# Audiovisual Perception in a Virtual World: An Application of Human-Computer Interaction Evaluation to the Development of Immersive Environments

**Carlos C. L. Silva**
Departamento de Informática, Universidade do Minho & HASLab / INESC TEC
Campus de Gualtar, 4710-057 – Braga
carlos.cl.silva@gmail.com

## ABSTRACT

Understanding the mechanisms underlying audiovisual perception is crucial for the development of interactive audiovisual immersive environments. Some human perceptual mechanisms pose challenging problems that can now be better explored with the latest technology in computer-generated environments. Our main goal is to develop an interactive audiovisual immersive system that provides to its users a highly immersive and perceptually coherent interactive environment. In order to do this, we will perform user studies to get a better knowledge of the rules guiding audiovisual perception. This will allow improvements in the simulation of realistic virtual environments through the use of predictive human cognition models as guides for the development of an audiovisual interactive immersive system. This system will encompass the integration of two Virtual Reality systems: a Cave Automatic Virtual Environment-like (CAVE-like) system and a room acoustic modeling and auralization system. The interactivity between user and the audiovisual virtual world will be enabled by the using of a Motion Capture system as a user position tracker.

## Author Keywords

Immersive Systems; User Interfaces; Audiovisual Perception; Predictive Human Cognitive Models.

## ACM Classification Keywords

H.1.2 **[User/Machine Systems]**: Human factors, Human information processing, Software Psychology; H.5.1. **[Multimedia Information Systems]**: Artificial, augmented, and virtual realities; H.5.2 **[User Interfaces]**: User-centered design.

## INTRODUCTION

Because of its inherent potential to directly interact with the human senses, immersive environments that make use of Virtual Reality (VR) or Augmented Reality (AR) have long being regarded as in line to become the next predominant human-computer interface [1]. However, in order to turn immersive environments into a serious candidate for the next predominant interface paradigm some current technological limitations have to be overcome and, additionally, developmental approaches more focused on human perception and action on immersive environments should be pursued. Accurate predictions about how users perceive and interact with a computerized environment are of foremost importance in the engineering of computer systems that emphasize usefulness and usability, as any attempt of developing an interactive system should put the human, the *user*, in a central position that defines all the subsequent discussion and design [2].

Computer generated immersive environments are normally classified into two different categories: VR environments and AR environments. The distinction between these two is not a procedural or technical one; rather it is more of a performance-based distinction. In the *reality-virtually continuum* of Milgram and Colquhoun [3] the fundamental distinction is between *Real Environments* and *Virtual Environments*, that are located on the continuum opposite ends (see Figure 1). The positioning of any immersive environment along this continuum coincides with its position along a parallel *Extent of World Knowledge continuum*. Hence, this definition highlights the importance of knowledge about both the physical world and the human mechanisms that allow the perception of these physical signals in order to develop satisfactory immersive systems.

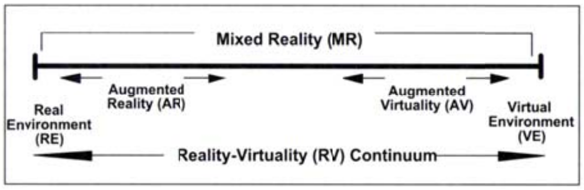

**Figure 1.** Reality-Virtuality Continuum (from Milgram & Colquhoun, 1999)

When we talk about immersive environments we are addressing any kind of environments that are capable of creating on users' the illusion of being in a place other than where they actually are, or of having a coherent interaction with objects that do not exist in the real world. In other words, we are alluding to all the software and hardware elements, needed to present stimuli to the users' senses, which will elicit this kind of effect in the user – normally referred to as *feeling of presence* [4]. There are some features that an immersive system should have and that are

*EICS'13*, June 24–27, 2013, London, United Kingdom.
ACM 978-1-4503-2138-9/13/06.

positively correlated with the capability of conveying an adequate feeling of presence: 1) *Unnoticeable hardware*; 2) *Real-time update* of the immersive environment to the user's position; 3) *Multimodal stimulation*; and 4) *User environment interactivity*.

