

**DEVELOPING EFFECTIVE VIRTUAL  
SIMULATIONS AND SERIOUS GAMES:  
THE EFFECT OF BACKGROUND SOUND CUES  
ON VISUAL QUALITY  
PERCEPTION**

by

David Arnulfo Rojas Gualdron

A Thesis Submitted in Partial Fulfillment  
of the Requirements for the Degree of  
Master of Science (MSc)  
in  
Computer Science

Faculty of Science  
University of Ontario Institute of Technology  
Oshawa, Ontario, Canada  
August 2012

Copyright © David Rojas, 2012

## **Author's Declaration**

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

David Arnulfo Rojas Gualdron.

# CONTENTS

<b>ABSTRACT</b>	<b>vi</b>
<b>KEYWORDS</b>	<b>vii</b>
<b>LIST OF PUBLICATIONS DIRECTLY ARISING FROM THIS THESIS</b>	<b>viii</b>
<b>ACKNOWLEDGMENTS</b>	<b>ix</b>
<b>LIST OF TABLES</b>	<b>x</b>
<b>LIST OF FIGURES</b>	<b>xi</b>
<b>CHAPTER 1 - INTRODUCTION</b>	<b>1</b>
<b>1.1 PREAMBLE</b>	<b>1</b>
<b>1.2 RESEARCH OBJECTIVES</b>	<b>7</b>
1.2.1 EXPERIMENT 1	7
1.2.2 EXPERIMENT 2	8
1.2.3 EXPERIMENT 3	8
1.2.4 EXPERIMENT 4	8
1.2.5 EXPERIMENT 5	9
1.2.6 EXPERIMENT 6	9
1.2.7 EXPERIMENT 7	9
<b>1.3 THESIS ORGANIZATION</b>	<b>10</b>
<b>2 BACKGROUND</b>	<b>11</b>
<b>2.1 HISTORY</b>	<b>11</b>
<b>2.2 LITERATURE REVIEW</b>	<b>18</b>
2.2.1 FIDELITY	18
2.2.2 REALISM	20
2.2.3 IMMERSION	22
2.2.4 PRESENCE	23
2.2.5 MULTIMODAL INTERACTION	24
<b>2.3 SUMMARY</b>	<b>32</b>
<b>3 CHAPTER 3 - EXPERIMENTS</b>	<b>35</b>
<b>3.1 GENERAL CONDITIONS</b>	<b>36</b>
3.1.1 PARTICIPANT RECRUITMENT	36
3.1.2 EXPERIMENTAL SETUP	37
3.1.3 EXPERIMENTAL VARIABLES.	38
3.1.4 VISUAL STIMULI	39
3.1.5 AUDITORY STIMULI	43
3.1.6 EXPERIMENTAL METHOD	44

<b>3.2 EXPERIMENTS</b>	<b>45</b>
<b><u>SENSORY INTEGRATION (AUDIO-VISUAL) WITH NON-CONTEXTUAL AUDITORY CUES.</u></b>	<b><u>45</u></b>
3.2.1 EXPERIMENT 1: VISUAL QUALITY DEFINED WITH RESPECT TO POLYGON-COUNT WITH NON-CONTEXTUAL AMBIENT AUDITORY CONDITIONS.	45
3.2.2 EXPERIMENT 2: VISUAL QUALITY DEFINED WITH RESPECT TO TEXTURE RESOLUTION WITH NON-CONTEXTUAL AUDITORY CONDITIONS.	50
<b><u>SENSORY INTEGRATION (AUDIO-VISUAL) GIVEN BY THE TYPE OF VISUAL PRESENTATION (STEREOSCOPIC 3D VS. NON-STEREOSCOPIC 3D)</u></b>	<b><u>56</u></b>
3.2.3 EXPERIMENT 3: VISUAL QUALITY DEFINED WITH RESPECT TO POLYGON COUNT IN A STEREOSCOPIC 3D VIEWING ENVIRONMENT WITH NON-CONTEXTUAL AMBIENT AUDITORY CONDITIONS	56
3.2.4 EXPERIMENT 4 : VISUAL QUALITY DEFINED WITH RESPECT TO TEXTURE RESOLUTION IN A STEREOSCOPIC 3D VIEWING ENVIRONMENT WITH NON-CONTEXTUAL AUDITORY CONDITIONS.	61
<b><u>SENSORY INTEGRATION (AUDIO-VISUAL) WITH CONTEXTUAL AUDITORY CUES DISCUSSION</u></b>	<b><u>69</u></b>
3.2.5 EXPERIMENT 5: VISUAL QUALITY DEFINED WITH RESPECT TO TEXTURE RESOLUTION WITH CONTEXTUAL AUDITORY CONDITIONS	69
3.2.6 EXPERIMENT 6: VISUAL QUALITY DEFINED WITH RESPECT TO TEXTURE RESOLUTION IN A STEREOSCOPIC 3D VIEWING ENVIRONMENT WITH CONTEXTUAL AUDITORY CONDITIONS.	75
<b><u>SENSORY INTEGRATION (AUDIO-VISUAL) WHILE PERFORMING GOAL DIRECTED ACTIONS</u></b>	<b><u>82</u></b>
3.2.7 EXPERIMENT 7: VISUAL QUALITY DEFINED WITH RESPECT TO LEVEL OF VISUAL BLURRING UNDER BOTH CONTEXTUAL AND NON-CONTEXTUAL AUDITORY CUES.	83
<b><u>4 CONCLUSIONS</u></b>	<b><u>92</u></b>
4.1 RESULTS	92
4.2 BACKGROUND	93
4.3 EXPLANATION OF THE RESULTS WITH RESPECT TO EXISTING THEORIES	98
4.4 SERIOUS GAMES	100
4.5 FUTURE WORK	101
<b><u>5 REFERENCES</u></b>	<b><u>102</u></b>

## **ABSTRACT**

Virtual simulations and serious games (video game-based technologies applied to teaching and learning) have been incorporated in the teaching and training curricula of a large number of professions including medicine/surgery. Despite their benefits, there are open, fundamental issues regarding simulation quality, multi-modal cue interaction, and the resulting effect on visual quality perception and ultimately on knowledge transfer and retention. Here the results of a series of seven studies that examined the effect of background sound (contextually related and non-related with respect to the visual scene) on the perception of visual quality (defined with respect to texture resolution, polygon count) presented in stereoscopic and non-stereoscopic 3D. Results indicate that the perception of visual quality is dependent on ambient (background) sound. The results of these studies have implications for designers and developers of serious games who typically strive for high quality virtual worlds despite the computational burden associated with doing so. The results of these studies also bring us closer to understanding the role of quality, multi-modal interactions, and their effect on visual quality perception. This thesis is part of a larger effort in developing an understanding of virtual environment rendering quality, multi-modal interactions, user-specific factors and their effect on knowledge transfer and retention.

## **KEYWORDS**

Serious games, virtual simulation, stereoscopic 3D (S3D), visual quality, multi-modal cue interaction, audio-visual interaction.

# LIST OF PUBLICATIONS DIRECTLY ARISING FROM THIS THESIS

## Refereed Journal Articles Submitted and Under Review

**D. Rojas**, B. Kapralos, A. Hogue, K. Collins, S. Cristancho, L. Nacke, C. Conati, and A. Dubrowski. Developing effective serious games: The effect of ambient auditory conditions on visual fidelity perception in stereoscopic 3D. *IEEE Transactions on System Man and Cybernetics part B, Special Issue on Modern Control for Computer Games*. Submitted April, 2012.

B. Cowan, **D. Rojas**, B. Kapralos, S. Cristancho, F. Moussa, and A. Dubrowski. The serious game surgical cognitive education and training framework (SCETF) and its use to examine the effect of ambient sound on visual fidelity perception and task completion time. *Virtual Reality*. Submitted March, 2012.

## Refereed Conference Articles Submitted and Under Review

**D. Rojas**, B. Cowan, B. Kapralos, K. Collins, S. Cristancho, and A. Dubrowski. Developing effective virtual simulations and serious games: The effect of contextual background sound cues on visual fidelity perception. *ACM SIGCHI Conference on Human Factors in Computing Systems*. Paris, France, April 27 – May 2, 2013.

## Refereed Conference Articles

**D. Rojas**, B. Kapralos, S. Cristancho, K. Collins, A. Hogue, C. Conati, and A. Dubrowski. Developing effective serious games: The effect of background sound on visual fidelity perception with varying texture resolution. Medicine Meets Virtual Reality 19, Newport Beach, CA, USA, February 9-11, 2012. Also appears in *Studies in Health Technology and Informatics*, 173:386-392, 2012.

**D. Rojas**, J. J. H. Cheung, B. Cowan, B. Kapralos, and A. Dubrowski. Serious games and virtual simulations debriefing using a social networking tool. In *Proceedings of the Computer Games Multimedia and Allied Technology (CGAT 2012) Conference*, Bali, Indonesia, May 7-8, 2012.

**Rojas, D.**, Kapralos, B., Cristancho, S. M., Conati, C., Collins, K., Dubrowski, A. (2011). The Effect of Background Sound on Visual Fidelity Perception. In *Proceedings of Audio Mostly 2011*, Coimbra, Portugal, September 7-9, 2011.

## **ACKNOWLEDGMENTS**

I would like to thank my supervisors Drs. Bill Kapralos and Adam Dubrowski for all the guidance and help provided during the realization of this thesis. I would also like to thank Dr. Andrew Hogue, Dr. Miguel Vargas Martin, Dr. Francois Desjardins and Dr. Carolyn McGregor. Dr. Hogue and Dr. Vargas Martin (both from the Faculty of Business and Information Technology, University of Ontario Institute of Technology) served on my supervisory committee and provided useful feedback regarding my thesis. Dr. Desjardins (Faculty of Education, University of Ontario Institute of Technology) served as the external thesis examiner and provided me with very and finally, Dr. McGregor served as Chair during my thesis defense.

I would also like to thank my family that has been my unconditional support at any moment. (My mom, my dad my brothers and my Aunt and cousins from Montreal, without them any of this would have been possible.)

This thesis work is also dedicated to those who always trust me and never doubt that I was going to accomplish this new step in my life. To my friends, my colleagues and everybody who one way or another contributed to this work.



## LIST OF TABLES

Table 1. The most relevant serious games developed and used up to 2006 (first of two tables). Reprinted from Atwell [2006].	14
Table 2. The most relevant Serious games developed and used up to 2006 (Second of two tables). Reprinted from Atwell [2006].	15
Table 3. Polygon count for each of the six surgeon head models considered.	40
Table 4. Texture Resolution for each model in each of the six images.	41
Table 5. Texture resolution for each model in each of the six images.	42
Table 6. Experiment 1. Summary of the video game attribute importance.	46
Table 7. Experiment 1 Results. Mean and Standard Deviation.	49
Table 8. Experiment 2. Summary of video game attribute importance.	51
Table 9. Experiment 2 Results. Mean and Standard Deviation.	54
Table 10. Experiment 3. Summary of video game attribute.	57
Table 11. Experiment 3 Results. Mean and Standard Deviation.	59
Table 12. Experiment 4. Summary of video game attribute importance.	62
Table 13. Experiment 4 Results. Mean and Standard Deviation.	66
Table 14. Experiment 5. Summary of video game attribute importance.	70
Table 15. Experiment 5 Results. Mean and Standard Deviation.	72
Table 16. Experiment 6. Summary of video game attribute importance.	77
Table 17. Experiment 6 Results. Mean and Standard Deviation	78
Table 18. Task completion time averaged across auditory condition.	88

## LIST OF FIGURES

Figure 1. Relationship among the multiple factors that influence transfer.....	3
Figure 2. The experimental design of Mastoropoulou 2004. Reprinted from [Mastoropoulou, 2004] .....	29
Figure 3. Visual stimuli. Six renderings of a surgeon's head, each varying with respect to polygon count.....	40
Figure 4. Surgeon's head textured with varying resolution texture. ....	41
Figure 5. Surgeon's upper body holding an orthopaedic drill model.....	42
Figure 6. Experiment 1 Results. Image quality (polygon count) vs. perceived quality (images presented individually).....	49
Figure 7. Experiment 2 Results. Texture resolution (x-axis) vs. perceived visual quality (y-axis) across each of the four non-contextual auditory conditions considered.....	53
Figure 8. Experiment 3. Polygon count (x-axis) vs. perceived visual quality (y-axis) across each if the four non-contextual auditory conditions considered.....	58
Figure 9. Experiment 4 results. Texture resolution (x-axis) vs. perceived visual quality (y- axis) across each of the four non-contextual conditions considered.....	65
Figure 10. Experiment 5 results. Texture resolution (x-axis) vs. perceived visual quality (y-axis) for each of the seven auditory conditions (three contextual and four non- contextual).....	71
Figure 11. Experiment 6 Results. Texture resolution (x-axis) vs. perceived visual quality (y-axis) across each of the seven (contextual and non-contextual) auditory conditions.....	77
Figure 12. Experiment 7. The levels of visual quality (defined with respect to image blurring) considered in this experiment. (a) Original (non-filtered) version. (b) - (f) Application of the blurring filter with an increase in the blurring effect.....	84
Figure 13. The experimental environment (a) Then view of the operating room environment at the start of each experimental trial. (b) Upon picking the drill, participants were asked to rate the visual quality perception .....	85
Figure 14. Experiment 7 results. Visual Quality perception (x-axis) vs. image blurring level under four ambient sound conditions. Image quality is defined with respect to a	

blurring of the entire scene with Level 1 representing the least blurring and Level 6  
the greatest blurring. .... 87

Figure 15. Experiment 7 results. Ambient auditory condition (x-axis) vs. time to complete  
the task (y-axis) in the presence of contextual and non-contextual auditory  
conditions..... 88

# CHAPTER 1 – INTRODUCTION

## 1.1 PREAMBLE

A serious game can be defined as an interactive computer application, with or without a significant hardware component, that i) has a challenging goal, ii) is fun to play and/or engaging, iii) incorporates some concept of scoring, and iv) imparts to the user a skill, knowledge, or attitude that can be applied to the real world [Bergeron, 2006]. Serious games are able to captivate and engage players/learners (and increased engagement within an educational setting has been linked to increased academic performance associated with academic achievement [Shute, 2009]), and allow users to experience situations that are difficult (even impossible) to achieve in reality due to factors such as cost, time, and safety concerns [Squire, 2003]. Furthermore serious games can allow learners/trainees to experience situations that are difficult (even impossible) to achieve in reality due to factors such as cost, time, and safety concerns, improve psychomotor skills, enhance retention of knowledge, enhance decision-making skills, interactive learning, options for immediate feedback, and retention of knowledge related to procedures [Mitchell and Smith, 2004]. Serious games have been incorporated within the teaching and training curricula of a large number of professions including medicine/surgery, and the military and due to the large appeal the present users/trainees, particularly the new generation of learners (i.e., the digital natives [Preseky, 2006]) who are accustomed to technology. A number of studies have demonstrated the effectiveness of serious games (an overview was provided in the report of the Summit on Educational Games [FAS, 2006], and further accounts on the effectiveness of serious games are available in the reviews of Bogost [2007] and Gee [2003, 2004, 2007] while the application of serious games in the healthcare field has been reviewed by Feingold [2004]). However, there are concerns that must be analyzed and addressed before serious games become more widespread. The Federation of American Scientists [FAS, 2006] in a summit held in 2005 converged on eight conclusions regarding serious games:

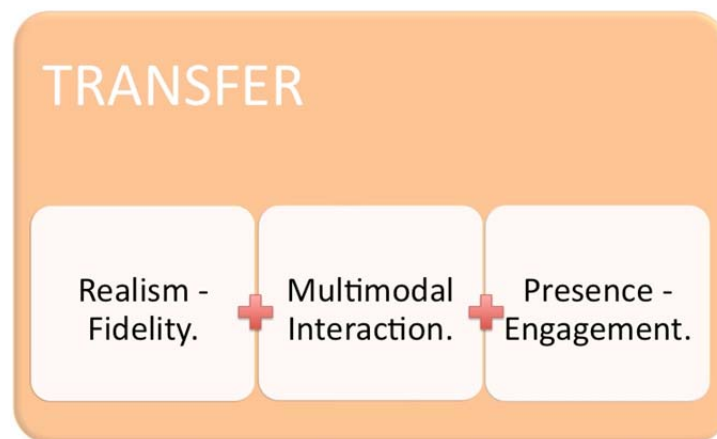
1. It is clear that the modern workforces of technology-oriented countries require the skills that many video games require players to master.
2. Attributes of games could be useful in applications in learning (contextual bridging, increased time on task, improved motivation and goal orientation, personalization of learning, feedback, cues and partial solutions).
3. Games for education differ from games for entertainment.
4. Rigorous research is required to help translate the art and technologies of gaming into teaching, learning, and assessment systems.
5. The video game and educational material industries are inhibited by high development costs and uncertain markets for educational innovations developed as learning games.
6. There are a variety of barriers that inhibit markets for educational games, including market fragmentation, faculty members' and parents' negative attitudes about video games, and lack of evidence for efficacy and evidence-based learning through gaming.
7. Educational institutions are slow to transform practices and organizational systems that take advantage of new technology, including gaming and simulations.
8. There is no serious evidence-based learning framework that currently exists for implementing large-scale evaluations of the outcomes of using educational games.

Furthermore, there is a lack of research that investigates the technological conditions under which learning can be maximized within a virtual simulation/serious game. There is also a lack of research that links virtual learning to proven educational theory and practice. Tashiro [2007], identified several major issues that must be addressed before widespread development of serious games. More specifically:

1. Instructional designers seldom conduct the research necessary to demonstrate their products actually improve learning or skills. In healthcare, an empirically-driven approach becomes especially critical in the context of the United States Institute of

Medicine's call for broadly based core competencies [AACN, 2003; Institute of Medicine, 2000; Institute of Medicine, 2003]. Similar problems exist throughout the educational games and simulations available at the undergraduate level.

2. With few exceptions, commercially available educational games and simulations have not shown to improve what some call critical thinking (including the important higher levels of declarative, procedural, and also metacognitive knowledge) of users while also improving disposition to engage in higher order thinking [Cacioppo, 1982; Caioppo, 1996; Sadowski, 2006]. Such simulations have remained elusive, despite many different types of simulations that are being evaluated.
3. Few commercially available simulations have been developed to mesh sensibly with the strategic needs of undergraduate curricula or with professional development, continuing education, and training programs.
4. There are no commercially available products related to improving learning outcomes or skills competencies that are designed to become part of an evidence-based education framework.



**Figure 1. Relationship among the multiple factors that influence transfer.**

In the context of a virtual learning environment or serious game, transfer can be defined as the application of knowledge, skills, and attitudes acquired during training to the environment in which they are normally used [Muchinsky, 1999]. However transfer is comprised of various other factors as well (see Figure 1). Transfer consists of the capacity to replicate what has been learned within the virtual environment/serious game to the real world environment. The situated learning model suggests that the virtual simulations/environments faithfully represent the real-world environment in order to provide the user the ability to practise under the conditions that they will encounter in the real-world [Brown et al., 1989; Dalgrano et al., 2010].

Fidelity denotes the degree of similarity between the training situation and the operational situation which is simulated [Hays, 1989]. Andrews [1996], defines fidelity as the extent to which the virtual environment emulates the real world. In other words, fidelity describes the extent in which similarities between the real and the virtual world can be measured.

Realism and fidelity are synonymous and can be used to measure similarities from the real to the virtual world [Alexander, 2005]. Tashiro [2010], describes levels of realism as the degree of realism with respect to the visuals/graphics as well as in the procedural rules of an educational game or simulation. However, [Feewerda, 2003], expands upon this appreciation and divides realism into three categories: i) physical, ii) photorealism, and iii) functional realism.

Physical realism analyzes the quality of the physical properties of the environment [Chalmers, 2008]. Photorealism is the term most commonly used for realistic computer imagery. The goal is a resultant rendered image, which is indistinguishable from a photograph of the real scene being portrayed [Chalmers, 2008]. Functional realism describes the ability of an image to provide sufficient information relating to the real scene to enable the viewer to perform an appropriate real task [Ferwerda, 2003].

Hodgkinson [2009], observed that as realism increases, so does the viewer's judgment, guiding them into a circle of "infinite acceptance", and therefore, total realism is unlikely to be reached. However in contrast to Hodgkinson, Chalmers et al. [2009] proposed that fully realistic environments are not needed in order to guarantee an effective learning process. He proposed an effect called *selective rendering*, where the environment will provide high quality rendering only with respect to the details where the user will be focused on, leading to computational resource savings while opening the possibility for a successful transfer process under lower quality virtual environments.

Slater [1994], proposed presence as a subjective measure of quality on virtual environments. Despite the fact that presence and immersion are sometimes used synonymous, immersion refers to the degree to which an individual feels absorbed by or engrossed in a particular experience [Witmer and Singer, 1998], while presence refers to the subjective experience of actually existing within the computer-mediated environment even when one is physically situated in another [Witmer and Singer, 1998]. Immersion is often seen as a measure of technical immersion, focused on the level of technical quality, especially in the areas of picture quality, field of view and level of interaction [Witmer and Singer, 1998]. Slater [1994], describes the difference between the sense of presence that can be reached with books and films, and the sense of presence resulting within, which is totally different than the sense of presence resulting from the immersion with respect to technology. Presence is thus not a direct function of an immersive technology, but rather the technology creates the stimuli which the human uses to build up a mental model of the virtual environment and his/her engagement with it [Schubert and Crusius, 2002]. Therefore, it seems that presence could be used as a tool to quantify a user's engagement within the virtual environment, and not as a measure of realism itself. The difference between presence and immersion is explained by Taylor [2002], who divides the term immersion into *diegetic* and *situated* immersion. Diegetic immersion refers to the immersion of the act of playing a game. On the other hand, situated immersion is not only the immersion resulting from playing the game but also arises by experiencing the illusion of existing within the virtual environment (presence).



The goal of a virtual simulation or a serious game when applied to education/training, is to achieve some amount of transferability that will allow the users/practitioners to perform the acquired skill in the real world on the same way it was performed within the virtual world. In order to achieve this goal, fidelity, immersion and presence can play an important role and therefore, further analysis of the factors that may influence the perception of quality or the feeling of presence is needed.

Storms [1998], found that high-quality sounds (higher sampling rate frequency manipulated) coupled with high-quality visual stimuli increase the perceived quality of a visual display [Storms, 1998], while Winkler and Faller [2005], showed that both audio and video quality contribute significantly to the perceived visual quality, and Mastoropoulou et al. [2003] showed the combination of tempo and emotional suggestiveness of music can affect users' visual perception of temporal rate and duration.

The work described in this thesis investigates multimodal interaction within a virtual environment/serious game and more specifically, the interactions and resulting effects of audio and visual cues. This work is part of a larger-scale effort to examine virtual environment quality, multi-modal interactions, user-specific factors (e.g., personality, learning style, existing knowledge, level of attention, motivation etc.) and their effects on knowledge transfer and retention within virtual simulations/serious games. The long-term goal of this work is to answer the following questions with respect to virtual simulations/serious games: i) "what effect do multi-modal interactions have on knowledge transfer and retention?", and ii) "how much quality is actually needed to maximize transfer and retention?" The question has a number of implications when considering that any training device—be it a virtual or physical simulator—will never be able to completely replicate the real world, and the fact that in virtual worlds we currently ignore the sense of smell altogether and limit the haptic senses (touch and movement). Complete (perfect) multi-sensory quality appears to be impossible to achieve, at least with our current technology. In addition, striving to reach higher levels of rendering quality can also lead to increased development costs, and it remains unclear if such quality is even needed for either enjoyment or knowledge transfer. Within this thesis, as described

below, a number of experiments were conducted that examine visual quality perception of a virtual scene in the presence of various ambient (background) auditory conditions. The results of these experiments have implications for designers, and developers of virtual simulations/serious games, and the educators that employ them. Ultimately, these results bring us closer to the long-term goal of developing effective virtual simulations and serious games that maximize knowledge transfer and retention.

## **1.2 RESEARCH OBJECTIVES**

As previously described, this thesis is composed of a series of seven experiments that examine the role that quality, and multi-modal interactions play with respect to knowledge transfer and retention for users of virtual simulations and serious games. Here, a brief description of each experiment is provided while greater details regarding each experiment are provided in Chapter 3. Each of the experiments builds upon the prior experiment in order to develop a better understanding regarding the roll that different ambient auditory cues play on the perception of visual quality of virtual simulations/serious games. Visual quality can be defined in many ways. Here, visual quality is represented with respect to the *polygon count* (i.e., the number of polygons (typically triangles) required to represent and render a 3D model) and *texture resolution*.

### **1.2.1 Experiment 1**

In this experiment, the perception of visual quality defined with respect to polygon count of a single (static) image of a surgeon's head under four auditory ambient (background) sound conditions: i) no sound, ii) white noise, iii) classical music, and iv) "heavy metal" music, was examined. Results indicate that perceived image quality is dependent on background sound particularly for the higher quality images where classical music leads to the greatest perceived visual quality.

### **1.2.2 Experiment 2**

In Experiment 1, visual quality was defined with respect to polygon count. However, visual quality can be defined in many ways. In contrast to Experiment 1, here polygon count remained constant while visual quality was defined with respect to the resolution of the texture that was mapped to the model varied (once again, the model consisted of a surgeon's head). Similar to the results of Experiment 1, the results of this experiment revealed that the perception of visual quality is dependent on ambient sound and more specifically, white noise can have detrimental effects on the perception of high quality visuals.

