Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Psychological and Physiological Acoustics
Session 3pPP: Multimodal Influences on Auditory Spatial Perception

3pPP1. Spatial sound and its effect on visual quality perception and task performance within a virtual environment

Brent Cowan, David Rojas, Bill Kapralos*, Karen Collins and Adam Dubrowski

***Corresponding author's address: Faculty of Business and Information Technology, University of Ontario Institute of Technology, Oshawa, L1H 7K4, Ontario, Canada, bill.kapralos@uoit.ca**

Immersive 3D virtual environments such as simulations and serious games for education and training are typically multimodal, incorporating at the very least both visual and auditory cues, each of which may require considerable computational resources, particularly if high fidelity environments are sought. It is widely accepted that sound can influence the other modalities. Our own previous work has shown that sound cues (both contextual and non-contextual with respect to the visual scene) can either increase or decrease (depending on the sound) visual fidelity (quality) perception in addition to the time required to complete a simple task (task completion time) within a virtual environment. However, despite the importance and benefits of spatial sound (sound that goes far beyond traditional stereo and surround sound techniques, allowing users to perceive the position of a sound source at an arbitrary position in three-dimensional space), our previous work did not consider spatial sound cues. Here we will build upon our previous work by describing the results of an experiment that will be conducted to examine visual fidelity (quality) perception and task performance in the presence of various spatial sound cues including acoustical reverberation and occlusion/diffraction effects, while completing a simple task within a virtual environment.

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.4798377]
Received 22 Jan 2013; published 2 Jun 2013
Proceedings of Meetings on Acoustics, Vol. 19, 050126 (2013)

INTRODUCTION

In the context of a simulation (physical and virtual), *fidelity* denotes the extent to which the appearance and/or behavior of the simulation matches the appearance and behavior of the real system (Farmer et al., 1999; Hays and Singer, 1989). Fidelity can be divided into two components: i) psychological fidelity, and ii) physical fidelity (Ker and Bradley, 2010). *Psychological fidelity* refers to the degree to which the skills inherent in the real task being simulated are captured within the simulation (Ker and Bradley, 2010) and may also include the degree of reality perceived by the user of the simulation (the trainee) (Rehmann et al., 1995). *Physical fidelity* covers the degree of similarity between the training situation and the operational situation which is simulated (Hays and Singer, 1989; Ker and Bradley, 2010). Physical fidelity can also be further divided into equipment fidelity that denotes the degree that the simulation replicates reality and *environmental fidelity* that denotes the degree that the simulation replicates the sensory cues (Ker and Bradley, 2010; Rehmann et al., 1995).

Immersive 3D virtual environments such as simulations and serious games for education and training often aim to mimic the real-world and are therefore typically multimodal, incorporating at the very least both visual and auditory cues. Each of these may require considerable computational resources, particularly if high fidelity environments are sought as they often are. However, it is currently beyond our capability to faithfully account for all of the human senses within a virtual simulation/serious game. In other words, complete (perfect) multi-sensory fidelity may be impossible to achieve, at least with our current technology. Even if we consider the audio and visual modalities only, real-time high fidelity audio and visual rendering (particularly of complex environments) is still not feasible, despite the computing hardware currently available (Hulusic, 2008). In addition, given the associated burden on computational resources, striving for such high fidelity environments increases the probability of lag and subsequent discomfort and simulator sickness (Blascovich and Bailenson, 2011). Moreover, it remains unclear if such fidelity is actually needed for either enjoyment or knowledge transfer and retention, and striving to reach higher levels of fidelity can also lead to increased computational requirements (processing time and memory resources) and increased development time and costs.

It has been shown that sound can influence our perception of a rendered graphics scene and vice versa (Hulusic, 2008). Various studies have examined the perceptual aspects of audio-visual cue interaction (amongst other multi-sensory interactions), and it has been shown that sound can potentially attract part of the user's attention away from the visual stimuli and lead to a reduced cognitive processing of the visual cues (Mastoropoulou et al., 2005). Sound can, for example, attract part of the viewer's attention away from any visual defects inherent in low frame-rate animations and lead to a reduced cognitive processing of the visual cues (Mastoropoulou et al., 2005). 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 and observed that the visual level of detail was perceived to be higher as the auditory level of detail was increased. Hulusic et al. (2008) showed that sound effects allowed slow animations to be perceived as smoother than fast animations and that the addition of footstep sound effects of footsteps to walking (visual) animations increased the animation smoothness perception. Our previous work examined the perception of visual fidelity defined with respect to polygon count and texture resolution within a non-stereoscopic and stereoscopic 3D environment under various ambient auditory conditions (both contextual and non-contextual auditory cues with respect to the visual environment) (Rojas et al., 2011; Rojas et al., 2012). Results from our previous work indicate that auditory cues can have a large effect on visual fidelity perception. More specifically, the perception of visual fidelity increased in the presence of classical music, while the perception of visual fidelity perception was greatly decreased in the presence of white-noise auditory cues. Results also indicated that contextual auditory cues can lead to a great increase in visual fidelity perception (Rojas et al., 2011; Rojas et al., 2012).

