H. Kim, T. Di Giacomo, A. Egges, E. Lyard, S. Garchery, N. Magnenat-Thalmann . Believable Virtual Environment: Sensory and Perceptual Believability. Believability in Virtual Environment. December 2004.

Believable Virtual Environment: Sensory and Perceptual Believability

HyungSeok Kim, Thomas DiGiacomo, Arjan Egges, Etienne Lyard, Stephane Garchery, Nadia Magnenat-Thalmann
MIRALab, University of Geneva
{kim|thomas|arjan|etienne|thalmann}@miralab.unige.ch
www.miralab.unige.ch

Abstract

Believability is a term to measure a level of realism in the interactive virtual environment. In this paper, the believability is described in two layered manner. The elementary layer consists of fundamental aspect of the system including immersion, presentation and interaction. These aspects are discussed in uni- and multi-sensory channel including visual, auditory and haptic. In a higher layer, emotion and personality is one of the main issues to provoke believability. The effects of elements are described in perception and generation of motions.

Keywords: Believability, Virtual Environment, Interaction, Perception, Motion Generation

1. Elements of Believability

The major goal of the virtual reality system is to simulate the sensory information such that the participant feels the generated experience as from the real-world. The believability of the virtual environment is a term for a measurement of achieving this goal.

The definition of believability is still an open issue. Zeltzer states that Autonomy, Interactivity, and Presence are important elements for the Virtual Environment [1]. These elements are one of the most essential ones to make the virtual world `realistic' but in terms of believability, a traditional definition of these terms is not sufficient. For the character representation, the believability is often discussed in context of generating behaviors [2]. Believable behavior covers not only realism but also emotions, personality, and intent [3]. We call these additional issues the perceptual believability in comparison to the sensory believability which represents realism in sensory channel. Here we describe the believability in terms of three elements of virtual environment: immersion, presentation, and interaction.

- Immersion

The user can believe that the experience in the virtual world is a real experience if he or she is totally immersed in the virtual environment. Immersion to the virtual environment can be categorized in terms of sensory immersion and perceptual immersion. Sensory immersion is provided by utilizing immersive devices including HMD or CAVE-like systems. Modeling and measuring this kind of immersion has been conducted by utilizing both cognitive surveys and performance indicators [4][5] In other points of view, users are immersed into the virtual world if its semantics are realistic. Sheridan called this element as `the active imagination in suppressing disbelief (and thus enhanced believability)' [6]. The most important element of the semantics is its consistency. The environment should be consistent to the expectation of the participant. Also, the semantics of a virtual environment consists of emotional elements, personalized elements and goal-oriented elements. If a set of objects and stories have these elements, participants believe the represented world. This level of immersion is often called ***presence***.

- Presentation

The believability of the virtual environment can be increased if the virtual world is presented as real as real world. Still it is true that non-realistic experience can give enough immersion and enough believability but if it is combined with realistic presentation, it will increase the believability. The realism in the presentation can be discussed in terms of sensory signal level realism and perceptual level realism.

- Interaction

One of the most important issues in the virtual environment is its interactivity. A realistic interactive system will result in higher believability in normal cases. The sensory feedback should be fast enough from its corresponding action input. Also it should be fast enough to the human visual sensor. In computer gaming environment, it is well known fact that slow visual refresh rate (for example 30 fps) will hurt performance of the gamer compared with fast refresh rate (for example 100 fps). In addition to the fast interactivity, the realism of the interactivity can be determined by its autonomous behavior. Realistic reactive behavior in interactivity is related but different from behaviors to induce perceptual immersion. The interactivity is increased if the behavior responds to actions of users in a life-like way. There many area overlap between in realistic interaction and immersion. Immersion largely depends on how well this is implemented, for example through goal-oriented artificial intelligence or emotional behavior simulation [2][7]. We believe that the perceptual immersion is invoked by goal-oriented intervention of intents, emotions, and personality. The realism of the interaction is defined by the involvement of the user in the virtual environment. For example, factors of presence as defined by Stevens et al. [8] can be re-categorized so that: 1) personal presence, intended aspect of social presence and task factors are components of immersion, 2) unintended aspect of social presence and environmental presence are components of realism in interactivity.

These effects of these elements are not independent. They influence each other in a complex way. In some cases a high level of realism for one area will elaborate the level of believability but if it is combined with a low level of realism on other area, it will decrease the level of believability. Even if the sensory channel has enough realism, it is not sufficient to make the VE believable if the VE does not have believable contents. From another point of view, a VE presented in written text (for example a novel or a book) depending on the quality of the stories.

In the section 2, issues in sensory channel are discussed. Perceptual believability is discussed in the section 3 and emotional issue is discussed in the section 4.

2. Believability and Interfaces

In terms of interface, believability can be discussed for each sensory channel. If each sensory channel can reproduce information with enough believable way, the whole virtual environment can be presented in believable way to the participant.

Among primary human sensory channels visual, auditory and haptic has been major elements in terms of interface. In this section, the believability issues of these three major channels are discussed.

2.1. Realistic visual sensory feedback

The visual sensory channel is a one of the most important channel to make virtual world believable. For example from the early version of movies, it has given successful believable experiences to audiences using mostly visual information only. Visual channel is the most investigated sensory channel in the virtual reality scene. Issues including modeling and re-producing visual information are investigated since the beginning of the computer graphics in 60's. They are started from the modeling and re-producing the virtual world itself and it is evolved to integrating real and virtual world altogether.

In the virtual environment, an immersion is a technical term describing the extent to which the sensory apparatus of the participant is exposed to the synthetic cues generated by the virtual environment. Depending on the types and number of the devices used one can arbitrary identify different types and levels of immersion.

