The Effects of 5.1 Sound Presentations on the Perception of Stereoscopic Imagery in Video Games

Brian Cullen[a,b], Daniel Galperin[a], Karen Collins[a], Andrew Hogue[b], Bill Kapralos[b]
[a]University of Waterloo, Waterloo, Canada; [b] University of Ontario Institute of Technology, Oshawa, Canada

ABSTRACT

Stereoscopic 3D (S3D) content in games, film and other audio-visual media has been steadily increasing over the past number of years. However, there are still open, fundamental questions regarding its implementation, particularly as it relates to a multi-modal experience that involves sound and haptics. Research has shown that sound has considerable impact on our perception of 2D phenomena, but very little research has considered how sound may influence stereoscopic 3D. Here we present the results of an experiment that examined the effects of 5.1 surround sound (5.1) and stereo loudspeaker setups on depth perception in relation to S3D imagery within a video game environment. Our aim was to answer the question: "can 5.1 surround sound enhance the participant's perception of depth in the stereoscopic field when compared to traditional stereo sound presentations?" In addition, our study examined how the presence or absence of Doppler frequency shift and frequency fall-off audio effects can also influence depth judgment under these conditions. Results suggest that 5.1 surround sound presentations enhance the apparent depth of stereoscopic imagery when compared to stereo presentations. Results also suggest that the addition of audio effects such as Doppler shift and frequency fall-off filters can influence the apparent depth of S3D objects.

Keywords: Stereoscopic, perception, video games, S3D, surround sound

1. INTRODUCTION

Video games have recently seen a rise 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 and 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 film and television 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 studies have investigated the relationship between S3D and 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 various audio effects such as Doppler frequency shift and frequency fall-off, have on our perception of stereoscopic imagery?

In this paper, we present the results of an experiment that was conducted to examine the effects of loudspeaker setups (stereo and 5.1 surround sound (5.1)), and audio effects 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.

Stereoscopic Displays and Applications XXIV, edited by Andrew J. Woods, Nicolas S. Holliman, Gregg E. Favalora,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8648, 864815 · © 2013 SPIE-IS&T
CCC code: 0277-786X/13/$18 · doi: 10.1117/12.2003136

Downloaded From: http://spiedigitallibrary.org/ on 11/05/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

2. BACKGROUND

2.1 Stereoscopic 3D

Spottiswoode, in [1], defines 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 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 a greater 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 experiences with the technology—an inefficient and potentially harmful way of developing best practices 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 practices 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 available to 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

Downloaded From: http://spiedigitallibrary.org/ on 11/05/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

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 regarding the interaction 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 spatial 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 is to examine some of these questions and ultimately create a set of user experience guidelines to assist game developers who are interested in developing S3D games. This experiment examines whether different sounds, audio effects and loudspeaker settings can be used to manipulate subjective depth judgements by asking participants to rate how close they believed an approaching S3D object appeared under various sound conditions.

3. METHODS

3.1 Technical Set-Up

Visuals for the experiment were presented with a 47" 3D capable HDTV (LG 47LW5700) positioned two meters directly in front the participant's chair. 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). The subwoofer was located on the floor to the left of the television. The left and right rear surround speakers were set to 1.37 m from the chair and although participants were instructed not to move backwards/forwards, it was observed that several of them did lean forward while carrying out their task. The left front and right front loudspeakers were set 2.16 m from the chair. In addition to the S3D visuals, participants were also presented with a variety of sound stimuli (a complex tone, an engine noise, and music), which were output with either a 5.1 surround sound or stereo loudspeaker configuration. Loudspeaker loudness levels were calibrated informally by the experimenter. The game used in this experiment was implemented and ran on an Apple 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.

