College of Life and Natural Sciences

7EL996 Independent Scholarly Activity

Name: Eoghan Tyrrell

Student Number: 100541857

Programme: Master of Science in Applied Acoustics

Course Code: MF8AB

IS Title: Sound Source Width Adjustment as a Cue for Auditory Distance

Supervisor: Dr. John Pritchard

Academic Year: 2020

Date: 01/12/2020

Sound Source Width Adjustment as a Cue for Auditory Distance

Eoghan Tyrrell

University of Derby
MSc Applied Acoustics
Independent Scholarly Article

A dissertation submitted to the University of Derby in partial fulfilment for the degree of Master of Science in Applied Acoustics

1st December 2020

Abstract

Sound source distance perception is poorly understood. Auditory distance perception is a subject area largely neglected by psychoacoustic researchers. Consequentially, immersive audio reproduction formats struggle in their ability to provide convincing sound source distance illusions. As sound moves around the virtual stage, the impression can quickly jump between a state of convincing simulated distance to one that occurs inside the listeners head. With a better understanding of distance hearing, distance illusions can be simulated more convincingly.

Cues used to resolve visual distance may inform the methods by which distance hearing is achieved. This supposition arose from the observation that several visual distance cues have an identical auditory distance cue equivalent. An object will appear visibly smaller with increased distance. A reduction in size may also indicate the retreat of a sound. This thesis reports an experimental examination of sound source width on perceived auditory distance. Sounds varying in width were presented to participants in an audio-based survey.

It was discovered that adjustments in sound source width can affect the perceived distance of a sound. Changes in width must be apparent and easily heard for the effect to occur. If the processing of sound involves the adjustment of width only, absent of all other distance cues, the direction of movement on the front-back axis is not always certain. Sounds should be processed to include as many distance cues as possible to increase the likelihood of a robust distance illusion.

Acknowledgements

I would like to thank my employer Jim Dunne for financing this project, as well as Dr John Pritchard and Tim Green, for their guidance. I would like to extend gratitude to my colleague Glen Plunkett for proofreading and critiquing the authorship.

Table of Contents

List of Figures

(All figures were drawn by the author.)

1. Introduction

1.1 Motivation

A sound close to a listener has an apparent size. For example, when seated at a piano, low notes will sound from the left, and high notes will sound from the right. The sound's size then collapses to a point with increased distance. As is the case with visual perspective, the apparent size of a sound object may contribute to its perceived distance. Despite this reasonable assumption, there is very little empirical evidence to indicate if this is true.

This prospect has implications for concert hall acoustics. As described by Holden (2016, pp.24-26), sidewall reflections increase apparent sound source width. Musical sources on stage will sound larger when the concert hall has been designed to produce strong lateral sonic reflections. In turn, this may contribute to an experience of greater intimacy due to a closer perceived sound source. Closer perceived sound sources may also contribute to an increased perceived loudness of the source, making for a more exciting performance.

This research is significant for the creation of "depth" during an audio mixing scenario. For example, when mixing audio for film, engineers often attempt to match a sound's perceived distance with the on-screen object creating that sound. Furthermore, the gaming industry is now driving the development and improvement of three-dimensional audio reproduction and immersive audio technologies in which the simulation of sound source depth is required. These virtual realities are also being embraced by the architecture, engineering and construction industry sectors for health and safety training, design review, and stakeholder engagement. It is the author's opinion that while such technologies are effective in reproducing most auditory cues, they fall short in the simulation of auditory depth. A better understanding of auditory distance perception means a more effective synthesis of sound source depth and more convincing virtual realities.

Auditory distance perception is a subject area largely neglected by psychoacoustic researchers. Even the most esteemed books written on the localisation of sound appear incomplete, with merely a few paragraphs describing distance hearing techniques. The impoverished state of the art alone provides enough rationale to justify this research topic.

1.2 Aim and objectives

Aim:

- To determine if the perceived distance of a sound is affected by changes in its width so that more effective distance illusions can be realised during audio reproduction.

Objectives:

- To critically review the research and literature relevant to the perception of auditory distance, making reference to visual distance perception and the cognitive processing of sensory information.
- By means of an audio-based survey, collect primary data which reflects the response of listener auditory distance perception to sounds that vary in width.
- To critically assess if changes in angular width alone (absent of acoustic reflections) alter the perceived distance of a sound.
- To critically evaluate the effect of changes in the angular spread of acoustic reflections on the perceived distance of sound.
- To determine if width-based distance illusions can be experienced via headphones and earbuds, despite sound being presented directly to the ear.

1.3 Method

An audio-based survey was presented to the survey participants. A pilot survey was first conducted, followed by a debriefing questionnaire. This debriefing suggested a better survey design ensuring proper collection of research data. The primary survey was then carried out.

Audio samples were organised into pairs. Except for angular size, each sample was identical to the other in the same pair. Participants were asked to indicate if a change in perceived auditory distance was observed between the pair of samples. The listener could freely audition each audio sample as many times as they wished. If a difference in depth was observed, the listener highlighted the sample with farthest perceived distance. If no change in perceived distance was observed, the participant indicated so.

The survey was administered online and in a controlled listening environment. Participants completing the survey online auditioned the audio samples via headphones or earbuds. With headphones or earbuds, the sound samples would not have interacted with the participant’s acoustic environment. Participants auditioning samples over loudspeaker did so in a controlled listening environment where sonic reflections were absorbed to an insignificant level. This ensured acoustic reflections did not interfere in unexpected ways with the samples presented.

1.4 Project overview

The study takes the form of a research project. Primary data was collected. This data was reviewed alongside secondary information in an attempt to address the project aim and achieve the objectives. The literature review, methodologies, findings, discussion, and conclusions are presented in the form of a thesis.

Chapter 1 provides an overview of the project. A rationale for the study is given. The aims and objectives are outlined. A brief description of the research method is also provided. Chapter 2 defines the key technical principles required for a proper understanding of the literature review, findings and discussion. Chapter 2 also critiques the literature directly related to the study. A hypothesis is then put forward based on observed similarities between auditory and visual distance perception. Chapter 3 describes the survey design and methodology employed for the collection of primary data. Chapter 4 presents the findings following an analysis of the primary data. An illustration of the data is also provided. Chapter 5 contextualises the findings and discusses their relevance. Chapter 6 concludes the research project. Limitations in the research method are reported. Paths for further research are suggested. The appendix of this thesis tabulates the survey results and primary data.

2. Literature Review

This chapter defines the fundamental technical principles underpinning this thesis. Following this chapter, the unaccustomed reader will be able to read and comprehend the subsequent chapters without difficulty. While this chapter presents and explains the *key* principles, some concepts extend beyond the scope of this work. An existing understanding of these concepts is beneficial, but it is not required. This chapter also critically analyses the research. A hypothesis is then put forward based on observed similarities between auditory and visual distance perception.

2.1 Sound source width

There are two terms used to describe the subjective width of sound: 'Apparent Source Width' and 'Envelopment'. Apparent Source Width (ASW) describes the spatial distribution of the sound's earliest arriving energy. That is to say, the sound's perceived size is determined by the angular spread of energy arriving within 80 milliseconds of its commencement. A wider spread of early energy causes a larger perceived sound, while narrower energy distributions result in a smaller perceived sound source. Envelopment describes the width of the sound's later occurring energy. With a wide spread of energy that occurs after 80 milliseconds, the listener feels encompassed by the sound (Holden, 2016, pp. 24-26).

This "energy" typically refers to the spread of reflected sound, in a concert hall, for example. However, the ASW and Envelopment definitions are not exclusive to reflected energy as sound in the absence of reflections can prompt a sense of width and Envelopment. Consider being seated at a piano. Low notes will sound from the left, high notes will sound from the right, and middle-range notes will sound from the centre of the piano – the piano alone has width.

Apparent Source Width and Envelopment describe the subjective, perceived width of a sound (BSI, 2009). These subjective parameters are linked to the objective, measurable parameters of Early Lateral Energy Fraction (Barron and Marshall, 1981) and Late Lateral Sound Level (Bradley and Soulodre, 1995) respectively. These objective parameters when measured will give an indication of the expected subjective impression of Apparent Source Width and Envelopment.

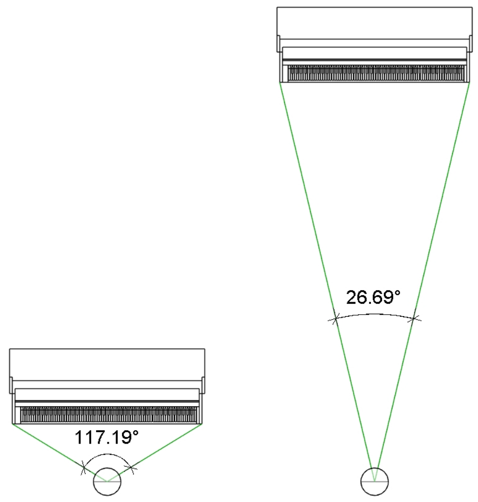

Figure 2.1.1: Sound source width

Figure 2.1.1 illustrates how a piano with increased distance will decrease in angular width. So too will the subjective width of the sound it produces. Via audition only, due to the subjectiveness of audition, the absolute width of the sound source cannot be heard nor can the objective parameters of Early Lateral Energy Fraction and Late Lateral Sound Level be determined.