Visual immersion is achieved through the use of shutter glasses, HMD or CAVE-like system. Various levels of the visual immersion are also achieved by adopting software technologies to simulate the visual cues: stereo- or monoscopic, number of colors, resolution, field of view, focal distance, etc. From the very early beginning of the virtual reality technologies, various immersive visual displays are developed. There have been many works to measure 'sense of presence' for difference visual immersion levels. These are measured in terms of distance/depth perception, task performance, and easy of use.

To achieve realism in the presentation, most of work has been done to generate images to have image level realism. The image level realism is defined as a state of realism in image with comparison in the real-image in terms of pixel-wise comparison. Realistic shape modeling and realistic illumination modeling fall into this category. Realistic shape modeling is investigated in various levels including capturing real shape using camera, special sensors including laser scanning and representing special features of the shape such as smoothness of the surface. To achieve more realism, various representations are investigated from parametric surface model, polygon surface model, point-based model, image-based model, and volumetric model. The realism in the presentation is primarily depends on the amount of data such as number of polygons or resolution of images. There have been many attempts to control the amount of data for real-time realistic visualization. There methods try to minimize the degradation of visual realism whilst reducing amount of data to be processed. There are also a set of work to model realistic illumination models. Starting from the simple point-light model, more complex light environment is investigated including modeling area lights, capturing light environments, simulating complex materials. The biggest issues of visual realism in the presentation are again the measurement. Mainly, the shape related geometric measurements, such as length, area, volume and curvature, are used to measure visual realism. Recently, some work has been done to consider human sensory limitation or perceptual issues such as give more detailed model where human visual sensor can perceive its delicate details.

2.2. Reproducing auditory information for immersive virtual environment

The audio is as or even more important than the video. The surrounding sound defines the environment all around you. Again, the problem of the auditory information generation is both on complexity and fidelity in modeling and rendering process. Here, the complexity is resolution of sound and number of sound source.

3D spatial audio in virtual environments is a relatively new and wide research topic, although spatial audio in general has been under investigation since the beginning of the last century. Rendering audible space with preserved three-dimensional sound illusion is called auralization according to Kleiner [9]. Virtual acoustics include virtual reality aspects such as dynamic listener, dynamic source and acoustic environment specificities [10][11][12]. Some fundamental elements already existent are necessary for a complete spatial audio system including transmission, reflections, reverberation, diffraction, refraction and head related transfer. As can be observed, some of these elements are affected by the position of the sound source relative to the listener (or receiver) and others are affected by the environments itself. Several propagation methods are proposed to simulated sound effect from the sound source to the listener [13][14][15]. Most of sound rendering techniques reproduce the sound field for a specific listening point. Binaural and transaural techniques directly attempt to model the sound field at both ears, while techniques based on loudspeaker arrays reconstruct the sound field at the center of the reproduction setup (and usually degrade quickly as the listener moves off-center). Multi-channel panning techniques are simple and efficient, but are more limited in imaging quality than Ambisonic techniques. Wave-field synthesis(WFS) is uniquely able to reconstruct the correct wavefronts everywhere inside the listening region and are thus a true multi-user reproduction system. However the inherent complexity of a WFS setup has, to date, prevented its use in virtual environment systems.

The presentation realism of auditory information can be represented based on subjective observations (exchange of opinions between acousticians and musicians in the case of concert halls); energy decay, clarity and binaural aspects. Based on subjective observations (exchange of opinions between acousticians and musicians in the case of concert halls), a relative set of objective criteria have been designed that describe how energy is distributed in the impulse response:

- Energy decay such as reverberation time, early decay time, etc.
- Clarity that measures a ration of early to late energy in the impulse response
- Binaural linked to our stereophonic perception, which measure the “sensation of space” or “envelopment” perceived by the listener.

Results of simulations can be compared against measurements in real spaces or using scale models by using graph which depends on time and frequencies.

2.3. Haptic devices in terms of believability

Until now, haptic sensory feedback is simulated in limited way. Although there have been discussion and illustration on full body haptic reproduction, for example data suite, the current technological level is still far away from that goal. Currently, most effort is devoted to simulate realistic presentation of haptic. Among the wide range of systems available, one can define classes of devices as; Arm-like devices, Exoskeletons and Tactile devices.

Figure 1 The Arm-like Device [16] : The user grasp a robotic arm with several degrees of freedom that can apply forces.

Figure 2 Example of Exoskeleton [17] : These devices are less disseminated than the previous class, but they offer a much higher level of realism

Figure 3 Tactile display [18]: These devices aren't force feedback but rather tactile feedback devices i.e. they give the sense of touch: shape, texture, temperature.

Arm-like devices: For this kind of devices, several factors will determine the realism of the simulations. First of all, the scenario of the simulation must include a pen like tool for interacting with the virtual world, as the device cannot reproduce any other kind of tool. The haptic device has to be light enough so that it corresponds to the virtual tool that is manipulated. The refresh rate of the simulation has to be extremely high (at least 500 Hz, 1kHz if better) in order to fulfill the high temporal accuracy requirements of our tactile sensors. The haptic device must be able to exert intensive forces to if one wants to simulate rigid objects. If the force isn't strong enough, then the user will fill that the objects' surface is soft.

