

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/228810839>

Can the perception of depth in stereoscopic images be influenced by 3D sound?

Article in *Proceedings of SPIE - The International Society for Optical Engineering* · February 2011

DOI: 10.1117/12.871960

CITATIONS

27

READS

274

3 authors, including:



Nick Holliman

Newcastle University

64 PUBLICATIONS 1,800 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Creative Fuse North East: Initial Report of the Project [View project](#)

Can the perception of depth in stereoscopic images be influenced by 3D sound?

Amy Turner, Jonathan Berry, Nick Holliman

Innovative Computing Group, School of Engineering and Computing Sciences,
Durham University, United Kingdom

ABSTRACT

The creation of binocular images for stereoscopic display has benefited from significant research and commercial development in recent years. However, perhaps surprisingly, the effect of adding 3D sound to stereoscopic images has rarely been studied. If auditory depth information can enhance or extend the visual depth experience it could become an important way to extend the limited depth budget on all 3D displays and reduce the potential for fatigue from excessive use of disparity.

Objective: As there is limited research in this area our objective was to ask two preliminary questions. First what is the smallest difference in forward depth that can be reliably detected using 3D sound alone? Second does the addition of auditory depth information influence the visual perception of depth in a stereoscopic image?

Method: To investigate auditory depth cues we use a simple sound system to test the experimental hypothesis that: participants will perform better than chance at judging the depth differences between two speakers a set distance apart. In our second experiment investigating both auditory and visual depth cues we setup a sound system and a stereoscopic display to test the experimental hypothesis that: participants judge a visual stimulus to be closer if they hear a closer sound when viewing the stimulus.

Results: In the auditory depth cue trial every depth difference tested gave significant results demonstrating that the human ear can hear depth differences between physical sources as short as 0.25 m at 1 m. In our trial investigating whether audio information can influence the visual perception of depth we found that participants did report visually perceiving an object to be closer when the sound was played closer to them even though the image depth remained unchanged.

Conclusion: The positive results in the two trials show that we can hear small differences in forward depth between sound sources and suggest that it could be practical to extend the apparent depth in a stereoscopic image by using 3D sound, providing a controlled way to compensate for the depth budget limits on 3D displays.

Keywords: Stereoscopy, Depth Perception, 3D Audio, 3D Images, Evaluation, 3D Cinema, 3D TV, 3D Games

1. INTRODUCTION

The challenge for stereoscopic image creators is that there is a limited range of image disparity,^{1,2} or depth budget, on all 3D displays. This is because, for a number of reasons,³ the entire geometric range of disparity available cannot be used without risking visual strain or fatigue. If we can find additional ways to emphasise the visual perception of depth we can hope to extend the impression of depth in a given 3D Cinema, 3D TV or 3D game system without exceeding the disparity depth budget.

In this paper we investigate whether 3D sound (i.e. an auditory depth cue in addition to the visual depth cue of disparity) can influence the apparent perception of depth in a stereoscopic image. When we use the terms 3D sound and auditory depth we are specifically interested in sound sources separated in depth in the forward direction from the listener. In the following study we use physical sound sources so that the findings are not confounded by artefacts from simulated 3D sound or headphones.

Further author information: Send correspondence to Nick Holliman
E-mail: n.s.holliman@durham.ac.uk, Telephone: +44 (0) 191 334 4287