This thesis is concerned with the subjective Apparent Source Width and Envelopment of sound only. Both of these subjective parameters are manipulated together in an attempt to provoke an increased sound source distance. To understand exactly how these parameters are manipulated, a general understanding of audio reproduction is required. This understanding is also required for comprehension of the methodology and findings chapters.

2.2 Audio reproduction

Monophonic audio refers to the reproduction of sound via a single channel. An example of monophonic audio is when a single channel of audio is sent to one loudspeaker. A second example of monophonic audio is when one audio channel is sent to multiple loudspeakers at once. Figure 2.2.1 illustrates these scenarios. Despite the increased number of loudspeakers, there is still only one channel of audio presented. Therefore, reproduction remains monophonic.

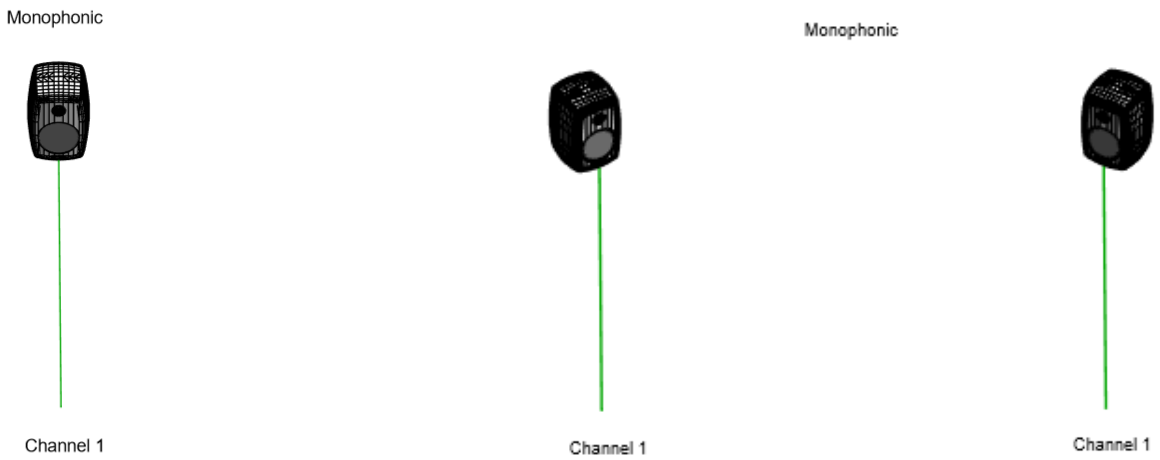

Figure 2.2.1: Monophonic reproduction

Credited to Alan Blumlein (1933), stereophonic audio refers to the presentation of multiple (differing) channels of audio over multiple loudspeakers. At least two loudspeakers are required for stereo listening, as illustrated in figure 2.2.2. Two-channel stereo is the most common form of stereophonic audio. This is the scenario that usually occurs when listening to music over headphones. The channel of audio presented to the left ear will differ to that of the right. Some elements may be common to both the left and right channels, such as the vocal in a pop song. However, any slight difference between the left and right channels means stereophonic audio is being reproduced.

Figure 2.2.2: Stereophonic reproduction

Technically, ‘stereo’ can refer to two or more channels of audio. However, the use of the word ‘stereo’ in informal circumstances has come to refer to two channels only. For simplicity, hereinafter, this thesis will also refer to two channels of audio when using the term ‘stereo’.

A monophonic playback system cannot reproduce a sound source characterised by width. For a source to sound wide, differing signals must be presented to the left and right ears. Consider the piano of Figure 2.1.1. The low notes will mostly arrive at the left ear, and high notes will mostly arrive at the right ear. A monophonic audio signal cannot spatially separate the low notes from the high notes.

A stereophonic playback system is capable of reproducing sound source width. Figure 2.2.3 illustrates an ideal positioning of loudspeakers for stereophonic reproduction, with an approximate thirty-degree angle between the speakers when viewed from the listening position (Everest and Pohlmann, 2015). In such a scenario, a piano's width can be reproduced. The lowest piano notes will emanate from the left speaker, and the highest piano notes will radiate from the right speaker. Both speakers will reproduce the piano's centrally located notes at once. Although produced by two speakers positioned left and right of the head, these middle-range notes will appear to emanate from the "phantom centre". These middle-range sound waves will arrive at both ears concurrently, as would happen if a third loudspeaker was located centrally. The greater the difference between the two stereo channels, the wider the audio appears.

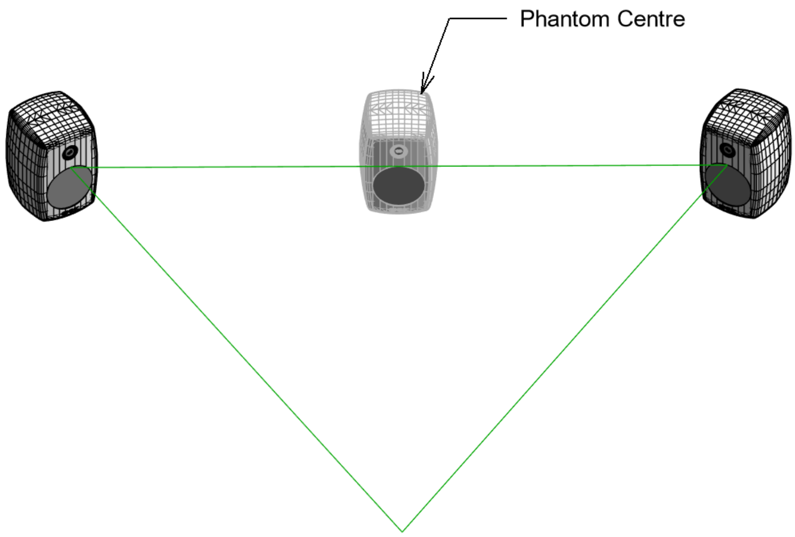

Figure 2.2.3: Stereo listening

Stereophonic signals can be 'folded' or 'collapsed' to monophonic signals. As illustrated in figure 2.2.4, all speakers outputting all channels at once will result in a summation of the channels. A monophonic sound will appear to emanate from the phantom centre as there is no longer any spatial separation between the sounds reproduced. Furthermore, in such scenarios where multiple channels are summed together, cancellations between the signals can occur; cancellations which adversely affect the integrity and timbre of the signal. For example, the low frequencies may cancel-out, leaving a thin, sharp and bass-less sound.

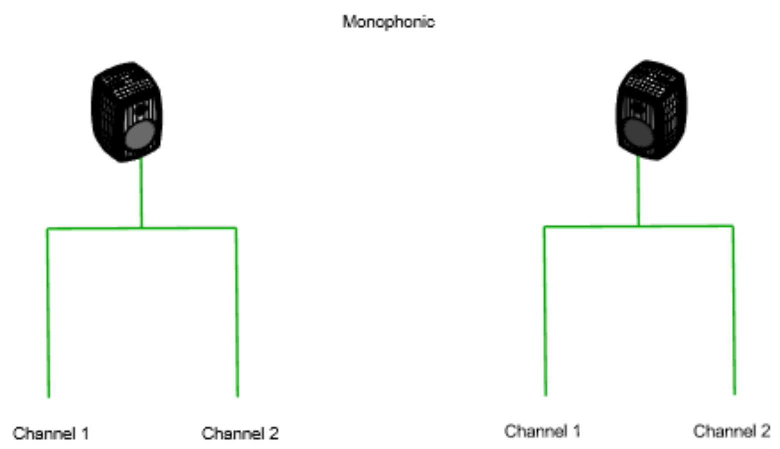

Figure 2.2.4: Collapsing a stereophonic signal to mono

In normal listening scenarios, sounds interact with both ears concurrently, as well as the torso, pinnae and head. This sophisticated filtering of sound assists us in the localisation of the source in three-dimensional space. Without these filtering processes, an effect referred to as "inside-the-head-locatedness" (IHL) can materialise whereby the sound appears to originate inside the listener's head, as explained by Blauert (1987). Listening to sound via headphones or earbuds can induce IHL as the complex filtering of the head and torso is bypassed.

Binaural audio is an extension of stereophonic audio. Through advanced recording and audio post-processing techniques, the filtering effects of the head, pinnae and shoulders can be simulated. These filtering effects of the head and torso can be described by a person's 'head-related transfer function' (HRTF). Everyone's head, shoulders and pinnae are uniquely shaped. In other words, everyone has a unique HRTF. Audio can be integrated or 'convolved' with a person's HRTF in order to reduce the effects of IHL. This allows for sound to appear at a distance in three-dimensional space, despite being presented directly to our ears via headphones or earbuds (Everest and Pohlmann, 2014).

The latest video game consoles and other modern immersive technologies employ binaural audio techniques to achieve '3D Audio'. Three-dimensional audio is achieved by processing the audio with a sample-average HRTF. People who are shaped more closely to the average HRTF will experience more convincing three-dimensional audio illusions. People who deviate from the average shape will be less convinced by such illusions.