3.2 S3D Parameters

A summary of the game camera and S3D settings for each experiment is provided in Table 1. The S3D settings were designed using the InitionStereoBrain Calculator [13] to limit participant eyestrain. The S3D visuals were calibrated with respect to creating an orthostereoscopic-like scene, where on-screen imagery acted as an extension of the participants' space. Consequently, the television screen appeared as a window into the world of the game environment used for this study. This technique provides a more natural stereoscopic experience and was mainly employed to avoid eyestrain. However, settings were altered from an exact orthostereoscopic setting to meet the requirements of the experiment. For example, the zero parallax was pushed back to 4 m behind the cameras to create a "pop out" effect. In general, S3D games and films treat such parameters in a similar manner, where such changes are a normal part of the stereographer's decision-making process.

For the experiment, the horizontal field of view of the virtual cameras was set to 28.1º, and the aspect ratio of the cameras was set to 1.78 to match that of the HDTV display (16.9). In addition, the camera was converged slightly toward the point/distance of zero parallax (this is often referred to as 'toed in'). The display (HDTV) supported passive (interlaced-based) S3D and as a result, participants were required to wear a pair of polarizing glasses.

Downloaded From: http://spiedigitallibrary.org/ on 11/05/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

Table 1. Game cameras and S3D settings

INTERAXIAL	ZERO PARALLAX	Z ANIMATION	ANIMATION DURATION
65 mm	4.0 m	-30 to +30 m	5 to 10 secs (variable)

The INTERAXIAL distance refers to the distance between the left and right virtual camera positions on the horizontal plane. The ZERO PARALLAX value refers to the distance from the virtual cameras where there is no displacement in the images. In general, visual displacement between the left and right eye increased as the object moved toward or away from this position. Z-ANIMATION refers to the distances travelled (in meters) by the moving objects along the Z axis, toward the camera/participant. ANIMATION DURATION refers to the amount of time it took for each animation of a particular speed setting where applicable.

3.3 Participants

Before conducting the experiment, participants were asked to complete a questionnaire that included demographic information such as age, sex, and information pertaining to video gameplay habits (how often they play video games, etc.). 34 participants (20 female, 14 male) took part in the experiments although the first three participants were omitted from our analysis because we implemented design changes based upon receiving some initial feedback from these 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. The experiments included the presentation of both sound and visual stimuli and therefore, we required that all participants had normal or corrected to normal vision and normal or corrected to normal hearing. The experiment abided by the University of Waterloo and the University of Ontario Institute of Technology Research Ethics Board process.

Participants were seated in a chair placed in front of a monitor (display) and were given a brief explanation of the experiment. Participants were encouraged to ask questions regarding the requirements of each task during the initial practises tasks for the experiment. These practise tasks were omitted from the final data analysis. Participants were divided into two categories, "odd" and "even". Participants in the "even" category completed all of the tasks in the experiment in a fixed but randomly arranged sequence. Participants in the "odd" category had to complete all of the 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.

3.4 Task Overview

3.4.1 Task

Participants were advised to go with their "intuition" if they had difficulty noticing the changes in depth. When the participant was ready to begin, the game presented the following text, 'Please rate how close the moving object appears to get from 1 to 5; where 1 is close and 5 is distant'. Recall that for each task the S3D settings remained identical. For each task, an object traveled from a position 30 m in front of the participant/virtual camera, which was positioned at the origin (0,0,0). The object continued to travel 30 m behind the camera/participant where it was audible through the rear 5.1 surround sound loudspeakers or through the frontal stereo loudspeakers depending on the sound mode setting. At the end of each task, the participant was asked to rate how close the moving object appeared to get before disappearing off-screen, even though it never changed how close it travelled for each task. The rating system ranged from 1 to 5, where 1 was close and 5 was distant. The purpose was to determine whether different sounds, loudspeaker configurations (5.1 surround sound or stereo loudspeaker configurations), and/or the inclusion of sound effects (Doppler frequency shift and/or frequency fall-off), affected the participant's perception of depth, which would result in different ratings of closeness depending on the combination of variables used. Results were then analyzed (See Section 4.1) to determine whether a particular combination of sound conditions had a significant influence on ratings. There were a total of 72 permutations of the variables used in this experiment.

