Development of a Perceptual Model for the Trade-off Between Interaural Time and Level Differences for the Prediction of Auditory Image Position

Nikita Goddard

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of MSc by Research

Applied Psychoacoustics Lab (APL)
The University of Huddersfield

Supervisor: Dr. Hyunkook Lee

January 2020

Copyright statement

ABSTRACT

In the current study, a simplistic and linear ITD-ILD trade-off model based on subjective data has been proposed to allow for various combinations of interaural time and level differences (ITD and ILD) for positioning auditory images at azimuth angles up to 60°. This can also be used to predict the azimuths of auditory images from input ILD and ITD. Independent ILD and ITD values used to generate the model were obtained through a lateral pointer task using a novel method which utilises HRIRs for defining azimuth position. From analysing the model, it was found that combination values were smaller compared to natural combinations extracted from HRTF in localisation. This was also apparent in a virtual localisation experiment where subjects used a HRIR pointer to report the azimuth of auditory image, laterally displaced by various ILD-ITD combinations from the model. It was found that the perceived angles were persistently narrower than their target angles. This underestimation was more significant for wider target angles above 45° and for ILD/ITD combination ratios with a larger weighting of ILD.

ACKNOWLEDGEMENTS

First and foremost, I would like to sincerely thank Dr. Hyunkook Lee for his excellent level of expertise and support throughout my studies. I value your enthusiasm in the realm of psychoacoustics.

I would also like to give many thanks to former and present members of the Applied Psychoacoustics Lab (APL); Dale Johnson, Bogdan Bacila, Maksims Mironovs, Alan Pawlak, Connor Millns, Andrew Parker, Maddie Nastasa, for your support and participation in my experiments, this study would have not been possible without. I would also like to thank the awesome staff at Pikcells, who provided social support and a way to fulfil my interest in web development outside of my studies. Finally, I would like to thank my friends and my beloved father, your memory will be with me always.

CONTENTS

List of Figures

List of Tables

List of Abbreviations

AMOA	Adaptive Method of Adjustment
ASW	Apparent Source Width
EC	Equalisation-Cancellation
EI	Excitation-Inhibition
GUI	Graphical User Interface
HRTF	Head Related Transfer Function
HRIR	Head Related Impulse Response
IACC	Interaural Cross-correlation Coefficient
IACF	Interaural Cross-correlation Function
ICLD	Interchannel Level Difference
ICTD	Interchannel Time Difference
ILD	Interaural Level Difference
ITD	Interaural Time Difference
JND	Just Noticeable Difference
MAA	Minimum Audible Angle
MOA	Method of Adjustment

1. Introduction

Spatial hearing is a common process of the auditory system and an essential part of everyday listening to help make sense of sounds in the surrounding environment. In the modern era, the study of spatial hearing is of most importance for emerging technologies such as Virtual Reality (VR) and Computation Auditory Scene Analysis (CASA). One of the main questions that arises in spatial hearing is how one perceives the direction of a sound source. In the current study, it has been observed that there is an abundance of models developed to predict sound source locations by simulating how the ears and the brain interpret incoming auditory signals and binaural cues, specifically interaural time and level differences (ITD and ILD), most of which is based on Head Related Transfer Functions (HRTF) or neurological findings. These are, however, limited by computational efforts and require neurological knowledge. Consequently, there is a gap for a simplistic and perceptually motivated prediction model that can derive ITD and ILD for various locations, and vice versa. Such a model can allow for various trading ratios of ILD/ITD which can be useful for designing new panning methods and microphone techniques. For example, such designs could trade less ILD and more ITD for increased spaciousness or less ITD or more ILD for increased localizability of sound sources. Unlike current panning and microphone techniques that rely on interchannel time and level differences (ICLD and ICTD) between microphones and loudspeakers, those based on the ITD/ILD model will not be limited to certain loudspeaker configurations and may exhibit improved localisation accuracy as it is known that ITD and ILD, the essential cues for localisation, are produced by the summed auditory images generated from ICTD and ICLD,

Within this study, Chapter 2 examines the fundamental psychoacoustic principles of sound localisation and lateralisation, along with the current models that have been developed to obtain binaural cues and/or to predict source locations. Chapter 3 describes a subjective experiment involving a novel method of obtaining independent ILD and ITD for azimuth positions. Chapter 4 presents the proposed interaural trade-off model based on experimental data obtained in Chapter 3 which is verified objectively through comparison with HRTF and subjectively through virtual localisation experiments.

2. Background Research

Within this chapter, existing literature will be reviewed involving the fundamentals of binaural hearing, the psychoacoustic principles of how humans localise sources and the associated models that have been proposed to understand and predict localisation. Analysis of this literature is also beneficial in gaining knowledge and finding potential gaps and questions that can influence the development of the current study.

2.1 Psychoacoustic Principles of Localisation and Lateralisation

Humans hearing is a process of the auditory system comprising of three parts: the outer ear, middle ear and the inner ear. Sound waves transmitted from an external sound source are modified by the outer ear, known as the pinnae, and are transported through the ear canal to the middle ear where acoustic vibrations are transferred to the inner ear via the tympanic membrane and the ossicles, the three tiny bones that transmit sound waves within the air to the fluid-filled cochlea. The inner ear is responsible for converting vibrations into electrical impulses that are fired to the brain for processing and interpretation. Frequency bands of the input signals are also analysed via the basilar membrane within the inner ear.

One of the basic functions of the auditory system is localisation, referred as the ability to perceive the direction of a sound source within a 3-dimensional space. Horizontal localisation is associated with binaural cues such as the interaural differences between the input signals of the two ears, further explored in section 2.1.1, whereas monaural cues from a single ear are primarily involved in vertical localisation, distance perception and front-back disambiguation, as discussed in section 2.1.3. Horizontal localization will be predominantly explored because vertical localisation will increase the complexity of the model within this study.

Another aspect of localisation is lateralization, often described as the perceived direction of source within a 1-dimensional space, along the axis between the ears (Jens Blauert, 1997a). As opposed to localisation, lateralisation produces intracranial (inside the head) images rather than external images that still rely on the same binaural cues. This is often prominent throughout early studies on binaural hearing and localisation where dichotic listening experiments, involving the use of headphones, are carried out to isolate localisation cues from the spectral and timbral characteristics of the environment. Previous researchers such as (Jeffress & Taylor, 1960) and (W. Hartmann & Wittenberg, 1996) have considered lateralisation and localisation as the same phenomenon. For example, Jeffress & Taylor (1960) stated that every point along the lateral axis of the ears is said to correspond to the azimuth of an auditory event. Now it can be said that their differences are the presence of information about the environment and spectral cues from the pinnae (Plenge, 1974), such of which is present in localisation. The main question is whether lateralisation data obtained from studies can be applied to real world localisation. In the real world, we are surrounded by room reflections which distort the temporal and spectral envelopes of the sound entering the ears, affecting interaural cues (Rakerd & Hartmann, 1985). Despite this, studies in lateralisation help us understand localisation in a controlled environment without biases from varying room acoustics.

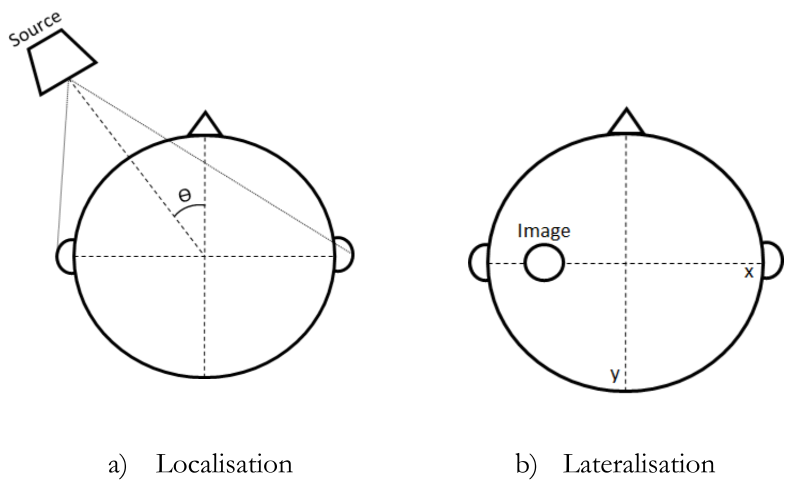


Figure 2.1: Visual representation of localisation with external sound sources (a) and lateralisation with internal auditory images (b)

2.1.1 Duplex Theory

An important explanation of horizontal localisation is the duplex theory, proposed by Lord Rayleigh (1907). It explains that sounds are localised through interaural time differences (ITD) and interaural level differences (ILD) at the ears, both of which are frequency dependent. ILD is said to play a major role in high frequency sounds above 1500Hz, and ITD for low frequency sounds below 735Hz. This can be proven from objective measurements performed by (Kuhn, 1977) who determined poor azimuth localisation from 1.4 to 1.6kHz and improvement of localisation above 3kHz with increasing ILD. Even though the theory is highly supported by subjective findings using pure tones (Lord Rayleigh, 1907; Feddersen et al 1957; Sayers, 1964) the duplex theory can also be confirmed in dichotic free-field experiments for noise bands whereby ITD is weighted strongly for low pass bands, and ILD is weighted strongly for high pass bands (Macpherson & Middlebrooks, 2002).

An extension of the duplex theory explains how the auditory system can extract ITD from high frequency envelopes, i.e. high frequency transient stimuli (Bernstein & Trahiotis, 1985). Above 1.6kHz, the auditory system detects lateral displacement as a function of the delays between envelopes (Leakey, 1959). In contrast, the role of ILDs in low frequency content has been shown in more current studies using transaural experiments where speakers are synthesised to reproduce real sources laterally between the ears. ILDs were as large as 8dB for 750Hz, 4dB more than predictions from the spherical head model (Hartmann, 2016)

The duplex theory along with its extensions provide insight into horizontal localization, however this does not regard the importance of spectral cues from the pinnae of the ear that are important for elevation and resolving front-back confusion (Gelfand, 2009). In addition, it does not consider the relationship between the interaural cues in complex real-world sources that contain both high and low frequency components.

2.1.1.1 Interaural Time Differences

ITDs are described as difference between the arrival time of ear input signals which increase as the source shifted from the median plane, i.e. the centre axis. As the source is displaced towards one of the ears, ITD increases linearly up to approximately 0.65ms, the maximum binaural time delay. This is also referred to as the maximum interaural phase difference, time delay as a function of frequency, whereby the wavelength is equated to the average distance between ears. Above 1500Hz, the wavelength decreases, reducing the resolution of the detection of ITD due to fine phase differences, therefore increasing localisation errors. However, this is more applicable for pure tones, whereas in broadband stimuli consisting of wider frequency bandwidths, ITD can be a dominant cue over ILD (Wightman & Kistler, 1992). Also, ITDs are also known to be detectable in high frequency envelopes (Bernstein & Trahiotis, 1985)

As a function of azimuth, ITDs can be assessed in multiple ways. Woodworth (1938), derived ITDs through the shape of a spherical head, known as the spherical head model (SHM), given as;

$$ITD = \frac{r\theta + rsin(\theta)}{\mathrm{c}} \tag{2.1}$$

where r is the radius of the head, c is the speed of sound in the air at approximately 340m/s, and θ is the source angle in radians. For the average diameter of the human head of 22-23cm, the ITD is approximately 0.66ms for a 90° azimuth angle. Precedingly, Fedderson et al. (1957) derived ITDs by placing microphones within the ear canal, resulting in an ITD of 0.68ms for pure tones at 90°, see Figure 2.2. Kuhn (1977) used a similar measuring technique but with a manakin, showing frequency dependency of ITD below 500Hz and above 3000Hz.

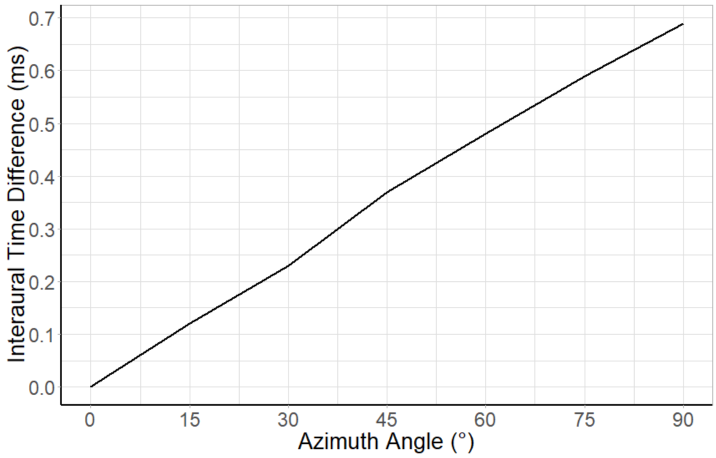


Figure 2.2: Interaural Time Difference (ITD) for pure tones up to 90° azimuth (Feddersen et al., 1957).

2.1.1.2 Interaural Level Differences

The difference of amplitude or sound pressure level between the ears as the source is positioned away from the median plane is referred to as ILD. This is influenced by the head shadowing effect. At low frequencies, the wavelengths are too large to be obstructed by the head, resulting in lower ILDs. However, for high frequencies, there is a trend involving the increase of ILD up to 20dB with the increase of frequency across all azimuths. This is supported by localisation studies using broadband noise and ⅓ octave band noise, where ILDs were extracted from ear canal measurements at multiple azimuths. For 90°, ILDs were around 20dB at 4kHz and around 35dB at 10kHz (Middlebrooks, 1989) or 2 dB at 200 Hz to over 20 dB above 6 kHz (E. a. G. Shaw, 1974). For pure tones, ILDs measured by Fedderson et al. (1957) showed values up to 10dB for frequencies below 1800Hz and up to 20dB up to 6kHz, see Figure 2.3.

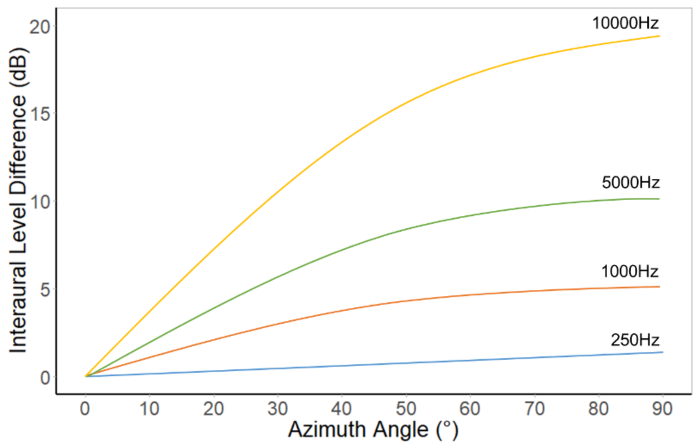


Figure 2.3: Interaural Level Difference (ILD) for pure tones at 250Hz, 1000Hz, 5000Hz and 10000Hz up to 90° azimuth (Gulick et al., 1989).

2.1.2 Interaural Trading

Due to the spectral and temporal complexity of real-world sources, it can be said that ILDs and ITDs are combined and not isolated in localisation (Banister, 1926). The main question is how these cues are weighted. In classic studies, the relationship between ITD and ILD is a time-intensity trading ratio, expressed in uS/dB, that derives the equivalent ITD of an ILD and vice versa (Hafter & Jeffress, 1968; Moushegian & Jeffress, 1959; Stecker, 2010a).

2.1.2.1 Variation of Trading Ratios

Literature has shown varying trading ratios from 2 to 200uS/dB (Jens Blauert, 1997a) that are biased by varying sound pressure levels and the use of different stimuli with varying frequency and duration. The difference in sound pressure levels across experiments signifies that trading ratios are dependent on loudness. For higher sound pressure levels, more ILD is required for a given ITD, whereas the contrary can be shown for lower levels (Deatherage & Hirsh, 1959). Hafter & Jeffress (1968) also verified this for ITD with a given ILD.

