Investigating the Feasibility of a Near-Field Binaural Loudspeaker System

Kat Young

PhD

University of York
Electronic Engineering

September 2019

Abstract

Binaural audio techniques have been used since the 1880s to create realistic and convincing virtual audio experiences. However, loudspeaker binaural reproduction generally involves an additional process to artificially increase the separation between the left and right audio channels. This process can affect the quality of the listening experience and complicate the reproduction setup. Therefore, a loudspeaker method which avoids this process would be of significant interest. The approach proposed in this thesis is to passively increase the left-right channel separation through suitable loudspeaker placement.

Acoustic simulation enables investigation of more loudspeaker directions and distances for more subjects than is feasible using acoustic measurement. However, simulation requires high-resolution meshes and considerable computational resources. Initial thesis work focused on the development and validation of a suitable 3D mesh model of the human form. A modified version of the mesh was created, comprising only the head and shoulders region. This served to reduce computation time and so enabled simulation of the performance of many loudspeaker positions without reducing the maximum valid simulation frequency.

Loudspeaker positions were identified which exhibit left-right channel separation greater than the threshold reported in the literature as required for robust binaural reproduction. To additionally characterise the perceptual impact of loudspeaker placement, the deterioration in interaural binaural cues associated with each loudspeaker pair was determined. Using a conservative model, positions were identified where deterioration was below the estimated threshold of perception for multiple subjects. A small number of loudspeaker pairs at close radial distances met this requirement, but it is likely that these are only a subset of the positions which would perform satisfactorily in a real-world loudspeaker system. This indicates that a simplified binaural loudspeaker reproduction system, capable of satisfactory performance for multiple listeners without adjustment, is a viable possibility.

I've done all that I can,
and so I say hello.

Tegan and Sara, *Hello*
Yellow Demo (1998)

List of Contents

List of Figures

List of Tables

Acknowledgements

This work was generously supported by Meridian Audio Ltd and MQA Ltd. My thanks to those colleagues current and former who gave me so much freedom to pursue this idea, in particular Michael Capp, Richard Hollingshead, Bob Stewart and Laurence Hobden.

I would like to thank my supervisors, Tony Tew and Gavin Kearney, for their wisdom, support, and guidance over the last few years. A team of three can be difficult to wrangle, but I think we managed it. I will never look at sentence structure in the same way again. My thanks also to Helena Daffern as my thesis (and general life) advisor - I'm sorry there are so many pages.

I am also grateful to the following people for their assistance with practical aspects of this PhD: Patrick Macey and John King for PAFEC-related support; Brian Hamilton for the introduction to Mesh2HRTF; Piotr Majdack and his team at ARI for answering our subsequent many questions about Mesh2HRTF; Joel Fergusson for his knowledge of Blender and his patience; Jonathan Rainer for helping me untangle set notation; Mark Hough and Dave Turner of TSS for accommodating my ridiculous requests; colleagues at GRAS Sound and Vibration for supplying the CAD file of KEMAR, and Andrew Chadwick for his seemingly endless willingness to help build contraptions.

I am also indebted to a number of people for their moral support. My officemates over the last few years, Alejandro, Evelyn and Mole - thank you for putting up with me; I'm sorry I talk to myself so much. Amelia and Naomi - the coffee breaks and the check-ins were much appreciated. Dave and Evelyn - thank you for the game-related light relief. There's always more studs to collect and holes to dig. Jude - thank you for, among many things, the cake. The many AudioLab members past and present - you've made this process far easier. Thank you for the encouragement, the ridicule, the proofreading and the distractions. My parents and my brother - thank you for not asking too many questions. To Tegan and Sara - for giving this weird kid self belief.

Finally, immeasurable thanks go to my dearest Laura. It's been a long and bumpy road, but I couldn't have done it without your support, your patience, and your willingness to make dinner.

Declaration

I, Kat Young, declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as references.

In addition, I declare that parts of this research have been presented at conferences during the course of the research degrees. The related publications are as follows:

Boundary element method modelling of KEMAR for binaural rendering: Mesh production and validation, K. Young, A. I. Tew, and G. Kearney, in Interactive Audio Systems Symposium, York, 2016.

Acoustic Validation of a BEM-Suitable 3D Mesh Model of KEMAR, K. Young, G. Kearney, and A. I. Tew, in Audio Engineering Society International Conference on Spatial Reproduction - Aesthetics and Science, Tokyo, 2018.

Loudspeaker Positions with Sufficient Natural Channel Separation for Binaural Reproduction, K. Young, G. Kearney, and A. I. Tew, in Audio Engineering Society International Conference on Spatial Reproduction - Aesthetics and Science, Tokyo, 2018.

A Numerical Study into Perceptually-Weighted Spectral Differences between Differently-Spaced HRTFs, K. Young, C. Armstrong, A. I. Tew, D. T. Murphy, and G. Kearney, in Audio Engineering Society International Conference on Immersive and Interactive Audio, York, 2019. *(Equal contribution of first two authors)*

Thesis

Chapter 1

Introduction

We don't know, so we only go forward.

Tegan and Sara, *Hype*
Under Feet Like Ours (1999)

All real-world sounds possess attributes which vary independently and over time. These include loudness, frequency content, duration, direction and distance. The auditory system is capable of interpreting these attributes and thus deriving meaning from them. Of particular interest to this thesis is the ability of the auditory system to determine the direction of a sound in space. Whilst this ability may have originated as a life-preserving one (for example, knowing the direction of a roaring lion), with modern technology it can be exploited to create extremely accurate and engaging virtual sound experiences for use in gaming, cinema, and virtual and augmented reality media.

The mechanisms behind determining the direction of sound (sound localisation) are generally well understood. Information from a number of cues is combined, including: the differences between the signals at the ears; the filtering applied by the torso, head and outer ear; and any available visual cues [1]. The cues which occur as a result of interactions between the sound waves and the subject (the interaural differences and spectral filtering) can be captured using in-ear measurement or simulation. They can then be used to create realistic virtual audio experiences. This technique, commonly referred to as binaural audio, aims to create the same signals at the eardrums as would occur if a real sound source were present, and so is capable of a high degree of realism.

The first use of binaural technology has been traced to the 1881 Paris World Expo. In a system later referred to as the *Théâtrophone*, Clément Ader placed 80 telephone transmitters across the front of a stage, and performances were relayed to distant headsets with a pair of tiny loudspeakers on each [2, 3]. The technology became very popular across much of Europe, although The Théâtrophone Company ceased operations in 1932 [4, 5]. The Chicago World Fair in 1933 saw the introduction of the first artificial head-based binaural system, courtesy of Bell Labs [3]. Nicknamed 'Oscar', the artificial head had a microphone mounted in each ear, and listeners were apparently astounded by the experience of listening to what Oscar could hear [5]. Further experimentation using binaural technologies continued through the next few decades, with the 1960s and 1970s sometimes considered a 'golden age' for binaural technology, particularly with regards to understanding key principles and aspects of localisation [3].

In 1978, Lou Reed released *Street Hassle*, said to be the first commercial pop album recorded using binaural techniques [6]. Manfred Schunke, sound engineer and owner of Delta Studio, had experimented with artificial head recordings and suggested using the recording technique to achieve a natural and live sound. Binaural audio went on to find favour across a range of material[1], with some record labels such as Audiostax [7] specialising in such recording techniques.

Today, use of binaural technology goes far beyond the reproduction of recorded music and the creation of short demonstrations. Understanding the underlying theory of binaural audio has given practitioners and researchers the tools to create virtual sound experiences from the ground up, rather than relying on recorded material which cannot subsequently be altered. This has enabled the inclusion of binaural audio in a wide variety of applications, including broadcast [8], gaming [9], assistive technologies [10] and many under the umbrella term of *virtual reality* for applications ranging from entertainment [11] to heritage [12] and training [13].

It is easiest to reproduce binaural audio over headphones, as the necessary degree of acoustic isolation between the left and right audio channels during playback can be more readily guaranteed to ensure the correct binaural cues reach the correct ear. However, headphones are not suitable in every listening situation, can be uncomfortable to wear for extended periods of time, and can lead to a feeling of internalisation, whereby the sound is perceived to come from inside the head [14].

Using a pair of loudspeakers to reproduce binaural audio avoids having to wear headphones, but introduces additional problems. In particular, there is now an unwanted acoustic path from the left loudspeaker to the right ear (and vice versa), referred to as crosstalk, which interferes with the binaural cues at each ear and degrades the listening experience. An additional processing step, referred to as crosstalk cancellation (CTC), is generally implemented to ensure that enough separation is present between the left-right channels to allow the binaural cues to reach the correct ear. However, there are downsides associated with the implementation of CTC, including spectral colouration (which may affect the experience of the listener) [15–17], and limitations imposed on the reproduction system itself (which may affect the implementation) [18, 19].

For much of binaural reproduction over loudspeakers, loudspeakers operating in the far-field have been used (those at a distance of greater than approximately 1 m from the head), as in this case sound propagates from the loudspeaker to the listener in an easily predictable manner. Recently, however, there has been an appreciable drive to place loudspeakers closer to the listener in a variety of applications (gaming, cinema, home theatre and vehicle seats, for example), but the potential of these types of configuration for binaural playback has not been fully explored. Given that such systems allow for a more compact and personalised listening experience without the invasiveness of headphones, exploration of the potential for binaural reproduction over near-field loudspeaker systems is a logical and worthwhile next step for immersive environments.

Due to near-field effects and the introduction of additional resonances, implementation of CTC within the near-field can be more difficult than in a far-field loudspeaker system [20]. There is potential, however, for further simplification of the reproduction

[1]For some highly-regarded examples, see: `https://hookeaudio.com/blog/music/best-binaural-albums/` and `https://youtu.be/IUDTlvagjJA`

system. Every loudspeaker position has an associated natural channel separation (NCS), whereby the sound from the loudspeaker arriving at the nearer ear is generally at a higher amplitude than it is arriving at the more distant ear due to the acoustic shadowing effect caused by the head [19, 21] (assuming the loudspeaker is not centrally positioned). CTC can be thought of as an active method for artificially increasing the NCS. However, a method of increasing the NCS passively is simply to bring the loudspeaker closer to the head to strengthen the acoustic shadowing effect . A small amount of existing literature has indicated that sufficient levels of channel separation for binaural reproduction can be achieved in this manner when the loudspeakers are approximately 20 cm from the head [22, 23], but the literature is missing a systematic investigation of the variation of NCS with both loudspeaker direction and loudspeaker distance.

It is possible that the increase in channel separation achieved inherently through the use of carefully-positioned loudspeakers in the near-field is sufficient to negate the requirement for CTC. This has the potential to bring about a far simpler binaural reproduction system. The identification of these loudspeaker positions forms the motivation for the work presented in this thesis, which is expressed in the following statement of hypothesis.

1.1 Statement of Hypothesis

It is hypothesised that:

It is possible to avoid the use of crosstalk cancellation when reproducing binaural audio using loudspeakers through suitable placement of the loudspeakers in the near-field.

1.2 Objectives

The main objectives of this thesis are therefore:

1. To acquire or create a 3D mesh model of the human form suitable for acoustic simulation of binaural loudspeaker reproduction;

2. To use the mesh in simulation of acoustic localisation cues for a large number of loudspeaker directions and distances;

3. To determine whether any pairs of loudspeaker positions are capable of satisfactory binaural reproduction without a CTC stage;

4. To extend this analysis to determine whether any loudspeaker pair positions exist which perform satisfactorily for multiple subjects.

1.3 Thesis Structure

The thesis is organised as follows. To provide background information relevant to the work presented in this thesis, Chapter 2 presents an overview of acoustic principles relevant to psychoacoustics and binaural technology, with an emphasis on the acquisition and reproduction of the acoustic cues required for localisation of sound sources. Topics covered include: the generation and propagation of sound waves; how the human auditory system detects, perceives and localises sound; various techniques for the capture of acoustic localisation cues; and methods of reproduction of these cues using headphones and loudspeakers.

The acoustic cues are embedded in a set of acoustic filters known as head-related transfer functions (HRTFs). A large number of these filters had to be generated in order to investigate the impact of loudspeaker position on binaural reproduction. Chapter 3 describes the workflow developed to create a 3D mesh model of the human head and torso suitable for use with acoustic simulation software and the subsequent numerical and acoustic validation steps needed to ensure satisfactory simulation performance. Chapter 4 focuses on the development of a simplified, but still topologically accurate, 3D mesh of the head and shoulders only, suitable for simulation using more limited computational resources. Chapter 5 describes the generation of acoustic filters for many loudspeaker directions and distances, using the mesh created in Chapter 4, and the subsequent calculation of NCS for each loudspeaker position.

Chapter 6 details the process of creating a virtual test environment and use of it in assessing the impact of loudspeaker placement on CTC-less binaural reproduction. Pairs of loudspeaker positions are identified which are predicted to be capable of satisfactorily reproducing binaural signals without CTC. The test environment is adapted in Chapter 7 to investigate the existence of loudspeaker pairs not requiring CTC for a variety of other head and ear shapes, and to identify any common pairs between them. Chapter 8 discusses some limitations of the work presented in the thesis and proposes directions for future research.

The thesis concludes in Chapter 9 with an overview of the research carried out and revisits the hypothesis in light of the findings presented.

1.4 Contributions

The work presented in this thesis has resulted in the following novel contributions to the field:

- Seven high-resolution 3D mesh models of the Knowles Electronics Manikin for Acoustic Research (KEMAR), numerically and acoustically validated for use with the BEM and valid up to audio frequencies not previously presented in the literature;

- NCS values for 655,216 loudspeaker positions over 15 radial distances, within both the near-field and the far-field;

- Identification of candidate pairs of loudspeaker positions capable of reproducing satisfactory binaural audio without requiring CTC across multiple subjects, determined using a simulated environment.

1.5 Conclusion

This chapter introduces the motivation for the work undertaken in this thesis and presents the statement of hypothesis. The structure of the remainder of the thesis is described and the novel contributions are listed. Building on the background discussed here, Chapter 2 presents relevant theory and literature in order to contextualise further the research undertaken.

Chapter 2

Literature Review

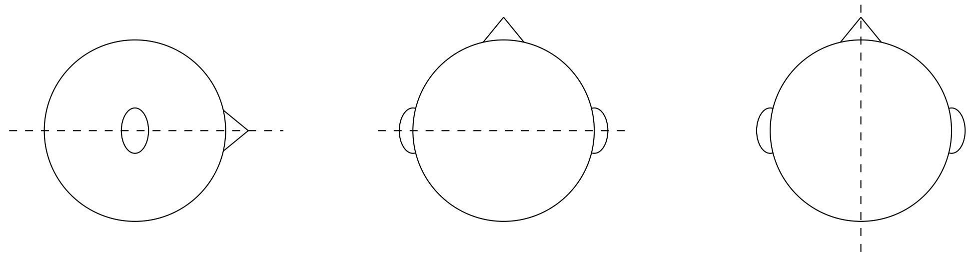
It's a top ten list of things that move me
the most.

Tegan and Sara, *The First*
This Business of Art (2000)

As indicated in Chapter 1, the work presented in this thesis builds on a number of key principles within binaural audio. This chapter presents relevant literature in several fundamental areas, including: the properties of sound and sound wave behaviour; the anatomy of the auditory system; and the formation, capture and reproduction of acoustic cues for the localisation of sounds in 3D.

2.1 Describing Auditory Space

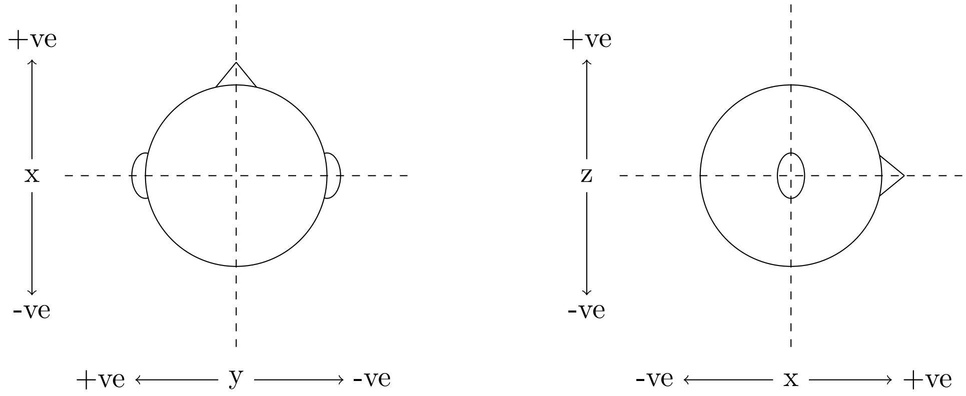
The 3D space surrounding the listener is referred to as the auditory space, which can be defined in terms of three perpendicular 2D planes passing through the origin at the centre of the head. These are referred to as the *principal anatomical planes*: the horizontal (also known as transverse) plane which separates top and bottom hemispheres, the frontal (also known as coronal or dorsal) plane which separates front and back hemispheres, and the median (also known as mid-sagittal) plane which separates left and right hemispheres, as shown in Fig. 2.1. Planes which lie parallel to the median plane but do not pass through the origin are referred to simply as sagittal planes.



(a) Horizontal plane (b) Frontal plane (c) Median plane

Figure 2.1: The three principal anatomical planes, indicated with dotted lines.

The auditory space surrounding the listener can be described using a number of coordinate systems, all with the origin at the centre of the head between the two ears. In Cartesian space, a position is defined by a value along each of three perpendicular axes: x, y, and z. The direction of each axis relative to the origin can vary, but in this thesis, x is back-to-front, y is right-to-left, and z is down-to-up (each in a positive-going direction). This is shown in Fig. 2.2. Therefore, the y-axis, the locus of points defined by $x = 0,\ z = 0$, is also the interaural axis, i.e. the axis between the two ears.



(a) Horizontal plane view (b) Median plane view

Figure 2.2: Views of the Cartesian coordinate system.

In a spherical coordinate system a position is again defined in terms of three numbers: a radial distance r, an azimuth (horizontal) angle θ and an elevation (vertical) angle ϕ. Two spherical coordinate systems exist. The vertical-polar coordinate system (Fig. 2.3a) is favoured by most researchers, where azimuth is measured in a full circle in the horizontal plane and elevation is measured in an arc between directly above and directly below. The interaural-polar coordinate system (Fig. 2.3b), used by Algazi *et al.* [24] and Brown and Duda [25] among others, defines azimuth as a frontal arc between the left and right, and elevation in a full circle about the interaural axis. This results in a common elevation value between the two systems, but differing azimuth values.

(a) Vertical-polar, where elevation angle is measured with respect to the origin.

(b) Interaural-polar, where elevation angle is measured with respect to the interaural axis.

Figure 2.3: The two spherical coordinate systems used in the literature.

Within the vertical-polar system, azimuth can be defined using one of two systems, as illustrated in Fig. 2.4. Black indicates azimuth defined as increasing in a counter-clockwise direction, such that $\theta = 90°$ is on the left and $\theta = 270°$ is on the right. Red indicates azimuth as increasing in a counter-clockwise direction and decreasing in a clockwise direction, both from zero straight ahead, such that $\theta = 90°$ is on the left and $\theta = -90°$ is on the right. For elevation, $-90°$ is below and $90°$ is above.

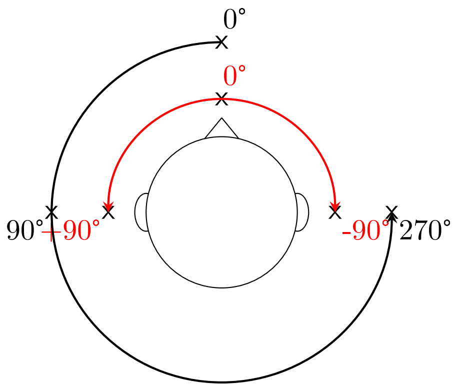


Figure 2.4: The two common ways of defining azimuth angle within the vertical-polar coordinate system. Black indicates azimuth defined as increasing in a counter-clockwise direction. Red indicates azimuth as increasing in a counter-clockwise direction and decreasing in a clockwise direction, both from straight ahead.

For the interaural-polar system, care must be taken to ensure that the correct definitions of azimuth and elevation are applied, that is, whether $\theta = 90°$ is defined as being on the left or the right, and whether $\phi = 90°$ is defined as being the top or the bottom, as it varies within the literature.

Both coordinate systems are useful for different purposes. For example, interaural-polar coordinates are a useful description of cones of constant azimuth[1] or planes of constant elevation.

In this thesis, the vertical-polar coordinate system will be used (Fig. 2.3a) with azimuth increasing in the counter-clockwise direction (indicated in black in Fig 2.4), as it is more appropriate here to relate each point to the centre of the head. In the case where an alternate system is used in the literature, this will be noted.

2.2 Properties of Sound

It is useful to understand the formation of sound waves, as well as a number of associated descriptive properties. Waves occur in a medium when the particles within that medium vibrate in an oscillatory motion as a result of an applied force. The oscillatory motion of one particle is coupled to the next according to the physical properties of the medium, and so on through subsequent neighbours. This creates regions within the medium where particles are squashed together (resulting in a higher pressure than equilibrium, known as compression) and regions where particles are stretched apart (resulting in a lower pressure than equilibrium, known as rarefaction) [26–28]. This behaviour is illustrated

[1]Section 2.5.3 explains that such cones are referred to as cones of confusion.

in the upper portion of Fig. 2.5. This type of propagation, where the vibration of the particles in the medium is parallel to the propagation of the wave, is described as longitudinal. The variation in pressure, also referred to as the pressure wave (illustrated in the lower portion of Fig. 2.5), is detected by the human auditory system as a sound when the variation occurs at a rate between approximately 20 Hz to 20 kHz [29].

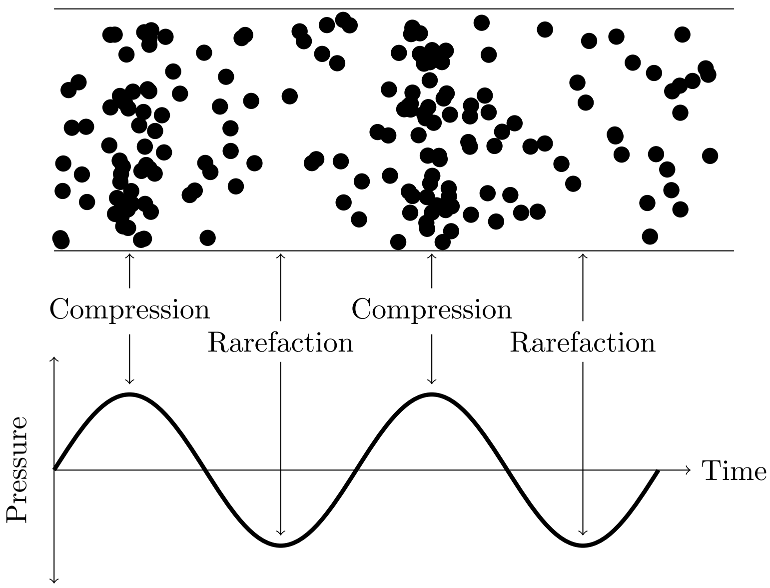


Figure 2.5: The vibration of particles in a medium results in regions of compression (higher than equilibrium pressure) and rarefaction (lower than equilibrium pressure). The variation in pressure, referred to as the pressure wave, is detected by the human auditory system.

The variation in pressure is directly proportional to the amplitude of the sound wave, where pressures higher than equilibrium are represented by a positive amplitude and pressures lower than equilibrium are represented by a negative amplitude [28]. The amplitude can be described using a number of measurements as illustrated in Fig. 2.6, but most commonly used is the root mean square (RMS) of the difference between the instantaneous pressure and the equilibrium pressure [26, 30]. This quantity is defined in IEC 801-21-20 [31] as the sound pressure p and has the standard unit of pressure, pascal (Pa). When considering the propagation of sound in air, the equilibrium state (where no displacement is present) is referred to as atmospheric, absolute or static pressure. Although atmospheric pressure constantly changes slowly, a standard pressure of one atmosphere is defined as 101,325 Pa.

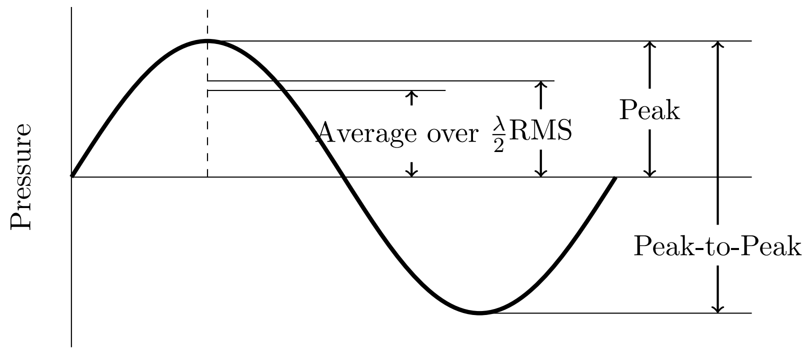


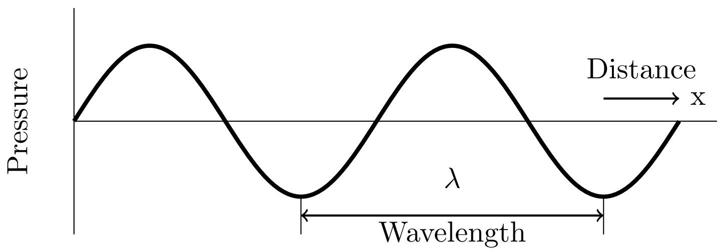
Figure 2.6: The amplitude of a sine wave can be described using a number of pressure metrics: average over half the period, RMS, peak and peak-to-peak.

Corresponding values of this sinusoidally varying sound pressure are found at regularly spaced points within the medium along the direction of wave propagation. These points indicate the time or distance taken to complete one cycle. When measured in terms of time, this quantity is the period T in seconds, and when measured in terms of distance, this quantity is the wavelength λ in metres. These quantities are illustrated in Fig. 2.7. T and λ are related to the number of cycles per second, the frequency f in Hz, by:

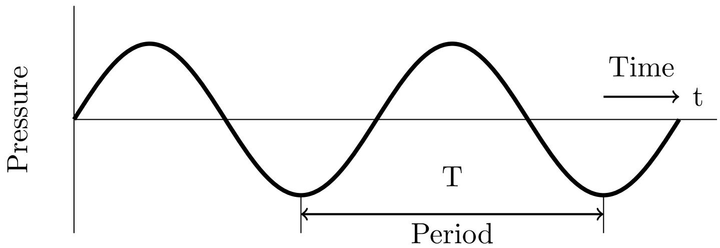
$$f = \frac{c}{\lambda} \tag{2.1}$$

$$f = \frac{1}{T} \tag{2.2}$$

where c is the speed of the wave in $\mathrm{m\,s^{-1}}$.



(a) The distance taken to complete one cycle is the wavelength, λ, measured in m.



(b) The time taken to complete one cycle is the period, T, measured in s.

Figure 2.7: Cyclic quantities of sound waves: wavelength and period.

In addition to the properties described above (frequency f, wavelength λ and period T), to fully characterise a sinusoid the starting phase Θ is also required. This denotes where in the cycle the variation pattern begins at time zero (as it may not start at equilibrium pressure), although the human auditory system is only sensitive to starting phase under certain conditions [32]. The instantaneous amplitude of a sinusoid at time t can be calculated using:

$$a(t) = A\sin(2\pi f t + \Theta) \tag{2.3}$$

where A is a value associated with the maximum displacement from zero, f is the frequency in Hz, t is the time point of interest in s, and Θ is the starting phase in °. Through rearrangement of Equations 2.1 and 2.2, the speed of the wave in the medium can be calculated if the distance travelled and time taken to travel this distance are known:

$$c = \frac{\lambda}{T} \tag{2.4}$$

$$c = \lambda f \tag{2.5}$$

The speed of the wave is also dependent on properties of the medium not explicitly described in Equations 2.4 and 2.5. In fluids (which includes air), the primary properties are the density ρ in $\mathrm{kg\,m^{-3}}$ and the bulk modulus K of the medium, both of which are volumetric measures and are affected by temperature. To include these dependencies, the speed can be given by the Newton-Laplace equation [33]:

$$c = \sqrt{\frac{K}{\rho}} \tag{2.6}$$

For air, ρ can be calculated using:

$$\rho_{air} = \frac{P_0}{\hat{R}_{air}(T_C + 273.15)} \tag{2.7}$$

where T_C is temperature in °C, P_0 is one standard atmospheric pressure (101,325 Pa) and $\hat{R}_{air}$ is the molar-weight-specific gas constant for dry air ($287\,\mathrm{J\,kg^{-1}\,K^{-1}}$), based on a mean molar mass of $28.97\,\mathrm{g\,mol^{-1}}$. This formulation is derived from [33]; the full derivation is given in Appendix A.

The speed of sound in air including temperature dependency can be calculated using the following approximation [27]:

$$c = 331.4 + (0.6 \times T_C) \tag{2.8}$$

The widely accepted value for the temperature of dry air used in acoustic calculations is 20 °C, giving a speed of $343.4\,\mathrm{m\,s^{-1}}$. Unless otherwise stated, this value of c will be used throughout this thesis. In acoustics, generally the medium of concern is air, and it is usually sufficient to consider the medium as having constant temperature.

The range of pressures that the human auditory system detects as sound is approximately 20 µPa to 64 Pa [29]. This wide variation means that expressing pressure p

directly can be inconvenient. The logarithmic scale, sound pressure level (SPL), defined by IEC 801-22-07 [34] and measured in decibels (dB), addresses this by comparing p to a reference pressure near the lower limit of human hearing in air:

$$SPL = 20 \log \left| \frac{p}{p_o} \right| \tag{2.9}$$

where p is the sound pressure in Pa and p_o is the reference value of 20 µPa. Measurements expressed in dB relative to p_o are indicated using the notation dB SPL, with a value in air of 0 dB SPL at 1 kHz.

2.3 Sound Wave Behaviour

In order to model and simulate the propagation of sound, it is useful to understand the mechanisms of sound wave behaviour.

2.3.1 Propagation

The propagation of pressure waves through the medium can be represented by the motion of wavefronts, which are lines of constant pressure that move with time. The differential equation that governs the relationship between pressure and time, and therefore the propagation of the wave, is known as the wave equation [27]:

$$c^2 \nabla^2 p = \frac{\partial^2 p}{\partial t^2} \tag{2.10}$$

where

$$c^2 = \kappa \frac{P_0}{\rho_0} \tag{2.11}$$

and κ is the adiabatic exponent (for air, $\kappa = 1.4$), ρ_0 is equilibrium density in kg m^{-3} and ∇^2 represents the Laplacian operator. If the acoustical quantities are assumed to depend only on time and a single direction (for example, the x direction of a Cartesian coordinate system), Equation 2.10 can be written as [35]:

$$c^2 \frac{\partial^2 p}{\partial x^2} = \frac{\partial^2 p}{\partial t^2} \tag{2.12}$$

The general solution of this is:

$$p(x, t) = F(ct - x) + G(ct + x) \tag{2.13}$$

where F and G are arbitrary functions for which second derivatives exist [35]. F represents a wave travelling in the positive x direction with a speed c and G represents a wave travelling in the negative x direction. These terms are both referred to as plane waves, where the sound pressure p is constant in any plane perpendicular to the direction of propagation (in this case, the x direction). The pressure amplitude does not remain constant over distance, but decreases according to the inverse square law. The ratio between sound pressure p and particle velocity v_p in the direction of propagation is frequency independent, and is referred to as the characteristic impedance z_0:

$$z_0 = \frac{p}{v_p} = \rho_0 c \tag{2.14}$$

For dry air at 20 °C, the characteristic impedance has a value of 416 kg m^{-3} s^{-1} [35]. Equation 2.10 can also be rewritten using the spherical polar coordinates system by assuming the acoustic quantities depend only on radial distance r and not direction [35]:

$$\frac{\partial^2 p}{\partial r^2} + \frac{2}{r}\frac{\partial p}{\partial r} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} \tag{2.15}$$

the solution of which is:

$$p(r,t) = \frac{\rho_0}{4\pi r}\dot{Q}\left(t - \frac{r}{c}\right) \tag{2.16}$$

This represents a spherical wave produced by a point source at the origin ($r = 0$) with the volume velocity Q (the rate in m^3 s^{-1} at which fluid is expelled by the source), where $\dot{Q}$ represents partial differentiation of Q with respect to time [35].

The ratio of sound pressure and particle velocity in a spherical wave is complex-valued and depends both on the radial distance r and the angular frequency $\omega = kc$, where k is the wave number (the number of wavelengths per unit distance). For large radial distances when compared to the wavelength ($kr \gg 1$), the ratio is frequency-independent and real, and tends asymptotically to the characteristic impedance z_0 [27]. In this region, the wave can be considered as an approximation of a plane wave [35] and the sound source can be treated as a distant point source. Although a plane wave is an idealised wave type which does not exist in the real world, the region where waves behave approximately as plane waves is commonly referred to as the far-field.

For small distances when compared to the wavelength ($kr \ll 1$), the wavefronts are spherical [27]. This region is commonly referred to as the near-field. The distance at which the near-field transitions to the far-field is debated in the literature, as it depends on the frequency content of the source, but the majority of the literature agree on 1 m, as defined by Brungart and Rabinowitz [36]. Spherical wavefronts can, therefore, be thought of as tending towards plane waves as the distance increases [37].

2.3.2 Interaction

It is almost certain that, as sound waves propagate through a medium, an interaction with a boundary to another medium will occur. The result of the interaction between the wave and the boundary depends on properties of the propagation medium, properties of the boundary material and the frequency of the wave. If there is an impedance mismatch between the two mediums, some of the energy of the wave will be transmitted into the boundary material and the remainder will be reflected back into the propagation medium.

If the boundary is flat and any irregularities in the surface are considerably smaller than the wavelength of the incident wave, specular reflections will occur, where the angle of reflection is equal to the angle of incidence [37]. The alternative to this is diffuse reflections, where the angles of reflection and incidence are not equal, and occurs if the boundary has larger irregularities on its surface [38]. The shape of the boundary also has an effect on the reflections. If the boundary is parabolic in shape, the energy will be focussed to a precise point. A concave boundary produces a point of variable size depending on the shape. Reflections from convex surfaces scatter the energy, creating diffusion of the wave [37]. A sound wave that repeatedly encounters boundaries will be reflected and absorbed until the acoustic energy has dissipated and turned into heat.

An interaction at the edge of the boundary or around an object will cause diffraction of the wave, as the variations in pressure cannot step abruptly to zero after passing an edge [30, 39]. The wave appears to bend, and this enables sound to still be heard even when the receiver has an obstructed path to the source. The amount of diffraction (that is, how far round the corner the sound can still be heard) depends on the wavelength of the sound wave compared with the size of the object. In general low frequencies tend to diffract more than higher frequencies which behave more like a ray [28, 30].

2.3.3 Superposition

Sound waves in air obey the principal of linear superposition, where waves combine such that the pressure at any given point is the sum of the pressures of the individual wave components at that point [30]. Depending on the relative phase between each of the components, constructive and destructive interference occurs, where the total peak pressure is greater than any single component and less than any single component, respectively. If two superimposed waves are of the same frequency and amplitude but are of opposite phase, the total pressure will always be zero.

Linear superposition can also be used to explain a phenomenon referred to as standing waves, which occur as a result of the interaction between a sinusoidal sound wave and two reflective boundaries. If there is a relationship between the distance of half a wavelength and the distance between the two boundaries, the wave repeatedly traces the same path as it travels between the two boundaries, resulting in a wave that appears to be stationary [30]. The superposition of these waves, travelling in opposite directions and thus with opposite phase, results in pressure nodes (where the amplitude is zero) and antinodes (where the amplitude is at a maximum). As the relationship between the wave and the boundaries is based on half-wavelengths, any multiple of half-wavelengths will cause the production of a standing wave. Theoretically, an infinite set of frequencies exist at which standing waves can occur. Standing waves can also occur between other arrangements of boundaries, including open and closed tubes, and between more than one pair of boundaries. The only requirement is that the path taken by the wave is a multiple of a half-wavelength. When the boundaries are walls in a room, standing waves are referred to as room modes.

2.4 Fundamentals of the Auditory System

As well as understanding the fundamental properties of sound behaviour, it is helpful for the purposes of this thesis to have a rudimentary understanding of how the human auditory system processes incoming sound waves, both physiologically and perceptually.

2.4.1 Physiology

The anatomical regions of the human auditory system are illustrated in Fig. 2.8. The auditory system consists of three anatomical regions, each with a different role:

- the outer (or external) ear, comprising the pinna, concha, ear canal and eardrum;
- the middle ear, comprising the malleus, incus and stapes bones (the ossicles) and the oval and round windows;
- the inner ear, comprising the cochlea and the auditory nerve.

Two other anatomical features are included in Fig. 2.8. The semicircular canals, whilst part of the inner ear region, are involved with balance and, therefore, are part of the vestibular rather than the auditory system. The Eustachian tube, which connects the middle ear to the throat and nasal cavity, controls the pressure within the middle ear to ensure it is equal to the pressure outside of the body. Of primary interest in this thesis is the function of the outer ear, although the functions of the middle and inner ear regions will be discussed briefly for completeness.

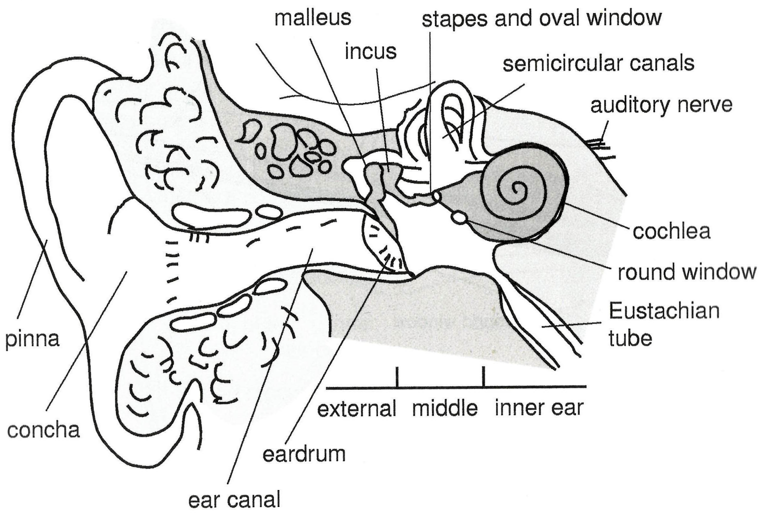


Figure 2.8: The anatomy of the auditory system. Reproduced from [40] with permission.

The primary function of the outer ear is to direct sound waves towards the eardrum, and to aid in determining the position of a sound source as the folds in the pinna variously amplify or attenuate some frequencies with respect to others as a function of source direction. This aspect is discussed more in Section 2.5.4. The ear canal is approximately 22.5 mm in length, and acts as a quarter-wavelength resonating tube to amplify the frequencies relevant to speech [40]. The eardrum is a conical elastic structure which converts the incoming pressure variations into mechanical vibrations. It forms the boundary between the outer and middle ears. The eardrum comprises three layers: the outer layer which continues the skin from the inside of the ear canal, the central layer (the umbo) which deflects under force of the sound waves, and the inner layer which continues the mucous lining of the middle ear [29, 41].

The middle ear acts as an impedance transformer to reduce the impedance mismatch between the air-filled ear canal and the fluid within the cochlea [41]. Three small bones (the ossicles) connect the eardrum to the membrane at the oval window (part of the boundary between the middle and inner ears). The lever moment of the bones about their axes of rotation converts airborne sound waves into mechanical movements needed for the inner ear [29].

The cochlea is a snail-like structure which sits within the inner ear, the function of which is to convert the mechanical vibrations at the oval window to nerve firings to be processed by the brain. A membrane within the cochlea (the basilar membrane) is displaced in response to the mechanical vibrations, and performs a frequency analysis as a consequence of its structure. It is roughly wedge-shaped, being narrow and thin at the base and wider and thicker towards the apex. As a result, the basilar membrane is displaced at different points along its length in response to waves of different frequencies. The basilar membrane can be thought of as a bank of band-pass filters placed alongside each other, where each filter responds to a different frequency region, depending on its position [42]. This is known as the *place theory of hearing*.

The displacements are then transformed into electrical signals through hair cells which trigger nerve firings when bent [29, 41]. The nerves from the hair cells form the auditory nerve, which connects the cochlea to the cochlea nucleus in the brainstem.

2.4.2 Frequency Resolution

The displacement profile of the basilar membrane in response to a specific frequency does not occur at an infinitely-small region. Some of the membrane on either side of that point is also displaced, as shown in Fig. 2.9. The amount of displacement increases as the amplitude of the input sound wave increases, and the displacement extends further on the higher frequency side than on the lower frequency side [43]. This limits the frequencies that can be detected above and below the original stimulus frequency, with different phenomena occurring depending on the relationship between the frequencies present.

When the frequency between two simultaneously-detected sounds is less than approximately 12.5 Hz, a phenomenon known as *beating* occurs where one sound is detected which appears to fluctuate in loudness [29]. As the difference between the two frequencies increases, the sensation of beats gives way to *flutter*, then *roughness*, and eventually two distinct smooth tones can be detected [29, 44]. These phenomena are due to the interaction of displacement peaks on the basilar membrane. Two separate sounds can be detected when two distinct peak displacements form rather than a single peak, and the perceived roughness is a result of interaction between the two peaks. The frequency difference at which two separate smooth tones can be heard is referred to as the critical bandwidth, which varies with frequency [45].

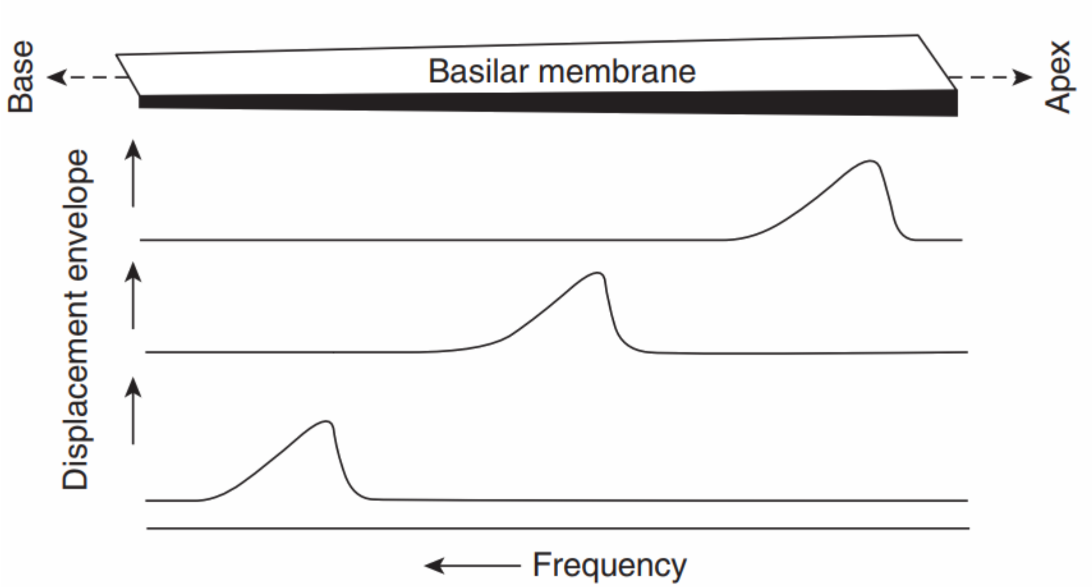


Figure 2.9: The idealised displacement of the basilar membrane at three frequencies, where higher frequencies are towards the left. Reproduced from [29] with permission.

2.4.3 Sensitivity

The human auditory system is not equally sensitive at all frequencies or amplitudes, and has a lower limit on the change in a stimuli that can be detected.

2.4.3.1 Sensitivity to Frequency

The frequency range of the human auditory system is often quoted as 20 Hz to 20 kHz. However, this varies between individuals and deteriorates with age. A young child may be able to hear up to 20 kHz; a 20-year old may have a reduced upper range of 16 kHz; in the elderly this upper limit may have reduced to as little as 8 kHz [29]. However, the auditory system is not equally sensitive to all frequencies. The threshold of hearing is defined as the minimum sound pressure variation able to be detected by the auditory system for a given frequency, as first measured by Fletcher [42] and later defined in ISO 226:2003 [46]. The threshold of hearing is lower (that is, signals of lower amplitude can be detected) in the middle of the frequency range than at lower or higher frequencies, as shown in Fig. 2.10. The lowest threshold occurs at approximately 4 kHz (chiefly due to the ear canal resonance [29]), and the thresholds of hearing at 100 Hz, 1 kHz and 10 kHz are approximately 20 dB, 0 dB and 10 dB, respectively [29]. The threshold of hearing increases sharply above approximately 14 kHz, with a number of possible causes proposed. These include inefficient acoustic energy transmission to the inner ear [47, 48], decreasing sensitivity of the basilar membrane to high frequencies [47], and the characteristics of the highest-frequency auditory channel of the cochlea [47, 49].

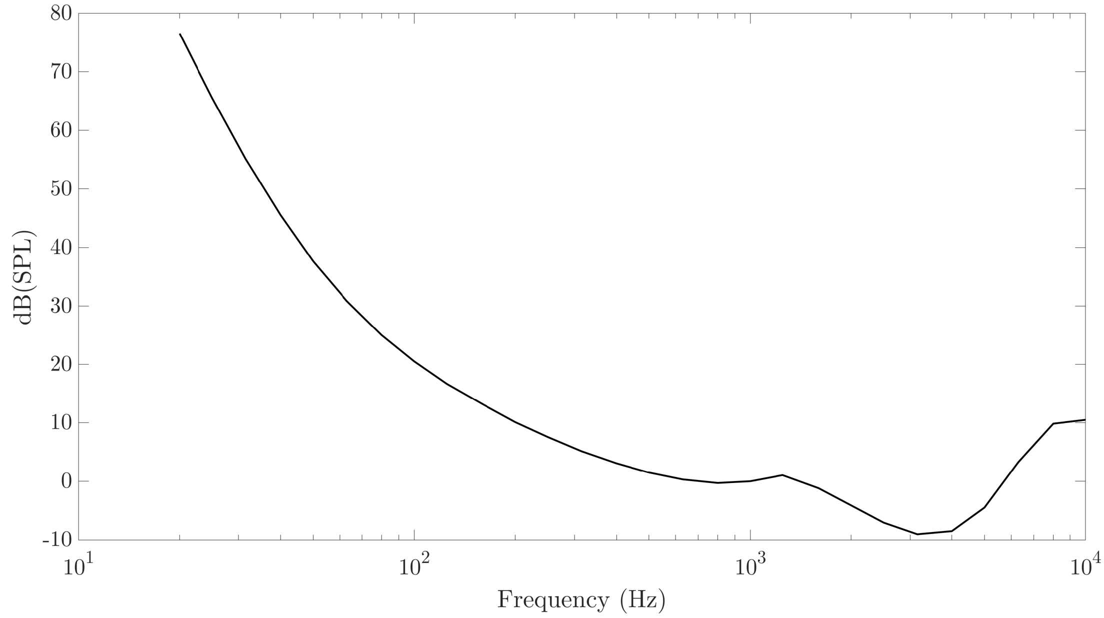


Figure 2.10: The threshold of hearing between 20 Hz and 10 kHz, as defined in ISO 226:2003 [46].

2.4.3.2 Sensitivity to Amplitude

The perceived amplitude of a sound (the loudness) can be expressed using the phon scale. The number of phons is equal to the sound pressure level in dB of a tone at a frequency of 1 kHz which sounds equally loud. For example, a sound which is perceived to be equal in loudness to a 1 kHz tone with an SPL of 30 dB SPL has a loudness of 30 phons. Tones at other frequencies which appear equally loud will, in general, have different sound pressure levels. Equal loudness contours, also defined in ISO 226:2003 [46] and shown in Fig. 2.11, map the dB SPL of a sound source at a given frequency to the perceived loudness in phons. The contours are a measure of sound pressure, such that a listener perceives a constant loudness across tones of varying frequency. From this, the amplitude limits of the auditory system are 0 phons (the threshold of hearing) to approximately 120 phons (the threshold of pain).

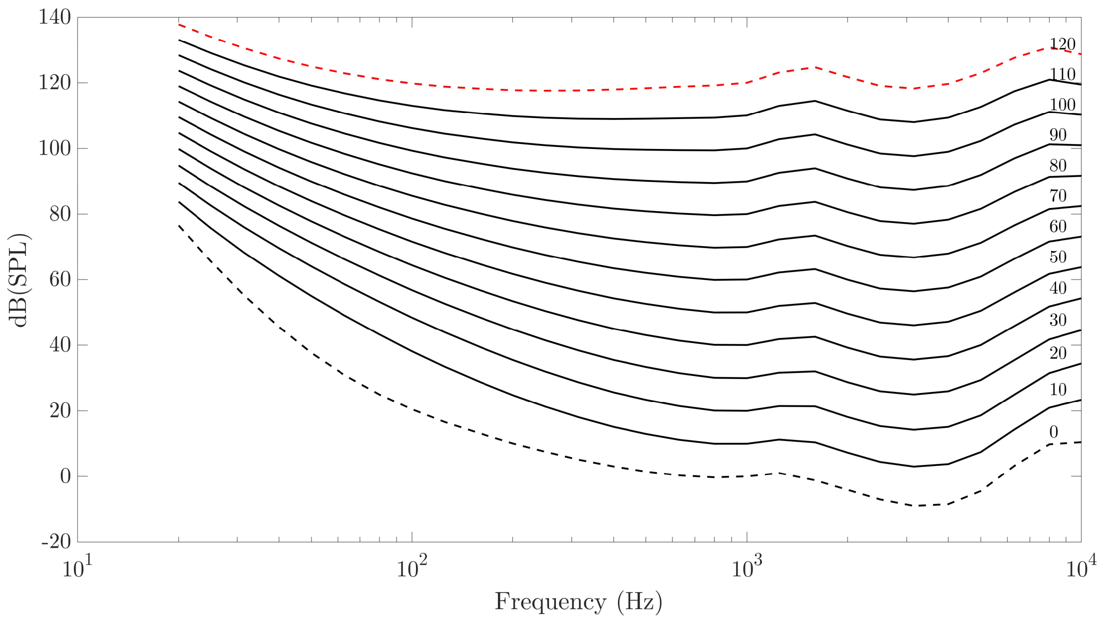


Figure 2.11: Equal loudness contours for phon levels of 0 to 120 in 10 phon steps, as defined in ISO 226:2003 [46]. 0 phons (indicated by the dashed black line) is equal to the threshold of hearing. 120 phons (indicated by the dashed red line) is equal to the threshold of pain.

2.4.3.3 Sensitivity to Change in Stimuli

Upper and lower thresholds of audibility for both frequency and amplitude exist for the human auditory system, but these thresholds do not describe the sensitivity of the auditory system to changes in stimuli. Ernst Weber observed in the 19th century that the amount of change required before that change could be detected was proportional to the initial magnitude. With Gustav Fechner, Weber established Weber's Law (sometimes referred to as the Weber-Fechner Law or the Weber fraction) [32]:

$$W = \frac{\Delta x}{x} \tag{2.17}$$

where Δx is the amount of change required in some stimulus value, x is the smaller of the two values being discriminated, and W is a constant which varies with the modality being studied, referred to as the value of the Weber fraction. More recently, the change in stimulus is also referred to as different threshold or the just-noticeable difference (JND) threshold. JND thresholds vary between individuals, and so values in the literature are often calculated as an average across the sample of the population tested. Difference thresholds are typically determined using a two-interval, two-alternative forced-choice listening test paradigm using a correct response rate of 50 % or 75 % [50]. Despite this, values vary across the literature.

With regards to discrimination of a change in frequency Δf, the variation of the Weber fraction changes between low and high frequencies. Zwicker *et al.* [51] reported JNDs for pure tones of 3.5 Hz below a lower frequency of 500 Hz, which then increased proportionally with respect to frequency at a Weber fraction of approximately 0.07. However, more recently the Weber fraction has been shown to be generally independent of frequency below approximately 2 kHz, increasing with frequency above this range [32, 52]. Frequency discrimination is also dependent on the amplitude and duration of the test signal, however, and requires extension of Weber's Law to include these additional dependencies [53]. Wier *et al.* [54] determined JNDs for sinusoids of varying amplitude, and reported values ranging between 1 Hz (for a tone at 200 Hz with a sound pressure level of 80 dB SPL) to 100 Hz (for a tone at 10 kHz with a level of 5 dB SPL). Moore [55] determined JNDs for sinusoids of varying duration, and reported values ranging from 500 Hz (for an 8 kHz tone of 10 ms duration) to 0.7 Hz (for a 250 Hz tone of 200 ms duration).

For amplitude discrimination, the threshold ΔI decreases steadily as the magnitude of the comparison increases [32, 43] and has a strong dependency on the type of signal, the frequency, and the amplitude [29]. The JND for broad-band noise is generally between 0.5 to 1 dB for signals between 20 to 100 dB SPL [29, 32, 56]. For pure tones, the JND is at its lowest in the frequency region 1 to 4 kHz. However, it has been suggested that once more than a few harmonics are present, the JND tends toward the broadband case rather than the pure tone case [29].

2.5 Acoustic Cues for Localisation of Sound Sources

The aspects of the auditory system discussed so far can be described as *monaural* abilities, in that the auditory system only requires a single input signal. However, a key aspect of interest in this thesis is the localisation ability of the auditory system, i.e. the ability to determine the direction of a sound source within the 3D space around the listener. This ability requires both monaural and binaural (using two ears) listening.

The acoustic localisation cues can be grouped into two categories: interaural difference cues which originate from the differences between the signals that arrive at the left and right ears (predominantly time and level, although there has also been some evidence of the additional role of interaural spectral difference (ISD) in localisation [57, 58]), and spectral cues which originate from the direction-dependent spectral filtering applied to the incoming sound by the outer ear, head and torso [1].

Whilst both groups contribute to localisation, the interaural difference cues are sometimes referred to as binaural cues (as the signals at both ears are required in order to determine the difference between them), whereas spectral cues are sometimes referred to as monaural cues (as they are still evident when a signal is present at only one ear). However, the term *binaural* can also be used in a more general sense to refer to reproduction systems which make use of both interaural and spectral cues. Therefore, for clarity, the acoustic localisation cues will be referred to as *interaural difference* cues and as *monaural spectral* cues.

At this stage, it is useful to be able to refer individually to each side of the head, and to refer to the head in relation to a point of interest. The terms *ipsilateral* and *contralateral* are used to refer to the side closer to and further from a point of interest, respectively.

2.5.1 Interaural Time Difference

2.5.1.1 Theory

The difference in time between the arrival of the sound at the two ears, referred to as the interaural time difference (ITD), occurs as a result of the increase in distance required for a sound to reach the contralateral ear [59]. In the median plane, ITD is approximately zero, as the distances to the two ears are similar. When the sound source deviates from the median plane, a non-zero ITD occurs. This is illustrated geometrically in Fig. 2.12, where plane waves from a sound source located to the left of the listener reach the left ear before the right ear. An additional distance d (indicated in red) is required to reach the contralateral ear:

$$\begin{aligned} d &= a\theta + a\sin\theta && (2.18) \\ &= a\,(\theta + \sin\theta) && (2.19) \end{aligned}$$

where a is the radius of the head in m and θ is the azimuthal incident angle (°) of the plane wave relative to the frontal direction. The time taken to travel the extra path distance (and, therefore, the ITD) is:

$$ITD = \frac{a}{c}(\theta + \sin\theta), \qquad 0 \leq \theta \leq \frac{\pi}{2} \tag{2.20}$$

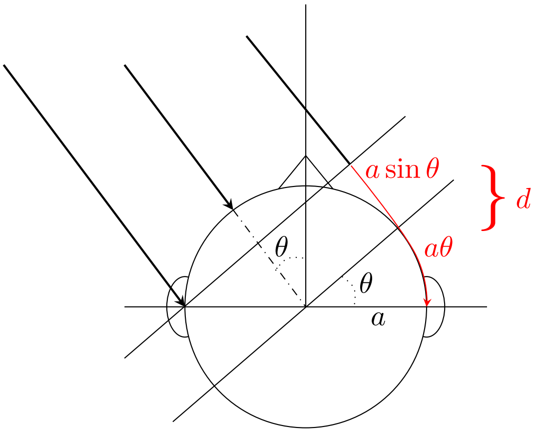


Figure 2.12: The difference in path length d (indicated in red) between the ipsilateral and contralateral signal paths creates an interaural time difference of arrival.

Equation 2.20 is known as the 'Woodworth' or 'Woodworth-Schlosberg' formula, after publication in textbooks of the same name under both authors [60, 61]. Fig 2.13 plots ITD, calculated using Equation 2.20, for the horizontal plane in the left hemisphere ($0° \leq \theta \leq 180°$) with $a = 8.75\,\text{cm}$, where a positive ITD indicates arrival at the left ear first. 8.75 cm is the standard head radius used across the literature, after Hartley and Fry [62] calculated the average value of 'a number of individuals' in the 1920s. Unless otherwise stated, this value of a will be used throughout this thesis. The maximum ITD (655 µs) occurs in this simple model at $\theta = 90°$. As Equation 2.20 is only valid for values of θ between zero and $\frac{\pi}{2}$, values from the frontal quadrant have been mirrored for the rear quadrant. ITD, therefore, decreases with sound sources approaching the rear, returning to zero at $\theta = 180°$.

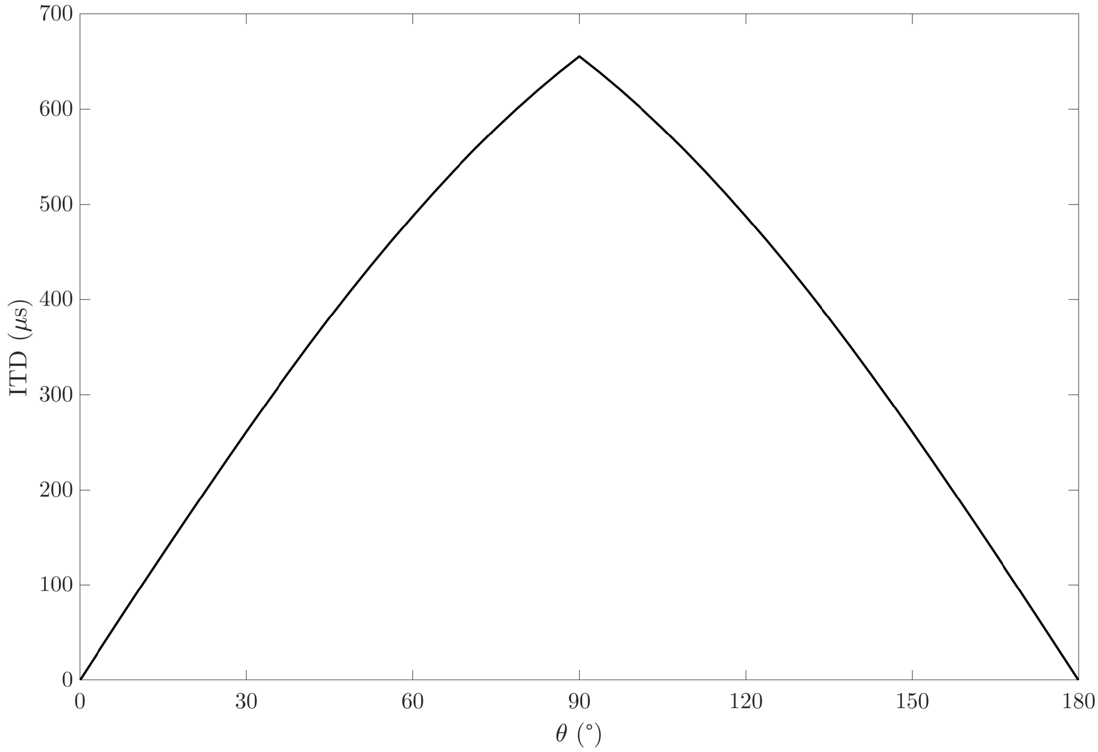


Figure 2.13: ITD calculated in the left hemisphere on the horizontal plane using Equation 2.20. The maximum ITD (655 µs) occurs at $\theta = 90°$. Values for the frontal quadrant have been mirrored for the rear quadrant.

