

# The Effect of Sound on Visual Fidelity Perception in Stereoscopic 3-D

David Rojas, Bill Kapralos, *Member, IEEE*, Andrew Hogue, Karen Collins, Lennart Nacke, Sayra Cristancho, Cristina Conati, and Adam Dubrowski

**Abstract**—Visual and auditory cues are important facilitators of user engagement in virtual environments and video games. Prior research supports the notion that our perception of visual fidelity (quality) is influenced by auditory stimuli. Understanding exactly how our perception of visual fidelity changes in the presence of multimodal stimuli can potentially impact the design of virtual environments, thus creating more engaging virtual worlds and scenarios. Stereoscopic 3-D display technology provides the users with additional visual information (depth into and out of the screen plane). There have been relatively few studies that have investigated the impact that auditory stimuli have on our perception of visual fidelity in the presence of stereoscopic 3-D. Building on previous work, we examine the effect of auditory stimuli on our perception of visual fidelity within a stereoscopic 3-D environment.

**Index Terms**—Audio-visual interaction, auditory fidelity, games, multimodal interaction, stereoscopic 3-D (S3-D), virtual environments, visual fidelity, visual quality.

## I. INTRODUCTION

As graphical computing power has increased, so has the desire to achieve higher visual fidelity in virtual environments and video games. However, the focus on increasing visual fidelity has downplayed the role of audio in these environments. Moreover, mainstream consumer-level technology now enables stereoscopic 3-D viewing in the home adding

Manuscript received April 9, 2012; revised December 8, 2012 and May 1, 2013; accepted June 13, 2013. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, the Canadian Network of Centres of Excellence, Graphics, Animation, and New Media (GRAND) Initiative, Ontario Media Development Corporation as part of their Entertainment and Creative Cluster Partnerships Fund, the Ontario Centres of Excellence, and the Social Sciences and Humanities Research Council of Canada. Paper recommended by Associate Editor S. Zafeiriou.

D. Rojas, B. Kapralos, A. Hogue, and L. Nacke are with the Faculty of Business and Information Technology, University of Ontario Institute of Technology, Oshawa L1H 7K4, Canada (e-mail: david.rojas@uoit.ca; bill.kapralos@uoit.ca; andrew.hogue@uoit.ca; lennart.nacke@uoit.ca).

K. Collins is with the Games Institute, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: collinsk@uwaterloo.ca).

S. Cristancho is with the Department of Surgery and Centre for Education Research and Innovation, Schulich School of Medicine and Dentistry, University of Western Ontario, London, ON N6A 3K7, Canada (e-mail: sayra.cristancho@schulich.uwo.ca).

C. Conati is with the Department of Computer Science, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: conati@cs.ubc.ca).

A. Dubrowski is with the Hospital for Sick Children Learning Institute, Toronto, ON M5G 2L3, Canada, and also with the Department of Pediatrics, Faculty of Medicine, and the Wilson Centre, University of Toronto, Toronto, ON M5S 1A1, Canada (e-mail: adam.dubrowski@gmail.com).

Digital Object Identifier 10.1109/TCYB.2013.2269712

to the potential for a high degree of visual fidelity. But, is it necessary to maintain a high degree of visual fidelity to engage players? How do other modalities play a role in user perception of visual fidelity? Understanding the relationships between visual, audio, and haptic modalities with respect to the fidelity of the experience is an under-explored area of research. In this paper, we explore the relationships between our perception of audio-visual fidelity in stereoscopic 3-D virtual environments. The results of this line of research have implications in the design and development of virtual environments, video games, military simulations, and virtual learning environments as they may aid designers in understanding what level of visual fidelity must be maintained in varying contexts to immerse and engage the player in the simulation.

We present the results of three experiments designed and conducted to examine the effect of auditory stimuli on our perception of visual fidelity within a stereoscopic 3-D environment. This paper is part of a larger-scale effort to examine virtual environment fidelity and multimodal interactions. The long-term goal of our work is to answer the questions: 1) “what effects do multimodal interactions have on game players/simulation users?”; and 2) “how much fidelity is actually needed to maximize certain gameplay elements (e.g., enjoyment, engagement, immersion, and knowledge transfer and retention)?” These questions have a number of implications when considering that despite the great computing hardware advances, particularly with respect to graphics rendering, real-time high fidelity audio and visual rendering particularly of complex environments, is still not feasible [1]. Furthermore, striving for such high fidelity environments increases the probability of lag and subsequent discomfort and simulator sickness [2]. Moreover, when some cues are too realistic, negative effects can occur, such as the uncanny valley effect, whereby near-real simulations can lead to feelings of disgust [3]. In addition, striving to reach higher levels of fidelity may lead to increased development costs. Does the effort needed to develop realistic graphics actually benefit the simulation? Are users more or less engaged in the task depending on the level of visual fidelity?

## II. BACKGROUND

### A. Stereoscopic 3-D and Virtual Environments

While stereoscopic 3-D has the potential for increasing engagement, there are many issues that must be properly

addressed when designing stereoscopic 3-D environments. The choice of display technology has an impact on the particular stereoscopic 3-D experience. For example, the size of the display will determine how far objects can appear to protrude out of the screen since the depth is proportional to the pixel disparity. The choice of stereoscopic 3-D technology for the display will also impact the user experience. For example, although autostereoscopic displays do not require users to wear glasses, there is a significant tradeoff between the number of views to generate the stereoscopic 3-D effect and the resolution of the screen. Active stereo requires the user to wear expensive LCD shutter glasses synchronized with the display refresh rate. Passive systems, on the other hand, require the user to wear inexpensive glasses, but are subject to higher amounts of ghosting. Ghosting or crosstalk (the user can see a remnant of the left eye image in their right eye or vice versa) can lead to detrimental effects on our perception of depth in these environments [4].

More specific to video games, users can comfortably handle only a relatively small range of disparity in a given scene precluding the use of large vistas or landscapes [5]. Full-screen special effects such as depth of field that are routinely used to force the user's gaze to a particular region of the screen do not work well in stereoscopic 3-D, thereby reducing the quality of the experience. These design issues must be taken into account when designing stereoscopic 3-D games. It is not sufficient to assume that a game designed for 2-D is equally effective and engaging in stereoscopic 3-D.

Although the use of stereoscopic 3-D within commercial video games is relatively new (mostly due to the availability of 3-D technology in the home), stereoscopic 3-D has been employed extensively in virtual reality-based training simulations, most notably pilot training. With respect to pilot training, the depth perception stereoscopic 3-D provides allows pilot trainees to assess complex 3-D structures and other aircraft positions and provides the ability to enhance group experience and group training, while motivating the trainee to more deeply explore the learning material [6]. In addition to pilot training, stereoscopic 3-D has been applied to virtual environments developed for surgical training [7]–[9], and plant operator training [10], among other applications. Stereoscopic 3-D was used in the science and math in an immersive learning environment (SMILE), a learning game that employs a fantasy 3-D virtual environment to engage deaf and hearing impaired children in math and science-based educational tasks [11]. In contrast to many of the existing virtual learning environments, the design of SMILE focused on enhanced user interaction techniques and game-play to ensure a motivating and appealing experience. Usability studies confirmed that SMILE is both enjoyable and easy to use [11].

### B. Real-Time Rendering of Audio-Visual Environments

Motivating the exploration of cross-modal interactions is the tradeoff between the (perceived) needs of the virtual environment/game designers, and the computational and algorithmic needs to achieve a high level of graphical fidelity. It is still common for designers to prioritize graphical quality over audio quality both in terms of design effort and computational time

allotted to the processes. This reduces the audio quality as it must be down-mixed from surround sound to stereo or compressed in other ways.

Auditory fidelity can refer to sample rate, bit rate, the use of various compression algorithms, and the number of channels involved (e.g., mono versus stereo). Each of these can be reduced to minimize the overall power required enabling real-time simulation performance. Graphical fidelity commonly refers to polygon count (the number of triangles that comprise a surface), or texture resolution (excellent overviews are provided by Provost [12], Guo *et al.*, [13], and Rubino and Power [14]). Polygon count was studied in relation to visual cognition by Mourkoussis *et al.* [15] and with respect to memory, they found that as long as the scene was relatively realistic, the level of visual realism had no effects on high-level cognition.