Downloaded From: http://spiedigitallibrary.org/ on 11/05/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

3.4.2 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) and was developed using the Unity 3D game development tool. For each task a floating platonic shape travelled from +30 m to -30 m in 5 s to 10 s. For each task, the virtual camera was positioned at the origin (0,0,0). Therefore at the halfway point, the object passed through the virtual camera's location, although it visually faded away between 2.5 m and 1.5 m in front of the camera.

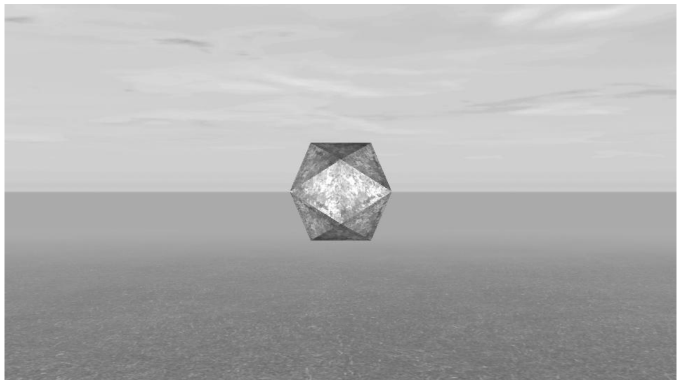

Figure 1. Visual game environment of the experiment.

3.4.3 Animation Parameters

To avoid over-familiarization with the task, the moving object had different size and animation speed 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 0.8 m, were as follows; 90%, 92.5%, 95%, 97.5% and 100%. The three speed settings are referred to as fast = 100% (5 s), medium = 75% (6.66 s), and slow = 50% (10 s). Both size and speed 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. 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.

3.4.4 Audio Stimuli Parameters

Three auditory conditions (referred to hereinafter as "sound files") were used: [T] a complex tone, [EN] an engine noise, and [MUS] a music track.

These particular conditions were chosen for their varying levels of complexity and due to the regular use of similar sounds in video games. The amplitudes of all of the audible files were normalized in Audacity then altered to have a similar perceived loudness level. 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 engine noise condition consisted of an automobile engine sound at slow speed. The music track consisted of a techno track sample. The techno track was chosen because it had a consistent rhythm and because it contained a good range of spectral content. 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. 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 engine noise and music sound files revealed that that they had a wide spread of spectral content ranging from 10 Hz to about 21 kHz.

Downloaded From: http://spiedigitallibrary.org/ on 11/05/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

The sound 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 at the beginning of the animation, the object was audible before it began to move toward the virtual camera/participant.

All of the frequency fall-off effects were linear and had short ranges with respect to how they might be setup in a videogame or their real-world equivalents. We chose exaggerated frequency fall-off effects 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 studies, the linear frequency and amplitude fall-off effects used were those provide by the Unity3D game development tool.

3.4.5 Summary of Parameters

The controlled variables were: the use of 5.1 surround sound and stereo loudspeaker modes, sound ([T] complex tone, [EN] an engine noise, [MUS] music), and the presence or absence of frequency fall-off and/or Doppler shift.

4. RESULTS

4.1 Analysis Method

The results analyzed in this section are based on the average ratings for each task, which were recorded for each participant. The mean estimated ratings in each table denotes a measure of closeness, where participants were asked to rate how close the moving object appeared to be before it disappeared off-screen (where 1 was close and 5 was distant). It should be noted that the object disappeared off-screen at the same location and with the same S3D depth settings for all tasks. 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 Mode and Sound File

We conducted repeated measures tests that examined the results for the following variables, mode (stereo, and 5.1 surround sound loudspeaker configurations) and sound file (complex tone, car engine noise, and music).

For the within-subjects effects analysis, the main effect of mode was significant, $F(1,30) = 29.631$, $p = .000$. However, these effects were qualified by a mode by sound file interaction, $F(2,60) = 37.831$, $p < .016$.