For a sinusoidal tone, an ITD can be expressed as a difference in phase between the two ears, the interaural phase difference (IPD) [63]:

$$IPD = \frac{\psi_L - \psi_R}{2\pi f} \tag{2.21}$$

where ψ_L and ψ_R represent the phase (°) of sinusoidal signals at the left and right ears, respectively. For example, for a 500 Hz tone (period of 2000 µs), an ITD of 250 µs is equal to an IPD of 45° (an eighth of a full cycle).

However, IPD (and, therefore, ITD) has an effective frequency range. High-frequency stimuli produce an ambiguous localisation cue, as the phase difference between the ears is greater than one phase cycle. For example, a 5 kHz tone has a period of 200 µs, and so an ITD of 250 µs is equivalent to an IPD of 1.25 cycles. The auditory system cannot determine between multiples of phase cycles [59], and so cannot differentiate between a phase difference of 90° (0.25 of a cycle) and 450° (1.25 of a cycle). The ambiguous nature of IPD starts to occur when the period of the stimulus is approximately twice the maximum possible ITD, that is, when the path difference d is approximately half a wavelength [1, 59].

Fig. 2.14 demonstrates this, where red (dashed line, right y-axis) indicates the difference in path length between the left and right ear signals (calculated using Equation 2.18), and black (full line, left y-axis) indicates the frequency at which the difference is equal to half a wavelength. The lowest frequencies occur for sound sources in the more lateral directions. For a sound source at 90°, the path difference is 22.5 cm and the corresponding half-wavelength frequency is 760 Hz. As the sound source approaches the median plane, the half-wavelength frequency increases.

Head or sound source movement can help to resolve the phase ambiguities, but the differences become highly ambiguous above 1.5 kHz [59, 64]. Therefore, the practical maximum effective frequency of ITD is approximately 700 to 800 Hz.

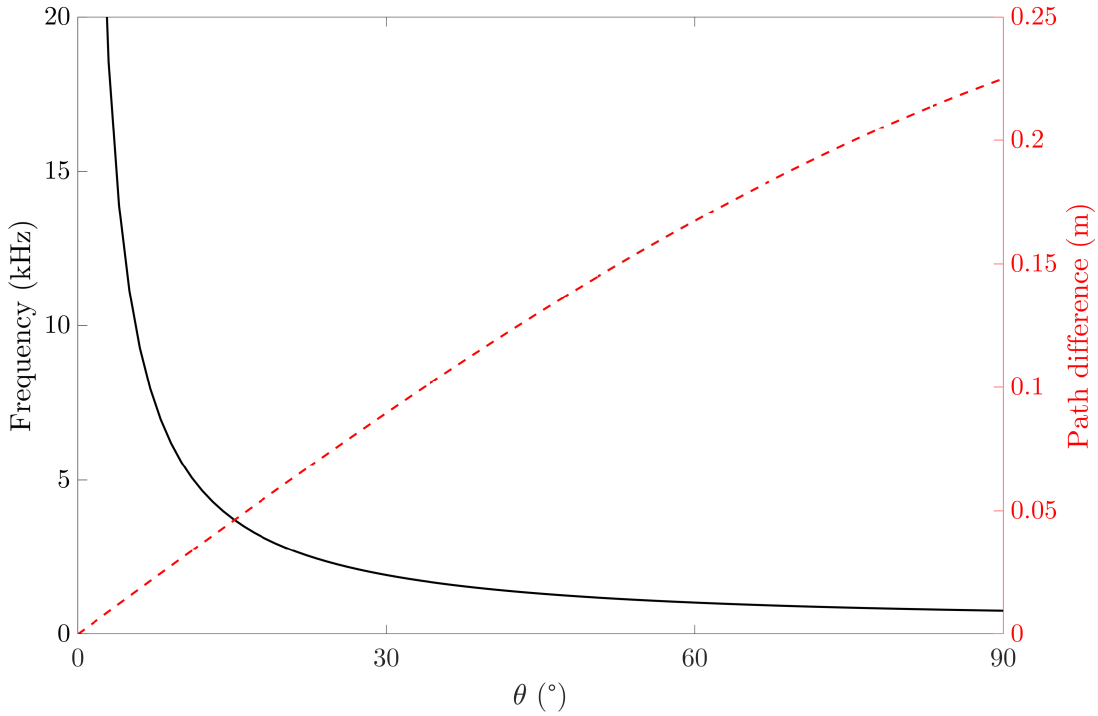


Figure 2.14: The difference in path length (red, right y-axis) at azimuth angles in the left frontal quadrant, and the corresponding frequency at which the distance equals half a wavelength (black, left y-axis). The lowest frequency corresponds to the most lateral sound source (760 Hz at $\theta = 90°$).

2.5.1.2 Estimation of Real World ITD

The equations discussed here are based on a geometric approximation of the head in the horizontal plane, and, therefore, do not accurately describe the full complexity of real world ITD variation. ITD can also be estimated from signals captured at the ears of subjects (methods for capture are discussed further in Section 2.6). There are three groups of methods used to estimate ITD from such signals, which all produce slightly different absolute values but which vary with azimuth in a similar way, as discussed by Andreopoulou and Katz [65]:

- Onset detection: the onset in each signal is selected as the first point which exceeds a threshold relative to a pre-determined level, and the time delay is calculated between the onsets (for example, as used by Kuhn [66] and Algazi *et al.* [67]);

- Cross-correlation: the time delay is determined by the lag at which the maximum value of coherence between the two signals is obtained (for example, as used by Kistler and Wightman [68] and Middlebrooks and Green [69]);
- Group delay: the phase difference between the two signals is estimated and converted to time delay (for example, as used by Jot *et al.* [70] and Minnaar *et al.* [71]).

The cross-correlation method was selected for use in this thesis, since physiological studies have reported that neural responses are linked to the correlation between binaural signals [72]. This method estimates the ITD as the time delay which produces the maximum coherence of one signal with respect to the other, which is performed using the interaural cross-correlation (IACC) function [73]:

$$ITD = \text{argmax}[IACC(\theta, \tau)], \qquad s.t.\, |\tau| < 1\,\text{ms} \tag{2.22}$$

$$IACC(\theta, \tau) = \frac{\int_{t_1}^{t_2} e_L(\theta, t) e_R(\theta, t + \tau) dt}{\sqrt{\int_{t_1}^{t_2} e_L^2(\theta, t) dt \int_{t_1}^{t_2} e_R^2(\theta, t) dt}} \tag{2.23}$$

where e_L and e_R are the time domain signals at the left and right ears, respectively, τ is the time shift in s, and the integration limits t_1 and t_2 are defined as zero and the maximum length of e_L and e_R, respectively. Equation 2.22 determines the time delay τ for which the two signals are most similar, where the magnitude of the IACC is the measure of similarity.

As ITD is an effective cue below approximately 800 Hz, signals are typically low-pass filtered before calculating the IACC. Fig. 2.15 shows ITD estimated for sound sources in the horizontal plane (in 1° increments at $r = 1.2$ m) using the cross-correlation method below 800 Hz (data from [74]). A positive ITD indicates that the left signal arrives first, a negative ITD indicates that the right signal arrives first. The largest absolute value of 735 µs occurs at $\theta = 270°$.

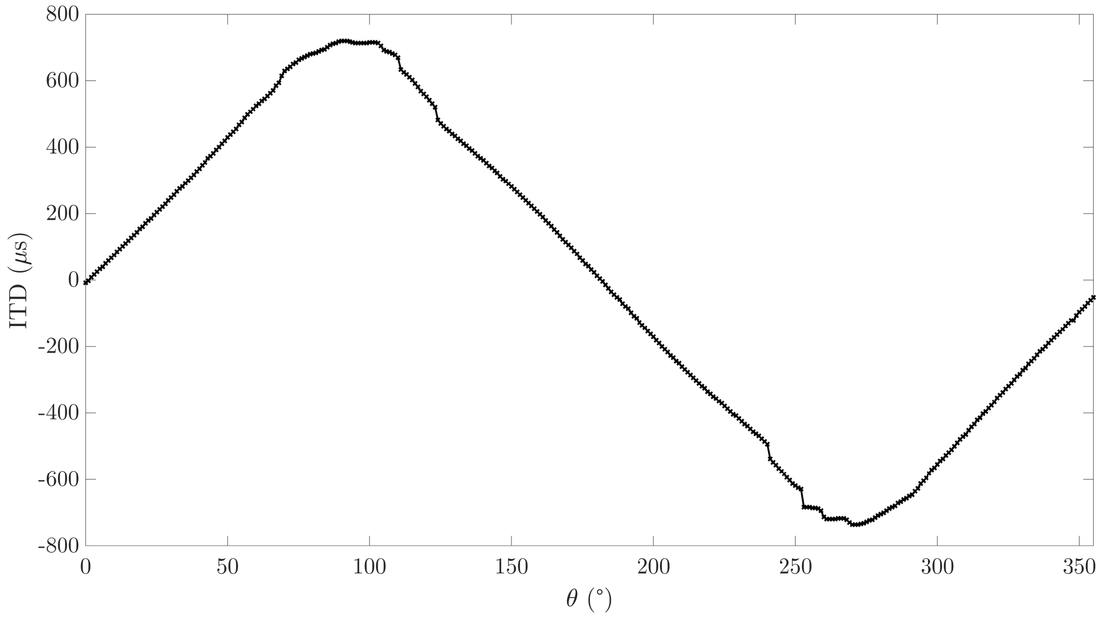


Figure 2.15: ITD estimated for sound sources in the horizontal plane (1° increments at $r = 1.2\,\mathrm{m}$) using the cross-correlation method below 800 Hz (data from [74]). The largest absolute value of 735 µs occurs at $\theta = 270°$.

2.5.2 Interaural Level Difference

2.5.2.1 Theory

The difference in sound pressure level between the sounds arriving at each of the two ears, referred to as the interaural level difference (ILD), is caused predominantly by the shadowing effect of the head [59]. This effect is illustrated in Fig. 2.16, where dotted lines indicate the shadowed region. ILD is frequency dependent; as with ITD, this is determined by the size of the human head. As wavelengths reduce below the approximate diameter of the head, the shadowing effect increases. At wavelengths greater than the size of the head (this corresponds to a frequency of less than approximately 2 kHz), the head becomes increasingly transparent. This results in a high frequency range of effectiveness for ILD, rather than a low frequency range as for ITD [1, 59].

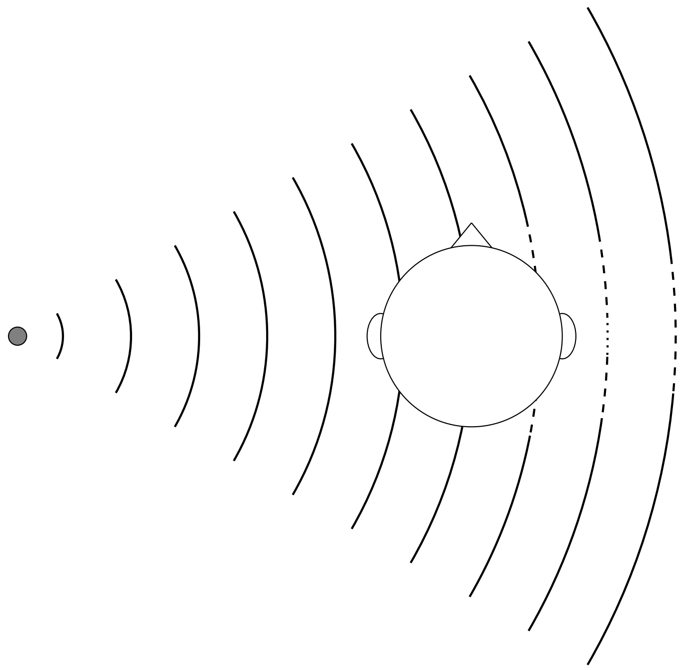


Figure 2.16: The head introduces an acoustic shadowing effect (indicated with dotted lines) which creates a difference in sound level between the two ears.

Xie [64] estimated ILD as a function of both frequency and azimuth angle using a spherical head model (discussed further in Section 2.6.2.1). Values ranged from 0.5 dB at low frequencies (approximately 300 Hz) up to a maximum of 17.4 dB at high frequencies (approximately 5 kHz). ILD increased with both azimuth angle and frequency, though the increase in ILD was not monotonic. The maximum value of ILD did not appear at the most extreme lateral source position due to constructive interference of the acoustic paths around the head creating an acoustic bright spot at the contralateral side. ILD has been shown to not act as an effective localisation cue until it varies with source direction in a stable manner, from approximately 1.5 kHz upwards [1, 59].

Practically, ILD is calculated as the arithmetic mean of the difference between the frequency spectra of the left and right ear signals ($E_L(f)$ and $E_R(f)$, more simply, E_L and E_R) within some frequency region of interest:

$$ILD_f = \frac{1}{N_f} \sum_{j=1}^{N_f} 20 \log_{10} \left| E_{L_j} - E_{R_j} \right| \tag{2.24}$$

where N_f is the number of points within the frequency band f. The ILD at a given single frequency point (that is, Equation 2.24 at a single value of f) can also be referred to as the interaural spectral difference (ISD).

2.5.2.2 Estimation of Real World ILD

As with ITD (see Section 2.5.1), ILD can be estimated from signals captured at the ears of subjects (methods for capture are discussed further in Section 2.6). This is done using Equation 2.24.

As ILD is a localisation cue predominantly above 1.5 kHz (in the far-field, discussed in Section 2.5.2), the data are often high-pass filtered before taking the FFT, or the lower portion of the frequency information is discarded after taking the FFT. The cut-off frequency is most typically around 1.5 kHz, and ILD calculated in this way is referred to here as $\text{ILD}_{>1.5kHz}$. When the entire audio frequency range is used (that is, including frequencies below 1.5 kHz), the ILD is referred to as *wideband*. Wideband ILD is less useful for describing localisation accuracy, but can be useful for describing spectral features within the signals [75].

Fig. 2.17 shows ISD in black (2048 frequency points) and ILD_f in blue (40 linearly-spaced frequency bands with the boundaries indicated by vertical lines) over 200 Hz to 20 kHz estimated for an acoustically measured source sound at $\theta = 45°$, $\phi = 0°$, $r = 1.2$ m (data from [74]). The wideband ILD (20 Hz to 20 kHz) is 15.1 dB and the ILD value between 1.5 kHz and 20 kHz ($\text{ILD}_{>1.5kHz}$) is 16.0 dB. An increase in both ISD and ILD_f with frequency is apparent, up to maximum values of 24.0 dB and 23.4 dB, respectively.

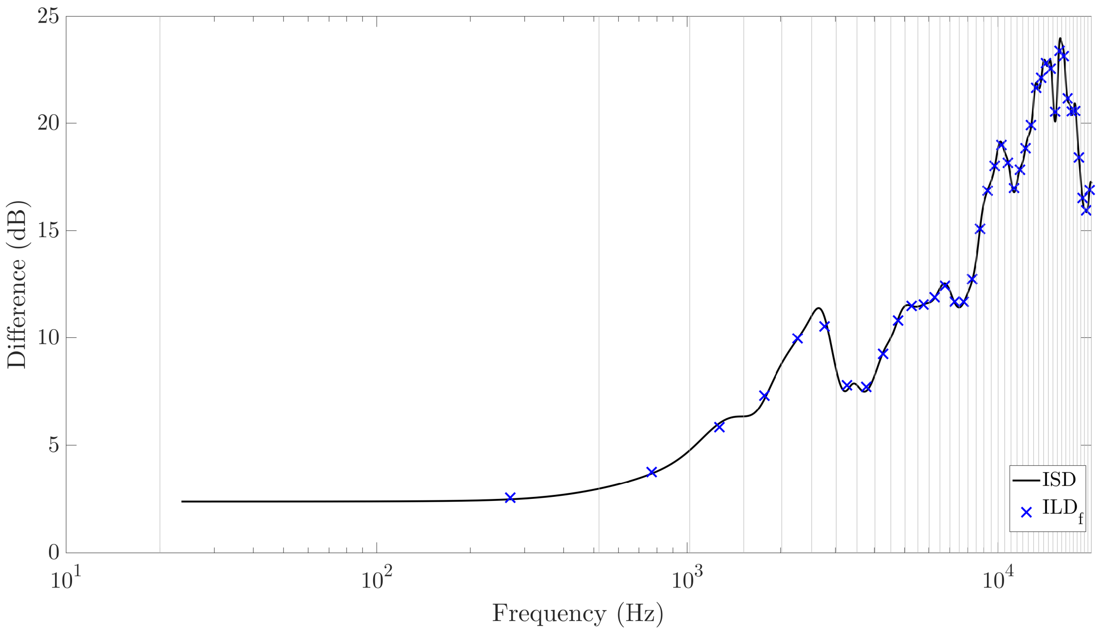


Figure 2.17: ISD and ILD_f estimated for 40 linear frequency bands for an acoustically measured sound source at $\theta = 45°$, $\phi = 0°$, $r = 1.2$ m (data from [74]). Frequency band limits (vertical lines) are indicated. An increase in both ISD and ILD_f with frequency is apparent.

$\text{ILD}_{>1.5kHz}$ for sound sources in the horizontal plane (1° increments at $r = 1.2$ m, data from [74]) is plotted in Fig. 2.18, with the value corresponding to Fig. 2.17 indicated in blue. $\text{ILD}_{>1.5kHz}$ increases as the sound source moves away from the median plane. The minimum value is 1.4 dB ($\theta = 0°$ and $\theta = 180°$), the maximum value is 23.9 dB ($\theta = 270°$). There appear to be somewhat anomalous values at the most lateral directions.

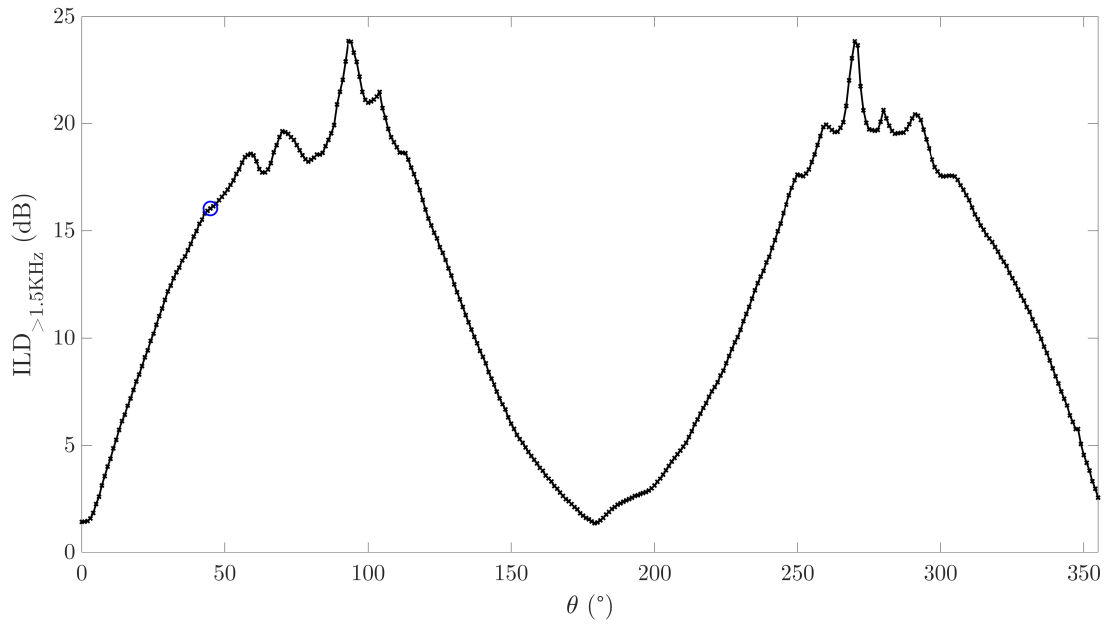


Figure 2.18: ILD$_{>1.5kHz}$ estimated for sound sources in the horizontal plane (1° increments at $r = 1.2\,\mathrm{m}$) (data from [74]). The value corresponding to Fig. 2.17 is circled in blue.

2.5.3 Resolving Source Directions with Identical Interaural Cues

As can be identified in Fig. 2.15 and Fig. 2.18, there exist multiple positions in auditory space for which the ITD and ILD cues are similar, that is, where the ITD at some point A is the same as that at a different point B, and the ILD at point A is also the same as that at point B. These positions lie on a surface, and form what are referred to as *cones of confusion*[2] extending outwards from the ear, where any sound source located on the cone of confusion gives rise to the same interaural differences as any other position on the cone[3] [76]. An extreme case of a cone of confusion is the median plane. Attempts to localise sound sources on these cones can result in front-back or up-down confusion errors. This is illustrated in Fig. 2.19, where sound sources located at points A and B can be mistaken for each other (a front-back confusion) due to their identical ITD and ILD cues. Rotation of the head helps to resolve these ambiguities, as the paths between the sound source and each ear change differently depending on the rotation [77]. For a sound source located at point A, a head rotation to the left would increase the path length between A and the left ear, and decrease the path length between B and the right ear. For a sound source located at point B, the converse is true.

[2]More technically, a cone of confusion is a hyperbolic surface of isometric ITD and ILD.

[3]This is true for sound sources located in the far-field, but does not hold for near-field sources due to distance-dependent variation. This is elaborated on in Section 2.5.5.

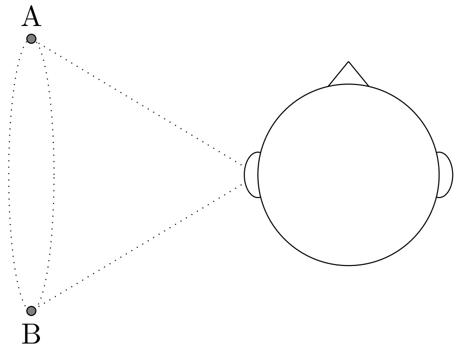


Figure 2.19: Sound sources at points A and B can be mistaken for each other (so called front-back confusion), as they both lie on the indicated cone of confusion.

2.5.4 Spectral Cues

Interaural differences are not the only source of sound localisation cues. The elevation localisation of the auditory system far exceeds that which can be predicted solely from interaural difference cues and head movements [59]. Additionally, the human auditory system can localise sound sources within a frequency region where neither interaural cue is particularly effective (as ITD is most effective below approximately 800 Hz and ILD is most effective above approximately 1.5 kHz). The dual nature of the interaural difference cues was first proposed by Lord Rayleigh in 1907 [78] and is now known as the duplex theory. To more fully describe the localisation ability of the auditory system, a discussion of spectral cues is required. Sound waves interact with the torso, head and ears, causing spectral filtering to occur before the waves reach the eardrums. The reflection and diffraction effects within the folds of the pinnae (see Section 2.4.1) are particularly influential. The filtering is direction-dependent (and, close to the head, distance-dependent), and so creates a different transfer function for each source direction in space. This provides the auditory system with additional information, particularly with regards to resolving elevation [79, 80] and front-back confusion [75]. It is important to note that the filtering is also ear-dependent, meaning that each listener has their own unique set of these filters.

The filtering for a given direction is described in the time domain by a head-related impulse response (HRIR), an example of which for both ears for a sound source at $\theta = 45°,\ \phi = 0°,\ r = 1.2\,\mathrm{m}$ is shown in Fig. 2.20. The frequency domain equivalent of the HRIR is the head-related transfer function (HRTF). The HRTFs corresponding to the HRIRs in Fig. 2.20 are shown in Fig. 2.21. The Fourier transform can be used to interchangeably express the signal in the time and frequency domains. When taken in pairs (and, therefore, maintaining the relationship between the left and right ear signals), HRIRs also encapsulate the time and level interaural difference cues. For clarity, in this thesis the terms *HRIR* and *HRTF* when used in the singular refer to a single head-related filter (for example, to refer to the left channel of a left-right pair). To refer to a left-right pair of such filters for a given source direction, the terms *HRIR pair* and *HRTF pair* will be used.

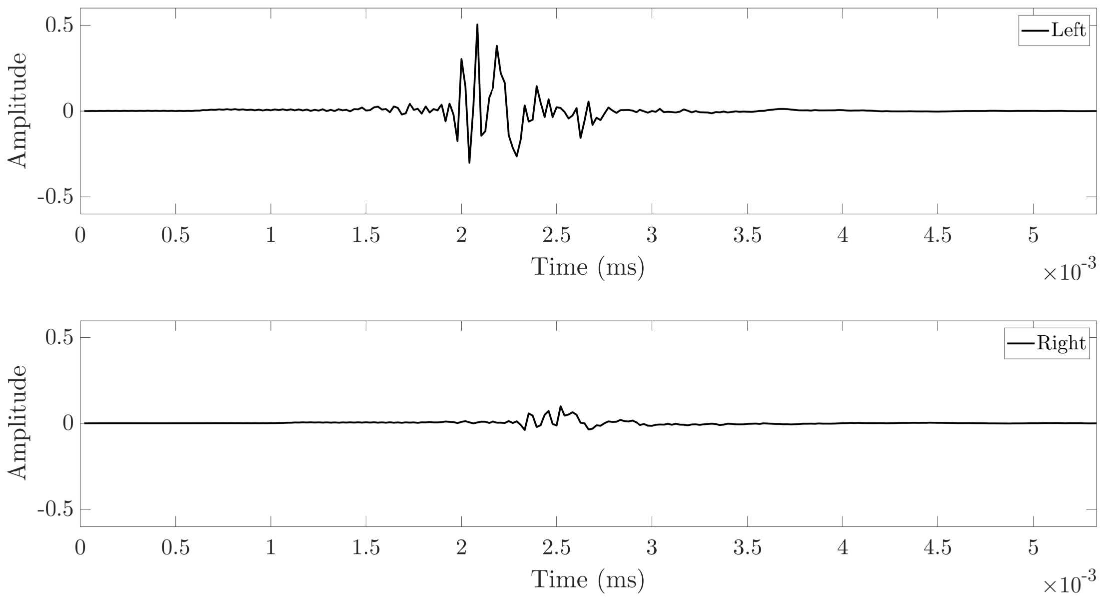


Figure 2.20: An example HRIR pair measured for a sound source at $\theta = 45°$, $\phi = 0°$, $r = 1.2\,\mathrm{m}$ by Armstrong *et al.* [74]. Top: left HRIR, bottom: right HRIR.

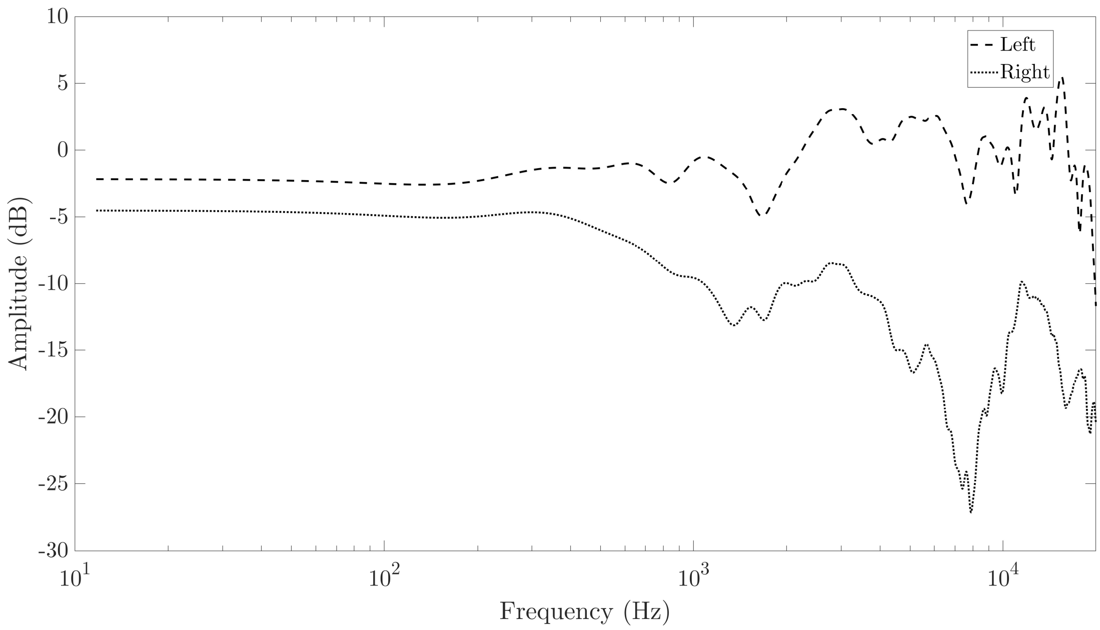
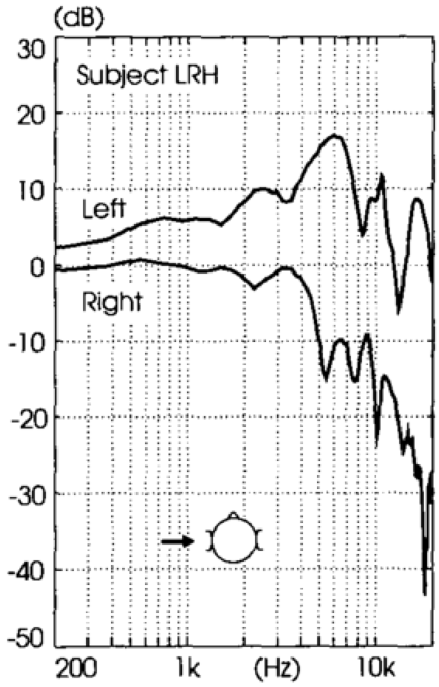


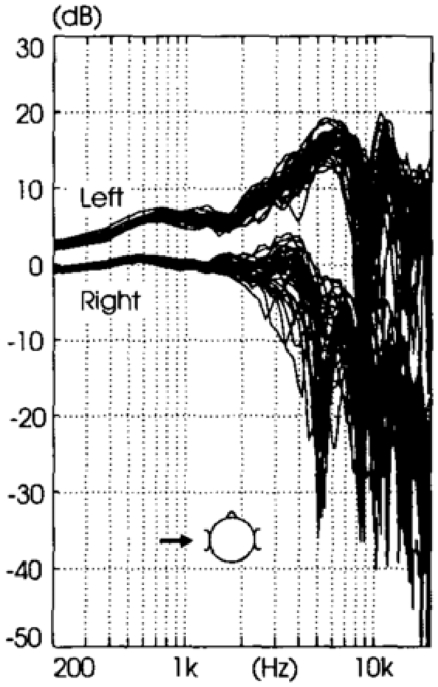
Figure 2.21: An example HRTF pair measured for a sound source at $\theta = 45°$, $\phi = 0°$, $r = 1.2\,\mathrm{m}$ by Armstrong *et al.* [74].

The features in HRTFs are directly related to the morphology of the pinnae, head, shoulders and torso. Each of these anatomical parts contributes differently to the features within a HRTF, e.g. to the peaks and notches visible in Fig. 2.21. The head and torso tend to have acoustic effects at lower frequencies, particularly the introduction of reflections due to the shoulders [79, 81, 82] and shadowing due to the torso at certain angles [83]. It is generally assumed that, due to their size, pinnae are responsible for HRTF features (and, therefore, localisation ability) above approximately 5 kHz [29]. However, the relationship between pinna morphology and HRTF feature is complex. The generation of peaks is generally attributed to resonant modes, whilst the generation of notches is attributed to a combination of interference and diffraction effects [84, 85]. The impact of some specific features with respect to vertical localisation has been investigated, particularly the contributions of the lowest frequency peak and notches [64, 79, 82, 86–88].

As the spectral filtering is due to the morphology, and morphology varies between individuals, each set of HRIRs and HRTFs is unique to the individual [59, 89]. Accordingly, HRTFs are referred to as *individualised* if derived from a particular subject. There are, however, similarities in spectral responses across listeners due to the underlying characteristic features of human morphology. Fig. 2.22 shows HRTF pairs for one listener (left) and for all subjects (right) for a sound source at $\theta = 90°$, $\phi = 0°$, as reported by Møller *et al.* [90]. Whilst much of the inter-subject variation is above 8 kHz, in accordance with the frequency range associated with the pinnae, there is a general consistency in the spectra across subjects.



(a) For one subject



(b) For all subjects

Figure 2.22: HRTF pairs for multiple subjects for a sound source at $\theta = 90°$, $\phi = 0°$. Both inter-subject variation and inter-subject consistency is evident. Reproduced from [90] with permission.

The auditory system of each individual has, over the course of their lifetime, adapted to the unique filtering arising from their morphology in order to determine the position of a sound source. If HRTF pairs individualised to one subject are used to render and present a binaural signal to another listener, a deterioration in localisation performance occurs, most commonly resulting in increased front-back confusion and elevation errors [59, 91]. This is due to the mismatch between the filters employed and those which the listener is familiar with. However, there has been increasing evidence that it is possible for a listener to adapt to a different set of filters, and that this ability is retained over time [92–95].

HRTFs derived from an artificial head (such as the Neumann KU100 [96]) or a head-and-torso simulator (HATS) (such as the Knowles Electronics Manikin for Acoustic Research (KEMAR) [97], shown in Fig. 2.23) are often referred to as *generic* HRTFs. KEMAR is reported to have acoustic properties approaching that of a median human, as the dimensions were derived from statistical research across over 4,000 human subjects [98]. Use of generic HRTFs is preferable to averaging, as creating a numerical average HRTF from measured HRTFs tends to smooth the features and, therefore, creates a filter response which does not resemble a HRTF [90].

Figure 2.23: An example of a HATS: the KEMAR model 45BC [97, 98]. Reproduced from [99] with permission.

2.5.5 Distance-Dependency of Localisation Cues

The far-field is defined in Section 2.3 as the region in which wavefronts approximate plane waves (typically at large distances compared to the wavelength). The near-field is defined as the region within which wavefronts are spherical. At audio frequencies, the distance where the near-field transitions to the far-field is considered to be roughly 1 m from the head, as defined by Brungart and Rabinowitz [36]. The near- and far-field regions can also be referred to with respect to the subject as *peripersonal* space (within reaching distance) and *extrapersonal* space (outside reaching distance) [100].

As the wavefronts are planar in the far-field, HRTFs measured in this region are assumed to be independent of distance [36, 101, 102], although the author has demonstrated elsewhere the existence of spectral changes with distance within the far-field that may challenge this assumption [103]. Within the near-field, however, HRTFs display substantial distance-dependent variation.

The magnitudes of the HRTFs generally increase with decreasing source distance, according to the inverse square law. However, there is an additional increase in magnitude with frequency on the ipsilateral side and a decrease with frequency on the contralateral side [36]. Near-field HRTFs also exhibit much larger ILD values than those found in the far-field as a result of increased head shadowing, particularly at lower frequencies [36, 104]. For example, Brungart [102] found ILD values of 5 to 6 dB at 500 Hz in the far-field, but greater than 15 dB in the near-field for the same source angle at the same frequency.

The peaks and troughs within the frequency spectrum are also different between near-field and far-field HRTFs for the same angle. Centre frequencies, amplitudes and bandwidths of spectral features found in the far-field counterpart show considerable variation with decreasing distance [105, 106], and additional peaks and notches appear due to the proximity of the shoulders and pinnae [102, 104, 106]. The presence of a parallax effect at close distances pushes HRTF features to more lateral positions [105, 107].

With regard to ITD, Brungart and Rabinowitz [36] demonstrated that ITD is roughly independent of distance, although at very close distances in the front hemisphere the facial features have an impact.

2.5.6 Localisation Accuracy

Localisation accuracy relates to the ability of the auditory system to determine the position of a sound source, and can be characterised using an absolute or a referential metric. Absolute localisation accuracy is tested when a listener is asked to indicate where they perceive a sound source to be (an identification task) [108]. Referential localisation accuracy is tested when the listener is asked to indicate when two stimuli are perceived as spatially distinct (a discrimination task) [108]. Whilst these metrics generally correspond, they are not truly interchangeable, as they make use of different aspects of the auditory system. Absolute localisation accuracy is based on the accuracy of the listener's internal map of auditory space, while referential localisation ability is based on the ability of the listener to discern the difference between two stimuli [109].

The principle aim of this thesis requires characterisation of the difference between two sets of data (discussed further in Chapter 6 and Chapter 7). Therefore, the referential approach to localisation ability was selected as most appropriate. This is commonly referred to as *localisation blur* or the *angular JND* (when quantifying localisation accuracy with respect to the direction of a source). When discussing localisation ability with respect to the binaural cues derived from a given sound source direction, it can also be quantified in terms of JND for changes in ITD and ILD.

Using either approach, localisation accuracy varies as a function of source direction. It is generally best in the frontal region, becoming worse as the sound source is moved away from the front in azimuth and/or elevation [110, 111]. In [63], Blauert reports the results of large-scale tests by Preibish-Effenberger [112] and Haustein and Schirmer [113], wherein mean localisation blur values of $\pm 3.6°$, $\pm 10°$ and $\pm 5.5°$ were found at the front, sides and rear of the head, respectively. Oldfield and Parker [114] observed that the 8 subjects in their study showed larger absolute localisation errors in the rear hemisphere ($110° \leq \theta \leq 170°$) and for the upper elevations tested ($20° \leq \phi \leq 40°$). However, localisation blur in terms of angle does not give an indication of the contribution of the interaural differences cues. The calculation of JND threshold values for ITD and ILD can give this further information. The lowest JND thresholds (and, therefore, defined as $1\,\mathrm{JND_{ITD}}$ and $1\,\mathrm{JND_{ILD}}$) are reached in the frontal region. For ITD, the smallest change which can be perceived is around 16 µs (equivalent to localisation accuracy of 1°) [111, 115–117], and whilst JND for ILD varies with frequency, the smallest perceivable changes are around 1 dB [118, 119].

The variation in JND threshold as a sound source moves away from the frontal region is different, however, between ITD and ILD. Whilst JND thresholds for both are at a minimum in the frontal region, and the smallest perceivable change increases as the sound source moves away from that region, the rate at which the thresholds increase is different between ITD and ILD. This can be quantified by discussing the change in source direction for which the JND threshold is a multiple of the minimum threshold.

For changes in ITD in the horizontal plane, Simon *et al.* [116] reported $2\,\mathrm{JND_{ITD}}$ to be reached by approximately 30°, and thresholds of up to $5\,\mathrm{JND_{ITD}}$ at a source direction of 90°. For changes in ILD, Tu *et al.* [120] demonstrated a roughly linear relationship between ILD and $\mathrm{JND_{ILD}}$, where a change in ILD of 1 dB saw an increase in $\mathrm{JND_{ILD}}$ of 0.1 dB. An increase in ILD is analogous to a source moving away from the median plane in the azimuth direction, and so $2\,\mathrm{JND_{ILD}}$ can be said to be reached by approximately 60° in azimuth [119, 120]. However, to the author's knowledge, a more complete analysis of the variation in $\mathrm{JND_{ITD}}$ and $\mathrm{JND_{ILD}}$ across many listeners is absent from the literature.

Additionally, localisation accuracy varies between individuals, and has been shown to be influenced by listening conditions, stimuli and the reporting method [63, 108]. Broadband stimuli typically produce a more accurate response as they are easier to localise [89]. Because of these variations, JND values are often quoted as an average value over the modality or subject group measured.

2.6 Capture of Acoustic Localisation Cues

A virtual sound source can be created by mathematically applying the localisation cues to a mono sound source. This is known as a *spatialised* sound source. The creation and reproduction of virtual spatialised sound sources are discussed further in Section 2.7.

However, in order to spatialise a sound source, the acoustic localisation cues must be known. Capturing these cues can be achieved in various ways, but the two most common methods are based on acoustic measurement and on numerical computation.

2.6.1 Measurement-Based Methods

The earliest methods for capturing localisation cues used acoustic measurement, where a known measurement signal is played through a loudspeaker and the result recorded at the ears of the listener. Mathematical processing is then applied to extract the HRIRs from the recorded signals. Usually, the acoustic measurement is performed in the time domain, so the acoustic cues will be described in this section as HRIRs rather than HRTFs.

2.6.1.1 Theory

A real-world acoustic filter is usually treated as a continuous linear-time invariant (LTI) system, whereby the relationship between the input signal and output signal is characterised as a convolution [121]:

$$g(t) = h(t) * f(t) = \int_{-\infty}^{\infty} h(x)f(t-x)dx \tag{2.25}$$

where $f(t)$ is the input signal, $g(t)$ is the output signal, and $h(t)$ fully characterises the system response. Convolution (denoted by the $*$ operator) is a mathematical operation performed on two time-domain functions to produce a third time-domain function, where the third function expresses how one function is modified by the other at each time step. The equivalent process in the frequency domain is to take the product of the two frequency-domain versions of the time-domain functions [122]. Within the measurement of HRIRs, $f(t)$ is the signal from the loudspeaker, $g(t)$ is the signal recorded at the ears, and $h(t)$ is the HRIR.

There are several ways to identify the system response $h(t)$. If a Dirac delta function $\delta(t)$ is input to the system, $g(t)$ will equal $h(t)$ and so $h(t)$ can be recorded directly. If the frequency response of both $f(t)$ and $g(t)$ are known ($F(f)$ and $G(f)$, respectively), the frequency response $H(f)$ (and, therefore, $h(t)$) can be determined through division. Alternatively, with an appropriate inverse version of the input signal (typically time-reversed and frequency-compensated), it is possible to extract only the characteristic of the system via:

$$h(t) = g(t) * f_{inv}(t) \tag{2.26}$$

where $f_{inv}(t)$ is the inverse of the input signal. Whilst this is strictly also a convolution process, it is often referred to as *deconvolution*, where the system response $h(t)$ is *deconvolved out of* the recorded signal $g(t)$.

2.6.1.2 Excitatory (Input) Signal Production

The excitation signal to be input to an LTI system can be one of several types, including sinusoids and sine sweeps, impulses, random noise, or pseudorandom noise such as maximal-length sequence (MLS) or a Golay code. An essential property of the signal is that it must contain energy at all frequencies of interest. For the measurement of HRIRs, the signal is reproduced using a loudspeaker at the desired position within an (ideally) anechoic room and the sounds arriving at the ears of the subjects are recorded.

The loudspeaker used to reproduce the excitation signal should have an approximately flat frequency response over the desired bandwidth of the measurement, and ideally should approximate a point source. Realistically, however, small or co-axial loudspeakers are used (such as by Kearney and Doyle [123]) to ensure that the central radiating point is common to all frequency components. If a non-co-axial loudspeaker is to be used, it should be placed at a far-field distance to achieve adequate plane wave propagation.

In order to obtain HRIRs at many different source directions, the relative position between the loudspeaker and the subject is altered. This can be done by: moving either the loudspeaker or the subject; using many loudspeakers simultaneously; or by some combination of these (for example, that used by Armstrong *et al.* [124]). An alternative method is to make use of the principal of reciprocity (which states that the transfer function from point A to point B is the same transfer function as that from point B to point A), and place the loudspeaker in the blocked ear and the microphones in the far-field, such as the measurement systems described by Zotkin *et al.* [125] and Matsunaga and Hirahara [126].

2.6.1.3 Response (Output) Signal Capture

During HRIR measurement, the response of the acoustic system to the excitation signal is usually captured via a pair of small microphones in the ears of the subject. The microphones are placed either close to the ear canal entrance or close to the eardrum, as the ear canal resonance affects the results if the measurement point is somewhere in between [121]. A probe microphone is used to measure at the eardrum, but these can be hazardous to the subject and difficult to use [121]. The ear canal does not provide directionally-dependent information [127, 128] and so does not strictly need to be included for HRIR capture. Techniques for measuring at the blocked entrance to the ear canal (so-called 'blocked meatus' techniques) were introduced by Møller [129], and subsequently became widely applied as the technique was safer and more convenient. In this technique, a miniature microphone is mounted in a plug or putty (as used by Armstrong *et al.* [74], see Fig. 2.24) and placed flush with the entrance to the ear canal such that the ear canal is blocked.

The recorded signals are then convolved with the appropriate inverse of the excitation signal to extract the raw HRIRs. HRIRs can also be measured using an artificial head, such as the Neuman KU100, or a HATS, such as KEMAR, which have built-in microphones in the blocked meatus position (see Fig. 2.25). The use of an artificial subject generally results in smaller errors due to the absence of subject movement and makes lengthier measurements possible as subject fatigue is not an issue.

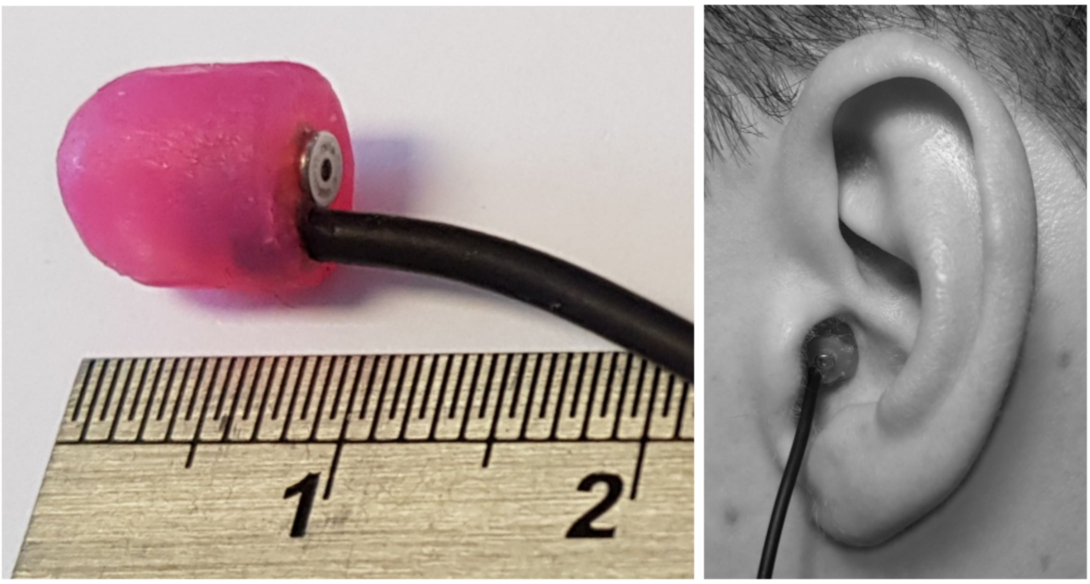


Figure 2.24: Knowles FG-23329-C05 microphone capsule in custom-designed 3D printed plug (left) and positioned within the ear of a participant (right), as used by Armstrong *et al.* [74]. Reproduced from [74] under CC-BY-4.0.

(a) Left ear of KEMAR.

(b) Left ear of Neumann KU100.

Figure 2.25: Microphones mounted inside artificial subjects in the blocked meatus position.

An increasing number of databases of acoustically-measured HRIRs exist, to the point where listing here every known instance is impractical. Databases are created with a variety of applications in mind, with both individualised and generic HRTFs measured under a variety of conditions and restrictions. Xie [121] describes and compares 15 databases from 1989 to 2012 across a range of methods, subjects, distances and angles, including [24, 90, 130–141]. Examples of more recent databases can be found in [74, 142, 143].

2.6.1.4 Near-Field vs. Far-Field Measurement

HRIRs are usually measured in the far-field, as the binaural cues are assumed to be independent of distance in this region due to the approximately planar nature of the wave fronts [36, 101, 102], and so only one measurement distance is required. If distance information is required in the reproduction stage, this can be implemented using amplitude and time shifts or, in more complex scenes, by a change in the direct-to-reverberant ratio or a high-frequency adjustment [144]. However, for true accuracy, near-field HRIRs are required if virtual sound sources are intended to be placed close to the head [102]. This presents several technical challenges in addition to those which exist when making far-field measurement. These relate to the differences between near-field and far-field acoustic cues (see Section 2.5.5).

As binaural cues vary greatly with distance when the source is within a metre of the head, and loudness has been demonstrated to increase non-monotonically [145], in order to represent distance information for close sources accurately more measurements must be made [102, 104]. This can substantially increase the overall duration of the measurements and the complexity of the system required. Additionally, any movement of the subject has a greater impact in the near-field, so it is preferable to immobilise the head of the subject as far as is practical [101].

Due to the proximity of the sound source to the head, a source smaller than a standard loudspeaker is required in order to approximate a point source adequately [102]. Various methods have been proposed to address this, such as the tube and compression system used by Brungart and Rabinowitz [36], the spherical dodecahedron source or 18 mm driver used by Yu *et al.* [104, 146] and the spark gap used by Qu *et al.* [147].

2.6.2 Model-Based Methods

The experimental procedure involved in acoustic measurement of HRIRs can be time-consuming, complex and repetitive. With recent increases in computational power, a viable alternative is to use computational methods. Model-based methods often operate in the frequency domain, so the acoustic cues will be described in this section as HRTFs rather than HRIRs.

2.6.2.1 The Spherical Head Model

The simplest approach for HRTF estimation is the spherical head model, where the head is approximated as a rigid sphere with radius a without pinnae and torso, as shown in Fig. 2.26. For a plane wave, the pressure $P(\Gamma, f)$ on the sphere surface can be calculated using the analytical solution for scattering on a sphere [148] (cited in [149]):

$$P(\Gamma, f) = -\frac{P_w}{(ka)^2} \sum_{l=0}^{\infty} \frac{(2l+1) j^{l+1} P_l(\cos\Gamma)}{\frac{\partial h_l(ka)}{\partial (ka)}} \tag{2.27}$$

where Γ is the angle between the incident direction and the ray from the sphere centre to the point on the sphere surface (in °, indicated in red), $k = \frac{2\pi f}{c}$ is the wave number, P_w is the pressure magnitude of the incident wave, $P_l(\cos\Gamma)$ is the Legendre polynomial of degree l, and $h_l(ka)$ is the spherical Hankel function of the second kind of order l. For practicality, this infinite series is often truncated to an order L.

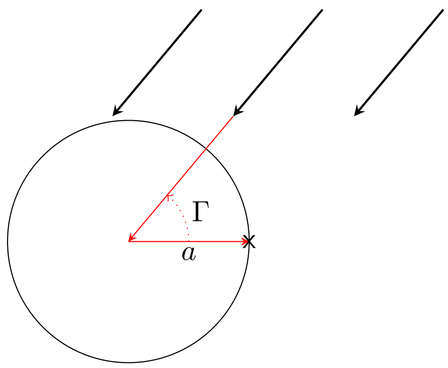


Figure 2.26: The spherical head model, where the head is treated as a rigid sphere with radius a. The pressure $P(\Gamma, f)$ as the result of an incident wave can be calculated at a point on the sphere surface (x) using Equation 2.27. Angle of incidence, Γ, is 50°.

Fig. 2.27 plots the frequency response at a point x on the sphere surface ($a = 8.75\,\mathrm{cm}$) for a sound source at increasing angles of incidence Γ in the horizontal plane at a radial distance of $r = 1\,\mathrm{m}$. A visual indication of the incident direction of the sound source for increasing values of Γ is plotted in corresponding colours in Fig. 2.28, where $\Gamma = 0°$ indicates that the point x is aligned with the angle of incidence.

At low frequencies (below about 400 Hz), the response is not directionally dependent and only exhibits a shift in amplitude. For higher frequencies, the response for low angles of incidence increases noticeably and the response for greater angles of incidence decreases. Interference ripples above 1 kHz are visible at all angles of incidence. The response does not create a minimum at $\Gamma = 180°$ due to an acoustic 'bright spot' caused by constructive interference of waves arriving at this point from multiple directions.

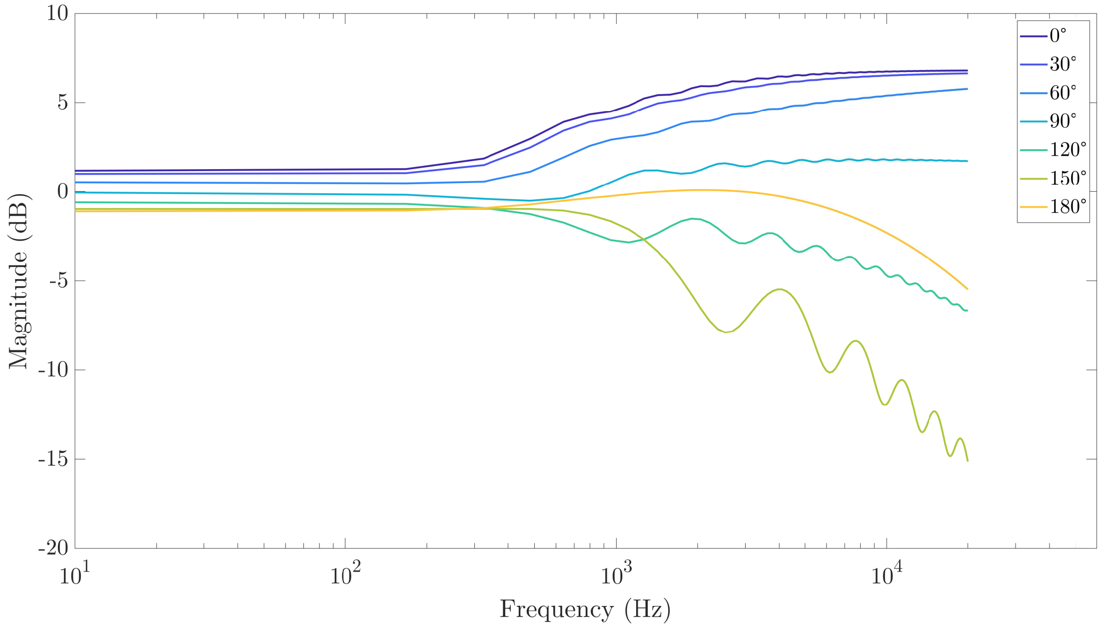


Figure 2.27: The frequency response of the spherical head for sound sources of varying angles of incidence Γ in the horizontal plane. Interference effects introduce ripples and head shadowing causes a general shift in level with direction at high frequencies.

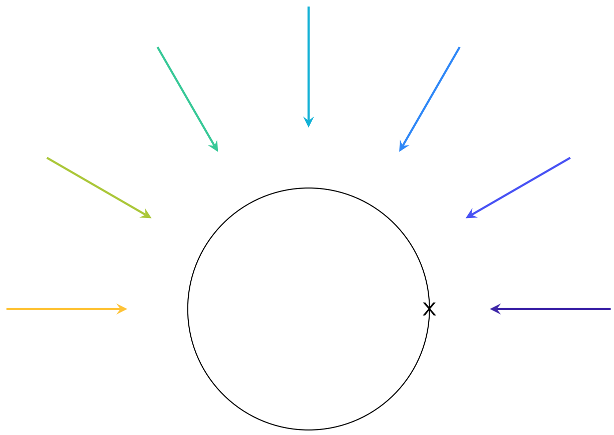


Figure 2.28: The direction of the sound source for increasing values of Γ in the horizontal plane with respect to the point x on the sphere surface. Colours correspond to those used in Fig. 2.27.

Xie [149] derives the equations for binaural signals at the left and right ears of a spherical head by defining the positions of the points of interest as diametrically across the head. For a sound source at incident angle θ (where 0° is the frontal direction), Γ can then be defined for each ear:

$$\Gamma_L = 90 - \theta \tag{2.28}$$
$$\Gamma_R = 90 + \theta \tag{2.29}$$

The angular values discussed in [149] have been adjusted to match the convention used in this thesis, where positive values increment in the counter-clockwise direction. The HRTFs are then calculated by substituting these expressions for Γ_L and Γ_R into Equation 2.27 in turn:

$$H_L(\theta, f) = -\frac{1}{(ka)^2}\sum_{l=0}^{\infty}\frac{(2l+1)j^{l+1}P_l(\sin\theta)}{\frac{\partial h_l(ka)}{\partial(ka)}} \tag{2.30}$$

$$H_R(\theta, f) = -\frac{1}{(ka)^2}\sum_{l=0}^{\infty}\frac{(2l+1)j^{l+1}(-1)^l P_l(\sin\theta)}{\frac{\partial h_l(ka)}{\partial(ka)}} \tag{2.31}$$

Without modelling the effect of the outer ear, however, the results are only valid for low frequencies, as the pinnae begin to have an influence above approximately 5 kHz [29]. Additionally, for a human subject, the ears are not located diametrically on the sphere, but rather sit approximately 10° behind the interaural axis [63]. Using more accurate ear positions increases the applicable frequency range of the model [149–151].

Duda and Martens [152] investigated the range dependence of the spherical head model response, as Equation 2.27 is defined for plane waves (that is, a source at an infinite distance). The pressure at a point on the sphere surface generated by a sinusoidal point source at radial distance r is given by [148] (cited in [149]):

$$P(r, \Gamma, f) = -j\frac{\rho_{air} c Q_0}{4\pi a^2}\sum_{l=0}^{\infty}\frac{(2l+1)h_k(kr)P_l(\cos\Gamma)}{\frac{\partial h_l(ka)}{\partial(ka)}} \tag{2.32}$$

where ρ_{air} is the density of air, and Q_0 is the intensity of the point sound source. Xie [149] plots the magnitude of HRTFs for sound sources varying in radial distance for two incident angles. The findings are generally in keeping with those discussed in Section 2.5.5 for near-field localisation cues. As the sound source approaches the sphere, the response increases on the ipsilateral side and decreases on the contralateral side, and ILD increases at low frequencies.