Exoskeletons: Most of the exoskeletons available nowadays are hand exoskeletons. Indeed, most of the time the user only have contacts with the virtual environment with his hands (and of course his feet). However, these devices only support one feedback per finger, and to experience a believable grasping, these devices should be able to provide one feedback per finger segment, so that complex shaped objects can be rendered accurately. Also, all the problems that occurred for the Arm-like devices are still valid for these ones: weight, accuracy of the force, render rate. For the weight of the device, the problem is even more complex because as the user moves his hand around, he will feel the inertia that is due to the device so except if this is partly handled by a secondary robotic arm (as with the CyberForce), this will contribute to lower the believability of the simulation.

Tactile displays: For this kind of devices, the most important issue is the resolution of the actuator. Indeed, for smooth surfaces, not many pins are necessary for simulating the texture, but for finer surfaces, the number of pins will have to increase with the complexity of the texture to be rendered.

Because of the dynamic aspect of haptics (i.e. without motion the concept of haptics is almost meaningless), the dynamic response of the system will determine its level of believability. For example, one can render realistically textures with a PHANTOM even if this device has only a pen as an interface. However, to gain a higher level of believability, the use of exoskeletons for actually grasping the virtual objects is crucial. Moreover, this kind of interfaces will have to gain more accuracy (by adding extra end-effectors for every finger segment) to be as efficient as they should. Eventually, even if this is a less important aspect of the simulation, adding a tactile actuator to the device help reaching the level of believable simulation..

2.4. Multisensory issues

A virtual environment is an interactive system in which the user manipulates and experiences a synthetic environment through multiple sensory channels. In a multimodal system communication channels are numerous: voice, gesture, gaze, visual, auditory, haptic etc. Integrating these modalities (multimodal inputs) improves the sense of presence and realism and enhances human computer interaction. Virtual environments using sight, sound and touch are quite feasible, but effects of sensory interaction are complex and vary from person to person. Nevertheless adding several communication channels leads to system complexity, cost, and of integration/synchronization problems. Sensory Fusion is a relatively new topic, for which we need to study two kind of human computer communication channels.

In addition to the issues in the uni-sensory channel, the multimodal interface introduces more complex situation to invoke believability of the environment.

Among issues of the believability of virtual environment, one of the key issues in the multimodal interface is synchronization of sensory channel. Some anomalies such as motion sickness are appeared when synchronization of sensory channel is not well provided. The basic question on this is the tendency of the believability to the synchronization level.

The rate of sensory feedback is another issue that affects believability of the interface. It is known that the visual sensory channel should provide feedback at a speed of around 60 frames per second to give enough realism. But when it combined with action channel sometimes speed of 100 frames per second is not enough in some cases. The question should be answered is identifying effective elements of sensory channel that affects to the other channel.

The most difficult but interesting issue is the

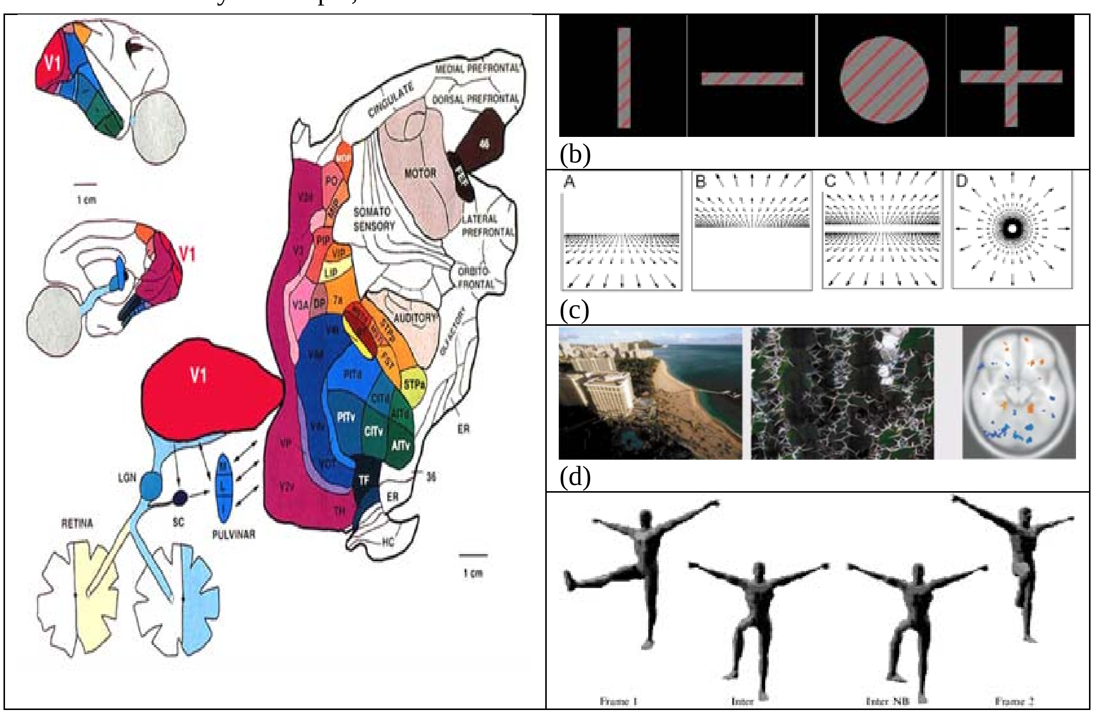

Figure 4 (a) Visual areas of the brain. (b) Four apertures. (c) Four optic flows. (d) Relationships between scene preferences and brain activity, measured by fMRI [27]. (e) Left and right images are the two prime views while middle images are views consistent or not with biomechanical constraints [28].

difference in the level of realism among sensory channels. Even if one sensory channel could provide very realistic informational feedback, other channels with low realism could prevent to create high level of believability. In other cases, there have been a set of research that one sensory channel could replace other sensory channel. It may suggest that there might be an discrimination on the differences of the realism between sensory channels. The question on this issue is discriminating the possible threshold where the higher realism in a specific sensory channel is desired.