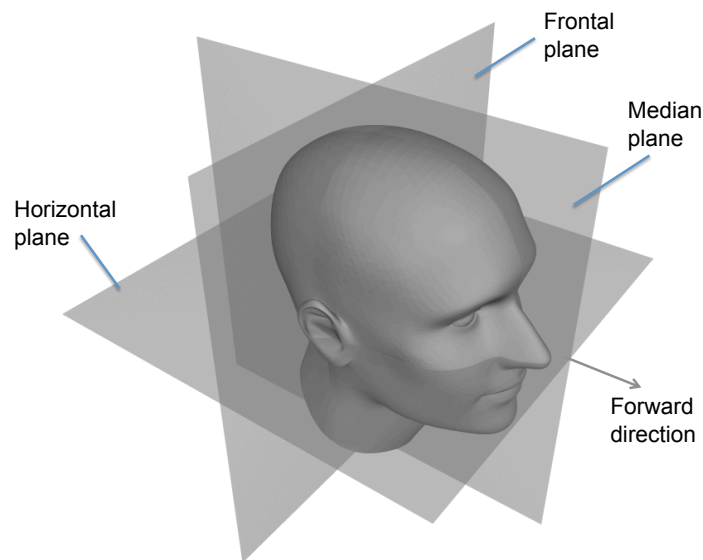


Figure 1. The division of space around the head using horizontal, median and frontal planes.⁵ Humans can detect angular difference between sounds as close as 1 degree of azimuth apart on the horizontal plane. However, in our investigation we are interested to know whether it is possible to detect a difference in depth (distance) between two sounds in front of the listener along the forward direction.

1.1 Auditory localisation

There have been a number of psychological studies that investigate how it is we are able to perceive elements of direction and angular location from hearing. Rosenzweig⁴ concluded that people find it easy to locate whether sounds are coming from the left or the right and how far left and right they are, but have far greater difficulty with up and down, and behind and in-front. The reasoning behind this is fairly obvious; the location of our ears are to the left and the right of our head. In describing our perception of sound sources, we will refer to the space around the head in terms of planes as shown in Figure 1.

Two cues we use to localise sound are the intensity of sound in each ear, and the time at which the sound reaches each ear. The first is known as the inaural intensity difference (IID) and the latter of these is known as the inaural time difference (ITD). A sound further to your right, in a normal environment, will reach the right ear before it reaches the left. In terms of judging distance, it has been reported that the intensity cue is the most important.⁴

Surprisingly, our ability to localise sounds is also dependent upon the sound itself. Stevens⁶ found that people find some tones far more difficult to locate than any other. When presented with a clicking noise, participants were able to locate the sound almost immediately, but, when presented with a tone, there were a far greater proportion of errors and a larger average in location time. Further analysis allowed them to conclude that there were significantly greater errors for tones of frequencies between 2000Hz and 5000Hz. The frequency of the tone also affects which cue, inaural time difference or intensity difference, we use to localise the sound. Low frequencies have a relatively long wavelength in comparison to the head and thus they are able to diffract around the head so that the sound reaches both ears. However, high frequencies have a short wavelength in comparison to the head and thus produce a shadow on the opposite side of the head where the sound is unable to diffract.⁵

Localisation on the horizontal plane is mostly reliant on the ITDs and IIDs produced by the sound. Most studies have shown that people find it easiest to locate sounds below 1000Hz and above 5000Hz on the horizontal plane. The Duplex theory stated by Rayleigh in 1907,⁵ explains that this is because ITDs are most effectively used for sound localisation at low frequencies and gradually becomes less useful as frequency increases. The opposite is true for IIDs, as the frequency decreases, IIDs become less useful. As binaural hearing relies on the use of both of these, we find we have a curve whereby midrange frequencies are less easily located. We measure

sound localisation using Minimum Audible Angle and the angle measured on the horizontal plane is referred to as an azimuth. That is the smallest angle of difference between two sound-source locations which can be binaurally recognised. Mills⁷ presented participants with sinusoids at 0, 30, 45, 60 and 75 degree azimuths and then a second sound slightly to the left or right. He found that the minimum audible angle for azimuths at 0 degrees azimuth could be as good as 1 degree.

1.2 Auditory Depth Perception

We refer to the depth of an object by either its absolute or its relative distance. This means we either calculate the distance of an object directly or we calculate it relative to another object. Moore⁸ reports there are four main cues to auditory distance perception:

1. Sound pressure level. The greater the level the shorter the perceived distance. The inverse square law stated that as the distance doubles, the sound pressure level decreases by about 6db.⁸
2. Amount of reverberation. The greater the ratio of direct to reverberant energy, the shorter the perceived distance. This usually gives an absolute cue.
3. Spectral shape of the received signal. The greater the high frequency content, the shorter the perceived distance. This usually gives a relative cue.
4. Binaural cues. The greater the ITD or IID the shorter the perceived distance.

