An Investigation into Immersion in Games Through Motion Control and Stereo Audio Reproduction

David Manuel
Glasgow Caledonian University
Cowcaddens Road
Glasgow
david.mediology@gmail.com

David Moore
Glasgow Caledonian University
Cowcaddens Road
Glasgow
j.d.moore@gcu.ac.uk

Vassilis Charissis
Glasgow Caledonian University
Cowcaddens Road
Glasgow
vassilis.charissis@gcu.ac.uk

ABSTRACT
This paper investigates the concept of immersion by means of motion tracking technology and novel gameplay experience. A system was developed that allows a user to navigate a virtual soundscape through motion-tracked movement around the bounds of a room. A subjective test was carried out to measure the perceived level of immersion for the motion-tracked systems versus a traditional hand-held control system. Results show that soundfield interaction through motion tracking is capable of producing a more enjoyable and immersive experience when compared to a traditional handheld analogue controller.

Categories and Subject Descriptors
H.5.0 [**Information Interfaces and Presentation**]: General

H.5.1 [**Information Interfaces and Presentations]**: Multimedia Information Systems – *Artificial, augmented, and virtual realities abstract data types, Audio input/output*

K.8.0 [**Personal Computing]**: General – *Games*

General Terms
Measurement, Performance, Design, Experimentation, Human Factors.

Keywords
Motion Control, Kinect Sensor, Audio Game, Immersion, Sound Interaction

1. INTRODUCTION
Computer games platforms are becoming increasingly more powerful through advancements in processing, storage and novel control technologies. Current generation systems are able to reproduce much higher quality audio than previous generations, and games developers are now taking sound more seriously than ever before. This, combined with greater expectations by consumers, has led to a significant increase in research on how to exploit technological advancements in order to increase player immersion and hence, improve the overall gaming experience.

People often play computer games as a form of escapism from the real world. Games can allow them to take on situations they cannot or would not face in real life. This being so, it is important that game developers create high degrees of realism in their games in order to create this illusion, and in modern game development, this illusion can be enhanced through photo-realistic graphics, expansive 3D environments and spatial audio. This state of false reality and the feeling that is often derived from this is known as "immersion".

The concept of immersion in computer games is not new and can be dated back to some of the earliest consoles and games [1]. Although immersion in computer games has improved rapidly over the last 20 years through technological advancements, it is only since the birth of third generation consoles (Xbox 360, PS3) that sound as a method of greatly increasing immersion has been fully recognised. Audio engines such as XACT, FMOD, and HAVOC have been key tools in improving in-game audio through the ability to dynamically change the audio in real-time during gameplay. Also, game developers are now able to manipulate auditory cues (Interaural level/time difference, spectral difference/timbre, HRTF etc) [2] to help bring the player from the real world and place them in a virtual environment.

Although current audio reproduction techniques can create very realistic sonic experiences, a key part of the auditory experience that is encountered every day is physical motion and interaction within our environments. Currently, the traditional methods of character movement consist of either an analogue game controller or mouse and keyboard. Whilst these methods provide the player with considerable control, it does not bring the player the auditory experience of their in-game counterpart and can be argued to miss fundamental parts of the games immersive experience [3] [4]. In light of this, there is a need for research into physical interaction with the game world. The logical progression for control of games would be to advance from stationary games control to full body motion control. This has been made possible today through motion tracking technologies developed by leading games companies (i.e. Nintendo Wii, Xbox Kinect, Sony PS Move) [5][6]. At the time of writing, the most recent of these technologies is Microsoft's device known as Kinect. Kinect uses an RGB camera and depth sensors to calculate the depth and precise location of joints of a user in space. Currently, this technology is utilised for control of GUI elements of games, and control of character movement, although currently more novel uses for the technology has been found in the research and development communities by means of open source libraries to utilise this hardware.