Subjacent to all of these features should be the most important principle in the engineering of immersive environments: The environment should convey an accurately replication of the geometric and temporal characteristic of the real world. This does not mean that computerized environments have to model the exact physical characteristics of a visual or an auditory real world's scenario – in fact this is, and is expectable it will remain being, technologically impracticable. What this principle really means is that the perception that a user has in a VR or AR environment, should be quantitatively indistinguishable of a correspondent scene perception in the real world.

Applying HCI techniques to the development and evaluation of immersive environments becomes an issue of the foremost importance, primarily when we think about the current lack of knowledge on some human perceptual mechanisms that are central to a proper interaction with the natural tridimensional world. The use of HCI research techniques, mostly focused in psychophysical experimentation, can both give us some insight about the human perceptual mechanisms that should be crucial in the engineering of a highly immersive VR or AR environment and, at the same time, allow us to quantitatively evaluate the human performance of an environment user, making possible comparisons with performance in real world situations.

This comparative evaluation had widespread use in the development of Predictive Cognitive Models that have boosted HCI influence in the design of computerized solutions (e.g. Fitt's law [5], Model Human Processor [6], EPIC [7]) and the same should happen in the development of immersive environments. Psychophysical experimentation can be used as a tool in HCI studies, allowing us to construct models of human perception and human performance capable of guiding the development process of a highly immersive and interactive VR or AR system.

## OUR GOALS

The project here presented has the generic goal of developing Human Predictive Cognitive Models to guide and evaluate the development of an interactive audiovisual immersive system. Our technological output will be the accurate implementation of aural-visual interaction in a multimodal and interactive immersive system. This development will be guided by psychophysical experimentation and usability tasks that will clarify how we perceive some audiovisual phenomena, such as audiovisual synchrony, audiovisual depth perception, and audiovisual recalibration phenomena.

The experimental data collected will be integrated in predictive cognitive models for the processes involved in the perception of audiovisual synchrony under different conditions of stimulation. These models will be the quantitative basis for the computational solutions to be integrated in our immersive system.

## IMMERSIVE SYSTEM DESCRIPTION

In order to carry out the work, the facilities of the Vision and Perception Laboratory at the University of Minho are being used. In this section we describe the main features of the relevant system to be used.

### Visualization System

The visualization system that we are using is composed by a cluster of 3 PCs with NVDIA® Quadro FX 4500 graphics boards, and works with custom projection software running on top of OpenGL and using VR/Juggler as a "virtual platform". Each of the PCs forming the cluster is connected to one image channel using 3chip DLP projectors Christie Mirage S+4K with a resolution of 1400x1050 pixels and a refresh rate of 60Hz up to 101Hz per channel (see Figure 2). The projectors are capable of stereoscopic projection, and the surface of projection can range from a PowerWall of 2.80 m high per 6.30 m wide to a three face (one frontal and two lateral) CAVE-like configuration with each face conveying a projection area of 2.80 m x 2. 10 m.

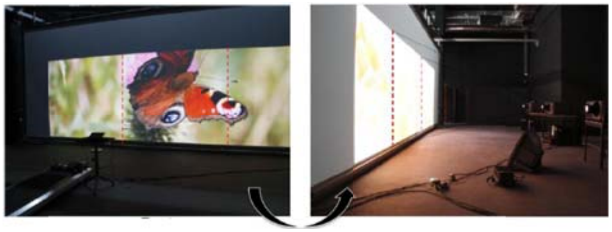

**Figure 2**. The visualization system in a PowerWall configuration. The red marks delimit the area of projection of each projector. Blending functions are used in order to give the sense of an uninterrupted projection surface.

### Room Acoustics Modeling and Auralization Software

The auralization system uses the *Image Source Method* in order to generate sounds corresponding to particular spaces (*Room Impulse Response* – RIR), taking into account the sound source and the listener positions (see Figure 3). Furthermore and depending on the source and listener position the program generates the correct temporal and frequency distribution for the sound presented at each ear (*Head Related Transfer Function)* applying it to a particular RIR. The final binaural sound with the simulated depth cues for a defined room is obtained by convolving the computed RIR with an anechoic binaural sound (a sound recorded in an anechoic chamber – a room design to absorb all the sound reflections and thus allowing the recording of only the direct sound of an auditory event).