There has been evidence to suggest that human auditory depth perception can be greatly aided by action tasks such as walking towards an object and generally produces a more accurate perception. This also applies in reverse, when the object moves it provides additional information which would not be available when static and thus produces more accurate perceptions.⁸

1.3 Virtual Audio Systems

Virtual audio systems can come in many forms, from very complex (and expensive) pieces of equipment, to careful calculations as to the layout of speakers. The Head-related transfer function (HRTF) refers to the calculations used to relate the ear to the sound source.⁹ As we have previously mentioned, the brain calculates the location of the sound using the intensity of sound to the ear and the inaural time difference (see Auditory Depth Perception). These factors alter as the individual moves their head and thus, if the sound system doesn't accurately portray the location of the sound from all angles, perception of sound could be incorrect. The binaural room impulse response (BRIR) represents the effect of the sound in a particular acoustical environment.⁹ It is important to highlight the effects the environment has on the sound. For example, a sound will sound different to the individual if it is played in an open field compared to an anechoic room.

For stereo loudspeaker based systems, we have to take into account that sounds from the left speaker are able to reach the right ear and vice versa. There are some non-true 3D systems already in existence which give the effect of 3D but are only used for entertainment based purposes. There are a number of different techniques used for developing 3D loudspeaker systems, each attempts to minimise crosstalk and the effects of HRTF. Transaural audio attempts to eliminate crosstalk signals by adding a delayed and inverted version of the crosstalk signal to the opposite loudspeaker.⁹ This is done recursively to eliminate the crosstalk of the crosstalk elimination etc. However, although theoretically this should completely eliminate crosstalk, the HRTF means that when the individual moves their head, the crosstalk changes.

In a normal surround sound speaker layout it is assumed that the sound can come from all directions around the listener. This gives the listener the impression of being immersed in the environment. However, visually the listener isn't able to see the full environment around him/her, just the screen area itself. Thus, in order to enhance 3D visualisations it is more important to give the impression of the image at the screen rather than through the listener. This new layout is preferable as it allows the sound to accurately be positioned within the visual scope of the listener.

In this study we only use physical loudspeakers placed at the position we want the sound to come from so that we can be sure of the exact location of the source.

2. RELATED WORK

We found a limited number of sources that investigate the influence that auditory cues have in depth perception, these include the following.

2.1 Comparison of Auditory, Visual, and Audio-Visual Navigation in a 3D space.

Grohn, Lokki and Takala¹⁰ investigate participants ability to locate moving sound sources in a virtual room and also the effect of a distracting stimulus. Participants were presented with a navigation task in a virtual environment whereby they had to search for gates in a molecular structure. For each gate they were either presented with an audio cue, a visual cue or both together. They were able to conclude from their results that it was easier for people to locate moving sources using both the audio and the visual cue in a virtual environment.

In evaluating the results, they also looked at the navigation paths of each participant. These indicated that people had different uses for each cue. People used the auditory cue to approximate the location and then the visual cue in the final approach. Overall, they were able to suggest that auditory highlighting of objects would work well in a virtual environment and best when combined with the visual cue for the final specific localisation.

2.2 Aurally aided visual search in three-dimensional space.

Bolia, D'Angelo and McKinley¹¹ conducted their investigation in an airforce base for military purposes (i.e. advantage to pilots etc) and thus had access to advanced apparatus. Participants were invited to sit in the centre of a 2.3 m radius geodesic sphere enclosed in a soundproofed 6.7m chamber. A speaker was located at all 272 vertices of the sphere and 4 LEDs were placed 5cm above the surface of each speaker. Each participant was presented with 3 auditory conditions; non-audio whereby there were only visual cues, free field audio whereby both the visual and audio cues were presented, and virtual audio whereby the audio cues were given through the use of headphones. Within these conditions they were presented with; 1, 5, 10, 25 or 50 visual distracters.

Overall they found that there was an average of over 95% accuracy rate for all conditions. However the most interesting findings were from the reaction times; they were able to further demonstrate the use of the auditory system for redirecting gaze. With the use of auditory cues, all searches, no matter the complexity, could be completed comparatively more quickly. The addition of auditory cues reduced the search times by factor 6 or more.

The conclusion they were able to draw was, "Thus it makes sense to consider the utility of spatial audio cueing when designing displays for use in environments in which efficient acquisition of visual targets is important."¹¹ Clearly this shows that there is a link between auditory and visual cues; although as all targets were equidistant from the viewer the result does little to directly answer our question about stereoscopic depth perception.

2.3 An experimental study on the role of 3D sound in augmented reality environment.

Zhou et al¹² undertook a study which aims to achieve similar results to this study although there are a number of key differences from our study. Firstly, the study is conducted using augmented reality where virtual objects are visualised in a real environment. They did this via use of a head mounted display and a tracker which recognised markers placed in the room, and drew images on them. The images themselves were monoscopic. Secondly, although the study investigated the effects of 3D sound on depth perception, they did this via use of headphones. When using headphones, the brain uses a different mechanism to locate the sound source called lateralisation whereby the sound is located inside the head.⁵

Participants used the head-mounted display to judge the relative depth of two virtual telephones on a desk. Half of the participants were given 3D sound and half were given no sound at all. They found that those who had 3D sound were 2.5 times better at judging the relative depth of two objects than those without 3D sound. Both the number of correct responses and the times to respond showed significant results, thus suggesting that depth perception was significantly enhanced by 3D sound. Also, the difference in times to respond suggested that the addition of 3D sound significantly reduced the level of uncertainty with depth judgements.

The results are encouraging in that they show 3D sound can be a depth cue, but the use of headphones and the monoscopic display mean the conclusions of this study cannot be transferred directly to a stereoscopic 3D viewing situation such as TV or cinema.

2.4 An analysis of cues to auditory depth perception in free space.

Coleman¹³ looked into a number of the cues used in assessing auditory depth and was able to draw some interesting conclusions. One of the cues looked into was amplitude; this can be modelled by a $1/R$ loss where R is the distance from point P . This subsequently means that loss in amplitude increases by 6db for each doubling of distance. Overall, amplitude loss is the most established cue for auditory depth perception. Frequency spectrum, the cue provided by the change in frequency due to the distance of an object, was concluded to be a good relative judgement of depth so long as sufficient experience of the stimulus had been accumulated. For binaural cues (ITDs and IIDs), the conclusions were very interesting. The binaural information presented to the brain did not alter for any other reason than a change in the depth of the object. This means that any affects on the stimulus did not affect the perception of depth, only the depth change itself could affect the brains perception.

Overall, this study acquired a lot of information about the cues to auditory depth perception but concluded, as is often the case, that significantly more research needed to be conducted. This was especially the case for pinna based cues as at the time of writing the paper, no quantitative evaluation of pinna effects on depth perception existed. However, Coleman suggested an interesting link between auditory depth perception using binaural cues, and visual depth perception using binocularity.

2.5 Depth perception-finding a design goal for sound reproduction systems.

Nielsen¹⁴ noted that depth is often badly reproduced in sound systems and undertook a controlled study using physical sound sources. Four loudspeakers were placed in the forward direction from the listener at distances of 1.0 m, 2.0 m, 3.5 m and 5.0 m respectively. Sounds were presented in random order from the loudspeakers and listeners were asked to indicate the “locatedness” of the sound they heard using a simple diagram on a graphic display. This diagram led to some quantisation of results as it indicated specific distances which guided responses.

Neilsen reported that there was substantial inter-listener difference in his results and this increased with source distance from the listener. However, the results indicated that participants seemed able to perceive sound source locations in depth up to at least 2.0 m distance. There was also indication of an acoustic horizon for depth perception at longer distances.

While the absolute distance reproduction studied by Neilsen is of interest we are equally interested to know how close two sound sources can be and still be distinguished relatively as being nearer or further. If we can confirm small differences are distinguishable at short distances, typical of viewing 3D displays, then we can then study the influence nearer and further sound sources have on visual depth perception in stereoscopic images.

2.6 Our Objectives

Given the limited amount of research we have found on 3D sound and auditory depth separation thresholds our objectives here are to begin to answer the question posed in the title of this paper, “Can the perception of depth in stereoscopic images be influenced by 3D sound?”, specifically we:

- investigate if relative differences in depth can be detected using auditory cues alone.
- investigate the influence of auditory cues combined with visual cues have on the perception of depth.

