Inducing audio-visual synaesthesia via augmented and virtual realities

A thesis presented to
The School of Information Technology
Deakin University

CAMERON HOLLINGSWORTH

Supervisor: Shaun Bangay

November 2018

Abstract

Gradually losing the fidelity of your senses, or losing your access to one of your senses entirely is an ever present part of the human experience. However, as technology advances, tools and strategies to delay, or circumvent this loss have arisen. Currently, partial solutions exist for both hearing and visual impairment, however the solutions for hearing impairment are often expensive, require a third party, and typically a conscious effort from individuals suffering from hearing impairment.

This research attempts to explore the potential of passive visual recognition of short phrases, simple tones, and spatial sound by inducing audio-visual synaesthesia via Augmented and Virtual realities, in pursuit of providing an alternate sensory pathway for audio information. The research is comprised of an evaluation of the visualization techniques and technologies utilized in related research areas and problems, and their respective applicability with regards to audio-visual synaesthesia.

Results are promising, yet further testing with a variety of sound characteristic to visualization mapping approaches is needed, along with a larger sample size, and a more significant repetition of testing with the same participants over time in order to build, a synaesthesia.

Acknowledgements

I'd like to express my gratitude and thanks to the following people:

My closest friends for keeping me upright, and being thoroughly critical of me during times of contentment, ensuring that I keep pushing myself to become better than who I was yesterday, and to do better than I have done up until now.

The academic staff at Deakin University, in particular (but in no particular order); Alexander Baldwin, Peter Eklund, Tim Wilkin, and Mick Hobbs for keeping me employed, thus keeping a roof over my head. And I suspect entirely without meaning to do so, for giving me the strongest source of motivation I've had to keep picking myself up, and telling myself to get moving, in the pursuit of sharing my knowledge in an academic environment.

A second time, to Peter Eklund for understanding that I needed to focus on my studies, and handling my leftover workload in recent weeks.

Shaun Bangay, for his insurmountable patience and tremendous if not misplaced faith in me. I'm still quite unsure what I've done to earn either, however I hope that this thesis is able to at least in part return that investment.

Finally, and perhaps most significantly, I'd like to thank Majid Mahdahvian, for imparting to me my sense of direction and purpose; To lead, to guide, to teach, and to help others.

Contents

Introduction

This thesis will evaluate inducing synaesthesia utilizing virtual and augmented reality, as a potential solution towards hearing impairment. In this pursuit, an evaluation of the current literature with particular regards to the following topics will be presented; Modality translation, what is synaesthesia, potential visualization strategies, and significant sound characteristics. Presently, no evaluation of virtual and augmented reality in this context has been described within the literature, and as such this thesis will attempt to fill this hole.

Following the literature review, this thesis will progress into the methodology of this study, describing its research question, constraints, design, and recruitment.

Continuing, this thesis will present its research findings, with discussion and evaluation of the individual tests, and then finally conclude.

Literature Review

The following chapter is an overview of the current literature in the surrounding and related areas to this thesis. The review will cover the following areas; Modality translation, Synaesthesia, Spatialised sound and sound characteristics, Visualization techniques available and their evaluation, training and its relevance to synaesthesia, and finally a conclusion.

Introduction

Humans rely on a variety of senses to interpret and navigate reality, with a heavy emphasis on our senses of touch, vision, and hearing. However as we age or experience traumas our senses often dull or can be entirely lost. Individuals are often born without access to or having one or more dulled senses. There are several approaches to solving this problem with reference towards vision and hearing loss; cochlea implants and hearing aids, guide dogs, Braille, sign language, and so on. These solutions can be either actively engaged; that is a solution that requires directed cognition and consideration in order to translate between modalities, such as a guide dog, or passively engaged as is the case with Braille. In the case of the later, it has been shown that passively engaged modality translation from the tactile modality can prompt neurological activity in the visual cortex of individuals with blindness (Sadato, Pascual-Leone et al. 1996), and further, that visual stimuli can prompt a similar response in the auditory cortex of deaf individuals (Finney, Fine et al. 2001). This passively engaged modality translation is Synaesthesia!

Several solutions for vision loss may be synaesthetic, utilizing the tactile modality to observe typically visual information in the case of walking canes, Braille, and other reading tools. However, hearing impairment lacks similar solutions, being left with only more actively engaged options, such as captioning; which is computationally expensive when automated or requires a human interface otherwise, and sign language; which requires also the assistance of third party individual. This results in these solutions often being unsuitable in real time and in real world environments. With recent developments in technology, another avenue of solution for vision and hearing impairment that may be passively synaesthetic has arisen; virtual and augmented reality (referred to as VR and AR respectively hereafter).

Existing research into this avenue has focused almost exclusively on visual impairment opportunities, with initial findings being promising for VR and AR spatial sound being a viable solution to partial or complete visual impairment (Rumiński 2015). Presently, little research has been conducted into the area of hearing impairment solutions via VR and AR technologies.

The surrounding and related literature will be discussed in the following capacity; The basics of modality, and modality translation, along with some examples. An overview of synaesthesia, its categorisation, and its history. A discussion on spatialised sound and its synaesthetic applications. And finally, a breakdown and review of visualization techniques associated with audio-visual modality translation.

Modality translation

Modality translation is the process of translating sensory information typically associated with one sensory means of interpretation to another. A straightforward example of this being written language; a visual representation of otherwise typically auditory information. Modality translation takes more complex forms however, such as complex visualizations derived from the characteristics of a sound field (Yatabe and Oikawa 2015), or a sound wave (Kaper, Wiebel et al. 1999) and (Ferguson, Moere et al. 2005).

Synaesthesia

Synaesthesia (sometimes written as Synesthesia) is a phenomena in which an individual experiences sensory information induced by mental events and external stimuli, a passive and involuntary instance of modality translation. For example, one may experience a sensation of colour when reading a letter, an example of visual->visual synaesthesia. (Grossenbacher and Lovelace 2001). Following on from the work of Grossenbacher and Lovelace, this experienced sensory information, will hereafter referred to as the 'concurrent', and the corresponding stimuli or event that inducing this concurrent will be referred to as the 'inducer'. Grossenbacher and Lovelace categorised synaesthesia into conceptual, and perceptive synaesthesia, determined by the nature of the inducer; sensory, or conceptual. Sensory synaesthesia occurs when the inducer is perceived via a sensory pathway, and Conceptual synaesthesia occurs when a thought or idea invokes synaesthetic outcomes. For example, if an individual were to see a Cow, inducing a concurrent "Moo" auditory experience, this would be an example of sensory synaesthesia. However if the same were to occur when simply imagining a cow, this would be a case of conceptual synaesthesia.

There are numerous cases of individuals with synaesthetic abilities, which have been covered extensively. These individuals known as 'Synaesthetes' typically experience synaesthesia unidirectionally (Grossenbacher and Lovelace 2001); such that inducer A induces concurrent B, but inducer B does not induce concurrent A. Synaesthetes have been documented exhibiting a range of synaesthesias, including visual-visual, audio-visual, and tactile-visual synaesthesia. It is however clear that synaesthesia is intimately linked with the visual modality.

Synaesthesia is often acquired via injury to the brain, or the surrounding sensory pathways, and while this can result in a permanent synaesthesia, if treated may be only temporary. (Afra, Funke et al. 2009).

Some initial research has been conducted into inducing synaesthetic outcomes, such as one experiment wherein participants experienced a visual 'illusion' flash when presented with an auditory stimuli, shortly after being presented with the flash and the same auditory stimuli simultaneously (Watkins, Shams et al. 2006)., however at the present, excluding a proposed prototype (Alex, Jose et al. 2013), there is no literature that covers audio-visual synaesthesia in an AR or VR environment.

Visualization techniques

There are several categories of visualization available to describe auditory information. The most commonly utilized audio visualizations I have categorised as follows; Speech to text, 2 dimensional grid based overlays, and 3 dimension geometrical overlays. As we are utilizing the visual field in order to provide an alternative sensory means of interpreting sound, we must evaluate the visualization techniques that are available for use in the context of enabling further sensory intake, without significantly inhibiting the information we are presently able to gather through our sensory pathways, to ensure we do not simply replace our hearing impairment with visual impairment.