Figure 1. Microsoft's Kinect Sensor

The recent public release of the software development kit for the Kinect has created opportunities for research to be conducted into the use of this device for interacting with sound. However, there is a lack of studies that explore the use of motion control via the Kinect for controlling soundfields with a view to increasing in-game immersion. Therefore, this paper outlines a use for motion tracking technologies to control virtual sound environments. A system was developed that allows users to successfully manoeuvre around a virtual sound environment through the use of physical body movements, in contrast to the traditional handheld controls. The main objective of this research was to investigate whether a user feels there is an increased level of immersion in an audio environment they can interact with through motion control hardware.

2. BACKGROUND

2.1 What is immersion?

Immersion is the subjective experience of being fully engrossed in a game and can be described as the point in which a user's full concentration is directed at a game with no knowledge of occurrences outside the game [7]. An understanding of the psychological concept of immersion is essential before the measurement of this subjective experience can be undertaken.

Csíkszentmihályi [8] describes a concept called "Flow", the mental state that leads to a feeling of immersion being perceived by a game participant. The requirements stated by Csíkszentmihályi for the development of a feeling of flow (and thus, immersion) can be broken down into the following elements:

- Clear Goals
- Concentration
- Distorted Sense of Time
- Direct and immediate
- The balance of Ability and Challenge
- Personal Control
- Rewarding

It should be noted that although each of these elements is important, Csíkszentmihályi states that not all are required to be met in order for a state of flow and immersion to be created.

More recently, Ermi and Mayra [9] define immersion under 3 categories – Sensory Immersion, Imaginative Immersion, and Challenged-Based immersion. Sensory immersion is that created by audiovisual stimuli in computer games. This sense of immersion can be improved through advances in graphics, game physics, audio quality, and more immersive screens e.g. HD, 3D enabled etc. Imaginative Immersion is how the player emotionally identifies with the story of a game, and the characters involved. This is when the player creates a sense of "empathy" with a storyline that is emotionally engaging, or with a character that the player can relate to. Nacke and Lindley [10] state that this is a similar definition to that of Brown and Cairns [11], who describe this as "absorption", and that Imaginative immersion is often encountered in role-playing games. Challenge-based Immersion is linked very closely to the idea of flow – the point at which a user's full concentration is directed at the game with no knowledge of occurrences outside the game, for example, the passing of time. This form of immersion comes from the player using his/her abilities to complete the challenges of the game, resulting in complete cognitive concentration at that moment in time. As with the definition by Csíkszentmihályi [8], the greater extent with which each of the above categories can be met, the more likely the player is likely to be immersed in the task at hand. Taking the above information into account, it is possible to design systems with the aim of intrinsically creating the feeling of flow and immersion

2.2 Measuring Immersion

To date, some of the most successful measures of immersion have come from attempting to measure the subjective scale to which a player feels immersed in their game experience, through means of video observations, SAM tests, and questionnaires. The Game Experience Questionnaire (GEQ), developed by Ljesselsteijn et al [12] is one such method of measuring immersion through administered questions. This form of questionnaire combines several game-related subjective measures and consists of 36 questions on immersion, tension, competence, flow, negative affect, positive affect, and challenge [10]. The GEQ employs simple statements such as "I forgot everything around me" and "I was good at it", followed by a 4 point scale to return the information from the subject, and was considered by Nacke et al [10] as of "sufficient quality to accurately report gameplay experience". The GEQ was used as a basis for the creation of the questionnaire used during testing in this project.

2.3 Sound Interaction

Biachi-Berthouze et al [13] propose that: "The contribution of full-body experience is threefold: (a) it facilitates the feeling of presence in the digital environment (fantasy); (b) it enables the effective aspects of human-human interaction; and (c) it unleashes the regulatory properties of emotion (affect)."

Their research suggests that an increase in body movement results in an increase in the player's engagement level, producing a more effective gaming experience. While they do not immediately suggest that this is an immersive state, it shows that the principle of physical motion can improve a gamer's engagement, and thus, by the definition of immersion being used for the purpose of this study, could be argued that in itself, motion can help create immersive states.