For the estimated marginal means for mode, 5.1 surround sound had a lower mean rating ($M = 2.148$) than stereo ($M = 2.692$). Therefore, participants rated the moving object as appearing closer before disappearing off-screen for 5.1 surround sound tasks than for the stereo tasks. Pairwise comparisons analysis revealed that there was a significant mean difference between the stereo and 5.1 surround sound conditions ($p = .000$). However, it should be noted that the estimated means incorporate all of the sound file conditions including ratings for various combinations of the Doppler frequency shift and frequency fall-off effects conditions. Therefore, this result should be understood as a general overview of the main effect of the mode condition.

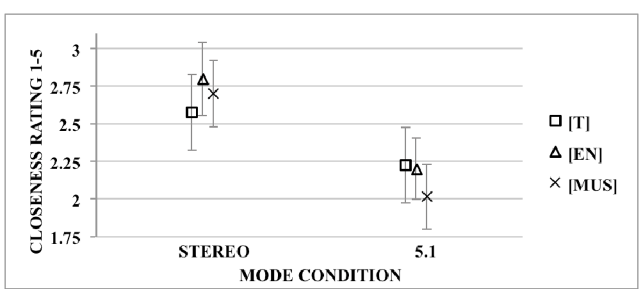

Figure 2. Mode and Sound file

Downloaded From: http://spiedigitallibrary.org/ on 11/05/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

Figure 2 shows the estimated marginal means for the interaction between mode and sound file. The vertical axis represents the participants' estimated mean rating for how close they felt the object travelled toward them before disappearing off-screen (where 1 was close and 5 was distant). The horizontal axis represents where the stereo and 5.1 ratings are located. Figure 2 illustrates how each sound file was rated as closer for the 5.1 surround setting. It should be noted that the estimated means incorporate ratings for the various combinations of the Doppler shift and frequency fall-off effects conditions.

Table 2. Mode and Sound file

MODE	SOUND	Mean	Lower Bound	Upper Bound
STEREO	T	2.577	2.371	2.783
	EN	2.799	2.555	3.043
	MUS	2.700	2.479	2.922
5.1	T	2.226	2.012	2.440
	EN	2.199	1.995	2.404
	MUS	2.018	1.803	2.233

Table 2 shows the estimated marginal means for the interaction between mode and sound file. For stereo, the moving object was rated as appearing closest for the complex tone condition, followed by the music condition, and then the engine noise condition. For the 5.1 surround sound loudspeaker configuration, the moving object was rated as appearing closest for the music condition, followed by the engine noise condition and then complex tone condition. The switching of the order of how close each combination of object and sound file was rated for each mode may relate to the fact that complex tone was least affected by the change in mode (see Figure 2). However, regardless of the order, all sound file conditions were attribute with closer ratings for 5.1 surround sound.

4.3 Mode, Sound file, Doppler Shift and Frequency Fall-off

We conducted repeated measures tests to examine the results for the following variables, mode (stereo, 5.1 surround sound loudspeaker configurations), sound file (complex tone, an engine noise, and music), frequency fall-off (ON, OFF), and Doppler frequency shift (ON, OFF).

For the within-subjects effects analysis, the main effect of mode was significant, $F(1,30) = 28.594$, $p = .000$. The main effect of Doppler frequency shift was significant $F(1,30) = 7.132$, $p = .012$. However, these effects were qualified by a significant sound file and Doppler shift interaction, $F(2,60) = 8.610$, $p =.002$, and a significant 3-way interaction between sound file, Doppler frequency shift and frequency fall-off, $F(2,60) = 4.441$, $p < .017$.

For the estimated marginal means for mode, 5.1 surround sound had a lower mean ($M = 2.157$) than stereo ($M = 2.708$). Therefore, participants rated the moving object as appearing closer, as predicted by the model, for 5.1 surround sound tasks than for the stereo tasks. Pairwise comparisons analysis revealed that there was a significant mean difference between the stereo and 5.1 surround sound mode ($p = .000$).

For the estimated marginal means analysis for Doppler shift, the moving object, for Doppler shift set to ON tasks, were rated as appearing closer ($M = 2.371$) than the OFF tasks ($M = 2.495$). Pairwise comparisons analyses of Doppler showed a significant mean difference between the ON and OFF conditions ($p = .012$).