Speech to text visualizations analyse an audio stream for vocalizations, and compute captions. In real time applications, this speech to text approach is too computationally expensive and limited to human vocalizations. I was unable to find any literature that

covered practical implementations of speech to text systems.
Our next option is a 2 Dimensional grid based overlay, hereafter "2DGBO". 2DGBOs use a grid displayed over an image or visual field to represent auditory information. Colour, geometric, or numeric information is displayed in the cells of grid, to represent details such as the source of a sound signal.(Laidlaw, Kirby et al. 2005) While 2DGBOs are a fantastic option, the information they enable to be presented can be limited, with significantly fewer opportunities to describe the audio stream due to the restrictive geometric consequences of a cell/grid based approach. Further; 2DGBOs are prone to obscuring the field of view, or details of the image. These limitations are particularly prominent with the colour and numerical approaches which fail to present much information without obscuring a similar level of information.
Similar to the 2DGBO, 3 dimensional geometric overlays, hereafter: "3DGO"s use a visualization that is displayed over a visual field or image to represent sound characteristics and attributes, however in contrast to the 2DGBOs, 3DGOs are typically geometric in nature, and are not restricted to a cell based grid, enabling a significantly higher level of detail to be represented, providing a level of visualization freedom not found in the 2DGBO model, and shown in (Kaper, Wiebel et al. 1999) and (Ferguson, Moere et al. 2005). The strongest argument for a 3DGO is that with the higher level of accessible potential detail, the same amount of information can be visualized, while occupying a smaller section of the visual field.

Sound spatialisation and wave characteristics

In order to visualize sound, we need to understand and represent the characteristics of a sound signal. The surrounding literature has described a VR/AR specific prototype for such a visualization, with size and colour of visualization artefacts representing amplitude and the dominant frequency of a given sound wave (Tepljakov, Astapov et al. 2016). However this approach has failed to address directional information which has been shown to be a significant advantage in environmental navigation (Rumiński 2015) (Dong and Guo 2016). Additionally building on this system may create one which when predominantly aimed at describing frequency varied information, as is the case for human speech, is unwieldy. Similarly sized artefacts may consistently obscure the colour of other nearby artefacts by overlapping them, and by using colour as a primary indicator of the most prominently varying data, this may make the visualization difficult to see in contrast in real world environments, and limit use for colour blind individuals. Instead using geometric manipulations to represent this data is more advantageous. Size, shape, smoothness, and so on, are capable of far more complexity, and thus more applicable to the vast range of frequency data that needs to be represented.

Training

One recurring element of the limited quantity of existing research is the lack of training of participants with the visualization, potentially impacting success rates with the various techniques. Individuals reported that they failed to understand visualization techniques utilized to describe sound characteristics, which may have impeded understanding and synaesthetic outcomes. It has been shown that individuals develop synaesthesia over a period of time, needing to learn to adapt to the sensory substitution method they utilize (Ward and Wright 2014) (Roberts and Shenker 2016), and this has failed to be addressed within the visualization and VR/AR literature entirely.

Conclusion

Synaesthesia enables individuals to experience sensory information via alternative sensory pathways, which may provide a means to interpret auditory information for individuals with hearing impairment. Simple auditory stimuli has been shown to induce synaesthetic outcomes for simple visual signals, and as such building a visualization representative of significant sound characteristics may produce synaesthetic outcomes for more complex sound signals. Of the categories of visualizations applicable, 3DGOs exhibit the largest number of the advantageous qualities we will require in order to create a viable passively synaesthetic solution for hearing impairment.

The singular synaesthesia VR/AR prototype that presently exists, uses colour to represent frequency, and size to represent amplitude. Instead, we will use size, shape, and smoothness to represent frequency, and colour amplitude, such that our predominantly frequency based variation is more fully supported and more accurately described by the visualization.

VR and AR provides a pathway to explore and potentially exploit inducing synaesthesia, by allowing individuals to expose themselves to virtual stimuli, in a predictable and repeatable manner, whilst utilizing visual, auditory, and spatial information to describe the VR and AR world, and I have been unable to find any literature pursuing this opportunity in full.

Methodology

The following chapter will describe the research question, the approach I have undertaken in order to answer this research question, and the associated data gathering process. This chapter will detail the experimental tools selected, with a discussion on how and why these tools were chosen. All custom built tools and software, their construction, and any integrated development environments utilized in construction will be described.

Research Question

Following on from the discussion at the end of the literature review, it is evident that a significant gap in audio-visual synaesthesia research exists, in particular research into utilizing VR and AR technology to achieve synaesthetic outcomes. This thesis will attempt to investigate the use of VR and AR in pursuit of these synaesthetic outcomes. It will do so by using a custom built virtual environment for the HTC Vive, which will collect data on modality translation and synaesthesia.

The research question is thusly:

"Is it possible to induce audio-visual synaesthesia via virtual and augmented realities?"

In order to answer this question, we have to investigate it constituent elements, as such, we are left with 2 and arguably 3 questions to answer. Firstly; Can we visualize sound such that a person can determine the location of a sound source visually? This will be achieved using a 3 dimensional spatial test, in which participants will use a visualization to determine the direction of a corresponding sound signal played from a series of fixed points in a computer generated random order. Secondly; Can we visualize sound to enable the appreciation of simple tones or dominant fundamental frequencies? This question will be addressed in a fundamental frequency test, in which participants will be asked to determine through a visualization, the relative pitch of a series of 4 simple, digitally generated tones, played in a computer generated random order. Finally, similarly to the second question, can we visualize a sound signal to enable the communication of simple phrases? Again, similarly to the previous question, this will be investigated by asking participants to use a visualization of a sound signal to determine the correct phrase from a small selection of potential phrases, played in a computer generated random order. Each of these tests will occur in a custom built virtual reality environment, utilizing HTC Vive VR technology. The aim of this research, is to determine if a VR or AR visualization is able to function as a sensory substitution device, enabling synaesthetic outcomes.
As it stands presently, there are few studies that evaluate inducing audio-visual synaesthesia, with a visual inducer and audio concurrent and no existing studies which evaluate VR or AR in this context. To this end, this study is unique.

Constraints

The design of this experiment is influenced by several constraints, detailed below.

Time Allocation

This research needed to be complete in October such that results would be able to be analysed and discussed in this thesis. However due to significant ethics approval delays, recruitment was delayed until the exam period, causing substantial recruitment problems, pushing back completion to the first week of November.

Sample size

As a further consequence of ethics approval delays, the recruitment has been severely limited, originally seeking a sample size of 30+, as presently it stands this research has a sample size of 7 participants. Additionally, experimental testing takes a significant amount of time, requiring participants for up to 30 minutes on average, meaning that the number of participants that are willing and may undergo testing on any given day is low. Fortunately, participants undergo a significant number of trials, such that each participant provides a reliable and robust data set.