For stimuli, it has been found that frequency impacts trading ratios. Low frequency clicks tend to exhibit smaller ratios than high frequency clicks which can be attributed to the increase level difference at high frequencies, as expressed by the duplex theory, which significantly alters the time difference (Deatherage, 1961; Harris, 1960). It has also been shown that clicks exhibit larger ratios of 80 to 100uSec/dB than tone stimuli from 0.3 to 2.5uSec/dB (Whitworth & Jeffress, 1961), signifying that amplitude envelopes have an effect on TR. However, this can also be attributed to response latency being greatly affected by short stimuli.

There is an abundance of literature that focuses on trading ratios primarily with single frequency pure tones, clicks or both. These trading ratios of high or low frequencies are in accordance to the duplex theory, i.e. high frequency tones favour ILDs so the trading ratio will be more favourable to ILD. On the other hand, limited research has shown that trading ratios of complex sources are based on the extensions of the duplex theory. For example, with high-pass white noise clicks, the trading ratios can be attributed to both ILDs at high frequencies and ITDs detected in the envelopes at high frequencies (David et al., 1959).

2.1.2.2 Measurement of Trading Ratios

Trading ratios are primarily obtained through lateralisation experiments with dichotic listening where the ear channels and stimuli parameters can be easily and directly controlled. A classic method is the centring method which involves centring a lateralized image, displaced by a given ILD or ITD, by adjusting the opposing cue in the opposite direction. This assumes that centring an image gives a point of equivalence between 2 cues, however, it disregards the idea that ITD and ILD are affected by lateral position (Domnitz & Colburn, 1977).

Moushegian & Jeffress (1959) altered the classic centring method by matching a pointer altering ILD and ITD to the opposing cue at a fixed lateral position. Although the results are less ambiguous compared to the centring method by incorporating an objective reference, the resulting trading ratio is still biased to the adjusted cue. According to

Trahiotis & Kappauf (1978), larger trading ratios are found in studies with an adjustable ILD pointer (Whitworth & Jeffress, 1961) than an ITD pointer (Domnitz & Colburn, 1977) for tonal stimuli at 500Hz. Subjects within the Domnitz & Colburn (1977) study reported a 'clouding' of the reference ITD image, which is most likely referred to as the spread of the image as the ITD is displaced away from the median plane. This makes responses prone to ambiguity.

In trading experiments, subjects have reported broad or split images (Hafter & Jeffress, 1968; Whitworth & Jeffress, 1961). Banister (1926) hypothesised that one image is affected by intensity difference and the other is affected by the time difference. In later studies, it has been found that the "time image" is dominated by ITD and the "intensity image" is dominated by both ILD and ITD. This assumes that trading ratios are inconsistent as subjects are focusing on one or the other, both exhibiting different ratios. Hafter & Jeffress (1968) found that TRs for the time image (TR < 10μs/dB for 500 Hz tones, 2 – 35μs /dB for high-pass clicks) are less than that of the intensity image (TR ≈ 20–50μs/dB for tones, 80 – 150μs/dB for clicks).

2.1.3 Head Related Transfer Function

One of the most prominent limitations of the trading ratio is that it does not reference a certain azimuth direction, therefore it is difficult to apply interaural relationships to the perceived direction of a sound source. The Head Related Transfer Function resolves this limitation by revealing knowledge on the natural combination of ILD and ITD for any position within a 3D space.

The transfer function *h* applied to an input signal *x* in the time domain for each ear can be written as:

$$x_L(t) = \int h_L(\tau)x(t-\tau)d\tau$$

$$\boldsymbol{x_R(t) = \int h_R(\tau)x(t-\tau)d\tau}$$

(2.2)

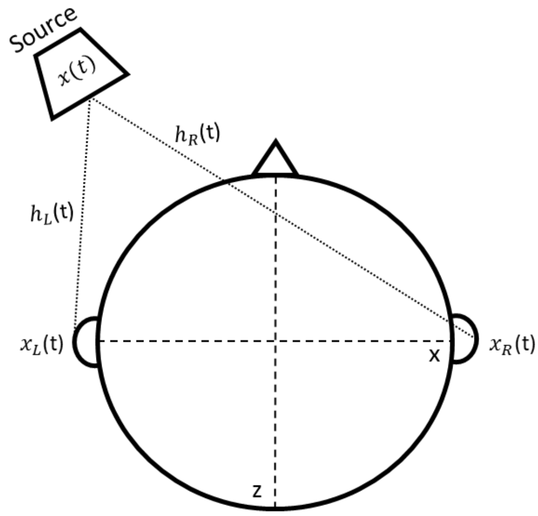


Figure 2.4: Signal *x(t)* filtered with separate transfer functions $\boldsymbol{h_L(t)}$ and $\boldsymbol{h_R(t)}$.

The HRTF contains both spectral and directional information from the pinnae of the ear for vertical and front-back localisation that cannot be explained by the duplex theory, which is limited to +/-90° (Jens Blauert, 1997a; J. C. Middlebrooks & Green, 1991). Without the pinnae, ILD and ITD cues will be constant for multiple locations, resulting in a cone of confusion with its apex positioned at the centre of the head and its axis lying between the ears. This supported by Fisher & Freedman (1968) whose subjects reported more localisation ambiguities when a tube was inserted into their ear to 'bypass' the pinnae.

Comparable to interaural cues, pinnae cues within HRTFs are also frequency dependent. These effects tend to appear around 3kHz, where the wavelength is equivalent to the size of the pinnae (Shaw, 1997). The HRTF can be represented as a third interaural cue and can be applied to the duplex theory for high frequencies to resolve directional ambiguity and front-back discrimination. At low frequencies, many studies have considered the torso to introduce directional and distance effects below 3kHz (Algazi et al., 2001).

2.1.3.1 Measurement Method

HRTFs are all distinguishable based on various characteristics; including the type of source, measuring technique, number of source locations, equipment and measuring procedures (Park, 2007). The type of subject is the most prominent factor that determines the equipment and the measurement procedure. An acoustic manakin or a Head and Torso Simulator (HATS) provides a straightforward procedure that reduces the time taken to measure impulse responses. However, for human subjects, the procedure is time-consuming and difficult. It also considers careful selection of flat frequency response microphones and handling when they are inserted into their ears. Nevertheless, the positioning of the subjects has the most critical impact on the resulting ILD and ITD based on localisation errors exhibited by head movements. Head tracking and using a backrest or headrest have been used to restrict movement and posture, but movement is inevitable and such methods can impact the sound field around the subject. In a study conducted by Denk et al (2017) head movements were recorded via a head tracker and subjects corrected their movement based on visual feedback. Results showed deviations as little as 0.3° in the horizontal axis. This is a significant improvement to that of unconstrained subjects within the Hirahara et al, (2010) study which showed deviations of 10°.

Once the measurement procedure has been identified from the type of subject and measurement equipment, a HRIR (Head Related Impulse Response), the time-domain representation of HRTF, is usually recorded at multiple positions based on the loudspeaker array and the rotation of the subject. Typically, the measurements are conducted in an anechoic chamber, free from early and late reflections that can greatly affect the reflections within the pinnae. Input signals from loudspeakers are usually sinusoidal sweeps or pink noise. Sweeps are often more beneficial due to high signal-to-noise ratios (SNR), limited harmonic distortion artefacts and time-variance effects (Carpentier et al, 2014). To compensate for the spectral effects imposed from the frequency response of the microphones and loudspeakers, an inverse filter of their responses can be convolved with the recorded HRTF, known as deconvolution (Armstrong et al, 2018).

2.1.3.2 Binaural Synthesis

Binaural synthesis is becoming increasingly popular along with the development of 3D technology which aims to accurately replicate real environments. This type of synthesis is used to render an auditory image into a 3D dimension space, providing an externalised image through headphones. One way to accurately match the auditory image to a real source, is by convolving the input signal with each of the 2 channels of an HRIR, each representing how the ear receives incoming free-field sources. Each HRIR is specific to a certain direction and are usually taken from a database containing many azimuths and/or elevations.

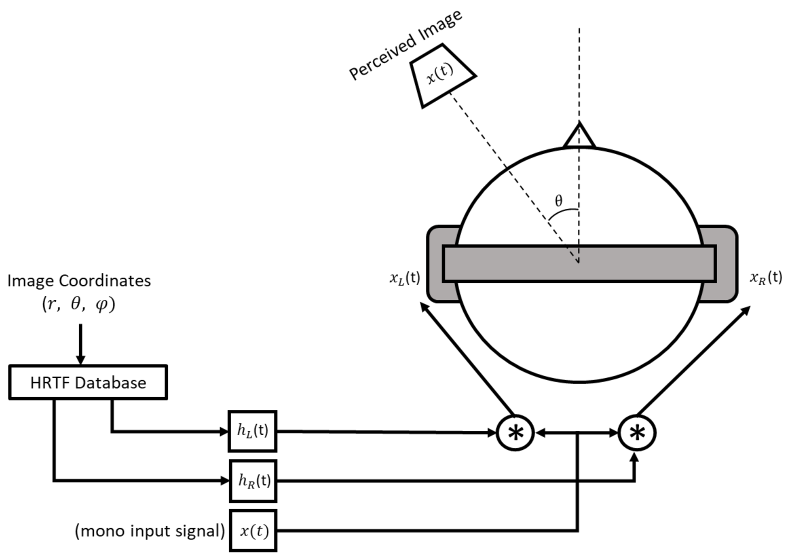


Figure 2.5: Schematic of an input signal convolved with HRIR. The resulting auditory image is perceived at the HRIR position between headphones.

When selecting a database, the choice of non-individualistic and individual HRIRs are important. Although non-individualised HRIRs can be generalised to every listener, auditory images tend to be intracranial, i.e. less externalised or not originating from outside the head. According to Armstrong et al. (2018) on the subjective evaluation of synthesized individual and non-individualized´HRTFs, externalisation did not hinder the subjects preferences. Subjects even preferred HRIRs from acoustic manikins, mostly the KU100. As discussed in Section 2.1.3.1, a less favoured individual HRTF could be a consequence of the microphone selection and subjects' movements during HRTF measurements. Besides preference, individualised HRIRs can help improve externalisation, with the addition of head tracking to correct front-back confusion and incorporating realistic room acoustics (Begault & Wenzel, 2001).

Other than 3D technology, binaural synthesis is also useful for psychoacoustic research, especially for pointer and matching tasks described in Section 2.1.2.2. As previously discussed, research mainly conducts lateralisation experiments for investigation into free-field localisation to limit interference of room reflections. By using HRIRs in lateralisation experiments, externalised sources can be simulated without interference of reflections. An example of this usage is within an experiment conducted by Park (2007) where an acoustic pointer was created using HRIRs which was used to specify the localisation of a lateralized image shifted by combined ILD and ITD.

2.2 Binaural Models of Auditory Position Estimation

To replicate the processes of the auditory system and to simulate binaural hearing, researchers have proposed various binaural models to explain simple and complex phenomenon including localisation. There has been a trend of computational models that provide neurological explanation such as how the brain interprets interaural signal differences rather than the modification processing of the signals entering or present within ears. The most influential of the models is the cross-correlation model of (Jeffress, 1948) and the equalisation-cancellation model of (Durlach, 1963) which have influenced many recent studies, as further discussed below.

2.2.1 Jeffress Cross Correlation Model

The cross-correlation model is the earliest and most profound model of binaural hearing that explains binaural interaction. It utilises a neural coincidence detection network using time-sensitive auditory neurons to determine ITD, therefore determining the azimuth of a sound source. This network comprises of 2 delay lines where the 2 ear signals are fed. At different taps along the delay lines, the signals are multiplied with coincidence detectors then a running integration and summation is performed (Jeffress, 1948), see Figure 2.6. For example, if the signal arrives at the right ear first, the coincidence detectors on the left delay line will be activated. When no time difference between the signals are employed, the middle coincidence detector is activated.
In accordance to the duplex theory, it can be assumed that the determined ITD will become more ambiguous for high frequencies. However, as stated before, neurons could also be highly sensitive to time differences in high frequency envelopes (Bernstein & Trahiotis, 1985).

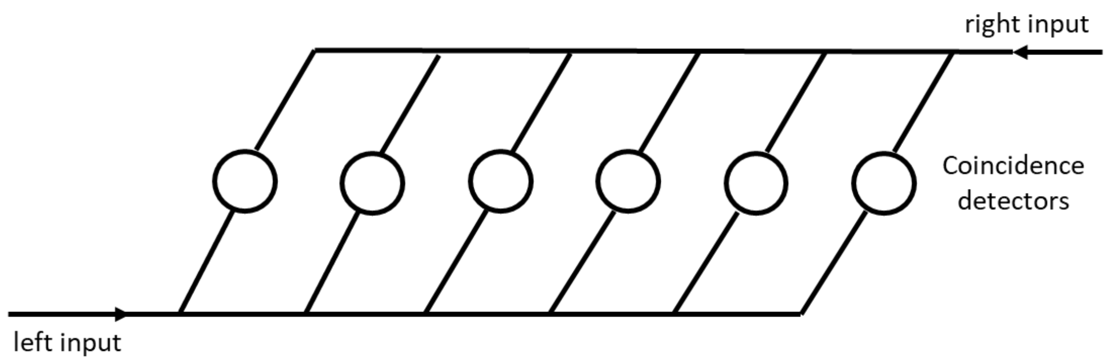


Figure 2.6: Schematic of the coincidence detection network of Jeffress 1948.

2.2.1.1 Extensions of the Jeffress Model

Consequently, after the Jeffress model, various researchers have developed extensions of the model that provide additional processes to account for ILD, the precedence effect, and the spectral complexity of real-world sources (W. A. Yost, 1993). The position-variable model, developed by (Stern & Colburn, 1978), provides a gaussian-shaped intensity weighting of the Interaural Cross Correlation Function (IACF) associated with the lateral position along the interaural delay axis which depends on the ILD of the ear signals. Although the model considers the interaction of ITD and ILD, only the lateral position of pure tones can be accurately described when compared to binaural data for 500Hz (Domnitz & Colburn, 1977) and up to 1200Hz (Shear & Stern, 1987). Therefore, imposing an arbitrary weighting function the model is seen as less favoured than the alternative Lindemann extension, which now accounts for many binaural phenomena including ILD and the precedence effect.

The Lindemann extension consists of 2 additional processes (Lindemann, 1986). The first being the monaural detector that responds if the signal at the opposite ear has an amplitude of zero. The most important is the second process involving a contralateral inhibition mechanism which suppresses secondary peaks and sharpens initial peaks of the coincidence detector. The interaction between these processes causes the peaks to shift along the interaural delay axis with changes in ILD. This is said to increase sensitivity of ILD and simulates that direct sound contributes mainly to localisation even with the presence of early reflections. When comparing the model with psychoacoustic data from trading experiments with pure tones, for 500Hz the Lindemann model predicts a trading ratio of 36us/dB whereas the psychoacoustic data shows a ratio of 50us/dB (Young Lamar L., 1976). Experiments using a pointing technique exhibited a much lower ratio than the trading experiment and the model with 25us/dB (Domnitz & Colburn, 1977). Lindemann explained the discrepancies due to the model parameters simulating different psychoacoustic data measured by the trading and pointer technique. According to(Jens Blauert, 1997a), even though the model is useful for simulating binaural hearing in controlled environments and incomplete fusion of binaural signals, the model will need to be adapted for binaural signals with natural combinations of ILD and ITD, such as those present in HRTFs.

Gaik (1993) proposed an extension of Lindemann's model to account for natural combinations of interaural cues by applying a secondary weighting function to the coincidence detector. Opposed to the signal dependent weighting function of the position-variable model, this weighting function is derived from HRTFs that portray naturally imposed ITD and ILD within critical bands across the entire audible frequency range. As a result, the model considers frequency dependency of interaural parameters as well as natural combinations. Although HRTFs were taken from dummy head measurements, increasing the generalisability of the model, it does not consider the inter-subject differences of natural combinations as evident in individualised HRTFs (Armstrong et al., 2018). However, it is more applicable to real world localisation than previous extensions which focus on lateralization.