Stereoscopic 3-D requires the system to render (draw) the scene twice (once for each eye), thus doubling the computational requirements of the graphical system. Given that the graphical scene is typically prioritized over sound processing, some compromises on audio versus visual fidelity must take place.

While the increasing processor power of modern computers and game consoles has alleviated some of the necessary tradeoffs between audio and visual fidelity, the move to streamed online content and smartphone games has reinvigorated the need to address these problems [16], [17]. The decision to reduce either auditory or visual fidelity is made by the game's developers, usually at the expense of audio. But whether the reduction of audio fidelity is necessary needs to be tested against our perception of fidelity in multimodal environments. Perceptual studies into cross-modal interactions have shown that it is possible to trick our brains into hearing sounds differently when paired with visuals [18]. How can we leverage the psycho-acoustic and cross-modal interactions to alleviate some of the difficulties of computational load? Which tradeoffs can be made remains unanswered, however, until we have a greater understanding of cross-modal perception in real-time virtual environments.

### C. Sound and Multimodal Interactions

The issues with stereoscopic 3-D in virtual environments, games, and other media can potentially be alleviated by exploiting cross-modal sensory effects. Cross-modal effects refer to the impact on the perceptual experience of one sensory input that the presentation of an additional sensory input can have [19]. Previous work has demonstrated that cross-modal effects can be considerable to the extent that large amounts of detail of one sense may be ignored in the presence of other sensory inputs (see [18] and [20], and for a more recent summary see also [21]).

Various studies have examined the interaction of sound with other sensory cues [22], [23]. Of interest to our study here are specifically audio-visual cross-modal effects. It has been shown that sound can potentially attract part of a user's attention away from the visual stimuli and lead to a reduced cognitive processing of the visual cues [24]. For most audio-visual events that are short in duration, we tend to respond to the visual stimulus [25]. The ventriloquist effect is a

well-known and studied effect where visual and auditory cues are perceived to be related to a single event when, in fact, they are spatially disjoint [26]. However, the auditory-induced flash [27] contradicts the assertion that sound is always driven by visuals.

Mastoropoulou *et al.* [24] examined the influence of sound effects on the perception of motion smoothness within an animation and, more specifically, on the perception of frame-rate. Their study involved 40 participants that viewed pairs of computer-generated walkthrough animations at five different frame rates. The visuals were consistent across the animation pairs although the pairs differed with respect to sound; one contained sound effects while the other did not (it was silent). The participants' task was to choose which animation had a smoother motion. There was a significant effect of sound on perceived smoothness and sound attracted a viewer's attention from the visuals, leading to a greater difficulty in distinguishing smoothness variations between animations containing sound cues displayed at different rates, than between silent animations [24]. Mastoropoulou *et al.* [24] infer that sound stimuli attract part of the viewer's attention away from any visual defects inherent in low frame rates. Hulusic *et al.* [28] examined audio-visual interactions for the purpose of reducing computational requirements of visual rendering with the use of motion-related sound effects. They observed that motion-related sound effects allowed slow animations to be perceived as smoother than fast animations and that the addition of footstep sound effects to walking (visual) animations increased the perception of animation smoothness. Hulusic *et al.* [28] conclude that for certain conditions, sound can be used to reduce the graphical rendering rate without the viewer being aware of this reduction. In a more recent study, motivated by the ability of sound (music) to create an emotional involvement with the simulation, Hulusic *et al.* [29] examined the effect of a sound's beat rate on the perception of visual frame rate. Their results indicate that a low beat rate sound could reduce the visual frame rate of a static scene without any reduction of perceived quality. The same effect was observed with a dynamic scene albeit without statistical significance [29].

Although sound's impact on stereoscopic 3-D remains underexplored, some research suggests that auditory depth cues can significantly impact visual depth perception cues [30]. A further study by our own research team explored the impact of stereo and surround sound on the perception of the visual field in stereoscopic 3-D games [31]. This series of experiments examined the impact of various sound conditions on the perception of depth, target acquisition, timing, and edge of screen discrepancies. Preliminary results suggest that sound can play a significant role in each of these factors. More specifically, spatial sound cues had an influence on depth perception, whereby 5.1 surround sound had an advantage over traditional stereo sound in increasing the users' sense of depth.

#### *D. Visual Fidelity Perception Under Different Auditory Conditions*

The perception of visual fidelity can affect the perception of sound fidelity and vice versa [32]. This perceptual phenomenon has important implications for designers of multi-

modal virtual simulations and games. As described by Larsson *et al.* [25], if the possibilities of enhancing the visuals within a virtual environment are economically or technically limited, one may consider increasing the quality of the audio channels instead. In this way, the illusion of increased fidelity can be maintained by improving auditory fidelity. Such cross-modal exploitation can be particularly important in gaming environments, where memory and processor abilities can be strained by real-time processing requirements.

Our previous work examined visual fidelity perception in the presence of various (background) sound conditions [33], [34]. The visuals consisted of a single 2-D image of a surgeon's head (a rendered 3-D model). In the first study, visual fidelity was defined with respect to the 3-D model polygon count while in the second study, polygon count was kept constant and visual fidelity was defined with respect to the 3-D model's texture resolution. In both studies, participants were presented with the static visual (a total of six visuals were considered in each study) in conjunction with one of four auditory conditions: 1) no sound at all (silence); 2) white noise; 3) classical music (Mozart); and 4) heavy metal music (Megadeth). For each of the visuals, the participant's task was to rate its fidelity on a scale from 1 to 7. The most important findings were as follows: 1) with respect to polygon count, visual fidelity perception increased in the presence of classical music, particularly when considering images corresponding to higher polygon count [33], and 2) when considering texture resolution, background sound consisting of white noise had very specific and detrimental effects on the perception of the quality of high-resolution images (i.e., the perception of visual fidelity of high fidelity visuals decreased in the presence of white noise) [34]. In contrast to the study that considered polygon count, background sound consisting of music (classical or heavy metal) did not have any effect on the perception of visual fidelity when visual fidelity was defined with respect to texture resolution.

However, these two studies used sound of a de-contextualized nature; that is, the sounds were not specific to the context of the surgeon presented in the visuals. To explore the impact of contextualization, we are currently exploring the influence of contextual sound cues on visual fidelity perception (with respect to texture resolution and polygon count) in a monoscopic viewing environment. Previous work has focused on monoscopic graphics. The purpose of our present study, therefore, was to explore whether or not the same results will occur for stereoscopic 3-D images.

### III. METHODS AND MATERIALS

#### *A. Hypotheses*

We anticipate results similar to our previous work and more specifically, that auditory conditions can influence the perception of visual fidelity within a stereoscopic 3-D viewing environment. Therefore, we assume the following null hypothesis.  $H_0$ : Auditory conditions will not influence stereoscopic 3-D visual fidelity (defined with respect to polygon count, or texture resolution) perception.

We also hypothesize the following.

**H<sub>1</sub>**: As with monoscopic viewing, stereoscopic 3-D visual fidelity perception level defined by polygon count will increase through sound conditions, specifically classical music.

**H<sub>2</sub>**: As with monoscopic viewing, stereoscopic 3-D visual fidelity perception level defined by texture resolution will be decreased by sound conditions, particularly white noise.

**H<sub>3</sub>**: Contextual auditory cues (i.e., sounds that are related to the visual environment) will increase visual fidelity perception in comparison to noncontextual auditory cues.

### B. Auditory Stimuli

For all of the experiments, the sound pressure level of the auditory conditions (independent variable) was 63 dB (about the same level as typical conversation [35]) measured with a sound level meter (A-weighting). All auditory stimuli were monophonic and were output with a pair of Sony MDR 110LP headphones.