Table 3. Sound File and Doppler Shift

SOUND	DOPP	Mean	Lower Bound	Upper Bound
T	ON	2.261	2.084	2.437
	OFF	2.609	2.391	2.827
EN	ON	2.445	2.234	2.655
	OFF	2.534	2.358	2.709
MUS	ON	2.406	2.200	2.612
	OFF	2.341	2.150	2.533

Downloaded From: http://spiedigitallibrary.org/ on 11/05/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

For the estimated marginal means for the interaction between sound file and Doppler shift (Table 3), the moving object was rated as closest for the complex tone and engine noise conditions when Doppler shift was set to ON. Conversely, it had a closer rating when set to OFF for the music condition.

Table 4. Sound File, Doppler Shift and Frequency Fall-off

SOUND	DOPP	FREQ	Mean	Lower Bound	Upper Bound
T	ON	ON	2.220	2.017	2.424
		OFF	2.301	2.102	2.500
	OFF	ON	2.637	2.426	2.849
		OFF	2.581	2.304	2.857
EN	ON	ON	2.457	2.232	2.682
		OFF	2.433	2.188	2.677
	OFF	ON	2.489	2.305	2.673
		OFF	2.578	2.382	2.774
MUS	ON	ON	2.524	2.283	2.765
		OFF	2.288	2.054	2.521
	OFF	ON	2.258	2.043	2.473
		OFF	2.425	2.213	2.636

Table 4 shows the estimated marginal means for the interaction between sound file, Doppler frequency shift and frequency fall-off. It should also be noted that the estimated means incorporate ratings for the both mode conditions.

When Doppler frequency shift is set to OFF, frequency fall-off set to ON attains closer average ratings than when frequency falloff is set to OFF except in the presence of a complex tone (Table 4). However, due to a concern relating to frequency fall-off and the ability of the filter's effects to be audible in relation to the complex tone condition, this exception may be due to chance.

When frequency fall-off is set to OFF, Doppler frequency shift set to ON attains closer average ratings than when Doppler frequency shift is set to OFF for all sound files.

When we take all combinations of Doppler frequency shift and frequency fall-off into account and look at the closest and furthest average ratings for each sound file condition, no obvious pattern emerges.

When both effects are switched to OFF, the music condition attains the closet estimated rating, followed by the sound effect and then the complex tone condition, although the latter two conditions have similar mean averages (M = 2.578, M = 2.581). It is worth examining this data in relation to the stereo and 5.1 mode separately.

For stereo and 5.1 surround sound modes, when Doppler frequency shift was set to OFF, frequency fall-off resulted in closer results when Doppler frequency shift was set to ON with the exception of the complex tone condition when mode was set to stereo (Table 5 and Figure 3). In addition, for the 5.1 surround sound mode and the music condition, the mean averages for frequency fall-off set to ON and OFF were similar ((ON) M = 1.935, (OFF) M = 1.946).

For stereo, when frequency fall-off was set to OFF, Doppler shift attains closer ratings when frequency fall-off was set to ON in all instances (Table 5). For 5.1 surround sound, when frequency fall-off was set to OFF, Doppler frequency shift resulted closer average ratings when frequency fall-off was set to ON for all sound file conditions except for the music condition, where the average rating, as with frequency fall-off, were similar (ON = 1.952, OFF = 1.946).

There was no clear pattern, in terms of the closet and furthest rating given to each sound file condition, in relation to each possible combination of Doppler frequency shift and frequency fall-off for either 5.1 surround sound or stereo loudspeaker configurations (Table 5). This suggests that when the effects appeared in combination the influence on perceived depth was inconstant, in other words, the strengths of the effects when presented together were unpredictable or may have caused confusion in relation to participant ratings.

