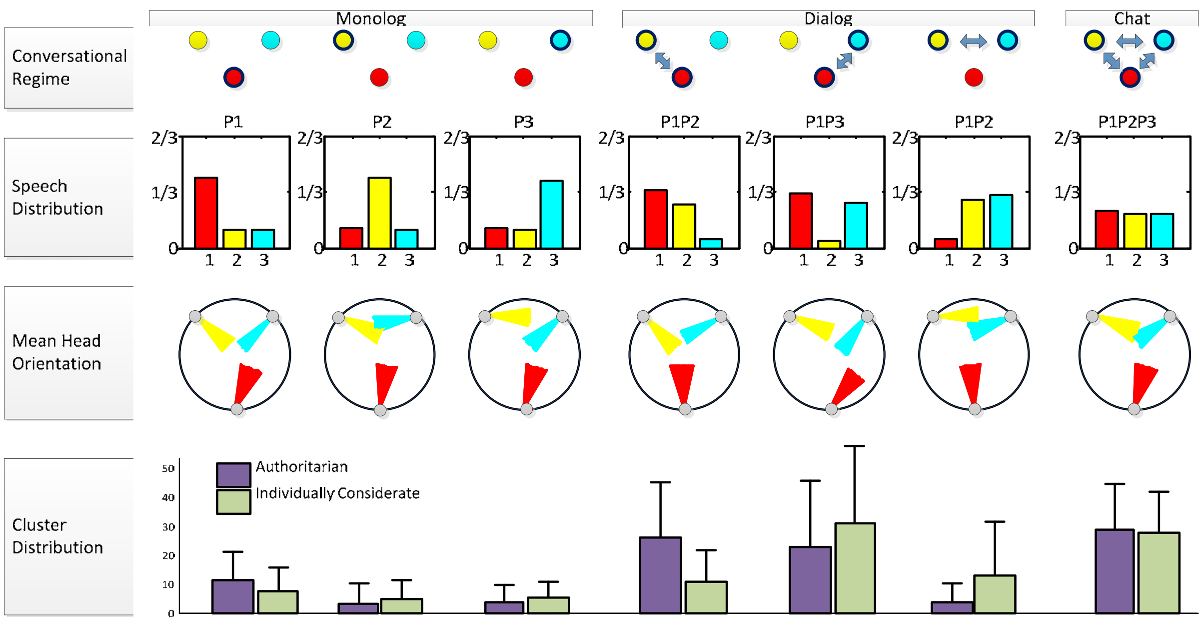


Figure 2. Rule-based conversational regimes with their corresponding speech distribution and mean head orientation. Only the speaking length of each group member per conversational slice was used for the clustering. Button row: Cluster distribution across discussions with two different leadership types.

Proceedings of Measuring Behavior 2012 (Utrecht, The Netherlands, August 28-31, 2012)
Eds. A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, and P.H. Zimmerman

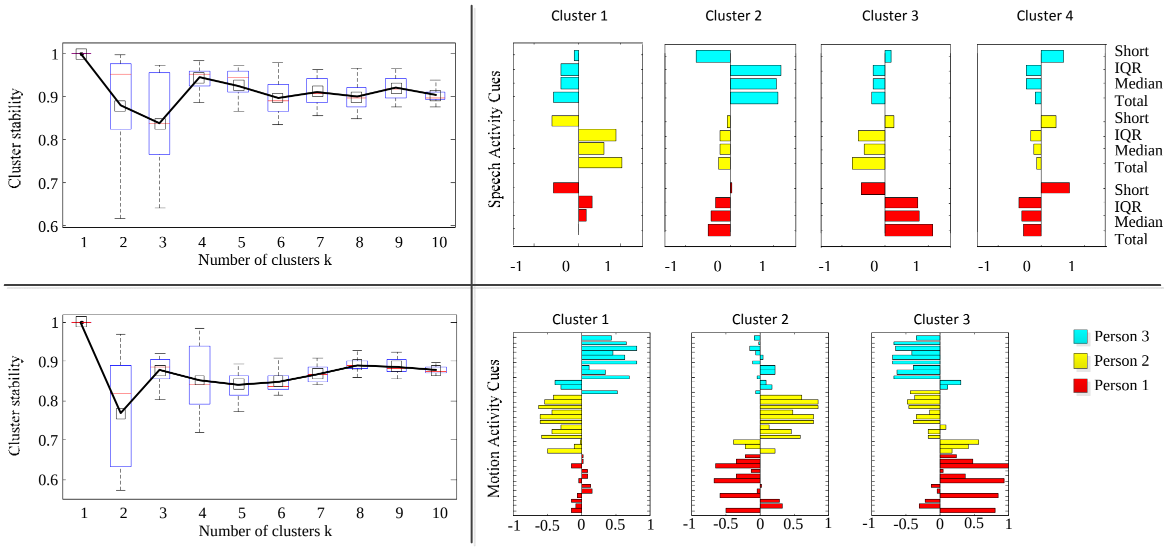


Figure 3. Top left: Cluster stability of speech activity cue based mixture modeling for different number of mixture components. Top right: Cluster means of the clustering solution for k = 4. Bottom left: Cluster stability using motion activity cues. Bottom right: Cluster means of the clustering solution for k = 3.

Conclusion

We have presented a rule-based method to cluster conversational slices into meaningful regimes and demonstrated their usefulness to discover possible reasons for higher group performance in a recent study on leadership. Furthermore, we have shown how speech and motion activity cues can be used to cluster conversational slices. While the clustering solutions based on speech activity appear to be stable and meaningful, the clusterings of simple motion activity cues appear to be less stable, suggesting that simple motion activity cues do not capture enough of the nonverbal communication expressed by body motion. In future, we will investigate other clustering algorithms and more meaningful motion cues such as head nodding, gesticulating and body postures.

References

1. Gatica-Perez, D. (2009). Automatic Nonverbal Analysis of Social Interaction in Small Groups: A Review. *Image and Vision Computing* **27**(12), 1775-1787.
2. Jayagopi, D.B., Gatica-Perez, D. (2010). Mining Group Nonverbal Conversational Patterns Using Probabilistic Topic Models. *IEEE Transactions on Multimedia* **12**(8), 790-802.
3. Otsuka, K. (2011). Conversation Scene Analysis. *IEEE Signal Processing Magazine* **28**(4), 127-131.
4. Pentland, A. (2008). Honest Signals, Bradford Books.
5. Frühwirth-Schnatter, S. (2006). Finite Mixture and Markov Switching Models, Springer.
6. Levine, E., Domany, E. (2001). Resampling Method For Unsupervised Extimation of Cluster Validity. *Neural Computation* **13**(11), 2573-2593.
7. Meyer, B., Jonas, K., Feese, S., Arnrich, B., Tröster, G., Schermuly, C.C. (2012). A social-signal processing view on leadership: Specific behaviors characterize considerate leaders. Paper accepted for presentation at the Academy of Management Annual Meeting 2012, Boston, MA.

Towards Sensing Behavior Using the Kinect

Wouter van Teijlingen[1,2], Egon L. van den Broek[1,3], Reinier Könemann[4], John G.M. Schavemaker[1]

[1] *Media and Network Services, Technical Sciences, TNO, Delft, The Netherlands*

[2] *Department of Computer Engineering, Faculty of Natural Sciences & Technology, Utrecht University of Applied Sciences, Utrecht, The Netherlands*

[3] *Human Media Interaction, Faculty of Electrical Engineering, Mathematics, and Computer Science, University of Twente, Enschede, The Netherlands*

[4] *Sustainable Productivity, Quality of Life, TNO, Hoofddorp, The Netherlands*

wouter@van-teijlingen.nl, vandenbroek@acm.org, reinier.konemann@tno.nl, john.schavemaker@tno.nl

Abstract

A method is proposed to validate Microsoft's Kinect as a device and, hence, to enable low fidelity, unobtrusive, robust sensing of behavior. The Xsens MVN suit is proposed as the measurements' ground truth. An overarching framework is introduced that facilitates a mapping of both devices upon each other. This framework includes a complete processing pipeline for both the Xsens and the Kinect data, recorded in parallel, which, at the end of both pipelines, are mapped upon reach other. Next, two strategies are presented that aim to interpret the data gathered. We close with a brief discussion on the pros and cons of the protocol proposed.

Introduction

In November 2010, Microsoft launched its Kinect, a motion sensing input device, as an alternative game controller for Xbox 360 video game consoles. Kinect competes with game controllers such as the Nintendo Wii's Remote Plus and Sony Computer Entertainment's PlayStation 3 Eye motion controllers. However, the Microsoft Kinect can also be used for various other applications than games. This has drawn the interest from science and engineering to this game controller. Also Kinect's low price (i.e., approx. €100,-), compared to that of other (traditional) scientific tools, has added to the interest it receives from science and engineering. Already within months from the introduction of the Kinect, various researches were presented and the first papers were published. The Kinect has already been used in various domains for several applications. For example, Gallo and colleagues [1] employed the Kinect to enable controller-free exploration of medical image data; Chang and colleagues [2] developed a Kinect-based system for physical rehabilitation; and Kamel Boulos and colleagues [3] introduced Kinoogle, a Kinect interface for interaction with Google Earth. This triplet illustrates the vast amount of research and development that has already been done using the Kinect, within two years since its introduction.

This article addresses yet another application domain for the Kinect: *sensing behavior*. This domain is not new [4,5]; however, the Kinect can make sensing behavior affordable for every day's (real world) practice. Next, in Section 2, we present our research rationale including its key elements. In Section 3, we discuss our quest towards low fidelity, robust, unobtrusive sensing of behavior. Its essence is a method to validate the Kinect for (real-time) behavior classification, using the Xsens suit (see also [6]). We close this article with Section 6, which provides a brief discussion and the conclusion.

On sensing of behavior

Traditionally (e.g., in experimental psychophysiology), "human motor activity is indexed by surface electromyography (EMG), which requires the placement of electrodes on the skin surface. This limits the mobility of the subject and reduces the possible contexts in which motor responses can be recorded and evaluated. In addition, EMG captures only the activity of specific muscle groups … ." [7]. With the rise of computing power and the steep progress of image processing and computer vision, behavior monitoring became possible via the visual domain [4,5]. However, as with EMG-based monitoring, computer vision-based monitoring of human behavior requires a very high level expertise and rather expensive apparatus [7,8]. With the Kinect these two issues seem to resolve and, additionally, traditional computer vision is augmented with an infrared (IR) depth sensor [9]. Sensing behavior is easier said than done. It starts with agreeing upon a definition of behavior. Although all of us have an intuitive understanding of what behavior is, providing a stipulate definition remains challenge. In particular, in engineering human-related concepts, such as behavior is, are often ill defined; in other words, such research has poor construct validity. To tackle this traditional flaw, we define behavior as an observable, measurable movement of some part of the body through space and time. Moreover, we further specify behavior by providing concise definitions of motion, action, and activity, as given in Table 1.

Validation of the Kinect: An overarching framework

In parallel, motion data is captured with the Kinect camera and the Xsens MVN body suit. An overarching framework is designed that includes two processing pipelines that are connected at their start (i.e., the data acquisition) and at their end (i.e., their mapping); see also Figure 1. First, we will describe the Xsens processing pipeline. Second, we will describe the Kinect processing pipeline. Third and last, we will describe their mapping. The Xsens body suit captured motion via 17 inertial motion trackers (i.e., 3D gyroscopes, 3D accelerometers, and 3D magnetometers). Their data is sent to the affiliated MVN suit, which saves it as a Biovision Hierarchy file (BVH) file, an animation file type (see also Figure 1). The Kinect captures an infrared signal (IR) and a video (RGB) stream in parallel. Both are sent to the newly developed ONI (OpenNI) recorder, which was founded on the Nestk library [10]. The ONI recorder generates an ONI file, which is sent to the newly developed Kinalyze module. The Kinalyze module converts the ONI file to a BVH file. Hence, both processing pipelines have a BVH file as output, which can be mapped upon each other. Having two BVH files, one that originates from the Xsens body suit and one that originates from the Kinect, a mapping between them can be realized. This requires the alignment of both data files. Next, a digital human model (DHM) through time has to be generated using both BHV files. To realize this, angles between joints have to be determined. Subsequently, the movements of the DHM have to be interpreted, which depends on the context of use. This phase of interpretation will be discussed next.

Table 1. Classification of four level of behavior: motion, action, and activity, adapted from Chaaraoui et al. [8]

	Time window	Description
motion	(parts of) seconds	Image segmentation, movement detection, and gaze and head-pose estimation.
action	seconds – minutes	Identify interaction of users with objects. Recognize primitives such as sitting, standing, walking, and running.
activity	minutes – hours	Sequence of actions with a particular order. Subsequently, recognition of daily activities (e.g., cooking or showering).

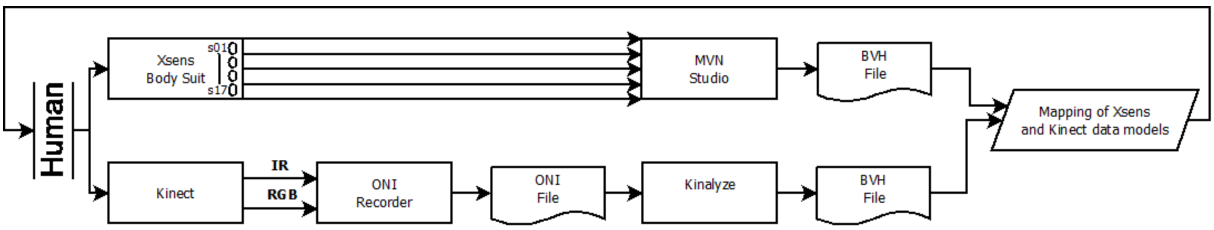


Figure 1. An overarching framework or processing pipeline for mapping the Xsens body suit data and the Kinect data upon each other.

Towards low fidelity, unobtrusive, robust sensing of behavior

Microsoft Kinect and the Xsens MVN exploit distinct modalities that can be used in parallel, hardly without any restrictions. This enables parallel data gathering for both devices. As such, an optimal set up can be achieved to validate the precision of the Kinect's measurements in practice. Although, the Kinect has been evaluated for kinematic measurement [9,11], to the authors' knowledge, it has not yet been compared directly with the Xsens MVN suit as a solid ground truth. When tracking of motion, action, and activities with a high temporal and spatial precision is needed, Xsens is the preferred device among this duo [6]. However, in practice, for various (real life) applications activity monitoring is at hand (see also Table 1), where precision in both the temporal and spatial domain is less important and ease of use is crucial. In this case, Kinect could provide an interesting low fidelity alternative, as it enables unobtrusive and robust sensing of activity and behavior [cf. 1,2,3,9,11].

One of the core differences between the output of Xsens and Kinect is the DOF in their measurement [6]. Xsens provides 45 DOF model (i.e., 23 segments and 22 joints) to characterize movements, where Kinect relies on 15 DOF (i.e., when using the ONI skeleton tracking algorithm). This is a crucial distinction when analyzing motion and action. The software packages available with both Xsens and Kinect support the use of skeleton models, which provide accurate modeling of motion and, subsequently, action. This enables knowledge-based dimension reduction. Depending on the task or context at hand, possibly only parts of the models are of interest. For example, when sitting behind a desk, the lower part of the body is of little interest. We will select the appropriate DOF of Xsens and, subsequently, select Kinect's DOF that map upon Xsens's DOF. Where no mapping is possible, we aim to determine mappings by way of interpolation.

An alternative for knowledge-based dimension reduction, as can be done using skeleton models, is data-driven dimension reduction. To exploit this, a set of standard motions, actions, and activities will be identified. Participants in experiments will be asked to execute these motions, actions, and activities, while wearing the Xsens MVN suit and while being observed by the Kinect. Offline signal processing and classification of the signals can be utilized to explore possible mapping between the signals generated by both devices. In a nutshell, of both devices the signals are captured, the signals are processed. A measurement space is defined, preprocessing (e.g., filtering and artifact removal), synchronization, and segmentation is applied. Next, features and, subsequently, their parameters are extracted from each of the signals. A pattern space is identified and appropriate parameters are selected, which results in a reduced pattern space. This pattern space is fed to a classifier (e.g., principal component analysis or a support vector machine, SVM). In case of supervised learning (i.e., the SVM), errors in the classification process are detected and the classification process is adapted. Note that the latter phase requires appropriate validation, which can be realized by separating training and test sets or by conducting cross validation.

Conclusion

This article merely proposed a method to validate Microsoft's Kinect as a device to enable low fidelity, unobtrusive, robust sensing of behavior. The Xsens MVN suit is presented as means to establish the measurements' ground truth (cf. [9,11]). To facilitate this, an overarching framework has been presented that facilitates the mapping of both devices' data streams. Moreover, the required alignment of both data streams is discussed to achieve this mapping.

The Kinect has tremendous advantages compared to most other devices. However, similar as most computer vision based approaches, the Kinect suffers from the physiognomies of body and faces. These vary considerably among individuals due to age, ethnicity, gender, facial hair, cosmetic products, and occluding objects (e.g., glasses and hair). Furthermore, both body and face can appear to be distinct from itself due to pose and/or lighting changes, or other forms of environmental noise. When the Kinect indeed would prove itself as a low fidelity, unobtrusive, and robust device for sensing of behavior, this could yield significant progress on measuring behavior in general [9,11] (cf. [1,2,3]). Its broad usage has the significant advantage that software packages are constantly both improved and extended (e.g., [10]). High level programming languages enable easy access, processing, and interpretation of Kinect data. Consequently, we expect that the measuring behavior community will embrace the Kinect and include it in its standard measurement setups.

Acknowledgments

This publication was supported by the Dutch national program COMMIT (project P7 SWELL). The authors also thank Joop Kaldeway (Utrecht University of Applied Sciences) and Tim Bosch (TNO) for their comments. Further, the authors gratefully acknowledge the two anonymous reviewers for their to the point and constructive remarks.

References

1. Gallo, L., Placitelli, A.P., and Ciampi, M. (2011). Controller-free exploration of medical image data: Experiencing the Kinect. In M. Olive and T. Solomonides (Eds.), *Proceedings of the 24th International Symposium on Computer-Based Medical Systems (CBMS)* (Bristol, UK, 27-30 June 2011), 1-6. Piscataway, NJ, USA: IEEE.

2. Chang, Y.-J., Chen, S.-F., Huang, J.-D. (2011). A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. *Research in Developmental Disabilities* **32**(6), 2566-2570.

3. Boulos, M.N.K., Blanchard, B.J., Walker, C., Montero, J., Tripathy, A. and Gutierrez-Osuna, R. (2011). Web GIS in practice X: a Microsoft Kinect natural user interface for Google Earth navigation. *International Journal of Health Geographics* **10**(1), 45.

4. Klein Breteler, M.D., Meulenbroek, R.G.J., Gielen, S.C.A.M. (1998). Geometric features of workspace and joint-space paths of 3D reaching movements. *Acta Psychologica* **100**(1-2), 37-53.

5. Noldus , L.P.J.J., Spink, A.J., Tegelenbosch, R.A.J. (2001). EthoVision: A versatile video tracking system for automation of behavioral experiments. *Behavior Research Methods*, *Instruments*, *& Computers* **33**(3), 398-414.

6. Roetenberg, D., Luinge, H, and Slycke, P. (2008). 6 DOF motion analysis using inertial sensors. In *Proceedings of Measuring Behavior 2008: 6th International Conference on Methods and Techniques in Behavioral Research* (Maastricht, The Netherlands, 26-29 August 2008), 14-15. Wageningen, The Netherlands: Noldus Information Technology.

7. Vousdoukas et al. (*in press*). SVMT: A MATLAB toolbox for stereo-vision motion tracking of motion reactivity. *Computer Methods and Programs in Biomedicine*.

   DOI: http://dx.doi.org/10.1016/j.cmpb.2012.01.006

8. Chaaraoui, A.A., Climent-Pérez, P., and Flórez-Revuelta, F. (2012). A review on vision techniques applied to human behaviour analysis for ambient-assisted living. *Expert Systems with Applications* **39**(12), 10873-10888.

9. Dutta, T. (2012). Evaluation of the Kinect sensor for 3-D kinematic measurement in the workplace. *Applied Ergonomics* **43**(4), 645-649.

10. Burrus N. (2012). Developing your own software based on the nestk library. <http://labs.manctl.com/rgbdemo/index.php/documentation/nestk>. Accessed 1 July 2012.

11. Ray, S.J., Teizer, J. (2012). Real-time construction worker posture analysis for ergonomics training. *Advanced Engineering Informatics* **26**(2), 439-455.

Sonification of Experimental Parameters as a New Method for Efficient Coding of Behavior

Andrea Ravignani[1], W. Tecumseh Fitch[2]

Department of Cognitive Biology, University of Vienna, Vienna, Austria.

[1]andrea.ravignani@univie.ac.at, [2]tecumseh.fitch@univie.ac.at

Abstract

Cognitive research is often focused on experimental condition-driven reactions. Ethological studies frequently rely on the observation of naturally occurring specific behaviors. In both cases, subjects are filmed during the study, so that afterwards behaviors can be coded on video. Coding should typically be blind to experimental conditions, but often requires more information than that present on video. We introduce a method for blind-coding of behavioral videos that takes care of both issues via three main innovations. First, of particular significance for playback studies, it allows creation of a "soundtrack" of the study, that is, a track composed of synthesized sounds representing different aspects of the experimental conditions, or other events, over time. Second, it facilitates coding behavior using this audio track, together with the possibly muted original video. This enables coding blindly to conditions as required, but not ignoring other relevant events. Third, our method makes use of freely available, multi-platform software, including scripts we developed.

Keywords. Sonification, blind coding, observational data, scoring behavior, experimental condition, behavioral observation, animal behavior, playback experiment, Python, ELAN.

Introduction

Experimental research in behavior and cognition often follows a typical methodological pattern. While a subject is exposed to several conditions, its reactions are videotaped. Afterwards, the researcher scores these data by blind coding specific behaviors from the video.

We developed a method that allows for an easy and more ergonomic video coding of behaviors and reactions, particularly suited for playback studies. While the person coding the video is busy with her visual task, she may need further information other than that visible in, or audible from, the original recording. Similarly, when a video is to be coded blindly, and its audio track must therefore be silenced, the coder nonetheless needs temporal guidance on what specific behaviors to code and at what point in time. In our approach, such information is broadcasted to the coder by the occurrence of specific sounds, played synchronously with the video.

Methods

The crucial innovation in our approach is to have an audio track where all aspects of the experiment relevant to video coding are mapped to sound (that is to say, "sonified"). Here we outline the step-by-step procedure that can be used to achieve this auditory-aided video coding (see Figure 1). The emphasis is on the methodology and the conceptual approach used, rather than the specific pieces of software employed. Therefore, in the rest of the section, the general description of each step is followed by a short description of the specific software solution we chose. Our choices, however, do not preclude alternative solutions.

1. While filming, the experimenter keeps a log of the events that, not being present on video, she would still like to include at the time of coding. Alternatively, the experimenter may use experimental software that automatically keeps a log with time and type of experimental conditions. For example, on Mac OS X platforms, Playback 3.0 [1] offers such a possibility for playback experiments (see reference [1] for download instructions), and its output are plain txt files. In any case, a log can be thought of as a list of different events whose time of occurrence is specified.

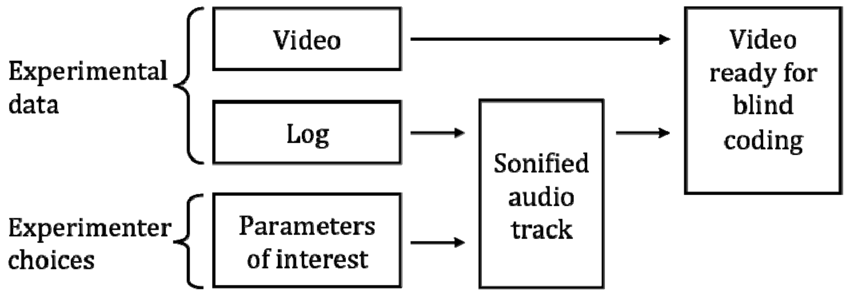


Figure 1. General procedure from experimental data to video annotation. The log of events collected during the study, together with the sounds the researcher decides to associate with particular parameters of interest, gives the sonified soundtrack of the experiment. This gets combined with the video and it is ready to be coded.

2. The experimenter decides what sound to associate with which type of event. Our first choice, dictated by perceptual saliency, has fallen on pure tones that differ by more than one octave in frequency. However, alternative solutions are easily implementable.

3. The log of events is transformed into an audio track, where each type of sound has a different meaning. Therefore the experiment has an “auditory counterpart” associated, spanning its entire duration in real time (see Figure 2). This is accomplished using a Python (Python 2.x, www.python.org) script we wrote, whose source code is available from the first author upon request. Python is a high-level, multi-platform programming language, built-in in Mac OS X and also available for download to Windows and Linux users. In particular, our script only makes use of libraries already included in Python, therefore avoiding additional installations.

   What follows is the pseudocode on which our sonification script is based, enabling interested researchers to create an alternative version of the script using their preferred programming language.

- Create a silent audio file spanning the entire duration of the experiment.
- Create a different sound for each relevant event.
- For each time step in the audio file.
  - If an event requiring sonification is present.
    - Insert the corresponding sound in the audio file.

4. Audio and video are easily synchronized by manually trimming the onset of the first event in the video. For the case of playback experiments, this is equivalent to trimming the video up to the first stimulus onset. Synchronization accuracy depends on the quality and accurateness of the event log (see point 1). In studies requiring high synchronization accuracy, researchers should opt for a computer-based log-keeping system. An upper bound for the precision of the synchronization, on the other hand, is determined by the video frame rate. Trimming can be accomplished using any video editing software. iMovie (included in Mac OS X, our choice for testing the system), Windows Movie Maker (built-in in Windows) and Apple QuickTime (both platforms) are, for instance, uncomplicated, free solutions.

5. Videos can be coded using a wide array of available software. We tested our system using the freely available video annotation software ELAN [2-3], which runs on Windows, Mac OS X and Linux platforms (see reference [2] for download). In ELAN, when simultaneously importing audio and video files, their onsets get automatically aligned. Therefore, once the video has been trimmed (see point 4), the audio-video synchronization is straightforward and automatic.

6. An additional Python script (also available on request) can be used to check whether the behaviors and reactions coded satisfy the conditions originally imposed by the researcher and dictated by the audio

track. For instance, if behaviors are supposed to be coded within a given time window, the script can trim, discard or leave unaltered the annotations that exceed that time span, depending on the needs of the study. Additionally, when data from more than one coder is present, our script calculates the inter-rater agreement, both as a naïve agreement coefficient (joint probability) and using Cohen's kappa statistic [4].

In the next sections, we outline two possible research contexts where our method is particularly suitable.

Playback Experiments

In a playback experiment, different categories of sounds are played to an infant or a non-human animal. The subject's differential reactions among stimuli types are used to investigate aspects of her auditory cognition, such as discrimination capabilities, individual recognition, generalization within classes, etc.

While coding the video recording of the experiment, the researcher needs to know when a stimulus was played, so to look for reactions to code. However, to avoid biasing the coding, she must be blind to the specific sound played. Therefore, the original audio track is removed from the video and substituted with placeholder sounds. These (pure sine waves at different frequencies) signal when the playback started, when the coder can actually begin coding reactions and when no coding is allowed anymore.

This method is currently used at the University of Vienna (together with ELAN [2-3]) to code behavior in cognitive experiments in several primate species.

Observations of Naturally Occurring Behavior

Researchers in ethology usually focus on naturally occurring behaviors. Either observations are coded in real time or a focal subject (or a group of individuals) is filmed and the behavior coded at a later time. In the second case, several environmental and social phenomena may escape the camera, but are recorded by the researcher at the site. Such events may include, for instance, fights among other members of an animal group, extraneous human interactions, appearance of a predator, etc. Our method offers a way of integrating, at the time of coding, this manually recorded information with the video.

Conclusion

While coding specific behaviors from a video, two types of information are important. On the one hand, there are sounds that were recorded in the original video but may need to be concealed from the blind coder. On the other hand, there may be information not present in the video that is nonetheless essential while coding. The method

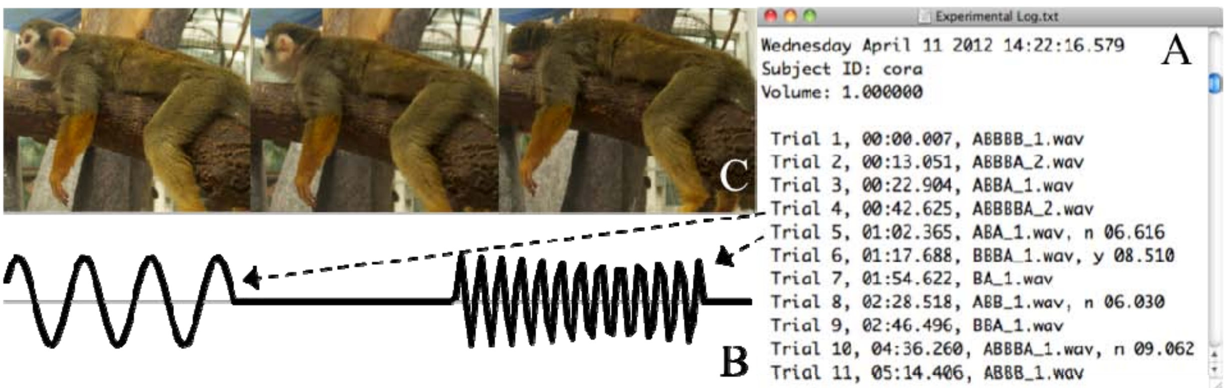


Figure 2. Sketch of the sonification procedure. Using a Python script, each event in the experimental log (A) is transformed into one or more sounds chosen by the researcher. The resulting soundtrack (B) is joined with the original video (C) and these are played simultaneously during behavioral coding. (Squirrel monkey pictures by Markus Boeckle)

we present here offers a simple solution to both issues at once. Information relevant to the experiment is sonified, that is, experiment-relevant events are transformed into sounds. This stream of sounds is then substituted to, or flanked by, the sound track of the original video. This allows coding blind to experimental conditions while keeping track of other events relevant to the study. We provided two practical examples of the many possible applications that this new method has in the study of human and non-human animal behavior and cognition. Finally, we show how the entire procedure, from performing the experiment to annotating behaviors, can be performed using freely available software.

Acknowledgments

The first author is supported by ERC Advanced Grant 230604 SOMACCA awarded to T. F. We thank Drasko Boko, Sarah Deventer, Maxime García, Bruno Gingras, Mauricio Martins, Stephan Reber and Ruth Sonnweber for testing the system and providing useful comments.

References

1. Jones, R., Fitch, T. (2009). Playback 3.0. Copyright 2009. Download requires git (multi-platform and freely available at http://git-scm.com/) and typing the following command: git clone git://playback.git.sourceforge.net/gitroot/playback/playback Accessed September 2011.

2. ELAN Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. http://www.lat-mpi.eu/tools/elan/ Accessed March 2012.

3. Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*, 1556-1559.

4. Cohen, J. (1960) A coefficient for agreement for nominal scales. *Education and Psychological Measurement* **20**, 37–46.

Video-Based Analysis of Fear Conditioning: A Validation Test

P.F. Fabene, G. Bertini, M. Pellitteri, E. Moscardo, B. Salvetti, F. Schio, A. Chakir, C. Tognoli

University of Verona, Faculty of Medicine, Dept. of Neurological, Neuropsychological, Morphological & Motor Sciences

Section of Anatomy and Histology Laboratory of Functional Neuroanatomy, Verona, Italy

paolo.fabene@univr.it, giuseppe.bertini@univr.it, michele.pellitteri@univr.it, elena.moscardo@univr.it, beatrice.salvetti@univr.it, federica.schio@univr.it, chakirasma@gmail.com, cristina.tognoli@univr.it

Introduction

Fear conditioning is a form of Pavovlian learning, where an electric shock is paired with an acoustic stimulus. Animals learn the association of the aversive stimulus (shock) with the auditory one, but also with the context where the shock takes place [1].

This type of learning is classified as a form of non-declarative memory and has several advantages. First and foremost is a form of one-trial learning, which is acquired rapidly and requires only one conditioning session, after which a robust and long-lasting behavioural change is produced. Moreover, it doesn't require food or water deprivation. Finally, fear conditioning is an important basic learning mechanism, which is common both to human and non-human animals and its neural mechanism appear to be relatively conserved across species [1-3].

The main behavioural measure in fear conditioning is the number and duration of freezing episodes, which may be scored by a trained observer or toward automated scoring system [1]. We would like to validate automatic scoring using the EthoVision® 8.5 software (Noldus Information Technology, Wageningen, The Netherlands) in both mice and rats, and further analysed by The Observer® XT (Noldus Information Technology).

Materials and Methods

Fear conditioning is carried out in both mice and rats, using the new Fear Condition System by Ugo Basile (Ugo Basile, Comerio, Italy). The system comprises a shock chamber, made of three/four black/white acrylic walls, the ceiling of white acrylic, the floor of the chamber is made of metal bars that deliver the electric foot shock. It also consists of a shock generator for the administration of shocks at various intensities, a tone generator, and an isolation cubicle.

We will test fear conditioning in two different protocols; in particular, the experiment consists of two sessions, a training session, and a test session. On day 1 the training is performed, the animal is placed in the conditioning chamber. The animal is left free to explore the chamber for

- Protocol A) 2 min before the presentation of the acoustic stimulus (30 s) followed by the presentation of the foot shock (2 s). [1]
- Protocol B) 1 minute before the presentation of two tone-shock pairings with an acoustic stimulus (20 s) and a 500 ms long electric shock [4]

The second day the test is performed.

During context test the subject is placed in the same conditioning chamber that is used for training. However, no shock and no tune cues were delivered.

During the cue test the animal is placed in a modified chamber, i.e. without the metal bars on the floor, which is replaced a normal floor, and with modified walls. The tone cue/s is/were presented as in the training session.

Rodent behaviour is recorded by a video camera and automated tracking is performed using EthoVision® XT 8.5 software (Noldus Information Technology, Wageningen, The Netherlands). The following behavioural elements were recorded: total freezing time, number and duration of freezing episodes, locomotor behaviour, escaping to one corner of the cage [5].

Ethical statement

The experiments received authorisation from the Italian Ministry of Health, and are conducted following the principles of the NIH Guide for the Use and Care of Laboratory Animals, and the European Community Council (86/609/ EEC) directive. Mice are housed and used according to current European Community rules. Experiments on mice are approved by the research committees from the University of Verona and Italian National Institute of Health.

References

1. Wehner J.M., Radcliffe R.A. (2004). Cued and Contextual Fear Conditioning in Mice. *Current Protocols in Neuroscience* 8.5C.1-8.5C.14.

2. Le Doux J.E. (2000). Emotion Circuits in the Brain. *Annual Review of Neuroscience* **23**, 155-184.

3. Maren S. (2008). Pavlovian fear conditioning as a behavioural assay for hippocampus and amygdala function: cautions and caveats. *European Journal of Neuroscience* **28**, 1661-1666.

4. Phillips, R.G. Le Doux, J.E. (1994). Lesions of the dorsal hippocampal formation interfere with background but not foreground contextual fear conditioning. *Learn. Mem*. **1**, 34-44.

5. Gerlai R. (1997). Contextual learning and cue association in fear conditioning in mice: a strain comparison and a lesion study. *Behavioural Brain Research* **95**, 191-203.

The Coordination between the Direction of Progression and Body Orientation in Normal, Alcohol- and Cocaine Treated Fruit Flies

Ilan Golani[1], Anna Gakamsky[1], Efrat Oron[1], Dan Valente[2], Partha P. Mitra[2],

Daniel Segal[3], Yoav Benjamini[4]

[1]*Tel Aviv University, Department of Zoology and Sagol School for Neuroscience, Israel. Ilan99@tau.ac.il, anna.chernogorova@gmail.com, oronefra@gmail.com*

[2]*Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA. mitra@cshl.edu, danvalente@gmail.com*

[3]*Tel Aviv University, Molecular Biology and Biotechnology Department and Sagol School for Neuroscience, Israel. dsegal@post.tau.ac.il*

[4]*Tel Aviv University, Department of Statistics and OR, and Sagol school for Neuroscience, Israel. ybenja@tau.ac.il*

Background

In studying the gene/brain/behavior relationship, detailed quantitative descriptions of the behavior of the animal's behavior is required. Especially important are descriptions of behavior in terms of the dynamics of the coordination between the direction of progression of the animal and the animal's body orientation, for these provide fundamental insight into motor control. For mouse locomotor behavior these measures have been well-developed, yet such descriptions are still sparse for Drosophila, the model system of choice for behavior geneticists. Here we present a dynamic description of the coordination between these two variables, as they change throughout different temporal segments of the motion.

Methodology/Principal Findings

Using intrinsic statistical and geometrical properties of a fly's movement, we have uncovered six fundamental modes of motion related to translation and rotation of a fly's body. Assignment of the frame-by-frame instantaneous movement into one of the six modes, followed by clustering and algorithmic classification, allows us to quantify the behavior in terms of the proportions and dynamic sequencing of each mode. The analysis uncovered that for normal flies, the angle between the direction the animal walks and the direction it faces is quite small. Under the influence of alcohol and cocaine, however, the angular interval between these two variables increases.

Conclusions/Significance

For several decades the representation of movement, indispensable for studying the interface between genes brain and behavior, suffered from the use of ad hoc building blocks such as "behavior patterns" or "response categories" assumed to be performed unblock by the whole organism, one at a time. The ethological and psychological schools using such representations, were based on expert (indeed, subjective) decisions. In the vast majority these alleged building blocks could not be shown to have a physiological reality in the brain; the school using them has been described as the "school of immaculate perception". These ad hoc units were used as "black boxes" and their variable kinematic content has been disregarded. Disregarding the coordination between translation and rotation, for example, sidestepped the problem of coordination, which is at the heart of the brain/behavior interface.

The huge progress in tracking and storing technology allowed the recording of continuous kinematic variables such as the location and orientation of the organism, thus bringing about a most significant change, allowing, for the first time, the study of coordination between these variables in long stretches of behavior lasting hours or even days. This time scale is necessary for the study of e.g., drug effects on behavior, or for the phenotyping of genetic mutants exploratory and locomotor behavior. The pioneers utilizing this technology use, however, the continuous data they obtain to reinstate the static old time behavior patterns, this time based on explicit criteria, but ignoring again their variable content. These patterns are useful for scoring the behavior of closely related phenotypes, or animal models and their corresponding wild types, or males and females. They are useless, however, for comparing organisms across remotely related taxonomic groups or across apparently very different drug effects on behavior.

What signifies the present study is that it not only tracks kinematic variables continuously, but also sticks to a dynamic representation of their coordination all the way. In other words, both initial and final results are formulated in a dynamic fashion, the building blocks of behavior are segmented on the basis of their intrinsic dynamics, and the results summarize their variable kinematic content. We select 3 apparently very different behavioral preparations – that of intact flies, of flies with alcohol and of flies with cocaine – and show that they lie on a continuum:

1. Examination of the dynamic management of the angular interval between the direction of translation of a fly and the direction it faces yields 6 modes of coordination (Figure 1). These modes are intrinsic. They reflect the actual dynamics of locomotor behavior and are not superimposed from without.

2. The 3 groups differ in the dynamics of management of the angular interval between these two variables. In intact flies the angular interval is small and short lived (high gain), with alcohol it increases (smaller gain) and with cocaine it is vastly increased (smallest gain and hence large misalignment between the direction the animal walks and the direction it faces.

Description of the six modes

Fixed-Front-on-Straight-Path (Figure 1a). This is one of the two most common modes of untreated movement. Flies usually move along relatively straight paths over long distances with close alignment of the body axis and the direction of progression. Net sideward/backward translation episodes are very short in duration and path length in untreated flies (~0.12 s and no more than 5 mm long).

Rotation-on-Straight-Path (Figure 1b). At the start of this mode, the fly's body is typically misaligned with the direction of progression. In the course of the mode, the body rotates toward the direction of progression, typically converging to the same direction. This rotation is superimposed on forward, sideways, or backward progression, all performed along a relatively straight path traced by the center of the fly’s body.

Fixated-Front-on-Curved-Path (Figure 1c). The fly maintains a more-or-less fixed orientation while moving on a curved path. We use the term fixated rather than fixed to highlight the active, compensatory fixation, as opposed to the fixed orientation of a fly whose body axis is aligned with the direction of progression on a straight path. This mode is typically performed during a short intermediate state between Fixed-Front-on-Straight-Path and Rotation-on-Curved-Path modes.

Rotation-on-Curved-Path (Figure 1d). In the course of this mode, the body of a fly typically rotates toward the direction of progression, the rotation and direction of displacement sign being the same. This is the second of the two most common modes in untreated fly locomotor behavior.

Lingering (Figure 1e). Lingering episodes include at least one arrest and may also include small below-threshold displacements. Lingering duration ranges between short interruptions in movement and long (presumably sleeping) episodes.

Rotation-in-place (Figure 1f). Rotation of the fly's body axis around a vertical axis located at the fly's body center is mostly performed in untreated flies between two lingering episodes.

While the results are presented in the context of fly behavior, the implications relate to any dynamic description of movement in any organism having a rigid body with bilateral symmetry.

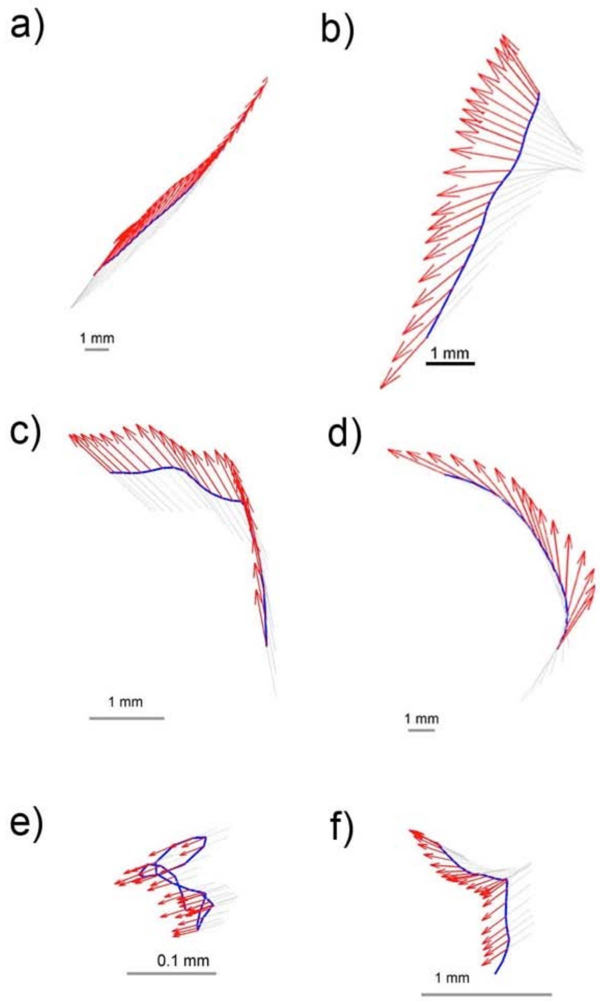


Figure 1. Examples of the six elementary modes of fly locomotor behavior. a) Fixed-front-on- Straight-Path ('A'), b) Rotation-on-Straight-Path ('B'), c) Fixated-Front-on-Curved-Path ('C'), d) Rotation-on-Curved-Path ('D'), e) Lingering ('L') and f) Rotation-in-place ('R'). Quiver plots: blue lines represent the path traced by the fly's centre. The arrows represent the orientation of the fly's body axis.

Acknowledgements

Flies were tracked by F-track[1]. The study was supported by an Israel Science Foundation (ISF) grant belonging to the Israel Academy of Science.

References

1. Valente, D., Golani, I., Mitra, P.P. (2007). Analysis of the trajectory of *Drosophila melanogaster* in a circular open field arena. *PLoS One* **2**, e1083.

Automated Measurement of Spontaneous Surprise

Bart Joosten[1], Eric Postma[1], Emiel Krahmer[1], Marc Swerts[1], Jeesun Kim[2]

[1]*TiCC, Tilburg University, Tilburg, The Netherlands*

[2]*MARCS Auditory Laboratories, University of Western Sydney, Sydney, Australia*

1. Introduction

Humans exhibit a rich variety of facial expressions reflecting their emotional state. As pointed out by [6], automatically measuring facial expressions to infer emotional states is a challenging task, amongst others because there are no uniquely defined representational units of facial expressions for the basic emotions. In natural circumstances, spontaneous facial expressions tend to be subtle and highly context-dependent, which makes their formal representation very difficult. We have adopted an approach that performs computational analyses on video data acquired from carefully designed experiments for the measurement of facial expressions (see, e.g. [4]). This paper reports on our approach to measure spontaneous surprise. Two main obstacles in automatically measuring spontaneous surprise are of a behavioral and a technical nature. The behavioral obstacle is to elicit and identify spontaneous surprise. The technical obstacle is to develop a computational method to measure the facial expressions associated with spontaneous surprise. In order to deal with the first obstacle two of the authors (MS and JK) developed a novel experimental paradigm to elicit spontaneous surprise in humans. The paradigm is described in Section 2. The second obstacle is addressed by the Spontaneous Surprise Measurement method presented in Section 3. The setup for the experimental evaluation of the method is described in Section 4 and the results are reported in Section 5. Finally, Section 6 is a concluding discussion.

2. Eliciting nonverbal expressions of surprise

According to [3], surprise is prototypically displayed by three facial actions: raising of the eyebrows, opening of the mouth, and widening of the eyes. However, when researchers try to elicit surprise expressions in participants in experimental studies, the prototypical visual actions are rarely shown (e.g., [7]). This may be due to the lack of a proper experimental setting. We tried to improve the setting by designing a memory task experiment in which lexically similar utterances are elicited in a neutral and in a surprise context. To keep participants unaware of our intentions, they are presented with a cover story that the experiment measures the how context and reading words aloud affects the number of words memorized. The experiment continues in three stages: imagination, verbalization, and recall.

1. Imagination stage. Participants imagine words that fit a specific context (e.g. “organs of the human body” in the neutral condition versus “favorite food items among Dutch children” in the surprise condition)
2. Verbalization stage. Participants read aloud each of a sequence of 10 words (of either of the two contexts) displayed on a screen. In both conditions, the target word “liver” is one of displayed words.
3. Recall stage. Participants have to recall as many words as possible.

Crucially, we elicit the target word (liver) in two conditions: a neutral condition in which the word clearly fits in the context “organs of the human body” and a surprise condition in which the word is highly unexpected “favorite food items among Dutch children” (liver is clearly not a favorite food item for Dutch children). Participants see the words on a screen using a hidden camera that is positioned behind the computer screen so that facial expressions can be clearly captured. We made such hidden recordings of 27 Dutch participants all of Dutch descent (students at Tilburg University who took part for course credit). In most participants, the paradigm results in clearly different behavior in both experimental conditions (neutral versus surprise).

3. The Spontaneous Surprise Measurement method

The Spontaneous Surprise Measurement (SSD) method aims at measuring surprise by focusing on the first of the three actions proposed by Ekman [3]: the raising of the eyebrows. To that end, the SSD method consists of three steps: the identification of landmarks, the texture analysis of the eyebrow region, and the classification of surprise. The landmark-identification step is realized by using the Constrained Local Model [8]. Given a video frame, it returns the locations of a number of predefined facial landmarks, such as, the nose tip, corners of the eyes, and the eyebrows. The texture analysis is restricted to an image patch covering the facial region of the eyebrows and is performed using a multi-scale Gabor filter-bank [1]. For each image position (pixel), the Gabor filter-bank returns N x M energy values representing the presence of oriented visual structure of a certain thickness (spatial frequency), where N represents the number of orientations and M the number of spatial frequencies. Finally, the classification maps the (aggregated) energy values onto the binary classes *frown* and *non-frown* and subsequently to *surprise* and *no surprise*.

4 Experimental set-up

In this section we describe the data set (4.1) and the experimental settings of the three components of the SSM method (4.2-4.4).

4.1 Data set

Our dataset was collected in the experiment described in Section 2. All participants were recorded on frontal-face video in 24-bit color (resolution 640 by 480 pixels at a frame rate of 30 fps). The video fragments selected for our data set started immediately following the presentation of the neutral and surprise-evoking target. This resulted in a set of 54 videos (27 for each condition). The recordings are approximately four-seconds long, varying from 119 to 122 frames in total. Figure 1 shows an example of one participant in both conditions, directly after verbalizing the target word.

Careful visual inspection of the video fragments revealed five distinctive facial expressions in the surprise condition, viz. (1) eyebrow frowning, (2) eyebrow raising, (3) widening eyes (4) mouth opening, and (5) brief head retraction. Of these expressions, (2), (3) and (4) correspond to Ekman’s actions. The presence and prominence of each of the expressions differed from participant to participant, but eyebrow frowning (1) was the most prevalent expression overall. Therefore, we manually annotated the frames for all participants who displayed the eyebrow frown (11 in total). The frowns had an average length of 17.1 frames σ^2=16.1. In addition to the labeling of frown and no-frown frames, we labeled consecutive sequences of frames as neutral or surprise. All frames are labeled neutral, except for those consisting of three consecutive frames labeled "frown". These sequences are labeled "surprise".

Figure 1. Illustration of a participant recorded in the neutral condition (left) and surprised condition (right).

4.2 Landmark identification

We employed FaceTracker as described in [8] on each video in the data set. The FaceTracker software contains a pre-trained CLM which automatically places a grid of 66 facial landmarks on each video frame. The output of the analysis is a collection of 66 coordinates per frame. After manual inspection 8 videos appeared to have insufficient (correct) fittings. These were excluded from further analysis. We selected a region of interest (ROI) based on the center between the inner eyebrow landmarks. At this center we extract a patch of 201 by 201 pixels. To compensate for in-plane rotations of the head, in each frame the patch is rotated so that both eyes are centered at the same row of pixels. Subsequently, the rotated patch is down-sampled to a size of 51 x 51 pixels.

4.3 Texture analysis

Frown detection is based on the output of Gabor filters convolved with the patches extracted around the inner eyebrow landmarks. For this purpose we constructed a 6 by 12 log-Gabor filter bank consisting of filters with 6 scales and 12 orientations. We used Kovesi's [5] Matlab tools to compute the convolutions (minimum wavelength 3 pixels, scaling factor of 1.8 between successive filters, and filter bandwidth 1.45). A margin of 5 pixels wide is excluded from the convolved image patch to remove image border artifacts. The Gabor energy for each scale and orientation combination is computed by summing over the entire convolved image. As a consequence, our feature space consists of 72 dimensions, each representing the total Gabor energy of a filter at a specific scale and orientation. A large value in any of these scale-orientation combinations signals the presence of a visual contour at a particular scale (thickness) and orientation. Given the specific thickness and orientation of frowns, we expect the Gabor filters with corresponding scale and orientation to generate large energy values.

4.4 Classification

The manual annotation of frown and no-frown frames resulted in 205 frown frames and 5343 no-frown frames. In order to automatically differentiate between the two classes of frames we explored two techniques for classification, viz. Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM). Training and testing was performed on a leave-one-subject-out validation scheme. This entails that we first trained our classifiers on a balanced number of frames (i.e., all 'frown frames' and an equal number of randomly selected 'non-frown frames') of both classes, excluding one participant's frames and subsequently tested the classifiers on the left out participant. This procedure is repeated for each participant and the summed score of all the iterations represents the classifier's predictive value. Since we have only 205 frown frames each training set contains at most 410 frames.

LDA linearly maps labeled instances from our feature space to a 1-dimensional space. The linear transformation matrix is constructed in such a way that it tries to maximize the separation between the classes. A maximal separation in one dimension, readily leads to the identification of a threshold value separating both classes. Unseen instances are mapped to the same 1-dimensional space and classified using the aforementioned threshold.