3. Animation Believability with Motion Perception

In this section, we discuss psycho-physical approaches of motion perception by the human visual system, and its potential impact on believability for Computer Graphics and Animation. After basic neurophysiologic mechanisms of perception and low-level motion detection and discrimination, higher-level concepts such as motion memory and learning, observer attention, are explored in the context of believability for Graphics.

3.1. Motion Detection

Three complementary approaches coexist in perception [19]: computational, psychophysical and neurophysiologic. Physiologically, the detection and analysis of motions are processed through a cascade of neural operations located in different areas of the brain (**Figure 4**): after registering local motions, interconnected neuron areas communicate to merge local signals into global descriptions of motion. Important terms from the motion perception literature are:

Aperture: Similarly to receptive fields, it is an opening within which neurons register motion. It can be seen as a spatially restricted window or a viewpoint in graphics.

Stimulus: From [20], "The complete definition of the stimulus to a given response involves the specification of all the transformations of the environment that leave the response invariant". When ensuring these invariant conditions, stimuli such as Gabor patterns (sinusoidal gratings and Gaussian functions), moving lines or dots, or computer animations, allow the understanding of response processing.

Optic Flow: It is a continuous sequence of images perceived by humans due to spatio-temporal changes. It provides information on shapes, distances, velocities, *etc.* and drives path-finding with collision awareness and avoidance, as well as depth segregation.

Additional results are provided in [21] and in other work. The lower limit for detection, *i.e.* motion acuity, varies with the number of objects. The detection of relative and biological motions, such as human gait, is far better than absolute or random motions. Motion detection varies with the size, exposure duration of the stimuli, but not with the direction. Exposures to coherent motion, *i.e.* motion with a general direction, also increases the motion discrimination, *i.e.* distinction between trajectories. The constant properties of velocity perception has been studied in [22], the smoothing of motion perception in [23],

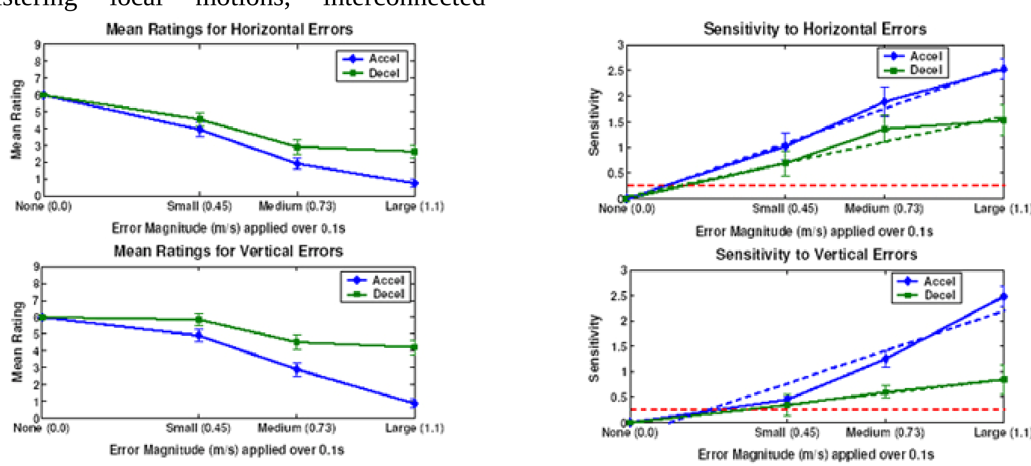

Figure 5 Mean ratings and sensitivity for errors perceived in human jumping animation sequences [36].

and it has been shown that motions based on translation are predictable by human brains. Most of the time, objects are subjected to different combinations of transformations, series of translations and rotations. [24] study the perception of objects rotating and translating, with lines as stimuli. They conclude that motion perception of rotation and translation is largely independent of the aperture shape configuration.

3.2. High-Level Perception and Believability

Memory for motion is a relatively new topic of research and though work has illustrated robustness for memory of velocity over 30 seconds or coherence in memory for direction, many questions still remains. For instance, though high level features have an impact on motion perception, this impact is minor and under restricted conditions of information semantics: [25] shows that semantic knowledge of objects is influencing motion correspondence. Other senses and information are as well modifying visual and motion perception, such as sound for instance, as detailed in [26]. Related to attention and interest, preferences for scenes are explored in [27]. Based on the assumption that we do not look at random spots, their experiments explore our internal mechanisms for the selection of regions of interest. Research has been conducted on recognition of human motion such as arm lifting or gait. For instance, the genre and the individuality are potentially recognized with gait as stated in [28]. Though it is highly articulated and deformable, the human body movements are very easily discriminated by the visual system of human observers. Following experiments on the integration of human multiple views in motion, [29] suggest that human movement perception might be based on the body biomechanical limitations, and potentially confirm that the perception of motion and the object-recognition process are closely linked. [30] have explored the recognition of human motion, with computer animated sequences as stimuli. Applied perception is often used as a criterion for level of details methods and therefore to impact on the representation believability. Early work of [31] proposes adaptive algorithms for visualization at stable frame rates, according to the size, focus, motion blur and semantics of perceived objects. [32] use perception for the believability of real-time collision detection. For improving believability, perception is also a mechanism to validate the level of complexity of physically-based animation, as explored in [33] and later in [34]. Generally, [35] detail some perception-based metrics for walkthroughs in credible virtual environments while [36] propose such an approach for believability of virtual human animation by the detection of non-plausible and plausible motions (Figure 5).