### **1.2.3 Experiment 3**

As in Experiment 1, visual quality was defined with respect to polygon count. However, here the visuals were presented to the participants in a stereoscopic 3D viewing environment (as with the previous experiments, a single (static) image of a surgeon's head comprised the visuals). Results indicate that sound does influence visual quality perception particularly for the higher quality visuals within a stereoscopic 3D viewing environment. For the lower quality visuals, the effect is less pronounced. Ambient sound consisting of white noise once again had a detrimental effect on visual quality perception across all of the image conditions.

### **1.2.4 Experiment 4**

This experiment was identical to Experiment 3 except that here visual quality was defined with respect to texture resolution while polygon count remained static. Results were similar to Experiment 3; sound does influence the perception of visual quality. Classical sound and heavy metal music showed to increase in different extent the visual quality perception and white noise showed to decrease the visual quality perception.

### **1.2.5 Experiment 5**

The results of the previous experiments have shown that background sound can affect our perception of visual quality. However, these previous experiments did not consider contextual auditory sounds. In other words, the auditory conditions considered were completely disjoint from the visuals that consisted of surgeons head. In this experiment, visual quality perception, defined with respect to texture resolution, in the presence of ambient auditory sounds was examined. However, in contrast to the previous experiments, here contextual sounds were considered. Results indicate that contextual sounds have a greater effect on visual quality perception over non-contextual sounds. As with the previous experiments, ambient white noise leads to a reduction of visual quality perception.

### **1.2.6 Experiment 6**

Here, Experiment 5 was repeated but within a stereoscopic 3D viewing environment. Results indicate that ambient sound does effect visual perception, particularly for the higher quality visuals and more specifically for the As with the previous experiments, ambient white noise led to a decrease of visual quality perception for all visuals considered.

### **1.2.7 Experiment 7**

The previous six experiments considered simple visual scenes comprised of a single static model and although the results have provided great insight into the interaction of visual and auditory cues, the experiments did not consider user interaction. This experiment builds upon the previous experiments by considering a complex visual environment in the presence of both contextual and non-contextual ambient (background) sound cues where the visual scene was blurred to approximate the effect of varying texture resolution, while conducting a simple task. The task involved navigating through a virtual operating room from a starting position to a point in the room which contained a tray and a surgical drill and picking up the drill. Results indicate that ambient (background) sound does not have any influence on the perception of visual quality irrespective of the level of blurring of the

visual scene or whether the auditory cues were contextual or non-contextual with respect to the visual scene. However, despite having no effect on visual quality, ambient sound did have an effect on the total time it took to complete the task. More specifically, in the presence of ambient white noise, the time to complete the task increased.

### **1.3 THESIS ORGANIZATION**

The first chapter (the current chapter) is the introduction and is comprised of two main sections: Preamble and Research Objectives. The Preamble provided a brief introduction to the thesis research topic. I have explained, in general terms, the constraints/limitations of the field I have been working in, and provided an overview of the terminology relevant to the remainder of the thesis chapters. The Research Objectives provided an explanation of the research expectations, and a brief explanation of the studies conducted within the scope of this thesis.

The remainder of this thesis is organized as follows. Chapter Two provides background information, providing support and relevance to the chosen research topic. The chapter is divided in two main sections. The first section is intended to provide the reader with strong background knowledge regarding the field that this thesis investigates. The second part of this chapter presents a literature review where the relevant and related previous research is described.

The third chapter provides an explanation of the sequence of studies that were conducted within the scope of this thesis. Finally, the fourth (and final) chapter provides a summary of this work, the implications of the experimental findings, and plans for future work.

## 2 BACKGROUND

### 2.1 History

Computer-based alternatives to live training (medical, army, flight, etc.) have become more common in recent years. These include computer-based training systems such as virtual environments, and video games [Alexander, 2005]. However, reaching high levels of skill competency requires novel approaches of training to ensure resources are allocated in a cost effective manner [Ricci, 2002], and virtual technologies such as virtual simulation and simulators, and serious games, offer strong promise and rich possibilities for learning innovation, capable of delivering training in a safe, cost-effective, and engaging manner.

As previously described, serious games have been referred to as “games that do not have any entertainment, enjoyment, or fun as their primary purpose” [Michael, 2006], and “leverage the power of computer games to captivate and engage players/learners for a specific purpose such as to develop new knowledge or skills” [Corti, 2006]. Fundamental to serious games and video games in general, is the concept of “play”. Krentz [1998], proposed that play and work are not opposed to each other since play can be serious and useful when it contributes to the educational process. “Play” is only one part of the functional exercise which take place during an individual’s development, and the other part is non-playful exercises in which the subject “learns how to learn”, not only on the context of play but in that of cognitive adaptation [Piaget, 1977].

Learners are allowed to explore (“to play”) with ideas that they may be too afraid to try in real life, and devise creative solutions to problems as a result [Attewell et al., 2006]. Also, within a serious game (or simulation in general), the player is in ultimate control of the environment and does not need to worry about the consequences, because although they ‘represent real-world systems, the associated costs of participant error is low, protecting them from the more severe consequences of mistakes [Garris et al., 2002].

The benefits provided by serious games are becoming more widespread and accepted within teaching/training curriculums. That being said, there are many questions regarding the effectiveness of serious games. More specifically, can effectiveness of a serious game be guaranteed and if so, how can effectiveness be maximized? However, this is an open problem affected by various factors including the level of engagement [Csikszentmihalyi, 1992].

The topic of serious game effectiveness has been studied for many years but results have been mixed. For example, Hays [2005], conducted a review of the literature (105 articles) on “instructional games” focusing on the empirical research on the instructional effectiveness of games and found that i) the research on the effectiveness of instructional games is fragmented, filled with ill-defined terms, and methodological flaws, ii) some games can provide effective learning for a variety of learners for different tasks but care should be taken not to generalize on the effectiveness of one game in a particular learning area for one group of learners, to all games, in all learning areas for all learners, iii) no evidence indicating that games are the preferred instructional method in all situations, iv) instructional games should include debriefing and feedback in order to provide the learner with an understanding of what happened in the game, and v) players should have access to support to help them understand how to properly use the game. That being said, and as summarized below, there is also plenty of research that indicates serious games do provide effective learning.

- Playing action-based video games can affect cognitive skills related to visual attention [Boot et al., 2005].
- Virtual environments could help to control social anxiety in a controlled manner [Know, 2009].
- Learners are encouraged to reflect more upon their learning when using serious games [Alden, 1999; Oberhofer, 1999].
- Serious games help students to identify particular “complicated points” in their understanding [Francis and Bryne, 1999].

- Serious games use can enliven the presentation of learning and teaching content, producing an increased relevance of textual-data to real-life context [Lowry, 1999; Neral and Ray, 1995].
- Serious games can simplify complex theories providing some level of experience within which the concept can be easily understood by the learner [Neral and Ray, 1995].
- Serious games use can help on the development of key and transferable skills with nominated benefits for communication and social skills [Sutcliffe, 2002], perceptual and motor Skills [Green and Bavelier, 2003], critical thinking skills [Jiwa and Lavelle, 2002], and psychomotor skills [Stone, 2003].
- Play does provide a significant role on formulating the core basics of the learning processes (analyzing the problem, finding the challenge, obtaining a solution) critical for young learners [Piaget, 1977].
- One of the reasons why serious games are really effective, particularly with young learners is due to the familiarity of them on using interactive interfaces (video game consoles) [Aldrich, 2004; Prensky, 2001, 2006].

Despite, the strong support regarding the effectiveness of applying this new tool within the curriculum of learning and training processes, there are also many factors that can be modified, and since learning can be applied to a large amount of cognitive and motor skills, there are several constraints that are worthy of discussion. Facer [2006], suggested that video game-based simulation effectiveness cannot be generalized among all groups of learners, and questioned the transferability of the results obtained after applying serious games to a specific learning group (with respect to age). For Facer, the main difference was between young learners and adult learners, basing his hypothesis's on the different learning processes and methods these groups use [Knowles, 1959; Merriam, 1991].

Griffiths [2002], supports Facer's idea proposing that there is a lack of analysis of the potential use of serious games by adult learners or by learners with specific difficulties. As previously described, this paucity is not due to a lack of interest regarding virtual



simulations and serious games which have been embraced by educators across a wide variety of areas including aviation, military, and surgery [Attewll et al., 2006].

Tables 1 and 2 summarize the most relevant serious games developed and being used up to 2006 [Attewll et al., 2006].

<b>Case study</b>	<b>Learner type</b>	<b>Start date for use</b>	<b>Number of people using game/sim</b>
Max Trax	Learners with numeracy difficulties	December 2001 (commissioned)	7288 learners (up to December 2003)
Skillswise	Learners with literacy and numeracy difficulties	Mar-02	More than 2 million page impressions per week (at May 2006)
Key Skills Trainer	Learners with numeracy problems	Jul-02	4500 registered learners (at December 2005)
MIST	Surgeons in training	1996	220 organizations – number of users unknown
America's Army	Army recruits and potential recruits	Jul-02	7,309,745 users in 2006
Strategy Copilot	Business learners	2000	3000–3500 learners (June 2006)
Krucible	Learners aged 14–19	Sep-03	350 users
Virtual School	Middle management leaders in schools	Sep-03	2000 middle leader participants and nearly 700 leadership coaches

**Table 1. The most relevant serious games developed and used up to 2006 (first of two tables).**

**Reprinted from Atwell [2006].**

<b>Type of game/sim</b>	<b>Description</b>	<b>Objectives of use</b>	<b>Subject</b>	<b>Context of use</b>
Game	Racing game with numeracy challenges included	To support the learning of numeracy	Maths	A service available at Learn Direct centres around the UK
Games	Skillswise provides a range of resources for training informal learners with specific literacy and numeracy needs	To support learners with literacy and numeracy needs	Literacy and numeracy	Web resource available at <a href="http://www.bbc.co.uk/skillswise">www.bbc.co.uk/skillswise</a>
Game	The game is embedded into a set of materials which take 200 hours to complete	To support learners developing their key skills (Levels 1 and 2)	Literacy and numeracy	The materials can be used in any learning context and alongside most post-16 courses
Simulation	Simulation used for training surgeons	To accelerate the learning time for surgeons	Medical surgery	In medical teaching hospitals and schools

Gamesim	Training simulation game for new recruits and potential recruits	To encourage young people to join the army and support teaching of military skills	Army training	Self-directed exploration of the army training environment
Simulation	Simulation with role playing capability.	To provide strategy training for business learners	Business studies	Business schools, universities and companies
Simulation	Krucible is an interactive learning tool for learners studying physics between ages 14 and 19	To support 14–19 physics teaching	Physics	Sixth form colleges and FE colleges
Training simulation	Simulation game based on scenarios in a school setting	Application of learning- centred leadership	School leadership	Part of the Leading from the Middle programme for middle managers. 5hrs 45 minutes allocated for Virtual School within other activities

**Table 2. The most relevant Serious games developed and used up to 2006 (Second of two tables).  
Reprinted from Atwell [2006].**

Despite the growing popularity of serious games and virtual simulations, before their use becomes more established and widespread, as described above, there are various open issues that must be addressed in order to ensure the development of effective serious games. Tashiro et al. [2007], developed a typology of serious games for healthcare education and explored the strengths and limitations of serious games for improving clinical judgment. They found that generally, serious games do not meet the standards for instructional materials set forth by the United States National Research Council, and the Federation of American Scientists and identified seven areas that require research and improvements for the effective development of serious games:

- Disposition to engage in learning.
- Impact of realism on learning.
- Threshold for learning.
- Process of cognitive development during knowledge gain.
- Stability of knowledge gain (retention).
- Capacity for knowledge transfer to related problems.
- Disposition toward sensible action within clinical settings.

Related to some of the points described by Tashiro [2003], Alexander [2005], proposed three main factors that could drive training transfer: i) quality, ii) immersion, and iii) presence. It is important to note that these constructs are not mutually exclusive, and more specifically, some of them appear to be functions of each other and exhibit similar effects. He also proposed that the higher the degree of transfer, the more successful a training system/method can be considered to be. However, what is transfer within the simulation/serious games context? What is needed to guarantee a higher degree of transfer?

Transfer as was mentioned on the introduction section have been defined as the application of knowledge, skills and attitudes acquired during training to the environment in which they are normally used [Muchinsky, 1991]. In other words, transfer is an outcome of a successful training process. Nevertheless, transfer can be either positive, negative, or null [Gick and Holyoak, 1987]. According to Alexander [2005], positive transfer occurs when the practitioner performs the acquired skill in a real-world setting better due to the practice performed within the learning/training virtual environment. Negative transfer occurs when the practitioner performs worse in the real-world as opposed to the virtual world, and this usually happens because trainees apply behaviours to the real-world that are appropriate only in the training environment. Null transfer occurs when there is no difference between the performances of the real and the virtual world, implying that the training had no effect.

Van Hulst [1998], defines transfer of complex cognitive skills such as: i) rule-based transfer, and ii) schema-based transfer. Guilligan [2010], explains rule-based transfer as a method that focuses on the familiarity of situations where skills acquired in the training setting can be used in the same way as in the target setting. This definition emphasizes the idea that the training and target setting should share the common elements that are relevant for the task. Rule-based transfer will therefore exhibit greater benefits in cases where the trainee has performed and extensive practice over a variety of training cases to develop a broad knowledge about the skill acquired [Guilligan, 2010].

Schema-based transfer employs the schema-based theory of learning [White, 1993], where the learner develops a generalized knowledge over a specific skill that allows him/her to make decisions when presented with unexpected situations based on the general understanding acquired. In addition, the more robust (i.e., more information) those cognitive schemata are, and the better they are integrated with other cognitive schemata, the more likely it is that transfer will actually occur [Van Hulst, 1998].

However, this raises the question of how one can measure transferability on an acquired skill? Wickens and Hollands [2000], proposed comparative tests of transfer, measuring post-training performance on a real-world task between two groups: an experimental group that trained with the new technology, and a control group that trained directly in the operational setting. From this point of view, the only way to test the effectiveness of a specific method in terms of transferability, is to conduct a randomized control trial (RCT) (or randomized comparative trial). An RCT is a specific type of scientific experiment, and the preferred design for a clinical trial. A RCT is often used to test the efficacy of various types of intervention within a patient population [Chalmers, 1981].

For the purpose of this thesis, the primary focus is not on the different types of transfer or how to measure them. Rather, within this thesis, the focus is on analyzing the perceptual effects of audio-visual (multimodal) interaction within a virtual simulation/serious game environment and in particular, how ambient auditory cues affect visual quality perception. It is anticipated that the findings resulting from this thesis will lead to a greater understanding of the role that multi-modal interactions play with respect to virtual simulations and serious games and ultimately lead to the development of more effective serious games and virtual simulations.

## **2.2 Literature Review**

### **2.2.1 Fidelity**

Fidelity denotes the degree of similarity between the training situation and the operational situation which is being simulated [Hays, 1989]. In other words, fidelity describes the extent to which the virtual environment emulates the real world or how realistic the virtual world is [Andrews, 1996]. Fidelity can be measured on multiple continua, and a large number of subcategories have been described in the literature [Allen et al., 1986; Hays and Singer, 1989; Lane and Alluisi, 1992; Lintern et al., 1990; Rehmann et al., 1995]. However for the purposes of this thesis, three main subcategories of fidelity will be discussed: i) psychological, ii) functional, and iii) physical.

Psychological fidelity is described as the degree to which the simulation replicates the psychological factors (e.g., stress, and fear) experienced in the real-world environment, engaging the trainee in the same manner as the actual equipment would in the real world [Kaiser and Schroeder, 2003]. Psychological fidelity is basically related to the feeling of presence, a concept that is described in the following section.

Functional fidelity has been described as the degree to which the simulation acts like the operational equipment in reacting to the tasks executed by the trainee [Allen et al., 1986]. This type of fidelity is strongly related with physical fidelity that is defined as the degree to which the physical simulation looks, sounds, and feels like the operational environment in terms of the visual displays, controls, and audio as well as the physics models driving each of these variables [Baum et al., 1982]. Physical fidelity encompasses a number of dimensions (visual, auditory, vestibular, olfactory, proprioceptive, etc.). A large literature base exists concerning these different dimensions of physical fidelity [Becker et al., 1995; Borah et al., 1977; McMillan et al., 1990; Rinalducci, 1996 ], but as will be described below, the research described in this thesis focuses on visual and auditory factors only.

There is a relationship between fidelity and transfer, but it is not as simple as ‘more is better.’ Based on Thorndike’s theory of elements [Gick and Holyak, 1987], the implications are that if the level of fidelity captures the critical elements/properties of the skills/tasks you wish to teach, that level of fidelity is sufficient *even if* it noticeably deviates from the real world [Alexander, 2005]. Despite the three sub-categories considered, the question still remains. How much fidelity is needed to maximize transfer? This question may be further complicated when consideration is given to each of the categories. In other words, one can consider each of the fidelity sub-categories as a variable within a simulation/serious game, each of which can be individually adjusted. This of course makes it difficult to determine the proper amount of “sub-category” fidelity to maximize transfer but it does open a large number of potential future studies.

Generally, it is assumed that a high level of fidelity is needed for each of the three fidelity sub-categories to fully replicate a real environment what would led into a successful performance in operational conditions [Alexander, 2005]. Turner et al. [2000], describe that trainees within a virtual simulation tend to demand high physical-fidelity because they think about previous exposure to high-fidelity simulators (physical) and are concerned that performance evaluation accurately reflects anticipated real-world performance. However, other characteristics within a rendered (virtual) environment such as the polygon count of the 3D models within the environment and realistic texturing (and texture resolution), may be noticed and appreciated by trainees generating an unknown effect on the outcome. However, Alexander [2005], describes that an appropriate level of fidelity is dependent on the skills or behaviours that are to be trained and more specifically, on the nature of the task itself that that is expected to be transferred. If both the training system and operational setting share properties with respect to the objectives of the training, other aspects of the training system may tolerate lower levels of fidelity without compromising training effectiveness [Alexander, 2005]. Furthermore, as previously described, striving to reach full fidelity (which typically involves high polygon count models, and higher texture resolutions, etc.) leads to increased computational processing in addition to higher development costs [Gilligan, 2010]. In addition, it remains unclear if such fidelity is actually needed for either enjoyment or knowledge

transfer [Hawkins and Orlady, 1993; Moroney and Moroney, 1999]. Thus, an important compromise must be made between physical fidelity, related costs (computational and development), and training effectiveness, such that an adequate match can be made between training and real environmental elements and the logical structure of tasks [Alexander, 2005]. The relationship between fidelity and transfer is further complicated by the finding that training effectiveness can be increased by adjusting the level of fidelity/difficulty over a sequence of training experiences [Lathan et al., 2002]; this is more commonly known as *scaffolding*.

### **2.2.2 Realism**

Realism is essentially another measure of fidelity since realism is an attribute that can be measured on each of the sub-factors that compose fidelity (visual, auditory, vestibular, olfactory, proprioceptive, etc.) [Lengyel, 1998]. After careful analysis of the literature, for the purposes of this thesis, fidelity and realism will be treated synonymous.

As with fidelity, how much realism is needed in order to provide a practical tool for education? Haders [2008], specifically raised this question with respect to simulation for clinical education.

Cittan [2008], developed an engine to produce the “highest realistic environment” including the use of stereoscopic 3D to produce visuals with depth as in the real-world, but questions whether such a high level of realism is actually needed. Initial testing has confirmed that the level of immersion is highest when interacting with a stereoscopic 3D scene.

In a study conducted by Brogni [2006], participants were placed within an immersive virtual simulation (a replica of a street with virtual people walking on the street) and their heart rate responses at various stages were measured and analyzed. Brogni was trying to develop an objective measure of realism/fidelity presented in a scene. The measures

employed by Brogni were all subjective (e.g., practitioner perception measured through the use of surveys/questionnaires after the completion of the study) he proposed stress level as a measure of realism. It has been demonstrated that both positive (e.g., entertainment-induced energy and/or motivation), and negative (e.g., conflict-induced fear and/or adverse physiological reactions), types of stress have been shown to enhance skill retention as well as the transfer of training from the simulation to the real-world [Mayer and Volanth, 1985; Williams, 1980]. The primary goal of “stress training,” is to provide practice under conditions similar to those likely to be encountered in the operational setting [Alexander, 2005].

More recently, studies have examined the effects of adding contextually-relevant stress to training paradigms on the improvement and transfer of skills concluding that in general, stress is an important factor on quality measure, although its effectiveness can be strongly related to the type of skills studied [Driskell et al., 2001; Morris et al., 2004]. However, further research is required to examine the relationships between the types of stress and skill transfer in general (and as influenced by individual differences), and to determine how long the beneficial effects of stress training persist [Morris et al., 2004].

As previously described, striving for perfect quality and realism may not be feasible given the high computational and development costs associated with doing so. That being said, there are a number of widely used computer graphics methods and techniques available that are available to allow for high quality/realism while minimizing computational requirements.

Alexander [2005], and Chalmers [2003], proposed “selective rendering” as a solution to the quest of high quality and the high cost involved in achieving it. Selective rendering involves rendering high quality visual/graphics for only the portion of the scene that the user is viewing. This idea is based on previous work of the human visual system and has been used to improve the quality of displayed images [Greenber, 1997; McNamara, 2000; Myszkowski, 2001], or to reduce the complexity of the models of objects in the scene without reducing the viewer’s quality perception of these models [Luebke and



Hallen, 2001; Watson, 2001]. More specifically, due to the *intentional blindness phenomena* that is presented when attention is focused onto an item in a scene, details regarding the remainder of the scene, including lighting, resolution, level of detail, and even the presence/absence of objects can, literally go unnoticed [Mack and Rock, 1998]. It can also be explained by the peripheral vision phenomena where the human eye processes detailed information from a relatively small part of the visual field only. The lack of visual acuity in the human peripheral vision system has also been exploited to reduce the overall computation time of an image by rendering the parts of a scene in the peripheral vision at a lower quality [Chalmers, 2003].

Although selective rendering is one solution to the high computational cost associated with generating high quality environments, it only focuses on one of the quality factors, and more specifically, visuals, despite the fact that virtual environments are multi-modal and at the very least, will, in addition to visuals, provide auditory cues/information.

### **2.2.3 Immersion**

Immersion refers to the degree to which an individual feels absorbed by or engrossed in a particular experience [Witmer and Singer, 1998]. Immersion is therefore based on the extent to which the virtual environment supports an illusion of reality that is inclusive (denoting the boundaries of physical reality), extensive (the range of sensory modalities accommodated), surrounding (the size of the field of view), and vivid (the display resolution, richness, and quality) [Slater and Wilbur, 1997]. Each of these dimensions may have a weight of how much implication it has on the perception of a particular reality. Based on these scales, Slater and Wilbur [1997], suggest that immersion can be an objective and quantifiable description of what a video game, or simulation provides.

## 2.2.4 Presence

Presence refers to the subjective experience of actually existing within the computer-mediated environment even when one is physically situated in another environment [Slater and Steed, 2000; Witmer and Singer, 1998]. Lombard and Ditton [1997], describe presence as very valuable in training tools because it increases motivation and provides a more engaging experience.

However, the question remains if there exists any specific way to guarantee the feeling of presence within an environment. Witmer and Singer [1994], conclude that video resolution and the connectedness and continuity of visual elements are highly important for generating perceived presence, while Slater [2010], states that visual display lighting within a virtual environment does not affect presence.

Cheng [2005], proposed a study to generate a better understanding of presence by examining how presence could be broken-down (by altering behavior and realism). Basically he wanted to measure presence (within a game environment) when the game behaves normal (coherent with real life) and what would happen if the behavior were not coherent. He concludes that understanding the facts of what produces an immersive experience (either a normal behaved environment or an environment with a non-coherent behavior), is difficult to determine. After analyzing real and non-realistic environments, Cheng [2005], proposed that the more realistic graphics/visuals of the virtual environment are, the more consistent the quality is perceived.

In general, it remains unclear how increased presence affects the transfer of skills or behaviors, or when increased presence is advantageous. A better understanding of the concept of presence may provide insight to its use as a mechanism for increasing training effectiveness [Alexander, 2005].

### 2.2.5 Multimodal Interaction

Recall that the definition of physical quality encompasses a number of dimensions, and more specifically, visual, auditory, vestibular, and olfactory. These dimensions can be correlated with the multisensory perception of the real world. Cross-modal effects relate to the impact on the perceptual experience of one sensory input that the presentation of an additional sensory input can have [Chalmers, 2008]. Previous experience, for example, demonstrates that cross-modal effects can be considerable, even to the extent that large amounts of detail in one of the senses may be ignored when in the presence of other sensory inputs [McGurk and MacDonald 1976; Ramic et al., 2007; Rimell et al., 2002].

Other studies have been conducted to determine the relationship between the senses and how this relation could affect the perception of reality within a virtual environment. Woods et al. [2011], discovered that background sound (noise) could have an effect on the perception of food gustatory properties (i.e., sugar level, salt level), food crunchiness and food liking. They found that background noise has three effects on food perception: i) food saltiness and sweetness was reduced in the presence of loud background noise, ii) food was “crunchier” in the presence of background noise, and iii) there was a correlation between the liking of the food and the liking of the noise.