The SVM classifier operates on the 72-dimensional Gabor feature space. We used the freely available LIBSVM [2] as our SVM implementation. We choose a radial basis function as a kernel and used a grid search to find optimal values for the parameters C and γ. In our case $C = 2^{15}$ and $\gamma = 2^{-5}$.

We evaluate the performance of surprise classification performance of the LDA and SVM classifier separately.

5 Results

Application of the SSM method to the videos in the data set yielded results at three levels, corresponding to the three components of the method. At the level of landmark detection, visual inspection revealed the FaceTracker to be quite successful in adequately positioning the 66 landmarks at the faces of the participants. In some cases, there were some misalignments along the cheek regions, but because the texture analysis does not rely on these

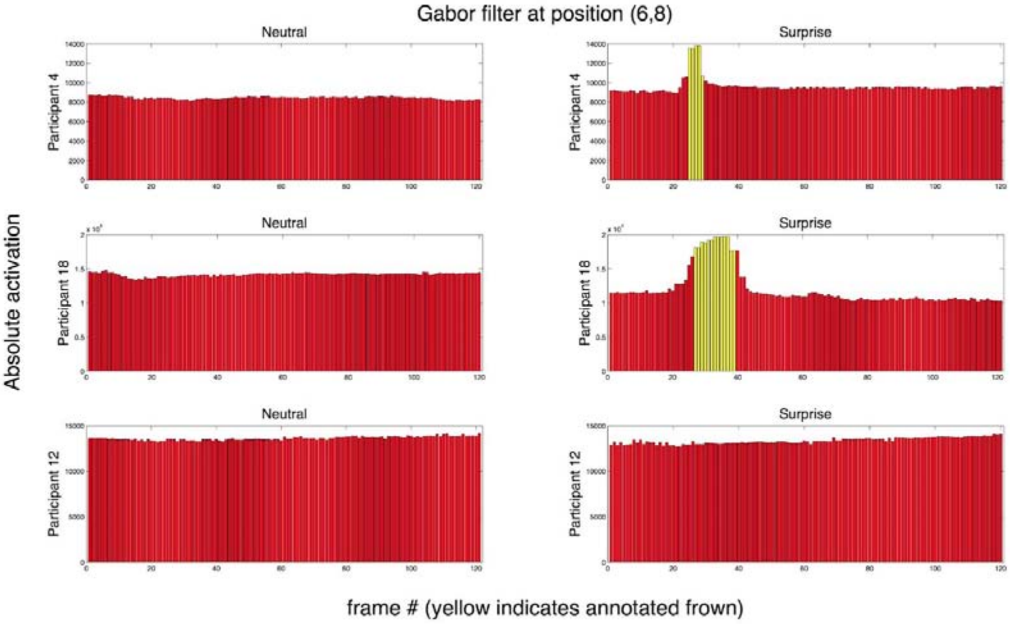


Figure 2. Gabor filter activation of filter at position (6,8) of three participants in both the neutral and surprise condition. Participants 4 and 18 clearly show a distinctive activation at the manually annotated 'frown frames' (depicted in yellow), while participants 12 lacks a frown of surprise.

landmarks, they do not affect the overall performance of the SSM method. The performance of the texture analysis component is illustrated in Figure 2 for the video sequences of three participants in both conditions: participant 4 (top row), participant 18 (middle row), and participant 12 (bottom row). The six panels in the figure show the energy (output) value of one of the Gabor filters (the filter at orientation 6 and scale 8) as a function of frame number of the video. The videos of the neutral and surprise conditions are displayed on the left and right, respectively. All frames during which participants frown have been manually labeled and are indicated by yellow bars, all others by red bars. The frowning of participants 4 and 18 is associated with an elevated energy value of the Gabor filter. As a consequence, the filter output appears to signal frowning. Participant 12 does not frown in the surprise condition and, hence, does not result in an elevated Gabor energy value.

An impression of the current leaving-one-subject-out performances of the SSM method as determined with the LDA and SVM classifiers is given in Tables 1 and 2. Tables 1a and 1b contain the confusion tables listing the frown-detection and surprise-detection results respectively for the LDA classifier. Similarly, tables 2a and 2b contain the detection results for the SVM classifier. Overall, the confusion tables show that no-frowns are often detected as frowns and that neutral sequences are often detected as surprise sequences. The higher scores on the main diagonal indicate that the SSM method is reasonably successful in detecting spontaneous surprise.

Tables 1a and 1b. Frown and surprise classification performance for the LDA classifier.

		predicted	
		no-frown	frown
actual	no-frown	4040	1303
	frown	78	127

		predicted	
		neutral	surprise
actual	neutral	13	10
	surprise	5	18

Tables 2a and 2b. Frown and surprise classification performance for the SVM classifier.

		predicted	
		no-frown	frown
actual	no-frown	3677	1666
	frown	72	133

		predicted	
		neutral	surprise
actual	neutral	13	4
	surprise	10	19

6 Concluding discussion

We have developed a method for the measurement of Spontaneous Surprise that is reasonably successful at detecting frowning as part of one of the key facial actions for surprise (raising the eyebrows). Admittedly, our method was evaluated solely on Western-European participants. (An interesting experiment would be to evaluate our method on subjects with a different ethnical background). Further optimizing the texture-analysis and extending our method with motion information of the facial actions may lead to an improvement in performance. However, our aim is to yield improvement by incorporating domain knowledge about the facial features signaling surprise. More specifically, we want to extend the SSM method by including measurements of the two other facial actions proposed by Ekman, i.e., opening of the mouth, and widening of the eye. In addition, "mouth opening" and "brief head retraction" will also be examined.

References

1. Berezhnoy, I.J., Postma, E.O., Herik, H.J. van den (2007). Computer Analysis of Van Gogh's Complementary Colours. *Pattern Recognition Letters* **28**, 703-709.
2. Chang, C.-C., Lin, C.-J. (2011). LIBSVM : a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* **2**,1-27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm
3. Ekman, P., Keltner, D. (1997). Universal facial expressions of emotion. In U. Segerstrale, P. Molnar (Eds.) *Nonverbal communication: Where nature meets culture*, 27-46.
4. Joosten, B., Amelsvoort, M.A.A. van, Krahmer, E.J., Postma, E.O. (2011). Thin slices of head movements during problem solving reveal level of difficulty. In G. Salvi, J. Beskow, O. Engwall, S. Al Moubayed (Eds.), *Proceedings of the international conference on audio-visual speech processing* (AVSP 2011) 85-88.
5. Kovesi, P.D. (2012): *MATLAB functions for computer vision and image analysis* http://www.csse.uwa.edu.au/~pk/research/matlabfns/. Accessed February 2012.
6. Pantic, M., Rothkrantz, L. (2000). Automatic analysis of facial expressions: the state of the art, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**, 1424-1445.
7. Reisenzein, R., Bördgen, S., Holtbernd, T., Matz, D. (2006). Evidence for strong dissociation between emotion and facial displays: The case of surprise. *Journal of Personality and Social Psychology* **91**, 295-315.
8. Saragih, J., Lucey, S., Cohn, J. (2011) Deformable model fitting by regularized landmark mean-shifts, *International Journal of Computer Vision* **91**, 200-215.

Controlled Game-Based Stress Manipulation

Benny van der Vijgh, Robbert Jan Beun, Peter Werkhoven

Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands
b.h.vandervijgh@uu.nl

Introduction

The goal of the research project presented in this paper is two-fold. First, we want to conduct fundamental research into the manipulation of stress using digital games, to which end we will construct an experimental set-up utilizing an adaptive digital game as a stressor. Second, we want to utilize this set-up to gain insight into the effects of different levels of stress on the fronto-striatal circuits of the brain (FSC). This insight is expected to contribute to the development of novel treatments of stress-related symptoms associated with disorders of the FSC[1]. Currently the project is in the first phase in which we pursue the first goal; fundamental research into stress manipulation. This is done by constructing and evaluating an experimental set-up allowing controlled manipulation of stress presented to human individuals by means of an adaptive digital game. To this end we first reviewed which biosignals and corresponding measurements met the requirements to be usable in our set-up. Next we conducted a meta-analysis analyzing which digital game and study methodology characteristics are related to the stressor function of a digital game. Finally, based on the data from both the review and meta-analysis, we are currently constructing a feedback model that automatically regulates the level of stress presented in the set-up, depending on a particular desired level. This is achieved by utilizing the continuous measurement of the player's stress related biosignals, comparing these signals with the desired level of stress and real-time adaptation of the digital game characteristics. In this paper, we will briefly discuss some of the underlying assumptions in the project, the main components of the experimental set-up that will be constructed and the current status of the project.

Stress and Gaming

In order to capture the diverse stressor characteristics a digital game can present and because the physiological stress response is the most prominent response in our set-up, we use a broad and physiologically-based definition. In our research stress is considered as any stimulus, internal or external, that presents a (perceived) challenge to the homeostasis of a biological system and the stress response, either physiological, psychological and/or behavioural, aimed at restoration or maintenance of the homeostasis (cf. Newport and Nemeroff (2002), [4]). The stimulus presenting the challenge to the homeostasis will be referred to as the stressor and the response to this challenge as the stress response. Moreover, we refer to variables influencing the relation between the stressor and stress response, such as age or gender, as intervening variables. In general, stressors can range from physical stimulation, such as a cold pressor task where participants place their hands in an ice water container, to psychological stressors, such as the performance of a public speech [3,5]. Physiological stress responses pertain to biosignals, such as cortisol levels, heart rate variability and electrodermal activity; psychological stress responses include alterations in affect and cognition, such as aggression and hostility.

Games have consistently been proven capable of eliciting both physiological and psychological stress responses (for example, see [1] and [2]). As we aim at real-time manipulations of the stressor, a digital game is a very suited stressor since it allows for real-time adaptations of the stressor parameters (without the participant directly noticing), which is harder in other stressor modes such as video, images or speaking tasks. Furthermore it is very flexible: we can present different stressor types such as aversiveness and social-evaluative threat through one construct, broadening our range of possible manipulations. Key elements in digital games are players, (inter)action, environment, goals and rules. An advantage of digital games is that many, or even all, of the key elements can be taken over by computer technology. A computer may simulate (parts of) the game environment,

[1] The presented project is part of the program *Brain & Cognition – Societal Innovation in Health Care, Education and Societal Safety* and funded by The Netherlands Organization for Scientific Research (NWO).

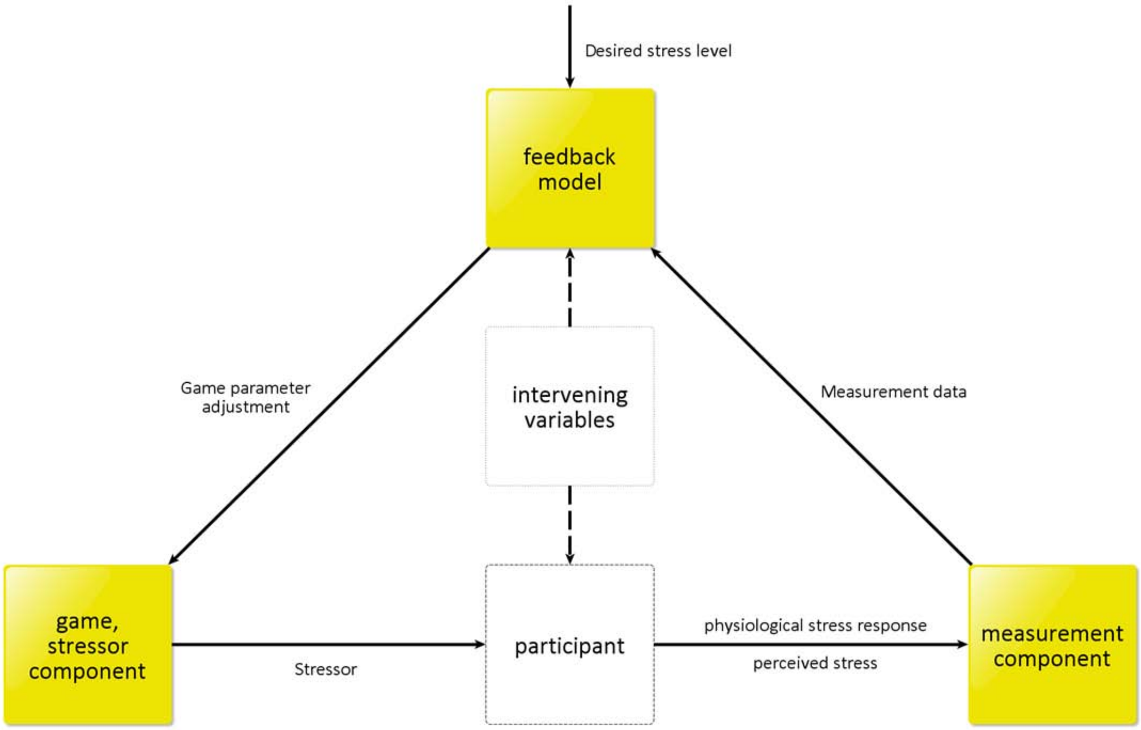


Figure 1. Overview of the set-up.

the interaction or even one or more players. Just like conventional games, digital games may offer various types of challenge that simulate the performance of particular activities, ranging from physical and mental activities (e.g. hitting another player, taking business decisions) to inactivity, such as meditation and relaxation for a particular period of time. In our project, we assume that only a particular set of game characteristics present a challenge for the homeostasis of the player and, therefore, function as a physiological or psychological stressor that causes a particular stress response.

Set-up

As depicted in Figure , the set-up we are constructing consists of three components. The first component, the digital game, represents the stressor and is presented by the lower left square. The second component, depicted by the lower right square, is the measurement component, which consists of a set of measurement devices continuously measuring diverse biosignals for detection of the stress responses and the perceived stress indicated by the participant[2]. The third component, that is the top square, depicts the feedback model. The model takes the current stress state of the participant in the form of the values of different biosignals and perceived stress, together with the desired level of stress, and adjusts the game characteristics based on this information. For example, when the measured values of the biosignals and perceived stress indicate that the participant is more stressed than the desired stress level, the game characteristics are adjusted in such a way that relieves the intensity of the stressor, hereby relieving the stress response in the participant. Throughout this process of manipulating the stress state of the participant we consistently safeguard the coherence between the stressor and stress response by using models of the underlying stress systems and by controlling for the relevant intervening variables for the specific stressor-stress response relations (i.e. variables influencing this relation, such as gender, psychological, age or psychological factors such as neuroticism), based on these underlying models.

[2] Perceived stress is included in the measurements to further investigate the relations between physiological stress responses and perceived stress. Since perceived stress is only one of the many signals being used as input, the unwanted phenomenon where participants infer the effects the reported perceived stress has on the game parameters are very slim.

The benefits from this set-up are manifold:

- By allowing continuous control and manipulation of the stressor in a real-time feedback loop different types of stress research can benefit from this set-up. Examples include research into prolonged stress exposure in which case the stressor can be adjusted to present a constant stressor over time, or memory and cognition research where data can be gathered in an extremely controlled environment.

- The set-up may be useful for stress therapy containing exposure to stressful stimuli, for example, such as used in certain Post Traumatic Stress Disorder (PTSD) therapies, since the presentation of stressors can now be manipulated and controlled to a far greater extent than was previously possible.

- The set-up elaborates on existing Brain-Computer-Interfaces (BCI's), since the set-up already contains the essential components of a BCI, it can be easily adapted to be incorporated in existing BCI's, giving essential information regarding the current stress state of the user.

- Gaming industry can utilize the set-up to for sensing of stress responses and automatic adjustment of game parameters to maximize the user experience, immersion and flow of the player.

Status of the project

We have first focused on the measurement component by determining the requirements on the biosignals and the corresponding measurements that need to be met in order for these signals to be usable in our set-up. These requirements are the following. First, the biosignals must respond to the stressor consistently over repeated trials, both quantitatively and temporally. Second, the measurement technique and the corresponding biosignal have to show low response latency, meaning that a change in stressor intensity should be reflected in a biosignal without too much time delay. Third, in order to have continuous insight into the stress state of the participant, the measurement should be applied continuously or at least as frequent as necessary for real-time adaptation by the feedback model. Fourth, the measurement should not induce an additional stress response or disturb the experiment. The fifth requirement is not a requirement on the set-up per se, but is needed because in the second phase of the research we will utilize this set-up to gain insight into the effect of different levels of stressor intensity on the fronto-striatal circuit of the brain (FSC) by using the set-up in combination with fMRI. Therefore, the fifth requirement is that the measurement must be applicable inside fMRI. From this set of requirements it follows, for instance, that measuring cortisol levels falls outside the range of measurable stress responses, because the signal often responds inconsistently and shows large time delays of about 20-40 minutes to a particular stressor. Stress response measures which will be used in our research are, for instance, heart rate and heart rate variability, blood pressure, electrodermal activity and electroencephalographic measures. Secondly we have looked at the stressor component to determine the relevant digital game characteristics for eliciting stress responses. To this end, we have conducted a meta-analysis on 59 studies utilizing a digital game stressor and measuring stress responses. The initial analyses show that several specific digital game characteristics such as the level of aversiveness, the level of realism and diverse methodological characteristics such as the data reduction method and baseline measurement type were found to be significant moderators of stress responses such as heart rate and blood pressure.

Currently we are finishing the meta-analysis and constructing the feedback model based on the results found in the analysis. For this feedback model we use the multiple linear regression model constructed in the meta-analysis, hereby utilizing the gathered data of a large body of studies to infer the relations between game characteristics and stress responses. It will be the first time that the adaptations in an adaptive stressor game will be based on such a detailed and extensive meta-analysis on game characteristics, as such an analysis has not been done before. This method provides added value, since it allows to create a more effective stressor and allows for the modelling of very specific relations between game characteristics and physiological stress responses. For example, the impact of the level of aversiveness in the digital game on heart rate is independently modelled of the impact of this characteristic on the other responses such as electrodermal activity.

References

1. Dickerson, S.S., Kemeny, M.E. (2004). Acute stressors and Cortisol Responses: a theoretical integration and synthesis of laboratory research. *Psychological Bulletin* **130**(3), 355-391.

2. Miller, S., Ditto, B. (1988). Cardiovascular responses to an extended aversive video game task. *Psychophysiology* **25**(2), 200-206.

3. Musante, L., Raunikar, A., Treiber, F., Davis, H., Dysart, J., Levy, M., et al. (1994). Consistency of children's hemodynamic responses to laboratory stressors. *International Journal of Psychophysiology* **17**, 65-71.

4. Newport, D.J. Nemeroff, C.B. (2002) Stress. In: *Encyclopedia of the Human Brain*, Vol. **4**, 449-462.

5. Stegeren, A.H., Wolf, O.T., Kindt, M. (2008). Salivary alpha amylase and cortisol responses to different stress tasks: Impact of sex. *International journal of Psychophysiology* **69**, 33-40.

A Comparison of Two Methods to Assess Mobile Hand-Held Communication Device Use

S. Berolo[1], I. Steenstra[2], B.C. Amick III[2,3,4], R.P. Wells[1,2,3]

[1]*Department of Kinesiology, University of Waterloo, Waterloo, Canada.* [1]*sberolo@uwaterloo.ca*

[2]*Institute for Work and Health, Toronto, Canada.*

[3]*Centre for Research Expertise for the Prevention of Musculoskeletal Disorders, University of Waterloo, Waterloo, Canada*

[4]*School of Public Health, University of Texas Health Science Center at Houston, Houston, USA*

Abstract

The purpose of this study was to: 1) examine the agreement between self-reported measures of mobile hand-held device use and direct measures, and 2) understand how respondents think about their device use when completing the questionnaire. 35 participants were recruited from a Canadian university to participate. The questionnaire was previously developed to estimate six types of usage potentially related to musculoskeletal disorders in cell phone users. Direct measures were collected on a custom software usage logging application resident on the participants' phone. Bland-Altman plots demonstrated that, generally speaking, participants' self-reports overestimated their logged usage; overestimates tended to be low at low mean usage times, and became more variable as mean usage time increased. Six themes were identified from the interviews: 1) types of mobile hand-held devices, 2) tasks attributed to different categories of device use, 3) their thought process used to arrive at usage, 4) ease of reporting usage, 5) physical interaction with device, and 6) completeness of list of categories of device use. The preliminary findings indicate that it may be challenging for respondents to provide good self-reports of usage. The root of this challenge does not appear lie with the design of the questionnaire; rather, it may be attributed to the respondents' difficulty in estimating usage, partly due to the variability of device use both within a day and a week.

Introduction

Despite concern that musculoskeletal symptoms are associated with mobile hand-held device use [2], there is a lack of knowledge concerning the use and exposures associated with these devices. To understand the use of mobile technology and determine if use is associated with any musculoskeletal or other health effects, accurate measures of how these devices are used is required. Previous research has examined agreement between self-reported computer, keyboard, and mouse use and direct measures of use collected using computer logging software or observation [3, 4], however no such literature exists comparing self-reported hand-held device use with directly measured use. Before progressing with further studies of hand-held device use, we thought it important to better understand the accuracy of the exposure measures collected by a questionnaire. The study described herein was designed to: 1) examine agreement between self-reports collected by a previously developed questionnaire [2] and direct measures of use collected using a custom software logging application, and 2) examine the usability of the questionnaire by conducting interviews to better understand how respondents think about their device use.

Methods

Data collection. 35 students/recent graduates (17 female and 18 male) under the age of 35 were recruited from a Canadian university to participate. Inclusion criterion included being an Android device user. There was no exclusion criterion. The study was approved by the Office of Research Ethics at the University of Waterloo, Ontario, Canada.

Mobile hand-held device use on a “typical” day last week was assessed using: 1) a paper and pencil version of the previously used questionnaire [2], and 2) the custom Android device logging software. The usage logging program recorded participants’ hand-held device use on second-by-second basis. It recorded generic information on the applications used but did not record any personal or identifying information. For example, it would show that an email or messaging application was used, but would not display contents of emails or messages, and would not record voice or keys pressed. It also recorded the number and duration of phone calls made and received, and the number of text messages sent and received. It was developed for the Android platform due to the relative ease of accessing usage information on this operating system.

The questionnaire was previously developed to estimate six types of hand and upper limb usage, potentially related to musculoskeletal disorders among cell phone users. These six categories of use were chosen to represent different intensities of thumb and upper limb usage [2]. The questionnaire assessed how much time on a typical day last week participants spent using their mobile hand-held device(s) for: 1) emailing, texting, and instant messaging; 2) scheduling (calendar, appointments); 3) internet browsing; 4) making phone calls and talking on the phone; 5) listening to music, watching videos, and taking pictures; and 6) gaming using a mobile device. A “typical” day referred to time at work, as well as time away from work, at home, and with friends. Participants met the researchers on two occasions; on their first visit they were shown how to download the logger software onto their device, and on their second visit, one week later, they uploaded the data log from their device and filled out the questionnaire. Participants also provided demographic data and information regarding the make and model of their device(s). A subgroup of respondents, 17 out of 35, then participated in an interview.

Interviews were carried out with an aim to better understand of how respondents think about their device use. An *intensive interview* framework was used, and followed these general steps: 1) the respondent was reminded of the question and the response he/she gave, 2) the respondent was asked what he/she thought the question meant, 3) the respondent was asked to describe the thought process used to arrive at his/her answer, and 4) the respondent listened to the question, again, and was asked whether he/she would change his/her answer, and if yes, what he/she would change it to [1, 5]. Each interview took approximately 20 minutes.

Analysis. A program was developed using Visual Basic to summarize logged device usage. Each participant’s data record was broken down by calendar day so average daily use could be compared to self-reported use on a “typical” day. There are currently over 400,000 Android applications available [6], and the applications used by each participant were quite different. Since the logged data was to be compared with self-reports of six kinds of usage, after the questionnaire was complete, participants were asked to categorize the applications recorded by the logging program into one of the six categories of device use specified in the questionnaire. A seventh category called ‘other’ was added to this mapping task for those applications that did not fit into one of the six categories. Using the results from the mapping task just described, the time spent performing each of the six measures of device use found in the questionnaire on each of the 7 days prior to the participants’ second visit was calculated. Bland-Altman plots were used to examine agreement between logged and self-reported usage. The 17 interviews were transcribed. A preliminary coding scheme of major themes was developed by reading through all interviews once. “Themes” were defined as concept trends or patterns. The coding scheme was confirmed by two members of the research team upon detailed examination of each interview, and was used to block sections of each transcript that corresponded to a theme.

Results

Comparison of self-reported and logged device usage

Bland-Altman plots for categories of device use emailing, texting, and instant messaging (Figure 1) and gaming (Figure 2) are shown. These two categories of device use demonstrated strong relationship to self-reported pain in the upper extremity [2]. If there were perfect agreement between self-reported and

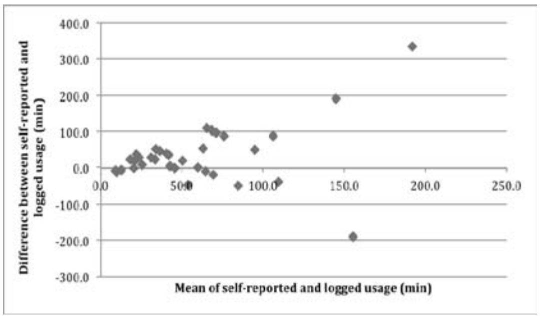


Figure 1. Bland-Altman plot for emailing, texting, and instant messaging on a "typical" day (N=35).

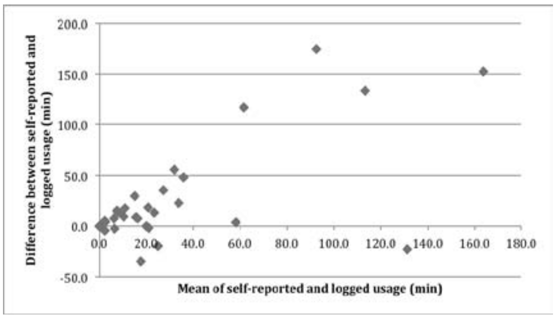


Figure 2. Bland-Altman plot for gaming on a "typical" day (N=35).

logged use, all data points in these plots would lie at 0 on the y-axis. Self-reports of zero usage time will lead to perfect agreement or an underestimation of logged use. In this population, generally speaking, for the six categories of device use, participants' self-reports overestimate their logged use; overestimates tend to be low at low mean usage times, and become more variable as mean usage time increases.

Themes from interviews

A total of six themes were identified from interviews: 1) types of mobile hand-held devices, 2) tasks attributed to different categories of device use, 3) their thought process used to arrive at usage, 4) ease of reporting usage, 5) physical interaction with device, and 6) completeness of list of categories of device use.

While all participants agreed those devices listed in the questionnaire are mobile hand-held devices (cell phones, personal digital assistants, and gaming devices), many had suggestions to be added to the list, tablets for example. A second theme was tasks participants attributed to each of the six categories of device use. For example, some participants considered use of social networking sites, like Facebook and Twitter, part of email, texting, and instant messaging, while others considered them part of Internet browsing. A third theme was the thought process to arrive at self-reported usage. A variety of processes were described by participants, including averaging, estimating based on numbers of events, thinking about time spent with or without a computer/laptop, and thinking about one's daily schedule. A fourth theme that surfaced was the difficulty some participants had in reporting device use on a typical day. One participant described this challenge by comparing phone use to computer use, and another explained it was difficult because each day was very different. A number of participants also spoke about the physical interaction they have with their device. For example, some participants indicated that they used both thumbs to type most of time, and others noted they do not use two thumbs for reasons related to device size and screen function.

Discussion

Participants' self-reports tended to overestimate their logged use; overestimates tended to be low at low mean usage times, and became more variable as mean usage time increased. This may suggest it is easier for respondents to provide self-reports that agree with logged usage when time spent performing a task is low. If respondents know, for example, that they do not perform a task, or that they spend little time performing a task, it may be more straightforward for them to provide a self-report than it would be if they performed that task more heavily. Some of the greater differences might be explained by inappropriate application matching. For example, the highest overestimate of logged use for gaming (Figure 2) likely results from the game Tetris being categorized by that participant as 'other' as opposed to 'gaming'. Such misclassification is a possible consequence of participants categorizing their applications; however, considering the number of applications available for download, the large variation in applications used by participants, and the fact that it is the agreement between self-reports and logged usage under study, application matching was the best way to set up the logged data for comparison with self-reports.

Themes from interviews were useful in identifying possible reasons for differences between logged and self-reported usage. Interviews highlighted that there can be blurring of boundaries between categories of device use. Should a respondent use both a mail application and an Internet browser application to check email, for example, the boundary between emailing, texting, and instant messaging and Internet browsing would blur. Although logged usage from such logging software has not been demonstrated as a gold standard, *per se*, the application likely provides a more refined representation of use compared to self-reports. That said, like similar exposure models in computing work, it might not simply be actual usage that is critical in predicting health outcomes. For example, holding a hand over a mouse may produce significant muscle loading that does not get captured in direct measurement.

The comparisons made in this study between self-reported and logged device usage are preliminary and made among a small, restricted university sample. A questionnaire may be the most convenient way to assess use in epidemiological studies. These findings, however, suggest it may be challenging for respondents of a questionnaire to provide self-reports of device use that do not overestimate logged use. The root of this challenge does not appear lie within the design of the questionnaire; instead, it may be attributed to the variability of time spent using a device. It remains unknown which method of estimating device usage is better predictive of musculoskeletal symptoms.

Conclusions

Going forward, research investigating the relationship between mobile hand-held device use and musculoskeletal symptoms and other health outcomes may want to consider using a logging application to examine exposure simultaneously with self-reported exposure and observational approaches to better understand the sources of any hazardous exposures.

References

1. Belson, W.A. (1981). *The Design and Understanding of Survey Questions*. Hants, England: Gower Publishing Co. Ltd.

2. Berolo, S., Wells, R.P., Amick, B.C.III (2011). Musculoskeletal symptoms among mobile hand-held device users and their relationship to device use: A preliminary study in a Canadian university population. *Applied Ergonomics* **42**, 371-378.

3. Douwes M., de Kraker H., & Blatter B.M. (2007) Validity of two methods to assess computer use: Self-report by questionnaire and computer use software. *International Journal of Industrial Ergonomics* **37**, 425-431.

4. Mikkelsen S., Vilstrup I., Lassen C.F., Kryger, A.I., Thomsen, J.F., & Andersen, J.H. (2007). Validity of questionnaire self-reports on computer, mouse and keyboard usage during a four-week period. *Occupational and Environmental Medicine* **64**, 541-547.

5. Presser, S., Couper, M.P., Lessler, J.T., Martin, E., Martin, J., Rothgeb, J.M. Singer, E. (2004). Methods for Testing and Evaluating Survey Questions. *Public Opinion Quarterly* **68**, 109-130.

6. Wikipedia. Google Play. <http://en.wikipedia.org/wiki/Android_Market>. Published 6 March 2012. Accessed 28 March 2012.

Assessment of Level of Professional Competence of Programmers

P.A. Orlov, V.M. Ivanov

Department of IT in design, St. Petersburg State Polytechnical University, Saint-Petersburg, Russia.
paul.a.orlov@gmail.com

Abstract

The objective assessment of the level of professional competence of programmers is an urgent task. In our research we try to find patterns in the behavioral aspect of programmer's work. Therefore we have conducted an experiment and collected gaze fixation data from professional programmers. During the analysis of the results we faced with different kinds of information representation for “back-end” and “front-end” developers. We assume that this factor may be an additional result of the experiment, namely, the programmers of varying specificity are formed of various behavioral situations. We showed that there is a general trend of the nature of the presence of sampling time gaze fixation under the supervision of the work of Web developers, suggesting the formation of a web developer similar behavioral aspects of the work. Finally we proposed the method to assessment of level professional competence of the programmers. As part of future research we propose to investigate the influence different factors(such as operational experience or IQ) to instrumental component of the competence of the programmer.

Keywords. Gaze fixation, eye-tracking, programmer, professional competence, SMI, information representation, Web developers, behavioral aspect of programmer's work.

Introduction

The quality level of the specialist is determined by its ability to solve problems of the real production. Employee's ability to solve actual problems in a rigidly fixed period of time (“here and now”) is what IT company are looking for. This ability is a combination of knowledge of personal and professional competencies. Professional competence as the facility and the ability to operate effectively within their profession, can be estimated by various methods: the biographical method, interviews, testing, and group methods for estimating staff, psycho-diagnostic technique. However, there is a problem to find the objective method of assessing the professional competence of a specialist in general and of a specialist in programming in particular. Competence, as the quality of the system can not be assessed on the basis of a simple summation of estimates of individual components, but there are integrating components of competence: instrumental (practical), communicational, axiological, knowledge-based.

The goal of this research is to identify patterns in the behavioral aspect of programmer's work in general and the assessment of the instrumental component of the competence of the programmer, in particular. A man meets a lot of problems for which solutions he has to use the conscious level of thinking in the profession. As soon as the employee will gain a positive experience of problem solving the process itself will be carried out automatically. It is assumed that the process of collecting information and to establish its compliance to problem will disappear from consciousness, such as the moments of doubt and hesitation. A chain of cognitive and motor operations will be run, instead of choosing process. Periods of doubt and hesitation will only occur in completely new and unexpected situations, during the user experience. Thus we can assume that professionalism can be estimated by the number of unexpected (extreme) situations throughout the working cycle. The question is how to objectively assess and hardware measure the amount of extreme situations in a person work. It could be the basis for the indirect estimation of instrumental (practical) component of competence.

Methodological study

Oculomotor activity is a necessary component of the mental processes which are associated with the procurement, conversion and use of visual information, as well as conditions and human activities. Therefore, recording and analyzing eye movements, we have the possibility of indirect estimates of internal mental activity. The nature of eye movements could help us to identify: the state of consciousness, the effectiveness of the decision by the user tasks and performance of individual stages of practice. Gippenreyter and her co-workers suggested that saccades restrict any "quantum" processes of regulation, even if the man does not use visual information [1]. According to the concept of automaticity, saccades do not occur in response to a stimulus, and are generated in a certain rhythm, like similar rhythm of the heart and respiration. This is confirmed by the fact that there are saccades with the eyes closed, and there are saccades even for blind people, when seemingly they are not needed. But in the case of mental work saccades followed by much less, intervals between them sharply, increasing by several times. When the mental activity level increase, intervals between saccades also increase (or intervals between saccades) as an arbitrary or involuntary nature. Many experimental studies suggest that most adequately the properties of cycles of mental regulation should reflect the duration of intervals between saccades (IBS or duration of eye fixation), and it can be used as a parameter to assess the complexity of mental processes. The problem boils down to, to determine what ranges of IBS correspond to different levels of mental regulation of the programmers in their daily work.

The experiment

Hypothesis. The positive programmers experiences of solving problems forms the similar behavioral aspects of their work(or automatic behavior).

Metrics. The degree of complexity of human activities associated with the duration of gaze fixation. The duration of gaze fixation it can be used as a parameter to assess the complexity of mental processes. Time of duration of gaze fixation (intervals between saccades – IBS).

Construction. Construction of the experiment was carried out with the following features:

- Participants in the experiment should be a professional domain, in this case is a web development;
- The context of the experiment should not change the habitual user environment as emotional and interior.

Providing the first aspect has been achieved, the fact that the experiment involved 10 programmers from the company FotoStrana (http://fotostrana.ru/). FotoStrana is one of the largest Russian Internet projects with an audience of more than 30 million. The web site is visited monthly by more than 12 million people, and the FotoStrana team creates its own new services. There are more than 50 programmers in the company whose staff are about 100 people. All volunteers are professionals in the field of programming (web development) from the perspective of the employer.

The problem for the providing usual context of the programmer was resolved by field experiment directly in the FotoStrana office (not in the laboratory). In our study, we use the mobile software and hardware system of registration of eye movements - iView X™ HED (Figure 1) produced by the German company SensoMotoric Instruments (SMI), owned by the department of Information technology in the design of the Institute of International Educational Programs St. Petersburg State Polytechnical University. The equipment was placed on a person's head with a helmet. An employee wearing a helmet does not feel the discomfort and get used to it within a few minutes. During calibration the test person was requested to look at the five points which were in the activity field (in our case, the corners of the monitor or monitors, if the programmer is working on two screens). System setup time and calibration usually takes about 5-10 minutes, the experiment itself 25 minutes (of which 5 minutes were subtracted from the environmental situation from the beginning or from the end of the experiment, so the clear recording time was 20 minutes). The case that the test programmers should solve were their current challenges of development, the employee did not pull much of the work timeline.

Figure 1. Field experiment.

Analysis. Analysis of the results was performed using a programming language for statistical data processing and graphics – "R". So, were constructed an IBS distribution histogram for each of the participants (see Figure 2). Visual assessment of the distribution histogram shows a strong deviation from Gaussian shape distribution (distribution of IBS is seems not to be normal). We use the Shapiro-Wilk test with params p-value > 0.05 and indicate that all samples are different from normal. The next step was to compare all samples. We use non-parametric Kruskal-Wallis test. As a result of Kruskal-Wallis test is the fact that our distributions are significantly different (p-value = 6.14e-32). There is a diagram with 10 distributions in Figure 3.

In the construction of diagram uses stable (robust) estimate of central tendency (median) and dispersion (interquartile scale). Despite the fact that samples are statistically distinguishable the diagram shows that the boxes of interquartile ranges overlap by more than 1/3 of its length. Distribution number 8 (see Figure 3) has a different configuration from most others. The median value of this sample is maximal. Next was an analysis of the two extreme (at the median value) of samples - sample number 8 and number 10. It was found that in the experimental group of programmers three people had a specific situation (sample № 8, № 9, № 10). Professional activities of these three developers, is associated with the development of graphical user interface. The seven remaining men occupy positions of the server developer (backend), and the nature of their work is associated more with the processing of textual information, rather than graphical. We assume that this factor may be an additional result of the experiment, namely, the programmers of varying specificity are formed of various behavioral situations. Our assumption is consistent with the results of the University of Eastern Finland. In "Analyzing and Interpreting Quantitative Eye-Tracking Data in Studies of Programming: Phases of Debugging with Multiple Representations" research has shown that the changes in eye-tracking measures reflect both the

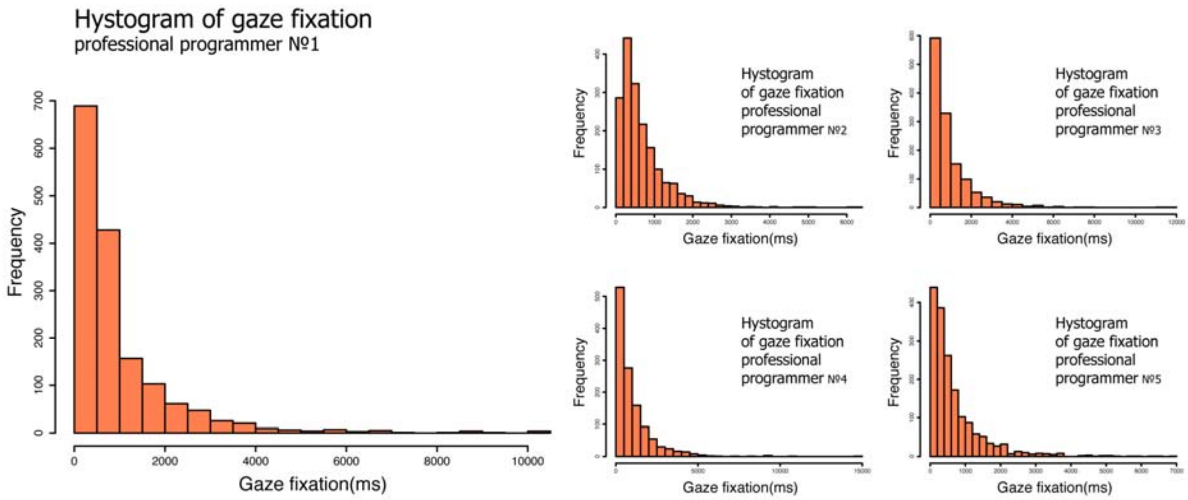


Figure 2. Hystogram of gaze fixation.

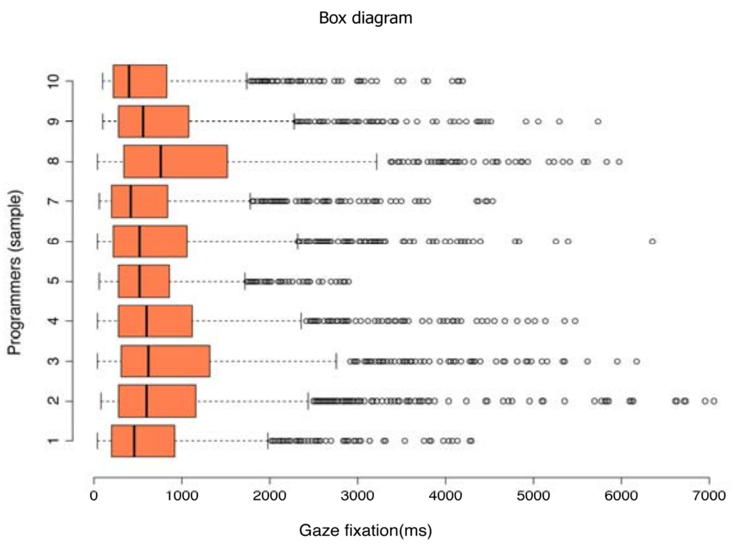


Figure 3. Box diagram.

importance of different representations(textual and graphical) during programming processes and differences in debugging strategies [2,3].

Thus, for further analysis we drop 3 samples. To clarify the statistical differences between the samples was carried out comparing trampled 7 remaining samples. As a result, the confidence level p = 0,05 were: 71,4% (15 out of 21) comparison of the samples showed statistically indistinguishable; 28,6% (6 out of 21) comparisons of samples showed a statistical difference. This suggests that there is a general trend of the nature of the presence of sampling time gaze fixation under the supervision of the work of Web developers, suggesting the formation of a web developer similar behavioral aspects of the work.

Results and Discussion

We use again the Shapiro-Wilk test with params p-value > 0.05 and indicate that the distribution of medians are normal (p-value = 0.39). On this basis were calculated by parametric features: mathematical expectation is 533.29; standard deviation is 76.35. We tend to believe that the obtained value of the mathematical expectation may be a reference, and the spread of it may be a criterion for deciding on the instrumental component of professional competence.

Next, similar experimental actions were carried out with four candidates for the programmer's position in company FotoStrana staff. All four men have already had one month training at the course of school web development Embria before the experiment. This school is a specialized unit, it main task is training in the field of programming for the company FotoStrana. Analysis of the results was constructed similar to the analysis of the gaze fixation of professional programmer. The results are presented in Table 1.

Table 1. Summary table.

Students id	Value of median from IBS sample, ms.	Reference value of IBS, ms.	Spread from reference, ms.	Reference spread, ms.	Correlation with inst. comp. of prof. comp.
Stud №1	480	533.29	53.29	76.35	true
Stud №2	439	533.29	94.29	76.35	false
Stud №3	499	533.29	34.29	76.35	true
Stud №4	699	533.29	165.71	76.35	false

The table shows that the two students (№ 1, № 3) deviation from the reference value of gaze fixation is placed in the reference, so we can talk about their high level of compliance instrumental component of professional competence. As a result, we tend to conclude that there is a general trend in the characters of the gaze fixation time of Web developers work. This fact confirms the assumption that web developers are formed similar behavioral aspects of the work. For an objective assessment of the behavioral aspects of the work could be applied the method of analysis taxonomy of eye movements, namely the value of the gaze fixation time. The main limitation of this study can be seen in the low number of participants (10 programmers and 4 students).

References

1. Gippenreyter, J.B. (1978). *Movement of the human eye*. Moscow State University, Moscow.
2. Bednarik, R., Myller, N., Sutinen, E., Tukiainen, M. (2006). Analyzing Individual Differences in Program Comprehension with Rich Data. *Technology, Instruction, Cognition and Learning* **3,** 205–232.
3. Bednarik, R., Tukiainen, M. (2007). Analysing and Interpreting Quantitative Eye-Tracking Data in Studies of Programming: Phases of Debugging with Multiple Representations. *PPIG'07, International Proceedings Series 7, University of Joensuu, Department of Computer Science and Statistics.*
4. Barabanshikov, V.A. (1997). Oculomotor structures of perception. Press “Institute of Psychology RAS,” Moscow (in Russian).
5. Buayas, T.M, Linde, N.D. (1976). The effect suppression of saccadic eye movements in the activity *Perceptions and activities*, 68-86.
6. Filin, V.A. (2002). Automaticity of saccades. Moscow State University (In Russian)
7. Filin, V.A. (1975). About the mechanism of involuntary eye jumps and their role in visual perception. *Motor components of vision*, 69-101.
8. Golikov, J.I., Kostin A.N. (2003). Analysis of multi-level mental processes in the regulation of operator activity. *Workshop on engineering psychology and ergonomics: Manual for higher students*, 113-135.
9. Jacob, R.J.K., Karn, K.S. (2003). Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary). In Hyona, J., Radach, R., Deubel, H., eds.: *The Mind’s Eye: Cognitive and Applied Aspects of Eye Movement Research*, Elsevier Science, 573-605.
10. Malyhin, A.A., Barabanshikov, V.A. (2000). The limits of applicability of mobile tracking system for the direction of gaze of the observer. *Experimental Psychology in Russia: Traditions and Prospects*, 127-132.
11. Sparks, D.L. (2002). The brainstem control of saccadic eye movements. *Nature Reviews Neuroscience* **3,** 952-964.
12. Zhukov, V.A. (2008). Engineering pedagogy. Concerns, experiences, and suggestions. St. Petersburg Politechnical University, Saint-Petersburg.

Posters

Performance of Rural and Urban Adult Participants in Neuropsychological Tests in Zambia

H. Zulu

Department of Physiological Sciences, University of Zambia, Lusaka, Zambia. happyoscarz@gmail.com

Formulation of normative sample for Zambia

Neuropsychological examination is an important way of formally assessing brain function. While there is so much documentation about the influence that some factors such as age and education has on neuropsychological tests (NP), not so much has been done to assess the influence that residency (rural/urban) may have. The specific objectives of this study were to establish if there is a significant difference in mean test scores on NP tests between rural and urban participants, to assess which tests on the Zambia Neurobehavioural Test Battery (ZNTB) are more affected by the participants' residency (rural/urban) and to determine the extent to which education, gender and age predict test performance on NP tests for rural and urban participants.

To ensure that the study met the ethical requirements, the research proposal was submitted to the University of Zambia Biomedical Ethics Committee and the Ministry of Health in Zambia after which it was approved. This was to ensure that the participants‟ interests and ethical issues such as confidentiality and informed consent were put into consideration.

The participants (324) were drawn from both urban and rural areas of Zambia (Rural = 152 and Urban = 172). However, only 234 participants (Rural = 152 and Urban 82) were used for all the analyses in this particular study. The 234 participants were used as the actual proportion of the rural vs. urban population in Zambia was 65%: 35%, respectively (CSO, 2003). The rural-urban ratio for the participants that were captured during the data collection process was 152: 172, respectively. Thus, all the rural participants (152) were included and 90 of the 172 urban participants were randomly excluded so that the rural/urban ratio reached the desired 65%: 35 % which was the required ideal statistic for appropriate representation of the actual population in Zambia. Data on NP tests were analyzed from 234 participants, rural (N=152) reflecting 65% and urban (N=82) reflecting 35%. T-tests indicated that urban participants had superior performances in all the seven NP tests domains and all the mean differences in all these domains were found to be statistically significant. The NP tests domains were visual episodic memory, verbal episodic memory, attention/working memory, verbal fluency, speed of information processing, executive functioning and motor functioning. Residency had a large or moderate effect in five domains, while its effect size was small only in two of the domains. A standard multiple regression revealed that education, age and residency as predictor variables made a significant contribution to variance in performance on various domains of the ZNTB. However, gender of participants was not a major factor in determining one's performance on neuropsychological tests.

As can be seen from table below, residency (rural/urban) had a large effect size with regard to performance on the domains of verbal fluency and speed of information processing. The effect of residency was moderate/medium on the domains of verbal episodic memory, visual episodic memory and attention/working memory. The effect size of residency was however, small in the domains of motor/dexterity and executive functioning. The influence of residency on the performance of the seven domains of the ZNTB was therefore, either large or moderate in five and was small only in two.

Standard multiple regression was also used to assess the ability of four independent/predictor variables (education – years/level of and quality of, gender, age, and residency – rural/urban) to predict NP tests performance on the seven domains of the Zambia Neurobehavioural Test Battery (ZNTB). Residency had a statistically significant contribution to the model in five of the domains, namely, visual episodic memory, verbal episodic memory, verbal fluency, speed of information processing motor/dexterity. The results in this study

Table 1. Mean Effect sizes of Residency on the NP Tests Domains of the ZNTB.

NP Tests Domains	Residency	N	Mean	Std. Deviation	*P*	Eta squared value
Verbal fluency	Rural Urban	152 82	9.75 12.10	2.18 1.67	<.005*	.238
Speed of information processing	Rural Urban	152 82	9.62 11.63	2.13 2.02	<.005*	.175
Verbal episodic memory	Rural Urban	152 82	9.63 11.72	2.74 2.73	<.005*	.118
Visual episodic memory	Rural Urban	152 82	9.64 11.47	2.88 3.00	<.005*	.083
Attention/Working memory	Rural Urban	152 82	9.80 11.20	2.19 2.49	<.005*	.079
Motor/dexterity	Rural Urban	152 82	9.72 11.04	2.56 2.82	<.005*	.054
Executive functioning	Rural Urban	152 82	152 82	1.99 2.28	<.005*	.053

* Results significant at $p < .05$ with 95% confidence.

indicated to a larger extent that residency as a predictor variable is an “umbrella” that has within it a number of associated factors such as education and acculturation that makes it a better predictor of performance on NP tests in developing countries as compared to a single independent/predictor variable of education, age, gender, etc., assessed individually. This particular report is part of an ongoing larger cutting-edge study aimed at formulating the normative data for Zambia with regard to performance on neuropsychological tests. This is necessary for appropriate, effective and efficient assessment or diagnosis of various neurocognitive and neurobehavioural deficits that a number of people may currently be suffering from, (Strauss et al., 2006). It has been shown in this study that it is vital to make careful analyses of the variables that may be associated with one’s performance on neuropsychological tests.

References

1. Central Statistical Office – CSO (2003). Zambia 2000 Census of Population and Housing: Zambia Analytical Report, Vol. 10. Lusaka, November 2003.

2. Strauss, D.T., Stethem, L.L., Hegenhoiltz, H., & Richard, M.T. (1989). Traumatic brain injury: A comparison of three clinical tests and analysis of recovery. *The Clinical Neuropsychologist* **3**, 145-156.

Development of a Scale for Fast Screening of Fatigue Syndrome from Mental Illness

Tang Tze-Chun, Huang Shih-Ying

Kaohsiung Medical University Chung-Ho Memorial Hospital, Kaohsiung, Taiwan. ttc1021@yahoo.com.tw

Background

Physical, mental, affective and vigor components of fatigue are often associated with depression and other Axis I psychiatric disorders. We developed a 9 item multi-dimensional scale to specifically evaluate symptoms of fatigue, and correlated it to a validated 30-items Chinese Version of Multidimensional Fatigue Symptoms Inventory-Short Form (MFSI-SF-C).