4. ECA, Emotion and Personality Simulation

The believability of the sensory input should be considered with the information contained in the virtual environment. Even if the sensory channel gives enough realism, it is not sufficient to make the world believable if the world does not give believable contents. In other points of view, some world presented in text-written novels in the form of the book can be believable with well composed stories.

Nowadays, a lot of interest from both industry and research exists for Virtual Environments (VEs) and Embodied Conversational Agents (ECAs). A lot of new techniques are being developed to improve the simulation in general, add more visual detail and make the interaction between human and VE/ECA more natural. *Believability* is a measure to help to determine how well these different techniques are *working*. Believability represents the ‘outsider’ point of view. It is thus a very powerful evaluation tool, since we can use it to make evaluations of different techniques/methods while the evaluations are independent from the underlying techniques. This allows us to compare different approaches and give a meaningful indication of their quality on the level of believability. In this section we will focus especially on believability and ECAs. Since ECAs are generally modeled after humans (even cartoon characters), one important aspect of their believability is how well an ECA succeeds in being like a human. We believe that the key to believable ECAs is the definition of their **personality** and their **emotions**. Although quite some research has been done to describe the influence of emotion and personality on ECAs, the results until now are not very convincing. We see several reasons for this:

- Psychological models of emotion/personality are not finalized. The exact structure of our emotions/personality is not certain, as well as the way in which emotions and personality interact with our perception, behavior and expression.
- When one wants to simulate emotions/personality computationally, one tends to take the model the most suitable for a computational simulation. However, this model is not necessarily the best representation for emotions/personality.
- Even if there exists a perfect emotion/personality model, it is very difficult to distinguish the resulting behavior from the emotion that is behind it. Also, other issues interfere with our impression of how emotional an ECA really is, such as its appearance, the surroundings, its capabilities, and so on.

In this section, we will attempt to give some ideas how believability of ECAs can be increased. We will especially focus on the *expressiveness* of ECAs. An important control mechanism for ECAs is a personality/emotion simulator. Personality and emotions have a significant effect on how one perceives, thinks and acts (Figure 6). In this section, we will give a short overview of the different existing techniques for including emotions into perception and reasoning. After this section, we will give some examples of how personality and emotion can play a role in **expression**.

4.1. Emotion, Personality and Perception

There are different scenarios that describe how an emotion is evoked from the perception of one or more events (see Figure 2 for an overview). The process of inducing an emotional response from perceptive data is called **appraisal**. One of the oldest theories, the James-Lange theory of emotion states that an event causes arousal first and only after our interpretation of the arousal, we experience an emotion. For example: You are approaching your house and you notice that the door has been forced open. You begin to tremble and your heart starts beating faster. You interpret these physiological changes as being part of fear. You then experience fear. The Cannon-Bard theory of emotion [37] states that emotion and the physiological response happen at the same time and unrelated from each-other. For example: You are approaching your house and you notice that the door has been forced open. You begin to tremble and your heart starts beating faster. At the same time you experience fear.

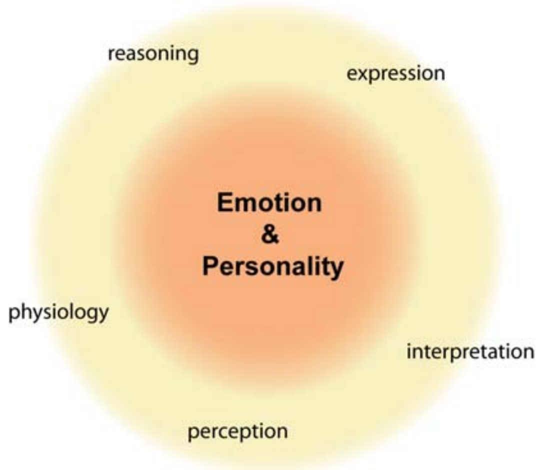

Figure 6 Personality and emotion and their link with ECA system parts.

The Schachter-Singer scenario [38] says that an event causes arousal, but that the emotion follows from the identification of a reason for the arousal. For example: You are approaching your house and you notice that the door has been forced open. You begin to tremble and your heart starts beating faster. You realize that there might be a burglar in your house, which is a dangerous situation. Therefore you experience fear.

The Lazarus theory of cognitive emotion [39] states that both arousal and emotion are invoked separately by a thought following an event. For example: You are approaching your house and you notice that the door has been forced open. You realize that there might be a burglar in your house, which is a dangerous situation. You begin to tremble and your heart starts beating faster. At the same time you experience fear.

Finally, the Facial Feedback hypothesis [40][41] says that emotion is the experience of changes in the facial muscle configuration. This result has also been shown by Ekman et al.[42]. For example: You are approaching your house and you notice that the door has been forced open. Your eyes widen and your mouth corners move backwards. You interpret this facial expression as fear. Therefore you experience fear.

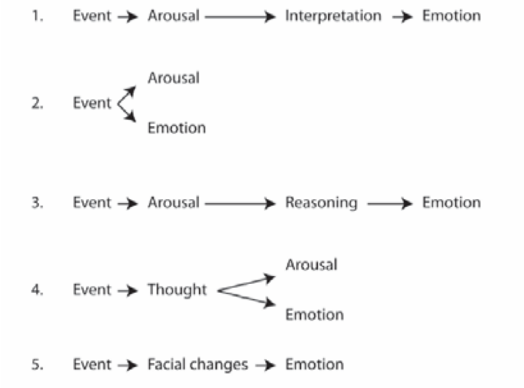

Figure 7 Five scenarios to describe the path from event to emotion: (1) James-Lange (2) Cannon-Bard (3) Schachter-Singer (4) Lazarus (5) Facial Feedback.