Motivated by these studies and the general lack of emphasis on audition in virtual environments, and games, where the emphasis is generally placed on visuals/graphics (Carlile, 1996), we have begun investigating multimodal interactions within virtual simulation, gaming, and serious gaming environments. Our long-term goal is to develop an understanding of simulation (and serious games in particular) fidelity, multimodal interactions, user-specific factors and their effects on knowledge transfer and retention. Although our previous findings indicate a strong effect between ambient auditory cues and visual fidelity perception, many questions remain. More specifically, in our previous studies, participants were presented with a static visual scene in the presence of mono (non-spatial) sound. *Spatial sound* goes far beyond traditional stereo and surround sound techniques, and allows users to perceive the position of a sound source at an arbitrary position in three-dimensional space. Adding realistic spatial sound to interactive applications such as a virtual environments, video games, and serious games can add a new layer of realism (Antani et al., 2012), contributes to a greater sense of "presence", or "immersion" (Pulkki, 2001), improve task performance (Zhou et al., 2007), convey information that would otherwise be difficult to convey using other

modalities (e.g., vision) (Zhou et al., 2007), and improve navigation speed and accuracy (Makino et al., 1996). Yet, despite the importance and benefits of spatial sound, previous work, including our own, did not consider spatial sound cues. Here, we build upon our previous work by examining visual fidelity perception and task completion time in the presence of contextual auditory cues (with respect to the visual scene) while carrying out a simple real-time interactive task within a virtual environment. In contrast to our previous work, here we consider both spatial and non-spatial sound, and rather than defining visual fidelity with respect to polygon count or texture resolution, here we define visual fidelity with respect to a "cartoon" ("cel-shading") effect. The non-photorealistic rendering effect is designed to make computer graphics appear to be hand-drawn and is widely used in video games but not with serious games. We hypothesize that the presence of spatial sound will lead to i) an increase in visual fidelity perception of both the originally (non-processed) and cel-shaded visual scenes, and ii) a decrease in the time required to complete the task. The cel-shading effect examined here is achieved by post-processing the originally rendered scene (at interactive rates using the graphics processing unit), and therefore, there is no computational savings as a result. However, if, as we hypothesize, spatial sound will lead to an increase visual fidelity, and a decrease in task completion time, it may provide further motivation to incorporate cel-shading effects into virtual environments (simulations and serious games), further investigate the effect of spatial sound on visual fidelity perception, and at the very least, further reinforce the importance of spatial sound within a virtual environment.

The remainder of the paper is organized as follows. Experimental details are provided in Section 2 ("Methods and Materials") while experimental results are provided in Section 3 ("Results"). Finally, a discussion of the results and plans for future research are provided in Section 4 ("Discussion and Future Work").

METHODS AND MATERIALS

Participants

Participants consisted of unpaid volunteers and were students from the University of Ontario Institute of Technology. A total of six (three female, and three male) volunteers participated in the experiment (average age was 21 years old). None of the participants reported any hearing or visual defects. The authors did not participate in the experiment and the experiment abided by the University of Ontario Institute of Technology Research Ethics Review process for experiments involving human participants.

Visual Stimuli

The visual scene consisted of five rendered (3D) versions of a virtual operating room with various tools, and equipment (see Figure 1). The virtual operating room is part of a serious game for total knee arthroplasty (Cowan et al., 2010), and was modified to allow for this experiment. Within the operating room were three non-player characters (nurses) which, for the purposes of this experiment, remained static and did not afford any interaction with the participants. The five conditions included the following (see Figure 1): i) original (no effect), ii) cel-shading with three levels (i.e., color is divided into three discrete levels), and iii) cel-shading with six levels (i.e., color is divided into six discrete levels). The experiment was carried out on an Acer Aspire laptop with a 15.6" screen size and a screen resolution of 1366 × 768. The operating room environment was viewed in "full screen" mode.