Methods

Adult subjects (n=125; Mean age 35 year old; >3 months after psychiatric treatment) completed the Beck Depression Inventory (BDI), and multidimensional Fast Screening of Fatigue Scale (FSFS), MFSI-SF-C and Sheehan Disability Scale (SDS) assessment. Internal consistency was determined by Cronbach's coefficient alpha calculations. Inter-factor correlations were also assessed. 125 subjects visiting a clinical trial site consented to participate in this reliability and validity study. Diagnoses included Major Depressive Disorder, Bipolar disorder, and schizophrenia. There was one cross-sectional clinic visit during which the FSFS rating instruments were administered in conjunction with the BDI , SDS and MFSI-SF-C. We correlated the score of FSFS, BDI to MFSI-SF-C and SDS. The study was verified with IRB consent and confidentiality was warranted.

Results

Internal consistency for all subscales was within the recommended ranges ($\alpha \geq 0.70$). We found moderate to strong correlations between BDI, FSFS and SDS (r=0.82 for BDI, r=0.76 for FSFS) and between BDI, FSFS and MFSI-SF-C (r=0.72 for BDI, r=0.87 for FSFS). There was high internal consistency of FSFS and BDI by one rater at visit 2 as demonstrated by Cronbach’s alpha = 0.86 and 0.92.

Conclusions

As shown above, the modified FSFS instruments for fatigue are reliable measures of fatigue and both measures are validated with the SDS and MFSI-SF-C. The multidimensional Fast Screening of Fatigue Scale (FSFS) appears to be feasible for population of major psychiatric disorders. More information is needed to determine whether the issues addressed by FSFS are also meaningful to fatigue syndrome for comorbid physical problems such as DM and cancer patients.

Multilevel Meta-Analysis of Single-Subject Experimental Designs

M. Ugille[1], M. Moeyaert[1], S.N. Beretvas[2], J. Ferron[3], W. Van den Noortgate[1]

[1]*Faculty of Psychology and Educational Sciences, University of Leuven, Leuven, Belgium. [1]maaike.ugille@ppw.kuleuven.be*

[2]*Department of Educational Psychology, University of Texas at Austin, Austin, USA.*

[3]*Department of Educational Measurement & Research, University of South Florida, Tampa, USA.*

Single-case or single-subject experimental designs (SSEDs) are used when one is interested in the effect of a treatment for one specific subject, a person or another entity [1]. In these kind of designs, the subject is observed repeatedly before the treatment is introduced, and during or after the treatment (Figure 1). Because it is difficult to generalize the results from such an experiment to other subjects, the experiment can be replicated within or across studies. The results of several single-subject studies can be combined by performing a multilevel meta-analysis. In a multilevel meta-analysis, the results of each subject are first converted to an effect size, and next, these effect sizes are combined, taking into account that measurements are clustered in subjects, and subjects are clustered in studies. Van den Noortgate & Onghena [2,3] proposed to use two regression coefficients as the effect size metrics: the effect of the treatment on the intercept and the effect of the treatment on the slope. They also proposed to standardize these regression coefficients, by dividing them by the residual within-phase standard deviation, so that these effects are comparable over subjects and studies. To combine the estimates of the effect on the intercept for multiple subjects in a meta-analysis, Equation 1 can be used:

$$\mathrm{b_{2jk}} = \gamma_{200} + \mathrm{v_{20k}} + \mathrm{u_{2jk}} + \mathrm{r_{2jk}} \tag{1}$$

$$\text{with } \mathrm{v_{20k}} \sim \mathrm{N}\left(0, \sigma^2_{\mathrm{v_{20k}}}\right),\ \mathrm{u_{2jk}} \sim \mathrm{N}\left(0, \sigma^2_{\mathrm{u_{2jk}}}\right),\ \mathrm{r_{2jk}} \sim \mathrm{N}\left(0, \sigma^2_{\mathrm{r_{2jk}}}\right)$$

With $\mathrm{b_{2jk}}$ the ordinary least squares estimate (OLS) of β_{2jk}, the treatment effect on the intercept for subject j (j = 1, 2, …, J_k) in study k (k = 1, 2, …, K). This estimate equals an overall effect γ_{200} of the intervention on the intercept, plus a random deviation for study k, $\mathrm{v_{20k}}$, a random deviation for subject j in study k, $\mathrm{u_{2jk}}$, and a residual deviation, $\mathrm{r_{2jk}}$ because the treatment effect for that person is not perfectly estimated. The same equation can also be used for the effect of the treatment on the trend.

We evaluated the performance of this approach, using simulation research. The results indicate that a multilevel meta-analysis of standardized effect sizes is only suitable when the number of measurement occasions for each subject is 20 or more. On the other hand, when there are only a few measurements per subject, the estimated effects were found to be biased (Figure 2, left boxplot).

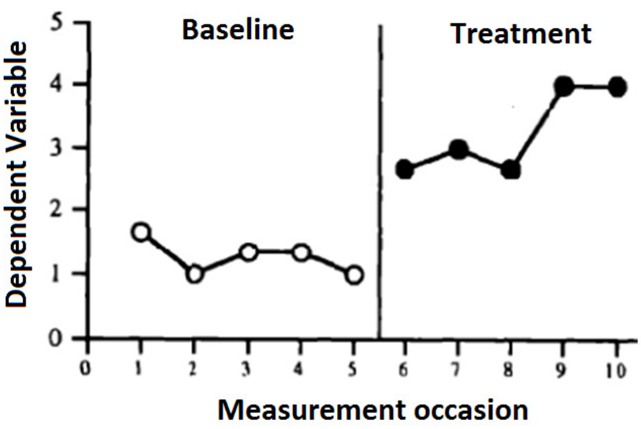


Figure 1. Graphical display of the basic single-case experimental design.

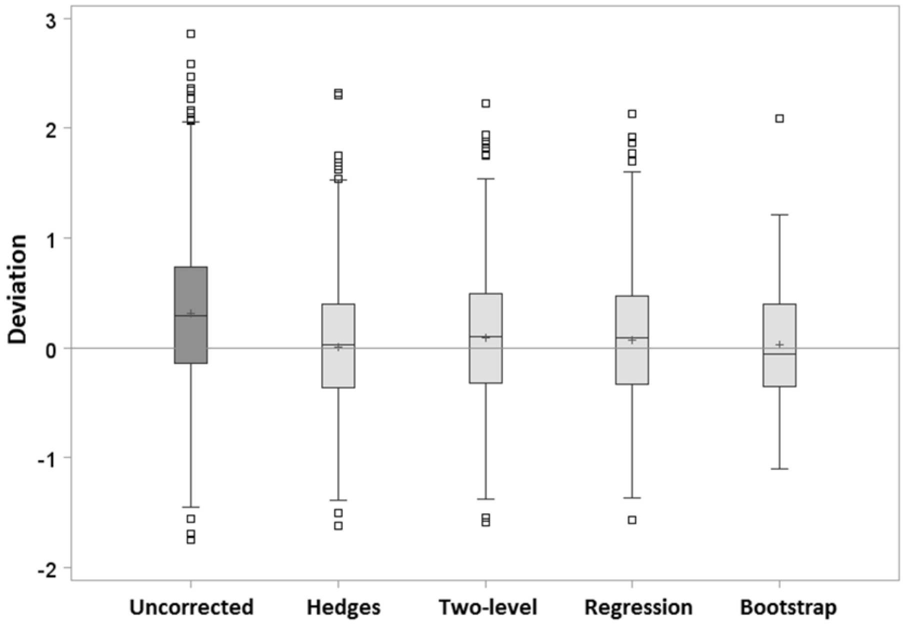


Figure 2. Distribution of the deviations of the estimated mean effect on the intercept of the treatment from its populations value, with the immediate treatment effect equal to 2, the effect on the trend equal to 0.2, the number of studies equal to 10, the number of subjects per study equal to 4, the number of measurements per subject equal to 10, the between-study and the between-subject variance equal to 2.

In a following step, we tested four ways to correct for the biased estimates of the effect. In the first approach, we used the formula of Hedges [4] to correct for small-sample bias, in the second approach, we estimated the residual within-subject standard deviation by performing a two-level analysis per study, in the third approach we estimated the residual within-subject standard deviation by performing a regression analysis per study, and in the fourth approach, we used iterative raw data bootstrapping to correct for bias [5]. Based on simulation studies, the first and the last approach seemed best suited to correct the bias (Figure 2).

References

1. Barlow, D. H., Hersen, M. (1984). Single-case experimental designs: Strategies for studying behavior change (2nd). New York: Pergamon Press.

2. Van den Noortgate, W., Onghena, P. (2003). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments, & Computers* **35**, 1-10.

3. Van den Noortgate, W., Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental design studies. *Evidence Based Communication Assessment and Intervention* **2**, 142-151.

4. Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics* **6**, 107-128.

5. Goldstein, H. (1996). Consistent estimators for multilevel generalised linear models using an iterated bootstrap. *Multilevel Modelling Newsletter* **8**, 3-6.

Effects of Maternal Infection on Anxiety and Depression-like Behaviours of Offspring

J. Solati[1], G.H. Moll[2], R.R. Dawirs[3], Y. Golub[4]

Dept. of Child and Adolescent Mental Health, University of Erlangen-Nürnberg, Erlangen, Germany

[1]Jalal.solati@uk-erlangen.de, [2]Gunther.moll@uk-erlangen.de, [3]Ralph.Dawirs@uk-erlangen.de, [4]Yulia.Golub@uk-erlangen.de

Introduction

LPS and some cytokines can activate the hypothalamic–pituitary–adrenal (HPA) axis and affect brain development in pregnant mice. Maternal infection during pregnancy is a risk factor for several psychiatric illnesses with neurodevelopmental origin. In this study we have evaluated the effects of exposure of pregnant mice to the bacterial lipopolysaccharide (LPS) on anxiety and depression-related behaviour of male offspring.

Methods

Pregnant NMRI mice were treated with intra-peritoneal administration of LPS (120, 240 and 480 µg/kg) at the 10^{th} gestational day. Induction of the pro-inflammatory cytokines, TNF-α, IL-1β and IL-6, was measured in maternal serum 1.5h following maternal LPS challenge. All of the pregnant rats were allowed to give birth and nurture their offspring normally. Number of animals in each litter was standardized (3 males and 3 females/ dam), in order to standard milk availability. Pups were weaned on postnatal day 21 (PD 21), and offspring housed (four animals from the same treatment/cage). Pregnant mice and male offspring divided to control and experimental groups (n = 7/ group for pregnant females; n = 7-10 with 2 pup each litter for the adulthood behaviors). Adult male offspring were chosen randomly for each test.

Anxiety-related behaviour of the male offspring (at postnatal day 70) was studied using elevated plus maze (EPM) test. EPM is a wooden, plus-shaped apparatus that elevated to the height of 50 cm above the floor. This maze is composed of two open arms (30cm×5cm), and two enclosed arms (30cm×5cm×15cm), each arm have an open roof. The maze was placed in the center of a quiet and dimly lit room. Following their respective treatment, mice were placed individually in the center of the plus-maze, facing one of the open arms. Behavioral data were collected by a “blind” observer who quietly sat 1 m behind one of the closed arms of the maze, using a time meter. The observer measured: (1) time spent in the open arms, (2) time spent in the closed arms, (3) number of entries into the open arms, and (4) number of entries into the closed arms during the 5-min test period. An entry was defined as all four paws in the arm. The maze was cleaned with distilled water after each mice was tested. For the purpose of analysis, open-arm activity was quantified as the amount of time that the mice spent in the open arms relative to the total amount of time spent in any arm (open/total×100), and the number of entries into the open arms was quantified relative to the total number of entries into any arm (open/total×100). The total number of arms entered, as well as the total number of closed arms entered was used as indexes of general locomotor activity. Anxiety test of pregnant mice was carried out 1.5 hours after injection of the LPS and anxiety test of offspring was carried out at postnatal day 60.

Forced swimming test was applied for evaluating the depression-like behaviour. Forced swimming test (FST) was performed as described, with minor modifications. Mice were forced to swim individually in a vertical glass cylinder (height 30 cm, diameter 15 cm) filled with water maintained at 24–26°C to a depth of 15 cm. After testing in the water, mice were removed and allowed to dry in a heated enclosure. Duration of immobility (making only minimal movements to keep the head above water or floating), swimming (movement of all four legs with body aligned horizontally in the water), and climbing (movement of all four legs with body aligned vertically in the water) were measured from videotapes by a trained blind observer. The total duration of each of the three behavioral parameters, immobility, swimming and climbing, was separately recorded by different observation sessions. To examine the robustness of this method, we compared the immobility time separately measured and that calculated by subtracting total time by the sum of swimming and climbing times.

Results and conclusions

Our results demonstrate that LPS administration induces a significant increase in TNF-α, IL-1β, IL-6 and corticosterone levels in maternal serum. However, in offspring, prenatal LPS administration has no significant effects on serum cytokines, corticostrone levels and depression-like behaviour, while decreasing the anxiety levels.

Ethical Statement

The study was approved by the Ethics Committees of *University of* Erlangen-Nürnberg and the experimental protocol is in compliance with the National Institutes of Health Guide for Care and Use of Laboratory Animals.

Some Common Indices of Group Diversity: Upper Boundaries

A. Solanas[1], R.M. Selvam[2], J. Navarro[3], D. Leiva[4]

[1]Department of Behavioral Sciences Methods, University of Barcelona, Barcelona, Spain. antonio.solanas@ub.edu

[2]Department of Behavioral Sciences Methods, University of Barcelona, Barcelona, Spain. rselvam@ub.edu

[3]Department of Social Psychology, University of Barcelona, Barcelona, Spain. j.navarro@ub.edu

[4]Department of Behavioral Sciences Methods, University of Barcelona, Barcelona, Spain. dleiraur@ub.edu

Diversity, which is defined as the collective amount of differences among members within a social unit [2], has been an important concept applied in various ways across fields like ecology, demography, information systems, sociology, economics, and psychology. In organizational research, diversity has been prominent in studies using team members' characteristics to predict performance [3]. A review of the past diversity research shows that most scholars choose the indices according to methodological priority, theory, and familiarity of the indices [4]. Some researchers have suggested that theoretical refinement of the conceptualization of diversity is necessary before selection of an index. For example, Harrison and Klein [1] proposed organizing indices according to three types of diversity *separation*, *variety*, and *disparity*. They defined diversity as the distribution of differences among the members of a unit with respect to a common attribute. Consequently, separation, variety, and disparity are respectively understood as differences in attitude or position, differences in categorical characteristics, and differences in power or status hierarchy. Concerning the maximum diversity for variety, separation, and disparity, they correspond to a discrete uniform distribution, a bimodal distribution at extremes of a continuum, and a positively skewed distribution, respectively.

The current study shows that the maximum value of a group diversity measurement is a function of the group size and the distribution of members within a group across the respective properties. The main purpose is to obtain proper upper boundaries for each of the commonly used indices of diversity for all conditions featuring group size characteristics within the concept of the diversity types. Following Harrison and Sin's [2] suggestion to normalize the diversity indices, which reduces the inflating effects on group size, a normalized range for the discrete random variables was obtained. These results are useful to applied researchers interested in comparing index values with respect to suitable maximum boundaries and also to use the normalized indices to predict group outcome at group level of analysis.

Table 1 shows the results obtained for three indices, that is, the maximum value for all of them (Blau's index, standard deviation, and coefficient of variation). Blau's index, standard deviation, and coefficient of variation respectively correspond to measure diversity as variety, separation, and disparity. The minimum value equals 0 for all diversity indices, meaning that individuals are absolutely homogeneous as regards the characteristic of interest. On the contrary, the maximum value varies depending on several properties. Our research focused on maximum values as they had not been solved for group size parity and varying group size and thus we obtained normalized indices for the possible cases (Table 1). In fact, we have also obtained maximum values for other indices (not shown here), such as the Teachman's index (for diversity as variety), mean Euclidean distance (for diversity as separation) and Gini coefficient (for diversity as disparity).

Having derived general upper boundaries, researchers can compare the values obtained to their suitable maximum values and thus make proper conclusions for the specific conditions (e.g., group size and group size parity). Additionally, applied researchers may consider that normalized diversity indices could be useful to improve outcomes at group level as they are expressed in the same metric. These normalized indices allow social researchers to make proper comparisons among groups or organizations as these measures provide measurements of dispersion that can be applied to collectives of differing sizes. It should be noted that the present study was constrained to descriptive analysis and thus no conclusions are made about the statistical properties of the indices as estimators.

Table 1. Upper boundaries and normalized indices for some common diversity measures. All normalized indices range from 0 to 1. n, k, and p_i respectively denote group size, the number of categories, and the proportion for the ith category. The value a is equal to n – k × int[n/k], where *int[]* denotes the integer function. By means of x_{min} and x_{max}, we denote the minimum and maximum values of scales.

Diversity indices	Standard formula	Maximum value	Normalized index
Blau's index (Variety)	$B = 1 - \sum_{i=1}^{k} p_i^2$	$B_{\max} = \frac{n^2(k-1) + a(a-k)}{kn^2}$	$B_N = \frac{B}{B_{\max}}$
Standard deviation (Separation)	$SD = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n}}$	$SD_{\max(even)} = \frac{x_{\max} - x_{\min}}{2}$ $SD_{\max(odd)} = \sqrt{\frac{n^2 - 1}{n^2}} \frac{x_{\max} - x_{\min}}{2}$	$SD_N = \frac{SD}{SD_{\max}}$
Coefficient of variation (Disparity)	$V = \frac{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2 / n}}{\overline{x}}$	$V_{\max(n)} = \frac{\sqrt{n-1}(x_{\max} - x_{\min})}{(n-1)x_{\min} + x_{\max}}$	$V_N = \frac{V}{V_{\max}}$

References

1. Harrison, D.A., Klein, K.J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review* **32**, 1199-1228.

2. Harrison, D.A., Sin, H.P. (2006). What is diversity and how should it be measured? In A.M. Konrad, P. Prasad, J.K. Pringle (Eds.), *Handbook of workplace diversity* (pp. 191-216). Thousand Oaks, CA: Sage Publications.

3. Horwitz, S.K., Horwitz, I. B. (2007). The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of Management* **33**, 987-1015.

4. Roberson, Q.M., Sturman, M.C., Simons, T.L. (2007). Does the measure of dispersion matter in multilevel research? A comparison of the relative performance of dispersion indexes. *Organizational Research Methods* **10**, 564-588.

Application of a Network-Analysis Algorithm for the Definition of Locations in Open Field Behavior: How Rats Establish Behavioral Symmetry in Spatial Asymmetry

S. Weiss[1], O. Yaski[1], E. Blumenfeld-Lieberthal[2], J. Portugali[3], D. Eilam[1]

[1]*Department of Zoology, George Wise Faculty of Life Sciences, Tel Aviv University, Tel-Aviv, Israel. eilam@post.tau.ac.il*

[2]*The David Azrieli School of Architecture, Yolanda and David Katz Faculty of the Arts, Tel Aviv University, Tel-Aviv, Israel. juval@post.tau.ac.il*

[3]*Department of Geography and the Human Environment, The Lester and Sally Entin Faculty of Humanities, Tel Aviv University, Tel-Aviv, Israel. efratbl@post.tau.ac.il*

Abstract

The present study applied network analysis to scrutinize spatial behavior of rats tested with symmetrical or asymmetrical layout of 4, 8, or 12 objects along the perimeter of a dark, round arena. We considered locations as the units of a network (nodes), and passes between locations as the connections of the network. In terms of Ethovision parameters, there were only minor differences between rats tested in either symmetrical or asymmetrical object layouts. However, network analysis revealed substantial difference in the behavior between the layouts. For the network analysis, we first defined locations in the environment, where each 'location' was a cluster of stopping coordinates (defined as no progression for at least 1 second) extracted from Ethovision. From the set of locations and the passes between them we extracted the network analysis parameters: for each node-degree,clustering coeficient and shortest mean path were calculate. In addition the average network degree,clustering coeficient and shortest mean path were extracted for each rat. It was found that behavior in either a symmetrical or asymmetrical layout of 4 objects, the key locations coincided with the objects. However, in the asymmetrical layout with 4 objects, additional key locations were spaced at a distance that was identical to the distance between other objects, forming behavioral symmetry among the key locations. In other words, it was as if the rats imposed behavioral symmetry in their spatial behavior in the asymmetrical environment. We suggest that wayfinding was easier in symmetrical environments, and therefore, when the physical attributes of the environment were not symmetrical, rats behaviorally established a symmetric layout of key locations, thereby gaining a more legible representation of the environment despite its more complex physical structure. Altogether, the present study adds a behavioral definition for a location, a term that so far has been mostly discussed according to its physical attributes or neurobiological correlates (place, border and grid neurons). Moreover, the network analysis enabled to identify a location and to assess its importance, even when that location did not have distinctive physical properties.

Materials and Methods

Ethical statement

This study was carried-out under permit # L-10-013 of the IACUC of Tel-Aviv University. For the purpose of this study, male Wistar rats (n = 16; age 3 months; weight 250–300 g) were housed in a temperature-controlled room (21°C) with 12/12 h light/dark cycle (dark phase 8:00 to 20:00). Rats were held in standard rodent cages (40 X 25 X 20 cm; 2 rats per cage) with sawdust bedding and were provided with free access to water and standard rodent chow. Each rat was marked with a waterproof marker on its tail, and accustomed to handling - 10 minutes a day for one week. Rats were maintained and treated according to the institutional guidelines for animal care and use in research.

Apparatus

Rats were tested in a round arena, 200 cm in diameter, surrounded by a 50 cm high tin wall. The arena was placed in a temperature-controlled (21 ±1 °C) and light-proofed room. The arena floor was covered with a navyblue PVC layer. During testing, the room was completely dark, illuminated only by infra-red light invisible to rats (Tracksys, IR LED Illuminator; UK, with a 830 nm wavelength filter). Trials were recorded by a video camera (Ikegami B/W ICD-47E, Japan) placed 2.5 m above the center of the arena, providing a top view of the entire arena. Footage was saved on a DVD device (Sony RDR-HXD 870, Japan). Each rat underwent three trials with 4, 8, and 12 objects respectively, in only one of the following two object layouts. Objects (black cement blocks; 6.5 X 6 X 6 cm), were placed in either a symmetrical or an asymmetrical layout along the arena walls. In the asymmetrical layout, distance between objects was established randomly, with at least 25 cm between objects, preventing the rats from touching two objects at the same time. Overall distance between objects was equal on average for both layouts in each trial.

Procedure

Sixteen rats were randomly assigned to two groups of eight rats, each undergoing three trials in an arena with either a symmetrical or an asymmetrical layout of objects. Each rat was individually tested on alternate days with an increasing number of objects, starting with 4, then 8, and finally 12 objects. We did not include the counterbalanced paradigm (18, 8 and then 4 objects) since our past studies indicate that this procedure did not affect the results [1-2]. At the beginning of each trial, a rat was placed at a fixed start location next to the arena wall, and its behavior was recorded for 20 minutes. The arena was wiped with detergent between successive trials. All testing took place during the dark phase when rats were most active.

Data acquisition and analysis

For path analysis, the arena was divided into the following virtual areas:

Perimeter - a 15 cm strip along the arena wall.

Center - the remaining central area of the arena (excluding the perimeter area).

Object area - a 25 X 25 cm square, around each object. Since objects were placed along the perimeter, object areas were of course located within the perimeter area.

The paths of movement of the rats in these areas were tracked from the video files using ‘Ethovision XT 7’ (Noldus Information Technologies, NL), a software that provided the coordinates of the center of mass of the rat five times per second. The following parameters were extracted for further analysis with 'Microsoft Excel 2007':

Distance traveled - the cumulative metric distance (m) traveled over 20 minutes.

Duration - the time spent (min) at each of the arena areas.

Travel between center and perimeter - incidence of crossing between center and perimeter areas.

Duration at an object - the time spent (min) in an object area.

Visits to an object - the number of entries to an object area.

Network analysis

For network analysis, behavior was considered as a set of locations and the transitions between these locations. In this representation, we defined local parameters that referred to the behavior within a specific location, and global parameters that referred to the behavior in the entire arena. Custom-designed software ('Huldot' by Michael Liberthal) was used to identify locations of interest for the rats, as reflected by their X-Y stopping coordinates inside the arena. We defined a stop as no progression for at least 1 second. ‘Huldot’ was based on an algorithm that was based on the stopping behavior [3-4], its mathematical principles much resemble the City Clustering Algorithm (CCA) [5-6], yet it represents a novel approach to the study of spatial behavior. Using the ‘Huldot’ algorithm, we identified the rat’s first stopping coordinate and defined it as node1. We then added to node1 all the stopping coordinates that were located at $d \leq l$ from the first stop (where d represents the measured

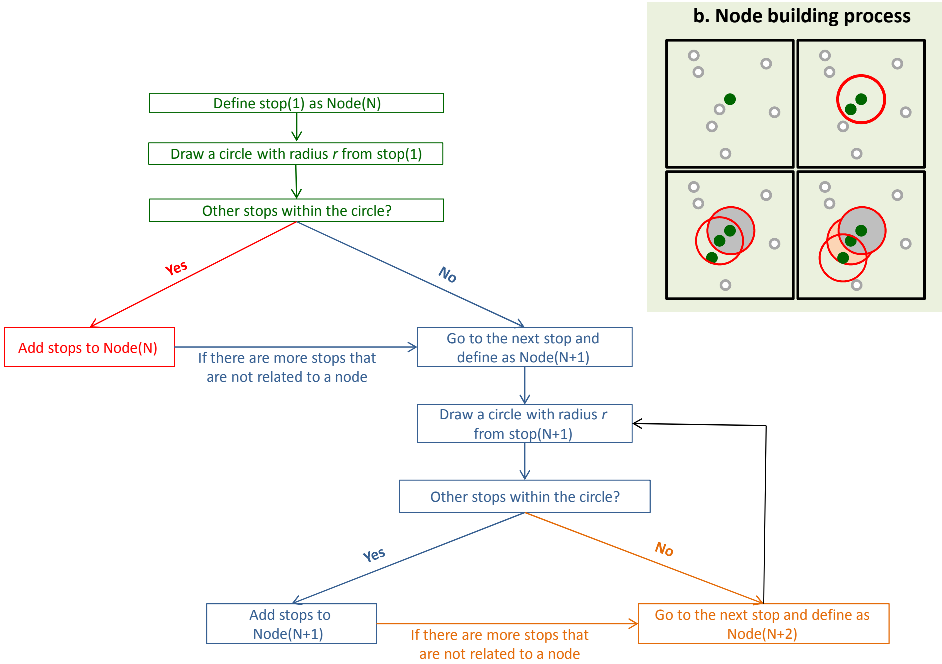


Figure 1. The algorithm for transforming stopping coordinates into a network node (**a**) and a visualized process of building a single node (**b**). (taken from Weiss et at. PlosOne 2012, under revision).

distance between the stops and l represents a “unifying criterion” that was set to 12 cm – about half a rat’s body-length). We continued adding new stops to node1 until there were no more stops at d≤l from any of the stops included in node1. We then identified the rat's next stop (not included in node1), defined it as node2, and repeated the process (see Figure 1).

The distance between stopping coordinates was calculated, and stopping coordinates within a 12-cm diameter were assigned to the same node. This diameter was found to be the best fit according to the following considerations: physically, this diameter (12 cm) had to be less than 14.5 cm (which is half the shortest distance between objects), as otherwise stopping coordinates at two adjacent objects could be attributed to one node at an intermediate distance between the two objects (see Figure 2). The 12-cm diameter also had to be greater than 9 cm in order to prevent the splitting of stopping coordinates at the same object into two separate nodes (see Figure 2). Within this diameter range, the 12-cm diameter was the best fit for all animals. Altogether, the algorithm provided a method by which to define locations in the spatial behavior of the rats, offering a useful definition irrespective of network analysis. In the present study, however, this definition of locations (nodes) was a prerequisite for network analysis, enabling us to view the behavior as a network comprised of nodes (node = clusters of stopping locations), and of links between these nodes (link = pass from one node to another). Moreover, once the nodes and the links between them were established, it was possible to refer only to the topology of the behavior while ignoring the metric distance between actual stopping coordinates, and to analyze the behavior only in terms of the nodes and links that constituted the network (see Figure 2e).

Stopping coordinates: - these are as the x-y coordinates of a single rat, as extracted from the tracking system (Ethovision). The large black circle represents the arena perimeter, each red dot represents a stopping coordinate at which the rat stopped for one second or longer, and the black squares represent the location of the objects.

Nodes under the application of a 12-cm circle around the additional stopping coordinate:- As shown, with this diameter the nodes (circles) coincide with the objects and behavior.

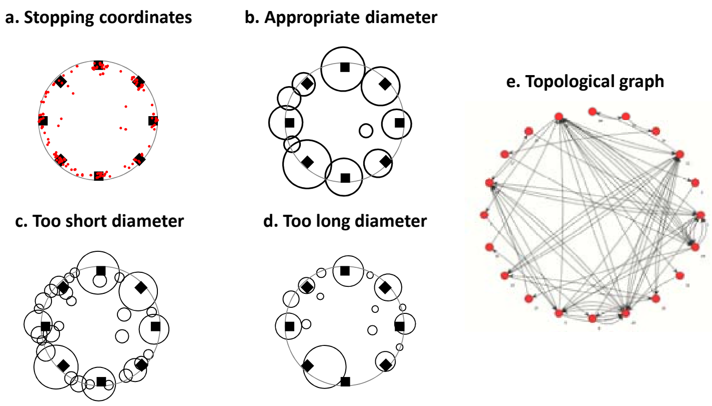


Figure 2. The rationale for establishing the criterion of 12 cm diameter and the transformation of stopping coordinates into a network is illustrated for one rat. (taken from Weiss et at. PlosOne 2012, under revision).

Nodes under the application of a 9-cm circle around the additional stopping coordinate:- As shown, with this diameter stopping coordinates around the same object split into several nodes, resulting in a mismatch between behavior and nodes.

Nodes under the application of a 14-cm circle around the additional stopping coordinate:- As shown, with this diameter the bottom node encompasses the stopping coordinates of two objects (see the red dots of these objects in a.).

Topological graph:- The presentation of the network after the transformation of stopping coordinates into nodes (red circles). Arrows between nodes represent the links (passes) between nodes. Note that the location of a red circle does not represent the physical location of that node. Likewise, the circles that represent the nodes in b-d do not represent the real size of the node but the number of stopping coordinates included in that node.

Local network parameters

Once nodes and links had been defined, the following parameters were provided by 'Huldot' for each node in the network:

Degree/Connectivity (**k**) - the number of links that a node has with other nodes (i.e. the number of neighbors that a node has).

Clustering coefficient (**C**) - the number of actual links between the neighbors of a specific node, divided by the total number of possible links that could occur between them ("how many of a node's neighbors are also each other's neighbors"). Clustering coefficient (C) is a value between 0 and 1, representing the level of connections between all of a node's neighbors (what portion of a node's neighbors are themselves neighbors). This was calculated as follows: $\boldsymbol{C} = \frac{2|e_{jk}|}{k_i(k_i-1)}$ where $\boldsymbol{e_{jk}}$ is the number of links between node $\boldsymbol{i}$ to other neighbors, and $\boldsymbol{k_i}$ is the number of node i neighbors.

Shortest path length (**l**) - the minimal number of nodes needed to be traveled in order to reach all the nodes in the network from a specific node.

The above parameters shed light on three different aspects of the nodes, indicating the relative importance of the nodes within the network. For example, a node with a high degree (**k**) and a low shortest path length (**l**) would be typical to a key node (hub) for travel within the network.

Global network parameters

While the above parameters refer to specific nodes, additional parameters were calculated for the entire network of each rat, as follows:

Total number of stops

Total number of nodes

Average network degree (**<k>**) - the average number of links per node (the average of the above degree values of all nodes). This parameter represents the connectivity of the network.

Average network clustering coefficient (**<C>**) - the average of the above local clustering coefficients of all nodes.

Average network shortest path length (**<l>**) - the average of the above shortest path lengths of all nodes. This parameter reflects the minimal number of nodes that needed to be traveled in order to reach from any node to any other node in the network.

Network density (**d**) - the ratio of the number of actual links divided by the number of theoretically possible links between all nodes.

"Key nodes" - nodes that encompass 10 or more stopping coordinates were defined as key nodes. The value of 10 stopping coordinates was set as a threshold based on ranking all nodes according to the number of stopping coordinates clustered within them. A two-fold difference from other nodes was noted in the layout with highest ranking nodes, which were thus separated and defined as key nodes with 10 as their minimal number of stopping coordinates.

Statistics

Data were compared by either a two-way ANOVA with repeated measures followed by a Tukey post hoc test, or by means of a Student's *t*-test. Alpha level was set to 0.05.

References

1. Eilam, D. (2003), Open-field behavior withstands drastic changes in arena size. *Behavioural Brain Research* **142**(1-2), 53-62.

2. Ben-Yehoshua, D., Yaski, O., Eilam, D. (2011). Global movement in a changing environment: spatial behavior in rats is shaped by global and not by local geometry. *Animal Cognition* **14** , 341-350.

3. Eilam, D., Golani, I. (1989).Home base behavior of rats (Rattus norvegicus) exploring a novel environment. *Behavioural Brain Research* **34**(3), 199-211.

4. Golani, I., Benjamini, Y., Eilam, D. (1993). Stopping behavior: constrains on exploration in rats (Rattus norvegicus). *Behavioural Brain Research* **53**, 21-33.

5. Rozenfeld, H.D., Makse, H.A. (2009). Fractality and the percolation transition in complex networks. *Chemical Engineering Science* **64**(22), 4572-4575.

6. Rozenfeld, H.D., et al. (2008). Laws of population growth. Proceedings of the National Academy of Sciences of the United States of America **105**(48) 18702-18707.

7. Weiss, S., Yaski, O., Eilam, D., Portugali, J., Blumenfeld-Lieberthal, E. (2012) Network analysis of rat spatial cognition: behaviorally establishing symmetry in a physically asymmetrical environment. PlosOne - under revision.

The Difference Between Sport Rituals, OCD Rituals, and Daily Routines: The Possible Adaptive Value of Seemingly Unnecessary Acts

O. Weiss[1], H. Keren[1], J. Mort[2], D. Eilam[1]

[1]*Department of Zoology, Tel-Aviv University, Tel-Aviv, Israel.* [1]*eilam@post.tau.ac.il*

[2]*Behavioral Modeling Branch, US Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio, USA.* [2]*joel.mort@wpafb.af.mil*

Abstract

Repetitive behaviors are common in daily life, constituting a seemingly non-functional component, manifested in excess in sport or compulsive rituals. The similarities and differences among various types of repetitive behavior remain unclear. Here we analyzed a daily task (donning a shirt, performed by 10 healthy volunteers, 3 tasks each) and a sport task (basketball free-throw, performed by 10 NBA players, 10 tasks each) by means of The Observer behavioral coding and analysis software, and found that both comprised pragmatic acts performed by all subjects, and idiosyncratic acts performed by some of the subjects. About half of the idiosyncratic acts were 'personal', for being typical to one individual but varying among individuals. In addition, in both tasks the pragmatic component of the task ('body') was preceded by an idiosyncratic component ('head') and was also followed by an idiosyncratic component ('tail'). We suggest that the seemingly nonfunctional acts have an adaptive value, with the 'head' serving as a preparatory phase and the 'tail' serving as a confirmatory phase. In sport rituals with definite end and high stakes, the head was long and the tail absent. In everyday tasks, the head and tail were relatively short, whereas OCD rituals featured a relatively long tail. These comparisons revealed that compulsive pathologic rituals share the same structural components with daily tasks but not with sport rituals, suggesting that normal and pathologic rituals could develop on different grounds. Focusing on form and structure, we distinguish between daily motor routines, sport and pathologic rituals. Further research is required to uncover the underlying mechanisms of these repetitive motor behaviors.

Ethical statement

This study was approved by the Institutional Helsinki Committee for Human Experimentation at Tel Aviv University.

Effect of Quercetin on the Short-term Impairment of Learning Induced by X-rays in Wistar Rats. Nonlinear Regression Analysis of Morris Water Maze Latencies

R. Marina[1], P. González[2], J.P. Barrio[3]

Departamento de Ciencias Biomédicas e Instituto de Biomedicina, Universidad de León, León, Spain.
[1]raquelina_soria@hotmail.com [2]mfgons@unileon.es [3]jpbarl@unileon.es

Introduction

Neurocognitive impairment is a serious adverse effect of radiation-based therapy [1]. It is well known that ionizing radiations induce oxidative stress on target tissues, mainly through the generation of reactive oxygen species (ROS) [2,3]. Spatial learning and memory are compromised after ionizing irradiation in rats using the Morris water maze test, and antioxidant-based diets are able to improve these radiation-induced behavioural deficits [4]. Quercetin is one of the most commonly found flavonoids in the diet with reportedly antioxidant and anti-inflammatory properties, and has been shown to have radioprotective actions in mice under gamma-radiation [5].

The analysis of learning curves obtained from Morris water maze tests has been traditionally carried out using repeated-measures analysis of variance [6], analysis of linear regression slopes [7], and analysis of non-linear exponential regression coefficients [8]. However, when the learning curve is altered by applying a disruptive agent (such as X-irradiation), these approaches may not be appropriate to best explain the behavioural changes induced. Piecewise (segmented) linear regression could be useful when a change in experimental conditions can be presumed to produce a sharp transition point in the response curve.

Objectives

The present study aimed to test the hypothesis that intragastric (IG) administration of quercetin shows a positive radiomodifier effect over spatial learning and memory dysfunction induced by X-irradiation in male Wistar rats, when the behavioural tests started either before or after X-irradiation.

Materials and methods

Experiments were designed to test the response of rats to a single sub-lethal X-irradiation (6 Gy, 0.4 Gy/min) on a Maxishot 200 X-ray machine, with a source-skin distance (SSD) of 50 cm. According Spanish legislation, irradiation procedures were performed by qualified technical staff. Male Wistar rats, two-months old (Harlan, Barcelona, Spain), were divided into four groups (CQ, CV, RQ, RV, n=8 each) according IG administration of quercetin (50 mg/kg body weight) in vehicle (propyleneglycol) (Q), vehicle IG-administered (V), whole-body irradiated with 6 Gy X rays after anaesthesia with pentobarbital (R), and sham-irradiated (C). Quercetin or vehicle IG administration started five days before behavior was first measured and lasted along all experimental period.

Learning was assessed using the Morris water maze test over the six days following X-irradiation (Experiment 1). We used a pool (120 cm in diameter and 60 cm height) made on glass fiber and painted on acrylic black, filled with water at 23 °C. In Experiment 2, with a new batch of similar, naïve rats, Morris tests started as previously four days before X-irradiation, to be resumed by the three days immediately after radiation exposure at the same conditions as above. All the tests were conducted during the morning, and the order in which animals were tested was randomized in sucessive days. Each rat was allowed to swim by 60 seconds in each of four sessions starting sequentially from each pool quadrant edge with 30 seconds resting time between trials. The black, circular escape platform was only visible to the rat on the first testing day.

One B/W camera (512x512 pixels) was set up 2 m high to record the trials. We used a Dell Optiplex GX280 computer running Noldus Ethovision 3.0 (Noldus Information Technology, Wageningen, The Netherlands). The behavioral parameter studied was the latency to reaching the escape platform when the animal was placed in the pool from any of the four quadrants.

Learning data were analyzed using nonlinear regression fitting to exponential and piecewise segmented linear equations. Segmented regression fits one line to all latency (L) points with time (t) less than some value t0, and a second line to all points with t greater than t0, ensuring that the two lines intersect at t0 [9]. The model was built according the following statements:

L1 = intercept1 + slope1*t
L at t0 = slope1*t0 + intercept1
L2 = L at t0 + slope2*(t – t0)
L = IF(t<t0, L1, L2)

There was also an imposed constraint so that X0 was greater than 4 days, that is, the first line should include the data from the four first days without irradiation, and the second line the data from the next Morris tests after irradiation.

Nonlinear fitting procedures were carried out using GraphPad Prism version 5.00 for Windows (GraphPad Software, San Diego CA, USA, www.graphpad.com).

The experimental protocol used was approved by the University of León Ethical Committee, and adhered to the European Community Guiding Principles for the Care and Use of Animals.

Results

When learning started after experimental X-irradiation, the relationship latency (L) vs. time (t) fitted an exponential decay curve given by the equation L=(L0 - Plateau)*exp(-K*t) + Plateau [9]. RV rats showed the highest latency plateau level following six test days after X-irradiation, which was indicative of a learning slowdown. However, learning was improved in X-irradiated animals administered quercetin (Figure 1). When the fitting model was a straight line, the slope was taken as a measure of the learning speed. RV rats showed the slowest learning speed, which was improved by quercetin treatment in a way similar to non-irradiated animals.

In Experiment 2, learning was interrupted by X-irradiation after four days, and behavioural tests were resumed on the next three post-irradiation days. All latency data were fitted to segmented linear regression.

Upon resuming the tests, the slope of the corresponding segment showed that learning speed was depressed in all irradiated animals, with a non-significant improvement by quercetin treatment (Figure 2).

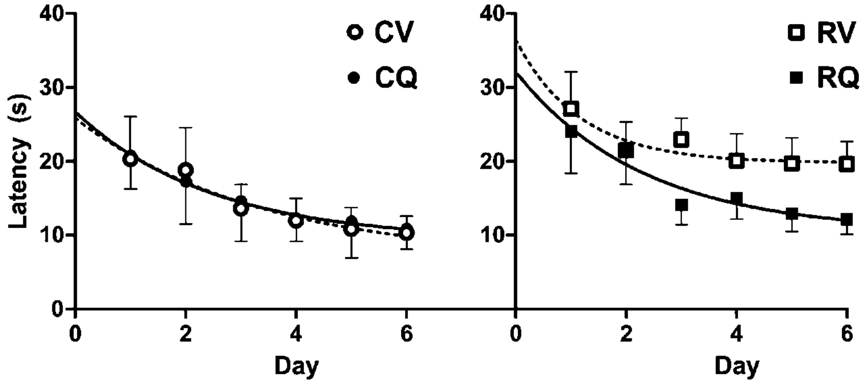


Figure 1. Exponential decay fitting of the latency to reach the escape platform vs. day for successive Morris water maze tests on rats after X-irradiation applied on day 0.

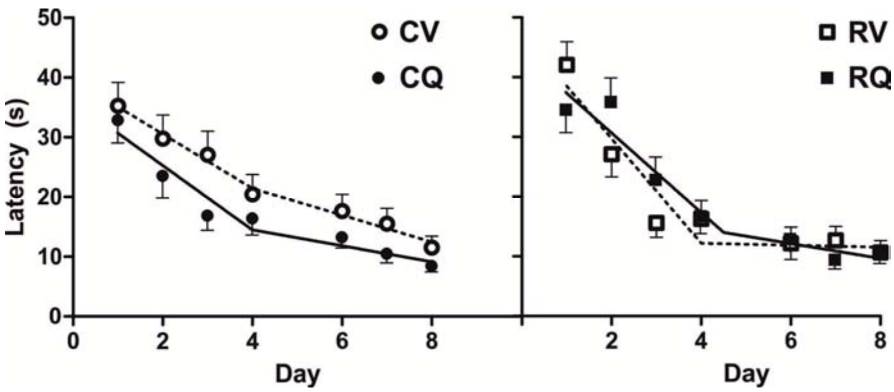


Figure 2. Segmented linear regression fitting of the latency to reach the escape platform vs. day for successive Morris water maze tests on rats before (days 1-4) and after X-irradiation (applied on day 5).

Discussion

Learning curves from Morris water maze experiments can be fitted to multiple equations, with the best results usually obtained using exponential decay equations [8]. However, when the experimental setup is such that disruption can be presumed to alter significantly the behavioral response, segmented linear regression may be a useful tool to quantify the difference in the time course of the observed process. In our work, the disruption was mainly due to X-irradiation taking place at the 5th day, but also there is an effect of the anaesthesia in CV and CQ animals (Figure 2, left). We are not aware of this fitting procedure being applied to Morris water maze data analysis in the literature.

The results of the present work show that quercetin administration before and after sub-lethal X-irradiation is able to counteract the learning slowdown induced by X rays only when the learning process starts once irradiation has been applied. These findings could be helpful to design clinical interventions aimed to reduce the impact of radiotherapy on cognitive function.

References

1. Yoneoka, Y., Satoh, M., Akiyama, K., Sano, K., Tanaka, R. (1999). An experimental study of radiation-induced cognitive dysfunction in an adult rat model. *Br. J. Radiol.* **72**, 1196-1201.

2. Halliwell, B. (1992). Reactive oxygen species and the central nervous system. *J. Neurochem.* **59**(5), 1609-1623.

3. Andrade, E.R., Cruz, I.B.M., Andrade, V.V.R., Piccoli, J.C.E., González-Gallego, J., Barrio, J.P., González, P. (2011). Evaluation of the potential protective effects of *ad libitum* black grape juice against liver oxidative damage in whole-body acute X-irradiated rats. *Food Chem. Toxicol.* **49**, 1026-1032.

4. Shukitt-Hale, B., Carey, A.N., Jenkins, D., Rabin, B.M., Joseph, J.A. (2007). Beneficial effects of fruit extracts on neuronal function and behavior in a rodent model of accelerated aging. *Neurobiology of Aging* **28**, 1187-1194.

5. Benkovic, V., Knezevic, H., Dikic, D., Lisicic, D., Orsolic, N., Basic, I., Kopjar, N. (2009). Radioprotective effects of quercetin and ethanolic extract of propolis in gamma-irradiated mice. *Arh. Hig. Rada Toksikol.* **60**, 129-138.

6. Vorhees C.V., Williams, M.T. (2006). Morris water maze: procedures for assessing spatial and related forms of learning and memory. *Nat Protoc.* **1**(2), 848–858.

7. Drouin, A., Bolduc, V., Thorin-Trescases, N., Bélanger, É., Fernandes, P., Baraghis, E., Lesage, F., Gillis, M.A., Villeneuve, L., Hamel, E., Ferland, G., Thorin, E. (2011). Catechin treatment improves cerebrovascular flow-mediated dilation and learning abilities in atherosclerotic mice. *Am J Physiol Heart Circ Physiol.* **300**, H1032–H1043.

8. Young, M.E., Clark, M.H., Goffus, A., Hoane, M.R. (2008). Mixed effects modeling of Morris water maze data: Advantages and cautionary notes. *Learning and Motivation* **40**, 160–177.

9. Motulsky, H. (2007). GraphPad Prism 5.0 Programming Guide. Graphpad Software Inc., San Diego CA, USA. <www.graphpad.com/downloads/docs/Prism5Regression.pdf >. Accessed 23 May 2012.

Measuring Pain-Related Behaviour in Four Inbred Rat Strains. Differences in Hot Plate Behaviour

M.W.H. Schaap[1], S.S. Arndt[2], L.J. Hellebrekers[1]

[1]*Faculty of Veterinary Medicine, Department of Clinical Sciences of Companion Animals, Utrecht University, Utrecht, The Netherlands.* [1]*m.schaap@uu.nl*

[2]*Faculty of Veterinary Medicine, Department of Animals in Science & Society, Utrecht University, Utrecht, The Netherlands.*

Very little is known about differences in behaviour of different strains during the hot plate test, and the strain-related effects on reported nociceptive thresholds. The hot plate is a commonly used test to study thermal (anti)nociception in rats and mice [1]. In this test, the animal is placed on a hot plate until a predetermined behavioural endpoint is observed, typically a hind-paw lick or jump response [2, 3]. Differences in latencies between animals are considered to reflect differences in nociceptive threshold. It is known that when studying drug-induced changes in nociceptive threshold, latencies of various behavioural endpoints are affected differently depending on the drugs used [2]. However, strain-related differences have to be taken into consideration as well.

The experimental protocol was approved by the Animal Experiments Committee of the Academic Biomedical Centre, Utrecht, The Netherlands. The Animal Experiments Committee based its decision on 'De Wet op de Dierproeven' (the Dutch 'Experiments on Animals Act', 1996) and on the 'Dierproevenbesluit' (the Dutch 'Experiments on Animals Decision', 1996). Both documents are available online at http://wetten.overheid.nl. Further, all animal experiments followed the 'Principles of Laboratory Animal Care' and refer to the Guidelines for the Care and Use of Mammals in Neuroscience and Behavioural Research (National Research Council 2003).

Four phylogenetically distant inbred rat strains (Elevage Janvier, Le Genet St. Isle, France; n = 14 for each strain), were used, including Wistar Kyoto (WKY), Fawn Hooded (FH), Brown Norway (BN) and Lewis (LE). Body weights (± SEM) were 225.07 (± 2.02), 202.43 (± 2.92), 163.64 (± 1.13) and 233.57 (± 2.04) gram at the time of arrival, respectively, and all rats were 8 weeks old at the time of arrival. All rats were housed individually in a clear 1500U Eurostandard Type IV S cage of 48 x 37.5 x 21 cm (Techniplast, Buguggiate, Italy). Rats were provided with bedding material (Aspen chips), *ad lib* access to food (CRM, Expanded, Special Diets Services Witham, United Kingdom), water and carton houses (Rat Corner House, Bio Services B.V., Uden, The Netherlands) and plastic tubes (Rat retreat Amber, Plexx, Elst, The Netherlands) as cage enrichment. The environment was temperature (21 ± 2 ºC) and humidity (47 ± 3%) controlled with an inversed 12:12 h light-dark cycle (lights off from 7.00 – 19.00 hrs) and a radio played constantly as background noise. Animals were handled at least three times a week by the experimenters. All testing procedures occurred in a room different than the animal housing room.

After an acclimatization period of three weeks, all animals were subjected to the hot plate test. The hot plate apparatus (Model 35100, Ugo Basile, Varese, Italy), was maintained at 50.0 ± 0.1 °C. Animals were placed in a glass cylinder of 24 cm diameter on the heated surface of the hot plate and removed immediately when either a hind paw lick or jumping response was observed [4]. The latency until one of these responses was scored. A cut-off time of 120 seconds was used to prevent tissue damage.

Two types of behavioural endpoints were measured, i.e. the hind-paw lick and the jumping response. None of the rats reached the 120 seconds cut-off time. Thermal nociceptive thresholds differed per strain ($F_{3,51} = 15.83$, $p < 0.000$; see Figure 1a). Post hoc test (Sidak) revealed that the latency time of the FH was significantly increased compared to WKY and LE, the BN showed a significantly increased latency time compared to the WKY. The type of behavioural endpoint observed differed per strain, $\chi^2(3, N = 56) = 12.36$, $p < 0.01$, see Figure 1b, but no effect of type of response was found on latency time (response x strain: $F_{3,48} = 1.23$, $p = 0.35$; response: $F_{1,51} =$

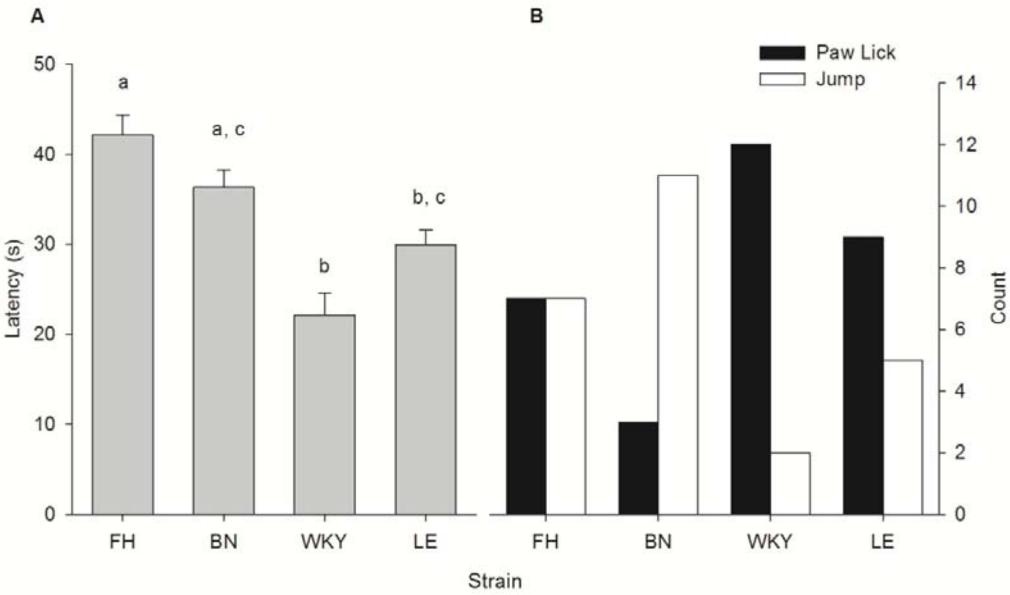


Figure 1. (A) The type of response did not significantly affect the latency time. Latency times differed between strains. Data are represented as mean ± SEM. Values which are different between strains are marked with different characters, i.e. FH = BN, FH > WKY = LE, BN = LE, BN > WKY. (Sidak; $p < 0.05$). (B) Strains differed in their behavioural endpoint. Data are represented as number of responses per strain. FH = Fawn Hooded, BN = Brown Norway, WKY = Wistar Kyoto and LE = Lewis.

