The Effects of Audio on Depth Perception in S3D Games

Brian Cullen[1,2], Daniel Galperin[2], Karen Collins[2], Bill Kapralos[1] and Andrew Hogue[1]

[1]University of Ontario Institute of Technology, Oshawa, Canada

{Bill.Kapralos|Andrew.Hogue}@uoit.ca

[2]University of Waterloo, Waterloo, Canada

{collinsk|bcullen}@uwaterloo.ca

ABSTRACT

Although studies have examined sound localization or stereoscopic perception, few have investigated how these phenomena work together. Studies that examine 2D imagery and sound interaction have highlighted numerous phenomena in the temporal, spatial, and the formal domains of each medium. With the resurgence of interest in stereoscopic 3D (S3D), research into the combined effects of S3D and sound is of importance.

Here we present the results of an experiment that examined the effects of sound on depth perception in relation to S3D video game imagery. Our aim was to answer the question: "can a sound's timbre and/or the addition of distance audio effects influence the user's depth perception accuracy?" Results suggest that depth perception is affected by sound, were sound can distort the apparent depth of audible S3D objects. Results also suggest that audio effects, specifically frequency fall-off over distance effects, can also influence the apparent depth of S3D objects.

Categories and Subject Descriptors

I.3.7 [**Computing Methodologies**]: Computer Graphics—*Three-Dimensional Graphics and Realism.* H.5.1 surround sound **[General]**: Multimedia Information Systems-*Audio input/output*.

General Terms

Experimentation, Human Factors.

Keywords

Stereoscopy, depth perception, video games, 3D, S3D, sound, audio, music, multimodality.

1. INTRODUCTION

Video games have recently seen a resurgence in the use of stereoscopic 3D (S3D) graphics. The drive for more exciting content in cinema, home cinema, and games has pushed the consumer-level market forward into adopting S3D technology into television sets. S3D TVs are now commonplace. However, the content development process has lagged behind the pace of S3D hardware. Furthermore, for game developers the S3D market is small in comparison to the standard market, reducing their incentive and increasing the potential financial risks when considering S3D game development. That being said, game publishers have been actively attempting to convince developers to create S3D content to engage players, providing developers with guidelines and suggestions for how to implement this content. While there is an increasing body of literature that explores the S3D user experience, few have investigated how S3D works in relation to sound. Evidence from the literature (primarily focusing on 2D visuals) suggests that sound has a significant impact on our visual perception but what effects will sound, including its spatial positioning and the addition of audio effects, have on our perception of stereoscopic imagery?

In this paper, we present the results of an experiment that examines the effects of sound on depth perception within an S3D-based game environment. This experiment is part of a larger project that includes a series of studies investigating S3D that aim to define an effective user experience within an S3D video game context, taking into account industry perceptions of S3D. With respect to sound, within this larger project, our goal is to investigate how various sounds, audio effects, audio sync settings, and loudspeaker setups interact with S3D visual perception in video games. In addition, we will suggest novel methods of visualizing information and/or propose novel game mechanics for S3D games based on our findings.

AM '12, September 26 - 28 2012, Corfu, Greece

2. BACKGROUND

2.1 Stereoscopic 3D

Spottiswoode and Spottiswoode [1] define S3D as encompassing techniques that present two views to the user (left and right images) simultaneously presented to the respective eyes to create a sense of depth in an image. This depth perception is accomplished by mimicking how the human visual system works [2]. When the images for the left and right eye are delivered to the corresponding eye in isolation, the brain will fuse these two slightly-offset images and provide the user with the sense of depth.

Parallax refers to the displacement in the two images induced by perspective projection onto an image plane. The amount of displacement is proportional to the distance that the object is in front of our eyes. The displacement between the two images enables our brain to interpret the images as being seen in depth rather than on the 2D screen. *Convergence* refers to the angle that our eyes need to rotate to in order to look at a particular object in S3D. As objects become more distant, our eyes become less converged. As objects come closer to us, our eyes must converge in order to merge both perspectives of the object into one image with depth. This depth is simulated in S3D environments by rotating each of the two virtual cameras on their own axes in order to emphasize a particular point in space. This enables the resulting images to have the appropriate amount of convergence and displacement to ensure that objects appear to have depth, pop out and/or appear to be located at distances behind the screen plane.

2.2 S3D in Games

Among game developers, S3D is still considered to be in its infancy, but the number of S3D-capable titles is gradually increasing. A difficult problem with S3D development lies with the need to make important design decisions prior to and during the development process. S3D adds an extra immersive element to 3D computer-generated worlds. However, discrepancies can become problematic for a variety of reasons, such as retinal rivalry occurring due to specular highlights, edge of screen discrepancies, crosstalk, and accommodation (focus) vs. convergence issues. In the simulation industry, these issues have long been associated with degraded performance, increased workload, and user discomfort [3]. Due to the interactive nature and unpredictable viewing scenarios inherent in games, these issues are even more of a concern than with S3D movies.

There are currently no standard set of design guidelines or a description of how specific design decisions impact the overall user experience in S3D, particularly in relation to S3D and sound interactions. Developers are largely basing their game design strategies on discussions with their peers, trial and error, or using previous experience with the technology—an inefficient and potentially harmful way of developing best practises for the industry and consumers. While developers are experts on designing challenging yet entertaining user experiences, there is a lack of communication between the industry and the scientific knowledge base.