Downloaded From: http://spiedigitallibrary.org/ on 11/05/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

Table 5. Mode, Sound File, Doppler Shift and Frequency Fall-off

MODE	SOUND	DOPP	FREQ	Mean	Lower Bound	Upper Bound
ST	T	ON	ON	2.409	2.175	2.642
			OFF	2.462	2.216	2.708
		OFF	ON	2.919	2.673	3.166
			OFF	2.742	2.319	3.165
	EN	ON	ON	2.645	2.338	2.953
			OFF	2.758	2.427	3.089
		OFF	ON	2.828	2.566	3.090
			OFF	2.849	2.580	3.119
	MUS	ON	ON	2.774	2.468	3.080
			OFF	2.624	2.358	2.889
		OFF	ON	2.581	2.217	2.945
			OFF	2.903	2.598	3.208
5.1	T	ON	ON	2.032	1.726	2.339
			OFF	2.140	1.908	2.371
		OFF	ON	2.355	2.082	2.628
			OFF	2.419	2.031	2.807
	EN	ON	ON	2.269	2.040	2.497
			OFF	2.108	1.860	2.355
		OFF	ON	2.151	1.923	2.378
			OFF	2.306	2.040	2.573
	MUS	ON	ON	2.274	1.961	2.588
			OFF	1.952	1.651	2.252
		OFF	ON	1.935	1.720	2.151
			OFF	1.946	1.699	2.193

Figure 3 shows the interaction between mode, sound file, Doppler frequency shift and frequency fall-off. The vertical axis shows the participant's closeness rating (where 1 was close and 5 was distant), the horizontal axis shows various combinations of Doppler frequency shift and frequency fall-off, with Doppler frequency shift setting first and the frequency fall-off second. For example, "ON ON" = Doppler shift set to ON and frequency fall-off set to ON.

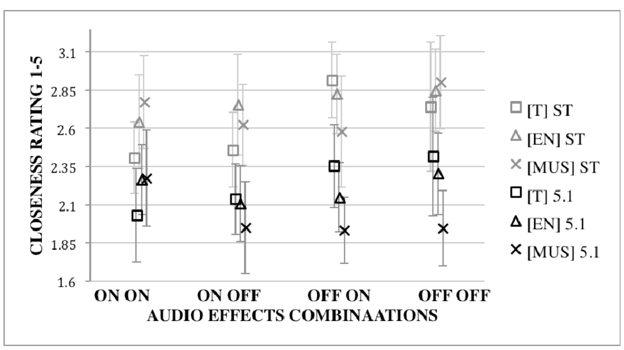

Figure 3. Mode, Sound File, Doppler Shift and Frequency Fall-off

In general, the stereo mode sound file tasks are rated as more distant than their 5.1 surround sound counterparts. In addition, all sound files are rated as closer than when effects are used in isolation than when no audio effects are present except for complex tone and the frequency fall-off in isolation setting. For the stereo condition, when we consider a three audio conditions or sound files, complex tone results in the two closest rating when Doppler is set to ON, followed by

Downloaded From: http://spiedigitallibrary.org/ on 11/05/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

music when the audio effects were applied in isolation, and then by the engine noise file when both audio effects were set to ON.

For the 5.1 condition, music had the three closest ratings, where the audio modes seem to have little effect, followed by complex tone and both effects set to ON, then by engine noise with Doppler in isolation.

For the complex tone in the stereo condition, tasks with Doppler frequency shift set to ON had a strong effect. This is also true for the complex tone condition in the 5.1 condition. For tasks with Doppler Shift set to OFF, Frequency fall-off did not have a consistent effect for the stereo and 5.1 modes. However, 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, this exception may be due to chance.

For the engine noise with a stereo loudspeaker configuration condition, the moving object was rated as closer for all audible combinations of the audio effects, especially when Doppler was set to ON. The engine noise sound with a 5.1 loudspeaker configuration follows a similar trend but is rated almost as distant to the OFF / OFF audio effect conditions when both audio effects are set to ON.

Music in the stereo loudspeaker configuration condition, results in closer ratings when both effects are in isolation. Unlike the complex tone and engine noise conditions, when both effects are set to ON, music is rated almost as distant as when both audio effects (Doppler frequency shift and frequency fall-off) are set to OFF. For music in the 5.1 condition, the audio effects in isolation appear to have little effect. However, both audio effects set to ON produce a sharp jump in results and produce the most distant rating.