2.2.2 Equalisation-Cancellation Model

Developed by (Durlach, 1963), the EC model is a non-cross correlation approach where the input ear signals are equalised then subtracted or 'cancelled' out. Prior to cancellation, the input signals are multiplied with internal noise that consists of amplitude and temporal jitter. The power of the internally jittered ear signals results in a decision variable or EC factor which increases with ITD and/or ILD (W. A. Yost, 1993). Even though the model was primarily developed for detecting binaural masking-level differences, it has also said to account for other binaural phenomena including localisation where the EC factor is identified with interaural differences or interaural correlations. According to (Durlach & Colburn, 1978), the linear function EC provides equivalent predictions to the cross correlation. However, the EC model is more effective at describing thresholds for fixed and variable ILD and ITD and describing sensitivity to ILD and ITD (Breebaart et al., 2001).

Park et al. (2008) developed a simple pattern-matching model for localising real and virtual auditory images by using the EC model within a binaural processor in place of the cross-correlation model. The EC process is used to obtain Excitation-Inhibition (EI) patterns across frequency within the activity of a neural cell. These EI-patterns are said to contain ITD and ILD for source location estimation. When comparing the model to a subjective pointer experiment with amplitude-panned broadband stereophonic images, subject responses agreed well with model prediction, especially predictions concerning ambiguity across frequencies. In addition, the pattern-matching model is seen as more effective at dealing with both ILD and ITD than additional weighting models or mapping models, such as the Lindemann Extension (Lindemann, 1986) or the Position Variable Model (Stern & Colburn, 1978).

2.2.3 Other Methods of Position Estimation

The above methods and models concerning neurological explanations of auditory position estimation, are often computationally expensive or requires neuroscientific background. Therefore, a simplistic model or method is required that can be used across many faculties, such as for sound engineers or acousticians. There already exists models that allow estimation on a simpler computational or analytical level which fall into two categories; models that combine separate estimates from ITD and ILD and models that consider a single parameter from ITD and ILD (Park, 2007). The former type of model involves finding the azimuth angle for ITD and ILD separately. (Macpherson & Middlebrooks, 2002) used this approach where estimates from ITD and ILD were weighted. This was based on experimental data where subjects rotated their head to the perceived position of a stimulus which was convolved with a HRTF with imposed delay or amplitude attenuation. It was found that for wideband stimuli, ITD weighting was greater or equal to ILD weighting. However, these were not representative of certain azimuth positions. Raspaud and Viste (2010), developed a parametric model that allowed estimation of azimuth by estimating ITD and ILD for each spectral coefficient from a short time Fourier transform spectra of input signals. These were jointly evaluated with a lookup table of HRTFs. It emerged that azimuth estimates based on ILD have a larger standard deviation than estimates from ITD and that ILDs can be used to improve location estimation by resolving ITD ambiguities. The method was also deemed more accurate than the probabilistic method where the probability of azimuths is based on ITD, ILD and spectral cues within a hierarchical system (Danfeng Li & Levinson, 2003).

By obtaining single estimate from each binaural parameter, it assumes that they are not closely related. There have been methods that emphasizes the interaction of ILD and ITD by obtaining a single estimate from both ITD and ILD. One example that has been previously described is the time-intensity trading ratio where a single parameter is obtained from the equivalence of ILD and ITD (Hafter & Jeffress, 1968). Again, this does not reference the azimuth relating to the parameter. Lim and Duda (1994) similarly converted ITD to an equivalent ILD which were placed into vector and compared to vectors containing data from HRTFs. An inter-subject characteristic curve model developed by (Park, 2007), also used this approach by finding the nearest-neighbour of the target ILD and ITD within a characteristic curve of natural combinations that is unique to each individual at each frequency. This was most effective at predicting lateralisation of low frequencies.

3. Effect of Independent ILD and ITD on Azimuth Position

3.1 Introduction

This chapter summarises the first set of subjective experiments to derive independent ILD and ITD for multiple azimuth directions for the model within the study. It is true that ILD and ITD can be determined using the binaural models mentioned in Section 2.1 (Jeffress, 1948; Gaik, 1993), however these only consider processed ear signals within the inner ear. HRTFs can provide ILD and ITD at certain azimuth positions upon extraction, but these values cannot be isolated. In previous lateralisation studies which focus on ILD or ITD individually, a perceived image position was derived with a set of interaural values in which an ITD or ILD pointer is adjusted to (Feddersen et al., 1957; Hafter & Jeffress, 1968; Stecker, 2010a), or the lateral position of imposed ILD or ITD is described using a visual marker (Sayers, 1964). However, values obtained from these studies do not reference azimuth but lateral position, usually described with a scale of 0-10 (W. Yost, 1981; Gaik, 1993) or 0-5 (Sayers, 1964). Therefore, ILD and ITD have not been extracted directly for a certain azimuth direction.

The aim of this section is to derive independent ITD and ILD values that will be useful for the trade-off model proposed in Chapter 4. A novel method has been proposed using the same acoustic pointer used in previous lateralisation experiments, but this is instead matched with a virtually localised source synthesised with Head Related Impulse Responses (HRIRs). This is based on Park's (2007) study where HRIRs are used to report positions of laterally placed images. Even though the task is deemed unnatural due to comparing localisation and lateralisation, it allows ILD and ITD to be obtained for various azimuths. This data can also be useful for establishing auditory position as a result of time and level panning within headphones.

3.2 Experimental Design

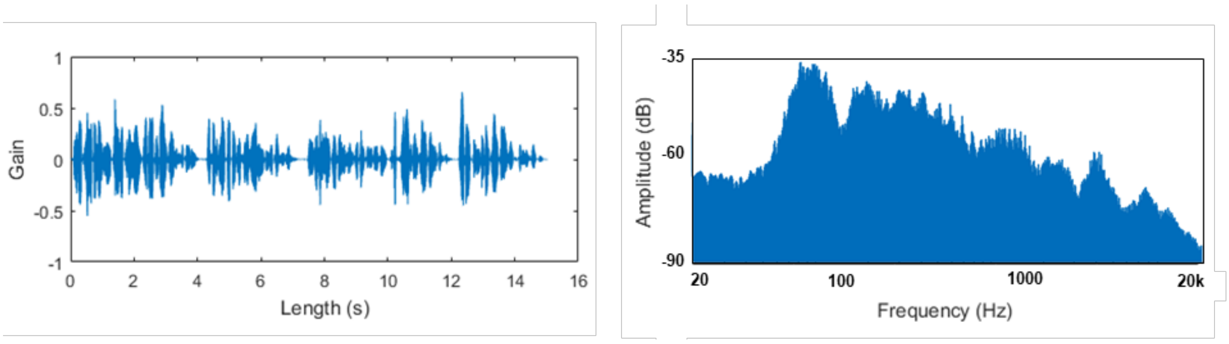
3.2.1 Methodology

One of the main questions emerged is how we can define azimuth position in a lateralisation experiment as an alternative to using the visual lateral marker used in previous lateralisation experiments. One potential method utilises a visual response method using a strip of LEDs, with an angular difference of less than 1°, positioned in an arc around the subject (Lee et al., 2016). This method was more effective than the visual marker method for localisation in terms of response time and intuitiveness. However, when pilot testing the method where a lateral ITD or ILD pointer is adjusted to the target LED position, difficulties have arisen when attempting to match an intercranial image with the externalised visual reference, which was placed 1m away. The issue of externalisation is mostly likely to produce measurement error and cross-modal bias when attempting to perceive the internal image to be radiating from the visual reference. The ventriloquist effect can also be induced, at which the auditory stimulus is pulled towards the visual stimulus, hindering localisation of the auditory stimulus. This has also been supported by Odegaard (2015) on visual and auditory bias, where visual stimuli are dominant in bi-sensory trials and tend to bias auditory localisation towards the median plane. To mitigate these errors and potential biases from a visual anchor, an acoustic anchor is preferred. In a study by Park (2007), HRIRs are employed as an acoustic pointer to report the position of laterally displaced images with combinations of ILD and ITD. As it is known that HRIRs are time-domain representations of HRTFs that contain pinnae cues and interaural cues from a certain spherical position, these will be most effective as an acoustic anchor to define azimuth position in a virtual space.

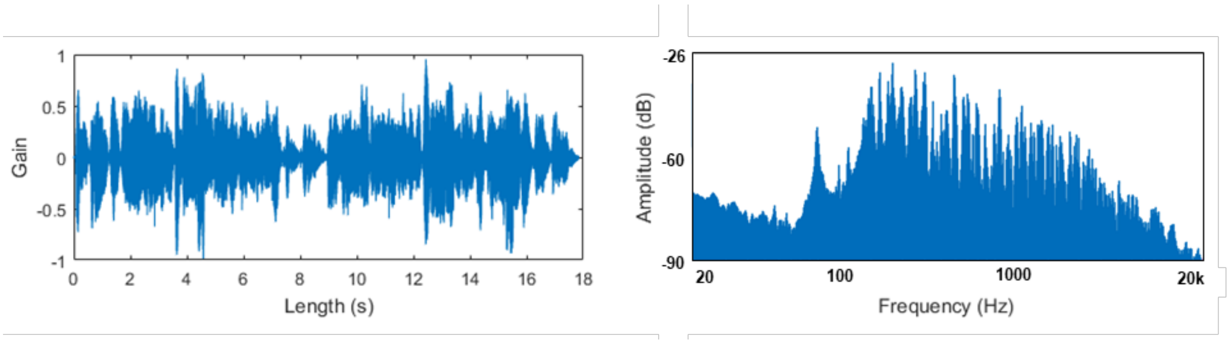
As commonly practised in past lateralisation and trading experiments (Feddersen et al., 1957; Whitworth & Jeffress, 1961; Domnitz & Colburn, 1977), an acoustic pointer will be used to adjust the independent variable, ITD and ILD, to the perceived position of an image in reference to azimuth. This is called 'Method of Adjustment' (MOA) (Bech & Zacharov, 2006) which is known to actively engage the subject, thus increasing concentration. Cardozo (1965) stated that MOA can be applied to situations of which perceptual attributes within stimuli vary. In this case, at wider azimuths, the perceived angle of the source may increase, especially with increasing ITD. This can be observed in data obtained from MOA than in forced choice methods for example. Another limitation of MOA is that subjects have more control of the parameter as opposed to the staircase method for example, this can lead to increase in errors due to large ranges. The staircase method allows for establishing thresholds between subjects (Levitt, 1971), though the procedure can be more tiring to subject if the experiment consists of many trials. Wallis & Lee (2017) implemented an adaptation of MOA called the 'Adaptive Method of Adjustment' where a threshold is approximated with a coarse step-size which is reduced a further two times. It was decided to incorporate a type of AMOA along with stand-alone MOA using only a single step size that can be incremented and decremented alongside the pointer. The step size is specified as the Just Noticeable Difference (JND) of ITD or ILD, respective to the pointer. For the ITD pointer, a JND of 0.01ms is chosen, and JND of 0.5dB for the ILD pointer (Stecker & Gallun, 2012). These values were obtained from broadband noise, which can be generalised to broadband stimuli used within the experiment. During pilot testing, it was revealed that the AMOA was effective in finetuning the interaural value to the perceived image position and reduced response error. The MOA slider was deemed difficult in accurately fine-tuning values at higher precision, therefore it the is used prior to AMOA to determine the threshold.

3.2.2 Stimuli

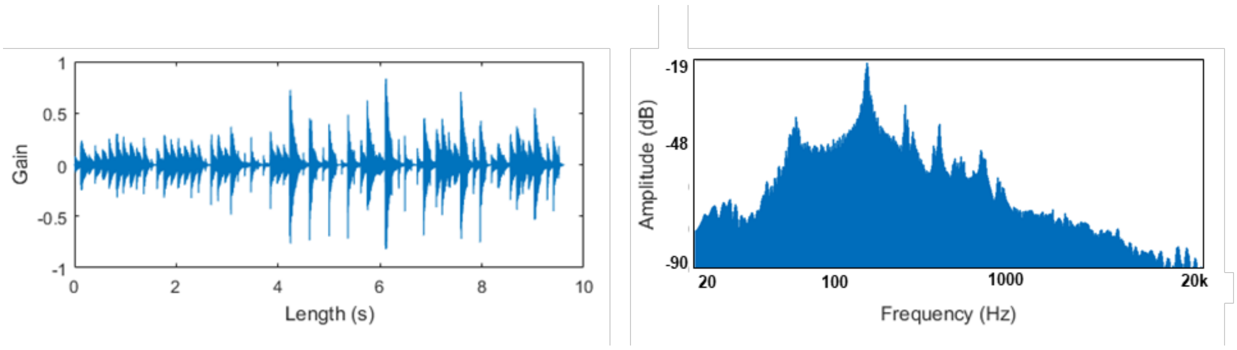
It is apparent that lateralisation studies based on independent interaural parameters mainly focus on the effects of frequency or temporal structure of stimuli, therefore pure tones and clicks have been widely used (Feddersen et al., 1957; Sayers, 1964; Hafter & Jeffress, 1968). However, results from these studies cannot be generalised to complex stimuli. In this experiment, generalisability is improved by using a broadband acoustic pointer. Three types of broadband stimuli were chosen. The first is speech, which is considered useful as it is what humans are familiar with and localise every day. The chosen speech sample will be an anechoically recorded sample in Danish, considering to be less distracting for English-speaking subjects. A cello sample is selected due to it continuous content and large frequency range, and conga drums are also selected for transient and low frequency content, as visualised in Figure 3.1. All samples are taken from the Bang and Olufsen 'Music for Archimedes' project (Hansen & Munch, 1991).



a) The waveform and frequency spectra of the Danish speech sample



b) The waveform and frequency spectra of the Cello sample



c) The waveform and frequency spectra of the Conga Drums sample

Figure 3.1: Waveforms and frequency spectra extracted from each stimuli shows their differences in terms of frequency range and temporal structure. These are used to increase generalisability of the data to other various sound sources.

For the acoustic HRIR anchor, pink noise bursts are used to prevent confusion between the anchor and the pointer. These are produced with an onset and offset of 1ms to prevent audible clicks and a duration of 270ms. The sustain is kept at the maximum level after the onset without any variable decay. Depending on azimuth position throughout each trial, the pink noise is convolved, using the 'multiconvolve~' object from the HISSTools Impulse Response Toolbox for Max 7 (Harker & Tremblay, 2012), with respective HRTFs from a KU100 Dummy Head taken from the SADIE II database (Armstrong et al., 2018).

3.2.3 Subjects

According to the ITU-R BS116 (2015) manual, the criteria for selecting subjects emphasises the use of experienced listeners who can disambiguate and analyse perceptual changes. This is also essential for obtaining reliable data for the model. Therefore, seven subjects were chosen who had mixed critical listening ability and experience in psychoacoustic experiments, including localisation tasks. All subjects, one female and 6 males, were undergraduate and postgraduate students, researchers and staff members within the Applied Psychoacoustic Lab (APL) at the University of Huddersfield and all confirmed to have normal hearing. Due to the small number of accessible and experienced subjects, an increase of four repetitions per 10 trials is required for more data points.