2.6.2.2 The Snowman Model

The spherical head was extended by Algazi *et al.* [153] to include the torso using the 'snowman' model. In this model, the head and torso are approximated by two spheres SA and SB with radius values of a_A and a_B, respectively. The wave equation can then be solved to find the pressure at a field point $\mathbf{r}'$ which is generated by a sinusoidal point source at $\mathbf{r}$ with intensity Q_0 [148] (cited in [149]):

$$\nabla\prime^2 P(\mathbf{r}', \mathbf{r}, f) + k^2 P(\mathbf{r}', \mathbf{r}, f) = -jk\rho_{air} c Q_0 \delta(\mathbf{r}' - \mathbf{r}) \tag{2.33}$$

where $\delta(\mathbf{r}' - \mathbf{r})$ is the Dirac delta function, and $\nabla\prime^2$ is the Laplacian operator with respect to field point $\mathbf{r}'$. The Sommerfeld radiation condition requires that a wave decays to zero as it travels infinitely far from the source [149]:

$$\lim_{r'\to\infty} r'\left[\frac{\delta P(\mathbf{r}', \mathbf{r}, f)}{\delta r'} + jkP(\mathbf{r}', \mathbf{r}, f)\right] = 0 \tag{2.34}$$

As with the spherical head model, only the two field points which correspond to the two ears are of interest (and only one of these if symmetry is assumed). The snowman model makes use of the principal of reciprocity, where exchanging the positions of the source and field point yields the same results. The pressure can therefore be calculated at an arbitrary source position $\mathbf{r}$ while the source is located on the head surface at $\mathbf{r}'$, which exchanges the positions of $\mathbf{r}'$ and $\mathbf{r}$ in Equation 2.33 and Equation 2.34 to give:

$$\nabla^2 P(\mathbf{r}, \mathbf{r}', f) + k^2 P(\mathbf{r}, \mathbf{r}', f) = -jk\rho_{air} c Q_0 \delta(\mathbf{r} - \mathbf{r}') \tag{2.35}$$

$$\lim_{r' \to \infty} r' \left[\frac{\delta P(\mathbf{r}, \mathbf{r}', f)}{\delta r} + jkP(\mathbf{r}, \mathbf{r}', f) \right] = 0 \tag{2.36}$$

where ∇^2 is the Laplacian operator with respect to field point $\mathbf{r}$. A HRTF can then be obtained by solving Equation 2.35 with the following boundary conditions, where the normal derivative of $P(\mathbf{r}', \mathbf{r}, f)$ should be zero on the surface of both SA and SB, as described by Xie [154]:

$$\left. \frac{\delta P(\mathbf{r}, \mathbf{r}', f)}{\delta n} \right|_{SA} = 0 \tag{2.37}$$

$$\left. \frac{\delta P(\mathbf{r}, \mathbf{r}', f)}{\delta n} \right|_{SB} = 0 \tag{2.38}$$

Algazi *et al.* [83, 153] calculated the HRTFs of a snowman model with a head and torso of radii $a_A = 8.75\,\text{cm}$ and $a_B = 23\,\text{cm}$, respectively. They reported notches for elevated sources which were not present in the results for the spherical head model, and attributed these to destructive interference between the direct sound and reflections off the torso (the so-called 'shoulder reflection'). The lowest frequency at which these notches occur (approximately 1 kHz) corresponds to the longest path difference between direct and reflected paths, which occurs for sources at approximately $\phi = 80°$. The notch centre frequency increases as the source elevation deviates downwards from $\phi = 80°$, and has been proposed as a cue for localisation of elevated sources [79]. For negative elevation angles, a shadowing effect is produced by the torso. Sources outside of what Algazi *et al.* [83] define as the 'torso shadow cone' produce a reflection, whilst those within are shadowed by the torso, as shown in Fig. 2.29.

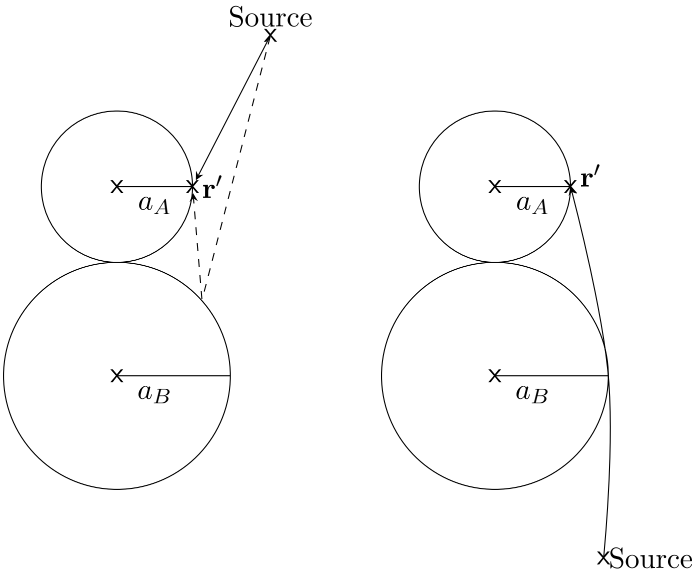


Figure 2.29: Examples of the snowman model [153]. Elevated sources (left) produce a torso reflection off the torso (indicated with a dashed line). Sources with lower elevation values (right) are shadowed by the torso.

Gumerov *et al.* [155] also calculated HRTFs of a snowman model using the algorithm discussed in [156, 157], with a ratio between the head and torso spheres of $\frac{a_B}{a_A} = 1.3253$. As Algazi *et al.* [153] did, they also reported the presence of notches and attributed them to reflection interference.

2.6.2.3 The Boundary Element Method

Whilst simplified numerical approaches such as the spherical head and snowman model give reasonable approximations of HRTFs, particularly at low frequencies (and, indeed, have been used for low frequency modelling of acoustic measurements [123]), these still lack the detail introduced by the pinnae at higher frequencies. Several computational techniques exist which permit the calculation of sound propagation around more complex geometries, among which the boundary element method (BEM) is the most commonly used in HRTF calculation. Wave propagation through the object is ignored, meaning that the BEM only requires subdivision of the boundary (as opposed to the subdivision of the volume as required by other methods) and thus reduces the dimensions of the problems from three to two [158]. Implementation of the BEM varies across research groups and software packages, but the underlying mathematical principles remain the same.

The BEM is suitable for use in any radiation and scattering problem where the wave equation is solved on the boundary of an object. The calculation of HRTFs is only one example of this; discussion of further applications of the BEM can be found in [159]. In the case of HRTF calculation, the boundary B is a surface mesh of the subject for whom HRTFs are desired, the field points of interest correspond to the ears, and the source is located at a position outside the mesh. A second boundary is located at an infinite distance, at which the pressure satisfies the Sommerfield radiation condition described by Equation 2.34 [149].

There are two BEM techniques used in acoustic computation: direct and indirect BEM (DBEM and IBEM, respectively). DBEM is based on the Helmholtz formula, which relates the pressure in the fluid domain (in this case, air) to the pressure and its normal on the boundary [160]. IBEM assumes that the pressure field is caused by a monopole distribution on the boundary surface [160]. The choice of technique is based on the problem size. The direct method is suited for meshes comprising more than a few hundred faces as it is optimised for speed (but requires more storage), whereas the indirect method is slower but requires less storage [128]. As the work in this thesis makes use of meshes comprising more than a few hundred faces, and the software packages used are based on the DBEM implementation, only the derivation for the DBEM-based solution will be presented here, based on that derived by Xie [149].

The pressure at a field point in a volume V is a combination of two pressures: that generated by the source and that caused by scattering and reflections off the boundary B. The solution can be expressed as a Kirchhoff-Helmholtz integral equation:

$$\begin{aligned} C(\mathbf{r}')P(\mathbf{r}',\mathbf{r},f) = {} & jk\rho_{air}cQ_0G(\mathbf{r}',\mathbf{r},f)+ \\ & \iint\limits_B \left[G(\mathbf{r}',\mathbf{r}'',f)\frac{\partial P(\mathbf{r}'',\mathbf{r},f)}{\partial n''} - P(\mathbf{r}'',\mathbf{r},f)\frac{\partial G(\mathbf{r}',\mathbf{r}'',f)}{\partial n''}\right] dB'', \end{aligned} \tag{2.39}$$

$$C(\mathbf{r}') = \begin{cases} 1/2 & \mathbf{r}' \in B \\ 1 & \mathbf{r}' \in V \\ 0 & other \end{cases} \tag{2.40}$$

where n'' is the outward normal direction, and $G(\mathbf{r}',\mathbf{r},f)$ is the free-space Green's function of a point source defined as:

$$G(\mathbf{r}',\mathbf{r},f) = \frac{1}{4\pi|\mathbf{r}'-\mathbf{r}|}\mathrm{e}^{(-jk|\mathbf{r}'-\mathbf{r}|)} \tag{2.41}$$

Equation 2.39 and Equation 2.40 indicate that, for a given source intensity and position, the pressure at the field point $\mathbf{r}'$ is determined by the free-space Green's function of a point source, and the pressure and its normal derivative on the boundary surface. However, the normal derivative of pressure is zero on a rigid boundary. More generally, on a boundary surface with acoustic admittance $Y(\mathbf{r}',f) = \frac{1}{Z(\mathbf{r}',f)}$, the boundary condition can be written as [148] (cited in [149]):

$$\left.\frac{\delta P(\mathbf{r}',\mathbf{r},f)}{\delta n'}\right|_B = -j2\pi f\rho_{air}v_{n'} \tag{2.42}$$

$$= -jk\rho_{air}cY(\mathbf{r}',f)P(\mathbf{r}',\mathbf{r},f) \tag{2.43}$$

where $v_{n'}$ is the normal velocity of the medium.

The boundary B is discretised into a mesh of β faces ($B_m, m = 1,2,...,\beta$), where the central position of the mth element is specified by the vector $\mathbf{r}'$ and within each element, the pressure and velocity are assumed to be constants. Equation 2.42 can be substituted into Equation 2.39, and Equation 2.39 can then be approximated as a summation over β faces:

$$C(\mathbf{r}')P(\mathbf{r}',\mathbf{r},f) = jk\rho_{air}cQ_0G(\mathbf{r}',\mathbf{r},f) - \sum_{m=1}^{\beta}\left[\iint\limits_{B_m}\frac{\partial G(\mathbf{r}',\mathbf{r}'',f)}{\partial n''}dB'' + jk\rho_{air}cY(\mathbf{r}\prime_m,f)\iint\limits_{B_m}G(\mathbf{r}',\mathbf{r}'',f)dB''\right]P(\mathbf{r}\prime_m,r,f) \tag{2.44}$$

This is then converted into β linear equations to determine the pressure at each face on the boundary surface. This is achieved by setting $\mathbf{r}' = \mathbf{r}'_q$ and $C(\mathbf{r}') = \frac{1}{2}$, where $\mathbf{r}\prime_q$ with $q = 1, 2, ..., \beta$ indicates the position of the centre of each element:

$$\frac{1}{2}P(\mathbf{r}'_q,\mathbf{r},f) = jk\rho_{air}cQ_0G(\mathbf{r}'_q,\mathbf{r},f) - \sum_{m=1}^{\beta}\left[G^n_{qm}(f) + jk\rho_{air}cY(\mathbf{r}'_m,f)G_{qm}(f)\right]P(\mathbf{r}'_m,\mathbf{r},f) \tag{2.45}$$

where

$$G^n_{qm}(f) = \iint\limits_{B_m}\frac{\partial G(\mathbf{r}'_q,\mathbf{r}'',f)}{\partial n''}dB'' \tag{2.46}$$

$$G_{qm}(f) = \iint\limits_{B_m}G(\mathbf{r}'_q,\mathbf{r}'',f)dB'' \tag{2.47}$$

$$q,m = 1,2,...,\beta \tag{2.48}$$

The following matrices can be defined to enable reformulation of the set of β equations defined in Equation 2.45:

$$\mathbf{P}_\beta = [P(\mathbf{r}'_1,\mathbf{r},f), P(\mathbf{r}'_2,\mathbf{r},f), ..., P(\mathbf{r}'_\beta,\mathbf{r},f)]^T \qquad [\beta\text{x}1] \tag{2.49}$$

$$\mathbf{G}_\beta = [G(\mathbf{r}'_1,\mathbf{r},f), G(\mathbf{r}'_2,\mathbf{r},f), ..., G(\mathbf{r}'_\beta,\mathbf{r},f)]^T \qquad [\beta\text{x}1] \tag{2.50}$$

$$[Y] = \text{diag}[Y(\mathbf{r}'_1,\mathbf{r},f), Y(\mathbf{r}'_2,\mathbf{r},f), ..., Y(\mathbf{r}'_\beta,\mathbf{r},f)] \qquad [\beta\text{x}\beta] \tag{2.51}$$

where $\mathbf{P}_\beta$ comprises the pressure on each face, $\mathbf{G}_\beta$ comprises the values of the free-space Green function's at the position of each face for a point source at $\mathbf{r}$, and $[Y]$ comprises the acoustic admittance value $Y(\mathbf{r}',f)$ at each face. Additionally, $[G^n]$ and $[G]$ are defined, which represent two βxβ matrices comprising G^n_{qm} and G_{qm}, respectively, and $[I]$ represents a βxβ identity matrix. Equation 2.45 can then be formulated as:

$$\left\{\frac{1}{2}[I] + [G^n] + jk\rho_{air}c[G][Y]\right\}\mathbf{P}_\beta = jk\rho_{air}cQ_0\mathbf{G}_\beta \tag{2.52}$$

From this, the pressure vector $\mathbf{P}_\beta$ can be calculated at each required frequency f, then substituted into Equation 2.44 to find the pressure $P\mathbf{r}', \mathbf{r}, f)$ at a field point $\mathbf{r}'$ by:

$$C(\mathbf{r}')P(\mathbf{r}', \mathbf{r}, f) = jk\rho_{air}cQ_0G(\mathbf{r}', \mathbf{r}, f) - \{\mathbf{G}^{\mathbf{n}}_{\mathbf{r}'} + jk\rho_{air}c\mathbf{G}_{\mathbf{r}}[Y]\}\,\mathbf{P}_{\beta} \tag{2.53}$$

where $\mathbf{G}^{\mathbf{n}}_{\mathbf{r}'}$ and $\mathbf{G}_{\mathbf{r}'}$ are 1xM vectors with elements given by:

$$G^{n}_{r',m}(f) = \iint\limits_{B_m} \frac{\partial G(\mathbf{r}', \mathbf{r}'', f)}{\partial n''} dB'' \tag{2.54}$$

$$G_{r',m}(f) = \iint\limits_{B_m} G(\mathbf{r}', \mathbf{r}'', f) dB'' \tag{2.55}$$

$$m = 1, 2, ..., \beta \tag{2.56}$$

The first term on the right side of Equation 2.53 is the free-field pressure at field point $\mathbf{r}'$ generated by the point source $\mathbf{r}$. The second term is the pressure at field point $\mathbf{r}'$ produced by the scattering and reflections off the boundary surface, and indicates that this contribution is produced by the sum of scattered and reflected waves due to the boundary faces.

Computational methods other than the BEM have been used for the simulation of HRTFs, albeit less often. Such methods include the infinite-finite element method (IFEM) [161], the finite-difference time-domain (FDTD) method [84, 162], and the ultra-weak variational formulation (UWVF) [163, 164].

As the contribution of each boundary element is of crucial importance to the overall result, the use of the BEM for HRTF simulation relies on high quality 3D meshes of the subject in question. While meshes of varying resolution have been shown to reduce computational requirements [165], the use of lower resolution meshes in the important pinna regions has been shown to produce HRTFs which do not match acoustically-measured results, particularly at higher frequencies [161, 166]. Scans of subjects have been obtained using laser [160, 167], scattered light [168], photogrammetry [169–171], infrared stereo imaging [172], magnetic resonance imaging (MRI) [80, 162] and computerised tomography (CT) [173], sometimes in combination [106, 166, 174]. The resolution of the mesh (more specifically, the number of faces used to represent the 3D object, and therefore the length of the edges in the faces used) determines the maximum valid frequency of the resulting simulation, according to:

$$f_{max} = \frac{c}{d_{max} \times \eta} \tag{2.57}$$

where f_{max} is the maximum valid frequency in Hz, d_{max} is the length of the longest face edge in m, and η is the number of edges per wavelength. Smaller faces, therefore, give a higher valid maximum frequency, but can lead to meshes comprising large numbers of faces and therefore requiring more storage. The optimum value of η for HRTF calculation is discussed further with respect to mesh processing workflows in Chapter 3. A value of $\eta = 6$ is most commonly used [175, 176], which gives a maximum edge length requirement of 2.86 mm for results to be valid up to 20 kHz.

However, the requirement for high mesh resolution leads to high computational and storage costs for the simulation. The size of the set of linear equations increases with maximum valid frequency, giving exponential relationships between the number of faces β and the computational and storage costs of β^3 and β^2, respectively [149].

To reduce computational cost, an alternative formulation of the BEM, referred to as the fast multipole BEM, has been proposed [166, 177, 178] and subsequently implemented in various software packages such as Mesh2HRTF [179–181]. FM-BEM also incorporates the principal of reciprocity, which further reduces computational cost [128]. Without the use of reciprocity, a separate calculation is required for each source position; exchanging the source and receiver positions reduces this to only two required calculations. The computational cost and storage for FM-BEM are (βN_{iter}) and β, respectively, where N_{iter} is the number of iterations in the calculation [149].

More recently, the added increase in computational speed of graphics processing units (GPUs) has also been investigated with regards to HRTF calculation. As a result of their parallel architecture, GPUs are capable of performing many calculations at once, as opposed to the concurrent processing of central processing units (CPUs) [182]. Such investigations have included both running the entire BEM solver on a GPU [183–185] and running only the matrix calculations on a GPU [186], with reported increases in speed of anywhere from 6 to 36 times, occasionally exceeding a 100-fold increase [183]. However, while the reported increases in speed are promising, use of GPUs for HRTF calculation does not appear to have spread across the wider research area.

As computational resources have increased and the computational requirement of the BEM has decreased, the use of the BEM within binaural research has grown, both as a means of obtaining HRTFs and as a means of investigations not feasible using acoustic measurements.

In the late 1990s, Kahana *et al.* [187] used the BEM to calculate HRTFs of a torso-less KEMAR up to 6 kHz and demonstrated that, despite the computational limitations imposed at the time, the results were consistent with measurements. In subsequent work [161], the same authors calculated HRTFs of spherical and ellipsoidal heads, and of KEMAR, to a higher maximum frequency of 10 kHz through simulation using half the model and assuming symmetry. Later work investigated the resonant modes within the pinnae [188] and the contribution of the pinnae to the HRTF [160]. The use of meshes, and the ability to modify them, is one of the benefits of using computational techniques such as the BEM. Investigations requiring removal of the pinnae, or the isolation of the ear canal, as by Walsh *et al.* [189], are easily achievable using the BEM but are clearly not practical with acoustic measurements using a real human subject.

Katz [128, 190] used the BEM to calculate individualised HRTFs with meshes obtained through laser scanning, focussing on the contribution of head and pinnae shape to the HRTF. The work was limited by available computational resources, with a maximum valid frequency of 5.4 kHz. Katz also looked at the acoustic properties of the skin and hair [191], finding the skin to be rigid and the hair to have an absorptive effect that could be included to improve the accuracy of the results, although propagation through hair is not able to be modelled using the BEM.

Gumerov *et al.* [166, 177] used FM-BEM to calculate HRTFs for several models, including the spherical head and a number of artificial heads, up to a maximum valid frequency of 20 kHz. Jin *et al.* [80] also used FM-BEM to calculate the HRTFs of a large database of subjects scanned using magnetic resonance imaging (MRI), culminating in the Sydney-York Morphological and Recording of Ears (SYMARE) database. The database contains meshes suitable for use within the BEM at a range of maximum valid frequencies. It also includes the HRTFs calculated from the meshes (valid to 5.6 kHz for the meshes including torso, valid to 16 kHz for the head-and-ears meshes), as well as acoustically measured HRTFs. Good agreement is demonstrated between acoustic and simulated data.

Not only has the BEM been used as a means of obtaining HRTFs, it is also being widely used to conduct investigations more generally where acoustic measurements are more difficult or time-consuming. Fels *et al.* [169] used the BEM to calculate the HRTFs of children up to a maximum frequency of 6 kHz using meshes derived from photographs. Otani *et al.* [106] simulated HRTFs of a Brüel & Kjær 4128C artificial head up to 20 kHz at 291 radial distances between 0.1 and 3.0 m to examine the source-distance-dependency of the features. Brinkmann *et al.* [192] used the BEM to simulate a high resolution database of different head-above-torso transfer functions for a laser scanned FABIAN head (maximum frequency of 22 kHz). Rui *et al.* [193] and Salvador *et al.* [194] both made use of the BEM to simulate individualised near-field HRTFs up to 16 kHz, with the meshes created from laser and CT scans. Much use of the BEM has also been made in examining the relationship between morphology and spectral features [195–199].

Steady increases in computational power, particularly in cloud computing, combined with the reduced computational demands of FM-BEM, have led to greater use of the BEM in acoustics research. It has become increasingly feasible in recent years to simulate full frequency range HRTFs using the BEM up to 20 kHz.

2.6.3 Other Capture Methods

A range of other techniques have been proposed as a means of capturing both generic and individualised localisation cues without acoustic measurement or computational modelling.

Genuit [200] (cited in [25]) proposed that, as incident sound waves arrive at the ears via multiple paths as a result of interactions with the body, these interactions could be modelled individually. Therefore, HRTFs could be represented by a combination of filter structures, where each filter accounts for the contribution of a corresponding morphological structure. It was also proposed that filter parameters could be adjusted based on anatomical measurements, therefore allowing a generic HRTF to be adapted to a particular listener. Brown and Duda [25] developed this further in creating a time-domain structural model which comprised three stages, where an IIR filter was used to represent the head shadow effect in the azimuth plane, and two FIR filters were used to represent each of the pinna and shoulder contributions to the elevation cue. This model was shown to give similar results to measured HRTFs, but was limited to the frontal region. Algazi *et al.* [81] also created a structural model comprising functional blocks for anthropometric values, with later work [201] focussed more specifically on modelling the contribution of the pinnae.

Other potential shortcuts have also been proposed such as matching a subject to existing HRTFs using morphological parameters [202, 203] and through auditory matching [204], and of the manipulation of artificial head HRTFs by adjusting ITD to match that of the listener [205]. Personalisation of existing databases via anthropometric measurements has also been proposed [206].

Neural networks and machine learning techniques are also being used to generate individualised HRTFs. For example, Yamamoto and Igarashi [207] proposed an algorithm which has been trained on the CIPIC database [24] to identify individualised HRTF features. The system is capable of calibrating a HRTF generator to create individualised HRTFs for a new subject through iterative user feedback. Additionally, Yao *et al.* [208] proposed an algorithm to best match a subject to an existing database using anthropometric features.

It has also been proposed that near-field HRTFs can be synthesised from those measured in the far-field. Kan *et al.* [101, 209] developed a distance variation function (DVF) which can be applied to far-field HRTFs to create a near-field representation of the same angular position. The presented method creates a DVF for each angular position and approximates the variation with distance across the frequency range. Although calculated using a rigid sphere head model, and therefore missing any parallax information introduced by the pinnae, the method was shown to be perceptually effective and an improvement over simple intensity adjustments.

2.7 Reproduction of Acoustic Localisation Cues

Once the acoustic localisation cues for a given direction are known, using the cues to spatialise a mono sound source gives a virtual audio experience similar to that which the listener would have when listening to a real sound source located at that position. This is performed using convolution (Equation 2.25) of a left-right HRIR pair and a mono sound source:

$$b_L(t) = h_L(t) * m(t) \tag{2.58}$$

$$b_R(t) = h_R(t) * m(t) \tag{2.59}$$

where m is the single-channel sound source, h_L and h_R are the left and right of a HRIR pair for a given direction, and b_L and b_R are the binaural signals to be presented to the left and right ears, respectively. In matrix notation, this becomes:

$$[\mathbf{B}] = [\mathbf{H}]M \tag{2.60}$$

$$\begin{bmatrix} B_L \\ B_R \end{bmatrix} = \begin{bmatrix} H_L \\ H_R \end{bmatrix} M \tag{2.61}$$

Multiple instances of convolution (or the frequency-domain equivalent) of sound source signals and HRIR pairs can be combined additively to create acoustic environments with multiple spatialised sound sources. Due to the computation of integration required for time-domain convolution, it can be computationally less expensive to compute the Fourier transforms, take the product, and take the inverse Fourier transform to obtain the same result.

An alternative technique to obtain binaural signals is to record the signals intended for reproduction directly using an artificial subject and the desired sound source at the desired position. This reduces the complexity of the reproduction system, but also reduces the flexibility of the content.

The signals containing the spatialised sound sources b_L and b_R are then presented to the listener using headphones or loudspeakers.

2.7.1 Reproduction over Headphones

The least complex method is to reproduce the virtual spatialised sound sources over headphones, as each channel of the left-right pair can be presented to the intended ear with only the effect of the headphones to compensate for. The effect of the headphones is referred to as the acoustic transfer function matrix $\mathbf{A}$, where the terms are the headphone to ear canal transfer functions (HpTFs) for the left and right ears. The signals at the ears e_L and e_R (E_L and E_R in the frequency domain) are therefore given by:

$$[\mathbf{E}] = [\mathbf{A}][\mathbf{B}] \tag{2.62}$$

$$\begin{bmatrix} E_L \\ E_R \end{bmatrix} = \begin{bmatrix} A_{LL} & 0 \\ 0 & A_{RR} \end{bmatrix} \begin{bmatrix} B_L \\ B_R \end{bmatrix} \tag{2.63}$$

$$\begin{bmatrix} E_L \\ E_R \end{bmatrix} = \begin{bmatrix} A_{LL} & 0 \\ 0 & A_{RR} \end{bmatrix} \begin{bmatrix} H_L \\ H_R \end{bmatrix} M \tag{2.64}$$

where A_{LL} and A_{RR} are the acoustic paths corresponding to the left headphone to left ear and right headphone to right ear, respectively (that is, the HpTFs), and B_L and B_R are the left-right pair of binaural signals. Ideal headphones HpTFs would have a flat frequency response and a linear phase response, allowing the binaural signals to reach the ears with no colouration. In reality, as no headphone has a perfectly flat frequency response, it is necessary to measure HpTFs and apply compensation using an inverse transfer function matrix $\mathbf{I}$ [210]:

$$[\mathbf{E}] = [\mathbf{I}][\mathbf{A}][\mathbf{B}] \tag{2.65}$$

$$\begin{bmatrix} E_L \\ E_R \end{bmatrix} = \begin{bmatrix} I_{LL} & 0 \\ 0 & I_{RR} \end{bmatrix} \begin{bmatrix} A_{LL} & 0 \\ 0 & A_{RR} \end{bmatrix} \begin{bmatrix} B_L \\ B_R \end{bmatrix} \tag{2.66}$$

$$\begin{bmatrix} E_L \\ E_R \end{bmatrix} = \begin{bmatrix} I_{LL} & 0 \\ 0 & I_{RR} \end{bmatrix} \begin{bmatrix} A_{LL} & 0 \\ 0 & A_{RR} \end{bmatrix} \begin{bmatrix} H_L \\ H_R \end{bmatrix} M \tag{2.67}$$

where the product of $\mathbf{I}$ and $\mathbf{A}$ should form an identity matrix:

$$\begin{bmatrix} I_{LL} & 0 \\ 0 & I_{RR} \end{bmatrix} \begin{bmatrix} A_{LL} & 0 \\ 0 & A_{RR} \end{bmatrix} = [1] \tag{2.68}$$

Measurements can be taken of $\mathbf{A}$ using techniques similar to those discussed in Section 2.6.1. Microphones are placed in the ears of the subject (whether human or artificial). The headphones under test are placed on the subject, over the microphones, and excitation signals are reproduced over the headphones. The microphones then record the results which are processed to calculate the left-right HpTF pair. HpTFs are characterised by both the headphone response and the coupling between the headphones and the ears [18, 210]. Removing and replacing the headphones typically alters the transfer functions, particularly at the higher frequencies, which can make them difficult to equalise reliably.

Møller *et al.* [211] measured HpTFs of 14 headphones on 40 human subjects, and found that HpTFs are smooth below approximately 5 kHz. More variation is present above this, as HpTFs are individualised in the same way as HRTFs, and both inter-headphone and inter-individual variation was found. HpTFs measured by Lindau and Brinkmann [212] were described to contain four approximate regions of frequency magnitude response. Below 200 Hz, differences of ±3 dB were attributed to leakage. From 200 Hz to 2 kHz, smaller differences of below 1 dB were observed, with these differences reaching a value of ±3 dB by 5 kHz. This is in agreement with the description given by Møller *et al.* [211]. Larger variations were observed above 5 kHz, as a result of pinnae effects. As large differences between subjects are observed in most measurements, Pralong and Carlile [213] suggested that individualised inverse filters should be used, although, this is not always feasible.

Repeat measurements are often used in an attempt to cover as much variation in headphone fit as possible, as inverse filters based on a single measurement have been reported to not equalise headphone characteristics effectively and, in some cases, create a worse result than with no equalisation [214]. The HpTF is then defined either as an average of the repeated measures (as in [215]) or as a perceptual combination (as in [18]). The terms of the inverse filter matrix **I** can then be derived using one of a number of methods [216], and the signals to be reproduced over the headphones are:

$$\begin{bmatrix} L' \\ R' \end{bmatrix} = \begin{bmatrix} I_{LL} & 0 \\ 0 & I_{RR} \end{bmatrix} \begin{bmatrix} B_L \\ B_R \end{bmatrix} \tag{2.69}$$

2.7.2 Reproduction over Pairs of Loudspeakers

Rather than using headphones, it is also possible to reproduce virtual spatialised sound sources using a pair (or more) of loudspeakers. This can give a more externalised experience (where the sound is perceived to come from a source located outside of the head) [16] and is less invasive, but more care must be taken to preserve the directional information and the quality of the audio.

2.7.2.1 Theory

When binaural signals are reproduced using loudspeakers, not only are the signals affected by the acoustic transfer function matrix **A** between each loudspeaker and its corresponding ipsilateral ear, but additional crosstalk paths are present due to the lack of isolation between the ears and the contralateral loudspeaker. This causes an additional pair of signals to reach the wrong ears. The signals due to the crosstalk paths, shown in Fig. 2.30 as A_{LR} and A_{RL} (that is, from left loudspeaker to right ear, and vice versa), interfere with the intended signals and must be removed in order to preserve the binaural information. The removal of the crosstalk signals is known as crosstalk cancellation (CTC).

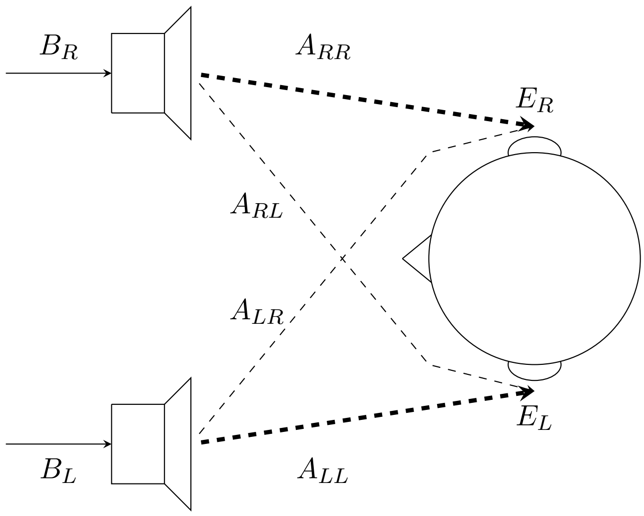


Figure 2.30: Acoustic stages in a loudspeaker reproduction system. When binaural signals **B** are reproduced over loudspeakers, unwanted crosstalk paths A_{LR} and A_{RL} are present. The paths over which the binaural signals are reproduced are indicated in bold.

CTC was first investigated by Atal and Schroeder [217] in the 1960s. Reproduction with a CTC stage is sometimes referred to as 'transaural' reproduction, after work by Cooper and Bauck [218]. The aim of CTC is to allow the pair of binaural signals to reach their intended ear without interference from crosstalk signal paths. It can also compensate for the unwanted colouration of the acoustic transfer function on the binaural signals; for example due to the response of the loudspeakers. In much the same way as headphone reproduction requires an equalisation stage to compensate for colourations due to the HpTFs, in loudspeaker reproduction the binaural signals are filtered prior to reproduction to perform CTC and signal path equalisation.

The underlying principle of CTC can be described as follows, based on the derivation by Xie [91]. The input binaural signals are denoted by B_L and B_R, the acoustic transfer functions between the loudspeakers and the ears are represented by A_{LL}, A_{LR}, A_{RL} and A_{RR} and the signals reproduced at the ears are E_L and E_R. For simplicity, the acoustic transfer functions A_{LL} to A_{RR} can be considered to have flat frequency magnitudes and linear phase responses, so that the paths do not apply colouration to the binaural signals. Without CTC, the binaural signals B_L and B_R do not reach the ears as intended, as they are contaminated by the signals created by the crosstalk paths A_{LR} and A_{RL}. In this case, the signals at the ears E_L and E_R are given by:

$$\begin{bmatrix} E_L \\ E_R \end{bmatrix} = \begin{bmatrix} A_{LL} & A_{LR} \\ A_{RL} & A_{RR} \end{bmatrix} \begin{bmatrix} B_L \\ B_R \end{bmatrix} \tag{2.70}$$

$$\begin{bmatrix} E_L \\ E_R \end{bmatrix} = \begin{bmatrix} A_{LL} & A_{LR} \\ A_{RL} & A_{RR} \end{bmatrix} \begin{bmatrix} H_L \\ H_R \end{bmatrix} M \tag{2.71}$$

To compensate for this, the binaural signals are filtered prior to reproduction using CTC filters C_{11}, C_{12}, C_{21}, and C_{22} to produce the pre-filtered loudspeaker signals L' and R', as illustrated in Fig. 2.31.

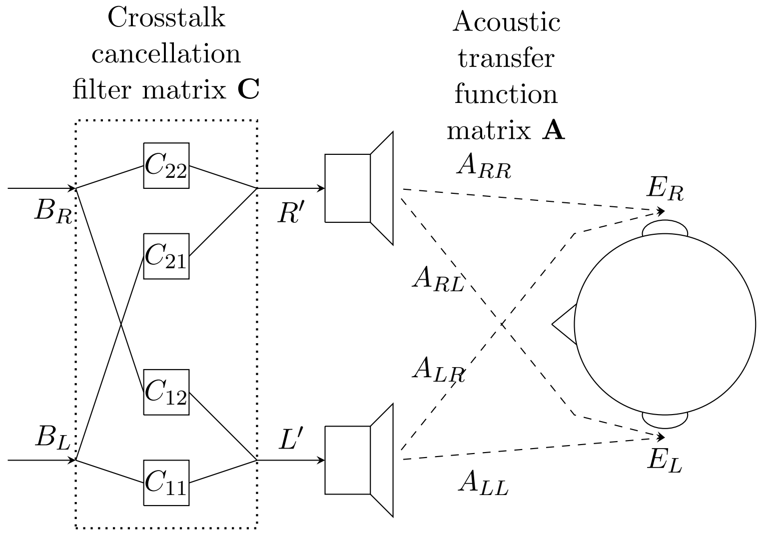


Figure 2.31: Acoustic stages in a loudspeaker reproduction system with crosstalk cancellation. A cancellation matrix **C** removes the acoustic transfer function matrix **A**, so that the signals produced at the ears **E** match the input binaural signals **B**.

This gives loudspeaker signals:

$$\begin{bmatrix} L' \\ R' \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} B_L \\ B_R \end{bmatrix} \tag{2.72}$$

The signals reproduced at the ears can then be expressed as:

$$\begin{aligned} \begin{bmatrix} E_L \\ E_R \end{bmatrix} &= \begin{bmatrix} A_{LL} & A_{LR} \\ A_{RL} & A_{RR} \end{bmatrix} \begin{bmatrix} L' \\ R' \end{bmatrix} \\ &= \begin{bmatrix} A_{LL} & A_{LR} \\ A_{RL} & A_{RR} \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} B_L \\ B_R \end{bmatrix} \end{aligned} \tag{2.73}$$

For **E** to equal **B**, the product of the two 2 x 2 matrices in Equation 2.73 must equal an identity matrix:

$$\begin{bmatrix} A_{LL} & A_{LR} \\ A_{RL} & A_{RR} \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{2.74}$$

Therefore, the required CTC filters are the inverse of the acoustic transfer functions between the loudspeakers and the ears, and the signals output from the loudspeakers are exactly reproduced at the ears. If **A** is non-singular, and therefore invertible, the CTC filters required are calculated by:

$$\begin{aligned} \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} &= \begin{bmatrix} A_{LL} & A_{LR} \\ A_{RL} & A_{RR} \end{bmatrix}^{-1} \\ &= \frac{1}{A_{LL}A_{RR} - A_{LR}A_{RL}} \begin{bmatrix} A_{RR} & -A_{LR} \\ -A_{RL} & A_{LL} \end{bmatrix} \end{aligned} \tag{2.75}$$

Therefore, the signals to be reproduced at the loudspeakers can be given as:

$$\begin{bmatrix} L' \\ R' \end{bmatrix} = \frac{1}{A_{LL}A_{RR} - A_{LR}A_{RL}} \begin{bmatrix} A_{RR} & -A_{LR} \\ -A_{RL} & A_{LL} \end{bmatrix} \begin{bmatrix} B_L \\ B_R \end{bmatrix} \tag{2.76}$$

If the binaural signals are to be obtained through convolution of a mono input signal with a HRIR pair, the spatialisation and CTC stages can be combined. Substitution of Equation 2.61 and Equation 2.75 into Equation 2.72 gives:

$$\begin{bmatrix} L' \\ R' \end{bmatrix} = \frac{1}{A_{LL}A_{RR} - A_{LR}A_{RL}} \begin{bmatrix} A_{RR} & -A_{LR} \\ -A_{RL} & A_{LL} \end{bmatrix} \begin{bmatrix} H_L \\ H_R \end{bmatrix} M \tag{2.77}$$

This can be expanded to:

$$L' = K_L M \qquad R' = K_R M \tag{2.78}$$

where

$$\begin{aligned} K_L &= \frac{A_{RR}H_L - A_{LR}H_R}{A_{LL}A_{RR} - A_{LR}A_{RL}} \\ K_R &= \frac{-A_{RL}H_L + A_{LL}H_R}{A_{LL}A_{RR} - A_{LR}A_{RL}} \end{aligned} \tag{2.79}$$

It is also possible to derive the CTC filters in the time domain, as done by Kirkeby *et al.* [219], which results in a decaying train of delta functions. This indicates that CTC filters are inherently recursive [16, 217, 219], although this is not generally a problem provided the inversion is causal and stable [219].

2.7.2.2 Channel Separation

A measure of the effectiveness of a CTC system is the achieved channel separation, that is, how much suppression has been applied to the crosstalk signal [220–222]. Channel separation (CS) is independent of the signals to be reproduced, and can be defined as the magnitude ratio between the direct signal and the crosstalk at a frequency point f [223]:

$$CS_{L_f} = \left| \frac{A_{LL_f}}{A_{LR_f}} \right| \tag{2.80}$$

$$CS_{R_f} = \left| \frac{A_{RR_f}}{A_{RL_f}} \right| \tag{2.81}$$

where A_{LR_f} and A_{RL_f} are the crosstalk paths at frequency point f, and A_{LL_f} and A_{RR_f} are the direct paths. As CS is frequency dependent, an additional metric describes the performance over a frequency region. The channel separation index (CSI) is defined as the average CS over the frequencies of interest [224]:

$$CSI_L = \frac{1}{N_f} \sum_{j=1}^{N_f} 20 \log_{10} |CS_{L_j}| \tag{2.82}$$

$$CSI_R = \frac{1}{N_f} \sum_{j=1}^{N_f} 20 \log_{10} |CS_{R_j}| \tag{2.83}$$

where N_f is the number of points in the frequency region of interest. In an ideal CTC system, the CSI is very large, as the signal energy in the contralateral path should be very small compared to the energy in the ipsilateral path. For a perfect CTC system, $CSI = \infty$. Individualised CTC, where the filters have been calculated based on the HRTFs of the subject, has been shown to achieve high levels of separation [225, 226]. However, as Masiero [227] notes, individualised CTC can actually be either matched or un-matched, for example in the case where individualised HRTFs have been used in CTC filter calculation but the acoustic transfer has changed slightly due to subject movement. Individualised but unmatched CTC was shown to give similar localisation performance to non-individualised CTC [227].

CS and CSI can either be quoted as a positive value, where 0 is no suppression and $+\infty$ is infinite suppression, or as a negative value, where 0 is no suppression and $-\infty$ is infinite suppression. In this thesis, the positive convention will be used, which is in agreement with Akeroyd *et al.* [226] but conflicts with the convention used by Lacouture Parodi and Rubak [228] and Bai and Lee [223].

Natural channel separation (NCS) is the inherent acoustic isolation arising between the ears for a sound source of a given direction, and occurs as a result of the acoustic shadowing introduced by the head [19]. Since NCS effectively describes the channel separation between the ears prior to the implementation of CTC, it is often the ground truth value against which the success of a CTC system is measured. NCS varies as a function of frequency, direction and distance [21, 229, 230]. The term has, however, been used by researchers to mean both NCS calculated at a single frequency point (analogous to CS) and NCS calculated as an arithmetic mean over a frequency range (analogous to CSI) [21, 228, 231]. Here, the distinction is explicitly made between NCS$_f$ (at a frequency point) and NCS (average over a frequency range). NCS$_f$ is calculated in the same way as CS and CSI, as the ratio between the ipsilateral and crosstalk signals. For a left-pair of loudspeakers (such as those shown in Fig. 2.30), NCS$_f$ is calculated as:

$$NCS_{L_f} = \left| \frac{A_{LL_f}}{A_{LR_f}} \right| \tag{2.84}$$

$$NCS_{R_f} = \left| \frac{A_{RR_f}}{A_{RL_f}} \right| \tag{2.85}$$

$$NCS_L = \frac{1}{N_f} \sum_{n=1}^{N_f} 20 \log_{10} |NCS_{L_n}| \tag{2.86}$$

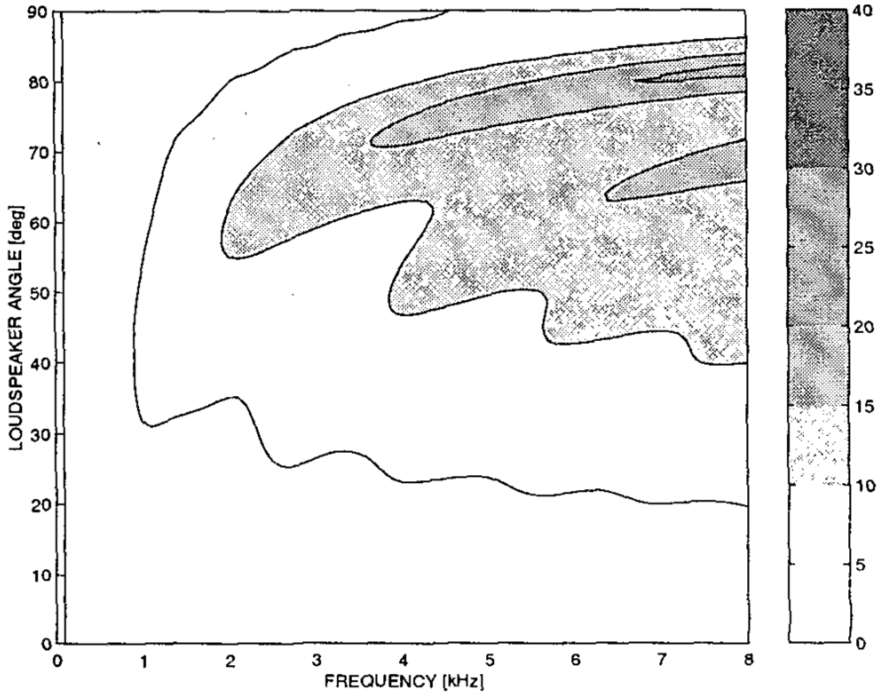
$$NCS_R = \frac{1}{N_f} \sum_{n=1}^{N_f} 20 \log_{10} |NCS_{R_n}| \tag{2.87}$$

In the log domain, this can be implemented as a subtraction:

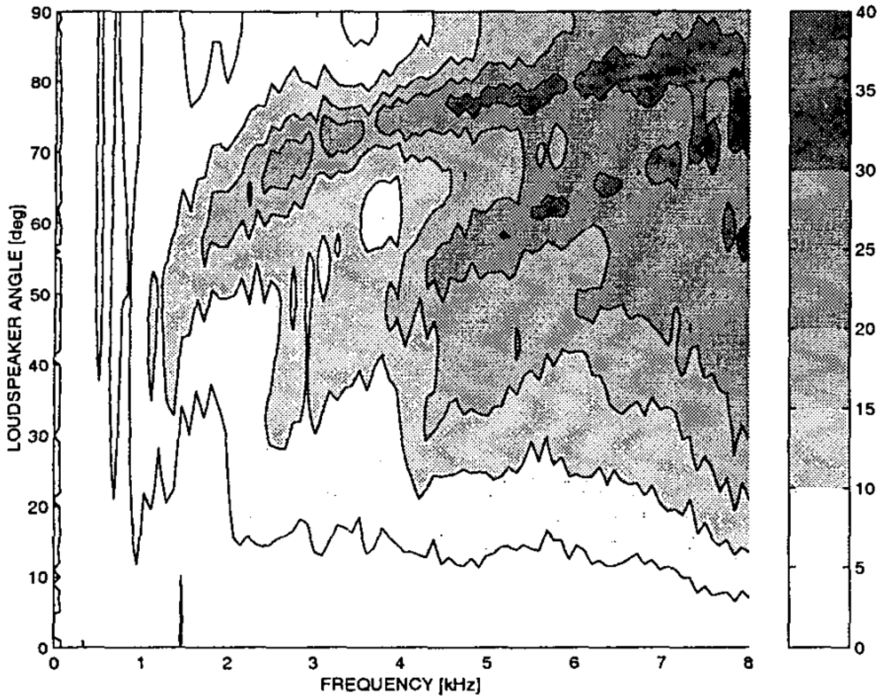
$$NCS_L = \frac{1}{N_f} \sum_{n=1}^{N_f} 20 \log_{10} |A_{LL_n} - A_{LR_n}| \tag{2.88}$$

$$NCS_R = \frac{1}{N_f} \sum_{n=1}^{N_f} 20 \log_{10} |A_{RR_n} - A_{RL_n}| \tag{2.89}$$

Fig. 2.32 shows NCS as a function of frequency and loudspeaker angle, as reported by Orduña-Bustamente *et al.* [21] for both a rigid sphere and for a Brüel & Kjær HATS. Whether NCS_f or NCS has been used is not reported; neither is distance. Low values occur for low frequencies and loudspeaker positions towards the front, and increase non-monotonically with both frequency and loudspeaker direction. Values for the Brüel & Kjær HATS are broadly similar to those calculated for the spherical head, with additional complexity as a result of the presence of pinnae. Masiero [227] calculated NCS values as the average within the frequency band 200 Hz to 8 kHz for 8 subjects across two sets of measured HRTFs for a loudspeaker at 45°. The values ranged from 13.8 to 16.4 dB, with the averages over the 8 subjects for the two HRTF measurements as 14.8 dB and 14.5 dB, respectively.



(a) Between two opposite points on a rigid sphere of radius 8.75 cm



(b) Between the ears of a Brüel & Kjær HATS

Figure 2.32: NCS as a function of frequency and loudspeaker direction, as presented by Orduña-Bustamente *et al.* [21]. Higher values are exhibited for larger loudspeaker angles and higher frequencies. Reproduced from [21] with permission.

Ideal CTC aims to completely remove the crosstalk paths and create theoretically infinite channel separation. However, it has been demonstrated that sufficient separation, rather than infinite, enables the binaural cues to be preserved [226, 231]. Ahrens *et al.* [17] used so-called 'gentle' CTC to create loudspeaker-based channel separation in the order of 20 dB, which informal listening tests suggested was sufficient to achieve lateralisation. In a headphone-based experiment which introduced crosstalk at increasing levels, Lacouture Parodi and Rubak [228] determined that a minimum separation of 15 dB is required to maintain the binaural information for most stimuli, increasing this threshold to 20 dB for broadband sources. These conclusions indicated that previous suggestions of minimum requirements of 10 to 12 dB [223] could be increased, where the 20 dB limit is more suitable due to the broadband nature of most binaurally-reproduced stimuli [228].

2.7.2.3 Limitations and Alternatives

There are downsides associated with the implementation of CTC, however. CTC can introduce spectral colouration which strains the reproduction system and affects the experience of the listener [15, 16]. Regularisation can be used to control this, but can introduce undesirable artefacts such as narrow-band boosts at high frequencies and a low frequency roll-off [16, 227]. To some extent, frequency-dependent regularisation can be used to avoid this, but results in a more complex solution [16]. Errors can also be caused by inversion problems, which can arise due to singular values in the filter design [19].

In the same way that binaural reproduction gives the best listener experience if individualised HRTFs are used, it has been shown that CTC is more effective if individualised acoustic transfer functions are used [19, 225, 226, 232], with the impact being more significant for sound sources located outside the loudspeaker span [227]. Akeroyd *et al.* [226] found that the system could not deliver enough cancellation to reliably preserve the binaural cues when mismatched HRTFs were used; the filters needed to be redesigned for each listener to be effective.

Exact CTC systems typically have a small 'sweet spot', the region in space where the cancellation performs as expected and the system is robust. If the listener's head position deviates too far from this area (more than approximately 75–100 mm [233], or 0.25–0.5 of a wavelength [91]) the difference between the cancellation filters and the actual transfer functions becomes too great and the system deteriorates, often in an unpleasant and distracting way [17]. However, head movement is important for resolving front-back confusions that can occur with non-individualised binaural audio [59] and so are important to account for.

Head tracking can be used to update the cancellation filters as required. This is typically known as dynamic CTC, and requires either the storing of multiple acoustic transfer functions to dynamically recalculate the filters, or interpolation between known filters [232–235]. Latency is inevitable in a system that requires updating with each movement, but Lentz [234] demonstrated that it is possible to keep latency at a low enough value not to be noticeable. It is also possible to negate the requirement for dynamic CTC by moving the sources with the head, such as in the wave field synthesis (WFS) system developed by Theile, Fastl and colleagues [236–238], or the beamforming approach by Guldenschuh and Sontacchi [239].

An alternative to a dynamic CTC system is to determine optimal positions for the loudspeakers to maximise robustness to head movement. Ward and Elko [240] demonstrated that the ideal loudspeaker angle varies with frequency, with a smaller span being more stable at high frequencies. This is the concept behind the stereo dipole system [219], and subsequently, the optimal source distribution (OSD) system [220, 241, 242].

The stereo dipole, developed by Kirkeby *et al.* [219], consists of two closely-spaced loudspeakers as shown in Fig. 2.33. This setup was shown to be robust to head movement, but Lopez and Gonzales [221] concluded that using loudspeakers in this formation removes any head shadowing effect present. In contradiction with Kirkeby *et al.* [219], however, Prodi and Velecka [231] concluded that a 30° span was more robust than a 10° span, although increasing the span further decreased robustness again. They therefore proposed that an optimum span existed, after which the robustness decreased. The OSD approach develops this further to create a practical solution for the dependence of the ideal loudspeaker angle on frequency, where multiple loudspeaker pairs at different spans are required, each reproducing a section of the frequency spectrum.

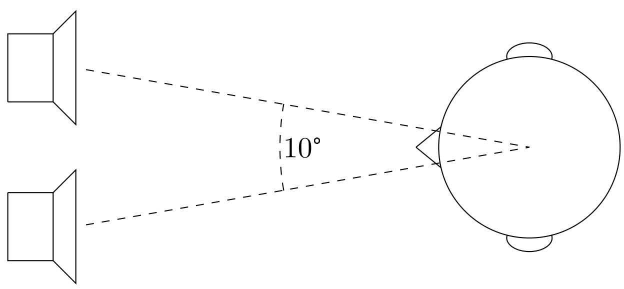


Figure 2.33: The stereo dipole loudspeaker configuration uses a pair of loudspeakers with a relatively small angular separation (shown here with a separation of 10°).

Han *et al.* [243] evaluated the robustness to head rotation of 52,650 loudspeaker pairs (every paired combination of 325 positions). Loudspeaker positions above the horizontal plane were found to be generally less susceptible to errors as a result of head rotation, with the most robust pair of two loudspeakers being at $\theta = 100°,\ \phi = 50°$ and $\theta = 220°,\ \phi = 80°$. In general, robustness to head movement improved if the two loudspeakers were positioned at azimuth values near to the listener's ear ($100° \leq \theta \leq 110°$ and $210° \leq \theta \leq 260°$). The least robust loudspeaker pair was found to be at $\theta = 0°,\ \phi = 80°$ and $\phi = 90°$ (directly overhead, so no azimuth angle). Whilst these results are interesting with regards to implementation, they bring into question the practicality of loudspeakers paired in such asymmetric positions, particularly at higher elevation angles.

It has been demonstrated, however, that it is possible to implement a binaural reproduction system without including a CTC stage. As the loudspeakers approach the head, NCS increases due to the greater head shadowing [19, 21, 22]. To take advantage of this, some researchers have investigated binaural reproduction using loudspeakers positioned close to the head.

The Internet Chair [244–246] is a networked rotating chair which uses positional information to inform input and output signal changes, such as vehicle driving simulators with spatialised navigational instructions, or as part of a networked conference call. Initial prototypes used headphone-based binaural reproduction, but later prototypes used ‘nearphones’: small loudspeakers attached to the headrest, as shown in Fig. 2.34a [245, 246]. CTC was not implemented as the shadowing provided by the head was large enough to preserve the binaural cues [247].

Jones *et al.* [22] and Elliott *et al.* [23] used a pair of small loudspeakers mounted approximately $\pm 90°$ to the head at a distance of approximately 20 cm (shown in Fig. 2.34b) to investigate the effect of head movement in a near-field binaural system. Two CTC approaches were implemented (exact and space-averaged), plus a gain adjustment to account for head movement. The CTC implementations did not perform significantly better than those using no processing at all, and listening tests indicated that CTC was not subjectively necessary. As part of an earlier implementation of the project, Jones *et al.* [22] briefly discuss the possible impact of headrest design on the results, and the fact that no attempt was made to increase the channel separation using headrest design in that particular study. It seems apparent that, for near-field binaural loudspeaker reproduction, consideration of the design of the loudspeaker system itself, including its mounting and the materials used, is of greater importance than it is for far-field loudspeaker reproduction.

(a) Second generation prototype of the Internet Chair, as used by Cohen *et al.* [245]. The white ‘nearphones’ can be seen either side of the headrest. Reproduced from [245] with permission.

(b) Near-field binaural experimental setup used by Elliott *et al.* [23]. The loudspeakers are placed approximately 20 cm from the head. Reproduced from [23] with permission.

Figure 2.34: Examples of near-field loudspeaker implementation.

Puomi *et al.* [20] investigated the potential of a commercial device for binaural reproduction, the HUMU Augmented Audio Cushion$^{\text{TM}}$ [248] (shown in Fig. 2.35). Several CTC strategies were implemented in combination with the binaural reproduction process, including CTC and ipsilateral-only cancellation. In this work, the use of near-field HRTFs in the derivation of the CTC filters was reported to create strong and unpleasant spectral colouration, and so the authors instead derived CTC filters using far-field HRTFs measured at the same angles. While this approach reportedly produced more stable filters, the assumption that far-field HRTFs and near-field HRTFs share similar pinnae effects is questionable given the distance-dependent differences discussed in Section 2.5.5, namely the variation in centre frequency, amplitude and bandwidth of features [105, 106], the presence of additional features in the near-field [102, 104, 106] and the presence of a parallax effect [105, 107].

Ipsilateral-only cancellation was additionally implemented on the basis that the head would provide adequate acoustic shadowing to maintain the required channel separation. Results indicated that CTC did not work effectively, probably because any mismatch between CTC filters and HRTFs reduces the channel separation [226] and, therefore, causes the binaural cues to break down. Ipsilateral-only cancellation moved the virtual sound source away from the loudspeaker positions for some stimuli and for some listeners. However, it did not have the desired effect of creating a frontal virtual sound source. As the authors note, this was likely due to only 9 dB of natural contralateral attenuation.

Figure 2.35: HUMU Augmented Audio Cushion investigated by Puomio *et al.* [20]. Reproduced from [20] with permission.

2.8 Conclusion

This chapter presents an overview of topics relevant to the work presented in this thesis within the fields of acoustics and psychoacoustics, including: the mechanisms for the production and propagation of sound; the anatomy and function of the human auditory system; and the origin, capture and reproduction of acoustic cues used in the localisation of sound sources. The principle aim of this thesis is to investigate the feasibility of implementing a near-field binaural loudspeaker reproduction system in which CTC is not required, with a particular focus on the impact of a listener's morphology on candidate loudspeaker positions. Therefore, the three key areas with respect to the work presented in the remainder of this thesis are:

- the formation of individualised HRTFs as a result of acoustic interactions between sound waves and the morphology of the listener;
- the theory and implementation of the BEM, and its use for calculating HRTFs;
- the possibility of avoiding the requirement for CTC through informed placement of loudspeakers in the vicinity of the head.

Chapter 3

Creation of a BEM-Suitable Mesh Model of KEMAR

Every morning, it's a cleanup.

Tegan and Sara, *Underwater*
If It Was You (2002)

3.1 Introduction

As discussed in Chapter 2, reproduction of binaural signals over loudspeakers has typically required the use of crosstalk cancellation (CTC) to artificially increase the channel separation and ensure that the correct information reaches the correct ear. However, the channel separation increases naturally as loudspeakers are brought closer to the head and this effect can be exploited. One of the aims of this research is to identify loudspeaker positions which have a large enough channel separation to support binaural reproduction without requiring CTC. In order to identify these positions, a large number of pairs of head-related transfer functions (HRTFs) at a substantial number of angles and distances is required. The large numbers involved necessitate the calculation of these HRTFs by simulation using the boundary element method (BEM). In these circumstances, the BEM is faster and more flexible than taking the direct approach of making many acoustic measurements. As discussed in Section 2.6.2, simulation using the BEM requires a 3D surface mesh of the subject for which the HRTFs are required. The mesh must meet several requirements, including how the mesh is defined, topological accuracy, and maximum edge length.

It was decided that it would be most useful to perform these simulations with the Knowles Electronics Manikin for Acoustic Research (KEMAR), rather than a mesh of a human subject or of a different artificial head such as the Neumann KU100 [96]. KEMAR is a mathematical median of over 4,000 humans [98] and, therefore, would give results that would be indicative of a variety of listeners rather than one specific listener. Additionally, the presence of shoulders and torso (which the KU100 does not possess) is likely to have an impact in the near-field.

Due to limitations present in existing KEMAR meshes resulting in the requirements for BEM simulation not being met, it was decided that the creation of a new mesh from which others could be derived was the most appropriate course of action. The goal was to create a mesh model which was consistent with a physical binaural manikin, was numerically valid for use within the BEM, and produced results consistent with acoustic measurements. However, two limiting factors exist when considering the creation of a mesh for computational simulation: the mesh acquisition method and the resolution.