0.56, $p = 0.46$). The BN showed mainly the jumping response, while within the WKY strain paw licks were most prevalent. The FH and LE showed both behavioural endpoints with equal prevalence.

As the strains differed in body weight, one concern is the potential influence of body weight on latency. However, a recent study showed that this effect is weak and therefore unlikely to fully explain differences in latency up to 50% in this study [5]. In conclusion, when selecting behavioural endpoints the animal strain should be considered. Since the type of observed behavioural endpoint did not influence the baseline latency (i.e. paw licks and jumping responses occurred at the same latency) it is recommended to take both types of behaviours as endpoint for the hot plate test as indicator for the baseline (i.e. non-drug related) thermal nociceptive threshold.

References

1. Le Bars, D., Gozariu, M. & Cadden, S.W. (2001). Animal Models of Nociception. *Pharmacological reviews* **53**, 597-652.

2. Carter, R.B. (1991). Differentiating analgesic and non-analgesic drug activities on rat hot plate: effect of behavioral endpoint. *Pain* **47**, 211-220.

3. Espejo, E.F. & Mir, D. (1993). Structure of the rat's behaviour in the hot plate test. *Behavioural brain research* **56**, 171-176.

4. Allen, J.W. & Yaksh, T.L. (2004). Assessment of acute thermal nociception in laboratory animals. In J.M. Walker (ed), *Pain Research: Methods and Protocols*. Humana Press Inc., Totowa, NJ, 111-123.

5. Gunn, A., Bobeck, E.N., Weber, C. & Morgan, M.M. (2011). The Influence of Non-Nociceptive Factors on Hot-Plate Latency in Rats. *The Journal of Pain* **12**, 222-227.

Burrowing as a Non-reflex Behavioral Readout for Analgesic Action in a Rat Model of Knee Joint Arthritis

Angela Robens, Kris Rutten, Klaus Schiene, Tobias Pasqualon, Thomas Christoph

Grünenthal GmbH, Global Preclinical Research and Development, Department of Pharmacology Pain, Aachen, Germany. angela.robens@grunenthal.com

Aim of investigation

Non-evoked readouts for assessment of analgesic efficacy against spontaneous pain have been proposed to improve the predictive validity of preclinical models for analgesia. As such, the development and validation of novel readouts which go beyond reflex-withdrawal assays and which rely more on innate animal behavior is necessary. Here we demonstrate that the normalization of innate rodent behavior suppressed by pain, such as burrowing, can be a useful alternative behavioral readout for assessment of analgesic efficacy. In a Complete Freund's Adjuvant (CFA)-induced model of knee-joint arthritis the effects of naproxen, ibuprofen and pregabalin were compared in a weight bearing, an open field and a burrowing assay.

Methods

Male Sprague Dawley rats were injected with 150µl CFA (2mg/ml) into the knee, three days before testing in weight bearing, open field and burrowing assays. Naproxen (0, 50, 100 mg/kg, ip), ibuprofen (0, 30, 100 mg/kg s.c.) and pregabalin (0, 10, 30 mg/kg, i.p.) were administered 30, 90 or 60 minutes, respectively, before testing. Weight bearing on each hind leg was determined using a rat incapacitance tester, horizontal locomotor activity and rearings were recorded in an open field for 5 minutes and burrowing performance was measured by the amount of gravel left in a hollow tube after 30 minutes of presentation to the rat. All experiments were performed in accordance with company, national and international regulations and laws for animal care and welfare as well as the recommendations and policies of the International Association for the Study of Pain (Zimmermann, 1983).

Results

CFA-induced knee-joint inflammation caused marked reductions in weight bearing in the inflamed leg, open field activity and burrowing. Naproxen, ibuprofen and pregabalin were efficacious in normalizing weight bearing, but in the open field horizontal locomotor activity was not normalized by any treatment condition and rearing behavior was reinstated only by ibuprofen (100 mg/kg) and not by the other treatment conditions. On the other hand, naproxen (100 mg/kg), ibuprofen (31.6 mg/kg) and pregabalin (10 mg/kg) effectively reversed CFA-induced deficits in burrowing behavior.

Conclusions

These experiments suggest that measuring burrowing performance is an alternative non-reflex readout relying on innate rodent behavior which is affected by pain and can be pharmacologically manipulated. The burrowing assay appears to be more sensitive to pharmacological treatment than locomotor activity assessments in an open field assay. Furthermore, as opposed to reflex withdrawal or weight bearing assays, burrowing has the benefit of dissociating selective analgesic doses of a drug from doses that induce locomotor impairment in the same animal.

Do Outbred ICR:CD1 Mice Form Attentional Sets in a Bait Digging Attentional Set-Shifting Procedure?

A.-M. Linden, C. Procaccini, M. Lehtomäki

Institute of Biomedicine, Pharmacology, University of Helsinki, Helsinki, Finland. anni-maija.linden@helsinki.fi

In order to analyze prefrontal function after pharmacological treatments, we wanted to use an attentional set shifting test for the outbred albino ICR:CD1 mice. The attentional set shifting test is successfully used in rats, and recently also in mice, although there has been controversy whether the mice can form attentional sets. Also, the methods used for mice vary a lot between laboratories. Here, we used a food bait digging procedure described for the rats [1] and later adapted for the C57Bl/6J mice [2].

In our procedure, the adult male ICR:CD1 mice (Harlan) were food restricted to 85-90% of their normal weight, habituated to testing arenas and trained to dig scented materials in cups before receiving a bait. During the testing phase, the mice were trained to receive the bait by digging materials in a correct cup based on cue which was either digging material (tactile, visual cue) or odor (olfactory cue) added in the materials. Thus, the bait was not hidden but given after a correct choice. The procedure consisted tasks of simple discrimination (day 1), compound discrimination and compound discrimination reversal (day 2), intradimensional shifts 1-5 (day 3-5) and intradimensional shift 6 and an extradimensional shift (day 6). The procedure took 6 days to perform, each day started with a repetition of the last stage of the previous day. A criterion to pass each stage was 8 out of 10 and last 6 consecutively correct choices. If the mouse did not complete stages of one day within 60 min or if the latency to make a choice was longer than 3 minutes, it was returned to home cage at least for one hour. The animal experimentation was approved by Animal Experiment Board of the Regional State Administrative Agency for Southern Finland.

The ICR mice could perform the discrimination tasks, and as expected the reversal task required significantly more trials than the preceding compound discrimination. We found intradimensional shifts easier when odors were used as relevant cues compared to material cues. Consistently, an extradimensional shift from odor to material was more difficult than the preceding intradimensional shift, while this was not the case for the extradimensional shift from material to odor.

Currently, we are testing for conditions to refine the protocol in order to be able to use the ICR:CD1 mice in pharmacological studies in this test measuring cognitive flexibility. As a conclusion, this test may be useful to analyze mouse prefrontal function. However, it is time consuming and a low throughput method limiting its applicability.

References

1. Birrell, J.M., Brown, V.J. (2000). Medial frontal cortex mediates perceptual attentional set shifting in the rat. *Journal of Neuroscience* **20**, 4320-4324.

2. Bissonette, G.B., Martins, G. J., Franz, T.M., Harper, E.S., Schoenbaum, G., Powell, E.M. (2008). Double dissociations of the medial and orbital prefrontal cortical lesions on attentional and affective shift in mice. *Journal of Neuroscience* **28**, 11124-11130.

A-KinGDom Program: Agent-Based Models for the Emergence of Social Organization in Primates

R. Dolado[1], V. Quera[2], F.S. Beltran[3]

[1]*Adaptive Behavior and Interaction Research Group, Department of Behavioral Science Methods, University of Barcelona, Barcelona, Spain.* [1]*ruth.dolado@ub.edu*

[2,3]*Institute for Brain, Cognition and Behavior (IR3C), Adaptive Behavior and Interaction Research Group, Department of Behavioral Science Methods, University of Barcelona, Barcelona, Spain.* [2]*vquera@ub.edu* [3]*francesc.salvador@ub.edu*

Abstract

Social organization in primate societies is a complex and self-organized phenomenon that integrates kinship, competition and cooperation behaviors, and which can be explained using simple rules according to the adaptive behavior approach. Although it is not currently a common approach in Primatology, some incipient agent-based simulations have been used in order to study the emergence patterns of social organization in primates. Hemelrijk [1] presented an agent-based model, called DomWorld, where dominance interactions (i.e., dyadic agonist encounters between two agents) determines both the dominance hierarchy and the spatial distribution of group members observed in macaque societies. More recently, and based on the co-variation hypothesis [2], Puga-González et al. [3] developed GrooFiWorld, an agent-based model that is an extension of DomWorld and includes agonistic and affiliative behaviors in order to reproduce the emergence of patterns of social organization observed in macaque societies.

Dolado & Beltran [4] obtained data from observation of a captive group of *Cercocebus torquatus* (a species that is phylogenetically close to macaques) and compared them with data from GrooFiWorld agent-based simulations [3]. The results suggested that, although the GrooFiWorld model can be used as a starting point to study the flexibility observed in behavior patterns in macaques and close species, some other parameters should be taken into account in order to determine the emergence patterns of social organization in primates. A good candidate to include as a factor in GrooFiWorld model is kinship. Many studies have identified kinship (i.e., the genetic relationship among group members) as a factor that can account for certain individual differences in aggressive and affiliative patterns observed among members of the same group [5]. Moreover, the females of the macaques, baboons or mangabeys remain in their natal groups and spend much time in contact with their relatives searching for food, in grooming sessions or by taking care of offspring. Therefore, including kinship in a model of social organization may modify the durations and frequencies of affiliative interactions in the group [6] and contribute to achieve a better fit of the model.

We have developed an agent-based program called A-KinGDom, which is written in C and Delphi and Open GL, and runs in Windows. The program implements DomWorld [1] and GrooFiWorld [3] models, plus a new KinWorld model that is an extension of GrooFiWorld which includes the kinship factor in order to modulate affiliatives interactions. The simultaneous application of dyadic interaction rules for all agents in the current model allows us to simulate the emergence of patterns of social organization in a group of primates. The KinWorld model includes dominance and affiliatives interactions, as well as kinship relations between agents (see Figure 1). According to the DomWorld and GrooFi World models, whenever an agent does not perceive any other agent within its personal distance, grouping rules come into effect. However, if one agent enters the personal distance of another, a social interaction may or may not take place. When two agents meet, one of them decides whether or not to engage in either a dominance or an affiliative interaction. Then, if an agent expects to win, dominance interaction takes place: if the agent's current dominance value is greater than a random value, it wins the interaction. If its dominance value is lower, it loses. An agent only considers grooming its partner if it expects to be defeated. Grooming behavior is modeled according to the level of *anxiety:* when the agent's *anxiety* value is greater than a random value, the agent grooms its partner; otherwise, the agent displays non-aggressive proximity without interacting socially with the other agent. In the KinWorld model, kinship is defined as $kinfactor=1+r^2$, where r is the coefficient of relatedness or the level of consanguinity between two given

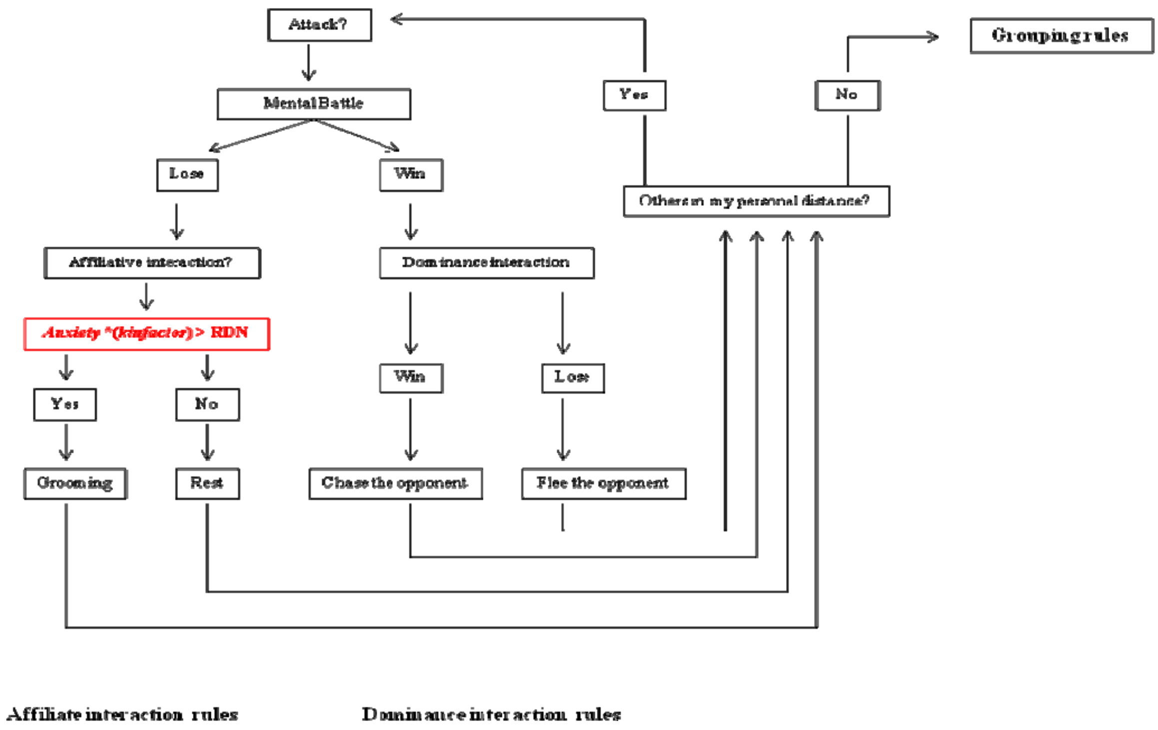


Figure 1. Social interactions according to the KinWorld model.

agents; *kinfactor* multiplies *anxiety*, thus increasing the likelihood of executing an affiliative interaction between two agents that have a close kinship.

In this study we also present results from a set of simulations comparing the GrooFiWorld and KinWorld models, and analyze how they fit with data obtained from a captive group of seven individuals of *Cercocebus torquatus* [4]. As expected, data obtained when the KinWorld model is simulated provide a better fit.

The agent-based models implemented in A-KinGDom program can be adapted to different biological conditions in order to compare results in different groups and species of primates. A-KinGDom simulations provide data that characterize the main quantitative measurements of social organization (i.e., gradient of hierarchy, undirectionality of aggression, grooming reciprocation or grooming up the hierarchy), which can be compared with empirical data. The comparison provides evidence for the plausibility of the KinWorld model and is useful for making predictions about patterns of the social organization in other groups of the subfamily Cercopithecinae (i.e., macaques, baboons or mangabeys).

Keywords. Agent-Based Models, Social Structure, Primates.

References

1. Hemelrijk, C.K. (2004). The use of artificial-life models for the study of social organization. In B. Thierry, M. Singh, & W. Kaumanns (Eds.), *Macaque societies: Amodel for the study of social organizations*, Cambridge, UK: Cambridge University Press, 295-318.

2. Thierry, B. (2004) Social epigenesis. In B. Thierry, M. Singh, & W. Kaumanns (Eds.), *Macaque societies: A model for the study of social organizations*, Cambridge UK: Cambridge University Press, 267-290.

3. Puga-González, I., Hildenbrandt, H., Hemelrijk, C. K. (2009). Emergent patterns of social affiliation in primates, a model. *PLoS Computational Biolology*, **5**(12): e1000630. doi:10.1371/journal.pcbi. 1000630

4. Dolado, R, Beltran, F.S. (in press). Emergent patterns of social organization in captive *Cercocebus torquatus*: Testing the GrooFiWorld agent-based model. *Journal of Biosciences.*

5. Chapais, B. (2004). How kinship generates dominance structures: a comparative perspective. In B. Thierry, M. Singh, & W. Kaumanns (Eds.), *Macaque Societies: a model for the study of social organizations*, Cambridge, UK: Cambridge University Press, 186-204.

6. Gouzoules, S. (1984). Primate mating systems, kin associations, and cooperative behaviour: Evidence of kin recognition? *Yearbook of Physical Anthropolopogy*, **27**, 99-134.

Automated Detection of Aberrant Behaviour of Mice on the Rotarod: Use of EthoVision® XT

M. Mahieu[1], R. Willems[1], L. Hoekstra[2], L. Ver Donck[1]

[1]Dept Neuroscience, Janssen Research and Development, a Division of Janssen Pharmaceutica NV, Beerse, Belgium. mmahieu@its.jnj.com

[2]Noldus Information Technology, Wageningen, The Netherlands

Introduction

Impaired motor coordination in man due to medication can have an important impact on private life and job performance. Antipsychotic drugs in particular are known to induce extrapyramidal motoric symptoms, causing disturbed coordination of body movements. The rotarod test is widely used to evaluate drug effects on motor coordination in rodents. The principle of this test is that rats or mice are first trained to walk on a rod rotating at a certain speed. Once the animals have learned this, the effect of a test-compound on their motor performance is evaluated. Animals experiencing impaired motor coordination are unable to cope with the rotating rod and will drop off when the rotation speed exceeds their motor coordination capacity (see Figure 1). The more disturbed the animals are, the sooner they fall off the rod. However, mice show a particular coping behaviour when experiencing motor coordination problems on the rotarod: to prevent dropping off, mice can grip themselves to the rod and turn around without falling off (see Figure 2).

This behaviour then results in late (or no) falling off the rotarod, which would incorrectly indicate that the compound tested, did not disturb motor coordination. The number of times mice turn around on the rotarod therefore represents a secondary measure of disturbed motor coordination. This turnaround behaviour can be determined by visual observation, but this is challenging when 5 mice are tested simultaneously on the rotarod.

The objective of this study therefore is to automate the determination of this turnaround behaviour using video recording of the experiment in combination with image analysis software.

Methods

Male NMRI-mice (body weight 22-24 g, Charles River) were habituated to the environment for 1 week. They were first trained to walk on the rotarod (Med Associates) at constant rotation speeds in 5 min long sessions at

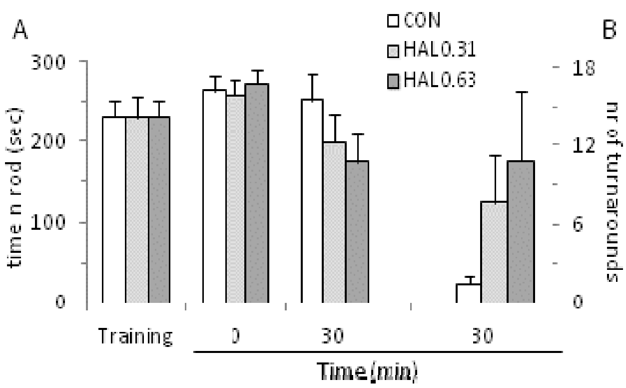


Figure 1. A: Effect of haloperidol (HAL, 0.31 or 0.63 mg/kg) on motor performance of mice on the rotarod (time spent on the rotarod); Time (min) is time after drug administration. B: number of times a mouse turns around on the rotarod. While HAL has only a very small effect on time on rod (A), the mice show a strong increase in number of turnarounds.

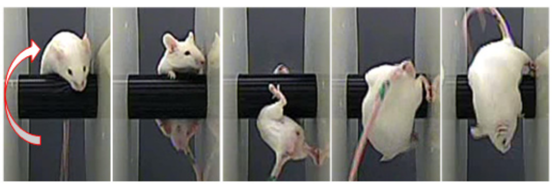


Figure 2. Time sequence (left to right) of a mouse showing coping behaviour on the rotarod by gripping itself to the rod and making a turnaround (front view). The arrow indicates the direction of rotation of the rod.

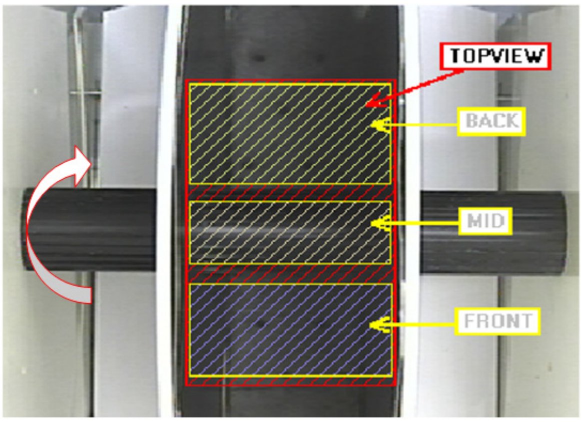


Figure 3. Top view image of one lane of the rotarod. Overlays show the area of interest (red) and the 3 distinct individual zones (yellow) covering the rod (mid) and the front and back zones as determined in EthoVison® XT. The width of the red area was set to the width of one lane of the rotarod and the length was set to 3.7 times the diameter of the rod. The arrow indicates turning direction of the rod.

16, 20 and 24 rotations per minute (rpm) at 30 min intervals; mice were placed back on the rod each time they fell off, until the 5 min session was completed. Then a session was performed with the rotarod rotating at an incremental speed starting at 4 rpm and accelerating over a 5 min period up to 40 rpm. The next day, a dose of the test-compound was administered, and mice were tested at 30 min intervals in 5 min sessions at accelerating speed. The time at which the mice fell off the rotarod was recorded, and the number of turnarounds was counted by visual observation.

To evaluate automated determination of turnaround behaviour, a high resolution video camera (Samsung SDC-435) was positioned 45 cm above the rod to record the session. EthoVision® XT (Noldus Information Technology, Wageningen, The Netherlands) was used to analyse the video images. A region of interest area was identified around the rod and this area was divided in 3 individual zones: the rod in the middle and the front and back area's just outside of the rod (Figure 3). The background was covered with black material such that a white mouse could be readily determined by EthoVision® XT as an object. The surface area and the center of gravity (CG) of the object were determined, as well as the occurrence of the latter in one of the zones. Also the sequential transition of the CG from middle to back to front and to middle again was determined: this sequence of events indicates a turnaround manoeuvre of the mouse. The Institutional Ethical Committee on Animal Experimentation approved the experimental protocol, in compliance with Belgian law (Royal Decree on the protection of laboratory animals dd. April 6, 2010) and the facilities are accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC).

Results

Figure 4A shows the surface area of the object (mouse) detected in the area of interest (cf. red area in figure 3). The downward peaks represent phases during which the surface area detected suddenly decreases: this is when the mouse makes a turnaround (marked by red squares) and thus disappears for a few moments behind the rod.

However, another example in figure 4B shows that the downward peaks do not always accurately represent a turnaround manoeuvre. Surface area tracking is therefore not an appropriate measure to determine turnaround behaviour.

The zone-transition option in EthoVision® XT determines when the CG of the object detected displaces from one zone to another. When the mouse turns around on the rod, the position of its CG should change from the front to the mid zone (i.e. the mouse can only appear in the front zone and then in the mid zone when it makes a complete turnaround). This routine indeed accurately detected 12 and 7 turnarounds in figures 4 A and B

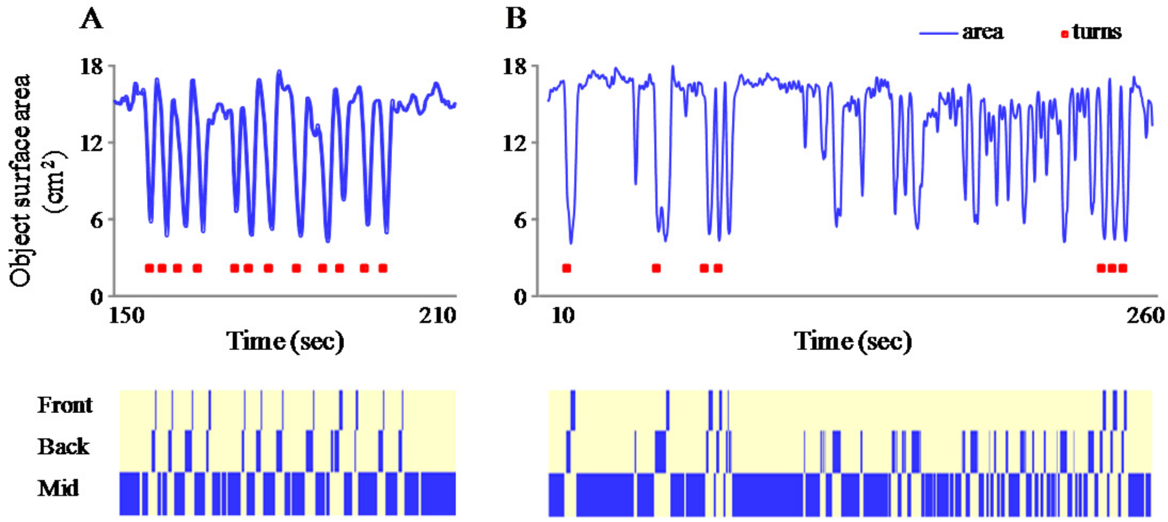


Figure 4. Top panels: examples of surface area tracking from 2 different mice treated with haloperidol; A: Sudden decreases of surface area detected by EthoVision® XT (downward peaks) correlate with visual observation of the mouse making turnarounds (red marks). B: Downward peaks recorded from this mouse do not correlate very well with turnarounds observed. Bottom panels are heat maps of occurrence of the CG of the object detected in the mid, back or front zone (blue marks). A sequential occurrence of the CG in the mid, back, front and mid zone again is consistent with visually observed turnarounds made by the mouse.

respectively. However, a limitation of this approach is that the exact timing of when these occur during the course of the experiment is not indicated.

A better option therefore is to determine the sequential transition of the CG from mid to back to front to mid again. Output from EthoVision® XT reporting the presence or absence (i.e. 1 or 0) of the CG in each of the 3 zones during a session was exported to Excel. Conditional formatting was then used to generate a heat map, colouring the field blue when the CG was detected in a zone (see bottom panels in figure 4). The horizontal axis represents the time dimension as in the graph above. A sequential appearance of the CG changing from the mid to the back to front and to the mid zone again is detected by an algorithm and reports this sequence as a turnaround. This procedure accurately detected the turnarounds as visually observed.

Conclusion

Monitoring the time that mice stay on the rotarod is not sufficient to determine impaired motor coordination in this species. EthoVision® XT can be readily used to automatically determine turnaround behaviour of mice on the rotarod. This measure is a critical parameter to assess disturbance of motor coordination in mice.

Measuring Bread Use in a French Restaurant. A Naturalistic Approach: Grid Analysis for the French Culture

C. Iborra-Bernad[1], O. Wathelet[2], A. Giboreau[3]

[1]*Universitat Politècnica de València, Valencia, Spain. maibber@aaa.upv.es*

[2]*SAS Seb, Selongey, France. owathelet@seb.fr*

[3]*Centre de Recherche de l'Institut Paul Bocuse, Château du Vivier, Ecully, France. agnes.giboreau@institutpaulbocuse.com*

Abstract

Bread is a staple food in France; its consumption in French culture is common during meals at home and in restaurants. However, the way bread is used in those settings is unknown, notably as a mean to perform non ingestive tasks. 117 people were observed in a restaurant, most of whom ate bread (n=106). Observation revealed that bread has non alimentary roles during the meal as well as being eaten. The main operations observed were: "mopping up" gravy or sauces and to support the picking up of other food. The knowledge of those operations conducted with a food product leads to better understanding of the many roles of bread during a meal. The aim of this study was to create a grid analysis of bread operations in a table in naturalistic situations based on French culture.

Keywords. bread, usage, operations, observational methodology, experimental restaurant.

Introduction

In France, wheat bread is a major contributor to the diet and a symbol of gastronomy [1]. Bread consumption per day in France has decreased during the twentieth century from 900 g in 1900 to 160 g in 1995 [2]. However 97.6% of French people still eat bread during the day [3]. Bread may be eaten as a single product, but it is usually consumed along with other dishes, during a meal. In francophone culture, breakfast could be composed of a slice of bread spread with butter or slices of bread filled with other ingredients as a sandwich or just the bread [4,5].

In all of cases the function of bread is its consumption. However, this product has different features allowing other uses during a meal such "mopping up" or use it as a base for a bar snack. Despite the rapid growth of new bakery products, research into bread usage by consumers is scarce. The decrease of its consumption is a cause for concern in the bakery industry, leading to new ways of understanding how food is used not solely as a ingredient but as a more complex object, permitting to dense and rich manipulation. In this case study, we rely on a design ethnogaphy approach which approximates the immersion methods of traditional ethnography, to deeply experience and understand the user's world for design empathy and insight [6]. In this aspect, Tony Salvador, Genevieve Bell, and Ken Anderson [7] describe design ethnography as being "a way of understanding the particulars of daily life in such a way as to increase the success probability of a new product or service or, more appropriately, to reduce the probability of failure specifically due to a lack of understanding of the basic behaviors and frameworks of consumers." In others words, design ethnography studies interactions between people and relevant parts of their environment and describes tension between planned use (the one expected by bakers or waiters) and ones performed in real life situations by actual consumers. In this research we hypothesize that bread is not only a foodstuff but is also a tool used during the meal situation for a wide range of purposes. To capture and describe those uses we based our data production on the full video recording of meals in a restaurant. In classical ethnographic effort, as stated by Pink [8], describtion ought to be based on an in-depth understanding of the meaning of action, which is made in our case by describing the interaction between eaters, bread and the others parts of the environment. However, we aimed to add a quantitative dimension to the anaysis by connecting this ethnographic observation to their frequency by combining these approaches. The objective of this study was to create a grid analysis of the bread operations on the table in a naturalistic situation based on french culture using a video recording permitting the capture of real-life complexity.

Methods

Participants and situation

117 clients of Experimental Restaurant (Institut Paul Bocuse, Ecully, France) participated in the study. 106 participants had physical contact with bread while sitting at tables. Clients were real customers who booked the table by themselves in the Experimental Restaurant and paid 20 Euros for a meal. Before starting, an ethical committee approved the study and participants were made aware of their participation into a study without revealing its objectives. A consent form for participation and video recording in the dining room was signed. Participants were adults between 20 and 85 year of age with on average of 44.7 years old.

Participants were received by the “maître d’hôtel” (headwaiter) as it is usual in a moderately priced restaurant. A total of eight sessions were conducted to study all subjects. Menus were composed of an “*amuse bouche*”, a starter, a main course and a sweet dessert. The style of the cookery was “haute cuisine” or gastronomic meal.

Products

White bread (WB) and whole wheat bread (WWB) were supplied daily by the same bakery. Both types of bread were proposed to favour the consumption by all consumers. WB was made following the traditional French method. WWB was made according to the recipe of GMS Meunerie (France), the supplier of flour used for “Pain complet”. A precision balance (Balance Aip, model NHB-1200) was used to calculate the approximate weight of each slice of bread: 7±1g for WB and 12±1 g for WWB. Bread slices were presented in dishes in the middle of the table.

Measurement

Cameras located on the ceiling of the experimental restaurant were used to monitor the participants during the meal. Six cameras were used to observe six different tables during the meals, and the number of participants was between 2 and 6 people per table. Recorded video films were employed to analyse customer behaviour towards the bread. Vlc media player and Excel 2007 were employed simultaneously to watch and code the behaviours observed, respectively. On consumption of each piece of bread, the following information was checked to create the grid analysis (Table 1):

1. At what moment of the meal was the bread consumed?
2. What is the kind and form of the dish?
3. Before eating the bread, was any operation carried out with the bread?
4. What are the elements taking part in the interaction?
5. What could be the objective of this usage?
6. Is there more than one usage?
7. Description of the sequence.

Results

Observations led us to define the categories in answer to the previous questions. Table 1 shows the codes to create the behaviour bread grid analysis and Figure 1 illustrates one of the multi-operations with the same piece of bread observed and an example of a grid analysis.

Table 1. Grid analysis about operations conducted with bread.

CRITERIA	CODES
(1) Stage of the meal	W: Waiting. A: Amuse bouche. S: Starter. MC: Main Course SD: Sweet Dessert
(2) Type of the dish	B: Bowl P: Plate SP: Soup dish
(3) Form of the dish	Ro: Round S: Square Re: Rectangular
(4) Elements taking part into the Interaction	BH: Bread-Eater's Hand BHP: Bread-Eater's Hand-Plate BHF: Bread-Eater's Hand-Foodstuff BHC: Bread-Eater's Hand-Cutlery BHFC: Bread-Eater's Hand-Foodstuff-Cutlery
(5) Operation with the bread	E: To eat. M: To mop up. Su: As a support to pick up other foods. Sp: As a base for spreading. Ck: To clean the knife. Sb: Function as a base for bar snack. D: To dip bread in soup. Cf: To clean the fork.

Times of different operations were counted and sequenced. Some operations chains were repeated. Of the 106 people who ate bread, the percentage of participants who carried out an additional operation with the bread was: 82.1% to mop up, 34.0% as a support to pick up other food, 20.8% to clean cutlery, 10.4% to spread other sauce and 4.7% use the bread as a base for a bar snack.

The operations repeated most times were "to mop up and to eat the bread" (M+E), followed by employing bread "as a support to pick food up from the dish and to eat the bread" (Su+E) (Table 2).

The chain of operations most repeated were the bread as a support and after employing it to mop up, another sequence repeated several times was cleaning the knife, mopping up and to eating the bread.

Criteria	(1)	(2)	(3)	(4)	(5)
Codes	S	SP	Ro	B;H;F;C	Ck,Su,M, E

Figure 1. Example of multi-operations with the same piece of bread.

Table 2. Sequence and number of operations repeated during the meal by participants (n=106).

Sequence	Times (%)
E	1082 (62.4)
M+E	436 (25.2)
Su+E	87 (5)
Su+M+E	34 (2)
Sp+E	28 (1.6)
Ck+E	25 (1.4)
Ck+M+E	20 (1.2)
Sb+E	10 (0.6)
Ck+Su+E	6 (0.3)
D+E	2 (0.1)
Ck+Su+M+E	1 (0.1)
Ck+Sp+E	1 (0.1)
Cf+Su+M+E	1 (0.1)

E: To eat; M: To mop up; Su: as a support to pick up other food; Sp: to spread; Ck: to clean the knife; Sb: Function as a base of snack bar; D: to dip bread in soup; Cf: to clean the fork.

This generic table identifies several uses of bread and point out some of its main features from the standpoint of French Customers. The qualitative analysis of video sequences show for instance some shift in priority of the quality of bread. As a tool to pick up other food, consistency should be dense; while at the oppositive its mopping up role rely on softer properties. Cleaning, dipping or spreading actions are in an intermediate state. Interestingly, there are limitations to the sequence of events. Of course, eating is always the end process, but mopping up is always next to last. On the contrary, cleaning a utensile (mostly the knife) appears at the begining of the process.

Conclusions

In this work, operations with bread by the customers in the experimental restaurant have been investigated. Observation of the different situations revealed some logical sequences in the clients’ behaviour over time. It was observed that consumption is not the only operation conducted in the dining room. Mopping up and being as support for picking up other food are operations that usually precede consumption. Bread could be related with the correct cleaning of other elements on the table: dish, fork and knife. Also, the bread was employed as a support to pick up other foodstuffs from the dish to avoid contact with fingers. This pattern of operations would be considered as a base to develop a more complex study about the operations regarding the kind of traditional bread of a region, culture and situations. This effort will also need extra time to collect oral data about the meaning of actions, such as for instance the use of video to ellicit the meaning of an action [9].

References

1. Steele, R. (2006).The French way: the keys to the behavior, attitudes, and customs of the French. McGraw-Hill Companies,.

2. EPI, <http://www.espace-pain-info.com/fiches/public.php>.Accessed 29 December 2011.

3. Observatoire du Pain. (2011). Quelle place pour le pain dans l’alimentation des Français en 2010? *La newsletter de l'Observatoire du Pain* **8**.

4. Abramson, J. (2007). Food culture in France. Greenwood Press.

5. Bellisle, F., Dalix, A.M., Mennen, L., Galan, P., Hercberg, S., de Castro, J.M., Gausseres N.. (2003). Contribution of snacks and meals in the diet of French adults: a diet-diary study. *Physiol Behav* **79**,183-189.

6. Hanington, B., Martin, B. (2012). *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. Rockport Publishers.

7. Salvador, T., Bell, G., Anderson, K. (1999). Design ethnography. *Design Management Journal (Former Series)* **10**, 35-41.

8. Pink, S. (2007). *Doing visual ethnography: images, media and representation in research.* Sage Publications.

9. Gore, G., Rix-Lièvre, G., Wathelet, O., Cazemajou, A. (2012). Can the body talk? Interviewing to understand bodily experience. In *Anthropology and the Interview* (Ed. J. Skinner). Berghahn Books, Oxford.

Lying Behaviour of Dairy Cows in Cubicles

E. van Erp-van der Kooij[1], O. Almalik[1], F.J.C.M. van Eerdenburg[2]

[1]*Department of Applied Biology, HAS Den Bosch University of Applied Sciences, 's Hertogenbosch, The Netherlands.*
[1]*EL@hasdb.nl*

[2]*Department of Farm Animal Health, Utrecht University, Utrecht, The Netherlands.*
[2]*F.J.C.M.VanEerdenburg@uu.nl*

Improving dairy welfare by improving management

In dairy barns, cows are restricted in their lying behaviour in several ways. A better designed lying area will ensure a better welfare and higher production levels for dairy cows [1, 4-10]. Besides the design of the cubicles, on-farm climate and other farm and herd characteristics such as stocking density and number of cubicles available per cow, determine much of the lying behaviour of cows [2].

Live observations of lying cows are difficult to obtain, because many cows are unused to unfamiliar observers and will react behaviourally (i.e. change their posture) to the presence of these observers. It is, therefore, difficult to observe lying behaviour and investigate relations between lying behaviour and other factors in a reliable manner.

In this study, we have recorded lying behaviour by using video recordings. On 14 farms, 24-hour recordings were made of lying behaviour of dairy cows by using 2 webcams per farm. No researchers were present during the filmed period, so that behaviour of the cows was undisturbed. Lying behaviour was analyzed using the Observer 10.5 (Noldus Information Technology bv). A simple ethogram was used including the behavioural elements standing, lying down, lying and standing up. Video files were coded using continuous sampling resulting in mean and frequencies of lying bouts, time used for lying down and for standing up, and time spent standing in the cubicle ('standing idle'). Standing and lying times are important indicators for cow welfare [3].

Temperature was recorded during 24 hours by using data loggers. Additional to these data, cubicle design, including bedding material, was recorded, other farm and herd characteristics were recorded using a cow comfort score protocol [11] and some climate measurements including air speed and humidity have been done, just before the cameras were installed.

On another 12 farms, no video recordings were made but a digital camera was used to take pictures of the posture of lying cows in cubicles, to see whether lying posture, recorded from the pictures, was related to farm and herd characteristics. We expected that taking pictures was a fast enough method not to influence cows too much. On these farms, cubicle design and other farm and herd characteristics were also recorded.

The relations of lying times, standing time and time used for standing up and lying down, with cubicle design parameters, were calculated using linear regression. The relations of lying posture with farm and herd characteristics was calculated using binary logistic regression. A comparison was made with available data of lying behaviour of cows kept outdoors in pasture. Furthermore, relations between lying behaviour and climate measurements were calculated using a repeated measures analysis.

These results can be used to further optimize cubicle design and management of dairy cows.

References

1. Fregonesi, J. A., von Keyserlingk, M.A.G., Viera, D.M., Weary, D.M. (2007a). Effects of bedding quality on lying behavior of dairy cows. *Journal of Dairy Science* **90**, 5732–5736.

2. Fregonesi, J.A., Tucker, C.B., Weary, D.M. (2007b). Overstocking reduces lying time in dairy cows. *Journal of dairy science* **90**, 3349-3354.

3. Haley, D.B., de Passillé, A.M., Rushen, J. (2001). Assessing cow comfort: effects of two floor types and two tie stall designs on the behaviour of lactating dairy cows. *Applied animal behavior science* **71**, 105-117.

4. Haley, D.B., Rushen, J., de Passillé, A.M. (2000). Behavioural indicators of cow comfort: activity and resting behaviour of dairy cows in two types of housing. *Canadian Journal of Animal Science* **80**(2), 257-263.

5. Lidfors, L. (1989). The use of getting up and lying down movements in the evaluation of cattle environments. *Veterinary Research communication* **13**, 307-324.

6. Manninen, E., de Passille, A.M., Rushen, J., Norring, M., Saloniemi, H. (2002). Preferences of dairy cows kept in unheated buildings for different kind of cubicle flooring. *Applied Animal Behaviour Science* **75**, 281–292.

7. Norrin, M., Manninen, E., de Passillé, A.M., Rushen, J., Munksgaard, L., Saloniemi, H. (2008). Effect of sand and straw bedding on the lying behaviour, cleanliness, and hoof and hock injuries of dairy cows. *Journal of dairy science* **91**, 570-576.

8. Rushen, J., Haley, D., de Passille, A.M. (2007). Effect of softer flooring in tie stalls on resting behavior and leg injuries of lactating cows. *Journal of Dairy Science* **90**, 3647–3651.

9. Tucker, C.B., Weary, D.M., Fraser, D. (2003). Effects of three types of free-stall surfaces on preferences and stall usage by dairy cows. *Journal of Dairy Science* **86**, 521–529.

10. Wechsler, B., Schaub, J., Friedli, K., Hauser, R. (2000). Behaviour and leg injuries in dairy cows kept in cubicle systems with straw bedding or soft lying mats. *Applied Animal Behaviour Science* **69**, 189-197.

11. Eerdenburg, F.J.C.M. van, Saltijera-Oaxaca, J. and Vazguez-Florez, S. (2009). Increasing milk yield by increasing cow comfort. *Proceedings of the European Forum Livestock housing for the future* (Lille, France, 22-23 October 2009), 143-148.

School Safety Architecture – How to Measure Perceptions of Safeness

M.D. Bredero

Saxion University of Applied Sciences, Deventer, The Netherlands. mail@architectmaartendouwebredero.com

Introduction

Housing children during their school-going years has been a challenge for the longest of times. However, our society is changing and becoming increasingly volatile with shorter attention spans due to, amongst other causes, social media. The associated cost of juvenile delinquency because of children dropping out of school is hard to estimate, but nevertheless very unwelcome. In this period of economic crisis, some educational programs are facing severe cuts. This demands an even more precise insight into the hardware of education, that is, the physical buildings themselves. Being made of bricks and mortar, they are not subject to down-sizing once constructed.

A specific design (Primary and Secondary School with Child Day Care and Community Centre for the Caribbean island of Bonaire) is the basis of this research proposal and allows research to be carried out on how to optimize the spatial setting for providing basic education [1].

The hypothesis states that optimum physical surroundings, in combination with its program, will compensate for the sense of danger, or better put, enhances the feeling of being comfortable and at home [2].

Research questions

The questions to be considered are:

- To what extent can a virtual design model reflect the actual building constructed?
- How can access control be implemented without hampering a free flow of activities?
- To which degree does the orientation of the building with respect to its surroundings enhance a sense of safety?
- Does overview of activities (visibility) with respect to the building in all cases improve the notion of safeness?
- In terms of safety (and durability!), can functions be distributed and mixed in diverse ways throughout the building?
- To what extent does limiting the interaction between different age groups during class hours and breaks induce a sense of safety due to the limited number of pupils present?
- Does optimizing the situation require complete or partial separation of the children and therefore zoning the building into specific entities [3] and [4]?

Type of research

The research involved is empirical with an inductive approach. Based on specific observations of behaviour, an attempt will be made to deduce general rules, on the one hand. On the other, it is assumed, as a deductive hypothesis, that clever separation of school-going children will enhance their feeling of safety. Also, overview in spatial surroundings generally seems to be appreciated.

The research will consist of questionnaires filled in at specific intervals, while subjects (children of various age groups) explore the virtual environment. A survey will therefore be combined with a case study in order to test

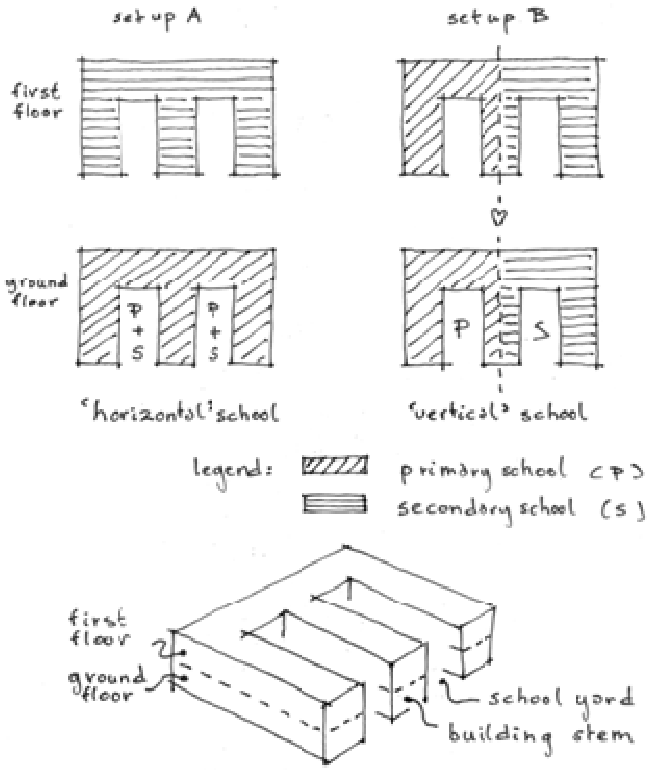


Figure 1. Distribution of classes to be examined.

the hypothesis. For instance, subjects will have to fill in forms on how safe (on a scale of five) they judge the space they virtually wander trough. Or how they appreciate having access to just one half or just one floor of the building. Recording of behaviour during the simulated process of passing through and exploring the virtual architectural space will provide an insight into the psychological perception of spatial information. In this process the subjects physical condition, that is, their eye movement, perspiration and heart rate are measured to register a corresponding level of anxiety. Depending on the number of measurements, the research will primarily be qualitative in nature in the sense that no absolute values can be determined. Only limited statistical calculations will therefore be necessary.

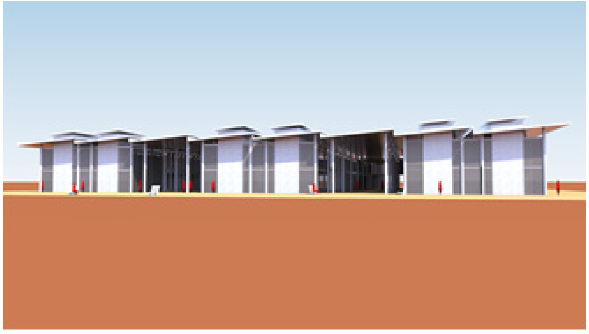
The objective is, of course, to achieve a neutral result by applying an appropriate methodology and precise focus.

Research methodology

Precise measurements can be achieved by working with distinct groups. The subjects will consist of:

1. Children attending primary school.
2. Children attending secondary school.
3. Children and teacher at primary school.
4. Children and teacher at secondary school.

Within a given timeframe, subjects within a group will explore the virtual models of the school and respond to specific questions put to them. Set-ups A and B will be investigated and then compared.



Figures 2 and 3. Draft school design including community centre (2549 m2).

Set-up A, the *horizontal* school: the horizontal first floor separates the school into Primary and Secondary sections. There will be no interaction between the two age groups during class. During the breaks both inner school yards are used by both groups, and all children have the possibility of interacting.

Set-up B, the *vertical* school: a vertical separation along the alley of the inner stem divides the school into the two age groups. In the stem, it is inevitable that a limited amount of interaction will take place during changes of class. During breaks the two age groups have their own square to themselves.

All subjects will work alone first and then together (with their teachers) during the test. It may be necessary to combine the different groups but this will depend on the results of the initial tests. Moreover, relevant questions will be put regarding comfort and orientation – i.e. the mental condition of subjects before and after the tests – with the answers being compared with the actual behaviour of the subjects [5] and [6].

Naturally, parents of young children will have to consent with the latter participation in this project.

Implementation of research

The present architectural Preliminary Design (see below) will be further converted into a semi-realistic environment using advanced visualization and gaming techniques, including sound characteristics of the visited artificial space.

Subjects will explore the interior environment – either individually or in groups – by means of computer interfaces. They will interact with computer screens while sitting on specially wired chairs at gaming consoles. This can be done, for instance, at the advanced gaming facilities at University of Twente (T-Xchange).

This method (registration of mental decisions and physical condition) will make it possible to measure the mental process of safety assessments.

General constraints

A realistic model is at stake, where the yet to be made Final Design will be converted into a real-life visualization;

A representative measurement (large enough number and appropriate groups of subjects) is required to establish sufficient accuracy;

Applicability should also be part of the outcome, i.e. the results must be convertible into a usable protocol for professional architects.

Keywords. architecture, school safety, comfort zones, privacy, children’s behavioural dynamics, orientation, perception, space

References

1. Doban, G., (2006) Designing for Safety. ENGINEERING.com LIBRARY. <http://www.engineering.com/library/articlesPage/tabid/85/articleType/ArticleView/articleID/85/Designing-For-Safety.aspx>. Accessed 5 May 2012.

2. CPTED (2012) (Crime Prevention Through Environmental Design) School Security. <http://www.cptedsecurity.com/school_security.htm >. Accessed 5 May 2012.

3. Larson Vaughan, E. (2011) Elementary School, Whole Building Design Guide, National Institute of Building Sciences, USA. <http://www.wbdg.org/design/elementary.php>. Accessed 5 May 2012.

4. Etemadi, M. (2011) A new outlook at Design of Secondary School. Canadian Journal of Scientific and Industrial Research, **2**(9), 312.

5. Etemadi, M., Kia, K., (2009). *Environmental Psychology and School Space Designing*. Islamic Azad University, Iran, 45.

6. Hunter, K., (2005) *Environmental Psychology in Classroom Design*. University of Cincinnati, USA, 92, 93.

Internal Low Dose Ionising Radiation in Pregnant Mice: Behavioral Effects in the Offspring

D. Lafuente[1], M. Bellés[1,2], S. Gonzalo[1], J.L. Domingo[1], V. Linares[1,2].

[1]*Laboratory of Toxicology and Environmental Health, School of Medicine, "Rovira i Virgili" University, Reus, Spain.*
daisy.lafuente@urv.cat

[2]*Physiology Unit, School of Medicine, "Rovira i Virgili" University, Reus, Spain*

Abstract

Most of the population living in areas close to the Chernobyl accident have been exposed to external and internal low doses of ionising radiation. High doses effects are well known, however low dose radiations had not been yet investigated. The aim of this work was to investigate the cognitive effects of internal low doses ionising radiation. We evaluate the behaviour in mice pups after that pregnant females were exposed to an acute internal low dose of ^{137}Cs (ionising radiation). ^{137}Cs was administrated on GD12 through the drinking water at radiation doses of 0, 500, 1000, 2000, 4000 and 8000 Bq/Kg. Behavioural tests were conducted at the age of 2 months to evaluate the spontaneous behavior, learning, memory capacities and anxiety. Although we observed an influence of the low dose radiation, in general terms, we have not found significant changes between groups in the different tests.