3.2.4 Interface

Figure 3.2, shows the test interface, developed with Max 7, presented to the subjects. The button marked as 'X' allows the subject to freely toggle between the acoustic pointer and the acoustic anchor which can also be triggered with key shortcuts for efficiency: Shortcuts are also available for the 'next' and 'audio on/off' buttons. The single slider along with the (+/-) buttons affect the lateral position of the acoustic pointer by increasing or decreasing ILD or ITD. This is configurable for the ITD or ILD session assigned to the subject. The ITD slider has a range of 0-4ms and the (+/-) buttons have a step size of 0.01ms, the JND of ITD for broadband sources. Milliseconds is converted to samples and is applied to the delay of the left or right audio signal depending on the positive and negative positions of the target angles, ranging from -90° to 90°. For the ILD slider, the range is 0-50dB, and the (+/-) buttons have a step size of 0.5dB, the JND of ILD. The response value is converted to theta θ given by:

$$\theta = \operatorname{atan}(ILD)$$

(3.1)

θ is used for constant power panning so minimal loudness changes are present when affecting ILD (Gaik, 1993):

$$g_1 = \cos(\theta)$$
$$g_2 = \sin(\theta)$$

(3.2)

g_1 and g_2 interchange for the left and right channels depending on negative or positive target angles. Confirmation of ILD is given by:

$$ILD = 20 \cdot \log 10 \frac{g_1}{g_2}$$

(3.3)

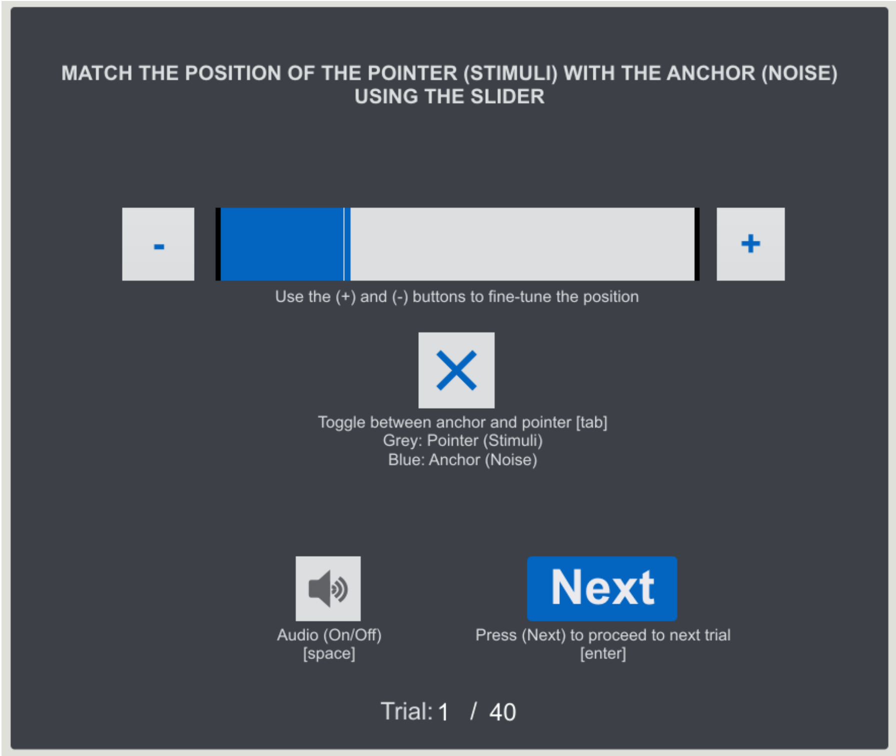


Figure 3.2: Screenshot of the test interface within Max 7 with a slider for MOA and (-/+) button for AMOA.

3.2.5 Physical Setup

The experiment was conducted in an ITU-R BS.1116 (2015) compliant critical listening room at the University of Huddersfield. Transparent curtains were placed around the subject, covering the loudspeakers in order to prevent visual bias, potentially induced by the speaker positions. The test stimuli were played through a Merging Horus audio interface configured to a sample rate of 44.1kHz, equivalent to the sampling rate of the HRIR convolution, audio samples, and the MAX 7 output. For playback, AKG 702 circum-aural type headphones were used. The Inverse filter of the AKG 702 headphone, recorded with a KU100 Dummy Head, the manakin used for the HRIRs, was convolved with the output of the HRIR renderer in order to reduce tonal effects of the headphone response that could alter image positions of the HRIRs. The playback level of the speech stimuli was set to an appropriate level of 68dB LA_{eq} by measuring the average level from one headphone cup using the Casella CEL-450 loudness meter. Other stimuli were matched by ear to the reference level of the speech stimuli with no interaural disparity and the anchor position at 0°. As presented in Table 3.1, different loudness levels measured are affected by varying dynamic range, temporal structure, and frequency content of the stimuli.

Table 3.1: Loudness levels of each stimulus measured as dBA.

Test Stimuli	Playback Level (dBA)
Speech	68
Cello	72.5
Conga	63.5
Pink Noise	69.1

3.2.6 Procedure

Ten-minute training sessions were carried out in advance to familiarise the subject with the interface and the HRTFs. It has been recommended that training sessions are to be carried out for non-individualised HRTFs, especially when the subjects have not had previous experience with the current HRTFs (Pernaux et al., 2002). Prior to the experiment, subjects were given specific instructions of the procedure. Subjects were required to slowly adjust the slider to match the perceived position of the pointer with the anchor. Once a threshold has been set by the slider, the (-/+) buttons are to be used to fine-tune the pointer position. The toggle button is to be used for switching between the pointer and anchor during adjustment and verification stages. Each subject was assigned to six 30-minute sessions, with 3 sessions involving ITD or ILD and each session differing in stimuli involving speech, cello and conga drums. A single session consists of 40 trials, containing 4 repetitions. Each repetition contains 10 different anchor angles: 0°, 10°, 20°, 30°, 40°, 50°, 60°, 70°, 80°, and 90°. Anchor angles were adjusted to either the left or right side of the head to prevent biasing of the data points to a single side of the head. In order to reduce contraction and anticipation bias, all sessions, trials and anchor angles were randomised.

3.3 Results

The data acquired for each target angle for the left and right hemisphere of the head is converted to positive values. Although this assumes symmetry of the head and ears, it allows for more data points for each target angle. It is also assumed that this will not affect localisation, as recent studies have suggested the insignificant effect of asymmetry in human localisation (Claes et al., 2015). Outlier removal was deemed as necessary as the data will have a significant impact on the predictive model within the study. Especially when it has been suspected that the model could be sensitive to extreme values associated with the large GUI slider limits, as apparent in some data points. Therefore, detection of outliers is carried out by observing data that lies beyond the 1.5 * Inter Quartile Range, below the first quartile and above the third quartile, as described below:

$$IQR = Q_3 - Q_1$$
$$High\ Outlier = Q_3 + 1.5IQR$$
$$Low\ Outlier = Q_1 - 1.5IQR$$

(3.4)

Instead of trimming the data, therefore reducing the amount of data points for the model, the data is capped to the first and third quartiles, known as Winsorisation (Dixon, 1960).

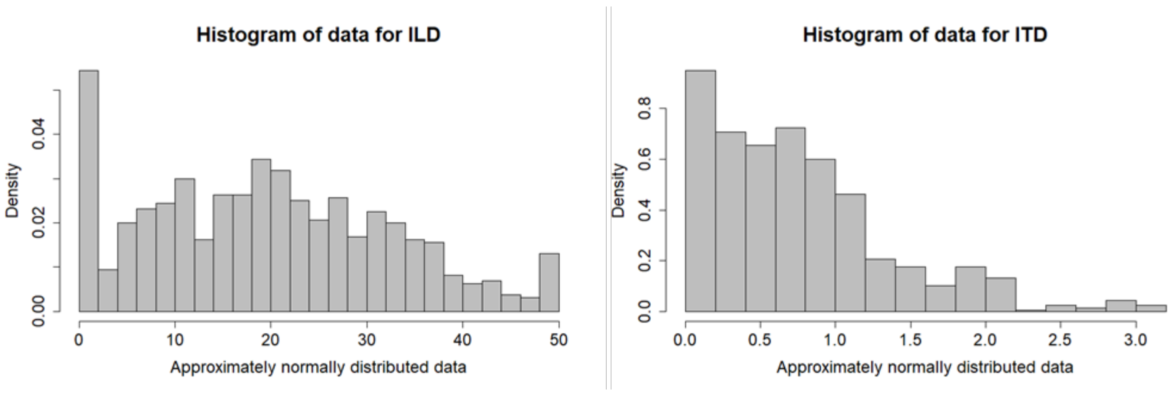


Figure 3.3: Histograms showing the range of the data across all azimuth angles for ITD in milliseconds and ILD in dB. For ILD the full range of the ILD pointer was used (0-50dB).

According to the Shapiro-Wilk Test for normality (Shapiro & Wilk, 1965), not all of the target angles had a normal distribution for the obtained ILD and ITD data. More ILD data points were normally distributed than ITD data points for target angles 10, 20, 50, 70 and 80. With the majority of the data having non-normal distribution, the data is safely represented with the median as the best measure of central tendency and must be analysed with non-parametric statistical methods.

Wilcoxon signed-rank tests were also performed on each target angle across each stimuli pair to determine the differences between each stimuli condition. Surprisingly, each stimuli pair showed no significant difference across all target angles for both subjective ILD and ITD ($p > 0.05$), despite the frequency and temporal differences of each stimuli. Furthermore, maximum values vary between 1.1ms and 1.4ms for ITD and 23.25dB to 34dB for ILD, with speech having the largest maximum values overall. In Figure 3.4, data values of the ITD pointer and the ILD pointer for each stimuli are presented as equivalent 95% confidence intervals for medians (Mcgill et al., 1978) given as:

$$\pm \frac{1.58IQR}{\sqrt{n}} \tag{3.5}$$

There is an apparent dispersion of the data between the conga drums and cello and speech, especially for ITD trials from 80°. On the other hand, similarities are shown between the cello and speech stimuli which have much larger error bars than the conga drums. This could be attributed to the continuous nature of the speech and the cello source as opposed to the conga drums which are dominated by transients. Fortunately, as indicated from the Wilcoxon tests, data from each stimuli condition can be merged to fulfil the overall data for the model due to no major significant differences.

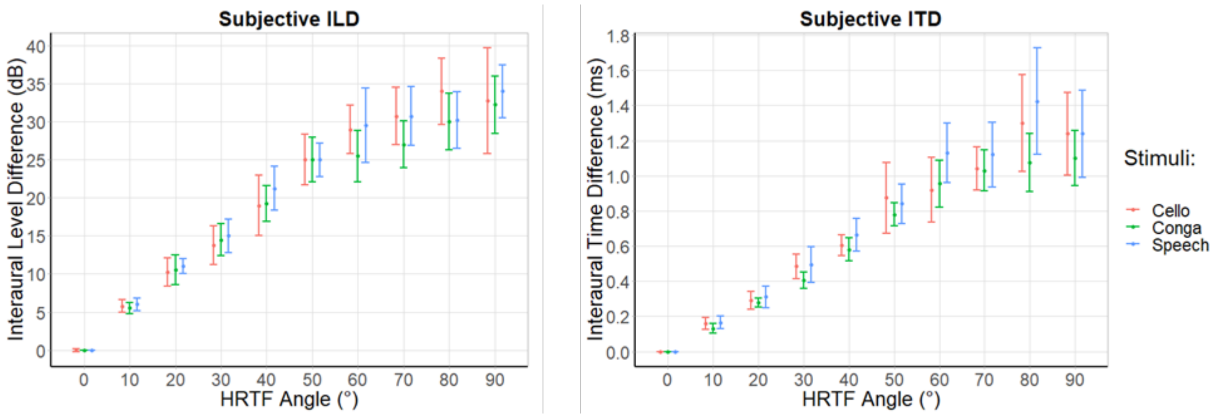


Figure 3.4: Medians and corresponding 95% confidence interval notch edges of each stimuli.

Because the stimuli were not statically different from each other, the overall medians and notch edges can be obtained and are shown in Table 3.2 and 3.3, and Figure 3.5 Overlapping of notches across target angles is apparent above 50° for subjective ILD and above 60° for subjective ITD. Below these angles, linearity is shown. Both plots display an increase of ILD and ITD at wider azimuth angles, with the error bar width increasing accordingly, even more so for subjective ITD. This coincides with the subjects reporting difficulty of matching the ITD pointer at wide target angles. Wilcoxon signed-rank tests are performed on each target angle pair, with the significance level of p having a cut-off value of 0.05. In sessions involving the ITD pointer, there is no significant difference between angle pairs above 50° ($p > 0.05$), except for pairs 70° and 80° with yet larger differences than pairs below 50°. 80° and 90° pairs showed the least significant differences compared to all pairs. Furthermore, effect sizes were calculated to establish how much each pair is correlated. Effect sizes decreased along with wider target angles, even drastically above 50° to approximately 0.05, labelled as the smallest effect size and therefore the least correlation between each paired angle (Cohen, 1988).

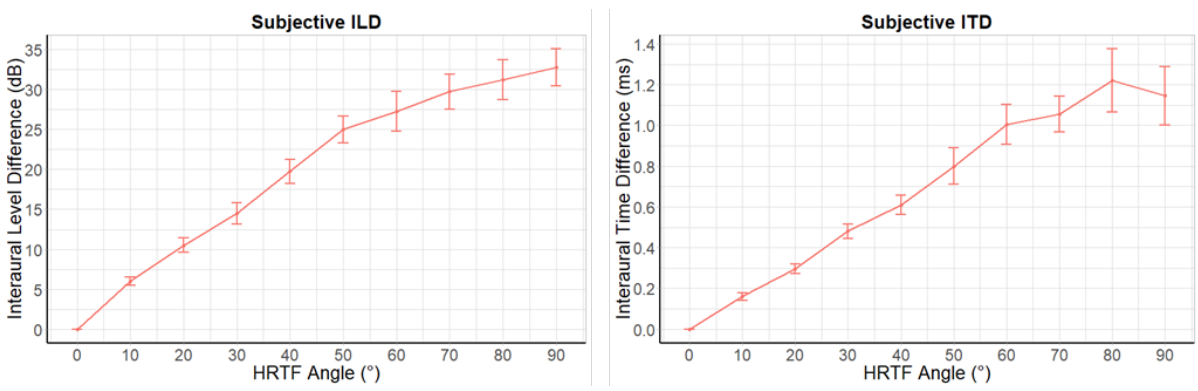


Figure 3.5: Medians and corresponding notch edges (i.e. non-parametric 95% confidence intervals) for all stimuli.

Table 3.2: Medians and upper and lower notch edges for each target azimuth for ILD (dB).

HRTF Angle (°)	Lower 95% CI	Median	Upper 95% CI
10	5.47	6	6.53
20	9.59	10.5	11.4
30	13.2	14.5	15.8
40	18.2	19.8	21.3
50	23.3	25	26.7
60	24.7	27.2	29.8
70	27.6	29.8	31.9
80	28.8	31.2	33.7
90	30.4	32.8	35.7

Table 3.3: Medians and upper and lower notch edges for each target azimuth for ITD (ms).

HRTF Angle (°)	Lower 95% CI	Median	Upper 95% CI
10	0.15	0.16	0.18
20	0.27	0.3	0.32
30	0.44	0.48	0.52
40	0.56	0.61	0.66
50	0.71	0.8	0.89
60	0.91	1.00	1.10
70	0.97	1.06	1.14
80	1.06	1.22	1.38
90	1.00	1.14	1.29

3.4 Discussion

In comparison to previous lateralisation experiments (Feddersen et al., 1957; Sayers, 1964), the results presented a linear trend relating to the increase of ILD and ITD values with increasing azimuth. However, the maximum values matched with a 90° azimuth is dissimilar to ILD for localisation. For ILD, approximately 20dB was suggested for 90°in (Durlach & Braida, 1969; Domnitz & Colburn, 1977; Howard & Angus, 2013; Feddersen et al., 1957) from ear canal measurements, whereas current results have shown a larger ILD of 32.8dB for 90°. This was also reported in the Fedderson et al. (1957) study where an ILD pointer was matched with measured ITD in a lateralisation task. It was found that pointer ILDs exceeded ear measurements of ILDs, with 60° having an ILD of approximately 25dB. This difference in ILDs across lateralisation and localisation could be due to the lack of head shadowing within the lateralisation task; therefore, it can be assumed that ILD values are much greater due to the lack of obstruction of the head. It is also important to note that studies which have obtained ILD through ear canal measurements are not isolated from pinnae and time difference cues, so ILDs are considerably smaller (Feddersen et al., 1957; J. Middlebrooks et al., 1989). This is especially the case for the KU100 HRTFs used as an anchor within this experiment where calculations have shown an ILD of approx. 22dB for 90°.