Scope

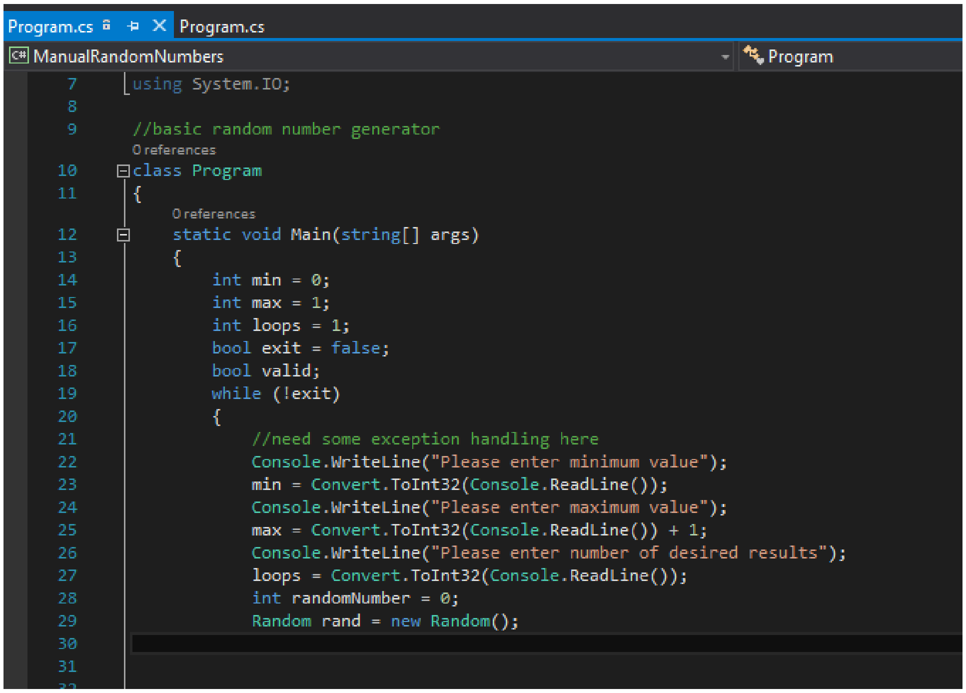
The scope of recruitment of this study is limited to participants with healthy vision, who are over the age of 18. Further, due to time constraints, the scope has been limited to testing only a single style of visualization after a one training session. However a series of tests enables us to rigorously test the visualization technique selected for is synaesthetic potential over a variety of sound signals, representing varying environments.

Design

This research will consist of a VR lab experiment with human participants. This VR experiment was selected as it was necessary to include the human component in order to test synaesthetic responses and outcomes. The inclusion of human participants carries along with it a series of ethical concerns. As such, a plain language and consent form is given to each participant prior to experimentation, with instructions attached describing the withdrawal of consent process, in addition to a set of written instructions and ethical guidelines that the experiment will adhere to. Further, a copy of identifiable data must be held temporarily, and the key that holds this stored in a password protected archive on a secure server at Deakin University. A written copy of verbal instructions is available for participants with hearing impairment, otherwise instructions are given verbally from the research team during experimentation. The experiment is broken into 2 phases, with 3 tests in each phase. Phase 1, the training phase, consists of the 3 tests with corresponding audio data, and excludes participant input and data collection. Phase 2 consists of the same 3 tests, however the corresponding audio is not provided, and participants are required to provide input to proceed through each trial of the tests. Each test utilizes a computer generated random sequence, and as such a small program which generates this sequence was developed for this study.

Random number generator

The following is a copy of the code of a small random number generator (figure 1), and it's console output (figure 2). This generator takes user input of a minimum and maximum values, and the number of desired results, before returning this number of random numbers within the specified range. This sequence of numbers is then manually entered into the trial sequences for the 3 tests, with both phase 1 and phase 2 using the same sequences.

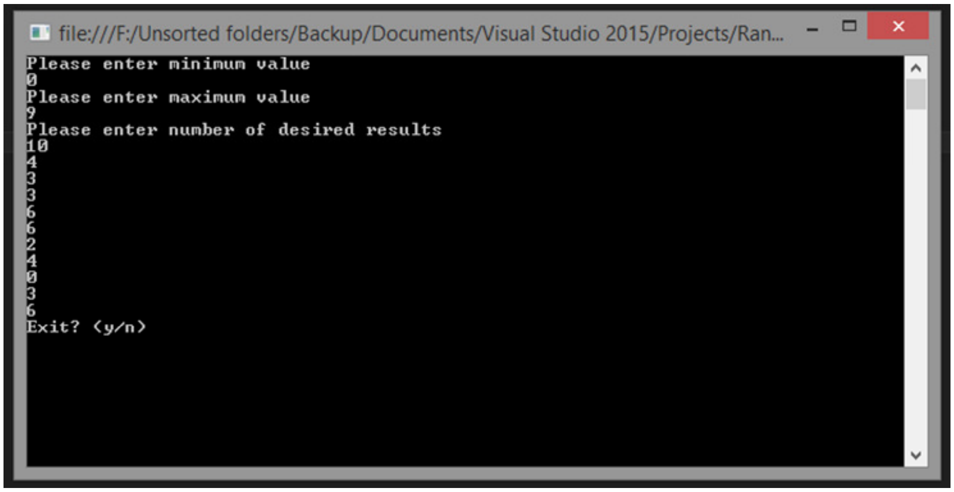


```
using System.IO;

//basic random number generator
class Program
{
    static void Main(string[] args)
    {
        int min = 0;
        int max = 1;
        int loops = 1;
        bool exit = false;
        bool valid;
        while (!exit)
        {
            //need some exception handling here
            Console.WriteLine("Please enter minimum value");
            min = Convert.ToInt32(Console.ReadLine());
            Console.WriteLine("Please enter maximum value");
            max = Convert.ToInt32(Console.ReadLine()) + 1;
            Console.WriteLine("Please enter number of desired results");
            loops = Convert.ToInt32(Console.ReadLine());
            int randomNumber = 0;
            Random rand = new Random();


```

Figure 1: Random number generator code.



```
Please enter minimum value
0
Please enter maximum value
9
Please enter number of desired results
10
4
3
3
6
6
2
4
0
3
6
Exit? (y/n)
```

Figure 2: Random number generator console output.

Virtual and lab environment

The 3 dimensional virtual environment created to facilitate this experiment is particularly simple, with the vast majority of the experimental functionality being derived from scripts. Participants are emerged into a virtual environment via a VR headset device, in this case a HTC Vive. Additionally participants will be required to use a set of over-ear headphones, in this case the Vive Deluxe audio strap or the HyperX Cloud II. The virtual environment contains a blue skybox, and green plane, but is otherwise empty (figure 3) The skybox is blue to emulate our physical sky, and the plane green to emulate a grass field, which forms a baseline for common real world visual contrast. The participant who is placed at and controls the rotation of the camera (figure 4) in the center of the environment, has a user interface attached to the camera. However this user interface only serves to transition between tests, display potential answers in phase 2 testing, and to be utilized in generating the visualization. The physical lab environment itself consists of the Deakin University Burwood campus VR lab, located in build B, room B4.08. The lab contains a dedicated VR room scale section of cleared space and floor, in which the testing takes place. While participants must stand for the first test in order to turn safely, they are able to be seated for the second and third test for each phase.

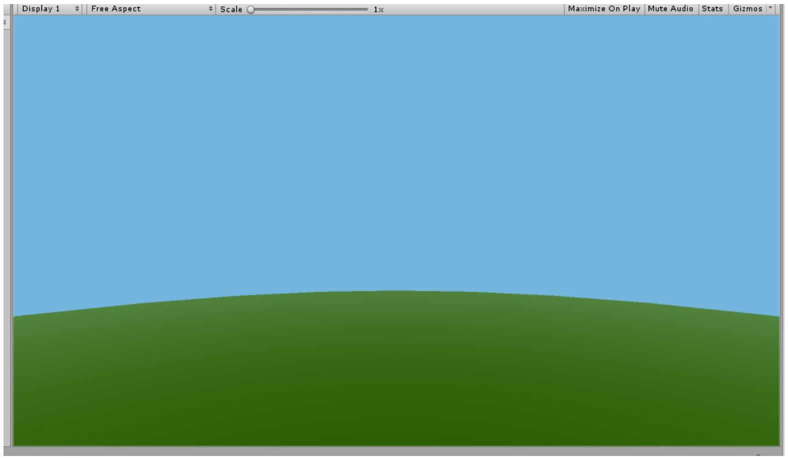


Figure 3: The participants view of the blue skybox and green plane. Note the UI is not currently being utilized, however this is uncommon during testing.

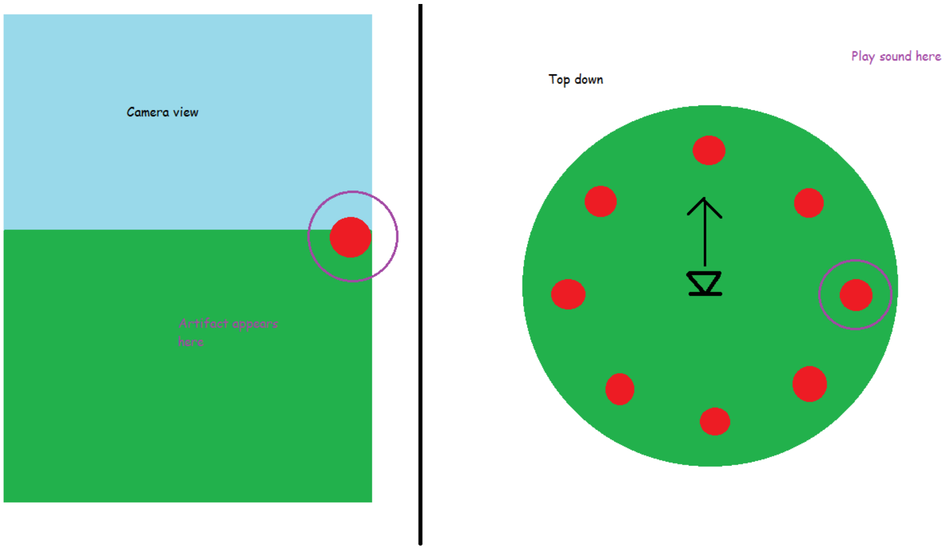


Figure 4: Right - Top down view of environment with the locations of potential sound sources marked in red.
Left - Camera view describing the relationship between the 3DGO and physical environment.

Visualization and audio assets

A graphical overlay displayed on top of the participants virtual viewing space will facilitate a visual representation of the audio information. The visualization itself will consist of a series of coloured spheres of varying size which will be created at the edge of the visual field, and move towards the center of the visual field. However, as these visual artefacts move towards the center of the visual field, they will be gradually reduced in size as to prevent them from increasingly obscuring the visual field as they move closer to the more vital regions of it, and are removed when they get to close to the center (figure 5). These visual artefacts are created at a rate of 10 per second, or at 100ms intervals. As such, they represent 100ms segments of audio data, in particular, the colour of the artefacts represents the amplitude or loudness of the audio segment, ranging from red to yellow, red representing yelling or talking loudly, and yellow a whisper. The initial size of the artefact represents the frequency or pitch of the audio segment. These visual artefact assets will be developed in unity itself, consisting of simple geometrical shapes such as spheres.
The corresponding audio assets will be recorded and developed via the personal recording equipment and appropriately licensed software of a member of the research team, and then manually analysed in 100ms segments, to produce the data used by the visualization to represent the sound signals. The sound signals for the spatial and fundamental frequency tests will be produced using the GNU licensed Audacity software chirp function, using frequencies that represent dominant or fundamental frequencies of a common range of musical notes (Michigan Technological University 2018).

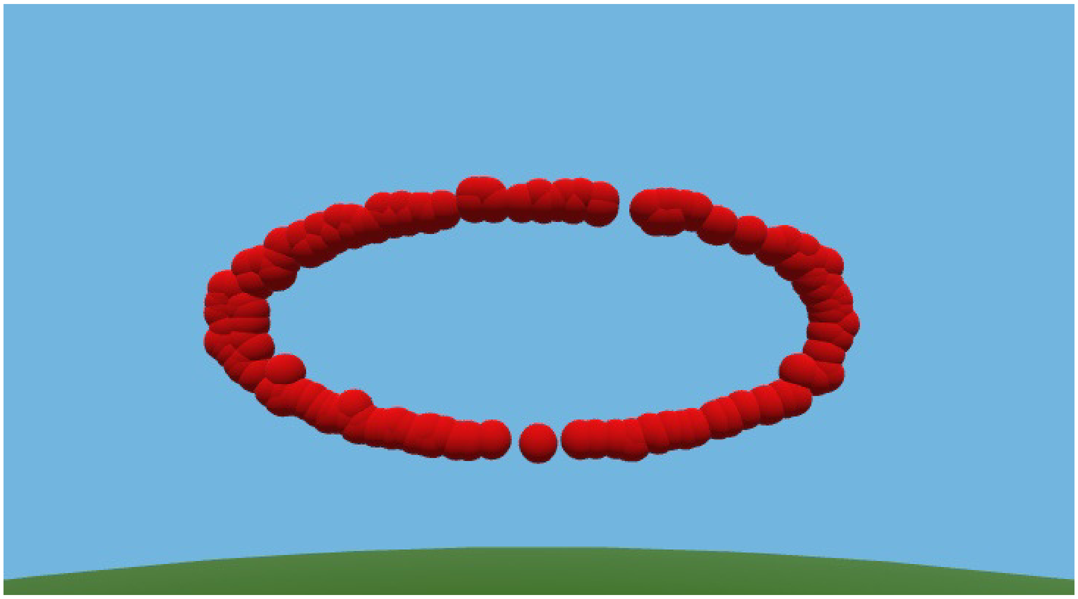


Figure 5: A ring of visual artefacts demonstrating the inner bound of the 3DGO visualization

Spatial sound

The first test is the Spatial sound test, in which users are asked to identify correctly the direction from which a sound originates from over a series of trials. The source of the sound will be determined by a fixed computer generated random series which will select which of the sound points it will originate from. In this test the participant will answer each trial by rotating their view towards the sound source (figure 6) and then selecting a button to record their answer (only applicable during phase 2 testing), thus beginning the next trial.

This experimental test answers the research question:
“Can we visualize sound such that a person can determine the location of a sound source visually?”

It does so by recording the rotation and response time of participants, enabling an analysis on both the accuracy of their answers, and if time is a factor in this accuracy, which indicates innate understanding or lack thereof.

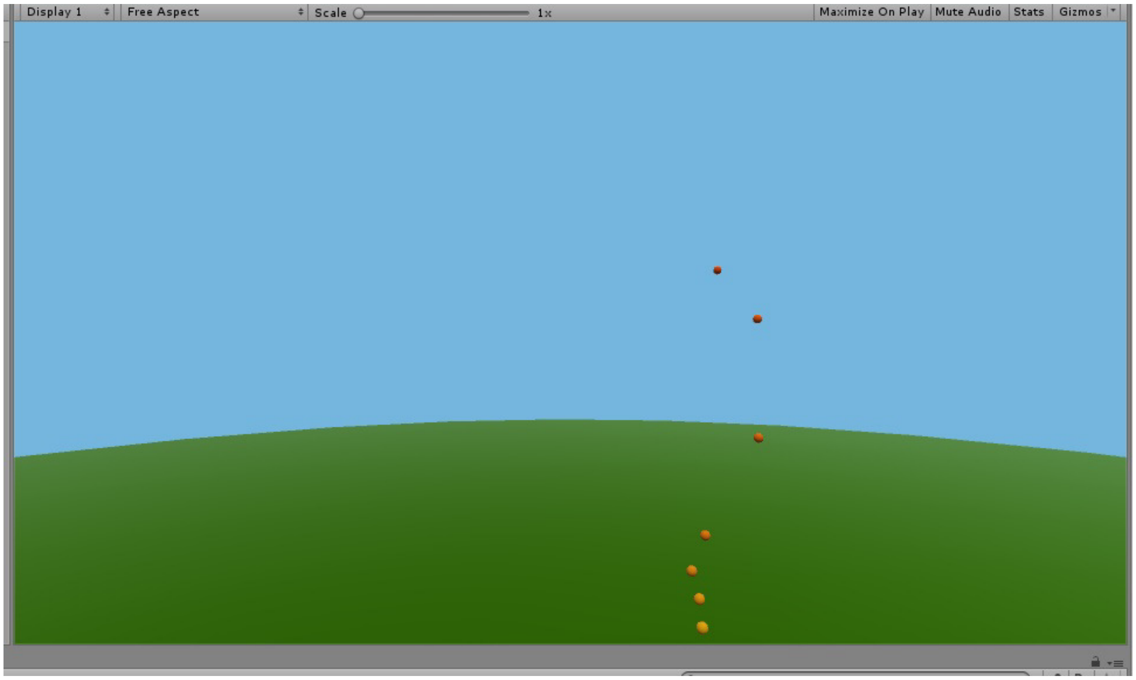


Figure 6: Spatial sound test, camera movement changes spawning point of artefacts

Fundamental frequency

The second experimental test will ask participants to determine the qualitative nature of a sound over a number of trials, by asking them to describe which of a set of fundamental frequency chirp tones they are ‘hearing’ via the 3DGO visualization. These fundamental frequency chirp tones which approximate musical notes, will be derived from a single musical octave, and each note will be produced using the chirp function in audacity using a single frequency. In phase 2, participants will answer the fundamental frequency trial by selecting the corresponding button in the environment using a HTC Vive controller following prompts from the UI (Figure 7), at which point the following trial will begin.

This experimental test answers the following question:
“Can we visualize sound to enable the appreciation of simple tones or dominant fundamental frequencies?"

It does so by recording the participant answer and response time, enabling an analysis on both the accuracy of their answers, and if time is a factor in this accuracy, which may indicate innate understanding or lack thereof, in determining the qualitative nature of a simple sound derived from singular frequency chirp tones.

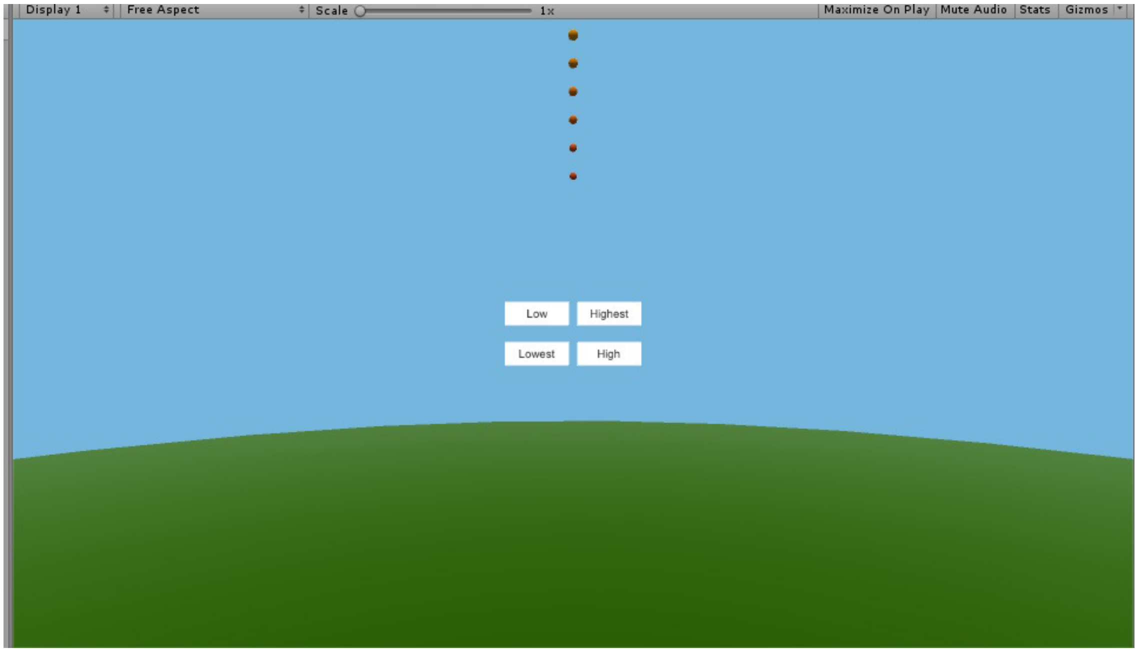


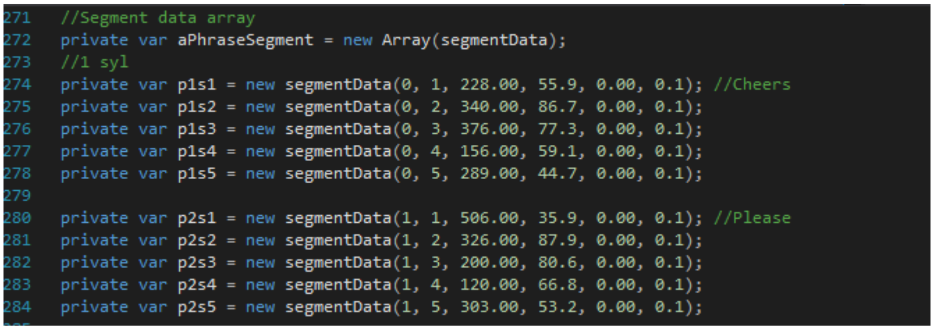
Figure 7: Fundamental frequency test, buttons correspond to potential answers

Phrase recognition

The third and final experimental test will ask participants to recognise simple common phrases from a pre-set list over a number of trials. During the phrase test there is significant variation in the artefacts generated within a singular sound signal. Each signal has been manually recorded, and analysed in 100ms segments, with the data being hard coded into the visualization script (figure 8). These phrases will be given to participants in a predetermined and computer generated random order consistent across all participants and sampled from a set list of a small number of phrases. In phase 2, each trial will require the participant to select which phrase that the participant believes they are hearing via the 3DGO visualization (figure 9).

This experimental test answers the question:
"can we visualize a sound signal to enable the communication of simple phrases"

It does so by recording the participant answer and response time, enabling an analysis on both the accuracy of their answers, and if time is a factor in this accuracy, which may indicate innate understanding or lack thereof, of individual phrases based on a visual channel interpretation of the audio characteristics of each individual phrase.



```
//Segment data array
private var aPhraseSegment = new Array(segmentData);
//1 syl
private var p1s1 = new segmentData(0, 1, 228.00, 55.9, 0.00, 0.1); //Cheers
private var p1s2 = new segmentData(0, 2, 340.00, 86.7, 0.00, 0.1);
private var p1s3 = new segmentData(0, 3, 376.00, 77.3, 0.00, 0.1);
private var p1s4 = new segmentData(0, 4, 156.00, 59.1, 0.00, 0.1);
private var p1s5 = new segmentData(0, 5, 289.00, 44.7, 0.00, 0.1);

private var p2s1 = new segmentData(1, 1, 506.00, 35.9, 0.00, 0.1); //Please
private var p2s2 = new segmentData(1, 2, 326.00, 87.9, 0.00, 0.1);
private var p2s3 = new segmentData(1, 3, 200.00, 80.6, 0.00, 0.1);
private var p2s4 = new segmentData(1, 4, 120.00, 66.8, 0.00, 0.1);
private var p2s5 = new segmentData(1, 5, 303.00, 53.2, 0.00, 0.1);
```

Figure 8: Sound signal segment data from the experimental environment script

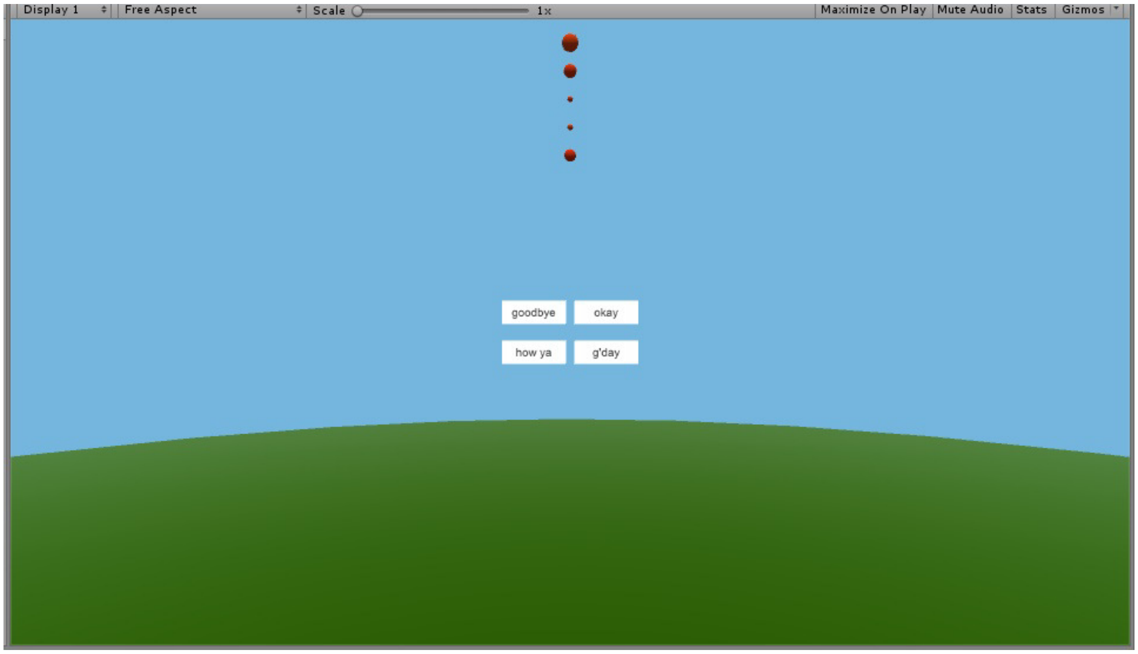


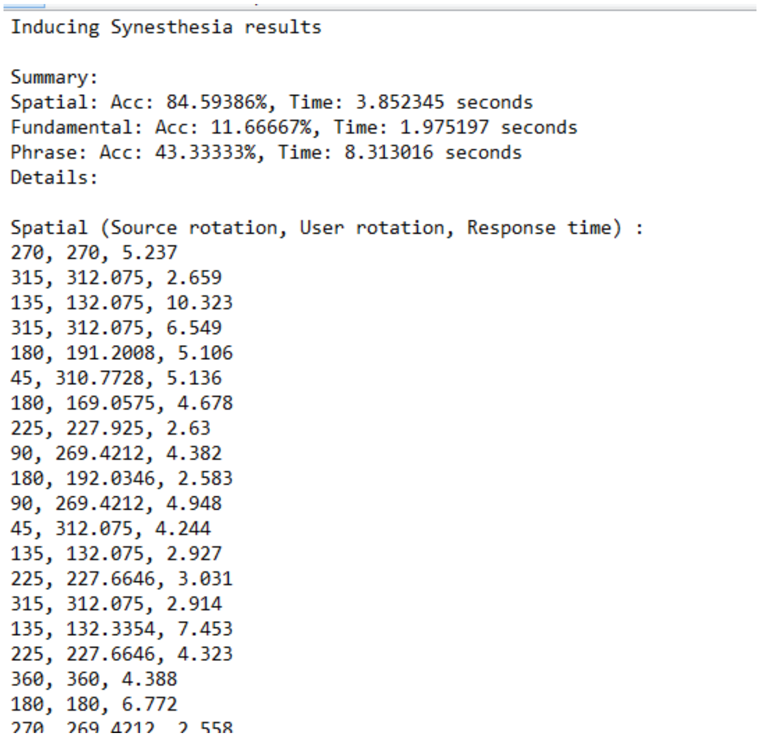
Figure 9: Phrase test, visual artefacts show significant variance, buttons correspond to potential answers

Data recording and processing

The virtual experimental code utilizes JavaScript to log all recorded data to file in a human readable format. However JavaScript is no longer supported by Unity, and as such the retrieval and processing of this logged data is handled by a custom script written in c# (C Sharp), and run within the experimental environment (figure 10), in order to process Quaternions data structures which are not effectively supported outside of the unity environment. This processed data is then logged to another file as a summary and raw data format (Figure 11), which can be imported to excel via the windows clipboard.

```
//Use : separators
switch (lineBuf[6])
{
    case 'S':
        {
            //Run spatial, rotation. and timestamp,
            sRotation[trial] = calcRot(secondLineBuf, testDataPosModifier);
            sAcc[trial] = calcRotAcc(sRotation[trial], SourceAngles[SpatialAnswers[trial]]);
            sTime[trial] = calcTime(lineBuf, secondLineBuf, testDataPosModifier);
            prevTrial = trial;
            break;
        }
    case 'F':
        {
            //Run fundamental
            testDataPosModifier = 2;
            fAns[trial] = calcInput(secondLineBuf, testDataPosModifier);
            fAcc[trial] = calcInputAcc(fAns[trial], FundamentalAnswers[trial]);
            fTime[trial] = calcTime(lineBuf, secondLineBuf, testDataPosModifier);
            prevTrial = trial;
            break;
        }
    case 'P':
        {
            //Run phrase
            testDataPosModifier = 3;
            pCorrectAns[trial] = readCorrectAns(secondLineBuf);
            pAns[trial] = calcInput(secondLineBuf, testDataPosModifier);
            pAcc[trial] = calcInputAcc(pAns[trial], pCorrectAns[trial]);
            pTime[trial] = calcTime(lineBuf, secondLineBuf, testDataPosModifier);
            prevTrial = trial;
            break;
        }
    default:
        {
            break;
        }
```

Figure 10: Data reading and processing code from the experimental environment, written in c#



```
Inducing Synesthesia results

Summary:
Spatial: Acc: 84.59386%, Time: 3.852345 seconds
Fundamental: Acc: 11.66667%, Time: 1.975197 seconds
Phrase: Acc: 43.33333%, Time: 8.313016 seconds
Details:

Spatial (Source rotation, User rotation, Response time) :
270, 270, 5.237
315, 312.075, 2.659
135, 132.075, 10.323
315, 312.075, 6.549
180, 191.2008, 5.106
45, 310.7728, 5.136
180, 169.0575, 4.678
225, 227.925, 2.63
90, 269.4212, 4.382
180, 192.0346, 2.583
90, 269.4212, 4.948
45, 312.075, 4.244
135, 132.075, 2.927
225, 227.6646, 3.031
315, 312.075, 2.914
135, 132.3354, 7.453
225, 227.6646, 4.323
360, 360, 4.388
180, 180, 6.772
270, 269.4212, 2.558
```

Figure 11: Processed output, including summary and raw data for excel

Recruitment

Recruitment was achieved via distributed flyers across the Deakin University Burwood campus, a series of email requests to approved organisations, and word of mouth. Recruitment was originally set to occur over a period of 2-3 weeks, however due to extensive delays in the ethics approval, recruitment persisted for 5 weeks. Only 2 exclusion criteria were utilized; participants must be adults such that they can provide consent, and must not have a visual impairment, as this would influence the results of the testing.

Results

The following chapter will present and discuss the data gathered from the experiment. The chapter will be broken down into the individual tests, with a final section discussing the results as a whole. Unfortunately only 7 participants were able to be included in testing, however each participant contributed 60 trials per test, for a total of 180 trials each, and 1260 trials total, with the data displayed and discussed below for each test representing around 420 trials. Below are the average results for each participant.

	Spatial		**Fundamentals**		**Phrase**	
Participant	**Av(%)**	**Av(Time)**	**Av(%)**	**Av(Time)**	**Av(%)**	**Av(Time)**
P1	83.5865	1.981883	16.6667	2.518833	31.6667	7.574900
P2	86.0038	3.916550	11.6667	2.008117	43.3333	8.451567
P3	79.6720	1.753967	20.0000	2.797567	33.3333	3.933383
P4	77.5081	4.176267	35.0000	2.136283	13.3333	7.571500
P5	82.2296	2.450617	28.3333	2.455850	35.0000	4.077667
P6	74.5701	1.330450	21.6667	6.037000	25.0000	7.604383
P7	82.1402	1.964800	43.3333	3.509233	43.3333	6.397833
Average	80.8157	2.510648	25.2381	3.066126	32.1429	6.515890

Figure 12: Average results for each test and each participant

Spatial sound

The spatial test requires participants to submit a directional or rotational answer each trial, once they believe they are facing the direction of the sound source that is being visualized. The experimental virtual environment records this direction or rotation and compares it to the known correct rotation. Participants are then given a % accuracy based on the comparison, e.g., Correct answer: 45 degrees, submitted rotation: 90 degrees, the difference between these angles is 45 degrees, or 12.5% of total rotation, thus the accuracy of the answer is 87.5%. The participants averaged accuracies ranging from around 74-86% (figure 12), with an overall accuracy average of 80.8% (figure 12)across all participants. There was small data loss estimated at approximately 3% due to an experimental fault associated with Unitys logging of Quaternions, which grants us an uncertainty of approximately ±3% or 10.8 degrees. This gives a total average accuracy of 69.12 ± 10.8 degrees.

