SPATIAL AUDIO QUALITY AND USER PREFERENCE OF LISTENING SYSTEMS IN VIDEO GAMES

Joe Rees-Jones,

Audio Lab, Department of Electronics,
University of York
York, UK
jrj504@york.ac.uk

Jude Brereton,

Audio Lab, Department of Electronics,
University of York
York, UK
jude.brereton@york.ac.uk

Damian Murphy,

Audio Lab, Department of Electronics,
University of York
York, UK
damian.murphy@york.ac.uk

ABSTRACT

Spatial audio playback solutions provide video game players with ways to experience more immersive and engaging video game content. This paper aims to find whether listening systems that are able to more accurately convey spatial information are preferred by video game players, and to what extent this is true for different loudspeaker configurations whilst engaged in video game play. Results do suggest that a listening system with high perceived spatial quality is more preferred.

1. INTRODUCTION

Spatial audio hardware and software solutions are quickly becoming the state of the art in video game audio playback. Whether this is through multichannel speaker configurations, such as stereo, 7.1 surround, Dolby ATMOS [1], or binaural headphone solutions such as DTS HeadphoneX [2] and Razer Surround [3], video game players can now experience immersive audio in a variety of ways.

The dynamic nature of video gameplay lends itself well to the fundamental qualities of spatial audio, arguably more so than cinema or music. A vast majority of what is perceived by the player, especially in first and third-person content occurs off-screen, either behind or to the side of the camera view [4]. This is where effective use of spatial audio can be especially useful, where spatialised sound cues can be used to warn players of impending threats, potentially giving a competitive advantage in multiplayer situations [5, 6]. Sound cues can also be used to reduce the information presented on screen, for example on the heads up display (HUD), to the extent where it is possible to develop audio only games [7].

This paper describes an experiment to ascertain whether the use of spatial audio in game audio playback rendering solutions, and in the games development process, has a positive impact on video game player quality of experience. Appropriate game content will be sourced and, in playing through a scene, participants will subjectively rate the spatial sound quality and preference of some typical game audio playback systems. It is hypothesised that an audio system with subjectively high spatial sound quality will be preferred to those with lower capabilities. Section 2 outlines the process of choosing appropriate video game content for the purposes of experimentation, and the experimental method. The spatial attributes used to assess said content are also given. Results and analysis are given in Section 3.

2. EXPERIMENTAL DESIGN

2.1. Content Selection

Video game content appropriate for the purposes of this study was chosen according to a specific set of criteria divided into two main categories:

Audio:

- Appropriate and effective use of spatial audio.
- 7.1 surround sound compatibility.
- Third-party critical acclaim for use of audio.

Gameplay:

- Repeatability.
- A limited number of 'fail-states', meaning it should be difficult for a player to fail as a result of death or not completing an objective.
- The ability to easily restart with little or no back-tracking.
- Simple controls.
- An easy to follow, preferably linear path.

Extensive play testing of a wide range of content revealed Naughty Dog's *The Last of Us: Remastered*, (on the Playstation 4) to be the most appropriate video game for the purposes of this study as it fulfills all the set criteria. The game is currently the most awarded in recorded history [8] and has won a multitude of audio related awards including a BAFTA for Audio Achievement [9] and G.A.N.G (Game Audio Network Guild) awards for Audio of the Year, Sound Design of the Year and Best Audio Mix in 2014 [10], showing considerable acclaim in the industry.

Not only does the spatial mix aesthetically enrich the virtual environments and settings presented to the player through effective use of spatial audio, but it is also essential to the gameplay. The audio in itself can be considered as one of the core gameplay mechanics, where players are able to locate potential threats purely through listening. Playing the game through a multi-channel audio system seems to significantly increase chances of survival where players are able to gain a tactical advantage over enemies by listening for their whereabouts.

The introductory sequence of *The Last of Us: Remastered* is especially appropriate for the purposes of experimentation. The player is required to follow a linear path, reducing the likelihood for inexperienced participants to get lost. The majority of the audio events are scripted and won't trigger until the player encounters a particular section, ensuring similar auditory experiences on multiple play-throughs. The number of fail-states in the sequence is low, where even if the player does fail they are able to quickly continue the play-through with minimal consequence.

It has also been shown that giving participants' too much control in an experiment such as this will have an adverse effect on their ability to rate sound quality [11], making this particular section in *The Last of Us: Remastered* ideal - all that is required of the player is to move their character and follow clear, on-screen button prompts for more complex interactions.