Other studies have shown that the introduction of smell increases the user’s sense of “presence” in the virtual environment [Dinh, 1999; Zybura, 1999]. Smell has been effectively used to treat soldiers returning from Iraq with post-traumatic stress disorder [Pair, 2006].

Dhin [1999], supported the premise of the importance of multisensory input on memory and the sense of presence in virtual environments. He also highlighted that audio cues produce a higher sense of presence and better retention compared to the cues of other modalities. Zhang [2005], analyzed the introduction of auditory and/or visual training feedback and found that feedback in general improves task performance and user

satisfaction, but it could not be determined if audio or visual exerted more or less influence on users satisfaction.

In order to provide a better understanding of the multimodal interaction phenomenon greater detail regarding the operation of the human visual system is provided. There exist two main aspects that define visual perception: i) spatial perception, and ii) temporal perception. Spatial perception is highly dependent on visual attention or in other words, spatial perception is the ability to perceive or otherwise react to the size, distance, or depth aspects of the environment. Temporal perception refers to the ability to perceive the passage of time and comes into play when you need to estimate or reproduce the length of events, compare the length of two or more events, or place events in the order in which they have occurred. However, these are not the only factors that can affect visual perception. More specifically, spatial frequency, which refers to the levels of detail or image sharpness also plays a role in visual perception. Furthermore, the visual sense has a frequency sensitivity that is three times lower than the audition.

Previous work has demonstrated that with respect to visual quality, higher quality graphics serves to focus trainees attention initially, however on a long period of time analysis graphics showed to be less important than other attribute for sustaining trainees attention [Raths, 2006].

With respect to audio-visual interaction, the ventriloquist effect (also known as ventriloquism) is a phenomenon that has been extensively explored in an attempt to develop an understanding of the relationship among audition and vision when both senses are working together.

There are two types of ventriloquism: i) temporal, and ii) spatial [Berlestone 2003]. In *temporal* ventriloquism, the audio and visual stimuli are delivered at different times but the user feels that they being delivered at the same time. With *spatial* ventriloquism, the audiovisual stimuli are delivered from different spatial positions, but the user perceives that the stimuli have both been delivered from the same location.

Bertelson [2003], examined the perception of audio-visual cues when both cues were presented at separate locations and at separate times and concluded that we perceived the audio cues to be synchronized with the visual stimuli (*temporal ventriloquism*). However, it was also concluded that depending on the location proximity (between the visual and audio sources), the perception of the time in which the stimuli are received might vary.

Aschersleben [2003], analyzed spatial ventriloquism and discovered a strong visual bias of the apparent auditory location and the occasional observation of a much smaller auditory bias of the apparent visual location. In other words, when the audio source location differs from the visual location, both the auditory and visual locations are affected, and participants perceived the sources as coming from different locations.

Vroomen [2004], demonstrated temporal ventriloquism in the flash-lag effect (FLE). FLE is a visual illusion in which a flash appears to lag relative to a moving object. Commonly, a sound presented in close temporal proximity to a visual stimulus can alter the perceived temporal dimensions of the visual stimulus (*temporal ventriloquism*). Vroomen's research demonstrated that a sound can alter visual perception in the temporal domain. More specifically, a sound sharpens and attracts the occurrence of a flash when it is in temporal conflict with it.

Morein-Zamira [2004], analyzed whether irrelevant sounds can influence the perception of lights in a visual temporal order judgment task (i.e., participants judged which of two lights appeared first). Study results were based on participant brain activity measurements. The study supports the phenomena proposed by the temporal ventriloquist phenomenon where in the temporal domain, the auditory system influences visual perception when the two sources of information are in conflict.

Shams [2000], proposed the following premise: "what you see is what you hear", implying that visual information dominates audition. However, various studies have shown that auditory information can affect the quality perception of an unambiguous

visual stimulus to create a striking visual illusion (e.g., the presentation of two auditory beats presented in conjunction with a visual cue (a spark) leads to the perception of two visual sparks instead of one [Shans, 2001]). Although it is generally believed that vision dominates the other senses, visual perception can be manipulated by other sensory modalities (in this case hearing).

Burr [2008], shows that auditory stimuli tend to dominate visual stimuli, but the domination is not total, and varies somewhat from individual to individual. An attempt was made to predict the dominance of audition over visual stimuli by applying some quantitative audio-visual model combination but the prediction was imperfect. Results showed that when auditory and visual stimuli are in conflict (presented with a small lag in between them), sound tends to dominate vision, but not totally. The pattern of results was roughly consistent with a model of optimal cue combination, but the quantitative predictions were not accurate.

Recanzone [2005], measured the interactions of visual and auditory stimuli in a temporally based task in normal human participants. The results showed that the auditory system has a pronounced influence over the visual temporal rate perception (how often the continuous visual stimulus were presented). This influence was independent of the spatial location, spectral bandwidth, and intensity of the auditory stimulus. However, the influence was strongly dependent on the disparity in temporal rate between the two stimuli, and more specifically, this influence is dependent on how similar (with respect to frequency) the patterns used to present the stimulus were.

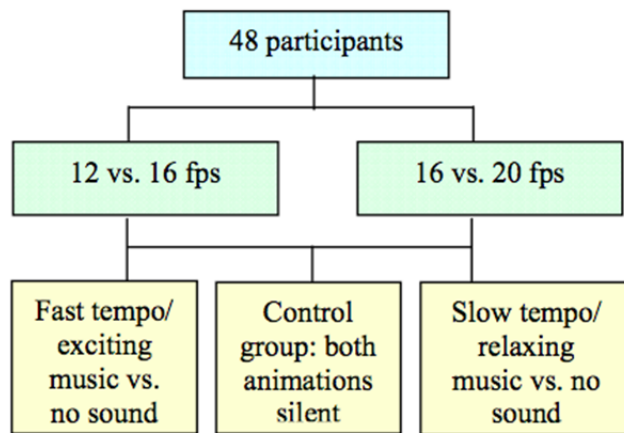
After analyzing the ventriloquism phenomena and demonstrating the audio-visual interactions within the real world, for the purpose of this thesis, the next step is to examine audio-visual interaction within a virtual environment (serious game).

Storms [1988], explored the auditory-visual cross-modal perception phenomena with a focus on visual quality perception through a series of experiments. He began with a pilot study that was designed to investigate the perceptual effects resulting from manipulating

visual display pixel resolution and auditory display sampling frequency. A second experiment was designed to investigate the perceptual effects of manipulating visual (static) display Gaussian noise level and auditory display Gaussian noise level. The visual display consisted of a radio, and the auditory stimuli consisted of an obscure portion of alternative music composed by a fragment of a song (“A Forest” from the CD “Mixed up” by “The Cure”), The song was chosen so participants would not have any preconception or preferences about the musical choice presented. Participants were asked to rate the quality of the visual and the audio presented in conjunction. Results showed significant auditory-visual cross-modal perception phenomena relevant to virtual environments and multimedia developers. A third experiment (referred to as “Static Resolution Non-Alphanumeric” was designed to investigate the perceptual effects of manipulating visual (static) display pixel resolution and auditory display sampling frequency. The visual display consisted of a fruit-flower scene, while the auditory display consisted of the fragment of the previously used song whose frequency was modified. This particular song also had no vocals and therefore, participants could not use any changes in speech as a cue. The task of the subject was to rate the perceived quality of auditory-only, visual-only, and auditory-visual displays via a Likert rating scale ranging from 1 (low) to 7 (high). Based on the outcomes of these three experiments, Storms [1988], suggests that medium and high auditory cue quality in conjunction with high quality visuals increases the visual quality compared to the visual alone conditions. These three experiments also indicate that low auditory quality in conjunction with high visual quality can lead to a decrease in the perception of audio quality compared to the audio alone condition.

Mastoropoulou [2004, 2005], has conducted plenty of experiments to develop an understanding of how auditory cues can affect smoothness perception of a virtual scene, focusing on the pursuit of the threshold point in which sound cues can compensate for the reduction of frame rate (number of frames per second) within a virtual animation. In one of her earliest studies, she hypothesized that the addition of auditory cues (sound of footsteps walking and running) to an animated sequence of images can create the impression that the temporal frame rate was higher than it actually was and therefore the duration of the audio-visual clip would seem shorter. She expected that the influence of

auditory cues to be more profound in the case of fast tempo music, but investigated the effect of both slow and fast tempo music. Results showed no significant difference on the perception of the length of the clip under the different auditory conditions. Within this study, she first conducted three pilot studies to test and refine her experimental design (Figure 2). In the first pilot study, participants had to watch an animation (first person perspective of a walk around a forest) with and without sound (the sound's tempo was either slow or fast and was randomly chosen). Participants completed a questionnaire after the study and were also asked to choose which of the two versions (the one with or without sound) of the animation they perceived lasted longer and in which of the two animations, the velocity of the motion involved in the scene was higher. Only two of the 12 participants realized that the visual part of both the animations they had watched was the same (with respect to duration and motion speed) and only another two of the participants replied that they found the motion faster in the silent animation.



**Figure 2. The experimental design of Mastoropoulou 2004. Reprinted from [Mastoropoulou, 2004]**

In her second pilot study, Mastoropoulou [2004], focused on the participants perception of the scene velocity of the animated sequences. To avoid any viewing order effects (preconditioning, training due to previous exposure, etc.), different animations of scenes with similar content were displayed in pairs. In each pair, one animation was silent and the other contained background music of high tempo and distinctive rhythmic quality. All of the animations used were based on two sequences of images that were slowed down to



obtain the various slower frame rate versions. The camera was moving at exactly the same speed (i.e., same scene velocity) and followed the exact the same motion path, thus the animations had exactly the same duration and the same motion characteristics. Only two of the 13 participants correctly recognized the animation with the faster motion implying that the audio did effect the velocity perception.

A third pilot study was conducted in 2004 to determine the two music clips that would be used within the main experiment. The candidate clips for the “fast tempo, exciting” music clip were excerpts from rap musical pieces, which were rated in terms of fast tempo and highly rhythmic quality. The candidates for the “slow tempo, relaxing” music clip were excerpts from music pieces used for relaxation and meditation and were rated in terms of slow tempo and relaxing quality.

In Mastoropoulou’s [2004], main study, participants were informed that they should carefully watch two computer-generated animations, one silent and one accompanied by music (except for the control group that watched two silent movies), in random order, and when finished, they would have to answer some questions. In the questionnaire, participants were asked which of the two versions of the animation lasted longer and which of the two versions contained a higher scene velocity. They also had to comment on the music and more specifically, whether they liked it and if so, how much, whether they found it relaxing or exciting, and whether it was familiar to them or not.

The results from Mastoropoulou’s [2004], work has shown that slow tempo-relaxing music has an opposite effect than her expected effect and that the coupling of an animated sequence with an arbitrary music of high tempo cannot universally create the impression that the scene temporal rate is higher and therefore that the duration of the audio-visual clip is shorter.

However, based on her initial study described above, Mastoropoulou [2005], conducted another study where she hypothesized that it would be more difficult for participants to distinguish differences in the motion smoothness between audio-visual composites than between silent animations. Two conditions were considered: i) “sound”, and ii) “no

sound” (control condition). It was also hypothesized that familiarity with animated computer graphics would help the participants perform the experimental task more efficiently than participants without any prior experience.

During each session, participants viewed pairs of clips depicting walkthroughs, one after the other, and after viewing each pair, their task was to indicate which of the two walkthroughs they perceived had smoother motion. The two clips in each test pair had the same visual content, but were rendered at varying frame rates (either different or the same within each pair) and one of the sounds condition was applied.

Results of this study showed that participants with computer graphics familiarity were generally better at distinguishing differences in motion across both conditions. However, in the presence of sound, not even those with computer graphics familiarity could escape the influence of the sound effects. More specifically, the type of sound effect can be used to differentiate between different frame-rates. The null hypothesis that stated there is no difference between silence and sound effects regarding the ability of the viewer to detect smoothness/jerkiness variations was rejected.

Hulusic [2010], presented a method for reducing the amount of effort required to render the visual aspects of a game by exploiting movement related-sound effects. A detailed psychophysical experiment was conducted to investigate how camera movement speed and sound affects the perceived smoothness of an animation. There were two-research hypotheses analyzed in Hulusic’s experiment, the first being that the camera movement speed will affect smoothness perception and the second being that sound effects arising from the virtual footsteps of a participant affect the perception of animation smoothness. Results indicate that walking (slow) animations were perceived as smoother than running (fast) animations. It was also found that the addition of sound effects, such as footsteps, to a walking/running animation, affects the animation smoothness perception. When a running footstep sound was applied to a walking (slow) animation, participants perceived the animation to be faster.

In a related study conducted by Jumisco [2009], it was observed that computer generated image quality is different than participant perceived visual quality. More specifically, participants did not perceive the quality of an image to be as good as the corresponding rendered version of it. He conducted a qualitative study in which participants had to rate a computer created image and the scores were compared against the rates of the real image. He also proposed that in the future, further research regarding stereoscopic 3D rendered images is needed. He also proposed that a better understanding of multimodal dynamic perception could lead to new possibilities for technical quality optimization.

Bonneel et al. [2010], examined the influence of the level of detail of auditory and visual stimuli in the perception of audio-visual material rendering quality. In each trial of their experiment, participants were presented with two sequences, each sequence of an object falling on a table, bouncing twice and producing audible bounce sounds. One of the sequences was a reference (highest quality with respect to sound and graphics), while for the other sequence, auditory and visual levels of detail varied (auditory level detail was defined with respect to modal synthesis while visual level of detail was defined with respect to the *bidirectional reflection distribution function* (BRDF)). The participants' task for each trial was to rate, on a scale between 0 and 100, the similarity of the falling objects in the two sequences. They observed significant interactions between visual and auditory level of details and the perceived material quality. In other words, visual level of detail was perceived to be higher as the auditory level of detail was increased.

## **2.3 Summary**

Perception in the real world is multi-sensory and thus any serious game and virtual simulation should include an appropriate level of realism associated with each of the senses, including any cross-modal effects [Chalmers, 2003]. Seitz [2006], demonstrated that a multisensory training procedure facilitates visual learning and results in significantly faster learning than unisensory visual training. Furthermore, Shams [2008], proposed that training protocols that only involve unisensory stimulus, are not as effective as those that involve multisensory stimuli. The benefits on multisensory training over

unisensory training have been demonstrated through various studies whose results indicate that a within multisensory training environment presence and immersion are increased [Seitz, 2006], and it may lead to an increase in the perception of quality, both of which may ultimately improve transfer.

The first attempt to provide a multisensory environment within a simulation was made by Morton Heilig's Sensorama of the early 1960's [William et al., 2003], which presented stereoscopic 3D images and included user movement, stereo sound, wind and aromas. Unfortunately Heilig was unable to find financial backing for his device and nothing further was developed [Chalmers, 2009]. As described above, there has been considerable previous work regarding the effects of audio-visual interactions. However, little work has considered the effect of sound on visual quality perception. Given the potential computational savings which may arise if sounds can in fact lead to an increase in visual quality perception, it was decided that within the scope of this thesis, the effect of visual quality perception in the presence of ambient (background) auditory stimuli. The decision to investigate this was supported by the fact that serious games and virtual simulations will typically include visual and the auditory cues (haptic cues are also becoming more widely used but are not considered here).

With respect to "image quality" there are few studies that have proposed methods to measure image quality and factors that could affect it. Sullivan [2009], proposed to measure visual quality based on the physic parameters of the animations presented to the participants (i.e., physics-based laws). Han [2005], measured brain activity under different visual conditions (i.e., human movies, cartoon clips, etc.), and observed that different neurological and functional processes were engaged with respect to perceiving quality in the real and virtual world. In terms of factors that could affect visual quality, Chen [2008], established that environment lighting is fundamental on the visual quality appearance.

The second dimension that will be analyzed within the scope of this thesis is audition. However, in contrast to the visual dimension, within the scope of this thesis, auditory quality will not be examined; auditory cues here will only be used to determine whether

or not they can affect visual quality perception. With respect to auditory cues, there are several studies that have examined their effect on other modalities [Woods 2010; Conrad et al., 2010]. It has also been well established that sound (and music in particular), can affect moods, emotions and behavior of both individuals and groups, and music is considered to lie on a continuum from highly stimulating and invigorating to soothing or calming [Gaston 1968; Hallam et al., 2002]. Finally sound and music can also influence task performance. For example, a study by Morsbach et al. [1986], determined that the cries of an infant distracted listeners who were instructed to pay attention to a simple cognitive task. Building on this study, Chang and Thompson [2011], proposed a study to test the effects of different sounds on a task performance. They applied different auditory cues (non-related to the task) and measured the level of distraction that the participants showed due to the auditory cues while performing a math-based task. Results showed that participants were more distracted by the sound that involves human vocalization than the ones that had no vocalization involved.

Conrad et al. [2010], examined the effect of background sound (including classical music (Mozart), and heavy metal music to induce “auditory stress”) on laparoscopic surgery. They found that stressful music (e.g., heavy metal) had a negative impact on the time required to complete a laparoscopic task (task completion time) but it did not impact task accuracy. They also found that classical music had a variable effect on task completion time but resulted in greater task accuracy across all participants (laparoscopic surgeons).

### 3 CHAPTER 3 – EXPERIMENTS

Within the scope of this thesis, seven experiments were conducted, each building upon the previous one. The first six experiments were conducted to develop a greater understanding of audio-visual cue interaction and more specifically, the effect of ambient auditory cues (both contextual and non-contextual with respect to the visual scene) on visual quality perception defined with respect to polygon count and texture resolution within both stereoscopic 3D (S3D) and non-S3D viewing environments. In addition, to examining the audio-visual interaction, the final experiment also considered the effect of auditory cues on the time required to complete a simple task within a virtual environment.

A summary of the experimental conditions are as follows:

- Sensory integration (audio-visual) with non-contextual auditory cues:
  1. Visual quality defined with respect to polygon-count with non-contextual ambient auditory conditions.
  2. Visual quality defined with respect to texture resolution with non-contextual auditory conditions.
  
- Sensory integration (audio-visual) given by the type of visual presentation (stereoscopic 3D):
  3. Visual quality defined with respect to polygon count (within a stereoscopic 3D viewing environment) with non-contextual ambient auditory conditions.
  4. Visual quality defined with respect to texture resolution (within a stereoscopic 3D viewing environment) with non-contextual auditory conditions.

- Sensory integration (audio-visual) with contextual auditory cues:
  5. Visual quality defined with respect to texture resolution with contextual auditory conditions.
  6. Visual quality defined with respect to texture resolution (within a stereoscopic 3D viewing environment) with contextual auditory conditions.
  
- Sensory integration (audio-visual) while performing goal directed actions:
  7. Visual quality defined with respect to scene blurriness under both contextual and non-contextual auditory cues.

Greater details regarding each of the experiments are presented below.

## **3.1 GENERAL CONDITIONS**

### **3.1.1 Participant Recruitment**

Participants recruited in all seven experiments were all students at the University of Ontario Institute of Technology (across various faculties including the Faculty of Business and Information Technology, faculty of Health Sciences, Faculty of Engineering and Applied Sciences, Faculty of Faculty of Social Science and Humanities, Faculty of Energy Systems and Nuclear Science, and Faculty of Education) but none of them were enrolled in any of the courses taught by Dr. Bill Kapralos (or any of the other researchers involved in this work). None of the students were also enrolled in any of the courses for which I was a Teaching Assistant for. Participants did not have to meet any special requirement to participate in our studies, in terms of previous knowledge or familiarity with the technology and they did participate on a strictly voluntary basis (i.e., they were not compensated in any way). All of the participants were verbally asked (by the experimenter) if they had any auditory or visual disabilities, and none of the participants reported any. In all of the seven experiments, the age range of the

participants was between 19 and 23 years. All of the experiments conducted within the scope of this thesis abided by the University of Ontario Institute of Technology Research Ethics Review process and the Helsinki Declaration.

### **3.1.2 Experimental Setup**

All visual conditions were presented to the participants on an Acer Aspire laptop with a 15.6" screen size and a resolution of  $1366 \times 768$ . The size of each image within the display was  $800 \times 630$ . The auditory stimuli were presented over a pair of Sony MDR 110LP headphones. The distance between the chair where the participants were seated and the table where the laptop was placed, was always the same (50cm from the back legs of the chair to the table). Participants were not allowed to move the chair but the distance of the participants to the screen was not always the same due to leaning backward and forward movements; although participants were asked to limit movement, such movements were not explicitly controlled. The experiments took place in a large university laboratory (room dimensions of  $40.0 \text{ m} \times 20.0 \text{ m} \times 9.5 \text{ m}$ ). Although the room itself contained a variety of equipment including workstations, tables, chairs, etc., for the duration of the experiment effort was taken to limit the amount of external noise (e.g., equipment not used in the experiment was turned off) and aside from the participant and experimenter, no one else was present in the room. The average background noise level, also measured at the location where the participant's head would be (and measured in the absence of the sound stimulus) was 57 dBA (the maximum and minimum background noise level was 63 dBA and 55 dBA respectively). Participants were not informed of how many trials they had to run during the experiments but prior to the experiment, they were told that the experiment could take up to approximately 40 minutes.

With respect to the 3D stereoscopic experiments, the stereoscopic 3D "depth" setting on the Acer Aspire laptop was set to "Default" which, on a scale of 0-15, represents a setting of 3. Depth is the only modifiable stereoscopic-related parameter within the NVIDIA stereoscopic 3D system; other stereoscopic-related



parameters were not available to modify. It was observed that participants went through an "adaptation stage" during the first three to four trials whereby they moved their heads trying to find the best spot in which the 3D stereoscopic effect could be perceived. Participants were allowed to do this and we did not attempt to control for such movements.

### **3.1.3 Experimental Variables.**

#### **3.1.3.1 Controlled Variables**

- Visual Variables

With respect to visual conditions, visual quality was defined by polygon count and texture resolution (see Section 3.1.4), visual quality was also defined with respect to "blurriness" in Experiment 7 to approximate texture resolution variation. The visual environment within the virtual environment used in Experiment 7 was too complex and would therefore require far too much time to vary the texture resolution of all the models within the environment.

- Auditory Variables

With respect to the auditory cues, only the type of sound delivered to the participants was varied (see Section 3.1.5). The "no sound" condition (where the visuals were presented alone, without sound) formed the control condition.

#### **3.1.3.2 Non-controlled Variables**

Aside from the controlled variables described above, there were other variables whose effect was not explicitly tested.

With respect to visuals, such non-controlled variables include: the size or brightness of the screen (laptop monitor), the distance between the participants and the screen, and the different types of display technology (LCD, LED, etc.). Furthermore, visual quality was varied with respect to polygon count and texture resolution only despite the fact that visual quality can also be affected by lighting and shadow effects for example which were not addressed within the scope of this thesis.

With respect to the auditory conditions, there were many non-controlled variables that were not addressed within the scope of this thesis include sound level (volume) , loudspeaker effects (we decided to use headphones over loudspeakers), quantization levels, sampling frequency (both of which were kept constant at 16-bit and 44.1 kHz respectively). Furthermore, the sounds in all of our auditory conditions were complex; simple sounds (e.g., pure tone, sound from a single instrument) were not considered.

### **3.1.4 Visual Stimuli**

#### **Rendered model of a surgeon's head (the rendering was varied with respect to polygon count).**

This condition consisted of six static images of a single surgeon's head (see Figure 3). As described in Table 3, the quality of each image varied with respect to polygon count only; everything else remained the same (texture resolution, etc.) but the number of polygons comprising the model of the surgeon's head was varied. Images were presented on an Acer Aspire laptop with a 15.6" screen size and a resolution of  $1366 \times 768$ . The size of each image within the display was  $800 \times 630$ . When displayed, the image remained static (i.e., participants were not able to interact with the image in any manner).



**Figure 3. Visual stimuli. Six renderings of a surgeon's head, each varying with respect to polygon count.**

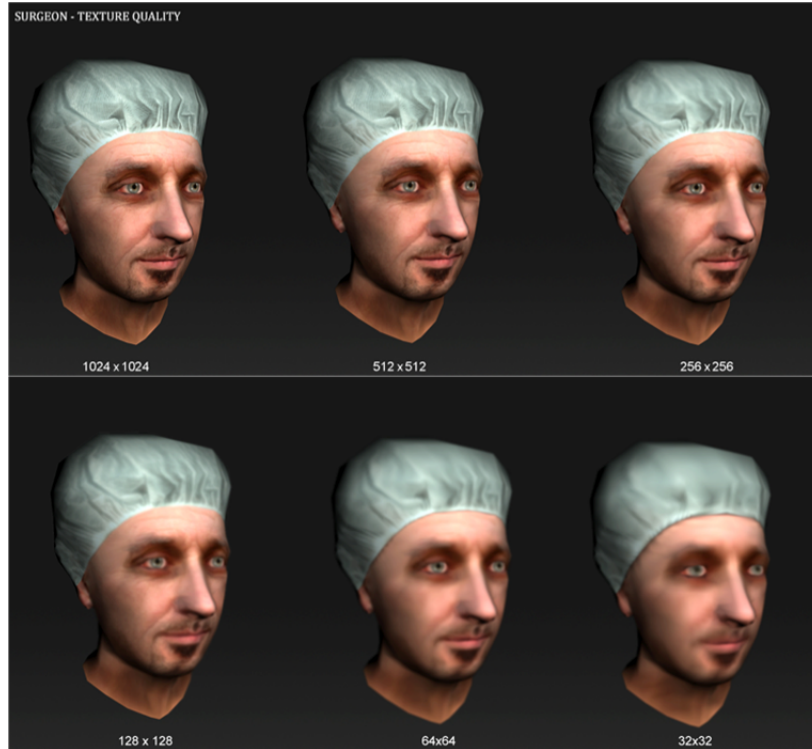
Image/Model	Polygon Count
1	17,440
2	13,440
3	1,250
4	865
5	678
6	548

**Table 3. Polygon count for each of the six surgeon head models considered.**

**Rendered model of a surgeon's head (the rendering was varied with texture resolution).**

This condition consisted of six images of a single surgeon's head (see Figure 4). The model of the surgeon's head was comprised of 17,440 polygons but for each of the six images, visual quality was varied with respect to texture resolution (see Table 4) only while all other parameters, including polygon count and image size, remained the same. Images were presented on an Acer Aspire laptop with a 15.6" screen size and a resolution of 1366 × 768. The size of each image within the

display was  $800 \times 630$ . When displayed, the image remained static (i.e., participants were not able to interact with the image in any manner).