This research project involves the presentation of monophonic, stereophonic and binaural audio files to listeners. The effects of sound source width adjustments on the perception of sound source distance are analysed through the use of all such reproduction formats.

2.3 Auditory distance perception

Auditory distance perception is a subject area somewhat neglected by psychoacoustic researchers. Even so, an in-depth explanation of all known auditory distance cues goes beyond the scope of this thesis. However, it is helpful to explain the most dominant distance cues so that the literature review and methodologies can be understood.

Sound level provides a crude indication of distance for familiar sounds (Moore, 2008). Changes in sound level allow for the discrimination of distance, with lower levels suggesting a sound source that is farther away. Absolute distances can be determined from level changes as the sound source is moving, albeit with crude accuracy (Ashmead *et al*., 1995). However, level cues are useful for hearing relative distance, particularly when multiple sound sources are playing at once so that a comparison of levels can be performed (Mershon and King, 1975).

The level of direct sound relative to reflected sound, often referred to as the direct to reflection ratio (DRR) is another primary cue for auditory distance discrimination (Mershon and Bowers, 1979). In rooms, we listen not only to the direct sound but to the reflections provided by the walls, floor and ceiling. Altering the magnitude relationship between the direct sound and its reflections changes the impression of sound source distance. When the reflections have a greater magnitude relative to the direct sound, the sound source appears at an increased distance, and vice versa (Von Békésy, 1960). Sounds with reflections that are relatively high in level can be described as "wet". Sounds with reflections at a relatively low level can be described as "dry". Figure 2.3.1 illustrates these wet and dry scenarios.

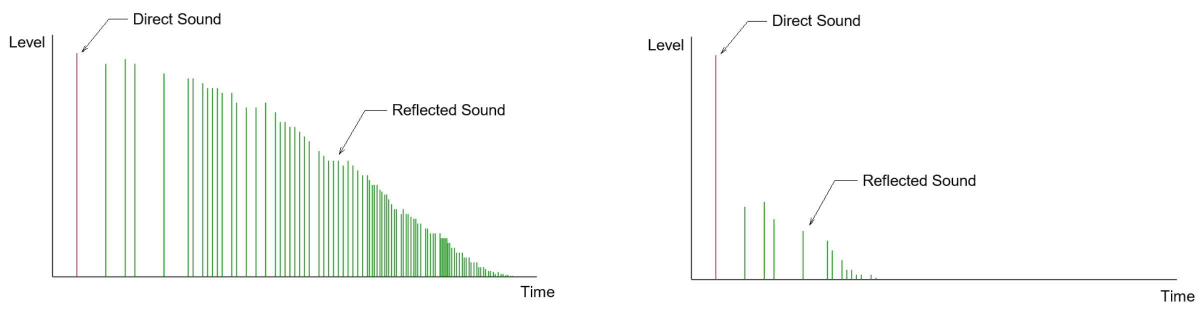

Figure 2.3.1: Left – Sound with reflections at a relatively high level. Right – Reflections at a relatively low level.

Air has the potential to absorb and scatter high frequencies over distance. Sounds travelling distances of fifteen meters or more will attain a "dull" characteristic as the high-frequency energy will be of a lower level upon arriving at the listener (Blauert, 1983, pp.93-137). Coleman (1962) found that effective use of this cue for distance deduction requires familiarity with the sound.

When a sound source is close to one side of a listeners head, a listener will receive signals that differ dramatically in level and frequency content between the ears (Brungart and Rabinowitz, 1999). Sounds at a distance cannot produce such low levels of inter-aural cross-correlation (IACC). A comparison of signals received at each ear is referred to as a binaural difference cue. The binaural disparity between the ears become enlarged in such scenarios.

The time delay between the direct sound and the first reflection, as illustrated in figure 2.3.2, provides another cue for distance hearing. The shorter this time delay, the farther the suggested sound source distance (Michelsen and Rubak, 1997). Long time-delays or "pre-delays" between the direct sound and the onset of reflected energy suggest a closer sound source – that both the listener and source are next to each other, away from any reflective surfaces.

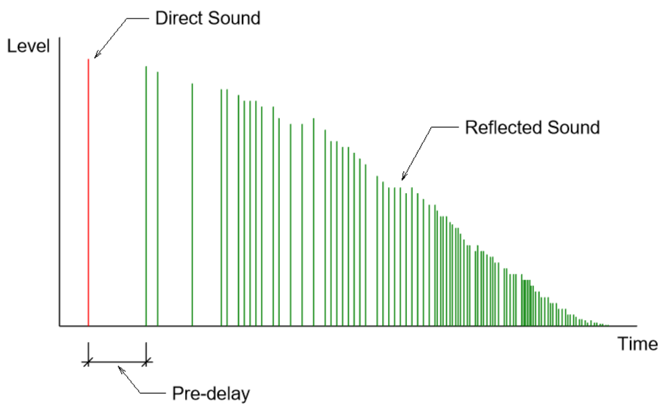

Figure 2.3.2: The onset time of reflected energy (pre-delay)

According to Zahorik (2002), judgements of distance depend on a multiplicity of cues operating at once. A single cue operating alone does not allow for effective distance hearing. Zahorik describes how the perceptual weight assigned to these cues may be flexible, adapting with changes in sound source properties and environmental conditions.

The Design of Distance Pan-pots (Gerzon, 1992) suggests that the angular sound source size can be used to discern the relative distance of a source. However, this has never been investigated empirically. The following section discusses distance perception in the context of vision. Later sections of this chapter make important comparisons between vision and audition. These comparisons are used to support Gerzon's proposal and to support the research aim – that relative sound source distance can be heard following changes in sound source width.

2.4 Visual distance perception

Unlike distance hearing, distance seeing has been explored exhaustively. Visual distance perception is far too extensive to describe in its entirety. This section explains the visual distance cues relative to this study only. These explanations are required for the critique of the literature to be understood.

Linear perspective or one-point perspective describes the scenario of a straight road narrowing to a single vanishing point on the horizon. The realisation of linear perspective is attributed to Italian Renaissance architect Filippo Brunelleschi who simulated visual distance in paintings as early as the year 1420 (Howard and Rogers, 2012, Vol I). Perspective is connected to object size as a cue for visual distance. As illustrated in figure 2.1.1, the piano's size (familiar to the onlooker) decreases as it proceeds into the distance. When there are two objects known to be the same size, the smaller will appear more distant. When an object's size is unfamiliar, smaller objects tend to appear at a greater distance (Howard and Rogers, 2012, Vol I).

Motion parallax describes how objects close to the viewer appear to move more quickly compared to those afar from the viewer when the viewer is in motion (Rogers and Graham, 1979) (Ferris, 1972). For example, when flying in an aeroplane, clouds in the near distance appear to pass by much more quickly than those in the far distance. Animals (including humans) may subconsciously tilt or shift their head, inducing motion parallax to localise better and interpret better the position of an object in the environment (Howard and Rogers, 2012, Vol II).

Texture Gradient describes how the granular detail of an object can be seen up close but conforms to a smoother texture with distance (Gibson, 1950). Consider the finely woven fibres of cloth that then lose their detail and structure with distance (Weinstien, 1957). Relatedly, Aerial Perspective describes how the atmosphere scatters light making distant objects have a reduced contrast and saturation. With distance, light is filtered and shifted towards the colder, bluer end of the spectrum (O'Shea, 1994).

Accommodation and defocus blur describes how the eye when in focus on the near field, then blurs the far-field, and vice versa (Mather, 1996).

Because of the horizontal separation between the eyes, a slightly different image will be projected onto each retina. This is referred to as Binocular Disparity or Stereopsis. Objects close to a viewer will cause greater disparity between the eyes. That is to say, the image presented to each eye will differ significantly. Similarly, objects far from a viewer will have a lower disparity as a similar image will be presented to each retina. Greater disparity suggests a closer object, with lower disparity indicating farther objects (Howard and Rogers, 1995, Vol II).

The following section begins to connect the sensory modalities of vision and audition. From here onward, an argument is made for the notion that the mind processes visual and auditory information in very similar ways – that similar processes are applied by the mind to both visual and auditory information for distance hearing.

2.5 Auditory stream analysis and stream segregation

The job of perception is to take sensory input and derive a useful representation of reality from it.

> "Your friend digs two narrow channels up from the side of the lake. Each is a few feet long and a few inches wide and they are spaced a few feet apart. Halfway up each one, your friend stretches a handkerchief and fastens it to the side of each channel. As waves reach the side of the lake they travel up the channels and cause the two handkerchiefs to go into motion. You are allowed to look only at the handkerchiefs and from the motions to answer a series of questions: How many boats are there on the lake? Which is the most powerful one? Which one is closer? Is the wind blowing? Has any large object been dropped suddenly into the lake?" (Bregman, 1994, p.5)

In this analogy, the channels represent the ear canals, and the handkerchiefs represent the tympanic membranes (or eardrums). Usually, multiple sound sources interact in complex ways within the surrounding environment. A complicated compound of sound energy reaches our ears all at once. Our minds are tasked with making sense of this "mush". The only information available to our auditory system is the vibration of our eardrums. Our mind is capable of deciphering and decoding these vibrations and answering the seemingly impossible questions of the analogy above: How many objects are creating sound? What are the sound sources? Which is louder? What is the apparent size of the sound source? Which source is further away?

Decoding this compound of energy is achieved by segregating and organising the compounded sound wave into separate "streams". An auditory stream is a perceptual grouping of the parts of a neurological response that seem to go together. Bregman (1994) uses the following text to present a visual example of auditory stream segregation:

Compounded Complex Sound Waves Arriving at the Ear: AI CSAITT STIOTOS

Stream Segregation: A$_{I}$ C$_{S}$A$_{I}$T$_{T}$ S$_{T}$I$_{O}$T$_{O}$S

This segregation and organisation is thought to conform to the Gestalt Laws of Perception (the laws of Similarity, Continuation, Proximity, Closure and Connectedness) whereby the closeness in pattern, timbre, time and pitch determine the grouping of parts into streams (Koffka, 2014). For example, when listening to music, the violin is not confused with the drums. The violin and the drums are organised into separate streams.

While these Gestalt laws of perception were derived on vision and visual illusions only, Bregman showed how such laws also directly apply to audition. This thesis also makes strong connections between vision and audition. The following section explains how cognitive processing bridges these seemingly different sensory modalities.

2.6 Connections between vision and audition

A comparison of visual and auditory localisation cues can be used to argue the case for sound source width adjustment as a cue for auditory distance. The mental processes used to discern visual distance may be identical to those used for the perception of auditory distance. Visual cues used to resolve an object's distance may suggest methods by which distance hearing is achieved. The same distance cue, when applied to either a visual or auditory stimulus, may provide object distance localisation information. This hypothesis is tenable given that all sensory modes are connected to the same processor; the mind. There are a plethora of distance cues identical to both vision and audition, as will be discussed in this section. It is not unreasonable to assume that the mind applies similar processes to the information received, regardless of the sensory mode type.

Visual perspective (sometimes referred to as linear perspective, or one-point perspective) describes the scenario of a straight road narrowing to a single vanishing point on the horizon. Illustrated in figure 2.1.1, from the perspective of the onlooker, the piano's visual size decreases as it proceeds into the distance. Auditory distance might be simulated in an identical manner; by narrowing and making smaller the apparent size of the sound. This hypothesis is supported by comparing and aligning the already proven auditory localisation cues with visual localisation cue counterparts. This comparison is as follows.

Changes in sound level allowing for distance hearing compares with object size changes for distance seeing. Mershon and King (1975) describe how changes in sound level allow for the relative judgment of sound source distance. Kilpatrick and Ittelson (1951) showed how changing visible size provides object distance localisation.

Hirsh (1971) describes how binaural confusion in the localisation of a sound object can be resolved with a tilt of the head (auditory parallax). Similarly, Palmer (1999) describes how a tilt or shift of the head inducing (visual) motion parallax can better inform on an object's visual distance.

High frequencies provide the detail in a sound. For example, consonants required for speech cognisance exist in the high frequencies as sibilance. Blauert (1983) describes how high frequencies are filtered and attenuated by the air over distances greater than fifteen meters. This cue is directly comparable to the effects of the texture gradient (Gibson, 1950) and aerial perspective visual cues whereby the luminosity (contrast and saturation) and detail of an object deteriorates with distance due to atmospheric effects.

Brungart and Rabinowitz (1999) found that at distances close to the head, significant differences in signal between the two ears indicate sound source closeness. Identically, Charles Wheatstone (1838) through the invention of the stereoscope (a device that presented slightly different images to each eye) showed how significant disparity between images presented to each retina also indicate the closeness of the object.

A well-known phenomenon of the auditory system is its ability to focus on one sound and defocus on all others. Dubbed by the Gestalt psychologists as "figure-ground", more commonly referred to as the "cocktail party effect", it describes the capacity of the auditory system to focus on a single person's voice in the presence of many others, in a bar or restaurant, for example. Neisser (1967) suggested that the in-focus "figure" sound is organised into one perceptual stream, while the "ground" backdrop sound is organised into another. This faculty is directly comparable to accommodation and defocus blur, which describes the eye's ability to change the depth of focus (Mather, 1996).

As demonstrated by this section, several auditory cues have a visual cue counterpart. From this, it is evident that the mind as the single processor treats visual and auditory information in similar and sometimes identical ways. Furthermore, in Auditory Stream Analysis (1994), Bregman's ability to bring the Gestalt Laws of Perception from vision to audition suggests stronger connections than normally realised; that all information converges on the one mind and is processed in similar ways; that visual cognitive processes can inform those used by the auditory system, and vice versa. From this, it is logical to connect apparent sound source width to an object's visible size for the discrimination of object distance.

It can be argued that while there are a number of comparisons to support this hypothesis, there are many visual and auditory distance cues that stand alone and cannot be linked to each other. Therefore, this hypothesis is, at the same time, rejected by the same evidence.

2.7 A critique of the closely associated literature

The Design of Distance Pan-pots (Gerzon, M., 1992) suggests that apparent angular sound source size can be used to discern the relative distance of a sound source. Gerzon does not formally reference this information, nor does he elaborate on this suggestion. However, as described in section 2.3 of this thesis, the level of direct sound relative to reflected sound (DRR) provides a cue for auditory distance. Additionally, the subconscious analysis of the onset (pre-delay) time between the direct sound and first early reflection provides another method for hearing distance. Both of these distance hearing techniques involve the use of reflected energy. It is reasonable to suggest that altering the configuration of reflections relative to the direct sound in other ways may realise other distance cues. Namely, a narrowing of the spatial reflection distribution may produce an increased sound source distance, as suggested by Gerzon.

During a research effort entitled 'The Distance Pan-pot: An Alternative Approach to the Distance Effect' (Tyrrell, E., 2016), anechoic recordings, digitally processed to have early lateral reflections were presented to listeners over headphones. Participants of the study indicated that no change in auditory distance was perceived when the early reflection spatial distribution pattern was narrowed. However, this research effort was mainly concerned with creating a distance effect via precisely timed early reflections. The assessment of changes in sound source width on perceived distance formed a small part of the research effort. Therefore, there are several potential reasons as to why no change in perceived distance was observed. These reasons are as follows.

Firstly, despite a large number of simulated reflections, the sounds Tyrrell (2016) presented to participants during testing were processed to have a very low spaciousness (or a very high IACC). That is to say; the virtual reflections were very evenly distributed in angular position and level across the stereo image. There was little audible difference in width between the various audio samples presented to survey participants.

Secondly, the survey was administered with participants listening via headphones and earbuds only. The experiments were not administered over loudspeakers. Binaural audio techniques were not employed to reduce the effects of inside-the-head-locatedness (IHL). This raises the question: how can distance be heard if the sound source appears to be located inside the listener's head?

Thirdly, the sound samples presented to listeners were processed to have early reflections only. Such a listening scenario is an unnatural and uncommon one. In the real world, sounds in a reflective environment occur with early reflections and a diffuse reverberant tail, even if such a reverberant tail is short.

Finally, the Distance Pan-pot refers to audio that has width due to spatially distributed reflections. A sound does not necessarily need reflections to be characterised by width. Consider being seated at a piano in the absence of acoustic reflections. Low notes occur on the left, and high notes occur on the right. As illustrated in Figure 2.1.1 of this thesis, with distance, the low notes and high notes would collapse to a point source. This scenario was not examined.

The author was unable to source any other research directly connecting a sound's apparent size to its perceived distance. This observation provides good rationale and motivation to implement the research topic and fulfil this work's aims and objectives.

3. Methodology

This chapter describes the survey design and data collection procedure. Illustrations are provided which show how the audio samples were processed and manipulated. Complications arose when designing the survey. The measures which solved these complications are outlined. A link to the survey can be found at the end of this chapter.

3.1 Survey design and data collection

An audio-based survey formed the method through which primary data was collected. Audio samples were presented to participants in pairs. Except for apparent width, each sample was identical to the other in the same pair. Participants completing the survey online were asked to audition the audio samples with headphones or earbuds. Participants auditioning samples over loudspeaker did so in a controlled listening studio environment. The auditioning room was characterised by almost anechoic conditions so that acoustic reflections would not interfere with the perceived width of the samples presented.

The samples were organised into two categories: anechoic and echoic. The anechoic samples were manipulated in width through amplitude panning – stereo recordings were made monophonic by redistributing both stereo channels to both speakers at once, as illustrated in figure 2.2.4, chapter 2. The echoic samples were monophonic recordings made to have width by synthesising lateral reflections. The widths of the samples were then narrowed by redistributing these simulated reflections to the centre of the stereo image. All echoic samples were processed with two different DR ratios. That is to say, wet samples processed to have different widths were paired against each other. Relatively dry samples were similarly paired against each other.