This chapter describes the workflow designed to create a BEM-suitable mesh model of KEMAR using various software packages. The mesh is numerically validated following a workflow adapted from that described by Jin *et al.* [80], and acoustically validated using comparisons between simulated and acoustically measured HRTFs.

3.2 Mesh Acquisition

Existing work using the BEM has used typically scanning techniques to create virtual meshes of the manikin or subject in question. As discussed in Section 2.6.2, these have usually been one of laser [106, 160, 167], scattered light [168], magnetic resonance imaging (MRI) [80, 162] or computerised tomography (CT) [106, 173] scans, sometimes in combination. Scanning techniques can have limitations, however. Those which rely on optical triangulation are restricted to line-of-sight, which results in occluded areas (such as pinna cavities and behind the pinna) being seen as filled unless extensive effort is made to overcome this [106, 161, 168, 190]. Some techniques can require altering the subject to improve the scanning results (for example, the application of white paint to KEMAR [249]), or taking separate scans of pinnae using plaster casts taken from the subject [168] or from multiple scans using different techniques [106, 161]. Scans can take a large amount of time, meaning any movement of the subject can have a detrimental effect on the result [80], and all have differing average accuracy and resolution specifications [250].

Scanning technologies can also require a large amount of processing to correct the output of the scanner before any subsequent processing is performed to produce a mesh suitable for the BEM. Much of this requires human input to confirm or correct mistakes, including joining and closing meshes where the scan is incomplete, removing nebulous vertices, filling holes, and coarsening or refining scan results [80, 128, 249], and in some cases the alignment and merging of multiple scans [168, 249]. Whilst the mesh can be processed to be numerically valid to a higher frequency by dividing the faces and, therefore, shortening the edge lengths, the limitations introduced by the initial scan may produce results which differ from those measured acoustically as the surface shape is not accurate.

With the aim of avoiding many of these problems, a CAD file in STEP format (.stp, defined by ISO 10303-21:2016 [251]) of the second generation KEMAR with large pinnae was obtained from GRAS Sound & Vibration, the current manufacturer of KEMAR. By circumventing the need to scan the physical manikin, the mesh model would not suffer the disadvantages associated with some scanning techniques, and higher consistency between reality and simulation would be possible[1].

[1]Due to unavoidable mechanical tolerances, physical manikins produced from the CAD file will

A mesh file describes the shape of a 3D object by breaking up the surface of the object into a series of regions referred to as *faces*, which can be any shape and have any curvature. The shape and curvature vary between mesh definitions and usage requirements. The coordinates which define the corners of each face are referred to as *vertices*, each vertex being shared between several faces. Thus, the boundary of a 3D shape can be described using a series of coordinates (the vertices) and the connectivity between them (the face edges). The CAD file defines the entirety of KEMAR using 2,812 large faces defined by Cartesian coordinate points and curved splines known as boundaries, as shown in Fig. 3.1. However, this type of definition is unsuitable for use with the BEM.

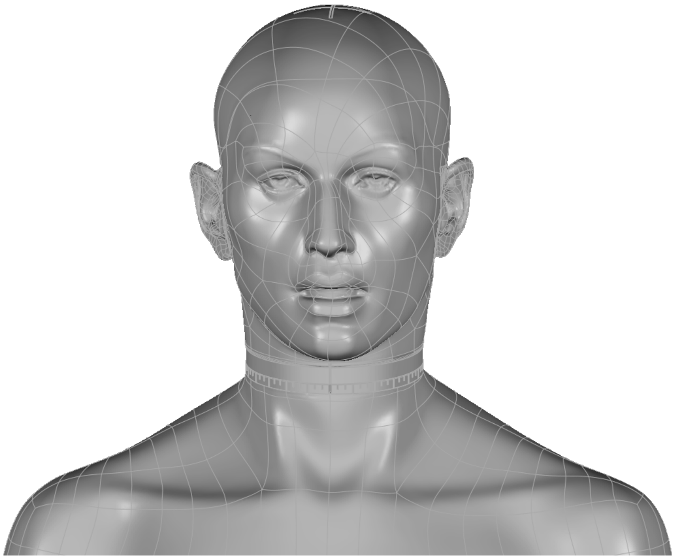


Figure 3.1: Head and shoulders portion of the KEMAR CAD file, defined using curved boundaries (the white lines) and large curved faces. Reproduced from [99] with permission.

3.3 Mesh Resolution Requirements

For a mesh to be suitable for use within a BEM calculation, the surface must be closed (no holes) and discretised into small, typically planar (having no curvature) faces. The size of the faces (more specifically, the length of the edges) determines the maximum valid frequency of the resulting simulation, according to Equation 2.57, reproduced here for reference:

$$f_{max} = \frac{c}{d_{max} \times \eta} \tag{3.1}$$

where f_{max} is the maximum valid frequency in Hz, c is the speed of sound in $\mathrm{m\,s^{-1}}$, d_{max} is the length of the longest element edge in metres, and η is the number of edges per wavelength. Smaller faces, therefore, give a higher valid maximum frequency but can lead to meshes containing large numbers of faces.

contain small additional differences. Importantly, however, the CAD file is likely to be the best description of the surface which can be obtained.

The optimal value of η is a subject of discussion within the literature. A value of 6 has long been established as the lower limit of acceptability for numerical accuracy [175, 176]. This likely came about from an extension of Shannon's sampling theorem, as discussed by Schmiechen [252] (cited in [175]). Shannon's sampling theorem states that two points per wavelength are necessary to detect the frequency, however Schmiechen stated that accuracy is improved with a factor of three to five, giving the lower limit of six points per wavelength [252]. More recently, values as low as $\eta = 3$ have been used, based on the numerical difference between $\eta = 3$ and $\eta = 5$ discretisation being below the perceptual threshold of 1 dB [106]. Non-uniform edge lengths are increasingly finding favour, particularly as a mechanism of reducing the time of computation [165, 253]. However, this complicates the calculation of maximum valid frequency over the whole mesh and requires a different meshing procedure. It can also be computationally more efficient to use multiple meshes at different maximum valid frequencies across the frequency range [192, 193]. $\eta = 6$ is used in this work based on the numerical validity shown in [175], and use of this value within Equation 2.57 is referred to as the *six-edges-per-wavelength rule of thumb*.

Equation 2.57 shows that the maximum valid frequency of a mesh is also dependent on the speed of sound within the medium. The speed of sound varies with temperature (see Section 2.2), therefore, either the temperature or the speed of sound must be known to correctly calculate the maximum valid frequency of a mesh. The generally-accepted value for air temperature used in acoustic calculations is 20 °C, giving a speed of $343.4\,\mathrm{m\,s^{-1}}$. Unless stated otherwise, this value of c is used throughout this work. For example, for a mesh to be valid to 20 kHz at $c = 343.4\,\mathrm{m\,s^{-1}}$, the maximum edge length anywhere in the mesh calculated using the six-edges-per-wavelength rule of thumb must be below 2.86 mm. For 16 kHz validity, the limit increases to 3.58 mm.

Consequently, processing is required to produce a mesh with edge lengths that meet a desired maximum frequency. A number of algorithms and software packages exist for this purpose; for example, Kahana *et al.* [160] used a decimation algorithm by Johnson and Hebert [254], and Jin *et al.* [80] used a series of software packages, including ACVD [255, 256] and Geomagic Studio (discontinued, replaced by Geomagic Wrap [257]).

However, the number of faces permitted within a mesh has often been constrained by the computational resources available to run the simulation. Meshes with smaller faces (and, therefore, more of them) require more random-access memory (RAM) at simulation. Historically, this has restricted the BEM to fairly low-frequency calculations: 5.4 kHz in the case of Katz [128], 5.6 kHz in the case of [80] and 10 kHz in the case of [160]. Recent advances in computational power, particularly in supercomputers and cloud computing services, have allowed calculation up to higher frequencies (for example, to 16 kHz [194] or 20 kHz [167]), in addition to faster formulations such as fast multipole boundary element method (FM-BEM) as discussed in Section 2.6.2. Therefore, as it is now possible to use the BEM to simulate HRTFs up to 20 kHz, meshes which are accurate to higher frequencies are needed.

3.4 Conversion Process Workflow

A workflow was designed to process the KEMAR CAD file into a suitable format for use within the BEM, with edge lengths short enough to allow a maximum valid frequency of 20 kHz. Following the work of Jin *et al.* [80], a series of software applications were used. Geomagic Wrap [257] was used to first convert the CAD file to a polygonal mesh (referred to as the *original-polygonal* mesh) with 200,014 vertices and 400,001 faces. However, this produced a mesh with problematic face shapes and an irregular distribution of the faces, with regions of larger and smaller faces. Whilst the majority of edge lengths are below 5 mm, the maximum edge length in this mesh is 28.34 mm, shown in Fig. 3.2. The presence of longer edges is due, in part, to the representation of the flat underside of KEMAR with a relatively small number of larger faces, as is shown in Fig. 3.3.

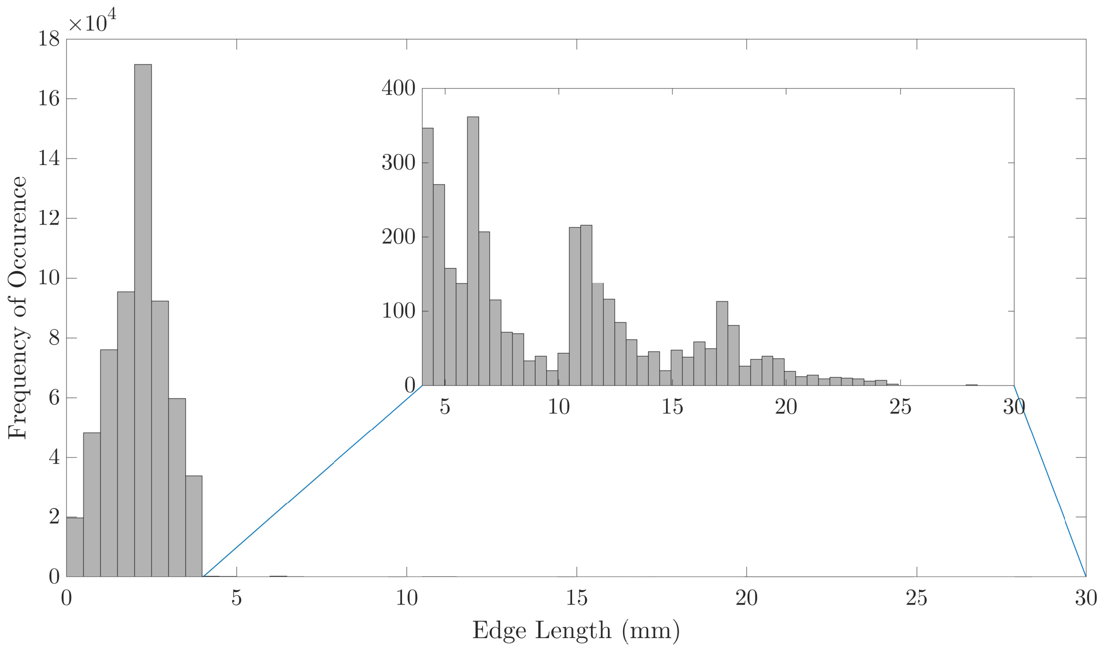


Figure 3.2: Distribution of edge lengths in the *original-polygonal* mesh. The shortest edge length is 1.2×10^{-3} mm, the longest is 28.34 mm and the median is 2.13 mm.

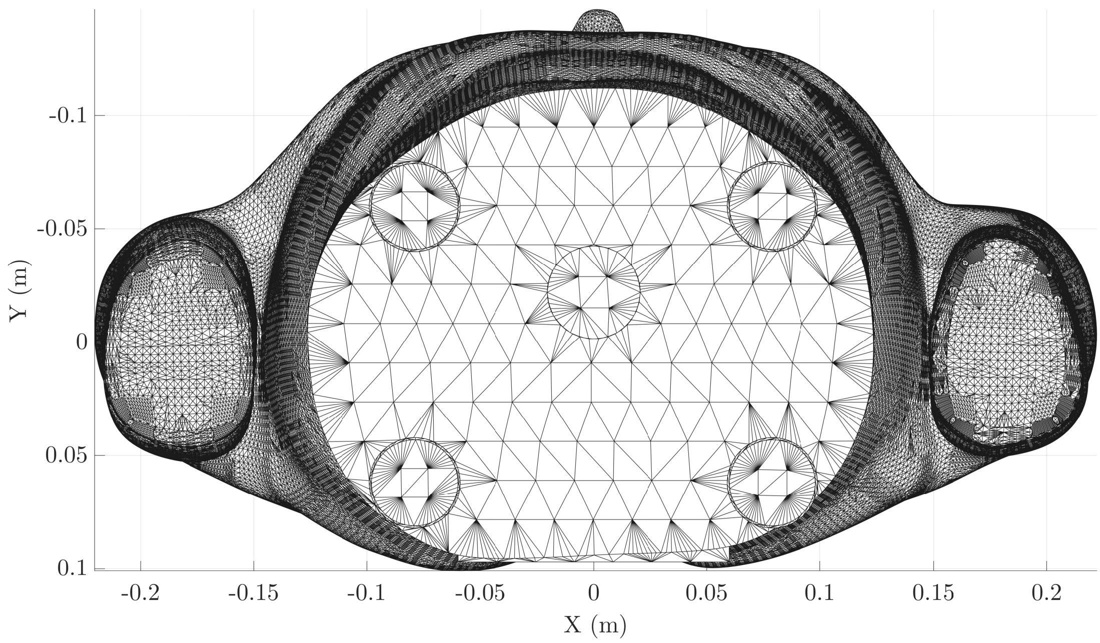


Figure 3.3: Underside view of the *original-polygonal* mesh. Large faces define the flat underside.

Fig. 3.4 shows the distribution of angles between edges in the *original-polygonal* mesh. The smallest angle is 0°, the largest is 180° and the median is 58.8°. A method for calculating the angle between edges can be found in Appendix B.1. The wide range of angles suggests the presence of skinny triangles which must be removed for optimal BEM simulation.

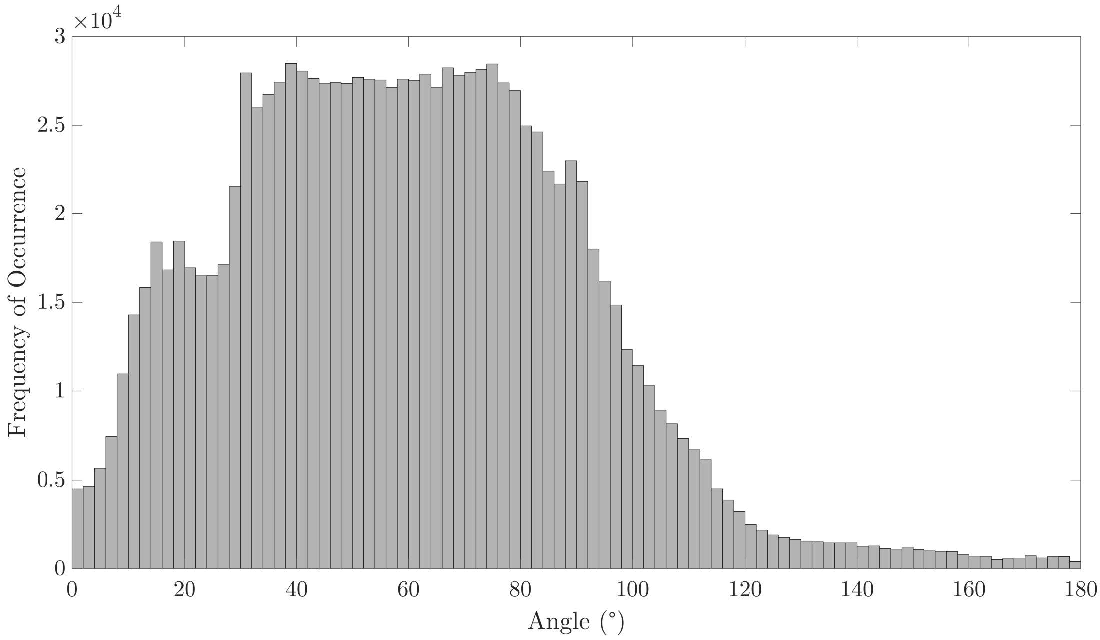


Figure 3.4: Distribution of angles between edges in the *original-polygon* mesh. The smallest angle is 0°, the largest is 180° and the median is 58.8°.

The 'Remesh' tool within Geomagic Wrap was used to redefine the surface mesh with a target edge length of 2 mm. This value was selected (rather than the slightly larger value of 2.86 mm required for 20 kHz validity) as the remesh process uses this as a target rather than a maximum, and so has the capacity to produce edge lengths over the specified value. The remesh process resulted in a mesh with 388,587 vertices and 566,777 faces, referred to as the *remeshed-2mm* mesh. However, the process was found occasionally to introduce artefacts, such as disconnected vertices and duplicated faces. To ensure an efficient mesh, the cleaning features of MeshLab [258, 259] were used to produce a mesh with a similar number of faces (566,759), but with fewer vertices (283,397).

The range of edge lengths was much improved in this mesh (Fig. 3.5); 99 % are under the desired maximum length of 2.86 mm. The longest is only 4.04 mm, giving a maximum valid frequency of 14.2 kHz calculated using the six-edges-per-wavelength rule of thumb. Additionally, edge lengths greater than 2.86 mm are distributed across the mesh (Fig. 3.6), as opposed to being concentrated in any particular region, as before. However, the shape of some faces is undesirable, as approximately equilateral triangles improve the computational accuracy of the BEM [260]. Additionally, the 'Remesh' tool gives no control over the relative distribution of vertices and so has a tendency to produce clusters of vertices across the mesh.

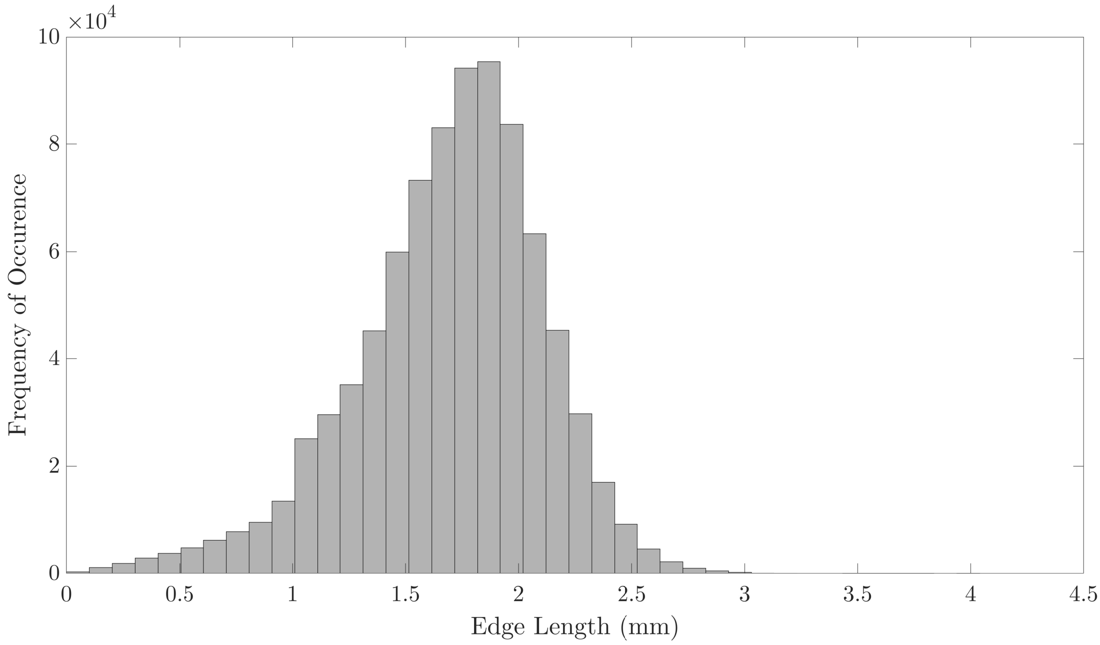


Figure 3.5: Distribution of edge lengths in the *remeshed-2mm* mesh. The shortest edge length is 6.1×10^{-3} mm, the longest is 4.04 mm and the median is 1.74 mm.

(a) Front (b) Left (c) Back (d) Right

Figure 3.6: Views of the *remeshed-2mm* mesh. Edges greater than 2.86 mm (indicated in black) are distributed over the whole mesh, rather than being concentrated in specific regions.

The open source software ACVD [255, 256] was used to create a more consistent distribution of approximately-equilateral triangles across the mesh. ACVD redistributes the vertices on a mesh using a target number of vertices and a 'gradation factor' (the influence of local curvature, where zero is uniform sampling). A uniform distribution was chosen despite evidence of computational gains from non-uniform distributions [165] as this mesh was intended to be a 'gold standard' from which other meshes could be derived. By maintaining a high uniform resolution and validity to 20 kHz using the six-edges-per-wavelength rule of thumb, this mesh would be easy to downsample for use at different analysis frequencies and to adapt to have a non-uniform distribution of faces in the future. Thus, gradation was set to zero.

Using a target value of 360,000 vertices gave a mesh with 360,018 vertices and 720,015 faces, with an improved range of edge lengths (Fig. 3.7). This mesh is referred to as the *full-torso-20* mesh. The maximum edge length was reduced to 2.49 mm (valid to 23.0 kHz using the six-edges-per-wavelength rule of thumb), with a median edge length of 1.51 mm and a minimum of 0.24 mm. This value was chosen rather than the intended value of 20 kHz as it allowed for a margin of error on future processing.

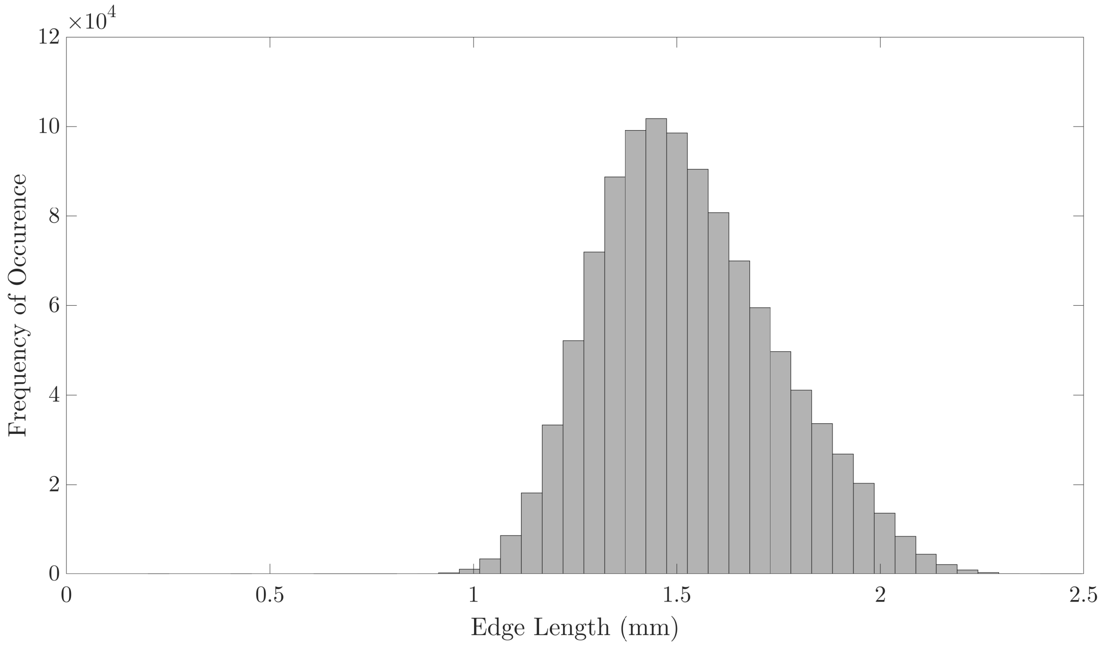
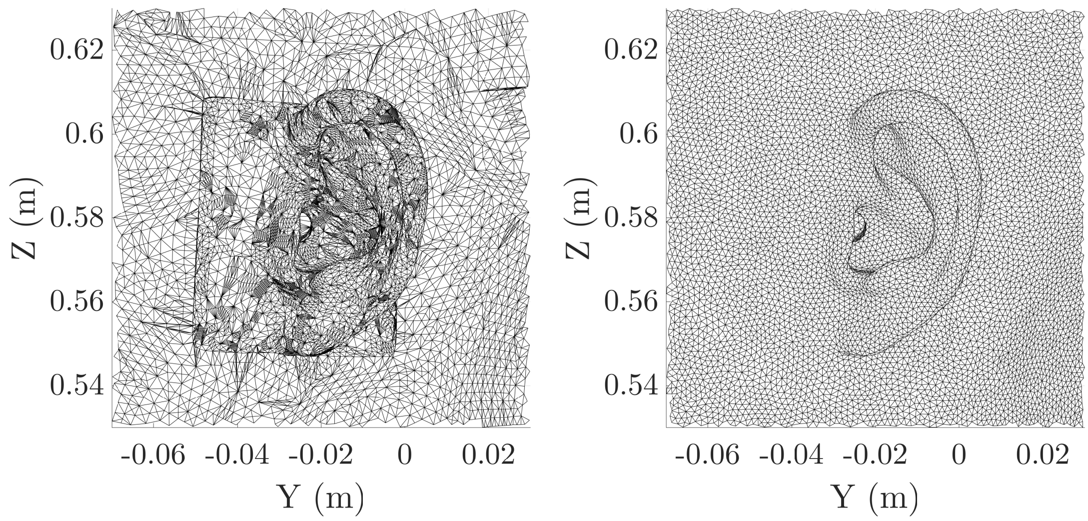


Figure 3.7: Distribution of edge lengths in the *full-torso-20* mesh. The shortest edge length is 0.24 mm, the longest is 2.49 mm and the median is 1.51 mm.

The region of the *full-torso-20* mesh containing the left pinna is shown in Fig. 3.8b. In contrast to the same region from the *original-polygon* mesh (Fig. 3.8a), the *full-torso-20* mesh contains consistently small equilateral triangles in a more uniform distribution. This mesh is stored in PLY format (.ply, also known as Stanford Triangle Format) as this permits easy definition of vertex coordinates and the connectivity between them.



(a) Prior to any processing

(b) After processing

Figure 3.8: Regions of the *original-polygon* and *full-torso-20* mesh around the left pinna. A non-uniform face distribution and shape is visible in 3.8a but not in 3.8b.

3.5 Numerical Validation

To be suitable for use within the BEM a mesh must satisfy certain trigonometric, numerical and topological conditions and it must be consistent with the actual physical manikin.

3.5.1 Angle Requirements for BEM simulation

BEM solvers often have a minimum and maximum internal angle requirement for mesh face edges to help avoid the inclusion of skinny triangles. For example, PAFEC-FE from PACSYS [261] requires angles between edges to be between 15° and 150°. Fig 3.9 shows the distribution of angles between edges in the *full-torso-20* mesh. The minimum is 16.0°, the maximum 146.7° and the median is 58.2°. This suggests that a majority of the triangles are approximately-equilateral.

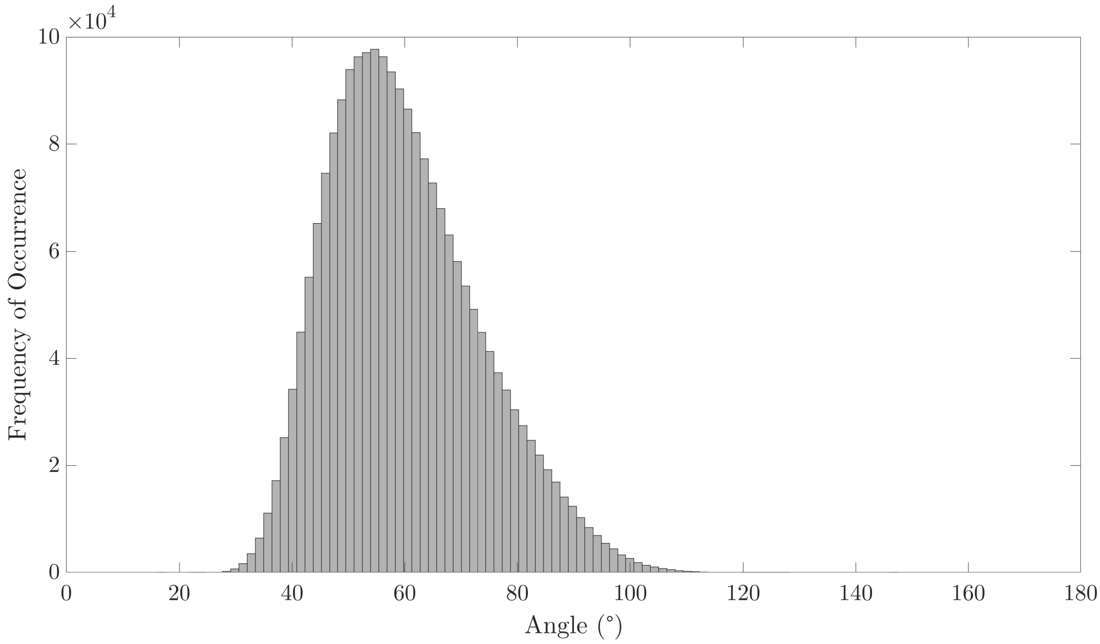


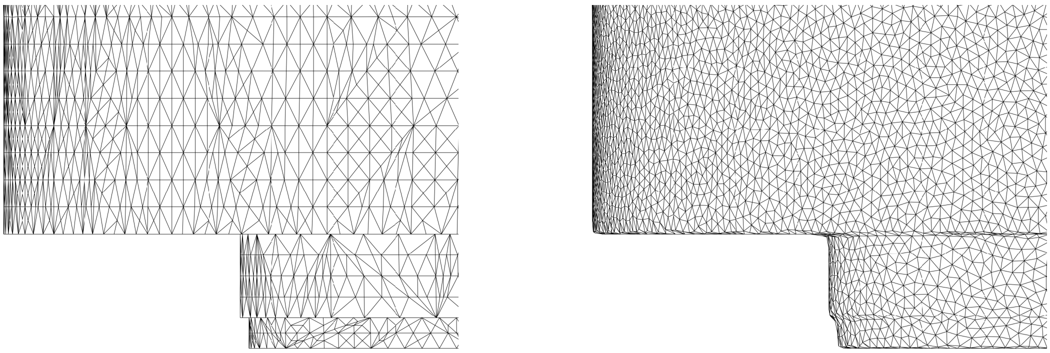
Figure 3.9: Distribution of angles between edges in the *full-torso-20* mesh. The smallest angle is 16.0°, the largest is 146.7° and the median is 58.2°.

3.5.2 Volumetric Consistency

To ensure that no stage of the conversion workflow introduced unwanted changes, the volume of the mesh at each stage was calculated using Autodesk Inventor [262] as a sanity check. Inventor was used as it allowed volume calculation of both the .stp and .ply format meshes without requiring conversion. The volumes (in mm^3) and their percentage differences from both the original CAD file and from the stage prior are shown in Table 3.1. There is a small loss of volume at each stage, with the largest as a result of the conversion from curved CAD file to planar *original-polygonal* mesh. The subsequent losses can be attributed to the slight rounding of sharp corners, as demonstrated in Fig. 3.10.

Mesh	Volume (mm^3)	From original (%)	From previous step (%)
Original CAD file	28,492,978	0	0
original-polygonal	28,489,372	−0.0127	−0.0127
remeshed-2mm	28,488,486	−0.0158	−0.0031
full-torso-20	28,486,546	−0.0226	−0.0068

Table 3.1: Volume of the mesh at each stage of conversion workflow, showing the small reductions in volume after each stage.



(a) *original-polygonal* mesh (b) *full-torso-20* mesh

Figure 3.10: The slight loss of volume is attributed to rounding at sharp corners during the remeshing process, shown here at the base.

3.5.3 Topological Consistency

Although no direct comparison with the physical manikin was possible, the consistency of the topology between the *original-polygonal* mesh and the *full-torso-20* mesh was calculated to ensure no region had been excessively distorted by the conversion workflow. The metric for comparing the shape of two meshes was defined as the perpendicular distance between the centre of each face on the *full-torso-20* mesh and the point of intersection of the centred face normal with the *original-polygonal* mesh. This was calculated using the pseudocode algorithm in Algorithm 1, which combines the mathematical steps described in Appendix B.2.

Using this method, no distance values greater than 0.63 mm were calculated, with 95 % less than 0.04 mm. Fig. 3.11 shows a region of the *full-torso-20* mesh with the perpendicular distances between the meshes plotted as a function of colour. Blue represents 0 mm and yellow represents the maximum value of 0.63 mm. The median and mean distances are 0.006 mm and 0.012 mm, respectively, with the difference as a result of the long tail comprising a small number of relatively large distances, as is shown in Fig. 3.12. These larger distances lie in regions such as the edge of the base and within the eyes, where the radius of curvature is be relatively small and the shape has been rounded during the workflow processes, as demonstrated in the example region in Fig. 3.10.

Load both meshes as matrices of vertices and of connectivity between those vertices (faces).

As the numbers of faces are large, define a set of faces in mesh A to calculate distances for. The process can be parallelised by doing this.

for *each set of faces in mesh A* **do**

  for *each face within the set* **do**

    Calculate the centroid and the centred normal vector of the face.

    Define a 3D range to calculate distance over (for example, $\pm 5\,\mathrm{cm}$).
    The intersection with the other mesh will be reasonably close to the testing face, so this is more computationally efficient than testing the entire mesh for each face.

    Determine the faces in mesh B that are in the region of interest.

    for *each face in mesh B within the region of interest* **do**

      Calculate the intersection of the normal of the test face with the plane of the face in mesh B.

      Determine whether the intersection point is within the triangle of the test face.

      if *intersection point is within the face* **then**

        Calculate the distance between the intersection and the centroid of the test face.

      else

        Set the intersection and the distance to **void**.

      end

    end

    Remove void intersections and distances.

    if *more than one distance is calculated per face* **then**

      Take the minimum distance.

      This occurs, for example, in the pinnae, where several folds of the boundary are close together in space.

    end

  end

end

Collate the outputs of parallel **for** loops to get distances for entirety of mesh A.

Algorithm 1: Pseudocode for calculating perpendicular distances between two meshes (referred to as mesh A and mesh B), combining mathematical steps described in Appendix B.2.

Figure 3.11: Perpendicular distances in mm between the *original-polygonal* and *full-torso-20* meshes, plotted on the final *full-torso-20* mesh as a colourmap, where blue is 0 mm and yellow is 0.63 mm.

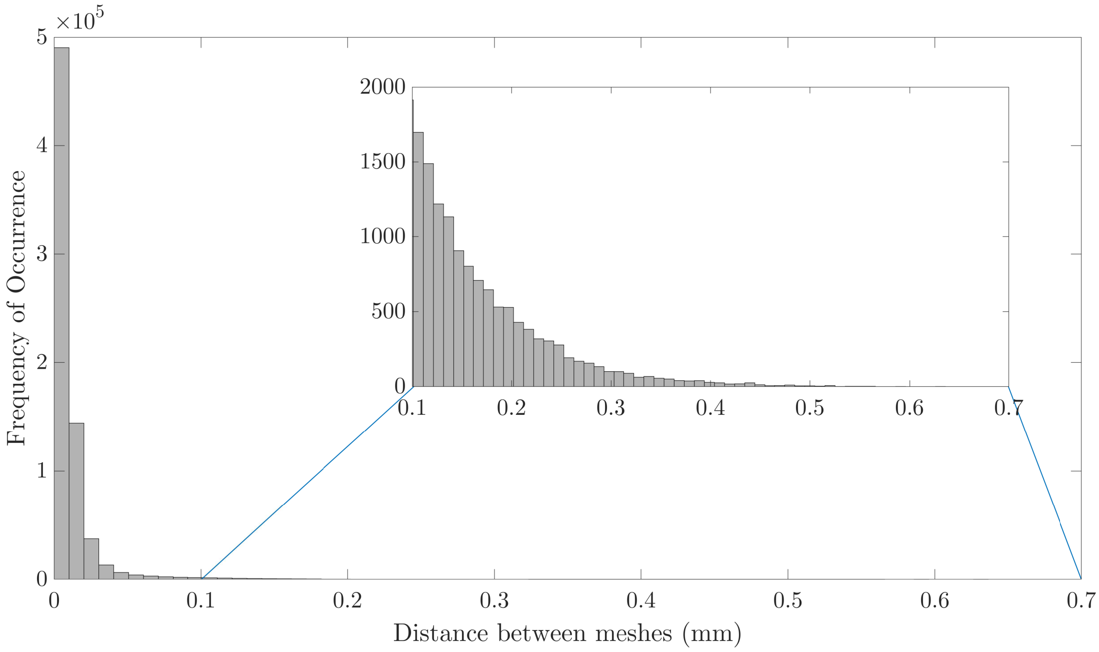
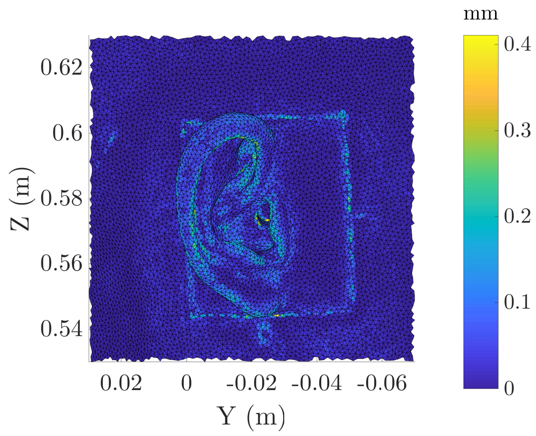
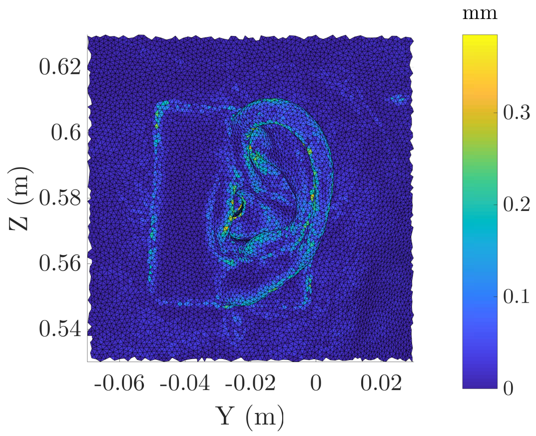


Figure 3.12: Distribution of the perpendicular distances between the *original-polygonal* and *full-torso-20* meshes. The shortest distance is 0 mm, the longest is 0.64 mm, and the median and mean are 0.006 mm and 0.012 mm, respectively.

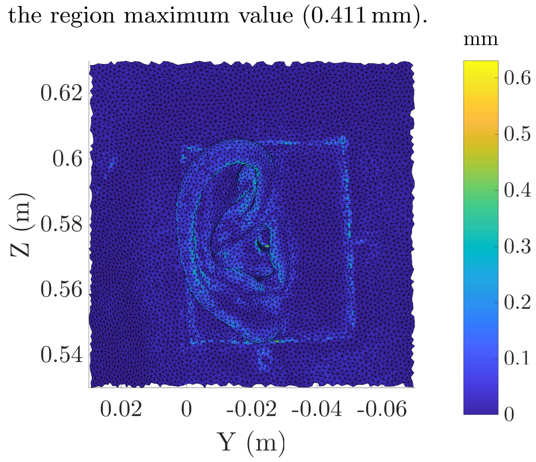
It is, arguably, most important with respect to HRTF simulation that the pinnae of the mesh are accurate. Fig. 3.13 shows these important pinnae regions, again with the perpendicular distances between the meshes plotted as a function of colour. Blue represents 0 mm. In Fig. 3.13a and Fig. 3.13b, yellow represents the maximum value within the region, whereas in Fig. 3.13c and Fig. 3.13d, yellow represents the maximum value in the full mesh (0.63 mm). The largest distances occur at the sharp edge at the entrance to the ear canal. The distribution of perpendicular distances in these regions for each ear is shown in Fig. 3.14. The maximum distances were 0.385 mm and 0.411 mm for the left and right pinnae, respectively, and 95 % of the values were below 0.142 mm and 0.138 mm, respectively.



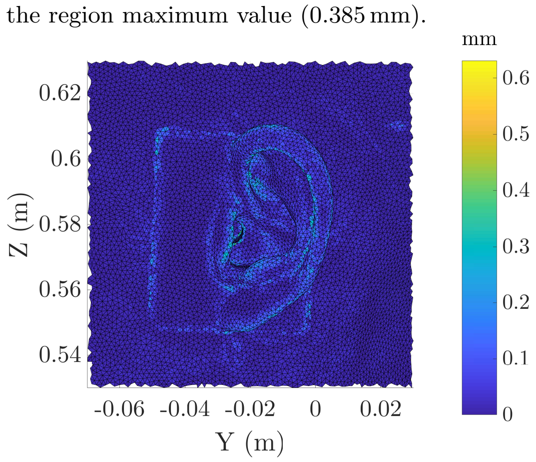
(a) Right ear, with a shading range normalised to the region maximum value (0.411 mm).



(b) Left ear, with a shading range normalised to the region maximum value (0.385 mm).



(c) Right ear, with a shading range normalised to the maximum value in full mesh (0.63 mm).



(d) Left ear, with a shading range normalised to the maximum value in full mesh (0.63 mm).

Figure 3.13: Perpendicular distances between the *original-polygonal* and *full-torso-20* meshes in the pinnae region, plotted on the *full-torso-20* mesh as a colourmap where blue is 0 mm and yellow is either the maximum value in the region (top row) or the maximum value in the full mesh (bottom row).

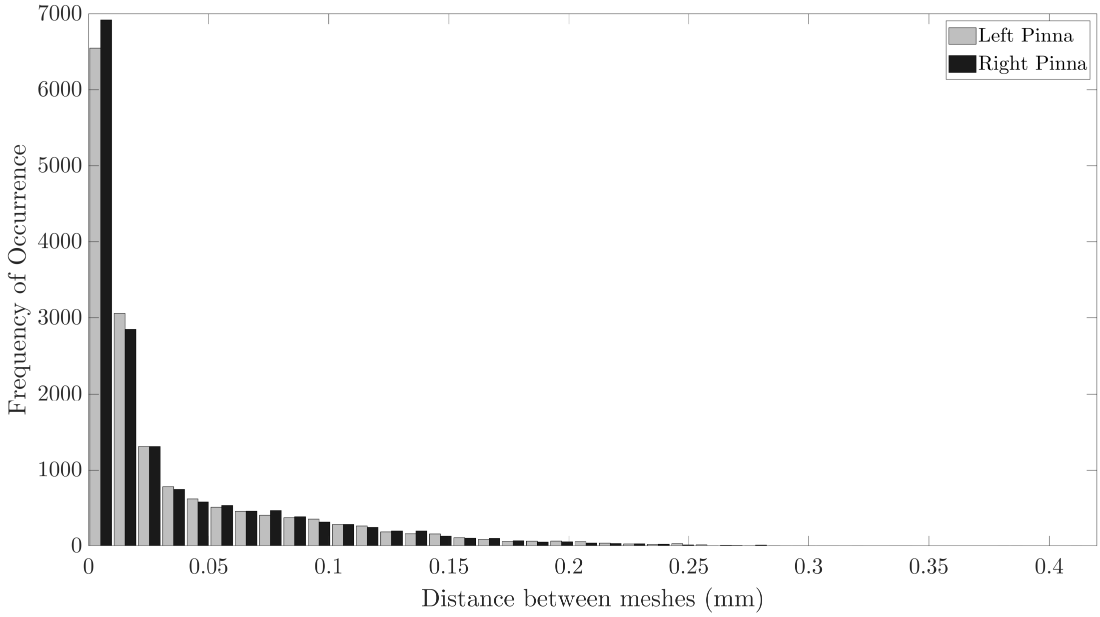


Figure 3.14: Distribution of the perpendicular distances between the *original-polygonal* and *full-torso-20* meshes in the pinnae regions, where grey bars indicate the left pinna and black bars indicate the right pinna.

These findings indicate that the workflow is capable of producing a BEM-suitable mesh which is consistent with the original CAD file. However, it cannot be yet said whether the mesh is accurate enough for the intended purpose of replacing direct acoustic measurements. Analysis of the difference between simulated and acoustically-measured results is needed to fully validate the mesh. This is considered in the following section.

3.6 Acoustic Validation: Method

The numerical analysis demonstrated that differences between the topologies of the mesh and the physical manikin are small. However, further validation is required to ensure simulations using the mesh produce results sufficiently similar to direct acoustic measurement.

3.6.1 Initial Mesh Adjustments

Within HRTF simulation, convention dictates that the origin of the coordinate system lies at the centre of the head and that the y-axis lies on the interaural axis. To adhere to these conventions, several adjustments were required, as the origin of the original CAD file is located at the centre of the base with the y-axis in the median plane. The mesh was rotated 90° around the z-axis, and translated -26 mm, -2 mm and -576 mm in the directions of the x-, y- and z- axes, respectively, to locate the interaural axis on the y-axis. Due to left/right asymmetries inherent within KEMAR, the points of intersection of the y-axis with the eardrums are not mirror images of each other. Native coordinates were converted from millimetres into metres.

The *full-torso-20* mesh includes ear canals, which may or may not be useful to researchers, depending on the preference for blocked or open meatus HRTF measurement techniques [129]. However, only blocked meatus-style measurements are possible with the physical KEMAR manikin, so removal of the ear canals from the mesh was required to match this.

The ear canals were removed from the *original-polygonal* mesh by selecting each ear canal (ensuring the selection was flush with the surface of the concha at the opening), deleting the faces and vertices, and filling the resulting hole using the 'Fill Hole $\rightarrow$ with Tangent'[2] tool in Geomagic Wrap. The numerical centres of the holes prior to filling were calculated and designated as the centres of the microphones (the coordinate values for which are listed in Table 3.2). It should be noted that, due to taking the local curvature into account when filling the hole, the numerical centres of the microphones lie slightly off the surface of the mesh. Additionally, these centres do not lie on the y-axis due to the left/right asymmetry of KEMAR. A visual comparison of meshes with and without ear canals is shown in Fig. 3.15.

Ear	Coordinates (m)		
	x	y	z
Left	-0.00035	0.06233	0.00175
Right	-0.00024	-0.06920	-0.00149

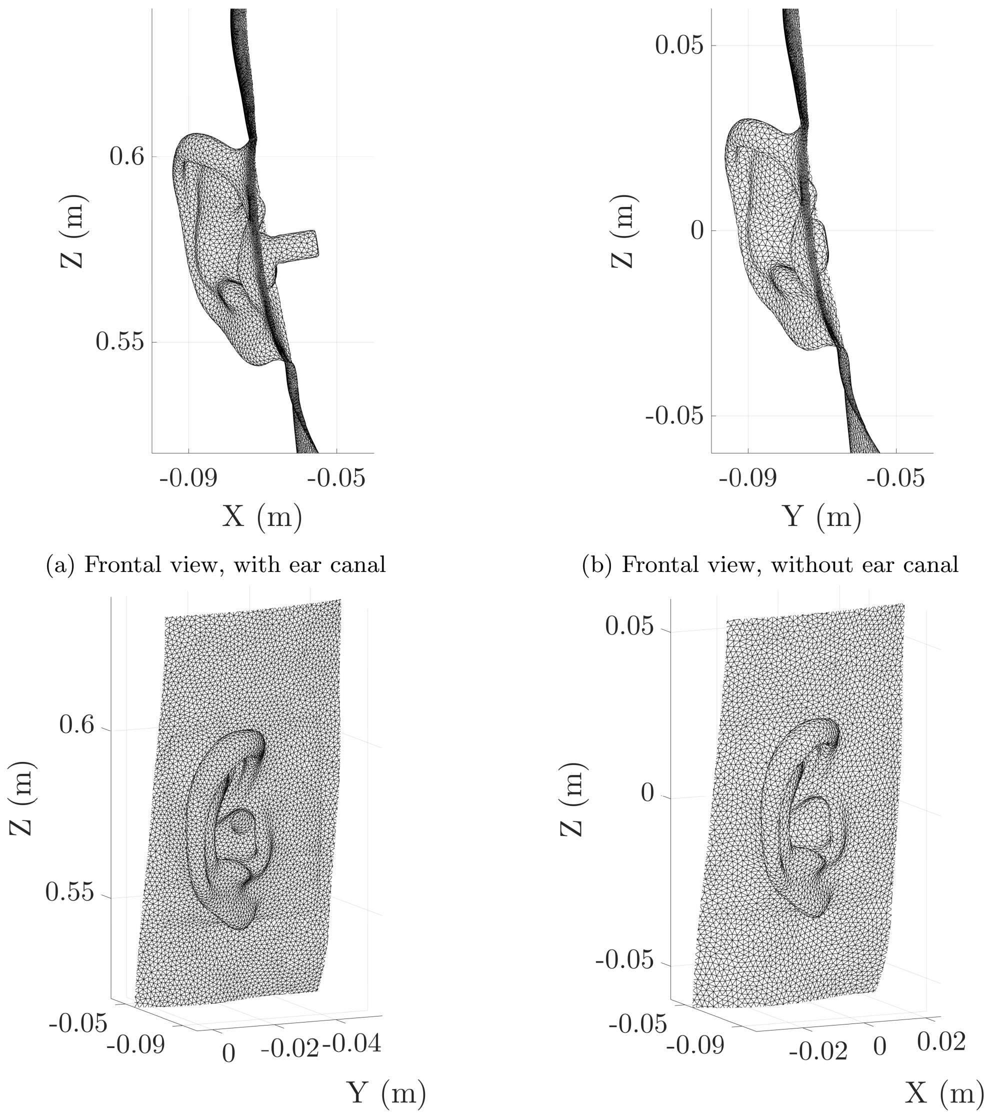
Table 3.2: Coordinates of the centre of the microphone for each of the left and right ears.

After removing the ear canals, the edited mesh was subject to the remeshing process as detailed in Section 3.4 to create two meshes without ear canals; one valid to 16.4 kHz (referred to as the *full-torso-noEC-16* mesh) and the other to 20.4 kHz (referred to as the *full-torso-noEC-20* mesh), both calculated using the six-edges-per-wavelength rule of thumb. Numbers of vertices and faces are detailed in Table 3.3. The difference in number of faces and vertices between the *full-torso-noEC-20* and *full-torso-20* meshes is a result of different maximum frequencies. Whilst both are valid to 20 kHz, the *full-torso-20* mesh is valid to the slightly higher frequency of 23.0 kHz and so has shorter edge lengths and more faces.

Mesh version	Vertices	Faces	Max edge length (mm)
full-torso-noEC-20	271,561	543,118	2.81
full-torso-noEC-16	157,250	314,496	3.48

Table 3.3: Numbers of vertices and faces and the maximum edge length in the full torso BEM-suitable meshes without ear canals.

[2]The '$\rightarrow$' icon is used to denote a choice of tool option within the software in question.



(a) Frontal view, with ear canal

(b) Frontal view, without ear canal

(c) Side view, with ear canal

(d) Side view, without ear canal

Figure 3.15: Right pinna region of the processed KEMAR mesh with the ear canal (*full-torso-20*, left) and without the ear canal (*full-torso-noEC-20*, right). The *full-torso-noEC-20* mesh has also been re-oriented such that the interaural axis lies on the y-axis.

3.6.2 Creation of the NEECK

At the time of simulation, computational resources in the AudioLab were not sufficient to simulate either the *full-torso-noEC-20* mesh or the *full-torso-noEC-16* mesh. During initial testing, it was determined that simulations using meshes containing more than approximately 250,000 faces did not successfully complete. It is not that computational resources do not exist anywhere which can handle meshes comprising so many faces (for example, Coustyx from Ansol [263] is able to handle larger meshes), it is only that such computational resources were not available for this work.

Since the head section of KEMAR is removable at the collar (shown in Fig. 3.16a), it is possible to perform acoustic measurements and the associated simulations on this section of KEMAR alone. Slicing the processed mesh at this point produced a mesh with approximately 60,000 faces; small enough to be simulated using AudioLab resources. Using the head section alone, a simulation could be performed using a derivative of the KEMAR mesh with a corresponding real-world counterpart to allow comparison between measured and simulated results, and thus validate the conversion workflow.

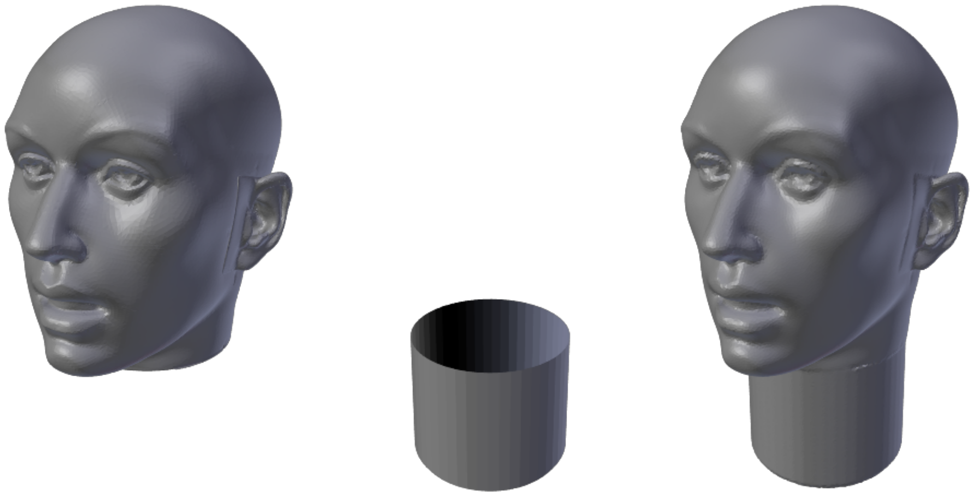
A mounting unit for the head section was designed in conjunction with the Technical Support Services (TSS) team in the Department of Electronic Engineering (Fig 3.16b). Technical details and diagrams are available in Appendix C. The unit covers the ribbed portion of the head unit that is shown in Fig. 3.16a, and allowed the head to be mounted on a standard microphone stand as shown in Fig. 3.16c. The ribbed portion is part of the mount for the mouth simulator inside the head of KEMAR; the mouth simulator was removed, but the weight of the bracket increased stability and so was retained. The head-and-neck mount configuration is known as the Neck-Extended Easily Computable KEMAR, or NEECK.

(a) Head section of KEMAR (b) Mounting unit. (c) Head section of KEMAR mounted on a microphone stand.

Figure 3.16: Mounting unit used to take acoustic measurements of the head only portion of KEMAR.

After initial testing indicated that a mesh representation of the NEECK would be small enough to simulate using AudioLab facilities, an accurate mesh was created using Geomagic Wrap and Autodesk 3DS Max [264]. Details of the creation of the mesh representation are available in Appendix C.

The mesh representation of the NEECK was remeshed using the process described in Section 3.4 to produce two NEECK meshes; *NEECK-16*, valid to 16.5 kHz, and *NEECK-20*, valid to 20.2 kHz. Both frequency limits were calculated using the six-edges-per-wavelength rule of thumb. ACVD target values of 40,000 and 62,000 vertices, respectively, were used for these two meshes. The numbers of vertices and faces are detailed in Table 3.4.



(a) Mesh of head section of KEMAR (b) Mesh of mounting unit. (c) Mesh of complete NEECK unit.

Figure 3.17: Mesh representation of the NEECK.

Mesh version	Vertices	Faces	Max edge length (mm)
NEECK-20	62,000	123,996	2.83
NEECK-16	40,000	79,996	3.46

Table 3.4: Numbers of vertices and faces and the maximum edge length in the NEECK meshes.

3.6.3 Measurement of HRTFs

To match the closed mesh *NEECK-20*, the open mouth of the physical NEECK was blocked with putty and shaped to correspond to that region of the mesh (Fig. 3.18a). HRTFs of the NEECK were measured using the sequential logarithmic sine sweeps method [265] (200 Hz to 20 kHz sweeps of 10 s duration) in a fully anechoic chamber at the AudioLab. Reaper [266] was used with an RME Fireface 800 and a MacBook Pro 2014 running macOS Sierra 10.12.6 for playout and recording. A Cougar C-500.6 amplifier provided loudspeaker signals to five Tectonic Elements BMR drivers (TEBM46C20N-4B) [267]. The drivers were mounted on an arc of five aluminium bars (technical details and diagrams are available in Appendix C) and suspended from the ceiling of the anechoic chamber to reduce vibrations and the impact of the suspended sprung-wire floor.

The NEECK was placed on an Outline ST2 electronic turntable and aligned using Stanley Cubix Self-Levelling Laser Levels. This setup is shown in Fig. 3.18b. The turntable was driven using an Outline ET2 automatic control unit and a pulse generator built by TSS to convert an impulse into the control data required by the ET2. This allowed for the playout system to also control the position of the turntable according to the increment set on the ET2 unit. All reflective surfaces were covered with acoustic foam.

(a) The mouth of the NEECK blocked with putty to emulate the *NEECK-20* mesh.

(b) The NEECK HRTF measurement rig in situ in the anechoic chamber.

Figure 3.18: HRTF measurement setup for the NEECK.

Using pink noise, the loudspeakers were level-matched to 70 dB SPL (A weighted) at the centre of the rig. In accordance with the vertical-polar system defined in Section 2.1, θ increases counter-clockwise from zero in the frontal direction, such that $\theta = 90°$ is on the left and $\theta = 270°$ is on the right. ϕ is measured with respect to the horizontal plane containing the interaural axis, where $\phi = 90°$ describes a source directly above the head and $\phi = -90°$ is directly below. Radial distances (r) are measured relative to the origin (the mid-point of the interaural axis) and each ear lies approximately 65 mm from the origin along the interaural axis. Five elevations ($\phi = -30°, 0°, 30°, 60°, 90°$) were measured at azimuthal steps of 5° in the frontal hemisphere ($0° \leq \theta \leq 90°, 270° \leq \theta \leq 359°$) at a radial distance of 1.2 m to give a total of 185 positions (shown in Fig. 3.19), referred to as the set $\mathbf{SP}^{\mathbf{NEECK}}_{\mathbf{sparse}}$. This can be written using set notation as:

$$\mathbf{SP}^{\mathbf{NEECK}}_{\mathbf{sparse}} = \{v \mid v \in \mathbb{R}^3\} \tag{3.2}$$

in that $\mathbf{SP}^{\mathbf{NEECK}}_{\mathbf{sparse}}$ is a set with members v, such that each member v is a three-dimensional, real-valued column vector[3].