2.2. Playback Scenarios

Three playback listening conditions were used in the experiment: mono, stereo and 7.1 surround-sound. These three conditions were chosen as they best represent current audio solutions available to the average gamer. Mono was used for the least spatially capable listening condition as it is near impossible to convey accurate spatial information over one loudspeaker. Stereo playback permits sound to be positioned horizontally left and right giving some amount of spatialisation. The 7.1 surround-sound loudspeaker arrangement is the best spatial solution of the three as it allows for full horizontal spatialisation, enveloping the player in sound and also allows for fairly accurate sound source localisation. In consumer gaming, it is also currently the limit as to what is offered in terms of surround-sound channel count.

In order to ensure participants are not biased in their responses by being able to detect the number of speakers actually active in the playback system, it was decided that both the mono and stereo listening conditions would make use of all of the 8 speakers to be used in the 7.1 surround-sound configuration. It is possible to set the AV receiver used in the tests, which handles all the audio and visual information from the Playstaion 4 and outputs it to the appropriate listening and viewing apparatus, to *Full Mono* where a single mono mix-down of the game audio stream is sent to all connected speakers. Similarly the listening mode *All Ch. Stereo* takes the stereo game audio stream (which may well be down-mixed automatically from a higher channel count by either the game engine or the receiver) and sends audio for the left channel to all three speakers to the left of the listener and the same for the right. The centre channel outputs a sum of the left and right channels. Typically it would not be expected that a listener would experience mono or stereo material in this fashion. A subwoofer was included in all three listening conditions for low frequency effects. Content was presented to participants in this way to make the transition between listening conditions less obvious. It can be expected that the effectiveness of the game's spatial audio would still be limited since the physical limitations of these listening modes would not allow for the exploitation of critical spatial information, especially in regards to sound cues from behind the listener.

To assess each playback scenario participants were split into three groups (A, B and C), each of which was exposed to two of the possible three listening conditions:

- Group A - Mono – Stereo.
- Group B - Mono – Surround.
- Group C - Stereo – Surround.

Due to the length of the chosen scene from *The Last of Us: Remastered* (approximately 12 minutes) it was decided participants should only be exposed to two of the three listening conditions, significantly reducing the amount of time required of each participant and also the risk of any learning bias that may be present after three play-throughs. A third play-through may also confuse subjects' judgement and therefore affect preference ratings. Participant identities were anonymised by assigning each a unique number and the order of exposure to the two listening conditions was randomised.

2.3. Spatial Attributes

The list of attributes chosen to rate the spatial sound quality of the game audio content is as follows:

- Depth.
- Distance.
- Sound Source Localisation.
- Sound Source Definition.
- Stability.
- Envelopment of Reverberation.
- Source Width.

These attributes and their accompanying descriptors (for the benefit of participants) were sourced from several different spatial attribute lists suggested by the SAQI (Spatial Audio Quality Inventory) [12], ITU recommendation BS.1284-1 [13], and other publications [14-16].
These spatial attributes were rated by participants on a 5-point quality scale ranging from Bad (1), Poor (2), Fair (3), Good (4) to Excellent (5) as suggested by ITU recommendation BS.1284-1 [13].

After completion of both play-throughs preference was measured on a 7-point paired comparison scale where participants were able to state which of the two listening conditions they were personally exposed to they preferred and to what extent. The scale ranged from Strong Preference for A (3), Preference for A (2), Slight Preference for A (1), No Preference (0), Slight Preference for B (1), Preference for B (2) and Strong Preference for B (3) [17].

Using paired comparison of this type it can be assumed that the preference rating for stimuli A will be the same value, but of opposite magnitude, to the rating given to stimuli B and vice-versa [18]. For example if surround-sound (A) were being compared to mono (B) and received a rating of 3 for strong preference, mono would then receive a -3 as it can be presumed it is strongly not preferred.

2.4. Apparatus

The Last of Us: Remastered was played on a Sony Playstation 4 connected via HDMI to an Onkyo TX-NR838 AV Receiver. Six Genelec 8040As, one Genelec 8040B (centre channel) and a Genelec 7060B Active Subwoofer were arranged according to the ITU specification [19] for 7.1 surround-sound listening and connected to the appropriate audio outputs of the receiver. The sound level of each speaker was measured with an SPL meter and test tone to ensure all speakers were at a consistent level. Overall volume was controlled by the receiver and set to a comfortable level for the duration of the experiment.

