methods

Automatic Detection of Gaze Convergence in Multimodal Collaboration: A Dual Eye-Tracking Technology

Anna Shvarts
Lomonosov Moscow State University, Moscow, Russia;
Utrecht University, Utrecht, The Netherlands

Andrey Stepanov
National Research University Higher School of Economics, Moscow, Russia

Dmitry Chumachenko
Lomonosov Moscow State University, Moscow, Russia

Abstract. The paper analyses the advantages and limitations of the current technical solutions for dual eye-tracking (DUET) in relation to the research questions from educational science about joint attention in a multimodal teaching/learning collaboration. The insufficiency of the current systems for the analysis of multimodal collaboration is stated as the reviewed systems do not allow researchers to relate a participant's eye movements to the video from their joint performance and accompanying gestures without time consuming manual coding. We describe a system of two low-cost Pupil-Labs eye-trackers and propose an open source utility *DUET for Pupil* that automatically produces synchronized gaze data in the shared system of coordinates. The data are available in the form of a video from the surface that is overlaid by gaze paths with supplementary sound waveforms and as textual data with synchronized coordinates of the two gazes. Our empirical evaluation of this technological solution reports 1.27° of visual angle as the spatial accuracy of the system after post-hoc calibration. The advantages, limitations, and further possible enhancments of the system are discussed.

Correspondence: Anna Shvarts, shvarts.anna@gmail.com, Utrecht University, Freudenthal Institute, Buys Ballot building third floor, 3.67, Princetonplein 5, 3584 CC Utrecht, the Netherlands; Andrey Stepanov, pisaga@yandex.ru; Dmitry Chumachenko, dmitry.chumachenko@gmail.com

Keywords: collaboration, dual eye-tracking (DUET), gesture, joint attention, multimodal communication, head-mounted eye-tracker, remote eye-tracker, Pupil-Labs

Acknowledgements. We would like to thank Phillip Barsky for the initial idea of dual eye-tracking technology that led us along the extensive route of technical investigations. Discussions with Anatoly Krichevets and his encouragement were highly valuable during the process of elaboration. The paper is supported by RFBR, Grant #18-013-00906.

Received June 17, 2018, accepted September, 28, 2018.

Introduction

Dual eye-tracking (DUET) is a currently developing technology that allows synchronous tracking of eye movements from two people. It was first elaborated a decade ago and still appears to be quite a rare technology despite the number of fascinating research questions that could be investigated with dual eye-tracking. One of the reasons for the rare engagement of this technology is a technical challenge that either 1) diminishes the ecological validity of the study by requesting a separation of two participants into different spaces and preventing them from multimodal interaction or 2) leads to very time-consuming manual data analysis.

In this paper, we explore how dual eye-tracking may serve the research goal of investigating the collaboration process and what kind of technical requests this research question entails. Our initial focus is on the teaching/learning collaboration, however we suppose that investigations of other forms of collaboration would have similar technical requests. We review the current technological solutions for the DUET, clarifying their advantages and limitations for the investigation of multimodal collaboration, and provide a description of our equipment that involves two inexpensive head-mounted eye-trackers. Our technological solution allows two participants to interact in the shared space in front of one display with the stimuli; the following analysis leads to a calculation of the gaze positions of the two participants in the same coordinates. The software produces a video from the screen with overlaid gaze paths from two participants that is supplemented by a sound waveform that allows the analysis of the accompanying verbal communication with temporal precision.

We suggest this paper serves two purposes: 1) it provides detailed information about existing DUET solutions and their limitations and advantages, depending on the research questions and 2) it suggests a low-cost accessible DUET system for those who are in search of such equipment and interested in the analysis of multimodal interactions.

Dual Eye-Tracking as a Research Instrument for Multimodal Collaboration

Investigations of referential behavior between a speaker and listener have revealed the spatial coordination between two subjects' gazes (Richardson & Dale, 2005). This kind of coupling between two communicating subjects was further confirmed in a dual eye-tracking study of a dialogue (Richardson, Dale, & Kirkham, 2007) and joint collaboration on program comprehension (Sharma, Jermann, Nüssli, & Dillenbourg, 2013). The immediate application of these results to collaborative tasks is a gaze-sharing paradigm, where one partner benefits from the depicted gaze of the other one (Schneider & Pea, 2013; Velichkovsky, Pomplun, & Rieser, 1996). In the quickly developing field of multimedia education, the gaze-sharing paradigm is promising in terms of training the perception of the learner. However, the situation appears to be more complicated than was expected from the beginning: while sharing experts' eye movements helps in conceptually low fields, such as biology (Jarodzka, Van Gog, Dorr, Scheiter, & Gerjets, 2013) or medical images (Gegenfurtner, Lehtinen, Jarodzka, & Säljö, 2017), in other domains such as logical problem-solving (van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009) or geometry (van Marlen, van Wermeskerken, Jarodzka, & van Gog, 2016) it does not lead to additional learning gains. The same concerns other ways of visual cuing (S. Schneider, Beege, Nebel, & Rey, 2018): direct guidance of students' perception does not lead to learning gains in mathematics. Deeper investigation of the teaching/learning collaboration is needed to understand the development of perceptual strategies in a variety of learning domains.

One hypothesis would be that logic and geometry belong to abstract domains and in these cases joint visual attention is not required, but perceptual strategies are taught through verbal guidance. However, according to the research in mathematics education, this is not the case: mathematical teaching/learning is a multimodal process, in which nuances of prosodic and verbal modalities are interrelated with gestures and visual information (e.g., Hwang & Roth, 2011). In student-tutor collaborations, a tutor does not guide a student but re-voices her multimodal utterances, steering the collaborative system towards conceptual understanding (Flood, 2018), and a student does not follow a tutor, but actively anticipates the tutor's moves as she aims to coordinate between multiple semantic presentations (Shvarts, 2018b). In these situations, the verbal utterances cannot clearly direct attention since the referential system itself is under construction during participatory meaning making (De Jaegher & Di Paolo, 2007).

This complicated multimodal process in which joint visual attention is established and sustained in a teaching/learning collaboration forced us to seek a system that would let us analyze the dynamics of joint attention in the flow of multimodal interactions. While we start from the teaching/learning situation, we assume that any collaboration benefits from the possibility to gesture in a shared space and the role of gestures needs to be investigated to enhance computer-supported systems for collaboration. Below we provide a review of existing dual eye-tracking systems and highlight the missing requirements for our research question.

Existing Technological Solutions for Dual Eye-Tracking

Head-mounted and remote eye-trackers are the main classes of the equipment for eye movement investigations. Both types of trackers may be used to track the eye movements of two or more participants simultaneously; however, these systems have different advantages and limitations. Analysis with technical details may be found in the literature (Nüssli, 2011; Anan'eva, Zhegallo, Kurakova, & Kharitonov, 2012; Anan'eva & Kharitonov, 2011). In this paper, we aim to review the many technical DUET solutions that have been offered so far from the point of their suitability for the analysis of multimodal collaboration, taking into account their ergonomics, spatial accuracy of a recording, synchronization procedure, supplementary multimodal information and a method of processing raw data. DUET is a new and quickly developing methodology, so many opportunities have emerged in the past few years.

Two Remote Eye-Trackers with Two Displays

A typical method of dual eye-tracking includes two participants who are sitting in front of two displays that present identical pictures; thus two people perceive and can discuss the same material (Bednarik & Shipilov, 2011; Belenky, Ringenberg, & Olsen, 2014; Carletta et al., 2010; Guo & Feng, 2013; Molinari, 2015; Richardson, Dale, & Kirkham, 2007; Sharma, Caballero, Verma, Jermann, & Dillenbourg, 2015; Basyul, Mongush, Anan'eva, Tovuu, & Kharitonov, 2017 and others). Eye movements in this case are recorded by remote eye-trackers that are usually connected through Ethernet or another interface for synchronization.

The accuracy of eye-tracking data is one of the most significant advantages of this method since precision is as good as in the case of eye-tracking with one person. The data analysis is simple since one easily acquires the coordinates of the gaze on the screen from both participants that are synchronized in time. Later, all the usual methods of eye movement analysis may be applied, such as attendance of particular areas of interest (Belenky et al., 2014; Jermann, Nüssli, & Li, 2010), as well as measures of cross-recurrence (Richardson et al., 2007) or gaze similarity (Sharma et al., 2015). Scan paths can be easily converged on a screen, as synchronized data report a coordinate for each tracker, although researchers do not report whether they used this representation at the preliminary stage of their quantitative analysis.

This technology also allows gaze sharing, which might be especially important in investigations of collaborative problem solving (Newn, Velloso, Carter, & Vetere, 2016) and learning (Bednarik & Shipilov, 2011; Guo & Feng, 2013; Schneider & Pea, 2013) and in the studies of collaborative visual search (Acartürk, Tajaddini & Kilic, 2017, Brennan, Chen, Dickinson, Neider, & Zelinsky, 2008; Dalmaijer, Niehorster, Holmqvist & Husain, 2017; Neider, Chen, Dickinson, Brennan, & Zelinsky, 2010).