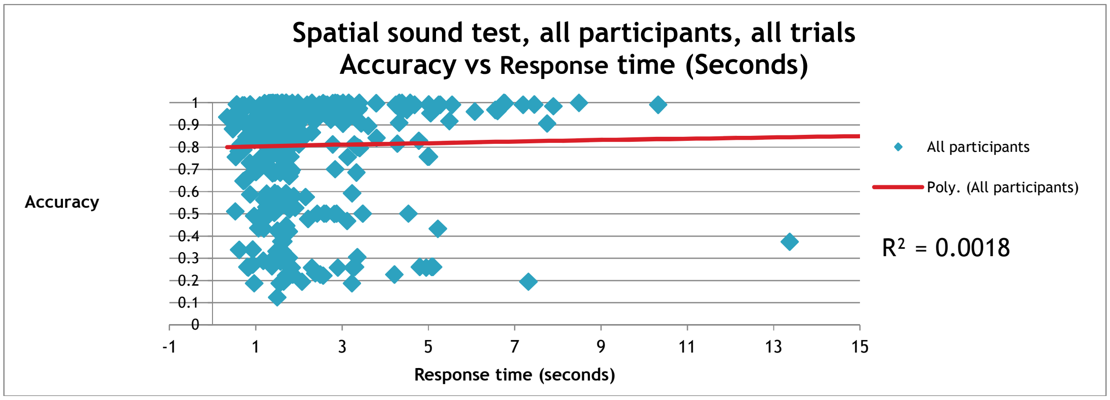


Figure 13: Spatial sound test, accuracy vs. response time, all trials with trend line

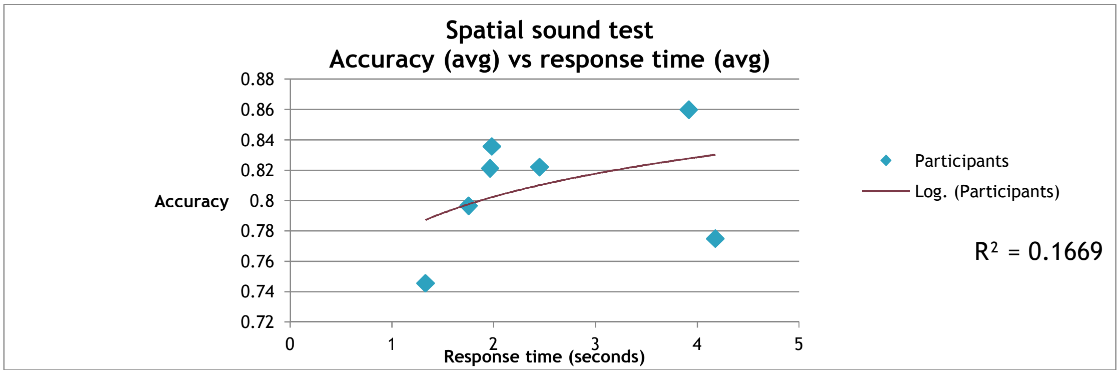


Figure 14: Spatial sound test, accuracy average vs. response time average, all participants

From the limited data we have been able to gather, when viewing all the trials individually, it is seemingly apparent that accuracy of response for this test does not increase with response time (figure 13). This is largely unchanged when examining individual performance averages (figure 14). This may be an indication that the data the visualization represents, in this case direction, is innately understood, and the visualization was able to produce a synaesthetic outcome.

Fundamental Frequency test

The fundamental frequency test requires the participant to select the correct relative pitch of a sound signal, given a corresponding visualization of the sound signal. The experimental virtual environment records the participants selection, and response time, comparing the selection to the known correct answer. Participants are scored as either 100% correct for the correct answer, or 0% correct for all other answers. Participants averaged accuracies ranging 16-43%. This is an extreme variance, and may indicate flaws in the experimentation process, likely the small sample size, or the nature of the visualization failing to represent the primary characteristic accessibly; frequency. Experience with VR technology does not account for the variance, as both the lowest and highest performing participants had extensive familiarity with VR and AR technology.

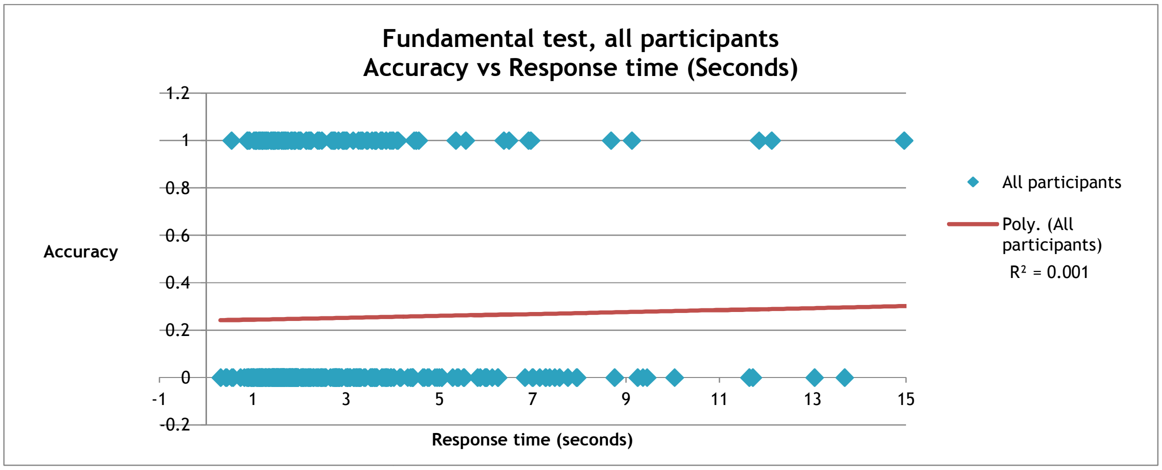


Figure 15: Fundamental frequency test, accuracy vs. response time, all trials with trend line

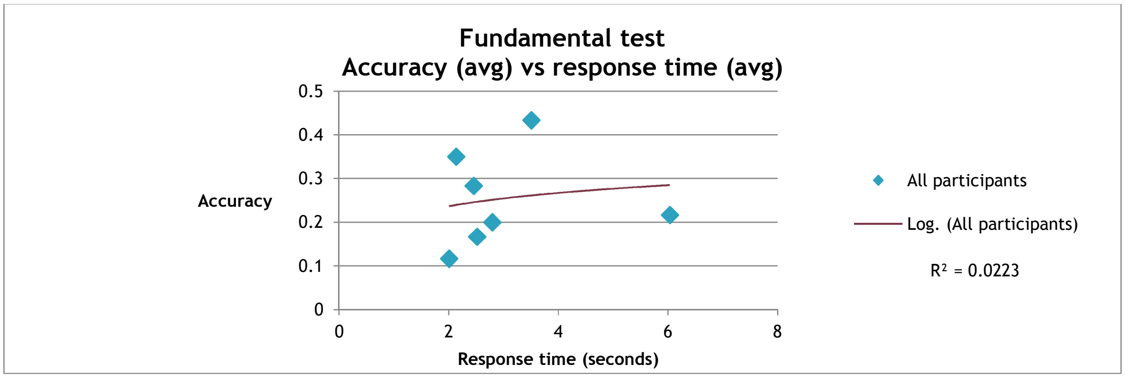


Figure 16: Fundamental frequency test, accuracy average vs. response time average, all participants

From the limited data we have been able to gather, when viewing all the trials individually, it is seemingly apparent that accuracy of response for this test slightly decreases with response time (figure 15). The results appear inconclusive when examining individual performance averages (figure 16). This result is not indicative of participants being able to appreciate simple tones of a singular dominant frequency.