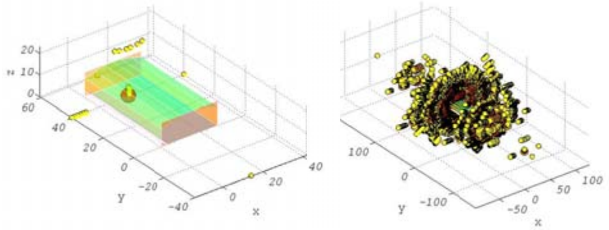

**Figure 3**.Graphical depiction of the external sound sources generated to simulate the wall reflected sound in a four orders of reflection simulation. Images from a simulation performed in the auralization software currently being developed.

## Motion Capture System

The motion capture (MoCap) system that we will use is a Vicon$^{TM}$ MX F20 MoCap system composed by 6 near-infrared cameras with a frame rate of up to 500 Hz, capable of tracking the three-dimensional position of retro-reflective markers with a temporal resolution of 240 HZ and an accuracy of 2 millimeters (see Figure 4). In order to use the Vicon$^{TM}$ MoCap system as a tracker for our interactive immersive environment, we will have to convert the data from the MoCap system into data capable of been read by the virtual platform. This will involve an equalization of the coordinate-axis orientation of both MoCap and visualization system as an equalization of the time stamps and frame rate of both systems. At the moment solutions for coordination between these two systems are being studied.

**Figure 4**. Details on one of the six cameras that integrate the MoCap system at the LVP (UM)

## PSYCHOPHYSICAL EXPERIMENTATION AND THE ITERATIVE APPROACH

The interactive audiovisual immersive system in development intends to be a perceptually validated integration between the three above described systems. We will have to develop a connection between the MoCap system and the visualization system in order to convey adequate visualization and auditory stimulation to the user's position. However there are several psychophysical problems that we need to account for.

### Audiovisual Synchrony Perception

The perception of audiovisual synchrony is central to the sense of a coherent audiovisual immersive environment and can be quite important in guiding user's action in the world. However this is still a quite intriguing phenomenon. Contradictory data exists on the aural-visual temporal relation that provides the best audiovisual synchrony perception and on its relation with depth perception [8, 9]. We intend to use Simultaneity Judgment Tasks in VR environments in order to develop a Predictive Cognitive Model for human synchrony perception. This model will guide the temporal relation between the visual and the auditory streams of audiovisual events in the final immersive system.

### Visual and Auditory Depth Perception

Here we are looking for the visual (pictoric and dynamic) cues and for the features of the acoustic signal (number of reverberation orders, air attenuation, sound pressure level decrement with distance) that are more important to give rise to an accurate judgment of distance in the audiovisual virtual world. In order to accomplish this we are preparing psychophysical experiments with auditory and visual distance judgments tasks, made through both comparisons between real world stimulation and computer generated stimuli, and through absolute estimation tasks. The rational underlying the distance judgment tasks made through real-virtual comparisons is that if we are accurately modeling the world and stimulating the user, there will be no difference on the distance judgment between a real and a virtual, computer generated, stimulus.

### Audiovisual Recalibration Processes

The delay between user's movement and audiovisual systems adjustment will have to be the minimum possible, mostly because systems latencies and its visible consequences are fundamental virtual environment deficiencies that can hamper user perception and performance [10]. In this scope, we will also explore the phenomenon of recalibration (i.e. the ability to handle, up to a measurable extend, temporal and spatial inconsistencies in the audiovisual world). Knowing the user recalibration capacity will give us an estimation of what could be the tolerable latencies between the user position and the audiovisual stream update. This is important for systems that have a limited computational capability or a certain network bandwidth.

### The Iterative Approach

All the implemented solutions will be guided by Predictive Cognitive Models on the above referred Perceptual phenomena and evaluated by psychophysical experimentation. The idea is to use the results of the real-virtual comparison experiences as a measure of good design and successful human computer interface implementation. This evaluation will be carried out first in each system individually and finally on the integrated interactive audiovisual immersive system. Along with the psychophysical validation others usability tests, like qualitative evaluation of comfort level and feeling of presence, will also be carried out.