One form of stimuli, which has been considered key in recent times for creating a feeling of immersion, is the use of spatial audio – audio playback that emulates the interaction of sound in an environment. One area of realistic soundscape recreation, which has not been extensively researched, is the physical sensation of movement in a sound environment, which is an essential part of the way we interact with things around us – especially for those who are visually impaired or blind who rely on their location in relation to sound emitting objects. This being so, the goal of this research project to determine to what extent it is possible to improve immersion in computer games through motion tracking technology and stereo audio processing and effects.

3. System

3.1 System Development

In order to investigate the goals set in this paper, a system was developed that uses Kinect motion tracking to allow a user to walk around a room, and hear a changing soundfield as if they were in that virtual location. All sounds are programmed to change dynamically in audio level and reverberation level depending on the user's position in the room, and the distance from sources in the game environment. Although it is possible to include visual elements in this game, only minimal visual aids were added in order to test the immersive effects of audio alone.

The system was developed using XNA, Microsoft's game engine, with XACT 3D audio engine, which works natively with XNA, and the Kinect for Windows SDK. This allowed us to use skeletal data from the Kinect sensor. The system interconnection is shown in Figure 2.

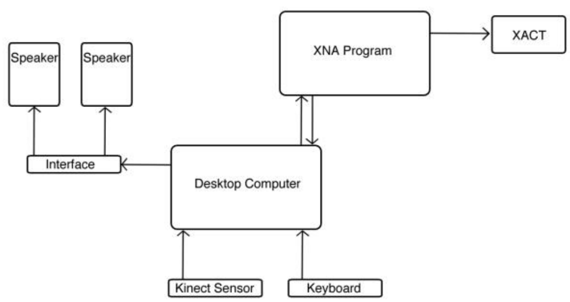

Figure 2. Developed System

XACT is an audio engine developed by Microsoft as a means of splitting the job of the programmer from the job of a sound designer, essentially removing the requirement for programming and allowing the sound designer to concentrate on the creative aspects of the sound rather than the mechanics of how the sound is played back during the game. XACT presents a graphical user interface that allows sound designers to load sound effects and music, manipulate these sounds through the use of DSP (Doppler, Pitch, Level, Modulation, Reverberation etc) to help create a sense of space and realism to the sound. These effects can be used to create varying sound effects to avoid repetition (e.g. changing the pitch of footsteps randomly within clamped limits to provide variation). XACT then outputs files that are integrated into to the programming and allow the programmer to implement these in a game by calling sounds from the created files. XACT does have limitations and is unable to output beyond a stereo speaker configuration.

The level to which the XACT audio engine manipulates the sound is minimal and is used as a method of reproducing mono audio files in 3D game space, and assisting in the dynamic control of the audio level and reverberation. XNA interacts with sound through audio emitters and audio listeners. An audio emitter is an object in the game world that emits sounds, and the listener is the equivalent of the player's ears in the game environment, hearing sounds in relation to the distance from the audio emitters. The further away the audio listener moves from the audio emitter, the lower the level of the audio emitter. This is manipulated through the use of graphs in XACT, where audio level is lowered over distance (distance is the measure of in-game distance, which was then mapped to match real-world space). It is possible to change the characteristics of how sound level changes over distance within the software, however, the default settings were used. XACT also controls the mix level of reverberation, a facility included in XACT as part of a selection of DSP controls, which again, can be controlled dynamically over distance. During the initial prototype programming, XACT's Doppler effect was used in an attempt to create more realistic moving sound sources, in this case, the moving speech of the conversation taking place in the game. It was found that the speed at which the sound source moved was too slow to warrant further use of the Doppler effect, but this effect should be re-considered should more moving sound sources be used in further development (i.e. birds, moving vehicles etc).