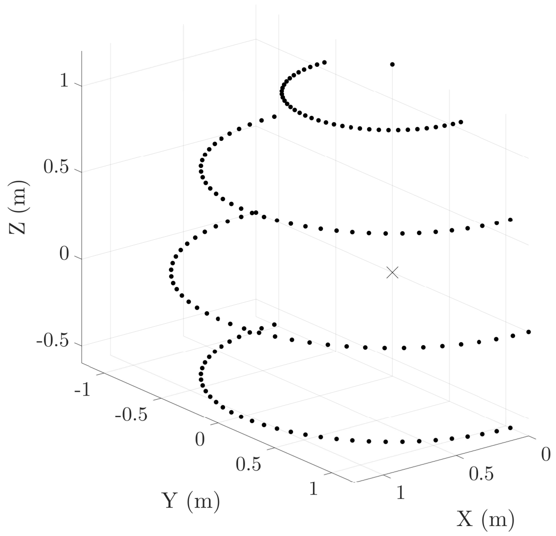


Figure 3.19: Cartesian coordinates of the $\mathbf{SP}^{\mathbf{NEECK}}_{\mathbf{sparse}}$ set of positions used in the acoustic measurement of HRTFs of the NEECK. The origin (x) is indicated.

The measurements were free-field equalised using minimum-phase inverse filters. The filters were calculated from measurements made with each microphone at the centre of the system in the absence of the NEECK using an implementation of the least-mean-square regularisation method outlined by Kirkeby *et al.* [269] with no octave smoothing. The frequency range of regularisation was 50 Hz to 16 kHz, with in-band and out-of-band regularisation values of 20 dB and −12 dB, respectively (when applying inverse filters for use in auralisation, a lower regularisation value may be used to avoid the introduction of audible artefacts. The aim here, however, is numerical accuracy, so the usual regularisation constraints were relaxed). The frequency spectra of two of the inverse filters used are shown in Fig. 3.20 for the left and right microphones for the loudspeaker at $\theta = 0°$, $\phi = 0°$. In the frequency domain, the sum in dB of the two magnitudes is approximately zero.

[3]For further information on set notation, the reader is directed to dedicated works such as [268].

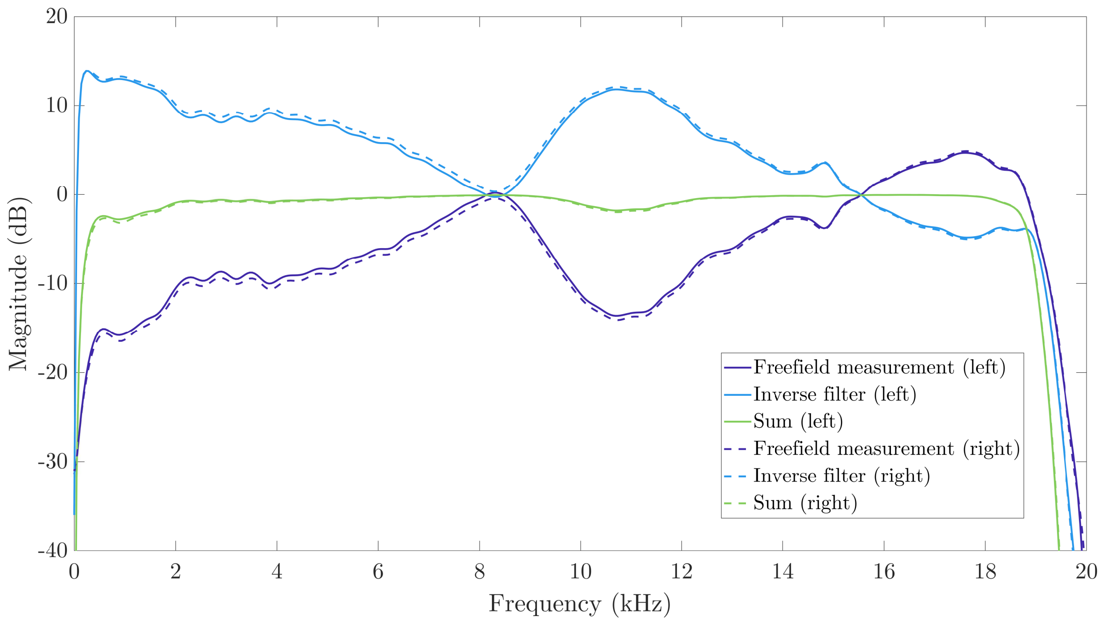


Figure 3.20: Frequency spectra of freefield measurements, inverse filters and their respective sums, generated for the left and right microphones for the loudspeaker at $\theta = 0°,\ \phi = 0°$.

3.6.4 Simulation of HRTFs

HRTFs for the NEECK were simulated using the *NEECK-20* mesh and multi-level FM-BEM, as implemented in the open source software Mesh2HRTF [179–181]. To match the ambient air temperature of 11 °C measured in the anechoic chamber, the simulation parameters for the speed of sound and air density were calculated using Equations 2.7 and 2.8 and set to $c = 338.0\,\mathrm{m\,s^{-1}}$ and $\rho = 1.24\,\mathrm{kg\,m^{-3}}$, respectively. As the change in temperature slightly affects the maximum valid frequency of the mesh, the maximum valid frequency is recalculated here using the six-edges-per-wavelength rule of thumb as 19.91 kHz.

To enable correction for possible misalignment in the measurement rig, additional simulation positions were defined on a 1° grid extending to ±5° in azimuth and in elevation around each position in $\mathbf{SP}^{\mathbf{NEECK}}_{\mathbf{sparse}}$. This gave a set of 10,205 positions (shown in Fig. 3.21) referred to as the set $\mathbf{SP}^{\mathbf{NEECK}}_{\mathbf{dense}}$. This can be written using set notation as:

$$\mathbf{SP}^{\mathbf{NEECK}}_{\mathbf{dense}} = \{v \mid v \in \mathbb{R}^3\} \tag{3.3}$$

in that $\mathbf{SP}^{\mathbf{NEECK}}_{\mathbf{dense}}$ is a set with members v, such that each member v is a three-dimensional, real-valued column vector. The evaluation grid of source positions was prepared using Delaunay triangulation in MATLAB [270, 271] and the preprocessing was performed in Blender [272] as required by Mesh2HRTF.

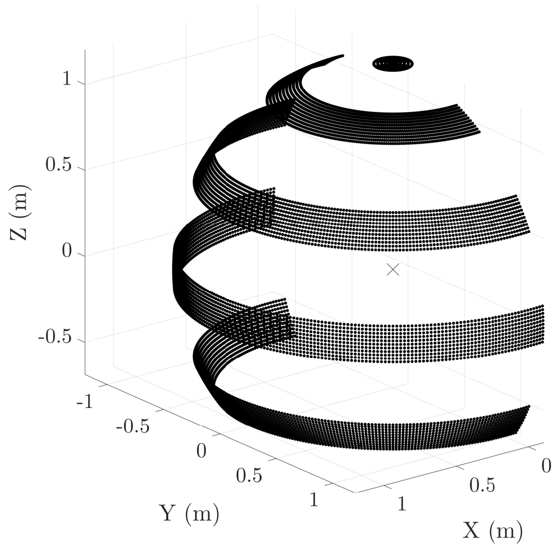


Figure 3.21: Cartesian coordinates of the $\mathbf{SP}^{\mathbf{NEECK}}_{\mathbf{dense}}$ set of positions used in the simulation of HRTFs. The origin (x) is indicated.

The 10,205 HRTF pairs were computed at 199 frequency points (200 Hz to 20kHz in 100 Hz steps). The simulation ran on an Ubuntu 16.04.2 machine with 24 central processing units (CPUs) clocked at 2.60 GHz, each accessing 10 GB of RAM. The vibrating element in the ear was chosen as the element nearest to the previously calculated centre of the microphone, since matching the acoustic and simulation set-ups as closely as possible was important to be able to compare the results. Using the principle of reciprocity, discussed in Section 2.6.1, the 24 concurrent threads cumulatively took approximately 120 hours at an average of 5 hours per thread. The maximum time for any thread, and therefore the actual length of time the simulation took to complete as all threads ran simultaneously, was 5.4 hours. The additional MATLAB processing stage as required by Mesh2HRTF (Output2HRTF) took approximately 5 minutes.

3.7 Acoustic Validation: Results

The measured and simulated data were compared for consistency using several metrics:

1. ΔITD: the unsigned difference between interaural time difference (ITD) values below 800 Hz at the same source position;
2. SD$_f$: the unsigned mean spectral difference between corresponding monaural responses, calculated using 40 linear frequency bands over the frequency range 200 Hz to 18 kHz;
3. ΔISD: the unsigned difference between interaural spectral difference (ISD) values calculated over the frequency range 200 Hz to 18 kHz in approximately 10 Hz steps at the same source position.

For positions at an elevation value of $\phi = 90°$, small values were found for the metrics as a result of the small interaural differences values. As these results did not provide particularly meaningful conclusions due to the small ranges, only source positions up to and including elevation values of $\phi = 60°$ are included in the following analysis.

3.7.1 Comparison of ITD

The ITD below 800 Hz for each $\mathbf{SP^{NEECK}_{sparse}}$ position was calculated using the cross-correlation method (see Section 2.5.1) for both measured and simulated head-related impulse responses (HRIRs). The ITDs have been divided by elevation band, plotted in Fig. 3.22. The ITD curves for measured and simulated responses show the same trend, but there appears to be an offset between them. Values with anomalously large magnitudes are apparent in the measured responses for four azimuth angles at an elevation value of $\phi = -30°$: $\theta = 60°,\ 65°,\ 290°,\ 295°$. Further investigation indicates that the ITD estimation method used a later peak in the measured contralateral response than in the simulated contralateral response, as shown in Fig. 3.23 for $\phi = -30°,\ \theta = 60°$. The IRs have been time-aligned for visualisation; the offset is due to the simulation software implementation. The peaks used for ITD estimation are indicated.

Values for the unsigned difference between ITD values, ΔITD, in each elevation band across azimuth are shown in Fig. 3.24, with the distribution of values shown as boxplots in Fig. 3.25. In this thesis, boxplots are defined as in [273], where the central mark indicates the median value. The median, mean and 95 % interval are listed for each elevation band in Table 3.5. The unexpectedly large values in $\phi = -30°$ are also apparent here, with a maximum ΔITD value of 229.2 µs appearing as an outlier.

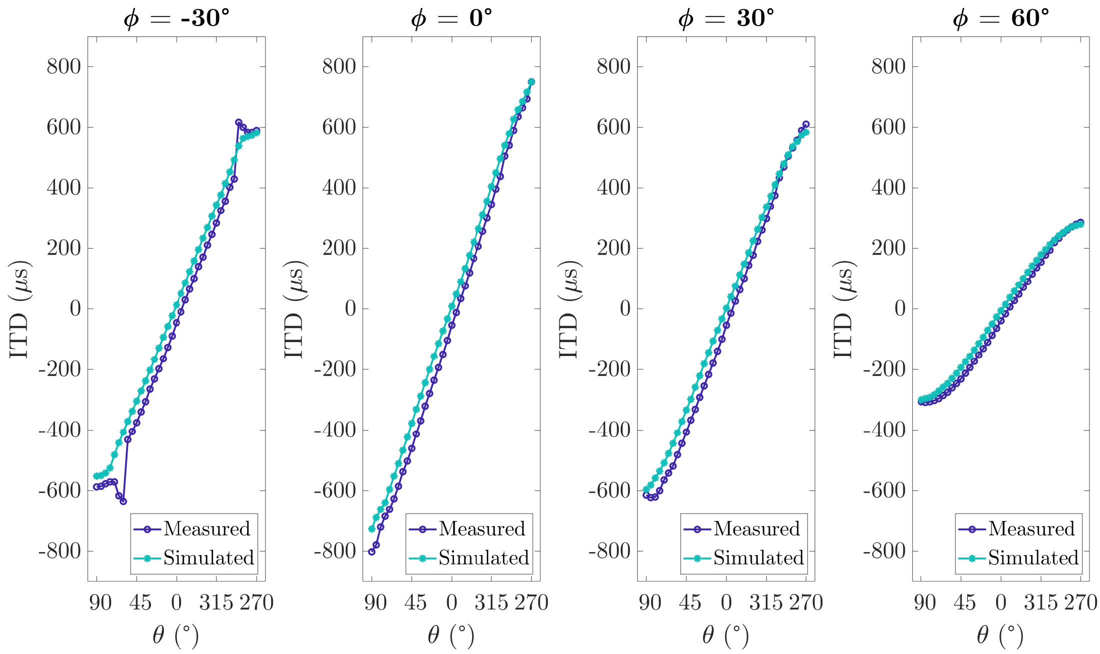


Figure 3.22: ITD plotted in 5° increments across azimuth for each elevation band analysed.

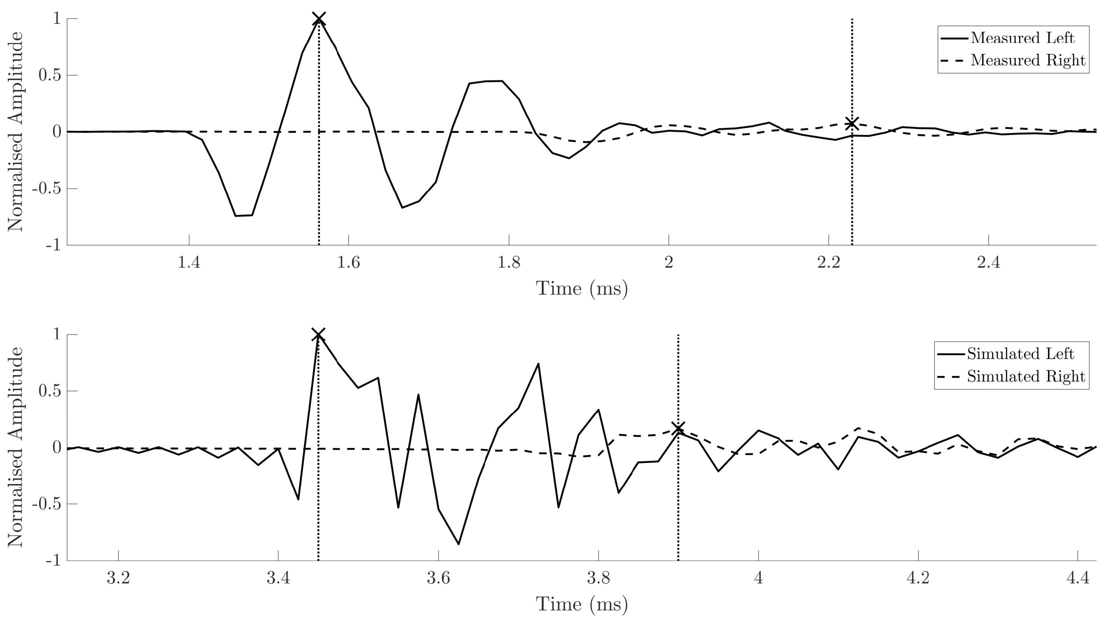


Figure 3.23: Measured (top) and simulated (bottom) IRs for $\theta = 60°$, $\phi = -30°$, with the peaks used for ITD calculation indicated with a marker and a line. The IRs have been time-aligned for visualisation. The peak used in the measured contralateral response is later than in the simulated contralateral response, producing a larger ITD value.

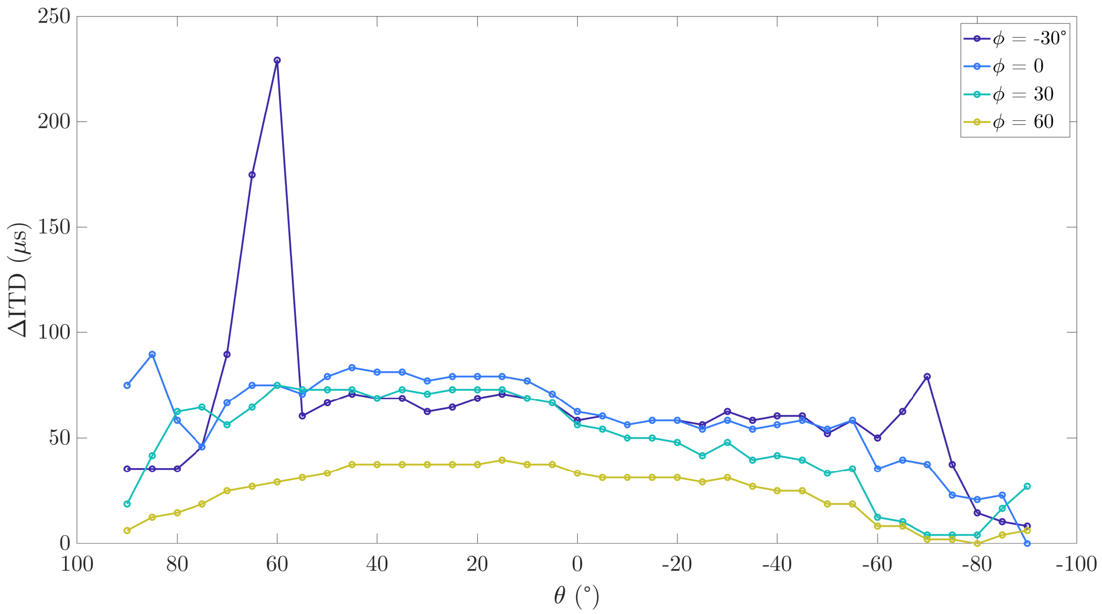


Figure 3.24: ΔITD plotted in 5° increments across azimuth for each elevation band analysed. Differences across elevation bands are apparent.

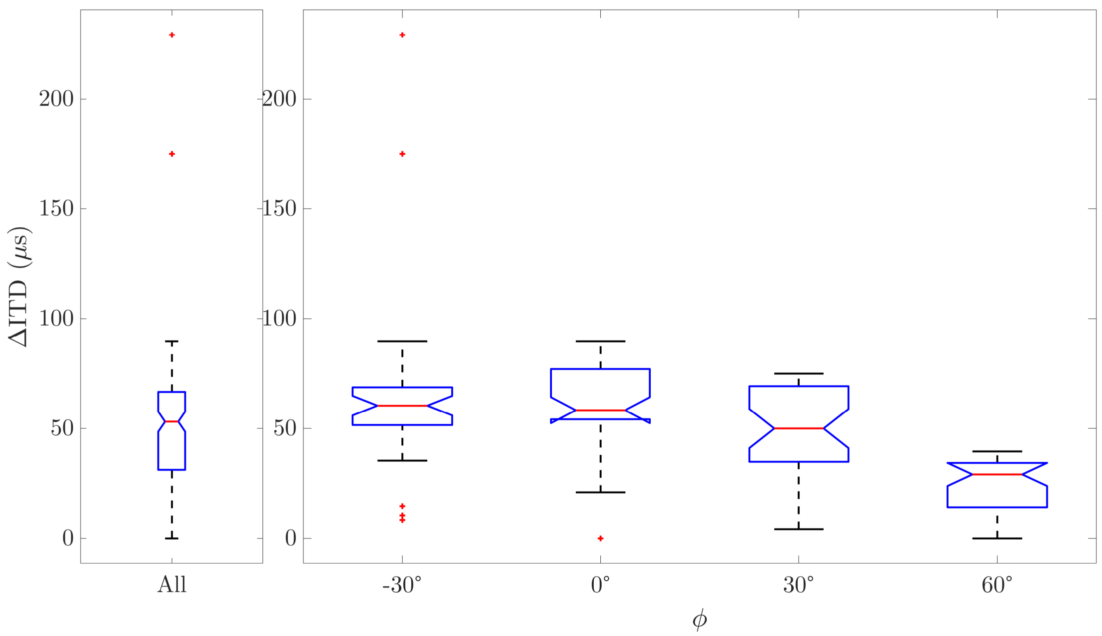


Figure 3.25: Distribution properties of ΔITD for all azimuth and elevation positions combined (left) and separated into elevation bands (right). The difference across elevation band is clearly visible.

	Median (µs)	Mean (µs)	95 % interval (µs)
$\phi = -30°$	60.4	63.4	127.1
$\phi = 0°$	58.3	59.8	82.3
$\phi = 30°$	50.0	48.3	72.9
$\phi = 60°$	29.2	24.4	37.5
All Positions	53.1	49.0	80.2

Table 3.5: Median, mean and 95 % interval for ΔITD for all azimuth positions, separated into elevation bands (first four rows) and for all azimuth and elevation positions combined (bottom).

Smaller ΔITD values exist at $\phi = 60°$ than at other elevation values. This may be attributable to the construction and support of the measurement rig. Whilst care was taken to minimise distortion of the aluminium arc under the load of the loudspeakers, it was only possible to support it at four points: at $\phi = 90°$, between $\phi = 90°$ and $\phi = 60°$, between $\phi = 60°$ and $\phi = 30°$, and between $\phi = 30°$ and $\phi = 0°$. These support points are shown in Fig 3.18b and indicated in Fig. C.1. It is, therefore, possible that the lower section of the arc distorted slightly under gravity (illustrated in Fig. 3.26). This may have created greater differences between measurement and simulation responses at low elevations. There also appears to be greater similarity for positions on the right-hand side ($270° < \theta < 360°$), perhaps due to small differences between seating of the two ears during the NEECK measurements.

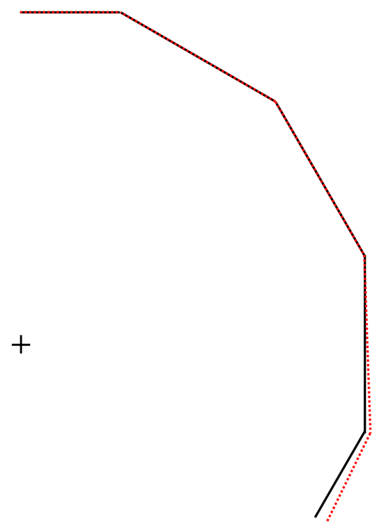


Figure 3.26: Lower segments of the measurement rig may have distorted under gravity (indicated in red) from its original position (indicated in black) and resulted in greater differences between measurement and simulation responses at low elevations. The origin (+) is also indicated.

The offset visible in Fig. 3.22 suggests the presence of a misalignment between measured and simulated responses. Analysis of the simulations made at the $\mathbf{SP}^{\mathbf{NEECK}}_{\mathbf{dense}}$ positions indicates that the highest similarity occurs when an azimuthal rotation of 5° to the left is applied to the simulation results. This rotational discrepancy was most likely caused by misalignment of the NEECK during the acoustic measurements. Fig. 3.27

and Fig. 3.28 show ΔITD values after correcting for this azimuthal error. Table 3.6 lists the median, mean, and 95 % interval values..

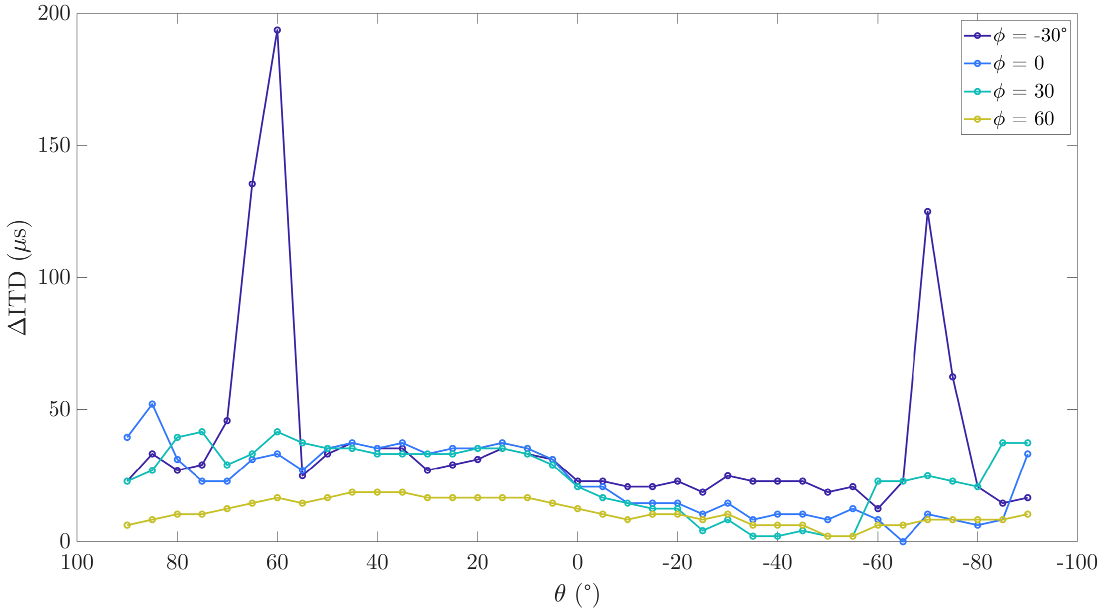


Figure 3.27: ΔITD plotted in 5° increments across azimuth for each elevation band analysed after applying an azimuthal correction of 5°. This correction reduces the differences.

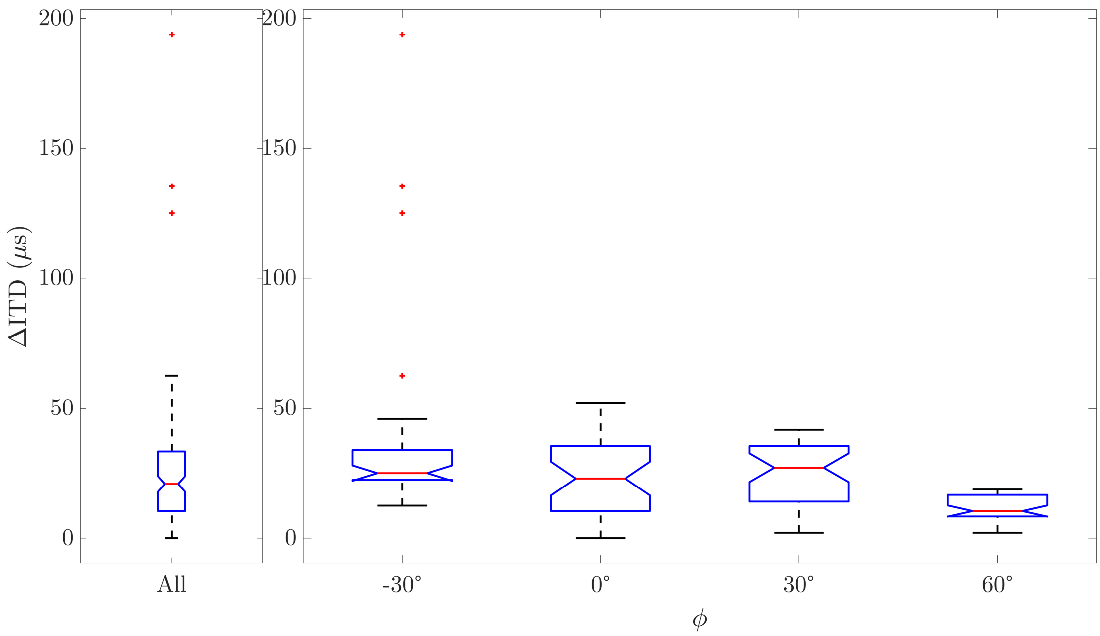


Figure 3.28: Distribution properties of ΔITD for all azimuth and elevation positions combined (left) and separated into elevation bands (right) after applying an azimuthal correction of 5°. The differences across elevation band have been reduced but are still visible.

	Median (µs)	Mean (µs)	95 % interval (µs)
$\phi = -30°$	25.0	37.3	99.0
$\phi = 0°$	22.9	23.0	38.5
$\phi = 30°$	27.1	24.4	39.6
$\phi = 60°$	10.4	11.2	17.7
All Positions	20.8	24.0	40.6

Table 3.6: Median, mean and 95 % interval for ΔITD for all azimuth positions, separated into elevation bands (first four rows) and for all azimuth and elevation positions combined (bottom), after applying an azimuthal correction of 5°.

The rotational correction reduces the differences, bringing the ΔITD values further into the range of just-noticeable-difference (JND) values, which typically range from 10 to 75 µs [111, 115–117]. This indicates that, whilst numerical differences do exist between the ITD values for measured and simulated response, these differences are unlikely to be perceivable by most listeners. These values are also similar to ITD differences found by others when looking at HRTF databases of KEMAR measured at different institutions over various lengths of time (30 to 235 µs according to Andrepoulou *et al.* [173]) and slightly larger than values found between databases of KEMAR measured at the same institution over a short period of time (maximum values of 30 µs and 25 µs reported by Andrepoulou *et al.* [173] and Rugeles Ospina *et al.* [274], respectively). The analysis in the following sections is performed including the azimuthal correction of 5°.

3.7.2 Directional Processing

Acoustic measurements of KEMAR have been shown to exhibit an exaggerated peak at approximately 4 kHz which has been attributed to ear canal resonance [74]. This physical phenomenon is not as apparent in the simulated data, as the vibrating element in the simulation is on the surface of the smoothly-occluded concha region of the mesh, whereas the microphone mounted in the physical KEMAR is set approximately 1 to 3 mm below the concha surface within the ear canal, depending on seating. In an attempt to account for this difference, the directional components alone of both the measured and simulated data were extracted. In the time domain, this is referred to as the directional impulse response (DIR). In the frequency domain, it is referred to as the directional transfer function (DTF). In order to minimise the introduction of distortion, the directional components were extracted in the first instance in the domain the results were obtained in. Hence, DIRs were calculated for the acoustic measurements and the fast Fourier transform taken to produce DTFs, whilst the DTFs were directly calculated for the simulated results as the BEM solver operates in the frequency domain.

The average response of any set of results contains the non-directional information. To produce a DIR or DTF, this average response must be calculated and removed from the results. In the time domain, this requires the calculation of an inverse filter from the average IR and the filtering of each measurement with this inverse filter [69]. In the frequency domain, the equivalent process is to calculate the average complex-valued response and divide each measurement by this value.

For the acoustic measurements, the average IR was calculated for each of the left and right ears. Minimum-phase inverse filters were generated from these using the same implementation of the least-mean-square regularisation method outlined by Kirkeby *et al.* [269] as for the measurements, using in-band regularisation of 20 dB between 50 Hz to 16 kHz and out-of-band regularisation of −12 dB with no octave smoothing. The inverse filters were applied to each free-field equalised left and right measurement before also calculating each corresponding DTF.

As only the spectral magnitude was of interest in this analysis, the averages of the magnitudes in decibels (rather than of the complex-valued responses) were calculated for the left and right simulated HRTFs. To ensure each magnitude contributed equally to the average, only $\mathbf{SP^{sparse}}$ source positions were included. Each HRTF magnitude was converted into the corresponding DTF by subtracting the appropriate left or right log average response. The average magnitudes are shown in Fig. 3.29, where the differences in frequency content as a result of the different environments is apparent.

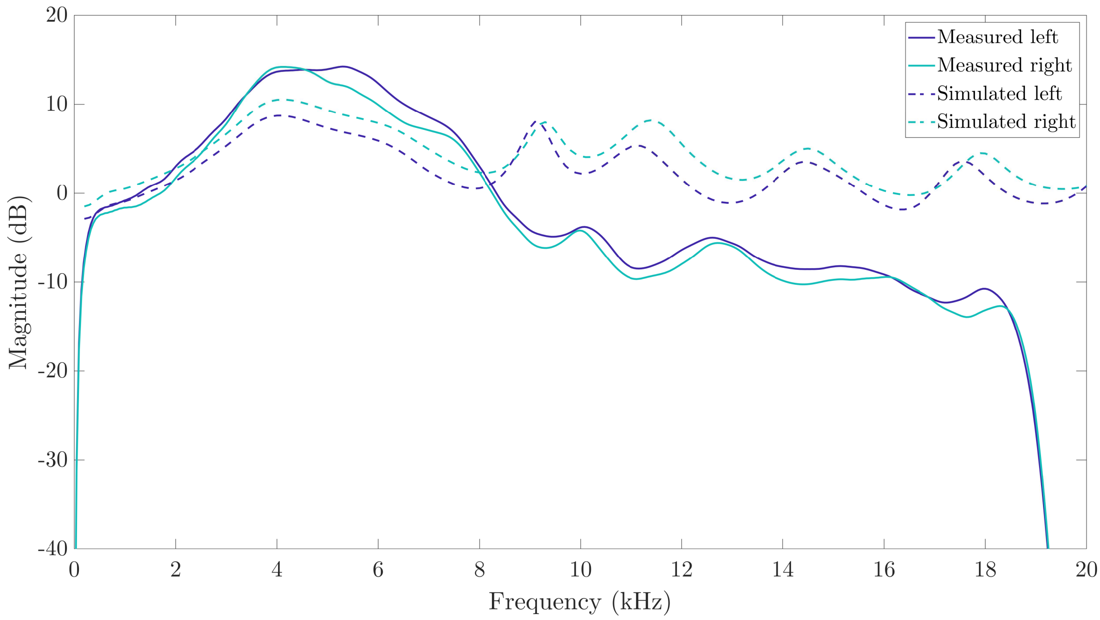


Figure 3.29: Average responses for measured (full line) and simulated responses (dashed line) used to calculate DTFs. The differences in frequency content as a result of the different environments is apparent.

The DTFs are used in subsequent analysis instead of the free-field equalised measured HRTFs and direct simulated data. The analysis was restricted to the frequency range 200 Hz to 18 kHz. Upon visual inspection, measured and simulated DTFs show good agreement. Fig. 3.30 shows the horizontal plane responses, with the corresponding results for other elevation values displaying the same level of similarity.

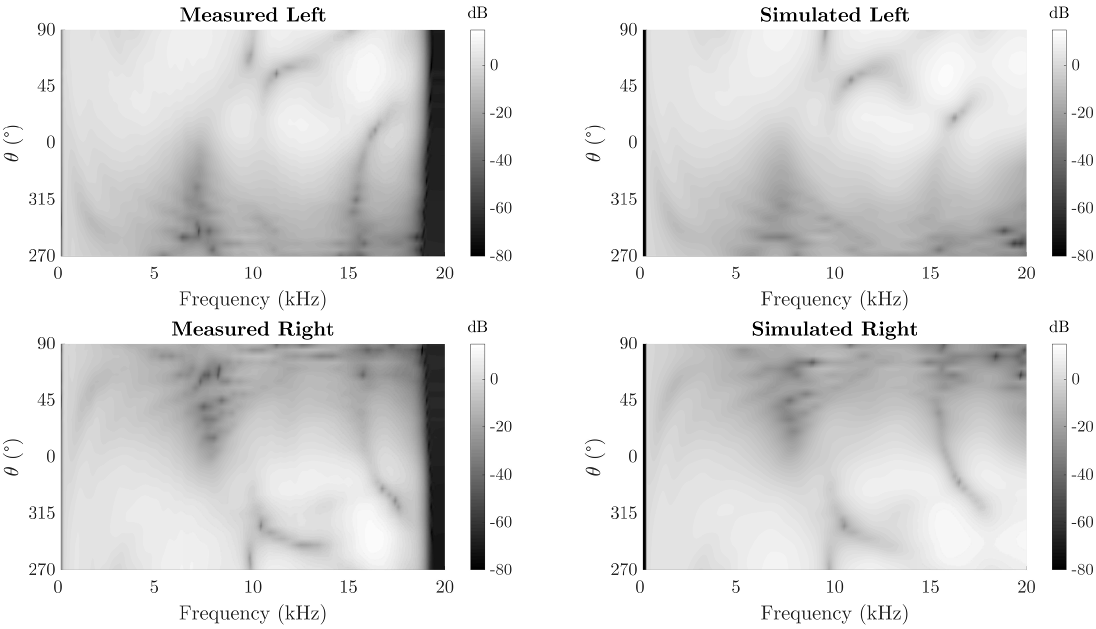


Figure 3.30: Measured (left) and simulated (right) DTF magnitude grayscale plots for both ears (left at the top, right at the bottom) in the horizontal plane. Dark bands indicate where data does not exist (above 18 kHz in measured, below 200 Hz in simulated).

3.7.3 Comparison of Monaural Spectra

The unsigned mean spectral difference between two signals within a frequency band, SD_f (see Equation 3.4), was used to evaluate the difference in monaural spectral content for corresponding directions between measured and simulated DTFs.

$$SD_f(\theta,\phi,\zeta) = \frac{1}{N_f}\sum_{j=1}^{N_f}\left|D_{meas(\theta,\phi,\zeta)_j} - D_{sim(\theta,\phi,\zeta)_j}\right| \tag{3.4}$$

where N_f is the number of points within the band f, ζ denotes whether the left or right signal is currently being used, and D_{meas} and D_{sim} are the DTFs (in dB) of the measured and simulated result.

SD_f was calculated using 40 constant bandwidth frequency bands (each with a bandwidth of 445 Hz) over 200 Hz to 18 kHz. Constant bandwidth bands were chosen to allow comparison with existing literature. This gives 11,840 data points across all DTFs, comprising all combinations of 2 ears, 4 elevation angles, 37 azimuth angles and 40 frequency bands. Values for all responses in each frequency band, plotted against centre frequency, are shown as a boxplot in Fig. 3.31. The values for mean, median and 95 % interval SD_f values are shown in Fig. 3.32. SD_f varies with frequency and displays a gradual rising trend, together with some local peaks, particularly in the 8 to 10 kHz region and above 15 kHz. The general increase in SD_f in upper-frequency bands is to be expected, as the generation of peaks and notches in this frequency region by the pinnae is highly sensitive to small errors [84, 173]. The mean and median values do not exceed 6 dB in any frequency band.

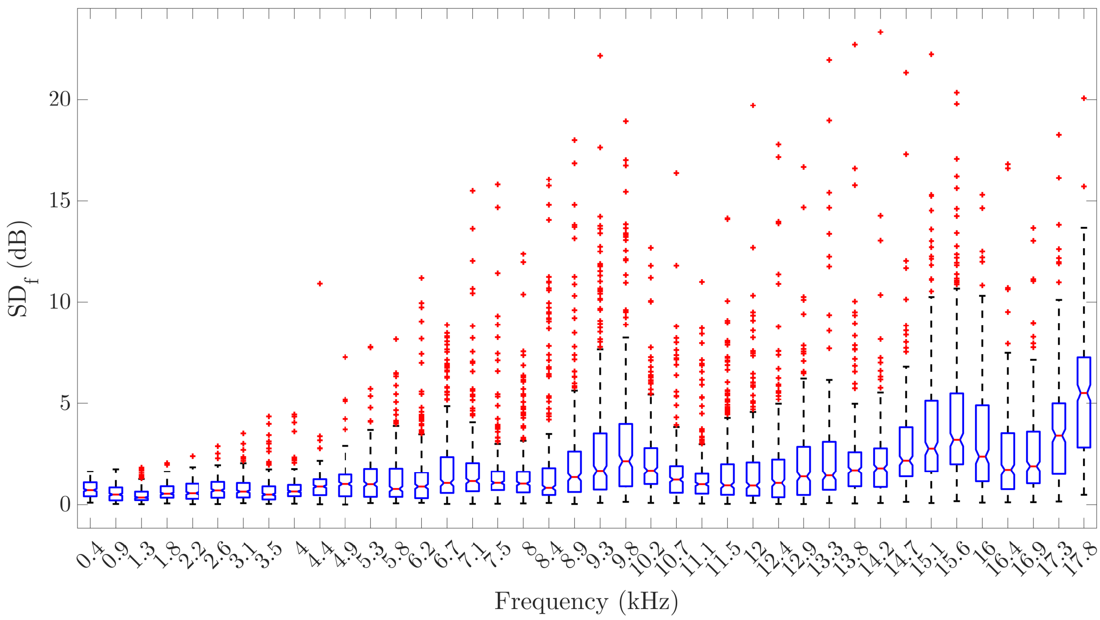


Figure 3.31: Distribution properties of SD_f for all responses in 40 frequency bands. A gradual rising trend is apparent, with some local peaks.

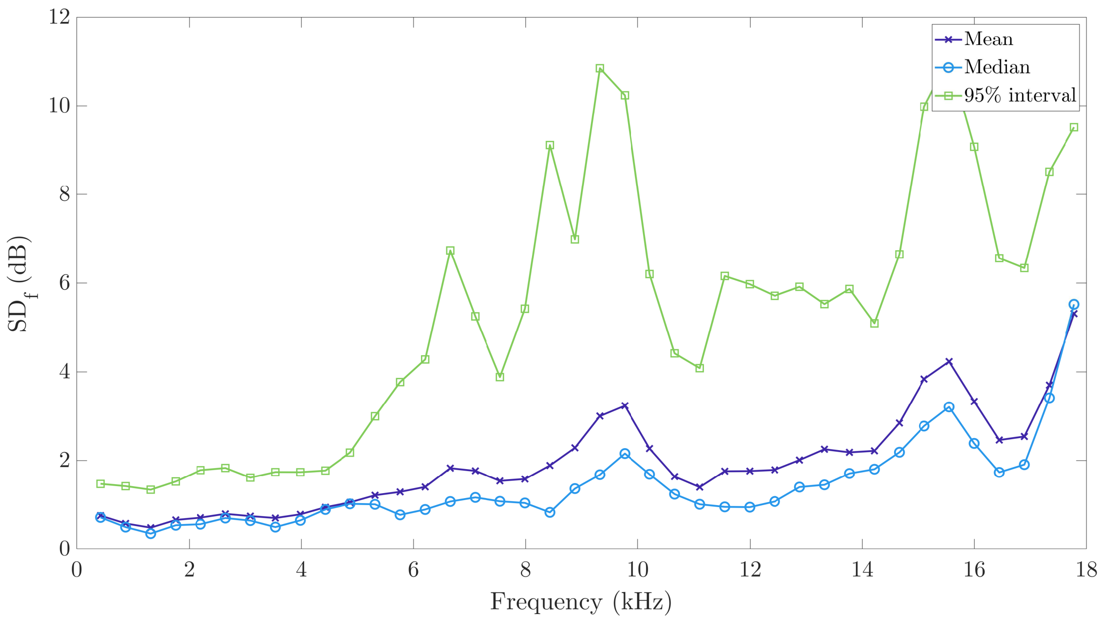


Figure 3.32: Mean, median and 95 % interval of SD_f for all responses in 40 frequency bands. Again, a gradual rising trend is apparent.

To generalise across frequency bands, the 11,840 data points are treated as one data stream. The mean SD_f across all bands is 1.9 dB (median 1.1 dB), and 95 % of the values are less than 6.7 dB. Fig. 3.33 shows SD_f values for all frequency bands in the left panel. In the right panel, the data is split at 6 kHz to enable comparison with the literature.

Below 6 kHz, SD_f values are close to the generally-accepted JND for spectral amplitude of 1 dB [44, 118]. The mean is 0.9 dB (median 0.6 dB), and the 95 % interval is 2.2 dB. Above 6 kHz, SD_f values increase: mean 2.5 dB, median 1.6 dB, 95 % below 7.8 dB. These results are similar to (and in some cases more favourable than) findings by others. For example, Andrepoulou *et al.* [173] reported inter-database variation of the same KEMAR unit of 2.5 to 6.7 dB below 6 kHz and 2.8 to 6 dB below 10 kHz. In a similar study, Zhong *et al.* [275] reported variation of 1 dB below 5 kHz and 5 dB above 5 kHz. The 90 % interval of differences between measured and simulated HRTFs reported by Brinkmann *et al.* [276] is 2 dB below 6 kHz and 10 dB above 6 kHz.

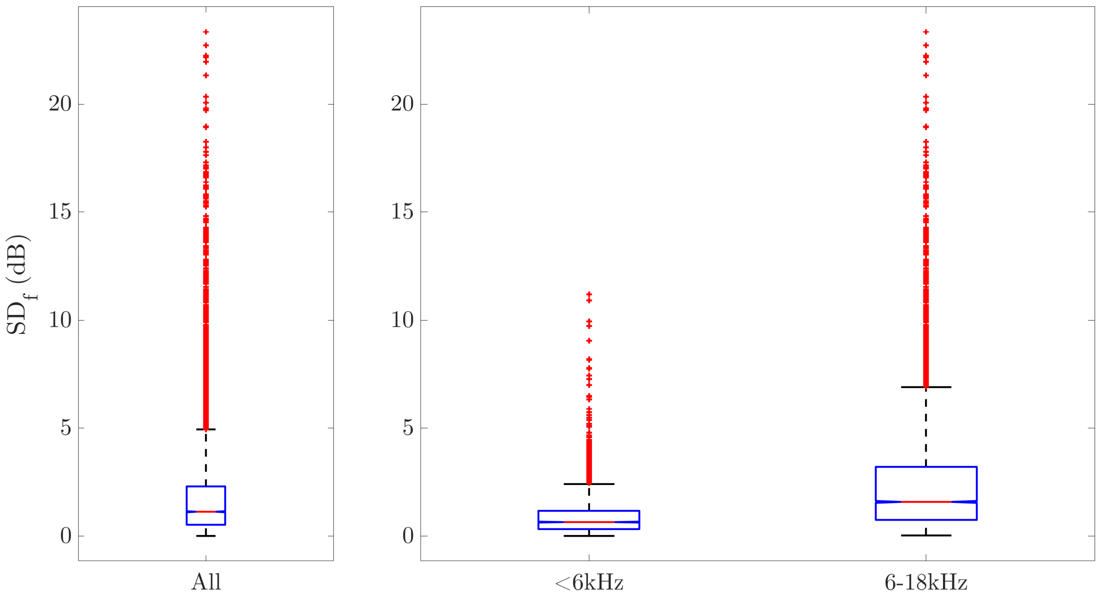


Figure 3.33: Distribution properties of SD_f across all responses, calculated for 40 frequency bands, for all frequency bands (left) and divided at 6 kHz (right). There is an appreciable difference between the values either side of the divide.

Fig. 3.33 indicates the existence of outliers with large SD_f values (maximum of 23.3 dB). These typically occur in deep notches where a small shift in notch frequency can result in a large numerical difference in spectral magnitude, despite its likely small perceptual impact [277]. Fig. 3.34 displays an example of this for the right ear signal at $\theta = 335°$, $\phi = 30°$. The SD_f in each frequency band is plotted on the right y-axis. The notch at approximately 9.7 kHz in the simulated data has been recorded as approximately 10 kHz in the measured response, producing a larger SD_f for that frequency band of 11.5 dB. It is worth noting that these outliers occur for a small number of notches only.

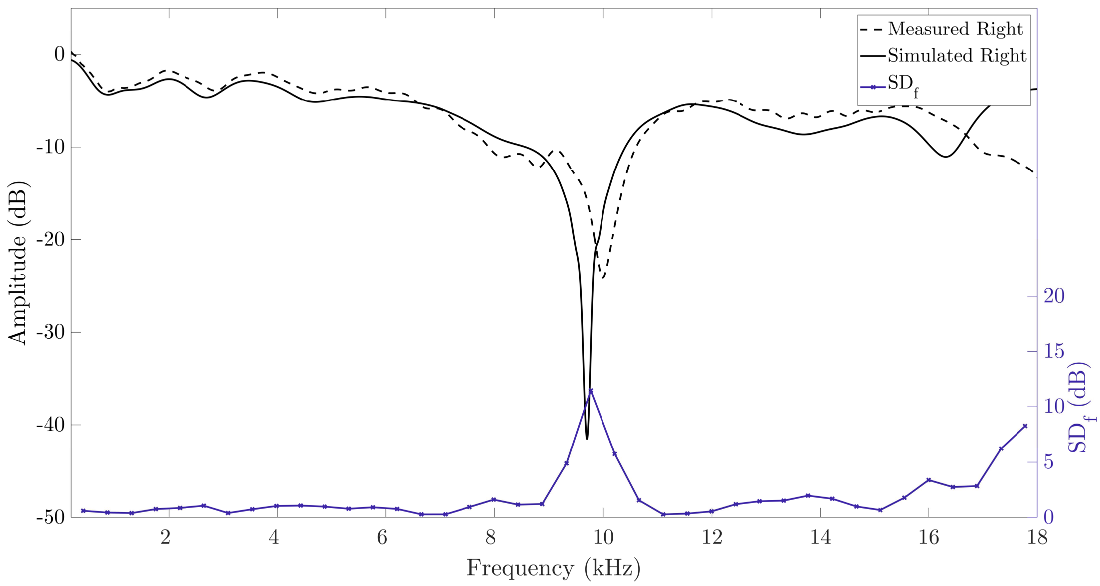


Figure 3.34: The misalignment of the notch at 9 to 10 kHz between measured and simulated right ear signals at $\theta = 335°$, $\phi = 30°$ causes a large numerical difference in spectral magnitude (indicated in purple), but likely a small perceptual difference.

As in the case of the ΔITD analysis in Section 3.7.1, SD$_f$ can be split by elevation band (shown in Fig. 3.35). The median, mean and 95 % interval values are listed in Table 3.7. A similar trend to that shown for ΔITD is evident, in that the highest elevation band ($\phi = 60°$) produces the smallest SD$_f$ values. The patterns of variation with increasing elevation are less consistent, with a fluctuation in mean values, a small rise in median values, and a drop in 95 % interval. It may be the case that spectral differences across frequency essentially dwarf the differences across elevation band.

Elevation Band	Median (dB)	Mean (dB)	95% interval (dB)
$\phi = -30°$	1.5	2.3	7.1
$\phi = 0°$	0.8	1.8	7.4
$\phi = 30°$	1.0	1.9	7.2
$\phi = 60°$	1.2	1.7	4.4

Table 3.7: Median, mean and 95 % interval for SD$_f$ for all directions, separated into elevation bands. Less of a trend across elevation is visible than in the ΔITD data.

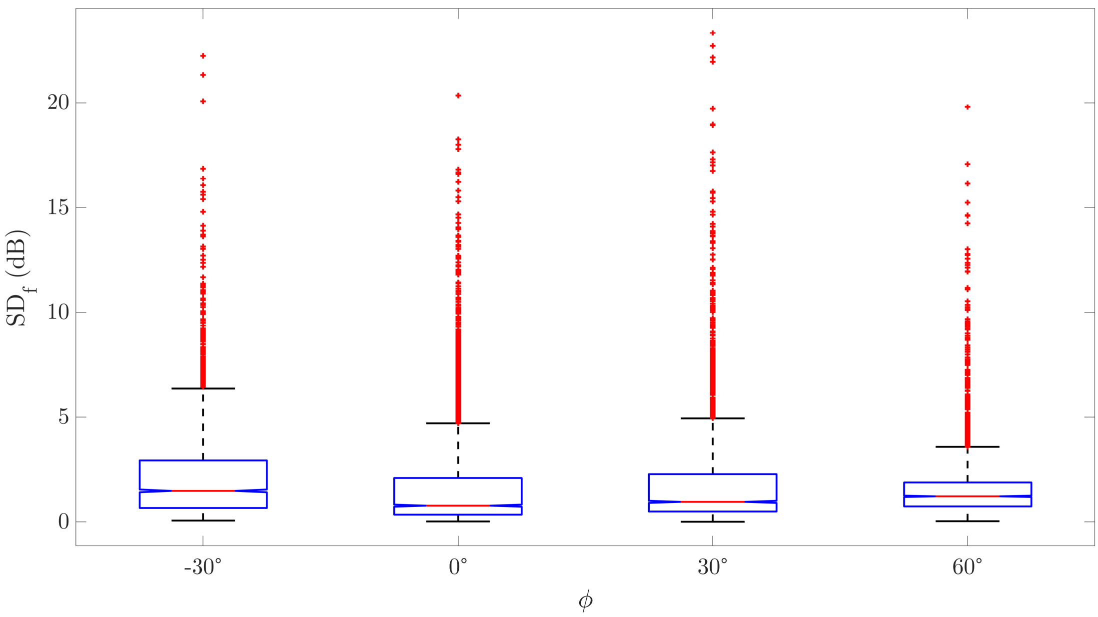


Figure 3.35: Distribution properties of SD_f for all directions, separated into elevation bands. Less of a trend across elevation is visible than in the ΔITD data (Fig. 3.28).

3.7.4 Comparison of ISD

As discussed in Section 2.5.2, the interaural level difference (ILD) at a single frequency point is often referred to as the interaural spectral difference (ISD). Hence, here the unsigned difference between the spectra of the DTFs at a frequency f, ISD_f is calculated by:

$$ISD_f(\theta, \phi) = \left| D_{L_f}(\theta, \phi) - D_{R_f}(\theta, \phi) \right| \tag{3.5}$$

where D_{L_f} and D_{R_f} are the DTFs at frequency point f in dB for the left and right ears, respectively. The unsigned ISD in dB was calculated for the measured and simulated results at each $\mathbf{SP}^{\mathbf{NEECK}}_{\mathbf{sparse}}$ position over the frequency range 200 Hz to 18 kHz in approximately 10 Hz steps (1765 frequency points), and the difference between the two, ΔISD, was used to determine similarity. This gives 261,220 data points comprising all combinations of 4 elevation angles, 37 azimuth angles and 1765 frequency points.

The distribution of results is shown in Fig. 3.36. 95 % of the values are below 9.0 dB and the overall mean is 2.4 dB (median 1.2 dB). The median is close to, but above, the reported JND of 1 dB [118]. As with the monaural spectral comparison, outliers are present with large ΔISD values, in this case, up to 38.9 dB. As before, these are attributable to the misalignment of notches between the measured and simulated responses, as shown in Fig. 3.34.

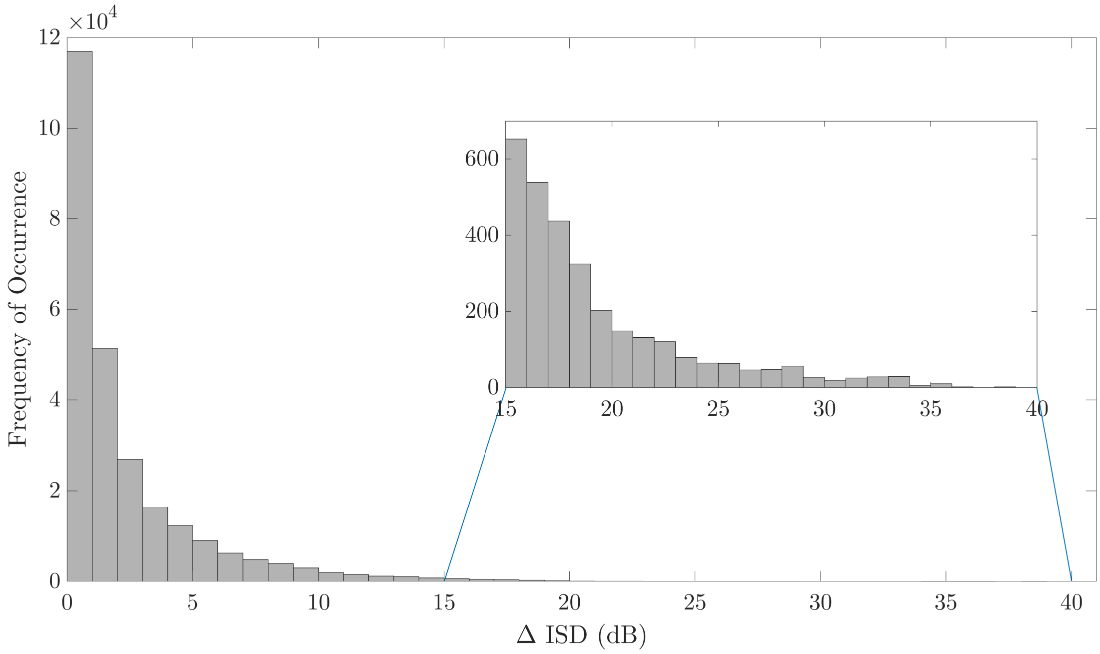


Figure 3.36: Distribution of ΔISD values for all directions and frequencies. The median, mean and 95 % interval are 1.2 dB, 2.4 dB and 9.0 dB, respectively.

Fig. 3.37 shows the mean, median and 95 % interval for ΔISD values for all directions as a function of frequency. Similar trends to those for monaural spectral comparison are apparent, with higher values in regions between 8 to 10 kHz and above 15 kHz. Again, mean and median values do not exceed 6 dB for any frequency point. Fig. 3.38 shows ΔISD values for all frequency points in the left panel, and split at 6 kHz in the right panel. The median, mean and 95 % interval values in the frequency range below 6 kHz are 0.5 dB, 0.7 dB and 2.2 dB, respectively. The median, mean and 95 % interval values in the frequency range above 6 kHz are 2.0 dB, 3.2 dB and 10.4 dB, respectively.

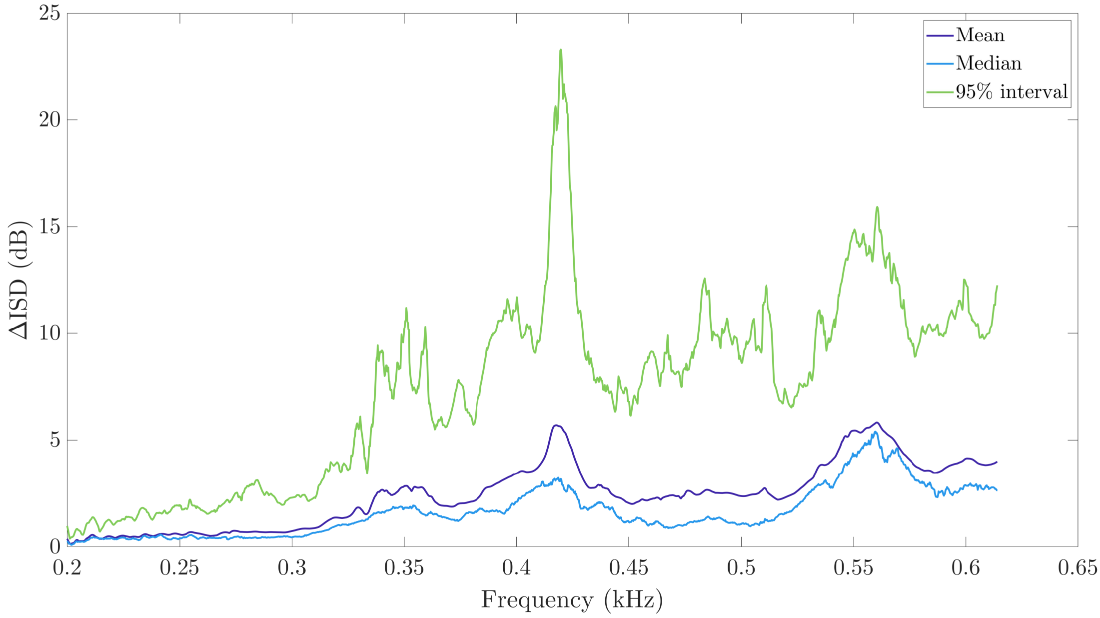


Figure 3.37: Mean, median and 95 % interval for ΔISD for all directions, as a function of frequency. Trends similar to those displayed for monaural spectral comparison (Fig. 3.32) are apparent.