For independent ITD, 1.14ms is required for 90° as opposed to real ITDs at approximately 0.63ms, the maximum binaural delay (Jens Blauert, 1997a) or approximately 0.66ms according to the spherical head model. In reference to the duplex theory, these values are only valid for pure tones up to 1500Hz, where ITD is a function of frequency. But for complex stimuli used within this experiment, ITD values can also be attributed to high frequency envelopes (Trahiotis & Kappauf, 1978). For real ITDs of 0.66-0.68ms obtained from ear canal measurements, ITDs are expected to be smaller as they are not isolated from ILD (Feddersen et al., 1957; Woodworth, 1938).

Another possible explanation of larger ITD relates to the precedence effect, where ITDs above 1ms causes an increase of the perceived width of the acoustic pointer, potentially causing subjects to overestimate ILD or ITD when matching the HRTF position. According to Blauert (1997a) the increase of perceived image width can also be a result of each frequency component displaced laterally at different amounts. These explanations suggest that lateral ambiguity of the acoustic pointer could have significant impact on the resulting ILD and ITD values. In previous lateralisation studies, ambiguity can be dependent on frequency due to the use of pure tones. For example, ambiguity has been reported at 12dB with a 600Hz tone (Sayers, 1964). With increasing frequency, the threshold of ILD at which ambiguity occurs increases. By using a broadband stimuli, this reduces the limitation of arbitrary cues due to frequency dependency (Park, 2007).

In observing the current results, wider error bars are more prominent beyond 1ms at 60°. This was observed by subjects who reported difficulty of matching the ITD pointer to wider angles. According to Blauert (1997a), at 0.8ms to 1ms, lateral displacement is provided at a slower pace. It may be that beyond 1ms, the image affected by the ITD pointer did not increase. Instead, the image became broader and more unstable, reducing the focal point of the image. This can be observed for phantom images panned with time-delay between loudspeakers (Pulkki et al., 1999). In Lee and Rumsey's (2013) study, larger error bars where found for interchannel time difference (ICTD) at wider angles in a 60° loudspeaker configuration. This uses a similar acoustic pointer technique to the one used in this study but with an ICTD and ICTD pointer matched with marker positions between loudspeakers.

Another explanation of the inconsistency of interaural values above 50-60°, is relevant to real-source localisation experiments, where localisation errors are known to increase with increasing azimuth (Makous & Middlebrooks, 1990; Carlile et al., 1997; Jens Blauert, 1997a). In this case, localisation performance could have been reduced at wider HRTF angles, increasing the difficulty of matching lateral positions affected by the acoustic pointer to HRTF positions. This also supported by virtual localisation tests where HRTF angles were either perceived wider (Pernaux et al., 2002) or HRTF angles over 60° were perceived to be near 90° (Park, 2007). Another explanation of angular distortion is related to the minimum audible angle (MAA), the just noticeable difference (JND) of direction. According to Mills (1958), the MAA tends to become larger with increasing azimuth between a loudspeaker pair towards the side of the listener.

3.4.1 Experimental Limitations

One of the main implications of the experiment is the unnaturalness of the task. Closed loop tasks, such as those involving MOA with a centring or pointing method, do not replicate the cognitions that involve localising sources in free-field environments and do not rely on spatial memory. According to Stecker (2010) 'sensory trace' processing is performed while the subject attempts to maintain an image without previous real-world context, i.e. comparing a target stimulus with reference stimuli (Durlach & Braida, 1969; Stecker, 2010). This explains the difficulty of the task reported by the researcher and subjects. In addition, along with 40 trials per session, fatigue is likely to be induced, affecting the validity of the subject's responses.

Another limitation is that subjects may have felt inclined to use the entirety of the GUI slider, despite being instructed to gradually increase the pointer position until it matched the HRTF position. The larger interaural values could possibly be due to the available ranges of the slider, up to 50dB for ILD and 4ms for ITD, as indicated in Figure 3.3. Even though outlier treatment has been considered, this could have affected the central tendency of the values.

3.5 Summary

The experiment within this chapter used an acoustic pointer technique like previous lateralisation studies to obtain independent ILD and ITD. But this is matched with a HRIR anchor representing azimuth positions rather than a visual lateral scale or a fixed interaural value. It was found that larger ITD and ILD values were found compared to real ITD and ILD obtained in localisation. For ITD, this is attributed to the instability of the image at wider azimuths above 0.8ms, and for ILD, this is attributed to the lack of head shadowing in the lateralisation task. Overall, the difference between results obtained from lateralisation and localisation is that interaural values obtained in terms of localisation are not isolated from each other. Furthermore, localisation ambiguity is shown by wider error bars at wider azimuths. This is attributed to the decrease in localisation performance common across many localisation studies and the increase of the minimum audible angle (MAA) at wider angles.

4. An Interaural Trade-off Model for Auditory Image Position

This chapter describes the development of a trade-off model based on interaural values obtained from the binaural matching task within Chapter 3. Using the model, any combination of ILD and ITD can be generated to create an auditory image in reference to azimuth. In Section 4.2 and 4.3, the model is verified objectively with comparisons of HRTFs and subjectively through a virtual localisation task.

4.1 Interaural Trading Function

In previous lateralisation studies, trading ratios are derived from finding the equivalent level differences to certain time differences, as discussed in Section 2.1.2. The fixed trading ratio, however, does not allow for flexible combination of these parameters for a target image position, and it does not describe the effect of each ITD and ILD parameter on the resulting image position . This is also apparent in localisation studies where fixed weighting of ITD and ILD are derived from HRTF (Macpherson & Middlebrooks, 2002; Park, 2007; Takanen & Lorho, 2012). One method for deriving any combination of ILD and ITD is related to the linear interchannel trade-off function proposed by Lee and Rumsey (2013), used to predict the localisation of phantom image sources in 2-0 stereophonic reproduction. This was based on Theile's (2002) hypothesis, which suggests that the resulting phantom image position between two loudspeakers is approximately the sum of each image position θ determined by individual time and level differences:

$$\theta\ (\Delta L, \Delta t) = \theta\ (\Delta L) + \ \theta\ (\Delta t) \tag{4.1}$$

where ΔL is the level difference, and Δt is the time difference. Implementation of this method will simplify the calculation of ITD and ILD for certain azimuth positions, assuming their relationship is linear. It will also be considered more adaptable and easy to generalise than that of some binaural prediction models where weighting of ILD and

ITD is limited to the HRTFs measured (Macpherson & Middlebrooks, 2002, Parks 2007), for example.

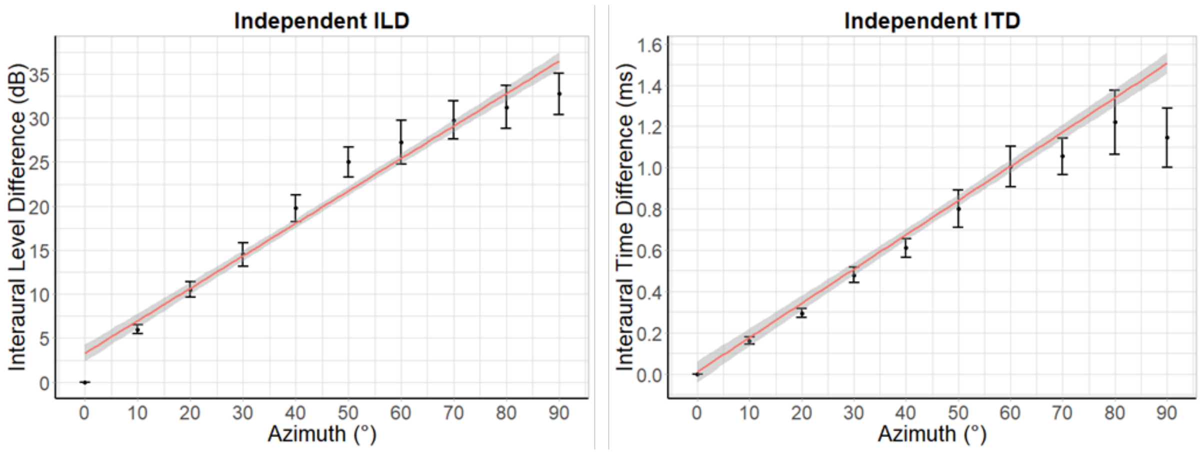


Figure 4.1: Linear regression lines indicating linearity of data for between 95% confidence intervals for ILD and ITD.

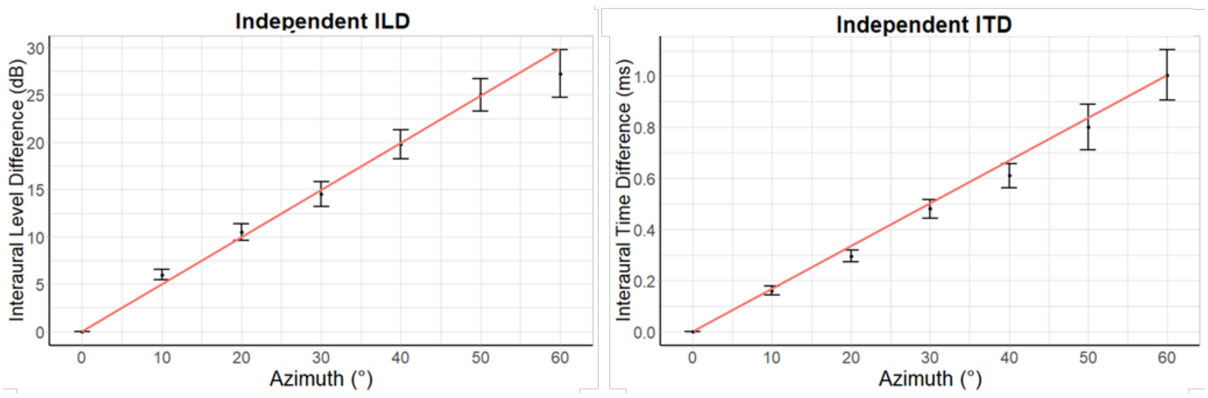


Figure 4.2: Linear regression lines for ILD and ITD arbitrarily fit between 95% confidence intervals for azimuths up to 60°.

Using unified data from all broadband stimuli obtained from the first experiment within Chapter 3, an interaural trading function will be developed to predict auditory images. To support Theile's (2002) hypothesis, interaural values along with their corresponding azimuths must be linearly correlated. This can be verified through a linear regression analysis to yield linear fit of the experimental data. Figure 4.1 plots the medians and their equivalent 95% confidence intervals with the linear regression line describing the linearity of the data. The plots show a line of best fit for ITD and ILD (R^2 = 0.99, p < 0.001) which lies between the notch edges up to 90°. Due to the results statistically presenting no significant differences between 60° and 90° for both ITD and ILD. The linear regression is arbitrarily fit within the 95% confidence intervals up to 60° so that the starting point is at 0ms and 0dB, as shown in Figure 4.2. Therefore, the maximum interaural values will be 1ms and 30dB for ITD and ILD respectively.

Interestingly, the positive slope of the arbitrary regression lines is very similar to the JND of the broadband stimuli, with ILD indicating a slope value of 0.5 and ITD indicating a slope value of 0.017. These values are the calculated shift factors for the region between 0° and 60°, described as dB/deg or ms/deg respectively. By confirming linearity of the data, the trading function can be drawn via the linear function given as:

$$Y = b0 + b1X$$

(4.2)

where $b0$ is the intercept or starting value of Y when $X = 0$, and $b1$ is the slope associated with the increase of X. Here it can be used to derive the equivalent ITD from ILD, as visualized in Figure 4.3:

$$ITD = 1ms + \frac{1ms}{30dB}\, ILD$$

(4.3)

ILD and ITD values can also be used to obtain the angular shift:

$$\theta = \frac{\Delta t}{0.017} + 2\Delta l$$

(4.4)

A trading ratio can also be computed from the linear regression by dividing the slope or shift factors from ITD and ILD (Stecker, 2010):

$$TR = \frac{bt}{bl}$$

(4.5)

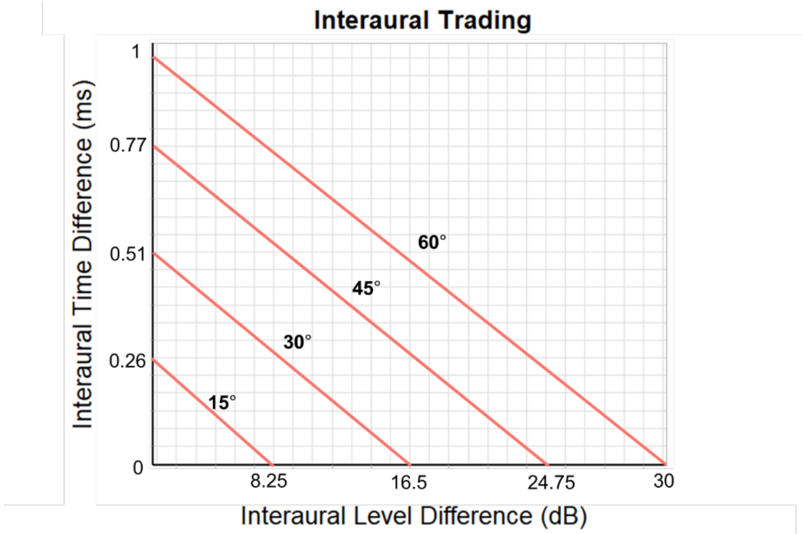


Figure 4.3: Proposed ILD and ITD trade-off graph with linear trading curves for 15°, 30°, 45° and 60° images.

The resulting trading ratio of 0.34ms/dB, or 34us/dB as a measurement in earlier lateralisation studies, is calculated. This is very much like the TR of 0.36ms/dB derived from the Lindemann extension of the cross-correlation model (Lindemann, 1986). In addition to 20-50us/db derived from 500Hz pure tones (Hafter & Jeffress, 1968), and nearer to 25us/dB derived from low-pass clicks below 1500Hz (Harris, 1960). As opposed to 80-100us/dB for 500Hz clicks (Whitworth & Jeffress, 1961) and 60us/dB for high-pass clicks above 1500Hz (Harris, 1960). In summary, the resulting TR from experimental data is comparable with TR for continuous stimuli and low frequency dominant stimuli from trading studies. This is applicable for the continuous broadband stimuli used in the first experiment; however, they contain a wider range of frequency components than those used in past studies. Studies using low frequency stimuli may have influenced the TR by a time image deemed most applicable for frequencies below 1500Hz, therefore the TR could be affected more by ITD (Whitworth & Jeffress, 1961). With broadband stimuli, the TR could be affected by both a time image dominated by ITD and an intensity image dominated by ILD and ITD. In addition, the TR between the first experiment within the study and previous lateralisation studies could be differentiated by the methodology. In early trading experiments, perception of the two auditory images was prone when interaural cues were placed in opposition to each other (Hafter & Jeffress, 1968) or an interaural cue is matched with an unnatural combination of interaural cues set by the researcher (Whitworth & Jeffress, 1961). However, in the current first experiment, individual interaural cues are matched with a natural combination of ITD and ILD imposed by the HRTF, therefore the TR is not biased to a time or intensity image. According to Gaik (1993), unnatural combinations lead to incomplete fusion of events, creating a split of auditory events which could explain why a separate time and intensity image is perceived in previous trading experiments.