Keywords. ^{137}Cs, internal low dose radiation, behavioral tests, mice.

Introduction

After the Chernobyl disaster, the population living in the contaminated areas was subjected to two types of radiation exposure: external radiation caused by the surface contamination of the environmental and internal contamination caused by food and water consumption. The main contributor to the exposure of people living in these areas is ^{137}Cs. Its chemical properties and especially the high solubility in water allow the diffusion of ^{137}Cs into plants and animals, resulting in the contamination of food chain in the years following the Chernobyl accident [1]. The biological consequences on the health status for internal exposure to ^{137}Cs are not completely understood. Several authors have described behavioural disorders and diseases of the central nervous system in populations living in radio contaminated areas.

It was generally admitted that ^{137}Cs must not have effect on the central nervous system, except perhaps at very high doses [2]. The brain is highly sensitive to ionising radiation during foetal and early post-natal period. That interval corresponds to the time of rapid proliferation and neuroblast migration from proliferative zones to the cerebral cortex. Furthermore, the risk of mental retardation is at least five-fold greater when foetuses are exposed during this period [3]. The evaluation of neurobehavioral function failed to identify significant effect of chronic exposure to ^{137}Cs [4]. Nevertheless, significant modifications in the metabolism of some neurotransmitters have been observed in rats fed with oats contaminated with ^{137}Cs at a very low dose (45 Bq/kg) for 1 month [5]. The aim of this study was to investigate the impact of internal low doses of ^{137}Cs during pregnancy on the postnatal development and behaviour of the offspring.

Methods

Animals and treatment

Pregnant female C57Bl6 mice (30 g) were randomly divided into 6 groups of 10 animals per group. At the GD12, animals of each group received a ^{137}Cs at oral doses of 0, 500, 1000, 2000, 4000 and 8000 Bq/Kg. After weaning, the pups were separated from the mothers and were raised up. The use of animals and the experimental protocol were approved by the Animal Care and Use Committee of the “Rovira i Virgili” University (Tarragona, Spain).

Experimental setup

At the age of 2 month, there were randomly selected 8 mice of each group of dose and the following behavioral tests were carried out: open–field activity, morris water maze test, radial arm maze and elevated plus maze.

Behavioral tests

Open-field activity. General motor activity was measured in an open-field apparatus. Mice were allowed to move freely around the open-field and to explore the environment for 15 min, divided into 3 sessions of 5 min each. The path of the animals was recorded by a video camera (Sony CCD-IRIS model) that was placed above the square and was connected to a VHS videocassette recorder (Panasonic AG-5700 model). The video tracking program Etho-Vision® XT from Noldus Information Technologies (Wageningen, The Netherlands) was used to measure the total distance traveled and the number of rearings as a measure of vertical activity [6].

Morris water maze test. Spatial learning and retention were tested in a water maze according to a test modified from the procedure of Morris (1984). The animals were tested in blocks of five trials during three consecutive days (each trial started from one of four points assigned on different arbitrary quadrants of the circular tank). The trial latency to reach the platform was 60s and the intertrail interval was 60s. Once the mice reach the platform, it remained there for 30s. If the mice did not locate the platform within 60s, the animal was then placed on the platform for 30s.

The movements of the animal in the tank (160 cm diameter and 60 cm height) were monitored with a video tracking system (EthoVision® XT, Noldus Information Technology, Wageningen, The Netherlands), were measured and recorded the total distance traveled and the latency time before reaching the platform.

Probe test. At the fourth day, a single probe trial was conducted. The platform was removed from the pool and each mice was allowed to swim for 60 seconds in the maze. It was recorded the time and the distance that the animal swam in the quadrant that was supposed to have the platform [7].

Radial arm maze. In this maze was measured the working memory. It has a circular platform (diameter 18 cm) and eight arms (6 cm wide and 35 cm long) radiating from it. At the final of each arm, they have a small food pellet (5 mg) behind a low barrier preventing the animal from seeing. The animals were tested on 3 consecutive days, one trial per day. The mice had free access to water but were deprived food 12h before the initial trial. The first 2 days of the radial arm maze was used to accustom the mice to the test environment and to the maze itself. Only data from the final performance day were used for analyses. The start of each trial began with the mouse placed on the central platform always facing the same direction. The trial was terminated after 10 min or as soon as the mouse had eaten all eight food rewards. To perform well at this task, the mice had to store information continuously about which arm(s) had already been visited during a particular trial and which had not (working-memory, storing trial-specific information). The behavioral were monitored with tracking system (EthoVision® XT) and it were measures the time to find all eight pellets and the number of errors. Error is defined here as reentering an arm where the food pellet had already been devoured.

Elevated plus maze. Anxiety-like behavior was measured in this test. This maze had two opposite open arms and two opposite closed arms placed 50 cm above the floor. The animals are transferred to the testing laboratory in their home cages at least 30 min before the test. A mice was placed on the central platform facing either of the closed arms. During 5 min were measured the number of entries into the open and enclosed arms and the time spent there. The animal’s behavior was recorded with a video camera (Sony CCD-IRIS model) and was used the video tracking program EthoVision® XT.

Statistical analysis

Homogeneity of variances was analyzed using the Levene’s test. If variances were homogeneous, ANOVA was used followed by the Tukey method to evaluate all dose groups simultaneously. If the variances were not homogeneous, the Kruskal–Wallis test was used. Differences between groups were analyzed using the Mann–Whitney U-test. The level of statistical significance for all tests was established at $p < 0.05$. All data were analyzed by means of the statistical package SPSS 15 (SPSS Sciences, Chicago, USA).

Results

The results of the present study indicate that, in general terms, exposure of female mice to low doses of cesium during gestation did not cause relevant dose-related adverse effects on the behavior of the offspring.

* Financial support for this study was provided by the FP7-Fission-2011, Grant Agreement Number 295552, CEREBRAD.

References

1. Bertho, JM., Louiba, S., Faure, MC., Tourlonias, E., Stefani J., Siffert, B., Paquet, F., Dublineau, I. (2010). Biodistribution of 137Cs in a mouse model of chronic contamination by ingestion and effects on the hematopoietic system. *Radiat Environ Biophys* **49,** 239-248.

2. Lestaevel, P., Grandcolas, L., Paquet, F., Voisin, P. (2008). Neuro-inflammatory response in rats chronically exposed to 137Cesium*.* *NeuroToxicology* **29**, 343-348.

3. Otake, M., Schull, WJ., Lee, S. (1996) Threshold for radiation-related severe mental retardation in prenatally exposed A-bomb survivors: a re-analysis. *Int J Radiat Biol.* **70**(6), 755-63.

4. Houpert, P., Bizot, JC., Bussy, C., Dhieux, B., Lestaevel, P., Gourmelon, P. et al. (2007) Comparison of the effects of enriched uranium and 137-cesium on the behaviour of rats after chronic exposure. *Int J Radiat Biol*, **83**, 99-104.

5. Bandazhevsky, YI., Lelevich, VV., (1995) *Clinical and experimental aspects of the effect of incorporated radionuclides upon the organism*. Belorussian Engineering Academy - Gomel State Medical Institute, Gomel, Belarus.

6. Bellés, M., Alonso, V., Linares, V., Albina, ML., Sirvent, JJ., Domingo, JL., Sánchez, DJ. (2010) Behavioral effects and oxidative status in brain regions of adult rats exposed to BDE-99. *Toxicology Letters* **194**(1-2), 1-7.

7. Bellés, M., Albina, ML., Linares, V., Gómez, M., Sánchez, DJ., Domingo, JL. (2005) Combined action of uranium and stress in the rat I. Behavioral effects *Toxicology Letters* **158,** 176-185.

In Vivo Characterization of the Role of Trpc1 Channel in Skeletal Muscle Function

Olivier Schakman[1] , Nadège Zanou[1], Georges Shapovalov[2], Chiara Gallo[2], Alexander Dietrich[3], Jean Lebacq[1], Urs Ruegg[2], Lutz Birnbaumer[4], Philippe Gailly[1]

[1]*University of Louvain, Louvain, Belgium. olivier.schakman@uclouvain.be*
[2]*University of Geneva, Geneva, Switzerland.* [3]*University of Marburg, Marburg, Germany.*
[4]*National Institute of Environmental Health Sciences, North Carolina, USA.*

Abstract

Skeletal muscle contraction is reputed not to depend on extracellular Ca^{2+}. Indeed, stricto sensu, excitation-contraction coupling does not necessitate entry of Ca^{2+}. However, we previously observed that, during sustained activity (repeated contractions), entry of Ca^{2+} is needed to maintain force production. The influx of Ca^{2+} through TRPC1 represents a minor part of the entry of Ca^{2+} into muscle fibers at rest, and the activity of the channel is not store dependent. The lack of TRPC1 does not affect intracellular Ca^{2+} concentration ($[Ca^{2+}]$i) transients reached during a single isometric contraction. However, the involvement of TRPC1-related Ca^{2+} entry is clearly emphasized in muscle fatigue. Indeed, muscles from $TRPC1^{-/-}$ mice stimulated repeatedly progressively display lower $[Ca^{2+}]$(i) transients than those observed in $TRPC1^{+/+}$ fibers, and they also present an accentuated progressive loss of force. Interestingly, muscles from $TRPC1^{-/-}$ mice display a smaller fiber cross-sectional area. In the present study, we evaluated the involvement of the canonical transient receptor potential TRPC1 ion channel in muscle function, basal activity and paws coordination. To study the role of TRPC1 in skeletal muscle functions, we investigated in vivo the ability of the $TRPC1^{-/-}$ mice to perform voluntary (running wheel and escape test) or endurance exercises (treadmill and wire test). If $TRPC1^{-/-}$ mice do not exhibit any difficulty to perform voluntary exercise, they show a predisposition to muscle fatigue. Indeed, the performances of the $TRPC1^{-/-}$ mice in the treadmill and wire test were lower than those observed in control mice. We also compared basal activity (home cage activity) and coordination (rotarod, balance-beam test). We conclude that TRPC1 ion channels modulate the entry of Ca^{2+} during repeated contractions and help muscles to maintain their force during sustained repeated contractions.

Introduction

Skeletal muscle contraction is reputed not to depend on extracellular Ca^{2+}. Indeed, stricto sensu, excitation-contraction coupling does not necessitate entry of Ca^{2+}. However, we observed that, during sustained activity (repeated contractions), entry of Ca^{2+} is needed to maintain force production. We previously reported that in myoblasts, TRPC1 ion channel is by far the most widely expressed channel of the TRPC subfamily (at least 100 times more than other isoforms). In soleus, EDL and tibialis anterior muscles, TRPC1 is also one of the most expressed TRPC isoforms. In several investigations, TRPC1 has been reported to be responsible for store-operated Ca^{2+} entry. In a recent study, we indeed confirmed that repression of TRPC1 in C2C12 myoblasts reduced the store-operated Ca^{2+} entry and that involvement of TRPC1-related Ca^{2+} entry is clearly emphasized in muscle fatigue (2).

Aim of the study

In the present study, we evaluated in vivo the possible involvement of TRPC1 in muscle function, basal activity and paws coordination in $TRPC1^{+/+}$ and $TRPC1^{-/-}$ mice.

Materials and methods

Generation of TRPC1−/− mice has been described previously (1). TRPC1−/− and TRPC1+/+ were obtained from heterozygous animals. TRPC1−/− were compared with their TRPC1+/+ control sex-matched littermates. Mice were placed in individual “Physiocages” (Panlab-Bioseb, Vitrolle France) in which they were accustomed for at least 24 h before starting the measurements. They had free access to food and water. Food and drink consumption and rearing were measured during 48h. Open field test. Mice were placed in a square arena (60 x 60 cm) on one side of the arena facing the wall. Mice were video tracked (Ethovision, Noldus) during 20 min. Distance moved and speed were recorded. Rotarod test. Mice were tested for their ability to keep their balance on a rotating rod. The time (latency) taken to fall off the rod rotating under continuous acceleration (e.g. from 4 to 40 rpm) was measured. Wheel running test. Each mouse was individually housed in a standard cage containing a low inertia running wheel. Wheel-running distances were measured during the active part of the day (from 6 pm to 6 am). Escape test. Each mouse was placed in front of a tube and a cuff was wrapped around the tail and connected to a fixed force transducer. In response to gentle pinching of the tail, the mice tried to escape into the tube and a short peak of force was recorded. Results are presented as the mean of the five highest peaks of force recorded, related to body weight (mN.g-1). Treadmill test. The mice were placed on a treadmill with an uphill inclination of 30°, at a speed of 5m/min for 5 min, followed by a progressive increase in speed up to 17 m/min. The test was stopped when the mouse remained on the shocker plate (back of the treadmill) for 20 s without attempting to reengage the treadmill, and the time to exhaustion was determined. Wire test. Mice were suspended by their forelimbs to a 1.5 mm thick, 60 cm long metallic wire at 45 cm above soft ground. The time until the mouse fell down was recorded. Three trials were performed per session with a maximum time per trial was set to 180 s. Results are presented as the mean ± SEM (n=5-10 for each strain). The experiments were conducted and the animals were cared in accordance with the directives of the institutional animal care and use committee of the University of Louvain.

Results and conclusion

TRPC1 −/− mice Present In Vivo Signs of Fatigue. Indeed, TRPC1+/+ and TRPC1−/− mice performed very differently in the two endurance tests (treadmill test and wire test). TRPC1+/+ mice ran easily during the progressive increase of treadmill speed and resisted beyond 45 min of the test. Interestingly, TRPC1−/− mice presented difficulties when speed was progressively increased and ran only <20 min (n = 10) (P < 0.001), suggesting a predisposition to muscle fatigue. A similar result was obtained with the wire test, we observed that TRPC1+/+ mice rarely fell down before the end of each trial. However, TRPC1−/− mice were unable to resist so long [57 ± 13 s (n = 11) in TRPC1−/− vs. 174 ± 4s (n = 10) in TRPC1+/+ mice. In contrast, similar performances were observed in TRPC1+/+ and TRPC1−/− mice in the tests evaluating basal activity (open field test), paw coordination (rotarod test) and voluntary exercise (running wheel test and escape test). In conclusion, the lack of TRPC1 has a repercussion on adult muscles in vivo: TRPC1 KO mice show more susceptibility to fatigue. But the lack of TRPC1 impaired neither basal activity nor paws coordination.

Acknowledgements

This work was supported by the “Association française contre les myopathies” (AFM), the “Association belge contre les maladies neuro-musculaires” (ABMM), by grant ARC 05/10-328 from the General Direction of Scientific Research of the French Community of Belgium, the “Programme d’excellence Marshall” (DIANE convention) from the “Région Walonne”, by grants of the Swiss National Science Foundation (SNSF), the Swiss Foundation for Research on Muscular Diseases and in part by the Intramural Research Program of the NIH (project Z01-ES101684 to LB).

References

1. Dietrich, A., Kalwa, H., Storch, U., Schnitzler, M., Salanova, B., Pinkenburg, O., Dubrovska, G., Essin, K., Gollasch, M., Birnbaumer, L., Gudermann, T. (2007). Pressure-induced and store-operated cation influx in vascular smooth muscle cells is independent of TRPC1. *Pflugers Arch* **455,** 465-477.

2. Zanou, N., Shapovalov, G., Louis, M., Tajeddine, N., Gallo, C., Van Schoor, M., Anguish, I., Cao, M.L., Schakman, O., Dietrich, A., Lebacq, J., Ruegg, U., Roulet, E., Birnbaumer, L., Gailly P. (2010) Role of TRPC1 channel in skeletal muscle function. *Am J Physiol Cell Physiol* **298**: C149-C162.

Approaching Real World Behavior: Enhancing External Validity of Psychological Research

F. Lange, S. Bosse, F. Eggert

Institute of Psychology, Department of Research Methods and Biopsychology, TU Braunschweig, Braunschweig, Germany. florian.lange@tu-bs.de

Introduction

After defining the subject's primary goal as describing and explaining certain aspects of behavior, contemporary introductory courses and textbooks in various psychological subdisciplines reveal an odd selection of behaviors intended to be investigated: filling out self-report scales and pressing keys in computerized experiments. Across the issues of the premier journal within the field of personality and social psychology, for instance, Baumeister, Vohs and Funder [1] reported a steady decline of the proportion of studies involving the measurement of actual behavior since the early 1980s down to barely 20% in 2006. This increasingly narrow focus in methodology comes along with a substantial restriction to the range of phenomena addressed by modern psychology [2]. By means of the exemplary experimental and observational settings introduced in the present paper, we attempt to expand this focus and, thereby, to overcome some of the classical problems associated with predominant approaches to human behavior.

Illegal pedestrian road crossing behavior

Identifying risk factors associated with pedestrians' tendency to cross against the lights is crucial to enhance traffic safety. A substantial part of psychological research within this field applies the theory of planned behavior [3] in order to investigate the influence of attitudes, subjective norm and perceived behavioral control on illegal crossing behavior. However, these studies [4] do not examine the constructs' impact on actual behavior, but on the reported intention to violate traffic rules in a fictitious scenario. Regarding the repeatedly observed inconsistency between (a) attitude and behavior and (b) hypothesized and actual behavior, respectively, this approach appears to be conceptually inappropriate. As an alternative, we developed a stimulus control model based on applied behavior analysis, identified discriminative stimuli conflicting with the inhibitory potential of the red traffic light, and examined the association between the occurrence of these stimuli and illegal pedestrian crossing behavior by means of a systematic observational study. We found that contradictory stimulus configurations strongly increase the odds of crossing against the lights. Investigating the determinants of this particular behavior in the field enables the development of tailored road safety measures which change actual behavior and not only some marginally correlated internal constructs.

Lie detection

Lie detection in an experimental context is usually investigated with the so-called "mock crime paradigm" [5]. It is, however, questionable whether an instructed lie is comparable to a real-life lie. To date, the adequacy of this experimental setting remains to be shown. If we want to make a scientific statement about lying behavior, we have to create experiments which reflect the crucial aspects of natural social interactions. Hence, within an alternative framework, we investigate deception and its detection with the aid of a multiplayer role-playing game. This setting allows us to dispense with instructing the subjects to lie because the contingencies requiring players to perform deceptive behavior are imposed by the game itself. In contrast to most of the predominant lying tasks, the subjects do not even have to know about the true background of the investigation. By suggesting a cover story, subjects' reactivity to the experimental procedure can be reduced and, thus, lying behavior can be observed in a more natural, unbiased context. Compared to the mock crime studies, the context of deception produced by a role-playing game is less artificial and the quality of a lie is of greater importance to the individual. Hence, arousal, emotions, verbal and nonverbal cues associated with lying are more likely to occur, behavioral and psychophysiological results are more likely to appropriately reflect the lying processes comparable to those operating in everyday situations.

Cooperative behavior

Whereas the increasing interest in human cooperation in general and indirect reciprocity in particular has inspired a large body of computer simulations and internally valid experiments under controlled laboratory conditions, conclusive field data is still lacking. In order to evaluate to which extent postulated mechanisms of reciprocity account for cooperation in "the real world", natural field experiments are crucially required [6]. Hence, we examined the appropriateness of a "real-world" setting to investigate the predictions of image-dependent indirect reciprocity [7]. In a German supermarket, one of a number of confederates of the experimenter lines up at the checkout appearing to buy a single item (a 0.5 liter bottle of mineral water or a corresponding amount of beer). The waiting person in front of the confederate (i.e., the experimental subject) can, as a potential donor in an altruistic act, let the confederate go ahead which obtains a benefit for the confederate and a cost for the experimental subject (defined as waiting time). The number of items bought by the experimental subject and his/her behavior, as well as the number of observers of the situation are recorded. First results indicate, consistent with the model of indirect reciprocity, that the decision to cooperate in one-shot interactions with strangers depends on the cost-to-benefit-ratio of the altruistic act and that the range of accepted cost-to-benefit-ratios constitutes a behavioral measure sensitive to variations of the recipient's image.

Conclusion

We believe that the empirical approaches illustrated above can pave the way for more externally valid inferences in each of the particular areas of research. More generally, we propose that behavior-oriented psychology, in order to enable a better understanding of human complexity, should focus again on its original interest: actual behavior.

References

1. Baumeister, R.F., Vohs, K.D., & Funder, D.C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science* **2**, 396-403.

2. Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review* **5**, 2-14.

3. Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes* **50**, 179-211.

4. Holland, C., Hill, R. (2007). The effect of age, gender and driver status on pedestrians' intentions to cross the road in risky situations. *Accident Analysis and Prevention* **39**, 224-237.

5. Ekman, P. (2009). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage* (3rd Edition). New York: Norton & Company.

6. Guala, F. (2012). Reciprocity: weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences* **35**, 1-59.

7. Nowak, M.A. & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573-577.

The 5 Choice Continuous Performance Task: Translation to a Touch-screen Paradigm

Pascal van Dorsselaer, Ewan Hughes Mcinnes, John Talpos, Thomas Steckler

Department of Neuroscience, Janssen Research & Development, Beerse, Belgium.
pvdorsse@its.jnj.com

Introduction

The 5 Choice Continuous Performance Task (5CCPT) is a test of attention and reaction time performed in rodents frequently used to test compounds for drug discovery. Traditionally, this test is executed in standard 5-hole operant boxes where the animal has to nose poke at briefly illuminated holes to earn a reward (“Go” trials). However in a small proportion of the trials all holes are illuminated signaling that the animal must refrain from responding in order to earn a reward (“NoGo” trials) [3]. However the 5CCPT has been criticized for the substantial difference in salience between Go and NoGo trials. Here we describe the training process for 3 different variations of this task (Figure 1). First is the “standard” 5CCPT as described by Young (Figure 1a). Next is a version highly similar to the standard 5CCPT, but translated to the touch-screen environment, the 4 choice continuous performance task (4CCPT; figure 1b). Finally, we trained animals to perform a 4CCPT where the Go/NoGo rule is dependent upon a visual discrimination; image “A” (e.g. plane) indicates that the animal should go, whereas image “B” (e.g. spider) means an animal should refrain from going (4CCPT-VD, figure 1c). This final procedural variation should control for differing saliencies associated with each trial type.

Materials and Methods

All experiments described in this abstract have been approved by the local Ethical Committee at Janssen Pharmaceutica.

C57BL6/J mice (Janvier, France) were group housed in ventilated cages with water available ad libitum and kept under 12/12h light/dark cycle in a temperature and humidity controlled room. During training and drug studies, the mice were food restricted and maintained at approximately 85% of their free-feeding body weight. A dedicated group of animals was used for each variation of the task.

Operant Chambers

All experiments were performed using the 5-CCPT procedure in five-hole operant chambers with an attached corridor leading to a pellet dispenser. Each of the chambers consisted of an array of five round holes (1.3 cm in diameter) that were arranged horizontally on a curved wall 5 mm above the grid floor and opposite the corridor. At the end of the corridor a food well was attached at floor level. Additionally, a house-light was installed at the ceiling of the main chamber. The whole set-up was located in a sound-attenuating chamber and vented by a fan.

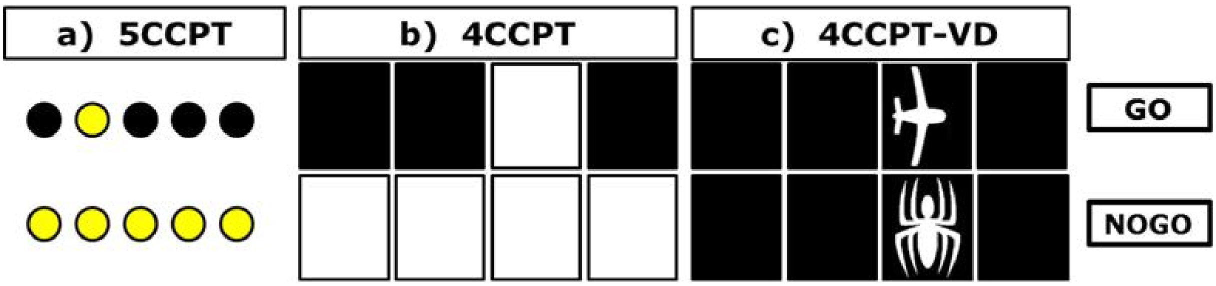


Figure 1. Go and NoGo stimuli for the 3 test variations: 5CCPT, 4CCPT and 4CCPT-VD.

In order to earn a food reward, mice were required to detect a brief light signal (0.5s) presented randomly in one of the five holes and respond at this location within 5s (Go-trials). However in 1/5th of trials, all five holes were illuminated and mice were required to withhold a response for 5s in order to earn a food reward (NoGo-trials).

Touch-screen – 4CCPT

Mice were tested in modified operant chambers of which the wall opposite to the pellet dispenser was replaced by a touch-sensitive computer monitor. In front of this screen was a black metal mask with 4 holes (5 x 7.5cm). Similar to the operant version, white rectangles were shown on the screen in holes of the mask. When a white rectangle was presented (1.5s) randomly in one of the four places on the screen, the mouse had to respond at this location within 5s (Go-trials). However in 1/5th of the trials, four rectangles were shown and mice were required to withhold a response for 5s in order to earn a food reward (NoGo trials).

Touch-screen – 4CCPT-VD

In this variant, one of two pictures is presented upon the screen for every trial (spider or plane). When picture A was presented (1.5s) randomly in one of the four places on the screen the mouse had to respond at this location within 5s (Go-trials). Again, in 1/5th of the trials, picture B was presented and mice were required to withhold a response for 5s (NoGo trials). Four response locations were used in the touch-screen procedure for practical purposes. The ability to discriminate visual stimuli is highly dependent upon the size of the stimuli [1] and these stimuli had previously been validated using a visual discrimination procedure. Smaller images would have dictated the correction of unique stimuli not currently being used in any of our tests. Therefore we choose to instead use fewer locations. To help make the critical comparison between the translated touch-screen version (4CCPT) and the version based upon the visual discrimination, we decided to use only 4 stimuli for this version as well.

Incorrect responses of either type, as well as omissions (Go-trials), resulted in de-activation of the house light and a short delay prior to the start of the next trial. Mice were tested for a total of 120 trials. Accuracies, omissions, and response latencies were measured.

Results

Animals were initially trained to associate a tone and a light in the food well with a reward. Afterwards animals were trained to respond to a stimulus (light on hole or image on a screen) to earn a reward. As soon as the animals reached a high level of performance NoGo trials were introduced. At this point pictures were introduced also in the 4CCPT-VD test. Initially the stimulus duration was set at 20 seconds. However this was decreased in a stepwise fashion every time the mice reached an accuracy of at least 70%. Animals acquired the different task variants at approximately equal rates until the stimulus duration was lowered to 1.5 seconds. We discovered that further decreasing of the stimulus duration only results in a decrease in performance in the touch-screen versions, although we were able to further reduce the stimulus duration to 0.5 seconds in the 5CCPT with minimal effect on accuracy measures.

Conclusion

We proved that it is possible to setup a touch-screen version of the 5-choice continuous performance task. Nevertheless the animals need a longer stimulus duration time, 1.5s versus 0.5s, to reach an acceptable performance. Animals trained in the 5 CCPT test reach an accuracy of 79% on the Go-trials and 86% on the NoGo trials while 4CCPT and 4CCPT-VD have a higher Go accuracy (95% and 96%) but a lower NoGo accuracy (62% and 78%). There’s also a very clear difference in Response Bias (RI) between the different tests. The animals trained in both touch-screen tests have a bias towards responding compared to the operant test where the animals have a bias towards not responding despite their high level of accuracy.

Table 1. Average performance during last week of training. 5CCPT with a stimulus duration of 0.5 seconds, 4CCPT and 4CCPT-VD with a stimulus duration of 1.5 seconds.

	GO accuracy	NOGO accuracy	Omissions	RI
4CCPT (n=12)	95 %	62 %	24 %	0.21
4CCPT-VD (n=19)	96 %	78 %	30 %	0.26
5 CCPT (n=11)	79 %	86 %	14 %	-0.30

Keywords. Behavior, attention, animal model, validation, 5-choice, touch-screen

References

1. Bussey, T.J., Padain, T.L., Skillings, E.A., Winters, B.D., Morton, A.J., Saksida, L.M. (2008). The touchscreen cognitive testing method for rodents: how to get the best out of your rat. *Learning Memory* **15**(7), 516-523.

2. Steckler, T., Sahgal, A. (1995). Psychopharmacological studies in rats responding at touch-sensitive devices. *Psychopharmacology* **118**(2), 226-229.

3. Young, J.W., Light, G.A., Marston, H.M., Sharp, R., Geyer, M.A. (2009). The 5-Choice Continuous Performance Test: Evidence for a Translational Test of Vigilance for Mice. *PLoS One* **4**(1), e4227.

Rapid Behavioral Effects of Sex Hormones in Rats

M. Malinova[1], B. Filova[1], J. Babickova[1], L. Tothova[1], D. Ostatnikova[2], P. Celec[1,2,3], J. Hodosy[1,2,*]

[1] *Institute of Molecular Biomedicine, Comenius University, Medical faculty, Bratislava, Slovakia.*

[2] *Institute of Physiology, Comenius University, Medical faculty, Bratislava, Slovakia.*

[3] *Department of Molecular Biology, Comenius University, Faculty of Natural Sciences, Bratislava, Slovakia.*

[*] *hodosy@gmail.com*

Background

Sex hormones have well-described organizational and activational effects on behavior. The mechanism of action of these effects is supposed to involve activation of intracellular steroid receptors acting as transcription factors in the nucleus. Beyond these genomic effects, rapid non-genomic effects have been described for both, testosterone and estradiol [1, 3]. Behavioral effects mediated by these rapid actions are largely unknown, however, it is expected to be within 30 minutes of testosterone or estradiol application. Proposed mechanisms based on Michels G. and Hoppe U. C. [2] can be seen in Figure 1. To our knowledge no studies regarding non-genomic effects of sex hormones and behavior were published so far. The aim of our experiment was therefore to describe the rapid behavioral effects of testosterone and estradiol in male castrated rats.

Methods

Twenty adult male Wistar rats (Anlab, Prague, Czech Republic) were obtained for the experiment. Upon delivery, the animals were allowed two weeks for acclimatization. Animals were kept in separate cages in a controlled environment (temperature 22ºC, humidity 50%) with 12:12 light-dark cycle, light period starting at 8:00 pm, with *ad libitum* access to water and food pellets. The research on animals was approved by the ethical committee of the Faculty of Medicine, Comenius University.

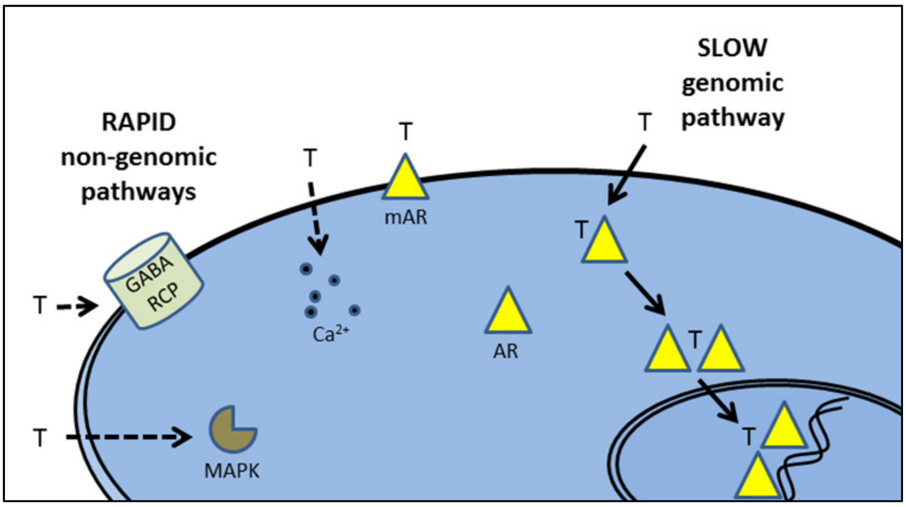


Figure 1. Rapid non-genomic and slow genomic pathway of testosterone (T) action. While in the slow pathway, T binds intracellular receptors and then it enhances the expression of various genes , in rapid pathway, T binds receptors on plasmatic membrane, and the action is mediated through different second messenger cascades (MAPK – Mitogen Activated Protein Kinase; Ca^{2+}- ionized calcium; A R – androgen receptor; GABA RCP – GABA-ergic receptor; mAR – membrane androgen receptor).

Surgery

After two weeks of acclimatization (12-weeks old animals), all rats underwent surgery under general anesthesia (ketamine 100mg/kg + xylazine 10mg/kg in the same syringe, applied intramuscularly). Except of the control sham group, the rest of the animals were castrated prior to experiment. Castration was performed through a small incision in the scrotum, where both testes and epididymis were ligated with absorbable suture and excised. The skin was sutured in two layers with absorbable silk size 4-0. All animals were allowed two weeks for recovery, after the surgery.

Hormonal supplementation

The castrated rats were injected with testosterone 5mg/kg, estradiol 0.5mg/kg or olive oil. Sham-operated group was injected with olive oil.

Behavioral testing

Five minutes after an injection, animals were tested in the open field test (5 min), simple novelty test (5 min), light-dark box (5 min) and forced swim test (3 min). The whole battery of tests was conducted within 30 minutes after injection, since 30 minutes is believed to be the non-genomic effect. The order of tests was chosen from the least to the most stressful. EthoVision XT version 8.5 obtained from Noldus was used as tracking software. Open field test consisted of a square arena of 100cm x 100cm. Animals were placed in the center of the arena, and subsequently were left to explore the maze for 5 minutes. Time in central zone, which was lit, speed, track and distance were recorded. Simple novelty test was performed in the same arena as the open field test; however, new object was inserted into the north-west quadrant of the open field arena. Light/dark box consisted of a box 80cm x 40cm that was divided in the half by a wall with entrance. Half of the box was covered and thus floor inside was in dark. Animals were inserted in the dark box and their behavior was tracked for five minutes. Time in light zone was recorded and evaluated. Forced swim test arena was obtained from Noldus. It consists of a cylindric water tank of 45cm height and 30cm in diameter. The arena was backlit by infrared light. Animals were allowed to swim in the water tank for 3 minutes, during which 3 types of behavior were recorded: A. Immobility - when the animal was not moving, B. Mobility – when the animal was partially moving, and C. Complete mobility – when the animal was swimming around, or was trying to get out of cylinder. Immobility time as a marker of depression-like behavior was calculated and further analyzed.

Statistical analysis was performed by XLStatistics 12.02.10 and GraphpadPrism v. 5.0. ANOVA with Bonferroni post hoc test were used. $P<0.05$ was considered to be statistically significant. Data presented as mean + standard deviation.

Results

Testosterone ($p<0.05$) and also estradiol ($p<0.05$) decreased the time spent in the light part of the light-dark box (see Figure 2). In the open field estradiol increased time spent in the central square ($p<0.05$) compared with castrated and testosterone group but not with control group (Figure 3). No differences in depressive behavior in the forced swim test measured as immobility time was observed between groups (Figure 4). No differences between the groups were found in the simple novelty test (data not shown).

Conclusion

Rapid behavioral effects of testosterone seem to include an anxiogenic effect. The rapid action of estradiol on anxiety differed according to the test used. As far as we know, we are first to describe such rapid testing by battery of tests, in order to evaluate non-genomic effects of sex hormones on behavior. Nevertheless, further studies using larger groups of animals to conquer inter-individual variability, which is high, should reproduce the results and analyze these effects in females as well as further validate the battery of tests.

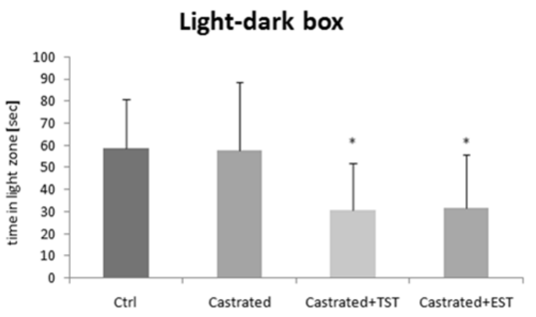


Figure 2. Light/Dark box results. Testosterone as well as estradiol decreased the time spent in the light part of the light-dark box. * denotes $p<0.05$ vs. control, data presented as mean + SD.TST – testosterone, EST – estradiol, Ctrl – control group.

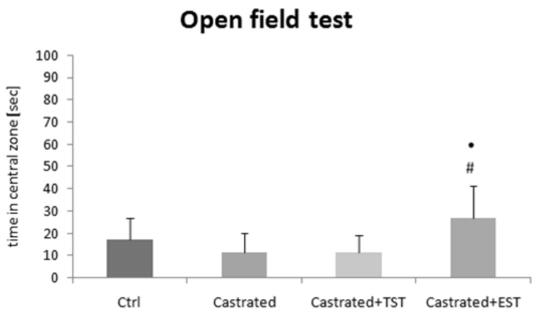


Figure 3. Open field results. Estradiol proved to have anxiolytic effect since, time in central lit zone was in estradiol group increased. # denotes $p<0.05$ in comparison to castrated group, • denotes $p<0.05$ in comparison to castrated+TST group. TST – testosterone, EST – estradiol, CTRL – control group.

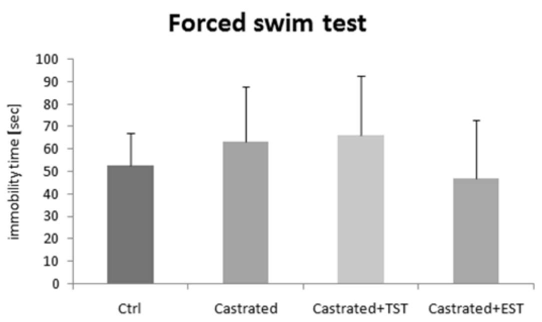


Figure 4. Results of forced swim test. No significant changes were found between groups. TST – testosterone, EST – estradiol, Ctrl – control group.

References

1. Cornil, C.A., Charlier, T.D. (2010). Rapid behavioural effects of oestrogens and fast regulation of their local synthesis by brain aromatase. *Journal of Neuroendocrinology* **22**, 664-673.

2. Michels, G., Hoppe, U.C. (2008). Rapid actions of androgens. *Frontiers in Neuroendocrinology* **29**, 182-198.

3. Moenter, S.M., Chu, Z. (2012). Rapid non-genomic effects of oestradiol on gonadotrophin-releasing hormone neurones. *Journal of Neuroendocrinology* **24**, 117-121.

Different Spatial Learning Performance of 5-Htt Knockout Mice on Land or Water

Lars Lewejohann[1,2], Sandra Grauthoff[1], Rebecca S. Heiming[1,2], Friederike Jansen[1,2], Sylvia Kaiser[1,2], Angelika G. Schmitt[3], Norbert Sachser[1,2]

[1]*Department of Behavioral Biology, University of Münster, Münster, Germany. Ljohann@phenotyping.de*

[2]*Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, Münster, Germany.*

[3]*Molecular and Clinical Psychobiology, University of Würzburg, Würzburg, Germany.*

Introduction

The serotonin transporter (5-HTT) as a key regulator of central serotonergic activity has been linked to neuropsychiatric disorders like anxiety and depression. The evidence of a connection between such mood disorders and genetic variations of the Serotonin transporter led to the generation of 5-HTT knockout mice with a targeted inactivation of the 5-HTT function. These mice, contingent upon their genotypes, differentially express functional 5-HTT. Homozygous knockouts completely lack the 5-HTT, heterozygous mice show a reduced 5-HTT density of about 50%, while wild-type controls express normal levels of 5-HTT. Behavioral phenotyping revealed that 5-HTT knockout mice display a number of phenotypic changes especially regarding anxiety related behavior. On the one hand it is known that 5-HTT as well as 5-HT receptor subtypes are abundant in the cortical areas involved in cognitive functions, implying also an important role in learning and memory. On the other hand, cognitive processes are well known to be strongly influenced by emotionality, possibly modulated by stress-related adrenal steroid hormones. In the present study we evaluated whether spatial memory is affected by the genotype per se or if there is an interference of differences in the aversiveness of the testing conditions.

We therefore conducted different learning and memory tasks in order to measure how the performance is affected by the 5-HTT genotype and differences in the aversiveness of testing conditions.

Methods

Male 5-HTT knockout mice (n=28), heterozygous 5-HTT mice (n=28), and wild-type controls (n=28) have been subjected to a 5 day series of repeated trials in either a water maze or a Barnes maze. An additional group of male mice of all three genotypes (KO n=32, HET n=36, WT n=30) were used to measure how the different testing procedures are related to plasma corticosterone concentrations, the main stress hormone of the murine hypothalamic-pituitary-adreno-cortical axis. These mice were sacrificed 15 minutes after a single session of the respective tests or, as a control, without any test experience. The water maze apparatus consisted of a circular pool (1 m in diameter) made of blue plastic. The pool was filled with water (21 °C) to a height of 33 cm. An escape platform made of transparent plastic was submerged 0.5 cm under the water surface. Three-dimensional visual cues were placed around the water maze. We measure the time until the mice found the submerged platform by means of an automated tracking system (http://www.phenotyping.de/digital.html). The test was conducted on four consecutive days with three trials per day using the same platform position. On the fifth day a different position had to be found. The Barnes maze apparatus consisted of a round white platform (1 m in diameter) that was elevated 127 cm above the floor. The platform was surrounded by three-dimensional visual cues. From the platform the mice could escape to their home cage (placed directly beneath the table, not visible for the mice) via one of 12 holes (3 cm diameter) at the border of the platform. The escape hole was connected to the home cage by a wire-mesh tunnel. The other holes led to little blind-ending mesh tunnels. We measured the time to find the correct hole that leads back to the home cage of the mouse by means of an automated tracking system (see above). The test was conducted on four consecutive days with two trials per day using the same escape hole. On the fifth day a different escape hole had to be found.

Results and Discussion

An ANOVA calculated for the areas under the learning curves as well as a non-linear mixed effect model did not reveal any significant difference between the genotypes in the Barnes maze ($p>0.1$). A search strategy analysis revealed also no differences regarding the strategy used to solve the maze between the genotypes. However, the same statistical analyses performed for the water maze identified significant differences between the genotypes ($p<0.001$) with regard to learning this test. Post hoc analysis revealed significant differences between 5-HTT knockout mice and both other genotypes (both: $p<0.0001$). Moreover, knockout mice significantly differed from both other genotypes regarding the strategy used to solve the water maze by relying on random search strategies rather than on direct navigation. Both learning tests led to significantly increased plasma corticosterone concentrations compared with basal values in all three genotypes. Corticosterone concentrations measured after a single trial in the water maze did not differ from concentrations measured after a single trial in the Barnes maze in heterozygotes and wild-types but corticosterone titers of 5-HTT knockout mice were noticeably higher in the water maze. We suggest that this exaggerated stress reaction contribute to the performance differences between the genotypes that were found in water maze learning.

Acknowledgements

This study was supplied by a grant from the German Research Foundation (DFG) (SFB/TRR58, Project A1).

Measuring Thermal Profile of Reptiles in Laboratory and Field

D.A. Paranjpe[1], R.D. Cooper[1], A. Patten[2], B. Sinervo[1]

[1]*Ecology and Evolutionary Biology, University of California, Santa Cruz, California, USA. dhanu@ucsc.edu*

[2]*Dept. of Biological Sciences, Humboldt State University, Arcata, California, USA.*

Introduction

Studying and measuring thermo-regulatory behavior of reptiles is central to understanding many aspects of their life history including but not limited to their growth rates [1, 2], physiological performance [3], reproductive patterns [4, 5, 6]. Monitoring thermal profiles and thermo-regulatory behavior of reptiles is becoming crucial for understanding impact of climate change on various species of reptiles and also for predicting whether they will adapt to local warming trends or not. Recently Sinervo et al. [7] reported reduction of lizard biodiversity due to local warming trends and altered thermal niches. The study also proposed a mathematical model to predict local extinctions of lizard populations. Data on thermal profiles of lizards is a crucial parameter for validating such models [7]. We have developed a simple, yet comprehensive approach to measure thermal profile of lizards in laboratory as well as in field. This protocol allows us to continuously monitor thermal profiles of lizards for extended periods of time for getting picture of their thermal microhabitats and thermo-regulatory behavior in unprecedented details.

Laboratory measurements

We monitor body temperature (Tb) of lizards in the lab continuously for two hours using T type thermocouple probes attached to an automated data logger (Eltek Squirrel model 1035). The ultra-thin probe is inserted in the cloaca or taped to lizard's belly during the run. The length of the wire attached is long enough to allow the lizard free movement in the laboratory thermal gradient. This method of continuous monitoring gives us a time-series of Tb with fine resolution (1 min.) for lizards instead of snapshots as done in earlier studies. The automated recording reduces amount of handling time which can influence Tb measurements. A detailed picture of movements of lizards in the thermal gradient and how they actively regulate their Tb emerges from the continuous record. From these data we can calculate temperature preference (Tp) and time spent at various temperatures. This method is safe for monitoring gravid females as well. In addition to monitoring adult lizards, this method can be adopted for estimating Tp of small hatchlings by attaching the probe to their ventral surface. The cloacal and belly temperature readings are comparable (data not shown). We, thus, successfully monitored Tb for both adult and hatchling lizards and calculated heritability of Tp (Paranjpe et al., in preparation). The Tb monitoring set up of data logger, probes and particle board tracks with lamps for creating thermal gradient is simple and portable enough to be used in field as well (see Figure 1). The lamps can potentially be replaced with

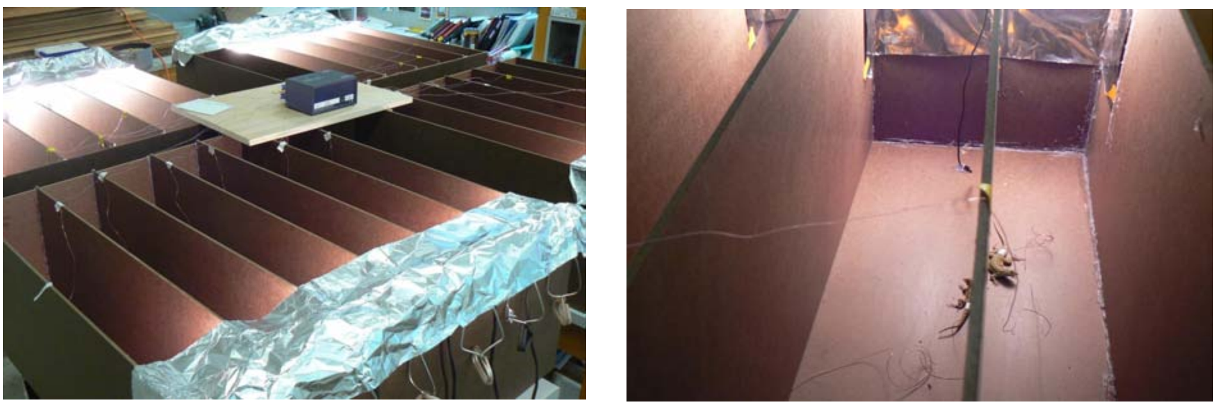


Figure 1. Experimental set up for recording Tb in laboratory using Eltek squirrel automated data logger. Up to 24 lizards can be recorded simultaneously in parallel tracks which are illuminated using lamps to create thermal gradients.

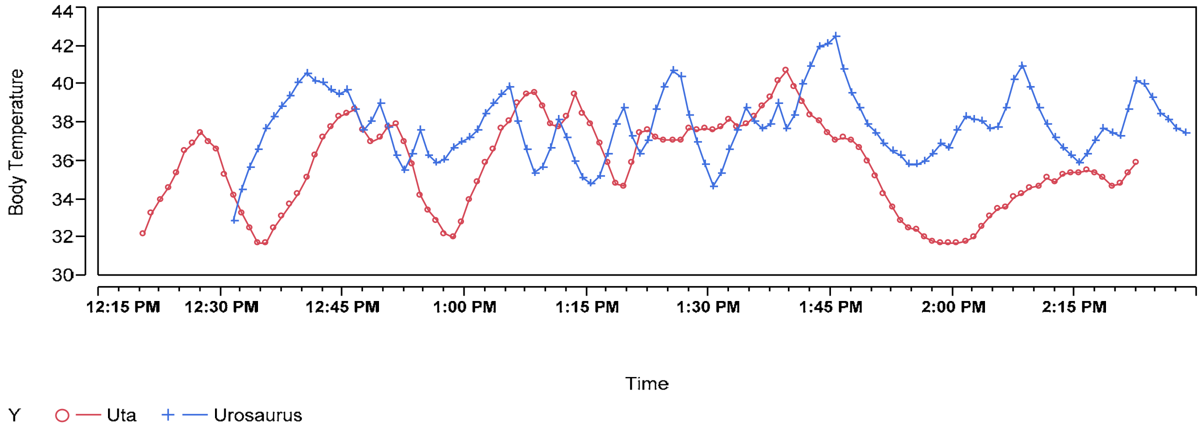


Figure 2. Body Temperature data of two species of lizards recorded using Eltek squirrel automated data logger.

heating tapes in the laboratory if we need to separate the effects of light and heat on Tb. We can record Tb of 24 lizards simultaneously using this set up. We have monitored Tb of various species such as *Uta stansburiana*, *Urosaurus*, *Elgaria* (see Figure 2) using varying lengths and thickness of the T type probe wires without causing any harm to the lizards.

Field measurements

To supplement this information with actual thermal niches available in the field we deploy external temperature data loggers in various lizard habitats. The HOBO external temperature data loggers (hobos henceforth) have two probes for measuring surrounding temperature. Each probe is embedded inside a lizard model made from PVC pipe of appropriate length and diameter, painted to mimic the lizard species under investigation. The hobos along with the lizard models are then deployed to actual habitats such that one of the probes is in the sunniest part while the other one is in the shaded part of the microhabitat. The data are gathered every minute or at longer intervals depending on how long we want to record the external temperature. With periodic download of data we were able to record detailed temperature profile of lizard's microhabitat year around. Using these detailed temperature profiles we could estimate not only the range of temperatures available for thermo-regulation but also the number of hours the lizards could be actually active (h restriction) in their micro-habitat. Such estimates are crucial for predicting potential extinction of the populations. In addition to hobos, we also use custom built "web" with type E thermocouple probes to monitor temperature of micro-habitats at even finer scale. The web consists of 6 radial arms each with 9 probes extending out attached to a data logger and multiplex reading board. Each of these probes has a lizard model attached that is placed in various attributes of a selected lizard territory. Models are placed in bushes and burrows, on rocks, trees and logs etc. so that recorded operative temperatures give a thorough representation of all available sites that can be used for behavioral thermoregulation. The web can also be adopted to monitor temperature of snakes and even amphibian models (Bufo sp.) in field.

Summary

This is a simple yet effective method of monitoring temperature profiles of lizards. The comprehensive approach of laboratory and field Tb monitoring gives a detailed picture of thermo-regulatory behavior in various lizard species. This approach can go a long way in gathering extensive data with relative ease for predicting effects of climate change on various species. Note: All protocols involving live animals were approved by Institutional Animal Care and Use Committee (IACUC).

References

1. Sinervo, B., Adolph, S.C. (1994). Growth plasticity and thermal opportunity in *Sceloporus* lizards. *Ecology* **75**(3), 776-790.

2. Sinervo, B., Dunlap, K.D. (1995). Thyroxine affects behavioral thermoregulation but not growth rate among populations of the western fence lizard (Sceloporus *occidentalis*). *Journal of Comparative Physiological B*. **164**, 509-517.