### C. Visual Stimuli

In each of the three experiments, visual stimuli (dependent variable) consisted of a static image presented to the participants in stereoscopic 3-D on an Acer Aspire laptop with a 15.6-inch screen size and a resolution of  $1366 \times 768$ . The size of each image within the display was  $800 \times 630$ . The Acer Aspire laptop employs NVIDIA's 3-D Vision technology (active stereoscopic system with wireless glasses). The stereoscopic 3-D depth setting was set to default where, on a scale of 0 to 15 (where 0 represents no stereoscopic 3-D effect and 15 represents the maximum stereoscopic 3-D effect); the default setting is approximately 15% of the maximum stereo separation. Depth (stereo separation) is the only modifiable stereoscopic-related parameter within the NVIDIA stereoscopic 3-D system; other stereoscopic-related parameters cannot be modified. When displayed, the image remained static and the participants were not able to interact with the image in any manner). However, participants were permitted (and encouraged) to move their heads (side to side, front and back) to find the sweet spot for the screen.

### D. Experimental Method

For each of the three experiments, participants were seated in front of the laptop computer that 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. Prior to the first trial of each experiment, participants were provided with two representative images (the lowest quality image followed by the highest quality image) to allow them to make meaningful comparisons during each trial.

For each experiment, each trial involved presenting the participants with one image and one sound combination. Their task was to rank the image presented to them with respect to their perceived visual fidelity on a scale from 1 (lowest perceived fidelity) to 7 (highest perceived fidelity). Figs. 1, 2, and 5 illustrate the images used in Experiments 1, 2, and 3, respectively. For Experiments 1 and 2, one of the six visual stimuli (images) was presented in conjunction with one of four

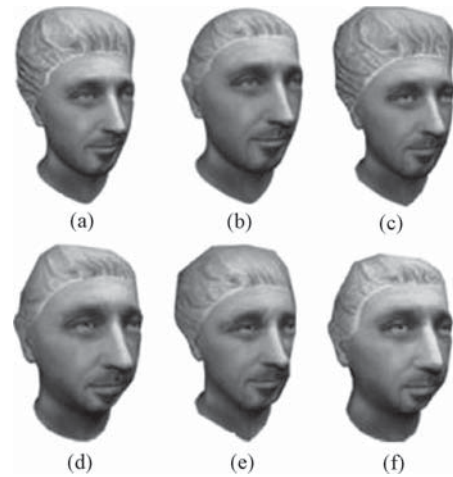


Fig. 1. Visual stimulus used in Experiment 1 consisting of a 3-D model of a surgeon's head, each varying with respect to polygon count. The polygon count (triangles) of each of the six models was defined as follows: (a) 17 440, (b) 13 440, (c) 1250, (d) 865, (e) 678, and (f) 548.

auditory conditions. For Experiment 3, one of the six visual stimuli (images) was presented in conjunction with one of seven auditory conditions.

In Experiments 1 and 2, for each participant, each of the six images and each of the four background sound combinations (24 combinations in total) was repeated three times for a total of 72 trials, presented in a randomized ordering. In Experiment 3, for each participant, 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 128 trials, presented in a randomized ordering. All three experiments followed a "within participants" whereby in each experiment, each participant was tested under all of the conditions of that particular experiment. For each experiment, the results from the three attempts for each condition were averaged and a between group analyses of variance (ANOVAs) were used to compare the results across the groups.

### E. Participants

Participants consisted of volunteer students and researchers from the University of Ontario Institute of Technology (UOIT). A total of 18 volunteers (13 females and 5 males) participated in Experiment 1 (average age was 20 years), 18 volunteers (17 females and 1 male) participated in Experiment 2 (average age was 20 years), and 18 volunteers (ten females and eight males) participated in Experiment 3 (average age was 22 years). Volunteers participated in one experiment only (i.e., Experiments 1, 2, or 3) and each of the three experiments was conducted separately. After a posthoc power analysis, it was determined that based on the mean, and between-groups comparison effect, a sample size of 15 was needed to obtain statistical power at (the recommended)  $d=0.95$  level [34] for each of the three experiments. None of the participants reported any hearing or visual defects. The authors did not participate in the experiments and the experiments abided by the University of Ontario Institute of Technology Research Ethics Review process.

TABLE I  
EXPERIMENT 1 RESULTS: MEAN AND STANDARD DEVIATION

Visual Fidelity (Polygon Count)	Auditory Condition	Mean	Std. Dev.
17,440	No Sound	4.44	0.32
	White Noise	2.67	0.32
	Classical	6.57	0.24
	Heavy Metal	6.00	0.32
13,440	No Sound	5.00	0.32
	White Noise	2.67	0.32
	Classical	6.43	0.44
	Heavy Metal	5.67	0.32
1,250	No Sound	5.33	0.32
	White Noise	1.71	0.29
	Classical	5.90	0.36
	Heavy Metal	5.62	0.29
865	No Sound	2.67	0.32
	White Noise	1.71	0.29
	Classical	4.29	0.29
	Heavy Metal	4.62	0.29
678	No Sound	2.67	0.32
	White Noise	1.76	0.24
	Classical	4.00	0.32
	Heavy Metal	3.90	0.31
548	No Sound	2.05	0.29
	White Noise	1.19	0.17
	Classical	4.33	0.32
	Heavy Metal	3.67	0.32

#### IV. EXPERIMENT ONE

In Experiment 1, we replicated our previous monoscopic study [33], but altered the image such that it was presented in stereoscopic 3-D. The visual stimuli consisted of six images of a single surgeon’s head against a white background (Fig. 1). Visual fidelity (quality) here was defined with respect to polygon count (the number of triangles that comprise a surface); everything else remained the same (texture resolution, etc.) while the number of polygons comprising the model of the surgeon’s head varied. The polygon counts were as follows: 1) 17 440; 2) 13 440; 3) 1250; 4) 865; 5) 678; and 6) 548. Through informal testing, it was determined that a polygon count of 17 400 resulted in a very smooth structure for the model of the surgeon’s face and increasing beyond this had no effect while lowering the polygon count below 548 resulted in various artifacts. Furthermore, the polygon counts considered here are representative of the range of polygon counts typically used in current games across all platforms. More specifically, with our currently available technology, for mobile devices, common polygon counts range from between 300 to 1500, while for desktops and consoles, polygon counts typically don’t exceed 4000 (as technology improves, undoubtedly polygon counts will also increase). Various methods and techniques are also available to emulate the proportions of high polygon count models. For example, through the use of a normal map, a model with 2 to 3 million polygons can be emulated with a model that contains 2500 polygons [37].

Four auditory conditions were examined: 1) no sound at all; 2) white noise; 3) classical music (Mozart); and 4) heavy metal (Megadeth). 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 cutoffs of

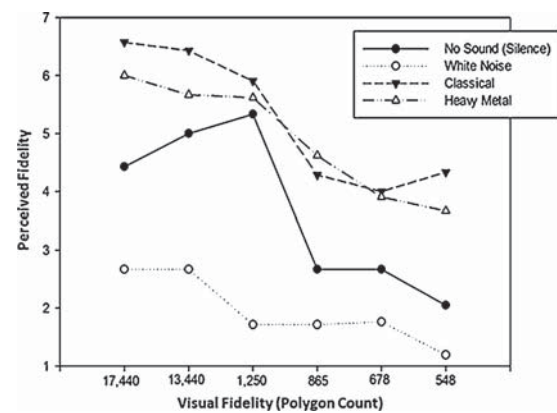


Fig. 2. Results for Experiment 1. Polygon count (x-axis) versus perceived visual fidelity (y-axis) across each of the four auditory conditions considered.

200 Hz and 10 kHz, respectively. The level of each of the four sounds was normalized (using the “normalize multiple audio tracks” option of the Audacity audio editor software) to ensure that all three sounds (tracks) had the same peak level. Each of these four auditory conditions was noncontextual (not-related) with respect to the visual scene presented to the participants.

A summary of the results is presented in Table I and Fig. 2 where the x-axis represents polygon count and the y-axis represents perceived visual fidelity (ranging from 1 to 7). 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 head, varying with respect to the number of polygons used to represent the model)  $\times$  4 auditory conditions: 1) no sound at all; 2) white noise; 3) classical music; and 4) heavy metal music. The data was tested for normality using the Shapiro–Wilks test and showed a normal distribution ( $SW=0.9716$ ,  $df=6$ ,  $p=0.4513$ ).