The Kinect analyses the location of the listener's head in real-space and returns positional information for it. The position is then used as the location for the "Audio Listener" in XACT. It was required that the system is tuned for the room size used during the testing phase, as the size of room effects the location of the in-game sound sources' positions. As stated earlier, the system outputs 2-channel stereo because of the limitations of XACT. Please note that OpenAL is also a popular choice for the creation of 3D virtual audio environments and has the capability of multi-speaker reproduction. However, OpenAL requires an understanding of low-level computer programming and does not come with a graphical interface like XACT or FMOD. It may be possible to reproduce the system described in this paper using a combination of OpenAL and XNA.

3.2 Game

A narrative-based game was developed for the system, where a participant is required to find a passcode to a secret room. The passcode was presented as part of a conversation, taking place between two people in an outdoor military setting with environmental sounds. Sound sources used in both game experiences were realistic soundscapes (field recordings of thunder, forests, rain, trees, animals etc), with the narrative of the game provided through voice acting. These sounds were all mono sound sources and were given their 3D position in XNA whilst XACT handled the relationship between the listener and emitter.

The location of the characters moves dynamically in the soundfield, requiring the user to try to follow them and "listen in". This presents a learning curve where the participant must find which directions the conversation moves in, and what positions in the room it passes through. Once the passcode has been obtained through a combination of focused listening and movement, the user has to type it into an onscreen keycode box using a computer keyboard. A correct passcode results in the successful completion of the game and the conclusion of the test. During the game design process each of the elements of the model detailed by Csíkszentmihályi [8], was taken into account.

The game was designed to incorporate Kinect motion control as well as analogue Xbox controller input – the controller represents a traditional method of game control and incorporating both allowed for direct comparison between them.

No music was used in the game, as it was felt that the subjective nature of music could detract from the overall goal of creating a realistic immersive auditory experience. Positioning music in this environment could also prove problematic, as it begs the question of where does music sit in a 3D environment and is this unlocalised sound detrimental to immersive experience (hearing sounds out width the natural environment of the game). This question was not answered in this research.

As this research is purely based around the effect that audio has on creating immersive audio environments, minimal visuals were used. The game participants were presented only with an input screen where they could input the passcode described in the previous section

3.3 Room

The room used for developing and testing the system was the Virtual Reality Lab in Glasgow Caledonian University (see Figure 3). This room provided a large television screen with the Kinect sensor positioned below, high-quality stereo playback speakers, and 15ft by 15ft of floor space. A reasonably large space was required in order to give users enough space to actively participate in the game with a high degree of freedom. The orange square shows the approximate area in the room where the Kinect would track the user's movement. During the analogue controller testing phase, participants were placed in the optimal stereo listening position for the room, which was directly between the stereo speakers, in the middle of the room.

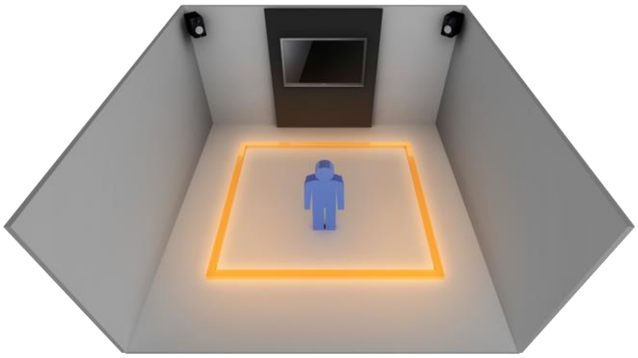

Figure 3. VR Lab

4. TESTING METHOD

For the test, 21 test participants were recruited based on the criteria that they were self-confessed gamers, with experience in playing narrative-based games. Most of these participants were students from the Games Development course and Audio Technology Courses at Glasgow Caledonian University. Lecturers and other researchers from the university also took part.

The average age of participants was 25 with ages ranging from 19 to 47. All participants were male apart from one, although gender was not considered an important factor in the recruitment process. The participants were asked their self-assessed ability to hear, with no participants declaring any medical conditions that they considered detrimental to their ability to localise sound accurately.