3. Angilletta, M.J., Hill, T., Robson, M.A. (2002). Is physiological performance optimized by thermoregulatory behavior?: a case study of the eastern fence lizard, *Sceloporus undulatus. Journal of Thermal Biology* **27**, 199-204.

4. Licht, P. (1971). Regulation of annual testis cycle by photperiod and temperature in lizard *Anolis carolinensis*. *Ecology* **52**(2), 240-&.

5. Licht, P. (1973a). Environmental influences on testis cycles of lizards *Dipsosaurus dorsalis* and *Xanthusia vigilis*. *Comparative Biochemistry and Physiology* **45**(1A), 7-20.

6. Licht, P. (1973). Influence of temperature and photo-period of annual ovarian cycle in lizard *Anolis carolinensis*. *Copeia* **3**, 465-472.

7. Sinervo, B. et. al. (2010). Erosion of lizard diversity by climate change and altered thermal niches. *Science* **328**(5980), 894-899.

Hand Movement Behavior with or without Speech Differs in its Kinetic Structure of Nonverbal Communication

I. Helmich, H. Lausberg

Institute of health promotion and clinical movement science, Department for Neurology, Psychosomatic Medicine and Psychiatry, German Sport University Cologne, Cologne, Germany. i.helmich@dshs-koeln.de

Introduction

In the present study we investigate the differences in hand movement behavior in the narration of everyday like scenes between a speech and a silent condition. Previous studies have shown differences between speech and silent conditions when investigating iconic hand movements [1,2]. In this study, we extend the analysis to all hand movements having occurred during the description of everyday like scenes. The analysis is performed according to the Neuropsychological Hand Movement Coding (NEUROGES)-ELAN-System, a tool for empirical hand movement research that combines a kinetic with a functional analysis of hand movement behavior [3].

Susan Goldin-Meadow proposed a model which assumes that hand movements are constrained by speech gaining the “full burden of communication” and grammatical properties without speech [2]. Studies so far, contrasting gestural output during speech and silence [1,2] could show that more hand movements are performed during silent conditions. Both of these studies have focussed on iconic hand movements not including the full manual repertoire of hand movement behavior. In this study, we look at the entire hand movement behavior of silent and verbal conditions to uncover differences of the complete manual range of nonverbal communication.

Hand movements convey meaning by different kinetic movement properties. For example, the kinetic structure of communicative hand movements is characterized by a preparation phase, a stroke phase and a retraction phase, whereas hand movements reflecting emotional states are of continuous/irregular body-focused manner without a preparation or a retraction phase [3]. Norbert Freedman has postulated that body-focused as compared to “object-focused” communicative hand movements indicates a withdrawal from communication for the purpose of self-stabilization [4]. Furthermore, hand movements that accompany speech are different in their “basic organization” [2] as the sequential nature of the information flow in speech makes the information flow in co-speech hand movements also sequential [5]. To gain insights about the kind of information transmitted by a certain kinetic property, we investigate in this study the hypothesis that the total amount of hand movements does not differ in situations with or without speech, but the kinetic hand movement structure changes in regard to mostly conceptual properties during silent conditions to a less conceptual but more self-stabilizational purposes during the speech condition. Furthermore, spontaneous hand preferences reflect the activation of the contralateral hemisphere [6]. As Kimura proposed that hand preference is determined by the lateralization of language [7], Lausberg and Kita argued that semantic aspects of the message influence the choice of hand in co-speech hand movements [1]. Because hand preference of co-speech hand movements is discussed controversially, with this study we seek to disentangle the relationship of the choice of hand and the respective neuronal correlates of hand movements by a tachistoscopic study design with a silent and a verbal condition.

Methods

Hand movement behavior of 11 healthy right-handed participants (5 female, 6 male; mean age: 41 years; mean WAIS score: 100.6; all participants living in Northern America) was videotaped for gestural analysis with the Neuropsychological Hand Movement Coding (NEUROGES)-ELAN-System [3]. The study was approved by the Ethics committee of the German Research Association, the Ethics committee of the Free University of Berlin Medical School, and the Ethics committee of the Montreal Neurological Institute. All subjects signed the Neuropsychology consent forms currently used at the Montreal Neurological Institute and Hospital. Participants sat in front of a computer and drawings of everyday life action scenes, e.g. a skipping girl, were presented unilaterally randomly in the left or right visual hemifield for 150ms. Subjects were asked on the first day of the experiment to demonstrate with hand movements but without speaking what they had seen on the drawing. The second day of the experiment participants were asked to describe verbally what they have seen on the drawing. Each subject's performance was taped with a video camera. Two independent blind raters were trained to analyze hand movement behavior according to the Neuropsychological Hand Movement Coding (NEUROGES)-ELAN-System [3]. Kinetic hand movement structures representing different cognitive processes are coded as communicative (phasic, repetitive), emotional and self-regulating (irregular), changes in postures (shifts) and aborted movements for the right and left hand without sound. One rater coded 100% of the data for further analysis while the second rater coded 25% of the data to establish interrater agreement according to Holle & Rein (modified Cohen's kappa coefficients as measures of inter-rater agreements) [8]. Repeated measures analyses of variance (ANOVA's) were used to report significant results of the frequency per stimulus of hand movements. Multiple post hoc pairwise comparisons were corrected with Bonferroni corrections.

Results

In all, 1672 movement units were analyzed. Modified Cohen's kappa coefficients for structure values were following: irregular, 0.66; phasic 0.88; repetitive 0.90; shift 0.59; aborted 0.39. Presenting stimuli in the right visual hemifield led to greater response in hand movement behavior at all ($F(1, 9)=8.681$. $p<0.05$). Concerning the structure of hand movements the two conditions (silent (GO), verbal (VG)) showed different distributions (*structure*experiment*. $F(4, 6) = 54.480$. $p< 0.001$) but no difference in the frequency of hand movements in general. During the GO condition significant more phasic (*mean difference* [*MD*] = 0.148, $p < 0.05$) and repetitive (*MD* = 0.140, $p < 0.001$) hand movements occurred as during the VG condition significant more irregular hand movements (*MD* = 0.131, $p < 0.05$) were executed. Further, we observed a significant interaction of hand movement structure and hand preference (*structure*hand*. $F(4, 6) = 13.017$. $p < 0.01$). Whereas the left hand was preferred for irregular, shift and aborted hand movements, phasic and repetitive hand movement structures were mainly executed with the right hand.

Discussion

The present study confirmed previous findings [1] that the choice of hand depends on the semantic aspect of the message and not the hemispheric specialization of language [7]. In this data it is shown by its tachistoscopical design that communicative conceptual hand movement structures (phasic, repetitive) are executed with a right-hand-preference whereas continuous/irregular, shift and aborted hand movements, the more emotional structures of hand movements, show a left-hand-preference. This counts for the silent as for the verbal condition diminishing the influence of speech on the choice of hand. Irregular hand movements are increased in stressful situations, negative emotional experiences or the more personal the topic gets [4]. The preferred use of the left hand for this movement structure demonstrates the influence of the "emotional" right hemisphere [9]. In contrast to previous studies [2] we found that the amount of hand movements during silent and verbal conditions, considering all occurring hand movements, does not change. We observed that it is the distribution of hand movements regarding their kinetic structure that is differing between verbal and silent conditions. During the silent condition the structure of hand movements is highly conceptual observable in mainly phasic and repetitive hand movements whereas during the verbal condition continuous/irregular hand movements are executed the most. While information is transmitted without speech by the classical communicative hand movement structures, speech takes over during verbal conditions changing the kinetic organization of the hand movements. We conclude that during the verbal description of everyday like scenes the sequential information flow of speech induces a consequent sequential gestural output [1, 5] shown in less conceptual hand movements as observed in the silent condition leading to more implicit hand movement behavior during speech.

References

1. Lausberg, H., Kita, S. (2003). The content of the message influences the hand choice in co-speech hand movements and in gesturing without speaking. *Brain and Language* **86**, 57-69.

2. Goldin-Meadow, S., McNeill, D. (1996). Silence is liberating: Removing the Handcuffs on Grammatical Expression in the Manual Modality. *Psychological Review* **103**, 34-55.

3. Lausberg, H. & Sloetjes, H. (2009), Coding gestural behavior with the NEUROGES - ELAN system. *Behavioral Research Methods* **41**, 841-849.

4. Freedman, N. & Bucci, W. (1981). On kinetic filtering in associative monologue. *Semiotica* **34** (3/4), 225-249.

5. Özyürek, A. & Kita, S. (1999). Expressing manner and path in English and Turkish: Differences in speech, hand movement, and conceptualization. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the twenty first annual conference of the cognitive science society* (Mahwah, NJ, London: Lawrence Erlbaum), 507–512.

6. Verfaellie, M., Bowers, D. & Heilman, K.M. (1988). Hemispheric asymmetries in mediating intention, but not selective attention. *Neuropsychologia* **26**, 521-531.

7. Kimura, D. (1973a). Manual activity during speaking—I. Right-handers. *Neuropsychologia* **11**, 45-50.

8. Holle, H., Rein, R. (2012). Assessing interrater agreement of movement annotations. IN: NEUROGES, Lausberg, H. (Ed.) [forthcoming].

9. Palmieri A., Naccarato M., Abrahams S., Bonato M., D'Ascenzo C., Balestreri S., Cima V., Querin G., Dal Borgo R., Barachino L., Volpato C., Semenza C., Pegoraro E., Angelini C., Sorarù G. (2010). Right hemisphere dysfunction and emotional processing in ALS: an fMRI study. *Journal of Neurology* **257**(12), 1970-1978.

Design and Validation of a Wireless Temperature Measurement System for Laboratory and Farm Animals

Marloes Hentzen[2], Dag Hovden[1], Mans Jansen[1], Gerard van Essen[2]

[1]*TeleMetronics Biomedcial bv, Wageningen, The Netherlands. dag.hovden@telemetronics.com*

[2]*MSD Animal Health, Boxmeer, The Netherlands. marloeshentzen@hotmail.com*

Background

In animal research, reliable monitoring of physiological body parameters is essential for obtaining good scientific results. One of the more fundamental of those parameters is core body temperature. The traditional way of measuring temperature is using a standard thermometer, inserted rectally. The value measured in this way gives a good indication of the core body temperature. There are, however a number of drawbacks: the measurement procedure is highly interfering and may even influence the temperature; the procedure is time consuming, labour intensive and not without risk; the value measured may be influenced by the skill and the 'touch' of the person performing the measurement. It's also easy to see that rectal measurements are not suitable if:

- The group of animals is large.
- The duration of the test is long (more than a couple of weeks).
- More frequent data than, say, once per hour is required.

A possible solution to all these points is using a measurement device permanently inserted into the animal and arrange it so that this device communicates wirelessly with some receiving unit in the vicinity. Later on we will describe just such a system but first we will take a closer look at traditional temperature measurements involving piglets.

A specific case: manually measuring temperature in young pigs

An extensive study was carried out to test the following hypothesis:

Stress is an important interfering factor when measuring temperatures. When animals are stressed by fixating and animal handling, this will quickly result in a rise in body temperature (stress induced temperature rise).

Assuming that this hypothesis is true, it's reasonable to assume that socializing and training of piglets would reduce the stress of handling, resulting in a decrease in temperature rise during rectal temperature measurements.

In total 100 piglets were used, divided in two treatment groups:

1. Experimental group: Socialization on day (age) 4- 25, training on day 26 – 40.
2. Control group: No socialization and no training

Socialization meant that the piglets were picked up and held by different persons, allowing them to get used to the procedure of taking a temperature reading if not the actual measurement itself. In the training period, days 26-40, actual measurements were made, 3 times a day: 08:30, 12:00, 15:30.

In the Figure 1 below, we see the difference in temperature between the control group and the experimental group during a period of 3 days, followed by an injection of XXX and the resulting pronounced temperature change.

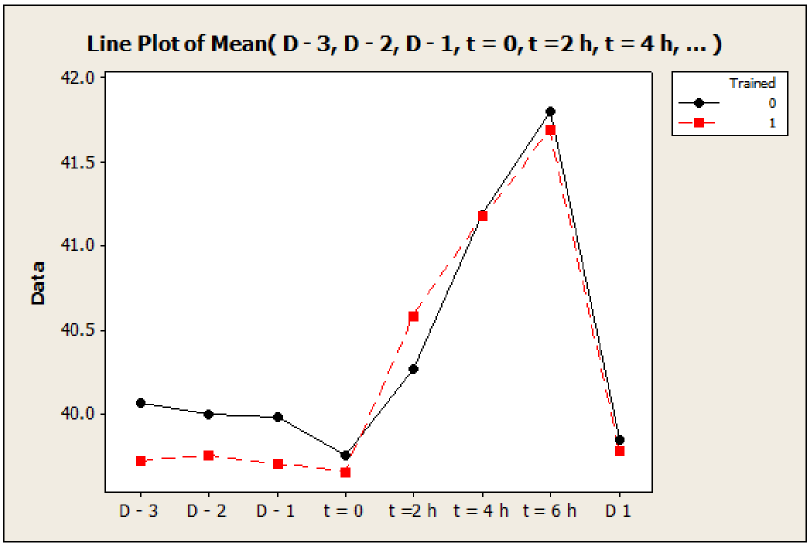


Figure 1. Mean temperature Experimental (1) vs Control (0) piglets.

Training particularly had an influence on the baseline temperature (mean temperature of 3 days prior to vaccination) and not on the temperature after vaccination.

Trained piglets: 39.7 ± 0.2 degrees Celsius
Control piglets: 40.0 ± 0.2 degrees Celsius $P<0.001$

At t=0 and after vaccination temperatures were alike. This means that also for the control group the temperatures were no longer elevated due to stress. It appears there is a training/habituation effect which occurs after three days.

The conclusion of this is that stress caused by handling and the rectal measurement itself only affects the body temperature for a short period of time, i.e. about 3 days. Furthermore, even during this habituation period, the temperature difference is small. Additionally, also an important finding, there was no correlation found between the rectal temperature and the order in which piglets were handled.

We will now move on to describe a similar study, albeit much smaller in terms of number of piglets involved, where telemetry was used side by side with rectal measurements.

Telemetry vs. rectal measurements: a validation study

The telemetry measurement system consisted of 4 implantable sensors, 1 receiver and software. By way of preparation, each implant had been programmed with a unique identification code and set to transmit each 60 seconds. After programming, the implants were encapsulated, resulting in 4 small cylinders, 35 by 15 mm.

Before insertion into the animals, each implant was calibrated in a warm water bath at 36, 39 and 42 degrees Celsius.

The test subjects were 4 young pigs, 5-8 weeks old. The implants were inserted subcutaneously (about 2cm deep) behind the ear in an operation taking roughly 45 minutes. A few days after the operation the scars had healed well.

Each piglet had its own pen, roughly 1.5 by 1.5 metres, but with the possibility of nose-to-nose contact with its neighbours. In two of the pens, receiving antennas were mounted and cabling run through the wall to an adjacent room where the receiver and PC were located. The antennas as well as the receiver have been designed and produced by TeleMetronics.

Figure 2. Encapsulated implant (SI ruler)..

Figure 3. Inserting the implant.

Figure 4. Back in the pen, with antenna.

The PC software, also developed by Telemetronics, was configured for 4 implants, using two antennas and one receiver. In fact, the software can handle any number of implants (limited by the speed of the PC) distributed across as many receivers as the PC can sustain. The total measurement ran for a period of more than two weeks, collecting some 20000 temperature readings per piglet. During the same period, about 100 manual readings per piglet were recorded. External ‘noise’ disturbed the measurements for parts of the period and for some of the implants. Why not all of the implants were affected in the same way is not clear but may be related to the different transmission frequencies used and the position of the antenna relative to the noise source (probable because implants sharing an antenna displayed roughly the same amount of disturbances).

After one week of measurements, one of the sensors stopped working. Examination afterwards showed that the wire to the battery had come loose. The following results therefore do not include this sensor since it wasn’t present in the more important phase of the measurement.

Table 1. Correlation between rectal and telemetry for total measurement period including one week of disturbances (data filtered to remove obvious ‘bad’ readings).

Implant	**Slope**	**Correlation**
1	0.84	0.72
2	0.89	0.80
3	0.96	0.81

Table 2. Correlation for the last 4 days, free of disturbances.

Implant	**Slope**	**Correlation**
1	0.95	0.85
2	0.89	0.80
3	1.04	0.92

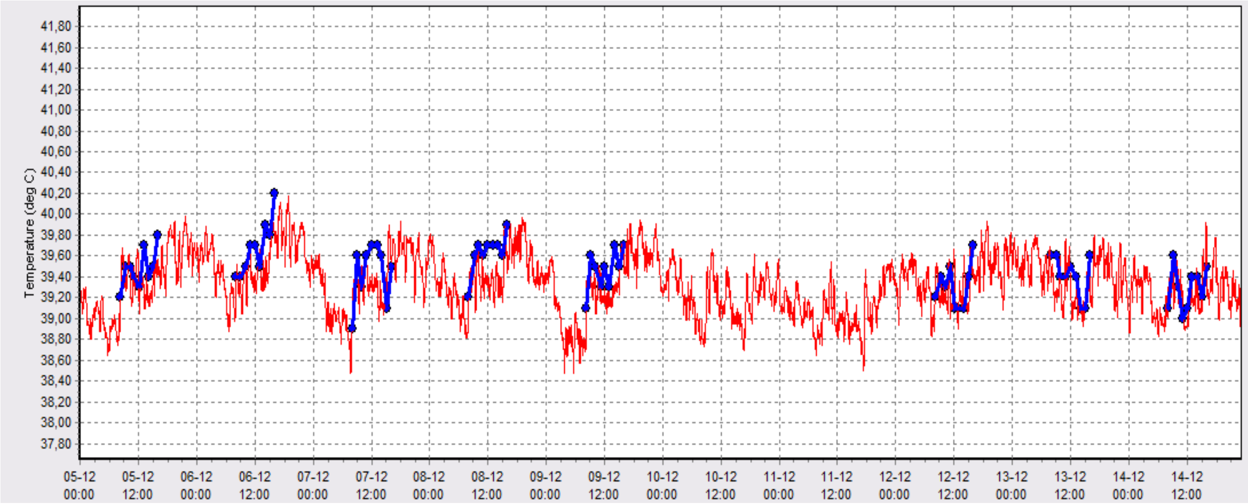


Figure 5. Sensor 3 (red) versus rectal measurements (blue).

As can be seen the slope is very close to 1, indicating that not only is rectal and telemetry correlating but producing the same absolute values, something further confirmed by figure 5

Conclusions

From the above we conclude that:

- The measurement principle used in the implant works well with a close to 1-to-1 relation and high correlation with rectal measurements
- Temperature measurements are as accurate as rectal probes
- Continuous measurement has several advantages: it is time-saving, shows effects that cannot be obtained with manual measurements, and goes on day and night 7 days a week

In the light of the first two points, i.e. the temperature measured by an implant in the neck agrees with rectal measurements, the last point becomes especially interesting. Using an implant and telemetry makes large scale studies over long periods of time possible, studies that would be almost impossible to carry out using only manual measurements.

We also identified a number of problems during the measurement period, the main ones being:

- The implants are not as user-friendly as they could be. During calibration and tuning, a period of 60 seconds between measurements is much too long. However, during measurement, 16 seconds is (mostly) too short. Since the broadcast interval cannot be changed after encapsulation, 60 seconds had to be selected with unpleasant consequences for the calibration and tuning work.
- The implants were sensitive to external disturbances. Because the communication implant-antenna-receiver is carried by an electromagnetic (EM) field, it has to compete with all other surrounding EM radiation to be detected by the receiver. The two ways of ensuring this is to 1) select a frequency far away from other frequencies and 2) broadcast a strong signal. As for point 1, there's no free choice, use of the EM spectrum is closely regulated. So the only way open for improvements is to transmit a stronger signal.

Based on these and other observations, it was decided to develop a new generation of the implant. The core of the implant, i.e. where the temperature is measured, will remain unchanged but several changes related to signal strength and practical handling will be made. Here is a summary of planned improvements:

- Redesigned electronics will result in a much higher signal output.
- The broadcast interval will be selectable from the list 15, 30, 60, 120 and 240 seconds and can be set externally with a 'remote control'.
- The implant can be made to 'go to sleep' during which it consumes current in the order of 1 μA. In 'sleep', the implant has a life span measured in years. The sleep mode is also controlled from the 'remote'.
- It will be possible to set the current used for transmission to be one of the following values: 0.5 mA, 1mA, 2mA, 4mA, 8mA. The lower value can be used where distances are small and/or external disturbances are low. In situation where there is much noise or where antennas cannot be mounted close to the animals, the higher current values may be selected, thereby increasing the signal strength and range. The flip side of the stronger signal is shorter battery life: a higher current will drain the battery sooner.

Ethical statement

All studies were performed in accordance with the Council of Europe Convention (ETS123)/ Directive (86/609/EEC) for the protection of vertebrate animals used for experimental and other scientific purposes, and with approval of the Animal Care Committee of MSD Animal Health.

Literature study

In animal research various systems are used to measure body temperature, and literature on comparisons of methods of body temperature measurement is extensive. However, almost none of these systems are suitable for continuous core body temperature measurement in pigs. Several studies [1,2,3] using infrared thermography to predict rectal temperature concluded that there is not a strong relationship between rectal and skin temperature and that infrared thermometry is not a good alternative to rectal temperature. Subcutaneously injectable transponders are widely used to measure body temperature [4,5,6]. However, this is not a continuous measurement and transponders need to be scanned with a handheld reader at a distance of 2-5 cm from the transponder to measure temperature. In ruminants [7] and horses [8] it is possible to use transponders that are administered orally and stay in the rumen or gastrointestinal tract. Although these 'pills' measure temperature continuously, from a distance, in unsedated freely moving animals, these orally administered transponders are not suitable for long term measurements in swine, because these transponders will pass through the gastrointestinal tract in about 48 hours before being pooped out [9]. Therefore, to our knowledge the telemetry implants used in this study are the best option for long term continuous core body temperature measurements in freely moving swine.

References

1. Sikoski, P., Banks, M.L., Young, R.W., Wallace, J.M., Nader, M.A. (2007). Comparison of rectal and infrared thermometry for obtaining body temperature in cynomolgus macaques (Macaca fascicularis). *Journal of Medical Primatology* **36**, 381-384.

2. Stephens Devalle, J.M. (2005). Comparison of Tympanic, Transponder, and Noncontact Infrared Laser Thermometry with Rectal Thermometry in Strain 13 Guinea Pigs (Cavia porcellus). *American Association for Laboratory Animal Science* **44**, 35-38.

3. Dewulf, J., Koenen, F., Laevens, H., de Kruif, A. (2003). Infrared thermometry is not suitable for the detection of fever in pigs. *Vlaams Diergeneeskindig tijdschrift* **72**, 373-379.

4. Torrao, N.A., Hetem, R.S., Meyer, L.C.R., Fick, L.G. (2011). Assessment of the use of temperature-sensitive microchips to determine core body temperature in goats. *Veterinary Record* **168**, 328.

5. Chen, P.H., White, C.E. (2006). Comparison of rectal, microchip transponder, and infrared thermometry techniques for obtaining body temperature in the laboratory rabbit (Oryctolagus cuniculus). *Journal of the American Association for Laboratory Animal Science* **45**, 57-63.

6. Quéva, C., Bremner-Danielsen, M., Edlund, A., Ekstrand, J., Elg, S., Erickson, S., Johansson, T., Lehmann, A., Mattsson, J.P. (2003). Effects of GABA agonists on body temperature regulation in GABA B(1) mice. *British Journal of Pharmacology* **140**, 315–322.

7. Signer, C., Ruf, T., Schober, F., Fluch, G., Paumann, T., Arnold, W. (2010). A versatile telemetry system for continuous measurement of heart rate, body temperature and locomotor activity in free-ranging ruminants. *Methods in Ecology and Evolution* **1**, 75-85.

8. Green, A.R., Gates, R.S., Lawrence, L.M. (2005). Measurement of horse core body temperature. *Journal of Thermal Biology* **30**, 370-377.

9. Brown-Brandl, T.M., Yanagi, T. Jr., Xin, H., Gates, R. S., Bucklin, R.A., Ross, G.S. (2003). A new telemetry system for measuring core body temperature in livestock and poultry. *Applied Engineering in Agriculture* **19**, 583-589.

Imposing Cognitive Load to Unmask Prepared Lies: A Recurrent Temporal Pattern Detection Approach

B. Diana, V. Zurloni, M. Elia

Department of Human Sciences for Education, University of Milan-Bicocca, Milan, Italy. b.diana2@campus.unimib.it

Abstract

According to Vrij, the differences between liars and truth tellers are greater when interviewees report their stories in reverse order than in chronological order. We intend to explore such influences on the organization of behavior within the framework of the T-pattern model. This model grants the possibility to measure and compare behavior patterns between liars and truth tellers, not easily detectable without the use of such tools. Recalling stories in reverse order will produce cognitive overloading in subjects, because their cognitive resources are already partially spent on the lying task; this should emphasize non verbal differences between liars and truth tellers. During the experiment, we asked female students to report specific autobiographical episodes. We videotaped them as they reported the stories in chronological order or in reverse order after asking to lie about one of the stories. We focused in analyzing how people organize their communicative styles during both truthful and deceptive interactions.. We coded the video recordings, after establishing the ground truth, using Theme Coder 5 software. We are currently analyzing datasets using Theme 6 software.

Keywords. T-patterns, Theme, non-verbal cues to deception, cognitive load, lie detection.

Introduction

Achieving high information assurance is complicated due to human fallibility in deception detection [1]. Two studies [2, 3] reveal that although people show a statistically reliable ability to discriminate truths from lies, overall accuracy rates average 54% or only a little above chance. Moreover, the average total accuracy rates of professional lie catchers (56%) is similar to that of laypersons.

A meta-analysis of cues to deception [4], which included above 130 studies published in English examining nonverbal cues to deception, reveals that many conflicting results have been found. Since no diagnostic cue to deception occurs, it could be that a diagnostic pattern does arise when a combination of cues is taken into account [3].

Moreover, cues to deception are typically faint and unreliable [4]. A contributing factor is that the underlying theoretical explanations for why such cues occur, like nervousness and cognitive load, also apply to truth tellers [5]. Studies in the past have focused on eliciting and amplifying emotions [3] for example by asking questions, but it is uncertain whether this procedure will necessarily raise more concern in liars than in truth tellers. Conversely, only a few efforts focused on unmasking the liars by applying a cognitive lie detection approach [6,7].

Instead of searching for universals in cues to deception, we focused in analyzing how people organize their communicative styles during both truthful and deceptive interactions. According to Vrij, the differences between liars and truth tellers are greater when interviewees report their stories in reverse order than in chronological order [6]. The innovativeness of the present work consists in using the methodological approach used in [8, 9], that is, the T-pattern model approach. We intend to explore the influences on the organization of behavior within the framework of the T-pattern model. This model grants the possibility to measure and compare behavior patterns between liars and truth tellers, not easily detectable without the use of such tools [1, 2, 3]. Recalling stories in reverse order will produce cognitive overloading in subjects, because their cognitive resources are already partially spent on the lying task [6]; this should emphasize non verbal differences between liars and truth tellers.

Methods

Participants

16 Students, all females, aged from 21 to 26. After being recruited, all participants gave their informed consent both to audio and video recording.

Instruments

Both audio and video recording equipment.

Setting

The present study was carried at the University of Milano-Bicocca in an audio-isolated laboratory room equipped with four cameras, set to video-record participants' full-lengths and close-ups. The cameras were connected with a 2 channel quad device (*split-screen* technique).

Procedure

We asked participants to consider two specific episodes, regarding the last time they had been to a party and the last time they went out for pizza. One of these two episodes had to be the truth, the other a lie. They were given 20 minutes to prepare the two stories, knowing that after they would have to report both episodes to another person (a confederate), who didn't know which was the lie. We asked half of the students to tell the pizza episode first, while the other half did the opposite. All of them were told to lie during the first episode. We controlled motivation by telling them that, if they succeeded in telling the lie (meaning, the other person couldn't tell which episode was a lie and which was the truth), they would receive extra credits.

In the first experimental condition, the interviewer asked the interviewee to report both episodes in chronological order. In the second condition, the interviewer asked to report both episodes in reverse mode (starting from the end of the story and going back to the beginning).

The time frame for the experimental task was 10 minutes for each participant (5 minutes per episode), marked by audio signals.

Manipulation check

To establish the ground truth and verify cognitive load manipulation and motivation, we asked the subject to complete a questionnaire after the experiment was finished. Later, we watched the video recordings with the participants and asked them when they lied (veracity status).

T-patterns and Theme 5

Data analysis is performed using Theme 5 software distributed by Noldus Information Technology [9]. Theme detects statistically significant time patterns in sequences of behaviors. The term T-pattern stands for temporal pattern; they are based on the timing of events, relative to each other. T-pattern detection [8, 12, 13] was developed for finding temporal and sequential structure in behavior. The algorithm implemented in the software detects repeated patterns of intra- or inter-individual behavior coded as events on one-dimensional discrete scales.

A minimal T-pattern consists of two event types. An event type is a category of observable behavior whereas an event is an instance of behavior occurring at a particular time unit without a duration [8]. Two event types are considered a T-pattern if they both occur at least twice in the behavior record in the same order and their occurrence times are invariantly distributed over time, i.e. their time distances are unlikely random [8].

Post-detection tools can be applied for filtering and analyzing patterns, for example based on their occurrence frequency (how often the pattern is repeated), length (number of events that comprise a pattern) or behavioral content.

Data analysis

The coding grid was built basing on literature review of non verbal cues in lie detection. [3, 4, 10]. We considered body movements (head, trunk, arms, hands, legs and feet), gestures (rythmic, iconic, metaphoric and deictic), self-contacts, gaze and facial micro-movements (FACS action units) [11].

2 minutes observation intervals were considered for subsequent analysis. Videos were coded frame by frame using Theme coder 5 software distributed by Noldus Information Technology [9]. Each behavior occurrence was regarded as a case of an event-type that occurs at a particular point on a discrete time scale, but has no duration otherwise. The occurrences of each event-type within the selected observation period form the so called "T-dataset". To assess inter-rater reliability of the T-dataset, Cohen's Kappa was calculated on 10% of the same video materials independently coded by two coders, using a "blind" coding procedure. Although differing through categories, inter-coder reliability was found to be good to satisfactory (ranging from .70 to .92; $p < .05$). When disagrements were identified or the agreement was not perfect, the specific cases were discussed and agreed by both coders.

Results

On-going work. We are currently analyzing datasets.

References

1. Qin, T., Burgoon, J.K., Nunamaker, J.F. (2004). An exploratory study on promising cues in deception detection and application of decision tree. *Proc. 37th Annual Hawaii International Conference on System Sciences* , 23-32.

2. Bond Jr, C.F., DePaulo, B.M. (2006). Accuracy of deception judgments. *Personality and social psychology Review* **10**(3), 214-234.

3. Vrij, A. (2008). Nonverbal dominance versus verbal accuracy in lie detection. *Criminal Justice and Behavior* **35**(10), 1323.

4. DePaulo, B.M., Lindsay, J.J., Malone, B.E., Muhlenbruck, L., Charlton, K., Cooper, H. (2003). Cues to deception. *Psychological bulletin* **129**(1), 74.

5. Vrij, A., Granhag, P.A., Mann, S., Leal, S. (2011). Outsmarting the liars: toward a cognitive lie detection approach. *Current Directions in Psychological Science* **20**(1), 28-32.

6. Vrij, A., Mann, S.A., Fisher, R.P., Leal, S., Milne, R., Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and human behavior* **32**(3), 253-265.

7. Vrij, A., Leal, S., Granhag, P.A., Mann, S., Fisher, R.P., Hillman, J., Sperry, K. (2009). Outsmarting the liars: The benefit of asking unanticipated questions. *Law and human behavior* **33**(2), 159–166.

8. Magnusson, M.S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods* **32**(1), 93-110.

9. Magnusson, M.S. (2004). Repeated patterns in behavior and other biological phenomena. In Griebel, U. *Evolution of communication systems: a comparative approach.* The MIT Press, 111-128.

10. Burgoon, J.K., Guerrero, L.K., Floyd, K. (2009). *Nonverbal Communication*. Allyn & Bacon, London, UK.

11. Ekman, P., Friesen, W.V. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Consulting Psychologists Press, Palo Alto, CA.

12. Borrie, A., Jonsson, G.K., & Magnusson, M.S. (2002). Temporal pattern analysis and its applicability in sport: an explanation and exemplar data. *Journal of Sports Sciences* **20**(10), 845-852.

13. Magnusson, M.S. (2005). Understanding Social Interaction: Discovering Hidden Structure with Model and Algorithms. In Anolli, L., Duncan, S., Magnusson, M.S., Riva, G. (eds.) *The hidden structure of interaction: from neurons to culture patterns*, 2-22. IOS Press, Amsterdam.

The Novel Object Recognition Test in Rodents: Which Are the Essential Methodological Aspects?

E. Moscardo, B. Salvetti, S. Becchi, G. Bertini, P.F. Fabene

Department of Neurological, Neuropsychological, Morphological & Motor Sciences, Section of Anatomy and Histology, Laboratory of Functional Neuroanatomy, University of Verona, Faculty of Medicine, Verona, Italy.

elena.moscardo@univr.it, beatrice.salvetti@univr.it; serena.becchi@gmail.com, giuseppe.bertini@univr.it; paolo.fabene@univr.it

Introduction

The Novel Object Recognition test (NOR) has been introduced by Ennanceur and Delacour in 1988 [1] and can be regarded as a spontaneous Delayed-Non-Matching-to-Sample (DNMS) [2] test. The test is based on a spontaneous behaviour: the main assumption at the base of this test is that access to novelty (e.g. an object or an environment) can elicit approach behaviours in animals. This apparent 'unconditioned preference' for novelty has been used in the NOR test in order to study memory functions, assessing the ability of animals to recognize a novel object in a familiar environment, because they maintain a representation of those is more familiar stored in memory

The Novel Object Recognition (NOR) test has been used in different variants, but typically consists in two trials. In the first trial (T1) the animal is exposed to one or two identical objects (sample object). Following the sample object exposure, the animal is returned to his home cage for a retention period. In the second trial (T2), which follows the retention time, the animal is returned to the environment (arena) and presented with a familiar (sample) and a novel object. When the subject 'remembers' the previous exposure to the familiar object, it will explore the novel object to a greater degree than that of the familiar one.

NOR test doesn't involve reference memory components (e.g. explicit rule learning), thus it can be considered a "pure" recognition memory test and a valid task to assess working memory. Finally the test doesn't involve positive or negative reinforces (e.g. food, electric shocks) and this makes NOR comparable to memory tests currently used in humans. All these advantages make NOR test quick and simple to be implemented and, therefore, it has been widely used for assessing mild cognitive impairment in pre-clinical research.

Analysis of available literature

Over the past 30 years or so, several hundred published papers report the use of the NOR test. Given NOR test wide use, the available literature on NOR test is abundant and there are quite lot differences in target aim of using this memory task, in the methodological approaches, and how results are interpreted, both in Academia and Industry. A comprehensive PubMed search on NOR test applied in Rodents from 1988 to 2010 retrieves 1389 papers. We restricted our in-depth analysis on published studies that explicitly and accurately describe the methodological aspects of the NOR task and the following parameters were identified, collected and compared:

- Animals - Species, strain, age, sex.
- Apparatus – large variety in shapes (circular, rectangular, Y and T maze) and in dimensions.
- Objects – large variety of shapes, materials, colours and odours associated.
- Time schedules – variety in durations.
- Behavioural measures – variety of data presentation.
- Other variables – habituation to test arena, handling, light-dark cycle, background noises, illumination of apparatus.

The aim of this exhaustive and detailed analysis of the available literature is to possibly provide guidelines on how NOR test could be better conduced in Rodents (rats and mice) in order to generate reliable and ethologically significant data, which can be easily compared between different laboratories. In addition, we intend to discuss what methodological aspects should be mentioned in NOR's papers and the reasons why they cannot be omitted.

Methodological aspects

1. **Apparatus** - There is a large heterogeneity among the analyzed literature for what concern the apparatus used in NOR task. The floor space of apparatus used were largely different across experiments in the different laboratories even if they had similar dimensions for rats and mice. The NOR task is generally performed in square or rectangular arenas or open fields and only few papers reported circular apparatus. It emerges that some variables of the apparatus are important for a successful NOR test: use age-appropriate apparatus size; allow sufficient environmental familiarization in order to minimize anxiety and stress related behaviors and decrease competition for environmental exploration; decrease more as possible extra-environment cues; the presence of bedding in the apparatus should be avoided to reduce distractor factors; the apparatus has to be built in washable materials to remove olfactory cues between test's pahses.

2. **Objects** - In the majority of studies analyzed, objects to be discriminated were made of odorless and easy to clean materials and there is a particular attention in changing them with identical copies when starting the choice phase, so that they could not readily be distinguished by olfactory cues. On the contrary there is a large variability between research groups on the characteristics of objects used, and they differ for shape, size, color, brightness, material, texture and combination of these variables. Within the NOR test, the researchers agree in using similar sized objects both in sample and choice phases and, as for the apparatus, also object sizes should be age-appropriate. First and foremost, the objects need to have no ethological significance for the animals and it is worth checking for any intrinsic preference for one of the two objects and its complexity (i.e. must be evaluated that animals spend similar amount of time interacting with both objects).

3. **Time schedules**

   - Sample phase: The evidences from reviewed literature suggest that a minimal amount of sample objects investigation is necessary to develop a novel object preference. It has to be taken into account that sample phase duration affects the length of subsequent retention, thus the use of a fixed duration is preferable over a limited cut-off duration, which doesn't always ensure a sufficient familiarization with the objects. On the other hand, sample objects are explored mostly in the first minutes of this phase; therefore a long sample phase (i.e. more than 10 min) may not be useful.

   - Retention phase: The length of the delay phase largely depends upon the type and aim of the experiment. Delay times have to be chosen in relation to the sample phase duration, keeping in mind that a shorter familiarization has been found to be related to a shorter retention period; it is worth not to choose delay over 24 hours and to avoid excessively short delays, which may cause a pavement effect.

   - Choice phase: The time of exposure more frequently used in choice phase is 1-3 minutes for rats and 3-10 minutes for mice, generally the same duration of the sample phase is maintained. Since it has been observed that preference for novelty in rodents is typically stronger during the first 2 minutes of the choice phase, but gradually diminishes during the last minutes, it is recommended to analyze the trial minute by minute during this phase.

4. **Other variables**: Most of the studies analyzed do not mention if the animals are undergone the NOR task during their light or dark period. Since rodents are nocturnal animals, they exhibit highest level of activity during the dark phase and, for this reason, we support papers which conduct the NOR test during the animals' dark/active phase, possibly using an inverted light/dark cycle. Considering that for

the NOR task a variety of habituation protocols has been used, it is recommended to increase environmental and apparatus familiarization in order to minimize the stress and novelty confounding factors, increasing as a consequence objects interaction and discrimination.

References

1. Ennanceur, A., Delacour, J. (1988). A new one-trial test for neurobiological studies of memory in rats. 1: Behavioral data. *Behavioural Brain Research* **31**(1), 47-59.

2. Aggleton, J.P., Hunt, P.R., Rawlins, J.N. (1986). The effects of hippocampal lesions upon spatial and non-spatial tests of working memory. *Behavioural Brain Research* **19**(2), 133-46.

Spatial Learning Characteristics of Transgenic Mice as Revealed by Detailed Video-Analysis Using EthoVision

Valeria Melis[1,2], Gernot Riedel[1,2], Charles R. Harrington[1,2]

[1]*Institute of Medical Sciences, University of Aberdeen, Aberdeen, United Kingdom. v.melis@abdn.ac.uk*

[2]*TauRx Therapeutics Ltd, Aberdeen, United Kingdom.*

The water maze (WM) test was first described in the early 1980s [1] and, currently, is one of the most widely used tasks in behavioural neuroscience to explore spatial memory in rodents. Methodological variations of the WM task have been developed and are being used by researchers in many different applications for instance, to test the impact of aging, to investigate various neurodegenerative diseases or for the screening of novel therapeutic drugs [2]. In order to take full advantage of the assessment of rodent models by means of the WM test, consideration of the behavioural paradigms employed is critical in investigating the phenotype under scrutiny. In addition, common proxies used to quantify cognitive abilities of animals are not always satisfactory and complete descriptors of behavioural anomalies. Standard performance measures include: path length (i.e. the length of the path swam from the release point to the platform) and latency (i.e. the time the animal takes to find the platform) as indices of learning achieved during training. Learning is usually defined by shorter latencies and by decreased path length. Furthermore, swimming velocity, as measure of genotype-related effects on motor activity and thigmotaxis, defined as the proportion of time spent swimming along the edge of the pool which is an index of animal's anxiety, are considered.

Nowadays, several reliable video tracking systems (e.g. EthoVision, Noldus Information Technology; AnyMaze, Ugo Basile; HomeCageScan, Clever Systems; etc.) for automated behavioural research are available. Although such software programs have the benefit of measuring a wide range of quantitative performance parameters, very basic indices are preferred to more complicated and time-consuming quantifications, resulting in an oversimplified description of the behaviour. For example, animals may use different search strategies; an animal that swims close to the platform during training may display the same latency or path length to the platform as animals that use a circular strategy, but do not show direct search toward the platform area [3]. As a consequence, investigation of path shape parameters would be helpful, for example, in discriminating between groups of animals that show otherwise no differences according to the usual parameters. Path shape or strategy may be determined as a) "cumulative distance to platform" (or "proximity") [4] which is the distance (cm) between the position of the animal and the platform location measured several times per second (the closer the animal swims to the platform, the shorter the cumulative distance); b) "turn angle" (degrees), that is the change in swimming direction; c) "angular velocity" (degrees/s) which indicates the change of direction of movement of the animal in the maze per time unit and d) "meander" (or " tortuosity") (degrees/cm) which is the amount of turning per distance unit.

The aim of this study was to provide a more detailed analysis of the water maze performance of a transgenic mouse model of Alzheimer's disease overexpressing mutated tau in brain, termed Line 1 [5], using EthoVision 3.1 (Noldus Information Technology, Wageningen, The Netherlands). Two different WM paradigms were used to investigate cognitive functions in 5-month old Line 1 and age-matched wild-type mice, namely the standard reference memory (wild-type: n = 10; Line 1: n = 13) and the problem-solving (wild-type: n = 10; Line 1: n = 12) WM tasks. The former requires the incremental learning of a constant platform location over multiple days of training [6]. By contrast, in the spatial problem-solving task, which was adapted from a protocol elaborated by Chen and co-workers [7], an initial visual pre-training is followed by a training to a hidden platform until a criterion is met (problem), then the platform location is changed and animals are trained until the same criterion of the next problem is met [8]. Mice were housed in groups of up to 10 until the beginning of the study and continuing through the completion of behavioural testing. Animals were allowed food and water ad libitum and were kept under standard conditions (temperature 20−21°C, 60−65% relative humidity) on a 12 hours light/dark cycle (light on at 7:00 a.m.). Tests took place during the light phase of the cycle. All experiments were conducted in strict accordance with the UK Home Office regulations outlined in the Animals (Scientific Procedures) Act 1986.

Path length, latency, swim speed and thigmotaxis were measured for both paradigms. In addition, the percentage of time the animal spent within the quadrant where the platform was located during training, as an index of spatial memory, was recorded for the standard reference memory WM task whereas the number of trials required to reach criterion for each problem was calculated as a measure of the animal’s learning capacity for the problem-solving WM paradigm. Standard parameters revealed a learning impairment for transgenic mice in both paradigms along with an increased swimming speed and a greater level of anxiety. For the standard reference memory WM, the learning impairment was also confirmed by the decreased proximity to the target platform location (see Figure 1). The analysis of the number of trials to reach criterion in the problem solving task highlighted a learning deficit for Line 1 mice in achieving the first problem only. On the other hand, the proximity parameter revealed a more robust phenotype for transgenic animals in achieving all the three problems (see Figure 2). This parameter reflects the searching pattern of the animal and a goal-directed searching strategy that centers on the area of the platform location results in a shorter cumulative distance as compared with random patterns. Therefore, the longer cumulative distance recorded for Line 1 mice suggests a less efficient searching strategy in the maze.

Furthermore, both paradigms revealed a decreased turn angle, angular velocity and path tortuosity for Line 1 mice when compared to wild-type animals. This may suggest either a different search strategy or a reduced exploratory response toward the environment. In support of this latter view is the observation that Line 1 mice exhibited increased level of anxiety measured as thigmotaxis. However, a further manual categorization of search strategy may help in the analysis of performance since it provides a qualitative description of complex explorative behavior.

In conclusion, the present study revealed that the rarely used shape parameters along with the traditional measures allow a more refined and comprehensive analysis of water maze behaviour.

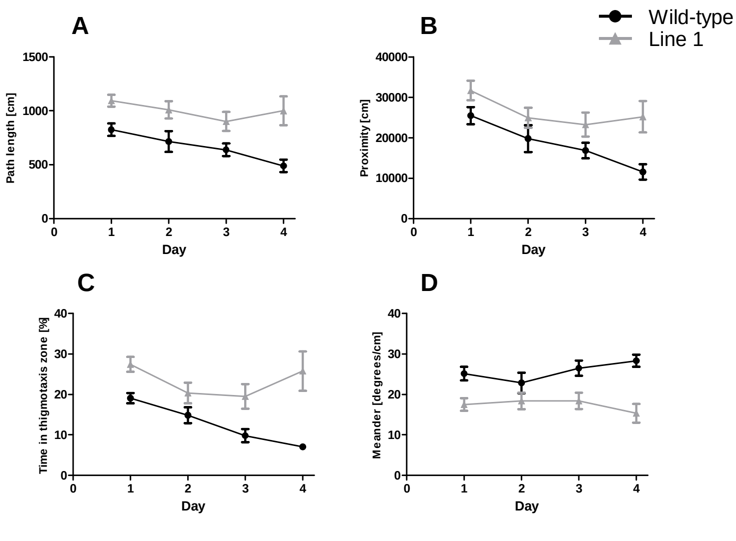


Figure 1. Spatial learning deficits in Line 1 mice in the standard reference water maze paradigm. Swim path length (A): Line 1 exhibited an inferior performance compared to wild-type mice [F(1,63) = 16.25; p = 0.0006]. Proximity (B): Learning deficit in transgenic animals was also confirmed by the decreased proximity to the target platform location [F(1,63) = 8.01; p =0.01]. Thigmotaxis (C): Line 1 mice presented with greater level of anxiety [F(1,63) = 13.23; p = 0.001]. Meander (D): the swim path tortuosity was significantly reduced for Line 1 mice [F(1,63) = 24.88; p < 0.0001]. Data were analysed with repeated measures two-way ANOVA with genotype and day as independent variables. For all comparisons, a 95% confidence level (p < 0.05) was set for the differences to be considered as significant. Values are expressed as a daily mean (± S.E.).

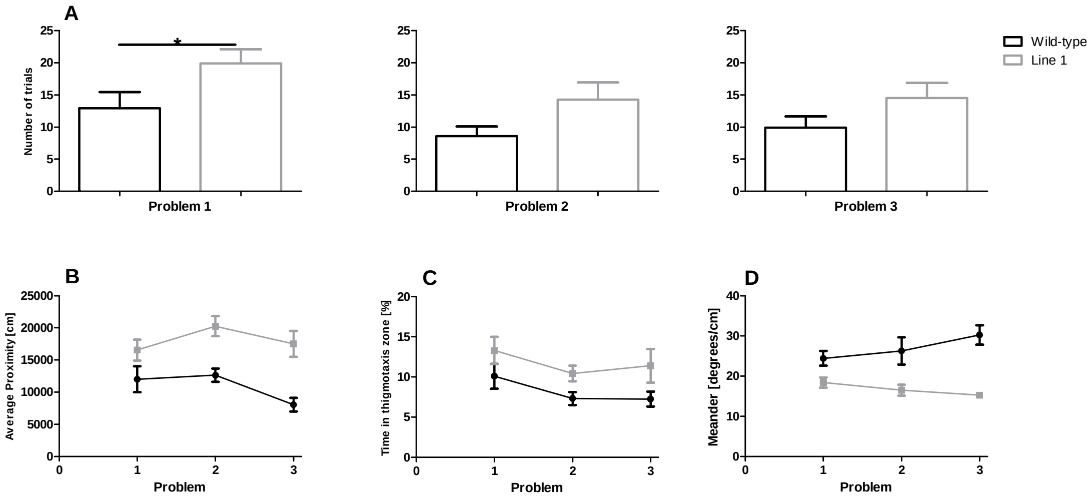


Figure 2. Spatial learning deficits in Line 1 mice in the problem-solving water maze paradigm. Number of trials to problems (A): Line 1 needed longer training to attain the first problem (Student t-test: $t = 2.1$; $p = 0.04$); no reliable difference was obtained in reaching criterion for the other two problems (p's > 0.05). * indicate statistically significant difference ($p < 0.05$). Proximity (B): Transgenic mice presented with greater proximity in achieving criterion of all problems [two-way ANOVA: $F(1,38) = 22.36$; $p = 0.0001$]. Thigmotaxis (C): Similarly to results obtained in the standard reference water maze task, Line 1 mice spent more time swimming along the edge of the pool [two-way ANOVA: $F(1,38) = 5.57$; $p = 0.02$]. Meander (D): Wild-type mice presented with a greater amount of turning per distance unit [two-way ANOVA: $F(1,38) = 33.57$; $p < 0.0001$]. For all comparisons, a 95% confidence level ($p < 0.05$) was set for the differences to be considered as significant. All values are expressed as a mean (± S.E.).

References

1. Morris, R.G., Garrud, P., Rawlins, J.N., O'Keefe, J. (1982). Place navigation impaired in rats with hippocampal lesions. *Nature* **297**(5868), 681-683.
2. D'Hooge, R., De Deyn, P.P. (2001). Applications of the Morris water maze in the study of learning and memory. Brain Research. *Brain Research Reviews* **36**(1), 60-90.
3. Schenk, E., Morris, R.G. Dissociation between components of spatial memory in rats after recovery from the effects of retrohippocampal lesions. *Experimental Brain Research* **58**, 11-28.
4. Gallagher, M., Burwell, R., Burchinal, M. (1993). Severity of spatial impairment in aging: Development of a learning index for performance in the Morris water maze. *Behavioral Neuroscience* **107**, 618-626.
5. Zabke, C., Dietze, S., Stamer, K., Rickard, J.E., Harrington, C.R., Theuring, F., Seng, K.M., Wischik, C.M. (2008) P1-054. Early and advanced stages of Tau aggregation in transgenic mouse models. *Alzheimer's and Dementia* **4**, T221-T222.
6. Deiana, S., Harrington, C.R., Wischik, C.M., Riedel, G. (2009). Methylthioninium chloride reverses cognitive deficits induced by scopolamine: Comparison with rivastigmine. *Psychopharmacology* **202**(1-3), 53-65.
7. Chen, G., Chen, K. S., Knox, J., Inglis, J., Bernard, A., Martin, S.J., *et al*. (2000). A learning deficit related to age and beta-amyloid plaques in a mouse model of Alzheimer's disease. *Nature* **408**(6815), 975-979.
8. Melis, V., Deiana, S., Zabke, C., Stamer, K., Harrington, C.R., Riedel, G., Theuring, F., Seng, K.M. & Wischik, C.M. (2008) P2-383. Reversal of spatial problem-solving cognitive deficit in a transgenic mouse line overexpressing the repeat domain of Tau by treatment with the Tau aggregation inhibitor methylthioninium chloride. *Alzheimer's and Dementia* **4**, T485.