2.3 S3D Research

There is a rich body of scientific literature on S3D and its physiological and psychological effects (e.g., [4], [5]). Much of the existing research was initiated and performed in the flight simulator, virtual reality, and psychology fields of study using isolated experiments that investigated the effects of S3D on humans. Much of the psychophysical literature has tried to isolate the effects of stereopsis (the perception of depth). In contrast, in games, the S3D cues are embedded in rich environments created by traditional 3D graphics. Generalizing results from highly controlled laboratory experiments to the unconstrained complex world of video gaming is not always possible or straightforward. Moreover, since much of the existing research was performed in isolation for specific tasks, it is important to critically explore the possibility that entertainment-focused games may be fundamentally different due to existing design practises and multi-modalities.

2.4 Audio-Visual Interactions

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 [6]. 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 (for example, see [7]). Various studies have examined the perceptual implications of audio-visual cross-modal effects and 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 [8]. For most audio-visual events that are short in duration, we tend to respond to the visual stimulus [9]. The "ventriloquist effect" is a well-known effect where spatially disjoint visual and auditory cues are perceived to be related to a single event [10]. Furthermore, the perception of visual display fidelity can affect the perception of sound quality and vice versa [11]. This cross-modal interaction has implications for designers of multi-modal virtual simulations and serious games. More specifically, as described by Larsson et al. [9], if the possibilities to enhance the visuals within a virtual environment are economically or technically limited, one may consider increasing the quality of the audio channels instead. Although sound's impact on S3D remains underexplored, some research suggests that auditory depth cues can significantly impact visual depth perception cues [12].

2.5 Aims

The lack of research on the interactions of S3D and sound raises many questions. If we know that sound significantly influences the perception of 2D animation (as previously described), we may surmise that similar interactions occur with S3D and sound. Moreover, can the use of sound alleviate some of the issues present traditionally in S3D systems? For example, if our perception of the quality of poor visual cues improves with the inclusion of auditory cues [9], can we use sound to compensate for some of the shortcomings associated with S3D? What implications would this influence have when considering the design of interactive environments to be displayed in S3D? The purpose of our experiment was to examine some of these questions and ultimately create a set of user experience guidelines to assist game developers.

3. METHODS

3.1 Technical Set-Up

The experiment used a 3D-capable HDTV (LG 47LW5700) positioned approximately 2 m directly in front the participant's chair. The experiment used a left and right stereo loudspeaker configuration, although the content they emitted was monaural (mono), since no left or right stereo information was required, and the task was completed before any rear auditory information would be needed. The game used in this experiment was implemented on a MacBook Pro laptop, facing away from the participant to avoid peripheral distraction. A PS3 controller was connected to the laptop for use in each experiment.

The game camera and S3D settings were as follows: interaxial 54 mm; zero parallax 1.5 m; field of view 38.1; aspect ratio 1.78; interlaced 1920 × 1080; toed in camera; -40/-30 to 0 z-animation range; animation duration of 3 to 4 seconds. The S3D settings were designed using the InitionStereoBrain Calculator [13] to limit participant eyestrain. In addition, one of the authors, through repeated informal trail runs, checked the game for uncomfortable levels of parallax.

3.2 Participants

34 participants took part in the experiment. There were 20 female and 14 male participants. The most common age range fell between 18 and 20 years of age at 46% followed by 21 to 23 at 24%. Four participants were between 30 and 40 years old. Most participants (56%) considered themselves casual gamers, playing less than 5 hours per week, and 41% claimed to play more than five. 21% of participants had played an S3D video game before, in comparison to 94% having seen an S3D movie. 18 of the 34

participants wore eyeglasses or prescription lenses at 53%. Participants were required to have normal or corrected to normal vision and normal or corrected to normal hearing. We did not ask participants to report if they used hearing aids, however we did ask potential participates not to participate if they had any known vision of hearing difficulties. The experiment abided by the University of Waterloo and University of Ontario Institute of Technology Research Ethics Review process.

3.3 Task Overview

3.3.1 Visual Stimulus Overview

The game environment consisted of a sky with clouds, a flat horizon and a ground plane with a grass texture attached (Figure 1). There was a geometric pointer located in the center of the screen pointing up from the ground. For each task, the virtual camera was positioned at the origin (0, 0, 0). For each task**,** a floating platonic object travelled toward the participant. The object travelled from set distances 40 m to 30 m away to a point where it visually faded away between 2.5 m and 1.5 m in front of the virtual camera.

Precautions were taken in the design of the experiment to avoid the use of monoscopic depth cues in determining the location of the floating object. The floating object was illuminated using a separate light that followed the floating object at a set distance, prohibiting the use of changes in illumination over distance as an aid in judging depth. In addition, the floating object did not cast a shadow, as a shadow on the same depth plane to the floating object could have been used as a cue also. Conversely, the geometric shapes and detailed textures of both the floating object and the pointer were designed to aid S3D depth perception.