Participants were asked to attempt to complete both the motion control version of the game and the analogue controller version of the game. They were given as long as they wanted to complete the goal of finding the passcode - this was important so as to not make the participant feel rushed, which could result in the loss of the feeling of immersion. After completing each test, the participant was asked to fill in a questionnaire. The questionnaire was designed to interpret the participant's feelings of flow immersion. Although multiple definitions of immersion were explored during the initial stages of the project, the definition by Csíkszentmihályi [8] was followed. Csíkszentmihályi's points of flow and immersion were incorporated as part of the questionnaire in order to gain quantitative data on the participants' feelings. Each point from the model was addressed in the test questionnaire as follows:

Clear Goals – Participants were asked if they felt the game had clear goals. This is an important point as it is a major definition in whether or not the experience can be described as a game, rather than an interactive experience.

Concentration – Participants were asked if they felt a sense of focus whilst interacting with the system and achieving the goal.

Distorted Sense of Time – Participants was asked to rate if they were aware of the amount of time that had past during each test method. This relates to the well-known colloquialism "Time flies by when you are having fun".

Direct and Immediate Feedback – Participants were asked whether or not they felt their actions (walking around the room/using the controller) were directly associated with their attempts to achieve the goal i.e. if their choice of location in the room was helping them achieve the task.

Balance of Ability and Challenge – A question was asked in order to determine whether or not the participant felt that they were capable of completing the tasks given to them, and rate how challenging they found these tasks.

Personal Control – Participants were asked if they felt they were in control of the system. This is the difference between watching a film and participating in a game. The participant was required to feel as though they were not only observing auditory events but that they were intrinsically part of the experience presented to them.

Rewarding – Participants were asked if they felt the game provided a rewarding experience. This refers to the sense of reward felt by participants during their test experience. This can be felt through different stages of the game, be it the first time a participant successfully locates the conversation, the direction in which it is moving, or hearing the passcode correctly after multiple attempts. They were also asked to rate how immersed they were if they felt the game was providing a rewarding experience.

The same questions were asked to participants when using both methods of control so that results could be compared later to determine whether the game provided everything required to create an immersive experience. The questions consisted of Yes/No response questions, rating question using the Likert Scale (aided using Self Assessment Manikin images), and open-ended questions for a qualitative opinion. Each game took on average 5 minutes to complete.

5. RESULTS

5.1 Test Results

For both the Kinect and analogue controller versions of the game, all participants felt there was a clear goal that helped create a sense of focus. Participants also agreed that they were capable of completing the tasks given to them. These points help validate the design goals of the game.

According to a t-test of rating responses, there was no significant difference between the level of challenge posed by the Kinect and controller versions of the game ($p = 0.23$).

With regards to direct and immediate feedback, 19/21 of subjects felt actions when using the Kinect were directly associated with their attempts to achieve the game objective. In the case of the controller, 18/21 of subjects felt their actions were directly associated with their attempts to achieve the goal. When asked whether subjects felt they were in full control of the system, 18/21 agreed they were in full control using the Kinect, whereas 19/21 agreed they were in full control when using the controller. The high number of subjects for both control methods also validates the developed system. When asked whether they found the overall experience rewarding, 19/21 stated they did for the Kinect, however, only 15/21 stated that had for the controller.

All participants were asked to rate the extent at which they were aware of time passing when playing the game. Results from a t-test show that there was no significant difference between the Kinect and controller versions of the game ($p = 0.16$). Upon reflection, the reliability of this result as a measure of immersion may be questionable as the time to complete the tests was relatively short.

When asked whether subjects would consider themselves as having been immersed in the game, 19/21 agreed they had for the Kinect version and only 13/21 for the analogue controller. All subjects agreed that the Kinect offered a more immersive experience overall. A t-test of the immersion ratings gathered from subjects revealed the difference between the Kinect and an analogue controller to be highly significant ($p \approx 0.00$). This difference is clearly shown in Figure 4 which displays mean rating values for both methods of control with 95% confidence intervals (a high mean value represents a high level of immersion).