For 5.1 surround sound, when the audio effects were set to OFF / OFF, the music condition resulted in the closet estimated rating, followed by engine noise, and then complex tone. This is in reverse order to that observed with stereo. This raises the question of whether the effect of passing from the front to the rear loudspeakers cause the sound files to have a different strength in terms of their ability to influence apparent depth.

4.4 Summary of Results

4.4.1 Impact of Mode

Mode exhibited a significant effect on participant ratings for all analyses preformed. The analysis for mode and sound file revealed that the 5.1 surround sound condition had a significantly lower estimated marginal mean than stereo condition. Therefore, participants rated the moving object as appearing closer under the 5.1 surround sound condition than for the stereo condition. In addition, all sound files were associated with closer mean ratings for the 5.1 surround sound condition than for the stereo condition.

The data was analyzed for mode, sound file, Doppler frequency shift, and frequency fall-off. Mode appeared to have an additional impact when the audio effects such Doppler shift and frequency fall-off, were applied. However the strength of this additional impact varied between sound files and different combinations of audio effect settings

Further research is required to ascertain if there is a significant relationship between the mode used and the audio effects that were applied. Mode may have had an influence on the sound files too, where stereo and 5.1 surround sound loudspeaker configurations have the opposite hierarchy in terms of which sound files achieved the closet ratings, when both audio effects are set to OFF. More research is required to determine the nature of the interactions between mode and the individual sound files and the audio effects in more detail. However, we can assert that 5.1 surround sound can increase the apparent depth of moving objects in S3D imagery.

4.4.2 Impact of Sound File

In our analysis, the interaction between mode and sound file exhibited a significant effect. However the nature of this interaction remains unclear. For example, for the stereo mode, the moving object was rated as appearing closest for the complex tone condition, followed by the music, and then the engine noise condition, while for 5.1 surround sound, the moving object was rated as appearing closest for the music condition, followed by the sound effect and then complex tone condition. For the pairwise comparisons for sound file, there was a significant mean difference exhibited between the music and engine noise only. It is worth noting that for all analyses, the sound effect condition never attained the closest average mean.

In general, no consistent patterns of behavior could be attributed to the way in which sound files influenced player ratings. However, mode (5.1 surround sound or stereo) and the combination of the presence or absence of Doppler frequency shift and frequency fall-off effects, may have dictated how strong an influence a particular sound file had on

Downloaded From: http://spiedigitallibrary.org/ on 11/05/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

S3D depth perception, therefore changing the order in which each sound file had the strongest influence in the process. Further research is required to determine whether a particular sound file in combination with stereo or 5.1 surround sound can have a specific effect on S3D depth perception.

4.4.3 Impact of Doppler Shift

Doppler shift, in isolation and/or in relation to sound file, exhibited a significant main effect on ratings for all analyses performed. In addition, when stereo and 5.1 tasks were considered together, Doppler shift had a stronger influence in isolation that frequency fall-off in isolation except for the music condition.

When we examined stereo and 5.1 surround sound individually and with frequency fall-off was set to OFF, Doppler shift resulted in closer ratings when set to ON in all instances, except for 5.1 and the music condition. However, the mean ratings for Doppler ON and OFF for music were very similar.

Examining the 5.1 surround sound mode reveals that all sound file conditions attained the closer ratings when Doppler shift was applied in isolation than in combination with frequency fall-off except for the complex tone condition, where Doppler shift set to ON attained a closer rating when the audio effects were set to ON / ON. However, due to the narrow range the complex tone inhabits on the frequency spectrum, frequency fall-off may have little effect on the complex tone sound file. Therefore, the apparent closer average ratings attributed to the audio effects set to ON / ON, rather than ON / OFF, may be incidental.

For stereo, the Doppler shift set to ON tasks attained closest ratings, with respect to each individual sound file, except for music.