The individual systems' evaluation will always be carried out, following some upgrading or development. In this way, we can say that the evaluation is part of the iterative nature of this work's development (Figure 5).

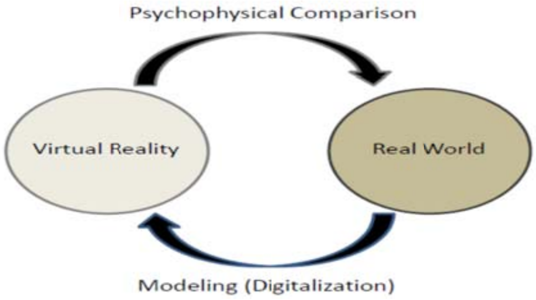

**Figure 5**. A representation of our iterative approach applied to the development of immersive environments.

After any improvement effort on the immersive system we should experiment and check the perceptual closeness between simulation and the real-world. If not satisfactory, a new cycle of psychophysical experimentation (with the goal of finding the error or lack in the modeling) should begin and then new simulation-reality comparisons should be carried out.

Once a perceptually satisfactory implementation of the interactive immersive system is accomplished, we intend to develop a final audiovisual and interactive system demonstration. This demonstration should consist in a simple task involving audiovisual perception and user's action, preferably capable of being accurately measured in different user's parameters as position, time reaction, and correct interactions. This could be accomplish by using, for instance, a catching task simulation, such as a baseball catch game or a service return in a tennis game, or any other interactive situation that involves accurate perception of the virtual object in order to effectively perform a task.

## CONCLUSION

This paper is intended to present a theoretically framed doctoral project description. Our key argument is that a better knowledge about both the physical world and the perceptual mechanisms underlying its perception can improve – and should guide – the development and implementation of interactive immersive systems. HCI has showed us several examples were the study of human cognition ended up in precious contributions for the development of computerized systems and the same should happen with the study of human perception and the improvement of immersive environments.

In the following years of this doctoral project we intend to put together visualization and auralization systems, and demonstrate their interactive capability using a MoCap system. We believe that in the end, enabling users to do a perceptually consistent interactive task using these three systems will provide a quite immersive experience.

## ACKNOWLEDGMENTS

The author acknowledges funding by ERDF - European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project FCOMP-01-0124-FEDER-015095.

## REFERENCES

1. Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. "Surround-screen projection-based virtual reality: the design and implementation of the CAVE". *Computer Graphics, SIGGRAPH Annual Conference Proceedings*, 1993. 135-142.
2. Dix, A. "Human-computer interaction: a stable discipline, a nascent science, and the growth of the long tail". *Interacting with Computers*, 22 (2010), 13-27.
3. Milgram, P, & Colquhoum, H. "A taxonomy of real and virtual world display integration". *Mixed Reality – Merging Real and Virtual Worlds*, 1999.Chapter 1. New York: Springer-Verlag.
4. Regenbrecht, H. T., Schubert, T. W., Friedman, F. "Measuring the sense of presence and its relations to the fear of heights in virtual environments". *International Journal of Human-Computer Interaction*, 10(3),1998, 233-249.
5. Fitts, P. M., & Peterson, J. R. "Information capacity of discrete motor responses." Journal of Experimental Psychology, 1964. 67, 103-112.
6. Card, S. K., Moran, T. P., & Newell, A. "*The psychology of human-computer interaction*." Lawrence Erlbaum Associates, Hillsdale, New Jersey. 1983.
7. Kieras, D., & Meyer, D. "An overview of the EPIC architecture for cognition and performance with application to human-computer interaction." *Human-Computer Interaction*, 1997. Vol. 12, 391-438.
8. Sugita, Y., & Suzuki, Y. "Implicit estimation of sound-arrival time." *Nature*, 2003. *421*, 911.
9. Lewald, J., Guski, R. "Auditory-visual temporal integration as a function of distance: no compensation for sound-transmission time in human perception." Neuroscience Letters, 2004. 357, 119-122.
10. Mania, K., Adelstein, B. D., Ellis, S. R., Hill, M. I. "Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity". *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization*, 39-47, New York, USA, 2004.