Also, DUET with this technical solution may easily be enhanced towards group eye-tracking, which may simply speed up data collection (Bielikova et al., 2018), track interactive processes within a group (Acartürk, Tajaddini & Kilic, 2017; Skuballa & von Suchodoletz, 2017; Räihä, Spakov, Istance & Niehorster, 2017) or provide a gaze sharing setup (Nyström, Niehorster, Cornelissen, & Garde, 2017). In the future, these systems are promising for eye-tracking classrooms that would inform a teacher about students' cognitive processes (Skuballa & von Suchodoletz, 2017; Räihä, Spakov, Istance, & Niehorster, 2017).

The limitation of this method is that participants do not have a shared space and their communication is limited to the audial channel. Consequently, no embodied interaction is possible between the partners: neither gestures nor joint performance and (physical) joint visual attention. This DUET method is inappropriate for investigations within any approach that stresses the importance of the embodied aspect of interaction and requires multimodal data, including analysis of teaching/learning collaborations for which gestures are very important.

Two Remote Eye-Trackers with One Display

Another technical solution consists of two remote eye-trackers and one display. To the best of our knowledge, the first time this solution was used was in 2008 when Tobii 1750 and ASL 504 remote eye-trackers were used to record programmers who worked collaboratively behind a display from the Tobii eye-tracker (Pietinen, Bednarik, Glotova, Tenhunen, & Tukiainen, 2008). A similar technical solution was presented at the ECEM conference in 2015 (Saunders, Melcher, & van Zoest, 2015): the researchers calibrated two Eyelink systems at one display and investigated oculomotor behavior in a joint visual search, which led to longer saccadic latencies in comparison to individual settings.

We merged the data from two remote eye-trackers on an image from a shared display as well (Shvarts, 2018b; Shvarts & Zagorianakos, 2016). Two remote eye-trackers (SMI RED and Eye Tribe) were calibrated on the same screen (infrared illumination of one of the eye-trackers were covered, but it did not influence the calibration accuracy) and then we overlaid the synchronized gaze paths above the presented stimulus. Video from an external camera of a SMI tracker was added synchronously into the video stream of eye movements along with the audio stream.

While shared space and eye-tracking accuracy are the advantages of this method, the limitation is determined by the ergonomics of the system: both participants need to sit very close to each other. They also cannot physically act on the same surface and are forced to gesture with a very thin pointer to avoid covering the infrared luminance.

Another limitation relates to analysis: this method did not allow processing a video on the screen, but only a series of stable images. Precise synchronization is also an issue, as it is done after the recording and is performed manually by detecting the same event in two independent eye-tracking systems. Pietinen and colleagues (2008) synchronized two eye-trackers at the moment when both participants would deliberately look away from a stimulus, but found this solution to be imprecise. Our manual synchronization was based on tracking the appearance of a stimulus on the screen: it was detected by Ogama, a software that served for the recording of eye movements by Eye-tribe, and by an external camera from the SMI eye-tracker. However, the unstable frame rate of the video from an external camera may be an obstacle towards precise temporal synchronization through video. It forced us to produce a synchronization between this video and eye movements separately for each stimulus (Shvarts, 2018b; Shvarts & Zagorianakos, 2016). The systems with two head-mounted eye-trackers, which we describe in the next section, meet the same challenge.

Two Head-Mounted Eye-Trackers

One more approach involves two or more head-mounted eye-trackers (Fedorova, Kibrik, & Yazykov, 2015; Hannula, Salminen-Saari, Garcia Moreno-Esteva, Toivanen, & Salonen, 2017; Ho, Foulsham, & Kingstone, 2015; Kurzhals, Hlawatsch, Seeger, & Weiskopf, 2017; Pfeiffer & Renner, 2014; Schneider et al., 2016a; Shvarts, 2018a; Yu & Smith, 2016). Synchronization may be achieved at the stage of analysis, for example through a hand clap which is recorded by both world cameras (Ho et al., 2015), or through a wireless connection (Lilienthal & Schindler, 2017; Shvarts, 2018a; Shvarts & Abrahamson, 2018) or Network Time Protocol (NTP) server/client (Broz, Lehmann, Nehaniv, & Dautenhahn, 2013) while recording. With this DUET equipment, a study can be conducted in ecological settings: participants share the space and can gesture, have eye contact and move beyond the laboratory settings and computer screens. In rare cases, more than two eye-tracking goggles may be involved (Hannula et al., 2017).

The disadvantage of this technology lies in the analysis of data: each participant's eye movements are represented in the coordinate system of her visual scene, and thus some additional procedure is needed to determine if they are looking at the same area or object. Researchers are forced to manually code at which area of interest the participants are focused at each fixation, and only then the episodes of cross gazes or other measures can be calculated (Fedorova et al., 2015; Ho et al., 2015; Sandgren, Andersson, Weijer, Hansson, & Sahlén, 2012; Yu & Smith, 2016).

There are automatized solutions that may replace manual coding in some cases. Schwarzkopf and colleagues (Schwarzkopf, Büchner, Hölscher, & Konieczny, 2017) introduced the gaze angle analysis that takes into account both the angle of the gaze as it is related to the body and the body angle. As a result, they calculated the absolute direction of the gaze, while information about the height of the object in focus is lost. Other approaches include an automatic association of a fixation zone with a predefined area of interest (Kurzhals et al., 2017), a face (Broz et al., 2013) or a human body and an object (De Beugher, Brône, & Goedemé, 2014). These systems, although promising due to the growing computational powers, are not elaborated well enough to be accessible for dual eye-tracking.

The most widespread solution for the moment is incorporation of black and white markers similar to a Quick Response code into the visual scene that later is automatically recognized from the scene video. In some cases, these markers allow distinguishing an area of interest, such as an object of reference in a dialogue (Gergle & Clark, 2011; Pfeiffer & Renner, 2014) or a participant's face (Renner, Pfeiffer, & Wachsmuth, 2014). Such data processing is applicable only when the objects of interest are known beforehand.

According to our research interests, we would like to capture all the complexity of the social interaction and produce a video of two gazes on a surface that later might be qualitatively analyzed frame by frame. In similar cases, the markers are used as points of reference to establish a shared ground truth on which the gaze paths are overlaid (Lilienthal & Schindler, 2017; Schneider et al., 2016a). In these studies, the gaze paths are drawn on the picture of the initial state of the surface and data from the eye movements are synchronized with the scene videos from the world cameras, the frames of which are depicted synchronously near the frames with the gaze information.

The technical solution that is proposed in this paper works essentially the same way. However, the transformations of the ground truth are important for the microethnographgic analysis that we conduct (Shvarts, 2018a; Shvarts & Abrahamson, 2018), so we overlay the eye movements of both participants on the video from the surface that is cut out from the scene video. We further describe the details of this method.

The solution with markers is a vital attainment for dual eye-tracking technology since two participants can communicate in a natural shared space and a researcher has access not only to the eye movements but also to the synchronized video and audio. All information about the multimodal interaction is aggregated in one video package that might be analyzed qualitatively while qualitative measures of gaze convergence might be calculated as well.

The obvious limitation of this solution is its inferior spatial resolution compared to the case of dual eye-tracking with two remote eye-trackers and two displays. The next section is dedicated to a detailed description of our equipment and merging technology and then to an experimental measurement of dual eye-tracking accuracy.

DUET for Pupil: Convergence of Eye Movements from Two Head-Mounted Eye-Trackers on One Surface

The system that we describe here involves two head-mounted eye-trackers from Pupil-Labs (Kassner, Patera, & Bulling, 2014) that are inexpensive eye-trackers with relatively good accuracy (up to 1 degree according to their specification). The sample rate of the eye-trackers may vary from 30 to 120 Hz, and the world scene camera works at 30–120 Hz frequency (depending on the spatial resolution). Two provided lenses grasp 100 or 60 degrees of visual angle, which being combined with the appropriate spatial resolution of the image may be used in a variety of conditions from general orientation in the space (the image is very wide but with serious distortions on the sides) to relatively narrow directed perception, such as on a computer display (the image has moderate distortion on the sides and the display can be approximated by a rectangle in 3D space; see Figure 1a). An additional advantage of this system is the open source software that any user may elaborate according to her preferences (see also Schindler, Lilienthal, Chadalavada, & Ögren, 2016 for an analysis of its benefits for research in mathematics education).

For a dual eye-tracking record, each device is connected to a laptop (preferably with Mac OS or Linux). The Pupil-Labs recording system (Pupil Capture) is already equipped with a plugin for synchronization (Time Sync) that allows recording data with coordinated timestamps. Another plugin (Pupil Groups) lets a user start the recording on a few devices from one computer; thus synchronized data from more than one eye-tracker will be obtained.

The Pupil-Labs system provides an opportunity to place black and white markers on some surface in the environment (for example, a display). Later during the analysis (which is done by Pupil Player), this surface may be recognized automatically (by plugin Offline Surface Tracker). The positions of an eye in the surface based coordinate system can be calculated by homological transformation from the initial coordinates in the visual scene and then stored with the corresponding timestamps.