**Figure 4. Surgeon's head textured with varying resolution texture.**

Image/Model	Texture Resolution
1	$1024 \times 1024$
2	$512 \times 512$
3	$256 \times 256$
4	$128 \times 128$
5	$64 \times 64$
6	$32 \times 32$

**Table 4. Texture Resolution for each model in each of the six images.**

**Rendered model of a surgeon's head, upper body, and surgical drill (the rendering was varied with respect texture resolution).**

This condition consisted of six images of a surgeon's upper body holding a surgical drill (see Figure 5). The 3D model of the surgeon's upper body was

comprised of 13,656 triangles, the drill consisted of 4,586 triangles, and the drill bit consisted of 616 triangles. For each of the six images, quality was varied with respect to texture resolution only (see Table 5) while all other parameters, including polygon count and image size, remained the same. Images were presented on an Acer Aspire laptop with a 15.6" screen size and a resolution of  $1366 \times 768$ . The size of each image within the display was  $800 \times 630$ .



**Figure 5. Surgeon's upper body holding an orthopaedic drill model.**

<b>Image/Model</b>	<b>Texture Resolution</b>
1	1024 x 1024
2	512 x 512
3	256 x 256
4	128 x 128
5	64 x 64
6	32 x 32

**Table 5. Texture resolution for each model in each of the six images.**

### 3.1.5 Auditory Stimuli

#### **Non-contextual auditory stimuli.**

The non-contextual auditory stimuli consisted of four ambient (background) sound conditions (non-related to the visual scene): i) no sound at all, ii) white noise, iii) classical music (Mozart), and iv) heavy metal music (similar to the work described in [Woods and Conrad, 2011]). The white noise sound was sampled at a rate of 44.1 kHz and band-pass filtered using a 256-point Hamming windowed FIR filter with low and high frequency cut-offs of 200 Hz and 10 kHz respectively. The level of each of the sounds was normalized (using the “normalize multiple audio tracks” option of the Audacity audio editor software) to ensure that all sounds had the same peak level. The sound pressure level of the output sounds was 66dB (about the same level as typical conversation [Morfey, 2005]) measured with a Radio Shack sound level meter (model 33-2055). All auditory stimuli were monophonic and were output with a pair of Sony MDR 110LP headphones.

#### **Contextual auditory cues.**

In addition to the four non-contextual cues described above, several of the experiments included contextual auditory cues (contextual with respect to the visuals). The three contextual auditory conditions were: v) surgical drill sound, vi) operating room ambiance mixed with a surgical drill sound, and vii) operating room ambiance without the drill sound. The operating room ambiance sound included machines beeping, doctors and nurses talking, and was purchased from AudioSparx.com and the operating room ambiance + drill sound was made by mixing the operating room ambiance sound with a recording of an actual drill sound (it was mixed using the Audacity audio editor software). The recording was made in an Eckel audiometric room to limit any external noise (air condition “hums”, etc.) and reverberation of the generated sounds within the environment, at a sampling rate of 44.1 kHz. As with the non-contextual auditory cues described above, the white noise sound was sampled at a rate of 44.1 kHz and band-pass filtered using a 256-point Hamming windowed FIR filter with low and high

frequency cut-offs of 200 Hz and 10 kHz respectively. The level of each of the sounds was normalized (using the “normalize multiple audio tracks” option of the Audacity audio editor software) to ensure that all sounds had the same peak level. The sound pressure level of the output sounds was 66dB (about the same level as typical conversation [Morfeey, 2005]) measured with a Radio Shack sound level meter (model 33-2055). As with the non-contextual auditory stimuli, all auditory stimuli were monophonic and were output with a pair of Sony MDR 110LP headphones.

### **3.1.6 Experimental Method**

#### **Non-contextual auditory cues experimental design (Experiments 1-4)**

Prior to the start of the experiment, participants were presented with a brief questionnaire and were asked to answer several questions regarding their video game play habits. The experiment began immediately following the completion of the questionnaire. For each trial, participants were presented with one of the six images shown on either Figure 3 or Figure 4 (depending on how visual quality was defined) in conjunction with one of the four non-contextual background sound conditions described in Section 3.1.1 and their task was to rank the image with respect to their perceived visual quality on a scale from 1 (lowest perceived quality) to 7 (highest perceived quality). Each of the six images and each of the four background sound combinations (24 combinations in total) was repeated three times (total of 72 trials).

#### **Contextual auditory cues experimental design (Experiments 5 and 6)**

As with the non-contextual auditory cues experiments described above, prior to the start of the experiment, participants were presented with a brief questionnaire regarding their video game play habits. The experiment began immediately following the completion of the questionnaire. In each trial, participants were

presented with one of the six images in conjunction with one of the seven contextual ambient sound conditions described in Section 3.1.2. Their task was to rank the image with respect to their perceived visual quality on a scale from 1 (lowest perceived quality) to 7 (highest perceived quality). Each of the six images and each of the seven background sound combinations (42 combinations in total) was repeated three times for a total of 126 trials. All participants completed the experiment within a single session.

## **3.2 Experiments**

### **Sensory Integration (Audio-Visual) with Non-Contextual Auditory Cues.**

#### **3.2.1 Experiment 1: Visual quality defined with respect to polygon-count with non-contextual ambient auditory conditions.**

##### *Participants*

Participants consisted of unpaid student volunteers from the University of Ontario Institute of Technology. A total of 18 (9 male and 9 female) volunteers participated in the experiment (average age was 21).

##### *Visual Stimuli*

The visual stimuli were given by the “rendered model of a surgeon’s head varied with polygon count” condition described in Section 3.1.1.

##### *Auditory Stimuli*

The auditory stimuli were given by the “non-contextual auditory stimuli” described in Section 3.1.2.



### *Experimental Method*

The experimental method used was the “non-contextual auditory cues experimental design” described in Section 3.1.3.

## **Results**

### *Questionnaire*

As previously described, the experiment began with a brief questionnaire that gauged participants’ video game habits. The first question asked whether they play video games or not and 14 of the 18 participants (78%) did report playing video games. For those that did report playing video games, the average number of hours per week spent playing video games was 2.4 hours, six of them reported they primarily played console-based video games, six of them primarily played PC-based video games, and two of them primarily played video games on their mobile phone. The questionnaire then asked participants to rate several video game attribute according to their importance on a 7-point scale (1 being least important and 7 being most important). A summary of the average ratings for each of the attributes is provided in Table 6 and as shown, graphics and sound received the highest scores (4.0 and 3.8 respectively) indicating participants believed they are the two most important attributes. Participant game preferences did not appear to impact experimental and therefore, will not be discussed further.

<b>Video Game Attribute</b>	<b>Avg. “Importance” ( /7.0)</b>
Graphics/visuals	4.0
Sound	3.8
Interaction	1.7
Story/narrative	3.1
Level of challenge/difficulty	2.9

**Table 6. Experiment 1. Summary of the video game attribute importance.**

### *Experiment*

A summary of the results is presented in Figure 6 where the x-axis represents polygon count and the y-axis represents perceived visual quality (ranging from 1-7) and on Table 7. The analysis of variance (ANOVA) was selected as the statistical model with two factors: 6 visual conditions (static images of a model of a surgeon's head, varying with respect to the number of polygons used to represent the model)  $\times$  4 auditory conditions: i) no sound at all, ii) white noise, iii) classical music, and iv) heavy metal music.

The results of the two way ANOVA (ambient sound by visual (image)) revealed significant main effects for sound ( $F=12.118$ ,  $p = .001$ ) and image ( $F=33.09$ ,  $p = .001$ ) in addition to the interaction between these two terms ( $F=14.374$ ,  $p = .005$ ). Since I was interested in the effect of ambient sound on the perception of visual quality, the two-way interaction was further analyzed by six subsequent one-way ANOVAs with sound as a factor, one for each of the visual conditions (polygon count).

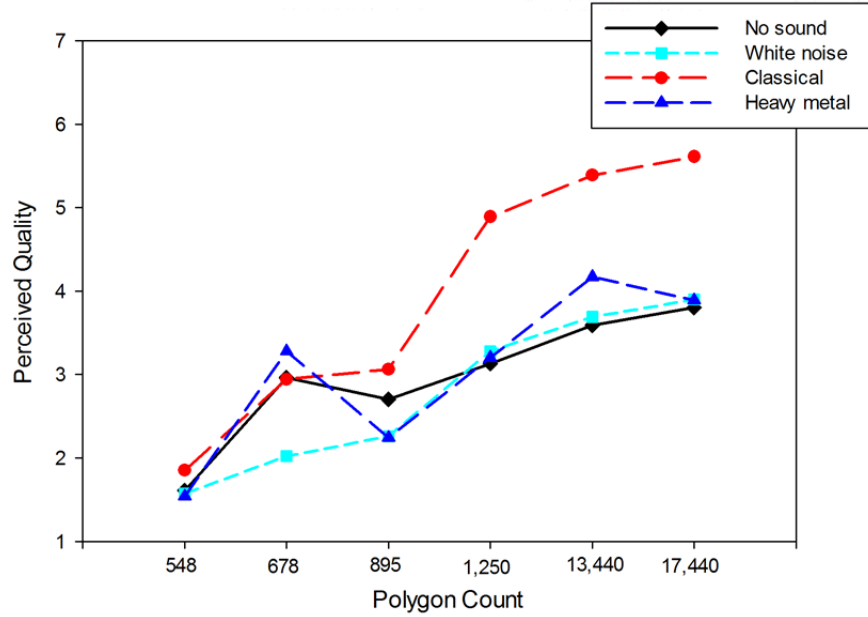
The analysis for the model with a polygon count of 17,440 revealed significant results ( $F(6)=34.503$ ,  $p=.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that classical music auditory condition influenced the perception of this visual; in contrast to the other three auditory conditions, the perception of visual quality increased in the presence of classical music auditory condition ( $p<.001$ ), which did not differ from each other ( $p<.05$  for all). The analysis for the model with a polygon count of 13,440 revealed significant results ( $F(6)=32.531$ ,  $p<.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the classical music auditory condition led to an increase in the perception of visual quality ( $p<.001$ ) when compared to other three auditory conditions (which did not differ from each other ( $p<.05$  for all)). The analysis for the model with a polygon count of 1,250 significant results ( $F(5)=8.752$ ,  $p<.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the classical music auditory condition resulted in an increase in the perception of visual image quality when compared to the other images

( $p < 0.001$ ) which did not differ from each other ( $p < .05$  for all). The analysis for the model with a polygon count of 865 showed significant results ( $F(5)=11.705$ ,  $p < .001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the classical music auditory condition influenced the perception of visual quality causing it to be higher in quality than the other sounds ( $p < .001$ ) (which did not differ from each other ( $p < .05$  for all)).

The analysis for the model with a polygon count of 678 revealed significant results ( $F(5)=34.503$ ,  $p = .001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that white noise auditory condition influenced the perception of this visual; in contrast to the other three auditory conditions, the perception of visual quality decreased in the presence of white noise ( $p < .001$ ), which did not differ from each other ( $p < .05$  for all).

The analysis for the model with a polygon count of 548 (the lowest polygon count) did not reveal any significance ( $F(5)=.372$ ,  $p = .073$ ) suggesting that sound did not have any influence on the perception of visual quality for the model with the lowest polygon count.

In summary, the heavy metal music and no sound ambient auditory conditions showed no influence on visual quality perception. Classical music led to an increase in visual quality perception. Finally, the white noise auditory condition had an attenuating effect on the perception of the visual quality. More specifically for the visuals with polygon counts between 17,440 and 885, the classical music auditory condition led to an increase in the perceived visual quality. Meanwhile, the white noise auditory condition had an attenuating effect on the perception of the quality for the visual with polygon counts of 885 and 548. The heavy metal music auditory condition was not statistically different from the no sound auditory condition.



**Figure 6. Experiment 1 results. Image quality (polygon count) vs. perceived quality (images presented individually).**

Image #	Sound	Mean	Std. Deviation
1	No Sound	3.759	0.439
	White Noise	3.833	0.574
	Classical Music	5.555	0.615
	Metal Music Sound	4.092	0.521
2	No Sound	3.574	0.339
	White Noise	3.685	0.620
	Classical Music	5.388	0.551
	Metal Music Sound	4.166	0.607
3	No Sound	3.074	0.436
	White Noise	3.370	0.469
	Classical Music	4.814	0.538
	Metal Music Sound	3.203	0.444
4	No Sound	2.629	0.441
	White Noise	2.296	0.426
	Classical Music	3.055	0.446
	Metal Music Sound	2.333	0.427
5	No Sound	2.944	0.707
	White Noise	1.925	0.405
	Classical Music	2.962	0.377
	Metal Music Sound	3.259	0.554
6	No Sound	1.388	0.171
	White Noise	1.611	0.328
	Classical Music	1.796	0.345
	Metal Music Sound	1.703	0.570

**Table 7. Experiment 1 results. Mean and Standard Deviation.**

### **3.2.2 Experiment 2: Visual quality defined with respect to texture resolution with non-contextual auditory conditions.**

#### *Participants*

Participants consisted of unpaid student volunteers from the University of Ontario Institute of Technology. A total of 18 (7 male and 11 female) volunteers participated in the experiment (average age was 20).

#### *Visual Stimuli*

The visual stimuli were given by the “rendered model of a surgeon’s head varied with respect to texture resolution” condition described in Section 3.1.1.

#### *Auditory Stimuli*

The auditory stimuli were given by the “non-contextual auditory stimuli” described in Section 3.1.2.

#### *Experimental Method*

The experimental method used was the “non-contextual auditory cues experimental design” described in Section 3.1.3.

### **Results**

#### *Questionnaire*

The experiment began with a brief questionnaire that gauged participants’ video game habits. The first question asked whether they play video games or not and 15 of the 18 participants (83%) did report playing video games. For those that did report playing video games, the average number of hours per week spent playing video games was 3.7 hours, eleven of them reported they primarily played console-based video games, three of them primarily played PC-based video games, and one of them primarily played video games on their mobile phone. The

questionnaire then asked participants to rate several video game attribute according to their importance on a 7-point scale (1 being least important and 7 being most important). A summary of the average ratings for each of the attributes is provided in Table 8 and as shown, graphics and sound received the highest scores (4.5 and 3.4 respectively) indicating participants believed they are the two most important attributes. Participant game preferences did not appear to impact experimental and therefore, will not be discussed further.

Video Game Attribute	Avg. "Importance" ( /7.0)
Graphics/visuals	4.5
Sound	3.4
Interaction	1.2
Story/narrative	2.7
Level of challenge/difficulty	3.2

**Table 8. Experiment 2. Summary of video game attribute importance.**

*Experiment*

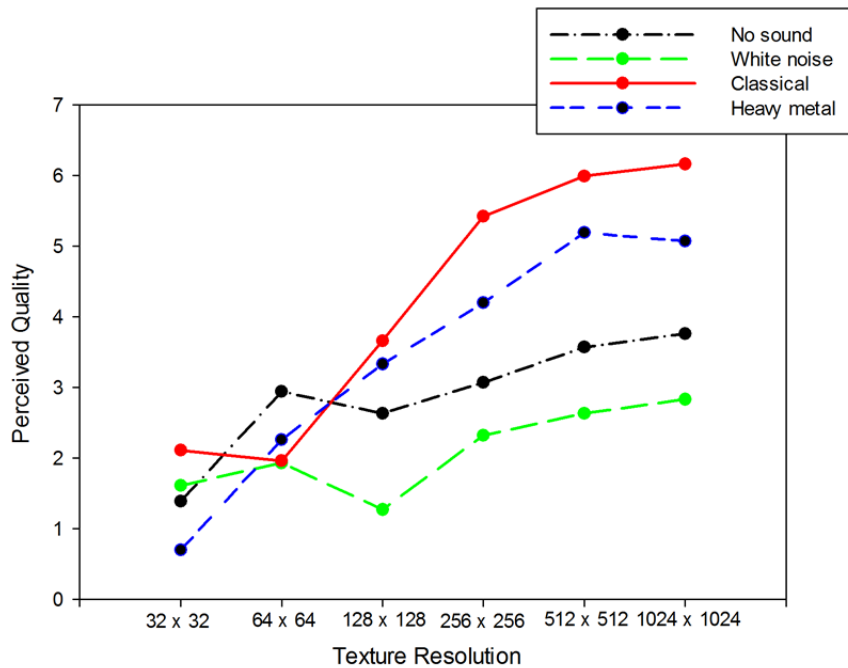
A summary of the results is presented in Figure 7 where the x-axis represents texture resolution and the y-axis represents perceived visual quality (ranging from 1-7) and on Table 9. The analysis of variance (ANOVA) was selected as the statistical model with two factors: 6 visual conditions (static images of a model of a surgeon's head, varying with respect to the number of polygons used to represent the model)  $\times$  4 auditory conditions: i) no sound at all, ii) white noise, iii) classical music, and iv) heavy metal music.

The results of the two way ANOVA (ambient sound by visual (image)) revealed significant main effects for sound ( $F=8.610$ ,  $p = .001$ ) and image ( $F=10.530$ ,  $p = .001$ ) in addition to the interaction between these two terms ( $F=4.694$ ,  $p = .003$ ). Since I was interested in the effect of ambient sound on the perception of visual quality, the two-way interaction was further analyzed by six subsequent one-way ANOVAs with sound as a factor, one for each of the visual conditions (polygon count).

The analysis for the model with a texture resolution of  $10124 \times 1024$  revealed significant results ( $F(5)=13.071$ ,  $p=.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed significant influence over the quality perception for all the conditions but no sound condition. The classical music and heavy metal music auditory conditions increased the perception of the visual quality; in contrast to the other three auditory conditions (which did not differ from each other ( $p<.05$  for all)). The white noise auditory condition led to a decrease in the perceived visual quality when compared to the other auditory conditions (which did not differ from each other ( $p<.05$  for all)). The analysis for the model with a polygon count of  $512 \times 512$  revealed significant results ( $F(5)=14.360$ ,  $p<.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the classical music and heavy metal music auditory conditions led to an increase in the perception of visual quality ( $p<.001$ ) when compared to other auditory conditions (which did not differ from each other ( $p<.05$  for all)). The white noise auditory condition led to a decrease in the perceived visual quality when compared to the other auditory conditions (which did not differ from each other ( $p<.05$  for all)). The analysis for the model with a polygon count of  $256 \times 256$  indicated significant results ( $F(5)=14.306$ ,  $p<.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the classical music and heavy metal music auditory conditions resulted in an increase in the perception of visual quality perception when compared to the other images ( $p<0.001$ ) (which did not differ from each other ( $p<.05$  for all)), while the white noise auditory condition led to a decrease in the perceived visual quality ( $p<.001$ ). The analysis for the model with a polygon count of  $128 \times 128$  revealed significant results ( $F(5)=11.705$ ,  $p<.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that white noise auditory condition influenced the perception of this visual; in contrast to the other three auditory conditions, the perception of visual quality decreased in the presence of white noise auditory conditions ( $p<.001$ ) (which did not differ from each other ( $p<.05$  for all)).

The analysis for the model with a polygon count of  $64 \times 64$  did not reveal any significant results ( $F(5)=34.503$ ,  $p=.001$ ), suggesting that sound did not have any influence on the perception of visual quality for this specific model. The analysis for the model with a polygon count of  $32 \times 32$  revealed significant results ( $F(5)=3.757$ ,  $p=.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the classical music and heavy metal music auditory conditions increased the perception of the visual quality; in contrast to the other three auditory conditions (which did not differ from each other ( $p<.05$  for all)).

In summary, the heavy metal music and classical music auditory conditions generally led to an increase in the perception of visual quality perception. More specifically, the influence is easily appreciated under the higher texture resolution conditions ( $1024 \times 1024$ ,  $512 \times 512$ ,  $256 \times 256$ , and  $128 \times 128$ ) in contrast to the effect shown on the lower resolution conditions. Finally, the white noise auditory condition had an attenuating effect on the perception of the visual quality across all texture resolutions considered.



**Figure 7. Experiment 2 results. Texture resolution (x-axis) vs. perceived visual quality (y-axis) across each of the four non-contextual auditory conditions considered.**



Image #	Sound	Mean	Std. Deviation
1	No Sound	4.44	0.316
	White Noise	2.67	0.320
	Classical Music	6.57	0.242
	Metal Music Sound	6.00	0.320
2	No Sound	5.00	0.320
	White Noise	2.67	0.320
	Classical Music	6.43	0.442
	Metal Music Sound	5.67	0.320
3	No Sound	5.33	0.320
	White Noise	1.71	0.288
	Classical Music	5.90	0.356
	Metal Music Sound	5.62	0.288
4	No Sound	2.67	0.320
	White Noise	1.71	0.288
	Classical Music	4.29	0.288
	Metal Music Sound	4.62	0.288
5	No Sound	2.67	0.320
	White Noise	1.76	0.242
	Classical Music	4.00	0.320
	Metal Music Sound	3.90	0.305
6	No Sound	2.05	0.288
	White Noise	1.19	0.171
	Classical Music	4.33	0.320
	Metal Music Sound	3.67	0.320

**Table 9. Experiment 2 results. Mean and Standard Deviation.**

**Sensory Integration (Audio-Visual) with Non-Contextual Auditory Cues Discussion (Polygon Count and Texture Resolution Variation vs. Non-Contextual Auditory Cues)**

Here the results of the first two experiments that examined perception of visual quality (with respect to polygon count and texture resolution) under four non-contextual related ambient (background) auditory conditions (with respect to the visual scene): i) no sound at all, ii) white noise, iii) classical music (Mozart), and iv) heavy metal music, were presented. Results indicate that perception of visual quality is influenced by background sound.

The findings presented here, and more specifically, that ambient (background) sound can effect perception is supported by previous studies. Woods et al. [2008], investigated the effect of background sound (noise) on the perception of food perception and more

specifically, on gustatory food properties (i.e., sugar level, salt level), food crunchiness and food liking. They found background noise has three effects on food perception: i) food saltiness and sweetness was reduced in the presence of loud background noise, ii) food was “crunchier” in the presence of background noise, and iii) there was a correlation between the liking of the food and the liking of the noise. Conrad et al. [2010], examined the effect of background sound (including classical music (Mozart), and dichotic (metal) music) on laparoscopic surgery. They found that dichotic music had a negative impact on time until task completion but did not impact task accuracy. They also found that classical music had a variable effect on time until task completion but resulted in greater task accuracy amongst all participants (laparoscopic surgeons).

Based on the results presented here, it can be concluded that background ambient music does affect visual quality perception. More specifically, when considering the higher quality visuals (defined both with respect to polygon count and texture resolution), the perception of visual quality increased in the presence of ambient sound consisting of either both classical music and heavy metal music auditory conditions. In addition, ambient sound consisting of white noise led to a decrease in visual quality perception across all visuals conditions defined with respect to polygon count or texture resolution. However, as shown on the results from the Experiment 1 only classical music condition showed to increase the visual quality perception, and on Experiment 2 both classical and heavy metal music showed to increase visual quality perception.

The results of the two studies presented so far are preliminary; although the results have shown that ambient sound can influence our perception of visual quality, there are many questions that must be considered before a greater understanding regarding the effect of audio-visual sensory integration and visual quality perception is developed. In addition to polygon count and texture resolution as considered in these two experiments, visual quality can be defined in many ways other ways including blurring of the scene. Visuals can also be presented in stereoscopic 3D and in fact, the following set of experiments replicate these two experiments within a stereoscopic 3D viewing environment.

## **Sensory Integration (Audio-Visual) Given by the Type of Visual Presentation (Stereoscopic 3D vs. Non-Stereoscopic 3D)**

In the experiments described in this section, the effect of stereoscopic 3D visuals on visual quality perception when presented under specific ambient auditory cues is examined. This is accomplished by replicating Experiments 1 and 2 previously described within a stereoscopic 3D viewing environment.

### **3.2.3 Experiment 3: Visual quality defined with respect to polygon count in a stereoscopic 3D viewing environment with non-contextual ambient auditory conditions**

#### *Participants*

Participants consisted of unpaid student volunteers from the University of Ontario Institute of Technology. A total of 18 (1 male and 17 female) volunteers participated in the experiment (average age was 21).