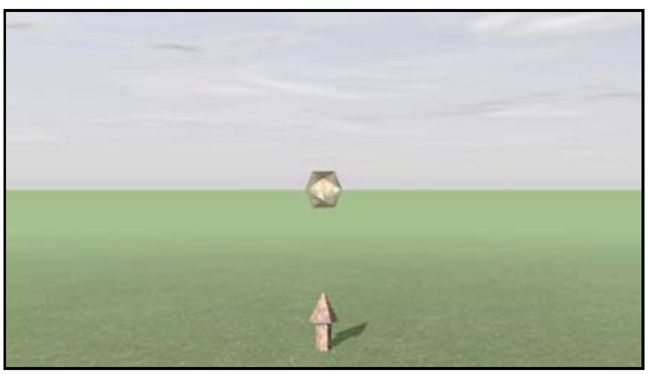

Figure 1. Visual game environment of the experiment

3.3.2 Animation Parameters

To avoid over-familiarization with the task, the moving object had different size, animation speed, and start distance settings for each task. Size here refers to the size of the moving object, while animation speed, refers to how fast the object approached the participant.

Object size was adjusted to alter the size of the moving object. The five sizes, with respect to an initial start size of about 0.8 m, were as follows; 90%, 92.5%, 95%, 97.5% and 100%. The three speed settings are referred to as fast = 100% (3 – 4 s), medium = 75% (4 – 5.33 s), and slow = 0.5 (6 – 8 s). Although each floating object moved at varying speeds, timing buffers were used to give each task the same overall duration. In addition, a two second pause at the beginning of each animation allowed the participants' eyes to adjust between both changes in the 3D scene and between changes between 2D and 3D.

In general, the floating object traveled at set distances; 40 m, 30.75 m, 30.5 m, 30.25 m and 30 m, from the virtual camera. Both speed and size values were randomly chosen for each task but without repeating any value until all five iterations for size, or three iterations for speed, had been exhausted. This process was repeated every five/three tasks. The random generator was given a specific seed value to ensure that each run of the experiment for each participant was identical. However, due to the fact that participants in the odd category had to complete the tasks in a reverse order, they received a different mix of tasks in relation to the size, start distance and speed settings of the moving object. Future iterations of the experiment will treat these variables differently and/or remove them altogether so as to negate their possible influence on results.

3.3.3 Audio Stimuli Parameters

Four auditory conditions ("sound files") were used: [S1] a complex tone, [S2] a sound effect, [S3] a music track, and [S4] silence. These particular conditions were chosen for their varying levels of complexity and due to the regular use of similar sounds in video games.

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. The amplitudes of all of the sound files were altered to have a similar perceived loudness level. The sound effect condition consisted of a looped automobile engine noise. The music track consisted of a techno track sample. All of the sounds seamlessly looped for the duration of each task and followed the exact position of the moving object.

If the frequency fall-off effect was enabled, the sound was passed through a low-pass filter. If the frequency fall-off effect was enabled, the sound was passed through a low-pass filter whose cut-off frequency varied (increased) from 0 Hz to 22 kHz as the object approached the camera/participant. Our aim was not to mimic real world frequency attenuation over distance but to examine if a frequency filter could affect distance perception. Customized frequency fall-off curves were originally designed for each sound file. Unfortunately, a limitation in the Unity 3D Game Engine and time constraints did not allow us to implement this feature in this experiment. Ultimately, each sound file had the same linear frequency fall-off applied to it. This setup may have had little effect on the complex tone sound condition but did have an audible effect on the sound effect and music condition. Spectrum analysis of the sound effect and music sound files revealed that that they had a good spread of spectral content ranging from 10 Hz to about 21 kHz.

The audio amplitude fall-off distance was altered to match each starting distance, therefore ending at the location of the virtual camera. A two second pause at the beginning of each animation allowed the participants' eyes to adjust between changes in the 3D scene for each task. Consequently, during the two-second pause, the object was silent and only became audible as it began to move toward the virtual camera/participant. This altering of the amplitude distance directly affected the frequency fall-off distance.

All of the fall-offs were linear and had short ranges with regard to how they might be setup in a videogame or their real-world equivalents. We chose exaggerated fall-offs over short distances to avoid ambiguous results regarding sound and S3D depth perception.

Since this was a basic experiment that was part of a larger set of preliminary experiments, the linear frequency and amplitude fall-off effects used were those provide by the Unity3D Game Engine.

3.3.4 Task

Participants were given a brief explanation of the task. To avoid undue concern, participants were warned that some tasks might not include S3D and/or audio. Participants were encouraged to ask questions regarding the requirements of the task during the initial ten practises tasks. These practise tasks were omitted from the final data analysis. When the participant was ready to begin, the Unity3D program presented the following text, 'Please press X when the moving object is directly above the arrow'. The participants' depth perception accuracy was measured. This data was attained by recording the distance between a moving object and a pointer at the moment the participant pressed the X button on the PS3 controller.

There were a total of 42 permutations of the variables used in this experiment and 42 tasks associated with each permutation. Six permutations were omitted because the frequency fall-off effect has no bearing on the silence parameter. In addition, the Unity3D game was split into odd and even categories. Participants in the even category completed each task in a fixed but randomly arranged sequence. Participants in the odd category had to complete tasks in a reverse, but still randomized, order. This strategy was implemented to counteract the effects of fatigue that may have distorted results over the duration of the experiment.

3.3.5 Summary of Parameters

In summary, our controlled variables were: the presence or absence of S3D, sound ([S1] complex tone, [S2] sound effect, [S3] music, [S4] silence), and the presence or absence of frequency fall-off.