Doppler shift seem to have a particularly strong effect on the complex tone condition. One possible reason for the strong interaction between complex tone and the Doppler effect is that the complex tone only takes up a small portion of the frequency spectrum and therefore its entire timbre appears to be modulated by the Doppler pitch shift effect where the other sound files still maintain other aspects for their formal structure, such as the rhythmic aspect of the music condition.

4.4.4 Impact of Frequency Fall-off

When tests were conducted that considered mode, sound file, Doppler shift and frequency falloff, the interaction between sound file, Doppler shift and frequency fall-off attained a significant result.

For stereo and 5.1 modes, when Doppler was set to OFF, frequency fall-off attains closer results when set to ON with the exception of complex tone. Although, for the 5.1 surround sound mode and the music condition, the mean averages for frequency fall-off set to ON and OFF were similar ((ON) $M = 1.935$, (OFF) $M = 1.946$).

In general, we could not determine if frequency fall-off had an effect on participant ratings, although the estimated means suggest that in using the effect in isolation may strengthen the apparent depth of the moving object.

4.4.5 Impact of Doppler Shift and Frequency Fall-off

There does not seem to be a consistent pattern evident in relation to each sound file and each possible combination of Doppler shift and frequency fall-off for either 5.1 surround sound or stereo modes in terms of the closet and farthest ratings. Further research is required to determine if this is always the case.

4.4.6 Possible Impact of Animation Size and Speed

We used a pivot table analysis, available in Microsoft Excel, to check if animation size and speed had an effect on results. Generally, the larger the size of the moving object, the closer it was rated. Speed had an inconsistent effect but higher speeds were most often regarded as closer. 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 if the distribution of the size 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.

Downloaded From: http://spiedigitallibrary.org/ on 11/05/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

5. DISCUSSION

The fact that participants provided different ratings of closeness to a moving in-game object that always disappeared at the same S3D settings demonstrates how open to influence S3D perception can be. Altering the loudspeaker settings and adding audio effects to sounds, such as Doppler frequency shift, causes moving object animations to be perceived closer on average. In addition, the type of sound file used and whether or not an audio effect was enabled seemed to influence the closeness rating, although, no logical pattern emerged.

It would be of value to examine if less or more exaggerated uses of Doppler frequency shift and/or amplitude and frequency fall-offs (which do not dynamically change distance) have the same effect on participant ratings.

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, amongst others. 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.

When considering the interaction of sound and S3D under more dynamic conditions than those examined by Turner et al. [12], the interactions between both modalities become more complicated. Unlike the experimental setup used in [12], when the loudspeaker positions are fixed, when different audio modes, sound files and effects are considered, and participants interact with moving audible S3D objects, the inclusion of sound can be detrimental to accurate S3D depth judgment. 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 is small, the experiment illustrates that 5.1 surround sound appears to increase the apparent depth within an S3D environment. Based on our findings, we recommend that game developers consider designing 5.1 surround sound audio for games with S3D in mind. Audio effects such as Doppler frequency shift and frequency fall-off, should also be considered in relation to the design process.

If, as our results suggest, 5.1 surround sound and certain combinations of sound files and audio sound effects can increase the sense of visual depth that a player experiences, and if S3D depth contributes to feelings of motion sickness in gamers, 5.1 surround sound and certain uses of audio sound effects 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 stereoscopic 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 accuracy in terms of judging distance. 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 stereoscopic 3D gaming.

ACKNOWLEDGMENTS

Funding for this project was provided by the Ontario Media Development Corporation, Ontario Centres of Excellence as part of the iGO3D initiative, and the Social Sciences and Humanities Research Council of Canada.

REFERENCES

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

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

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

Downloaded From: http://spiedigitallibrary.org/ on 11/05/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

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

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

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

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

[8] Mastoropoulou, G., Debattista, K., Chalmers, A. and Troscianco, T., "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, (2005).

[9] Larsson, P., Västjäll, D. and Kleiner, M., "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, (2003).

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

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

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

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

Downloaded From: http://spiedigitallibrary.org/ on 11/05/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx