Sound and stereoscopic 3D:

Examining the effects of sound on depth perception in stereoscopic 3D

Brian Cullen, Karen Collins
University of Waterloo
Waterloo, Ontario, Canada.
{collinsk@uwaterloo.ca}

Andrew Hogue, Bill Kapralos
University of Ontario Institute of Technology
Oshawa, Ontario, Canada.
{andrew.hogue, bill.kapralos}@uoit.ca

Abstract—**Prior work focusing on 2D visuals primarily suggests that sound has a significant impact on visual perception. However, little work has considered what, if any, effects sound, including its spatial positioning and the addition of various auditory effects, have on our perception of stereoscopic 3D (S3D) imagery. Here we present the results of two experiments that were conducted to examine the effects of sound and various auditory conditions on stereoscopic 3D depth perception within a virtual (game) environment. Our results reveal that sound can have a significant effect on S3D depth perception. Results suggest that asynchronous audio-visual interactions, the type of sound, and various audio effects can influence distance perception within a virtual environment that incorporates S3D viewing. Our results have implications for game designers, who, with the appropriate use of sound and S3D interactions, can improve the player's experience within stereoscopic 3D-based virtual environments.**

Keywords—spatial (3D) sound; stereoscopic 3D (S3D); depth perception; virtual reality; games;

I. INTRODUCTION

While there is an increasing body of literature that explores stereoscopic 3D (S3D) within virtual environments and games, few studies have investigated the multi-modal interaction of stereoscopic 3D with sound. Evidence from the literature (primarily focusing on 2D visuals) suggests that sound has a significant impact on visual perception, but what effect does sound have on our perception of stereoscopic 3D imagery? Moreover, can sound compensate for some of the shortcomings associated with S3D (e.g. limited disparity ranges)? If so, what implications would this have for the design of interactive S3D virtual environments and games? Could game designers use sound to 'trick' players into thinking an object is further or closer (biasing their perception of depth/disparity)? Can sound be used to overcome or distract from some of the common side effects (e.g., motion sickness, ghosting[1], window violations), inherent with S3D viewing? If so, designers may be provided with the ability to extend content further out of the screen perceptually, thus creating a more comfortable experience for users in situations where they are presented with large stereoscopic disparities (e.g. situations with objects in the near foreground and a distant landscape). Ultimately the goal of this work is to develop user experience guidelines that can assist content designers to create better and more convincing virtual environments and games. This paper describes two user experiments designed to study the interaction between audio-visual cues in a video game environment. Results of these studies suggest that asynchronous audio-visual interactions (offsets of 60.0 ms between the sound and the corresponding visual), certain sounds, and effects (i.e., amplitude attenuation, frequency attenuation, and Doppler shift), can influence a users' depth perception when viewing S3D imagery. Our findings have implications for game designers and developers, who, with the appropriate use of sound can improve the player's experience within S3D-based virtual environments including games.

II. PRIOR WORK

Creating compelling S3D experiences is important for designers and developers of games and virtual environments who wish to take full advantage of the technological capabilities of current displays. There is a rich body of literature regarding S3D and its physiological and psychological effects (e.g., [1,2,3]). However, the psychophysical literature has primarily tried to isolate the effects of stereopsis (i.e., the perception of depth and three-dimensional structure deriving from binocular vision). In contrast, S3D games provide a rich playing field for multi-modal interaction due to their complex dynamic and interactive nature.