4. RESULTS

4.1 Analysis Method

The results analyzed in this section are based on the average distance for each task that was recorded for each participant. The results in each table donates an error distance in meters from a target distance of zero, where the moving object would be located directly above the pointer. The following results were implemented with SPSS, using a repeated measures analysis, with a 95% confidence interval. In addition, the Bonferroni confidence level adjustment was applied to the estimated means. All repeated measures main effects and interactions were statistically assessed for their adherence to the assumption of sphericity. Where the sphericity assumption was violated, a Greenhouse-Geisser correction was applied to the relevant main effect or interaction.

4.2 S3D, Sound File and Frequency Fall-off

We conducted repeated measures tests that examined the results for the following variables S3D (ON, OFF), sound file (complex tone, sound effect, music), and frequency fall-off (ON, OFF). It should be noted that a fourth silence condition, which we will examine later, was not included for this analysis as it had no relevance in relation to the frequency fall-off conditions.

For the within-subjects effects analysis, the main effect of sound file was significant, $F(2,66) = 9.029$, $p < .001$. The main effect of frequency fall-off was significant $F(1,33) = 37.831$, $p < .000$. However, these effects were qualified by a significant sound file by frequency fall-off interaction, $F(2,66) = 3.629$, $p = .037$, and a significant 3-way interaction between S3D, sound file and frequency fall-off, $F(2,66) = 7.364$, $p < .003$.

When results were analysed in relation to S3D set to ON only, the main effect of frequency fall-off, $F(1,33) = 26.493$, $p < .000$ was significant. This effect was qualified by a sound file by frequency fall-off interaction, $F(2,66) = 5.536$, $p = .008$.

When results were analysed in relation to S3D set to OFF only, the main effect of sound file, $F(2,66) = 5.808$, $p = .007$, and of frequency fall-off, $F(1,33) = 16.426$, $p < .000$ was significant. These effects were qualified by a sound file by frequency fall-off interaction, $F(2,66) = 5.752$, $p < .008$.

Due to a concern relating to frequency fall-off and the ability of the filter's effects to be audible in relation to complex tone, additional analyses that excluded complex tone tasks were conducted. Similar significant main effects and interactions were observed, although for the analysis in relation to S3D set to OFF the interaction between sound file and frequency fall-off was not significant. Therefore, even when of complex tone was excluded from our analysis, frequency fall-off exhibited a significant main effect.

We will now return to our repeated measures analysis that includes the complex tone condition. For the estimated marginal means analysis in relation to S3D, the S3D set to OFF (M = 3.763, 95% CI [3.10, 4.43]) tasks had a slightly lower average mean accuracy than the ON tasks (M = 3.815, 95% CI [3.31, 4.31]. In other words, when S3D was set to OFF participants were more accurate on average at judging the distance between the moving object and the pointer.

Table 1. S3D and Sound File

S3D	SOUND	Mean	Lower Bound	Upper Bound
ON & OFF	S1	3.956	3.392	4.520
	S2	4.002	3.443	4.561
	S3	3.409	2.773	4.045
ON	S1	3.857	3.362	4.352
	S2	3.918	3.391	4.445
	S3	3.669	3.146	4.191
OFF	S1	4.055	3.340	4.769
	S2	4.086	3.456	4.715
	S3	3.149	2.240	4.058

Table 1 shows the estimated marginal means for the interaction between S3D and each sound file (excluding silence). It should be noted that the mean values for each sound file incorporate both frequency fall-off conditions. When 2D and S3D tasks are considered together, the music condition resulted in the lowest, and therefore most accurate average results, followed by complex tone, then sound effect (Table 1). Pairwise comparisons analysis of sound file revealed that the music condition had a significant mean difference when compared to complex tone ($p = .004$) and sound effect ($p < .008$). The same significant mean difference was evident with S3D set to OFF tasks ($p = .033$, $p < .034$). However, when we conducted the same analysis, but in relation to S3D set to ON only, the music condition had a significant mean difference when compared to sound effect ($p = .006$) but not complex tone.

The estimated marginal means for each sound file, for S3D set to ON and OFF in isolation followed the same pattern of accuracy (Table 1) where the music condition had the lowest mean average, followed by complex tone, then sound effect. We will examine this hierarchy in relation to both frequency conditions later.

For the estimated marginal means analysis in relation to frequency fall-off for S3D (ON, OFF) including all sound file conditions, frequency fall-off set to OFF (M = 3.491) had a lower estimated mean, or more accurate average score, than frequency fall-off set to ON (M = 4.087). A pairwise comparisons analysis revealed a significance mean difference between both conditions ($p < .000$).

When we analysed the results in relation to S3D set to ON only, frequency fall-off set to OFF (M = 3.597) had a lower mean average than ON (M = 4.033). In addition, frequency fall-off set to ON and OFF had a significant mean difference ($p < .000$).

When we analysed the results in relation to S3D set to OFF only, frequency fall-off set to OFF (M = 3.385) had a lower mean average than ON (M = 4.141). In addition, frequency fall-off set to ON and OFF had a significant mean difference ($p < .000$).