Our innovation concerns the analysis of the eye movements. To conduct a microethnographic analysis (Streeck & Mehus, 2005), we aimed to receive information that is as rich as possible about multimodal communication. Built on the open source Pupil-Labs system, we created an independent utility *DUET for Pupil*. This utility cuts the image of the surface from each frame of the scene video of one participant and transforms it into a rectangle. Having recognized the same surface in the scene videos from each of the two participants, we collect information about gaze of both participants within the same coordinate system. Then we draw their gaze positions on a frame that corresponds to this very moment and add the gaze positions from the past few seconds to depict gaze paths. Additionally, on each frame we draw a waveform of the sound that accompanied

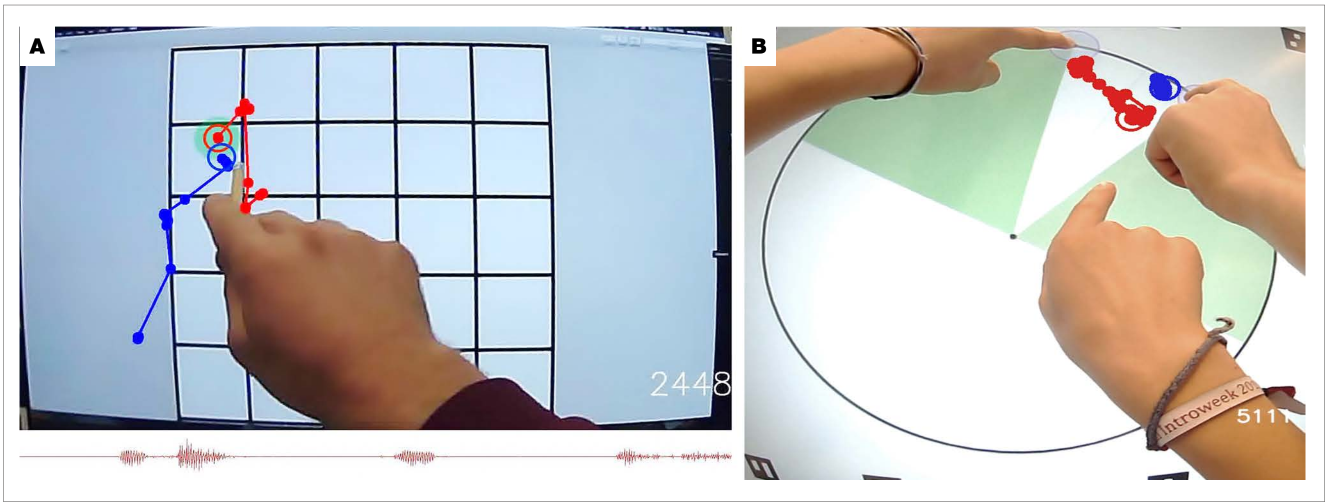

Figure 1. Examples of dual eye-tracking data frames that were produced by the utility *DUET for Pupil:* (a) each dyad's eye movements are depicted together with a gesture; (b) participants collaborate on one touch screen.

it during the past few seconds. Then all the frames with the gaze paths are combined into a video. As a result, the video consists of two gazes that are overlaid on a video from the initial surface together with any movements (e.g., gestures) that happen in front of the surface. See Figure 1 for an example of a frame (a) on which the dyad's eye movements are depicted together with a gesture and (b) on which the two participants collaborate together on one touch screen.

The configuration file includes the settings specified in Appendix A. Instructions on how to conduct a recording and how to use the utility are provided in Appendix B.

Empirical Evaluation of Spatial Accuracy

Aiming to evaluate the spatial accuracy of our dual eye-tracking system, we ran a small experiment where participants were asked to look synchronously at the same point on a screen.

Method

Participants. We recorded ten pairs of participants, aged from 18 to 69, with visual acuity ranging from normal to that corresponding to −3 diopters, which is the typical range for participants of eye-tracking studies in educational science research. All recordings were conducted without glasses.

Material and procedure. The task for the participants was to recognize a digit on the screen. In 2000 ms after a participant would start a trial by pressing any key, a white square appeared for 500 ms. Then a random digit appeared in the very same place for the next 500 ms. As soon as the participants have given their answers, they could start the next trial by pressing any key. The first participant was naming the digits, while the second one was confirming agreement or not.

There were 40 stimuli: 8 digits were placed in the center of the screen; 16 digits were at a medium distance from the center (in the corners and middles of the sides of a square with the side 400 px, approximately 5° and 11.3 cm); and 16 digits were at a larger distance from the center (in the corners and middles of the sides of a square with the side 800 px, approximately 10° and 22.5 cm). The sequence of the digits' positions was quasi-random, while the values were attributed randomly.

The experiment started with a practice series that included five trials with the same task. This series aimed to let participants become familiar with the procedure and to set the font size of the digits to the minimal size that the participants would be able to recognize (varied from 10 pt to 20 pt fonts, which corresponded approximately to 0.15–0.30 of the visual angle).

After the practice series, we run a 5-point calibration for each participant, and then started the main series.

Equipment. The participants were seated approximately 700 mm from a 22 inch flat display (473×296 mm with resolution 1680×1050). Each participant was wearing a Pupil-Labs eye-tracker (with 60 Hz for both eyes) world camera; the world camera had 60 degrees of visual angle. The eye-trackers were connected to a MacBook Pro with the processors 2.7 GHz and 2.9 GHz. The stimuli were presented by Psychtoolbox 3.0.12 that run in the Matlab 8.4.0.

The records of two participants in a pair were synchronized by the plugin Time Sync of Pupil Capture 1.5–12. The data were processed by Pupil Player 1.5–12 and then by our utility *DUET for Pupil*. We analyzed data with Excel and SPSS 20.0.

Data Analysis and Results

One pair's data were not analyzed due to the unstable recording of eye movements from one of the participants. Some trials were missing in the records of three other pairs due to technical issues.

In order to distinguish saccades, we used a velocity-based algorithm and calculated a fixation position as the center of the sample points between saccades. This algorithm was chosen as the simplest for implementation while still having relatively good quality (Salvucci & Goldberg, 2000). As we provide further an estimation of accuracy, the imprecise qualification of fixations might lead to an underestimation of the dual eye-tracking quality, not an overestimation. Assuming that shortly after a digit appeared onscreen two participants would look at the same place, we calculated the distance between the fixations of the two participants starting from a time point of 100 ms after the digit presentation in order to take into account the possible corrective saccades. This distance was calculated in visual angles and became a measure of the spatial accuracy of the dual eye-tracking. The trials where the visual angle between participants' gazes was larger than two standard deviations from this pair's mean were treated as outliers and excluded from the analysis.

According to the results, the mean spatial accuracy of our DUET technical solution is 2.46° of visual angle. There was no significant difference in the accuracy at the different distance from the center (see Table 1).

Table 1. Descriptive Statistics of the Spatial Accuracy

The position of stimuli	*N*	Mean, degree	Std. Deviation	Minimum	Maximum
Center of the screen	58	2.32	1.27	0.30	5.72
200 px from the center	123	2.47	1.22	0.41	7.78
400 px from the center	112	2.52	1.25	0.15	5.82
Total	293	2.46	1.24	0.15	7.78

In the videos with gaze paths we saw that there is a lack of calibration in many cases: all gaze samples were shifted towards one direction. We applied post-hoc calibration and moved the fixations towards one direction for each participant according to the vector from a central fixation to the center of the screen. After this recalibration, the mean accuracy became 1.27° of visual angle independently from the distance (see Table 2).

The mean accuracy varied from 0.7° to 2.2° among the participants (see Table 3).

Table 2. Descriptive Statistics of the Spatial Accuracy After Post-Hoc Calibration

The position of stimuli	*N*	Mean, degree	Std. Deviation	Minimum	Maximum
Center of the screen	50	1.23	1.23	0.11	6.53
200 px from the center	124	1.06	1.00	0.08	4.85
400 px from the center	110	1.52	1.07	0.15	6.10
Total	284	1.27	1.09	0.08	6.53

Table 3. A Variation of the Spatial Accuracy After Post-Hoc Calibration

	N	Mean	Std. Deviation	Minimum	Maximum
pair1	38	1.54	1.05	0.20	4.12
pair2	37	1.44	0.87	0.11	3.22
pair3	38	0.85	0.48	0.13	1.90
pair4	37	1.27	1.01	0.29	4.09
pair5	37	2.20	1.82	0.13	6.53
pair6	36	0.71	0.44	0.08	1.76
pair7	27	0.86	0.46	0.09	2.04
pair8	18	1.15	0.73	0.20	2.60
pair9	16	1.14	1.18	0.22	4.85

Discussion

Our results show that dual eye-tracking with Pupil-Labs trackers is less accurate than eye-tracking of one person or dual eye-tracking with two remote trackers. However, the precision is adequate for many tasks: 1.27° of visual angle is approximately 3 cm on a 22 inch display and about 100 px in case of 1680×1050 screen resolution. This limit is important to take it into account as a prerequisite to the experimental design. In many cases, it might be sufficient for the research task. For example, it would allow researchers to distinguish if the participants were attending together a face in the movie, or a particular representation in the learning materials (such as a figure or formula) and other material with Areas of Interest analysis (e.g., Belenky et al., 2014; Gergle & Clark, 2011; Jermann, Nüssli, & Li, 2010). Accuracy also needs to be considered when setting a threshold for the measures of gaze similarity in case it is calculated directly from the closeness of two gazes. While the influence of the main factor (3D or 2D manipulative in learning in problem solving) on joint visual attention was found independently from such a threshold (Schneider et al., 2016b), the chosen threshold of 50 px (B. Schneider et al., 2018) may be too narrow. However, there are also measures that allow a comparison of scan paths without spatial alignment, which would avoid the issue of special accuracy (Dewhurst et al., 2012).