Cross-modal effects refer to the impact of one sensory input on another [4], and previous work demonstrates that cross-modal effects can be considerable [5]. Various studies have shown that sound can attract a user's attention, leading to a reduced cognitive processing of visual cues [6]. Sound can improve the detection of visual targets amongst distractions [7], and visual localization can affect cross-modal bias when localizing audio-visual events [8]. For short audio-visual events, we tend to respond to the visual stimulus [9]. Cross-modal interaction has implications for game designers; if the possibilities to enhance the visuals within a virtual environment are limited or constrained, one may consider increasing the quality of the sound channels instead [9]. Consequently, it is important to examine if different sounds, auditory effects, and synchronization settings influence S3D depth perception. Although sound's impact on S3D remains underexplored, some research suggests that auditory distance cues can significantly impact visual depth perception [10]. In

---

[1] Ghosting occurs when imagery intended for one eye appears in the imagery intended for the other eye.

this paper, we examine how different characteristics and the manipulations of sound can affect S3D depth perception. We also examine the effects that the synchronization of the sound stimulus with the visual stimulus has on a user's judgement of depth in S3D.

III. EXPERIMENTAL METHODS

A. Participanmts

34 participants (20 female, 14 male) took part in both experiments. The age of the majority of the participants was between 18 and 20 years (four of the participants were between 30 and 40 years old). 21% of participants had played an S3D video game in the past, while 94% of the participants had watched a stereoscopic 3D movie. All participants had normal or corrected to normal vision and hearing (no formal listening tests were conducted; participants were only asked). Participants were provided with an explanation of the tasks involved and were encouraged to ask questions during the initial practice tasks at the beginning of each experiment. The experiments abided by the University of Waterloo and the University of Ontario Institute of Technology Research Ethics Board review process.

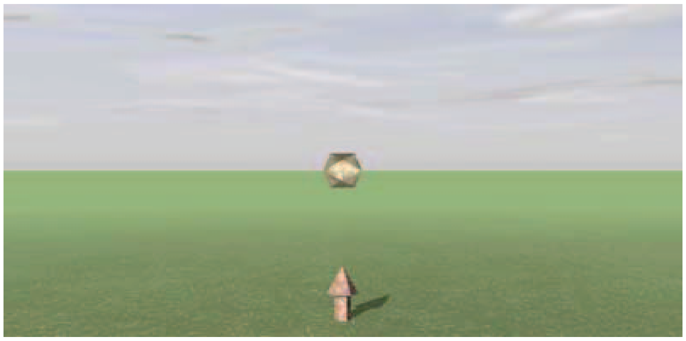
B. Experimental Procedure

In both experiments, the participants were asked to judge how far an approaching audio-visual object was from them under different manipulations of the audio-visual sources. We hypothesized that different sounds and auditory effects (i.e., amplitude attenuation, frequency attenuation, and Doppler shift), combined with different audio-visual synchronization timings will affect how close participants judge the approaching S3D object to be. Moreover, we also hypothesized that a delay in the onset of the sound stimulus (with respect to the visual stimulus), will cause participants to judge the object as appearing further back than when the sound precedes the object (visual stimulus). The experiments were conducted in an acoustically treated room (5.54 m × 5.1 m × 3.0 m). Participants sat in front of a 3D-capable HDTV (LG 47LW5700) positioned two meters directly in front the participant's chair, while playing a specifically designed video game. The display supported passive S3D and as a result, participants were required to wear polarized S3D glasses. The chair was surrounded by a 5.1 surround sound loudspeaker setup comprised of five loudspeakers (JBL LSR 2300 series) and a subwoofer (Polk Audio, RM6750). The subwoofer was located on the floor to the left of the television. The low-pass cut-off on the subwoofer was set to 110 Hz. The loudspeaker placement was designed according to the Polk Audio user manual specifications and the recommendations of [11]. Levels for each loudspeaker were adjusted informally using a white noise burst sent to each loudspeaker. The 5.1 surround and stereo channel sound outputs were calculated by the game engine in relation to a listener position (the virtual camera) and the position and attenuation settings of an audio source in the virtual environment. The Unity3D game engine employs power-normalized *vector-based amplitude panning* (L2-norm VBAP) for 5.1 surround and stereo, resolving only the lateral component for stereo loudspeaker presentations.