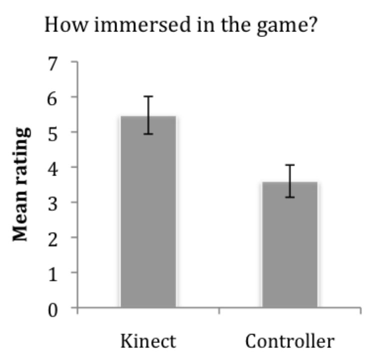

Figure 4. Level of perceived immersion

Finally, 76% of participants felt the Kinect provided a more natural method of control. Furthermore, all participants felt that the Kinect system created a higher degree of realism in the narrative-based game, which is important in creating a feeling of immersion.

5.2 Summary of Comments

Subjects were also asked to comment on the Kinect system and immersion. The following shows a summary of comments made by participants for the open-ended question:

"What did you like/not like about the Kinect experience, and how would you like to see it improved?"

Many participants declared that the motion control system presented felt "fresh", "new" and "novel", perhaps indicating that new technologies, such as the Microsoft Kinect, have the potential to create new and innovative game mechanics and experiences. Some participants mentioned they would have liked to have experienced surround sound in the game to improve the overall listening experience. Surround sound would allow sounds to be placed behind the participant and could make effects such as the Doppler effect being used more effective when compared to stereo. Furthermore, the soundfield would have the potential to be more enveloping because of sound arriving from all directions.

Some participants also stated that they occasionally found the localisation of moving sound sources difficult, and it is theorised that 5.1 surround sound could make this easier.

Although surround sound may present a significant improvement in creating immersive audio environments, the results of this research do show that a stereo configuration of loudspeakers are able to create highly immersive audio environments, with a number of participants amazed when told that the experience they had been presented with had not been made using surround sound techniques. (often participants looked towards the 5.1 speaker systems in the rear of the room and were shocked to discover they were not actually on during their experience.

A number of participants mentioned that the addition of a visual system along with the sound would probably help improve their immersive experience and create a more viable game.

Some participants described how a dynamic soundfield controlled using the Kinect is a good way of enveloping the participant in sound, a feeling that does not always come across when using stationary control.

6. CONCLUSIONS

This paper investigated immersion through a novel audio game that incorporates motion control and stereo sound source positioning. The soundfield was controlled and dynamically changed by movement of a listener in a space. Results show that when compared to a traditional handheld analogue controller it is possible for motion tracking to significantly improve the level of immersion.

It should be mentioned that the use of stereo over surround sound audio potentially limits the immersive potential for this system, and that the use of stereo output was purely a limitation of the current technology and programming environments offered by Microsoft in regards to the XNA game development framework (i.e. the lack of surround sound support available in XNA/XACT). Should the techniques reported in this research project be advanced in future research, it is hoped that surround sound output is available through the XACT software or other software development systems such as Unity3D and that these be stably compatible with the Kinect SDK.

It was found that narrative based audio games are capable of creating enjoyable, realistic experiences when combined with full body movement, but can be found lacking with analogue controls. This could be because of a lack of realism, or a lack of personal involvement when playing from a stationary position. This research has stimulated many ideas for future research exploiting spatial audio and motion sensor technologies.

7. FUTURE WORK

Some consideration has been put into potential uses for the system developed during this project.

Use in Current Games - this system could be incorporated into current games as part of QuickTime events, interactive cut scenes, or a new way to listen to in-game narratives. If rotational information were to be incorporated into the Kinect SDK, a fully realistic audio environment could be created which could improve audio in smaller rooms or stationary locations.

Virtual Reality Experiences - as this system develops a virtual auditory environment, it is theorised that it may be used to create audio-realistic environments, for example recreating certain areas of the world, or possibly, using internet audio streaming and surround sound microphones, allow a participant to manoeuvre around an environment located somewhere else in the world.