#### *Visual Stimuli*

The visual stimuli were identical to the visual stimuli of Experiment 1 (rendering of the surgeon's head with varying polygon count (Section 3.1.1) but here, the visuals were presented in stereoscopic 3D. The Acer Aspire laptop used in this experiment (see Section 3.1.1) employs NVIDIA's 3D Vision technology (active stereoscopic system with wireless glasses). The stereoscopic 3D "Depth" setting was set to "Default" which, on a scale of 0-15 (where 0 represents no stereoscopic 3D effect and 15 represents the maximum), the "Default" setting is set to 3. Depth is the only modifiable stereoscopic-related parameter within the NVIDIA stereoscopic 3D system; other stereoscopic-related parameters are not available. Each 3D visual was presented with its corresponding background sound until the participant made a choice. When displayed, participants were able to rotate the 3D model.

### *Auditory Stimuli*

The auditory stimuli were given by the “non-contextual auditory stimuli” described in Section 3.1.2.

### *Experimental Method*

The experimental method used was the “non-contextual auditory cues experimental design” described in Section 3.1.3.

## **Results**

### *Questionnaire*

The first question asked whether or not participants regularly played video games (in order to determine their regular use of computer-animated figures): 9 of the 18 participants (50%) played video games regularly. Of these participants, the average number of hours per week spent playing video games was 2.7 hours (minimum one hour and maximum 10 hours), six of them reported they primarily played console-based video games, two of them primarily played PC-based video games, and one of them primarily played video games on their mobile phone. When asked to indicate their video game genre preference (from a list of given genres), three of the participants prefer “Strategy” games, five “Sports” games and one “Shooter” games. The questionnaire then asked participants to rate several video game attributes according to their importance on a 7-point scale (1 being least important and 7 being most important). A summary of the average ratings for each of the attributes is provided in Table 10. Participant game and musical genre preferences did not appear to impact experimental results and therefore, will not be discussed further.

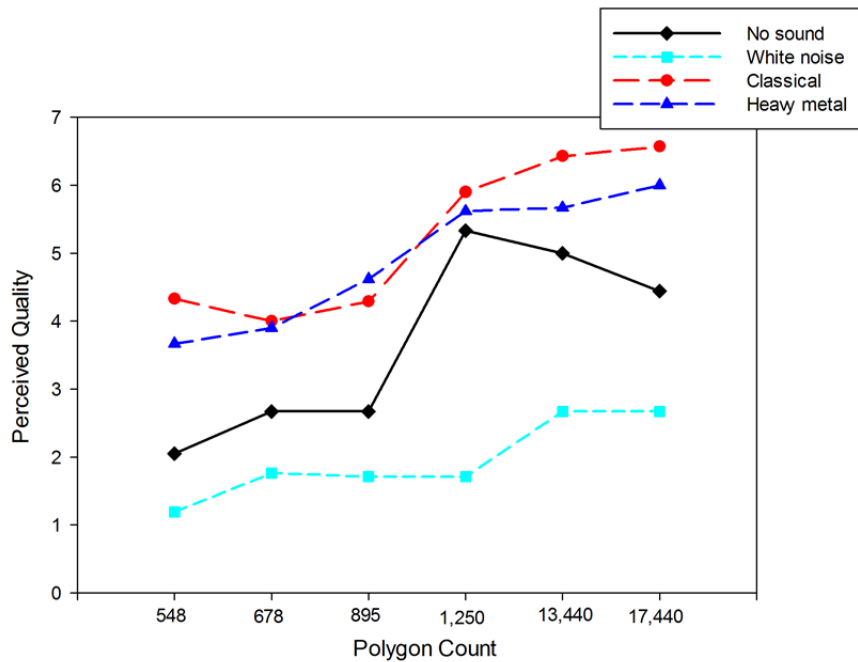
<b>Video Game Attribute</b>	<b>Avg. “Importance” (/7.0)</b>
Graphics/visuals	4.5
Sound	3.3
Interaction	1.4
Story/narrative	2.3
Level of challenge/difficulty	3.5

**Table 10. Experiment 3. Summary of video game attributes.**

### Experiment

A summary of the results is presented in Figure 8 where the x-axis represents polygon count and the y-axis represents perceived visual quality (ranging from 1-7) and on Table 11. The analysis of variance (ANOVA) was selected as the statistical model with two factors: 6 visual conditions (static images of a model of a surgeon's head, varying with respect to the number of polygons used to represent the model)  $\times$  4 auditory conditions: i) no sound at all, ii) white noise, iii) classical music, and iv) heavy metal music.

The results of the two way ANOVA (ambient sound by visual (image)) revealed significant main effects for sound ( $F(2)=23.2$ ,  $p = .001$ ) and image ( $F(2)=35.84$ ,  $p = .001$ ) in addition to the interaction between these two terms ( $F(2)=2.13$ ,  $p = .008$ ). Since the effect of ambient sound on the perception of visual quality was of interest, the two-way interaction was further analyzed by six subsequent one-way ANOVAs with sound as a factor, one for each of the visual conditions (polygon count).



**Figure 8. Experiment 3. Polygon count (x-axis) vs. perceived visual quality (y-axis) across each if the four non-contextual auditory conditions considered.**

Image #	Sound	Mean	Std. Deviation
1	No Sound	5.055	0.826
	White Noise	3.740	0.371
	Classical Music	5.648	0.820
	Metal Music Sound	4.888	0.921
2	No Sound	4.574	1.571
	White Noise	3.611	0.383
	Classical Music	5.388	0.818
	Metal Music Sound	4.851	0.998
3	No Sound	4.000	1.357
	White Noise	3.592	0.554
	Classical Music	5.388	0.785
	Metal Music Sound	3.870	1.588
4	No Sound	3.981	1.345
	White Noise	3.518	0.460
	Classical Music	5.277	0.937
	Metal Music Sound	4.055	1.496
5	No Sound	3.666	1.409
	White Noise	3.462	0.486
	Classical Music	3.907	1.261
	Metal Music Sound	4.370	1.537
6	No Sound	2.444	1.778
	White Noise	2.500	0.383
	Classical Music	2.777	1.850
	Metal Music Sound	2.222	1.832

**Table 11. Experiment 3 results. Mean and Standard Deviation.**

The analysis for the model with a polygon count of 17,440 revealed significant results ( $F(2)=19.562$ ,  $p=.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the white noise auditory condition influenced the perception of this visual; in contrast to the other three auditory conditions, the perception of visual quality decreased in the presence of white noise auditory condition ( $p<.001$ ), which did not differ from each other ( $p<.05$  for all). The analysis for the model with a polygon count of 13,440 revealed significant results ( $F(5)=9.3$ ,  $p<.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the white noise auditory condition led to a decrease in the perception of visual quality compared to other three auditory conditions ( $p<.001$ ), all three of which did not differ from

each other ( $p < .05$  for all). The analysis for the model with a polygon count of 1,250 revealed significant results ( $F(5)=8.752$ ,  $p < .001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the classical music auditory condition resulted in an increase in the perception of visual image quality when compared to the other images ( $p < 0.001$ ) which did not differ from each other ( $p < .05$  for all). The analysis for the model with a polygon count of 865 revealed significant results ( $F(5)=7.912$ ,  $p < .001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the classical music auditory condition influenced the perception of visual quality causing it to be higher in quality than the other sounds ( $p < .001$ ), which did not differ from each other ( $p < .05$  for all).

The analysis for the models with a polygon count of 678 and 548 (the two lowest polygon counts) did not reveal any significance ( $F(5)=1.77$ ,  $p=0.16$  and  $F(5)=.372$ ,  $p=.773$  respectively) indicating that ambient sound did not have any influence on the perception of visual quality for these two conditions (i.e., the two models with the lowest polygon counts).

In summary, classical music led to an increase in visual quality perception. The heavy metal music auditory condition showed to increase the visual quality perception when compared against the no sound and the white noise ambient auditory conditions, however the effect is not as strong as classical music auditory condition. Finally, the white noise ambient auditory condition had an attenuating effect on the perception of the visual quality although this effect was evident only for the models whose polygon count was greater than 678. Ambient sound does influence higher quality visuals when presented in stereoscopic 3D.

### **3.2.4 Experiment 4 : Visual quality defined with respect to texture resolution in a stereoscopic 3D viewing environment with non-contextual auditory conditions.**

#### *Participants*

Participants consisted of unpaid student volunteer from the University of Ontario Institute of Technology (UOIT). A total of 18 volunteers (5 male and 13 female) participated in the experiment (average age was 20 years old).

#### *Visual Stimuli*

The visual stimuli were identical to the visual stimuli of Experiment 2 (rendering of the surgeon's head with varying texture resolution (Section 3.1.1)) but here, as in Experiment 3, the visuals were presented in stereoscopic 3D. As with Experiment 3, the stereoscopic 3D "Depth" setting on the Acer Asipre laptop was set to "Default" which, on a scale of 0-15 represents a setting of 3. Depth is the only modifiable stereoscopic-related parameter within the NVIDIA stereoscopic 3D system; other stereoscopic-related parameters are not available. Each 3D model was presented with its corresponding background sound until the participant made a choice. When displayed, participants were able to rotate the image.

#### *Auditory Stimuli*

The auditory stimuli were given by the "non-contextual auditory stimuli" described in Section 3.1.2.

#### *Experimental Method*

The experimental method used was the "non-contextual auditory cues experimental design" described in Section 3.1.3.



## Results

### *Questionnaire*

The first question asked whether or not participants regularly played video games (in order to determine their regular use of computer-animated figures): 16 of the 18 participants (89%) played video games regularly. Of these participants, the average number of hours per week spent playing video games was 3.8 hours (minimum one hour and maximum 10 hours), six of them reported they primarily played console- based video games, seven of them primarily played PC-based video games, and three of them primarily played video games on their mobile phone. When asked to indicate their video game genre preference (from a list of given genres), six of the participants prefer “Strategy” games, five “Sports” games, four “Shooter” games, and one responded with “Other”. Participant game and genre preferences did not appear to impact experimental results and therefore, will not be discussed further. The questionnaire then asked participants to rate several video game attributes according to their importance on a 7-point scale (1 being least important and 7 being most important). A summary of the average ratings for each of the attributes is provided in Table 12 and as shown participants believed that graphics and sound are the two most important attributes. Participant game and musical genre preferences did not appear to impact experimental results and therefore, will not be discussed further.

<b>Video Game Attribute</b>	<b>Avg. “Importance” ( /7.0)</b>
Graphics/visuals	4.7
Sound	4.3
Interaction	2.9
Story/narrative	2.9
Level of challenge/difficulty	2.3

**Table 12. Experiment 4. Summary of video game attribute importance.**

### *Experiment*

A summary of the results is presented in Figure 9 where the x-axis represents texture resolution and the y-axis represents perceived visual quality (ranging from 1-7) and on Table 13.

The results of the two way ANOVA (ambient sound by visual (image)) revealed significant main effects for sound ( $F=21.2$ ,  $p < .001$ ) and image ( $F=9.6$ ,  $p < .001$ ) in addition to the interaction between these two terms ( $F=5.18$ ,  $p < .008$ ). Since the effect of the ambient sound on the perception of visual quality is of interest here, the two-way interaction was further analyzed by six subsequent one-way ANOVAs with ambient sound as a factor, one for each of the six visuals (images). These analyses showed that for model with the highest texture resolution ( $1028 \times 1028$ ) the ANOVA yielded significant results ( $F(5)=4.823$ ,  $p=.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that all ambient sound conditions influenced the perception of visual quality. The classical music ambient auditory condition ( $p<.001$ ) led to an increase in the perceived visual quality when compared to the other conditions ( $p<.05$  for all). The heavy metal music auditory condition led to an increase in the perceived visual quality when compared to the white noise and no sound auditory conditions ( $p<.05$  for both), but lead to a decrease in the perceived visual quality when compared to the classical music condition ( $p<.001$ ), implying that the heavy metal music condition, despite it increases the visual quality perception is not as effective as the classical music auditory condition. The white noise auditory condition led to a decrease in the perceived visual quality ( $p<.001$ ) when compared to the other conditions ( $p<.05$  for all).

The analysis for the model with a texture resolution of  $512 \times 512$  showed significant results ( $F(5)=29.3$ ,  $p<.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that all four ambient auditory conditions influenced the perception of visual quality compared to the other sounds ( $p<.05$  for all). The classical music and heavy metal music auditory conditions led to an increase in perceived visual quality ( $p<.001$ ) when compared to the other auditory conditions ( $p<.05$  for all). In contrast, the white noise auditory condition led to a decrease in the perceived visual quality ( $p<.001$ ) when compared to the other conditions ( $p<.05$  for all).

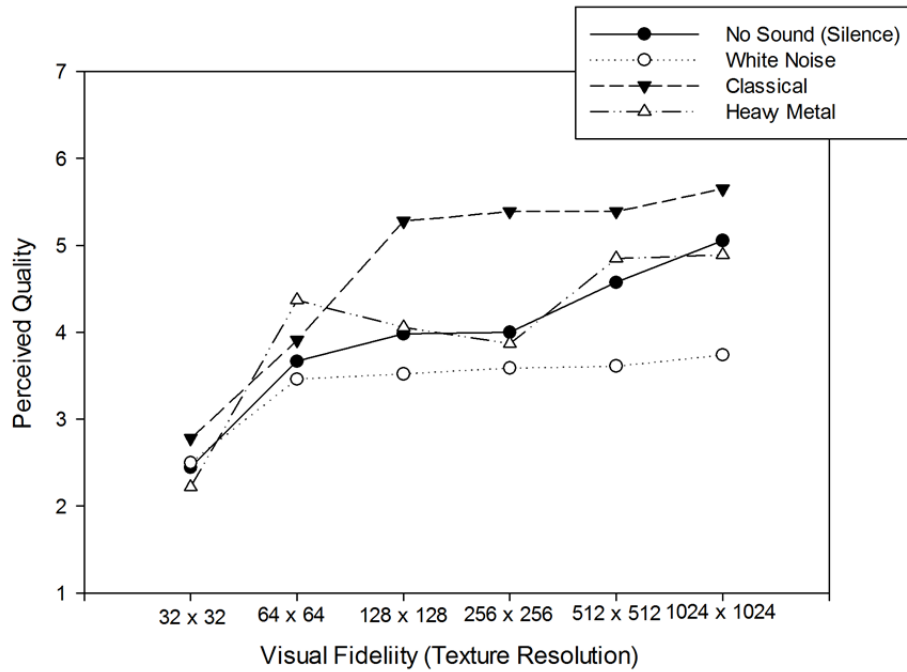
The analysis for the model with a texture resolution of  $256 \times 256$  showed significant results ( $F(5)=5.43$ ,  $p<.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the white noise auditory condition led to a decrease in visual quality perception compared to the other sounds ( $p<.001$ ), which did not differ from each other ( $p<.05$  for all).

The analysis for the model with a texture resolution of  $128 \times 128$  showed significant results ( $F(5)=16.52$ ,  $p<.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that all four auditory conditions influenced the perception of visual quality ( $p<.001$ ), which did not differ from each other ( $p<.05$  for all). The classical music auditory condition led to an increase in the perceived visual quality ( $p<.001$ ). The heavy metal music auditory condition also led to an increase in the perceived visual quality ( $p<.022$ ) when compared to the other auditory conditions while the white noise auditory condition led to a decrease in the perceived visual quality ( $p<.001$ ).

The analysis for the model with a texture resolution of  $64 \times 64$  showed significant results ( $F(5)=17.983$ ,  $p<.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the white noise auditory condition led to a decrease in the perception of visual quality compared to the other auditory conditions ( $p<.001$ ), which did not differ from each other ( $p<.05$  for all).

The analysis for the model with a texture resolution of  $32 \times 32$  showed significant results ( $F(5)=29.3$ ,  $p<.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that all four auditory conditions influenced the perception of visual quality compared to the other auditory conditions ( $p<.05$  for all). The classical music and heavy metal music auditory conditions led to an increase in perceived visual quality ( $p<.001$ ) and ( $p<.022$ ) respectively. The white noise auditory condition led to a decrease in the perceived visual quality ( $p<.001$ ).

In summary, there was a significant difference between the classical music auditory condition (resulted in the greatest perceived quality) and the other three auditory conditions. There was also a significant difference between the white noise auditory condition and the other three auditory conditions (white noise led to a decrease in visual quality perception). The heavy metal music and no sound ambient auditory conditions had a similar influence on visual quality perception (i.e., no effect).



**Figure 9. Experiment 4 results. Texture resolution (x-axis) vs. perceived visual quality (y-axis) across each of the four non-contextual conditions considered.**

Image #	Sound	Mean	Std. Deviation
1	No Sound	3.759	0.439
	White Noise	3.833	0.574
	Classical Music	5.555	0.615
	Metal Music Sound	4.092	0.521
2	No Sound	3.574	0.339
	White Noise	3.629	0.635
	Classical Music	5.388	0.551
	Metal Music Sound	4.185	0.596
3	No Sound	3.074	0.436
	White Noise	3.314	0.491
	Classical Music	4.814	0.538
	Metal Music Sound	3.203	0.444
4	No Sound	2.629	0.441
	White Noise	2.296	0.426
	Classical Music	3.055	0.446
	Metal Music Sound	2.333	0.427
5	No Sound	2.944	0.707
	White Noise	1.925	0.405
	Classical Music	2.962	0.377
	Metal Music Sound	3.259	0.554
6	No Sound	1.388	0.171
	White Noise	1.611	0.328
	Classical Music	1.833	0.431
	Metal Music Sound	1.703	0.570

**Table 13. Experiment 4 results. Mean and Standard Deviation.**

### **Audio-visual Sensory Integration Given by the Types of Visual Presentation Discussion (Stereoscopic 3D vs. Non-Stereoscopic 3D)**

The use of stereoscopic 3D within a gaming (and serious gaming) environment has been limited but it has been used extensively in virtual reality-based training simulations, most notably pilot training, where the depth perception it provides allows pilot trainees to assess complex 3D structures, and other aircraft positions while providing the ability to enhance group experience and group training while motivating the trainee to more deeply explore the learning material [Schiefele et al., 1999]. In addition to pilot training, stereoscopic 3D has been applied to virtual environments developed for surgical training [Tendick et al., 2000; Henn et al., 2002; Reitingner et al., 2006], and plant operator training [Fisk and Hill, 2010], amongst others. While S3D has the potential for increasing

engagement by providing the users with additional visual information (depth into and out of the screen plane), there are many issues that must be properly addressed when designing S3D serious games.

Within a non-stereoscopic viewing environment, the results of Experiments 1 and 2 suggest that our perception of visual quality is influenced by ambient auditory stimuli. However, in these two experiments the influence that stereoscopic 3D has on our perception of visual quality in the presence of auditory stimuli yet was not examined.

Here the results of two experiments that examined the perception of visual quality within a stereoscopic 3D viewing environment under various ambient (background) auditory conditions. In Experiments 3 and 4, visual quality was defined with respect to polygon count and texture resolution (of a model of a surgeon's head) respectively and the four non-contextual ambient auditory conditions (with respect to the visual scene) were: i) no sound at all, ii) white noise, iii) classical music, and iv) heavy metal music (see Section 3.1.2).

In Experiment 3, the classical music auditory condition led to an increase in visual quality perception, heavy metal music auditory condition showed to increase the visual quality perception (less than the classical music auditory condition) when compared to the other auditory conditions, while the white noise auditory condition had an attenuating effect on the perception of the visual quality perception although both of these effects were evident only for the visual models with polygon counts greater than 678 (i.e., the classical music and heavy metal music ambient auditory conditions had no effect on the two smallest polygon count models). In Experiment 4 the classical music auditory conditions led to an increase in visual quality perception. In contrast, the white noise auditory condition led to a decrease in visual quality perception. The heavy metal sound and no sound auditory conditions showed no effect on the visual quality perception.

As the results for both Experiments 3 and 4 indicate, the difference between the visual presentations in 3D stereoscopic vs. non-stereoscopic 3D is present only for the higher quality visual models (polygon count greater than 1,250, and texture resolution greater

than  $256 \times 256$ ). Results indicate that stereoscopic 3D may result in a more sensitive perception of visual quality for higher quality visuals. More specifically, ambient sound can influence visual quality when presented in conjunction with the high quality visuals, but it showed to not have the same effect when presented to the low quality visual conditions; this can be explained by the extra information that the participant is presented in stereoscopic 3D (i.e., depth), and therefore lack of quality might not be influenced on the same way than before.

The results of Experiments 3 and 4 together with the results of Experiments 1 and 2 suggest that ambient sound can affect various aspects of a virtual simulation/serious game. More specifically, distracting sounds such as white noise can lead to a decrease in the perception of visual quality; classical music (and heavy metal music in some cases) can cause our perception of visual quality to increase. Although previous work has examined the interaction of auditory cues and the other senses with respect to visual cues, in contrast to Experiments 1 and 2 in this thesis, the experiments presented here specifically examined the interaction of ambient auditory cues and visuals within a stereoscopic 3D viewing environment. The significance of these experiments lies in the simple fact that sound is shown to influence the perception of visuals with respect to stereoscopic 3D. This impact of sound on visuals will have implications for any studies into stereoscopic 3D in relation to audio-visual media, and particularly in studies that explore the role of stereoscopic 3D in serious games, training and simulations. Any future studies into quality or stereoscopic 3D must take into consideration the importance of multi-modal interactions.

In addition to polygon count and texture resolution, visual quality can be defined in many other ways including blurring of the visual scene. Visuals can also be presented in either stereoscopic or non-stereoscopic 3D as considered in Experiments 3 and 4 described here. Furthermore, with respect to the visuals, the ambient sounds considered here lacked any context (with respect to the visuals). What effect, if any, will contextual auditory cues have on our perception of visual quality? Moreover, it is unclear if specific attributes of the music (the classical music and heavy metal music auditory conditions) were

responsible for the results (such as the presence of lyrics, the frequency band, rhythm, etc.). The next set of experiments considered contextual auditory cues (with respect to the visual scene).

## **Sensory Integration (Audio-Visual) with Contextual Auditory Cues Discussion**

In the last four experiments, the four ambient auditory conditions were non-contextual with respect to the visual scene. Here, the influence of contextual ambient auditory cues on the perception of visual quality is examined. To be consistent to the previous experiments described in this thesis, contextual auditory cues were considered both in a stereoscopic 3D and non-stereoscopic 3D viewing environment.

### **3.2.5 Experiment 5: Visual quality defined with respect to texture resolution with contextual auditory conditions**

#### *Participants*

Participants consisted of unpaid student volunteers from the University of Ontario Institute of Technology. A total of 18 (3 male and 15 female) volunteers participated in the experiment (average age was 22).

#### *Visual Stimuli*

The visual stimuli were given by the “rendered model of a surgeon’s head varied with respect to texture resolution” condition described in Section 3.1.1.

#### *Auditory Stimuli*

The auditory stimuli were given by the “contextual auditory stimuli” described in Section 3.1.2.



### *Experimental Method*

The experimental method used was the “contextual auditory cues experimental design” described in Section 3.1.3.

## **Results**

### *Questionnaire*

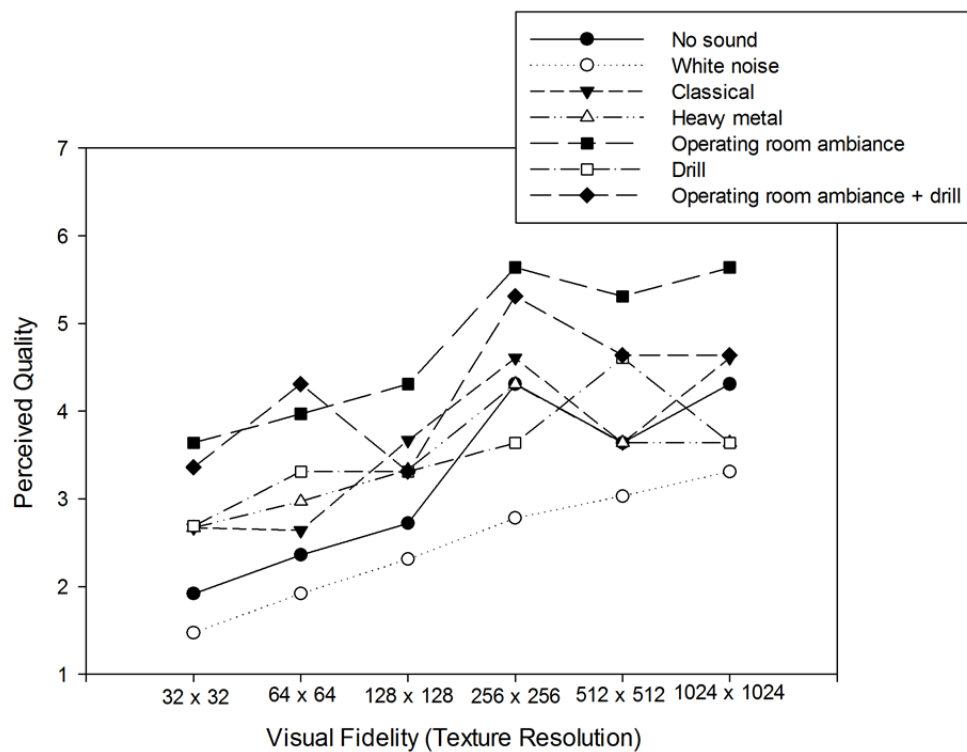
The first question asked whether or not participants regularly played video games (in order to determine their regular use of computer-animated figures): 12 of the 18 participants (66%) played video games regularly. Of these participants, the average number of hours per week spent playing video games was 3.7 hours (minimum one hour and maximum 10 hours), nine of them reported they primarily played console- based video games, two of them primarily played PC-based video games, and one of them primarily played video games on their mobile phone. When asked to indicate their video game genre preference (from a list of given genres), two of the participants prefer “Strategy” games, seven “Sports” games and three “Shooter” games. The questionnaire then asked participants to rate several video game attributes according to their importance on a 7-point scale (1 being least important and 7 being most important). A summary of the average ratings for each of the attributes is provided in Table 14. Participant game preferences did not appear to impact experimental results and therefore, will not be discussed further.