4.2 Comparison with HRTF Parameters

A way to objectively verify the linear trade-off model is by comparing with objective measurements of HRIRs containing natural interaural parameters associated with azimuth. The HRIRs used is the same used for the HRIR anchor within the first experiment, from a KU100 Dummy Head. Firstly, ITD and ILD are extracted via the Interaural Cross Correlation Coefficient (IACC) and the summed energy differences between the left and right channels using MATLAB. According to the Jeffress Cross Correlation Model on Cross Correlation, as discussed in Chapter 1, the IACC is derived to calculate ITD between the left and right ear signals (Jeffress, 1948) or left and right signals of an HRIR in this case. It is described as the maximum of the Interaural Cross Correlation Function (IACF) given by

$$IACF(\tau) = \frac{\int_{t_1}^{t_2} h_L(t)\, h_R(t+\tau)dt}{\sqrt{\int_{t_1}^{t_2} h_L^2\,(t)dt\ \int_{t_1}^{t_2} h_R^2\,(t)dt}}$$

$$IACC = |IACF|_{max} \tag{4.6}$$

where h_L and h_R are the left and right signals of the HRIR. The time-lag m at which the IACC is computed is used to calculate the ITD associated with the sampling frequency of 44.1kHz and the maximum lag of 44.

$$\Delta t = \frac{m-44}{f_s} \tag{4.7}$$

Δl derived from the left and right channels of the HRIR is given by

$$\Delta l \; = 10 \cdot \; \log_{10} \frac{\sum_{k=1}^{n} h_L(k) \cdot h_L(k)}{\sum_{k=1}^{n} h_R(k) \cdot h_R(k)}$$

(4.8)

where *n* is the length of the HRIR in samples. Figure 4.4 shows the extracted ILD and ITD for the following azimuth angles; 0°, 15°, 30°, 45°, 60°, 75°, 90°.

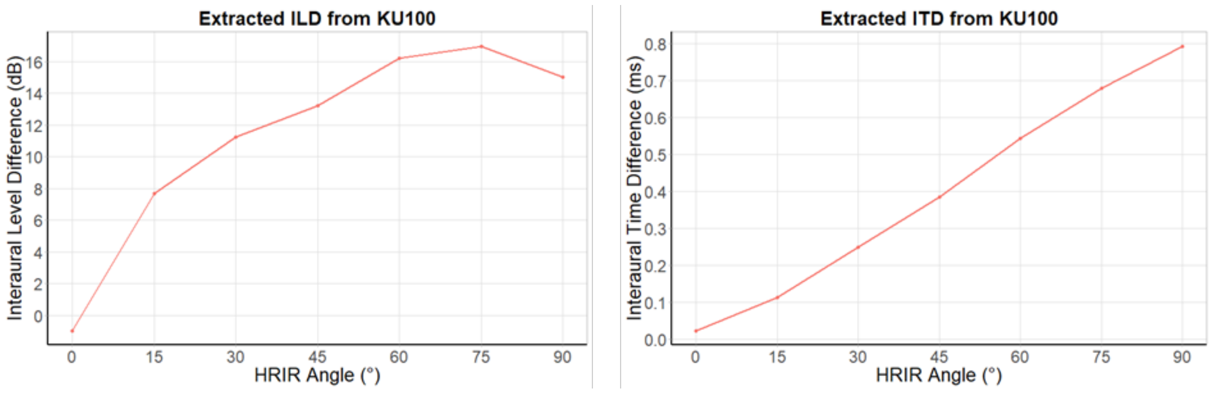


Figure 4.4: ILD and ITD extracted from HRIRs of the KU100 Dummy Head for the corresponding azimuths; 0°, 15°, 30°, 45°, 60°, 75°, 90°.

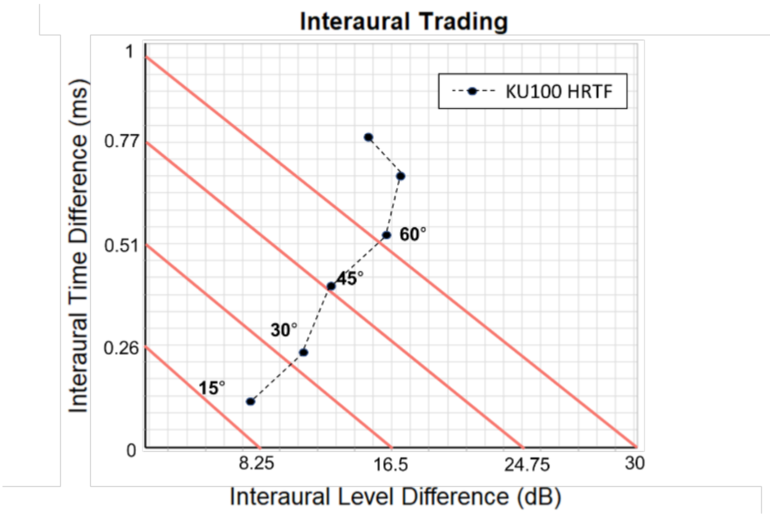


Figure 4.5: Combination graph with extracted ITD and ILD data points for HRTF angles; 15°, 30°, 45°, 60°, 75°, 90°.

Within Figure 4.5, the trade-off model is compared to extracted ILD and ITD values. It is shown that HRTF ILD and ITD is closest to the linear trade-off at 45°. For the remaining HRTF angles, variably more ILD and ITD is determined than what is calculated from the model for each target angle. This could indicate that the trading between interaural parameters within HRTF is non-linear for each azimuth, with an increased weighting of ITD at wider angles. This non-linearity can also be verified by inputting HRTF ILD and ITD into the auditory image prediction model, using Equation 4.4.

Table 4.1: Comparison of HRTF angle with predict azimuth using HRTF ILD and ITD.

HRIR Angle (°)	Predicted Azimuth (°)
15	22
30	37
45	49
60	64
75	74
90	77

As depicted in Table 4.1, disparity between model predictions and the HRIR angles ranges between 1°- 13°. For the HRIR angle of 75°, only a difference of 1° shown. The trend of errors tends to be associated with the curve of extracted ILD in Figure 4.4, as opposed to extracted ITD where a linear trend is shown. The linearity of ITD is comparable to that of independent ITD gained from the first experiment within Chapter 3. For ILD, this is not the case. This could be due to the lack of acoustic head shadowing and crosstalk between the ears during the first experiment due to the use of headphones, reducing the amount of ILD required for a lateral shift. In terms of HRIR measurement and localisation, the head obstructs the radiating signals from a source, increasing the difference of level between the ears. This is significant in early studies and literature such as Lord Rayleigh (1907) and Blauert (1997a), where it was suggested that the binaural ratio of sound level could be dependent on the diffraction of sound around the head.

4.3 Experimental Evaluation

Once the trade-off model has been developed to derive ITD and ILD for a specified azimuth, the next task is to verify the model for multiple combinations in terms of localisation accuracy. This will also verify the effectiveness of summing angular shifts from imposed ILD and ITD (Wittek & Theile, 2002). If larger localisation errors occur throughout each combination for each target angle, it can be assumed that a linear function is ineffective for localisation. Based on the first experiment for obtaining ILD and ITD, it has been hypothesised that a time dominated image is prone to ambiguity due to increased image width. By imposing ILD, this is said to resolve these ambiguities (Sayers, 1964; Raspaud et al., 2010). This can be established by increasing ILD along with ITD until a threshold is met that improves azimuth estimation.

4.3.1 Methodology

In order to record the perceived position of a given combination of ILD and ITD, the HRIR pointer method by Parks (2007) was used. As this is a reversal of the previous MOA method, where an ITD or ILD acoustic pointer is matched to the perceived position of a HRIR anchor, this inhibits other potential biases, whereas another imposed method could induce additional biases and discrepancies. For this localisation task, the HRIR pointer is expected to be perceived naturally while maintaining an intercranial position which is effective for indicating the position of lateral images. Using the combination function in Section 4.1, the following combination ratios with the respective interaural values are shown in Table 4.2. A 100% ILD or ITD trial will be used to verify the ILD and ITD values obtained in the first experiment to help establish discrepancies from the opposing test methods. It was also decided to provide combinations above 90° azimuth, using the trading model which is limited to angles up to 60°, to see if the linearity of the model can also account up to the full azimuth position at the ears, 90°. Even though statistically there was no significant difference between 60°-90, medians were still higher and increasing at a slower rate. The wider errors bars shown in Figure 3.5 could indicate room for a larger value.

Table 4.2: Calculated ILD and ITD values from the trade-off model for each combination, in relation to the following target angles; 15°, 30°, 45°, 60°, 75°, 90°.

	100% ITD	100% ITD
15°	7.5dB	0.23ms
30°	15dB	0.51ms
45°	22.5dB	0.77ms
60°	30dB	1.02ms
75°	37.5dB	1.28ms
90°	45dB	1.53ms

	75% ILD + 25% ITD	50% ILD + 50% ITD	25% ILD + 75% ITD
15°	5.63dB + 0.06ms	3.75dB + 0.13ms	1.88dB + 0.19ms
30°	11.25dB + 0.13ms	7.5dB + 0.26ms	3.75dB + 0.38ms
45°	16.86dB + 0.19ms	11.25dB + 0.38ms	5.63dB + 0.57ms
60°	22.5dB + 0.26ms	15dB + 0.51ms	7.5dB + 0.77ms
75°	28.13dB + 0.32ms	18.75dB + 0.64ms	9.38dB + 0.96ms
90°	33.75dB + 0.38ms	22.5dB + 0.77ms	11.25dB + 1.15ms

4.3.2 Stimuli

For this experiment, different broadband stimuli were used for the lateral images which were laterally adjusted with a set combination ITD and ILD using sample delay and constant power panning, like in Chapter 3. Different stimuli are used to ensure that the model can be generalised other broadband stimuli with various temporal and frequency components. The previous experiment showed no significant difference of results between broadband stimuli; therefore, generalisability is deemed to be feasible. A speech source is still used to prevent stimuli from being completely different from the first test to allow for comparison. The sample is an anechoic recording of a German speaking male taken from Sound Quality Assessment Material (SQAM) by EBU (2008). Another sound source is an excerpt of Capricco Arabe on the acoustic guitar taken from the Bang and

Olufson (B&O) 'Music for Archimedes' project (Hansen & Munch, 1991). This anechoic recording was chosen based on its large dynamic range, use of low and high notes, as well as long and short note durations in order to account for many types of broadband stimuli. For the HRIR pointer, the same pink noise bursts used for HRIR anchor within the previous experiment are used. These are also convolved with the HRIRs of the KU100.

4.3.3 Subjects

Six subjects, including two females and four males who are members of the Applied Psychoacoustics Lab (APL) at the University of Huddersfield, participated in this experiment. Five of the subjects have already participated in the previous experiment, therefore they have had prior experience in localising HRIR, using the response method. The remaining subject had less critical listening experience and has not participated in the previous experiment but has engaged in the training sessions, however, their data can help represent the naïve population.

4.3.4 Physical Setup

The setup for this experiment, conducted in an ITU-R BS116 (2015) compliant critical listening room at the University of Huddersfield, is very similar to that of the first experiment. Loudspeakers were hidden by curtains at the front of the listener to reduce the likelihood of visual bias. A Merging+Anubis audio interface running at a sample rate of 44.1kHz, the same for HRIR rendering within MAX 7, was used to provide audio playback through AKG 702 headphones. Playback levels of the speech sample were adjusted to 68dB LA_{eq} and the guitar and noise stimuli were matched accordingly, with the guitar and pink noise having a playback level of 63dBA and 68dBA, respectively.

4.3.5 Procedure

Prior to the localisation task, participants were familiarised with the user interface within MAX 7, which is the same as the previous experiment in terms of the response method, labelling and shortcuts, see Figure 4.6. They were also debriefed on the response task which was to gradually position the HRIR pointer to match the position of the stimuli. This slider had a range of +/-90° with an angular resolution of 1°. Subjects were also instructed to listen to the laterally displaced stimuli with imposed combination of ILD and ITD for extended periods of time. It was found from pilot testing and observations of the first experiment that listening to the stimuli longer than 5 seconds enabled easier establishment of the perceived position which prolonged as soon as the HRIR pointer has been toggle to the report the position of the image. This is related to investigations using noise band signals, whereas with the increase of duration of the stimuli, ongoing cues became more effective for localisation (Tobias & Zerlin, 1959; Perrott & Baars, 1974).

All subjects participated in 4 sessions, 2 for each stimulus. Each session lasted approximately 45 minutes with 60 trials containing 4 repetitions. After 30 trials, subjects were instructed to take a 10-15-minute break to minimise the effects of fatigue. All trials were randomised to prevent contraction bias.

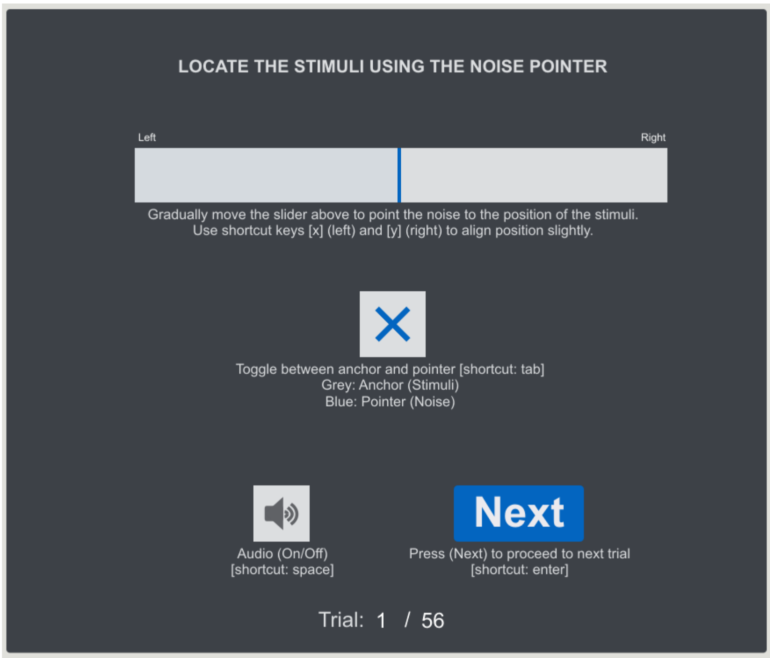


Figure 4.6: : Screenshot of the test interface within Max 7 for the localisation task to verify the trade-off model A slider is provided for adjusting the position of the HRIR acoustic pointer.

4.3.6 Results

Experimental data containing the perceived HRIR angles for each combination ratio and target angle was tested for normality. Using a Shapiro-Wilk test, it was found that most of the data was significantly different to normal distribution ($p < 0.05$), therefore non-parametric statistical methods can be performed, and the data will be best represented with medians. The medians and their equivalent 95% confidence intervals are calculated for the speech source and the guitar source, as shown in Figure 4.7. Throughout each combination ratio up to 60°, the perceived angles of speech were consistently more underestimated compared to the guitar condition which was more adjacent to the target angles. This underestimation becomes larger as the ratio of ITD to ILD increases. Differences between the speech and guitar were statistically identified from Wilcoxon signed-rank tests. For many of the target angles for each combination ratio, no significant differences were found ($p < 0.05$). The most disparity was found for the 50/50 combination ratio between 15-45°, and the response angle of 15° for ratios with more than 50% ITD.