IV. GAME ENVIRONMENT

The game environment consisted of a sky with clouds, a flat horizon and a ground plane with a grass texture (see Fig. 1). Precautions were taken to avoid the use of monoscopic depth cues (such as perspective, and relative depth judgments with other objects), in determining the location of the floating object. The object used in the experiments was a regular icosahedron with 20 equilateral triangular faces and was chosen as its numerous sides and edges provided additional stereoscopic depth cues. The S3D visuals were calibrated to create an orthostereoscopic scene, where on-screen imagery acted as an extension of the participants' space and the interaxial distance of the virtual cameras were consistent with the participants' inter-pupillary distances to avoid eyestrain.

Fig. 1. A floating platonic solid with geometric arrow positioned on the ground.



V. AUDITORY STIMULI

Four auditory stimuli were presented: [T] a complex tone, [E] an engine noise, [M] music (Techno music), and [S] silence. These stimuli were chosen for their varying levels of complexity and for their similarity to those used in video games. All of the sounds looped for the duration of each task and followed the exact position of the moving object. The amplitudes of all of the sound stimuli were normalized to have similar perceived loudness levels. The complex tone was comprised of a 160 Hz fundamental and harmonics at 320 Hz, 480 Hz, and 640 Hz. The amplitudes of the harmonics were attenuated in order to unify the sound into an audible whole. More specifically, the amplitude of the 320 Hz and 480 Hz harmonics were attenuated by -2.6 dB -6.7 dB respectively. The engine noise stimulus consisted of an automobile engine sound running at a slow speed. A Techno music sample was chosen because of its consistent rhythm and wide range of spectral content. The controlled variables for manipulating these sound sources were: the presence or absence of S3D, the sound of the approaching object ([T] complex tone, [E] an engine noise, [M] music, [S] silence), Doppler shift (ON/OFF) and frequency attenuation (ON/OFF).

A. Amplitude Attenuation

The amplitude of the sound source in each experiment was attenuated over a set distance from the listener. If the virtual camera/listener moved to a position inside this distance, the amplitude of the sound source was interpolated accordingly, becoming louder as they approach the sound source or quieter as they moved away.

B. *Frequency Attenuation*

When the frequency attenuation effect was activated, the frequency components of the sound source were attenuated based on the sound source’s location and the virtual camera/listener's position. This sound effect attempts to mimic how the higher frequency components of sound are attenuated the further away they are located from a listener in the real world. Our aim was to examine whether frequency attenuation can affect distance perception. We employed a 4th-order Butterworth filter to attenuate sound sources. The Q setting or 'resonance' was set to 1. This setting scales the Qs of the two-cascaded 2nd-order Butterworth sections. The cut-off frequency of the low-pass filter was varied using the linear attenuation graph from 0 Hz to 22 kHz based on the listener's distance from the audio source and a maximum attenuation distance.

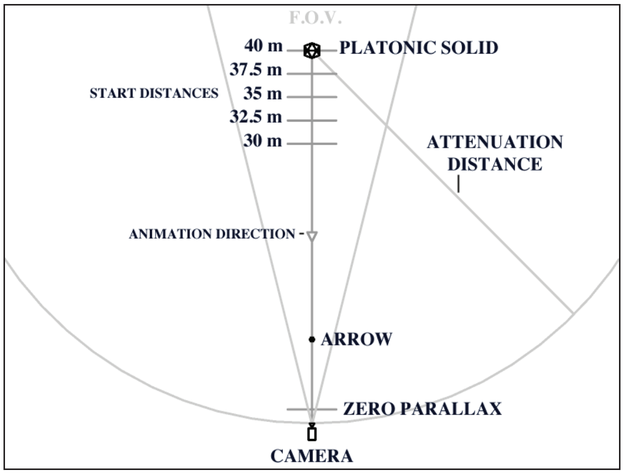
C. *Doppler Shift*

The Doppler shift was implemented as a standard playback-rate change affecting the pitch and duration of the looping sound source. If the Doppler shift effect was activated, the participant was presented with a change in perceived pitch based on the relative velocity between the listener and sound source.