The results of the two way ANOVA (sound by visual (image)) revealed significant main effects for sound ( $F=23.2$ ,  $p=0.001$ ) and image ( $F=35.84$ ,  $p=0.001$ ) in addition to the interaction between these two terms ( $F=2.13$ ,  $p=0.008$ ). Since we were interested in the effect of sound on the perception of visual fidelity, 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 of results for the model with a polygon count of 17 440 revealed significant results ( $F(5)=19.562$ ,  $p=0.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the white noise auditory condition influenced the perception of these visuals ( $p < 0.001$ ); in contrast to the other three auditory conditions, the perception of visual fidelity decreased in the presence of white noise, which did not differ from each other ( $p < 0.05$  for all). The analysis for the model with a polygon count of 13 440 revealed significant results ( $F(5)=9.3$ ,  $p < 0.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the white noise auditory condition led to a decrease in the perception of visual fidelity compared to other three auditory conditions ( $p < 0.001$ ), all three of which did not differ from each other ( $p < 0.05$  for all). The analysis for the model with a polygon count of 1250 significant results ( $F(5)=8.752$ ,  $p < 0.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the classical music auditory

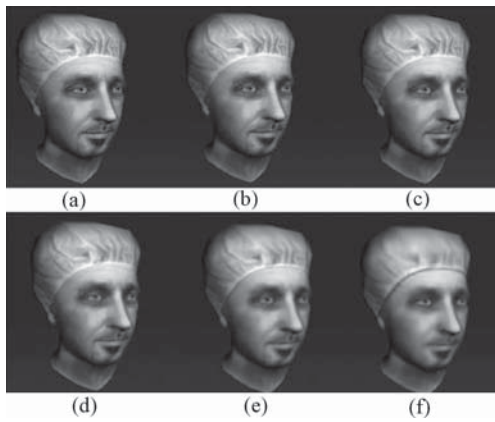


Fig. 3. Visual stimulus used in Experiment 2 consisting of a 3-D model of a surgeon's head, comprised of 17 440 polygons. The texture resolution of each of the six renderings were defined as follows: (a)  $1024 \times 1024$ , (b)  $512 \times 512$ , (c)  $256 \times 256$ , (d)  $128 \times 128$ , (e)  $64 \times 64$ , and (f)  $32 \times 32$ .

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 ( $p < 0.05$  for all). The analysis for the model with a polygon count of 865 showed significant results ( $F(5)=7.912$ ,  $p < 0.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the classical music auditory condition influenced the perception of visual fidelity causing it to be higher in quality than the other sounds ( $p < 0.001$ ), which did not differ from each other ( $p < 0.05$  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  $F(5)=0.372$ ,  $p=0.773$ , respectively) suggesting that sound did not have any influence on the perception of visual fidelity for the two models with the lowest polygon count.

## V. EXPERIMENT TWO

In Experiment 2, we replicated the study of Rojas *et al.* [34], but altered the image such that it was presented in stereoscopic 3-D. The visual stimuli consisted of six images of a single surgeon's head (the same model as in Experiment 1) against a black background (Fig. 3). The 3-D model of the surgeon's head was comprised of 17440 polygons but for each of the six images, as shown in Fig. 2, fidelity was varied with respect to texture resolution only while all other parameters, including polygon count and image size, remained the same. The texture resolutions were  $1024 \times 1024$ ,  $512 \times 512$ ,  $256 \times 256$ ,  $128 \times 128$ ,  $64 \times 64$ , and  $32 \times 32$ . The range of texture resolutions considered here is typical of the texture resolutions found in current games; despite the fact that some of the hardware/gaming consoles can support larger resolutions (e.g., the Microsoft Xbox supports textures up to  $8192 \times 8192$ ), texture resolution is generally limited to  $512 \times 512$  due to performance constraints [38]. The auditory stimuli were the same as in Experiment 1.

A summary of the results is presented in Table II and Fig. 4, where the  $x$ -axis represents texture resolution and the  $y$ -axis represents perceived visual fidelity (ranging from

TABLE II

EXPERIMENT 2 RESULTS: MEAN AND STANDARD DEVIATION

Visual Fidelity (Texture Resolution)	Auditory Condition	Mean	Std. Dev.
$1024 \times 1024$	No Sound	5.06	0.83
	White Noise	3.74	0.37
	Classical	5.65	0.82
	Heavy Metal	4.89	0.92
$512 \times 512$	No Sound	4.58	1.58
	White Noise	3.62	0.38
	Classical	5.39	0.82
	Heavy Metal	4.85	1.00
$256 \times 256$	No Sound	4.00	1.36
	White Noise	3.59	0.55
	Classical	5.39	0.79
	Heavy Metal	3.87	1.59
$128 \times 128$	No Sound	3.98	1.35
	White Noise	3.52	0.46
	Classical	5.28	0.94
	Heavy Metal	4.06	1.50
$64 \times 64$	No Sound	3.67	1.41
	White Noise	3.46	0.49
	Classical	3.91	1.26
	Heavy Metal	4.37	1.54
$32 \times 32$	No Sound	2.44	1.78
	White Noise	2.50	0.38
	Classical	2.78	1.85
	Heavy Metal	2.22	1.83

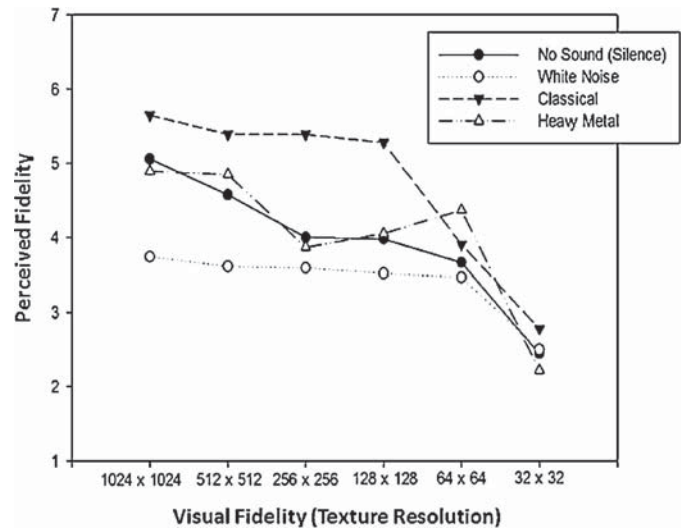


Fig. 4. Experimental results for Experiment 2. Texture resolution ( $x$ -axis) versus perceived visual fidelity ( $y$ -axis) across each of the four auditory conditions considered.

1 to 7). A visual inspection of Fig. 5 clearly indicates that there is generally a decreasing trend in perceived quality with decreasing texture resolution although when considering the white noise background sound condition, this does not hold for the two images with a resolution of  $128 \times 128$  and  $64 \times 64$  where there is a rise in perceived quality. The data was tested for normality using the Shapiro-Wilks test and showed a normal distribution ( $SW=0.843$ ,  $df=6$ ,  $p=0.321$ ).

The results of the two-way ANOVA [sound by visual (image)] revealed significant main effects for sound ( $F=21.2$ ,  $p < 0.001$ ) and image ( $F=9.6$ ,  $p < 0.001$ ) in addition to the

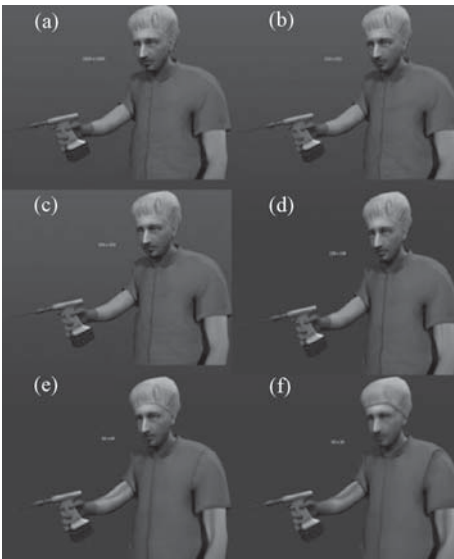


Fig. 5. Visual stimulus used in Experiment 3 consisting of a 3-D model of a surgeon holding a drill. The 3-D model of the surgeon's head was comprised of 13 656 triangles, the drill consisted of 4586 triangles, and the drill bit consisted of 616 triangles. The model (surgeon's head and drill) was rendered with six different texture resolutions ranging from  $1024 \times 1024$  to  $32 \times 32$  as shown in (a)–(f), respectively.

interaction between these two terms ( $F=5.18$ ,  $p < 0.008$ ). Since we were interested in the effect of the sound on the perception of visual fidelity, the two-way interaction was further analyzed by six subsequent one-way ANOVAs with 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 ( $F(5)=4.823$ ,  $p=0.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that all sound conditions influenced the perception of visual fidelity.