Games for Health - this system could also be used to create games directly suited to blind and partially sighted gamers, or educational tools in order to help younger blind users to navigate themselves around rooms using audio. This could also lead to research into making games a more interactive, viable experience than what is currently on offer.

Novel Music Listening Experiences - this technology could be utilised to create novel music listening experiences, through means of placing virtual instruments throughout location of a room and participants being able to physically interact with the sound emitters. For example, this could allow games to be created that allow participants to act as the conductor of an orchestra, and walk around all the different sections and hear mixes from different perspectives.

8. REFERENCES

[1] McMahan, A. 2003 Immersion, Engagement, and Presence: A Method for Analyzing 3-D Video Games, in Wolf, M.J.P., and Perron, B. (eds.). The Video Game Theory Reader. Routledge, New York

[2] Roads, C., 1996 The Computer Music Tutorial. Cambridge: MIT Press. Pg 451 – 472, 1054 – 1068.

[3] J. J. López, A. González. 1999 “3-D audio with dynamic tracking for multimedia environments” Digital Audio Effects Workshop (DAFX99) Trondheim, Norway.

[4] Serafin, S, 2004. Sound Design To Enhance Presence In Photorealistic Virtual Reality. In International Conference on Auditory Display. Sidney, 6-9 July. Australia: Unknown. 4.

[5] Intro to the Kinect SDK–Drawing Joints in XNA « XImplosionX. 2012. Intro to the Kinect SDK–Drawing Joints in XNA « XImplosionX. [ONLINE] Available at http://www.ximplosionx.com/2011/06/19/intro-to-the-kinect-sdkdrawing-joints-in-xna/. [Accessed 02 April 2012].

[6] PlayStation(R)Move Motion Controller Delivers a Whole New Entertainment Experience to PlayStation(R)3 -- TOKYO, March 10 /PRNewswire/ --. 2012.PlayStation(R)Move Motion Controller Delivers a Whole New Entertainment Experience to PlayStation(R)3 -- TOKYO, March 10 /PRNewswire/ --. [ONLINE] Available at: http://www.prnewswire.com/news-releases/playstationrmove-motion-controller-delivers-a-whole-new-entertainment-experience-to-playstationr3-87288777.html. [Accessed 16 January 2012].

[7] Brown, E. and Cairns, P., 2004, A Grounded Investigation of Game Immersion. In CHI '04 extended abstracts on Human factors in computing systems.

[8] Csíkszentmihályi, M., 1990. Flow: The Psychology of Optimal Experience. HarperPerennial, New York.

[9] Ermi, L. and Mäyrä, F. 2005 Fundamental components of the game play experience: Analysing immersion. In: S. de Castell & J. Jenson (eds.), Changing Views: Worlds in Play. Association’s (DiGRA) Second International Conference. Selected papers of the 2005 Digital Games Research.

[10] Grimshaw ,M. Lindley C. A., and Nacke, L. 2008 Sound and immersion in the first-person shooter: Mixed measurement of the player’s sonic experience. Proceedings of the AudioMostly, 3rd Conference on Interaction with Sound, Pitea, Sweden.

[11] Brown, E., Cairns, P. 2004 A grounded investigation of immersion in games. ACM Conf. on Human Factors in Computing Systems, CHI 2004, ACM Press, 1297-1300.

[12] IJsselsteijn, W., Poels, K., de Kort, Y.A.W., 2008. The Game Experience Questionnaire: Development of A Self-report Measure to Assess Player Experiences of Digital Games. TU Eindhoven, Eindhoven.

[13] Bianchi-Berthouze, N. W, Kim. D, Patel, 2007. Affective Computing and Intelligent Interaction - Does Body Movement Engage You More In Digital Game Play? and Why?. 1st ed. Unknown: Springer Berlin/Heidelberg.