<b>Video Game Attribute</b>	<b>Avg. “Importance” (/7.0)</b>
Graphics/visuals	5.2
Sound	3.5
Interaction	1.8
Story/narrative	2.8
Level of challenge/difficulty	4.3

**Table 14. Experiment 5. Summary of video game attribute importance.**

*Experiment*

A summary of the results is presented in Figure 2 where the x-axis represents visual quality (texture resolution) and the y-axis represents perceived visual quality (ranging from 1-7) and on Table 15. The analysis of variance (ANOVA) was selected as the statistical model with two factors: six visual conditions (static images of a model of a surgeon's upper-torso, varying with respect to texture resolution)  $\times$  seven auditory conditions (i) no sound at all, ii) white noise, iii) classical music, iv) heavy metal music, v) operating room ambiance sound, vi) surgical drill sound, and vii) operating room ambiance mixed with the surgical drill sound).



**Figure 10. Experiment 5 results. Texture resolution (x-axis) vs. perceived visual quality (y-axis) for each of the seven auditory conditions (three contextual and four non-contextual).**

Image #	Sound	Mean	Std. Deviation
1	No sound	4.305	0.481
	White Sound	3.305	0.481
	Classical Music sound	4.611	0.445
	Metal Music Sound	3.638	0.481
	Surgical room ambiance	4.638	0.481
	Drill Sound	3.638	0.481
	Surgical Room + Drill Sound	5.638	0.481
2	No sound	3.638	0.481
	White Sound	3.027	0.481
	Classical Music sound	3.638	0.481
	Metal Music Sound	3.638	0.481
	Surgical room ambiance	4.638	0.481
	Drill Sound	4.611	0.445
	Surgical Room + Drill Sound	5.305	0.481
3	No sound	4.305	0.481
	White Sound	2.777	0.518
	Classical Music sound	4.611	0.509
	Metal Music Sound	4.305	0.481
	Surgical room ambiance	5.305	0.481
	Drill Sound	3.638	0.481
	Surgical Room + Drill Sound	5.638	0.481
4	No sound	2.722	0.467
	White Sound	2.305	0.481
	Classical Music sound	3.666	0.512
	Metal Music Sound	3.333	0.471
	Surgical room ambiance	3.305	0.481
	Drill Sound	3.305	0.481
	Surgical Room + Drill Sound	4.305	0.481
5	No sound	2.361	0.437
	White Sound	1.916	0.429
	Classical Music sound	2.638	0.437
	Metal Music Sound	2.972	0.481
	Surgical room ambiance	4.305	0.481
	Drill Sound	3.305	0.481
	Surgical Room + Drill Sound	3.972	0.481
6	No sound	1.916	0.351
	White Sound	1.472	0.521
	Classical Music sound	2.666	0.449
	Metal Music Sound	2.666	0.449
	Surgical room ambiance	3.361	0.437
	Drill Sound	2.694	0.437
	Surgical Room + Drill Sound	3.638	0.481

**Table 15. Experiment 5 results. Mean and Standard Deviation.**

The results of a two-way ANOVA (ambient/background sound by visual quality (texture resolution)) revealed significant main effects for sound ( $F=19.02$ ,  $p = .001$ ) and visual quality ( $F=20.95$ ,  $p = .001$ ) in addition to the interaction between them ( $F=8.14$ ,  $p = .001$ ). Since I was interested in the effect of ambient sound on the perception of visual quality, the two-way interaction was further analyzed by six subsequent one-way ANOVAs with ambient sound as a factor, one for each of the visual quality (texture resolution) conditions.

The results of a one-way ANOVA for the model corresponding to a texture resolution of  $1028 \times 1028$  yielded significant results ( $F(5)=33.974$ ,  $p=.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the operating room ambiance auditory condition caused the perception of visual quality for the image with a texture resolution of  $1028 \times 1028$  to increase ( $p<.001$ ). This increase was not observed for the other six auditory conditions (which did not differ from each other ( $p<.05$  for all)).

The analysis for the model with a texture resolution of  $512 \times 512$  showed significant results ( $F(5)=33.395$ ,  $p<.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the white noise auditory condition led to decrease the perception of visual quality ( $p<.037$ ), when compared to other six auditory conditions condition (which did not differ from each other ( $p<.05$  for all)). Furthermore subsequent post-hoc comparisons also revealed that the operating room ambiance auditory condition caused the perception of visual quality for the image with a texture resolution of  $512 \times 512$  to increase as well ( $p<0.016$ ). This increase was not observed for the other six auditory conditions (which did not differ from each other ( $p<.05$  for all)).

The analysis for the model with a texture resolution of  $256 \times 256$  showed significant results ( $F(5)=46.678$ ,  $p<.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the white noise auditory condition led to decrease the perception of visual quality ( $p<.001$ ), when compared to other six auditory

conditions (which did not differ from each other ( $p < .05$  for all)). Furthermore subsequent post-hoc comparisons also revealed that the operating room ambiance mixed with the surgical drill auditory condition caused the perception of visual quality for the image with a texture resolution of  $256 \times 256$  to increase as well ( $p < 0.022$ ). This increase was not observed for the other six auditory conditions (which did not differ from each other ( $p < .05$  for all)).

The analysis for the model with a texture resolution of  $128 \times 128$  showed significant results ( $F(5)=16.52$ ,  $p < .001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the operating room ambiance auditory condition caused the perception of visual quality for the image with a texture resolution of  $128 \times 128$  to increase ( $p < .001$ ). This increase was not observed for the other six auditory conditions (which did not differ from each other ( $p < .05$  for all)).

The analysis for the model with a texture resolution of  $64 \times 64$  showed significant results ( $F(5)=41.502$ ,  $p < .001$ ) Subsequent post-hoc comparisons (Tukey HSD) revealed an interesting combination. The white noise auditory condition led to a decrease in the perception of visual quality ( $p < .001$ ). This detriment was also observed under the heavy metal sound auditory condition for the image with a texture resolution of  $64 \times 64$  ( $p < .001$ ). These detriments were not observed for the other five auditory conditions (which did not differ from each other ( $p < .05$  for all)). Furthermore subsequent post-hoc comparisons also revealed that the operating room ambiance auditory condition increased the perception of visual quality ( $p < 0.022$ ). This increase was also observed under the operating room ambiance auditory condition coupled with the drill sound auditory condition for the image with a texture resolution of  $64 \times 64$  ( $p < .001$ ). These increases were not observed for the other six auditory conditions (which did not differ from each other ( $p < .05$  for all)).

The analysis for the model with a polygon count of  $32 \times 32$  showed significant results ( $F(5)=33.764$ ,  $p < .001$ ) Subsequent post-hoc comparisons (Tukey HSD)

revealed an interesting combination. White noise auditory condition caused a decrease on the perception of visual quality ( $p < .001$ ). This detriment was also observed under the No sound auditory condition for the image with a texture resolution of  $32 \times 32$  ( $p < .001$ ). These detriments were not observed for the other five auditory conditions (which did not differ from each other ( $p < .05$  for all)). Furthermore subsequent post-hoc comparisons also revealed that the operating room ambiance sound coupled with the drill sound auditory condition increased the perception of visual quality ( $p < 0.035$ ). This increase was also observed under the operating room ambiance auditory condition for the image with a texture resolution of  $32 \times 32$  ( $p < .021$ ). These increases were not observed for the other five auditory conditions (which did not differ from each other ( $p < .05$  for all)).

### **3.2.6 Experiment 6: Visual quality defined with respect to texture resolution in a stereoscopic 3D viewing environment with contextual auditory conditions.**

#### *Participants*

Participants consisted of unpaid student volunteers from the University of Ontario Institute of Technology. A total of 18 (8 male and 10 female) volunteers participated in the experiment (average age was 22).

#### *Visual stimuli*

The visual stimuli used in this experiment were the same as the visual stimuli used in Experiment 5. However, here the visuals were presented in a stereoscopic 3D. As with Experiment 5, the stereoscopic 3D “Depth” setting on the Acer Aspire laptop was set to “Default” which, on a scale of 0-15 (where 0 represents no stereoscopic 3D effect and 15 represents the maximum), the “Default” setting is 3. Each 3D model was presented with its corresponding background sound until the participant made a choice. Participants were able to interact with the model (i.e. rotate it using the mouse).

### *Auditory Stimuli*

The auditory stimuli were given by the “contextual auditory stimuli” described in Section 3.1.2.

### *Experimental Method*

The experimental method used was the “contextual auditory cues experimental design” described in Section 3.1.3.

## **Results**

### *Questionnaire*

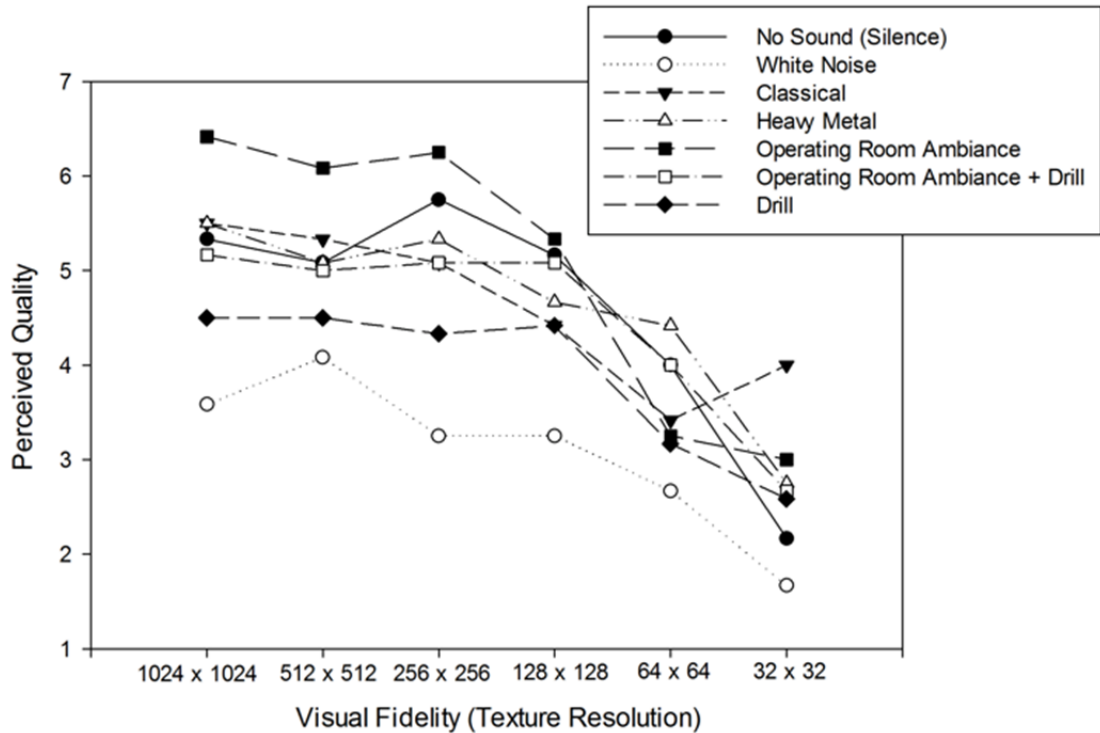
The first question asked whether or not participants regularly played video games (in order to determine their regular use of computer-animated figures): 16 of the 18 participants (89%) played video games regularly. Of these participants, the average number of hours per week spent playing video games was 2.6 hours (minimum one hour and maximum 10 hours), eight of them reported they primarily played console-based video games, six of them primarily played PC-based video games, and two of them primarily played video games on their mobile phone. When asked to indicate their video game genre preference (from a list of given genres), five of the participants prefer “Strategy” games, four “Sports” games and seven “Shooter” games. The questionnaire then asked participants to rate several video game attributes according to their importance on a 7-point scale (1 being least important and 7 being most important). A summary of the average ratings for each of the attributes is provided in Table 16. Participant game and musical genre preferences did not appear to impact experimental results and therefore, will not be discussed further.

Video Game Attribute	Avg. "Importance" (/7.0)
Graphics/visuals	4.4
Sound	3.7
Interaction	2.2
Story/narrative	2.5
Level of challenge/difficulty	3.1

**Table 16. Experiment 6. Summary of video game attribute importance.**

*Experiment*

A summary of the results is presented in Figure 11 where the x-axis represents poly-count variance and the y-axis represents perceived quality (ranging from 1-7) and on Table 17.



**Figure 11. Experiment 6 results. Texture resolution (x-axis) vs. perceived visual quality (y-axis) across each of the seven (contextual and non-contextual) auditory conditions.**



Image	Sound	Mean	Std. Dev.
1	Surgical room ambiance	6.416	0.419
	Surgical room ambiance + Drill	5.166	0.430
	Drill	4.500	1.000
	Heavy Metal Music Sound	5.500	0.430
	No Audio	5.333	0.384
	Noise	3.583	0.569
	Classical Music	5.500	0.962
2	Surgical room ambiance	6.083	0.630
	Surgical room ambiance + Drill	5.000	0.471
	Drill	4.500	1.036
	Heavy Metal Music Sound	5.083	1.370
	No Audio	5.083	1.166
	Noise	4.083	0.687
	Classical Music	5.333	0.272
3	Surgical room ambiance	6.250	0.739
	Surgical room ambiance + Drill	5.083	0.876
	Drill	4.333	0.720
	Heavy Metal Music Sound	5.333	1.054
	No Audio	5.750	0.957
	Noise	3.250	0.569
	Classical Music	5.083	0.687
4	Surgical room ambiance	5.333	1.247
	Surgical room ambiance + Drill	5.083	0.917
	Drill	4.416	0.569
	Heavy Metal Music Sound	4.661	0.769
	No Audio	5.166	1.170
	Noise	3.250	0.569
	Classical Music	4.416	0.957
5	Surgical room ambiance	3.250	0.687
	Surgical room ambiance + Drill	4.000	0.720
	Drill	3.166	1.000
	Heavy Metal Music Sound	4.416	0.957
	No Audio	4.000	0.816
	Noise	2.666	0.942
	Classical Music	3.416	1.228
6	Surgical room ambiance	3.000	1.333
	Surgical room ambiance + Drill	2.666	1.563
	Drill	2.583	1.641
	Heavy Metal Music Sound	2.750	2.132
	No Audio	2.166	1.914
	Noise	1.666	1.122
	Classical Music	4.000	1.276

**Table 17. Experiment 6 results. Mean and Standard Deviation**

The analysis of variance (ANOVA) was selected as the statistical model with two factors: six visual conditions (model of a surgeon's head, visual varying with respect to texture resolution)  $\times$  seven auditory conditions: i) no sound at all, ii) white noise, iii) classical music, iv) heavy metal music, v) surgical drill sound, vi) operating room ambiance mixed with the surgical drill sound and vii) operating room ambiance sound without the surgical drill sound. Main effects and interactions were further analyzed using Tukey HSD post-hoc comparisons. The results revealed only that only the main effects for visual ( $F=9.83$ ,  $p<.001$ ) and auditory ( $F=27.19$ ,  $p<.001$ ) conditions, were significant and there was no significant interaction between the sound and image ( $F=7.14$ ,  $p<.761$ ). More specifically, on average the participants were relatively accurate at discriminating the quality of the visual images, where the lower quality visuals were judged as lower and higher quality visuals were judged as higher. With the exception of the white noise auditory condition, visual quality was not affected by ambient sound. Ambient sound consisting of white noise led to a significant reduction in the perception of the visual quality across all of the visuals considered.

### **Visual Quality Perception with Contextual and Non-Contextual Auditory Cues Discussion**

Experiments 5 and 6 examined the perception of visual quality (defined with respect to the texture resolution of a visual model) under both contextual and non-contextual ambient auditory conditions (with respect to the visual scene). Results suggest that contextual ambient auditory cues increase our perception of visual quality while non-contextual cues in the form of white noise may lead to a decrease in visual quality perception particularly when considering lower quality visuals. These results are consistent with the results of Experiments 1-4 and the work of others. More specifically, Experiments 1-4 indicate that perception of visual quality of a virtual model is dependent on ambient (background) sound. In those experiments, it was observed that ambient white noise resulted in a decrease of visual quality perception for visual quality defined

with respect to polygon count and texture resolution. In contrast, ambient sound consisting of either classical music or heavy metal music led to an increase in the perception of visual quality when considering both polygon count, and texture resolution.

Although not specific to multimodal interactions, it has also been well established that sound (and music in particular), can effect moods, emotions and behavior of both individuals and groups, and music is considered to lie on a continuum from highly stimulating and invigorating to soothing or calming [Gaston, 1968; Hallam et al., 2002]. Finally sound and music can also influence task performance. For example, a study by Morsbach et al. [1986], determined that the cries of infant distracted listeners instructed to pay attention to a simple cognitive task. Building on this study, Chang and Thompson [2011], demonstrated that whines, cries, and motherese (child-directed speech) distracted listeners completing simple mathematical (subtraction) problems.

Here, in Experiments 5 and 6, the visual stimuli consisted of a single surgeon (upper body only) holding a surgical drill. There were two key parts on the model presented: i) the (upper body of) the surgeon, and ii) the surgical drill that the surgeon was holding. In order to be congruent with the idea of contextual cuing, included three contextual ambient auditory conditions: i) operating room ambiance, ii) operating room ambiance + drill sound, and iii) drill (alone). Results indicate that contextual ambient sounds can lead to an increase in the perception of visual quality. However in Experiment 5 (which also examined the effect of contextual ambient auditory conditions on visual quality perception but in a non-stereoscopic 3D viewing environment), this increase was observed for only two of the three contextual ambient auditory conditions and more specifically, for the operating room ambiance, and operating room ambiance + drill auditory conditions and not for the drill auditory condition despite the fact that the surgeon within the visual scene was holding a surgical drill. In Experiment 6, contextual related auditory cues showed to influence visual quality but this influence was not statistically significant, this might be explained by the extra information provided in experiment 6 and more specifically, the addition of stereoscopic 3D (depth), and this additional information may have reduced the ability of the participants to pay attention to

the auditory cues presented.

Assuming the premise that contextual auditory cues increase visual quality perception, the results presented here suggest that participants perceived the operating room ambiance cue as providing the greatest contextual component, thus leading to the observed increase for this condition. This discrepancy leads to the question of just how contextual auditory cues may be defined. Contextual cuing is a concept in psychology that refers to the manner in which the human brain gathers information from external elements and their surroundings [Chun and Jiang, 1998]. External elements refer to the different sensory inputs (audio, visual, haptics, etc.) that will drive the attention based on past experiences with the regularities of the world. Regularities of the world can be the objects around a given target object, their spatial layout, their trajectories, sound involved, etc.

One explanation is provided by Scott and Canter [1997], who compared the perceived similarity of a diverse set of familiar places based on a photographic sorting task to similarity measures based on a second task that immediately followed. In the second task, the same participants were instructed to now regroup the labeled pictures based on imagined actions, thoughts and feelings they might have at the places represented. The authors reported that the photograph-based similarity mapping indicated a focus on physical features and compositions in the pictures (e.g., landform and water). The second grouping also showed the effects of the landform and to a lesser extent the water factor, but the participants' verbal descriptors focused more on 'activities, people, sounds, emotions and physical conditions'. It was concluded that the type of task that was asked to perform drove subject's attention being able to modify subject's focus over the same action. Here, the task of the participants was to rate the visual quality of the presented image. Participants were not instructed to focus on any part of the visual scene including the drill held by the surgeon. Given the results presented here, participants did not focus on the drill at all but rather assumed that the surgeon provided the context of the visual scene (i.e., the operating room ambiance led to the greatest increase in visual quality perception). The drill ambient auditory condition appears to operate as a distractive cue alone and when mixed with the operating room ambiance (i.e., although there was an

increase in visual quality perception with the operating room ambiance + drill condition, this increase was less than the increase with the operating room ambiance condition.

The significance of Experiments 5 and 6 lies in the fact that contextual sound is shown to influence the perception of visual quality (at least when presented in non-stereoscopic 3D). However, there are many questions that must be considered before a greater understanding regarding the effect of visual quality presented in conjunction with auditory cues on visual quality perception.

In addition to the different types of auditory cues (contextual dependent and non-dependent) and the different types of visual presentations (non-stereoscopic 3D and stereoscopic 3D) in order to obtain a greater understanding on the audio-visual multimodal interaction phenomenon our next experiment will also consider the effect of user interaction on visual quality perception and, more specifically, actively involving the participant by having them perform a simple task within the virtual world under various auditory conditions (contextual and non-contextual). A simple measurement of performance could be defined and this will allow for the comparison of sounds that are unnatural to an interaction with a virtual object being acted upon with more natural sounds.

## **Sensory Integration (Audio-Visual) While Performing Goal Directed Actions**

The previous six experiments examined the influence of contextual and non-contextual ambient auditory conditions of visual quality within both a stereoscopic and non-stereoscopic viewing environment. Experiment 7 builds upon the first six experiments by examining the influence of both contextual and non-contextual ambient auditory conditions on visual quality and also on task completion time. In contrast to the previous six experiments, participants in Experiment 7 had to interact with the environment and more specifically, they had a task to complete (navigate the environment to reach and pick up a tool).

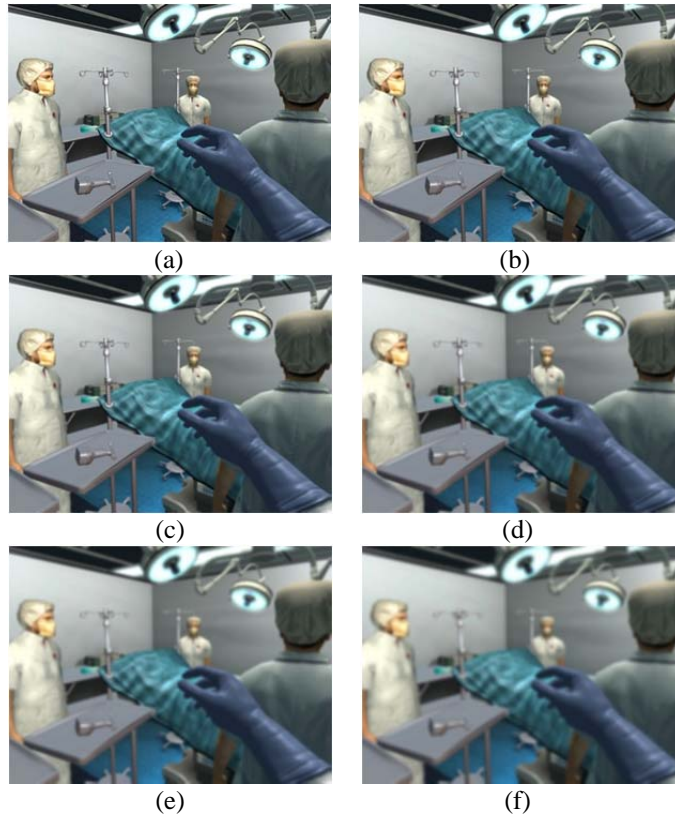
### **3.2.7 Experiment 7: Visual quality defined with respect to level of visual blurring under both contextual and non-contextual auditory cues.**

#### *Participants*

Participants consisted of unpaid student volunteers from the University of Ontario Institute of Technology (UOIT). A total of 12 volunteers (6 male and 6 female) participated in the experiment (average age was 20 years old).

#### *Visual Stimuli*

The visual scene consisted of six rendered (3D) versions of an operating room with various tools, and equipment (see Figure 3). Within the operating room were three non-player characters (nurses) which, for the purposes of this experiment, they remained static and did not afford any interaction with the participants. Each of the rendered versions of the scene was filtered (dynamically) with a blurring filter, which caused the scene to be blurred. The level of blurring introduced to each of the six scenes varied linearly from 0 to 1 with 0 (no blurring introduced; i.e., the original, reference scene) 1 (the highest level of blurring). The blurring effect was accomplished dynamically using an OpenGL GLSL-based shader. The blurring was accomplished by (on a per-pixel basis) sampling several discrete areas around a center point (each pixel) and averaging the result. An example of the six levels of blurring is provided in Figure 12.