3. EVALUATING AUDITORY INFORMATION AS A DEPTH CUE

The results from our literature survey suggest that sound can be a cue to depth but none of the investigations we have so far found give detailed information on the threshold distance that sound sources must be apart for them to be perceived to be at different depth locations. Therefore, before we begin to look at the combination of sound with stereoscopic images, we investigate human depth perception using auditory cues alone. We will do this using a simple setup of two loudspeakers placed at a range of distances from each other as shown in Figure 2. Subjects will be asked to judge whether a sound played came from the nearer or further speaker. Our hypotheses for the experiment are:

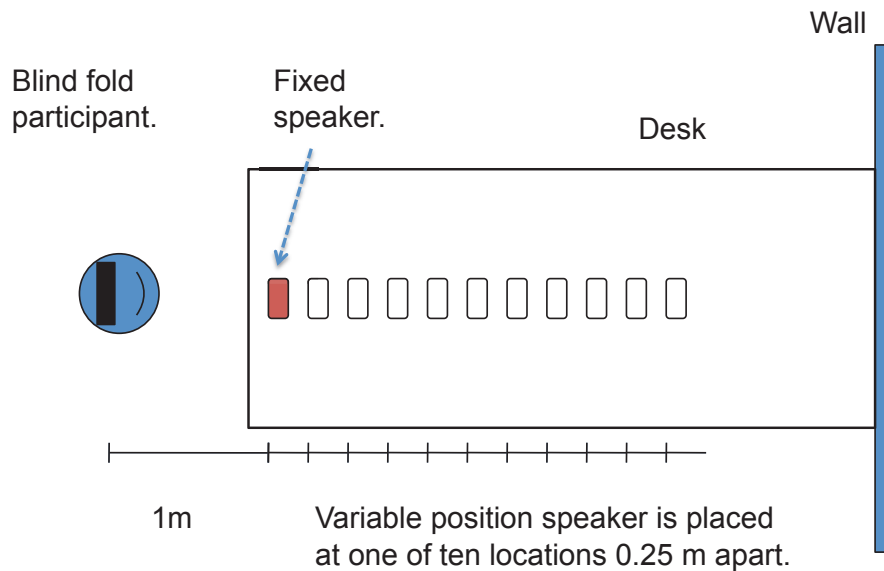


Figure 2. The layout of the apparatus used for evaluating auditory information as a depth cue. The participants sit blindfolded in a stationary chair. A first loudspeaker is placed 1m from the participant and a second loudspeaker is placed randomly at one of ten positions 0.25 m apart. A sound is played three times from each speaker in random order; each time the sound is played participants are asked to judge if the sound has come from the nearer or the more distant speaker.

- H_0 : null hypothesis, participants will perform no better than chance at judging the depth difference between two sources at a set distance apart.
- H_1 : experimental hypothesis, participants will perform better than chance at judging the depth difference between two sources at a set distance apart.

3.1 Apparatus

The system used includes:

- Creative 64MB Blaster X-Fi Elite Pro soundcard
- Logitech Z-5500 Digital speakers including 24.4cm (10) ported sub-woofer driver.

The participant sits 1.0 m from the first of two loudspeakers. The second loudspeaker is then placed at varying distances from the first speaker on the opposite side to the participant. The distances are 0.25 m, 0.5 m, 0.75 m, 1.0 m, 1.25 m, 1.5 m, 1.75 m, 2.00 m, 2.25 m and 2.50 m. Ensuring the speakers were located at the same azimuth as each other avoided the possibility of the results being as a consequence of ITDs and IIDs.

The Open AL library was used with C to create a program which would play the sound out of each of the speakers in a random order. For each of the 10 distances, a dog bark sound is played: one bark out of each speaker 3 times; so a total of 6 per distance. The program internally randomises the order in which the sound is played out of each speaker, thus preventing the participant being able to work out the order, and avoiding any order effects or effects from familiarisation. The dog bark was chosen because of its multi-frequency content, single tone sources are known to be difficult to localise.

To make the sound come out of the correct speaker, we use OpenAL to position the sound at the computers perceived location of the speaker. So once the speaker has been randomly selected, the sound source is set to the speaker location and played just once. It was very important to ensure that the location the sound was set to was precisely the centre point of both speakers in OpenAL otherwise the sound would be different. This would

Speaker separation (metres)	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50
Mean score $\bar{x}$	4.3	4.6	5.3	4.95	5.05	5.4	5.55	5.6	5.8	5.7
d.o.f.	19	19	19	19	19	19	19	19	19	19
t-value	3.901	6.024	11.89	7.610	9.180	10.25	12.07	17.08	30.51	15.06
p-value	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005
Accept H1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Accept H1 after BC	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 1. The results of the trial evaluating whether auditory information could be used as a depth cue. The independent variable is the speaker separation, i.e. the distance in metres between the fixed and variable speaker positions. The dependent variable is the number of attempts (out of 6) where participants correctly identified which of the two speakers the sound came from.

give the participant an advantage in being able to detect the difference in sound between the speakers because any difference detected would not be due to the change in depth but frequency and volume.