The classical music auditory condition ( $p < 0.001$ ) showed to increase the perceived visual fidelity when compared to the other conditions ( $p < 0.05$  for all). The heavy metal music condition led to an increase in the perception of visual fidelity when compared to white noise and no sound auditory conditions ( $p < 0.05$  for both), but led to a decrease in the perceived visual fidelity when compared to the classical music condition ( $p < 0.001$ ), implying that despite the increase in visual fidelity perception associated with the heavy metal music this increase is smaller than the classical music condition. The white sound auditory condition led to a decrease in perceived visual fidelity ( $p < 0.001$ ) when compared to the other conditions ( $p < 0.05$  for all).

The analysis of results for the model with a texture resolution of  $512 \times 512$  showed significant results ( $F(5)=29.3$ ,  $p < 0.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that all four auditory conditions influenced the perception of visual fidelity compared to the other sounds ( $p < 0.05$  for all). The classical music and heavy metal music auditory conditions led to an increase in perceived visual fidelity ( $p < 0.001$ ) when compared to the other sounds ( $p < 0.05$  for all). In contrast, the white noise auditory condition led to a decrease in the perceived visual fidelity ( $p < 0.001$ ) when compared to the other conditions ( $p < 0.05$  for all).

The analysis for the model with a texture resolution of  $256 \times 256$  showed significant results ( $F(5)=5.43$ ,  $p < 0.0001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the white noise auditory condition led to a decrease in visual fidelity perception compared to the other auditory conditions ( $p < 0.001$ ), which did not differ from each other ( $p < 0.05$  for all).

The analysis for the model with a texture resolution of  $128 \times 128$  showed significant results ( $F(5)=16.52$ ,  $p < 0.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that all four auditory conditions influenced the perception of visual fidelity ( $p < 0.001$ ), which did not differ from each other ( $p < 0.05$  for all). The classical music auditory condition led to an increase in perceived visual fidelity ( $p < 0.001$ ). The heavy metal music auditory condition also led to an increase in perceived visual fidelity ( $p < 0.022$ ) when compared to the other sounds while the white noise auditory condition led to a decrease in the perceived visual fidelity ( $p < 0.001$ ).

The analysis for the model with a texture resolution of  $64 \times 64$  showed significant results ( $F(5)=17.983$ ,  $p < 0.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that the white noise auditory condition led to a decrease in the perception of visual fidelity compared to the other auditory conditions ( $p < 0.001$ ), which did not differ from each other ( $p < 0.05$  for all).

The analysis for the model with a texture resolution of  $32 \times 32$  showed significant results ( $F(5)=29.3$ ,  $p < 0.001$ ). Subsequent post-hoc comparisons (Tukey HSD) revealed that all four auditory conditions influenced the perception of visual fidelity compared to the other auditory conditions ( $p < 0.05$  for all). The classical music condition led to an increase in perceived visual fidelity ( $p < 0.001$ ). The heavy metal music auditory condition also led to an increase in perceived visual fidelity ( $p < 0.022$ ) when compared to the other auditory conditions. The white noise auditory condition led to a decrease in perceived visual fidelity ( $p < 0.001$ ).

In summary, there was a significant difference between the classical music auditory condition (which resulted in the greatest perceived quality) and the other three auditory conditions. There was also a significant difference between the white noise and the other three auditory conditions, where the white noise led to a decrease in visual fidelity perception.

## VI. EXPERIMENT THREE

In Experiment 3, we replicated the study of Rojas *et al.* (in preparation), again altering a monoscopic image to stereoscopic 3-D. The visual stimuli for Experiment 3 consisted of six images of a single surgeon (upper body only) holding a surgical drill, against a black background (Fig. 5). The 3-D model of the surgeon's head and torso was comprised of 13 656 triangles, the drill consisted of 4586 triangles, and the drill bit consisted of 616 triangles. For each of the six images, fidelity was varied with respect to texture resolution only (both the surgeon and the drill), while all of the other parameters, including polygon count and image size, remained the same (Fig. 5).

In addition to the four noncontextual cues of the first two experiments, another three cues were considered. These three additional auditory cues provided contextual information

TABLE III  
EXPERIMENT 3 RESULTS: MEAN AND STANDARD DEVIATION

Visual Fidelity (Texture Resolution)	Auditory Condition	Mean	Std. Dev.
1024 × 1024	No Sound	5.33	0.38
	White Noise	3.58	0.57
	Classical	5.50	0.96
	Heavy Metal	5.50	0.43
	OR Ambiance	6.42	0.42
	OR Ambiance + Drill	5.17	0.43
	Drill	4.50	1.00
512 × 512	No Sound	5.08	1.17
	White Noise	4.08	0.69
	Classical	5.33	0.27
	Heavy Metal	5.08	1.37
	OR Ambiance	6.08	0.63
	OR Ambiance + Drill	5.00	0.47
	Drill	4.50	1.04
256 × 256	No Sound	5.75	0.96
	White Noise	3.25	0.57
	Classical	5.08	0.69
	Heavy Metal	5.33	1.05
	OR Ambiance	6.25	0.74
	OR Ambiance + Drill	5.08	0.88
	Drill	4.33	0.72
128 × 128	No Sound	5.17	1.17
	White Noise	3.25	0.57
	Classical	4.42	0.96
	Heavy Metal	4.67	0.77
	OR Ambiance	5.33	1.24
	OR Ambiance + Drill	5.08	0.92
	Drill	4.42	0.57
64 × 64	No Sound	4.00	0.82
	White Noise	2.67	0.94
	Classical	3.42	1.23
	Heavy Metal	4.42	0.96
	OR Ambiance	3.25	0.69
	OR Ambiance + Drill	4.00	0.72
	Drill	3.17	1.00
32 × 32	No Sound	2.17	1.91
	White Noise	1.67	1.12
	Classical	4.00	1.28
	Heavy Metal	2.75	2.13
	OR Ambiance	3.00	1.33
	OR Ambiance + Drill	2.67	1.56
	Drill	2.58	1.64

with respect to the visual scene presented to the participants: 1) operating room ambiance sound; 2) surgical drill sound; and 3) operating room ambiance mixed with a surgical drill sound (operating room ambiance + drill sound). The operating room ambiance sound included machines beeping, doctors and nurses talking, and was purchased from AudioSparx.com. The operating room ambiance with the drill sound was made by mixing the operating room ambient sound with a recording of an actual drill sound. The level of each of the seven sounds used in Experiment 3 was normalized to ensure that all sounds (tracks) had the same peak level.

A summary of the results is presented in Table III and Fig. 6 where the  $x$ -axis represents texture resolution and the  $y$ -axis represents perceived fidelity (ranging from 1 to 7). The data was tested for normality using the data was tested for normality using the Shapiro–Wilks test and showed a normal distribution ( $SW = 0.925$ ,  $df = 6$ ,  $p = 0.486$ ).