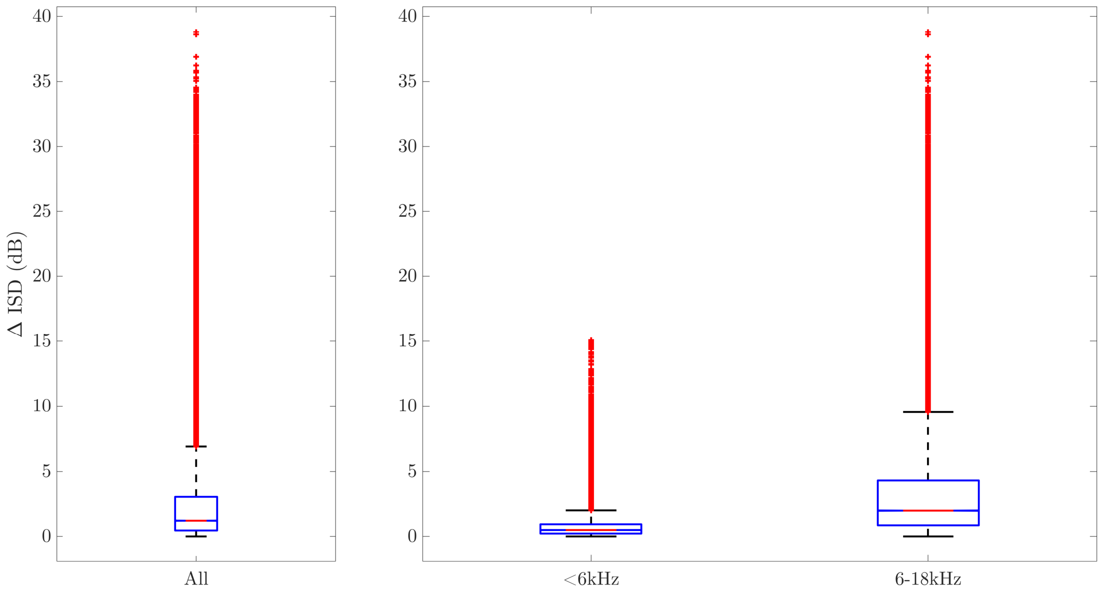


Figure 3.38: Distribution properties of ΔISD across all directions, for both all frequency points (left) and divided at 6 kHz (right). As for monaural spectral comparison (Fig. 3.33), there is an appreciable difference between the values either side of the divide.

As with ΔITD and SD_f, ΔISD has also been split by elevation band. This is shown in Fig. 3.39 and median, mean and 95 % interval values are listed in Table 3.8. As previously observed with ΔITD and SD_f, the highest elevation band ($\phi = 60°$) produces the smallest ΔISD values. As with SD_f, the variation with increasing elevation is less consistent than that observed with ΔITD.

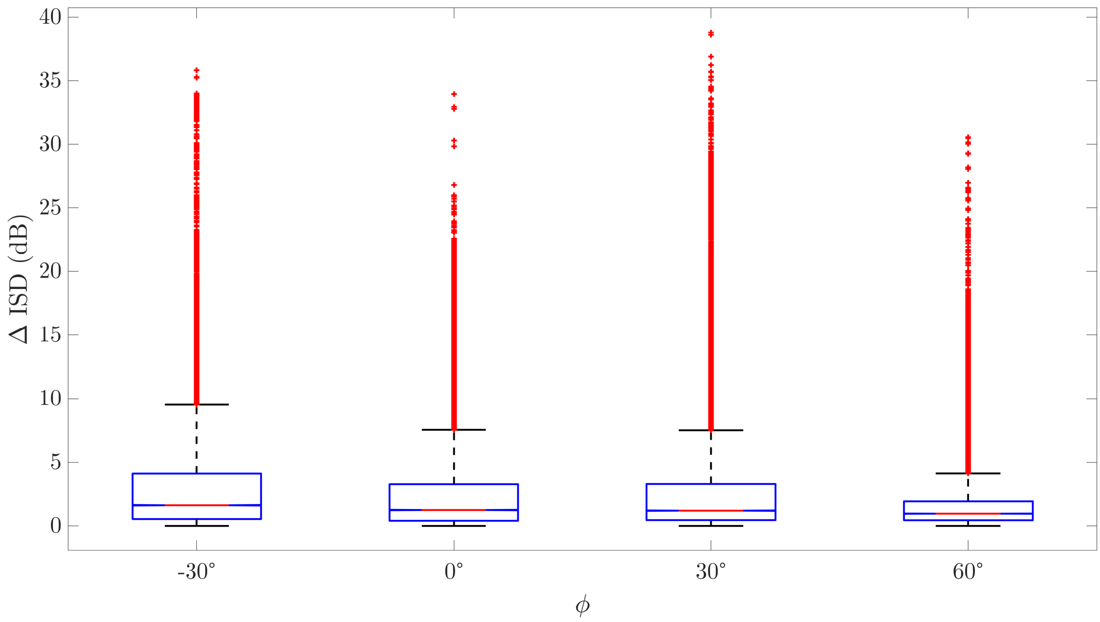


Figure 3.39: Distribution properties of ΔISD for all directions, separated into elevation bands. As with monaural spectral comparison (Fig. 3.35), less of a trend across elevation is visible than in the ΔITD data (Fig. 3.28).

Elevation Band	Median (dB)	Mean (dB)	95% interval (dB)
$\phi = -30°$	1.6	2.9	9.7
$\phi = 0°$	1.2	2.5	9.3
$\phi = 30°$	1.2	2.6	9.3
$\phi = 60°$	1.0	1.7	6.8

Table 3.8: Median, mean and 95 % interval for ΔISD for all frequencies, separated into elevation band. As with monaural spectral comparison, less of a trend across elevation is visible than in the ΔITD data.

3.8 Conclusion

This chapter describes the process of converting a CAD file of KEMAR into a format suitable for use within the BEM. Numerical and acoustic validation is described which validates the workflow for generating BEM-suitable meshes and confirms the consistency between simulation data from the processed mesh and acoustic measurements using the physical manikin. Five meshes have been created as a result of this process:

1. *full-torso-20*: full torso mesh with ear canals, valid for BEM simulation up to 23.0 kHz
2. *full-torso-noEC-20*: full torso mesh without ear canals, valid for BEM simulation up to 20.4 kHz
3. *full-torso-noEC-16*: full torso mesh without ear canals, valid for BEM simulation up to 16.5 kHz
4. *NEECK-20*: NEECK mesh, valid for BEM simulation up to 20.2 kHz
5. *NEECK-16*: NEECK mesh, valid for BEM simulation up to 16.5 kHz

These maximum valid frequencies were calculated for a dry air temperature of 20 °C (therefore a speed of sound of $c = 343.4\,\mathrm{m\,s^{-1}}$) using the six-edges-per-wavelength rule of thumb. Two intermediate-stage meshes have also been created during the conversion workflow:

1. *original-polygonal*: polygonal mesh form of the original CAD file of KEMAR
2. *remeshed-2mm*: *original-polygonal* mesh, remeshed to a target edge length of 2 mm

Numerical topological comparisons between the *full-torso-20 mesh* and the polygonal form of the original CAD file (*original-polygonal*) suggest the conversion workflow is adequate. Processing through all stages resulted in the accumulation of small errors in volume (an overall reduction of 0.02 %) and topology of the mesh (a median distance between the meshes of 0.006 mm and maximum distance of 0.63 mm).

To assess the perceptual impact of these errors, an acoustic validation was conducted using a head-and-shoulders representation of KEMAR referred to as the NEECK. The NEECK allowed comparison between acoustic measurement and the equivalent computational simulation using the same subject. A good match was generally observed between the simulations and the measurements. A rotational misalignment between measured and simulated results was discovered: applying a compensatory rotation of 5° in the horizontal plane to the simulated results reduced the systematic error.

The median of the differences between ITD values, ΔITD, is 20.8 µs, with 95 % of differences below 40.6 µs. This is in the range of 1 JND for ITD (typically between 10 and 75 µs [117]), suggesting that these differences are unlikely to be perceived by most listeners. A small number of outliers exist for the lowest elevation band, which can be attributed to an unexpectedly large ITD calculation for the small number of measured results. Additionally, an increase in ΔITD with a decrease in elevation is apparent, possibly due to a mechanical distortion of the measurement arc.

The median monaural spectral difference, SD_f, is 1.1 dB across all frequency bands, with 95 % of the values below 6.7 dB. Whilst this is above the lowest JND value for spectral amplitude of 1 dB [32, 118], a considerable variation across frequency is apparent. Below 6 kHz the median and mean values are similar (0.6 dB and 0.9 dB, respectively), and 95 % of the values are below 2.2 dB. Above 6 kHz, a gradual rising trend is apparent, with local peaks around 8 to 10 kHz and above 15 kHz. In this region, the median and mean values are more different (1.6 dB and 2.5 dB, respectively) and the 95 % interval is much larger at 7.8 dB. A small number of outliers exist as a result of misaligned notches, where a small shift in notch frequency results in a large numerical difference. The largest of these is 23.3 dB. However, the perceptual relevance of differences at higher frequencies is questionable. The use of a perceptual spectral difference model such as that proposed by Armstrong *et al.* [277] may go some way to analysing these differences perceptually rather than just only numerically. Separating the SD_f values by elevation band does not reveal as much of an influence of the measurement rig as with the ΔITD values. The highest elevation band produces the smallest SD_f values, but the pattern of variation with increasing elevation is less consistent.

The median of the difference between ISD values, ΔISD, is 1.2 dB, with 95 % of the values below 9.0 dB. This is also above the lowest reported JND value of 1 dB [32, 118]. ΔISD values display a similar trend to the SD_f results, with low values in the frequency range below 6 kHz (median of 0.5 dB), a gradual increase with frequency, and local peaks around 8 to 10 kHz and above 15 kHz. The outliers present in the SD_f analysis are also produced in ΔISD analysis, where a small shift in notch frequency results in a large numerical difference. In this instance, the largest difference is 38.9 dB. Separation across elevation band reveals a trend similar to that shown with the SD_f values: the highest elevation band produces the lowest ΔISD values, but variation with increasing elevation is less consistent.

The differences shown between measured and simulated results using these three metrics, whilst not wholly within perceptual limits, are mostly within the ranges reported by other studies [173, 274–276]. This suggests that a simulation using this mesh is unlikely to produce discrepancies from the ‘true’ HRTF larger than those introduced using direct acoustic measurement. For example, in this work, the presence of a rotational misalignment and the differences found across elevation bands suggests that the human operator and/or the mechanics of the system has affected the measured HRTFs. While the impact of external forces can be reduced with more elaborate measurement systems with higher precision and validation mechanisms (for example, that used by Armstrong *et al.* [124]), simulation still has the potential to avoid such problems.

Overall, these findings suggest that simulations using derivations of the *full-torso-20* mesh can safely be used in place of acoustic measurement when small discrepancies close to the threshold of perception are acceptable. The inability to simulate HRTFs for the *full-torso-20* mesh is an obvious drawback of this work. However, through numerical confirmation of the workflow and acoustic validation using the NEECK, it is a reasonable step to extend the assumption of validity from just the most salient portion of the mesh to the full torso. It may also be the case that the observed discrepancies are as a result of differences between the CAD file and the physical KEMAR due to manufacturing tolerances. However, these differences would be difficult to quantify and are likely to be small.

Despite the usefulness of the NEECK for acoustic validation here, it is not an accurate representation of a human listener as it has no shoulders or torso and has an unrealistic extended neck. Therefore, an alternative mesh is required to allow simulation of HRTFs that more closely match those from a real listener, whilst also being small enough to be simulated using the AudioLab computational facilities. The creation of such a mesh is discussed in Chapter 4.

Chapter 4

Creation of a Half Torso KEMAR Mesh

What do you do with the leftover you?

Tegan and Sara, *Where Does The Good Go*
So Jealous (2004)

4.1 Introduction

Evaluating the acoustic properties of a physical system using computational simulation techniques can be preferable to undertaking real measurements, particularly as simulation provides the ability to perform large numbers of measurements with small alterations to one or more parameters. In this research, one of the aims is to identify pairs of near-field loudspeaker positions that can satisfactorily reproduce binaural signals without the use of crosstalk cancellation (CTC). This involves the analysis of a very large number of loudspeaker positions and the computation of a variety of performance measures for each one; a task for which simulation is well suited.

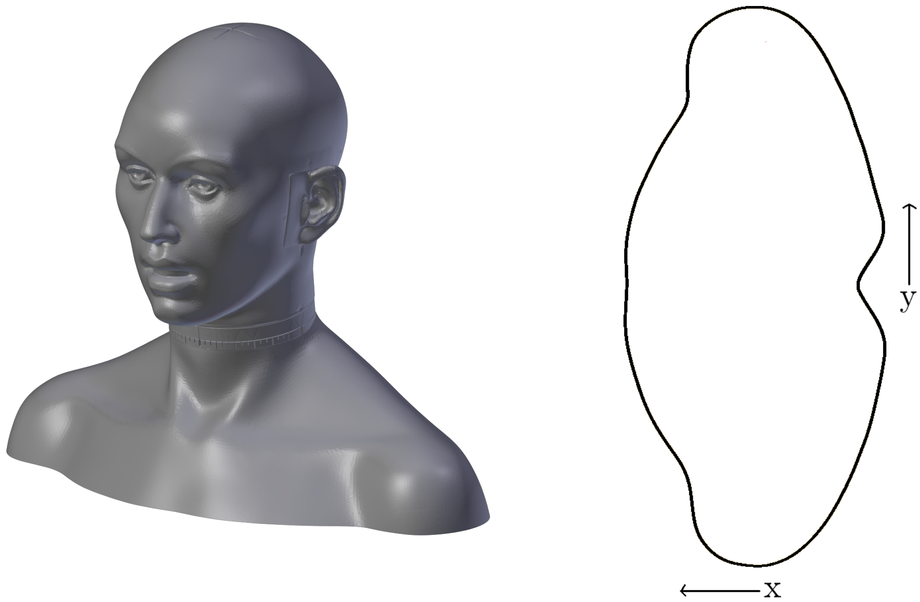
A mesh of the Knowles Electronics Manikin for Acoustic Research (KEMAR), suitable for simulation using the boundary element method (BEM), was created in Chapter 3. However, as discussed, at the time of writing the full torso mesh of KEMAR was too large to be used in BEM simulation software at the AudioLab, and so a head-only section known as the NEECK was used in acoustic validation. However, the NEECK is not an accurate representation of a listener as it has no shoulders and an unrealistic extended neck. An alternative mesh is, therefore, required to generate results more similar to those from a human listener, as the shoulders play an important role in the localisation of sound, particularly for elevated sound sources [79].

This chapter describes the creation of a mesh of KEMAR which is small enough to be computed using the facilities available in the AudioLab whilst maintaining topological accuracy. As the lower torso has a relatively weak effect [79, 153, 178], removal of the lower torso to reduce the overall size of the mesh would retain the important shoulder region and avoid interference patterns observed with the sharply-truncated full-torso KEMAR [153, 278]. After initial testing, it was determined that a so-called ‘half torso’ mesh of KEMAR would have a small enough number of faces to be simulated using the available BEM facilities (an estimated limit of approximately 250,000 faces). This mesh would also require processing using the workflow described in Chapter 3 to produce a mesh suitable for use within the BEM. The following steps are discussed in this chapter in relation to the creation of the half torso mesh:

1. Slicing the mesh to remove the torso;
2. Defining a Bézier curve with which to round the abrupt edge to reduce the diffraction artefacts;
3. Creating an additional rounded mesh section using the Bézier curve (more specifically, performing this within Blender [272]);
4. Joining the two sections together to form one mesh;
5. Processing the new mesh to make it suitable for use in the BEM.

4.2 Mesh Slicing

The *original-polygonal* mesh was used as the starting point for this process, as in the workflow described and validated in Chapter 3. After translation of the mesh and removal of the ear canals (Section 3.6.1), the 'Trim with X-Y Plane' tool in Geomagic Wrap [257] was used (with 'create boundary' enabled and 'close intersection' disabled) to slice the full torso mesh at $z = -0.2376$, where the z-axis is vertical. This produced the *head-and-shoulders-sliced* mesh with 192,851 faces and 97,018 vertices, shown in Fig. 4.1a. The z position of the plane was chosen to align with the top of the underarms to create a single edge loop (shown in Fig. 4.1b). In addition, slicing at this z position produced a mesh with a similar shoulder and torso region to those in the SYMARE database [80] which was intended for inclusion in later work in this thesis.



(a) *head-and-shoulders-sliced* mesh

(b) Edge loop of the *head-and-shoulders-sliced* mesh

Figure 4.1: The *original-polygonal* mesh, sliced with an X-Y plane at $z = -0.2376$ to create a head-and-shoulders portion (a) with a single open edge loop (b).

4.3 Bézier Curve Definition

To avoid the introduction of excessive edge artefacts, such as those discussed by Ziegelwanger *et al.* [278] and Algazi *et al.* [153], a Bézier curve was used to close the open base of the *head-and-shoulders-sliced* mesh with a rounded additional mesh section rather than an abrupt edge (the final result of this is shown in Fig. 4.6b). A 2D 3rd-order 50-segment Bézier curve was defined using Blender [272], the basic mathematics of which are outlined next before discussion of the Blender-specific implementation.

4.3.1 Mathematical Description

A Bézier curve is derived from a set of coordinates referred to as control points, and can be calculated from these using [279]:

$$\begin{aligned} B(\alpha) &= \sum_{n=0}^{N} b_{n,N}(\alpha)\boldsymbol{CP}_n \\ b_{n,N}(\alpha) &= \binom{N}{n}\alpha^n(1-\alpha)^{N-n}, \qquad n = 0, ..., N \end{aligned} \tag{4.1}$$

where $B(\alpha)$ is the Bézier curve, the polynomials $b_{n,N}(\alpha)$ are Bernstein basis polynomials of degree N where $\binom{N}{n}$ are the binomial coefficients and the points $\mathbf{CP}_n$ are the coordinates of the control points for the Bézier curve.

The control points can be separated into two groups based on function: anchored end control points (of which there are two) and intermediate control points (of which there can be any number). The order of the Bézier curve is given by the number of control points minus one. Unlike a polynomial curve, a Bézier curve does not pass through all the points used to define it. Instead, the intermediate control points affect the curvature of the line as it moves from the starting point CP_0 to point CP_n. The polygon formed by connecting the control points starting at CP_0 and ending at CP_n is referred to as the Bézier (or control) polygon. The convex hull is formed by joining the first and last points in the Bézier polygon.

The coordinates for the Bézier curve used in this work were determined using an iterative testing process. A curve was required that was compatible with both the steep gradient at the front of the mesh and the more shallow gradient at the back of the mesh, that is, a curve that did not create a sharp change in surface topology at the join between the *head-and-shoulders-sliced* mesh and the additional rounded portion. A visual representation of the Bézier curve using 50 segments is shown in Fig. 4.2. The coordinates used are listed in Table 4.1.

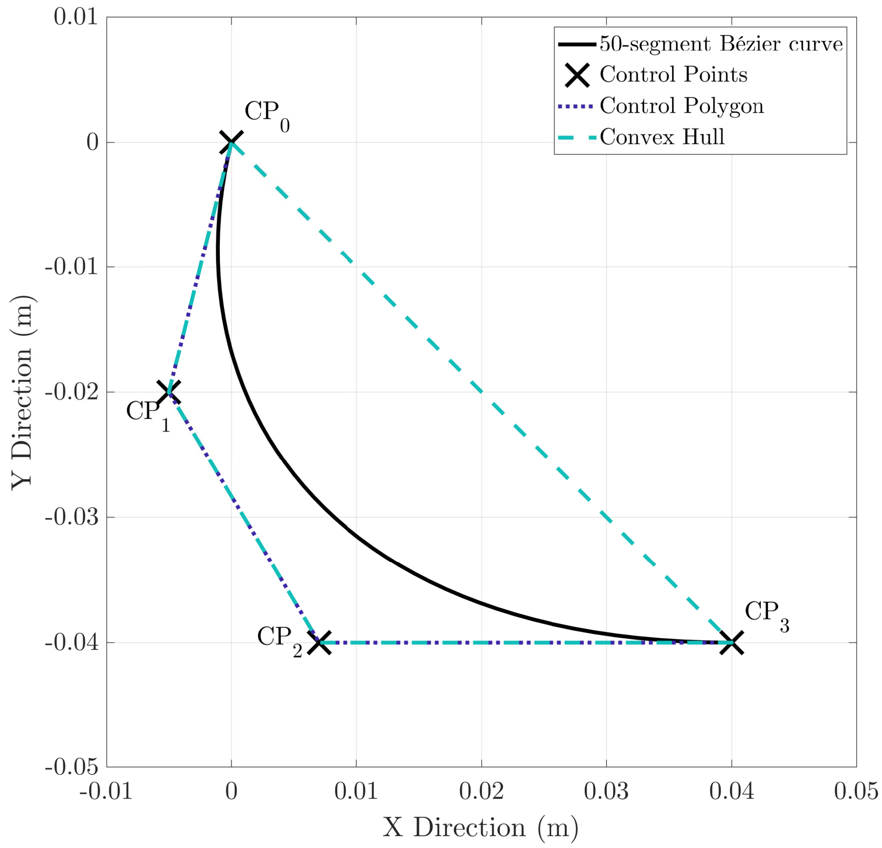


Figure 4.2: Visualisation of the 2D 3rd-order 50-segment Bézier curve used to round the *head-and-shoulders* mesh. Control points CP_0 to CP_3 (x) are indicated, in addition to the control polygon (dotted line) and the convex hull (dashed line).

Control Point	Coordinates (m)	
	x	y
CP_0	0.000	0.000
CP_1	−0.005	−0.020
CP_2	0.007	−0.040
CP_3	0.040	−0.040

Table 4.1: 2D coordinates for the control points used to define the Bézier curve.

4.3.2 Blender Implementation

For clarity and completeness, the process of creating a 3rd-order Bézier curve in Blender is included here. The 'Add' → 'Bézier Curve'[1] operation was used to create the initial curve. The resolution of the curve is given by the 'Resolution' value: in this case, a value of 50 was used to create a high-resolution curve that did not introduce additional diffraction artefacts due to corners arising between segments of the curve.

Blender uses slightly different terminology when implementing the control points, $\mathbf{CP}_n$ (as listed for this curve in Table 4.1). In Blender, $\mathbf{CP}_n$ are implemented as points which are referred to as *control points* and *handles*. Control points are the anchored end points, whilst the handles determine the curvature of the line between the end points (and are, therefore, functionally the same as the intermediate control points described above). Handles are visually connected to their respective end control point and operate as a pair of points with one on either side. Moving one handle of a pair moves the other, with the pivot at the control point. The handles within the convex hull have an influence on the curvature; handles outside the convex hull are merely for visual feedback as to the position of those inside the convex hull.

The Bézier curve used in this work is shown in Fig. 4.3; the similarity with Fig. 4.2 is apparent. For clarity, the control points are indicated with squares, the left handles with diamonds, and the right handles with circles. Control points CP_0 to CP_3 are labelled for comparison with Fig. 4.2, although these labels have no meaning within Blender. The trajectory of the line behind the end points is indicated with arrows. The left handle of CP_0 and the right handle of CP_3 are outside the convex hull, and so have no effect on the curvature. A particularly long right handle on CP_3 was used to ensure that the Bézier curve would make contact smoothly with a flat plane.

[1]The '→' icon is used in this chapter to denote a choice of tool option within the software in question.

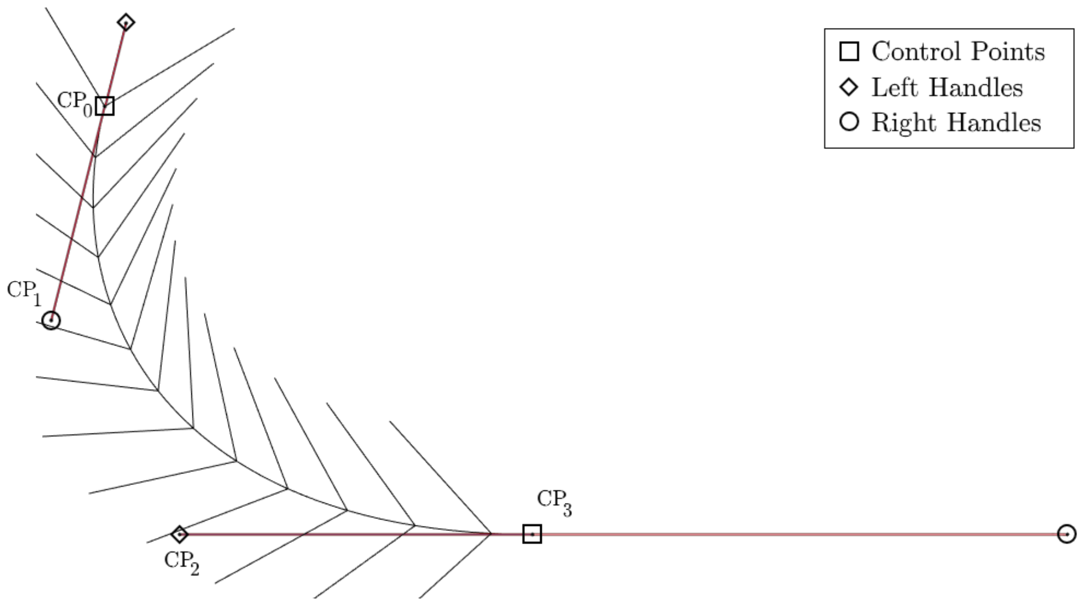


Figure 4.3: Visualisation of the 2D 3rd-order 50-segment Bézier curve as implemented in Blender. Control points are indicated with squares, left handles with diamonds, and right handles with circles. Control points CP_0 to CP_3 are labelled in accordance with the algebraic derivation, although these labels have no meaning within Blender.

4.4 Rounded Mesh Creation

To create the rounded section which was required to close the base of the mesh, the Bézier curve in Fig. 4.3 was applied to each vertex in the loop shown in Fig. 4.1b. The 'bevel' modifier in Blender is normally used to apply a bevel to the corner of an object. However, in this instance, it can be used to create a 3D entity with the desired shape taken from a 2D loop. The Bézier curve is applied to each vertex in the loop in turn, and new vertices are created along the curve. This required several steps, outlined below for clarity and completeness.

The edge loop of vertices of the *head-and-shoulders-sliced* mesh (Fig. 4.1b) was selected and a new entity created so as to apply the modifier only to the loop. It is not possible to apply a modifier to a mesh, so the separated loop was converted to an object. The Bézier curve was then applied to the object using the 'Bevel Object' modifier to create the additional rounded section, which was then converted back to a mesh to create the additional faces. The user interface operations required for these steps are described in Fig. 4.4.

Additional processing using the 'Mesh Doctor' tool in Geomagic Wrap was required to clean the mesh. In some regions with a large radial change between neighbouring vertices, the length of the Bézier curve meant that some self-intersections and duplicate vertices were created (for example, between the shoulder blades as shown in Fig. 4.5) which required removal for mesh optimisation.

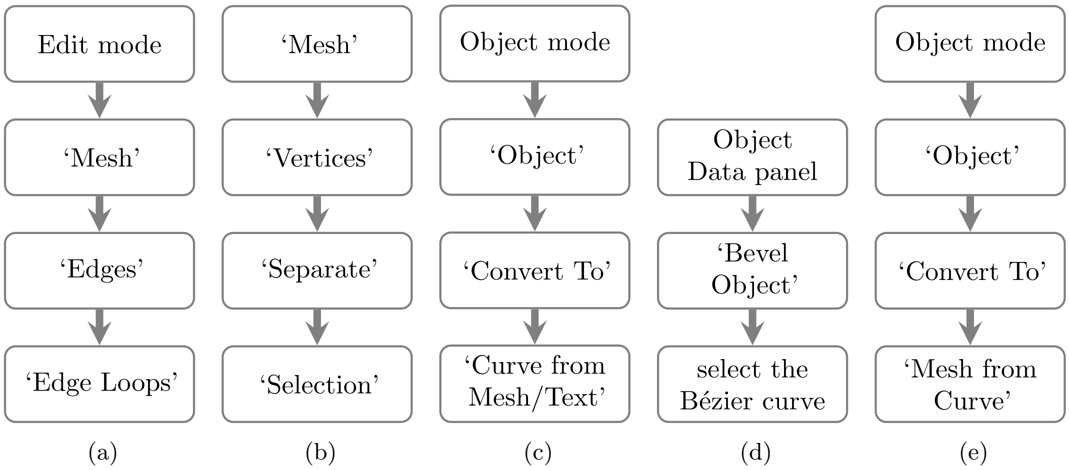


Figure 4.4: The user interface operations used in Blender to create the initial rounded mesh section: a) selecting the edge loop, b) separating the edge loop, c) converting separate loop to a object, d) applying the Bézier curve using the 'Bevel Object' modifier, e) converting the rounded curve object to a mesh.

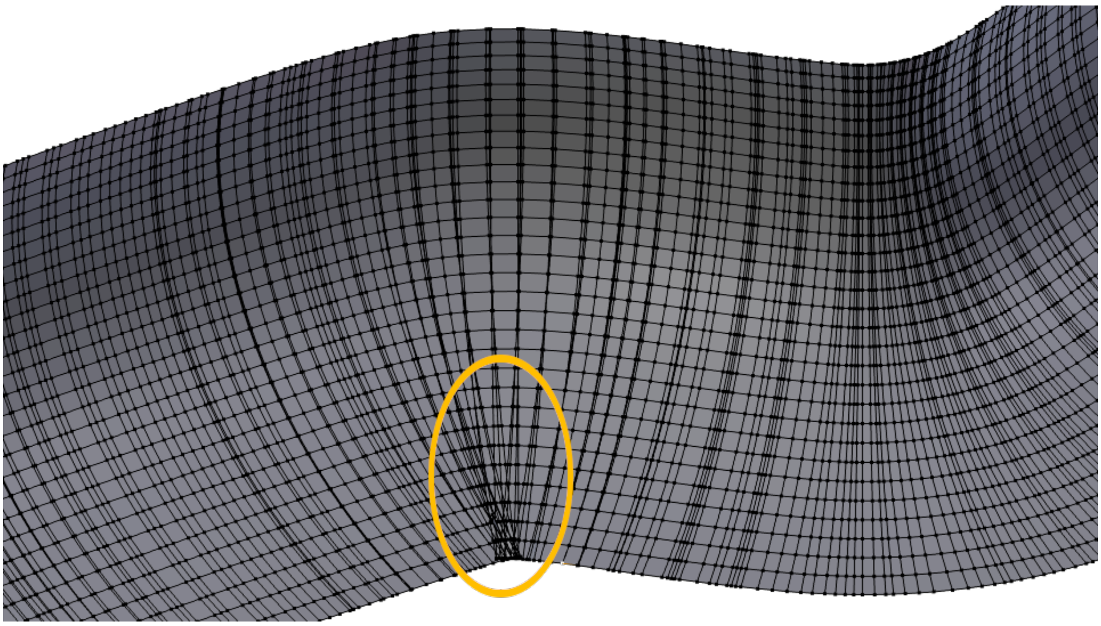
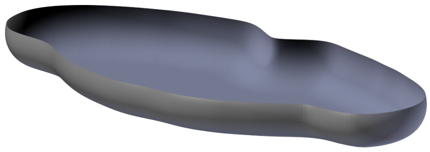


Figure 4.5: Problematic self-intersections (indicated within the ring) created at regions of high radial curvature as a result of overlapping neighbouring Bézier curves. These were removed for optimal BEM performance.

The 'Fill Hole → with Tangent' tool in Geomagic Wrap was used to fill the planar hole on the underside to create the complete rounded section with an open top shown in Fig. 4.6a, referred to as the *rounded-portion* mesh. This section has 85,437 faces and 48,008 vertices. Face normals were corrected to be all exterior rather than interior, as, although the face normal has no influence on the appearance on the mesh, the BEM simulation requires that face normals point towards the acoustic medium. The *rounded-portion* mesh was aligned with the *head-and-shoulders-sliced* mesh in Blender and the two sections joined using the 'Edit Geometry → Attach' tool to create a closed, rounded, half-torso mesh consisting of 278,288 faces and 145,026 vertices (shown in Fig. 4.6b, referred to as the *joined* mesh). The *rounded-portion* mesh extended the *head-and-shoulders-sliced* mesh in the negative z direction by 40 mm.



(a) *rounded-portion* mesh

(b) *joined* mesh. The join can be seen as a result of a small surface orientation discontinuity at the join.

Figure 4.6: The *rounded-portion* mesh (a) created from the edge loop of the *head-and-shoulders-sliced* mesh, attached to the *head-and-shoulders-sliced* mesh to form (b) the *joined* mesh.

4.5 Processing for BEM Use

The *joined* mesh is not optimised for BEM simulation with regards to edge length, face shape and face distribution. The maximum edge length is 28.34 mm, with 53,112 edges longer than the 2.86 mm limit for simulation to 20 kHz (12.7 %) and 11,466 longer than the 3.58 mm limit for 16 kHz (2.7 %). The distribution of edge lengths is shown in Fig. 4.7, where the median value is 1.98 mm.

The distribution of angles between edges is shown in Fig. 4.8. The large frequency of occurrence in the region of 90° is a result of the arrangement of triangles in the added rounded portion. The smallest angle is 0°, the largest is 180° and the median is 59.8°. These values indicate the presence of long skinny triangles, and so processing is required for BEM suitability as uniformly-distributed approximately equilateral triangles are computationally advantageous [260]. The variation in edge length and face shape and size are visible in the left shoulder portion of the *joined* mesh shown in Fig. 4.9, in addition to the high number of right-angled triangles in the rounded portion.

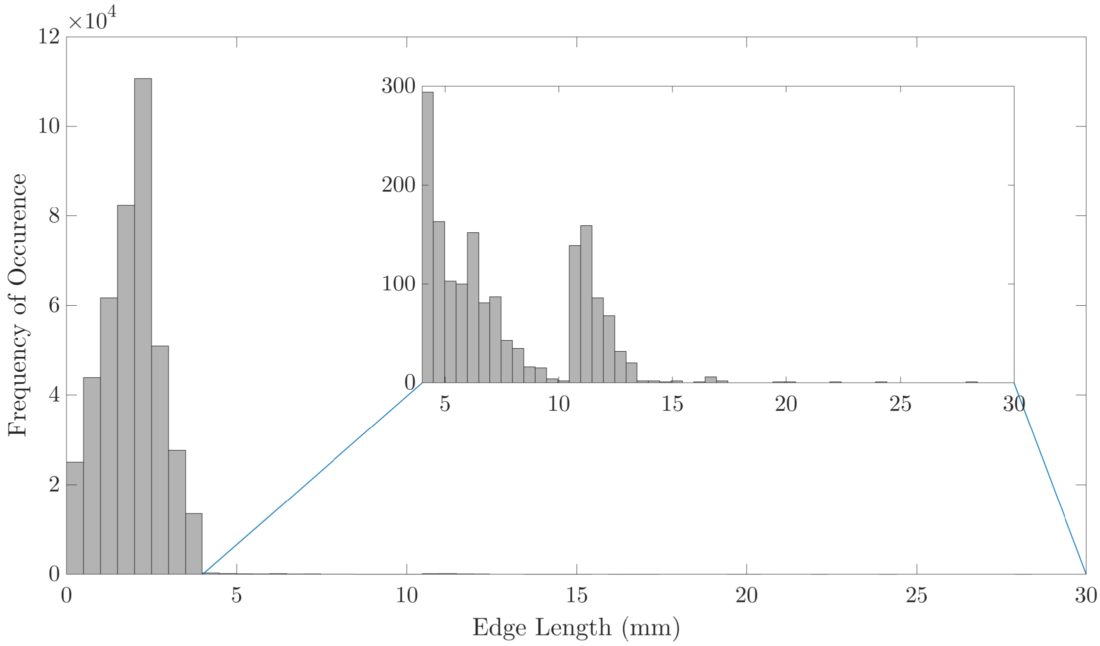


Figure 4.7: Distribution of edge lengths in the *joined* mesh. The shortest edge length is 9.98×10^{-4} mm, the longest is 28.34 mm and the median is 1.98 mm.

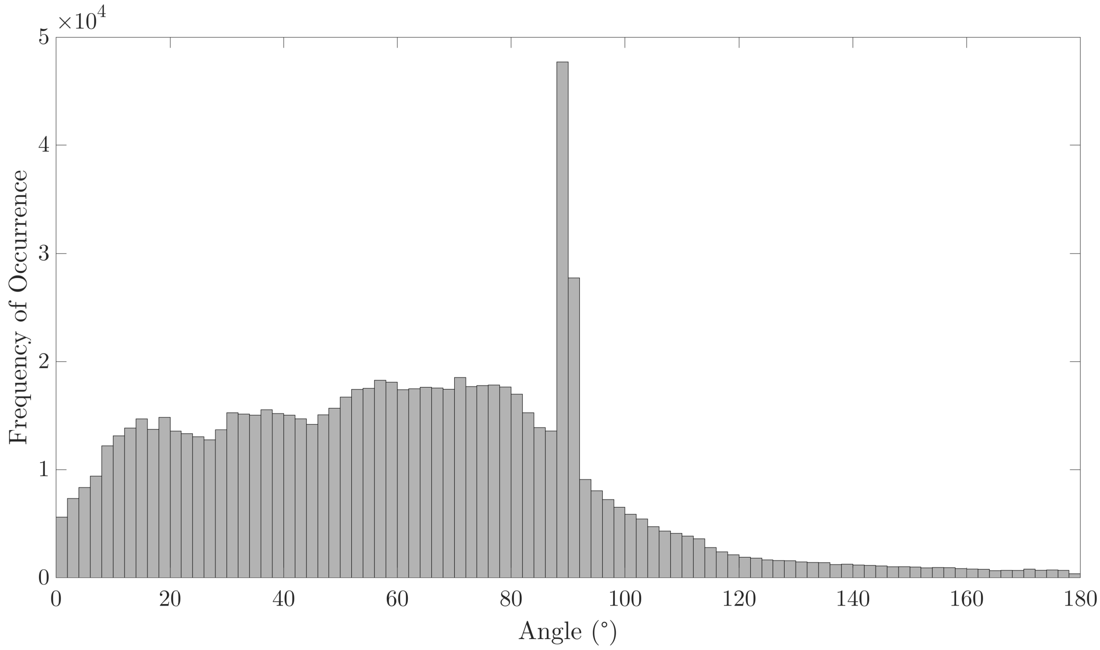


Figure 4.8: Distribution of angles between edges in the *joined* mesh. The smallest angle is 0°, the largest is 180° and the median is 59.8°.

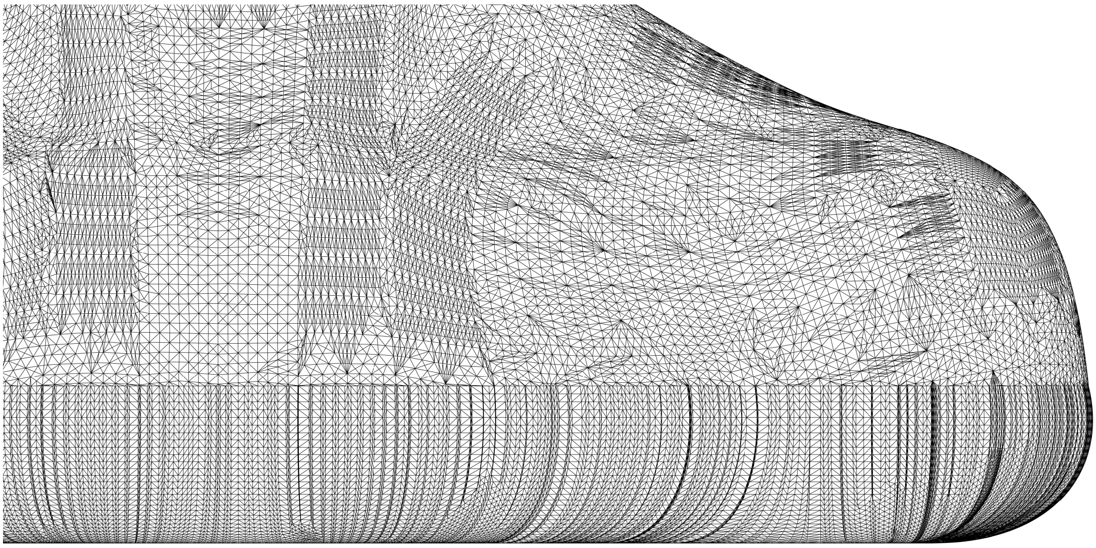


Figure 4.9: The left shoulder region of the *joined* mesh, showing the variation in edge length and face shape and size.

The workflow described and validated in Chapter 3 (and recapped briefly here) was used to process the *joined* mesh for use within the BEM. The 'Remesh' tool within Geomagic Wrap was used to redefine the surface with a target edge length of 2 mm, resulting in a mesh with 248,932 faces and 124,468 vertices with a maximum edge length of 3.19 mm. The open source software ACVD [255, 256] was then used to redistribute the vertices more uniformly across the surface. Two versions of the half torso mesh were created using target vertices values of 140,000 and 90,000, respectively: *half-torso-20* mesh valid to 20.7 kHz and *half-torso-17* mesh valid to 17.1 kHz, both calculated for a speed of sound of $c = 343.4\,\mathrm{m\,s^{-1}}$ and the six-edges-per-wavelength rule of thumb. Numbers of faces, vertices and maximum edge lengths for these two meshes are listed in Table 4.2. Fig. 4.10 shows the same region as that shown in Fig. 4.9, but for the *half-torso-17* mesh. The faces are more uniform in both shape and distribution across the surface. The distributions of edge lengths and angles between edges for both meshes are shown in Figs. 4.11–4.14. The full *half-torso-17* mesh is shown in Fig. 4.15.

Mesh version	Vertices	Faces	Max edge length (mm)
half-torso-20	140,100	280,196	2.77
half-torso-17	90,000	179,996	3.35

Table 4.2: Numbers of vertices, faces and maximum edge lengths in the half torso BEM-suitable meshes.

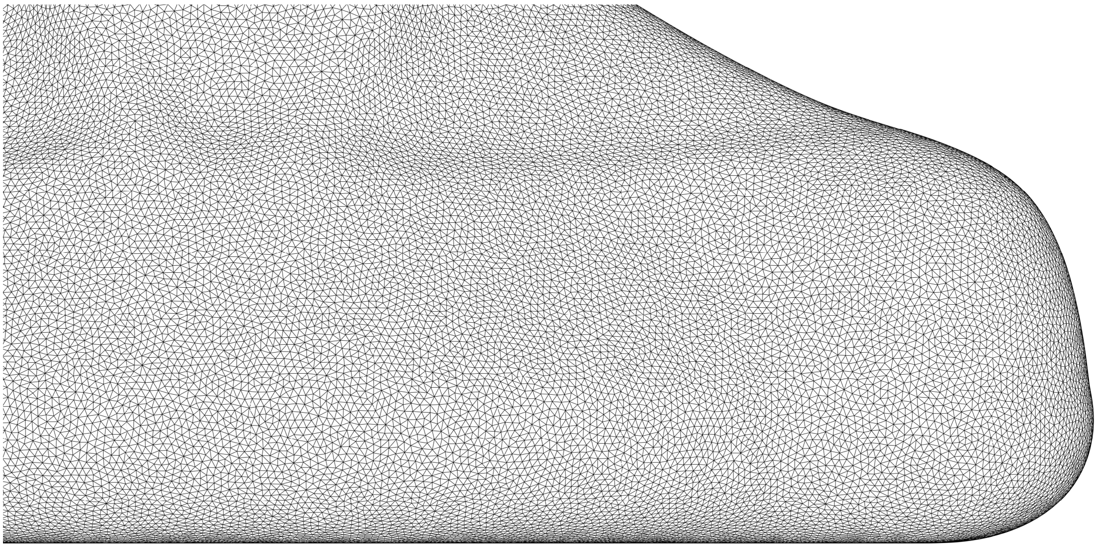


Figure 4.10: The same left shoulder region as shown in Fig. 4.9 from the *half-torso-17* mesh, showing the more regular distribution of face shape and size after processing for use within the BEM.

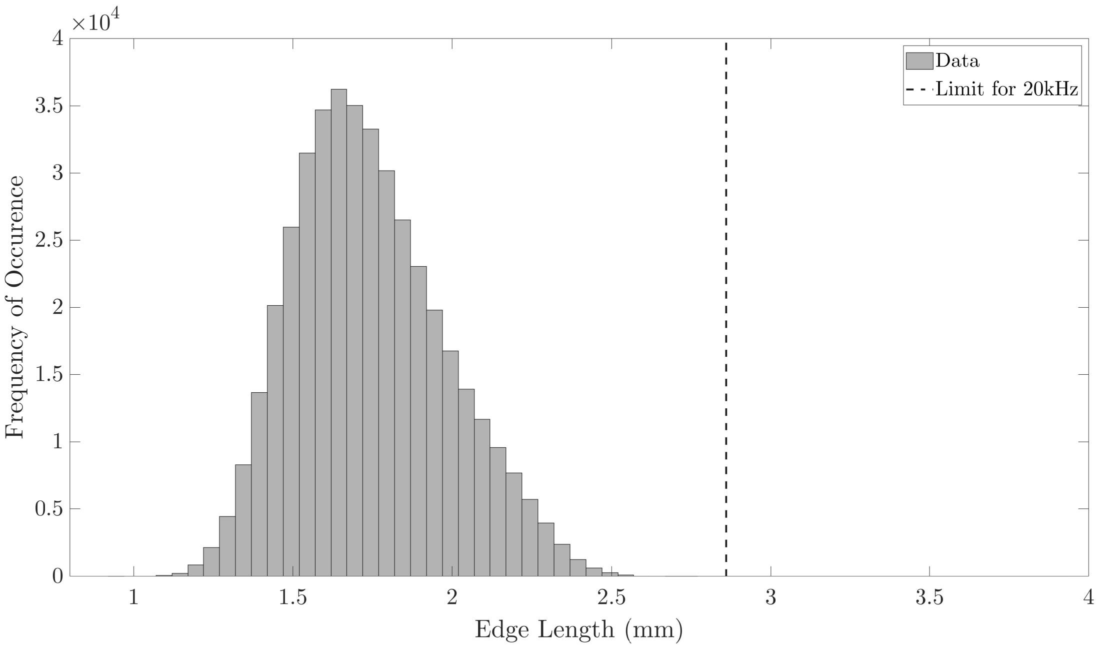


Figure 4.11: Distribution of edge lengths in the *half-torso-20* mesh. The shortest edge length is 0.93 mm, the longest is 2.77 mm and the median is 1.72 mm. The maximum edge limit of 2.86 mm for 20 kHz validity is indicated (dashed line).

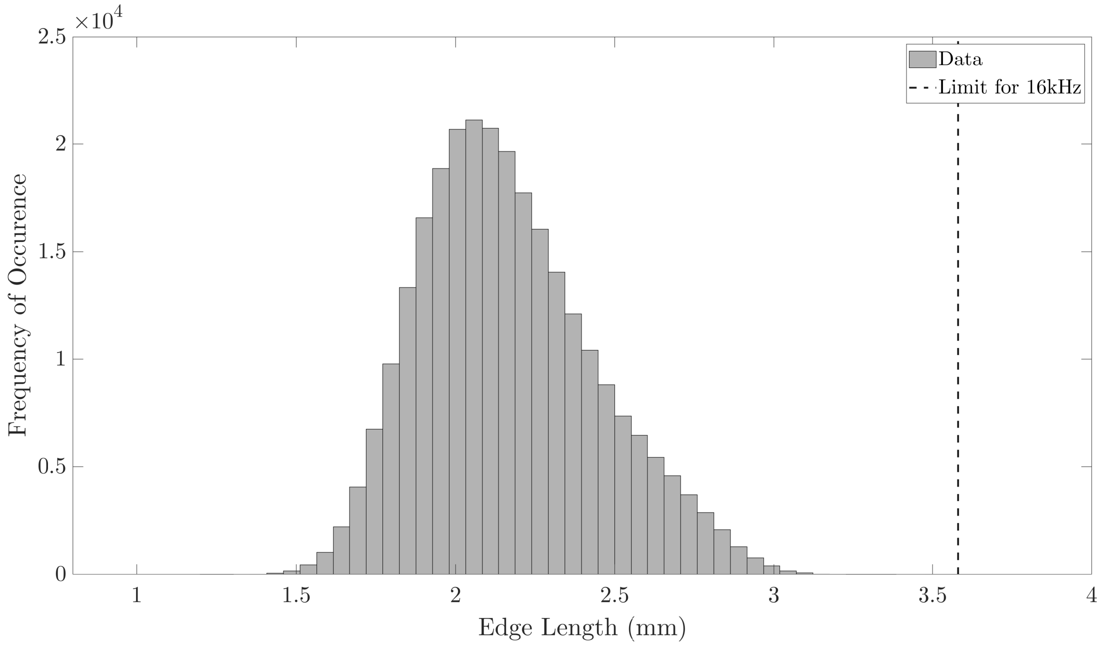


Figure 4.12: Distribution of edge lengths in the *half-torso-17* mesh. The shortest edge length is 1.25 mm, the longest is 3.35 mm and the median is 2.13 mm. The maximum edge limit of 3.58 mm for 16 kHz validity is indicated (dashed line).

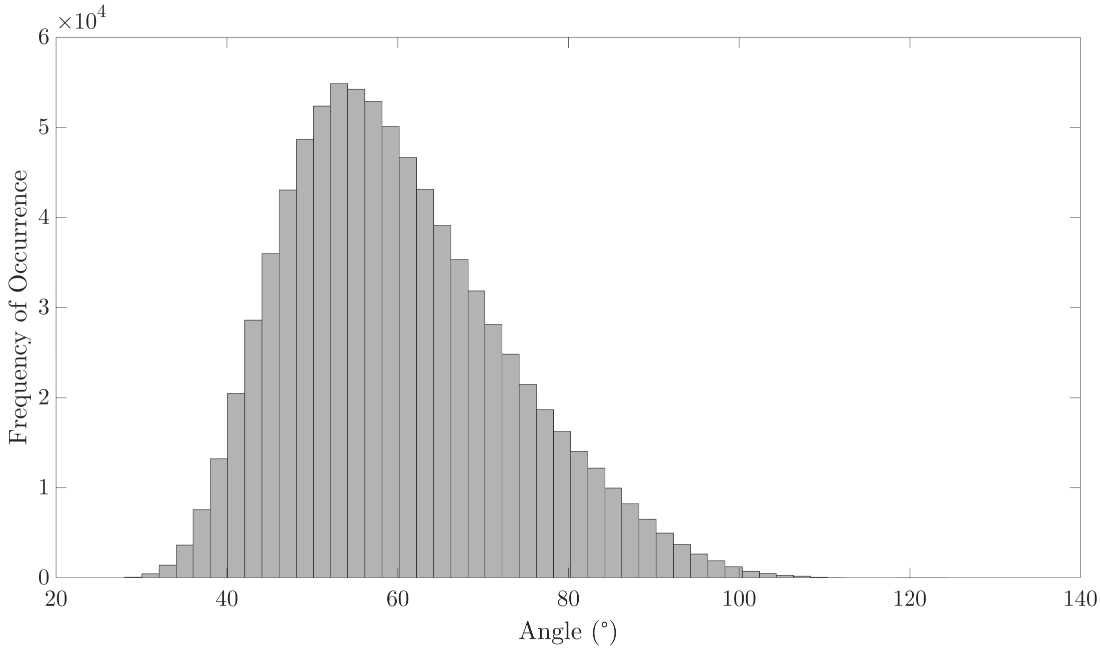


Figure 4.13: Distribution of angles between edges in the *half-torso-20* mesh. The smallest angle is 26.2°, the largest is 124.1° and the median is 58.3°.

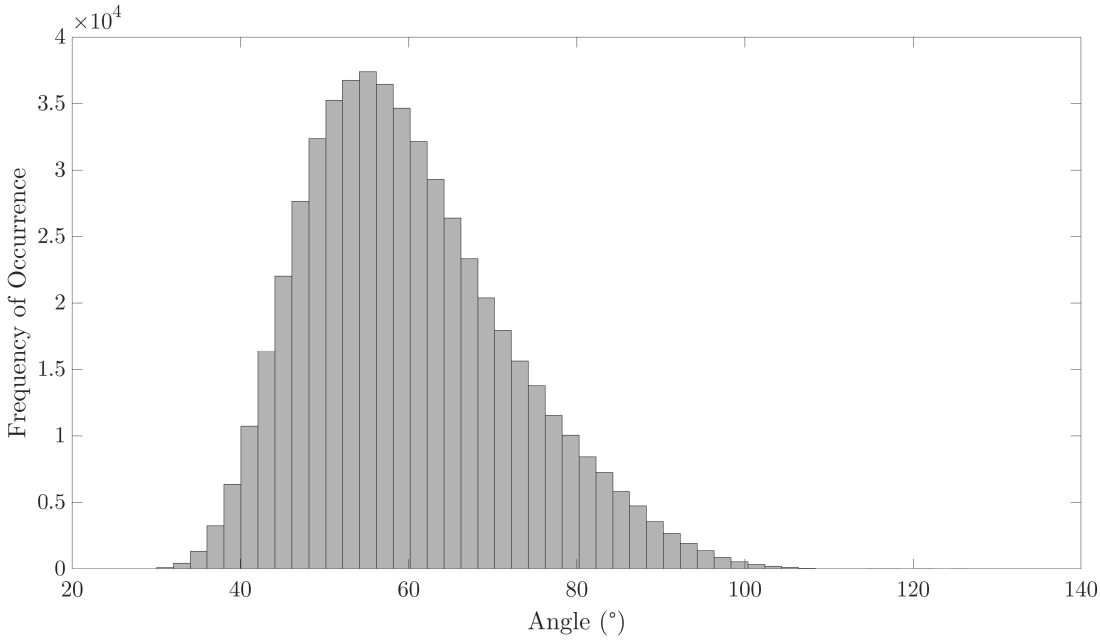
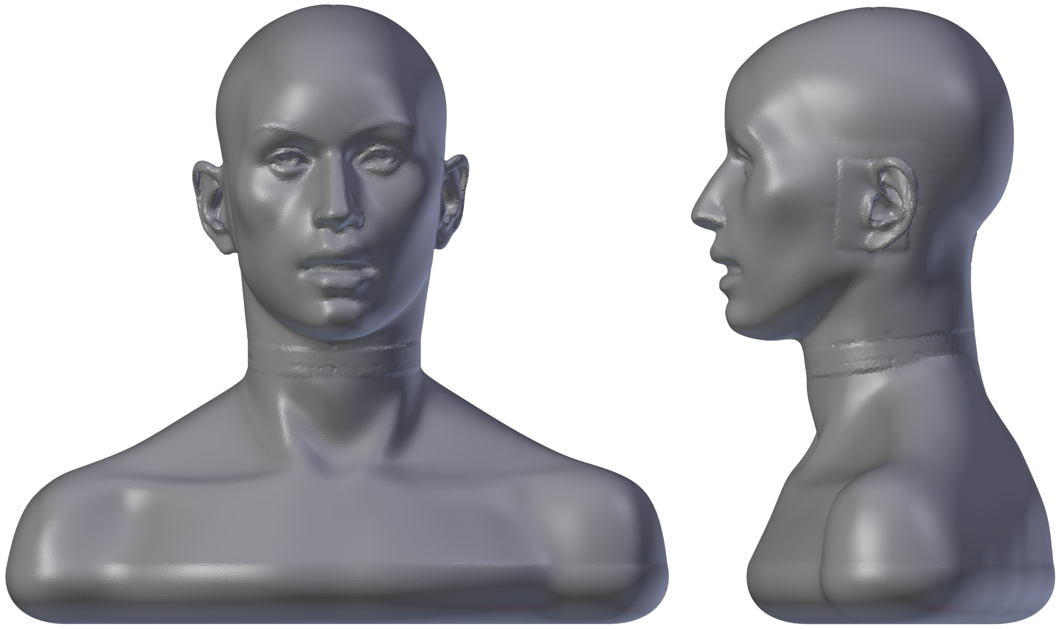


Figure 4.14: Distribution of angles between edges in the *half-torso-17* mesh. The smallest angle is 26.1°, the largest is 126.0° and the median is 58.3°.



(a) Front view

(b) Left view

Figure 4.15: The *half-torso-17* mesh.

4.6 Conclusion

This chapter describes the process of creating a half torso mesh of KEMAR for the purposes of reducing computational requirements for BEM simulation. Geomagic Wrap and Blender were used to slice the mesh and create an additional rounded mesh portion to reduce diffraction artefacts such as those described by Ziegelwanger *et al.* [278] and Algazi *et al.* [153]. This additional portion is based on a Bézier curve. The two portions have been aligned, attached and processed to produce a mesh suitable for simulations using the BEM. Two meshes have been created as a result of this process:

1. *half-torso-20*: head and shoulders mesh, processed using the workflow validated in Chapter 3, valid for use in BEM simulation up to 20.7 kHz.
2. *half-torso-17*: head and shoulders mesh, processed using the workflow validated in Chapter 3, valid for use in BEM simulation up to 17.1 kHz.

These maximum valid frequencies were calculated for a dry air temperature of 20 °C (therefore a speed of sound of $c = 343.4\,\mathrm{m\,s^{-1}}$) using the six-edges-per-wavelength rule of thumb. Three intermediate-stage meshes have also been created during the conversion workflow:

1. *head-and-shoulders-sliced*: the *original-polygonal* mesh as described in Section 3.4, sliced with an X-Y plane positioned at the top of the underarms.
2. *rounded-portion*: an additional rounded section of mesh created using a Bézier curve.
3. *joined*: the *rounded-portion* mesh and *head-and-shoulders-sliced* mesh, aligned and attached.

Reducing the mesh by removing the lower torso in this manner was intended to reduce the number of faces to allow simulation using AudioLab BEM facilities. The three full torso meshes discussed in Chapter 3 all contain more faces than the approximate limit for simulation found during initial testing, as discussed in Section 3.6.2 (720,015, 543,118 and 314,496 faces, respectively). Correspondingly, these meshes are not suitable for simulation using the BEM facilities at the AudioLab. The NEECK mesh contains 123,996 faces, and is therefore suitable for simulation, but is not an accurate representation of a human listener and so has limited use. Of the two head and shoulders meshes discussed in this chapter, only the *half-torso-17* mesh has been successfully used in BEM simulation, as the *half-torso-20* mesh has more faces than can be handled by the AudioLab's simulation software. This is not regarded as a hindrance, however, as 16 kHz is on the limit of perception for the adult human hearing system [280, 281]. Accordingly, the *half-torso-17* mesh is used in this work from this point onwards. In order to begin to examine the impact of loudspeaker position on the binaural reproduction, Chapter 5 describes using the *half-torso-17* mesh to simulate a large number of HRTFs using the BEM.

Chapter 5

Variation of Natural Channel Separation with Source Position

I'm up and doing circles.

Tegan and Sara, *Are You Ten Years Ago*
The Con (2007)

5.1 Introduction

As discussed in Chapter 2, binaural reproduction over loudspeakers typically requires crosstalk cancellation (CTC) to produce the necessary channel separation [231, 233]. This is sometimes known as 'transaural' reproduction [218], and ensures that the left channel signal arrives only at the left ear and that the right channel signal arrives only at the right ear. It does this by cancelling out the interfering signal reaching the left ear from the right loudspeaker and the interfering signal reaching the right ear from the left loudspeaker, therefore increasing the channel separation between the two ears. However, inclusion of CTC can introduce unwanted artefacts (see Section 2.7.2).