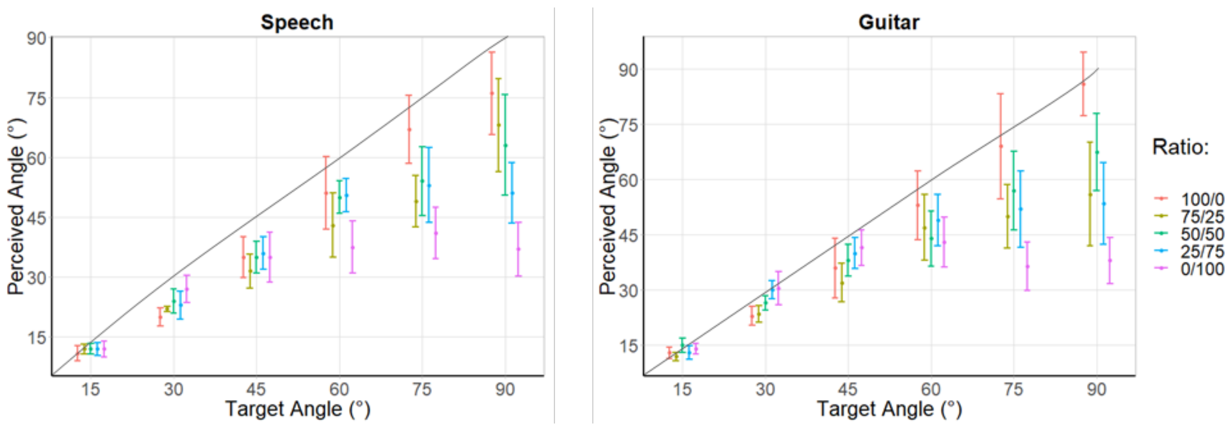


Figure 4.7: Subject responses depicted as medians and their equivalent 95% confidence intervals for speech and guitar throughout each combination ratio. Ratio labelling is described as ILD/ITD.

As the stimuli were deemed as statistically alike, the overall medians and 95% confidence intervals are plotted for each combination ratio, see Figure 4.8. Plots show a more positive trend with larger amounts of ILD. As the ratio of ITD to ILD increases, the response angle decreases accordingly. At 100% ITD, the response angle is maintained between 37-40° beyond the 45° target angle. Across all ratios, underestimation is shown, the amount of which increases with wider target angles, except for 100% ILD which showed less underestimation of up to 10° throughout each target angle up to 90°. Ratios with 75% and 100% ITD had responses closer to the target angle up to 30°, however, these are dramatically underestimated from 45°. Furthermore, significant differences for each combination ratio are shown between angles up to 60°, the limit of the ILD and ITD shift factors, beyond of which overlapping of the notches occur up to 90°.

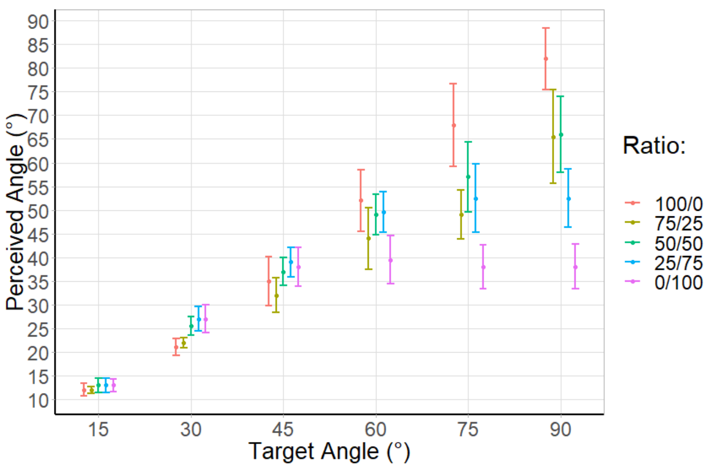


Figure 4.8: Overall subject responses depicted as medians and their equivalent 95% confidence intervals for each combination ratio. Ratio labelling is described as ILD/ITD.

Verifications results obtained for 100% ILD and ITD ratios are compared with model predictions, as shown in Figure 4.9. For ILD, responses are parallel with the linear function but is mainly underestimated by approximately 10°. This is contrasted with ITD where responses are closer to model predictions up to a cut-off point of approximately 0.75ms at 45° from which the angle is reluctant to change significantly.

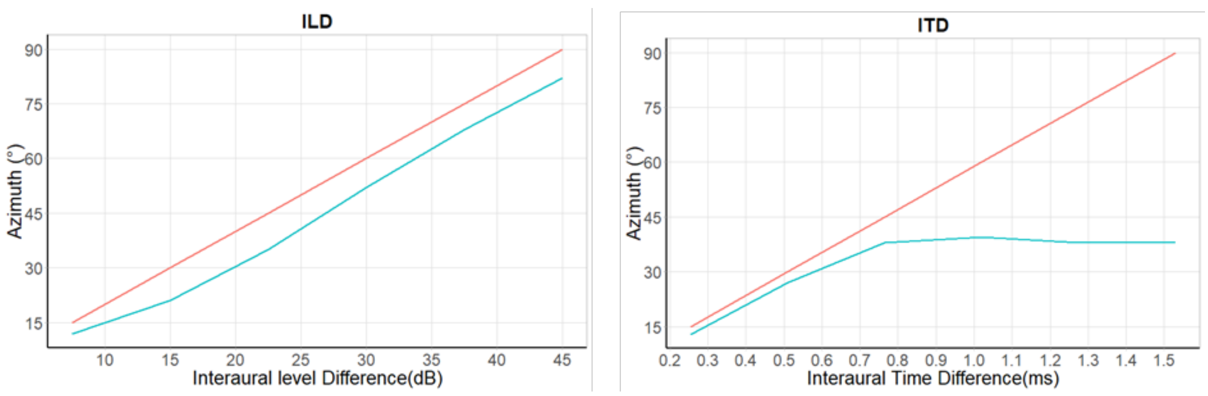


Figure 4.9: Model predictions (orange) vs experimental data from verification tests (blue) for independent ILD and ITD.

4.3.7 Discussion

During the analysis of results compared to model data, there is a significant amount of underestimation of the perceived angle throughout each combination ratio, especially at wider target angles. This could be due to systematic biases involving the response method chosen for derivation of interaural values and verification. The first experiment where a lateralised ILD or ITD pointer is matched with a HRIR anchor was deemed as unnatural and difficult compared to the second experiment where the HRIR pointer was naturally perceived and involved localising the lateral ITD-ILD image. Parks (2007), who also used the HRIR pointer method, reported difficulty not solely because of the HRIR pointer but the ambiguity of the internalised image locations. This was reported in the verification experiment where images displaced with ILD seemed closer to the head compared to images displaced with ITD which were perceived more externalised. One explanation is the increase spaciousness and increased apparent source width (ASW) induced by fluctuations of ITD in lateralisation. This is suggested by Blauert and Lindemann (1986), where a positive correlation was shown in subjective tests along with spaciousness. According to Barron and Marshall (1981), ASW is increased by reflections within 80ms. This can be apparent in broadband stimuli used where individual frequencies components may be displaced at different amounts, creating varying time differences (Jens Blauert, 1997a). In the context of the verification experiment, the guitar source could have been perceived wider as depicted by wider 95% confidence intervals, shown in Figure 4.7, which could be as a result of fluctuations of ITD in various note changes.

The observed externalisation of ITD could indicate why response angles for 100% ITD and 75% ITD was closer to the target angles up to 45°. It could be that they were easier to match with the HRIR pointer which was perceived as more external than ratios with more ILD. According to Armstrong's (2018) study on the perceptual evaluation of the

SADIE HRTF database, the extent of externalisation can vary across subjects, especially for the KU100 used in the current experiments where varied subjects' ratings were shown for externalisation. This could explain the variation of response angles within the error bars, assuming this is affected by localisation ambiguity between the HRIR pointer and the lateral anchor.

From another viewpoint, the difficulty of the localisation task and ambiguity of images could be as a result of unnatural combinations of ILD and ITD. In Gaik's (1993) study, psychoacoustic experiments show that the human auditory system can recognise whether ITD and ILD match. Unmatching combinations may cause incomplete fusion of auditory images or duals images, as reported in previous trading experiments, therefore ambiguity may occur (Raspaud et al., 2010). Within the current results, the unnaturalness of combinations can be observed with increasing combination of ITD at wider angles, especially above 45° where substantial underestimation is found. This is associated with the ASW and instability of images with increasing ITD.

As previously discussed, 100% ITD and 75% ITD were more accurate compared to other combination ratios up to target angle of 45°. This is also evident within the Macpherson and Middlebrooks (2002) study where weighting of ITD was larger or equivalent to ILD for wideband stimuli. This emphasises the importance of ITD in localisation as stated in early previous studies (Lord Rayleigh, 1907; Leakey, 1959) and the suggestion of ILD being a weak cue for wideband stimuli (Wightman & Kistler, 1997). However, ILD is still important for resolving ambiguities induced by ITD (Sayers, 1964). This is contrasted with current data above 45°, where response angles were closer to target angles with more ILD to ITD. Above 45°, this is where 100% ITD hits a cut-off point at which hardly any change in the target angle is shown, this is also apparent with Blauert (Blauert, 1997) where linearity shown up to 0.63ms, but between 0.8ms and 1ms, lateral displacement is reduced at a slower pace. The limit of approximately 0.75ms at 45° is lower than the full lateral displacement from model predictions but still larger than real ITD in localisation with a value of 0.68ms (Feddersen et al., 1957) or 0.66ms from the spherical head model (Woodworth, 1938). However, these values are naturally smaller because of the influence of head size.

One limitation of the response method is the localisation accuracy of the HRIRs themselves that may have cause underestimation of response angles and wrong predictions from the trade-off model. In a study by Pernaux et al. (2002), left and right judgements of HRTFs were perceived much wider with average angles errors of 16°- 42°. Errors up to 12.2° for trained subjects and up to 20° towards peripheral locations were also found from Majdak and Goupell (2010) and Makous and Middlebrooks (1990) respectively. Park (2007) also found that for angles above 60°; subjects reported positions of 90° which has caused uncertainty of responses. If subjects within this study have perceived HRIR position to be wider in a similar way to the studies stated, this would have caused a bias in independent ILD and ITD where more ILD and ITD is required to match the wider HRIR angles, affecting the resulting trading-model. However, if the HRIR pointer is perceived wider, this would cause an underestimation of the response angles, therefore it would balance out the bias. Yet, underestimation persisted in the current results. As stated before, broadband stimuli are perceived to have a larger ASW which may have caused the localisation errors.

4.3.7.1 Experimental Limitations

One of the main limitations in terms of the experiment is the potential of fatigue induced by a large amount of trials, with each session lasting 45mins. Even though, a 10-15min break was recommended halfway through the session, higher concentration may have caused more fatigue when attempting to accurately matching the pointer with the anchor. Larger localisation errors may have been prominent during the last trials of the sessions where subjects may have rushed their responses to finish the test.

Even through broadband stimuli was used for more generalisability of the model for various real-world and musical sources, generalisability is also affected by using subjects with experience in localisation tasks and critical listening. Therefore, it is unsure whether the trade-off model is also valid for naïve subjects who are most common to larger localisation errors

4.4 Summary

A trade-off-model based on independent interaural parameters was developed to allow linear weighting for the desired target azimuth. This assumes that calculated angles from each weighted parameter can be summed to create the resulting auditory image, based on Thiele's (2002) hypothesis. In comparison to HRTF data, model predictions are larger than the resulting HRTF angles. This is due to the non-linearity of extracted ILD from HRTF as a result of the 'head shadow' effect that was not present in the first experiment for obtaining independent ILD and ITD.

The model predictions were also verified through a virtual localisation experiment involving a reversed method of the first experiment where responses were based on HRIR pointer positions localising laterally displaced images rather than ITD and ILD pointer position matched with a HRIR anchor. It was found that for broadband stimuli, combinations ratios with larger ITD to ILD showed closer responses to the target angle up to 45°. Throughout each target angle, underestimation was significant. This was due to systematic biases and ambiguities concerning the response method when attempting to match a virtually localised source with a lateralised image. This was also greatly affected by the perceived width of the source, and the HRIR being perceived at wider angles.

5. Conclusion

In the current study, a simple linear trade-off model has been proposed to allow trading of primary binaural cues involving interaural time and level differences (ILD and ITD) for azimuth positions to up to 60°, and to predict auditory images from input ILD and ITD. A novel method was used to obtain subjective independent ILD and ITD for the model using HRIRs to define azimuth direction in headphones. It was found that that interaural values are larger than real values in localisation.

When comparing the model to HRTF data, it was found that HRTF combinations consisted of larger ILD and ITD compared to linear combinations within the trade-off model. In addition, the model was ineffective at predicting images from extracted interaural values that are induced by the effects of the shape of the head, including as the 'head shadow' effect.

A second psychoacoustic experiment in Chapter 4 was conducted to verify the model. This involved a reversal of the proposed novel method where azimuth responses were reported with a HRIR pointer. It was found that selected combination ratios calculated from the trade-off model showed underestimation of the target angle, even more so for ratios with larger weighting of ILD across all target angles and ratios with larger weighting of ITD above 45°. This discrepancy is attributed to the systematic bias and difficulty of matching internal lateralised images to virtual localised images imposed by HRIRs. Therefore, dissimilarities between lateralisation and localisation are observed. Furthermore, underestimation is also attributed to localisation ambiguities imposing from HRIRs being perceived wider, and the increase of the minimum audible angle (MAA) and apparent source width (ASW) at wider angles. This is especially the case for ITD where after approx. 0.75ms at 45°, the images become more unstable and difficult to localise due to increased ASW and MAA.

5.1 Future Work

To improve the model so it is more effective at calculating ILD and ITD values for a certain azimuth position, a new shift factor is required. A way of deriving the shift factor is by using data from the verification experiment which is seen as more valid due to the naturalness of the task where subjects were required to localise images rather than to match them. However, this still has biases between lateralisation and localisation, as previously discussed. Instead, another way of deriving independent ILD and ITD without these biases is through transaural synthesis (W. M. Hartmann et al., 2016). This involving synthesizing loudspeaker signals so crosstalk is eliminated, and binaural cues can be mimicked externally. That way, stimuli can be presented in a free field environment and cues at the ears can be arbitrary controlled, these can therefore be obtained from real-world localisation than lateralisation.

Data from the model and current experiments can be used to understand localisation in terms of independent ILD and ITD. This is beneficial for developing microphone techniques and panning methods were perceived images can be customised. For example, the trade-off model can be used to incorporate more ITD for spaciousness or more ILD for directional and localisable auditory images.

Bibliography

Algazi, V. R., Avendano, C., & Duda, R. O. (2001). Elevation localization and head-related transfer function analysis at low frequencies. *The Journal of the Acoustical Society of America*, *109*(3), 1110–1122. https://doi.org/10.1121/1.1349185

Armstrong, C., Thresh, L., Murphy, D., & Kearney, G. (2018). A Perceptual Evaluation of Individual and Non-Individual HRTFs: A Case Study of the SADIE II Database. *Applied Sciences*, *8*(11), 2029. https://doi.org/10.3390/app8112029

Banister, H. (1926). Three Experiments on the Localization of Tones1. *British Journal of Psychology. General Section*, *16*(4), 265–292. https://doi.org/10.1111/j.2044-8295.1926.tb00390.x

Barron, M., & Marshall, A. H. (1981). Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure. *Journal of Sound and Vibration*, *77*(2), 211–232. https://doi.org/10.1016/S0022-460X(81)80020-X

Bech, S., & Zacharov, N. (2006). *Perceptual Audio Evaluation—Theory, Method and Application*. https://doi.org/10.1002/9780470869253.ch1

Begault, D., & Wenzel, E. (2001). *Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source*. 904–916.

Bernstein, L. R., & Trahiotis, C. (1985). Lateralization of low-frequency, complex waveforms: The use of envelope-based temporal disparities. *The Journal of the Acoustical Society of America*, *77*(5), 1868–1880. https://doi.org/10.1121/1.391938

Blauert, J., & Lindemann, W. (1986). Auditory spaciousness: Some further psychoacoustic analyses. *The Journal of the Acoustical Society of America*, *80*(2), 533–542. https://doi.org/10.1121/1.394048

Blauert, Jens. (1997a). *Spatial Hearing: The Psychophysics of Human Sound Localisation* (Revised Edition). The MIT Press.