Table 2. S3D, Sound File and Frequency Fall-off

S3D	SOUND	FREQ	Mean	Lower Bound	Upper Bound
ON & OFF	S1	ON	4.407	3.864	4.951
		OFF	3.504	2.884	4.125
	S2	ON	4.200	3.655	4.745
		OFF	3.804	3.194	4.414
	S3	ON	3.654	3.031	4.277
		OFF	3.164	2.475	3.852
ON	S1	ON	4.073	3.522	4.625
		OFF	3.640	3.132	4.149
	S2	ON	3.948	3.372	4.523
		OFF	3.889	3.368	4.409
	S3	ON	4.077	3.564	4.590
		OFF	3.261	2.700	3.821
OFF	S1	ON	4.741	4.115	5.367
		OFF	3.368	2.498	4.239
	S2	ON	4.452	3.829	5.074
		OFF	3.720	2.954	4.486
	S3	ON	3.231	2.310	4.152
		OFF	3.067	2.080	4.054

Table 2 shows the estimated marginal means for the interaction between S3D, sound file and frequency fall-off. When results for S3D set to ON and OFF tasks are considered together and in isolation, frequency fall-off set to OFF resulted in the most accurate average results for all three sound file conditions and in the following order, music, complex tone and sound effect (see Table 2).

For S3D set to ON tasks, with frequency set to ON the order of the sound file in relation to accuracy was reversed as follows, sound effect, complex tone, then music. The complex tone and music condition had similar means (M = 4.073 and M = 4.077).

For S3D set to OFF only tasks, with frequency fall-off set to ON the order of participant accuracy was as follows, music, sound effect, and complex tone. This does not conform to any of the previous patterns.

In Figure 2, the vertical axis shows the participants' estimated mean accuracy for complex tone tasks in meters. The horizontal axis shows where S3D set to ON and OFF results are located. Estimated mean data for silence [S4] was added from an analysis of the interaction between sound file and S3D for comparison proposes.

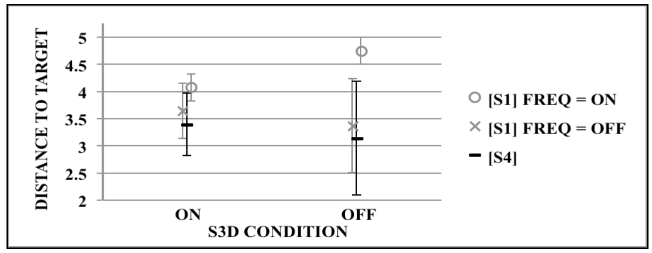

Figure 2. S3D, Complex Tone and Frequency Fall-off

For complex tone, frequency fall-off set to ON tasks had the least accurate participant results especially for S3D off tasks (Figure 2). However, due to a design oversight in the experiment, the frequency fall-off filter had little audible effect on the complex tone sound file. Therefore, the lack of accuracy for the S3D on ON and OFF tasks may have been due to chance or, in relation to the OFF condition, to a lack visual depth cues for tasks using this condition. In general, silence tasks had better results than complex tone tasks.

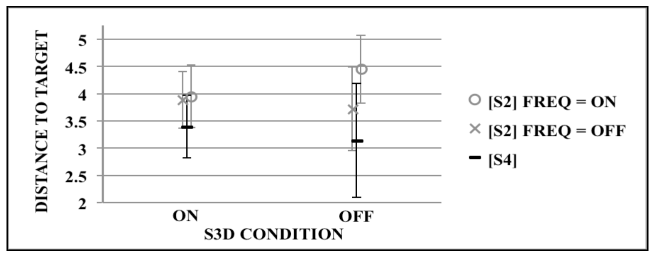

Figure 3. S3D, Sound File and Frequency Fall-off

Figure 3 shows the interaction between S3D, sound effect and frequency fall-off. Frequency fall-off set to ON tasks had the least accurate participant results especially for S3D set to OFF tasks (Figure 3). However, the lack of accuracy for the S3D OFF tasks is likely due to chance or to a lack of visual depth cues using this condition. In addition, frequency fall-off set to ON did not seem to have a strong effect on accuracy for S3D set to ON tasks. Once again, silence had more accurate average results.

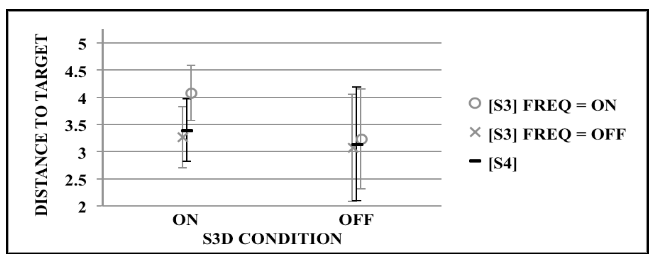

Figure 4. S3D, Music and Frequency Fall-off

The interaction between S3D, music and frequency fall-off was almost opposite to complex tone and sound effect, where frequency fall-off set to ON tasks had the more accurate participant results for S3D OFF tasks (Figure 4). Although due to the limited number of depth cues available in relation to the S3D OFF condition this difference may not be of significance. Unlike the sound effect condition, frequency fall-off set to ON seemed to have a strong effect on accuracy in relation to the music condition. Unlike the other sound file conditions, music resulted in a more accurate average score than silence, but only for frequency fall-off set to OFF tasks. However, similar to the sound effect file and complex tone conditions, music followed a similar pattern of accuracy as silence for S3D set to ON and OFF tasks when no frequency fall-off filter was applied.