Development of a Visual-guided Probabilistic Selection Task for Rats

A. Trecker[1], T.W. Robbins[2], A.C. Mar[2]

[1]*Department of Comparative Psychology, Institute of Experimental Psychology, Heinrich-Heine-University Düsseldorf, Germany. andrea.trecker@email.de*

[2]*Department of Experimental Psychology, Behavioural and Clinical Neuroscience Institute, University of Cambridge*

Reinforcement learning can be defined as a trial and error learning of how to maximize rewards and/or minimize punishments through repeated sampling or interaction with the environment. This implicit, unsupervised form of learning is an essential component of daily adaptive behaviour and is observed to be disrupted in several neuropsychiatric disorders such as Parkinson's disease and schizophrenia [1, 2, 3]. The observation that phasic firing of midbrain dopamine neurons may encode reward prediction errors during reinforcement learning [4] has fueled considerable interest in the neural mechanisms of reinforcement.

While probabilistic reinforcement learning paradigms using visual objects are commonly-used in research with humans and non-human primates, to our knowledge, no fully-validated paradigm currently exists for the rodent. The aim of the present study was thus to develop a reinforcement learning paradigm (Probabilistic Selection Task) for rats with based on the experimental methods and computational framework of Frank and colleagues [1].

Experiments were conducted using 16 male lister hooded rats. Rats were tested in Med Associate Inc. operant chambers. One wall of a standard operant chamber was removed and replaced with an infrared touch sensitive flat-screen monitor at the front. House light and food magazine containing built-in photocells was located on the rear wall.

Behavioural training proceeded as follows:

In phase 1, rats were trained to press a white rectangle (3 x 15 cm) at the bottom centre of the screen in order to receive immediate delivery of a food pellet reward. After rats completed 180 trials within a session (max. 30 minutes), the size of the rectangle was reduced to its final size of 3 x 6 cm.

In phase 2, touching the rectangle resulted in trial initiation rather than a food reward. The rationale for using the initiation rectangle on the screen was to keep attention directed toward the screen and to prevent biases (sometimes observed in other touchscreen tasks) due to stereotyped turning responses from the food magazine at the rear of the chamber. Upon trial initiation, rats were presented with two distinct visual stimuli (e.g., spider and plane, each 8 x 8 cm) in two of three possible positions (y from bottom = 3 cm, x from left = 0, = 7,6, = 15,3 cm) on the touchscreen. The positions of the stimuli were randomized on every trial. Stimuli were also counterbalanced across the reward probability and presentation order of the reinforcement pairs. For designated AB pairs, touching stimulus A resulted in 90% chance of reward (10% chance of non-reward), while touching stimulus B only a 10% chance of reward (90% non-reward). Feedback for non-rewarded choices were followed by a 5s timeout with the house light illuminated. Similarly, designated CD pairs were associated with 80%:20% reward ratios and designated EF pairs with 70%:30% reward ratios. The trial initiation rectangle was presented immediately after reward collection or the timeout period to begin the next trial. Sessions lasted a maximum of 30 min or after completion of 180 trials.

In Phase 3, after achieving the learning criteria in Phase 2, subjects are then presented with the original stimuli in novel pairings to probe the contributions of positive and negative feedback to probabilistic learning by assessing whether rats displayed a bias for choosing frequently-reinforced stimuli, or for avoiding frequently-punished stimuli.

As expected we observed a significant effect of designated pair in phase 2, with the highest choice ratios (e.g., A/(A+B)) seen for AB (90:10) pairs and the lowest for EF (70:30) pairs. These differences in reinforcement-guided discrimination learning were apparent at the group level after only 7-8 sessions of training. Moreover, novel use of the initiation rectangle resulted in rapid choice latencies and a relative absence a priori stimulus or position biases. Trial by trial data were fit using standard reinforcement learning models to further characterise the differences in learning [5].

In sum, we have developed a visually-guided reinforcement learning task for rodents with parallels to the human Probabilistic Selection Task. This paradigm holds promise for the translation of human reinforcement learning in rodents, and may be of utility in future studies examining its neuropharmacological underpinnings.

References

1. Frank, M.J., Seeberger, L.C., O'Reilly, R.C. (2004). By Carrot or by Stick: Cognitive Reinforcement Learning in Parkinsonism. *Science* **306**, 1940-1093.

2. Waltz, J.A., Frank, M.J., Robinson, B.M., Gold, J.M. (2007). Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal-cortical dysfunction. *Biol Psychiatry* **62**, 756-764.

3. Yilmaz, A., Simsek, F., Gonul, A.S. (2012). Reduced reward-related probability learning in schizophrenia patients. *Neuropsychiatr Dis Treat* **8**, 27-34.

4. Schultz, W., Dayan, P., Montague, P.R. (1997). A neural substrate of prediction and reward. *Science* **14**, 1593-1599.

5. Sutton, R.S., Barto, A.G. (1998). *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA.

Rodent Fear Conditioning: Development and Pharmacological Validation of a Video-Based Freezing Detection System

H. Van Craenendonck[1], L. Ver Donck, D.J. Pemberton

Dept Neuroscience, Janssen Research and Development, a Division of Janssen Pharmaceutica NV, Beerse, Belgium.
[1]hvcraene@its.jnj.com

Abstract

Rodent fear conditioning is a relatively simple, cognitive-based paradigm extensively used to study the neurobiological mechanisms underpinning learning and memory. Prior to the introduction of video-based acquisition systems, manual observer scoring was the only available method to capture the freezing responses of animals. There are however considerable downsides to this approach including the risk of inter-animal and/or study variability as well as the risk of unintentional operator bias. These issues can lead to inaccurate results and difficulties in replicating studies across laboratories. The relatively recent introduction of video-based behavioral acquisition systems now allows investigators to record and analyse fear conditioning studies using locomotor detection algorithms. Our preference for a video based system over other fear conditioning recording systems is that this allows the investigator to compare the system generated data with the actual animal behaviour. We entered into a joint risk-sharing collaboration with Biobserve, Germany, with the objective to develop a bespoke fear conditioning video acquisition and analysis system.

This collaboration started in spring 2010 and after a four month design and data exchange period, followed by trial running a beta-version, the final system was delivered to us in December 2010. The software package does not require object identification, but analyses differences between subsequent video images to quantify the activity (or lack of) of the rat. It also includes unique features in its functionality including individual software based camera sensitivity filters that can be altered dependent on the types of the cameras used. This functionality also allows the operator to individually adjust each camera's sensitivity so that small differences in the test cage environment (e.g. subtle differences in illumination) can be balanced across the test chambers. The software can also drive additional hardware (e.g. shock grids, speakers) using basic programming skills. In addition, the package also contains statistical analysis capabilities (details are available at http://www.biobserve.com/products/fear-test/fear-test.html).

The fear conditioning test involves a conditioning phase, during which animals are placed in a test chamber for 4 min and receive an electric foot shock after 2 min (conditioning stimulus, CS). The second phase 24h later consists of a recall test, where the animals are placed in the same test chamber (context, or unconditioning stimulus, US) for 8 min, but do not receive a foot shock. After the foot shock in phase 1, and during phase 2, the animals show freezing behaviour. This freezing is represents a fear response to the foot shock, and the duration of freezing during phase 2 is taken as a measure of how well the animals can still associate the context (test chamber, US) to the foot shock (CS). Test compounds were administered just before phase 1 to assess their effect on memory consolidation. The Institutional Ethical Committee on Animal Experimentation approved the experimental protocols, in compliance with Belgian law (Royal Decree on the protection of laboratory animals dd. April 6, 2010) and the facilities are accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC).

Identification of freezing identified by the algorithm showed a high correlation with that by an observer blinded to treatment. The pharmacological validation studies using compounds that target the cholinergic system, (scopolamine, a muscarinic agonist and donepezil), revealed some unexpected effects during the conditioning phase (CS/US pairing). For example, whilst there was no effect of scopolamine during the initial habituation phase before the foot shock, significant dose-dependent reductions in shock-induced freezing were observed in scopolamine treated animals, suggestive of hyperactivity. During the recall phase after 24h, freezing was reduced compared to previously unshocked rats, indicating recall impairment. Interestingly, while donepezil attenuated this hyperactive response and reversed the scopolamine-induced recall impairment, the selective M1

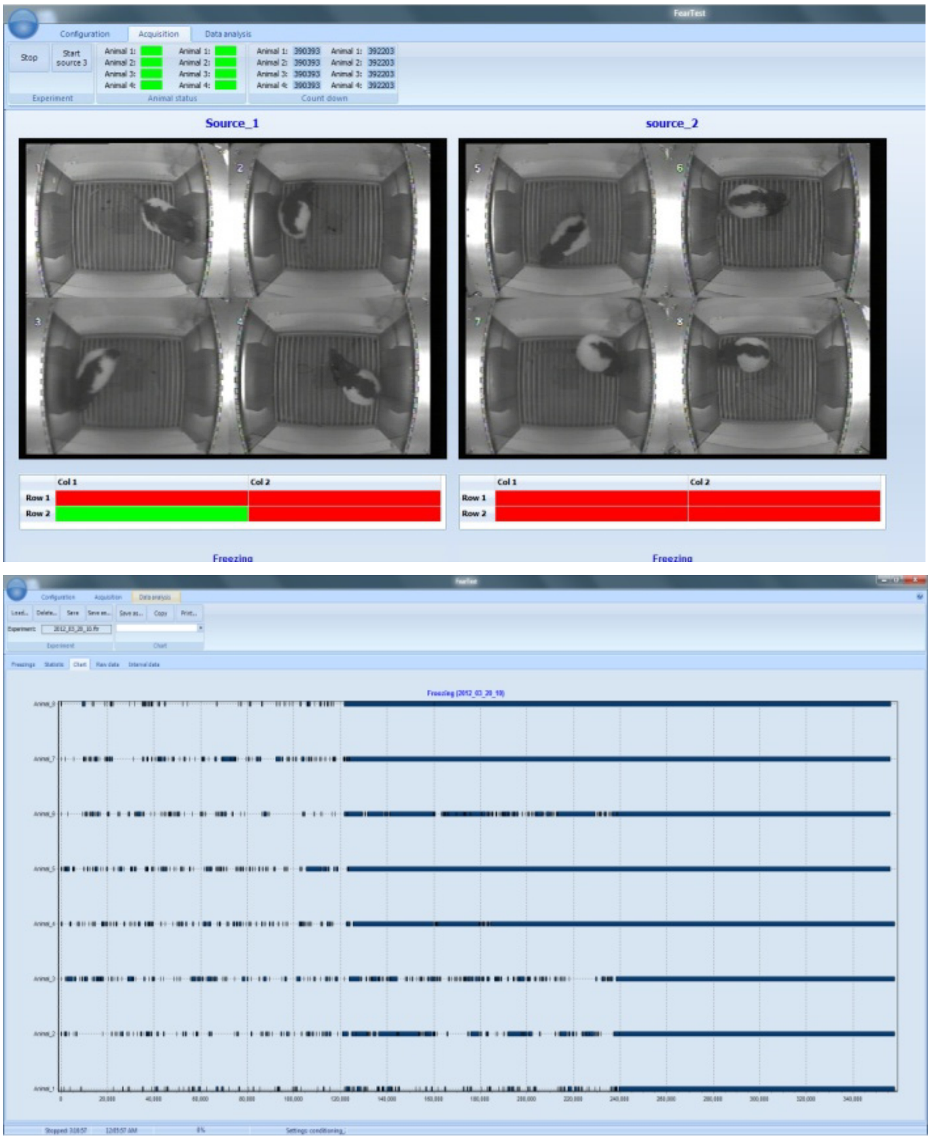


Figure 1. Screen captures of the fear conditioning test software. Top panel shows the image acquisition window; bottom panel shows the ethogram comparison: each horizontal line represents automated scoring of freezing behaviour (bold segments) of a rat.

muscarinic agonist compound-A (3-{1-[4-(2-Methoxyethoxy)-1-methylcyclohexyl]piperidin-4-yl}-2-oxo-2,3-dihydro-1,3-benzoxazole-5-carbonitrile) [1] only reversed the effects of scopolamine on conditioning, but not on recall.

In summary, we have developed a bespoke fear conditioning acquisition software package that allowed us to detect previously unobserved differences in cholinergic compounds. These studies underscore the importance of collecting and interrogating response data during the conditioning phase as interesting differences in pharmacology can be detected.

Reference

1. Cooper et al. (2009). *Compounds which have activity at M1 Receptor and their uses in medicine*. WIPO Patent Application WO/2009/037296.

Obstacle Detection for Autonomous Vehicles in Agriculture

Frits K. van Evert, Ard T. Nieuwenhuizen

Plant Research International, Wageningen, The Netherlands. frits.vanevert@wur.nl, ard.nieuwenhuizen@wur.nl

Precision agriculture is an important technology which can help make agricultural production more sustainable by applying the right kind of inputs, at the right place and at the right time. This 3R concept almost certainly entails more passes through the crop, while driving slower (in order to apply the input at the right place). High labour costs can be avoided by deploying autonomous vehicles. Examples of such autonomous vehicles are the Intelligent Autonomous Weeder [1], a robot to detect and control broad-leaved dock [2], a pesticide applicator for strawberry fields (see Figure 1), and an autonomous golf course mower [3]. These vehicles use GPS to follow a pre-defined path. The broad-leaved dock uses additional sensing to detect and control weeds. The autonomous golf course mower is currently the only one which halts when it senses an obstacle.

Safety of autonomous vehicles in agriculture is obviously important, but has received less attention to date than issues such as navigation, efficient paths, detection of crop condition, and development of actuators which are suitable for autonomous operation. Development of a safety system for autonomous vehicles in agriculture is complicated as the outdoor agricultural environment is often heterogeneous: the soil surface may be rutted, crop rows may curve or run together, and crop plants have variable size and shape even in one field. Detection of obstacles and especially of people is more difficult against this variable background than it would be, e.g., on a factory floor.

Experiments will be conducted in spring and summer 2012 to assess the usefulness for obstacle detection of several kinds of sensors, in several agricultural environments. Sensors included are laser scanners, ultrasonic sensors, a time-of-flight camera, and colour and infrared cameras. Environments included are a strawberry field and an apple orchard. Experiments will be conducted with various kinds of obstacles: for example, a spade left in the field; a child (or crouching adult); an adult standing in the field or among trees; machinery such as irrigation equipment left in the field; a fence; but also bunches of weeds, tall grass, and overhanging branches (in the orchard) that are obstacle-like but need not be avoided by an autonomous vehicle. Experiments will be conducted with a tractor which is driven through the field or orchard.

Figure 1. Autonomous tractor used to apply pesticides in strawberry fields. At present, the only safety feature of this tractor is the front-mounted bar with emergency stop buttons at either end.

Analysis of the data will focus first on individual sensors. Data from each sensor will be filtered to derive stable signals from which the presence of obstacles may be detected. While literature suggests that this approach will give good results, we hypothesize that even better detection of obstacles can be achieved by including knowledge about the semi-structured agricultural environment. Therefore we will explicitly model the environment in a second step of the analysis. In the strawberry field, the strawberry beds, individual plants, and the tractor path are modelled; in the orchard, trees are modelled individually when they are growing separately, or as a "wall" when they are growing very close together. Model parameters include height and width of strawberry plants, and height and shape parameters of the trees. Model parameters will be estimated online with a particle filter. For each kind of sensor deployed, a physical model will be used to describe formation of the signal, and a likelihood function will be developed to relate belief to signal. Data from two or more sensors can be used to update the particle filter. With a clear estimate of what the environment looks like when it is free of obstacles, it is expected that obstacles -deviations from the environment- can be detected with high accuracy.

References

1. Bakker, T., Van Asselt, K., Bontsema, J., Muller, J., Van Straten, G. (2011). Autonomous navigation using a robot platform in a sugar beet field. *Biosystems Engineering* **109**, 357-368. DOI: 10.1016/j.biosystemseng.2011.05.001.

2. Van Evert, F.K., Samsom, J., Polder, G., Vijn, M., Van Dooren, H.-J., Lamaker, A., Van der Heijden, G.W.A.M., Kempenaar, C., Van der Zalm, T., Lotz, B. (2011). A robot to detect and control broad-leaved dock (Rumex obtusifolius L.) in grassland. *Journal of Field Robotics* **28**, 264-277. DOI: 10.1002/rob.20377.

3. Driverless mowing (2011). Retrieved from <http://www.probotiq.com>. Accessed 30 March 2012.

Social Transmission of Food Preference in C57BL/6 Mice

Kaja Plucinska[1], Lianne Strachan, Dominique Peeters, Bettina Platt, Gernot Riedel

School of Medical Sciences, College of Life Sciences and Medicine, University of Aberdeen, Aberdeen, UK. [1]r02kp0@abdn.ac.uk

Abstract

Motivation/Aims

Our understanding of memory processes has greatly benefited from animal research. The assessment of learning and memory in rodents has so far intensely relied on simple tests with easily translatable experimental methods, and thus has concentrated largely on spatial tasks. Socially communicated food bias as a semantic knowledge task has been developed in our laboratory as a modified version of the Galef model of food bias induced through the social interaction with a conspecific [1,2]. This constitutes a natural odour-odour association and fulfils criteria of semantic knowledge including the acquisition during an episodic-like learning event, but the information then becomes unrelated to this mouse-mouse interaction and conceptually independent from the learning episode; it is finally declared in a different environmental setting. Such ‘learning from others’ can be employed in rats and mice, is not induced through a simple presentation of the odour cue, is transmitted by an anaesthetised animal and thus not reliant on the social encounter as such. Furthermore, memory for the novel odour is long-lasting (at least 24 hours) and dependent on elements of the temporal lobe as confirmed through hippocampal lesion studies (for example [3]) and more recently through measurement of activated immediate early genes [4]. Olfactory deficits and memory impairments have been linked to several neurological disorders (such as Alzheimer’s disease, Parkinson’s disease, autism, schizophrenia) with respect to detection, discrimination, recognition and identification [5], hence the need for identifying an early marker for disease prior to clinical manifestations.

In the present study, we carefully examined and constructed new methodologies of investigating olfactory memory in mice using the STFP paradigm. Principle manipulations introduced are the automation of the test which allows the recording of a large number of subjects simultaneously during all phases of the test (habituation, social interaction, test of food preference) and improved methods of measuring activity of the subjects by analysing time spent in particular zones of the arena.

Materials and Methods

Behavioural apparatus

Olfactory discrimination and STFP were assessed in the PhenoTyper cages (30cm x 30 cm x 35cm). The PhenoTyper system (Noldus, Wageningen, Netherlands) is a video-based observation system used for long-term continuous monitoring of behavioural activity in mice. The top unit of the cage contains a built-in digital infrared sensitive video camera and infrared lighting sources for video tracking. The infrared sources provide a constant and even illumination of the floor of the cage that enables continuous recording of the subject’s activity during light and dark periods. Video tracking was performed at a rate of 12.5 samples/ second and recorded by the computer-based tracking software EthoVision 3.0 (Noldus Information Technology, Wageningen, the Netherlands).

Animals

Male (N=24; 12 Observers, 12 Demonstrators) and female (N=24, 12 Observers, 12 Demonstrators) C57BL/6 mice at the age of 6-8 weeks were used for this study (purchased from Harlan, UK). Animals were housed in groups up to 10 subjects per cage prior to testing. Animals were allowed food and water ad libitum and were kept under standard laboratory conditions (temperature 20-21°C, 60-65% relative humidity) on a 12h light/dark cycle (light on at 07:00). The testing took place during the light period.

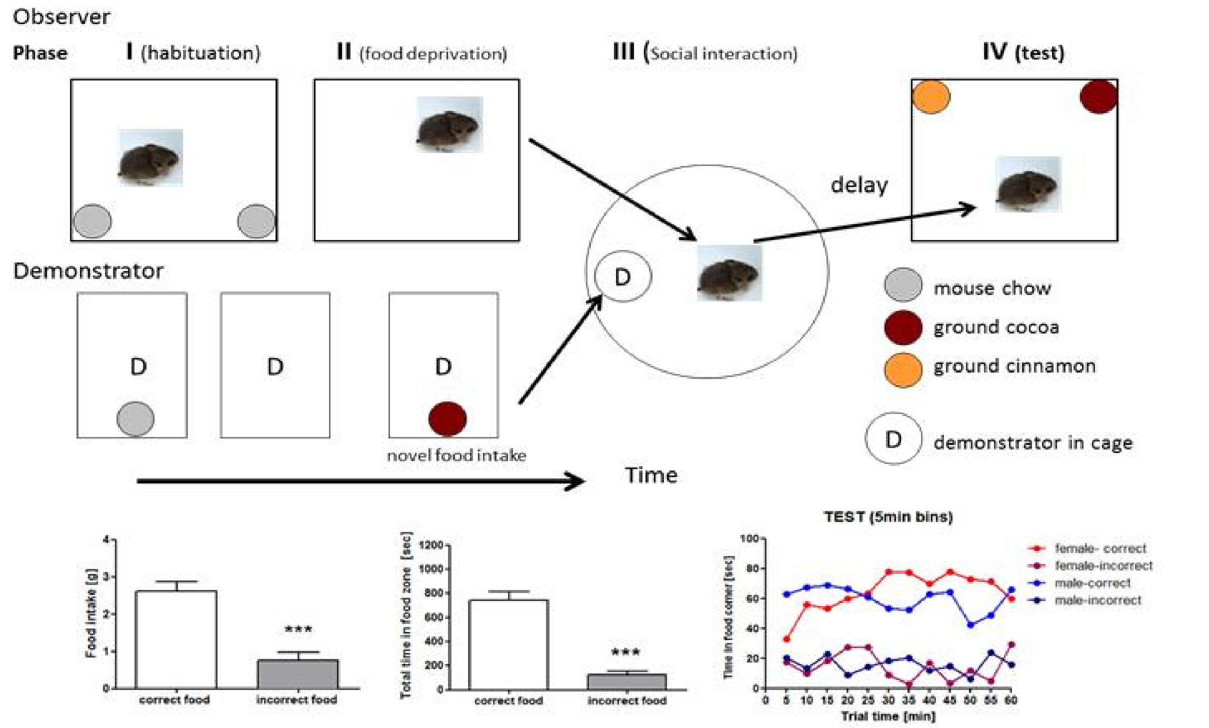


Figure 1. *Semantic-like knowledge task: protocol and pilot data.* Top: Protocol timeline of *observers* and *demonstrators* undergoing 4 phases each. In common are phase I (habituation) and II (food withdrawal over night), respectively. *Demonstrators* (D) then experience a novel food and are paired with observer mice in a novel environment. Only observers are tested after a delay for flavour preference (test) between for example cocoa (+) and cinnamon (-). All phases of observer are recorded; this enables the selective determination of brain activation (WP3 and 5) at any given time during testing. Bottom: Initial data confirming flavour bias at test (left) and corresponding times spent in correct and incorrect food zones (middle) as well as the time course of this response throughout a 1 hour recording period in both genders of C57Bl/6 mice (right). Data are group means (+SEM) and asterisks denote reliable differences (P<0.0001).

Procedure (see Figure 1)

The Observer (N=24) animals were placed in the PhenoTypers and given 3 days of habituation (96-h recording) to the novel environment, with free access to food pellets and water. Once mice were fully habituated to the PhenoTyper, the STFP task was employed. At the start of STFP test food hoppers were emptied and replaced with food jars containing ground mouse chow. Mice were exposed to the ground mouse chow for a period of 24 hours in order to allow for habituation and determination of subjects spatial preference for food jar location. Prior to testing animals were food deprived overnight (maximum 16 hours). The following morning the observer animals were exposed to a stranger mouse (demonstrator previously presented with a particular food scent) placed in a social interaction cylinder positioned in the corner of the PhenoTyper cage. The social interaction (SI) phase was recorded for 30 minutes (simultaneous recording of 24 animals in individual cages). The measurement of observer x demonstrator interaction was based on time spent and distance moved of the observer in the 'interactive zone' drawn in the vicinity of the cylinder. After SI, the demonstrator mice were removed from the cages and following the determined delay (3 conditions: a) 15 min delay; b) 4-h delay; c) 24-h) the observer mice were exposed to two food jars, each containing a different scented food. It is worth to note that the 24-delay-test was performed on the subjects from condition a) and b) with no food deprivation prior to test. The jar containing cued food was placed in the least preferred corner for each subject, in order to avoid experimental bias. The activity of the subjects during food preference test was recorded for 30 minutes (or over night in the 24hr delay). Time spent in the food jars zones (squares 6cm x 6cm) recorded by EthoVision as well as the amount of food eaten from each jar comprise multiple measures of food preference. A longitudinal cross-over design was employed with the experiment performed over a period of 2 weeks using different delays between demonstrator interaction and food exposure and food scents (cocoa, cinnamon, nutmeg and coffee).

Statistical analyses

Group mean comparisons were performed with Prism software (GraphPad Software Inc., San Diego, USA) using parametric or non-parametric analysis of variance (one way or two way ANOVAs as appropriate) followed by suitable post-tests for comparison of selected data pairs. Normalized (%) amount of correct food eaten (in grams, with precision to 2 decimal places) served as a main index of food preference. Further behavioural activity data recorded and extracted by EthoVision software, in 1 hourly bins (for 96-h habituation recording) or 5 minute bins (SI and test) and the following parameters were analysed: time spent, distance moved and frequency of occurrence in social interaction zone, jar vicinity zones (square), water zone; overall distance moved, velocity and immobility. P's< 0.05 were considered reliable.

Results

C57BL/6 mice displayed a strong preference for the cued food under all conditions (15-min delay 70.5% preference; 4h delay 65.1% preference; and 24-h delay 69.7% preference; *p's<0.05 above the level of chance), with no effect of gender (p= ns). Further analysis revealed significantly more time spent and increased frequency in correct compared with incorrect food jar zones (15-min and 4-h delay groups: ***p<0.001; 24-h delay group **p<0.01). Latency to correct food jar reached significance in the 15-min delay group only (*p<0.05) and was much shorter for the correct scent. Moreover, food intake correlated significantly with time (%) (15-min delay- ** r=.42; 4h delay- ***r=.72; and 24-h delay- *r=.37) and frequency in jar vicinity zones (15-min delay- ** r=.45; 4h delay- ***r=.95; and 24-h delay- *r=.40). Further correlation analysis of demonstrators and observers food intake was found positive for all conditions.

Conclusions/Implications

Our data confirm that C57BL/6 mice display social and contextual learning, as indicated by intact preference for the cued food, and concurrently prove the ability of these animals to maintain the olfactory memory for a period of 24 hours. Additionally, the introduction of time delays between the acquisition (SI) and retrieval (test) seems to broaden the scope of STFP and allows for assessment of memory at different time intervals. More importantly however, the additional behavioural measures designed in the system (PhenoTyper and the tracking software) employed in this study facilitated the analysis and quantification of STFP-related behaviours, which until now have been restricted to manual recording in other laboratories. The combination of STFP with automated video-observation might be a step forward in measuring behaviours related to olfaction and memory in mice.

References

1. Galef, B.G.Jr., Wigmore, S.W., Kennett, D.J. (1983). A failure to find socially mediated taste-aversion learning in Norway rats (R. norvegicus). *J. Comp. Psychol.* **97**(4), 358-363.

2. Kogan, J.H., Frankland, P.W., Blendy, J.A. (1997). Spaced training induces normal long-term memory in CREB mutant mice. *Curr. Biol.* **7**(1), 1-11.

3. Alvarez, P., Wendelken, L., Eichenbaum, H. (2002). Hippocampal formation lesions impair performance in an odor-odor association task independently of spatial context. *Neurobiol. Learn. Mem.* **78**, 470-476.

4. Gal-Ben-Ari, S., Rosenblum, K. (2011). Molecular mechanisms underlying memory consolidation of taste information in the cortex. *Front. Behav. Neurosci.* **5**, 87.

5. Mesholam, R.I., Mobergm, P.J., Mahrm, R.N., Doty, R.L. (1998). Olfaction in neurodegenerative disease: a meta-analysis of olfactory functioning in Alzheimer's and Parkinson's diseases. *Arch. Neurol.* **55**, 84-90.

Multimodal Sensing System to Enhance the Safety of Infants in the Home Environment

K. de Groot[1], S. Boughorbel[1], N. Ambroise[2], M. Buter[3], J. Kang[4], B. Loke[5], L. Spreeuwers[6] , J. Vandenabeele[7]

[1]*Philips Research Laboratories, Eindhoven, The Netherlands. koen.de.groot@philips.com, sabri.boughorbel@philips.com*

[2]*Actimage, Luxembourg, Luxembourg (Grand-Duchy). nicolas.ambroise@actimage.com*

[3]*Sound Intelligence, Amersfoort, The Netherlands. mbuter@clb.nl*

[4]*TP Vision Netherlands, Eindhoven, The Netherlands. jeffrey.kang@philips.com*

[5]*Noldus Information Technology bv, Wageningen, The Netherlands. b.loke@noldus.nl*

[6]*Department of Signals and Systems, University of Twente, Enschede, The Netherlands. l.j.spreeuwers@utwente.nl*

[7]*Eagle Vision Systems, Naarden, The Netherlands. JVandenabeele@eaglevision.nl*

Abstract

We propose a system which consists of a multimodal sensing-, reasoning- and actuation-stage to monitor the infant's behavior and sleeping environment, particularly to improve the infant's safety. Moreover, the system advices parents on the child's development. The multi-parametric monitoring system can detect physiological and behavior parameters derived from the infant. Furthermore, people visiting the infant's sleeping environment are identified upon entrance and tracked during their stay in the room. The sensors systems include video, pressure and audio based technologies. High level sensor events are extracted from the raw sensor data, fused and fed into a reasoning system. Depending on the decision outcome, relevant in-home services are automatically actuated to provide feedback to the parents or other caregivers on the behavior of the infant.

Keywords. Multimodal sensing systems, contextual decision making, context awareness, video tracking, infant safety.

Introduction

Many young parents are concerned about the wellbeing of their sleeping child. The presented architecture addresses parent's desire for support in supervision when the infant is unattended in the bedroom. SIDS (Sudden Infant Death Syndrome) is a well-documented fear of many parents and it is the leading cause of death among infants below the age of 12 months [1,2]. Also the comfort of the sleeping child is of course important.

The presented architecture is an ambient intelligent system in the home environment, with the main goal to provide peace of mind to the parents when there is no direct parental supervision. In case of an unsafe situation in the infant's sleeping environment or infant's distress, the system signals the parents. Depending on the significance of the alarm, parents in turn could respond by giving the infant some attention. Moreover, the system includes methods to give the parents insight on the child's development.

The system is developed within the ITEA2 GUARANTEE (Guardian Angel for the Extended Home Environment) project. The project provides technical solutions for personal safety in the home environment. GUARANTEE introduces local and network-supported decision making for safety applications on the basis of sensor input and with immediate response and feedback to the people concerned.

Method

As starting point of the system design, use case scenarios have been made which define the scope and application context and underline the benefits of the system from a safety enhancing point of view. The use case scenarios were mapped on a high level system architecture that consists of three main stages, i.e. sensing, decision making and actuation. The system takes into account design aspects proposed for Ambient Intelligence (AmI) environments [3]. Each use case scenario contains a set of distinct situations that cause a condition change which may influence a sub-process due to user interaction or situation change in the observed area. For each stage, requirements have been specified which evolved into the architecture illustrated in Figure 1. The system transfers different types of information, i.e. events which could be either safety events (raised by a critical or dangerous situation) or standard events (raised by subsystems or nodes for further processing). A second form of information is data flows which could be either raw measured values or audio/video streams.

Sensing

The sensing stage extracts high level events from the raw data acquired by the sensor platform installed in the infant‘s bedroom. The various sensors comprise vision, audio and pressure based modalities: *(i) Side-view face recognition*: Cameras mounted in the doorpost of the infant’s sleeping environment identify individuals visiting the room. Besides identification of an individual, the walking direction is detected to determine if a person enters or leaves the room. Based on the authentication result, the parent may decide if an individual is allowed to be with the infant. The analysis algorithm comprises several processing steps to acquire images suitable for feature extraction. The system uses binary local patterns extracted from gray level images as features for authentication [4]. *(ii) Zone detection:* A stereo vision system installed in the ceiling of the baby room tracks the position of individuals during their stay. In order to compute the tracking coordinates, the stereo cameras first need to be cross-calibrated. Once the cameras are aligned, pixels that are part of the same object are matched across multiple views to build a disparity map [5]. From this information, coordinates of the visible objects are computed which are subsequently fed to a system which performs real-time zone detection, i.e. it detects if this person enters or leaves a predefined zone. In these cases the system generates a zone enter or zone leave event, respectively. The zone events are used in the reasoning stage. The system could for instance detect whether a second child enters the infant’s sleeping environment without adult supervision which may provoke an insecure situation or may bother the infant while sleeping. *(iii) Infant’s emotion analysis:* This module aims to extract the infant’s emotion from an auditory scene. The sound analysis system consists of three main stages. Firstly, audio is captured from a microphone. Next a time frequency analysis is performed which incorporates information from a physical model of the human inner ear (cochlea). Several algorithms extract cues from the cochleogram. These cues are further analyzed to distinguish baby sounds (e.g. crying) from the audio stream. The signaled audio events are interpreted by the decision making process. *(iv) Posture and restlessness detection:* The activity of infants and more specifically the posture of infants is an important aspect in their safety and development. Several independent risk factors for SIDS have been identified, and a combination of preexisting conditions and

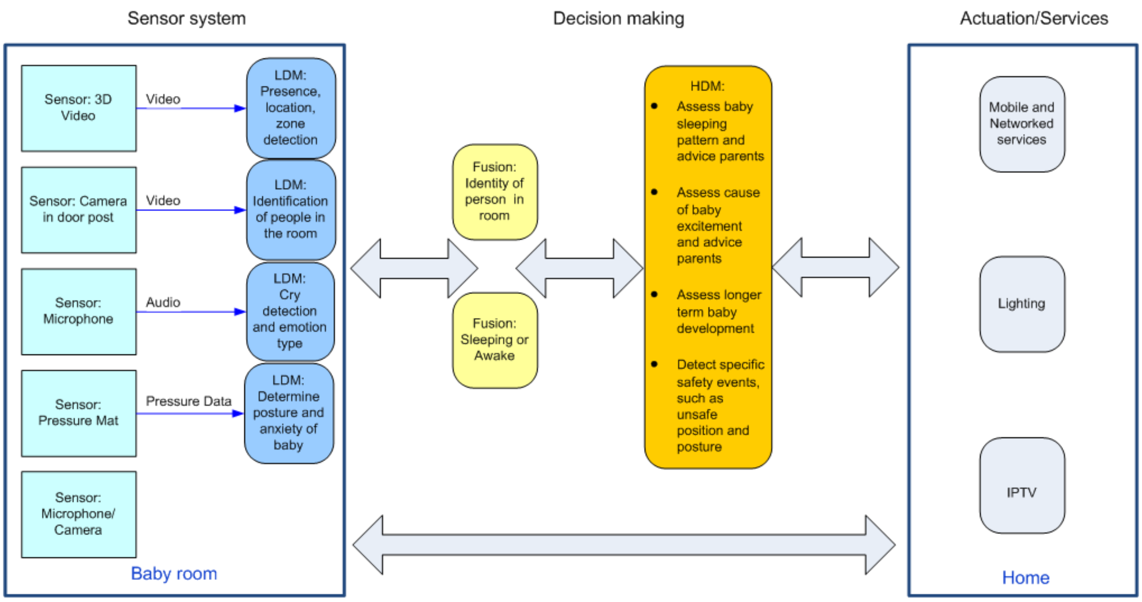


Figure 1. High level infant monitoring architecture consisting of three stages: sensing, reasoning and actuation.

initiating events may lead to SIDS. In view of safety, the supine sleeping position is widely regarded as advised sleeping position for a infant opposed to the prone position which is considered as an individual SIDS risk factor [1][6]. In order to enhance the safety of the infant, we propose a pressure sensitive mat which is positioned under the mattress in the infant's bed, to determine different postures and restlessness levels of the infant when present in the bed [7]. Postures that could be detected by the pressure mat module include: sitting, standing, supine- and prone-position. The infant's activity level is determined by analyzing the variation in applied pressure over time. The observed behavior of the infant is transferred to the decision engine.

Decision making

The sensor events generated by the different modalities are aggregated, fused and translated into safety events. The actual interpretation of the safety events is done in the second stage by means of an autonomous decision making process. The process is event driven and depends on the current received sensor event message, current state of the process and previously logged data. The decision making process models the current behavior of the infant and other occupants in the infant's room and updates the information based on the received sensor events. A received sensor event comprises typically several attributes, e.g. parameter values, priority and timestamp. High level sensor information, relevant for storage, is logged in memory. Different situation dependent services, defined in the third stage are actuated by the decision making module.

Actuation

The actuators and services provide insight on the current situation in the baby room. Most of the baby monitoring systems rely on audio signaling to notify the restless of infants or inactivity (no-breathing). Such signaling can be achieved by transmitting baby sound or audio warnings. Such audio notifications are in some situations not adequate and alternative 'silent mode' notifications are desired. Moreover, modern video baby monitors add a constant video stream to provide visual information about the infant's state. In the proposed architecture, the in-home services are three fold and are excited by the decision engine: *(i) Smartphone*: The system uses a smartphone application to visualize the status in the infant's room, in particular the child's emotion and posture, and the schedule of entrance/exit of individuals visiting the room. According to the type of event that is detected by sensors, the application displays an alert. *(ii) IPTV*: An IP-based baby camera attached to the baby crib is proposed that can send its video stream to different connected display devices in the home, e.g. a TV or smartphone. In case the behavior of the infant desires visual feedback to the parent, a notification service is pushed by the decision engine which overrides the TV image or smartphone current view. *(iii) Lighting*: The infant's activity level or emotional state is associated to a certain color within the light spectrum and is represented by a color LED lamp. Such lighting service could be integrated with the home light atmosphere and provide the parents a way to remotely keep an eye on their child. As an example, blue would indicate that the infant is calm and red would indicate that the child is very agitated. Different levels of restlessness are represented by color grades between these two colors. The level of restlessness is determined by available sensor nodes such as pressure sensor mat, microphones and cameras.

Communication

In order to exchange safety events and data flows, the Universal Plug and Play (UPnP) concept was selected as preferred communication mechanism. UPnP offers seamless device-to-device communication and allows automated device discovery and configuration. A UPnP device can be any entity on the network that implements the protocols required by the UPnP architecture. Devices in a UPnP network announce their presence and defined services to control points present in the network. A control point is an entity in the network that works with the services provided by a device [8]. In GUARANTEE, the different sensor nodes are modeled as UPnP devices, whereas decision making and actuator entities are modeled as UPnP control points.

Conclusions and future work

We present the architecture of an automated system to enhance the safety of infants in the home environment. Behavior analysis and decision making are the key technologies in the proposed architecture. The behavior of the infant and other individuals, e.g. children and adults which are not necessary directly interact with the infant, is monitored by means of a multimodal sensing platform. In order to minimize the false alarm rate, sensor fusion is performed to obtain a robust estimation of the infant's emotion and identity of individuals approaching the child. Based on the high level safety events extracted from the raw measurement data, reasoning is performed to define the situation context. The reasoning process takes into account aspects such as: infant's sleeping position, sleeping status, anxious and type of anxiety and pattern thereof. The output of the decision-making process offers to the end-user good situational awareness of the infant's behavior and current situation in its room which are potentially insecure. The proposed system mainly focuses on observing and mining the behavior of the infant and other occupants in the infant's room. No direct interaction is performed between the infant and the system; therefore it primarily enhances the role of the parent or caregiver as opposed to other concepts like robot nannies which ultimately functions as a child's caregiver [9]. Although many commonalities exist between our proposed system and robot nannies, especially from a sensory, context awareness and reasoning point of view, the later approach poses significant ethical issues (e.g. emotional attachment and connection with the caregiver) as a robot would mainly substitute the primary caregiver instead of improving its role [10].

In a domestic AmI environment, respecting the privacy of individuals interacting is an important aspect. During the design process of the system, a proper balance between the benefits of an automatic monitoring and invasion of privacy was taken into account. Privacy of individuals is preserved as the reasoning is based on high level sensor events which do not contain personal data. For future work we planned to verify the proposed system from a user perspective. This could be for instance by exposing groups of users to test usage (parts of) the system.

Acknowledgements

This work is supported by the GUARANTEE (ITEA 2) 08018 project.

References

1. Task Force on Sudden Infant Death Syndrome (2005). American Academy of Pediatrics: The Changing Concept of Sudden Infant Death Syndrome: Diagnostic Coding Shifts, Controversies Regarding the Sleeping Environment, and New Variables to Consider in Reducing Risk. *Pediatrics* **116**, 1245-1255.

2. Elizabeth, A., MacDorman, M.F., Donna, M.S, Bernard, G. (2002). Annual Summary of Vital Statistics. *Pediatrics* **112**, 1215-1230.

3. Nakashima, H., Aghajan, H., Augusto, J.C. (2010) *Handbook of Ambient Intelligence and Smart Environments*, Springer.

4. Santemiz, P., Spreeuwers, L.J., Veldhuis, R.N.J. (2011). Side-View Face Recognition. *32nd Symposium on Information Theory in the Benelux*, 305-312.

5. Hirschmuller, H. (2008). Stereo Processing by Semiglobal Matching and Mutual Information, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**, 328-341.

6. Mitchell, M., Hutchison, L., Stewart, A. (2007). The continuing decline in SIDS mortality. *Archives of disease in childhood* **92**, 625-626.

7. Boughorbel, S., Bruekers, F., Breebaart, J. (2010). Baby-posture classification from pressure-sensor data. *Proc. 20th international conference on pattern recognition (ICPR)*, 556 - 559.

8. Jeronimo, M., Weast, J. (2003). *UPnP Design by Example*. Intel Press.

9. Van den Broek, E. L. (2010), Robot nannies: Future or fiction? *Interaction Studies* **11**, 274-282.

10. Sharkey, N., Sharkey, A. (2010). The crying shame of robot nannies: An ethical appraisal. *Interaction Studies* **11**, 161-190.

Game Based Physiotherapy for Treatment of Children with Juvenile Idiopathic Scoliosis

P. Feistritzer-Gröbl[1], A. Nischelwitzer[2], V. Saraph[3], H. Holzer[4]

[1]*Department of Physiotherapy, FH Joanneum, Graz, Austria.* [1]*petra.feistritzer-groebl@fh-joanneum.at*

[2]*Department of Information Management, FH Joanneum, Graz, Austria.* [2]*alexander.nischelwitzer@fh-joanneum.at*

[3]*Department of Adolescent Surgery, Medical University, Graz, Austria.* [3]*vinay.saraph@meduni-graz.at*

[4]*Institute of Sports science, KF University, Graz, Austria.* [4]*hans-peter.holzer@uni-graz.at*

Introduction

The performance of daily exercises of patients with “juvenile idiopathic scoliosis” (JIS) during the home-based exercise period is difficult to monitor, therefore the collection of evidence is a challenge. Although a correct execution of these exercises are of particular importance for the treatment of *JIS* there is, until now, no easy tool to monitor and guide these home-exercises [1]. Furthermore, because such exercises have to be performed many times in the same way, they tend to be boring. This may lead to a negligent execution of the home-based exercises, consequently reducing their therapeutic effect.

The aim of this study was to develop a computer game to guide, monitor and evaluate therapeutic exercises along with a way to strengthen the motivation of children with *JIS*.

Methods

Software and interfaces of a 3D input device were used to detect and measure smallest movements within a predefined, body fixed, 3D space. As input device a Sony© Gametrak for Playstation II© with two strings were used. One string was fixed on the body (near the COG) to get the “stabilisation point” and the movement of the trunk, the other string was fixed on leg or arm depending on the exercise used (in figure 2, the string was fixed at the ankle). To maintain motivation three different game screens have been developed (see Figure 1).

To play one of these games you have to move your arm or leg within the predefined *Movement Space* and get points for reaching special targets. When playing the game the exact movements of the extremities in respect to a body fixed “stabilisation point” can be measured along with additional parameters like exercise time or time outside the *Movement Space (MS).*

During the performance of an exercise the game offers feedback information to keep the right position: should one part of the body (arm, leg) go beyond the *Movement Space*, the game stops and indicates the position of the correct movement. As soon as the arm or leg is once again within the *Movement Space* the game continues.

Figure 1. Game interfaces (Bee, Helicopter and Balloon).

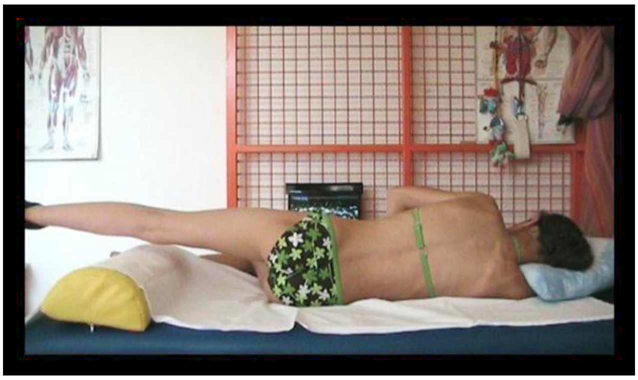


Figure 2. Standardized exercise of Schroth concept *MUSCLECYLINDER IN SIDEPOSITION*.

In this way the children are guided to perform their exercises in a precise way. Additionally, more data are recorded for further evaluation by the therapist. Furthermore, the level of difficulty can be adjusted by the therapist to the individual skill of the child.

A single case study was performed with a 12 year old girl with *JIS*, which exercised six months according to the Schroth concept [2]; two standardized exercises of this concept (*MUSCLECYLINDER IN SIDEPOSITION*, *MUSCLECYLINDER IN STANCE*) were carried out guided by the game (see Figure 2).

Results & Discussion

The available computer game for patients with *JIS* seems suitable for evaluating therapeutic exercises. The collected quantitative data give valuable information on the exercise regime and can be used to monitor and evaluate treatment process. Due to the real time Feedback of precise performance of exercises, exercise times are used more efficiently.

In this single case Study a Cobb curve (measured in brace) of 22 deg. (thoracic curve left convex) and 26 deg. (lumbar curve right convex) could be reduced to 16 deg. respectively 20deg.

A randomized study with the same target group is currently carried out to evaluate possibly different results of treatment with or without computer game.

Ethical statement

The experiment was approved by the ethical committee of the Medical University in Graz, Austria. Date: 18.03.2011, EK-Number:23-235 ex 10/11.

References

1. Negrini, S., Zaina, F., Romano, M., Negrini, A., Parzini, S. (2008). Spezific exercises reduce brace prescription in adolescent idiopathic scoliosis: a prospective controlled cohort study with worst-case analysis. *Rehabil Med* **40**, 451-455.

2. Lehnert-Schroth, C. (2007). *Dreidimensionale Skoliosebehandlung: Atmungs- Orthopädie System Schroth* (7th ed.). München, Urban & Fischer.

Effects of Age, Sex and Anxiety on Spatial Learning and Memory in Rats

M. Belviranlı[1], K.E. Nurullahoğlu Atalık[2], N. Okudan[1], H. Gökbel[1]

[1]*Department of Physiology Faculty of Medicine Selcuk University, Konya, Turkey.* [1]*mbelviranli@yahoo.com*

[2]*Department of Pharmacology Meram Faculty of Medicine Konya University, Konya, Turkey*

Abstract

The main objective of study was to investigate the effects of age, sex and anxiety differences on learning and memory in rats. We also investigate whether repeated elevated plus maze (EPM) tests induce anxiety in rats. Eighty Wistar rats were divided into the eight groups according to their sex, age and anxiety status. Spatial learning and memory were evaluated with Morris Water Maze (MWM). All behavioral tests were recorded online and analyzed offline with analytical software. In conclusion, age, sex and anxiety affect learning and memory in different aspects.

Keywords. Behavioral measurement; anxiety; aging; sex difference.

Introduction

Aging is a physiological process which adversely affects many vital functions, including cognitive functions [1]. Aging induced cognitive dysfunctions are well documented in humans [2,3] and several animal species including rats and mice [4]. Behavioral sex differences are observed in many examples of anxiety and memory tests. These differences depend on the sex hormones and other mediators in the brain [5]. However, there is no consensus about the effects of sex differences on spatial learning and memory and possible effects of aging and anxiety. In the present study we aimed to investigate the effect of aging and sex differences on learning and memory in anxiety induced rats.

Materials and methods

Animals

The study protocol was approved by the Ethics Committee of the Selcuk University Experimental Medicine Research and Application Center. A total of 80 adult male (n: 40) and female (n: 40) Wistar rats were obtained from the same center. Animals were divided into eight groups (ten animals in each group) as follows:

- Group 1: 7-month-old control young males (CYM).
- Group 2: 7-month-old control young females (CYF).
- Group 3: 15-month-old control aged males (CAM).
- Group 4: 15-month-old control aged females (CAF).
- Group 5: 7-month-old experimental young males (EYM).
- Group 6: 7-month-old experimental young females (EYF).
- Group 7: 15-month-old experimental aged males (EAM).
- Group 8: 15-month-old experimental aged females (EAF).

The principle of laboratory animal care of the National Institute of Health (NIH) guideline was followed in all these experiments.

Behavioral Tests

All behavioral test devices were placed in an isolated room and test protocols were performed here to eliminate the factors that affect animal's behavior. MWM was carried out followed by EPM (5 min) for four consecutive days. During the training period, four starting points were randomized in everyday to avoid the horizontal navigation to the platform. Each trial was terminated in 90 s, if the rats were unable to locate platform in this time interval and then rats were placed on the platform for 30 s. Following the four days of training period, 120 s of probe trial was conducted after the hidden platform was removed. Control groups were performed same protocols with the experimental groups, except the anxiety protocols.

Tracking System

All tests were recorded online using a video camera. The videos of the behavioral tests were analyzed offline using analytical software (Noldus Information Technology, Ethovision XT 8.0, Wageningen, The Netherlands).

Statistical Analysis

All data are presented as mean ± SD. Changes of the MWM variables through the experiment were analyzed using three-way analysis of variance (ANOVA) with repeated measures. EPM values were analyzed using two-way ANOVA with repeated measures. Probe trial data were analyzed by three-way ANOVA. A p value less than 0.05 was considered to be statistically significant.

Results and discussion

During the repeated EPM tests, total distance traveled, numbers of entries in center, open arms and closed arms significantly affected by time interaction (F=22.907, P=0.000; F=6.093, P=0.001; F=4.033, P=0.009 and F=7.157, P=0.000, respectively). These variables were decreased by repeated measurements in all groups.

The latency to find platform and total distance traveled significantly affected by time interaction (see Table 1). In all groups latency to reach platform and total distance traveled significantly decreased with time. Average swimming speed was affected by both time and time x age interactions (see Table 1). Average swimming speed was decreased in all groups with time and young rats were faster than aged rats.

In probe trial, total distance traveled and average swimming speed affected by anxiety and age factors (see Table 2). Total distance and swimming speed were higher in experimental and aged groups compared to control and young groups, respectively. Although number of platform crossing was not affected by anxiety, age and sex factors, time spent in platform was significantly affected by age (see Table 2). Time spent in platform zone was higher in young groups than aged groups. During the probe trial, the latency to reach the platform was not different among the groups.

Table 1. Morris water maze. Summary of the three-way repeated measures ANOVA results in all groups.

	Time		Time x Anxiety		Time x Age		Time x Sex	
	F	P	F	P	F	P	F	P
Latency (s)	35.087	0.000	1.514	0.130	0.862	0.608	0.826	0.645
Total distance traveled (cm)	35.438	0.000	1.466	0.149	1.127	0.354	0.966	0.501
Average swimming speed (cm/s)	16.909	0.000	1.486	0.141	2.278	0.013	1.153	0.334

Table 2. Morris water maze probe trial. Summary of the two-way ANOVA results in all groups.