Blauert, Jens. (1997b). *Spatial Hearing: The Psychophysics of Human Sound Localisation* (Revised Edition). The MIT Press.

Breebaart, J., van de Par, S., & Kohlrausch, A. (2001). Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters. *The Journal of the Acoustical Society of America*, *110*(2), 1105–1117. https://doi.org/10.1121/1.1383299

Cardozo, R. N. (1965). An Experimental Study of Customer Effort, Expectation, and Satisfaction. *Journal of Marketing Research*, *2*(3), 244–249. JSTOR. https://doi.org/10.2307/3150182

Carlile, S., Leong, P., & Hyams, S. (1997). The nature and distribution of errors in sound localization by human listeners. *Hearing Research*, *114*(1), 179–196. https://doi.org/10.1016/S0378-5955(97)00161-5

Carpentier, T., BAHU, H., Noisternig, M., & Warusfel, O. (2014, September). Measurement of a head-related transfer function database with high spatial resolution. *7th Forum Acusticum(EAA)*. https://hal.archives-ouvertes.fr/hal-01247583

Claes, P., Reijniers, J., Shriver, M. D., Snyders, J., Suetens, P., Nielandt, J., De Tré, G., & Vandermeulen, D. (2015). An investigation of matching symmetry in the human pinnae with possible implications for 3D ear recognition and sound localization. *Journal of Anatomy*, *226*(1), 60–72. https://doi.org/10.1111/joa.12252

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge. https://doi.org/10.4324/9780203771587

Danfeng Li, & Levinson, S. E. (2003). A Bayes-rule based hierarchical system for binaural sound source localization. *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03).*, *5*, V–521. https://doi.org/10.1109/ICASSP.2003.1200021

David, E. E., Guttman, N., & van Bergeijk, W. A. (1959). Binaural Interaction of High-Frequency Complex Stimuli. *The Journal of the Acoustical Society of America*, *31*(6), 774–782. https://doi.org/10.1121/1.1907784

Deatherage, B. H. (1961). Binaural Interaction of Clicks of Different Frequency Content. *The Journal of the Acoustical Society of America*, *33*(2), 139–145. https://doi.org/10.1121/1.1908602

Deatherage, B. H., & Hirsh, I. J. (1959). Auditory Localization of Clicks. *The Journal of the Acoustical Society of America*, *31*(4), 486–492. https://doi.org/10.1121/1.1907740

Denk, F., Heeren, J., Ewert, S., Kollmeier, B., & Ernst, S. (2017, March 17). *Controlling the Head Position during individual HRTF Measurements and its Effect on Accuracy*.

Dixon, W. J. (1960). Simplified Estimation from Censored Normal Samples. *The Annals of Mathematical Statistics*, *31*(2), 385–391. https://doi.org/10.1214/aoms/1177705900

Domnitz, R. H., & Colburn, H. S. (1977). Lateral position and interaural discrimination. *The Journal of the Acoustical Society of America*, *61*(6), 1586–1598. https://doi.org/10.1121/1.381472

Durlach, N. I. (1963). Equalization and Cancellation Theory of Binaural Masking-Level Differences. *The Journal of the Acoustical Society of America*, *35*(8), 1206–1218. https://doi.org/10.1121/1.1918675

Durlach, N. I., & Braida, L. D. (1969). Intensity perception. I. Preliminary theory of intensity resolution. *The Journal of the Acoustical Society of America*, *46*(2), 372–383. https://doi.org/10.1121/1.1911699

Durlach, N. I., & Colburn, H. S. (1978). *Handbook of Perception* (Vol. 4). Academic Press.

Feddersen, W. E., Sandel, T. T., Teas, D. C., & Jeffress, L. A. (1957). Localization of High-Frequency Tones. *The Journal of the Acoustical Society of America*, *29*(9), 988–991. https://doi.org/10.1121/1.1909356

Fisher, H. G., & Freedman, S. J. (1968). The role of the pinna in auditory localization. *Journal of Auditory Research*, *8*(1), 15–26.

Gaik, W. (1993). Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling. *Journal of the Acoustical Society of America*, *94*(1), 98–110. https://doi.org/10.1121/1.406947

Gelfand, S. A. (2009). *Hearing: An introduction to psychological and physiological acoustics* (5. ed). Informa Healthcare.

Gulick, W. L., Gescheider, G. A., & Frisina, R. D. (1989). *Hearing: Physiological acoustics, neural coding, and psychoacoustics*. Oxford University Press.

Hafter, E. R., & Jeffress, L. A. (1968). Two-Image Lateralization of Tones and Clicks. *The Journal of the Acoustical Society of America*, *44*(2), 563–569. https://doi.org/10.1121/1.1911121

Hansen, V., & Munch, G. (1991). Making Recordings for Simulation Tests in the Archimedes Project. *Journal of the Audio Engineering Society*, *39*(10), 768–774.

Harker, A., & Tremblay, P. A. (2012). The HISSTools Impulse Response Toolbox: Convolution for the Masses. In M. Marolt, M. Kaltenbrunner, & M. Ciglar (Eds.), *Proceedings of the International Computer Music Conference* (pp. 148–155). The International Computer Music Association. https://github.com/HISSTools/HISSTools_Impulse_Response_Toolbox

Harris, G. G. (1960). Binaural Interactions of Impulsive Stimuli and Pure Tones. *The Journal of the Acoustical Society of America*, *32*(6), 685–692. https://doi.org/10.1121/1.1908181

Hartmann, W. M., Rakerd, B., Crawford, Z. D., & Zhang, P. X. (2016). Transaural experiments and a revised duplex theory for the localization of low-frequency tones. *The Journal of the Acoustical Society of America*, *139*(2), 968–985. https://doi.org/10.1121/1.4941915

Hartmann, W., & Wittenberg, A. (1996). On the externalization of sound images. *The Journal of the Acoustical Society of America*, *99*, 3678–3688. https://doi.org/10.1121/1.414965

Hirahara, T., Sagara, H., Toshima, I., & Otani, M. (2010). *Head movement during head-related transfer function measurements*. https://doi.org/10.1250/ast.31.165

Howard, D. M., & Angus, J. (2013). *Acoustics and Psychoacoustics (4th Edition)*. Focal Press. https://www-dawsonera-com.libaccess.hud.ac.uk/abstract/9780080961873

International Telecommunication Union. (2015). *BS.1116. Methods for the subjective assessment of small impairments in audio systems*.

Jeffress, L. A. (1948). A place theory of sound localization. *Journal of Comparative and Physiological Psychology*, *41*(1), 35–39. https://doi.org/10.1037/h0061495

Jeffress, L. A., & Taylor, R. W. (1960). Lateralization vs Localization. *The Journal of the Acoustical Society of America*, *32*(7), 936–936. https://doi.org/10.1121/1.1936556

Kuhn, G. (1977). Model for the interaural time differences in the azimuthal plane. *Journal of The Acoustical Society of America - J ACOUST SOC AMER*, *62*. https://doi.org/10.1121/1.381498

Leakey, D. M. (1959). Some Measurements on the Effects of Interchannel Intensity and Time Differences in Two Channel Sound Systems. *The Journal of the Acoustical Society of America*, *31*(7), 977–986. https://doi.org/10.1121/1.1907824

Lee, H., Johnson, D., & Mironovs, M. (2016, May 26). *A New Response Method for Auditory Localization and Spread Tests*. Audio Engineering Society Convention 140. http://www.aes.org/e-lib/browse.cfm?elib=18144

Lee, H., & Rumsey, F. (2013). Level and Time Panning of Phantom Images for Musical Sources. *Journal of the Audio Engineering Society*, *61*(12), 978–988.

Levitt, H. (1971). Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2B), 467–477. https://doi.org/10.1121/1.1912375

Lim, C., & Duda, R. O. (1994). Estimating the azimuth and elevation of a sound source from the output of a cochlear model. *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*, *1*, 399–403 vol.1. https://doi.org/10.1109/ACSSC.1994.471484

Lindemann, W. (1986). Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. *The Journal of the Acoustical Society of America*, *80*(6), 1608–1622. https://doi.org/10.1121/1.394325

Lord Rayleigh, L. R. O. M. P. (1907). XII. On our perception of sound direction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *13*(74), 214–232. https://doi.org/10.1080/14786440709463595

Macpherson, E. A., & Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *The Journal of the Acoustical Society of America*, *111*(5), 2219–2236. https://doi.org/10.1121/1.1471898

Majdak, P., Goupell, M. J., & Laback, B. (2010). 3-D Localization of Virtual Sound Sources: Effects of Visual Environment, Pointing Method, and Training. *Attention, Perception & Psychophysics*, *72*(2), 454–469. https://doi.org/10.3758/APP.72.2.454

Makous, J. C., & Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. *The Journal of the Acoustical Society of America*, *87*(5), 2188–2200. https://doi.org/10.1121/1.399186

Mcgill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of Box Plots. *The American Statistician*, *32*(1), 12–16. https://doi.org/10.1080/00031305.1978.10479236

Middlebrooks, J. C., & Green, D. M. (1991). Sound Localization by Human Listeners. *Annual Review of Psychology*, *42*(1), 135–159. https://doi.org/10.1146/annurev.ps.42.020191.001031

Middlebrooks, J., Makous, J., & Green, D. (1989). Directional sensitivity of sound-pressure levels in the human ear canal. *The Journal of the Acoustical Society of America*, *86*, 89–108. https://doi.org/10.1121/1.398224

Mills, A. W. (1958). On the Minimum Audible Angle. *The Journal of the Acoustical Society of America*, *30*(4), 237–246. https://doi.org/10.1121/1.1909553

Moushegian, G., & Jeffress, L. A. (1959). Role of Interaural Time and Intensity Differences in the Lateralization of Low-Frequency Tones. *The Journal of the Acoustical Society of America*, *31*(11), 1441–1445. https://doi.org/10.1121/1.1907647

Odegaard, B., Wozny, D. R., & Shams, L. (2015). Biases in Visual, Auditory, and Audiovisual Perception of Space. *PLOS Computational Biology*, *11*(12), e1004649. https://doi.org/10.1371/journal.pcbi.1004649

Park, M. (2007). *Models of binaural hearing for sound lateralisation and localisation* [Phd, University of Southampton]. https://eprints.soton.ac.uk/162297/

Park, M., Nelson, P. A., & Kang, K. (2008). A Model of Sound Localisation Applied to the Evaluation of Systems for Stereophony. *Acta Acustica United with Acustica*, *94*(6), 825–839. https://doi.org/10.3813/AAA.918102

Pernaux, J.-M., Emerit, M., Daniel, J., & Nicol, R. (2002, June 1). *Perceptual Evaluation of Static Binaural Sound Synthesis*. Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio. http://www.aes.org/e-lib/browse.cfm?elib=11152

Perrott, D. R., & Baars, B. J. (1974). Detection of interaural onset and offset disparities. *The Journal of the Acoustical Society of America*, *55*(6), 1290–1292. https://doi.org/10.1121/1.1914699

Plenge, G. (1974). On the differences between localization and lateralization. *The Journal of the Acoustical Society of America*, *56*(3), 944–951. https://doi.org/10.1121/1.1903353

Pulkki, V., Karjalainen, M., & Huopaniemi, J. (1999). *Analyzing virtual sound source attributes using a binaural auditory model*.

Rakerd, B., & Hartmann, W. M. (1985). Localization of sound in rooms, II: The effects of a single reflecting surface. *The Journal of the Acoustical Society of America*, *78*(2), 524–533. https://doi.org/10.1121/1.392474

Raspaud, M., Viste, H., & Evangelista, G. (2010). Binaural Source Localization by Joint Estimation of ILD and ITD. *IEEE Transactions on Audio, Speech, and Language Processing*, *18*, 68–77. https://doi.org/10.1109/TASL.2009.2023644

Sayers, B. McA. (1964). Acoustic-Image Lateralization Judgments with Binaural Tones. *The Journal of the Acoustical Society of America*, *36*(5), 923–926. https://doi.org/10.1121/1.1919121

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, *52*(3/4), 591–611. JSTOR. https://doi.org/10.2307/2333709

Shaw, E. a. G. (1974). Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *The Journal of the Acoustical Society of America*, *56*(6), 1848–1861. https://doi.org/10.1121/1.1903522

Shaw, E. A. G. (1997). *Acoustical features of the human external ear*.

Shear, G. D., & Stern, R. M. (1987). Extending the position-variable model: Dependence of lateralization on frequency and bandwidth. *The Journal of the Acoustical Society of America*, *81*(S1), S27–S27. https://doi.org/10.1121/1.2024167

Sound Quality Assessment Material recordings for subjective tests. (2008). https://tech.ebu.ch/publications/sqamcd

Stecker, G. C. (2010a). Trading of interaural differences in high-rate Gabor click trains. *Hearing Research*, *268*(1), 202–212. https://doi.org/10.1016/j.heares.2010.06.002

Stecker, G. C. (2010b). Trading of interaural differences in high-rate Gabor click trains. *Hearing Research*, *268*(1), 202–212. https://doi.org/10.1016/j.heares.2010.06.002

Stecker, G. C., & Gallun, F. (2012). *Binaural Hearing, Sound Localization, and Spatial Hearing*.

Stern, R. M., & Colburn, H. S. (1978). Theory of binaural interaction based in auditory-nerve data. IV. A model for subjective lateral position. *The Journal of the Acoustical Society of America*, *64*(1), 127–140. https://doi.org/10.1121/1.381978

Takanen, M., & Lorho, G. (2012, March 1). *A Binaural Auditory Model for the Evaluation of Reproduced Stereophonic Sound*. Audio Engineering Society Conference: 45th International Conference: Applications of Time-Frequency Processing in Audio. http://www.aes.org/e-lib/browse.cfm?elib=16204

Tobias, J. V., & Zerlin, S. (1959). Lateralization Threshold as a Function of Stimulus Duration. *The Journal of the Acoustical Society of America*, *31*(12), 1591–1594. https://doi.org/10.1121/1.1907664

Trahiotis, C., & Kappauf, W. E. (1978). Regression interpretation of differences in time-intensity trading ratios obtained in studies of laterality using the method of adjustment. *The Journal of the Acoustical Society of America*, *64*(4), 1041–1047. https://doi.org/10.1121/1.382087

Wallis, R., & Lee, H. (2017). The Reduction of Vertical Interchannel Crosstalk: The Analysis of Localisation Thresholds for Natural Sound Sources. *Applied Sciences*, *7*(3), 278. https://doi.org/10.3390/app7030278

Whitworth, R. H., & Jeffress, L. A. (1961). Time vs Intensity in the Localization of Tones. *The Journal of the Acoustical Society of America*, *33*(7), 925–929. https://doi.org/10.1121/1.1908849

Wightman, F. L., & Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, *91*(3), 1648–1661. https://doi.org/10.1121/1.402445

Wightman, F. L., & Kistler, D. J. (1997). Monaural sound localization revisited. *The Journal of the Acoustical Society of America*, *101*(2), 1050–1063. https://doi.org/10.1121/1.418029

Wittek, H., & Theile, G. (2002, April 1). *The Recording Angle—Based on Localisation Curves*. Audio Engineering Society Convention 112. http://www.aes.org/e-lib/browse.cfm?elib=11307

Woodworth, R. S. (1938). *Experimental psychology*. Holt.

Yost, W. (1981). Lateral position of sinusoids presented with interaural intensive and temporal differences. *Journal of The Acoustical Society of America - J ACOUST SOC AMER*, *70*. https://doi.org/10.1121/1.386775

Yost, W. A. (1993, June 1). *Perceptual Models for Auditory Localization*. Audio Engineering Society Conference: 12th International Conference: The Perception of Reproduced Sound. http://www.aes.org/e-lib/browse.cfm?elib=6242

Young Lamar L. (1976). Time-Intensity Trading Functions for Selected Pure Tones. *Journal of Speech and Hearing Research*, *19*(1), 55–67. https://doi.org/10.1044/jshr.1901.55