VI. Experiment 1: The Effects of Sound and Sound Effects on Dept Perception

The purpose of this experiment was to examine the effect of different sounds and various auditory effects on depth (distance) perception within a stereoscopic 3D viewing environment. The participants’ task in this experiment was to press a button on a game controller at the moment they believed that an oncoming virtual object (a floating platonic solid) was directly above the virtual arrow (see Fig. 1). The floating platonic solid moved toward the camera/participant, positioned at the origin (0, 0, 0), from one of five starting positions at one of three speeds (see Fig. 2). When audible, the sound of the platonic solid grew louder as it moved toward the camera/participant. In addition, the cut-off frequency of the frequency attenuation effect (when set to ON) interpolated from 0 Hz to 22 kHz. A geometric arrow was located 9 m from the camera in the center of the screen pointing up from the ground. The participants’ distance perception accuracy was attained by recording the distance between the moving platonic solid and the arrow at the moment the participant pressed the “X” button on the game controller. To avoid over-familiarization and to create a more dynamic game-like environment, the moving platonic solid had different radii, start distances, and speeds for each task. The five radii of the platonic solid were: 0.432 m, 0.444 m, 0.456 m, 0.468 m and 0.480 m. The five start distances were: 40.0 m, 37.5 m, 35.0 m, 32.5 m and 30.0 m. The three speeds were: 10 meters per second (mps), 7.5 mps and 5.0 mps.

Fig. 2. Top-down view of game environment of Experiment 1.



The sound's amplitude was attenuated based on a maximum attenuation distance from the camera. To avoid the loudness of the platonic solid's sound at the start of each task being used as a distance cue, the maximum attenuation distance was altered to match each of the five starting distances, always ending at the location of the camera. Consequently, the platonic solid only became audible after a two-second pause at the beginning of each task as it began to move toward the virtual camera/participant. The maximum frequency attenuation distance matched any changes in the maximum amplitude attenuation distance.

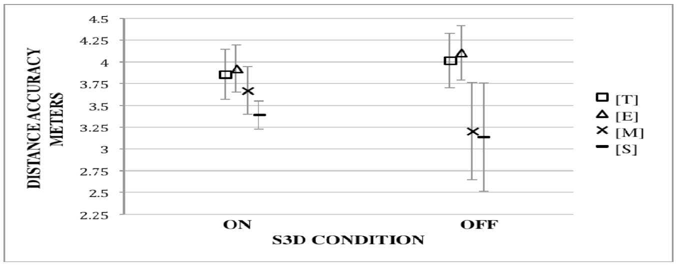
The speed and radius values of the approaching platonic solid were randomly chosen for each task but without repeating any value until all five iterations for radius, or three iterations for speed, were exhausted. This process was repeated for each of the five tasks. The random generator was given a seed value to ensure that each run of the experiment for each participant was identical.

The controlled variables were: the presence or absence of stereoscopic 3D (ON, OFF), the sound of the approaching object ([T] complex tone, [E] engine noise, [M] music, [S] silence), and the presence or absence of the frequency attenuation audio effect that was applied to the sound of the approaching object (ON, OFF). There were a total of 16 permutations. The analysis of the participants’ distance judgements was based on the average results of three run-throughs of the 16 permutations. Two permutations were omitted because the frequency attenuation audio effect had no bearing on the silence parameter.