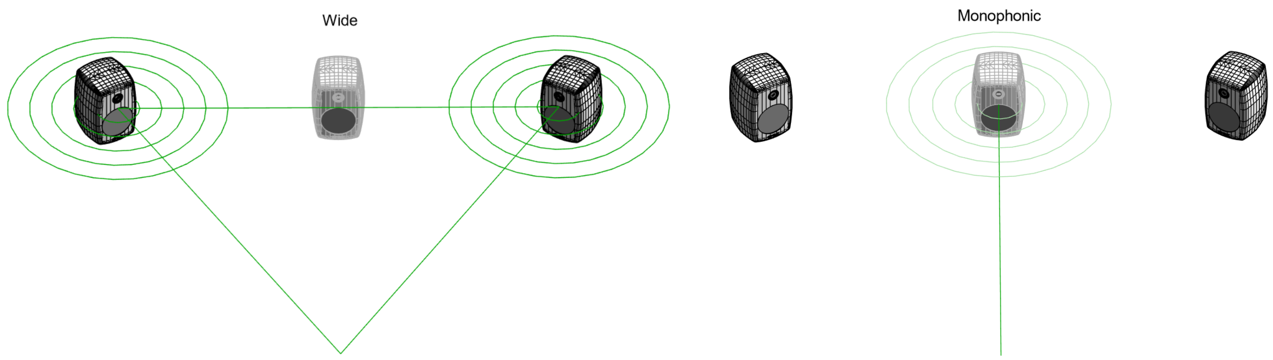

Figure 3.1.1: Illustration of how survey samples were processed.

Figures 3.1.1 and 3.1.2 illustrate how each sound source was processed to have one of three different width-scenarios: wide, narrow and mono. When developing the survey, the author found the monophonic samples to sound somewhat unnatural. To investigate whether a natural-sounding reduction in width is required for distance hearing, a 'narrow' scenario (Figure 3.1.2) characterised by a more realistic timbre was introduced into the survey.

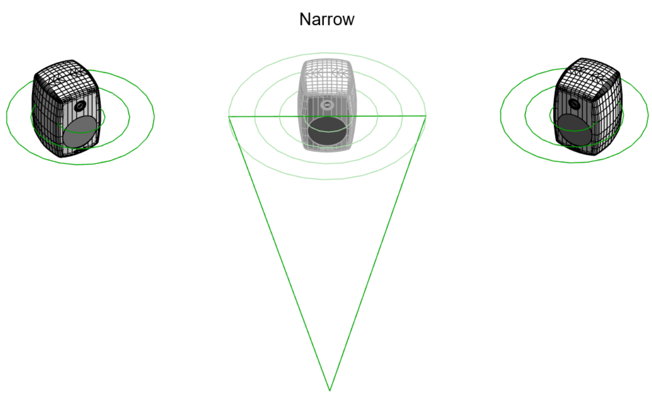

Figure 3.1.2: Narrow samples with a more natural timbre were included in the survey.

As described in section 2.2 of this thesis, a redistribution of the stereo signal to both the left and right audio channels allowed for the manipulation of stereo width. A narrowing of the stereo image towards a monophonic signal creates a scenario where the sound appears to emanate from the phantom centre, i.e. from a point source. Caution had to be exercised so as not to introduce a level difference during this processing of the samples. Had there been a change in the level between the samples in a pair, level distance cues would have been introduced into the experiment. In order to prevent such interference, loudness meters were utilised to ensure no change in level occurred between the samples. All samples were normalised to a loudness of -23dB LUFS (Loudness Units Full Scale).

Similarly, when redistributing the reflections across the stereo image, it was important not to introduce a change in the ratio of direct to reflected sound (DRR), which in turn would have introduced another unintentional distance cue. This was achieved with a software package called Exponential Audio Phoenix Verb. This package allowed for the narrowing of both the early reflections (ASW) and the diffuse reverberation tail (Envelopment) without altering the direct to reflection DRR ratio.

The samples presented to listeners were recordings of familiar sounds. All samples comprised of music and speech sounds. As suggested by Moore (2008), the localisation of a sound source is more easily achieved if the sound in question is familiar. All samples were rich in harmonic content and characterised by transients, which, according to Newman and Stevens (1936) also assists the in localisation of sound in three-dimensional space.

A preliminary pilot survey was first conducted. This pilot exercise allowed for the refinement and optimisation of the survey design in order to prevent compromise of the data collected. This pilot survey unveiled several flaws which were subsequentially fixed for the primary survey and data collection process.

During the pilot survey, an auditioning system was trialled whereby listeners had control to instantaneously switch between the two comparison samples during playback. This mechanism allowed listeners to compare the samples against each other in real-time, swiftly and effortlessly. In practice, this approach was flawed. The majority of the samples are dynamic. That is to say, their loudness changes from moment to moment. The sample perceived to be more distant was determined by this dynamic and the timing of the switch. As participants randomly switched between the samples, the audio with the quieter dynamic at that particular moment was perceived to be more distant – a random level distance cue was introduced into the experiment. This made for a set of compromised data from which no conclusions could be drawn. Following this realisation, the survey was revised to remove this instantaneous sample-switching functionality. During the primary survey, samples were instead presented sequentially which removed this unintended random distance cue.

Some survey participants found that due to the quality of their internet connection, the longer three-minute music samples would take time to load, or would not load at all. A second alteration to the survey involved shortening all audio files so that the audio files would load promptly and properly for participants. A drop-down option was also included in the survey whereby participants could indicate if errors occurred when loading the audio files. Such data could then be excluded from the data analysis process.

In the pilot survey, some of the questions required the participants to compare three audio samples against each other. Participants noted that this was quite a difficult task. The audible difference between samples was not always obvious. Participants had to initiate the samples numerous times and make multiple comparisons which required too much concentration and effort. As a result, some participants were not confident that their answers were true to their perception. For the primary survey, it was ensured that all survey tasks were simplified and easy to complete. The survey questions that once involved three audio samples were reduced to have two audio samples only.

The primary survey was conducted on the week of the 21st September 2020. Both the online and in-person surveys were administered simultaneously. Both online and in-person participants received an identical survey. Thirty participants located across four different countries completed the survey online. The aim was to have a study sample as wide and far-reaching as possible. Findings from a diverse sample set can more easily extend from the sample to the population.

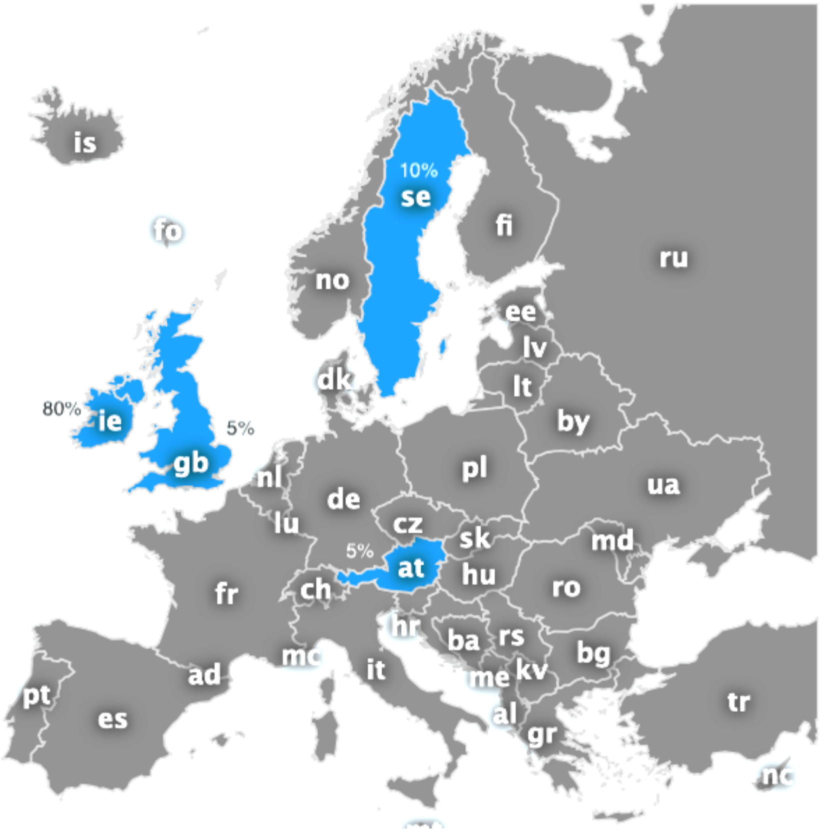

Figure 3.1.3: Survey response distribution

Seven participants completed the survey in the controlled listening studio environment. Expert listeners who operate as acousticians and studio designers were tested alongside non-expert listeners such as carpenters, accountants and marketing executives. Twenty-three participants completed the survey online. The response distribution is illustrated in figure 3.1.3. It takes approximately fifteen minutes to complete the survey. The survey is still active for review at auditorydistanceonline.questionpro.com (Tyrrell, 2020).

4. Findings

This chapter presents the survey findings. Appendix I presents the raw data from which these findings were derived. These findings assume that the laboratory listening scenarios presented over loudspeakers, headphones and earbuds extend to real-world listening scenarios. The findings assume that the perception of reproduced audio represents the perception of sound experienced first-hand. It also assumes that the electro-acoustically simulated reflections accurately represent real-world acoustic reflections. The findings assume that all listening devices used were of adequate quality – that headphones and earbuds which may have been limited in frequency response and accuracy did not influence the results.