**Figure 12. Experiment 7. The levels of visual quality (defined with respect to image blurring) considered in this experiment. (a) Original (non-filtered) version. (b) - (f) Application of the blurring filter with an increase in the blurring effect.**

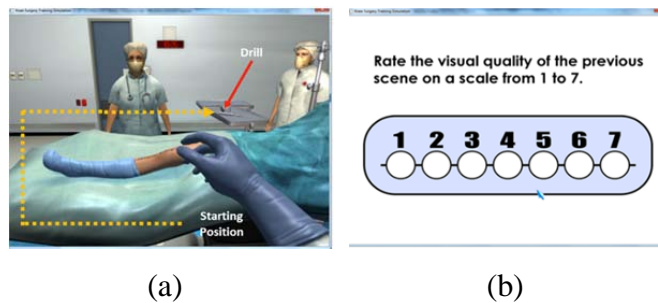
### *Auditory Stimuli*

Four ambient (background sound) conditions were examined: i) no sound, ii) white noise (see Section 3.1.2), iii) operating room ambiance mixed with a surgical drill sound (see Section 3.1.2), and iv) operating room ambiance without the drill sound (see Section 3.1.2). For all trials, the auditory stimulus (when present) began playing at the start of the trial and was stopped once the participant chose the surgical drill. The sound pressure level of the output sounds was 60dB (about the same level as typical conversation [Morfev, 2005]) measured with a Radio Shack sound level meter (model 33-2055). All auditory stimuli were monophonic and were output with a pair of Sony MDR 110LP headphones.

### *Experimental Method*

Participants were seated in front of the laptop computer, which was used to

conduct the experiment. Participants were provided with an overview of the experiment followed by a description of their required task by one of the experimenters. In each trial, participants were presented with one of the six versions of the rendered operating room shown in Figure 12, in conjunction with one of the four ambient sound conditions described above. Figure 13 illustrates the view of the operating room presented to the participants at the start of the experiment. Their task was to navigate through the operating room from their starting position to a point in the room which contained a tray and a surgical drill and pick up the drill (as outlined in Figure 13(a), they had to navigate around the bed and one of the NPC nurses to reach the drill).



**Figure 13. The experimental environment (a) Then view of the operating room environment at the start of each experimental trial. (b) Upon picking the drill, participants were asked to rate the visual quality perception**

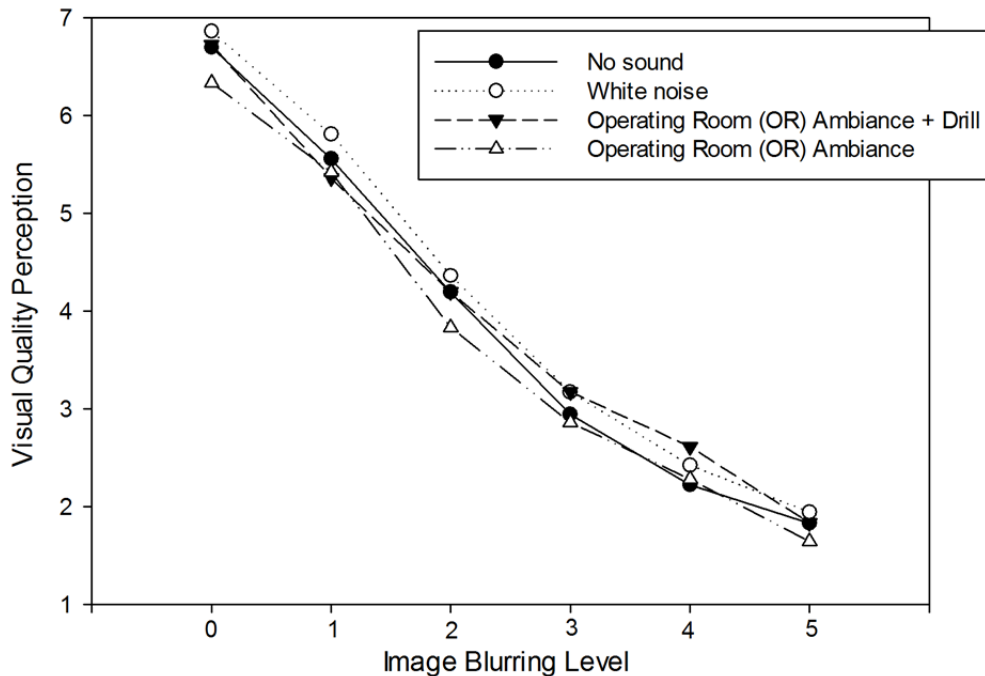
Navigation through the environment was accomplished in a first-person perspective (taking on the role of the surgeon) using the standard arrow keys (to move the “player”) and mouse (to move the camera). Within this first-person view, the hand and lower arm of the participant’s avatar was displayed. Choosing the surgical drill involved moving their avatar’s hand over the drill and clicking the left mouse button. Once the drill was chosen, it appeared in the hand of the “surgeon” and the participant was prompted to rank the visual scene with respect to their perceived visual quality on a scale from 1 (lowest perceived quality) to 7 (highest perceived quality); see Figure 13(b). Aside from interacting with the drill, there were no other interactions permitted (e.g., the participants could not interact with any of the NPCs or other objects in the room). Entering their choice



signaled the end of the trail; the following trial began after the user clicked the “Continue” button that appeared on the visual quality ranking screen after the participant entered their choice. Each of the six rendered versions of the operating room and each of the four ambient sound combinations (24 combinations in total) were repeated three times for a total of 72 trials, all of which were presented in a randomized ordering. The experiment took approximately 30 minutes to complete and all participants completed it in a single session. As with the previous six experiments, the Acer Aspire laptop was used here as well. The operating room environment was viewed in “full screen” mode.

## **Results**

Two different variables were analyzed as outcomes for this experiment. The first one was the participants’ perceived quality of the visual scene in the presence various ambient auditory cues and the second was the total task completion time. Results of the first analysis (perceived visual quality) are summarized in Figure 14 where the x-axis represents visual quality (amount of visual blurring) and the y-axis represents perceived quality (ranging from 1-7; error bars represent standard deviation). Inspection of Figure 14 and clearly indicate that visual quality perception decreases as the level of blurriness increases.

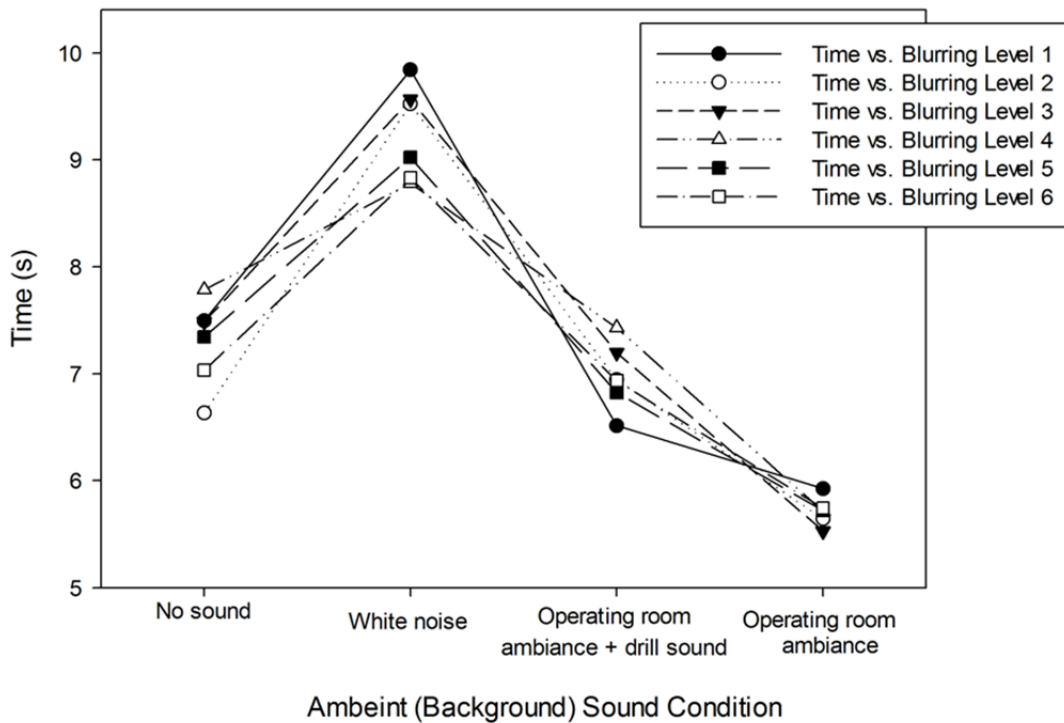


**Figure 14. Experiment 7 results. Visual Quality perception (x-axis) vs. image blurring level under four ambient sound conditions. Image quality is defined with respect to a blurring of the entire scene with Level 1 representing the least blurring and Level 6 the greatest blurring.**

The analysis of variance (ANOVA) was selected as the statistical model with two factors: six visual conditions  $\times$  four ambient (background) auditory conditions. Main effects and interactions were further analyzed using the Tukey HSD post-hoc comparisons. In addition to the main effects for visual conditions ( $p < .000$ ) and ambient sound ( $p < .130$ ), there was no significant interaction between the ambient sound and image ( $p < 1.00$ ). Post-hoc tests revealed that the type of ambient sound did not influence the perceived quality of the visual scene irrespective of the level of blurring. Although there was no difference in visual quality perception with respect to auditory conditions, there was a significant difference between the each of the six visual conditions examined. In other words, participants did perceive a difference in visual quality across each of the six scenes they were presented with.

The second variable analyzed was the task completion time (i.e., the time taken for the participant to reach the surgical drill and “choose” it from the initial starting

point). Results of the second variable analyzed (task completion time) are summarized in Figure 15 where the x-axis represents the ambient sound condition and the y-axis represents time (in seconds). A summary of the task completion time averaged across each of the six visual quality conditions for each of auditory conditions is provided in Table 18.



**Figure 15. Experiment 7 results. Ambient auditory condition (x-axis) vs. time to complete the task (y-axis) in the presence of contextual and non-contextual auditory conditions.**

AVERAGE TASK COMPLETION TIME	
Ambient Sound Condition	Time (s) ± Std. Dev. (s)
No sound	7.30 ± 0.41
White noise	9.26 ± 0.44
Operating room ambiance + Drill	6.97 ± 0.32
Operating room ambiance	5.71 ± 0.13

**Table 18. Task completion time averaged across auditory condition.**

The analysis of variance (ANOVA) was selected again as the statistical model with two factors: 6 visual conditions × 4 ambient (background) sounds. Main effects and interactions were further analyzed using Tukey HSD post-hoc comparisons.

Furthermore from the main effects for visual conditions ( $p < .902$ ) and ambient sound ( $p < .000$ ), there was no significant interaction between the sound and image ( $p < .936$ ). Post-hoc tests revealed that the type of ambient sound influenced the task completion time. The auditory condition consisting of white noise sound resulted in the largest task completion times while auditory condition consisting of the operating room ambiance alone resulted in the least amount of time to complete the task. Interestingly, the auditory condition consisting of the operating room ambiance + drill sound showed a similar effect (with respect to task completion time) to the no-sound condition, both of which resulted in task completion times that were greater than task completion times of the operating room ambiance auditory condition.

### **Audio-Visual Sensory Integration while Performing Goal Directed Actions Discussion.**

Experiment 7 examined the effect of contextual ambient (background) sound on visual quality perception (defined with respect blurring of the scene), and task completion time while conducting a simple task (navigating through a virtual operating room to reach a tray of surgical instruments) was examined. Results indicate that ambient (background) sound has no influence on the perception of visual quality irrespective of the level of blurring or whether the auditory cues were contextual or non-contextual with respect to the visual scene. However, results also indicate that ambient sound did have an effect on task completion time and more specifically, white noise led to a large increase in task completion time while contextual ambient sound consisting of operating room ambiance resulted in the lowest task completion time.

Experiments 1-6 revealed that the perception of visual quality of a virtual model is dependent on ambient (background) sound. Experiments 1-6 examined visual quality perception of a single object with visual quality defined with respect to polygon count and texture resolution in the presence of various ambient auditory conditions (that were non-contextual with respect to the visual scene) including white noise, classical music, and

heavy metal music. In Experiments 1-6, it was observed that auditory cues did in fact influence the perception of visual quality. White noise resulted in a decrease of visual quality perception and classical music increased the perception of visual quality for visual quality defined with respect to polygon count, particularly for the images corresponding to the highest polygon counts.

In Experiment 7, both contextual and non-contextual ambient auditory cues were considered. Furthermore, the visual scene was modeled after an actual operating room and contained many objects (the operating room itself, the bed, lights, NPCs, etc.) as opposed to a single model considered in Experiments 1-6. Also, in Experiment 7, visual quality was defined with respect to blurring of the scene which was chosen to approximate the effect of reduced texture resolution as in Experiments 2, and Experiments 4-6 (the virtual scene employed in Experiment 7, there were far too many objects/models making it impractical to individually reduce the texture resolutions of each). In addition, in Experiment 7, participants were required to complete a task that involved navigating through the operating room to reach the tray of surgical instruments and choose the surgical drill and this may have led to overloading of their mental workload with differential effects of sensory manipulations on perception and action particularly when considering that the participants were undergraduate students in non-medical areas of study and not familiar with surgery and operating rooms in particular. As speculated by Milner and Goodale [1995], in their two-stream hypothesis of visual processing, visual information is processed in different areas of the brain when the intention of the visual information is to make a judgment (the ventral stream) or to make an action (the dorsal stream). Future work will repeat this experiment but with participants within the medical field who are familiar with an operating room to test the effect of ambient sound on both visual quality perception and task completion time given the prior experience with the surgical environment. Future work will also repeat this experiment in stereoscopic 3D to examine multi-modal cue interaction within a stereoscopic 3D environment and what effect this may have on task completion time.

Although ambient sound did not affect visual quality perception, it did have an effect on the total time required to complete the task (“task completion time”) and more specifically, ambient sound consisting of white noise resulted in the largest task completion times while the operating room ambiance sound resulted the lowest task completion time. Interestingly, the ambient sound condition consisting of the operating room ambiance with drill sound showed a similar effect (with respect to completion time) to the no-sound condition, both of which were greater than the auditory condition consisting of the operating room ambiance only. This result is supported by previous work. Conrad et al. [2010] examined the effect of background sound (including classical music (Mozart), and heavy metal music to induce “auditory stress”) on laparoscopic surgery. They found that stressful music (e.g., heavy metal) had a negative impact on task completion time but did not impact task accuracy. They also found that classical music had a variable effect on time until task completion but resulted in greater task accuracy amongst all participants (laparoscopic surgeons). Although task accuracy was not examined here, it has been shown that ambient sound consisting of white noise has a negative impact on task completion time similar to the “stressful” music of the Conrad et al. [2010], study.

Collectively, the results presented here and in previous experiments, suggest that sound can affect various aspects of a virtual simulation/serious game. Distracting sounds such as white noise can not only decrease the perception of visual quality, but also can detrimentally affect task completion time. Designers and developers of virtual simulations and serious games should work closely with educators and content experts to explore and devise proper ways to help trainees in learning how to perform under the presence of potentially distracting sounds which, in many situations, characterize the real-world environment and cannot be eliminated. Perhaps virtual simulations and serious games can be used to explicitly acquaint trainees with such “real-world” distracting sounds while performing technical tasks to minimize any negative effects when distracting sounds are encountered in the real-world.

# 4 Conclusions

## 4.1 Results

In this thesis, the results of a series of seven experiments that examined the effect of background sound (contextually related and non-related with respect to the visual scene) on the perception of visual quality defined with respect to (texture resolution, and polygon count) presented in stereoscopic and non-stereoscopic 3D were presented. The results of Experiments 1-4 reveal that the perception of visual quality of a virtual model is dependent on ambient (background) sound. In these particular experiments, it was observed that ambient white noise resulted in a decrease of visual quality perception for visual quality defined with respect to polygon count and texture resolution. In contrast, ambient sound consisting of either classical music or heavy metal music led to an increase in the perception of visual quality when considering both polygon count, and texture resolution under different visual presentations (stereoscopic 3D and non-stereoscopic 3D).

Experiments 5 and 6 considered contextual auditory cues. Results indicate that contextual auditory cues (with respect to the visual scene) and more specifically, the operating room ambiance, and operating room ambiance + drill auditory conditions lead to the greatest increase in visual quality perception. This influence was greater than the influence exhibited by the classic music and heavy metal music (non-contextual) auditory conditions also considered in Experiments 5 and 6.

Experiment 7 analyzed the effect of multi-modal interaction (audio-visual) while performing goal directed actions. Surprisingly, an influence was evident with respect to the time required to complete the task (the “task completion time”) and not on the visual quality perception as demonstrated in Experiments 1-6. More specifically, white noise ambient sound caused the task completion time to increase while ambient sound consisting of operating room ambiance led to a reduction in the task completion time. The results of these seven experiments indicate that ambient (background) sound does have some influence over visual quality perception.

## 4.2 Background

The world we live in is full of information and our senses are constantly receiving this information and processing it. The central nervous system integrates these various sensory inputs to arrive at a coherent representation of the external world, and this is termed multimodal sensory interaction. Various models of sensory integration have been proposed; describing how we pay attention to the different cues we received from the external world and how these cues are processed.

For example, Broadbent's [1958], model hypothesized that information from all of the stimuli presented to our senses at any given time enters a *sensory buffer*. One of the inputs is then selected on the basis of its physical characteristics for further processing by being allowed to pass through a filter. Because we have only a limited capacity to process information, this filter is designed to prevent the information-processing system from becoming overloaded. The inputs not initially selected by the filter remain briefly in the sensory buffer, and if they are not processed, they decay rapidly. Broadbent assumed that this filter rejected the unattended message at an early stage of processing.

One of the experiments Broadbent conducted involved sending a message (a 3-digit number) to a participant's right ear, and a different message (a different 3-digit number) to their left ear simultaneously. Participants were asked to listen to both messages at the same time and repeat what they heard (this is known as a "dichotic listening task"). Results showed that participants repeated what was heard in the right ear, followed by what was heard in the left ear and this led Broadbent to produce his "filter" model to explain how selective attention operates (Figure 16). Broadbent concluded that we could pay attention to only one channel at a time hence this is referred to as a single channel model.



## Broadbent's Filter Model of selective attention

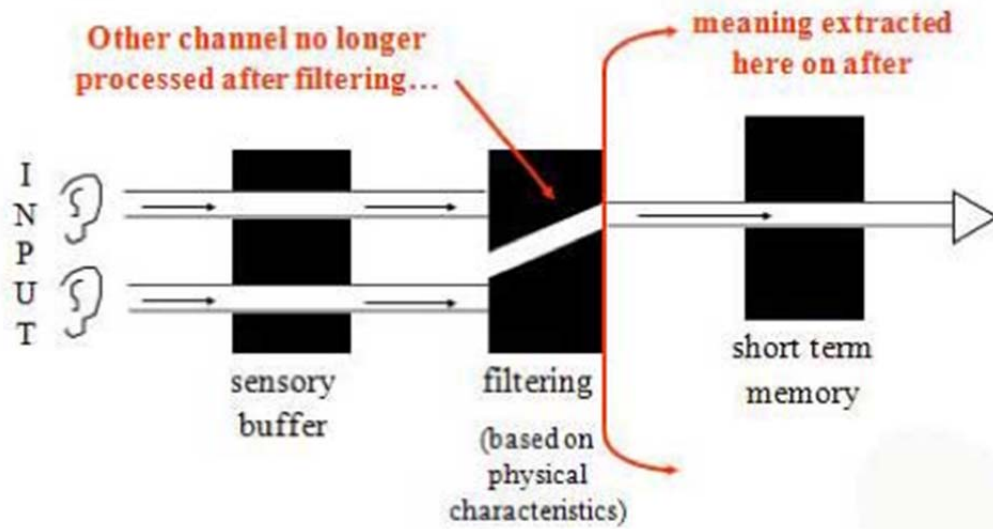
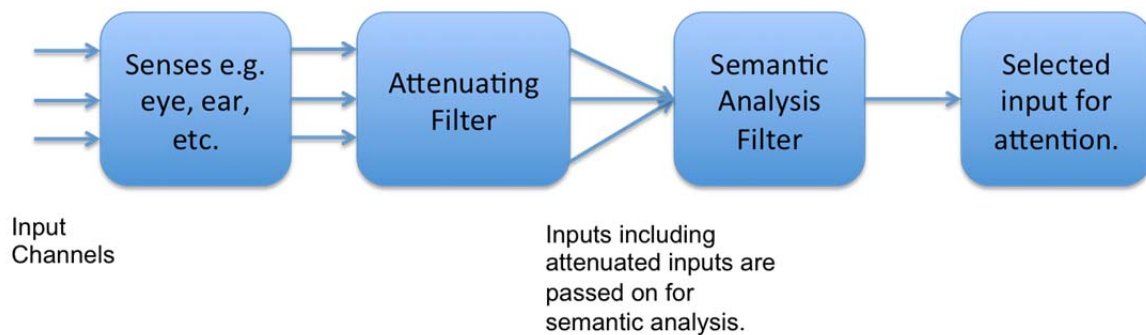


Figure 16. Broadbent's filter model of attention. Reprinted from [McLeod, 2011].

Broadbent's model has been analyzed for many years, and more recent research has found that it does not cover all the variability present in multimodal sensory interactions. For example, Broadbent's model predicts that hearing your name when you are not paying attention should be impossible because unattended messages are filtered out before you process the meaning. This is in contrast to the "cocktail party effect", which refers to the ability to focus one's auditory attention on a particular stimulus while filtering out a range of other stimuli, much the same way that a partygoer can focus on a single conversation in a noisy room [Bronkhorst and Adelbert, 2000]. This effect is what allows most people to "tune in" to a single voice and "tune out" all others but not cancelling them at all. It may also describe a similar phenomenon that occurs when one can immediately detect words of importance originating from unattended stimuli, for instance hearing one's name in another conversation [Woods et al., 1995; Conway, 2001]. This implies that people usually process the meaning before filtering out the cue, opposite to what Broadbent's model proposes.

As an alternative to Broadbent's model, Treisman [1964], proposed a model also based on selective attention. Selective attention requires that stimuli are filtered so that attention is

directed. However, the difference between Treisman's and Broadbent's models is that the filter in Treisman's model attenuates rather than eliminates the unattended material. However, if a non-attended channel includes your name, for example, there is a chance you will hear it because the material, although attenuated, is still there (see Figure 17).



**Figure 17. Treisman's attention filter model. Reprinted from [McLeod, 2011].**

Treisman agreed with Broadbent in that there is a bottleneck, but disagreed with its location. Treisman conducted experiments using the speech shadowing method. Typically, in this method participants are asked to simultaneously repeat aloud speech played into one ear (called the attended ear) while another message is spoken to the other ear. In one type of Treisman's shadowing experiments, identical messages were presented to the two ears but with a slight delay between them. If this delay was too long, then participants did not notice that the same material was played to both ears. In an experiment with bilingual participants, Treisman presented the attended message in English and the unattended message in a French translation. When the French version lagged only slightly behind the English version, participants could report that both messages had the same meaning. Therefore, the unattended message was being processed for meaning and Broadbent's filter model, where the filter extracted information only with respect to physical characteristics, could not explain these findings.

In terms of visual attention, several studies have analyzed how humans focus their visual attention. Posner [1978], found that when people are told to fixate on one part of the visual field, it is still possible to attend to stimuli approximately seven degrees on either

side of the fixation point. Also, attention can be shifted more quickly when a stimulus is presented in an “expected” rather than “unexpected” location (i.e., when the visual stimulus is presented on the same position where the subject attention is located). Posner [1980], calls this covert attention (also known as the spotlight theory).

Evidence in support of the spotlight theory came from a study by LaBerge [1983]. He presented participants with five-letter words and a target. The target could appear in the position of any of the five letters in a word. In one condition participants task was to respond to the whole word by categorizing it. In another condition, the task was to respond to the middle letter only by categorizing it. As a secondary task on both of the conditions mentioned before a target was presented when participants were performing the primary task and participants were asked to hit the target. It was expected that, when being asked to respond to the middle letter, a narrow attentional beam would be employed. The results indicate that when participants were asked to categorize the word their attentional beam was broad, so it did not matter at which point on the display the visual target was presented, response times were equally fast [LaBerge, 1983]. However, when participants focused on the middle letter, the position of the target was critical. Response times were significantly slower when the target was not presented in the center of the five-letter display [LaBerge, 1983].

Although the models described above were proposed for multiple senses (i.e., sensory interaction), most cases were tested by stimulating one of the senses only. However, a number of studies did examine multimodal sensory interaction based on these previous models [Shaffer, 1975; Allport et al., 1972].

McLeod [2011], called the multimodal interaction phenomena as “divided attention” and refers to the ability to perform two tasks simultaneously. Eysenck and Keane [1995], identify three factors that affect the performance of dual tasks (DT).

1. **Task difficulty** (the more difficult the less successful the DT performance).
2. **Task practice** (improve DT performance).
3. **Task similarity** (non-similar tasks are easier, perhaps use different resources).

In terms of task similarity, several studies have found that task similarity is an important factor in determining the ability to perform two tasks at the same time. Researchers such as McLeod [1977], Treisman and Davies [1973], have found that dual-task performance is greatly improved when the two tasks are dissimilar (i.e., in different sensory modalities). However, while task similarity is relatively easy to measure and manipulate in a laboratory environment, it is much harder to measure the similarity of more everyday tasks (i.e., driving or playing the piano) [McLeod, 2011].

Practice is another factor determining dual-task performance. Spelke et al. [1976], have found that, with practice, participants can greatly improve their dual-task performance (e.g., dictation and reading for comprehension). They even go so far as to suggest that with practice, we can perform two tasks together equally well. However, this claim has been challenged and further studies have shown that although practice can increase dual-task performance, performance is not as good as when each task is performed alone [McLeod, 2011].