An Optoma HD200X projector was used for visual feedback and connected via HDMI to the main video out of the receiver. An office chair was positioned in the sweet spot for participants to be seated whilst partaking in the experiment. The listening room was surrounded by a thick absorbing drape with foam acoustic panelling above the listener. The extra speakers above, below and to the side of listener, shown in Figure 2, that do not conform to the 7.1 surround sound speaker configuration were not active.

3. RESULTS

In total 21 participants took part in the experiment (17 male and 4 female). 20 of these participants were aged 20-35 with one over 50. In regards to gameplay experience 5 of the participants play games on a daily basis, 10 weekly, 3 monthly, 2 yearly and 1 never. Participants were split evenly into 3 groups (A, B and C) corresponding to the 3 sets of listening conditions giving 7 per group and 14 for each listening condition.

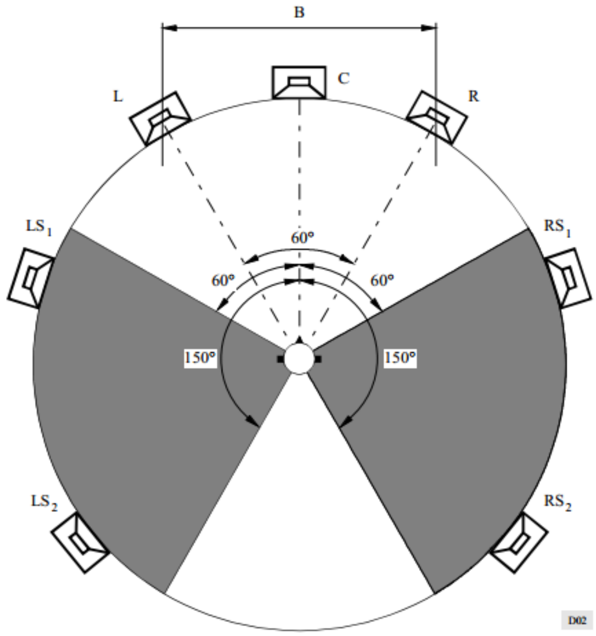

Figure 1: *The full experimental surround-sound loudspeaker configuration used, conforming to the 7.1 standard - from [19].*

Participants were first given an experiment pack containing an information sheet with consent form, event timeline, control scheme, spatial attribute list with descriptors and questionnaire. Spatial attributes were further explained to participants by the experimenter upon request. Subtitles were turned on and the vertical axis for the game's camera control inverted if requested by the participant.

Depending on group allocation, and unknown to the participant, one of two listening conditions was randomly chosen and set on the receiver by the experimenter. After completing the first play-through participants were asked to exit the listening space and fill in the first spatial quality questionnaire whilst the experimenter reset the game and set the receiver to the second listening condition. Upon completion of the second play-through participants filled in the remainder of the questionnaire. The experimenter stayed in the room throughout the test to offer help in navigating the game if needed.

Comments regarding the experiment were also noted by participants.

Figure 2: *The listening room used throughout the experiment. Loudspeakers not part of the defined 7.1 playback system were non-active during the test.*

3.1. Analysis

Before analysis it was decided the stability attribute would be removed as one participant did not give a score due to a lack of understanding. It was felt it would be less detrimental to the experiment if the attribute was removed from the analysis process rather than omitting a participant's entire set of results. Individual spatial attribute scores for each participant were first normalised by generating z-scores [20] and then averaged by participant to give overall spatial quality ratings for each listening condition. All statistical analysis was conducted in MATLAB.

Independent t-tests revealed there to be a statistically significant difference between the spatial quality ratings for all three listening conditions given by groups A, B and C at $p<.05$. This suggests that perceived spatial quality is influenced by the listening condition a participant is exposed to. Results from the t-tests are given in Table 1.