A. *Results*

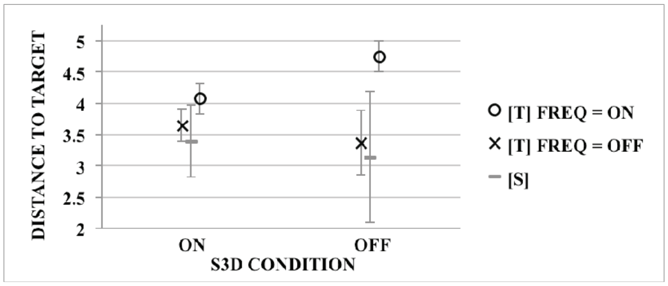
The estimated marginal means in Fig. 3 and 4 are the mean of the difference between the virtual distance to the arrow and the virtual distance to the “object” when the subjects pressed the “X” button in relation to a particular set of auditory conditions. The lower the mean distance accuracy, the more accurate participants were. Hypothetically, if all participants were 100% accurate the mean distance accuracy would equal zero. Participants underestimated the distance to the moving platonic solid, since the majority of distances were recorded after the platonic solid had passed above the arrow. Out of the 714 S3D tasks, only 15 of the distances were recorded when the platonic solid was behind the arrow. This is in agreement with previous studies that suggest human listeners underestimate distances to sound sources that are far away [13]. Although participants also underestimated the distance for the silent tasks, the level of underestimation was not as strong (Fig. 4). Alternatively, the underestimation may have been caused by the S3D settings, delayed reflex times, or a delay in the Unity game engine. Even if we consider that the experiment had inaccuracies in simulating distance, the arrow and the moving platonic solid would have presented an identical amount of disparity for the left and right eyes when at equal distances from the virtual camera. Therefore, the task of identifying equivalent distances between the two game objects based on stereoscopic depth cues remains valid.

Fig. 3. Experiment 1 results: Sound condition vs. accuracy.



Participants were most accurate when the moving object was silent for both 2D and 3D tasks (Fig. 4). For audible tasks, participants were most accurate when the moving object emitted the music sound followed by the complex tone, and then the engine noise for both 2D and 3D (Fig. 4). With respect to S3D set to ON tasks (Fig. 4), participants were significantly more accurate when the approaching platonic solid was silent than when it emitted the complex tone (p = .02) and the engine noise (p = .001). Participants were significantly more accurate when the approaching platonic solid emitted music than when it emitted the engine noise (p = .014).

Fig. 4. Experiment 1 results: S3D condition vs. distance to target.

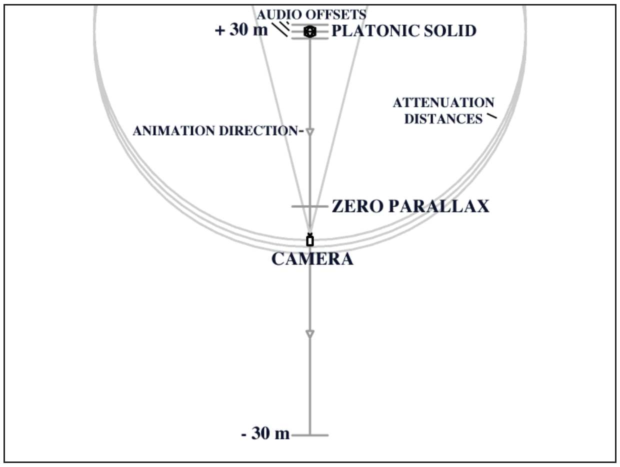


When stereoscopic 3D was set to ON, tasks were analysed, participants were more accurate for each audible sound condition when frequency attenuation was set to OFF, F(2,66) = 5.536, p = .008. Participants were more accurate for the frequency attenuation set to OFF (M = 3.597) tasks than set to ON (M = 4.033), F(1,33) = 26.493, p = .000. For example, participants were more accurate when frequency attenuation was set to OFF when the approaching platonic solid emitted the complex tone (Fig. 5).