4.3 S3D and Sound File (including Silence)

We conducted repeated measures tests that examined S3D (ON, OFF) and sound file (complex tone, sound effect, music, silence). We must keep in mind that the mean results here for the audible sound file conditions incorporate frequency fall-off set to ON and OFF conditions. Therefore the results should only be understood to give a general impression of how each audible sound condition relates to silence.

The main effect of sound file was significant $F(3,99) = 7.233$, $p = .001$. S3D and the interaction between S3D and sound file did not exhibit a significant effect. Although there was no significant main effect for S3D, when frequency fall-off was included in a repeated measures test that excluded the sound condition there was a significant 3-way interaction between S3D, sound file and frequency fall-off.

Returning to our repeated measures analysis of S3D and sound file, when results were analysed in relation to S3D set to ON only, the main effect of sound file was significant, $F(3,99) = 7.854$, $p < .000$.

When results were analysed in relation to S3D set to OFF only, the main effect of sound file was significant $F(3,99) = 4.607$, $p = .011$.

Table 3. S3D and Sound File (including Silence [S4])

S3D	SOUND	MEAN	Lower Bound	Upper Bound
ON & OFF	S1	3.937	3.371	4.503
	S2	4.014	3.458	4.569
	S3	3.439	2.800	4.078
	S4	3.263	2.486	4.041
ON	S1	3.857	3.361	4.353
	S2	3.920	3.394	4.447
	S3	3.672	3.150	4.194
	S4	3.390	2.818	3.963
OFF	S1	4.017	3.302	4.731
	S2	4.107	3.483	4.731
	S3	3.206	2.288	4.125
	S4	3.137	2.088	4.185

Table 3 shows the estimated marginal means for the interaction between S3D (ON, OFF) and each sound file (including silence [S4]). When S3D set to ON and OFF results were considered together, silence exhibited in the best average results, followed by music, complex tone, then sound effect (Table 3). The same pattern was evident for S3D set to ON and OFF in isolation. However we should remain aware that music attained a better average score than silence when the frequency fall-off filter was set to OFF (Figure 4)

For the following pairwise comparisons analyses we should keep in mind that the significant previously observed interaction effects with S3D and frequency fall-off have not been considered, as we cannot analyze silence in relation to frequency fall-off. For the pairwise comparisons analyses of sound file, with S3D ON and OFF considered together, the silence condition had a significant mean difference in relation to complex tone ($p = .024$) and sound effect ($p < .039$). The music condition also had a significant mean difference when compared to complex tone ($p = .029$) and sound effect ($p = .024$). The silence and music conditions did not exhibit a significant mean difference when compared to each other. In addition, the complex tone and sound effect conditions did not have a significant mean difference.

For the pairwise comparisons analysis of sound file, with S3D set to ON only, the silence condition retained a significant mean difference when compared to sound effect and complex tone. However the music condition had a significant mean difference to sound effect only. For the pairwise comparisons analyses of sound file, with S3D set to OFF only, the silence condition displayed a significant mean difference when compared to complex tone only, and music had a significant mean difference when compared to sound effect only.

In general the silence and music conditions had a similar effect on results, where both attained more accurate results, than sound effect and complex tone. Although in some instances the music condition had a similar effect on results as complex tone (Figures 2 and 4) where for frequency fall-off set to ON and S3D set to ON they attained similar mean averages (complex tone $M = 4.073$, music $M = 4.077$).

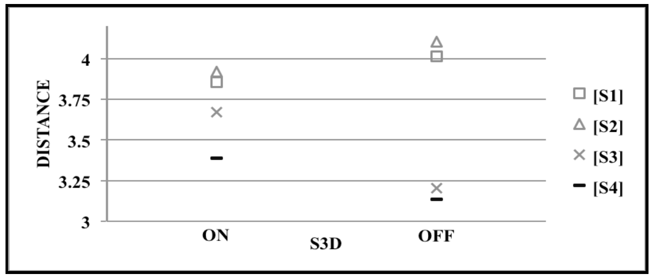

Figure 5. S3D and Sound File (including Silence [S4])

In Figure 5, the vertical axis shows the participant's estimated mean accuracy, the horizontal axis shows where S3D ON and OFF results are located. However, it should be noted that the results for each sound file condition consists of a mean average that incorporates both frequency fall-off conditions.

Results for the complex tone and sound effect condition were lower, or more accurate, for the S3D ON condition. Unlike complex tone and sound effect, the results for the music condition were more accurate for S3D off. In general, the music condition tasks were more accurate than the other two audible conditions. The silence condition had the most accurate results imply that sound, especially the complex tone and sound effect conditions, distorted participant accuracy for both S3D ON and OFF tasks. For the S3D OFF tasks, where there was a limited number of monoscopic depth cues, sound may have had on an affect on the participants' accuracy when guessing distance.

4.4 Summary of Results

4.4.1 Impact of S3D

S3D exhibit a significant effect on results when analyzed in relation to its interaction with sound file and frequency fall-off. S3D set to OFF has a slightly lower average mean accuracy than S3D set to ON. In relation to the impact sound seems to have on participant accuracy, the lower mean for S3D set to OFF might suggest that sound can have a stronger influence on distance perception when imagery is presented in S3D. Further investigation is needed to understand the methods that participants use to judge depth for 2D and S3D with respect to audio-visual interactions. The lower mean value for S3D set to OFF could be due to chance, but may also suggest that participants were just as good at judging depth in 2D tasks and may have used factors such as sound, frequency fall-off, monoscopic cues, even familiarization with the S3D tasks they had experienced, to approximate distance.

For the estimated marginal means in relation to individual sound files (excluding silence), when 2D and S3D tasks are considered together, music resulted in the best results; followed by complex tone, then sound effect. The same pattern was evident when 2D and S3D tasks were considered in isolation. The recurring pattern in both S3D conditions suggests that sound seemed to have a consistent influence in determining depth irrespective of the S3D setting.

4.4.2 Impact of Sound File

Sound file and the interaction between sound file and frequency fall-off, exhibited statistically significant effects for S3D, both ON and OFF together (Table 1), and in isolation.

For the estimated marginal means in relation to individual sound files (excluding silence), when 2D and S3D tasks are considered together and in isolation, the music condition resulted in the most accurate results, followed by complex tone, then sound effect. Further investigation is needed to understand why music tasks achieved the most accurate average results, and the why sound effect tasks had least accurate results. It should be noted that attaining the least accurate results might be a sign that the sound effect file was the most effective at influencing depth perception including S3D depth perception. It is especially relevant in relation to the fact that silence had the lowest mean score out of all the analyses performed. From this perspective, the music condition may have been viewed as more distinct from the object and therefore had less of an effect on participant accuracy. In terms of direct embodied threat, the sound effect condition, similar to that of an approaching automobile, might be perceived to be the most important/urgent of the sound file conditions.

The pairwise comparisons analysis showed that in general music had a significant mean difference to sound effect and complex tone for S3D OFF tasks and a significant mean difference to sound effect for S3D ON tasks. Further investigation is required to explain why a particular sound file has a particular influence on participant accuracy.

4.4.3 Impact of Frequency Fall-off

Although frequency fall-off exhibited a significant main effect on results, a concern was that the effect was not as audible in relation to the complex tone condition. However, when additional tests were run which excluded this sound file condition, frequency fall-off continued to exhibit a significant main effect for all S3D conditions.

In our main analysis (including complex tone), frequency fall-off, and the interaction between sound file and frequency fall-off, had a significant effect for S3D set to ON and OFF results combined, and for S3D set to ON and OFF results in isolation. For the estimated marginal means analysis in relation to frequency fall-off, frequency fall-off had a better average score when set to OFF, for both S3D set to ON and OFF considered together and in isolation. Pairwise comparisons analyses revealed a significance mean difference between both frequency fall-off conditions for all S3D conditions. When results for S3D ON and OFF tasks are considered together and in isolation, frequency fall-off set to OFF resulted in the most accurate results for all of the three sound files conditions (Table 4) and (Table 5). When frequency fall-off was set to ON the order in which each individual sound file had better results became erratic, this may suggest that frequency fall-off set to ON caused confusion amongst participants when judging depth.

Although frequency fall-off exhibited a significant effect, it was unclear from our experiment if the addition of a frequency fall-off effect can strengthen or weaken a sound's influence on the participants' experience of visual depth, for S3D and 2D conditions. A redesign of this aspect of the experiment is needed to examine in detail if frequency fall-off filters can affect the perception of depth for particular types of sound file in a manner unique to each sound file. It would be of worth examining if less exaggerated fall-off distances and cut-off ranges also exhibit an effect.

4.4.4 Impact of Silence.

For repeated measures analysis examining S3D (ON, OFF), sound file (complex tone, sound effect, music, silence), the sound file condition displayed a significant main effect. The within-subjects effects analysis shows significant effects for the sound file condition (including silence). When S3D set to ON and OFF results were considered together, the silence condition scored the most accurate results, followed by music, complex tone, then sound effect. The same pattern was evident when S3D was set to ON and OFF. For the pairwise comparisons analyses for S3D ON and OFF together, and for ON only, silence had a significant mean difference when compared to sound effect and complex tone but not when compared to music. In fact, when we compared music tasks to the silence tasks in relation to frequency fall-off, it became evident that music had more accurate results when the frequency fall-off filter was set to OFF (Figure 4). The better mean accuracy scores associate with silence further suggests that sound has a direct influence on depth perception for both S3D and 2D tasks, although the music file with the frequency fall-off effect included seemed to contradict this pattern. Further investigation is needed to understand why music and silence tasks resulted in more accurate average results.

4.4.5 Possible Impact of Animation Size and Speed

We used a pivot table analysis in Excel to check if animation speed and scale had an effect on results. Generally speaking, the faster the speed of the moving object, the less accurate the results. Moreover, the larger the size of the moving object, the more accurate the results. It should be noted, however, that size was linked to start distance due to a programming error that attributed them with the same random seed value. Due to an uneven dispersion of these variables in relation to our main independent variables we were unable to include these speed and scale in our repeated measures analyses. Without further investigation, it is impossible to determine whether the longer time afforded by the longer start distance or the distortions in perception caused by variations in size, or a combination of both, lead to more accurate results. The distribution of scale and speed setting in relation to the other variables may have had an influence on our statistical analysis. It is necessary to rule out the influence of these parameters by omitting them in future iterations of this experiment.

5. DISCUSSION

In general, our participants seemed to underestimate the depth of an object, since the vast majority of distances were recorded when the object had already passed above the arrow. For example, for the 714 S3D set to ON tasks, only 15 of the distances were recorded when the object was behind the pointer. This is in agreement with previous studies on auditory distance perception [14] that suggest that human listeners systematically underestimate distances to faraway sound sources. Alternatively, this may have been caused by the S3D settings or delayed reflex times. Further research is needed to determine the cause.