Table 1: *Table of t-test results across all listening conditions and groups. A value of '1' represents a rejection of the null hypothesis at p<.05.*

	Mono (a)	Stereo (a)	Surround (b)
Mono (b)	X	1	1
Stereo (c)	1	X	1
Surround (c)	1	1	X

One-way ANOVA tests revealed there was no statistically significant difference between spatial quality scores given by different groups assessing the mono and surround listening conditions at $p<.05$. For mono groups A and B [$F(1,6) = 4.34$, $p = 0.06$], and surround groups B and C [$F(1,6) = 3.69$, $p = 0.08$]. These results suggest that individuals from different sample groups will have a similar subjective opinion regarding spatial sound quality of the same listening condition when engaged in playing a video game. However, this cannot be said for the stereo condition where [$F(1,6) = 16.22$, $p = 0.002$]. Through further analysis it was shown that the spatial quality scores given for stereo when compared to mono were significantly

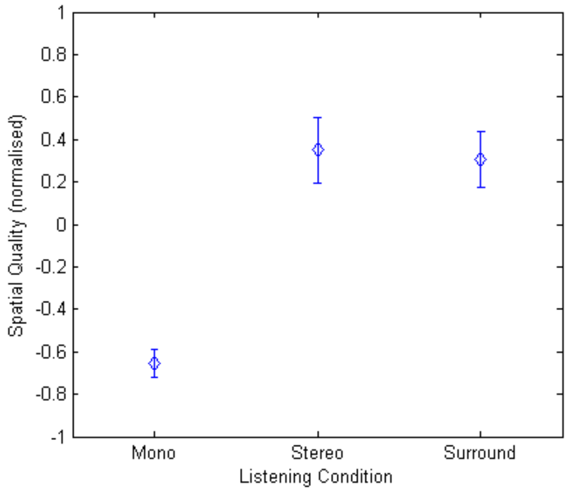

Figure 3: *The mean of the normalised spatial quality scores for the mono, stereo and 7.1 surround-sound listening conditions with standard error.*

higher than those when compared to 7.1 surround-sound.

A comparison between the group means of the three listening conditions indicates a statistically significant difference between the subjective spatial quality of mono, stereo and 7.1 surround sound. The mean spatial quality scores for each listening condition are presented in Figure 3. Analysis by one-way ANOVA determined [$F(2,13) = 21.17$, $p<0.001$] at the $p<.05$ level, suggesting that listening condition has a significant effect on subjective impression of spatial quality. An honest significance difference (HSD) post-hoc test revealed that both stereo and 7.1 surround-sound total spatial quality scores were considerably higher than those for mono. Spatial quality scores for stereo and surround are however not significantly different to one another at $p<.05$.

It was anticipated that mono would receive the lowest spatial quality score of the three listening conditions since the physical capabilities of a single, down-mixed audio signal make conveying accurate spatial information difficult. However, it was also expected that 7.1 surround-sound would be perceived by individuals to be the more spatially capable system which, according to the results, is not the case.

Analysis of preference ratings given to the same listening condition by different groups revealed there to be no statistical difference between mono groups A and B, [F(1,6) = 0.96, p = 0.34] and surround groups B and C [F(1,6) = 3.43, p = 0.09]. There was however a much more noticeable difference between the preference ratings given by groups A and C for the stereo condition where [F(1,6) = 8.65, p = 0.01] at p<.05. HSD output revealed stereo received higher preference ratings from group A (Mono – Stereo) and lower ratings from group C (Stereo – Surround).

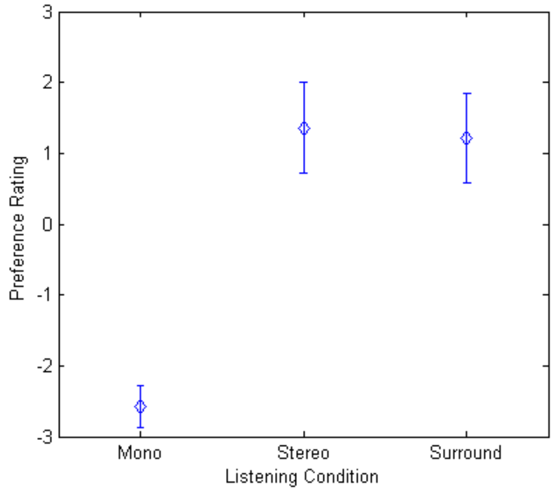

Figure 4: *The mean preference ratings for the mono, stereo and 7.1 surround-sound listening conditions with the standard error.*

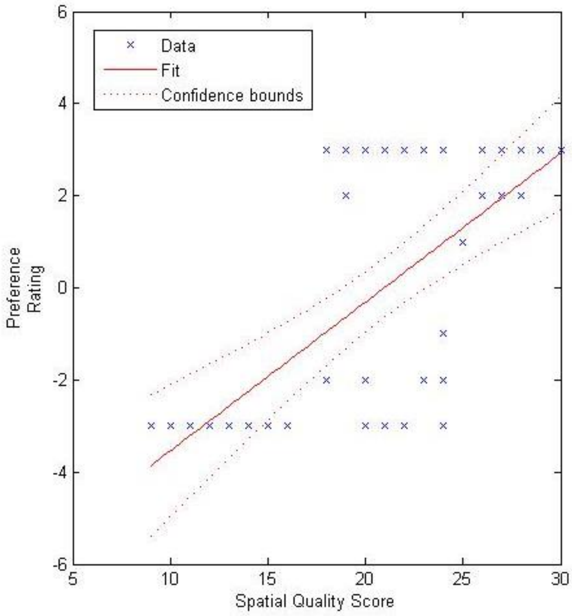

Figure 5: *Regression model of spatial quality scores vs preference ratings.*

Much like the spatial quality scores given for stereo, this implies that to some extent preference is being affected by the combination of conditions a participant is exposed to.

Analysis of preference ratings grouped by listening condition determined that [F(2,13) = 16.63, p<0.001] at p<.05, suggesting it has a significant effect on preference. The mean preference ratings for each listening condition are given in Figure 4. A Tukey's HSD post-hoc test of the preference ratings yielded a similar result to that of spatial quality, where mono was preferred the least but stereo and surround received statistically similar ratings at p<.05.

Again this result was not expected as it was hypothesised that 7.1 surround-sound would be more preferred than both the mono and stereo listening conditions. However, these results do show that as subjective impression of spatial quality increases, so does preference. This is illustrated by the regression model of non-normalised data in Figure 5 where the line of best fit shows how preference (Y axis) for a listening condition will increase as its spatial quality (X axis) improves.

3.2. Discussion

From the analysis of results it can be seen that a listening condition that is subjectively perceived to be of high spatial quality will be more preferred when engaged in video game play. This is especially clear in regards to the mono listening condition which received the lowest spatial quality scores and was also the least preferred.

However, the analysis also suggests that stereo is as spatially capable as 7.1 surround sound and equally as preferred. This outcome was not expected and is most likely due to the fact that all-channel stereo was used in place of traditional 2-channel frontal stereo in an attempt to make the transition between listening conditions less obvious. Stereo allows for some amount of spatialisation between the left and right channels, which will have been exaggerated by the use of all-channel stereo. This extreme panning to the left and right may have been perceived by listeners to be more spatial than regular stereo, resulting in a more positive opinion of the condition. It is also important to note that full mono and all-channel stereo are naturally going to feel more enveloping than their more traditional one and two-channel counterparts, since audio will be perceived to be all around the listener, even if the spatial information is not accurate.

The game's onscreen visuals coupled with the potential envelopment of full mono and all-channel stereo, may have also played a part in convincing participants into thinking they were experiencing a more spatialised soundscape. Studies have shown that visual stimuli can have significant effects on an individual's ability to perceive spatial aspects of a soundscape, especially relating to sound source localization [21, 22]. This is apparent in some of the participants' comments. After it was revealed to a participant which listening conditions they were personally exposed to it was implied that they were able to localise the sounds of two passing police cars even in the full mono environment. Another stated they could hear a helicopter passing over their head in both the stereo and mono conditions even though loudspeakers to simulate elevation were not included in this experiment.

It was decided frontal two-channel stereo should be added as a fourth listening condition and directly compared to 7.1 surround sound in a second set of tests to find if the use of all-channel stereo influenced the original results. It was hypothesised that stereo would receive a lower spatial rating than 7.1 surround-sound and be less preferred.

3.3. Experimental Procedure

A fourth group of participants (D) was assembled to assess the spatial sound quality and give preference ratings for normal two-channel frontal stereo compared to 7.1 surround-sound. Participants were subject to the same procedure as the previous experiment. 6 new participants took part (4 male, 2 female) aged between 18 and 30.

3.4. Analysis

Spatial quality scores were normalised and averaged by participant, then analysed by one-way ANOVA. Analysis of the spatial quality scores for surround-sound and full frontal stereo determined [$F(1,5) = 20.59$, $p = 0.001$] at $p<0.05$. As expected the spatial quality of 7.1 surround-sound was significantly higher than that of traditional, two-channel stereo. The mean spatial quality scores for both conditions are illustrated in Figure 6.

Furthermore, analysis of the preference ratings given to frontal stereo and surround revealed [$F(1,5) = 320$, $p<0.001$] at $p<0.05$. By inspecting of the mean preference ratings presented in Figure 7 it can be seen that there is a clear difference between the two conditions.

3.5. Discussion

As was hypothesised, 7.1 surround-sound received higher spatial quality scores than the regular stereo listening environment and as a result was more preferred by participants. This further reinforces the idea that, if given the choice, players will prefer to experience their games through more spatially able listening environments. It can also be said that the unexpected results previously observed were most likely down to the use of full channel stereo, rather than regular two-channel frontal stereo, as a playback condition, giving a more realistic insight into playback solutions available to the average gamer.

However, a crucial question still remains unanswered: how perceivable are the differences between 7.1 surround-sound and all-channel stereo and where does the threshold of discrepancy lie?

4. CONCLUSIONS

Results from these experiments do suggest that video game players have a more preferable experience when listening to their games through spatial audio speaker systems. The first set of tests indicate that stereo and 7.1 surround-sound gameplay demonstrated clear perceptual spatial quality improvement in comparison to mono, thus being more preferred. The subjective similarity between stereo and 7.1 was not expected and shows no significant difference in player preference, however further tests using regular, two-channel frontal stereo suggest that this was due to the experimental conditions used in the original experiment.

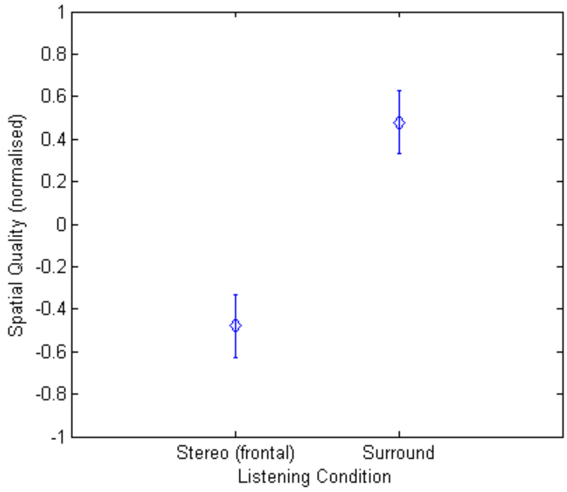

Figure 6: *The means of normalised spatial quality scores for the frontal stereo and 7.1 surround sound listening conditions with the standard error.*

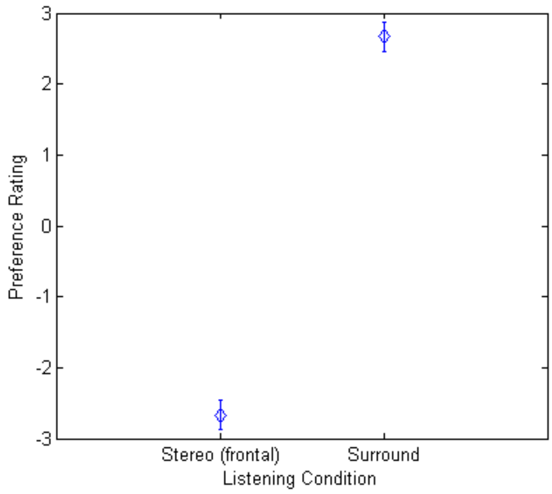

Figure 7: *The mean preference ratings for the frontal stereo and 7.1 surround sound listening conditions with the standard error.*

5. FURTHER WORK

Even where 7.1 surround-sound demonstrates improvement, a loudspeaker setup of this caliber isn't always the most practical configuration, especially for gamers where price, free space and living conditions have to be taken into account. Virtualising surround-sound headphones are the current alternative available to video game players on both next generation consoles and PC. Experiments designed to not only assess the spatial quality of such solutions, but also test whether players actually get better at their games as a result of using them would be highly beneficial.

The Last of Us: Remastered falls very firmly into the action-adventure genre so it can easily be seen how games like this and of similar a genre (such as FPS (first person shooter) and horror) could benefit from effective spatial audio. However, it is not clear how spatial audio may benefit players of less obvious genres such as puzzle games or side-scrollers. It would be interesting to see how spatial audio could be used in a more creative sense to provide auditory feedback in these types of games, rather than to create 'lifelike' audio simulations as is often the case.