VII. EXPERIMENT 2: AUDIO-VISUAL SYNCHRONIZATION

The purpose of this experiment was to examine the effect of audio-visual synchronization on depth perception within a S3D viewing environment. Participants were required to judge the depth of an approaching audio-visual object using a five-point Likert Scale, under various audio-visual synchronization conditions to test whether a delay in the sound (with respect to the visual stimulus) will cause the object to be perceived further back as opposed to when the sound precedes the object. For each task, an object traveled from an initial position 30.0 m in front of the virtual camera to a position 30.0 m behind the virtual camera while emitting sound (see Fig. 5). To avoid over-familiarization and create a more dynamic game-like environment, the moving platonic solid had different radius settings for each task. The five radii considered were: 0.432 m, 0.444 m, 0.456 m, 0.468 m and 0.480 m. A sound source that followed the path of the moving object was either i) synchronized with the moving object; ii) positioned either 0.6 m behind the object corresponding to a -60.0 ms delay; or iii) preceded the object by 0.6 meter corresponding to +60.0 ms in front of the object. As the object passed the zero parallax plane (4.0 m in front of the camera), the object appeared to pop-out of the screen. When the object was behind the camera/participant it was audible through the rear loudspeakers. For each task, the S3D settings remained identical. At the end of each task, the participant judged how close the moving object travelled toward them, even though the final S3D disparity remained identical for all tasks.

Fig. 5. Top-down view of game environment in Experiment 2.

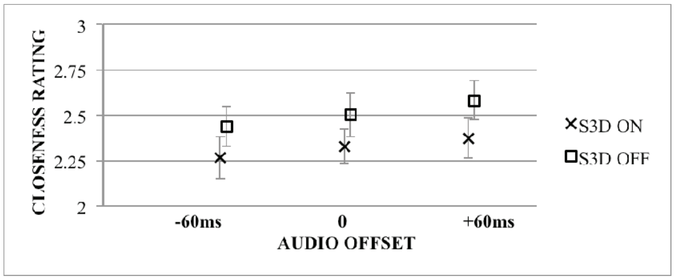


The maximum amplitude attenuation distance was 32.0 m for all tasks. At the beginning of each task the maximum amplitude attenuation distance ended 1.4 m, 2.0 m or 2.6 m behind the location of the virtual camera depending on whether the sound source was positioned behind, at the same location or preceded the object. Consequently, during the 2.0 s pause at the beginning of the animation, the object was audible before it began to move toward the virtual camera/participant. A time delay of 60.0 ms was chosen with respect to the findings of [14]. The three audio offset conditions of the approaching object were [-60.0 ms, 0.0 ms, +60.0 ms]. The radius values of the platonic solid were chosen randomly for each task but without repeating any value until all five iterations had been exhausted. This process was repeated every five tasks. Participants completing the tasks in reverse order, received a different mix of tasks with regard to the radius settings of the moving platonic solid. The controlled variables were: the presence or absence of S3D, the sound offset condition of the approaching visible object (-60.0 ms, 0.0 ms, +60.0 ms), the sound of the approaching object ([T] complex tone, [E] an engine noise, [M] music), the audio effects, Doppler shift (ON, OFF) and frequency attenuation (ON, OFF), which were applied to the sound of the approaching object. There were a total of 36 permutations of the variables. The analysis of participants' ratings was based on the average results of two run-throughs of the 36 permutations.

A. *Results*

Two analyses were conducted on participant closeness ratings for tasks with, 1. S3D set to ON, 2. S3D set to OFF. The sound of approaching object had a significant effect on participant closeness ratings for both analyses (1. F(2, 62) = 6.637, p = .003, 2. F(2, 62) = 3.817, p = .034). Participants judged the approaching object as significantly closer (1. p = .003, 2. p = .005) when the approaching object emitted the music sound (1. *M* = 2.164, 2. *M* = 2.365), than when the approaching object emitted the engine noise, (1. *M* = 2.414, 2. *M* = 2.620). The audio offset also had a significant effect on participant closeness ratings for both analyses (1. F(2,62) = 3.243, p = .048, 2. F(2,62) = 3.834, p = .028). Participants rated the approaching object as significantly closer (1. p = .039, 2. p = .050) when sound was delayed by 0.6 m/-60.0 ms (1. *M* = 2.268, 2. *M* = 2.440) than when the sound preceded the approaching object by 0.6 m/-60.0 ms (1. *M* = 2.375, 2. *M* = 2.583) (Fig. 5).