(a) (b) (c)

FIGURE 1. Visual stimuli considered in this experiment. (a) Original (non-filtered), (b) cel-shading with three levels (i.e., color is divided into six discrete levels), and (c) cel-shading with six levels.

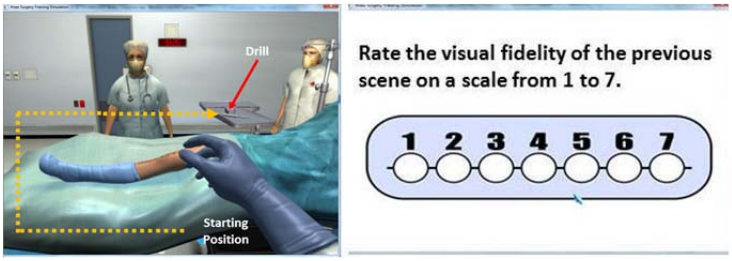
Auditory Stimuli

Three auditory conditions were examined: i) no sound (visuals only), ii) monaural (non-spatial) surgical drill sound, iii) spatialized surgical drill sound. The surgical drill sound was obtained by recording an actual drill sound. The (monaural) 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, and sampled at a rate of 44.1 kHz. All auditory stimuli were output with a pair of AKG Acoustics A240 headphones. The spatialized sound of condition iii) included head-related transfer function (HRTF) processing and environmental (occlusion and reverberation) processing. HRTF processing was using the software-based, HRTF approximation provided by FMod audio library (default settings). The FMod software HRTF approximation applies a low-pass filter to sounds that are behind the listener. A maximum and minimum angle defines conical areas behind the listener with the listener's head at the cone's apex. The minimum angle is a narrow cone defining the area where the low-pass filter effect is applied at maximum strength (4,000 Hertz). The maximum angle defines a wider cone. The strength of the effect is interpolated for angles between the maximum and minimum angle allowing the effect to fade in and out as the head is rotated relative to the sound source. Environmental processing was accomplished using the acoustical occlusion modeling method introduced by Cowan and Kapralos (2010), and the acoustical reverberation modeling method also introduced by Cowan and Kapralos (2011), both of which employed the graphics processing unit and therefore operated at interactive rates.

Experimental Method

Participants were seated in front of the Acer Aspire 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 four versions of a virtual operating room in conjunction with one of the three sound conditions previously described. Their task was to navigate through the operating room from their starting position to a point in the room which contained a tray with surgical instruments and pick up a surgical drill (they had to navigate around the bed and one of the NPC nurses to reach the tray that contained the surgical instruments; see Figure 2(a)).

Navigation through the environment was accomplished in a first-person perspective (were participants took 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 surgeon 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 user's avatar and the participant was prompted to rank the visual scene with respect to their perceived visual fidelity on a scale from 1 (lowest perceived fidelity), to 7 (highest perceived fidelity); see Figure 2(b). Aside from interacting with the surgical instruments, 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 trial; the following trial began after the user clicked the "Continue" button that appeared on the visual fidelity ranking screen after the participant entered their choice. Each of the four rendered versions of the operating room and each of the three sound combinations (12 combinations in total) was repeated three times for a total of 36 trials, all of which were presented in a randomized ordering. The experiment took approximately 15 minutes to complete, and all of the participants completed the experiment in a single session.



(a) (b)

FIGURE 2. The virtual operating room environment. (a) Then view of the operating room environment at the start of each experimental trial. The task of each participant was to navigate the environment from the starting position to the position of the surgical drill and then "choosing" the drill. (b) Upon choosing the drill, participants were prompted to rank their perception of the visual fidelity.

RESULTS

Within the experiment, two dependent variables were collected: i) visual quality perception, and ii) task completion time (which was measured from the moment the participant started at trail until the moment they picked up the drill). A summary of the results (average mean and standard deviation across all participants) for visual fidelity perception and task completion time are provided in Table 1 and Table 2 respectively. Examination of the results in Table 1 (average visual fidelity perception) reveals that the difference between the average values for each of the three visual conditions is very small irrespective of the auditory condition. More specifically, the difference between the minimum and maximum fidelity value for the "original", "3-level cel-shaded" and "6-level cel-shaded" visual conditions were: 0.1, 0.2, and 0.2 respectively. However, the average fidelity is larger when considering the "original" visual condition. A one-way between participants ANOVA was conducted to compare the effect of the auditory conditions over the perception of visual fidelity for each of the three renderings of the virtual operating room. There was no significant effect of auditory cues on visual fidelity perception at the $p < .05$ level for any of the conditions [$F(2,4) = .080$, $p = .812$].