The third main factor determining dual-task performance is task difficulty. Sullivan [1976], and Duncan [1979], have found that increasing the difficulty of the tasks reduces performance.

These key factors for multimodal interaction can also be explained by a concept called *cognitive load*. The term cognitive load is used in cognitive psychology to illustrate the load related to the executive control of working memory (WM). Theories contend that during complex learning/training activities, the amount of information and interactions that must be processed simultaneously can either under-load, or overload the finite amount of working memory one possesses. All elements must be processed before meaningful learning can continue [Pass et al., 2004]. This supports Treisman's [1964], attention filter model and therefore can also be used to explain McLeod's [2011], "divided attention" model previously described.

### **4.3 Explanation of the Results With Respect to Existing Theories**

The results of Experiments 1-7 can be explained with the models described above. As the experiments within this thesis have shown, ambient (background) sound influences visual quality perception. By using Tresiman's bottleneck attenuation model, it can be hypothesized that the participants of Experiments 1-7 were analyzing the different stimuli they were presented with (auditory and visual), one at the time, and depending on the task they were asked to perform, their cognitive load differed. More specifically, the difference between the results of the experiments can be explained by cognitive load. Experiments 1-6 showed that ambient sound influences visual quality perception. However, this influence differed across the various conditions examined. For example, in contrast to a non-stereoscopic 3D viewing environment, when considering a stereoscopic 3D viewing (Experiments 3, and 4), the user is presented with additional information (depth) and this additional information requires greater cognitive resources to process, due to that difference between the visual presentations in 3D stereoscopic vs. non-stereoscopic 3D was present only for the higher quality visual models (polygon count greater than 1,250, and texture resolution greater than  $256 \times 256$ ). Results indicate that stereoscopic 3D might result in a more sensitive perception of visual quality for higher quality visuals. As a further example, in contrast to Experiments 1-4, Experiments 5 and 6 considered contextual ambient auditory stimuli. Contextual ambient auditory cues led to a greater increase in visual quality perception than the increase with non-contextual auditory stimuli. Recalling the "task similarity" factor proposed by Keane [1995], described above, it is hypothesized that participants did not require using the same amount of cognitive resources when presented with contextual auditory stimuli due to the contextual similarity for both of the cues. Contextual related cues (audio-visual) can be easily integrated by the attention model and due to that caused participants to perceive visual quality higher. That is, the participants can pay more attention to the visual quality when cognitive resources are not being fully used. When participants were presented with the non-contextual auditory conditions, due the no contextual relation between the cues, it is hypothesized that participants needed to use additional cognitive resources and this resulted in a smaller increase of visual quality perception.

In contrast to Experiments 1-6, in Experiment 7, participants were required to navigate through a virtual operating room and pick up a drill (using a mouse and keyboard to do so) that was located in the opposite end of the starting position. In order to navigate through the room, participants needed to synchronize mouse movements, which rotated the avatar head, and pressing the arrow keys on the keyboard, which moved the avatar around the environment. At the same time participants were also present with ambient auditory cues and they were asked to rate the visual quality perception of the scene (visual quality was defined by blurriness of the visual scene).

The task of Experiment 7 was different (more complex because of the navigation component) from the tasks of Experiments 1-6. Given the greater complexity associated with the task of Experiment 7, it is hypothesized that greater cognitive resources may have been required to complete the task, leaving less processing for ambient auditory cues and therefore ambient sound did not influence visual quality perception.

The main task for the participants in Experiment 7 was to navigate through the operating room and pick up the drill. Once the drill was picked up, a pop-up window was displayed prompting the participants to rate the perceived visual quality (the pop-up window covered the majority of the visual scene and the ambient auditory cues were stopped once the pop-up window appeared). In contrast, in Experiments 1-6, the visual and auditory stimuli were present while the participants were prompted to rate the visual quality. In other words, in Experiment 7, participants were rating visual quality perception after being presented with the auditory and visual stimuli, not at the same time. Therefore, since synchronizing the movements required to move their avatar within the virtual operating room, involved some increased amount of cognitive resources, and as a result, participants were not able to differentiate the different visual quality conditions. This is supported by [Pass, 2003], who proposed that high cognitive load usages reduce short memory retention. However, despite the fact that ambient sound did not exhibit any influence over visual quality perception, sound did influence task competition time. This can be explained by the task practice factor proposed by Keane [1995], who proposed that the more a specific task is practiced, the better you become at doing that task in conjunction with some other task. Since each condition of Experiment 7 was repeated

three times, participants performed a total of 72 trials of navigating through the operating room from the starting position to the position of the drill in order to choose/pick the drill. Repeating this action 72 times this can be considered as training [Mayer, 2003], and can therefore lead to a decrease in the required cognitive resources. As a result, since participants were not using the same amount of cognitive resources during the latter trials as required in the earlier trials, they were able to unconsciously (Tresiman's attention filter model [1964]) start paying attention to the auditory cues presented, resulting in a reduction in the task completion time when presented with contextual dependent auditory cues. I would like to clarify that the fact that participants due to the practice start using less cognitive resources to accomplish the task results in a capability to pay attention to the other cues (e.g., auditory cue) rather than an increase of the performance quality. Influence is showed only on completion time rather than on visual quality perception because of the nature of the asked task. Since participants were asked to navigate the virtual environment and pick up a drill, they were always more focused on the motor task rather than in rating visual quality.

#### **4.4 Serious Games**

As previously described, auditory cues can influence visual quality perception and this influence was explained above by a phenomenon called "cognitive load" which describes why in some cases, depending on the type of auditory cue presented (contextual or non-contextual with respect to the visual scene) and the task difficulty, the influence over visual quality perception differs.

The work presented in this thesis is part of a larger-scale effort to examine virtual environment quality, multi-modal interactions, user-specific factors (e.g., personality, learning style, existing knowledge, level of attention, motivation etc.) and their effects on knowledge transfer and retention when using virtual simulation/serious games. The results presented here indicate that multimodal interaction is fundamental to perceiving the external world and therefore, any virtual environment (serious games, etc.) should include and account for multimodal interactions. However, multimodal interactions are constrained by attention models that describe how our cognitive resources may become

loaded depending on the senses that are being stimulated and the characteristics of the task performed [Pass, 2003]. In addition, these models also describe how these cognitive resources can decrease through practice of a particular skill. With this in mind, it is hypothesized that this will lead us to the point where knowledge transfer and retention may be constrained by cognitive resource demands required by the audio-visual interactions of users of serious games.

“No one learns to run before learning how to walk”. If a user of a serious game is presented a game that requires large cognitive resources, they will be so focused on performing the required task that they may ignore some or all of the cues presented to him/her. If, instead of presenting the user to a difficult task from the onset, they are initially presented with a simple task and only when they reach a certain level of proficiency do they move to the next level, the practise that will arise from completing the simpler tasks may limit the cognitive resources required at the next (more difficult) level.

## **4.5 Future Work**

Future work will further analyze the cognitive load phenomenon when presented with a goal-oriented task and examine what effect this will have on training/learning, retention, and transferability of an acquired skill. An example of an experiment that will be conducted to examine this is repeating Experiment 7 in stereoscopic 3D. The stereoscopic 3D environment should add additional cognitive resources and thus should lead to a greater task completion times than that observed in Experiment 7.

Here, only audio and visual cue interactions were analyzed despite the availability of other sense, most notably touch (haptics) and olfactory. Therefore, future work will investigate the effect of audio, visual, audio and haptic cues on the perception of visual quality. Future work will also examine multimodal interactions within a virtual environment that incorporates a learning task. This will allow us to examine multimodal effects on learning and knowledge transfer and retention more specifically.



## 5 REFERENCES

- AACN - American Association of Colleges of Nurses. *The Role of the Clinical Nurse Leader*. American Association of Colleges of Nursing, Washington, DC, 2003.
- Alden D. Experience with scripted role-play in environmental economics. *Journal of Economic Education*, 30(2):127–132, 1999.
- Aldrich C. *Simulations and the future of learning*. San Francisco, CA, USA: John Wiley, 2004.
- Alexander M., Brunyé T., Sidman J., and Shawn A. From Gaming to Training: A Review of Studies on Quality, Immersion, Presence, and Buy-in and Their Effects on Transfer in PC-Based Simulations and Games. *DARWARS* (3) 135-146, 2005.
- Allen J. A., Hays R. T., and Bufford L. C. Maintenance training, simulator fidelity, and individual differences in transfer of training. *Human Factors*, 28(5):497-509, 1986.
- Andrews D. H., Carroll L. A., and Bell H. H. The future of selected fidelity in training devices. *Educational Technology*, 35(6):32-36, 1996.
- Becker A. B., Warm J. S., Dember W. N., and Hancock P. A. Effects of jet engine noise and performance feedback on perceived workload in a monitoring task. *International Journal of Aviation Psychology*, 5(1):49– 62, 1995.
- Bergeron B. *Developing serious games*. Hingham, MA. USA: *Thompson Delmar Learning*, 2006.
- Boot W. R. *Enhancing warfighter training and performance*. Naval Training Principal Investigator Program Review. 2-4 August 2005, Arlington, VA, 2005.
- Brogni A., Vinayagamoorthy V., Steed A., Slater M. Variations in Physiological Responses of Participants During Different Stages of an Immersive Virtual Environment Experiment. *Virtual Reality Software and Technology'06*, November 1–3, 2006, Limassol, Cyprus, 2006, pp. 376-382.
- Brown J. S., Collins A., and Duguid S. Situated cognition and the culture of learning. *Educational Researcher*, 18(1):32-42,1989.
- Cacioppo J. T., Petty R. E., and Quintanar L. R. Individual differences in relative hemispheric alpha abundance and cognitive responses to persuasive communications. *Journal of Personality and Social Psychology*, 43(3):623-636, 1982.

Cacioppo J. T., Petty R. E., Feinstein J. A., and Jarvis W. B. G. Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, 19(2):197-253, 1996.

Chalmers A. and Debattista K. Level of realism for serious games. *International Journal of Virtual Reality*, 25(3):1101–1108, 2009.

Chalmers T.C., Smith H. Jr., Blackburn B., Silverman B., Schroeder B., Reitman D., Ambroz A. A method for assessing the quality of a randomized control trial. *Control Clinical Trials*, 2(1):31–49, 1981.

Chalmers A., Debattista K., Mastoropoulou G., dos Santos L.: There-reality: selective rendering in high fidelity virtual environments. *International Journal of Virtual Reality*, 6(1):1–10, 2007.

Chang R. S. and Thompson N. S. Whines, cries, and motherese: Their relative power to distract. *Journal of Social Evolutionary and Cultural Psychology*, 5(2):10-20, 2011.

Conrad C., Kobuk Y., Werner P., Cao C. G., Warsaw A., Ratner D., Jones D. B., and Gee D. The effect of defines auditory conditions versus mental loading on the laparoscopic motor skill performance of experts, *Surgical Endoscopy Journal*, 24(6):1347–1352, 2010.

Corti K., Game-based learning: a serious business application, Coventry, WM, UK: PIXE-Learning, 2006.

Dalgarno B., and Lee M. J. W. What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology*, 41(1):10-32, 2010.

Dinh H. Q., Walker N., Kobayashi A., and Hodges L., Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In *Proceedings of IEEE Virtual Reality 1999*, March 13-17, 1999, pp. 222–228.

Facer K. why do we think it's worth talking about computer games and learning in the same breath? NESTA Futurelab discussion paper, at [www.nestafuturelab.org/research/discuss/02discuss01.htm](http://www.nestafuturelab.org/research/discuss/02discuss01.htm), 2006 (accessed June 29, 2012).

FAS (Federation of American Scientists). Harnessing the power of video games for learning. In *Proceedings of Summit on Educational Games*, October 25, 2005.

Feingold C., Calaluce M., and Kallen M. Computerized patient model and simulated clinical experiences: Evaluation with baccalaureate nursing students. *Journal of Nursing Education*, 43(2):156-163, 2004.

Fewerda J., Pattanaik S.N., Shirley P., and Greenberg D.P. A model of Visual Adaptation for Realistic Image Synthesis, In *Proceedings of SIGGRAPH 1996*, August 4-9, 1996, pp. 249-258.

Fisk T. and Hill D. 3D simulation brings immersive dimension to operator training, Dedham, MA, USA: ARC Insisght 2010-06MP, 2010.

Francis P. and Byrne A. Use of role-play exercises in teaching undergraduate astronomy and physics, *Astronomical Society of Australia*, 6(2):203-211, 1999.

Garris R., Ahlers R., Driskell JE. Games, motivation, and learning: a research and practice model. *Simulation and Gaming*, 33(4):441–467, 2002.

Gaston E. T. Music in Therapy, New York, NY. USA: MacMillan Publishing, 1968.

Gee J. P. Good video games + good learning – Collected essays on video games, leaning, and literacy. New York, NY. USA: Peter Lang Publishing, 2007.

Gee J. P. Situated Language and Learning: A critique of traditional schooling. London, UK: Routledge, 2004.

Gee J. P. What video games have to teach us about learning and literacy. Palgrave, New York, USA: MacMillan, 2003.

Gick M. L. and Holyoak K. J. The cognitive basis of knowledge transfer. New York, USA: S. M. Cormier and J. D. Hagman (Eds.), 1987.

Green C., Bavelier D. Action video game modifies visual selective attention. *Nature*, 423(6939): 534–537, 2003.

Greenberg D. P., Torrance K. E., Shirley P., Arvo J., Fewerda J., Pattanaik S.N. Lafortune A.E., Walter B., and Trumbore B. A framework for realistic image synthesis, In *Proceedings of SIGGRAPH 1997 (Special Session)*, 3 - 8 August, 1997, pp. 4777-4945.

Griffiths M. The educational benefits of videogames. *Education and Health*, 20(3):47–51,2002.

Hawkins F. and Orlady H. W. Human factors in flight. Gower, Wales: Brookfield, 1993.

Hays R. T. and Singer M. J. Simulation fidelity in training system design. New York, USA : Springer-Verlag,1989.

Hays R.T. The effectiveness of instructional games: literature review (Technical Report 004) Alexandria, VA: USA Army Research Institute for the Behavioral and Social Sciences, 2005.

Hodgkinson G. The Seduction of Realism. In *Proceedings SIGGRAPH Asia 2009*, December 16–19, 2009, pp. 567-576.

Institute of Medicine. Health professions education – A bridge to quality. Washington, DC, USA: National Academy Press, 2003.

Institute of Medicine. To err is human –Building a safer health care system. Washington, DC, USA: National Academy Press, 2000.

Henn J. S., Lemole G. M., Ferreira M. A. T., Gonzalez L. F., Schornak M., Preul M. C., and Spetzler R. F. Interactive stereoscopic virtual reality: a new tool for neurosurgical education, *Journal of Neurosurgery*, 96(1): 144-149, 2002.

Jiwa S. and Lavelle D. Evaluating the quality of learning through gaming and simulations. In Percival, F, Godfrey, H., Laybourn, P., and Murray, S. (Eds.) *The international simulation and gaming yearbook*, 11(3): 233–242, 2002.

Krentz A. Play and education in Plato’s Republic. In *Proceedings of the 20th World Congress of Philosophy*, August 10–16 1998, pp. 1–7.

Lane N. E. and Alluisi E. A. Fidelity and validity in distributed interactive simulation: Questions and answers (Report No. IDA-D-1066). Alexandria, VA, USA: Institute for Defense Analysis, 1992.

Lathan C. E., Tracey M. R., Sebrechts M. M., Clawson D. M., and Higgins G. A. Using virtual environments as training simulators: Measuring transfer. Mahwah, NJ, USA: Lawrence Erlbaum Associates, 2002.

Lintern G., Roscoe S. N., Koonce J. M., and Segal L. D. Transfer of landing skills in beginning flight training. *Human Factors*, 32(3): 319-327, 1990.

Lombard M. and Ditton T. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2):25-50, 1997.

Lowry P. E. Model GATT: A role-playing simulation course. *Journal of Economic Education*, 30(2):119–126, 1999.

Lubeke D., Reddy M., Watson B., Cohen J. and Varshney A. 2001. Advanced Issues in Level of detail. *From Course #41 at SIGGRAPH 2001*. August 12-17, 2001.

Mack A. and Rock I., Intentional Blindness. Boston, MS, USA: *Massachusetts Institute of Technology Press*, 1998.

Mastoropoulou G. and Chalmers A., The effect of music on the perception of display rate and duration of animated sequences: an experimental study, In *Proceedings of Theory and Practice of Computer Graphics 2004 (TPCG '04)*. June 8-10, 2003, pp. 128–134.

Mastoropoulou G., Debattista K., Chalmers A., and Troscianko T. Auditory bias of visual attention for perceptually-guided selective rendering of animations. In *Proceedings of GRAPHITE 2005, sponsored by ACM SIGGRAPH*. December 1-10, 2005, pp. 96-105.

Mastoropoulou G., Debattista K., Chalmers A., and Troscianko T., The influence of sound effects on the perceived smoothness of rendered animations. In *Proceedings of ACM Symposium on Applied Perception*, August 26-28, 2005, pp. 9–15.

Mastoropoulou G., Chalmers A. The effect of music on the perception of display rate and duration of animated sequences: an experimental study. In *Proceedings of Theory and Practice of Computer Graphics 2004*, June 8-10, 2004, pp. 128–134.

Mayer J. D. and Volanath A. J. Cognitive involvement in the mood response system. *Motivation and Emotion*, 9:261-275, 1985.

McGurk H. and MacDonald J. Hearing lips and seeing voices, *Nature*, 264:746–748, 1976.

McMillan G. R., Cress J. D., and Middendorf M. S. Dynamic seat cuing with wide versus narrow field of view visual displays. In *Proceedings of the AIAA Flight Simulation Technologies Conference*, September 17-19, 1990, pp. 53-62.

McNamara A., Chalmers A.G., Troscianko T., and Gilchrist I. Comparing Real and Synthetic scenes using human Judgments of Lightness. In *Proceedings 12<sup>th</sup> Eurographics Workshop on Rendering*, July 18-21, 2000, pp. 207-219.

Michael D., and Chen S. *Serious games: Games that educate, train and inform*. Boston, MA. USA: Thomson Course Technology, 2006.

Mitchell A., and Smith S., *The use of computer and video games for learning: A review of the literature*. London, UK: LSDA, 2004.

Moroney W. F. and Moroney B. W. Flight simulation. In D. Garlund, J. Wise, and V. D. *Handbook of Aviation Human Factors*. Mahwah, NJ, USA: Lawrence Erlbaum, 1999.

Morsbach G., McCulloch M., Clark A. Infant crying as a potential stressor concerning mothers' concentration ability. *Psychologia: An International Journal of Psychology in the Orient*, 29:18-20, 1986.

Muchinsky P. M. *Persuasive Games – the expressive power of videogames*. Stamford, CT, USA: Wadsworth, 1999.

Myszkowski K., Tawara T., Akamine. Nad Seidel H-P. Perception-Guided Global Illumination Solution for Animation Rendering. In *Proceedings of SIGGRAPH 2001*, August 12-17, 2001, pp. 221–230.

Neral J. and Ray M. Experiential learning in the undergraduate classroom: two exercises. *Economic Inquiry*, 33(2):170–174, 1995.

Oberhofer T. Role-playing in the history of economic thought. *Journal of Economic Education*, 30(2):119–126, 1999.

Pair J., Allen B., Dautricourt M., Treskunov A., Liewer M., Graap K., Reger G., and Rizzo A. A virtual reality exposure therapy application for Iraq war posttraumatic stress disorder. In *Proceedings of IEEE Virtual Reality 2006*, March 25-29, 2006, pp. 235-250.

Prensky M. Digital games-based learning. New York and London, USA and UK: McGraw Hill, 2001.

Prensky M. Don't bother me mom! I'm learning. St. Paul, MN, USA: Paragon House Publishers, 2006.

Ramic B., Chalmers A., Hasic J., and Rizvic S., Selective Rendering in a Multimodal Environment: Scent and Graphics, In *Proceedings Spring Conference on Computer Graphics (SCCG 2007)*, February 13-15, 2007, pp. 189–192.

Reitinger B., Bornik A., Beiche R., and Schmalstieg D.. Liver surgery planning using virtual reality, *IEEE Computer Graphics*, 26(6):36-47, 2006.

Ricci K. E., Salas E., and Cannon-Bowers J. A. Do computer-based games facilitate knowledge acquisition and retention? *Military Psychology*, 8(4): 295-307, 2002.

Rimell S., Howard D., Tyrrell A., Kirk P., and Hunt A., Cymatic: restoring the physical manifestation of digital sound using haptic inter- faces to control a new computer based musical instrument. In *Proceedings of the International Computer Music Conference (ICMC02)*, May 11-14 2002, pp. 125-131.

Rinalducci E. J. Characteristics of visual fidelity in the virtual environment. *Presence: Teleoperators and Virtual Environments*, 5(1): 330-345, 1996.

Schiefele J., Dörr K.-U., Olbert M., Kubbat W. Stereoscopic projection screens and virtual cockpit simulation for pilot training, In *Proceedings of Immersive Projection Technology (IPT) Workshop*, March 13-17, 1999, pp. 45-53.

Shute V., Ventura M., Bauer M., and Zapata-Rivera D., “Melding the power of serious games and embedded assessment to monitor and foster learning,” in *Serious games. Mechanisms and effects*: U. Ritterfeld, M. J. Cody, P. Vorderer (Eds.), 2009, pp. 295-321.

Slater M. and Wilbur S. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6):254-255, 1994.

Squire K. and Jenkins H. Harnessing the power of games in education, *Insight*, 3(1):5–33, 2003.

Stone R. Computer simulations for training and education. Paper presented at the LSRC Seminar *Learning by playing*. June 2003, Available at: [www.bbk.ac.uk/ccs/elearn/Learning by playing/LSRC Paper BOB STONE.doc](http://www.bbk.ac.uk/ccs/elearn/Learning%20by%20playing/LSRC%20Paper%20BOB%20STONE.doc) (accessed June 29, 2012).

Storms R. Auditory–visual cross-modal perception phenomena. PhD thesis, Naval Postgraduate School, Monterey, California, 1998.

Sutcliffe M. Simulations games and role-play. In P. Davies (Ed.) *The handbook for economics lecturers*. September 2002, Available at: [www.economics.ltsn.ac.uk/handbook/games](http://www.economics.ltsn.ac.uk/handbook/games) (accessed June 29, 2012).

Tashiro J. and Dunlap D. The impact of realism on learning engagement in educational games. In *Proceedings of the ACM Future Play 2007 Conference*, November 15-17, 2007, pp. 113–120.

Tashiro J. T., Sullins E. S., and Long G. Virtual clinical excursions for fundamental concepts and skills for nursing. Mosby, St. Louis, USA: Paragon House Publishers, 2010

Tashiro J. T., Sullins E. S., Long G., and Kelly M. Virtual clinical excursions for basic nursing: essentials for practice. Mosby, St. Louis, USA: Paragon House Publishers, 2003.

Taylor L. N. Video games: Perspective, point-of-view, and immersion. Master's Thesis, University of Florida: <http://purl.fcla.edu/fcla/etd/UFE1000166>, 2002 (accessed June 29, 2012).

Tendick F., Downs M., Goktekin T., Cavusoglu M. C., Feygin D., Wu X., Eyal R., Hegarty M., Way L. W. A virtual environment testbed for training laparoscopic surgical skills, *Presence: Teleoperators and Virtual Environments*, 9(3): 236-255, 2000.

Turner S., Turner P., Dawson L., and Munro A. Discovering the impact of reality. in *Proceedings of the Third International Conference on Collaborative Virtual Environments*, September 10-12, 2000, pp. 209-210.

Van der Hulst A., de Hoog R., and Wielemaker J. BOOT: decision support for the selection of facilities for education and training. The Hague, SH, NTH: FEL-99-A188, 1999.

Van Merriënboer J. G., and Jeroen J. Training complex cognitive skills: A four-Component instructional design model for technical training. Educational Technology, NY, USA: Englewood Cliffs, 1997.

Visschedijk G. What is needed in 3D Military Serious Games? Master's Thesis, University of Twente. Soesterberg, April 2010.

Watson B., Friedman A., and McGaffey A. Measuring and predicting visual fidelity. In *Proceedings of SIGGRPAH 2001*, August 12-17, 2001, pp. 213 -220.

Wickens C. D. and Hollands J. G. *Engineering Psychology and Human Performance*. Upper Saddle River, NJ, USA: Prentice Hall, 2000.

Williams J. M. Generalization in the effects of a mood induction procedure. *Behaviour Research and Therapy*, 18:565-572, 1980.

Winkler S. and Faller C. Audiovisual quality evaluation of low bit rate video. *SPIE/IS&T Human Vision and Electronic Imaging*, 5666 (7):139–148, 2005.

Witmer B. and Singer M. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3):225-240, 1998.

Witmer B. and Singer M.. Measuring immersion in virtual environments. Alexandria, VA, USA: United States Army Research Institute for the Behavioral and Social Sciences, 1994.

Woods A. T., Poliakoff E., Lloyd D. M., Kuenzela J., Hodsona J. R., Gondaa H., Batchelora J., Dijksterhuis G. B., and Thomas A. Effect of background noise on food perception, *Food Quality Preference*, 22(1): 42-47, 2011.

Zybura M. and Eskeland A. Olfaction for virtual reality, *Quarter Project Industrial Engineering*, 12(4): 543-550, 1999.