Once the program played the sound, it recorded the speaker used in a file. This was so that later the correct speaker could be compared to the responses given by the participants to calculate their scores.

3.2 Procedure

Before participating in the experiment, participants took part in a hearing test in order to ensure they had normal hearing. The hearing test used was the online hearing-aid test provided by the British Society of Hearing Aid Audiologists.¹⁵

Twenty participants sat in a chair 1.0 m from the nearest speaker which was on the desk top. Participants were then blindfolded; this was to ensure that participants were not using any visual information about the speaker location which may allow them to use previous knowledge of sounds at that distance. Overall, this ensures that results are an effect of auditory information only.

Throughout the experiment, the participant does not move or remove the blindfold, and the near speaker is not moved. The second speaker is moved to one of the 10 locations which previously have been randomly ordered; this was to avoid any order effects between participants. The locations were randomised by placing them into an array and shuffling that array for each participant, ensuring complete randomisation. The dog bark was then played a total of 6 times at each distance, 3 from the near speaker and 3 from the far speaker, in a random order. Participants were asked to say aloud whether they thought the sound came from the near speaker or the far speaker, and this response was recorded.

3.3 Results

The results produced from the sound trial were a set of ten scores, marked out of 6, for each of the twenty participants. If the participant correctly identified the near speaker in each test at one speaker separation they would have a score of 6 for that separation. A one-tailed single-sample t-test was used to determine if the mean score for each speaker separation was different to chance.

Taking chance to be a result of $\mu = 3$ the null hypothesis (H_0) is that the participants scores are not significantly different to chance. For a result to be significant, the calculated t-value must be greater than the t-critical value, $t_{crit} = 3.883$, found at the chosen p-value, $p = 0.0005$ for degrees of freedom $v = n - 1 = 19$. As can be seen from Table 1, this is true for all distances and so the null hypothesis can be rejected in all cases with 99.9% confidence.

However, as we have conducted numerous t-tests on the same data, it is possible to cause a type 1 error. A type 1 error is when we reject the null hypothesis when it is in fact true, the probability of which is increased the more tests are performed on the data.¹⁶ To overcome this, we performed the Bonferonni correction on the data

which alters the p-value thus making it more difficult to reject the null hypothesis. Even when this is applied to all the groups, all still accept H_1 and thus for all distances used, the results are better than chance.

4. EVALUATING THE INFLUENCE OF 3D SOUND ON VISUALLY PERCEIVED DEPTH

Having demonstrated that we can perceive depth differences between sound sources in the forward direction we are now interested to investigate if this can be used to influence the perceived depth of visual stimulus. Our hypotheses for the experiment are:

- H_0 : null hypothesis, participants will not perceive a difference in the visual stimulus regardless of the position of the sound they hear.
- H_1 : experimental hypothesis, participants judge a visual stimulus to be closer if they hear a closer sound when viewing the stimulus.

4.1 Apparatus

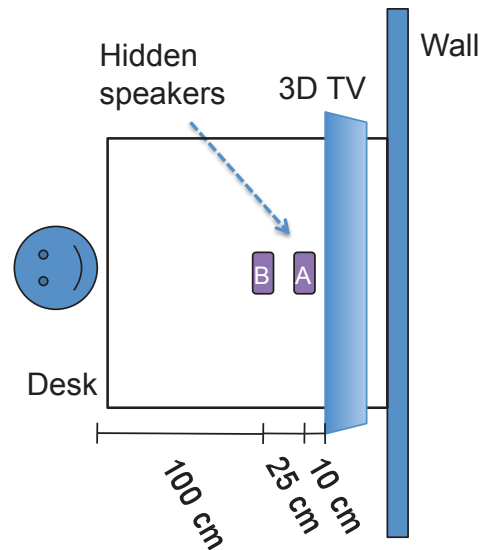


Figure 3. The apparatus used for the combined auditory and visual stimulus trial. The speaker A was placed 0.1 m from the 3D TV and speaker B 0.25 m beyond that. The speakers were hidden from the participant's view using a black cloth. Speaker A was raised slightly so that the physical presence of speaker B did not block speaker B's sound from the participant, as recommended by Neilsen.¹⁴

The apparatus used was the same computer system as in the previous trial but with the addition of a 3D display, a Hyundai 46" 3D LCD Xpol Virtual 3D monitor. This implements horizontal line interleaving and uses polarising glasses which participants wear throughout the trial. The monitor was positioned against a back wall and two speakers were positioned in front of it as shown in Figure 3. The speakers were set 0.25 m apart and hidden beneath a black cloth so that the participants could not see their positions. Speaker A was raised slightly so that speaker B did not block the sound from the participant.

A stereoscopic image of a well known brand of cell phone was displayed on the screen using OpenGL and the disparity adjusted so that it appeared to be situated in depth just over speaker A 0.1 m from the display surface. The participant was positioned 1.0 m from speaker B.

The sound used in this experiment was a telephone ring played using OpenAL, again a sound with multi-frequency components.

4.2 Procedure

In total there were fifteen participants ($n = 15$) and before taking part in the investigation each participant was required to pass three checks of their hearing, vision and stereo-acuity.

- The British Society of Hearing Aid Audiologists online hearing test: this plays 4 tones of varying pitches and all must be heard.
- Snellen eye test: a row of letters 9mm high must be read correctly from a distance of 6 meters (20/30 vision).
- Titmus stereoscopic vision test: all nine tests must be answered correctly.

In the trial each participant was shown 10 *alpha* tests and 10 *beta* tests in a different random order. Each test presented the same visual stimulus twice (the stereoscopic image of a cell phone) with a black screen shown between each stimulus presentation. The only difference between the two presentations was the sound played in the hidden speakers:

- In the *alpha* test the sound was played from speaker A for the first image presentation and from speaker B for the second image presentation.
- In the *beta* test the sound was played from speaker B for the first image presentation and from speaker A for the second image presentation.

After each test the participants were asked: “Which image displays the telephone nearest to you?”

Participants were allowed a small number of practice runs to familiarise themselves with the experimental procedure and the trial began.

4.3 Results

Participant ID	1	16	14	2	4	8	12	11	9	3	5	7	15	6	10
Score (%)	50	50	50	55	55	60	65	70	70	70	70	70	75	80	85

Table 2. The results of the trial evaluating whether 3D auditory information influences the apparent depth of a visual stimulus. The score is the percentage of forced choice decisions where participants reported a visual stimulus as being closer when the sound was presented in the closer speaker.

Results, shown in Table 2, were recorded as the percentage of correct answers. Where a correct answer was taken to be when a participant identified the image as being nearer when the sound was played through speaker B.

By chance a mean score in this trial would be $\mu = 50\%$, corresponding to the participants simply guessing the answer for every forced choice.

From Table 2 we calculate the sample mean to be $\bar{x} = 65\%$ with a sample standard deviation $s = 11.18$, using a one-sample t-test we find the computed t-value of $t = 5.196$, the degrees of freedom being $v = n - 1 = 14$. Using the one-tailed t-distribution table we find that for a p-value $p = 0.05$ the table t-critical value $t_{crit} = 1.76131$. As this is less than the computed t-value we can reject the null hypothesis with at least 95% confidence.

Therefore we can accept the experimental hypothesis H_1 : participants judge a visual stimulus to be closer if they hear a closer sound when viewing the stimulus.

In calculating these results we eliminated two participants who scored 100% and who, it became clear from the post-trial questionnaire, had prior knowledge of the experiment.

5. DISCUSSION

In setting out to understand the effect of 3D audio on stereoscopic image perception we found limited literature investigating the topic of 3D sound in relation to forward depth discrimination using auditory cues alone. This motivated our study to establish whether forward depth perception is possible using auditory cues alone. In addition while we found studies reporting positive influences of sound on vision we found no studies looking at the combination of binaural and binocular depth cues.

Our first trial investigated auditory depth perception and demonstrated that it is possible to judge a depth difference between two sound sources as close as 0.25 m at 1 m from the listener. The data suggest that the greater the depth difference between the sources the more reliably a depth difference can be judged.