4.1 Illustrating the project data

Figure 4.1.1 illustrates the data collected. To read and interpret this illustration correctly, consider the following:

There were 16 survey questions presented to each participant. These questions are numbered 1 to 16 in figure 4.1.1. A percentage bar represents the response to each question.

Every question involved the presentation of two audio samples. These paired samples were identical in every manner, except for stereo-width. For example, question number 6 involved the presentation of the "Drum A" sample. A wide version and a monophonic version of the "Drum A" sample was presented during question 6. Participants were to indicate whether the "Drum A" (wide) or the "Drum A" (mono) sample was perceived to be farther away. "Drum A" was processed with simulated reflections; hence, it belongs to the "echoic" category. 28% of participants perceived the wide sample as farther away. 28% of participants perceived the monophonic sample as farther away. 44% of participants indicated that there was no difference in perceived distance between the samples.

Questions 1 to 4 involved the presentation of anechoic samples. That is to say, dry stereo recordings were manipulated in width via amplitude panning, as described in chapter 2. Questions 5 to 16 involved the presentation of samples processed to have different widths via the redistribution of simulated reflections about the stereo image. Questions 11 to 16 paired wide samples against narrow samples (as opposed to wide samples against monophonic samples). All questions were presented to listeners in random order. In figure 4.1.1, the questions have been organised sequentially so that patterns in the data can be observed. Appendix I presents the questions in the random order the survey participants received it.

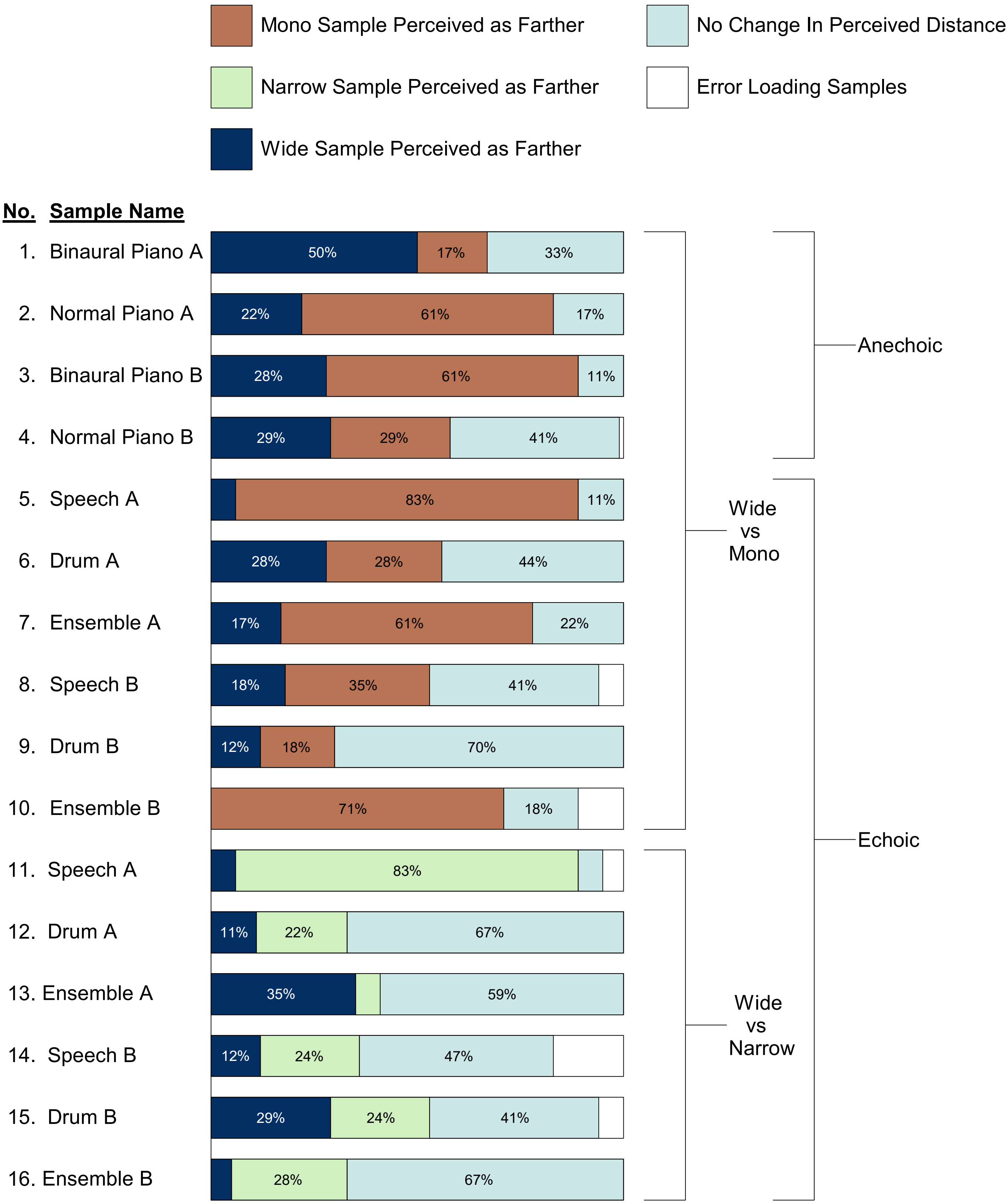

Figure 4.1.1: Illustrating the project data

4.2 A high-level assessment of the data

56% of all listeners regarded themselves as expert listeners. 44% of listeners regarded themselves as non-expert listeners. There were no observable differences in trends between these different listener types.

As described in chapter 4, narrow samples were introduced into the survey as the author was worried the unnatural sounding monophonic samples would adversely impact the findings. It transpired that listeners (including expert listeners) found it difficult to hear the difference between the wide and narrow samples. As a result, comparisons pairing a wide sample against a narrow sample mostly indicated “no change in perceived distance”. The majority of comparisons which paired a wide sample against a monophonic sample usually indicated some change in perceived distance. This implies that substantial changes in width are required for a change in the perceived auditory distance.

With the narrow samples removed, patterns observed in the data became much more conclusive. From the remaining sample pairs (wide samples versus monophonic samples), 67% of all comparisons positively indicated some change in perceived distance following a change in sound source width. 69% of these favourable comparisons pointed to the monophonic sound as the more distant sound. 31% of these positive comparisons indicated the stereo sample as the more distant sound. This suggests that sound source width alone is not a strong or sturdy cue as the direction of the perceived movement is not always certain.

As mentioned, participants found it difficult to hear differences in width between the wide and narrow samples. Proper judgements could not be made. Therefore, so as not to dilute the results with meaningless data, tests that involved the narrowly processed samples are excluded from this chapter hereinafter.

4.3 Addressing the project objectives

Do changes in angular size alone (in the absence of acoustic reflections) affect the perceived distance of a sound?

A change in perceived distance was observed during 73% of anechoic sample sets. 27% of comparisons indicated no change in perceived distance. These results indicate that a sound absent of reflections will likely change in perceived distance following a change in width. The results show that width as a cue *can* be used in the absence of reflections. However, 57% of these positive results indicate the monophonic sample at a farther distance. 43% of the positive comparisons identify the stereo sample as the more distant sample. Therefore, the perceived movement of anechoic sounds on the front-back axis is not certain following a width adjustment. For anechoic sounds, the binaural samples more often induced a change in perceived distance compared to the non-binaural stereo sounds. This suggests that width-based distance illusions are more successful with binaural listening, but the synthesis of distance is also possible during normal stereo reproduction.

Do changes in the angular spread of acoustic reflections affect the perceived distance of a sound?

In 65% of comparisons, a change in perceived distance was observed. 35% of comparisons indicated no change in perceived distance. While favouring the proposition that a narrowing of reflections will prompt a change in perceived distance, the data also implies that there is little confusion in the direction of sound source movement. 82% of these favourable comparisons indicated that the sound proceeded into the distance (perceptually speaking) following a narrowing of the simulated early and late reflections.

As described in chapter 4, every "wet" sound was assessed with reverberation simulated at two different DR ratios. It was discovered that the wetter sample comparisons more often induced a change in perceived distance, suggesting that width-based distance illusions are more robust when the width changes are apparent and easily heard.

Tyrrell (2016) found that changes in width by narrowing the simulated reflections did not cause a change in the perceived distance of the sound. However, as described in chapter 3, these experiments involved a reflection pattern that had a very low spaciousness. Furthermore, only early reflections were simulated. The findings of this study conflict with the findings of the 2016 study, suggesting that a change in spaciousness must be evident for a change in the perceived distance, i.e. it must be relatively easy to hear the width changes. This also suggests that reverberation should be simulated in a natural way involving both early reflections and some form of a reverberant tail.

Can sound source distance be experienced via headphones and earbuds, despite the potential for inside-the-head-locatedness (IHL) to arise?