Participants were generally inaccurate at judging depth in the S3D set to ON tasks. 436 out of 714 S3D set to ON tasks had inaccuracies greater than three virtual meters. Tasks involving silence resulted in the best average accuracy when S3D was set to ON and OFF. In relation to our findings, sound seems to have an effect on depth perception accuracy in S3D gaming. The type of sound used, and whether or not a frequency fall-off filter was enabled, also seemed to have an effect on accuracy.

Although the least accurate results occurred when sound was enabled, further research is required to ascertain if this is always the case, or if our particular sound settings or selections led to these results. It would be of value to examine if less exaggerated, more consistent amplitude and frequency fall-offs, which do not dynamically change distance, can aid depth perception accuracy. In addition, the inclusion of less abstract objects that do not switch size or speed, such as computer graphic models of cars or mobile phones paired with the types of sounds they are associated with in everyday living, would also be worth investigating. In terms of the S3D settings, it would be preferable to begin with exact orthostereoscopic settings that take into account viewer distance from screen, screen size, the game camera's focal length and depth of field, and so on. These visual settings in combination with naturalistic audio cues would help to rule out the distortion effects that may come about through the use of exaggerated representations of sound, image and/or depth. They would also function as a more neutral starting position for such research.

When considering the interaction of sound and S3D under more dynamic conditions than those examined by [12], the interactions between both modalities become more complicated. Unlike the experimental setup used in [12], when the loudspeaker positions are fixed and participants interact with moving virtual sound sources and S3D objects, the inclusion of sound can be detrimental to accurate S3D depth judgement. The results suggest that there may be a complex ecology of audio-visual interrelationships regarding how we interpret and interact with S3D imagery.

6. CONCLUSIONS

Although these results are preliminary, in the sense that our sample of participants and sound parameters was small, the experiment illustrates that sound and the inclusion audio effects, specifically frequency fall-off over distance effects, does impact our perception of depth in S3D environments. Until further work is completed, we are unable to make specific recommendations to game developers based on these results. However, we do suggest that any quality assurance testing undertaken by developers of S3D games incorporates both audible and silent game testing, including the testing of various distance related effects such as the frequency fall-off effect.

If, as our results suggest, sound can impact the sense of depth that a player experiences, and if S3D depth contributes to feelings of motion sickness in gamers, sound could be used to increase the sense of depth without implementing extreme S3D settings. In this way, we may be able to reduce the actual S3D depth and compensate for this reduction through the use of sound. However, further research is required to develop a proper understanding of the way that sound may affect player depth judgement accuracy. Such research could provide a definitive list of the combinations of audio and S3D interactions that work best. In other words, such findings could be used to improve design techniques, and ultimately, the player's experience of S3D gaming.

7. ACKNOWLEDGMENTS

Funding for this project was provided by the Ontario Media Development Corporation, Ontario Centres of Excellence and the Social Sciences and Humanities Research Council of Canada.

8. REFERENCES

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

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

[3] Mikšícek, F. 2006. Causes of Visual Fatigue and Its Improvements in Stereoscopy. *Engineering*.

[4] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross. 2010. Nonlinear disparity mapping for stereoscopic 3D. In *ACM SIGGRAPH 2010 papers,* Hugues Hoppe (Ed.). Article 75, DOI=10.1145/1833349.1778812

[5] H-S. Lin, S-H. Guan, C-T. Lee, and M. Ouhyoung. 2011. Stereoscopic 3D experience optimization using cropping and warping. In *SIGGRAPH Asia 2011 Sketches* (SA '11). Article 40. DOI=10.1145/2077378.2077428

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

[7] H. McGurk, and J. MacDonald. 1976. Hearing lips and seeing voices. *Nature,* 264*,* 746-748.

[8] G. Mastoropoulou, K. Debattista, A. Chalmers, and T. Troscianco. 2005. The influence of sound effects on the perceived smoothness of rendered animations. In *Proc. 2nd Symposium on Applied Perception in Graphics and Visualization*, La Coruña, Spain, August 26 - 28, 9–15.

[9] P. Larsson, D. Västjäll, and M. Kleiner. 2003. On the quality of experience: A multi-modal approach to perceptual ego-motion and sensed presence in virtual environments. In *Proc. First International Speech Communications Association Tutorial and Research Workshop on Auditory Quality of Systems*, Akademie Mont-Cenis, Germany.

[10] P. Bertelson, and Radeau. 1981. Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept. Psychophys*, 29,6, 578-584.

[11] S. L. Storms, and M. J. Zyda. 2000. Interactions in perceived quality of auditory-visual displays. *Presence*: *Teleoperators and Virtual Environments*, 9, 6, 557-580.

[12] Turner, A., Berry, J., & Holliman, N. 2011. Can the perception of depth in stereoscopic images be influenced by 3D sound? *Displays*, *7863* (February).

[13] Inition StereoBrain Calculator from Inition. (n.d.). Retrieved April 28, 2012, from http://www.inition.co.uk/3D-Technologies/inition-stereobrain-calculator

[14] Zahorik, P., Brungart, D. S., and Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present research. *Acta Acustica United With Acustica*, *91*(3), 409-420. S. Hirzel Verlag.