TABLE 1. Average visual fidelity and standard deviation (averaged across all six participants).

Auditory Condition	Visual Condition	Avg. Visual Fidelity	Std. Dev.
No sound	Original	5.7	1.2
	Cel-shading (3 levels)	3.2	1.1
	Cel-shading (6 levels)	3.9	1.3
Monaural Sound	Original	5.6	1.0
	Cel-shading (3 levels)	3.0	1.1
	Cel-shading (6 levels)	3.9	1.2
Spatial Sound	Original	5.7	1.2
	Cel-shading (3 levels)	3.0	1.3
	Cel-shading (6 levels)	3.7	1.2

Examination of the average values of Table 2 (average task completion time) reveals that the difference between average task completion time was generally very small across each of the conditions although the average task completion times for the "monaural" and "spatial sound" auditory conditions were slightly lower than the average task completion times for the "no sound" auditory condition, indicating that the participants did take longer to complete the task in the absence of sound. A one-way between participants ANOVA was conducted to compare the effect of the auditory conditions on task completion time for the three visual renderings of the virtual operating room revealed that there was no significant difference at the $p < .05$ level for any of the conditions [$F(2,4) = 1.795$, $p=.170$].

TABLE 1. Average task completion time and standard deviation (averaged across all six participants).

Auditory Condition	Visual Condition	Avg. Task Completion Time	Std. Dev.
No sound	Original	11.8	11.0
	Cel-shading (3 levels)	9.5	4.6
	Cel-shading (6 levels)	13.5	8.9
Monaural Sound	Original	8.5	2.5
	Cel-shading (3 levels)	10.0	4.4
	Cel-shading (6 levels)	9.0	2.9
Spatial Sound	Original	9.9	4.7
	Cel-shading (3 levels)	10.1	5.3
	Cel-shading (6 levels)	10.2	7.2

DISCUSSION AND FUTURE WORK

Contrary to our original hypothesis, the presence of sound (spatial and non-spatial) did not have any effect on visual fidelity perception and task completion time. With respect to task completion time, our results are also contrary to previous work that has examined performance in the presence of sound and music. For example, a study

by Chang and Thompson (2011) demonstrated that whines, cries, and "child-directed speech" distracted listeners completing simple mathematical (subtraction) problems. Similarly, Woods et al. (2011) discovered that sound (noise) can have an effect on the perception of food gustatory properties, food crunchiness and food liking, while Conrad et al. (2010) discovered that stressful music (e.g., heavy metal music) had a negative impact on the time required to complete a laparoscopic surgery task but did not impact task accuracy while classical music had a variable effect on the time required to complete the task but resulted in greater task accuracy. That being said, the virtual operating room environment considered here was small and with few objects obstructing the path between the sound source and the listener and as a result, the spatial sound cues may have been limited. In addition, the task itself (i.e., navigate through a small operating room and pick up the surgical drill), was rather simple and the participants could see the surgical drill from the starting position; hence, they may have simply relied on visual cues to complete this simple task. With respect to visual fidelity perception, once again, our results are also contrary with our own previous work that revealed the perception of visual quality of a virtual model is dependent on sound (Rojas et al., 2011; Rojas et al., 2012). In those studies, it was observed that white noise resulted in a decrease of visual fidelity perception for visual fidelity defined with respect to polygon count and texture resolution of a static object consisting of a 3D rendered model of a surgeon. In contrast, sound consisting of either classical music or heavy metal music led to an increase in the perception of visual fidelity when considering both polygon count, and texture resolution. However, here we considered only cel-shading effects (two levels: three and six) and our sound consisted of a surgical drill sound that was either spatialized or non-spatialized. The drill sound itself was contextual with respect to the visual scene (i.e., related to the drill which the participants had to reach and pick up) and may not have had any distracting effects as illustrated in our previous work where the sounds (white noise, and heavy metal music in particular) were non-contextual with respect to the visual scene (static 3D rendering of a surgeon) and led to a decrease in the perception of visual fidelity.

Finally, results presented here are very preliminary and represent a very small sample size. Greater work must be carried out before any conclusions regarding the interaction of spatial sound and visual fidelity and task completion time can be made. Future work will include repeating the experiment with a larger, and more diverse, participant population. Future work will also consider of additional definitions of graphical/visual fidelity, including the polygon count and texture resolution of the models within the virtual environment in addition to a more complex environment with more objects that occlude the direct path between the sound source and the listener.