77% of listeners completed the survey on headphones or earbuds. 23% of listeners completed the survey via loudspeakers. There was no noticeable difference in the data trends between those completing the survey on headphones and those completing the survey on loudspeakers. All headphone/earbud listeners indicated that a change in perceived distance had occurred at some stage during the survey. This demonstrates that despite the potential for inside-the-head-locatedness (IHL) to arise, distance illusions can be experienced via headphones and earbuds.

4.4 A summary of the findings

To summarise the findings of this research:

- Sound source width adjustment can be used as a cue for auditory distance.
- Sound source width adjustment as a distance cue can be used in the absence of reflections.
- A change in the reflection spread can also prompt a change in perceived auditory distance.
- Following a width adjustment, the perceived movement of anechoic sounds on the front-back axis is not certain.
- Following a narrowing of reflections, the sound is more likely to retreat into the distance, rather than move closer.
- Apparent changes in width are required for a change in the perceived auditory distance - the change in width must be easily heard.
- A change in the reflection spread will more likely provoke a distance change if the sound is relatively wet as the width change will be more apparent. That is to say, a change in spaciousness must be distinct for a change in the perceived distance to occur.
- Sound source distance can be experienced via headphones and earbuds despite the potential for inside-the-head-locatedness (IHL) to arise.
- The perception of sound source distance does not change between expert and non-expert listeners.
- For distance illusions prompted by width-changes in the reflection spatial distribution, reverberation should be simulated in a natural way involving both early reflections a reverberant tail.

5. Discussion

The results from the author's previous investigations (Tyrrell, 2016) meant the findings that materialised in this research effort were not expected. It was the author's attempt with this thesis to disprove any notions that sound source width changes could induce a change in auditory depth. Instead, a more comprehensive examination of sound source width resulted in compelling evidence *for* its influence on distance hearing.

While the findings favour the proposition that perceived distance is affected by sound source width, the evidence appears refutable. According to the data, a change in perceived depth does not always occur, and an increase in perceived depth does not always follow a narrowing of the sound. The author would argue that distance cues (and localisation cues in general) rarely operate in isolation. In the real-world, a sound source proceeding into the distance will always trigger numerous cues at once. As described by Zahorik (2002), judgements of distance depend on a multiplicity of cues operating at once. Single cues operating alone do not allow for effective distance hearing. Even Lord Rayleigh in his duplex theory of sound localisation (Strutt, 1909) had suggested this case in the early 1900s. In this manner, sound source width as a distance cue is no different to any other already established auditory distance cue – that distance cues in isolation are not effective.

The outcome of the findings support the hypothesis that visual cues can be used to suggest the methods by which distance hearing is achieved. The findings support the notion that the mind applies similar processes to the sensory information it receives, regardless of the sensory mode type. In addition to the comparisons made in section 2.6 of this thesis, one more matched distance cue pair is realised – that the sound source width cue is directly equivalent to the object visual size cue for distance discrimination.

This thesis implies techniques for better simulating the position of a sound object in three-dimensional virtual space. The sound source width cue should be incorporated alongside a multiplicity of other auditory distance cues. Furthermore, virtual sound objects will have their simulated distance reinforced by the corresponding visual information. As described by the term 'cross-modal integration' (Wallach, 1940), effective and accurate localisation of an object is achieved when visual and auditory information is combined. To simulate the position of an object in virtual three-dimensional space, one should present as many possible (visual *and* auditory) localisation cues to the mind in order to maximise the effectiveness of the illusion.

Unfortunately, in the reproduction of audio, there can be technical requirements which hinder the effectiveness of such audio object placement illusions. As illustrated in figure 2.2.3, there is an ideal positioning of loudspeakers for effective stereophonic listening. Often, audio is presented to listeners over less than ideal loudspeaker configurations. As is the case with a mobile phone, television speakers, or a public address (PA) system in a public building, for example, these sub-optimal conditions often involve poorly positioned stereo loudspeakers or an array of loudspeakers capable of monophonic playback only. As a result, audio produced and published for reproduction over a broad range of playback devices usually requires "monophonic compatibility". This means the audio can be produced in stereo but must collapse to a monophonic signal without the cancellation of audio information. To achieve this, mixing engineers make use of width effects to a much lesser extent than would be preferred. The findings of this research suggest that these width-based limitations reduce the potential depth-of-field from the early stages of mixing, as mixing engineers do not have full freedom to experiment with the wideness of sounds. Furthermore, monophonic playback systems cannot make use of any width-based distance effects to enhance sound source depth as such systems are incapable of the spatial separation of sound.

This research has significance for the design of performance spaces. As described by Holden (2016), strong lateral reflections increase the apparent source width (ASW) of the sound. Strong lateral reflections, good ASW and good Envelopment, are already design aims considered in the development of performance spaces. However, following this research, it can be argued that a closer perceived sound source will promote performance intimacy, thus bettering the subjective impression of the performance hall. This research supports the argument for strong lateral reflections for the enjoyment of a performance in a performance space.

The concept of stream segregation would have provided a reasonable explanation had no change in perceived distance occurred following a narrowing in the spread of reflections. Had such a result emerged, it could have been argued that the direct sound and reflected sounds get organised into separate streams – that the narrowing of one stream (the reflections) does not affect the perceived distance of the other stream (the direct sound). The findings instead indicate the exact opposite – that a narrowing in the spread of reflections alters the interpretation of sound source depth. This implies that the direct and reflected sounds are grouped into the same perceptual stream. It is suggested that both the direct and reflected sounds are processed in the mind as one single entity, perhaps due to the Gestalt law of similarity and proximity.

6. Conclusion

This research set out to determine if the perceived distance of a sound changes with the sound's angular width. Convincing evidence was produced to support the use of sound source width for the manipulation of perceived sound source distance. While it can be concluded that the research aim was achieved, the study was not without its limitations. This chapter evaluates the research approach, presents the limitations, and discusses methods to further strengthen the research findings.

6.1 Evaluation of the Research Approach

Thirty listeners participated in the survey. While trends in the data had been established by the thirtieth participant, additional listeners would have been welcomed in order to bolster these trends further. Greater diversity in the geographical and demographical spread in participant contribution would have better extended the findings from the sample to the human population.

So as not to misconceive and misunderstand the findings extracted from the data, the data analysis procedures were kept simple. The data processing simply involved observation of the percentage majorities. All data is included in the appendix of this thesis. A more thorough and advanced analysis of this data may reveal additional findings missed by the author.

During their participation in the survey, listeners were asked to listen out for changes in sound source distance. Knowing the desired outcome of the survey, participants who heard a change in timbre between the samples might have indicated a change in distance, even if a distance change was not perceived. Participants may have believed they were completing a quiz, rather than a survey. It was made clear to participants that there are no wrong answers – that their perception cannot be wrong. However, having participated in such surveys in the past, the author understands the pressure on participants to answer "correctly". While these risks are common to most surveys of this nature, and while such potential risks cannot be avoided as participants needed to be instructed to listen for distance changes, it is important to declare that this phenomenon may have compromised the data.

Overall, the approach to the research was focused and methodical. The project did not deviate much from the initial research proposal to the final thesis. Time management was effective, which allowed for a thorough review of the relative literature.

6.2 Further Work

This work involved the analysis of a single distance cue in isolation. As mentioned numerous times throughout this thesis, the perception of depth is enhanced when numerous cues operate simultaneously. In fact, in the real world (away from reproduced sound and virtual reality), sound source width as a distance cue cannot be provoked in isolation. A reduction in sound level will always accompany the narrowing of a sound source as it increases in distance. This work can be developed further to examine the effectiveness of sound source width as a distance cue alongside sound level, for example. It may be the case that sound source width as a distance cue is more effective in the presence of a loudness cue, or other distance cues.

This work presented the experiments over the two-channel stereo. While two-channel stereo is still the most used listening format, the popularity in multi-channel reproduction is growing. Dolby are promoting their multi-channel Atmos format for use across cinemas and home theatre systems. Due to the encouragement from YouTube, the video game industry, and other mainstream media creators, immersive three-dimensional audio has emerged from a specialist audiophile listening format into the mainstream. Further investigations into the influence of width on distance over more modern reproduction formats should be carried out. While some sample comparisons in this research involved binaural listening, these assessments formed a small part of the overall experiment. Furthermore, listeners may have deviated significantly from the average HRTF used to process these binaural audio files. Also, collapsing the binaural files to mono will likely have affected the HRTF in unknown ways. These considerations would be required in future width-based investigations involving binaural listening.

One of the Gestalt Laws of Perception is ‘proximity’. It describes how elements may be grouped (perceptually speaking) based on their closeness to each other in time and timbre. As discussed in chapter five, it is believed that the direct sound and the reflected sounds are organised into the same perceptual stream, perhaps due to the law of proximity. As described in section 2.2, the onset time between the direct and reflected sound can influence the perceived distance of the sound source. Perhaps the reason for the sound source appearing closer with longer pre-delay times is because the proximity law “breaks”. That is to say, the greater the separation in time between the direct and reflected sound, the more likely they organise into separate streams. With the reflected sound organised into its own stream, the direct sound, in essence, becomes dry (perceptually speaking) and appears closer. The reflections may no longer influence the perceived distance of the sound source. Assessing the existence of a “breaking-point” would test this suspicion.