Channel separation also occurs inherently, as a result of the presence of a head between two ears. This natural channel separation (NCS) is a measure of the acoustic isolation which exists between the ears for a sound source in any given direction [19]. It is produced by the acoustic shadowing provided by the head and torso, and is often used as the ground truth value against which the success of CTC systems are measured. Irrespective of the method of generation, a channel separation of 20 dB has been reported to be sufficient for preserving the ability of a listener to localise sounds in reproduced binaural signals [17, 228].

Since acoustic shadowing increases as the radial distance of the source sound decreases, so does NCS [230]. It follows, therefore, that as a sound source approaches the head, the associated channel separation may become large enough to support binaural reproduction without the need for CTC. A number of studies have investigated NCS at specific directions and distances (for example: [19, 22, 23, 231]), but to the author's knowledge, a systematic study of the variation of NCS with both direction and radial distance has not previously been published.

This chapter describes such a study into the variation of NCS with direction and radial distance. The study utilises the *half-torso-17* mesh created in Chapter 4 in simulating a large number of head-related transfer functions (HRTFs) across 15 radial distances, from which the NCS is calculated and the variation with loudspeaker direction and radial distance is discussed. Source positions are identified which exhibit NCS greater than the 20 dB threshold and which are, therefore, candidate positions for binaural reproduction without the use of CTC.

5.2 Methods

The intention of this study was to examine the variation of NCS over a very large number of directions and a range of radial distances in both the near-field and far-field. In order to calculate NCS, a pair of HRTFs between each desired sound source and the two ears is needed. Simulation using the boundary element method (BEM) lends itself to this type of task.

5.2.1 Simulation of Near-Field HRTFs

HRTFs were simulated using the *half-torso-17* mesh of the Knowles Electronics Manikin for Acoustic Research (KEMAR) described in Chapter 4 and multi-level fast multipole boundary element method (FM-BEM) as implemented in the open source software Mesh2HRTF [179–181]. 655,214 HRTF pairs were simulated in 14 spherical domes with 46,801 points at each radial distance (listed in Table 5.1). As in previous chapters, azimuth (θ) increases counter-clockwise from zero in the frontal direction, such that $\theta = 90°$ is on the left and $\theta = 270°$ is on the right. Elevation (ϕ) is measured with respect to the horizontal plane containing the interaural axis, where $\phi = 90°$ describes a source directly above the head and $\phi = -90°$ is directly below. Radial distances (r) are measured relative to the origin (the mid-point of the interaural axis) and each ear lies approximately 65 mm from the origin along the interaural axis.

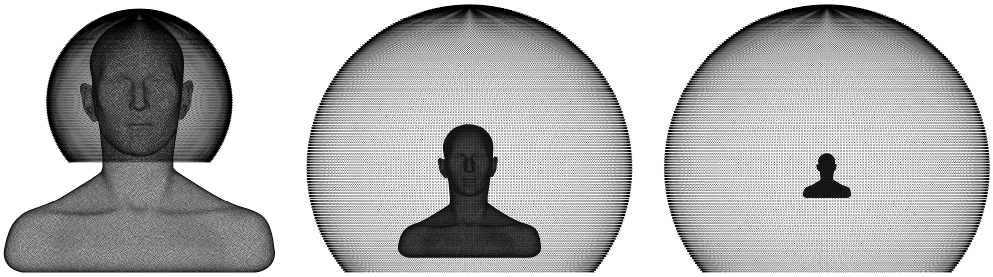
Value	Bounds	Increment
θ	0 to 359°	1°
ϕ	-40 to 90°	1°
	0.15 to 0.50 m	0.05 m
r	0.60 to 1.00 m	0.10 m
	1.50 m	N/A

Table 5.1: The 655,214 source positions in domes used in the simulation. The source position at $\phi = 90°$ exists only at $\theta = 0°$ for each radial distance for computational efficiency.

For computational efficiency, the source position at $\phi = 90°$ exists only at one azimuth value ($\theta = 0°$) for each radial distance, and therefore each dome used in the simulation consists of $(360 \times 130) + 1$ positions. The 14 radial distances were chosen to allow analysis within the region commonly referred to as the near-field [36, 102] as well as a small number of radial distances in the far-field. An additional two positions were simulated at $r = 0.10$ m, directly outside each ear ($\phi = 0°,\ \theta = 90°$ and $\phi = 0°,\ \theta = 270°$).

These were included to act as somewhat of a reference, as these positions are likely to be subject to the greatest amount of head shadowing.

Elevation angles below $-40°$ were omitted as many of these positions lie inside the *half-torso-17* mesh for radial distances less than $r = 0.40\,\mathrm{m}$ and are, therefore, not valid for simulation or analysis. Examples of the spherical domes for the cases of $r = 0.15\,\mathrm{m}$, $r = 0.50\,\mathrm{m}$ and $r = 1.50\,\mathrm{m}$ are shown in Fig. 5.1. Inner domes have been removed for the purpose of visualisation. A dome was not used at $r = 0.10\,\mathrm{m}$, as many of these positions were also not valid for simulation or analysis. The evaluation grid of 655,216 source positions was prepared using Delaunay triangulation in MATLAB [270, 271] and the pre-processing steps were performed in Blender [272] to satisfy the requirements of Mesh2HRTF.



(a) $r = 0.15\,\mathrm{m}$ (b) $r = 0.50\,\mathrm{m}$ (c) $r = 1.50\,\mathrm{m}$

Figure 5.1: Three of the spherical domes, each comprising 46,801 positions, used in the simulation of HRTFs of the NEECK for three radial distances: a) 0.15 m, b) 0.50 m and c) 1.50 m. Where they exist, inner domes have been removed for the purpose of visualisation.

As discussed in Section 2.6.2, BEM solvers require knowledge of the speed of sound and the density of the medium. The results from this simulation were not intended for comparison against real-world, acoustically-measured results, as was done in Chapter 3. Therefore, software default values were used for the speed of sound and the density of air ($c = 343.0\,\mathrm{m\,s^{-1}}$ and $\rho = 1.21\,\mathrm{kg\,m^{-3}}$, respectively). This gives a maximum analysis frequency for the mesh of 17.1 kHz, calculated using Equation 2.57 with a maximum edge length of $d_{max} = 3.35\,\mathrm{mm}$ and the six-edges-per-wavelength rule of thumb ($\eta = 6$). The vibrating element in the ear was chosen as the element nearest to the centre of the microphone as calculated in Section 3.6.1.

The 655,216 HRTF pairs were computed at 159 frequencies in 100 Hz increments from 200 Hz to 16 kHz. As in Chapter 3, the simulation ran on an Ubuntu 16.04.2 machine with 16 central processing units (CPUs) clocked at 2.60 GHz, each accessing 16 GB of random-access memory (RAM). Using the principle of reciprocity (discussed in Section 2.6.2), the 16 concurrent threads cumulatively took approximately 964 hours at an average of 60.3 hours per thread. The maximum time for any thread, and therefore the actual simulation time as all threads ran simultaneously, was 62.4 hours. The additional MATLAB processing stage as required by Mesh2HRTF (Output2HRTF) took approximately 69 hours, giving an overall time of 130 hours.

In [190], Katz suggests that increasing the number of simulated HRTF positions has a negligible impact on computation time. Whilst this may be true for smaller numbers of additional positions, it is evident here that substantial computational overhead is involved when increasing the number of positions to 655,216. This may be due, in part, to the simulation software used. Of the three calculation stages within Mesh2HRTF ('assembling', 'solving' and 'post processing'), only the third stage substantially increased in duration with larger numbers of simulation positions. The length of time for the required additional MATLAB processing stage also increased linearly with the addition of more positions.

5.2.2 Calculation of Natural Channel Separation

NCS as a function of direction and radial distance has been calculated as the arithmetic average spectral magnitude difference within a frequency range for each source position, as defined in Section 2.7.2.2:

$$NCS = \frac{1}{N_f} \sum_{j=1}^{N_f} \left| A_{ipsi_j} - A_{contra_j} \right| \tag{5.1}$$

where N_f is the number of points in the frequency range, and A_{ipsi} and A_{contra} are the ipsilateral and contralateral HRTF magnitudes in dB for a source at direction $\theta,\ \phi,\ r$. 20 dB of channel separation has been reported to be the lower threshold for robust binaural reproduction [17, 228]. Since the NCS calculations performed by Lacouture Parodi and Rubak [228] (and in other related work) used a frequency range of 200 Hz to 8 kHz, and the inclusion of higher frequencies can lead to a larger value of NCS [229], the same range has been used here to facilitate comparison. The values of NCS calculated using this method and frequency range are comparable with corresponding values reported by others, as listed in Table 5.2.

	Source Position ($\theta,\ \phi,\ r$)	Reported NCS (dB)	Calculated Here (dB)
Prodi and Velecka [231]	5°, 0°, 1 m	<5	1.7
Prodi and Velecka [231]	15°, 0°, 1 m	6	5.8
Prodi and Velecka [231]	30°, 0°, 1 m	10	11.6
Masiero [227]	45°, 0°, N/A	13.8 to 16.4	16.7
Elliott *et al.* [23]	90°, 0°, ≈ 0.2 m	20	30.3

Table 5.2: NCS reported by others and calculated here under comparable conditions. Radial distance is not provided in [227], therefore, a far-field distance of 1 m was used for the comparable calculation. Radial distance in [23] is indicated as approximate, as the exact distance is not provided.

5.3 Results

The variation in NCS has been analysed in three ways in the following sections: as a function of direction, as a function of radial distance, and within each of the three anatomical principal planes (horizontal, median and frontal).

As Mesh2HRTF calculates only one value at $\phi = 90°$ for computational efficiency, an azimuth of $\theta = 0°$ was used and the simulated data for that source position has been duplicated across all 360 azimuth positions to aid in visualisation and analysis. This gives a total number of 47,160 positions at each radial distance. This set of 660,240 source positions (47,160 × 14) is referred to as the set $\mathbf{SP}^{\mathbf{KEMAR}}$, which can be written using set notation as:

$$\mathbf{SP}^{\mathbf{KEMAR}} = \{v \mid v \in \mathbb{R}^3\} \tag{5.2}$$

in that $\mathbf{SP}^{\mathbf{KEMAR}}$ is a set with members v, such that each member v is a three-dimensional, real-valued column vector.

5.3.1 Variation as a Function of Direction

As expected, the two largest NCS values occurred at the reference directions: 43.2 dB at $\theta = 90°$, $\phi = 0°$, $r = 0.10$ m and 45.9 dB at $\theta = 270°$, $\phi = 0°$, $r = 0.10$ m. The global minimum is 0.2 dB, which occurs at $\theta = 183°$, $\phi = -16°$, $r = 0.15$ m. This gives a range of 45.6 dB across all source positions. To look at the variation as a function of direction, intensity surfaces of NCS for all domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ (therefore excluding $r = 0.10$ m) are plotted in Fig. 5.2. Larger versions are included in Appendix D, Fig. D.1.

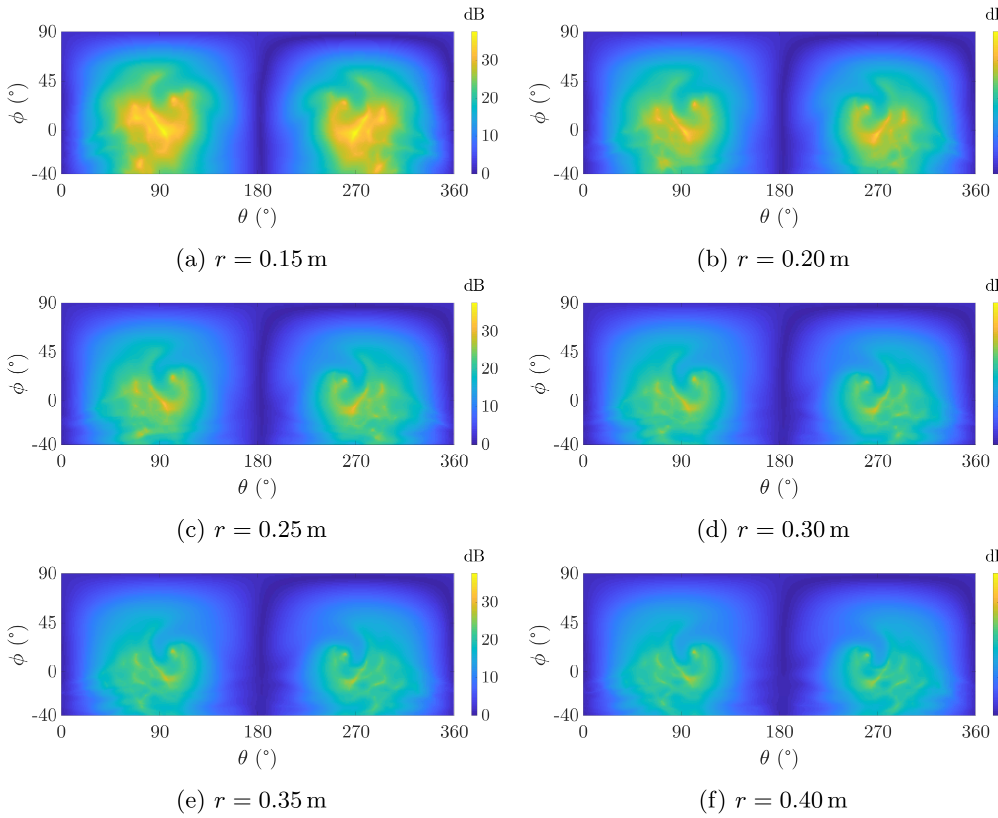
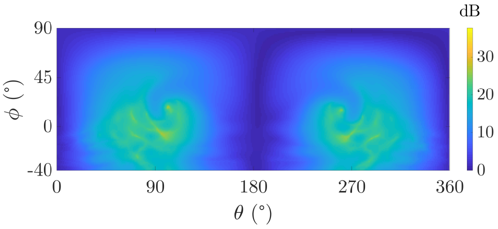
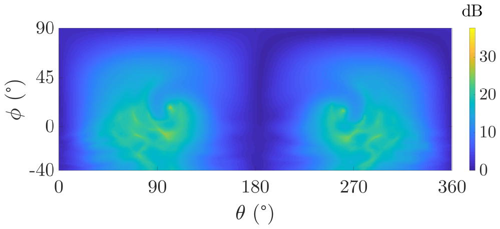


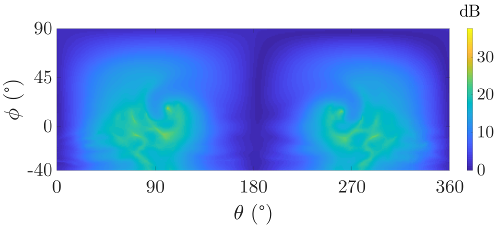
Figure 5.2: NCS plotted as intensity surfaces for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 0.15 m to 0.40 m. *Figure continued overleaf.*



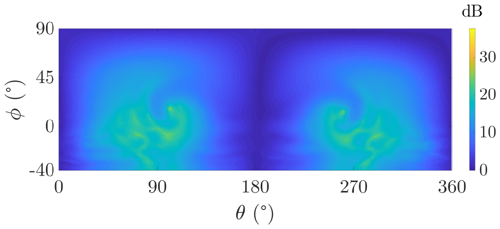
(g) $r = 0.45$ m



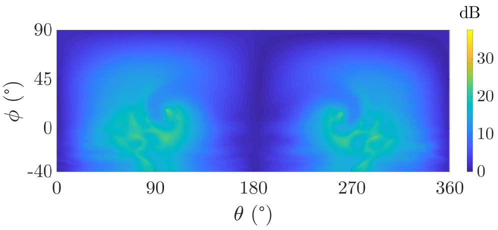
(h) $r = 0.50$ m



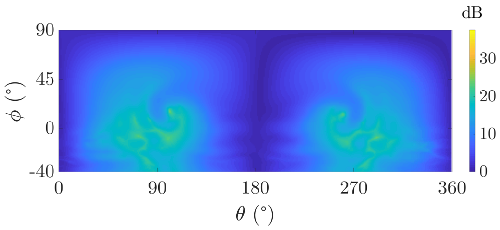
(i) $r = 0.60$ m



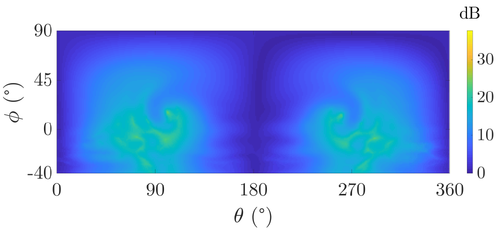
(j) $r = 0.70$ m



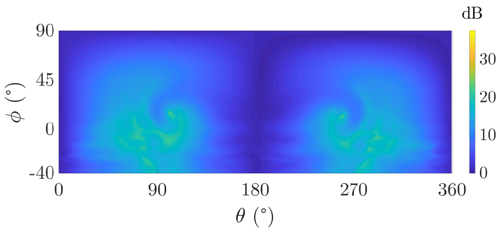
(k) $r = 0.80$ m



(l) $r = 0.90$ m



(m) $r = 1.00$ m



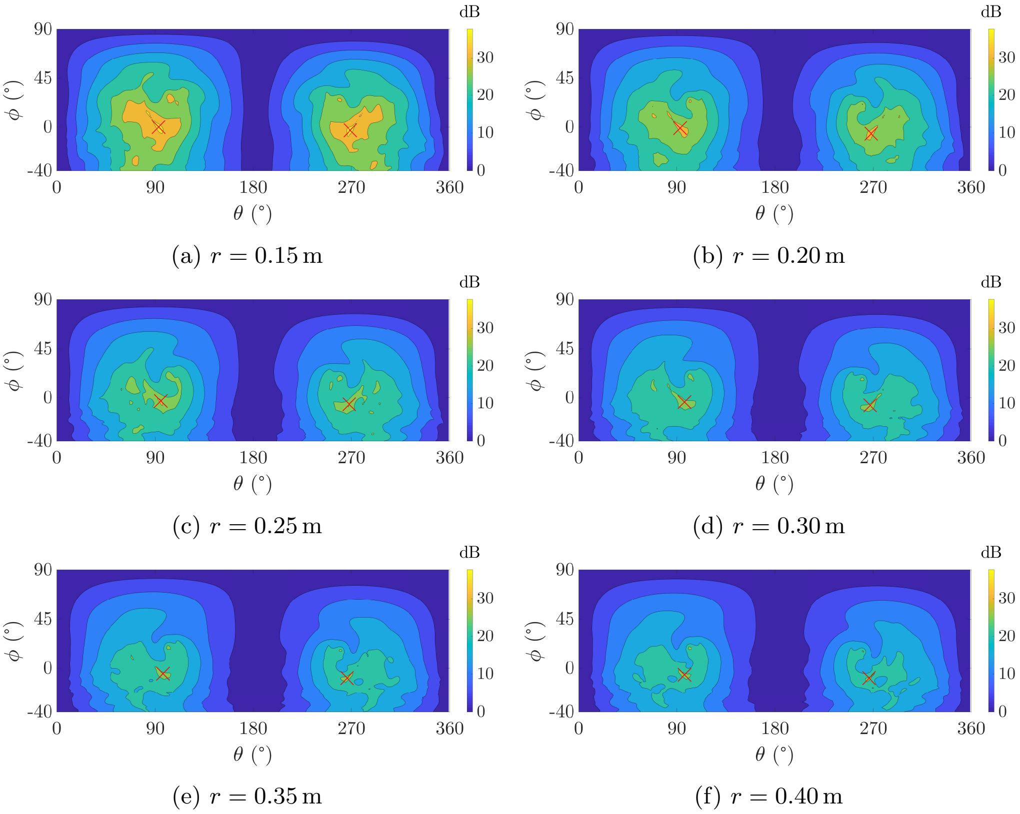
(n) $r = 1.50$ m

Figure 5.2: NCS plotted as intensity surfaces for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 0.45 m to 1.50 m. The variation as a function of azimuth and elevation is apparent in each surface.

Whilst variation as a function of radial distance is evident (discussed further in Section 5.3.2), general patterns with direction remain. Smaller NCS values occur when the source lies on or close to the median plane. In this region, the transfer functions from the source to each eardrum are very similar and are characterised by an absence of head shadowing. Below elevation angles of around $\phi = 50°$ and in the azimuth range of $\pm 40°$ of the interaural axis, larger values of NCS appear. The largest values occur in the vicinity of the interaural axis, where ipsilateral and contralateral differences are at their greatest. This is in agreement with Lundkvist *et al.* [19], who found that loudspeakers positioned behind or above the head had larger values of NCS than loudspeakers positioned in front.

The position of the single maximum value within a dome is not particularly informative. A more useful position might be the centroid of each region for which NCS lies above a specified threshold based on the values of the data, as generated by a contour plot. Fig. 5.3 shows contour plots of NCS across each dome within $\mathbf{SP^{KEMAR}}$, plotted in increments of 5 dB. Larger versions are included in Appendix D, Fig. D.2. The centroid of the region of highest threshold value in each plot, calculated using the MATLAB 'regionprops' function [282], is indicated with a red cross and listed in Table 5.3. Where two or more separate regions exist for that threshold value, the centroid for the largest region has been listed. For radial distances greater than $r = 0.50$ m, the variation pattern divides into two distinct regions of approximately-equal size. Thus, two centroid values are included at these radial distances.

The position of the centroid is stable within these two ranges of radial distance. For distances up to and including $r = 0.50$ m, the centroid is slightly behind and below the interaural axis. Beyond $r = 0.60$ m, there is one region below and slightly behind the interaural axis, with a second region below and 20° in front of the interaural axis.



(a) $r = 0.15$ m

(b) $r = 0.20$ m

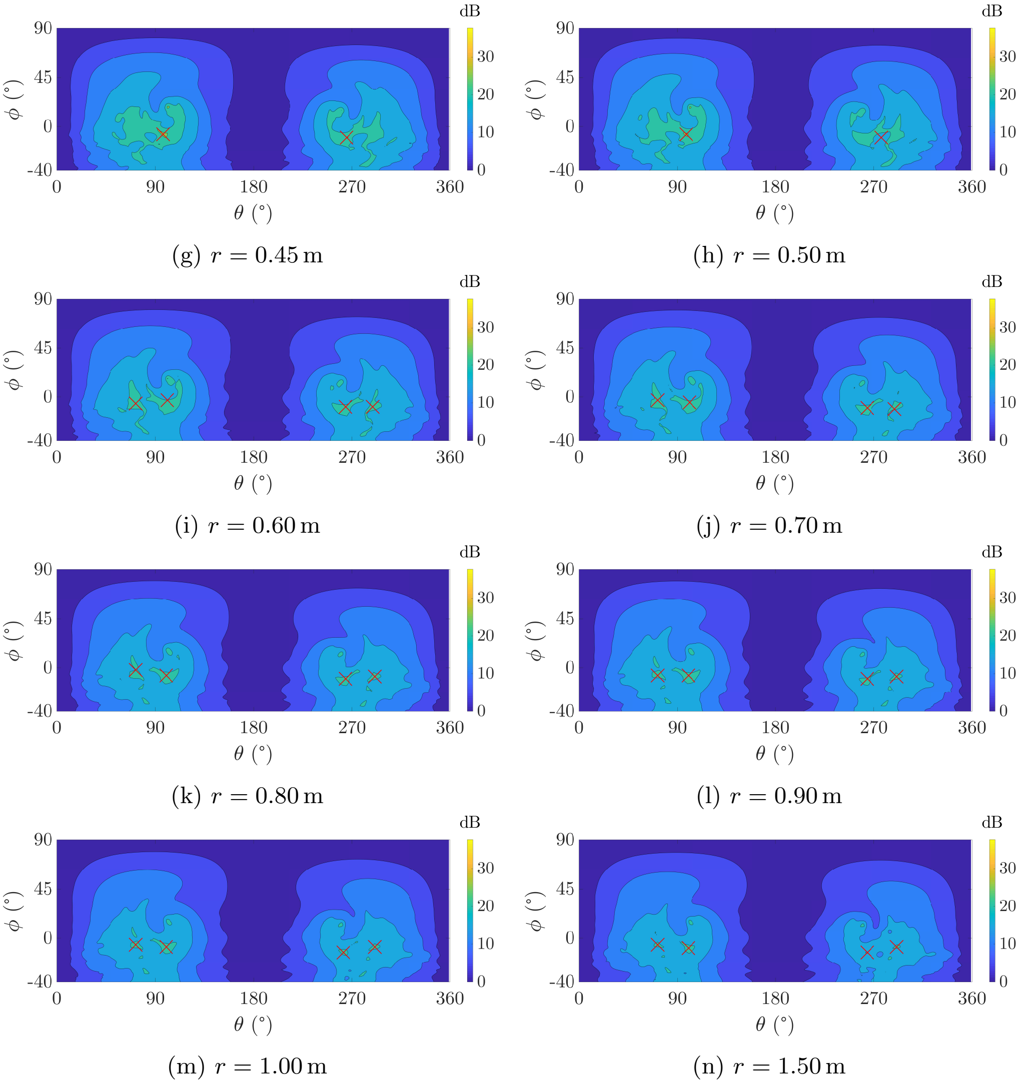
(c) $r = 0.25$ m

(d) $r = 0.30$ m

(e) $r = 0.35$ m

(f) $r = 0.40$ m

Figure 5.3: NCS plotted as contour plots (increments of 5 dB) for domes within $\mathbf{SP^{KEMAR}}$ for radial distances 0.15 m to 0.40 m. The centroid of each region with the largest threshold value is indicated with a red cross. *Figure continued overleaf.*



(g) $r = 0.45\,\mathrm{m}$

(h) $r = 0.50\,\mathrm{m}$

(i) $r = 0.60\,\mathrm{m}$

(j) $r = 0.70\,\mathrm{m}$

(k) $r = 0.80\,\mathrm{m}$

(l) $r = 0.90\,\mathrm{m}$

(m) $r = 1.00\,\mathrm{m}$

(n) $r = 1.50\,\mathrm{m}$

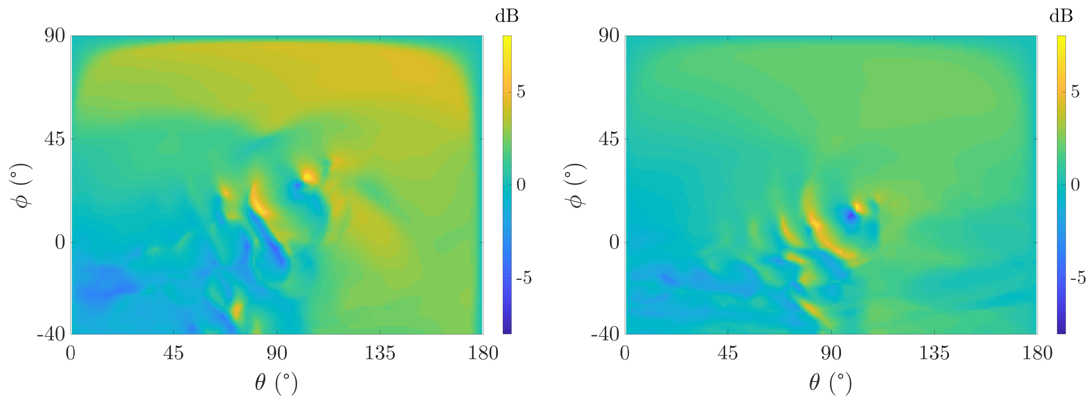
Figure 5.3: NCS plotted as contour plots (increments of 5 dB) for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 0.45 m to 1.50 m. The centroid of each region with the largest threshold value is indicated with a red cross. Where two separate regions exist, two centroids are indicated.

r (m)	Left Hemisphere		Right Hemisphere	
	Centroid (θ, ϕ)	Threshold (dB)	Centroid (θ, ϕ)	Threshold (dB)
0.15	93°, 0°	35	269°, −3°	35
0.20	93°, −1°	35	268°, −6°	30
0.25	95°, −3°	30	268°, −6°	30
0.30	97°, −4°	25	267°, −7°	25
0.35	97°, −5°	25	266°, −9°	25
0.40	97°, −6°	25	266°, −9°	25
0.45	97°, −7°	25	265°, −10°	25
0.50	98°, −7°	25	277°, −10°	20
0.60	72°, −6° 101°, −3°	20	264°, −9° 289°, −9°	20
0.70	72°, −3° 101°, −5°	20	264°, −10° 289°, −11°	20
0.80	72°, −2° 100°, −7°	20	264°, −10° 291°, −8°	20
0.90	72°, −7° 100°, −7°	20	264°, −10° 291°, −8°	20
1.00	72°, −6° 100°, −8°	20	262°, −13° 291°, −8°	20
1.50	72°, −6° 100°, −9°	20	264°, −13° 291°, −8°	20

Table 5.3: Positions of the centroids of each region of maximum threshold at each radial distance. The positions are consistent across two ranges of radial distance: $r \leq 0.50\,\mathrm{m}$ and $r \geq 0.60\,\mathrm{m}$, as indicated by the dotted line.

There is also a difference between the left and right hemispheres at each radial distance, which can be most likely attributed to asymmetry in the morphology of KEMAR. Fig. 5.4 shows the signed difference between the left and right hemispheres for the cases of $r = 0.15\,\mathrm{m}$ and $r = 1.50\,\mathrm{m}$, where a positive value indicates that the left hemisphere value is larger and a negative value indicates that the right hemisphere value is larger.

The global maximum difference is 7.9 dB, and the global minimum difference is −6.7 dB. The NCS in regions above $\phi = 45°$ and behind $\theta = 120°$ is generally larger in the left hemisphere (a range of approximately 2 to 4 dB for $r = 0.15\,\mathrm{m}$, and a range of approximately 1 to 2 dB for $r = 1.50\,\mathrm{m}$), whilst the region in front and below (the bottom left quadrant of the figure) is generally larger in the right hemisphere (a range of approximately −2 to 1 dB for $r = 0.15\,\mathrm{m}$, approximately −1.5 to 0 dB for $r = 1.50\,\mathrm{m}$). The signed difference between hemispheres generally reduces with an increase in radial distance, with the exception of bright and dark regions in the vicinity of the interaural axis. These regions maintain their values over radial distance (approximately 8 dB and −4 dB, respectively) and are mostly likely due to the slight difference in the angular positions of the highest NCS values in each hemisphere, as a result of the anthropometric asymmetries in KEMAR.



(a) $r = 0.15\,\mathrm{m}$

(b) $r = 1.50\,\mathrm{m}$

Figure 5.4: Difference between NCS in the left and right hemispheres for the cases of: a) $r = 0.15\,\mathrm{m}$ and b) $r = 1.50\,\mathrm{m}$. Larger values occur in the upper and rear regions, smaller values occur in the frontal lower region, and the values generally decrease with increasing radial distance.

5.3.2 Variation as a Function of Radial Distance

Fig. 5.2, Fig. 5.3 and Table 5.3 show that the maximum NCS within each dome decreases with an increase in radial distance. This is shown in Fig. 5.5, which plots the distribution properties of NCS at each radial distance. For ease of comparison, the mean, median and maximum values for each radial distance are shown in Fig. 5.6.

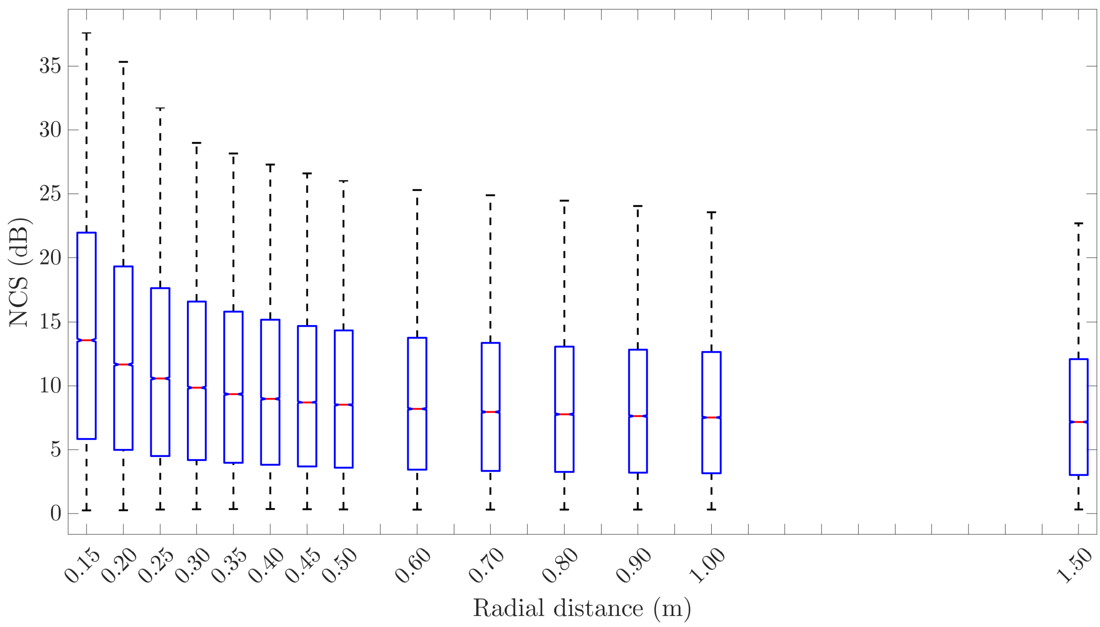


Figure 5.5: Distribution properties of NCS at each radial distance. The decrease in maximum NCS with increasing radial distance are apparent.

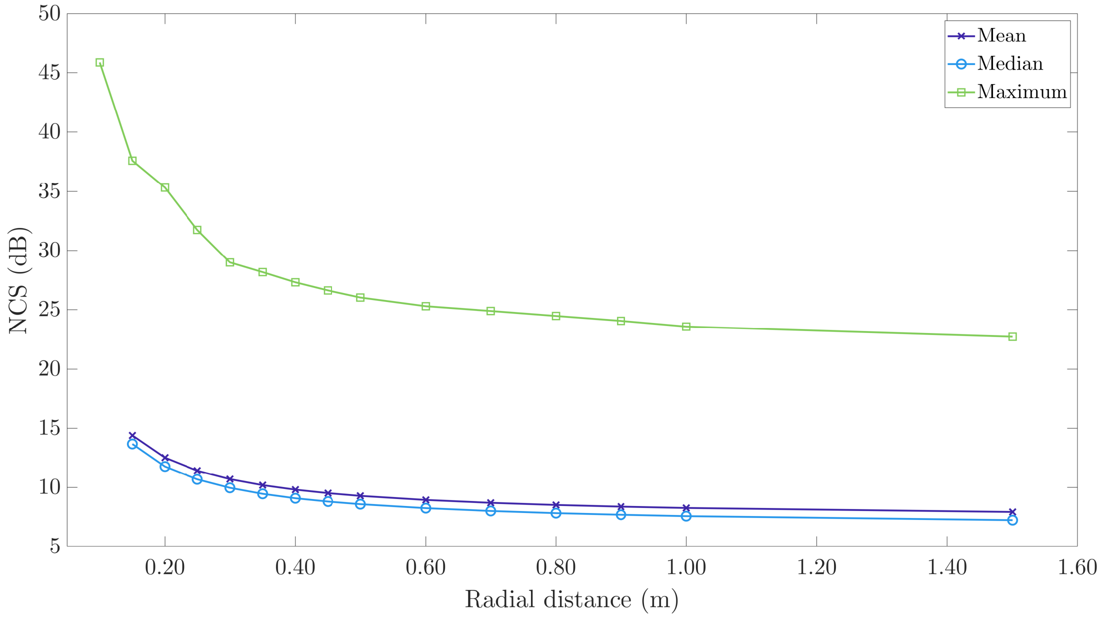


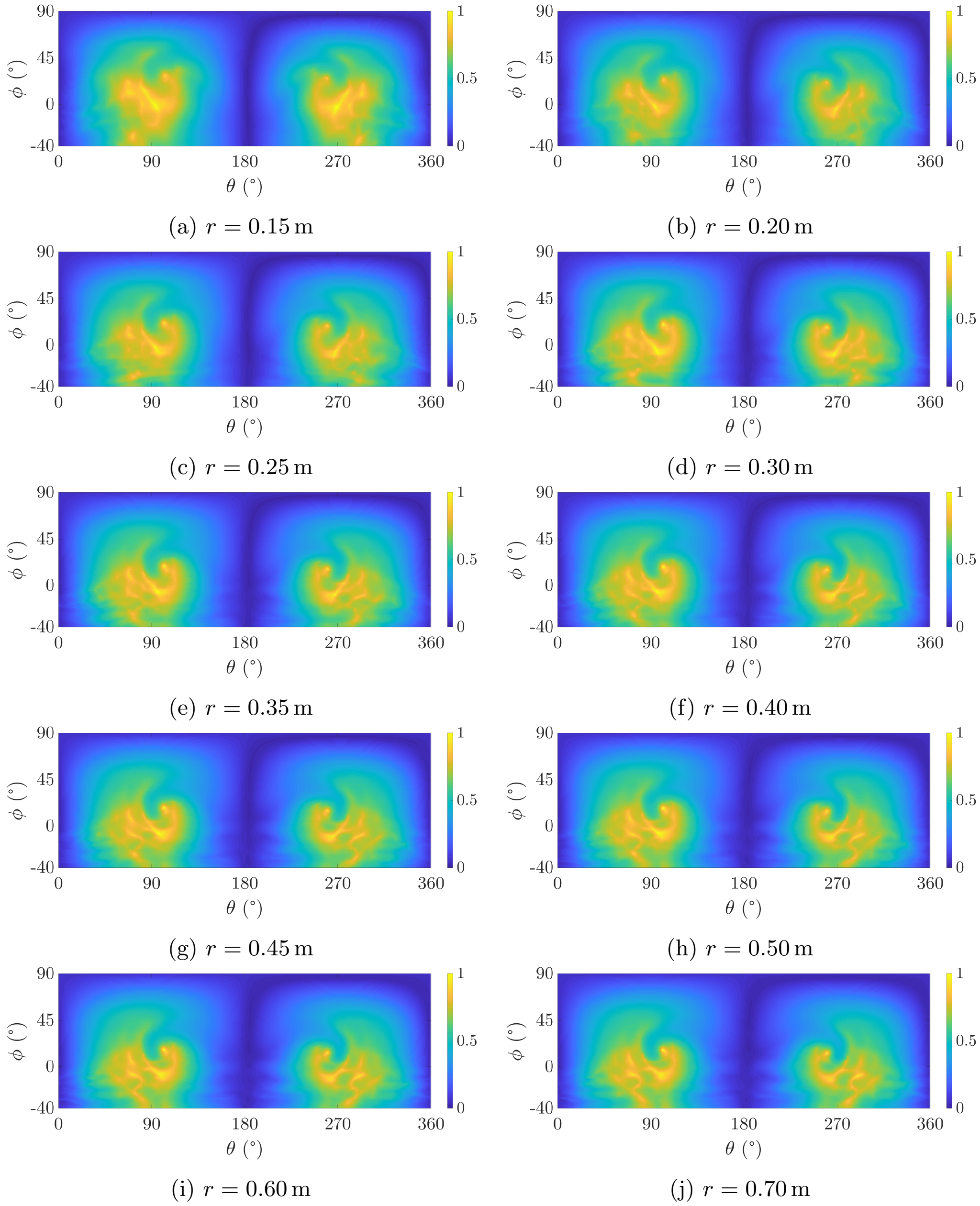
Figure 5.6: Maximum, mean and median NCS as a function of radial distance. All three measures plateau after approximately $r = 0.50$ m.

All three measures rise at an increasing rate as the source approaches the head, with the largest values calculated for the smallest radial distance: 44.5 dB, 44.5 dB, 45.9 dB for mean, median and maximum, respectively. The mean and median for $r = 0.10$ m are meaningless as only two positions have been calculated at this radial distance and they have, therefore, been omitted from the plot. The mean and median values for the smallest radial distance at which many positions have been calculated ($r = 0.15$ m) are 14.3 dB and 13.5 dB, respectively.

All three measures plateau beyond approximately 0.50 m. Between $r = 0.15$ m and $r = 0.50$ m, there are changes of -11.6 dB, -5.0 dB and -5.0 dB for maximum, mean and median values, respectively. An independent-samples t-test indicated that NCS values at the smaller radial distance were significantly larger than those at $r = 0.15$ m: $t(94, 318) = 98.02$, $p < 0.001$ with an effect size of $d = 0.64$. Between $r = 0.50$ m and $r = 1.50$ m there are much smaller changes of -3.3 dB, -1.3 dB and -1.3 dB for maximum, mean and median, respectively,however, an independent-samples t-test between the data at these radial distances also indicated a significant difference: $t(94, 318) = 35.60$, $p < 0.001$ with an effect size of $d = 0.23$. The minimum value at each radial distance is stable, ranging from 0.2 dB to 0.4 dB.

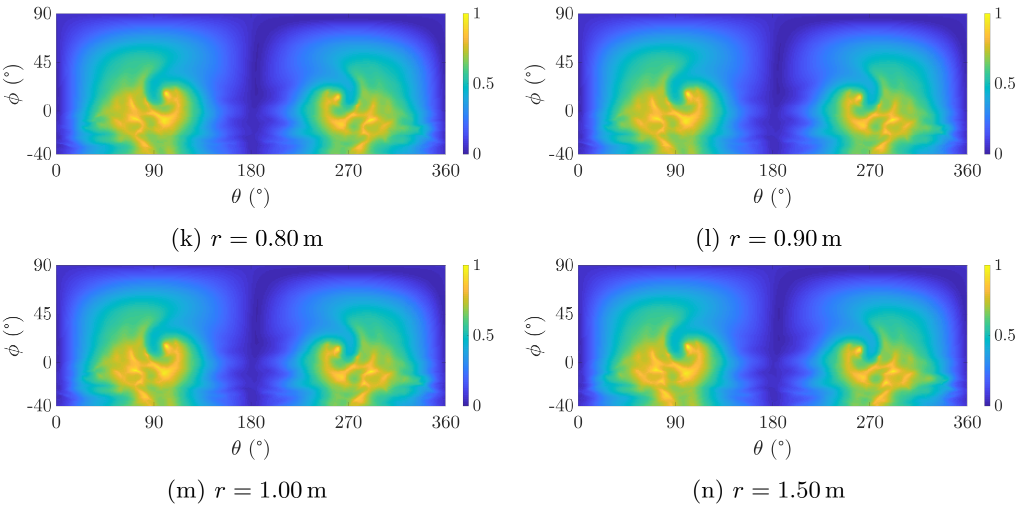
Variation in absolute NCS shows the global effect of radial distance on NCS. However, it is also useful to consider the variation in the normalised NCS value. Normalising each intensity surface in Fig. 5.2 with respect to the corresponding maximum value highlights whether radial distance has an impact other than a global decrease with distance.

Fig. 5.7 shows intensity surfaces of NCS normalised with respect to the corresponding maximum value in each surface. Larger versions are included in Appendix D, Fig. D.3. Whilst the patterns of variation in NCS are similar across all surfaces, there are some differences. The regions of larger NCS are more continuous at smaller radial distances and become more fragmented as radial distance increases. After normalising for the decrease in maximum value, there is still some reduction in NCS with increasing radial distance, as shown in Fig. 5.8. The dashed line indicates the median value of the largest radial distance (0.3 at $r = 1.50\,\text{m}$). This suggests that the relationship between the maximum value and all other values in a dome is not invariant with distance, that is, it is not only a simple global reduction.



(a) $r = 0.15\,\text{m}$ (b) $r = 0.20\,\text{m}$ (c) $r = 0.25\,\text{m}$ (d) $r = 0.30\,\text{m}$ (e) $r = 0.35\,\text{m}$ (f) $r = 0.40\,\text{m}$ (g) $r = 0.45\,\text{m}$ (h) $r = 0.50\,\text{m}$ (i) $r = 0.60\,\text{m}$ (j) $r = 0.70\,\text{m}$

Figure 5.7: NCS normalised with respect to the corresponding maximum value, plotted as intensity surfaces for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 0.15 m to 0.70 m. *Figure continued overleaf.*



(k) $r = 0.80\,\mathrm{m}$ (l) $r = 0.90\,\mathrm{m}$

(m) $r = 1.00\,\mathrm{m}$ (n) $r = 1.50\,\mathrm{m}$

Figure 5.7: NCS normalised with respect to the corresponding maximum value, plotted as intensity surfaces for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 0.80 m to 1.50 m. Variation exists between surfaces despite normalising for radial distance.

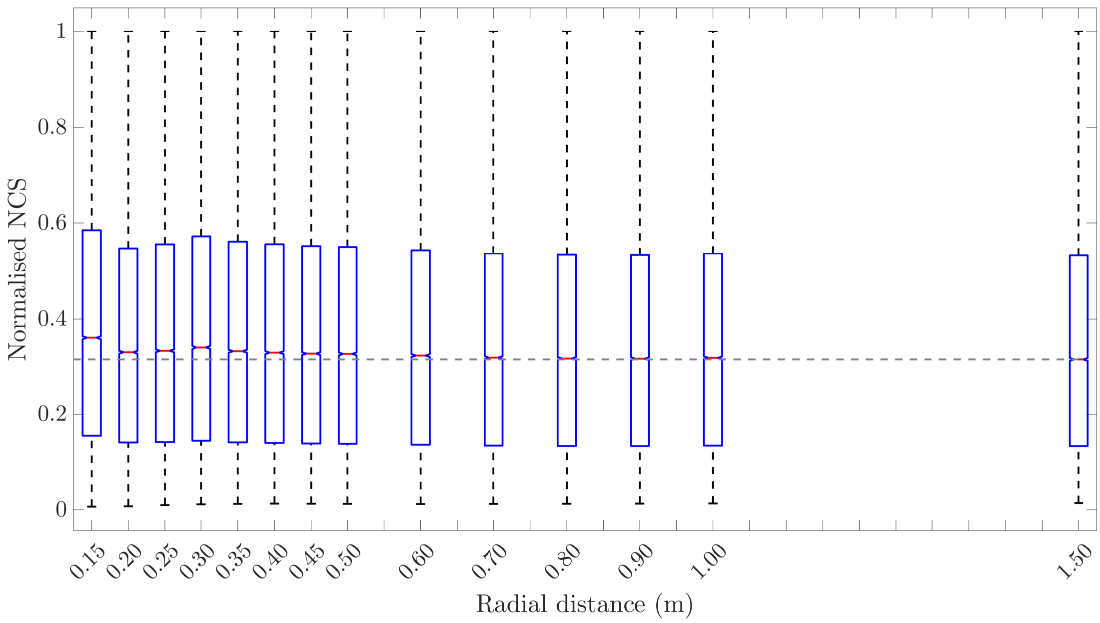


Figure 5.8: Distribution properties of normalised NCS at each radial distance, showing the non-invariant relationship between maximum value and all other values in a dome. The dashed line indicates the median value of the largest radial distance (0.3 at $r = 1.50\,\mathrm{m}$).

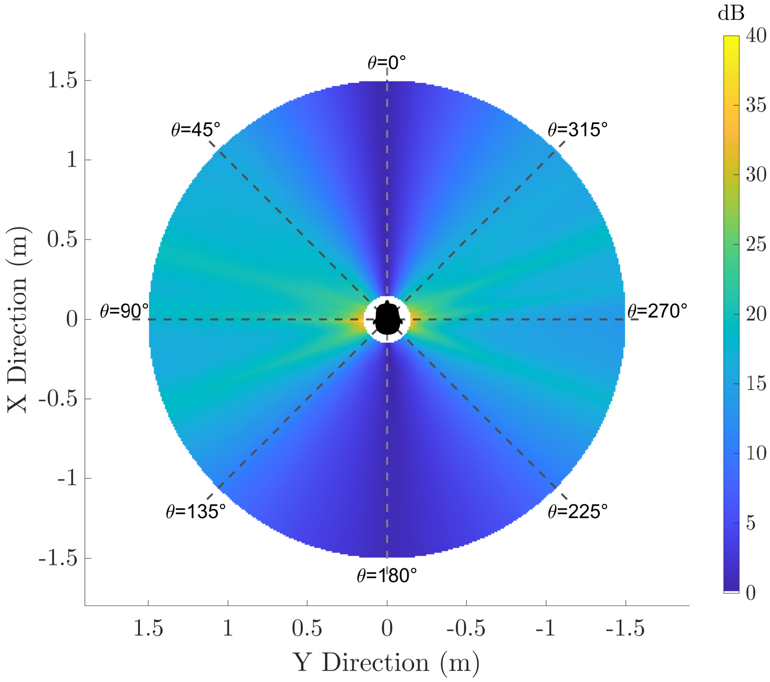
5.3.3 Variation within Principal Anatomical Planes

NCS has been shown to vary as a function of both radial distance and direction. However, the contributing factors to the more subtle variations in NCS are also worthy of investigation. These variations can be conveniently observed in the three principal planes: horizontal, median and frontal.

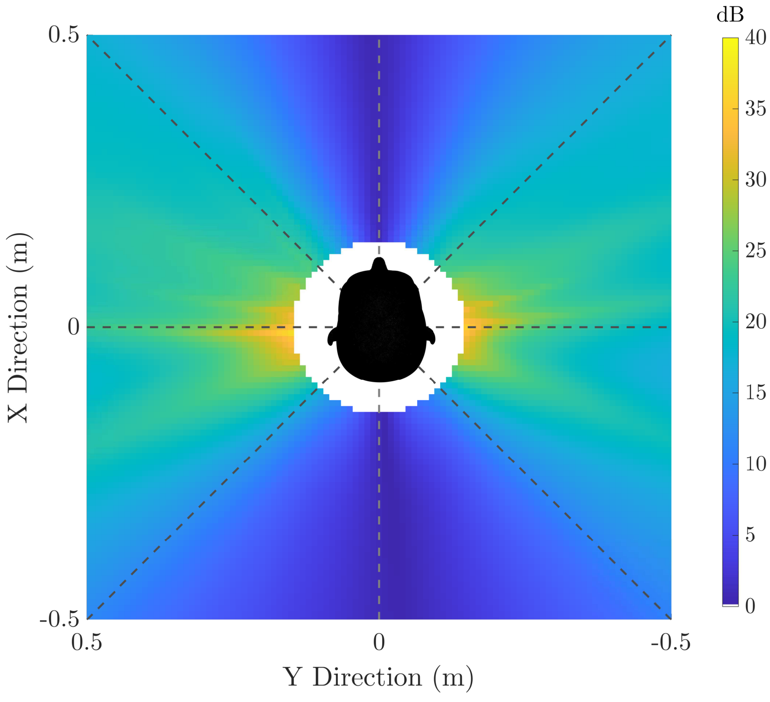
5.3.3.1 Horizontal Plane

Fig. 5.9 shows NCS in the horizontal plane at all radial distances within $\mathbf{SP}^{\mathbf{KEMAR}}$. To assist with visualisation, results are interpolated between scattered data points using the MATLAB 'griddata' function [283]. The region within ± 0.50 m of the origin is also presented. A pattern of local maxima and minima is apparent which remain stable across radial distance. The largest NCS values for radial distances up to $r = 0.50$ m occur close to the interaural axis: slightly behind for the left hemisphere and slightly in front for the right hemisphere. This left-right difference is most likely due to the difference in pinna *flare angle* (as defined by Algazi *et al.* [24]) between the two pinnae. The global largest NCS value is 37.1 dB at $\theta = 93°$. However, the azimuth angle of the largest local maximum appears to change with radial distance. For radial distances beyond $r = 0.50$ m, the local maximum located approximately 20° behind the interaural axis becomes larger than the local maximum close to the interaural axis.

An asymmetry between front and rear hemispheres is also particularly evident, with higher NCS values (more yellow and green than blue) appearing in the front hemisphere. This indicates that NCS for rear $\mathbf{SP}^{\mathbf{KEMAR}}$ positions ($90° \leq \theta \leq 270°$) decreases more as a function of azimuth than for corresponding frontal directions, with this effect being more pronounced at smaller radial distances. As an example, in the left hemisphere at $r = 0.15$ m, an azimuthal increase of 42° (towards the rear) is required to reach half the maximum value, whereas a numerically-larger azimuthal change of 60° is required to reach the same value in the frontal hemisphere. This is most likely due to the increase in acoustic shadowing caused by extreme proximity to the facial features, in combination with the slightly rearward placement of the outer ear (on KEMAR, the interaural axis is 0.13 m from the tip of the nose, as opposed to 0.08 m from the nape of the neck). This has the effect of increasing shadowing at the contralateral ear for close frontal sources and therefore generates larger NCS.



(a) Over the full range of radial distances.



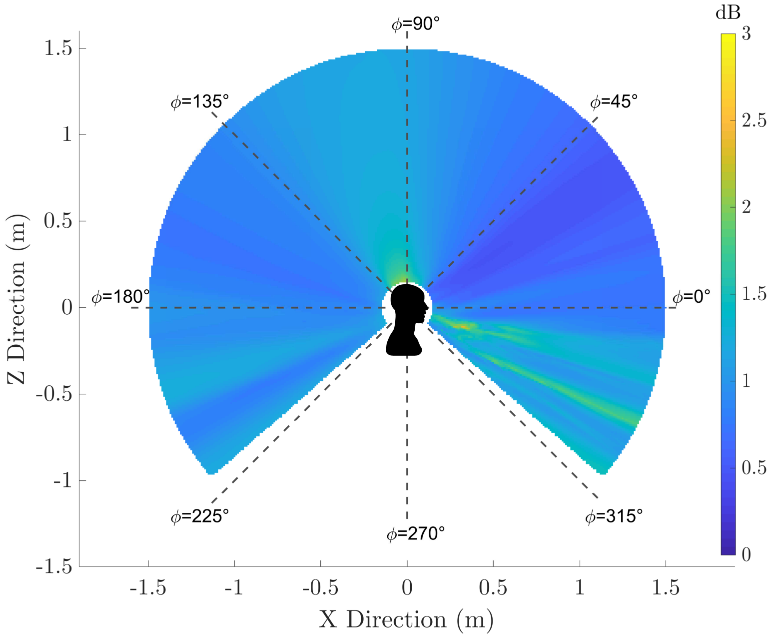
(b) A close up of the region within ± 0.50 m of the origin.

Figure 5.9: Interpolated NCS in the horizontal plane. θ in increments of 45° are also shown. The largest NCS values occur in the vicinity of the interaural axis. Local maxima and minima are present which remain stable over radial distance.

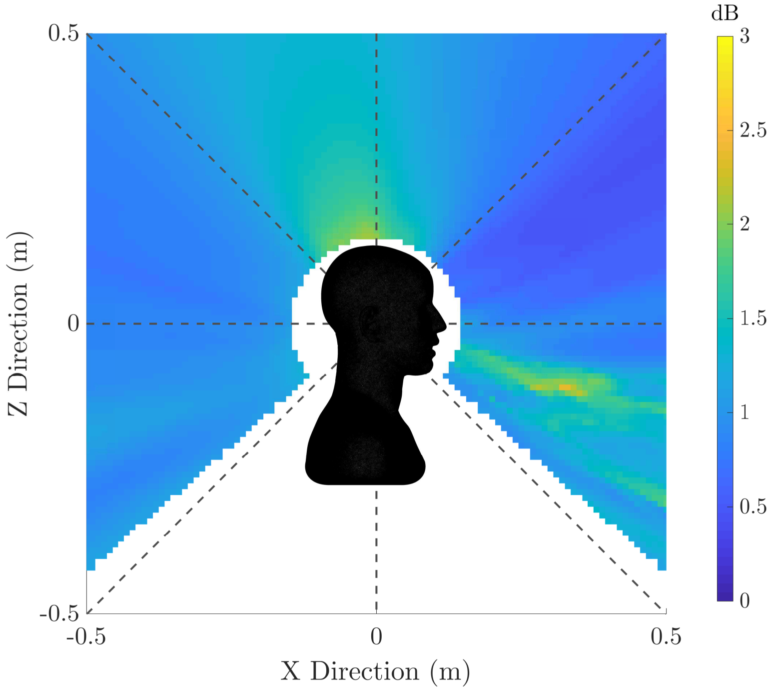
5.3.3.2 Median Plane

Fig. 5.10 shows NCS in the median plane for all radial distances within $\mathbf{SP}^{\mathbf{KEMAR}}$. It has been produced using the same interpolation method applied in the horizontal plane analysis of Section 5.3.3.1. For ease of discussion, elevation angles here will be discussed starting at zero in the frontal direction and increasing through 90° directly above, 180° at the rear and 270° directly below. As discussed in Section 5.3.1, generally low values of NCS exist in the median plane as a result of the similarity between left-right transfer functions in this region. However, some variations are still apparent. Three regions of relatively larger values exist ($85° \leq \phi \leq 110°$, $190° \leq \phi \leq 210°$ and $320° \leq \phi \leq 359°$), with the maximum value of 2.5 dB occurring at $\phi = -19°$ at $r = 0.35$ m.

In the region between $\phi = 320°$ (40° below the horizontal plane) and $\phi = 0°$, the pattern is unstable for radial distances under $r = 0.50$ m. NCS ranges from 0.7 dB to 2.5 dB, and there are a high number of local maxima and minima. This is perhaps due to interaction with the facial features in this region, mainly the lips and chin, in combination with the left-right asymmetry inherent within KEMAR. Above the horizontal plane, the maximum value is 2.1 dB, which occurs at $\phi = 96°$. This slight increase in NCS just above the head at small radial distances may be attributable to the influence of the curvature of the top of the pinnae, which might be expected to become negligible at larger radial distances.



(a) Over the full range of radial distances.



(b) A close up of the region within ±0.50 m of the origin.

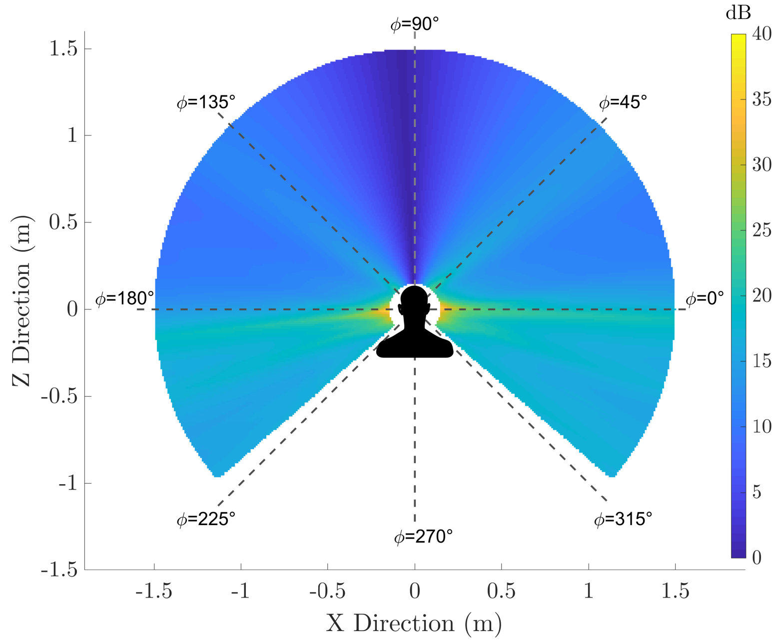
Figure 5.10: Interpolated NCS in the median plane. ϕ in increments of 45° are also shown. Local maxima and minima are present which do not remain stable over radial distances under 0.50 m.