In emotion simulation research so far, appraisal is popularly done by a system based on the OCC model [43]. This model specifies how events, agents and objects from the universe are used to elicit an emotional response depending on a set of parameters: the goals, standards and attitudes of the subject. Since the emotional response is generated from a cognitive point of view, this type of appraisal is called cognitive appraisal and it corresponds closely with Lazarus' emotion theory (not taking into account the physiological response). When one wants to develop a computational model of appraisal, not all of the above mentioned scenarios are *suitable* to take as a basis, especially those scenarios where arousal plays a crucial role in the determination of the emotional response (ECAs do not yet have a physiology). This rises the question if it is possible to develop a computational model of appraisal that has a high believability. On the level of personality, one could consider the goals, standards and attitudes of the OCC model as a domain dependent 'personality'. However, personality can also be modeled in a more abstract, domain-independent way [44][45]. Egges et al. [46] discusses how a link between multidimensional personality models and the OCC appraisal model can be established.

4.2. Emotion, Personality and Reasoning

The effect of personality and emotion on agent behavior has been researched quite a lot [47], whether it concerns a general influence on behavior [48], or a more traditional planning-based method [49]. Also, rule-based models [50], probabilistic models [51][52] and fuzzy logic systems [53] have been developed. In the case of real human beings there are still many questions regarding how emotion influences our behavior, but in the field of Neuroscience, work has been done that partly describes the relationship between emotions and the brain [54][55].

4.3. Emotion, Personality and Expression

The expression of emotions has been widely researched, and the most well known research is the work done by Ekman [56]. Not only will personality and emotion have an effect on expressions by the face or body; also physiological changes can be measured according to different emotions. Furthermore, emotions and personality have an important effect on speech [57]. In the following two sections, we will concentrate on the relationship between emotions, personality and face/body animation. Also, we will give some examples on how to improve the believability of an ECA using emotions and personality.

4.4. Believable Facial Animation

When communicating with an ECA, the dialogue itself is only a small part of the interaction that is actually going on. In order to simulate human behavior, all the non-verbal elements of interaction should be taken into account. An ECA can be defined by the following parts:

- Appearance (face model, age, race, etc.)
- Behavior (choice of non-verbal behavior, accompanying speech)
- Expressiveness of movement (amplitude, tempo, etc.)

Other information can also be important like the cultural background or the context.

Facial animation synchronized with speech can be improved by different factors such as non-verbal actions, speech intonation, facial expression consistent with the speech and the context, and also facial expressions between speech sequences. All this information helps to increase the believability of facial animation for ECAs. The main problem is to determine

when and how this kind of non-verbal behavior should be expressed.
Finally, one of the most important points for increasing believability of facial and body animation is the synchronization between verbal and non-verbal expressions [58][59]. The following types of non-verbal behaviors have a notable influence on the believability of ECAs:

• Gaze: Eyes and head movement play an important role in non-verbal communication. There are rules that describe how eye and head movements are related to the action that is performed. A lot of study has be done in this field [60] that proves the importance of gaze in the communication process. Smid and al. [59] has studied recorded sequences of real speaker for building a statistical model. This study reveals the importance of head motions during speech.

• Eyebrows: Eyebrow movements are be very important because specific movements during speech are made to stress parts of the contents. Also, eyebrow movements are used in emotions and other expressions, such as uncertainty.

• Expression dynamics: Finally, facial expression timing and dynamics contain a lot of information. Facial expression dynamics change depending on the emotion or personality. In order to synthesize facial motions, we use a facial animation technique based on the MPEG-4 standard [61]. For defining the visemes and expressions, we use the technique described by Kshirsagar et al. [62]. Here, a statistical analysis of the facial motion data reflects independent facial movements observed during fluent speech. The resulting high level parameters are used for defining the facial expressions and visemes. This facilitates realistic speech animation, especially blended with various facial expressions. In order to generate believable facial animation, the following steps are taken:

• Generation of speech animation from text: a text-to-speech (TTS) software provides phonemes with temporal information. Then co-articulation rules are applied [63].

• Expression blending: proper expressions are selected according to the content of the speech. Each expression is associated with an intensity value. An attack sustain decay- release type of envelope is applied for the expressions and it is blended with the previously calculated co-articulated phoneme trajectories. This blending is based on observed facial dynamics, incorporating the constraints on facial movements wherever necessary in order to avoid excessive/unrealistic deformations (Figure 8).

• Periodic facial movements and gaze: Periodic eye-blinks and minor head movements are applied to the face for increased believability.

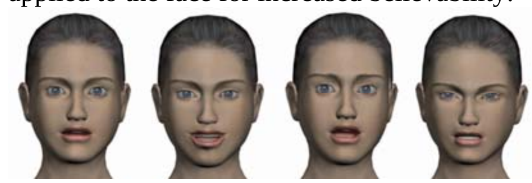

Figure 8 Some examples of facial expressions mixed with speech.

5. Conclusion

We have presented in this paper a discussion related to believability and virtual environment and specially in terms of Embodied Conversational Agents. A clear description are presented in section 1, presented the different elements of believable environment like immersion, presentation and interaction. We have also presented interfaces related to believability and specially visual sensory feedback, audio and haptic devices. Personality and emotions should be part of any ECA simulation system. Psychological models of the human mind can help us to determine how we should proceed to include personality and emotions in our ECAs. This aspect are developed in section 3.
The final implementation of believable virtual world may or may not resemble the human personality/emotion system, but from the point of view of believability this does not matter since the evaluation is independent of underlying technology. In that sense, believability can be seen as the psychology for ECAs.