The procedure of post-hoc calibration appears to be highly effective, so we suggest including some points of reference during the recording that would allow later re-calibration during the data analysis in any experiment.

Advantages and Limitations of *Duet for Pupil* and Futute Directions

The main advantage of our technical solution is an automated analysis of eye movements from two head-mounted eye-trackers. Unlike the other head-mounted DUET solutions (Gergle & Clark, 2011; Pfeiffer & Renner, 2014), our system provides not only the detection of synchronous attendance of areas of interest but a gradual estimation of the distance between the gazes and precise information about gaze position on a recognized surface. Similar systems were elaborated by two other research groups (Lilienthal & Schindler, 2017; Schneider et al., 2016a); however, they do not supplement eye-tracking data with a video from the surface and with a sound waveform, which are essential for the qualitative analysis of multimodal communication.

Spatial accuracy is the most challenging issue for head-mounted eye-tracking systems and the precision of our system is worse than what two remote eye-trackers would provide, but still around the physiological threshold of eye movements. The Pupil-Labs eye-trackers are sensitive to the minimal movements of the tracker on a forehead; however, these moves lead to a parallel shift and could be compensated by post-hoc calibration.

In our empirical studies (Shvarts, 2018a; Shvarts & Abrahamson, 2018) and during evaluation of precision, the participants were sitting shoulder to shoulder or standing near each other. They could gesture and look at each other's eyes, but convergence of eye movements on one surface is available only when they look at the screen or the interactive white board. However, it was sufficient for the qualitative analysis of intersubjective coordination dynamics in a teaching/learning process.

There are a few enchantments that we would like to develop in the future. The data from the surface lack synchronous information about the general scene. As a first and straightforward solution, we plan to incorporate the information from the world cameras of the participant into the final video. A more complicated and more valuable step is to synchronize eye-tracking data with information from an external camera. However, achieving good temporal resolution is a challenge, since an ordinary camera would not report timestamps, and may be out of sync for 100–200 ms, as, for example, an external camera in SMI RED, which we used in the previous studies (Shvarts, 2018b; Shvarts & Zagorianakos, 2016).

Another improvement might concern the variety of surfaces. It is possible to recognize more than one surface in an environment and detect if these surfaces were attended at the same time and where the participants were focused. We also would like to estimate the precision of the system in case participants are not sitting shoulder to shoulder and might communicate in less restricted settings.

For the moment, the analysis of gestures is limited to a qualitative description or manual coding. In the future, the system shall be enhanced by a motion tracking system (see e.g. Burger, Puupponen, & Jantunen, 2018 on the synchronization issues) and/or analysis of sensory-motor coordinations from the tracking actions on sensory screens.

The system is relatively cheap (two Pupil eye-trackers and two laptops — preferably MacBooks — are required) and easy to use since Pupil Player and Pupil Capture have user-friendly interfaces with many settings and plugins (see also instructions in Appendix B). Open source Pupil-Labs software allows creation and incorporation of custom plugins. You may find the utility *DUET for Pupil* at https://github.com/Arhisan/DUET-for-Pupil and instructions in Appendix B. Please contact the authors of this paper if you have any questions.

Conclusions

The paper reviews dual eye-tracking research technological solutions with remote and head-mounted eye-trackers that may involve one or two displays for the presentation of stimuli. In the investigations of dialogues, referential communication, collaborative problem solving, teaching and learning processes and joint visual attention between parents and toddlers, the researchers tend to use head-mounted eye-trackers as they wish to track communication in ecological settings. In these cases, the issue of data processing is challenging since each tracker reports the gaze coordinates in the coordinate system of its world camera. Besides time-consuming manual coding, automated solutions are based on machine learning algorithms of object recognition or take into account head movements. Currently, the most reliable method is the recognition of black and white markers that enable recalculation of the gaze position relative to the fixed points in the environment.

The proposed system relies on this algorithm of marker detection. It consists of two Pupil-Labs head-mounted eye-trackers and two laptops. The markers are put on a surface (a display or another flat surface in the environment) and our utility *DUET for Pupil* overlays the gaze paths of the participants above the video from the surface. Additionally, a sound waveform is provided as it simplifies the qualitative multimodal analysis. Synchronized data of eye movements on the marked surface are also provided in a text format for quantitative data analysis. In the paper we described the principals of this utility usage, and detailed instructions are provided in Appendix B.

According to our empirical evaluation of the spatial accuracy of the equipment, the distance between the calculated gazes while participants are looking at the same point is about 1.27° of visual angle. This precision is achieved after post-hoc calibration, which is necessary due to the high sensitivity of the eye-trackers to the movements on a forehead. This spatial accuracy is sufficient for answering many research questions, and our technological solution enables both quantitative and qualitative analysis of synchronized eye movements from two participants in a multimodal and dynamic environment.

References

Acarturk, C., Tajaddini, M., & Kilic, O. (2017). Group Eye Tracking (GET) applications in gaming and decision [abstract]. In R. Radach, H. Deubel, C. Vorstius, & M. Hofmann (Eds.), *Abstracts of the 19th European Conference on Eye Movements* (Vol. 10, p. 103). Wuppertal.

Anan'eva, K.I., & Kharitonov, A.N. (2011). Metodologicheskie problemy issledovaniya okulomotornoi aktivnosti v kommunikativnykh situatsiyakh [Technical problems in inverstigation of oculomotor activity in communicative situations]. In *Psikhologicheskie i psikhoanaliticheskie issledovaniya. Ezhegodnik 2010–2011 [Psychological and psychoanalytical studies. Year 2010–2011]* (pp. 37–49). Moscow: Moscow Institute of Psychoanalysis. (In Russian).

Anan'eva, K.I., Zhegallo, A.V., Kurakova, O.A., & Kharitonov, A.N. (2012). Registratsiya dvizhenii glaz v kommunikativnykh situatsiyakh: metodicheskie problemy [Registration of eye-movements in communicative situations: Technical problems]. In A.L. Zhuravlev, & V.A. Koltsova (Eds.), *Razvitie psikhologii v sisteme kompleksnogo chelovekoznaniya. Materialy Vserossiiskoi nauchnoi konferentsii, posvyashchennoi 40-letiyu instituta psikhologii i 85-letiyu ego osnovatelya B.F. Lomova [Development of psychology within the system of complex human study. Proceedings of the Institute of Psychology 40th anniversary / B.F. Lomov 85th anniversary conference]* (pp. 141–143). Moscow: Institute of Psychology, RAS. (In Russian).

Basyul, I.A., Mongush, A.M., Anan'eva, K.I., Tovuu, N.O., & Kharitonov, A.N. (2017). Reshenie graficheskoi zadachi v traditsionnoi i tekhnogennoi kommunikativno-kognitivnykh sredakh [Graphical problem solving in traditional and technologically enchanced communicative-cognitive enviroment]. In *Evolyutsionnaya i sravnitelnaya psikhologiya v Rossii. Teoriya i praktika issledovanii [Evolutionary and comparative psychology in Russia. Theory and practice of research]* (pp. 312–321). Moscow: Cogito Centre. (In Russian).

Bednarik, R., & Shipilov, A. (2011). Gaze cursor during distant collaborative programming: A preliminary analysis. In *Dual Eye Tracking in CSCW, 2011.*

Belenky, D., Ringenberg, M., & Olsen, J. (2014). Using dual eye-tracking to evaluate students' collaboration with an intelligent tutoring system for elementary-level fractions. In *Proceedings of 36th Annual Meeting of the Cognitive Science Society (CogSci 2014)* (pp. 176–181). Quebec City, Canada,: Retrieved from https://mindmodeling.org/cogsci2014/papers/041/paper041.pdf.

Bielikova, M., Konopka, M., Simko, J., Moro, R., Tvarozek, J., Hlavac, P., & Kuric, E. (2018). Eye-tracking en masse: Group user studies, lab infrastructure, and practices. *Journal of Eye Movement Research, 11*(3), 6:1–15. doi:10.16910/JEMR.11.3.6

Brennan, S.E., Chen, X., Dickinson, C.A., Neider, M.B., & Zelinsky, G.J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition, 106*(3), 1465–1477. doi:10.1016/j.cognition.2007.05.012

Broz, F., Lehmann, H., Nehaniv, C.L., & Dautenhahn, K. (2013). Automated analysis of mutual gaze in human conversational pairs. In *Eye gaze in intelligent user interfaces* (pp.41–60). London: Springer. doi:10.1007/978-1-4471-4784-8_4

Burger, B., Puupponen, A., & Jantunen, T. (2018). Synchronizing eye tracking and optical motion capture: How to bring them together. *Journal of Eye Movement Research, 11*(2), 5:1–16.

Carletta, J., Hill, R.L., Nicol, C., Taylor, T., de Ruiter, J.P., & Bard, E.G. (2010). Eyetracking for two-person tasks with manipulation of a virtual world. *Behavior Research Methods, 42*(1), 254–265. doi:10.3758/BRM.42.1.254

De Beugher, S.D., Brône, G., & Goedemé, T. (2014). Automatic analysis of in-the-wild mobile eye-tracking experiments using object, face and person detection. In *Proceedings of the 9th International Conference on Computer Vision Theory and Applications* (pp.625–633). SCITEPRESS — Science and and Technology Publications. doi:10.5220/0004741606250633

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences, 6*(4), 485–507. doi:10.1007/s11097-007-9076-9

Dewhurst, R., Nyström, M., Jarodzka, H., Foulsham, T., Johansson, R., & Holmqvist, K. (2012). It depends on how you look at it: Scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach. *Behavior Research Methods, 44*(4), 1079–1100. doi:10.3758/s13428-012-0212-2

Fedorova, O.V., Kibrik, A.A., & Yazykov, S.A. (2015). Perspektivy ispolzovaniya ochkov-aitrekerov v kognitivnykh issledovaniyakh multimodalnosti [The prospects for the usage of eye-tracking glasses in inverstigation of multimodality]. In V.A. Barabanshchikov (Ed.), *Aitreking v psikhologicheskoi nauke i praktike [Eyetracking in psychological science and practice]* (pp.79–86). Moscow: Cogito Centre. (In Russian).

Flood, V. (2018). Multimodal revoicing as an interactional mechanism for connecting scientific and everyday concepts. *Human Development, 61*(3), 145–173. doi:10.1159/000488693

Gegenfurtner, A., Lehtinen, E., Jarodzka, H., & Säljö, R. (2017). Effects of eye movement modeling examples on adaptive expertise in medical image diagnosis. *Computers & Education, 113*, 212–225. doi:10.1016/j.compedu.2017.06.001

Gergle, D., & Clark, A.T. (2011). See what I'm saying? Using dyadic mobile eye tracking to study collaborative reference. In *Proceedings of the ACM 2011 conference on Computer Supported Cooperative Work — CSCW '11* (pp.435–444). New York: ACM Press. doi:10.1145/1958824.1958892

van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior, 25*(3), 785–791. doi:10.1016/j.chb.2009.02.007

Guo, J., & Feng, G. (2013). How eye gaze feedback changes parent-child joint attention in shared storybook reading? In Y. Nakano, C. Conati, & T. Bader (Eds.), *Eye gaze in intelligent user interfaces. Gaze-based analyses, models and applications* (pp.9–21). London: Springer. doi:10.1007/978-1-4471-4784-8

Hannula, M., Salminen-Saari, J., Garcia Moreno-Esteva, E., Toivanen, M., & Salonen, V. (2017). Tracking visual attention during collaborative problem solving. In B. Kaur, W. Ho, T. Toh, & B. Choy (Eds.), *Proceedings of the 41st Conference of the International Group for the Psychology of Mathematics Education (*Vol.2, p.28). Singapore: PME.

Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and listening with the eyes: Gaze signaling during dyadic interactions. *PLoS ONE, 10*(8), e0136905: 1–18. doi:10.1371/journal.pone.0136905

Hwang, S., & Roth, W.-M. (2011). *Scientific and mathematical bodies. The interface of culture and mind.* Rotterdam: SensePublishers. doi:10.1007/978-94-6091-567-3_2

Jarodzka, H., Van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction, 25*, 62–70. doi:10.1016/j.learninstruc.2012.11.004

Jermann, P., Nüssli, M.-A., & Li, W. (2010). Using dual eye-tracking to unveil coordination and expertise in collaborative Tetris. In *BCS '10 Proceedings of the 24th BCS Interaction Specialist Group Conference* (pp.36–44). British Computer Society.

Kassner, M., Patera, W., & Bulling, A. (2014). Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication — UbiComp '14 Adjunct* (pp.1151–1160). New York: ACM Press. doi:10.1145/2638728.2641695

Kurzhals, K., Hlawatsch, M., Seeger, C., & Weiskopf, D. (2017). Visual analytics for mobile eye tracking. *IEEE Transactions on Visualization and Computer Graphics, 23*(1), 301–310. doi:10.1109/TVCG.2016.2598695

Lilienthal, A., & Schindler, M. (2017). Conducting dual portable eye-tracking in mathematical creativity research. In B. Kaur, W. Ho, T. Toh, & B. Choy (Eds.), *Proceedings of the 41st Conference of the International Group for the Psychology of Mathematics Education (*Vol.1, pp.231). Singapore: PME.

van Marlen, T., van Wermeskerken, M., Jarodzka, H., & van Gog, T. (2016). Showing a model's eye movements in examples does not improve learning of problem-solving tasks. *Computers in Human Behavior, 65*, 448–459. doi:10.1016/j.chb.2016.08.041

Molinari, G. (2015). From learners' concept maps of their similar or complementary prior knowledge to collaborative concept map: Dual eye-tracking and concept map analyses. *Psychologie Francaise, 62*(3), 293–311. doi:10.1016/j.psfr.2015.11.001

Nüssli, M.-A. (2011). Dual eye-tracking methods for the study of remote collaborative problem solving. Unpublished doctoral dissertation, École Polytechnique Fédérale De Lausanne. doi:10.5075/epfl-thesis-5232

Neider, M.B., Chen, X., Dickinson, C.A., Brennan, S.E., & Zelinsky, G.G.J. (2010). Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin and Review, 17*(5), 718–724. doi:10.3758/PBR.17.5.718

Newn, J., Velloso, E., Carter, M., & Vetere, F. (2016). Exploring the effects of gaze awareness on multiplayer gameplay. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts — CHI PLAY Companion '16* (pp.239–244). New York: ACM Press. doi:10.1145/2968120.2987740

Nyström, M., Niehorster, D.C., Cornelissen, T., & Garde, H. (2017). Real-time sharing of gaze data between multiple eye trackers — evaluation, tools, and advice. *Behavior Research Methods, 49*(4), 1310–1322. doi:10.3758/s13428-016-0806-1

Pfeiffer, T., & Renner, P. (2014). EyeSee3D: a low-cost approach for analyzing mobile 3D eye tracking data using computer vision and augmented reality technology. In *Proceedings of the Symposium on Eye Tracking Research and Applications — ETRA '14* (pp.195–202). New York: ACM Press. doi:10.1145/2578153.2578183

Pietinen, S., Bednarik, R., Glotova, T., Tenhunen, V., & Tukiainen, M. (2008). A method to study visual attention aspects of collaboration. In *Proceedings of the 2008 Symposium on Eye tracking Research & Applications — ETRA '08* (p.39). New York: ACM Press. doi:10.1145/1344471.1344480

Renner, P., Pfeiffer, T., & Wachsmuth, I. (2014). Spatial references with gaze and pointing in shared space of humans and robots. In C. Freksa, B. Nebel, M. Hegarty, & T. Barkowsky (Eds.), *Spatial Cognition IX. Spatial Cognition 2014. Lecture notes in computer science* (Vol.8684 LNAI, pp.121–136). Springer, Cham. doi:10.1007/978-3-319-11215-2_9

Richardson, D.C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science, 29*(6), 1045–60. doi:10.1207/s15516709cog0000_29

Richardson, D.C., Dale, R., & Kirkham, N.Z. (2007). The art of conversation is coordination. *Psychological Science, 18*(5), 407–413. doi:10.1111/j.1467-9280.2007.01914.x

Salvucci, D.D., & Goldberg, J.H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the symposium on Eye tracking research & applications — ETRA '00* (pp. 71–78). New York: ACM Press. doi:10.1145/355017.355028

Sandgren, O., Andersson, R., Weijer, J.V.D., Hansson, K., & Sahlén, B. (2012). Timing of gazes in child dialogues: A time-course analysis of requests and back channelling in referential communication. *International Journal of Language and Communication Disorders, 47*(4), 373–383. doi:10.1111/j.1460-6984.2012.00151.x

Saunders, D.R., Melcher, D.P., & van Zoest, W. (2015). A dual eyetracker design for measuring shared attention in dynamic oculomotor selection. *Abstracts of the 18th European Conference on Eye Movements 2015. Journal of Eye Movement Research, 8*(4), 253. Retrieved from https://bop.unibe.ch/JEMR/article/view/2408/3604.

Schindler, M., Lilienthal, A., Chadalavada, R., & Ögren, M. (2016). Creativity in the eye of the student: Refining investigations of mathematical creativity using eye-tracking goggles. In C. Csíkos, A. Rausch, & J. Szitányi (Eds.), *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education (PME)* (Vol. 4, pp. 163–170). Szeged, Hungary.

Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning 8*(4), 375–397. doi:10.1007/s11412-013-9181-4

Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2016a). Detecting collaborative dynamics using mobile eye-trackers. In C.K. Looi, J.L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming learning, empowering learners: The International Conference of the Learning Sciences (ICLS) 2016* (Vol. 1, pp. 522–529). Singapore: International Society of the Learning Sciences.

Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2016b). Using mobile eye-trackers to unpack the perceptual benefits of a tangible user interface for collaborative learning. *ACM Transactions on Computer-Human Interaction, 23*(6), 39–62. doi:10.1145/3012009

Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2018). Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning, 13*(3), 1–21. doi:10.1007/s11412-018-9281-2

Schneider, S., Beege, M., Nebel, S., & Rey, G.D. (2018). A meta-analysis of how signaling affects learning with media. *Educational Research Review, 23*, 1–24. doi:10.1016/J.EDUREV.2017.11.001

Schwarzkopf, S., Büchner, S.J., Hölscher, C., & Konieczny, L. (2017). Perspective tracking in the real world: Gaze angle analysis in a collaborative wayfinding task. *Spatial Cognition & Computation, 17*(1–2), 143–162. doi:10.1080/13875868.2016.1226841

Sharma, K., Caballero, D., Verma, H., Jermann, P., & Dillenbourg, P. (2015). Looking AT versus Looking THROUGH: A dual eye-tracking study in MOOC context. In O. Lindwall, P. Häkkinen, T.T. Koschman, & S.P. Ludvigsen (Eds.), *Exploring the material conditions of learning: the Computer Supported Collaborative Learning (CSCL) Conference 2015* (pp. 260–267). Gothenburg, Sweden: International Society of the Learning Sciences. Retrieved from https://www.isls.org/cscl2015/papers/MC-0250-FullPaper-Sharma.pdf.

Sharma, K., Jermann, P., Nüssli, M.-A., & Dillenbourg, P. (2013). Understanding collaborative program comprehension: Interlacing gaze and dialogues. In N. Rummel, M. Kapur, M. Nathan, & S. Puntambekar (Eds.), *To see the world and a grain of sand: Learning across levels of space, time, and scale: CSCL 2013 Conference Proceedings Volume 1 — Full Papers & Symposia* (pp. 430–437). Madison, WI: International Society of the Learning Sciences.

Shvarts, A. (2018a). A dual eye-tracking study of objectification as student–tutor joint activity appropriation. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.), *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 171–178). Umeå, Sweden: PME.

Shvarts, A. (2018b). Joint attention in resolving the ambiguity of different presentations: a dual eye-tracking study of the teaching-learning process. In N. Presmeg, L. Radford, W.-M. Roth, & G. Kadunz (Eds.), *Signs of signification: Semiotics in mathematics education research* (pp. 73–102). Dordrecht: Springer.

Shvarts, A., & Abrahamson, D. (2018). *Towards a complex systems model of enculturation: A dual eye-tracking study.* Paper presented at the annual conference of the American Educational Research Association (Special Interest Group: Learning Sciences), New York.

Shvarts, A., & Zagorianakos, A. (2016). Theoretical perception of the Cartesian plane: A dual eye-tracking study through double theoretical lenses. In C. Csíkos, A. Rausch, & J. Szitányi (Eds.), *Proceedings of 40th Annual Meeting of the International Group for the Psychology of Mathematics Education (PME 40), 3–7 August 2016* (p. 287). Szeged, Hungary.

Streeck, J., & Mehus, S. (2005). Microethnography: The study of practices. In K.L. Fitch, & R.E. Sanders (Eds.), *Handbook of language and social interaction* (pp. 381–404). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers. doi:10.4324/9781410611574.ch15

Velichkovsky, B., Pomplun, M., & Rieser, J. (1996). Attention and communication: Eye-movement-based research paradigms. *Advances in Psychology, 116*, 125–154. doi:10.1016/S0166-4115(96)80074-4

Yu, C., & Smith, L.B. (2016). Multiple sensory-motor pathways lead to coordinated visual attention. *Cognitive Science, 41*(S1), 1–27. doi:10.1111/cogs.12366

Appendix A. Customizing *DUET for Pupil*

Table A1. Configuration Settings

Name	Default setting	Commentary
base_radius	10	The radius of the point in the current position of the first gaze, in pixels
base_line_width	3	The width of a line that connect the first gaze's samples, in pixels
base_color_red	0	The value of a red channel, first gaze
base_color_green	0	The value of a green channel, first gaze
base_color_blue	255	The value of a blue channel, first gaze
base_inner_radius	2	The radius of the points in the previous positions of the first gaze, in pixels
second_radius	10	The radius of the point in the current position of the second gaze, in pixels
second_line_width	3	The width of a line that connects the second gaze's samples, in pixels
second_color_red	255	The value of a red channel, second gaze
second_color_green	0	The value of a green channel, second gaze
second_color_blue	0	The value of a blue channel, second gaze
second_inner_radius	2	The radius of the points in the previous positions of the second gaze, in pixels
gazes_limit	20	The length of the gaze paths, in gaze samples
base_gaze_adjustment_x	0	Horizontal adjustment of the first gaze, in a fraction of the surface size (positive value is down)
base_gaze_adjustment_y	0	Vertical adjustment of the first gaze, in a fraction of the surface size
second_gaze_adjustment_x	0	Horizontal adjustment of the second gaze, in a fraction of the surface size
second_gaze_adjustment_y	0	Vertical adjustment of the second gaze, in a fraction of the surface size (positive value is down)
with_audio	1	If the audio is present, true (1) / false (0)
decomposition_to_set_of_frames	0	If the video should be split to the frames, true (1) / false (0)
frames_quality	31	The quality of the frames, varies from 1 to 31, 1 is the best quality
waveform_length	3	The duration of a sound waveform's part on one frame, in seconds

Appendix B. Instructions

Preparation:

1. Attach the Markers (https://pupil-labs.com/blog/2013-12/pupil-v0-3-6-marker-tracking/) to the corners and sides of the display (the markers should not cover the corners of the screen to allow the calibration process).
2. Turn on both computers and check if they are plugged in: the recording process is energy consuming and we do not recommend running it on battery power.
3. Check that the computers are connected to the same wireless network.
4. Connect both eye-trackers.
5. Turn on Pupil Capture on both computers. Turn on the required plugins: Time Sync, Pupil Groups, Surface Tracker, Audio Capture.
6. Check if:
   a. the computers can be connected and synchronized (each one appears in the list of the Other Group Members for the other one in the plugin Pupil Groups and the plugin Time Sync displays the synchronization);
   b. there is enough space for the recording on each computer (we recommend 20 Gb), otherwise provide the path to the external drive for the Recorder;
   c. that Audio Capture has access to a microphone; and
   d. all the markers are within the visual scene of the world camera from the two participants' places and they are recognized by the Surface Tracker plugin (add additional light if needed). Then turn off the plugin.

Recording:

1. Put on an eye-tracker that is connected to the computer with the stimuli on a participant and adjust the eye camera: acutance, size of the pupil, its position, intensity range, aperture until the pupil is recognized with confidence (note, it may take several minutes for the software to get used to various angles of the eye-ball). Check that all the markers are within the visual scene the acutance of the display is obtained (if necessary adjust the world camera), and Markers are recognized (turn on and turn off the Surface Tracker plugin).
2. Repeat the same with the second participant.
3. Connect the display to the computer of the second participant and calibrate the participant.
4. Connect the same display to the computer of the first participant and calibrate the participant. Alternatively use Manual Marker Calibration (see https://docs.pupil-labs.com/#calibration for more information).
5. Start the record.
6. Ask the participants to follow your pointer (repeat this from time to time during the record to be able to adjust calibration after the record if needed).

Analysis:

1. Open each of the records using Pupil Player:
   a. add manual corrections if needed;
   b. add a surface of the screen in the plugin Offline Surface Tracker (adjust it to fit the screen as well as possible); and
   c. export the video and data on the surface.
2. Choose the record in which the surface is better recognized. This one will become the base record. Copy the exported folder into the folder with utility DualMerger. The folder should have the name 000.
3. Copy the exported folder from the other tracker into the same folder. The exported folder should have the name 001.
4. Adjust the configuration file is necessary. Run the utility with Python 3.
5. You will receive the video from the surface that is overlaid by the gaze paths of the participant and the .csv file with the synchronized data from two participants.

методы

Запись движений глаз участников мультимодальной совместной деятельности в единых координатах: технология двойного айтрекинга

Анна Юрьевна Шварц
Институт Фройденталя научного и математического образования, Факультет науки, Утрехтский университет, Утрехт, Нидерланды;
Факультет психологии, МГУ имени М.В. Ломоносова, Москва, Россия

Андрей Николаевич Степанов
Департамент программной инженерии, факультет компьютерных наук, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

Дмитрий Валерьевич Чумаченко
Факультет психологии, МГУ имени М.В. Ломоносова, Москва, Россия

Аннотация. В статье приводится анализ преимуществ и недостатков существующих систем двойной записи движений глаз в контексте изучения совместного внимания при мультимодальном сотрудничестве в ситуации обучения. Отмечается несовершенство существующих систем, требующих длительного ручного кодирования для соотнесения движений глаз участников с видеоданными об их жестикуляции и о том, как они совместно выполняют задания. Далее мы описываем систему из двух бюджетных айтрекеров Pupil-Labs и разработанную нами утилиту, отображающую синхронизированные пути движения глаз двух испытуемых в единой системе координат. Движения глаз накладываются на видеозапись поверхности, выделенной в перцептивном поле обоих испытуемых; это видео сопровождается осциллограммой звука. Также координаты синхронизированных движений глаз по поверхности доступны в текстовом формате. Согласно нашей эмпирической оценке, точность данного технологического решения для двойного айтрекинга составляет 1.27 угловых градусов после post-hoc калибровки. В статье обсуждаются преимущества и недостатки данного решения, а также намечаются пути его дальнейшего развития.

Контактная информация: Анна Юрьевна Шварц, shvarts.anna@gmail.com, Utrecht University, Freudenthal Institute, Buys Ballot building, third floor, 3.67, Princetonplein 5, 3584 CC Utrecht, the Netherlands; Андрей Николаевич Степанов, pisaga@yandex.ru; Дмитрий Валерьевич Чумаченко, dmitry.chumachenko@gmail.com.

Ключевые слова: двойной айтрекинг, жестикуляция, мобильные системы для записи движений глаз, мультимодальная коммуникация, совместное внимание, сотрудничество, стационарные системы для записи движений глаз, Pupil-Labs

Благодарности. Мы благодарны Филиппу Барскому за исходную идею двойного айтрекинга, которая запустила длительный процесс поиска оптимального технологического решения. Обсуждения с Анатолием Николаевичем Кричевцом и его поддержка были неоценимы на всех стадиях разработки оборудования и подготовки статьи. Работа выполнена при поддержке РФФИ, грант № 18-013-00906.

Статья поступила в редакцию 17 июня 2018 г. Принята в печать 28 сентября 2018 г.

Литература

Ананьева К.И., Жегалло А.В., Куракова О.А., Харитонов А.Н. Регистрация движений глаз в коммуникативных ситуациях: методические проблемы // Развитие психологии в системе комплексного человекознания. Материалы Всероссийской научной конференции, посвященной 40-летию института психологии и 85-летию его основателя Б.Ф. Ломова / Под ред. А. Л. Журавлева, В. А. Кольцовой. М.: ИП РАН, 2012. С. 141–143.

Ананьева К.И., Харитонов А.Н. Методологические проблемы исследования окуломоторной активности в коммуникативных ситуациях // Психологические и психоаналитические исследования. Ежегодник 2010–2011. М.: Московский институт психоанализа, 2011. С. 37–49.

Басюл И.А., Монгуш А.М., Ананьева К.И., Товуу Н.О., Харитонов А.Н. Решение графической задачи в традиционной и техногенной коммуникативно-когнитивных средах // Эволюционная и сравнительная психология в России. Теория и практика исследований. М.: Когито-центр, 2017. С. 312–321.

Федорова О.В., Кибрик А.А., Языков С.А. Перспективы использования очков-айтрекеров в когнитивных исследованиях мультимодальности // Айтрекинг в психологической науке и практике / Под ред. В.А. Барабанщикова. М.: Московский институт психоанализа, 2015. С. 79–86.

Acarturk C., Tajaddini M., Kilic O. Group Eye Tracking (GET) applications in gaming and decision [abstract] // Abstracts of the 19th European Conference on Eye Movements / R. Radach, H. Deubel, C. Vorstius, M. Hofmann (Eds.). Wuppertal: 2017. Vol. 10. P. 103.

Bednarik R., Shipilov A. Gaze cursor during distant collaborative programming: A preliminary analysis // Dual Eye Tracking in CSCW, 2011.

Belenky D., Ringenberg M., Olsen J. Using dual eye-tracking to evaluate students' collaboration with an intelligent tutoring system for elementary-level fractions // Proceedings of 36th Annual Meeting of the Cognitive Science Society (CogSci 2014). Quebec City, Canada: 2014. P. 176–181. URL: https://mindmodeling.org/cogsci2014/papers/041/paper041.pdf.

Bielikova M., Konopka M., Simko J., Moro R., Tvarozek J., Hlavac P., Kuric E. Eye-tracking en masse: Group user studies, lab infrastructure, and practices // Journal of Eye Movement Research. 2018. Vol. 11. No. 3. P. 6:1–15. doi:10.16910/JEMR.11.3.6

Brennan S.E., Chen X., Dickinson C.A., Neider M.B., Zelinsky G.J. Coordinating cognition: The costs and benefits of shared gaze during collaborative search // Cognition. 2008. Vol. 106. No. 3. P. 1465–1477. doi:10.1016/j.cognition.2007.05.012

Broz F., Lehmann H., Nehaniv C.L., Dautenhahn K. Automated analysis of mutual gaze in human conversational pairs // Eye gaze in intelligent user interfaces London: Springer London, 2013. P. 41–60. doi:10.1007/978-1-4471-4784-8_4

Burger B., Puupponen A., Jantunen T. Synchronizing eye tracking and optical motion capture: How to bring them together // Journal of Eye Movement Research. 2018. Vol. 11. No. 2. P. 5:1–16.

Carletta J., Hill R.L., Nicol C., Taylor T., de Ruiter J.P., Bard E.G. Eyetracking for two-person tasks with manipulation of a virtual world // Behavior Research Methods. 2010. Vol. 42. No. 1. P. 254–265. doi:10.3758/BRM.42.1.254

De Beugher S.D., Brône G., Goedemé T. Automatic analysis of in-the-wild mobile eye-tracking experiments using object, face and person detection // Proceedings of the 9th International Conference on Computer Vision Theory and Applications. SCITEPRESS — Science and and Technology Publications, 2014. P. 625–633. doi:10.5220/0004741606250633

De Jaegher H., Di Paolo E. Participatory sense-making // Phenomenology and the Cognitive Sciences. 2007. Vol. 6. No. 4. P. 485–507. doi:10.1007/s11097-007-9076-9

Dewhurst R., Nyström M., Jarodzka H., Foulsham T., Johansson R., Holmqvist K. It depends on how you look at it: Scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach // Behavior Research Methods. 2012. Vol. 44. No. 4. P. 1079–1100. doi:10.3758/s13428-012-0212-2

Flood V. Multimodal revoicing as an interactional mechanism for connecting scientific and everyday concepts // Human Development. 2018. Vol. 61. No. 3. P. 145–173. doi:10.1159/000488693

Gegenfurtner A., Lehtinen E., Jarodzka H., Säljö R. Effects of eye movement modeling examples on adaptive expertise in medical image diagnosis // Computers & Education. 2017. Vol. 113. P. 212–225. doi:10.1016/j.compedu.2017.06.001

Gergle D., Clark A.T. See what I'm saying? Using dyadic mobile eye tracking to study collaborative reference // Proceedings of the ACM 2011 conference on Computer Supported Cooperative Work — CSCW '11. New York: ACM Press, 2011. P. 435–444. doi:10.1145/1958824.1958892

van Gog T., Jarodzka H., Scheiter K., Gerjets P., Paas F. Attention guidance during example study via the model's eye movements // Computers in Human Behavior. 2009. Vol. 25. No. 3. P. 785–791. doi:10.1016/j.chb.2009.02.007

Guo J., Feng G. How eye gaze feedback changes parent-child joint attention in shared storybook reading? // Eye gaze in intelligent user interfaces. Gaze-based analyses, models and applications / Y. Nakano, C. Conati, T. Bader (Eds.). London: Springer, 2013. P. 9–21. doi:10.1007/978-1-4471-4784-8

Hannula M., Salminen-Saari J., Garcia Moreno-Esteva E., Toivanen M., Salonen V. Tracking visual attention during collaborative problem solving // Proceedings of the 41st Conference of the International Group for the Psychology of Mathematics Education / B. Kaur, W. Ho, T. Toh, B. Choy (Eds.). Singapore: PME, 2017. Vol. 2. P. 28.

Ho S., Foulsham T., Kingstone A. Speaking and listening with the eyes: Gaze signaling during dyadic interactions // PLoS ONE. 2015. Vol. 10. No. 8. P. e0136905: 1–18. doi:10.1371/journal.pone.0136905

Hwang S., Roth W.-M. Scientific and mathematical bodies. The interface of culture and mind. Rotterdam: SensePublishers, 2011. doi:10.1007/978-94-6091-567-3_2

Jarodzka H., Van Gog T., Dorr M., Scheiter K., Gerjets P. Learning to see: Guiding students' attention via a model's eye movements fosters learning // Learning and Instruction. 2013. Vol. 25. P. 62–70. doi:10.1016/j.learninstruc.2012.11.004

Jermann P., Nüssli M.-A., Li W. Using dual eye-tracking to unveil coordination and expertise in collaborative Tetris // BCS '10 Proceedings of the 24th BCS Interaction Specialist Group Conference. British Computer Society, 2010. P. 36–44.