Phrase test

The phrase test requires the participant to select the correct phrase, given a corresponding visualization of the sound signal of that phrase. The experimental virtual environment records the participants selection, and response time, comparing the selection to the known correct answer. Participants are scored as either 100% correct for the correct answer, or 0% correct for all other answers. Participants averaged accuracies ranging 13-43%. This is an extreme variance, and may indicate flaws in the experimentation process, however this seems unlikely. Despite the small sample size, many participants scored very highly, with only one participant whom was previously unfamiliar with VR and AR, in addition to exhibiting a self reported 80% hearing loss in their right ear, scoring below the expected value of guessing (25%).

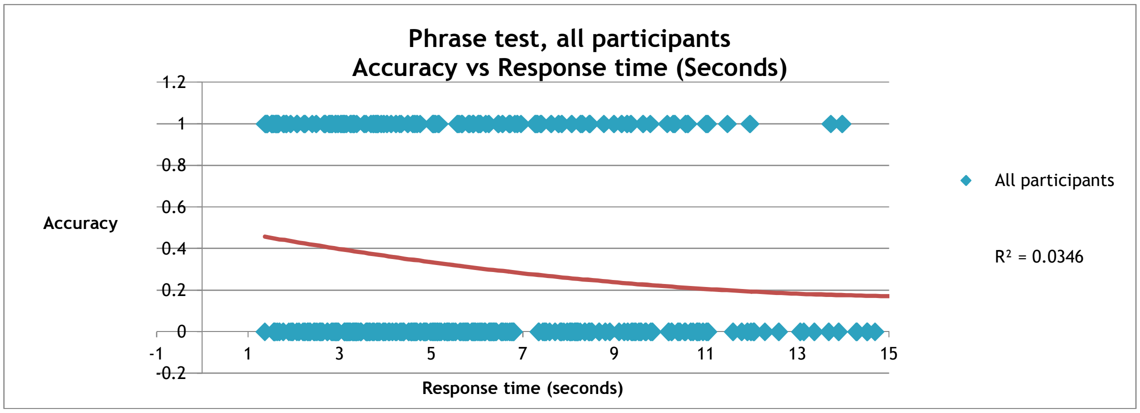


Figure 17: Phrase test, accuracy vs. response time, all trials with trend line

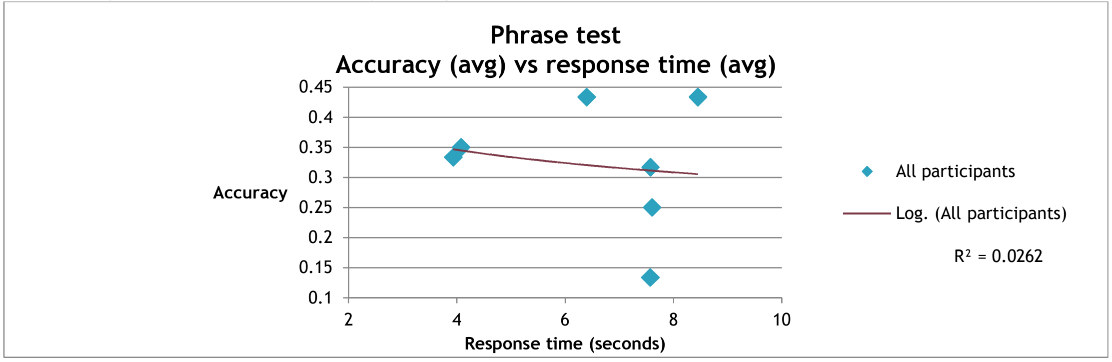


Figure 18: Phrase test, accuracy average vs. response time average, all participants

From the limited data we have been able to gather, when viewing all the trials individually, it is seemingly apparent that accuracy of response for this test decreases with response time (figure 17). The results appear inconclusive when examining individual performance averages (figure 18), likely as a consequence of the small sample size. Ordinarily, the later would suggest that there is no indication that participants were able to communicate simple phrases using the visualization, however when looking at individual accuracies, the significant majority of participants drastically exceeded the guess average of 25%, which suggests the opposite; that participants were able to communicate simple phrases using the visualization. Additionally, as accuracy decreased with response time, this suggests that active engagement, or conscious consideration was not a significant factor in success, and may have even been detrimental. Combined, it is reasonable to conclude, that since participants were able to successfully communicate via the visualization, and seemingly passively so, that this visualization application may have achieved synaesthetic outcomes.

Conclusion

Of the three tests, the spatial sound and phrase test results appear to indicate the possibility of synaesthetic outcomes, whereas the fundamental frequency tests did not indicate any possible synaesthetic outcomes. This may be an indication that the style of visualization while effective for directional and dominant frequency varied sound signals, is not suitable for singular dominant frequency sound signals. However due to the limitations of the sample size, the data is not statistically significant, and it is tenuous to draw any concrete conclusions. Instead the results should serve as an hint or suggestion that VR and AR sound signal visualizations, may be capable of producing synaesthetic outcomes. Several participants reported that the experimental tests were difficult, and expressed an interest in learning their success rate. This eagerness was refreshing in contrast to previous studies where participants expressed that the visualization was confusing, and reflects the training element of the testing. Expanding on this research should include a comparison of the differing visualization styles with a more significant number of participants over several testing sessions. This would ensure that the most suitable sound characteristic mapping to the visualization was utilized, that the data would be statistically significant, and that much like it occurs naturally, the participants would have time to develop synaesthesia.

References

Afra, P., et al. (2009). "Acquired auditory-visual synesthesia: A window to early cross-modal sensory interactions." Psychology research and behavior management **2**: 31-37.

Alex, C., et al. (2013). Sound localization and Visualization device. 2013 IEEE Global Humanitarian Technology Conference (GHTC).

Dong, M. and R. Guo (2016). Towards understanding the capability of spatial audio feedback in virtual environments for people with visual impairments. 2016 IEEE 2nd Workshop on Everyday Virtual Reality (WEVR).

Ferguson, S., et al. (2005). Seeing sound: real-time sound visualisation in visual feedback loops used for training musicians. Ninth International Conference on Information Visualisation (IV'05).

Finney, E. M., et al. (2001). "Visual stimuli activate auditory cortex in the deaf." Nature Neuroscience **4**: 1171.

Grossenbacher, P. G. and C. T. Lovelace (2001). "Mechanisms of synesthesia: cognitive and physiological constraints." Trends in Cognitive Sciences **5**(1): 36-41.

Kaper, H. G., et al. (1999). "Data sonification and sound visualization." Computing in Science & Engineering **1**(4): 48-58.

Laidlaw, D. H., et al. (2005). "Comparing 2D vector field visualization methods: a user study." IEEE Transactions on Visualization and Computer Graphics **11**(1): 59-70.

Roberts, M. H. and J. I. Shenker (2016). "Non-optic vision: Beyond synesthesia?" Brain and Cognition **107**: 24-29.

Rumiński, D. (2015). "An experimental study of spatial sound usefulness in searching and navigating through AR environments." Virtual Reality **19**(3): 223-233.

Sadato, N., et al. (1996). "Activation of the primary visual cortex by Braille reading in blind subjects." Nature **380**: 526.

Tepljakov, A., et al. (2016). Sound localization and processing for inducing synesthetic experiences in Virtual Reality. 2016 15th Biennial Baltic Electronics Conference (BEC).

University, M. T. "Frequencies of Musical Notes, A4 = 440 Hz." from http://pages.mtu.edu/~suits/notefreqs.html.

Ward, J. and T. Wright (2014). "Sensory substitution as an artificially acquired synaesthesia." Neuroscience & Biobehavioral Reviews **41**: 26-35.

Watkins, S., et al. (2006). "Sound alters activity in human V1 in association with illusory visual perception." NeuroImage **31**(3): 1247-1256.

Yatabe, K. and Y. Oikawa (2015). Optically visualized sound field reconstruction based on sparse selection of point sound sources. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).