6. Acknowledgements

This work is partially supported by the European Project ENACTIVE. Thanks to the all the partners of the project ENACTIVE especially, to Annie Luciani and John Stewart who had intensive discussion with authors on the issues of the believability.

7. References

[1] D. Zeltzer. Autonomy, interaction, and presence. Presence: Teleoper. Virtual Environ., 1(1):127–132, 1992.

[2] J. C. Lester and B. A. Stone. Increasing believability in animated pedagogical agents. In Proceedings of the first international conference on Autonomous agents, pages 16–21. ACM Press, 1997.

[3] P. Doyle. Believability through context using ”knowledge in the world” to create intelligent characters. In Proceedings of the first international joint conference on Autonomous agents and multiagent systems, pages 342–349. ACM Press, 2002.

[4] R. Pausch, D. Proffitt, and G. Williams. Quantifying immersion in virtual reality. In Proceedings of the 24th annual conference on Computer graphics and interactive techniques, pages 13–18. ACM Press/Addison-Wesley Publishing Co., 1997.

[5] G. Robertson, M. Czerwinski, and M. van Dantzich. Immersion in desktop virtual reality. In Proceedings of the 10th annual ACM symposium on User interface software and technology, pages 11–19. ACM Press, 1997.

[6] T. B. Sheridan. Interaction, imagination and immersion some research needs. In Proceedings of the ACM symposium on Virtual reality software and technology, pages 1–7. ACM Press, 2000.

[7] C. Martinho and A. Paiva. Pathematic agents: rapid development of believable emotional agents in intelligent virtual environments. In Proceedings of the third annual conference on Autonomous Agents, pages 1–8. ACM Press, 1999.

[8] B. Stevens, J. Jerrams-Smith, D. Heathcote, and D. Callear. Putting the virtual into reality: assessing object-presence with projection-augmented models. Presence: Teleoper. Virtual Environ., 11(1):79–92, 2002.

[9] M. Kleiner, B.-I. Dalenbck, and P. Svensson. Auralization - an overview. Journal of the Audio Engineering Society, 41:861–875, 1993.

[10] T. Takala, R. Hnninen, V. Vlimki, L. Savioja, J. Huopaniemi, T. Huotilainen, and M. Karjalainen. An integrated system for virtual audio reality. In 100th Convention of the Audio Engineering Society, 1996.

[11] L. Savioja, J. Huopaniemi, T. Lokki, and R. Vnnen. Creating interactive virtual acoustic environments. Journal of the Audio Engineering Society, 47(9):675–705, 1999.

[12] T. Funkhouser, P. Min, and I. Carlbom. Real-time acoustics modeling for distributed virtual environments. In SIGGRAPH 1999 Conference Proceedings, 1999.

[13] J. Borish. Extension of the image model to arbitrary polyhedra. Journal of Acoustics of America, 75:1827–1836, 1984.

[14] A. Kulowski. Algorithmic representation of the ray tracing technique. Applied. Acoustics, 18:449–469, 1984.

[15] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J. West, G. Pingali, P. Min, and A. Ngan. A beam tracing method for interactive architectural acoustics. Journal of the Acoustical Society of America, 2004.

[16] [PHANTOM] The Phantom webpage: http://www.sensable.com/products/phantom_ghost/phantom.asp (This is the root link to all the Phantom product line, from the cheapest one (the Phantom Omni Device) to the most expensive one (the Phantom Premium 6DOF Device))

[17] [FORCE] The CyberForce webpage at the Immersion website: http://www.immersion.com/3d/products/cyber_force.php

[18] [TOUCH] The CyberTouch webpage at the Immersion website: http://www.immersion.com/3d/products/cyber_touch.php (On this webpage, you will all the information concerning the CyberTouch device.)

[19] R. Sekuler, S. Watamaniuk, R. Blake. Perception of Visual Motion. In Stevens' Handbook of Experimental Psychology, Sensation and perception, Wiley & Sons, 2002.

[20] S. Stevens. Mathematics, Measurement and Psychophysics. In Handbook of Experimental Psychology, Wiley & Sons, 1951.

[21] E. Adelson. Mechanisms for Motion Perception. Journal of Optics and Photonics News, 2(8):24-30, 1991.

[22] H. Distler, K. Gegenfurtner, H. VanVeen, M. Hawken. Velocity constancy in a virtual reality environment. Perception, 29(12):1423-1435, 2000.

[23] R. Rao, D. Eagleman, T. Sejnowski. Optimal Smoothing in Visual Motion Perception. Neural Computation, 13(6):1243-1253, 2001.

[24] Z. Yang, A. Shimpi, D. Purves. Perception of objects that are translating and rotating. Perception, 31(5):925-942, 2002.

[25] K. Yu. Can semantic knowledge influence motion correspondence? Perception, 29(6):693-707, 2000.

[26] R. Sekuler, A. Sekuler, R. Lau. Sound alters visual motion perception. Nature, 385:308, 1997.

[27] E. Vessel, I. Biederman. Why do we prefer looking at some scenes rather than others? OPAM, 2001.

[28] N. Troje. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. Journal of Vision, 2(5):371-387, 2002.

[29] Z. Kourtzi, M. Shiffrar. Dynamic representations of human body movement. Perception, 28(1):49-62, 1999.

[30] J. Hodgins, J. O'Brien, J. Tumblin. Perception of Human Motion with Different Geometric Models. IEEE Transactions on Visualization and Computer Graphics, 4(4), 1998.

[31] T. Funkhouser, C. Sequin. Adaptive Display Algorithm for Interactive Frame Rates During Visualization of Complex Virtual Environments. SIGGRAPH 1993.