One of the drawbacks of the experiments presented in this paper is the reliance on participants to be able to accurately rate the spatial quality of the listening conditions presented to them. There are a number of factors that can affect the reliability of such subjective measures such as:

- Participants' ability to recall their opinions on specific elements of the in-game soundscape after extensive playing time.
- The clarity of the spatial attributes to be rated.
- The assumption that playing a video game is not going to distract from the core purposes of the test.

An experiment is currently in development to find whether the introduction of spatial audio in a gaming environment has an effect on a player's physically measureable emotional output. Through skin conductance and heart rate measures it is possible to quantify an individual's arousal and valence, which when collectively observed can give an idea as to their current emotional state [23]. A positive outcome from an experiment such as this could add considerable weight to data gathered by subjective methods.

REFERENCES

[1] Dolby.com, 'Dolby Atmos', 2015. [Online]. Available: http://www.dolby.com/in/en/brands/dolby-atmos.html. [Accessed: 28- May- 2015].

[2] Headphonex.com, 'DTS HeadphoneX', 2015. [Online]. Available: http://www.headphonex.com/. [Accessed: 28- May- 2015].

[3] Razer, 'Razer Surround Personalized 7.1 Gaming Audio Software', 2015. [Online]. Available: http://www.razerzone.com/gb-en/surround. [Accessed: 28- May- 2015].

[4] Collins, Karen. *Playing with sound: A theory of interacting with sound and music in video games*. MIT Press, pp. 47, 2013.

[5] Richardson, John, Gorbman, Claudia, and Vernallis, Carol, eds. *The Oxford Handbook of New Audiovisual Aesthetics*. Oxford University Press, pp. 585-602, 2013.

[6] Kerins, Mark. *Beyond Dolby (stereo): cinema in the digital sound age*. Indiana University Press, pp. 322-324, 2010.

[7] Oldenburg, Aaron. "Sonic mechanics: Audio as gameplay." *The International Journal of Computer Game Research [Online journal]* 13, 2013.

[8] G. Nelva, 'Naughty Dog's The Last of Us Is the Most Awarded Game in Recorded History by Critics | DualShockers', Dualshockers.com, 2015. [Online]. Available: http://www.dualshockers.com/2014/01/24/the-last-of-us-is-the-most-awarded-game-in-history-by-critics/. [Accessed: 28- May- 2015].

[9] Awards.bafta.org, 'Games in 2014 | BAFTA Awards', 2015. [Online]. Available: http://awards.bafta.org/award/2014/games. [Accessed: 28- May- 2015].

[10] Audiogang.org, '2014 Awards', 2015. [Online]. Available: http://www.audiogang.org/awards/2014-awards/. [Accessed: 28- May- 2015].

[11] Zielinski, Slawomir K., et al. "Computer games and multichannel audio quality-The effect of division of attention between auditory and visual modalities."*Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*. Audio Engineering Society, 2003.

[12] Lindau, Alexander, et al. "A Spatial Audio Quality Inventory (SAQI)." *Acta Acustica united with Acustica* 100.5, pp. 984-994, 2014.

[13] Recommendation, I. T. U. "General methods for the subjective assessment of sound quality." *ITU-R BS* (2003): 1284-1.

[14] Rumsey, Francis. *Spatial audio*. Taylor & Francis, 2001.

[15] Rumsey, Francis. "Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm." *Journal of the Audio Engineering Society* 50.9, pp. 651-666, 2002.

[16] Bech, Søren, and Zacharov, Nick. *Perceptual audio evaluation-Theory, method and application*. John Wiley & Sons, 2007.

[17] Scheffé, Henry. "An analysis of variance for paired comparisons." *Journal of the American Statistical Association* 47.259, pp. 381-400, 1952.

[18] David, Herbert Aron. *The method of paired comparisons*. Vol. 12. London, 1963.

[19] Recommendation, CCIR Draft New. "Multi-channel stereophonic sound system with and without accompanying picture." *International Telecommunication Union, BS*: 775-1.

[20] Jordan, Dominic William, and Peter, Smith. *Mathematical techniques*. Oxford University Press, pp. 898-899, 1997.

[21] Moore, Brian CJ, ed. *An introduction to the psychology of hearing*. Brill, 2012.

[22] Werner, Stephan, Liebetrau, Judith, and Sporer, Thomas. "Audio-visual discrepancy and the influence on vertical sound source localization." *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*. IEEE, 2012.

[23] Russell, James A. "A circumplex model of affect." *Journal of personality and social psychology* 39.6 1161, 1980.