Having established this we then undertook our second trial to determine if there is an influence on the perception of depth in a stereoscopic image from 3D sound. In this case we demonstrated that a 0.25 m difference between the image stimulus depth and the auditory stimulus depth led participants to report they were seeing the stimulus closer than in the case when the image stimulus and auditory stimulus were at the same depth. Note in both cases the image stimulus was unchanged.

Our results suggest that depth in a stereoscopic image can be made to appear greater if a nearer 3D sound is played with the image. This could be of real practical benefit because stereoscopic image creators must work within the strict limits of the depth budget available on each type of 3D display they use. Images viewed for significant periods of time at or above these limits can cause fatigue and the appropriate use of 3D sound could allow significantly more depth to be perceived without using visual disparities beyond the limits of the depth budget.

For this effect to be implemented commercially a reliable 3D sound cue is needed, this could come from an array of speakers reaching out towards the audience, or from synthetic 3D sound created in headphones, although as noted earlier this is a different phenomena. There remain many unknowns about the relationship between 3D sound and stereoscopic images and this should be a fruitful area for future research and development.

6. ACKNOWLEDGEMENTS

Amy Turner was supported in part by an EPSRC summer studentship to work in the Durham Visualization Laboratory (DVL) at Durham University, she is currently working for Microsoft UK. Jonathan Berry is currently an undergraduate student working on his Bachelors degree thesis at DVL. Our thanks to Julie Harris for statistical advice during her Carnegie Trust funded visit to DVL.

All trademarks used in this script are acknowledged to be the property of their respective owners.

REFERENCES

- [1] Woods, A. J., Docherty, T., and Koch, R., "Image distortions in stereoscopic video systems," in [*Stereoscopic Displays and Applications IV. Proceedings of the SPIE*], **1915**, 36–47 (Sept. 1993).
- [2] Jones, G. R., Lee, D., Holliman, N. S., and Ezra, D., "Controlling perceived depth in stereoscopic images," in [*Stereoscopic Displays and Virtual Reality Systems VIII*], *Electronic Imaging* **4297A**, SPIE (2001).
- [3] Hoffman, D., Girshick, A., Akeley, K., and Banks, M., "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of Vision* **8**(3), 1–30 (2008).
- [4] Murch, G., [*Visual and Auditory Perception*], Bobbs-Merrill, Indianapolis (1973).
- [5] Moore, B., [*An introduction to the psychology of hearing*], Academic Press, London, fourth ed. (1997).
- [6] Stevens, S. and Newman, E., "The localization of actual sources of sound," *American Journal of Psychology* **48**, 297–306 (1936).
- [7] Mills, A., "On the minimum audible angle," *Journal of the Acoustic Society of America* **30**, 237–246 (1958).
- [8] Moore, B., [*Hearing*], Academic Press, London (1995).
- [9] Kapralos, B., Jenkin, M., and Milios, E., "Virtual audio systems," *Presence: Teleoperators and Virtual Environments* **17**, 527–549 (December 2008).
- [10] Grohn, M., Lokki, T., and Takala, T., "Comparison of auditory, visual and audio-visual navigation in a 3D space," *ACM Transactions on Applied Perception* **2**(4), 564–570 (2005).
- [11] Bolia, R., D'Angelo, W., and McKinley, R., "Aurally aided visual search in three-dimensional space," *Human Factors: The Journal of the Human Factors and Ergonomics Society* **41**(4), 664–669 (1999).
- [12] Zhou, Z., Cheok, D., Yang, X., and Qiu, Y., "An experimental study on the role of software synthesized 3D sound in augmented reality environments," *Interacting with Computers* **16**, 989–1016 (October 2004).
- [13] Coleman, P., "An analysis of cues to auditory depth perception in free space," *Psychological Bulletin* **60**(3), 302–315 (1963).
- [14] Neilsen, S., "Depth perception - finding a design goal for sound reproduction systems," in [*AES Convention 90*], 9193–9102, Audio Engineering Society (1991).
- [15] "BSHAA online hearing test," (August 2009).
- [16] Abdi, H. and Salkind, N., [*Encyclopaedia of Measurement and Statistics*], Sage, Thousand Oaks, CA (2007).