This research used distance seeing techniques to inform on the methods by which distance hearing is achieved. Namely, with increased distance, an object will decrease in visible size, suggesting that a reduction in sound size also indicates a retreat of the sound source. This was shown to be the case through experimental analysis of sound source width on perceived auditory distance. An understanding of auditory distance perception might be furthered by assessing other processes used to sense visual distance.

List of References

Ashmead, D.H., Davis, D.L. and Northington, A., 1995. Contribution of listeners' approaching motion to auditory distance perception. *Journal of Experimental Psychology - Human Perception and Performance*, 21(2), pp.239-256.

Barron, M. and Marshall, A. H., 1981. Spatial impression due to early lateral reflections in concert halls: the derivation of physical measure. *Journal of Sound and Vibration*, 77, pp.211-232.

Blauert, J., 1983. *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge: MIT Press.

Blumlein, A., 1933. *Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing Systems*, British Patent Specification no. 394325.

Bradley, J.S. and Soulodre, G.A., 1995. The influence of late arriving energy on spatial impression. *Journal of the Acoustical Society of America*, 97, pp.2263-2271.

Bregman, A., 1994. *Auditory Scene Analysis; The Perceptual Organization of Sound*. Cambridge: MIT Press.

British Standards Institution, 2009. *BS 8233-1:2009 Acoustics — Measurement of room acoustic parameters Part 1: Performance Spaces.* Milton Keynes: BSI.

Brungart, D. and Rabinowitz, W., 1999. Auditory localisation of nearby sources. Head- related transfer functions. *The Journal of the Acoustical Society of America*. 106, pp.1465-1579.

Coleman, P., 1962. Failure to Localise the Source Distance of an Unfamiliar Sound. *The Journal of the Acoustical Society of America*, 34, pp.345-346.

Everest, A., and Pohlmann, K. C., 2015. *Master Handbook of Acoustics*. 6th ed. New York: McGraw-Hill Education.

Ferris, S.H., 1972. Motion parallax and absolute distance. *Journal of Experimental Psychology*, 95(2), pp.258–263.

Gerzon, M., 1992. The Design of Distance Pan-pots. *Audio Engineering Society*, 92, pp.1- 33.

Gibson, J.J., 1950. *The perception of the visual world*. Cambridge: The Riverside Press.

Hirsh, I., 1971. Masking of Speech and Auditory Localization. *Audiology*, 10, pp.110-114.

Holden, M., 2016. *Acoustics of Multi-Use Performing Arts Centres*. Florida: CRC Press.

Howard, I.P. and Rogers, B.J., 1995. *Binocular vision and stereopsis*. New York: Oxford University Press.

Howard, I.P. and Rogers, B.J., 2012a. *Perceiving in Depth: Volume I*. New York: Oxford University Press.

Howard, I.P. and Rogers, B.J., 2012b. *Perceiving in Depth: Volume II*. New York: Oxford University Press.

Kilpatrick, F.P., and Ittelson, W.H., 1951. Three demonstrations involving the visual perception of movement. *Journal of Experimental Psychology*, 42(6), pp.394–402.

Koffka, K., 2014. *Principles of Gestalt Psychology*. Oxford: Routledge.

Mather, G., 1996. Image Blur as a Pictorial Depth Cue. *Proceedings of the Royal Society: Biological Sciences*. 263, pp.169–172.

Mershon, D. and Bowers, J., 1979. Absolute and relative cues for the auditory perception of egocentric distance. *Perception*, 8, pp.311-322.

Michelsen, J. and Rubak, P.,1997. Parameters of Distance Perception in Stereo Loudspeaker Scenario. *Audio Engineering Society*, 102, pp.1-14.

Moore, B.C.J., 2008. *An Introduction to the Psychology of Listening*. 5th ed. Bradford: Emerald Group Publishing Limited.

Neisser, U., 1967. *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Newman, E.B. and Stevens, S.S., 1936. The localisation of actual sources of sound. *American Journal of Psychology*, 48(2), pp.297-306.

O'Shea, R.P., Blackburn, S.G. and Ono, H., 1994. Contrast as a depth cue. *Vision Research*. 34(12), pp.1595–1604.

Palmer, S.E., 1999. *Vision science: Photons to phenomenology*. Cambridge: Bradford Books/MIT Press.

Strutt, J. W., 1909. On our perception of the direction of sound. *Royal Society of London*, 83(559), pp.61-64.

Tyrrell, E., 2016. *The Distance Pan-pot: An Alternative Approach to the Distance Effect*. Trinity College Dublin. Available at: <https://eoghantyrrell.com/the-distance-pan-pot-an-alternative-approach-to-the-distance-effect/> [Accessed 10 October 2020].

Tyrrell, E., 2020. *Sound Source Width Adjustment as a Cue for Auditory Distance.* [online] (21st September 2020) Available at: <https://auditorydistanceonline.questionpro.com> [Accessed 23rd November 2020].

Von Békésy, G., 1960. *Experiments in Hearing*. New York: McGraw-Hill.

Wallach, H., 1940. The role of head movements and vestibular and visual cues in sound localisation. *Journal of Experimental Psychology*. 27(1), pp.339-368.

Weinstein, S., 1957. The Perception of Depth in the Absence of Texture-Gradient. *The American Journal of Psychology*. 70(4), pp.611–615.

Wheatstone C., 1838. Contributions to the physiology of vision - Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision'. *Philosophical Transactions of the Royal Society of London,* 128, pp.371-394.

Zahorik, P., 2002. Assessing auditory distance perception using virtual acoustics. *Acoustical Society of America*, 111, pp.183-1846.

Appendix I – Survey Data

Audition both samples. Which sample (Sample [X] or Sample [Y]) sounds more distant?				
Question Number	**File Name**	**Sample Number**	**Sample Width**	**Percentage (%)**
1	Binaural Piano 1 (Mono)	Sample 1	Mono	17%
	Binaural Piano 1 (Stereo)	Sample 2	Wide	50%
	No Difference			33%
2	Normal Piano 1 (Stereo)	Sample 3	Wide	22%
	Normal Piano 1 (Mono)	Sample 4	Mono	61%
	No Difference			17%
3	Drum Wet 1 (Wide)	Sample 5	Wide	28%
	Drum Wet 1 (Mono)	Sample 6	Mono	28%
	No Difference			44%
4	Ensemble Wet 1 (Mono)	Sample 7	Mono	61%
	Ensemble Wet 1 (Wide)	Sample 8	Wide	17%
	No Difference			22%
5	Speech Wet 1 (Wide)	Sample 9	Wide	6%
	Speech Wet 1 (Narrow)	Sample 10	Narrow	83%
	No Difference			6%
6	Binaural Piano 2 (Stereo)	Sample 11	Wide	28%
	Binaural Piano 2 (Mono)	Sample 12	Mono	61%
	No Difference			11%
7	Drum Wet 1 (Narrow)	Sample 13	Narrow	22%
	Drum Wet 1 (Wide)	Sample 14	Wide	11%
	No Difference			67%
8	Ensemble Wet 2 (Wide)	Sample 15	Wide	5%
	Ensemble Wet 2 (Narrow)	Sample 16	Narrow	28%
	No Difference			67%
9	Speech Wet 1 (Mono)	Sample 17	Mono	83%
	Speech Wet 1 (Wide)	Sample 18	Wide	6%
	No Difference			11%
10	Normal Piano 2 (Mono)	Sample 19	Mono	29%
	Normal Piano 2 (Stereo)	Sample 20	Wide	29%
	No Difference			41%
11	Drum Wet 2 (Wide)	Sample 21	Wide	29%
	Drum Wet 2 (Narrow)	Sample 22	Narrow	24%
	No Difference			41%
12	Drum Wet 2 (Mono)	Sample 23	Mono	18%
	Drum Wet 2 (Wide)	Sample 24	Wide	12%
	No Difference			70%
13	Ensemble Wet 1 (Narrow)	Sample 25	Narrow	6%
	Ensemble Wet 1 (Wide)	Sample 26	Wide	35%
	No Difference			59%
14	Speech Wet 2 (Narrow)	Sample 27	Narrow	24%
	Speech Wet 2 (Wide)	Sample 28	Wide	12%
	No Difference			47%
15	Ensemble Wet 2 (Wide)	Sample 29	Wide	0%
	Ensemble Wet 2 (Mono)	Sample 30	Mono	71%
	No Difference			18%
16	Speech Wet 2 (Mono)	Sample 31	Mono	35%
	Speech Wet 2 (Wide)	Sample 32	Wide	18%
	No Difference			41%

This appendix presents the survey questions and results in a tabular form. Thirty participants completed the survey. The results are expressed as a percentage. Some participants indicated that an error occurred when loading the audio samples. Therefore, some test answers do not sum to 100%.

The piano samples were anechoic. The piano samples were changed in width via amplitude panning, i.e. via redistributing both stereo channels to both speakers at once. The “wet” samples are those processed to have reflections; their width was altered by changing the spread of the reflections across the stereo stage.

The question presented to participants: “Audition both samples. Which sample (Sample [X] or Sample [Y]) sounds more distant?”