Kassner M., Patera W., Bulling A. Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction // Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication — UbiComp '14 Adjunct. New York: ACM Press, 2014. P. 1151–1160. doi:10.1145/2638728.2641695

Kurzhals K., Hlawatsch M., Seeger C., Weiskopf D. Visual analytics for mobile eye tracking // IEEE Transactions on Visualization and Computer Graphics. 2017. Vol. 23. No. 1. P. 301–310. doi:10.1109/TVCG.2016.2598695

Lilienthal A., Schindler M. Conducting dual portable eye-tracking in mathematical creativity research // Proceedings of the 41st Conference of the International Group for the Psychology of Mathematics Education, Vol. 1 / B. Kaur, W. Ho, T. Toh, B. Choy (Eds.). Singapore: PME, 2017. P. 231.

van Marlen T., van Wermeskerken M., Jarodzka H., van Gog T. Showing a model's eye movements in examples does not improve learning of problem-solving tasks // Computers in Human Behavior. 2016. Vol. 65. P. 448–459. doi:10.1016/j.chb.2016.08.041

Molinari G. From learners' concept maps of their similar or complementary prior knowledge to collaborative concept map: Dual eye-tracking and concept map analyses // Psychologie Francaise. 2015. Vol. 62. No. 3. P. 293–311. doi:10.1016/j.psfr.2015.11.001

Nüssli M.-A. Dual eye-tracking methods for the study of remote collaborative problem solving. PhD dissertation. École Polytechnique Fédérale De Lausanne, 2011. doi:10.5075/epfl-thesis-5232

Neider M.B., Chen X., Dickinson C.A., Brennan S.E., Zelinsky G.G.J. Coordinating spatial referencing using shared gaze // Psychonomic Bulletin and Review. 2010. Vol. 17. No. 5. P. 718–724. doi:10.3758/PBR.17.5.718

Newn J., Velloso E., Carter M., Vetere F. Exploring the effects of gaze awareness on multiplayer gameplay // Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts — CHI PLAY Companion '16. New York: ACM Press, 2016. P. 239–244. doi:10.1145/2968120.2987740

Nyström M., Niehorster D. C., Cornelissen T., Garde H. Real-time sharing of gaze data between multiple eye trackers — evaluation, tools, and advice // Behavior Research Methods. 2017. Vol. 49. No. 4. P. 1310–1322. doi:10.3758/s13428-016-0806-1

Pfeiffer T., Renner P. EyeSee3D: a low-cost approach for analyzing mobile 3D eye tracking data using computer vision and augmented reality technology // Proceedings of the Symposium on Eye Tracking Research and Applications — ETRA '14. New York: ACM Press, 2014. P. 195–202. doi:10.1145/2578153.2578183

Pietinen S., Bednarik R., Glotova T., Tenhunen V., Tukiainen M. A method to study visual attention aspects of collaboration // Proceedings of the 2008 Symposium on Eye tracking Research & Applications — ETRA '08. New York: ACM Press, 2008. P. 39. doi:10.1145/1344471.1344480

Renner P., Pfeiffer T., Wachsmuth I. Spatial references with gaze and pointing in shared space of humans and robots // Spatial Cognition IX. Spatial Cognition 2014. Lecture notes in computer science / C. Freksa, B. Nebel, M. Hegarty, T. Barkowsky (Eds.). Springer, Cham, 2014. Vol. 8684 LNAI. P. 121–136. doi:10.1007/978-3-319-11215-2_9

Richardson D. C., Dale R. Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension // Cognitive Science. 2005. Vol. 29. No. 6. P. 1045–1060. doi:10.1207/s15516709cog0000_29

Richardson D. C., Dale R., Kirkham N. Z. The art of conversation is coordination // Psychological Science. 2007. Vol. 18. No. 5. P. 407–413. doi:10.1111/j.1467-9280.2007.01914.x

Salvucci D. D., Goldberg J. H. Identifying fixations and saccades in eye-tracking protocols // Proceedings of the symposium on Eye tracking research & applications — ETRA '00. New York: ACM Press, 2000. P. 71–78. doi:10.1145/355017.355028

Sandgren O., Andersson R., Weijer J. V. D., Hansson K., Sahlén B. Timing of gazes in child dialogues: A time-course analysis of requests and back channelling in referential communication // International Journal of Language and Communication Disorders. 2012. Vol. 47. No. 4. P. 373–383. doi:10.1111/j.1460-6984.2012.00151.x

Saunders D. R., Melcher D. P., van Zoest W. A dual eyetracker design for measuring shared attention in dynamic oculomotor selection // Abstracts of the 18th European Conference on Eye Movements 2015. Journal of Eye Movement Research. 2015. Vol. 8. No. 4. P. 253. URL: https://bop.unibe.ch/JEMR/article/view/2408/3604.

Schindler M., Lilienthal A., Chadalavada R., Ögren M. Creativity in the eye of the student: Refining investigations of mathematical creativity using eye-tracking goggles // Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education (PME) / C. Csíkos, A. Rausch, J. Szitányi (Eds.). Szeged, Hungary: 2016. Vol. 4. P. 163–170.

Schneider B., Pea R. Real-time mutual gaze perception enhances collaborative learning and collaboration quality // International Journal of Computer-Supported Collaborative Learning. 2013. Vol. 8. No. 4. P. 375–397. doi:10.1007/s11412-013-9181-4

Schneider B., Sharma K., Cuendet S., Zufferey G., Dillenbourg P., Pea R. Detecting collaborative dynamics using mobile eye-trackers // Transforming learning, empowering learners: The International Conference of the Learning Sciences (ICLS) 2016 / C. K. Looi, J. L. Polman, U. Cress, P. Reimann (Eds.). Singapore: International Society of the Learning Sciences, 2016a. Vol. 1. P. 522–529.

Schneider B., Sharma K., Cuendet S., Zufferey G., Dillenbourg P., Pea R. Using mobile eye-trackers to unpack the perceptual benefits of a tangible user interface for collaborative learning // ACM Transactions on Computer-Human Interaction. 2016b. Vol. 23. No. 6. P. 39–62. doi:10.1145/3012009

Schneider B., Sharma K., Cuendet S., Zufferey G., Dillenbourg P., Pea R. Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups // International Journal of Computer-Supported Collaborative Learning. 2018. Vol. 13. No. 3. P. 1–21. doi:10.1007/s11412-018-9281-2

Schneider S., Beege M., Nebel S., Rey G. D. A meta-analysis of how signaling affects learning with media // Educational Research Review. 2018. Vol. 23. P. 1–24. doi:10.1016/J.EDUREV.2017.11.001

Schwarzkopf S., Büchner S. J., Hölscher C., Konieczny L. Perspective tracking in the real world: Gaze angle analysis in a collaborative wayfinding task // Spatial Cognition & Computation. 2017. Vol. 17. No. 1–2. P. 143–162. doi:10.1080/13875868.2016.1226841

Sharma K., Caballero D., Verma H., Jermann P., Dillenbourg P. Looking AT versus Looking THROUGH: A dual eye-tracking study in MOOC context // Exploring the material conditions of learning: the Computer Supported Collaborative Learning (CSCL) Conference 2015 / O. Lindwall, P. Häkkinen, T. T. Koschman, S. P. Ludvigsen (Eds.). Gothenburg, Sweden: International Society of the Learning Sciences, 2015. P. 260–267. URL: https://www.isls.org/cscl2015/papers/MC-0250-FullPaper-Sharma.pdf.

Sharma K., Jermann P., Nüssli M.-A., Dillenbourg P. Understanding collaborative program comprehension: Interlacing gaze and dialogues // To see the world and a grain of sand: Learning across levels of space, time, and scale: CSCL 2013 Conference Proceedings Volume 1 — Full Papers & Symposia / N. Rummel, M. Kapur, M. Nathan, S. Puntambekar (Eds.). Madison, WI: International Society of the Learning Sciences, 2013. P. 430–437.

Shvarts A. A dual eye-tracking study of objectification as student–tutor joint activity appropriation // Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education, Vol. 4 / E. Bergqvist, M. Österholm, C. Granberg, L. Sumpter (Eds.). Umeå, Sweden: PME, 2018a. P. 171–178.

Shvarts A. Joint attention in resolving the ambiguity of different presentations: a dual eye-tracking study of the teaching-learning process // Signs of signification: Semiotics in mathematics education research / N. Presmeg, L. Radford, W.-M. Roth, G. Kadunz (Eds.). Dordrecht: Springer, 2018b. P. 73–102.

Shvarts A., Abrahamson D. Towards a complex systems model of enculturation: A dual eye-tracking study // Paper presented at the annual conference of the American Educational Research Association (Special Interest Group: Learning Sciences). New York, 2018.

Shvarts A., Zagorianakos A. Theoretical perception of the Cartesian plane: A dual eye-tracking study through double theoretical lenses // Proceedings of 40th Annual Meeting of the International Group for the Psychology of Mathematics Education (PME 40), 3–7 August 2016 / C. Csíkos, A. Rausch, J. Szitányi (Eds.). Szeged, Hungary: 2016. P. 287.

Streeck J., Mehus S. Microethnography: The study of practices // Handbook of language and social interaction / K.L. Fitch, R.E. Sanders (Eds.). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 2005. P. 381–404. doi:10.4324/9781410611574.ch15

Velichkovsky B., Pomplun M., Rieser J. Attention and communication: Eye-movement-based research paradigms // Advances in Psychology. 1996. Vol. 116. P. 125–154. doi:10.1016/S0166-4115(96)80074-4

Yu C., Smith L. B. Multiple sensory-motor pathways lead to coordinated visual attention // Cognitive Science. 2016. Vol. 41. No. S1. P. 1–27. doi:10.1111/cogs.12366