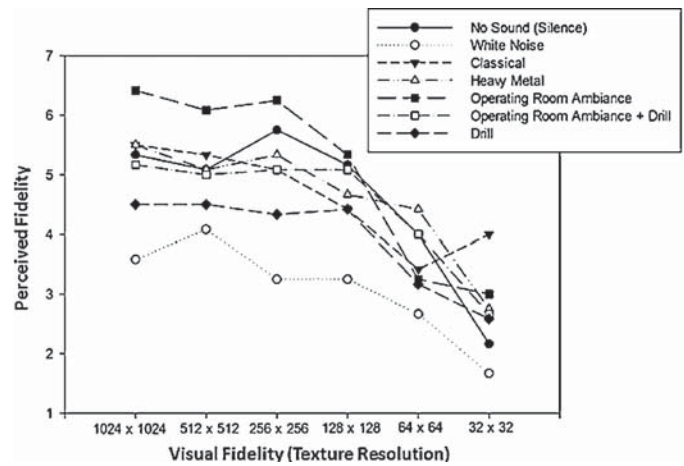


Fig. 6. Experimental results for Experiment 3. Texture resolution ( $x$ -axis) versus perceived visual fidelity ( $y$ -axis) across each of the seven (contextual and noncontextual) auditory conditions considered.

The analysis of variance (ANOVA) was selected as the statistical model with two factors: 6 visual conditions (model of a surgeon's head, visual varying with respect to texture resolution)  $\times$  7 auditory conditions: 1) no sound at all; 2) white noise; 3) classical music; 4) heavy metal music; 5) operating room ambiance sound; 6) surgical drill sound; and 7) operating room ambiance mixed with the surgical drill sound). Main effects and interactions were further analyzed using post-hoc comparisons (Tukey HSD). The results revealed that only the main effects for visual ( $F = 9.83$ ,  $p < 0.001$ ) and auditory ( $F = 27.19$ ,  $p < 0.001$ ) conditions were significant, but there was no significant interaction between the sound and image ( $F = 7.14$ ,  $p < 0.761$ ). Given the lack of interaction between the factors, a single ANOVA for each condition was not required and therefore not conducted. More specifically, on average the participants were relatively accurate at discriminating the quality of the visual images, where the low were judged as lower and high as higher. Overall, these percepts of visual fidelity of the images were not affected by the presence of background sound, with the exception of white noise, which causes a significant degradation in the perception of visual fidelity across all of the images.

## VII. DISCUSSION OF STEREOSCOPIC 3-D AND MONOSCOPIC COMPARISON RESULTS

In Experiment 1, the classical music auditory condition led to an increase in visual fidelity perception while the white noise auditory condition had an attenuating effect on the perception of the visual fidelity; both of these effects were evident for only the visual models whose polygon count was greater than 678 (i.e., auditory condition had no effect on the two smallest polygon count models). We could assume that 678 is a polygon count threshold after which the visual distinction is not great enough to be negatively influenced by white noise. In addition, the heavy metal music auditory condition did not have any effect on the perception of visual fidelity. The results revealed that white noise led to a reduction of visual fidelity perception while classical music led to an increase in



visual fidelity perception. It would be interesting to follow up with a study investigating whether this increased perception is due to increased mental processing that could be triggered by classical music. These effects supported our previous studies relating to monoscopic cues, whereby auditory cues aside from white noise led to an increase in visual fidelity perception while white noise led to a decrease in the perception of visual fidelity [33].

We conducted a three way ANOVA to compare the results of Experiment 1 and the results from our previous corresponding monoscopic-based experiment. The results of the ANOVA (sound by visual (image) by experiment), revealed significant main effects for sound ( $F=15.2$ ,  $p<0.001$ ), image ( $F=13.5$ ,  $p<0.001$ ), and experiment ( $F=8.1$ ,  $p<0.001$ ) in addition to the interaction between image and experiment ( $F=7.3$ ,  $p<0.001$ ), image and sound ( $F=5.6$ ,  $p<0.001$ ), and sound and experiment ( $F=6.8$ ,  $p<0.001$ ). The significant interactions were further analyzed by evaluating the main effect of image from our previous corresponding (monoscopic) experiment and Experiment 1 and revealed that Experiment 1 (stereoscopic 3-D) had a greater effect on visual fidelity perception for the visual conditions containing a polygon count of 1250 or higher. We further analyzed the interaction between sound and experiment and the results revealed that sound had a greater effect on visual fidelity perception when the visuals were presented in stereoscopic 3-D (Experiment 1) as opposed to monoscopic.

In Experiment 2, the classical music and heavy metal music auditory conditions led to an increase in visual fidelity perception. In contrast, the white noise auditory condition led to a decrease in visual fidelity perception. When compared to the previous monoscopic experiments, the results showed the same trend [33]. However, there was a difference with respect to the manner that sound affected visual fidelity perception.

We conducted another three-way ANOVA to compare the results of Experiment 2 and the results from our previous corresponding monoscopic-based experiment. The results of the ANOVA (sound by visual (image) by experiment) revealed significant main effects for sound ( $F=17.5$ ,  $p<0.001$ ), image ( $F=13.5$ ,  $p<0.001$ ), and experiment ( $F=12.3$ ,  $p<0.001$ ), in addition to the interaction between image and experiment ( $F=5.4$ ,  $p<0.001$ ), image and sound ( $F=9.7$ ,  $p<0.001$ ), and sound and experiment ( $F=8.4$ ,  $p<0.001$ ). The significant interactions were further analyzed by evaluating the main effect of images from our previous monoscopic experiment and Experiment 2 and revealed that our previous corresponding (monoscopic) experiment had a greater effect on visual fidelity perception particularly when considering the visual conditions with a texture resolution of  $128 \times 128$  or higher. We further analyzed the interaction between sound and experiment and the results revealed that sound had a greater effect on visual fidelity perception when the visual were presented in monoscopic as opposed to stereoscopic 3-D (Experiment 2).

In Experiment 1, visual fidelity was defined with respect to polygon count while in Experiment 2, visual fidelity was defined with respect to texture resolution. The difference in fidelity between the visuals defined with respect to polygon count may be more pronounced and thus more noticeable

than the visuals defined with respect to texture resolution which appear to be smoother (see Figs. 1 and 3, respectively). This difference may be greater emphasized when presented in stereoscopic 3-D and when presented in combination with certain auditory cues, the auditory cues will lead to a greater increase in visual fidelity.

In Experiment 3, the white noise auditory condition led to a decrease in visual fidelity perception across each of the six visuals considered. The results of Experiment 3 considered the effect of auditory cues that provided contextual information with respect to the visual scene. As with our previous studies that considered monoscopic viewing, the white noise auditory condition led to a decrease in visual fidelity perception. However, although as shown in Fig. 6 there does appear to be an increase in visual fidelity perception with the addition of classical and heavy metal music and the other contextual auditory conditions, this apparent increase is not statistically significant. This is contradictory to the results of previous experiments that considered a monoscopic viewing environment with noncontextual auditory cues [34] (the effect of contextual auditory cues on visual fidelity perception was not examined in our previous work). In Experiments 1 and 2 described here, the visual stimuli consisted of a surgeon's head/face while in addition to the surgeon's head, the visual stimuli of Experiment 3 included the surgeon's torso with a surgical drill in the hand of the surgeon. Although the contextual cues were intended to correspond to the visual scene, the resulting correspondence may not have been as great as initially assumed. More specifically, the participants were not surgeons or medical practitioners and may not have been familiar with an operating room and the sounds contained within an operating room. In other words, the notion of contextual auditory cues may also be subjective and may depend on prior experience. Furthermore, auditory cues in general may be subjective and therefore, the resulting effect on visual fidelity perception may differ across individuals. Further testing is required to develop a greater understanding of the subjective effect of sound on visual fidelity perception.

## VIII. DISCUSSION AND CONCLUSION

In this paper, we presented the results of three experiments that examined the perception of visual fidelity in stereoscopic 3-D viewing under various (background) auditory conditions. In the first and second experiments, visual fidelity was defined with respect to polygon count and texture resolution (of a model of a surgeon's head) respectively, and the four noncontextual (with respect to the visual scene) auditory conditions were: 1) no sound at all; 2) white noise; 3) classical music; and 4) heavy metal music. Building upon these two experiments, in the third experiment, in addition to the four noncontextual auditory conditions, three contextual auditory conditions were also considered: 1) operating room ambiance; 2) operating room ambiance mixed with a surgical drill sound; and 3) surgical drill sound.