Fig. 6. Experiment 2 results: Sound offset vs. accuracy.



The sound offset settings maintained a consistent order, in terms of their ability to affect participant ratings, for both S3D and non-S3D tasks. Therefore, the type of audio offset used may have had a universal ability to affect the perceived distance of approaching platonic solid, for both S3D and non-S3D and presentations. The order in which the moving platonic solid was judged to be closer for each sound remained consistent between S3D and non-S3D tasks, where participants judged the approaching platonic solid as appearing closest in the presence of music, followed by the complex tone, and then the engine noise condition. Therefore, the type of sound used may have had a universal ability to affect the perceived distance of objects in gaming environments, for both S3D and non-S3D presentations. In addition, there was a statistically significant difference between the music and engine noise conditions. These results confirm our hypothesis that different audio-visual synchronization timings and different sounds will affect how close participants rate an approaching S3D object. However, results reject our hypothesis that a 60.0 ms delay in the audio will cause participants to rate the S3D object as appearing further back than when the audio precedes the S3D object by 60.0 ms, where in fact the opposite effect was observed.

VIII. DISCUSSION

In this paper we presented the results of two experiments that were conducted to examine the influence of sound on depth perception within a virtual environment that employed S3D viewing. In Experiment 1, the effect of different sounds and auditory effects on depth perception within a S3D viewing environment were considered. In Experiment 2, the effect of audio-visual synchronization on depth perception within a S3D viewing environment was considered.

In Experiment 1, participants were least accurate when the approaching platonic solid was audible and emitted the engine noise, followed by the complex tone and then music. Furthermore, participants were least accurate for each audible sound condition when the frequency attenuation audio effect was set to ON. Participants were most accurate when the approaching platonic solid was silent. These results reject our hypothesis that the addition of auditory cues aid participant distance accuracy perception in stereoscopic 3D environments. However, results show that the addition of sound did influence participant distance judgements. The fact that participants were least accurate for tasks where the approaching object emitted the engine noise implies that the engine noise was most effective at influencing distance perception. The extreme amplitude and frequency attenuation settings used (with respect to everyday listening) may have led to participants having less accurate results, contradicting the findings of Gröhn [15], who found that performance increased in auditory navigation tasks when additional auditory cues relating to the distance and elevation of the target sound source were provided to the participant.

In Experiment 2, and contrary to our hypothesis and the findings of [5], tasks where the approaching object was presented with a delayed sound pushed the visual appearance of the approaching object forward in space (closer to the participant) while the tasks where the sound preceded the approaching object pushed the appearance of the approaching object further back. This effect was observed in participant ratings of closeness for both S3D and non-S3D tasks. If the visual object approached the participant after the sound source had passed into the rear loudspeakers (+60.0 ms) then the visible object was judged to be more distant. On the other hand, if the visible object approached the participant before the sound source had passed into the rear loudspeakers (-60.0 ms), the visual object was judged to be closer. Therefore, closeness ratings may have been decided upon by comparing the visual cue of the object leaving the screen to the auditory cue of the sound passing into the rear speakers.

Overall, the type of sound effect employed impacted the perceived distance of the approaching object for both S3D and non-S3D presentations. Participants judged the approaching object as appearing closest in the presence of music, followed by the complex tone, and then the engine noise condition. In addition, there was a statistically significant difference between the music and engine noise conditions. These results confirm our hypothesis that different audio-visual synchronization timings and different sounds will affect how close participants rate an approaching S3D object. Participants judged the approaching object as closest when it emitted music. It may be the case that representational sounds or sounds that could be perceived as a threat such as the engine noise may be most effective at influencing participant judgements.