	Anxiety		Age		Sex	
	F	P	F	P	F	P
Distance traveled (cm)	5.476	0.022	11.116	0.001	2.139	0.148
Average swimming speed (cm/s)	5.141	0.026	11.023	0.001	2.246	0.138
Number of platform crossing	2.747	0.102	1.112	0.295	1.112	0.295
Time spent in platform (s)	0.016	0.889	4.997	0.029	1.241	0.269

Conclusions

In conclusion, age, sex and anxiety differentially affect spatial and emotional memory in rats. More detailed researches are needed about the cognitive functions to clarify the exact role of estrogen in sex difference and exact roles of neurotransmitters and oxidative stress in aging process.

References

1. Gottfries, C.G. (1990) Neurochemical aspects on aging and diseases with cognitive impairment. *Journal of Neuroscience Research* **27**, 541-547.

2. Almaguer, W., Estupiñán, B., Uwe Frey, J., Bergado, J.A. (2002). Aging impairs amygdala-hippocampus interactions involved in hippocampal LTP. *Neurobiology of Aging* **23**, 319-324.

3. Wetherell, J.L., Reynolds, C.A., Gatz, M., Pedersen, N.L. (2002). Anxiety, cognitive performance, and cognitive decline in normal aging. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* **57**, 246-255.

4. Bowman, R.E., Maclusky, N.J., Diaz, S.E., Zrull, M.C., Luine, V.N. (2006). Aged rats: Sex differences and responses to chronic stress. *Brain Research* **1126**, 156-166.

5. Simpson, J., Kelly, J.P. (2012). An investigation of whether there are sex differences in certain behavioural and neurochemical parameters in the rat. *Behavioural Brain Research* **229**, 289-300.

Age, Sex and Anxiety Affect Locomotor Activity in Rats

Muaz Belviranlı[1], Kısmet Esra Nurullahoğlu Atalık[2], Nilsel Okudan[1], Hakkı Gökbel[1]

[1]*Department of Physiology Faculty of Medicine Selcuk University, Konya, Turkey. nokudan@selcuk.edu.tr*

[2]*Department of Pharmacology Meram Faculty of Medicine Konya University, Konya, Turkey.*

Abstract

The main objective of study was to investigate the effects of age, sex and anxiety differences on locomotor activity in rats. Locomotor activity was assessed in open field (OF). Repeated anxiety tests were performed in elevated plus maze (EPM). All behavioral tests were recorded online and analyzed offline with analytical software. In conclusion, age, sex and anxiety affect locomotor activity in different aspects.

Keywords. Behavioral measurement; anxiety; aging; sex difference

Introduction

Aging is a physiological process which adversely affects many vital functions, including cognitive functions [1]. It has been reported that there may be many different reasons for this impairment such as impairment in long-lasting, synaptic plasticity processes, like long-term potentiation [2]. Changes in neurotransmitter systems might also reason of the loss of memory [3]. Behavioral sex differences depend on the sex hormones and other mediators in the brain [4]. However, there is no consensus about the effects of sex differences on locomotor activity and possible effects of aging and anxiety. In the present study, we aimed to investigate the effect of aging and sex differences on locomotor activity and exploratory behavior in anxiety induced rats.

Materials and methods

Animals

A total of 80 adult male (n: 40) and female (n: 40) Wistar rats were obtained from Selcuk University Experimental Medicine Research and Application Center (Konya, Turkey). Animals were divided into eight groups (ten animals in each group) as follows:

- Group 1: 7-month-old control young males (CYM, n:10).
- Group 2: 7-month-old control young females (CYF, n:10).
- Group 3: 15-month-old control aged males (CAM, n:10).
- Group 4: 15-month-old control aged females (CAF, n:10).
- Group 5: 7-month-old experimental (anxiety-induced) young males (EYM, n:10).
- Group 6: 7-month-old experimental (anxiety-induced) young females (EYF, n:10).
- Group 7: 15-month-old experimental (anxiety-induced) aged males (EAM, n:10).
- Group 8: 15-month-old experimental (anxiety-induced) aged females (EAF, n:10).

The study protocol was approved by the Ethics Committee of the Selcuk University Experimental Medicine Research and Application Center. The principle of laboratory animal care of the National Institute of Health (NIH) guideline was followed in all these experiments.

Behavioral Tests

On day 1, an evaluation test in OF was performed. On the following days (2-5) EPM was carried out. On day 6, second evaluation test in OF was performed. After each animal's behavioral tests, EPM and OF were cleaned with an alcohol and dried to prevent the olfactory cues.

Tracking System

All tests were recorded online and the video of the behavioral tests were analyzed offline using analytical software (EthoVision XT 8.0, Noldus Information Technology, Wageningen, The Netherlands).

Open Field

Spontaneous locomotor activity of the rats was assessed in OF. The animals individually placed in the centre of the box and allowed to freely explore for 5 min.

Elevated Plus Maze

The EPM consisted of a plus shaped platform with two open and two enclosed arms. Rats were placed individually on the center of the platform facing one of the open arms and were allowed to freely explore the maze for 5-min testing period.

Statistical Analysis

Statistical analyses of data were carried out by using SPSS 15.0 for Windows. All data are presented as mean ± SD. Changes of the OF variables through the experiment were analyzed using three-way analysis of variance (ANOVA) with repeated measures. EPM values were analyzed using two-way ANOVA with repeated measures.

Results and discussion

Open Field

The total distance traveled and average speed were significantly affected by the time and time x sex interactions (see Table 1). They were significantly decreased in second trial compared to first trail in all groups; also total distance traveled and average speed were higher in females than males. Time spent in center and walls were affected by time x sex interaction (see Table 1). Time spent in center increased in males and decreased in females however, time spent in walls decreased in males and increased in females in second trial compared to first trial. Time spent as a mobile and immobile were significantly affected by time and time x sex interactions (Table 1). Time spent as a mobile significantly lower and time spent as an immobile significantly higher in second trial than first trail in all groups. Furthermore, decrease in time spent as a mobile and increase in time spent as an immobile were higher in females compared to males. The number of defecation was significantly affected by time, time x anxiety and time x sex interactions (see Table 1). Changes of number of defecation were higher in experimental groups and males compared to controls and females, respectively. Grooming behavior was significantly affected by time, time x anxiety and time x age interactions (see Table 1). Changes of number of grooming were significantly decreased in second trial compared to first trial in all groups. Grooming behavior was lower in experimental groups and aged rats compared to controls and young rats, respectively. Rearing behavior was significantly affected by time, time x anxiety and time x sex interactions (see Table 1). The number of rearing events was significantly decreased in the second trial compared to first trial for all groups. Incidence of rearing was lower in experimental groups and males compared to control groups and females.

Elevated Plus Maze

Total distance traveled, numbers of entries in center, open arms and closed arms significantly affected by time interaction (see Table 2). These variables were decreased by repeated measurements in all groups.

Table 1. Open Field. Summary of the three-way repeated measures ANOVA results in all groups.

	Time		Time x Anxiety		Time x Age		Time x Sex	
	F	P	F	P	F	P	F	P
Total distance traveled (cm)	170.889	0.000	0.095	0.759	0.243	0.624	11.618	0.001
Average speed (cm/s)	170.855	0.000	0.099	0.754	0.225	0.636	11.617	0.001
Time spent in center (s)	0.128	0.722	0.617	0.435	0.242	0.625	4.863	0.031
Time spent in walls (s)	0.128	0.722	0.617	0.435	0.242	0.625	4.863	0.031
Number of rearing	179.530	0.000	5.691	0.020	3.000	0.088	7.007	0.010
Number of defecation	0.330	0.014	4.308	0.042	0.479	0.491	6.330	0.014
Number of grooming	11.903	0.001	8.561	0.005	4.579	0.036	0.154	0.695
Time spent as a mobile (s)	86.657	0.000	0.669	0.416	0.328	0.569	36.595	0.000
Time spent as an immobile (s)	86.447	0.000	0.605	0.439	0.247	0.621	36.350	0.000

Table 2. Elevated plus maze. Summary of the two-way repeated measures ANOVA results in all groups.

	Time		Time x Age		Time x Sex	
	F	P	F	P	F	P
Total distance traveled (cm)	22.907	0.000	1.321	0.283	1.923	0.130
Number of entries in center	6.093	0.001	0.506	0.732	0.460	0.764
Number of entries in open arms	4.033	0.009	1.659	0.183	0.874	0.490
Number of entries in closed arms	7.157	0.000	0.657	0.626	1.872	0.139
Time spent in center (s)	1.380	0.262	0.161	0.956	0.186	0.944
Time spent in open arms (s)	2.046	0.111	0.427	0.788	0.717	0.587
Time spent in closed arms (s)	0.293	0.881	0.201	0.936	0.322	0.861
Time spent as a mobile (s)	0.887	0.482	0.772	0.551	0.547	0.702
Time spent as an immobile (s)	0.180	0.947	1.106	0.370	0.119	0.975

Conclusions

In conclusion, age, sex and anxiety affect locomotor activity and exploratory behavior. More detailed researches are needed about the cognitive functions to clarify the exact role of estrogen in sex difference and exact roles of neurotransmitters and oxidative stress in aging process.

References

1. Gottfries, C.G. (1990). Neurochemical aspects on aging and diseases with cognitive impairment. *Journal of Neuroscience Research* **27**, 541-547.

2. Bergado, J.A., Almaguer, W., Rojas, Y., Capdevila, V., Frey, J.U. (2011). Spatial and emotional memory in aged rats: A behavioral-statistical analysis. *Neuroscience* **172**, 256-269.

3. McEntee, W.J., Crook, T.H. (1993). Glutamate: its role in learning, memory, and the aging brain. *Psychopharmacology (Berlin)* **111**, 391-401.

4. Simpson, J., Kelly, J.P. (2012). An investigation of whether there are sex differences in certain behavioural and neurochemical parameters in the rat. *Behavioural Brain Research* **229**, 289-300.

Measure of Anxiety and Exploration Behaviours in Cerebellar Mice Using a Simple and Unique Apparatus: Influence of Habituation on Behavioural Disinhibition

P. Hilber, T. Lorivel

Laboratoire de Psychologie et Neurosciences de la Cognition et de l'Affectivité, Université de Rouen, Rouen, France.
pascal.hilber@univ-rouen.fr

Abstract

Stress is one of the prominent factors driving exploration showed by rodents in behavioural tests. Such an influence is exacerbated in mice like the Lurcher cerebellar mutants which are hyper reactive to experimental handling. In order to precisely delineate this influence of stress on explorative behaviours, we designed an innovative apparatus (the I maze) existing under two configurations (aversive and less aversive) and permitting to habituate animals more or less to experimental conditions before being tested. Exemplifying the interest of this device, the results we obtained in the Lurcher mutant mice permitted to differentiate between the roles played by the cerebellum in motivation for exploration proper and stress-related behaviours.

Keywords. anxiety, emotional behaviour, exploration, cerebellum, Lurcher mutant, mice.

Introduction

The cerebellar Lurcher mutant mice (+/Lc) exhibit severe and obvious motor deficits and also cognitive and emotional-related behaviours disturbances [1]. Thus, when tested in the elevated plus maze, these mutants did not avoid the open spaces (the time they spent in the open arms is similar to that measured in the closed arms). As revealed by corticosterone measures, they are also more stressed than their littermate controls when placed in a novel environment (i.e. in a novel cage or tested in the elevated plus maze). For this reason, we proposed that the total degeneration of the cerebellar cortex caused by the Lurcher mutation led to pathological behavioural disinhibition [2]. On the other hand, some authors showed the mutants were less motivated to explore their environment in the hole board test and concluded from this that the cerebellar cortex plays a role in exploration motivation [3]. In the present study we evaluated the impact of experimental manipulation (environmental changes) on anxiety- and spontaneous exploration-related behaviours in the Lurcher mice. For this purpose we used a unique apparatus with two potential configurations which permits in particular to maintain mice in a starting arm for a habituation period.

Material and methods

Apparatus: the I maze (see Figure 1)

All the components of the apparatus were made from black-painted IffaCredo® individual cages 20 x 7 x 8 (length x width x height in cm). It consisted of a starting closed alley facing either an open alley (open configuration) or another similar closed alley (closed configuration). The starting and testing alleys were separated by a guillotine door. The whole apparatus was elevated to a height of 50 cm above floor level.

Animals

Adult male Lurcher mutant and control mice from the same strain (B6CBA) were tested in the open and the closed configuration, with or without a habituation period (H or NH; see Table 1). We also verify the validity of the test (open versus closed configuration) with mice from other strains (BALB/C and C57BL/6) as well as its sensitivity to a pharmacological agent (the benzodiazepine chlordiazepoxide, CDP, 5 mg/kg).

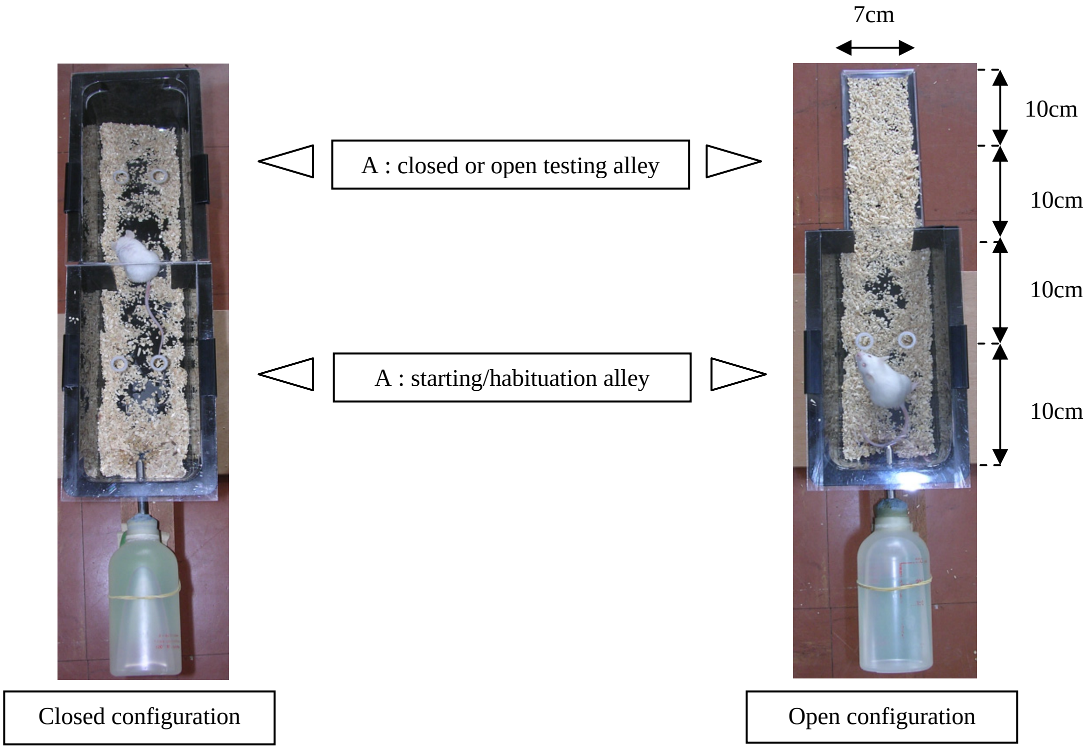


Figure 1. The I maze under its two configurations.

Table 1. Groups and protocols.

Genotype	Treatment	Closed configuration	Open configuration
C57BL/6	NaCl, NH	N=10	N=9
BALB/C	NaCl, NH	N=10	N=9
BALB/C	Chlordiazepoxide: 5mg/kg, NH	N=10	N=10
Lurcher	NH	N=10	N=12
Lurcher	H	N=9	N=11

Procedures

All the mice tested were randomly subjected to the I maze in either its open or closed configuration. In the classical procedure (no habituation), at the beginning of the test each mouse was placed in the starting closed alley and its behaviours were then measured during 5 minutes with the free EthoLog® Software. Additional Lurcher mice and littermate controls were submitted to the habituation procedure in which each mouse was placed during 20 minutes in the starting closed alley with water available (as it can be seen in Figure 1) before the opening of the separation door and the assessment of behaviours for 5 minutes. For each condition, we measured the locomotor activity, the time spent and the number of entries into the different areas, the number of SAPs and rears. The floor of the apparatus (starting and testing alleys) was covered with sawdust.

Ethical statement

All the experiments reported here were conducted in accordance with the ethical recommendations of the Directive 2010/63/EU of the European Parliament and of the Council for the protection of animals used for scientific purposes.

Statistical Analysis

Data were analysed with Statistica® software. Two ways ANOVAs with genotype (or drug) and configuration as factors were performed, followed by LSD tests for post hoc comparisons when allowed.

Results and discussion

Consistently with the data obtained in the elevated plus maze, we showed that all the mice we tested (B6CBA, BALB/C, C57BL/6 mice) in the open configuration of the I maze spent less time in the open testing alley than in the closed starting alley (see Figure 2). Such avoidance behaviour was reduced in mice injected with benzodiazepine. On the contrary, mice whether administered with the anxiolytic or not did not show any place preference in the closed configuration of the apparatus.

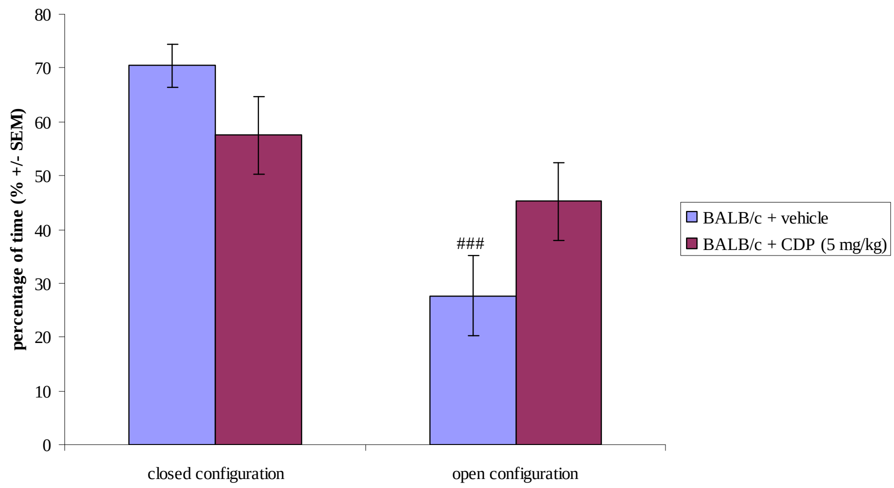


Figure 2. Percentage of time spent into the testing arm under the closed and open configuration by BALB/c mice injected or not with chlordiazepoxide (### = closed vs open configuration in BALB/c + vehicle, $p<0.001$).

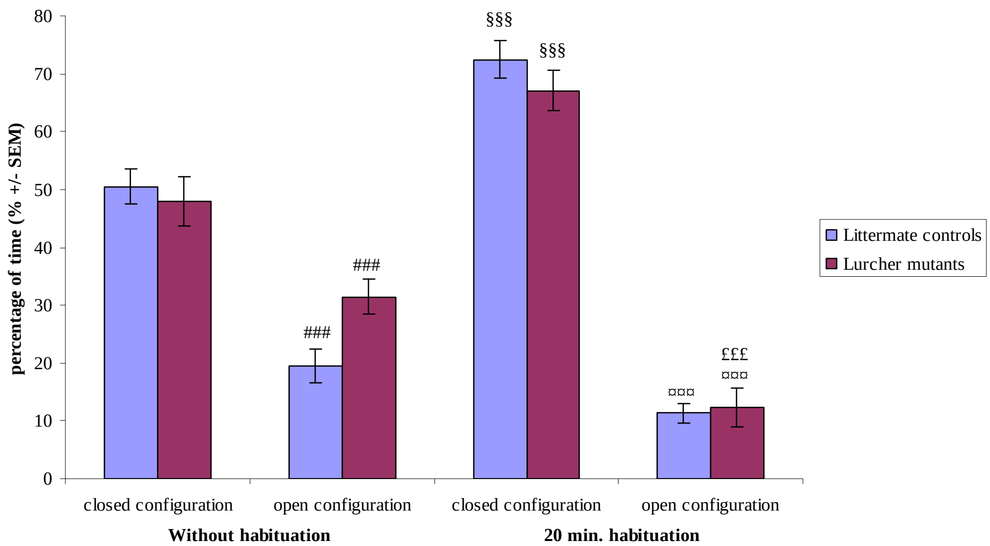


Figure 3. Percentage of time spent in the testing arm by Lurcher mutants and littermate controls without prior habituation or after a 20 min. habituation period (### = open vs closed configuration without habituation, $p<0.001$; §§§ = closed configuration without habituation vs closed configuration after 20 min. of habituation, $p<0.001$; ¤¤¤ = open vs closed configuration after 20 min. of habituation, $p<0.001$; £££ = open configuration without habituation vs open configuration after 20 min. of habituation, $p<0.001$).

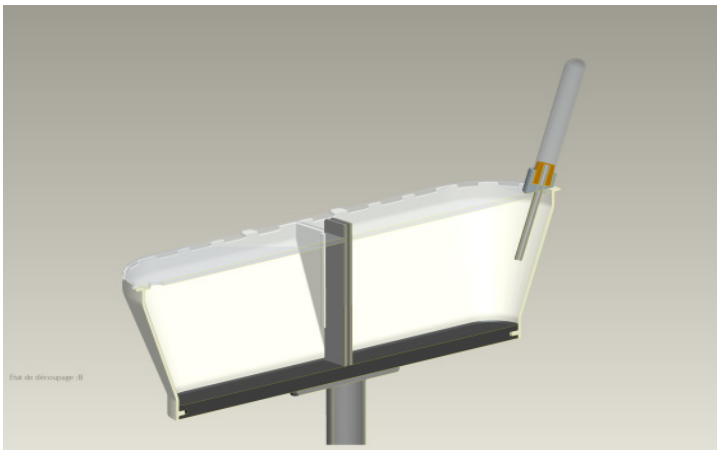


Figure 4. Longitudinal section of the I maze (closed configuration)

As previously noticed in the elevated plus maze and in accordance with the disinhibition hypothesis, we observed that compared to their littermate controls , the cerebellar Lurcher mice did not avoid significantly the testing alley when placed in the open configuration of the I maze (see Figure 3). As expected, we showed that 20 minutes habituation to experimental design were sufficient to abolish the stress-related behavioural disinhibition of the mutants. In contrast with precedent results we did not observe reduced exploration motivation in the Lurcher mice when tested in the closed configuration of our device.

Conclusion and perspectives

Our results clearly suggested that cerebellar damages potentially altered behavioural expression of anxiety in stressful conditions but changed neither anxiety-related behaviours in neutral conditions nor spontaneous exploration motivation in mice.

The novel I maze prototype is now developed (in collaboration with Intellibio® and Novall® see Figure 4) and new protocols are in process to investigate the influence of habituation and drug treatments in both spontaneous and anxiety-related exploration behaviours in mice. The results are compared to those obtained with the elevated + maze. With this novel apparatus behaviours are videotracked during 5 minutes (ANYMAZE®, Stoelting).

Acknowledgments

Special thanks to the students C. Philippot, B. Quignon and C. Campart for their contribution to this work.

Grant: Technology Transfer Funds.

References

1. Vogel, M. W., Caston, J., Yuzaki, M., Mariani, J. (2007). The Lurcher mouse: Fresh insights from an old mutant. *Brain Research* 1140, 4-18.

2. Hilber, P., Lorivel, T., Delarue, C., Caston, J. (2004). Stress and anxious-related behaviors in Lurcher mutant mice. *Brain research*, **1003**(1-2), 108-112.

3. Caston, J, Chianale, C., Delhaye-Bouchaud, N., Mariani, J. (1998). Role of the cerebellum in exploration behavior. *Brain research*, **808**(2), 232-237.

Workshops

WORKSHOP

Measuring Behaviour in Open Spaces

Hayley Hung[1], Ronald Poppe[2], Janienke Sturm[3]

[1]*University of Amsterdam., Amsterdam, The Netherlands. hayleyhung@gmail.com*

[2]*University of Twente, Enschede, The Netherlands. R.W.Poppe@utwente.nl*

[3]*Eindhoven University of Technology / Fontys University of Applied Sciences, Eindhoven, The Netherlands. j.sturm@tue.nl*

Open spaces such as lobbies, playgrounds, museums and streets are common locations where people meet and interact with the environment. For various reasons, we might be interested in the behaviour of individuals and groups. Detection of suspicious behaviour or traffic control are typical applications in the areas of security and surveillance. But we can also analyse how people navigate through crowded places, whether children play co-operative or competitive on a playground or whether people are attending items on sale in a clothing store.

When the behaviour of people is analysed, we can also try to influence this behaviour as an interactive system. When displays or actuators are present in the open space, these can be used to navigate people through crowded halls or try to stimulate interactions between people that appear to have similar interests in a museum. An interactive system can even try to seduce people to explore how the system responds to them. In all these examples, there is a need to robustly identify and track people around the space. Sensors such as cameras, microphones and Kinects are employed and their signals processed to obtain information about the people's locations and movements. Subsequently, we can analyse these to understand the behaviour, either for individuals or for groups. This interpretation is non-trivial as it strongly depends on the spatial and social context in which the behaviour occurs. Moreover, unexpected behaviour and inconsistencies in the tracking make this task a challenging one. Besides the challenges in proper sensor technology and behaviour interpretation, research on measuring behaviour in open spaces also depends on the availability of labelled behaviours for the training and evaluation of various aspects of the system. This workshop addresses the challenges and opportunities in the analysis and understanding of (groups of) people in open spaces.

Speakers at the workshop include Marco Cristani (University of Verona), Liesbeth Jans (Fontys Sporthogeschool Tilburg), Andries Lohmeijer (KITT Engineering), Ben Schouten (Eindhoven University of Technology / Fontys University of Applied Sciences) and Mettina Veenstra (Amsterdam University of Applied Sciences).

WORKSHOP CONTENTS

Creating Interactive Public Spaces

Mettina Veenstra (Amsterdam University of Applied Sciences)

Human behavior analysis for the design of playful interactions

Ben Schouten (Eindhoven University of Technology / Fontys University of Applied Sciences)

Social Computer Vision for Group Behavior Analysis

Marco Cristani (University of Verona)

25 years Motion Interpreted Media Interface Control (MIMIC)

Andries Lohmeijer (KITT Engineering)

Observing physical activity and play in open spaces

Liesbeth Jans (Fontys Sporthogeschool Tilburg)

Social Computer Vision for Group Behavior Analysis

Marco Cristani

Università di Verona, Verona, Italy. marco.cristani@univr.it

Detecting human interactions represents one of the most intriguing frontiers of the automated surveillance since more than a decade. Recently, sociologic and psychological findings have been considered into video surveillance algorithms, especially thanks to the advent of Social Signal Processing work, a recent multi-disciplinary area where computer vision and social sciences converge.

This chapter follows this direction and proposes a detailed overview of the work we are conducting to detect and characterize social interactions, with particular reference to groups. In particular, we will present three scenarios where a group of interacting people is first detected, using positional and orientation features, and then characterized by extracting voice activities from solely visual cues. Finally, we classify existent social relations among the participants.

The first contribution is devoted to detect social interactions using statistical analysis of spatial-orientation arrangements that have a sociological relevance. As social interactions we intend the acts, actions, or practices of two or more people mutually oriented towards each other. In general, any dynamic sequence of social actions between individuals (or groups) that modify their actions and reactions by their interaction partner(s).

We analyze quasi-stationary people in an unconstrained scenario identifying those subjects engaged in a face-to-face interaction, i.e., a scene monitored by a single camera where a variable amount of people (10-20) is present. We import into the analysis the sociological notion of F-formation as defined by Adam Kendon in the late ’70s [1]. Simply speaking, F-formations are spatial patterns maintained during social interactions by two or more people. Quoting Kendon, “*an F-formation arises whenever two or more people sustain a spatial and orientational relationship in which the space between them is one to which they have equal, direct, and exclusive access.*”.

In practice, an F-formation is a set of possible geometrical configurations in space that people may assume while participating in a social interaction There can be different F-formations: in the case of two participants, typical F-formation arrangements are vis-a-vis, L-shape, and side-by-side; when there are more than three participants, a circular formation is typically formed. Our approach aims at automatically detecting these visually significant configurations taking as input a calibrated scenario, in which the position of the people and their head’s orientations have been estimated. In particular, we design an F-formation recognizer based on a Hough-voting strategy, which lies between an implicit shape model, where weighted local features vote for a location in the image plane, and a mere generalized Hough procedure where the local features have not to be in a fixed number as in the implicit shape model. This approach provides the estimation of the position of the social space subject to the interaction, so as the persons involved, thus individuating the people which are socially interacting.

In such regard, our approach is the first to use F-formations detection in order to discover social interactions solely from visual cues. It has been tested on about a hundred of simulated scenarios, and two real annotated datasets. In the latter two cases, tens of individuals are captured while they were enjoying coffee breaks, in indoor and outdoor environment, giving rise to heterogeneous real crowded scenarios. Our method obtains convincing results and it compares favorably against other similar work aimed at group detection.

Following the analysis of non-verbal cues for the detection of social signals, our second contribution is related to the characterization of the group interactions by proposing a Voice Activity Detection approach only based on the automatic measurement of the persons’ gesturing activities.

This work takes inspiration from the fact that it is common experience to observe that people accompany speech with gestures, the *"... range of visible bodily actions that are, more or less, generally regarded as part of a person`s willing expression ..."* [2]. Far from being independent phenomena, speech and gestures are so tightly intertwined that every important investigation of language has taken gestures into account, from *De Oratore* by Cicero (1st Century B.C.) to the latest studies in cognitive sciences [3,4] showing that the two modalities are *"... components of a single overall plan ..."* [2].

Hence, this work presents a method for estimating the level of gesturing as a means to perform Voice Activity Detection (VAD), i.e. to automatically recognize whether a person is speaking or not. The main rationale is that audio, the most natural and reliable channel when it comes to VAD, might be unavailable for technical, legal, privacy related issues or simply for a noisy scenario. A condition that applies in particular to surveillance scenarios where people are monitored in public spaces and are not necessarily aware of being recorded.

Previous works take advantage of restrictive experimental setups, such as a meeting room where the audio and video signals can be extracted with a strong degree of reliability. This allows to perform an accurate diarization of the audio signal (e.g., [Ba2010]) in a smart meeting room, which remarkably simplifies the VAD by using multiple cameras capturing each person individually at close distance. Our work breaks such constraints and deploys a system "in the wild" designing a more credible setup for a video surveillance system. We use solely visual cues obtained from only one camera positioned 7 meters above the scene. In particular, the experiments focus on people involved in standing conversations, with an automatic person tracking system that follows each individual. Our VAD method is based on a local video descriptor that extracts from each individual body the optical flow, encoding its energy and "complexity" using an entropy-like measure. This allows one to discriminate between body oscillations or noise introduced by the tracker, where the optical flow is low and homogeneous, and genuine gestures, where the movement of head, arms and trunk produces a local flow field which is diverse in both intensity and direction. The descriptor extracted for each participant produces a signal that can be used for VAD.

The proposed approach is interesting for three main aspects. First, the relationship between speech and gestures has been widely documented and studied, but relatively few quantitative investigations of this phenomenon have been made. Second, approaches similar to ours might help to infer information about privacy protected data (speech in this case) from publicly accessible data (gestures in this case): this is also important for establishing whether the simple absence of a certain channel is sufficient to protect the privacy of people and how much. Finally, inferring missing data from available ones can make techniques dealing with challenging scenarios more effective and reliable.

The results show that our gesturing level estimation approach proposed achieves, on a frame-by-frame basis, an accuracy of 71% in distinguishing between speech and non-speech for the tested cases.

The last contribution is about the characterization of the group interaction aimed at the recognition of the relations among the interlocutors by using proxemic cues.

Proxemics can be defined as "[...] the study of man's transactions as he perceives and uses intimate, personal, social and public space in various settings [...]", quoting Hall [1], [2], the anthropologist who first introduced this term in 1966. In other words, proxemics investigates how people use and organize the space they share with others to communicate, typically outside conscious awareness, socially relevant information such as personality traits (e.g., dominant people tend to use more space than others in shared environments), attitudes (e.g., people that discuss tend to seat in front of the other, whereas people that collaborate tend to seat side by side), etc..

This work focuses on one of the most important aspects of proxemics, namely the relationship between physical and social distance. In particular, the paper shows that interpersonal distance (measured automatically using computer vision techniques) provides physical evidence of the social distance between two individuals, i.e. of whether they are simply acquainted, friends, or involved in a romantic relationship. The proposed approach consists of two main stages: the first is the automatic measurement of interpersonal distances, the second is the automatic analysis of interpersonal distances in terms of proxemics and social relations.

The choice of distance as a social relation cue relies on one of the most basic and fundamental findings of proxemics. In fact, people tend to unconsciously organize the space around them in concentric zones corresponding to different degrees of intimacy [5], [6]. The size of the zones changes with a number of factors (culture, gender, physical constraints, etc.), but the resulting effect remains the same: the more two people are intimate, the closer they get. Furthermore, intimacy appears to correlate with distance more than with other important proxemic cues like, e.g., mutual orientation. Hence, it is reasonable to expect that the distance accounts for the social relation between two persons.

One of the main contributions of the paper is that the experiments consider an ecological scenario (standing conversations) where more than two people are involved. This represents a problem because in this case distances are not only determined by the degree of intimacy, but also by the need of ensuring that every person can participate to the interaction. This leads to the emergence of stable spatial arrangements, the already mentioned F-formations, that impose a constraint on interpersonal distances and need to be detected automatically solely using unobtrusive computer vision techniques. Furthermore, not all distances can be used because, in some cases, they are no longer determined by the degree of intimacy, but rather by geometric constraints.

Our approach is to consider only the distances between people adjacent in the F-formation and, unlike other works in the literature, the radii of the concentric zones corresponding to different degrees of intimacy are not imposed a-priori, but rather they are learned from the data using an unsupervised approach. This makes the technique robust with respect to the factors affecting proxemic behavior, like culture, gender, etc., as well as environmental boundaries. In particular, the experiments show how the organization into zones changes when decreasing the space available to the subjects and how the unsupervised approach is robust to such effect.

Standing conversations are an ideal scenario not only because they offer excellent examples of proxemic behavior, but also because they allow one to work at the crossroad between surveillance technologies, often applied to monitor the behavior of people in public spaces, and domains like Social Signal Processing [7] that focus on automatic understanding of social behavior. All the presented techniques draws a path towards a new paradigm for analyzing social interactions where social signals can be extracted in unconstrained scenarios. This is expected to lead, on the long-term, to socially intelligent surveillance and monitoring technologies [8].

References

1. Kendon, A. (1990). *Conducting Interaction: Patterns of behavior in focused encounters*. Cambridge University Press.
2. Kendon, A. (2000). Language and gesture: unity or duality? In *Language and Gestures*, D. McNeill Eds., Cambridge University Press, 47-63.
3. McNeill, D.(1992). *Hand and mind: What gestures reveal about thought*. Chicago University Press.
4. Kendon, A. (1980). *Gesticulation and speech: Two aspects of the process of utterance*. In: The Relationship of verbal and nonverbal communication, M.R.Key Eds., Mouton Publishers, The Hague, 207-227.
5. Hall, E.T. (1974). *Handbook for proxemic research. Studies in the anthropology of visual communication series. Society for the Anthropology of Visual Communication*.
6. Hall, E.T. (1966). *The hidden dimension*. Doubleday Eds. New York.
7. Vinciarelli, A., Pantic, M., Bourlard, H. (2009). Social Signal Processing: Survey of an emerging domain. *Image and Vision Computing* **27**, 1743-1759.
8. Cristani, M., Murino, V., Vinciarelli A. (2010). Socially intelligent surveillance and monitoring: analysing social dimensions of physical space, *Int’l Wks on Socially Intelligent Surveillance and Monitoring*, 51-58.

Itour: Using Ambient Intelligence to Support Tourism

S. Alizadeh[1], M. Kanis[2], M. Veenstra[3]

Create-IT, Amsterdam University of applied sciences, Amsterdam, The Netherlands.
[1]sean.alizadeh@hva.nl, [2]marije.kanis@gmail.com, [3]Mettina.Veenstra@novay.nl

Abstract

Technology is becoming omnipresent in public spaces: from CCTV cameras to smart phones, and from large public displays to RFID enabled travel cards. Although such technology comes with great potential, it also comes with apparent (privacy) threats and acceptance issues. Our research focuses on realizing technology-enhanced public spaces in a way that is acceptable and useful for the public. This paper gives a brief overview of the research that is aimed to unlock the positive potential of public spaces. This paper's main focus is on the acceptance of sensor technology in the realm of tourism. The ITour project which investigates the potential and acceptance of using (sensor) technology and ambient media to collect, uncover and interpret data regarding tourists' movements, behavior and experiences in the city of Amsterdam is particularly discussed as an example.

Introduction

Technology offers great potential for making public spaces more interesting, social, beautiful and effective [16]. However, the use of pervasive technology, such as sensors may sometimes be perceived as a violation of privacy, and large displays in public spaces might not always contribute to activities and needs of passers-by. Even more so, these can be perceived as useless or even annoying by some, instead of being a source of information or having social benefits. Thus, in order to fully realize this potential, examples are needed that are aimed to minimize the potential drawbacks and dangers that ambient technology withholds. For this purpose, the ITour project investigates the potential and acceptance of using (sensor) technology and ambient media to collect, uncover and interpret data regarding tourists' movements, behavior and experiences in the city of Amsterdam. The goal of this data collection and interpretation is to provide tourists with better and more targeted (ambient media) services. It focuses on developing personas and visualizing user scenarios (using techniques such as photo-based storyboards) to explore and assess these in close collaboration with the targeted groups.

ITour

In 2010, the city of Amsterdam counted more than 5,000,000 tourist arrivals [10] and a tourism growth rate of 14% compared to 2009 [11]. Indeed, according to its inhabitants and tourists Amsterdam is a highly rated city [14, 15], enabling visitors from all around the world to engage in all sorts of endeavors. Such activities range from visiting the vast variety of cultural and historical museums that Amsterdam has to offer, to canal cruising and attending art-exhibitions or concerts. Upcoming technologies, particularly sensor-mediated systems, have great potential to better identify and support such touristic activities and services. However, the wide range of tourists and possibilities make it difficult to determine the most suitable services to design and cater for, also from the perspective of the stakeholders in the tourism sector. Therefore, the focus of this study is to identify the different potential users and investigate their behavior and activity patterns for designing and developing tourist-targeted technology services. In order to increase understanding toward these new potential services, the potential end-users of such services are closely involved in the study process, through the creation and discussion of personas, (photo-based) user scenarios (such as [9]) and interviews with tourists and other stakeholders.

Related work

Research (e.g. [1,2,4]) has shown that various sensors in the environment can be utilized for mobile applications and guides that help supporting and personalizing the touristic experience. Sensors, such as GPS, electronic compasses and wireless networks, create the opportunity to determine the location of tourists, and as such, enable personalized services on location. Research (e.g. [2-4]) suggests that offering such personalized services can be done within the real time-frame that tourists are undertaking activities, such as visiting a museum or an art gallery. Additionally, researchers (such as [5-8]) consider the data-mining of social media and other web platforms a viable basis for offering personalized information to tourists. Such data could subsequently be used to generate an overview of a tourist's personal interests, needs and intentions with regard to undertaking activities in a foreign city such as Amsterdam. Thus, a large body of research has shown that there is a wide variety of ambient technology that could support tourists in undertaking their activities. However, when regarding such ambient intelligence –a vision that combines concepts of ubiquitous technology, intelligent systems, responsive environments and advanced user interfaces– and particularly its usage and privacy implications, developing applications that are suitable and accepted by tourists remains a challenge.

There have also been a number of studies in various fields, which focus on evaluating user acceptance toward the use of technology. Some studies in health care suggest [12,13] that despite the benefits of various technologies that are specifically designed to aid patients in their daily life, these applications fail in providing the desired results due to lack of acceptance by the user. Also acceptance of technology by health care professionals is considered [19] important to maintain a high quality of service provided by health care application. Acceptance of technology is also considered important in various educational settings [20,21] such as e-learning and traditional learning. Other interesting fields of study are Law enforcement [23] and predicting consumer behavior [22] with regard to acceptance and adoption of technology.

Study

General approach to gauge public's needs and acceptance

The general approach of our research is human-centered, in which the needs of the users and the acceptance of technology of public space is key. For this research purpose, data is collected in two ways: by hand and with sensors. We use a living lab approach where we place technology in a public space for a longer period, sometimes up to several years [17].

For the collection of data by hand, well-known techniques such as focus groups, interviews, observations and surveys are being applied. In some projects targeted instruments are being developed for instance for charting the activities and needs of users of a specific public space. For the automatic collection of data we use technology such as interactive mats, RFID, GPS, computer vision, logging of touches of categories of information on a touch screen and social media. Furthermore, interactive models [18] are developed for discussing and visualizing the potential of ambient media.

This data collection approach is aimed to serve different purposes. Firstly, systems such as public displays are made more intelligent and responsive by using sensors. Secondly, the software or services, such as services for tourists, studied by collecting and analyzing data are improved by humans based on the research results. And finally, the data is used for visualizations, for instance a visualization of people's moods on an interactive map.

ITour study

In order to address the goal of supporting Amsterdam's tourist needs through utilizing (sensor) technology, the strategy is to:

- Identify and categorize different sources of data with regard to touristic activities, so to generate an overview of the available sensors to collect these data, and of available and future services that can benefit from these data.
- Develop (visual) scenarios for user-targeted and personalized ambient media, based on data that is acquired through a variety of available ambient sensors.
- Discuss these scenarios with user groups in order to match the information and service needs of tourists with the generated data and to address acceptance and privacy issues.

- Investigate whether and in what way the collected data could be applied for understanding and supporting the needs of possible stakeholders, such as tourists and organizations in the tourist sector. The resulting knowledge could then enable companies and organizations in the tourist sector to enhance and improve their existing and future services.

Engaging the targeted user group is considered key for the ITour project. Hence, a user-centered study (including interviews and the development and discussion of scenarios) is planned to be conducted to: (1) Gauge tourists' acceptance issues toward the use of various ambient intelligence and technology, and (2) Clarify what kind of (sensor) technology is considered viable and desired for the purpose of gathering touristic data.

The different types of data

A data-driven approach is chosen for the purpose of creating an inventory of needs with regard to tourists and ICT-services in the tourist sector. Examples of the types of different data that can be collected are categorized as follows:

- **Data about movements of tourists**: This kind of data depicts the movement pattern of tourists in Amsterdam. For instance; how does one transit from a hotel room to a museum nearby? What kind of transportation is used? What routes? In which order are they visiting different attractions in Amsterdam?

- **Data about activities of tourists**: This data that can be acquired through the use of available information services (such as reservation and payment systems) in the city. Analyzing such data will create an impression of how these services are being used by the tourists in Amsterdam. This type of data includes, for example, the type of service that is used, the amount of time spent by tourists using the available services and the type of locations visited.

- **Data used to gauge the total amount of time spent in Amsterdam**: a few interesting types of data are for example Hotel check-ins, check-outs and length of stay in hotels or other temporary accommodations.

- **Tourist demographics:** Tourist demographics are also a valuable type of data. Via categorizing this type of data, one could better understand different types of tourists and their individual needs.

- **Data about tourists' experiences:** A few suggested methodologies for data collection with regard to experiences are; (1) Experience sampling through mobile phone applications; (2) Data-mining the social media and web applications, such as Twitter and Flickr, for mood and emotion sampling; (3) Using text-mining on tourist forums and social media to gauge people's opinion with regard to touristic Amsterdam; (4) Experience sampling on touristic locations via cameras and similar devices.

Data segmentation, denoting the analysis and categorization of various tourist related data, could subsequently lead to more accurate understanding of creating (personalized ICT-) services. This could then catalyze more relevant and targeted communication between different stakeholders. In addition, segmentation of tourist data helps in constructing unique profiles of each tourist archetype based on their needs and characteristics.

Discussion and Conclusion

This paper presented the ITour project that investigates the employment of ambient intelligence for services for tourists. Research has shown the interesting potential of using (sensor) technology and ambient media to support tourism. However, to provide tourists with better, acceptable and more targeted services that use ambient-intelligence, these need to be discussed and assessed with the user in mind. For this purpose, personas were developed, which will be used for further user-centered study, including the development and discussion of desired scenarios. Such discussion is needed in order to fully understand the implications and requirements of ambient intelligence from the tourist perspective. Eventually, the presented study approach will lead to increased understanding of people's attitudes and privacy concerns when regarding ambient personalized systems so that better suited services can be offered.

Acknowledgements

The research reported in this paper has been supported by the Foundation Innovation Alliance (SIA) with funding from the Dutch Ministry of Education, Culture and Science (OCW), as part of the 'Smart Systems for Smart Services' project.

References

1. Kenteris, M., Gavalas, D., Economou, D. (2011). Electronic mobile guides: a survey. *Pers. Ubiq. Comp.* **15**, 97-111.

2. Gavalas, D., Kenteris, M. (2011). A web-based pervasive recommendation system for mobile tourist guides. *Pers. Ubiquit. Comput.* **15**, 759-770.

3. Kuflik, T., Stock, O., Zancanaro, M., Gorfinkel, A. et al. (2011). A visitor's guide in an active museum: presentations, communications, and reflection. *ACM J. Comput. Cult. Herit.* **3**(3), 11.

4. Schöning. J, Hecht. B, Starosielski. N. (2008). Evaluating Automatically Generated Location-Based Stories for Tourists. *Proceedings CHI 2008*, 2937-2942.

5. Popescu. A, Grefenstette. G (2011). Mining Social Media to Create Personalized Recommendations for Tourist Visits. *Proceedings of the 2nd International Conference on Computing for Geospatial Research & Applications*, 1-6.

6. Naaman, M. (2011). Geographic Information from Georeferenced Social Media Data. *SIGSPATIAL Special* **3**(2), 54-61.

7. Cao, T., Nguyen, Q., Nguyen, A., Le, T. (2011). Integrating open data and generating travel itinerary in semantic-aware tourist information system. *Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services*, 214 - 221

8. Kurashima, T., Iwata, T., Irie, G., Fujimura, K. (2010). Travel route recommendation using geotags in photo sharing sites. *Proceedings of the 19th ACM international conference on Information and knowledge management*, 579-588.

9. Greenberg, S., Carpendale, S., Marquardt, N., Buxton, B. (2012). The narrative storyboard: telling a story about use and context over time. *Interactions* **19**(1), 64-69.

10. Fedorova, T. (2011). *Amsterdam fact sheet*. Amsterdam, research and statistics economic development, Geemente Amsterdam. http://www.os.amsterdam.nl/pdf/2011_factsheets_6.pdf Accessed 30 June 2012.

11. Fedorova, T. (2010). *Amsterdam fact sheet*. Amsterdam, research and statistics economic development, Geemente Amsterdam.

12. Or, C., Karsh, B., Severtson, D., Burke, L., Brown, R., Brennan, P. (2011). Factors affecting home care patients' acceptance of a web-based interactive self-management technology. *JAMIA* **18**(1), 51-59.

13. Robinson, C. (2007). Clinical Adoption of IT. *Canadian Nursing Informatics Journal* **2**(1), 4- 21.

14. Van der Veer, J., Kan, S. Y. (2010). Wonen in Amsterdam 2009, Leefbaarheid (quality of life). Factsheet 2010.

15. Euromonitor (2011). http://blog.euromonitor.com/2011/01/euromonitor-internationals-top-city-destinations-ranking.html. Accessed January 2011.

16. Veenstra, M., Kanis, M., Groen, M., Meys, W., Slakhorst, W. (2011). Beyond Advertising: Large Displays for Supporting People's Needs and Activities in Public Space. *Proceedings of the CHI'11 Workshop on Large Displays in Urban Life*.

17. Kanis, M., Groen, M., Meys, W., Veenstra, M. (2012). Studying screen interactions long-term: The library as a case. *International Symposium on Pervasive Displays 2012,* Porto, Portugal.

18. Kanis, M., Robben, S., Kröse, B. (2012). Miniature play: Using an interactive dollhouse to demonstrate ambient interactions in the home. *Proceedings of DIS 2012,* June 11-15, 2012, Newcastle, UK.

19. Chau, P., Jen-Hwa, Hu. P. (2002) Investigating healthcare professionals' decisions to accept telemedicine technology: an empirical test of competing theories. *Information & Management* **39**, 297-311.

20. Zhang, C. (2010). Technology acceptance in learning settings from a student perspective: a theoretical framework. *Proceedings of the 2010 ACM conference on IT education*. ACM, NY, USA, 37-42.

21. Alshare, K., Grandon, E., Miller, D. (2004). Antecedents of computer technology usage: considerations of the technology acceptance model in the academic environment. *J. Comput. Small Coll.* **19**(4), 164-180.

22. Chen, L., Gillenson, M.L., Sherrell, D.L. (2004). Consumer acceptance of virtual stores: a theoretical model and critical success factors for virtual stores. *SIGMIS Database* **35**, 8-31.

23. Hu, P. J., Chen, H., Hu., H.F. (2009). Law enforcement officers' acceptance of advanced e-government technology: a survey study of COPLNK mobile. *Proceedings of the 11th International Conference on Electronic Commerce (ICEC '09)*. ACM, New York, NY, USA, 160-168.

Acknowledgements

Scientific Program Committee

- Dr. Andrew Spink (chair, scientific program committee), Noldus Information Technology bv, Wageningen, The Netherlands.
- Prof. Remco Veltkamp (conference chair), Media Studies, Utrecht University, Utrecht, The Netherlands .
- Prof. Gernot Riedel (conference chair), Department of Biomedical Sciences, School of Medical Sciences, University of Aberdeen, Aberdeen, United Kingdom.
- Prof. Berry Spruijt (honorary chair), Behavioral Biology Group, Department of Biology, Utrecht University, Utrecht, The Netherlands.
- Dr. Boris de Ruyter, Cognitive & Behavioral Insights User Experiences Department, Philips Research Europe, Eindhoven, The Netherlands.
- Dr. Emilia Barakova, Department of Industrial Design, Technical University of Eindhoven.
- Dr. Egon van den Broek, Professor in Man-Machine Interaction, Technical University of Vienna, Vienna, Austria.
- Prof. Wim Crusio, Research Director, French National Research Council, Talence, France.
- Prof. Robert Gerlai, Department of Psychology, University of Toronto, Mississauga, Ontario, Canada.
- Dr. John Krantz, Department of Psychology, Hanover College, Hanover, IN, USA.
- Dr. Lucas Noldus, Noldus Information Technology bv, Wageningen, The Netherlands.
- Dr. Fabio Paternò, Laboratory on Human Interfaces in Information Systems, Institute of Information Science and Technologies, C.N.R., Pisa, Italy.
- Prof. Vicenç Quera, Department of Behavioral Science Methods, University of Barcelona, Barcelona, Spain.
- Prof. Matthias Rauterberg, Department Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands.
- Dr. Hugh Marston, TPP Global Development, Edinborough, Scotland.
- Prof. Ilan Golani, Department of Zoology, Faculty of Life Sciences, Tel Aviv University, Israel.

Local Organizing Committee

- Natasja Bogers, Chair, Local Organizing Committee
- Andrew Spink, Co-chair, Program Committee
- Melvin de Bruijn, Finances
- Sabine van der Meijde, Conference secretariat

Conference Secretariat

- Sabine van der Meijde
- Nathalie Oostendorp
- Nancy van den Dam
- Mariska Walter
- Sidney Hooijer, Finances

Author Index