[32] C. O'Sullivan, J. Dingliana. Collisions and Perception. ACM Transactions on Graphics, 20(3), 2001.

[33] R. Barzel, J. Hughes, D. Wood. Plausible Motion Simulation for Computer Graphics. Eurographics Workshop on Computer Animation and Simulation, pp.183-197, 1996.

[34] C. O'Sullivan, J. Dingliana, T. Giang, M.K. Kaiser. Evaluating the Visual Fidelity of Physically Based Animations. ACM Transactions on Graphics, 22(3), 2003.

[35] K. Myszkowski. Perception-based global illumination, rendering, and animation techniques. Spring Conference on Computer Graphics, pp.13-24, 2002.

[36] P. Reitsma, N. Pollard. Perceptual Metrics for Character Animation: Sensitivity to Errors in Ballistic Motion. SIGGRAPH 2003.

[37] W. B. Cannon. The james-lange theory of emotion: A critical examination and an alternative theory. American Journal of Psychology, 39:10–124, 1927.

[38] S. Schachter and J. Singer. Cognitive, social and physiological determinants of emotional state. Psychol. Rev., 69:379–399, 1962.

[39] R. S. Lazarus. Emotion and Adaptation. Oxford University Press, New York, 1991.

[40] R. Buck. Noverbal behavior and the theory of emotion: The facial feedback hypothesis. Journal of Personality and Social Psychology, 38:811–824, 1980.

[41] J. Lanzetta, J. Cartwright-Smith, and R. Kleck. Effects of nonverbal dissimulation on emotion experience and autonomic arousal. Journal of Personality and Social Psychology, 33:354–370, 1976.

[42] P. Ekman, R. W. Levenson, and W. V. Friesen. Autonomic nervous system activity distinguishes among emotions. Science, 221:1208–1210, 1983.

[43] A. Ortony, G. L. Clore, and A. Collins. The Cognitive Structure of Emotions. Cambridge University Press, 1988.

[44] H. J. Eysenck. Biological dimensions of personality. In L. A. Pervin, editor, Handbook of personality: Theory and research, pages 244–276. New York: Guilford, 1990.

[45] P. T. Costa and R. R. McCrae. Normal personality assessment in clinical practice: The NEO personality inventory. Psychological Assessment, (4):5–13, 1992.

[46] A. Egges, S. Kshirsagar, and N. Magnenat-Thalmann. Generic personality and emotion simulation for conversational agents. Computer Animation and Virtual Worlds, 15(1):1–13, 2004.

[47] P. Piwek. An annotated bibliography of affective natural language generation.

Technical report, University of Brighton, July 2002.
[48] S. Marsella and J. Gratch. A step towards irrationality: Using emotion to change belief. In Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems, Bologna, Italy, July 2002.
[49] M. Johns and B. G. Silverman. How emotions and personality effect the utility of alternative decisions: a terrorist target selection case study. In Tenth Conference On Computer Generated Forces and Behavioral Representation, May 2001.
[50] E. Andr´e, M. Klesen, P. Gebhard, S. Allen, and T. Rist. Integrating models of personality and emotions into lifelike characters. In Proceedings International Workshop on Affect in Interactions. Towards a New Generation of Interfaces, 1999.
[51] G. Ball and J. Breese. Emotion and personality in a conversational character. In Proceedings of the Workshop on Embodied Conversational Characters, pages 83–84 and 119–121, October 1998.
[52] L. Chittaro and M. Serra. Behavioural programming of autonomous characters based on probabilistic automata and personality. Computer Animation and Virtual Worlds, 15(3–4):319–326, 2004.
[53] M. El-Nasr, T. Ioerger, and J. Yen. A pet with evolving emotional intelligence. In Proceedings of Autonomous Agents99, 1999.
[54] J. E. LeDoux and W. Hirst. Mind and Brain: Dialogues in Cognitive Neuroscience. Cambridge University Press, 1986.
[55] R. J. Davidson, P. Ekman, C. D. Saron, J. A. Senulis, and W. V. Friesen. Approach-withdrawal and cerebral asymmetry: emotional expression and brain physiology. Journal of Personality and Social Psychology, 58:330–341, 1990.
[56] P. Ekman. Approaches to emotion, chapter Expression and the nature of emotion, pages 319–343. Lawrence Erlbaum, Hillsdale, N. J., 1984.
[57] K. R. Scherer. Music, language, speech, and brain, chapter Emotion expression in speech and music, pages 146–156. MacMillan, London, 1991.
[58] C. Pelachaud and M. Bilvi. Computational Model of Believable Conversational Agents, chapter Communication in MAS: background, current trends and futur. Marc-Philippe Huget, Springer-Verlag, 2003.
[59] C. Smid, I. Pandzic, and V. Radman. Autonomous speaker agent. In CASA 2004, 17th International Conference on computer nimation and Agent, pages 259–266, 2004.
[60] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse. The impact of avatar realism and eye gaze control on the perceived quality of communication in a shared immersive virtual environment. In SIGCHI, 2003.
[61] S. Kshirsagar, S. Garchery, and N. Magnenat-Thalmann. Deformable Avatars, chapter Feature Point Based Mesh Deformation Applied to MPEG-4 Facial Animation, pages 33–43. Kluwer Academic Publishers, 2001.
[62] S. Kshirsagar, T. Molet, and N. Magnenat-Thalmann. Principal components of expressive speech animation. In Computer Graphics International, pages 59–69, 2001.
[63] M. M. Cohen and D. Massaro. Modelling co-articulation in synthetic visual speech. Springer-Verlag, 1993.