ACKNOWLEDGMENTS

The financial support of the *Canadian Network of Centres of Excellence* (NCE), *Graphics, Animation, and New Media* (GRAND) initiative, the *Social Sciences and Humanities Research Council of Canada* (SSHRC) and the *Natural Sciences and Engineering Research Council of Canada* (NSERC) is gratefully acknowledged.

REFERENCES

Antani, L., Chandak, A., Savioja, L., and Manocha, D. **(2012)**. "Interactive sound propagation using compact acoustic transfer operators," ACM Transactions on Graphics, **31**.

Bergeron, B. *Developing Serious Games*, **(2006)**. Thomson Delmar Learning, Hingham MA USA.

Blascovich, J., and Bailenson, J. *Infinite Reality,* **(2011)**. Harper Collins, New York NY USA.

Carlile, S. *Virtual auditory space: Generation and application*, **(1996)**. R. G. Landes, Austin TX USA, (1996).

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

Conrad, C., Konuk, Y., Werner, P., Cao, C.G., Warshaw, A., Rattner, D., Jones, D.B., and Gee, D. **(2010)**. The effect of define auditory conditions versus mental loading on the laparoscopic motor skill performance of experts. Surgical Endoscopy 24, 1347–1352.

Cowan, B., and Kapralos, B. **(2011)**. "A real-time, GPU-based method to approximate acoustical reverberation effects," Journal of Graphics, GPU, and Game Tools, **15**, 210-215.

Cowan, B., and Kapralos, B. **(2010)**. "GPU-based real-time acoustical occlusion modeling," Virtual Reality, **14**, 183-196.

Cowan, B., Sabri, H., Kapralos, B., Porte, M., Backstein, D., Cristancho, S., and Dubrowski, A. **(2010)**. "A serious game for total knee arthroplasty procedure education and training," Journal of Cybertherapy and Rehabilitation, **3**, 285-298.

Farmer, E. Rooij, J. von., Riemersma, J., Joma, P., and Morall, J. **(1999)**. *Handbook of Simulator Based Training*. Ashgate Publishing, Surrey UK.

Hays, R. T., and Singer, M. **(1989)**. *Simulation Fidelity in Training System Design*. Springer, New York, NY, USA.

Hulusic, V., Aranha, M., and Chalmers, A. **(2008)**. "The influence of cross-modal interaction on perceived rendering quality thresholds," *16th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision '2008*, Plzen - Bory, Czech Republic, pp. 41-48.

Ker, J., and Bradley, P. **(2010)**. "Simulation in medical education," in U*nderstanding Medical Education: Evidence, Theory and Practice* edited by T. Swanwick, Wiley-Blackwell, Ch. 12, pp. 164-180.

Makino, H., Ishii, I., Nakashizuka, M. **(1996)**. "Development of navigation system for the blind using GPS and mobile phone communication," *Proceedings of the 18th Annual Meeting of the IEEE Engineering in Medicine and Biology Society*, Amsterdam, the Netherlands, pp. 506-507.

Mastoropoulou, G., Debattista, K., Chalmers, A., and Troscianco, T. **(2005)**. "The influence of sound effects on the perceived smoothness of rendered animations," *Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization*, La Coruña, Spain, pp. 9–15.

Pulkki, V. **(2001)**. Spatial sound generation and perception by amplitude panning techniques. PhD Thesis, Electrical and Communications Engineering, Helsinki University of Technology, Finland.

Rehmann, A. J., Mitman, R. D., and Reynolds, M. C., **(1995)**. *A handbook of flight simulation fidelity requirements of human factors research*. Crew System Ergonomics Information Analysis Center, Wright-Patterson Air Force base, Dayton, OH, USA.

Rojas, D., Kapralos, B., Crsitancho, S., Collins, K., Conati, C., and Dubrowski, A. **(2011)**. "The effect of background sound on visual fidelity perception," *Proceedings (Extended Abstracts) of ACM Audio Mostly 2011*, Coimbra, Portugal.

Rojas, D., Kapralos, B., Crsitancho, S., Collins, K., Conati, C., and Dubrowski, A. **(2012)**. "Developing effective serious games: The effect of background sound on visual fidelity perception with varying texture resolution," Studies in Health Technology and Informatics **173**, 386-392.

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

Zhou, Z. Y., Cheok, A. D., Qiu, Y., and Yang, X. **(2007)**. "The role of 3-D sound in human reaction and performance in augmented reality environments," IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, **37**, 262-272.