IX. CONCLUSION

Here we have demonstrated that different sounds, and/or the inclusion of various auditory effects (e.g., Doppler shift, intensity and frequency attenuation), can impact distance perception in S3D viewing environments. This has implications for designers and developers of virtual environments and games who wish to employ S3D. If S3D depth contributes to feelings of motion sickness in gamers, the appropriate use of sound can be used to increase the sense of S3D depth without implementing extreme S3D settings and thus reduce (if not eliminate) motion sickness. Although further research is required to determine the extent to which the depth bias can be manipulated, combining the methods proposed in our experiments for making the S3D imagery appear closer or participants more accurate might boost the strength of the each method into a range that is significant for gaming.

X. ACKNOWLEDGEMENTS

The financial support of the *Ontario Media Development Corporation* (OMDC)*, Ontario Centres of Excellence* (OCE), *Natural Sciences and Engineering Research Council of Canada* (NSERC), and the *Social Sciences and Humanities Research Council of Canada* (SSHRC) is acknowledged.

REFERENCES

[1] R. Spottiswoode and N. Spottiswoode, “The Theory of Stereoscopic Transmission & Its Application to the Motion Picture,” University of California Press, 1953.

[2] I. P. Howard, “Perceiving in Depth, Volume 1: Basic Mechanisms,” Oxford University Press, 2012.

[3] F. Mikšícek, “Causes of Visual Fatigue and Its Improvements in Stereoscopy,” Technical Report, Department of Computer Science and Engineering, University of West Bohemia in Pilsen, Czech Republic, 2006.

[4] A. Chalmers and K. Debattista, “Levels of realism for serious games,” *IEEE Games and Virtual Worlds for Serious Applications*, Coventry, UK, Pp. 225-232, 2009.

[5] H. McGurk and J. MacDonald, “Hearing lips and seeing voices,” *Nature,* vol. 264*,* pp. 746-748, 1976.

[6] G. Mastoropoulou, et al. “The influence of sound effects on the perceived smoothness of rendered animations,” *2nd Symposium on Applied Perception in Graphics and Visualization*, La Coruña, Spain, pp. 9–15, 2005.

[7] J. Vroomen and B. Gelder, “Sound enhances visual perception: Cross-modal effects of auditory organization on vision,” *J. Exp. Psychol.-Hum. Percept. Perform.*, vol. 26, no. 5, pp. 1583-1590, 2000.

[8] W. D. Hairston, et al, “Visual Localization Ability Influences Cross-Modal Bias,” *J. Cognitive Neurosci.* vol. 15, no.1, pp. 20-29, 2003.

[9] 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,” *1st International Speech Communication Association Research Workshop on Auditory Quality Systems*, Akademie Mont-Cenis, Germany, 2003.

[10] A. Turner, J. Berry and N. Holliman, “Can the perception of depth in stereoscopic images be influenced by 3D sound?,” *SPIE Stereoscopic Displays and Applications XXII*, San Francisco, CA, USA, 2011.

[11] H. Massey (Ed), “The Recording Academy Producers and Engineers Wing: Recommendations for Surround Sound Production,” The National Academy of Recording Arts & Sciences, Inc., 2004.

[12] F. Rumsey, P. Ward, and S. K. Zielinski, “Can playing a computer game affect perception of audio-visual synchrony?" *117th Convention of the Audio Engineering Society*, San Francisco, CA, USA, 2004.

[13] P. Zahorik, D.S. Brungart and A. W. Bronkhorst, “Auditory distance perception in humans: A summary of past and present research,” *Acustica*, vol. 91, no. 3, pp. 409-420, 2005.

[14] J. Vroomen and B. Gelder, “Sound enhances visual perception: Cross-modal effects of auditory organization on vision”, *J. Exp. Psychol.-Hum. Percept. Perform.,* 26(5), 1583-1590, 2000.

[15] M. Gröhn, “Application of spatial sound reproduction in virtual environments – experiments in localization, navigation, and orientation”, Doctoral Thesis, Department of Computer Science and Engineering, Helsinki University of Technology, 2006.