The series of experiments presented here are of course merely first steps in understanding the impact of multimodal interactions. The results of these experiments indicated that

sound does indeed have an effect on visual fidelity perception within a stereoscopic viewing environment. More specifically, white noise consistently led to a reduction of visual fidelity perception while classical music and heavy metal music can sometimes lead to an increase in visual fidelity perception.

The results presented here and in previous work suggest that sound can affect various aspects of a virtual simulation/ game. Distracting sounds such as white noise can lead to a decrease in the perception of visual fidelity; classical music (and heavy metal music in some cases) can cause our perception of visual fidelity to increase. Although previous work has examined the interaction of auditory cues and other senses, the experiments presented in this paper specifically examined the interaction of auditory cues and visuals within a stereoscopic 3-D viewing environment.

The significance of this paper lies in the simple fact that sound was shown to influence the perception of visuals with respect to stereoscopic 3-D, and that this influence was relatively consistent with monoscopic cues, with some exceptions. This impact of sound on visual fidelity will have implications for any studies into stereoscopic 3-D in relation to audio-visual media, and particularly in studies that explore the role of stereoscopic 3-D in games, virtual environments and simulations. Any future studies into fidelity or stereoscopic 3-D must take into consideration the importance of multimodal interactions.

Our results provided a greater understanding of multimodal interactions within a stereoscopic 3-D virtual environment. Specifically, noncontextual sound (with respect to the visual scene) influenced monoscopic and stereoscopic fidelity perception in roughly the same way. Based on these results, special consideration on the part of composers of music for games and other virtual environments was not, therefore, recommended as necessary in changing from monoscopic to stereoscopic 3-D graphics. However, our results suggested that some discrepancy may occur in the perception of visual fidelity between monoscopic and stereoscopic viewing with respect to contextual auditory cues (sonic environments or ambient sound beds), and it was recommended that in the case of visual fidelity reductions for processing requirements, contextual auditory cues should be included in user quality assurance testing.

The discrepancy between monoscopic and stereoscopic environments can be explained by the cognitive load theory. This term was used in cognitive psychology to illustrate the load related to the executive control of working memory (WM) [39]. Theories contend that during complex perceptual and 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 [39]. A possible explanation for the differences in visual fidelity perception between monoscopic and stereoscopic 3-D viewing is that higher cognitive resources are required to process the additional information (depth) presented to viewers in stereoscopic 3-D information that they were being presented.

The results presented here indicated that multimodal interaction was fundamental to perceiving the external world

and therefore, any virtual environment (games, serious games, etc.) should include, and account for multimodal interactions. However, multimodal interactions were constrained by our cognitive resources that may become loaded depending on the way that our senses were being stimulated [39]. With this in mind, these results will ultimately pave the way to a better understanding of how knowledge transfer and retention may be constrained by the cognitive resource demands required to perceive events in a multimodal environment.

Although greater work remains to be done with respect to the use of stereoscopic 3-D within virtual environment/gaming environments and its interaction with other modalities (sound, haptics, etc.), based on the results presented here, designers and developers of virtual environments, particularly those aimed at training and education (e.g., 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 that, in many situations, characterize the real-world environment. Perhaps virtual environments 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.

The participants considered here were young (average age 21 across all three experiments) and their musical preferences were not considered; hence, we cannot draw any conclusions regarding what, if any, role musical preference had on the results. Future work will include repeating these experiments with a larger participant size across a wider age range and will include examining musical preferences and prior music-based knowledge. Moreover, the experiment should also be run on different stereoscopic 3-D settings. Future work will also examine whether specific attributes of the music were responsible for the results (such as the presence of lyrics, the frequency band, and rhythm, amongst others). Future work will also consider the effect of user interaction on visual fidelity 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 us to compare sounds that are unnatural to an interaction with a virtual object being acted upon with more natural sounds and will take us closer to answering our questions regarding fidelity and its effect on knowledge transfer and retention.

## REFERENCES

- [1] V. Hulusic, M. Aranha, and A. Chalmers, "The influence of cross-modal interaction on perceived rendering quality thresholds," in *Proc. Comput. Graph. Visualization Vision*, 2008, pp. 41–48.
- [2] J. Blascovich and J. Bailenson, *Infinite Reality*. New York, NY, USA: Harper Collins, 2011.
- [3] M. Mori. "The uncanny valley," *Energy*, vol. 7, no. 4, pp. 33–35, 1970.
- [4] I. Tsirlin, L. M. Wilcox, and R. S. Allison, "Disparity biasing in depth from monocular occlusions," *Vision Res.*, vol. 51, no. 14, pp. 1699–1711, 2011.

