A COMPARISON OF PLAYER PERFORMANCE IN A GAMIFIED LOCALISATION TASK BETWEEN SPATIAL LOUDSPEAKER SYSTEMS

Joe Rees-Jones

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

Damian Murphy

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

ABSTRACT

This paper presents an experiment comparing player performance in a gamified localisation task between three loudspeaker configurations: stereo, 7.1 surround-sound and an equidistantly spaced octagonal array. The test was designed as a step towards determining whether spatialised game audio can improve player performance in a video game, thus influencing their overall experience. The game required players to find as many sound sources as possible, by using only sonic cues, in a 3D virtual game environment. Results suggest that the task was significantly easier when listening over a 7.1 surround-sound system, based on feedback from 24 participants. 7.1 was also the most preferred of the three listening conditions. The result was not entirely expected in that the octagonal array did not outperform 7.1. It is thought that, for the given stimuli, this may be a repercussion due to the octagonal array sacrificing an optimal front stereo pair, for more consistent imaging all around the listening space.

1. INTRODUCTION

As an entertainment medium, interactive video games are well suited to the benefits of spatial audio. Spatialisied sound cues can be used to fully envelop the player in audio, creating immersive and dynamic soundscapes that contribute to more engaging gameplay experiences. In addition to this, an international survey carried out by Goodwin [1] in 2009 suggests that video game players consider spatialised sound to be an important factor of a game experience. At the time of writing, the majority of video game content is able to output multichannel audio conforming to home theater loudspeaker configurations, such as 5.1 and 7.1 surround-sound. More recently, Dolby Atmos has been employed in a handful of titles such as *Star Wars Battlefront* (2015) [2] and Blizzard's *Overwatch* (2016) [3].

A potential benefit arises when considering the influence a multichannel listening system may have on a player's performance in the game world. If a player is able to better determine the location of an in-game event, such as the position of a narrative-progressing item or a potential threat, will this inform their gameplay decisions? From this concept, a novel idea for a comparative listening test was derived, in an attempt to assess the influence a spatial rendering system may have on a player's ability to localise a sound source in an interactive game environment. The test presented in this paper was based on the gamification of a simple localisation task, designed according to three core principles:

1. The player's objective is to locate as many sound sources as possible in the given time limit.
2. The player does not receive any visual feedback regarding the position of a sound source.
3. The player receives a final score determined by how many sound sources are found.

Three loudspeaker configurations commonly used for spatial audio rendering were compared in this study: stereo, 7.1 surround-sound and an octagonal array with a center channel placed at 0° relative to a front-facing listening position. Stereo and 7.1 were chosen as they represent popular, commercially available, rendering solutions used by video game players. The octagonal array was chosen for its relatively stable sound source imaging all around the listener, although it is not a standard configuration currently used for game audio. Player performance was quantified based on the number of correct localisations in each condition, represented by their final score.

The paper is structured as follows: In Section 2, the phantom image stability of loudspeaker arrays is discussed with an emphasis on surround-sound standards. Game design and the implementation of the three listening conditions is described in Section 3. Results, analysis and a discussion of the results are presented in Section 4. The paper conclusion is given in Section 5.

2. BACKGROUND

A listener's ability to infer the location of a sound source, presented using a physical loudspeaker array, is reliant on stable phantom imaging between two adjacent loudspeakers. By manipulating the relative amplitude of two loudspeakers (g_1 and g_2 in Figure 1), the illusion of a phantom (or virtual) sound source emanating at some point between them can be achieved [4].

The ratio of gain values (g_1 and g_2) between the two loudspeakers shown in Figure 1 is given according to Bauer's sine panning law [6]:

$$\frac{sin\theta}{sin\theta_0} = \frac{g_1 - g_2}{g_1 + g_2} \tag{1}$$

where θ is the perceived angle of the virtual source and θ_0 is the angle of the loudspeakers relative to a listener facing forward at 0°. If the listener's head is to be more formally considered then it is suggested that replacing the sin term with tan in (1) will provide more consistent imaging [5]. The actual gain values with a constant loudness can then be derived using:

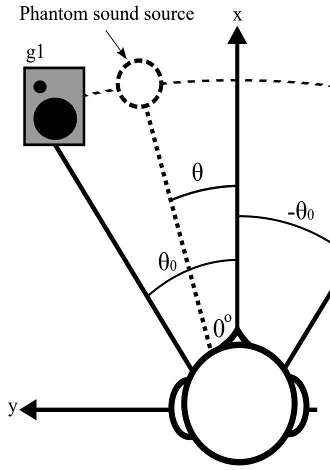

Figure 1: *The relationship between the desired angle of a phantom sound source (θ) and two loudspeaker angles ($\pm\theta_0$) [5].*

$$\sqrt{\sum_{n=1}^{N} g_n^2} = 1 \tag{2}$$

where N is the number of loudspeakers. The wider the angle between the two loudspeakers, the more unstable the phantom imaging becomes, in that sounds are perceived to 'pull' towards the closest speaker relative to the intended position of the sound source. It is widely accepted that the angle between two loudspeakers should not exceed 60° if stable imaging is to be preserved, where $\pm 30^\circ$ relative to a forward facing listener is optimal [7]. Imaging can be further improved by placing a centre loudspeaker, in-between the pair, at 0°.

For spatialised game audio, it is often the case that the mix will be rendered according to cinema listening standards, such as 5.1 and 7.1 surround-sound. Figure 2 shows the positioning of loudspeakers for 5.1 surround sound listening as suggested in ITU-R BS: 775 [8]. When considering the left and right surround channels (LS and RS in Figure 2), the angles between them exceed the recommended 60° for stable imaging. Such standards are appropriate for film/television viewing in that left, right and center loudspeaker channels are reserved for dialogue, music and diagetic effects whereas the surround channels are more generally used to emphasise the ambiance of a scene. Although mixing strategies vary between different video games, it is not uncommon for game audio to be rendered to every channel of a surround-sound system [9]. One aim is to give the player the impression of a reactive soundscape that dynamically responds to their physical input (i.e. through using a control pad). Consistent and stable phantom imaging around the listening space would therefore be desirable.

A review of the literature suggests that phantom imaging becomes unstable due to the wide angle between loudspeakers as is standard in surround-sound systems, especially when considering lateral sound sources. Cabot [10] assessed the localisation of both rectangular and diamond quadrophonic (4 loudspeaker) systems, finding phantom imaging to be most stable in the front quadrants but very unstable to the sides and rear of the listener. This result is further emphasised by Martin et. al [11] who considered image stability in a 5.1 surround sound system. Their results suggest phantom imaging is both reliable and predictable using the front three loudspeakers but is highly unstable when a sound at a position greater than 90° relative to a front facing listener is desired. An improved panning law for lateral virtual sound sources in 5.1 surround-sound was derived by Kim et. al and gave promising results, although the authors admit their relatively small subject base (5 participants) is insufficient to draw general conclusions [12]. Theile and Plenge [13] suggest that for more stable lateral imaging, sound sources intended to be perceived at $\pm 90^\circ$ relative to the listening position should be represented by a real sound source i.e. a loudspeaker. They propose an equally spaced arrangement of six loudspeakers to get a suitable 'all-around' effect. This configuration was extended by Martin et. al [14] to an equally spaced octagonal array with a front center speaker placed at 0° relative to the listener. The array was found to give relatively stable imaging around the listening space for amplitude-based panning algorithms.

The conclusions drawn from these studies provide evidence that a listener's ability to successfully localise a sound will be influenced by the phantom image stability of the loudspeaker array used. However, none of these studies asked participants to directly interact with audio stimuli by playing a game. Therefore it is of interest to investigate whether similar comparisons can be made between different loudspeaker arrays, with varying degrees of phantom sound stability, in the context of an interactive game task. In order to retain consistency with these studies, amplitude panning of interactive game audio was compared over stereo, surround-sound and octagonal loudspeaker arrays. Stereo is standard in all modern video game content and, according to Goodwin [1], has the highest user base amongst gamers. The configuration gives strong frontal phantom imaging due to the placement of the left and right loudspeaker at $\pm 30^\circ$ relative to the central listening position. However, imaging to the sides and rear is not possible due to the lack of loudspeakers at these positions. It would therefore be expected that a listener would find it difficult to locate a sound anywhere but within the $\pm 30^\circ$ of the stereo pair.

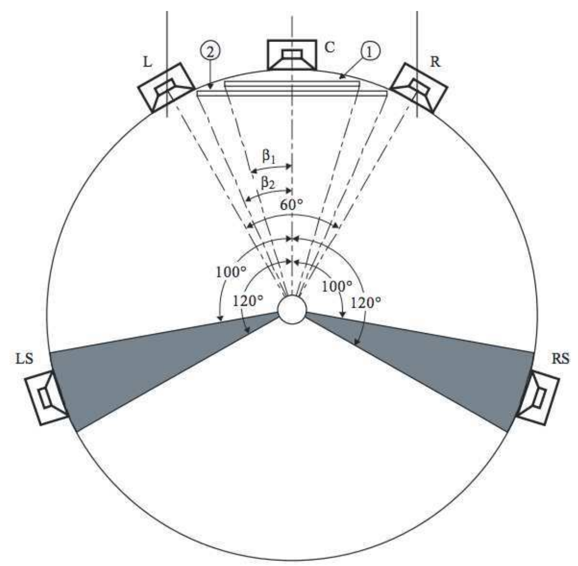

Figure 2: *The loudspeaker configuration recommended by ITU-R BS:*775 *for* 5.1 *surround-sound listening [8]. The placement of LS and RS is not optimal for stable imaging.*

7.1 surround-sound represents the current state-of-the-art in spatial game audio playback, hence its inclusion in the study. The format is an extension of 5.1 surround-sound through the addition of two loudspeakers either behind or to the sides of the listener [15], giving an improvement in spatialisation capability compared to stereo. The $\pm 30°$ stereo arrangement is retained, with an additional center channel placed in-between. For this study the subwoofer ('.1' channel) was not included, as it is intended for further defining low frequency effects which were not used.

As stated previously, an equidistant array of 8 loudspeakers arranged as an octagon gives stable phantom imaging all around the listener. This is an improvement on both 5.1 and 7.1 surround-sound where the inconsistent placement of loudspeakers leads to instability at the sides and rear. However, unlike stereo and 7.1 surround, octagonal arrays are not used to render audio in consumer gaming. Also, the loudspeakers in-front of the listener need to be spaced at a wider angle than those in stereo and 7.1 configurations if equidistant placement is to be achieved, with two loudspeakers placed at $\pm 90°$ for reliable lateral imaging. Therefore the trade-off in ease of localisation between more consistent imaging all around the listener, and the potential for higher resolution frontal imaging in 7.1, is of interest.

3. METHODS

This section outlines the localisation task participants were asked to complete and how it was implemented using a game-like virtual environment. The methods used to render game audio to the three listening conditions - stereo, 7.1 surround-sound and an octagonal array - are then covered. It was decided early in the design process that a custom-made game environment would be used. Previous experiments by the authors have used program material taken from commercially available video game content for current-generation gaming consoles. However, it is not possible to access the source code of such content making it difficult to determine the exact audio rendering methods used, beyond the way in which loudspeakers should be placed. The repeatability between participants is also questionable, along with potential learning effects that may occur due to multiple play-throughs of the same piece of game content. Creating a custom video game gave more control over the underlying mechanics/systems and the effectiveness of an octagonal loudspeaker array could be more easily explored.

For clarity, some game specific terms used in reference to the game design are defined here:

Game engine: A basic framework of code and scripts for creating video games, usually bundled as a complete software package. Handles the game's visual rendering, physics systems and underlying rules/mechanics.

Game object: Conceptually, objects refer to the game's building blocks. They act as containers for all the systems and code required to construct anything needed to make the game operate as intended, such as walls, characters, weapons or on-screen text [16].

Game world: The virtual environment/space in which the game is played. In the present study this refers to a virtual 3-dimensional space.

Player avatar: The player's virtual representation within the game world. The avatar's actions are directly controlled by the player, allowing the player to interact with and navigate through the game world.

Sound source: A game object placed at some position in the game world, from which sound is emitted.

3.1. Game Design

The virtual environment and underlying systems for the localisation task were designed and implemented using the Unity game engine [17]. Sound spatialisation and rendering was done separately in Max/MSP [18]. A single sound source was used in the game, the position of which changed as soon as it was successfully located by the player. The sound source was represented by a spherical Unity game object with a radius of 0.5 metres and its visual renderer turned off, ensuring that the source would be invisible to participants. The position of the sound source was always determined randomly within the boundaries of the game world, represented by a 20x20 metre square room. Random positioning was implemented so that players would not learn sound source positions after playing the game multiple times. The virtual room comprised of four grey coloured walls, a floor and a ceiling to serve a visual reference regarding the player's position within the game world. According to Zielinkski et al. [19], visuals can distract significantly from an audio-based task, therefore visuals were deliberately simplified.

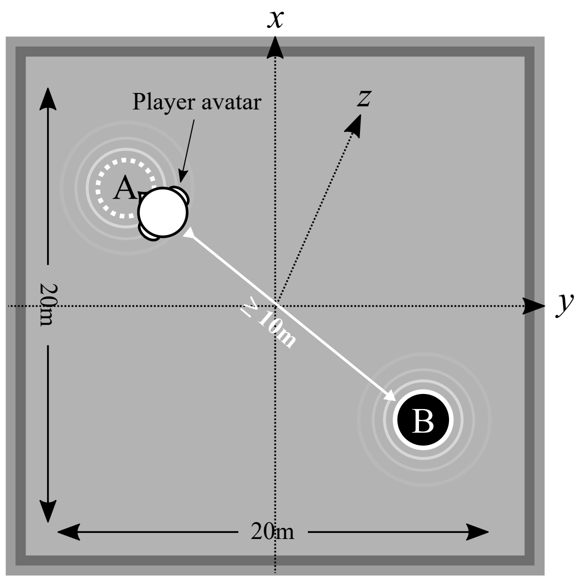

Figure 3: *A conceptual illustration of a player correctly locating the sound source in its current position (A) by entering its radius and pressing 'x' on the gamepad. The sound source then moves to a new random position (B) at least 10m away from the player's current position.*

Players were able to navigate the game world through the eyes of a virtual avatar, using a control system similar to those found in the majority of first-person point of view games. The position and rotation of the avatar, within the boundaries of the game world, could be controlled by the player using the left and right joysticks of a standard Playstation 4 gamepad. This allowed for full 360° movement in all directions on a horizontal plane. The gamepad's 'x' button was used to trigger a simple if statement within the game's code to determine whether the player had successfully found the sound source. If, upon pressing the 'x' button,

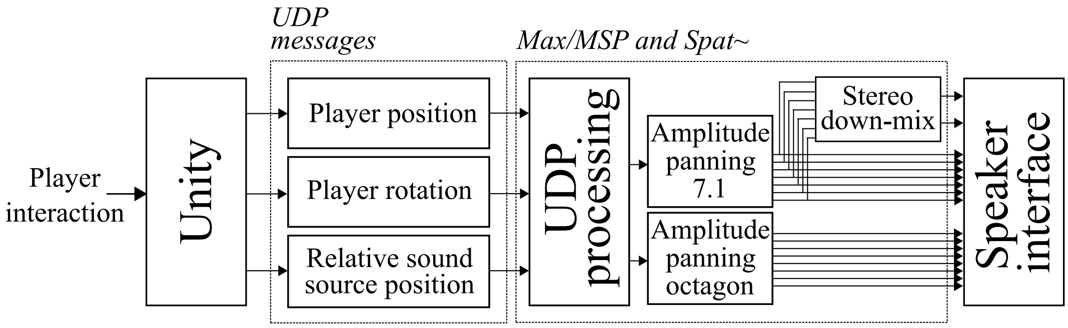

Figure 4: *Outline of data flow from Unity to Max/MSP to the loudspeaker interface. Coordinates from Unity were sent to the Max/MSP patch via UDP. UDP messages were then processed to pan a sound source at different positions for the three listening conditions.*

the player avatar was within the radius of the sphere representing the sound source's current location, the sphere would move to a random new location at least 10 metres away from the player, within the room's boundaries. Upon triggering this event, an on-screen value depicting the player's score increased by one. A top-down interpretation is illustrated in Figure 3, where position A represents the current position of the sound-source and position B is the new position. If the 'x' button was pressed and the player was not within the radius of the sound source then the current position was maintained with no increase in score. A count-down timer set to 2 minutes 30 seconds was also implemented. The timer was not displayed to players and once it reached 0, "Game Over" was displayed to the player, along with their final, score to signify the end of the game session. The game was played three times by each participant.

3.2. Game Audio Rendering

Game audio was rendered separately to the main game using the *Spatialisateur* (Spat~) object library for Max/MSP provided by IRCAM [20]. Communication between Unity and Max/MSP was achieved using the User Datagram Protocol (UDP). The player avatar's x, y, and z coordinates in the game world were packed and transmitted over UDP on every frame update of the game. This ensured the Max/MSP patch would be synced to the game systems and visuals. The x, y and z coordinates of the sound source relative to the player were sent to Max/MSP in the same way. A diagram of the data flow from Unity to Max/MSP is given in Figure 4. Players were asked to locate a sine tone at a frequency of 440Hz repeating every half a second with an attack time of 5 milliseconds to give a hard onset. A short delay was also applied to the tone, giving a sonar-like effect. An ascending sequence of tones were played to the player upon every correct localisation to give some auditory feedback as to their success, in-line with the increase in score. If the player was incorrect, a descending sequence was played. It was decided other effects commonly found in video games, like music, ambiance and footsteps, would not be included for this test, so as to not confuse the listener.

3.2.1. Sound Spatialisation

The amplitude panning law given in (1) can be extended for more than two loudspeakers using pairwise amplitude panning algorithms [7]. In this work, pairwise panning was implemented for the 7.1 and octagonal loudspeaker configurations using a '$spat.pan{\sim}$' Max/MSP object. The '$spat.pan{\sim}$' object takes a sound source (in this case the 440Hz repeating sine tone) as its input, and pans it according to x, y and z coordinates around the pre-defined loudspeaker layout. The x, y and z coordinates used for panning corresponded to the relative position of the sound source to the player, as transmitted from Unity via UDP. The number of loudspeakers and their placement around the listening area were defined for the panners as follows:

7.1 surround-sound: $0°$, $30°$, $90°$, $135°$, $-135°$, $-90°$, $-30°$

Octagon: $0°$, $45°$, $90°$, $135°$, $180°$, $-135°$, $-90°$, $-45°$

Angles used in the 7.1 condition were defined such that they conformed to the ITU-R BS: 775 for surround-sound listening [8]. The octagonal array was arranged in the same configuration used by Martin et al. [14], inclusive of a front centre loudspeaker at $0°$ relative to a forward-facing listener. Adjacent loudspeakers were defined equidistantly with an angle of $45°$ between them. The angles used for both conditions are reflected in Figure 5 by the 7.1 surround-sound and Octagon labelled loudspeakers.

3.2.2. Distance Attenuation

Since the player was able to move around the game world freely, it was necessary to include distance attenuation in the audio rendering. This made the sound appear louder as the player moved towards its source and quieter as they moved away. This was achieved by taking the inverse square of the relative distance (in metres) between the sound source and the player. This can be expressed in decibels (dB) using:

$$10 \log_{10} \left(\frac{1}{d^2} \right) \tag{3}$$

where d is the distance between the sound source and the listener. The same distance attenuation was used across the three conditions in order to keep changes in amplitude consistent. The amplitude of the sound remained constant as the player stayed within the radius of the sound source. This was implemented after informal testing, as it was found that otherwise, the sound would only ever reach maximum amplitude if the player was stood directly in the centre of the sound source position.

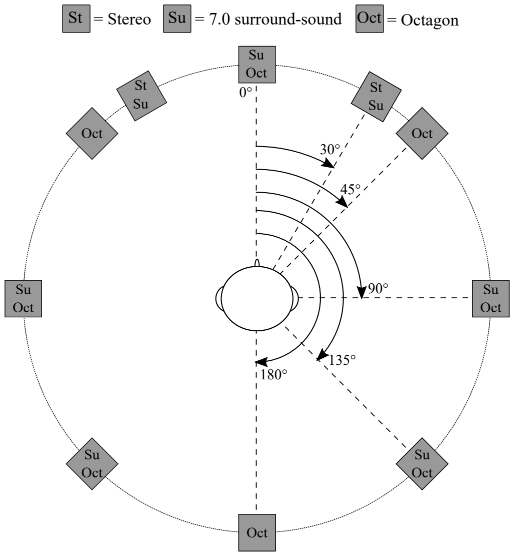

Figure 5: *Loudspeaker angles used for all three listening conditions. Angles are symmetrical to the left and right of a front-facing listener.*

3.2.3. Stereo Down-mix

It is not uncommon for a stereo down-mix to be generated from a game's 7.1 audio output. Audio content intended for surround-sound listening can be presented over any regular two-channel stereo system by down-mixing non-stereo channels. Additional channels are combined with the front left and right channels to ensure sound cues mixed for the centre and surround channels are not lost. As illustrated in Figure 4, the down-mix used for this experiment was done using the same 7 audio channels (not including the low-frequency effects) used in the 7.1 surround-sound listening condition, according to ITU-R BS.775-3 [14]. The left/right surround (Ls/Rs) channels and left/right back surround (Lbs/Rbs) channels were attenuated by 3dB and sent to the respective left and right front channels. The centre channel was also attenuated by 3dB and sent to the front left and right channels. This can be expressed as:

$$\begin{aligned} L_D &= L + 0.707C + 0.707Ls + 0.707Lbs \\ R_D &= R + 0.707C + 0.707Rs + 0.707Rbs \end{aligned} \tag{4}$$

where L_D and R_D are the down-mixed left and right channels, respectively.

4. EXPERIMENTAL PROCEDURE

A total of 24 subjects participated in the experiment, 16 of which were male, 6 female, and 2 non-binary. All subjects were aged between 18 and 35. Before participating, all potential subjects were asked if they were familiar with using a gamepad to control a game. If not, they were asked not to participate in order to reduce the amount of time needed to learn the game's control system. All provided a signature to confirm their consent.

Participants played the game in all three of the listening conditions - stereo, 7.1 surround-sound and octagon - but were not made aware of any of the conditions prior to, or during, the test. Repeated-measures test designs such as this are susceptible to learning effects, in that participant results may be influenced through being exposed to the same program material multiple times. To reduce this risk, the order of listening conditions was counterbalanced as suggested in [21]. This design with 24 participants and 3 listening conditions gave 4 participants in each counterbalanced group. Furthermore, a training session was provided, as described below.

Table 1: *Counterbalanced participant groupings for the three listening conditions.*

Allocation	Condition		
Group 1	Stereo	7.1	Octagon
Group 2	Stereo	Octagon	7.1
Group 3	7.1	Stereo	Octagon
Group 4	7.1	Octagon	Stereo
Group 5	Octagon	Stereo	7.1
Group 6	Octagon	7.1	Stereo

Each of the three game sessions lasted 2 minutes 30 seconds. A 'Game Over' message and the player's score (i.e how many times the sounds source was correctly located) were displayed on-screen at the end of each session. The number of correct localisations was output to a separate text file after each game session, giving each subject a final score for each of the three listening conditions. Once a subject had been exposed to all of the listening conditions, they were asked to state which of the three conditions they preferred, and provide any comments regarding the experiment.

Before the formal test began, participants were asked to complete a training session based on a simplified version of the game, allowing them to become familiar with the control scheme. The training version of the game took place in the same virtual room, with the addition of 5 coloured rings placed in its center and each corner. During the training, the sound source would only ever appear at one of these pre-defined locations. Participants were asked to move the in-game avatar to each of these locations and press the gamepad's 'x' button if they believed that to be the origin of the sound source. Once each of the pre-defined sound sources had been found once, the coloured rings were removed, and participants were asked to find the sound sources again, without a visual cue. Training was done in mono to eliminate the possible learning effect due to playing the game in an experimental condition more than once. The distance attenuation was preserved, allowing participants to familiarise themselves with amplitude changes as they moved closer to and further away from the sound source. The training session was not timed and only finished once a subject had found each of the 5 sound sources twice.

4.1. Apparatus

10 Genelec 8040a loudspeakers were arranged as shown in Figure 5, 1.5 metres from the central listening position. Those intended for 7.1 surround-sound listening conformed to ITU-R BS: 755 [8]. The Unity game and Max/MSP patch were run from the same Windows PC. Participants interacted with the game using a standard Playstation 4 gamepad connected to the PC via USB. Loudspeakers were driven by a MOTU PCI-424 soundcard. Visuals were

Table 2: *Wilcoxon signed-rank output for the mean-adjusted player scores. T is the signed-rank and p is the significance value. The z value is used to determine the significance value (p) and the effect size (r). A value of* 1 *in the h column signifies a rejection of the null hypothesis.*

Conditions compared		Median		*T*	*p(<.05)*	*z*	*r*	*h*
Stereo	7.1 Surround-sound	-1.167	1.167	17.5	<0.001	-3.559	0.514	1
Stereo	Octagon	-1.167	-0.667	87.5	0.202	-1.277	0.184	0
7.1 Surround Sound	Octagon	1.167	-0.667	173.5	0.043	2.023	0.292	1

presented using an Optoma HD200X projector, projecting onto an acoustically transparent screen. Loudspeakers positioned at 0° and ±30° were located behind the screen.

5. RESULTS

This section presents the results from statistical analysis of the player scores obtained during the experiment. Player scores (i.e. the number of correct localisations) were compared between pairs of the three listening conditions. Relationships between participants' success at the game and their preference for a listening condition are also given. All statistical analysis was performed using the statistics and machine learning toolbox in MATLAB.

5.1. Player scores

Player scores for the three listening conditions were first checked for normal distribution using a Kolmogorov-Smirnov goodness-of-fit test. Scores were found to be non-normally distributed (non-parametric), therefore Wilcoxon signed-rank tests were used to check for significances between pairs of conditions, as suggested by [21]. Scores were standardised before analysis due to the overall differences in scores between participants. This was done by subtracting a subject's mean score from their three individual condition scores. This ensured the relative distances between a player's own scores would be preserved. The null hypothesis for analysis was: *There is no statistically significant difference in the number of correct localisations between pairs of listening conditions*. The output from the Wilcoxon signed-rank tests are presented in Table 2. A value of 1 in the column labeled h of Table 2 signifies a rejection of the null hypothesis at the $p < 0.05$ significance level.

Analysis suggests there was a statistically significant difference in scores between stereo and 7.1 surround-sound as well as between 7.1 surround-sound and octagon conditions. Upon viewing the boxplot given in Figure 6 it can be seen that participants achieved higher localisation scores in the 7.1 surround-sound condition compared to both stereo and the octagonal array. This result implies that players were better at the game when listening to the audio using 7.1 surround-sound, in that they were able to more successfully localise sound sources. When considering stereo and 7.1 surround-sound, the effect size ($r = 0.514$) also signifies that listening condition had a large effect on player scores. This is higher than the moderate effect size observed between the 7.1 and octagon conditions ($r = 0.292$). This suggests scores achieved in the 7.1 condition were consistently higher in comparison to stereo than when compared to the octagonal array.

The null hypothesis could not be rejected rejected for the comparison between the stereo and octagon conditions, suggesting there was no statistically significant difference in player scores between the two at the $p < 0.05$ significance level. This is reflected by the boxplot in Figure 6, where it can be seen a similar range in values

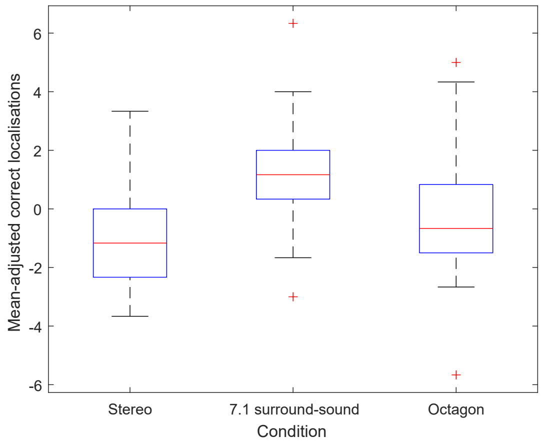

Figure 6: *Mean-adjusted distribution of player scores for the three listening conditions: stereo,* 7.1 *surround-sound and octagonal array. Analysis suggests the highest scores were achieved during the* 7.1 *surround-sound condition.*

is spanned by the stereo and octagon plots. The result implies that participant performance neither improved nor worsened between the two conditions, in that the number of correct localisations was similar.

5.2. Player preference

Once participants had played the game using all three listening conditions, they were asked to state which of the three they preferred, and were also encouraged to provide comments regarding their decision. In general, 7.1 surround-sound was the most preferred of the three conditions, as chosen by 70.8% of participants. Both stereo and the octagonal array were preferred by significantly fewer subjects. The preference scores for each condition are illustrated in Figure 7.

A number of participants commented that their preference was influenced by the the condition in which their highest score was achieved. Table 3 shows the percentage of highest player scores attained in each condition, alongside the corresponding percentage of overall preference. 60.4% of the highest scores were obtained in the 7.1 surround-sound condition, which was also most preferred by 70.8% of players. This implies that there was a preference for the condition in which the game was found to be easiest, which in the majority of cases was 7.1 surround-sound. Although the stereo condition contributed to 12.5% of the highest scores, the majority of subjects who achieved those scores stated that, perceptually, sounds were easier to localise in 7.1 surround-sound, hence it was

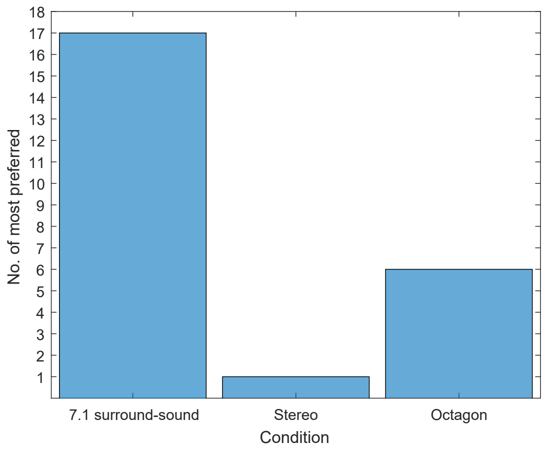

Figure 7: *Preference ratings given for each listening condition.* 7.1 *surround-sound was preferred by the majority of participants.*

Table 3: *The percentage of highest scores, alongside the percentage of preference ratings, for each condition.*

Condition	% highest score	% most preferred
Stereo	12.5%	4.2%
7.1 surround-sound	60.4%	70.8%
Octagon	27.1%	25.0%

more preferred. This may explain the minor discrepancy between the score and preference percentages for the stereo and surround-sound conditions.

5.3. Discussion

In general, players had greater success in the game when listening to audio over a 7.1 surround-sound loudspeaker array. The number of correct localisations was consistently higher for the majority of participants than when compared to stereo and octagonal loudspeaker systems. 7.1 was also the most preferred of the three listening conditions, with participant comments suggesting this was due to it being the condition in which the highest scores were achieved. It was expected that 7.1 would outperform stereo due to the increased number of channels available in the system, and from these results it can be said that players did benefit from using a more spatial listening array. However, the same could not be said in regard to the octagonal array.

As stated in section 2, it would be expected that the localisation of sound sources, especially those positioned laterally and to the rear of the listener, would be easiest when listening over an octagonal array of loudspeakers. However, the more consistent and stable phantom imaging that can be achieved using such a system seems to have had little impact on the results obtained in this experiment. Visuals were presented to the player using a stationary screen, therefore players were only ever required to look forwards. For this reason it may have been that those loudspeakers located directly infront of the listener's forward facing position were of most use in the localisation task. The front left and right loudspeakers of the octagonal array were spaced wider than the $\pm 30^\circ$ used in the 7.1 arrangement. Although both conditions made use of a centrally spaced loudspeaker at 0°, the increased resolution generated by the narrower angles between loudspeakers had in 7.1 surround-sound may have been more helpful than consistent imaging from all directions. Also, the directionality that can be achieved with a 7.1 array, although not perfect, would at least allow a listener to gain a vague sense of a sound source's general direction. It may therefore be the case that once a player had positioned their in-game avatar such that the sound was perceived to emanate at some point straight ahead, triangulating its specific location was then easiest using a more optimally spaced stereo pair. This observation is reflected in the comments given by participants, where it was stated on multiple occasions that it was easiest to triangulate/focus on the sound source in the 7.1 surround-sound condition.

It is important to note, however, that in comparison to commercial games, the game used in this experiment was a relatively simple example. Generally, modern games include more in-depth sound design and visual effects that work together in forming the entire game experience. It would therefore be of interest to determine whether the results obtained from this experiment could be replicated using a more complex game task, inclusive of more 'true-to-life' game systems. This would provide clarity as to whether the results from this study were dependent on the stimulus used, and is the proposed next step for this work. In comparison to previous work by authors, the use of a custom game environment was found to allow for far more control over experimental variables, and is therefore recommended for such studies.

6. CONCLUSION

This paper presented an experiment designed to determine whether enhanced spatial audio feedback has an influence on how well a player performs at a video game. Player performance was quantified by how many correct localisations of a randomly positioned sound source were achieved within a time limit of 2.30 minutes. This was compared between three listening conditions: stereo, 7.1 surround-sound and an octagonal array of loudspeakers. Results suggest that by using a more spatial listening system, player performance was improved, in that significantly higher localisation scores were achieved when using a 7.1 system in comparison to stereo. 7.1 was also consistently the most preferred of the three conditions by participants. However, the same result was not observed for the octagonal array. A possible explanation is that the angles between the front three loudspeakers used in the octagonal array (-45°,0°,$+45^\circ$) were wider than those in the 7.1 surround-sound system (-30°,0°,$+30^\circ$). Therefore frontal sound source imaging may not have been as well defined in comparison to the 7.1 condition.

7. REFERENCES

[1] Simon N Goodwin, "How players listen," in *Audio Engineering Society Conference: 35th International Conference: Audio for Games*. Audio Engineering Society, 2009.

[2] "*Star Wars*™ Battlefront™ in Dolby Atmos for PC, note=Available at https://www.dolby.com/us/en/categories/games/star-wars-battlefront-in-dolby-atmos.html, year=accessed March 28, 2017," .

[3] "Dolby Atmos Unleashes the Power of Blizzard's Overwatch®," Available at https://www.dolby.com/us/en/categories/games/overwatch.html, accessed March 28, 2017.

[4] Francis Rumsey and Tim McCormick, *Sound and Recording: An Introduction*, Focal Press, Oxford, UK, fifth edition, 2006.

[5] Ville Pulkki and Matti Karjalainen, "Localization of amplitude-panned virtual sources 1: stereophonic panning," *Journal of the Audio Engineering Society*, vol. 49, no. 9, pp. 739–752, 2001.

[6] Benjamin B Bauer, "Phasor analysis of some stereophonic phenomena," *The Journal of the Acoustical Society of America*, vol. 33, no. 11, pp. 1536–1539, 1961.

[7] Francis Rumsey, *Spatial Audio*, Focal Press, Oxford, UK, 2001.

[8] ITU-R, "Recommendation BS. 775-3: Multichannel stereophonic sound system with and without accompanying picture," *International Telecommunications Union*, 2012.

[9] Mark Kerins, *The Oxford Handbook of New Audiovisual Aesthetics*, chapter Multichannel Gaming and the Aesthetics of Interactive Surround, pp. 585–604, Oxford University Press, Oxford, UK, 2013.

[10] Richard C Cabot, "Sound localization in 2 and 4 channel systems: A comparison of phantom image prediction equations and experimental data," in *Audio Engineering Society Convention 58*. Audio Engineering Society, 1977.

[11] Geoff Martin, Wieslaw Woszczyk, Jason Corey, and René Quesnel, "Sound source localization in a five-channel surround sound reproduction system," in *Audio Engineering Society Convention 107*. Audio Engineering Society, 1999.

[12] Sungyoung Kim, Masahiro Ikeda, and Akio Takahashi, "An optimized pair-wise constant power panning algorithm for stable lateral sound imagery in the 5.1 reproduction system," in *Audio Engineering Society Convention 125*. Audio Engineering Society, 2008.

[13] Gunther Theile and Georg Plenge, "Localization of lateral phantom sources," *Journal of the Audio Engineering Society*, vol. 25, no. 4, pp. 196–200, 1977.

[14] Geoff Martin, Wieslaw Woszczyk, Jason Corey, and René Quesnel, "Controlling phantom image focus in a multichannel reproduction system," in *Audio Engineering Society Convention 107*. Audio Engineering Society, 1999.

[15] Tomlinson Holman, *Surround Sound: up and running*, CRC Press, FL, US, 2014.

[16] "GameObjects," Available at https://docs.unity3d.com/Manual/GameObjects.html, accessed March 28, 2017.

[17] "Download Unity Personal," Available at https://store.unity.com/download?ref=personal, accessed April 7, 2017.

[18] "Cycling 74: Tools for sound, graphics, and interactivity," Available at https://cycling74.com/products/max/, accessed April 7, 2017.

[19] Slawomir K Zielinski, Francis Rumsey, Søren Bech, Bart De Bruyn, and Rafael Kassier, "Computer games and multichannel audio quality-the effect of division of attention between auditory and visual modalities," in *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*. Audio Engineering Society, 2003.

[20] Jean-Marc Jot and Olivier Warusfel, "Spat: a spatial processor for musicians and sound engineers," in *CIARM: International Conference on Acoustics and Musical Research*, 1995.

[21] Andy Field and Graham Hole, *How to Design and Report Experiments*, Sage, London, UK, 2003.