- [5] I. Bickerstaff. "Case study: The introduction of stereoscopic games on the Sony PlayStation 3," in *Proc. Stereoscopic Displays Appl. XXIII*, 2012, pp. 8238–8288.
- [6] J. Schiefele, K. U. Dorr, M. Olbert, and W. Kubbat, "Stereoscopic projection screens and virtual cockpit simulation for pilot training," in *Proc. 3rd Int. Immersive Projection Technol. Workshop*, 1999, pp. 211–222.
- [7] F. Tendick, M. Downs, T. Goktekin, M. C. Cavusoglu, D. Feygin, X. Wu, R. Eyal, M. Hegarty, and L. W. Way, "A virtual environment testbed for training laparoscopic surgical skills," *Presence Teleop. Virt.*, vol. 9, no. 3, pp. 236–255, 2000.
- [8] J. S. Henn, G. M. Lemole, M. A. T. Ferreira, L. F. Gonzalez, M. Schornak, M. C. Preul, and R. F. Spetzler, "Interactive stereoscopic virtual reality: A new tool for neurosurgical education," *J. Neurosurg*, vol. 96, no. 1, pp. 144–149, 2002.
- [9] B. Reitingger, A. Bornik, R. Beichel, and D. Schmalstieg, "Liver surgery planning using virtual reality," *IEEE Comput. Graph.*, vol. 26, no. 6, pp. 36–47, Nov.–Dec. 2006.
- [10] T. Fisk and D. Hill, "3-D simulation brings immersive dimension to operator training," *ARC Insight* 2010-06MP, Feb. 2010.
- [11] N. Adamo-Villani and K. Wright, "SMILE: An immersive learning game for deaf and hearing children," in *Proc. ACM SIGGRAPH-Educators*, Article 17, 2007.
- [12] G. Provost. *Beautiful, Yet Friendly Part 2: Maximizing Efficiency* [Online]. Available: <http://www.ericchadwick.com/examples/provost/byf2.html>
- [13] S. Guo, P. Gerasimov, and B. Aona, "Practical game performance analysis using Intel graphics performance analyzers," Intel Corporation White Paper, 2011.
- [14] C. Rubino and J. Power, "Level design optimization guidelines for game artists using the epic games: Unreal editor and unreal engine 2," *Comput. Entertain.*, vol. 6, no. 4, article 55, 2008.
- [15] N. Mourkoussis, F. M. Rivera, T. Troscianko, T. Dixon, R. Hawkes, and K. Mania, "Quantifying fidelity for virtual environment simulations employing memory schema assumptions," *ACM Trans. Appl. Percept.* vol. 8, no. 1, article 2, 2010.
- [16] W. Hürst and M. Helder, "Mobile 3-D graphics and virtual reality interaction," in *Proc. 8th Int. Conf. Advances Computer Entertainment Technol.*, Article 28, Nov. 2011.
- [17] X. Ma, M. Dong, L. Zhong, and Z. Deng, "Performance and power consumption characterization of 3-D mobile games," *Computer*, vol. 99, pp. 76–82, 2012.
- [18] H. McGurk and J. MacDonald, "Hearing lips and seeing voices," *Nature*, vol. 264, no. 5588, pp. 746–748, Dec.
- [19] A. Chalmers and K. Debattista, "Levels of realism for serious games," in *Proc. IEEE Games Virtual Worlds Serious Appl.*, Mar. 2009, pp. 225–232.
- [20] B. Ramic, A. Chalmers, J. Hasic, and S. Rizvic, "Selective rendering in a multimodal environment: Scent and graphics," in *Proc. Spring Conf. Comput. Graph.*, 2007, pp. 189–192.
- [21] V. Hulusic, C. Harvey, K. Debattista, N. Tsingos, S. Walker, D. Howard, and A. Chalmers, "Acoustic rendering and auditory-visual cross-modal perception and interaction," *Comput. Graph. Forum* vol. 31, no. 1, pp. 102–131, 2012.
- [22] A. T. Woods, E. Poliakoff, D. M. Lloyd, J. Kuenzela, J. R. Hodson, H. Gondaa, J. Batchelora, G. B. Dijksterhuis, and A. Thomas, "Effect of background noise on food perception," *Food Qual. Prefer.*, vol. 22, no. 1, pp. 42–47, 2011.
- [23] C. Conrad, Y. Konuk, P. Werner, C. G. Cao, A. Warshaw, D. Rattner, D. B. Jones, and D. Gee, "The effect of define auditory conditions versus mental loading on the laparoscopic motor skill performance of experts," *Surg. Endosc.*, vol. 24, no. 6, pp. 1347–1352, 2010.
- [24] G. Mastoropoulou, K. Debattista, A. Chalmers, and T. Troscianko, "The influence of sound effects on the perceived smoothness of rendered animations," in *Proc. 2nd Symp. Appl. Perception Graph. Visualization*, Aug. 2005, pp. 9–15.
- [25] P. Larsson, D. Västjäll, and M. Kleiner, "On the quality of experience: A multi-modal approach to perceptual ego-motion and sensed presence in virtual environments," in *Proc. 1st Int. Speech Commun. Assoc. Tutorial Research Workshop Auditory Quality Systems*, 2003, pp. 97–100.
- [26] P. Bertelson and M. Radeau, "Cross-modal bias and perceptual fusion with auditory-visual spatial discordance," *Percept. Psychophys.*, vol. 29, no. 6, pp. 578–584, 1981.
- [27] L. Shams, Y. Kamitani, and S. Shimojo, "Visual illusion induced by sound," *Cogn. Brain Res.* vol. 14, no. 1, pp. 147–152, 2002.
- [28] V. Hulusic, K. Debattista, V. Aggarwal, and A. Chalmers, "Maintaining frame rate perception in interactive environments by exploiting audio-visual cross-modal interaction," *Vis. Comput.* vol. 27, no. 1, pp. 57–66 2011.
- [29] V. Hulusic, K. Debattista, and A. Chalmers, "Smoothness perception: Investigation of beat rate effect on frame rate perception," *Vis. Comput.*, published Online Nov. 9, 2012.
- [30] A. Turner, J. Berry, and N. Hollman, "Can the perception of depth in stereoscopic images be influenced by 3-D sound," in *Proc. Stereoscopic Displays Appl. XXII*, 2011, pp. 7806–7863.
- [31] B. Cullen, D. Galperin, K. Collins, B. Kapralos, and A. Hogue, "The effects of audio on depth perception in S3D games," in *Proc. ACM Audio Mostly*, Sep. 2012, pp. 32–39.
- [32] S. L. Storms and M. J. Zyda, "Interactions in perceived quality of auditory-visual displays," *Presence Teleop. Virt.*, vol. 9, no. 6, pp. 557–580, 2000.
- [33] D. Rojas, B. Kapralos, S. Crsitancho, K. Collins, C. Conati, and A. Dubrowski, "The effect of background sound on visual fidelity perception," in *Proc. ACM Audio Mostly*, Sep. 2011, pp. 32–39.
- [34] D. Rojas, B. Kapralos, S. Crsitancho, 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," *Stud Health Technol Inform.* vol. 173, pp. 386–392, 2012.
- [35] C. Morfey, *The Dictionary of Acoustics*. San Diego, CA., USA. 2001.
- [36] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ, USA: Lawrence Erlbaum, 1988.
- [37] J. Dargie. "Modeling techniques: Movies vs. games," *ACM SIGGRAPH Comput. Graph.* vol. 41, no. 2, Article 2, 2007.
- [38] Autodesk. (2011). Autodesk animation academy curriculum: Level 1: Texture basics [Online]. Available: [http://students.autodesk.com/ama/orig/Curriculum/ADA2011/2010/HTML/AA\\_Level\\_1/PDF\\_Tutorials/Texture\\_Basics/AA\\_L1\\_Texture\\_Principles.pdf](http://students.autodesk.com/ama/orig/Curriculum/ADA2011/2010/HTML/AA_Level_1/PDF_Tutorials/Texture_Basics/AA_L1_Texture_Principles.pdf)
- [39] F. Paas, A. Renkel, and J. Sweller, "Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture," *Instructional Sci.*, vol. 32, no. X, pp. 1–8, 2004.

**David Rojas** is currently pursuing the Ph.D. degree in the Medical Science Program, University of Toronto, Toronto, ON, Canada.

His current research interests include multimodal virtual interaction, influence of auditory events on visual fidelity perception, and use of online collaborative tools to teach psychomotor skills. He is also interested in the merging of virtual reality/serious games and social networking environments for medical education applications where these new technologies could enhance the current curriculum.

**Bill Kapralos (M'13)** is currently an Associate Professor at the Game Development and Entrepreneurship Program, University of Ontario, Institute of Technology, Oshawa, Canada. His current research interests include serious games and multimodal virtual environments/reality, the perception of auditory events, and 3-D (spatial) sound generation for interactive virtual environments and video games.

**Andrew Hogue** is currently an Assistant Professor at the University of Ontario Institute of Technology, Oshawa, Canada, within the Game Development and Entrepreneurship Program, Faculty of Business and IT. His current research interests include the development and evaluation of game design techniques for education, the use of serious game technology to make business education more effective, immersive VR technology, stereoscopic 3-D game design, and computer vision.

**Karen Collins** is the Canada Research Chair in Interactive Audio and Advisory Board Chair of the Games Institute, University of Waterloo, Waterloo, ON, Canada. Her current research interests include an embodied cognition theory to the sound in video games.

**Lennart Nacke** is currently an Assistant Professor for human–computer interaction and game science at the University of Ontario Institute of Technology (UOIT), Oshawa, Canada. His current research interests include three main themes: augmenting player control with physiological sensor information and contextual data, using applied game design techniques in non-game systems to increase user engagement, and affective evaluation of games, user interfaces, and entertainment computing systems.

**Sayra Crsitancho** received the Ph.D. degree in biomedical engineering from the University of British Columbia, Vancouver, BC, Canada.

She is currently an Assistant Professor with the Department of Surgery and the Department of Medical Biophysics, and a Scientist with the Centre for Education Research and Innovation, Schulich School of Medicine and Dentistry, UWO. She is specifically interested in studying how surgical experts deal with operative difficulties and how one can translate differences in the surgical strategies of trainees and experts into instruction and training. By exploring how experienced surgeons use interactions with the surgical environment to acquire a better understanding of the challenging situation at hand, she intends to provide evidence that will allow moving the research on surgical decision-making from a task-related to a systems-related conceptualization.

**Cristina Conati** received the M.Sc. degree in computer science from the University of Milan, Milan, Italy, in 1988, and the M.Sc. and Ph.D. degrees in artificial intelligence from the University of Pittsburgh, Pittsburgh, PA, USA, in 1996 and 1999, respectively.

She is currently an Associate Professor of computer science at the University of British Columbia, British Columbia, Vancouver, BC, Canada. Her current research interests include adaptive interfaces, intelligent tutoring systems, user modeling, and affective computing. She published over 50 strictly refereed articles.

Dr. Conati received Best Paper Awards from the International Conferences on User Modeling, AI in Education, Intelligent User Interfaces, and the Journal *User Modeling and User-Adapted Interaction*.

**Adam Dubrowski** is currently a Scientist at the SickKids Learning and Research Institutes and an Associate Professor at the Department of Paediatrics, University of Toronto, Toronto, ON, Canada. His current research interests include three main areas: explorations of conditions of practice that optimize simulation-based education, the use of virtual learning environments including educational networking and serious and educational games, and validation of assessment methods.