5.3.3.3 Frontal Plane

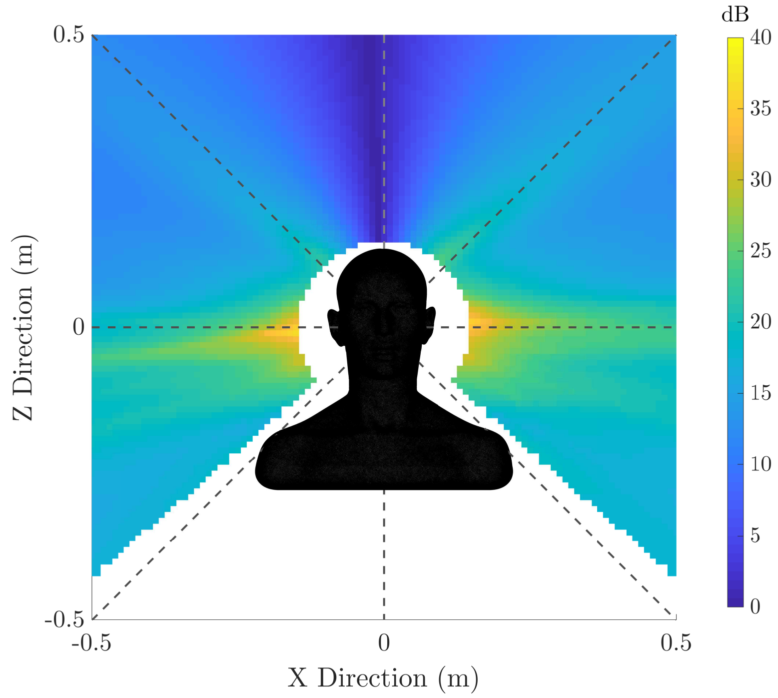
Fig. 5.11 shows NCS in the frontal plane at all radial distances within $\mathbf{SP^{KEMAR}}$. Again, the interpolation processing applied in Section 5.3.3.1 has been used here. For ease of discussion, elevation angles here will be discussed starting at zero at the right and increasing through 90° directly above, 180° at the left and 270° directly below. The two halves of the figure are roughly symmetrical, as would be expected. The largest NCS values in each of the left and right hemisphere are 37.6 dB and 36.5 dB, respectively, and are generated in the vicinity of the interaural axis. The slight difference in angle of the maximum values in each hemisphere may be due to the orientation of the pinna as, in this plane, the right pinna appears to have a greater *protrusion* (as defined by Nishino *et al.* [135]).

A similar pattern of variation with elevation is evident for all radial distances, with local maxima and minima occurring at approximately the same elevation angles across radial distances. A second local maximum in each hemisphere can be seen at approximately $35° \leq \phi \leq 55°$ and $125° \leq \phi \leq 145°$, respectively. These regions may be related to the curvature of the top region of the pinnae increasing the shadowing slightly at the ipsilateral ear and therefore increasing the NCS.

It is evident from observing NCS in the three principal anatomical planes that smaller-scale morphological features (including the rear offset of the ears, the pinnae folds, angle of the pinnae relative to the head, and the facial features) contribute to the generation of NCS at each source position. However, a more detailed analysis of the specific contribution of each feature to the NCS at any given $\mathbf{SP^{KEMAR}}$ position is outside the scope of this thesis, although a brief investigation is described in Chapter 8.



(a) Over the full range of radial distances.



(b) A close up of the region within $\pm 0.50\,\mathrm{m}$ of the origin.

Figure 5.11: Interpolated NCS in the frontal plane. ϕ in increments of 45° are also shown. The largest NCS values occur in the vicinity of the interaural axis. Local maxima and minima are present which remain stable over radial distance.

5.4 Source Positions with Sufficient NCS

NCS has been shown to vary with radial distance, azimuth angle and elevation angle. At smaller radial distances, there is greater variation in NCS and relatively large values can be achieved. At larger radial distances, NCS values are more stable and generally smaller. Of particular interest for this thesis are those $\mathbf{SP^{KEMAR}}$ positions with NCS exceeding the reported 20 dB threshold required for binaural robustness [17, 228]. This subset is referred to as $\mathbf{SP^{KEMAR}_{20}}$, and can be defined using set notation as follows.

As the set $\mathbf{SP^{KEMAR}}$ has been previously defined in Equation 5.2 as:

$$\mathbf{SP^{KEMAR}} = \{v \mid v \in \mathbb{R}^3\} \tag{5.3}$$

a function $f : \mathbf{SP^{KEMAR}} \to \mathbb{R}$ can be defined such that, if $sp \in \mathbf{SP^{KEMAR}}$:

$$f(sp) = NCS \tag{5.4}$$

where the function operating on each member sp in $\mathbf{SP^{KEMAR}}$ is analogous to the calculation of the NCS value at that source position. The subset $\mathbf{SP^{KEMAR}_{20}}$ can then be defined as:

$$\mathbf{SP^{KEMAR}_{20}} = \left\{sp \in \mathbf{SP^{KEMAR}} \mid f(sp) < 20\right\} \tag{5.5}$$

The percentage of $\mathbf{SP^{KEMAR}_{20}}$ within $\mathbf{SP^{KEMAR}}$ at each radial distance is plotted in Fig. 5.12 and listed in Table 5.4. Many of these positions are found closer to the head: 30.9 % of source positions (14,576) at $r = 0.15$ m exist in $\mathbf{SP^{KEMAR}_{20}}$, while only 0.3 % (158) do at $r = 1.50$ m. This echoes the plateauing observed in Fig. 5.6; a percentage change of 26.8 % from 0.15 m to 0.50 m, but only 3.8 % from 0.50 m to 1.50 m.

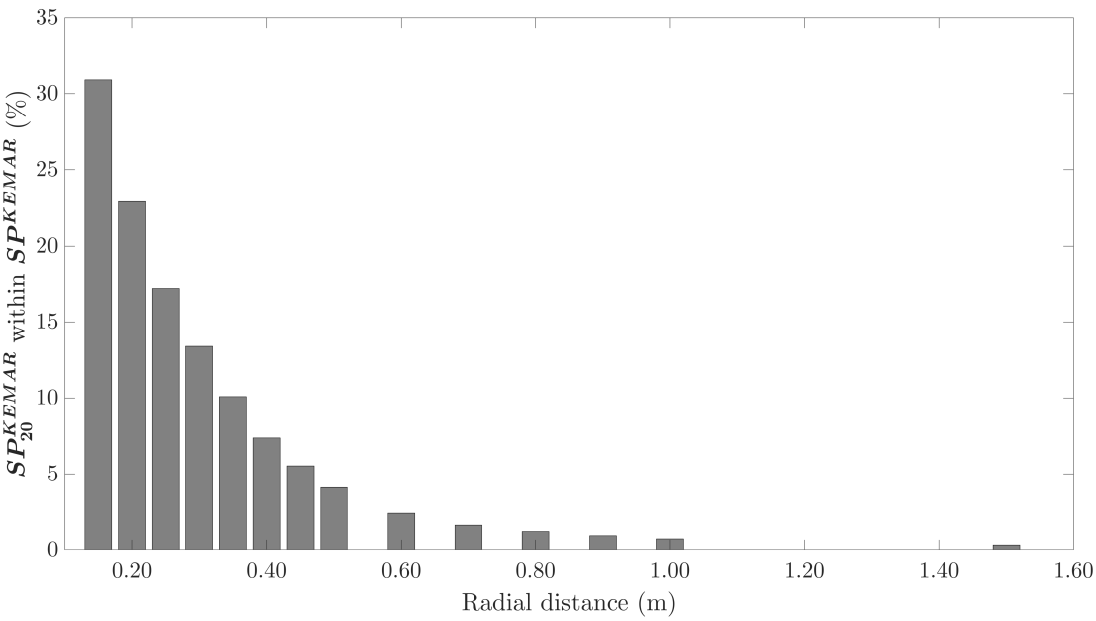
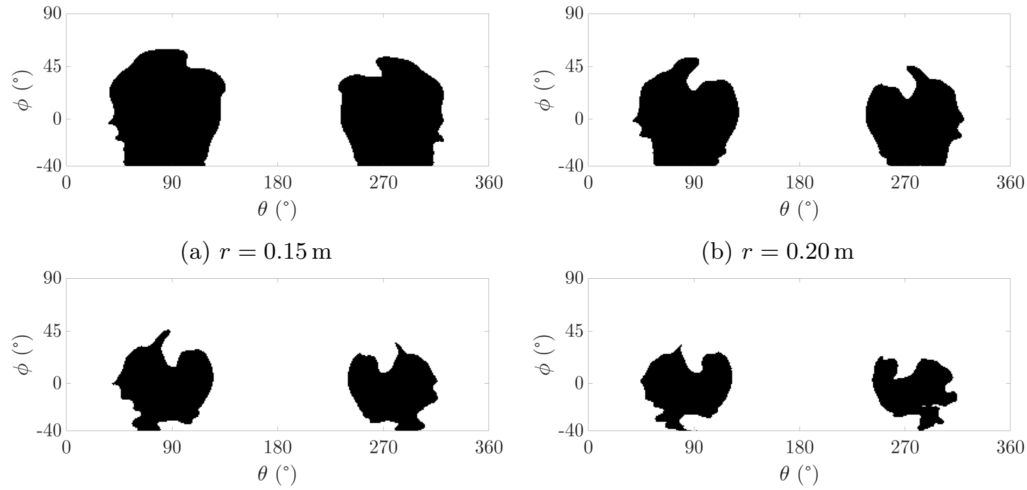


Figure 5.12: Percentage of $\mathbf{SP^{KEMAR}_{20}}$ positions which exist in $\mathbf{SP^{KEMAR}}$ at each radial distance. Many more positions exist in the region closest to the head.

r (m)	$\mathbf{SP_{20}^{KEMAR}}$ Number	%
0.15	14,576	30.9
0.20	10,808	22.9
0.25	8,115	17.2
0.30	6,326	13.4
0.35	4,759	10.1
0.40	3,481	7.4
0.45	2,611	5.5
0.50	1,942	4.1
0.60	1,146	2.4
0.70	774	1.6
0.80	575	1.2
0.90	445	0.9
1.00	348	0.7
1.50	158	0.3

Table 5.4: Number and percentage of $\mathbf{SP_{20}^{KEMAR}}$ positions within $\mathbf{SP^{KEMAR}}$ at each radial distance.

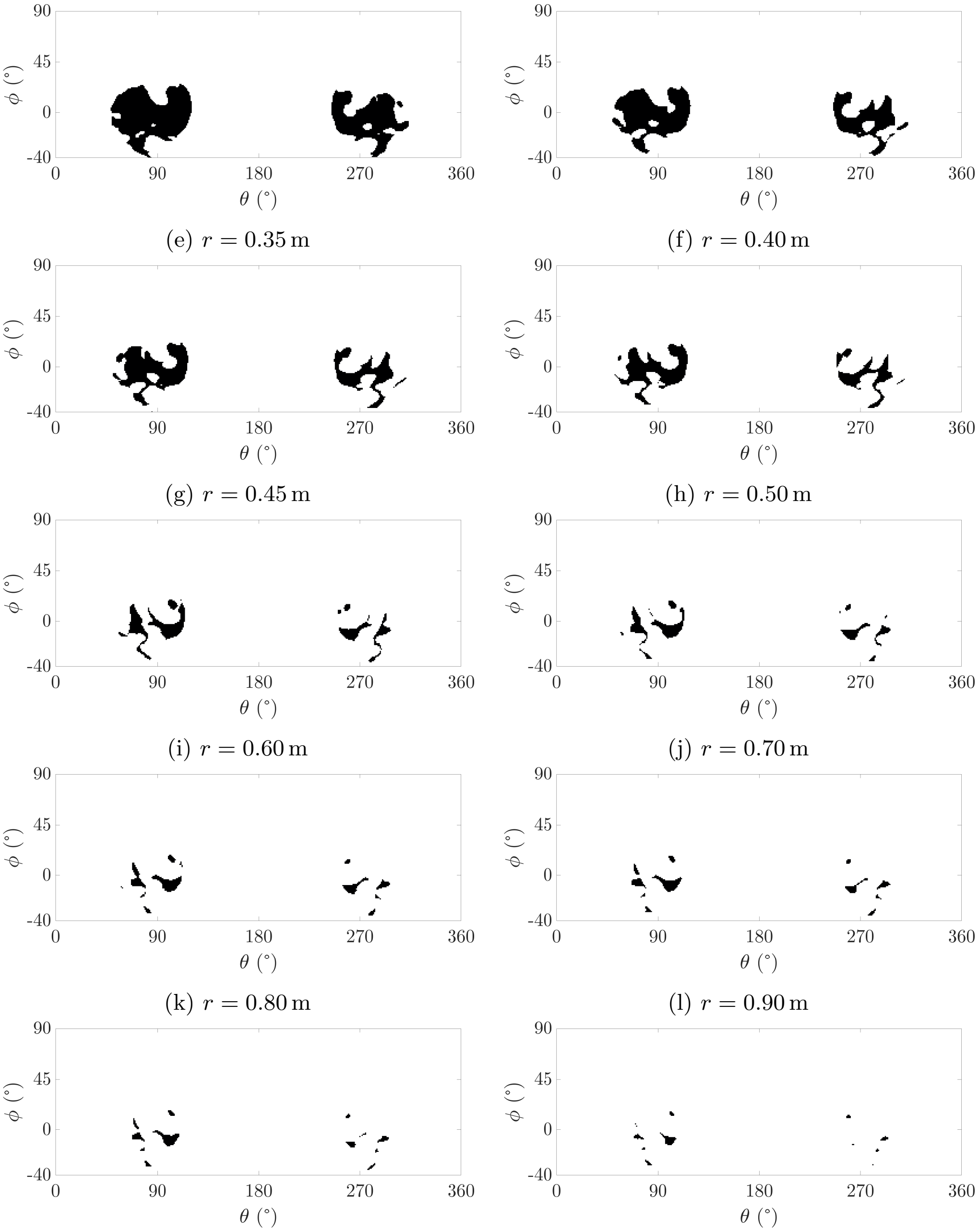
Fig. 5.13 shows the regions of $\mathbf{SP_{20}^{KEMAR}}$ positions (indicated in black). Larger versions are included in Appendix D, Fig. D.4. For $r = 0.15\,\mathrm{m}$, $\mathbf{SP_{20}^{KEMAR}}$ positions lie in the azimuth regions $\theta = 90 \pm 40°$ and $\theta = 270 \pm 40°$ and in elevation below $\phi = 60°$. Differences between front and rear hemispheres are apparent. These are to be expected due to the large morphological differences between the front and back of the head. There are also more subtle differences between the left and right hemispheres due to the small morphological asymmetries of KEMAR (discussed in Section 5.3.3). As suggested by previous figures, the dark regions reduce in size with an increase in radial distance.



(a) $r = 0.15\,\mathrm{m}$ (b) $r = 0.20\,\mathrm{m}$

(c) $r = 0.25\,\mathrm{m}$ (d) $r = 0.30\,\mathrm{m}$

Figure 5.13: Regions of $\mathbf{SP_{20}^{KEMAR}}$ positions (indicated in black) for domes within $\mathbf{SP^{KEMAR}}$ for radial distances 0.15 m to 0.30 m. *Figure continued overleaf.*



(e) $r = 0.35\,\mathrm{m}$ (f) $r = 0.40\,\mathrm{m}$

(g) $r = 0.45\,\mathrm{m}$ (h) $r = 0.50\,\mathrm{m}$

(i) $r = 0.60\,\mathrm{m}$ (j) $r = 0.70\,\mathrm{m}$

(k) $r = 0.80\,\mathrm{m}$ (l) $r = 0.90\,\mathrm{m}$

(m) $r = 1.00\,\mathrm{m}$ (n) $r = 1.50\,\mathrm{m}$

Figure 5.13: Regions of $\mathbf{SP^{KEMAR}_{20}}$ positions (indicated in black) for domes within $\mathbf{SP^{KEMAR}}$ for radial distances 0.35 m to 1.50 m. $\mathbf{SP^{KEMAR}_{20}}$ positions exist at all radial distances.

In addition to decreasing in size with increasing distance, the regions also decrease in homogeneity. A continuous region is present in each left-right hemisphere for radial distances up to $r = 0.30\,\mathrm{m}$, then several ‘holes’ develop up to $r = 0.50\,\mathrm{m}$. Beyond that, the regions become increasingly fragmented either side of the interaural axis. The number of separate regions at each radial distance is listed in Table 5.5. At $r = 0.60\,\mathrm{m}$ and beyond, the number of fragments stabilises. The fragments then reduce in size around their respective central points. Whilst $\mathbf{SP_{20}^{KEMAR}}$ positions do exist at larger radial distances, it is suspected that reliably identifying a source position within such a sparse set would be impractical in a real-world system.

r (m)	Regions
0.15	2
0.20	2
0.25	2
0.30	3
0.35	3
0.40	4
0.45	5
0.50	6
0.60	11
0.70	13
0.80	12
0.90	11
1.00	13
1.50	12

Table 5.5: Number of separate regions of $\mathbf{SP_{20}^{KEMAR}}$ positions at each radial distance. A sudden increase in the number of fragments occurs at $r = 0.60\,\text{m}$.

The existence of source positions which exhibit channel separation that is large enough for successful binaural reproduction indicates that, when used in combination, left-right pairs of loudspeakers at these positions may be capable of reproducing binaural signals without the need for CTC.

5.5 The Impact of Calculation Frequency Range on NCS

NCS has been calculated over the frequency region 200 Hz to 8 kHz to enable comparison with existing literature. However, since simulated HRTF data has been obtained up to 16 kHz, it is of interest to look at the consequences of extending the upper frequency bound of the NCS calculation. It is also of interest to determine source positions with sufficient NCS for binaural reproduction with NCS calculated using a number of frequency bands.

5.5.1 Extending the Upper Frequency Bound

Contralateral HRTFs exhibit a general decrease in magnitude with increasing frequency, resulting in larger values of NCS calculated for the same source position when using a larger frequency range [21, 229]. Fig. 5.14 shows the HRTF pair for a source at $\theta = 90°$, $\phi = 0°$, $r = 0.15\,\text{m}$, with the 8 kHz limit for NCS calculation indicated in purple. The contralateral (in this case, right) HRTF has a general downward trend with increasing frequency. For a visual aid, the linear trend line (the equation for which is $y = -1.4681x + 8.5730$) is indicated in green. For this source position there is a 3.2 dB difference between the NCS values calculated using the frequency range 200 Hz to 8 kHz ($NCS^{8k} = 34.3$ dB), compared with the range 200 Hz to 16 kHz ($NCS^{16k} = 37.5$ dB).

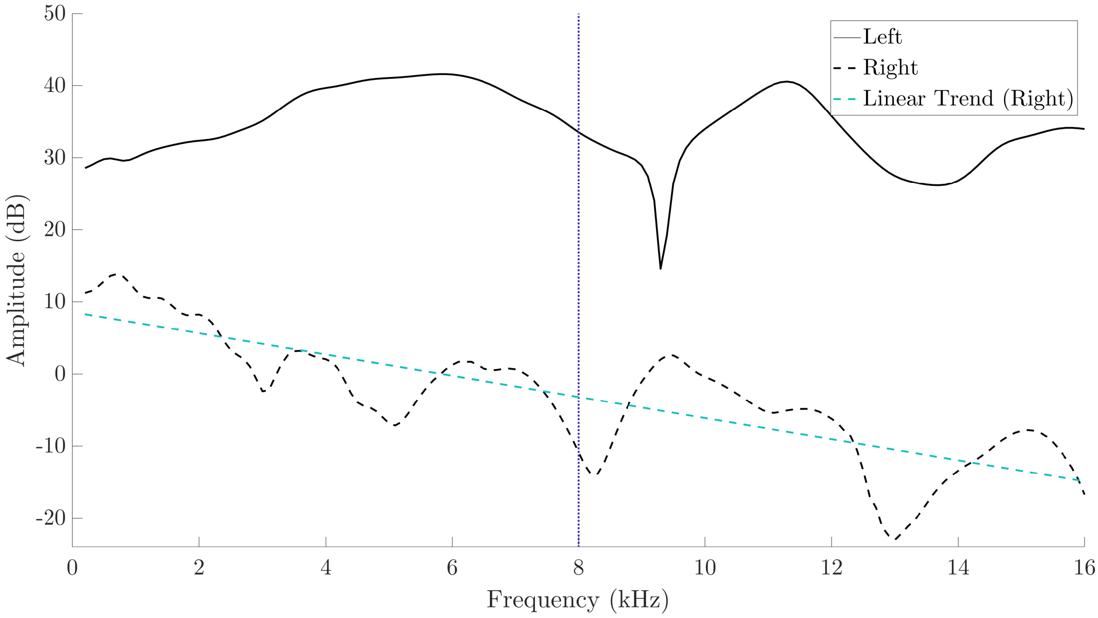
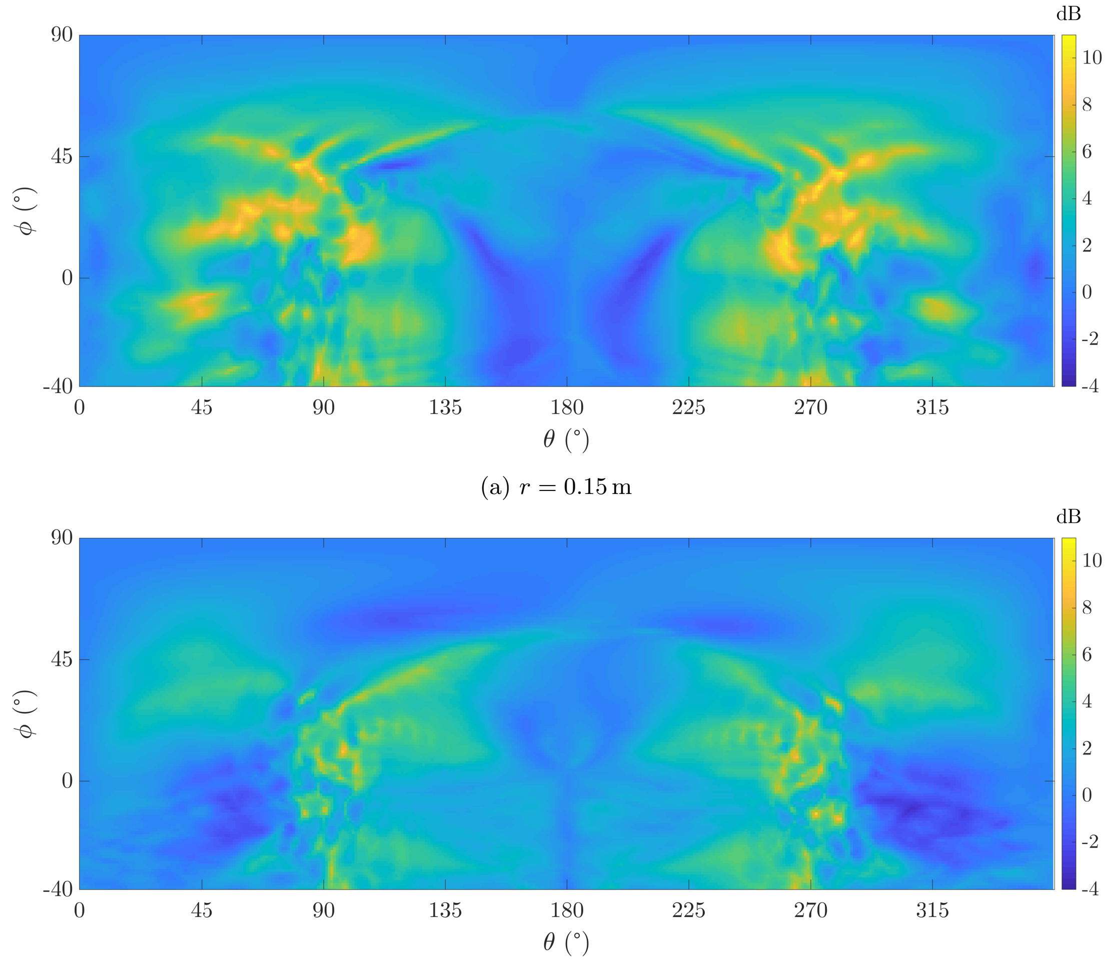


Figure 5.14: The HRTF pair for a source at $\theta = 90°, \phi = 0°, r = 0.15$ m. The 8 kHz limit used in NCS calculation (dotted purple) and the general downward trend of the contralateral HRTF (dashed green) are indicated.

To investigate the impact of using a larger frequency range on the value of NCS obtained for a given source direction, NCS was additionally calculated for all $\mathbf{SP^{KEMAR}}$ positions using a frequency range of 200 Hz to 16 kHz. The signed difference between the two NCS values at each $\mathbf{SP^{KEMAR}}$ position is referred to as ΔNCS:

$$\Delta NCS = NCS^{16k} - NCS^{8k} \tag{5.6}$$

Fig. 5.15 shows the signed variation in ΔNCS for the cases of $r = 0.15$ m and $r = 1.50$ m, where a positive value indicates a larger value in NCS^{16k} and a negative value indicates a larger value in NCS^{8k}. The numerical difference between NCS values calculated with different frequency ranges varies in a complex way. The range of values at $r = 0.15$ m is -2.2 dB to 10.8 dB, and at $r = 1.50$ m is -3.2 dB to 9.4 dB. The distribution of values at each radial distance also varies. The largest ΔNCS values at $r = 0.15$ m occur in arcs beginning approximately 10° behind the interaural axis at $\phi = 0°$ and ending approximately 40° in front of the interaural axis at $\phi = 45°$, as well as around $\theta = 80°$, $\phi = 20°$. However, at $r = 1.50$ m the larger ΔNCS values are predominantly in the rear hemisphere.



(a) $r = 0.15\,\mathrm{m}$

(b) $r = 1.50\,\mathrm{m}$

Figure 5.15: ΔNCS plotted as intensity surfaces for the cases of: a) $r = 0.15\,\mathrm{m}$ and b) $r = 1.50\,\mathrm{m}$. Regions of larger ΔNCS values are visible in the vicinity of the interaural axis for the smaller radial distance, but predominantly in the rear hemisphere for the largest radial distance.

Fig. 5.16 shows the distribution properties of ΔNCS at each radial distance. The distribution of values changes slightly with radial distance, showing a small reduction in maximum and median values as the distance increases. The dashed line indicates the median of the furthest radial distance ($1.5\,\mathrm{dB}$ at $r = 1.50\,\mathrm{m}$). Nevertheless, the range of ΔNCS values at the largest radial distance, $r = 1.50\,\mathrm{m}$, remains substantial, ranging between $-3.2\,\mathrm{dB}$ and $9.4\,\mathrm{dB}$. This emphasises the importance of using the frequency range which corresponds to that used in the literature if comparisons are to be made.

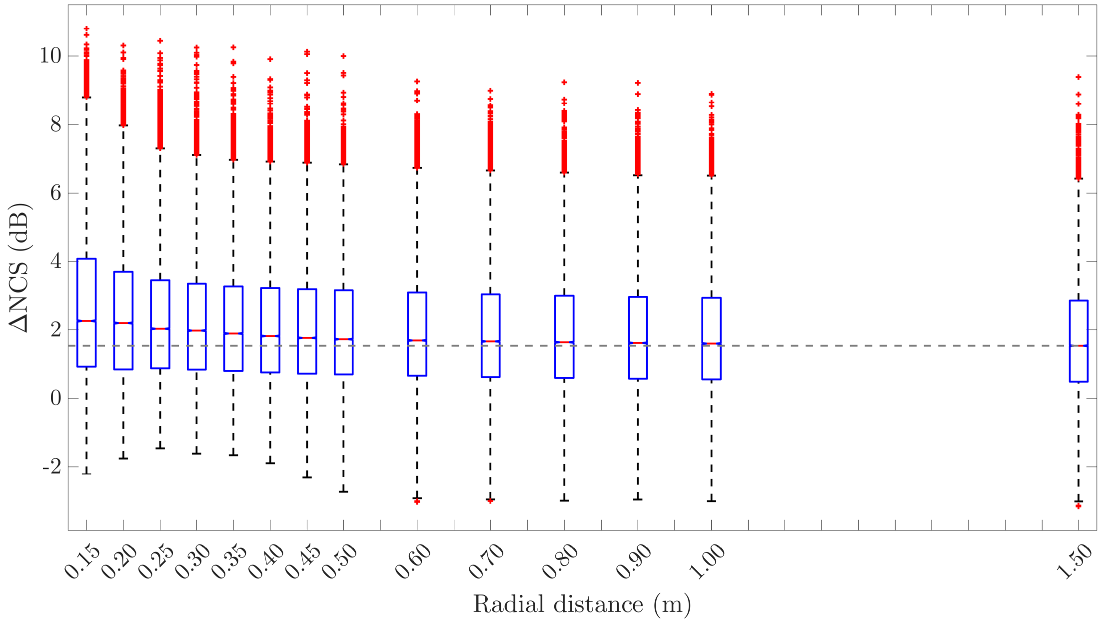


Figure 5.16: Distribution properties of ΔNCS at each radial distance within $\mathbf{SP^{KEMAR}}$. A small reduction in median and maximum value is visible with an increase in radial distance, but ΔNCS values between −3.2 dB and 9.4 dB still exist at the largest radial distance. The dashed line indicates the median of the furthest radial distance (1.5 dB at $r = 1.50$ m).

5.5.2 Source Positions with Sufficient NCS in Frequency Bands

To investigate the existence of $\mathbf{SP^{KEMAR}_{20}}$ positions within different frequency bands, NCS was additionally calculated using four 4 kHz frequency bands (excluding the first frequency band, which has a bandwidth of 3.8 kHz). Fig. 5.17 shows the distribution properties of NCS for each frequency band at each radial distance. A non-linear x-axis has been used to avoid unnecessarily squashed boxes with large amounts of white space. The decrease in NCS with an increase in radial distance is visible (as discussed in Section 5.3.2) and, in accordance with the variation in NCS discussed in Section 5.5.1, NCS increases across the frequency bands at each radial distance. The maximum NCS (55.8 dB) occurs within frequency band 12 to 16 kHz at radial distance $r = 0.15$ m.

Using the analysis methods described in Section 5.4, source positions with NCS exceeding 20 dB were determined for each frequency band (defined as the $\mathbf{SP^{KEMAR}_{20_f}}$ sets). The percentage of $\mathbf{SP^{KEMAR}_{20_f}}$ within $\mathbf{SP^{KEMAR}}$ at each radial distance is plotted in Fig. 5.18 and listed in Table 5.6. Excluding the lowest frequency band, all percentage values are larger than the corresponding values in Table 5.4 for each radial distance by between 8.3 % and 21.8 %.

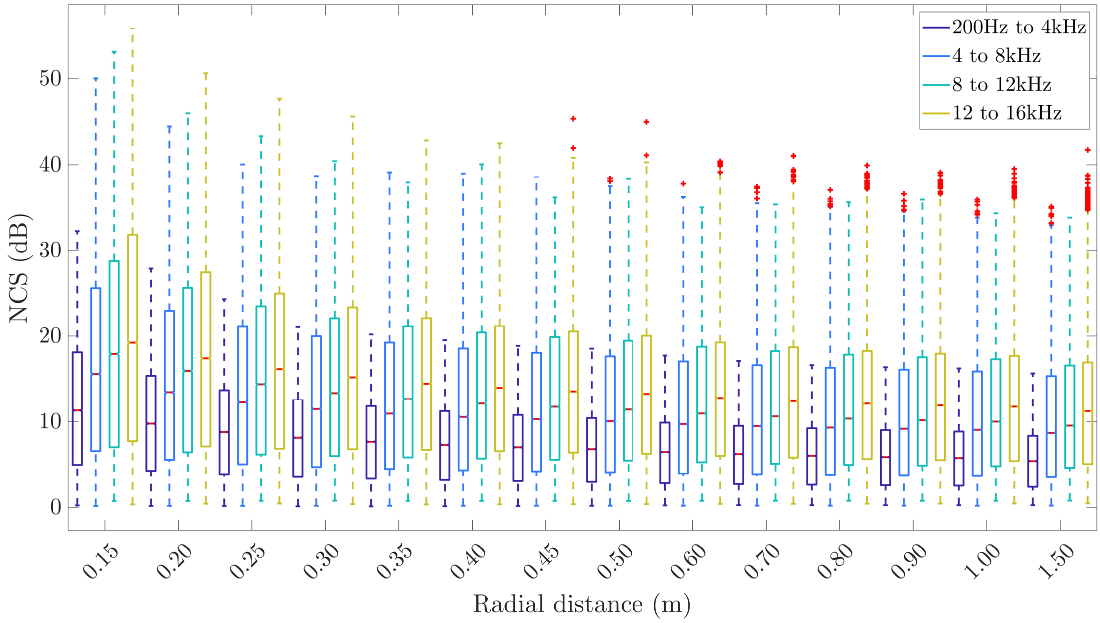


Figure 5.17: Distribution properties of NCS for each frequency band at each radial distance. Note the non-linear axis to avoid unnecessarily squashed boxes.

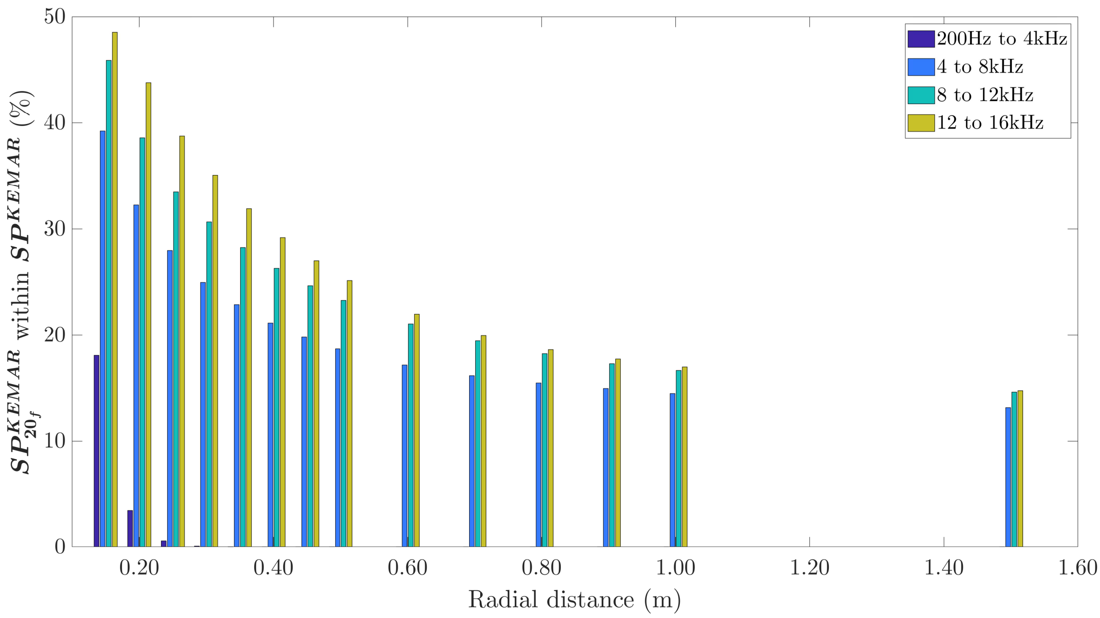
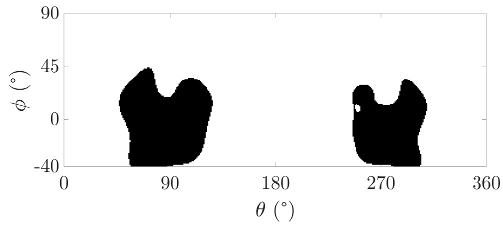


Figure 5.18: Percentage of $\mathbf{SP_{20_f}^{KEMAR}}$ positions which exist in $\mathbf{SP^{KEMAR}}$ for each frequency band f at each radial distance.

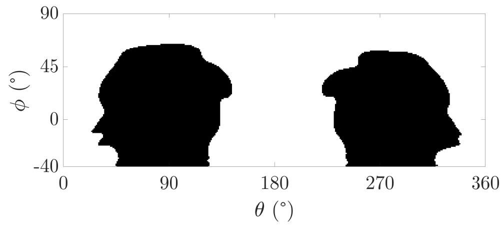
r (m)	$\mathbf{SP^{KEMAR}_{20_f}}$ (%)			
	200 Hz to 4 kHz	4 kHz to 8 kHz	8 kHz to 12 kHz	12 kHz to 16 kHz
0.15	18.1	39.2	45.9	48.6
0.20	3.5	32.2	38.6	43.8
0.25	0.6	28.0	33.5	38.8
0.30	0.1	25.0	30.7	35.1
0.35	0	22.9	28.3	31.9
0.40	0	21.2	26.3	29.2
0.45	0	19.8	24.6	27.0
0.50	0	18.7	23.3	25.1
0.60	0	17.2	21.1	22.0
0.70	0	16.2	19.5	20.0
0.80	0	15.5	18.3	18.6
0.90	0	15.0	17.3	17.8
1.00	0	14.5	16.7	17.0
1.50	0	13.1	14.6	14.8

Table 5.6: Percentage of $\mathbf{SP^{KEMAR}_{20_f}}$ positions within $\mathbf{SP^{KEMAR}}$ for each frequency band f at each radial distance.

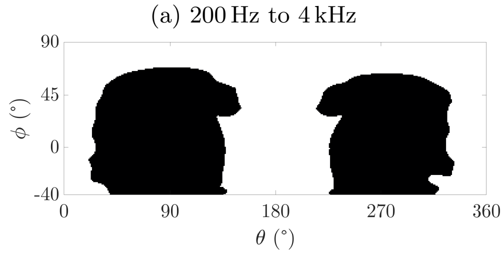
Fig. 5.19 shows the regions of $\mathbf{SP^{KEMAR}_{20_f}}$ positions for the smallest radial distance (indicated in black). The regions for all radial distances analysed are included in Appendix E. The increase in number of source positions with frequency band is visible. As radial distance increases, the regions of $\mathbf{SP^{KEMAR}_{20_f}}$ positions for each frequency band reduce in size and become less homogeneous with increased numbers of holes and fragments, particularly in the uppermost frequency band (12 to 16 kHz). Beyond radial distance $r = 0.35$ m, zero $\mathbf{SP^{KEMAR}_{20_f}}$ positions exist for the lowest frequency band.



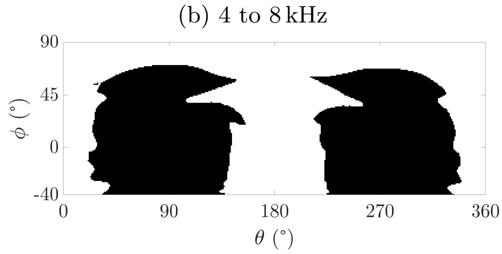
(a) 200 Hz to 4 kHz



(b) 4 to 8 kHz



(c) 8 to 12 kHz



(d) 12 to 16 kHz

Figure 5.19: Regions of $\mathbf{SP^{KEMAR}_{20_f}}$ positions (indicated in black) within $\mathbf{SP^{KEMAR}}$ at $r = 0.15$ m for the four frequency bands used in analysis.

Binaural spectral cues are known to operate above approximately 5 kHz (see Section 2.5.4). The $\mathbf{SP^{KEMAR}_{20_f}}$ sets for frequency bands above this value are larger than the corresponding $\mathbf{SP^{KEMAR}_{20}}$ set at each radial distance. This indicates that, if reproduction of only certain frequency regions is desired, more left-right loudspeaker pairs may be capable of accurate binaural reproduction without the need for CTC than suggested when using the frequency range described in the literature (200 Hz to 8 kHz). However, without values of sufficient NCS within these frequency ranges with which to compare, this is only speculation.

5.6 Conclusion

This chapter describes a study into the variation of NCS with loudspeaker direction and radial distance. In order to calculate NCS, a HRTF pair between each desired loudspeaker position and the two ears is required. Due to the large numbers of positions involved in this study, simulation using the BEM was the most appropriate method of HRTF acquisition. HRTFs of the *half-torso-17* mesh of KEMAR created in Chapter 4 were computed using the BEM for 655,216 source positions (referred to as the set $\mathbf{SP^{KEMAR}}$) organised predominantly in spherical domes across 15 radial distances between 0.10 m and 1.50 m. NCS, defined as the average spectral difference between the ears over the range 200 Hz to 8 kHz and caused by the acoustic shadowing of the head and torso, was calculated for each $\mathbf{SP^{KEMAR}}$ source position. Inclusion of frequencies above this threshold leads to NCS values which are not comparable to values discussed in the literature, where a channel separation of 20 dB has been reported as the threshold for binaural robustness [17, 228].

The variation in NCS has been investigated as a function of both direction and radial distance. In general, larger NCS is observed for sound sources close to the head, in the vicinity of the interaural axis and away from the median plane. This is to be expected, as this is where acoustic head shadowing effects are greatest. The global maximum NCS was 45.9 dB, and occurs for a sound source at $\theta = 270°,\ \phi = 0°,\ r = 0.10$ m. The global minimum was 0.2 dB, and occurs at $\theta = 183°,\ \phi = -16°,\ r = 0.15$ m.

Distinct patterns of NCS are observed which vary in a stable and consistent way with increasing radial distance. Whilst values generally decrease with increasing distance, the direction of highest NCS remains approximately the same: slightly behind and slightly below the interaural axis. Left-right hemisphere differences can be attributed to the small asymmetries inherent in KEMAR. The much larger front-back hemisphere differences are due to the differences in head morphology between the front and the rear of the head of KEMAR (and, indeed, of any human head), for example the rear offset of the ears and the concave nape of the neck.

The maximum, mean and median values for each spherical dome within $\mathbf{SP^{KEMAR}}$ all decrease with increasing radial distance, effectively reaching a plateau beyond approximately 0.50 m from the head. The maximum NCS at $r = 0.15$ m is 37.6 dB, at $r = 0.50$ m is 26.0 dB and at $r = 1.50$ m is 22.7 dB. The median and mean values follow the same trend. The minimum value at each radial distance remains approximately the same, ranging from 0.2 dB to 0.4 dB. An additional impact of radial distance can be examined by normalising the NCS values in each spherical dome with respect to the maximum value for that dome. There is not only a global decrease as radial distance increases: larger radial distances contain more of the lower NCS values whilst smaller radial distances possess a greater proportion of the higher NCS values.

There are many positions at smaller radial distances which exceed the 20 dB of channel separation reported to maintain binaural robustness (referred to as the set $\mathbf{SP_{20}^{KEMAR}}$). 30.9 % of source positions at the smallest radial distance (0.15 m) exhibit NCS values greater than 20 dB, but only 0.3 % of positions at the furthest distance (1.50 m) do so. For the smallest radial distance, these positions are located approximately at azimuth angles of $\theta = 90 \pm 40°$ and $\theta = 270 \pm 40°$, and below $\phi = 60°$ in elevation.

These findings suggest that source positions exist which, numerically, produce channel separation that is large enough for successful binaural reproduction. When used in combination, left-right pairs of loudspeakers at these positions may be capable of reproducing binaural signals without the need for CTC. However, these results alone do not indicate success in perceptual terms, only in numerical terms. Simulation and analysis of the perceptual performance of left-right pairs of loudspeakers at these candidate source positions via the use of a virtual reproduction system for binaural reference signals is discussed in Chapter 6.

Chapter 6

Simulation of the Perceptual Performance of Candidate Loudspeaker Pairs with respect to KEMAR

Discard [it] and discover a brand new way.

Tegan and Sara, *Hell*
Sainthood (2009)

6.1 Introduction

Loudspeaker-based binaural reproduction typically requires the use of crosstalk cancellation (CTC) to ensure that information from the relevant binaural channel reaches each ear from the closest loudspeaker without degradation as a result of interference from the other loudspeaker. The success of a CTC system is often measured in terms of the channel separation (CS) achieved; that is, how much the interfering channel is suppressed [220–222]. However, CS occurs naturally as a result of the acoustic shadowing produced by the head and torso, and increases as the sound source is brought closer to the head [19, 21]. The values of this natural channel separation (NCS), determined in Chapter 5, indicate the likely existence of loudspeaker positions which do not require a CTC stage to achieve the CS necessary for binaural reproduction. However, some form of perceptual testing of these candidate loudspeaker positions when used within a binaural reproduction system without CTC is required. This method of binaural reproduction (reproduction over a left-right loudspeaker pair without the use of a CTC stage) is referred to in this work as *CTC-less* binaural reproduction.

This chapter describes the process of creating a virtual test environment and its use in assessing the impact of loudspeaker placement on CTC-less binaural reproduction. As it is not yet known in the context of this work whether NCS is an accurate indicator of perceptual binaural performance, all loudspeaker positions originally defined in

Section 5.2.1 are tested, not only the subset identified in Section 5.4 as candidate positions.

Due to the large number of loudspeaker positions under test, it was not feasible to conduct this study using acoustic reproduction and human subjects in a listening test. Objective binaural cue analysis, as opposed to a series of listening tests with human subjects, also has the benefit of avoiding the introduction of additional errors due to differences between the transfer functions used in the reproduction system and those of each subject. Therefore, a virtual reproduction system (referred to as the VRS) was implemented to simulate the signal paths that occur in a two-loudspeaker acoustic binaural reproduction system.

Simulation using the VRS enabled the comparison of two pairs of left-right head-related impulse response (HRIR) pairs for each loudspeaker pair under test, referred to as *reference* HRIRs and *modified* HRIRs. The reference HRIRs are those which would be used to spatialise a mono sound source to create a binaural sound source (at a specific direction and distance) under ideal circumstances (for example, over equalised headphones as discussed in Section 2.7.1). In Equation 2.67, H_L and H_R would be the reference HRIR pair.

The modified HRIR pairs are the result of reproducing the reference HRIR pairs over a simulated loudspeaker pair, as discussed in Section 2.7.2. In Equation 2.71, the modified HRIR pair is represented by the product of the two matrices $\mathbf{A}$ (the acoustic transfer functions between the loudspeakers and the ears) and $\mathbf{H}$ (the reference HRIR pair), in that the modified HRIR pairs are the result of the reference HRIR pairs having been modified by the system-specific transfer functions, including any crosstalk paths that may be present:

$$\begin{bmatrix} A_{LL} & A_{LR} \\ A_{RL} & A_{RR} \end{bmatrix} \begin{bmatrix} H_L \\ H_R \end{bmatrix} \tag{6.1}$$

In this work, these modified HRIRs are created as a result of inputting the reference HRIRs to the VRS. The modified HRIR pairs for 18 reference HRIR pairs were simulated for each of the 325,780 loudspeaker pairs under test.

To determine the impact of the loudspeaker placement on the binaural cues reproduced using CTC-less binaural reproduction, the difference between the estimated interaural binaural cues for the reference HRIR pair and for each corresponding modified HRIR pair was calculated for each loudspeaker pair under test. A perceptual model is devised here to relate the numerical differences to perceptual impact. This approach provides a means of identifying positions where loudspeaker pairs exhibit deterioration between the reference and modified HRIR pairs below the threshold of perception. These loudspeaker pair positions are, additionally, compared to positions which exhibit an NCS of greater than 20 dB to investigate the degree to which NCS is a predictor of satisfactory binaural reproduction performance.

6.2 Methods

A means of approximating and analysing the perceptual performance of the loudspeaker positions in a virtual test environment was required. A virtual reproduction system (henceforth referred to as the VRS) was implemented using the simulated HRIRs discussed in Chapter 5 as the loudspeaker acoustic paths. This allowed comparison between 18 left-right reference HRIR pairs and corresponding left-right modified HRIR pairs created as a result of reproducing the reference HRIR pair using each loudspeaker pair under test. No audible stimuli were used in this reproduction system.

For clarity, and in accordance with the conventions set out in Section 2.7, the impulse responses (IRs) and transfer functions are referred to using the following:

- the system IRs for the loudspeaker pair under test are referred to as $\mathbf{a}$, where the intended paths are a_{LL} and a_{RR}, respectively, and the crosstalk paths are a_{LR} and a_{RL}, respectively (in the frequency domain, $\mathbf{A}$ is used);

- the reference HRIR pairs are referred to as $\mathbf{h^N}$, where the left and right signals are h_L^N and h_R^N, respectively (in the frequency domain, $\mathbf{H^N}$ is used), and the integer N is the index of the reference HRIR pair from 1 to 18;

- the modified HRIR pairs are referred to as $\hat{\mathbf{h}}_\mathbf{M}^\mathbf{N}$, where the left and right signals are $\hat{h}_{M_L}^N$ and $\hat{h}_{M_R}^N$, respectively (in the frequency domain, $\hat{\mathbf{H}}_\mathbf{M}^\mathbf{N}$ is used), the integer N is the index of the reference HRIR pair from 1 to 18 and the integer M is the index of the loudspeaker pair under test from 1 to 325,780.

The effect of the position of each loudspeaker pair on the binaural reproduction could then be determined by calculating the differences between the estimated interaural binaural cues of $\mathbf{h^N}$ and $\hat{\mathbf{h}}_\mathbf{M}^\mathbf{N}$. If the loudspeaker pair under test has a sufficiently numerically small effect on the reproduction, the differences observed will be of negligible perceptual consequence.

6.2.1 Reference HRIRs

As the performance of each loudspeaker pair position across a range of directions was of interest, 18 reference HRIR pairs were selected to give good coverage of the sphere. They lie on the vertices of an icosahedron (where the vertices are defined by circular permutations of the basis vector $[1, 0, \chi]$ where χ is the golden ratio [284]) to give a regular distribution over the sphere, with an additional six points for the extremes of each principal anatomical plane (above, below, front, back, left and right). These reference directions are listed in Table 6.1 and plotted in Fig. 6.1. Reference HRIR pairs were taken from the KEMAR measurements in the Spatial Audio Domestic Interactive Entertainment (SADIE) database v1 [123] (RAW format, 48 kHz, 24 bit).

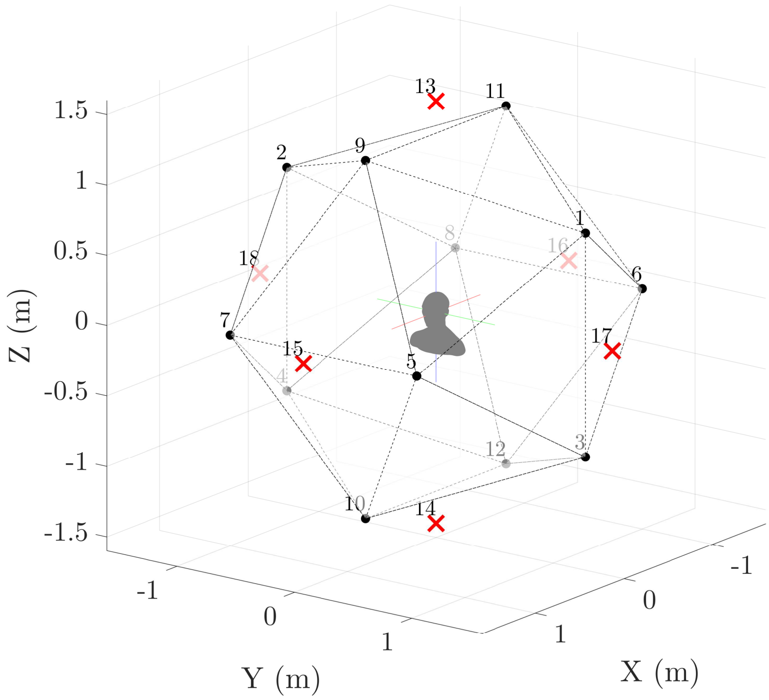


Figure 6.1: The 18 reference HRIR pair directions $\mathbf{h^N}$, plotted using Cartesian coordinates. The numbers correspond to the values of N in Table 6.1. The icosahedron is indicated in black, the additional points are red.

N	In icosahedron?	θ (°)	ϕ (°)	r (m)
1	Yes	90	32	1.5
2	Yes	270	32	1.5
3	Yes	90	-32	1.5
4	Yes	270	-32	1.5
5	Yes	32	0	1.5
6	Yes	148	0	1.5
7	Yes	328	0	1.5
8	Yes	212	0	1.5
9	Yes	0	58	1.5
10	Yes	0	-58	1.5
11	Yes	180	58	1.5
12	Yes	180	-58	1.5
13	No	0	90	1.5
14	No	0	-90	1.5
15	No	0	0	1.5
16	No	180	0	1.5
17	No	90	0	1.5
18	No	270	0	1.5

Table 6.1: The 18 reference HRIR pair directions $\mathbf{h^N}$ in spherical coordinates.

6.2.2 Virtual Reproduction System

The VRS was implemented in MATLAB [285]. The HRIRs simulated in Chapter 5 using the *half-torso-17* mesh of the Knowles Electronics Manikin for Acoustic Research (KEMAR) were used as the system IRs. These describe the acoustic paths $\mathbf{a}$ between each loudspeaker position and the two ears. The path between each loudspeaker position and the ipsilateral ear is referred to as the *intended* path (that which either the left or right binaural channel should be reproduced over) and the path between each loudspeaker and the contralateral ear is referred to as the *crosstalk* path (that which creates the unwanted interference). This is shown in Fig. 6.2, where an intended path (indicated in bold) and a crosstalk path exist for each loudspeaker as listed in Table 6.2. h_L^N and h_R^N represent the left and right channels of the reference HRIR, respectively, and $\hat{h}_{M_L}^N$ and $\hat{h}_{M_R}^N$ represent the left and right channels of the modified HRIR, respectively.

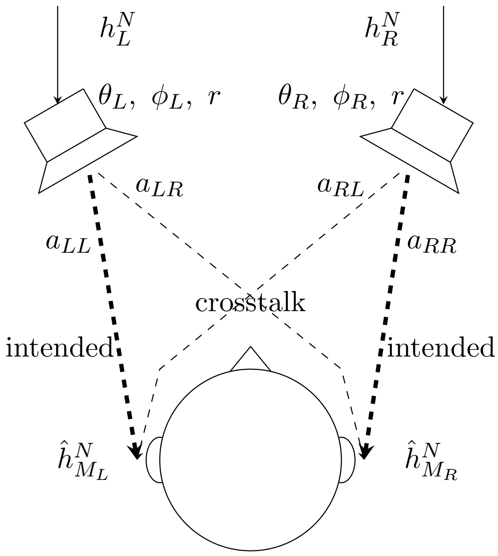


Figure 6.2: For each loudspeaker, two acoustic paths exist: an intended path (indicated using bold dashes) and a crosstalk path (plain dashes). a_{LL} and a_{RR} are the intended paths, a_{LR} and a_{RL} are the crosstalk paths. h_L^N and h_R^N represent the left and right channels of the reference HRIR, respectively, and $\hat{h}_{M_L}^N$ and $\hat{h}_{M_R}^N$ represent the left and right channels of the modified HRIR, respectively.

Loudspeaker	Intended path	Crosstalk path
Left ($\theta_L,\ \phi_L,\ r$)	a_{LL}	a_{LR}
Right ($\theta_R,\ \phi_R,\ r$)	a_{RR}	a_{RL}

Table 6.2: Labels used to denote the intended and crosstalk path for each loudspeaker of the pair under test.

All loudspeaker positions were tested using the VRS, not only those indicated as possible candidate positions in Chapter 5, as it was not yet known in the context of this work whether NCS was an accurate predictor of binaural perceptual performance.

A set of 325,780 loudspeaker pairs (23,270 per radial distance), referred to as $\mathbf{LP^{KEMAR}}$, was defined by pairing laterally symmetrical $\mathbf{SP^{KEMAR}}$ positions (that is, positions which are mirrored in the median plane and so excluding those actually in the median plane). For example, $\mathbf{SP^{KEMAR}}$ positions at $\theta = 30°,\ \phi = 0°,\ r = 0.15\,\text{m}$ and $\theta = 330°,\ \phi = 0°,\ r = 0.15\,\text{m}$ are paired. In line with convention, and to keep computational requirements within practical limits, pairings of asymmetrical left-right loudspeaker pairs were excluded from this analysis[1]. Using set notation, $\mathbf{LP^{KEMAR}}$ can be defined as:

$$\mathbf{LP^{KEMAR}} = \{(v, w) \mid v, w \in \mathbb{R}^3\} \tag{6.2}$$

which states that $\mathbf{LP^{KEMAR}}$ is a set of pairs of vectors v and w, such that v and w are each a three-dimensional, real-valued column vector, related through being mirrored in the median plane. For brevity, each member of the set $\mathbf{LP^{KEMAR}}$ (consisting of two $\mathbf{SP^{KEMAR}}$ positions) is referred to using the angle of the left hemisphere $\mathbf{SP^{KEMAR}}$ position. Later in this chapter, regions of loudspeaker pair positions will be discussed and plotted with reference to the left hemisphere.

For each loudspeaker, an ITD and ILD exist between the intended path and the crosstalk path. As a result of the asymmetry inherent within KEMAR, these ITD and ILD values are slightly different for the two loudspeaker positions within each member of $\mathbf{LP^{KEMAR}}$. The two loudspeaker positions, although symmetrical in angle and radial distance from the origin, are not geometrically symmetrical with regards to their respective ipsilateral ears. Ideally, there should be no difference in the time and level of arrival of a_{LL} compared to a_{RR}, but, due to the asymmetries in KEMAR, this is not the case. This disrupts the crucial relationship between the ipsilateral and contralateral binaural channels.

Whilst small asymmetries such as these may be present in a real-world reproduction system, at this stage of the work it is desirable to create an idealised symmetrical system. Therefore, the IRs for each right hemisphere loudspeaker position are aligned to match the arrival time and level for the corresponding left hemisphere IRs. The phase differences between a_{LL} and a_{RR} and between a_{LR} and a_{RL} are calculated and applied to the right hemisphere IRs. The difference in root mean square (RMS) value between the adjusted right hemisphere IRs and the left hemisphere IRs are calculated and applied to the adjusted right hemisphere IRs to produce time- and level-aligned IRs, whilst retaining the original spectral fine structure.

Inverse filters (denoted in the time domain by i_{LL} and i_{RR}) are required to equalise the intended paths for each loudspeaker position. Although the VRS implemented here does not use a CTC stage, the acoustic paths between the loudspeakers and ears (a_{LL} and a_{RR}) must be accounted for to maintain binaural fidelity. This stage is referred to as *ipsilateral inversion*. A future simplification could be to omit the ipsilateral inversion stage (discussed further in Section 8.3.5), but this is outside the scope of the work presented here.

The inverse filters were calculated using the same implementation of the least-mean-square regularisation method outlined by Kirkeby *et al.* [269] as in Chapter 3.

[1] Whilst Han *et al.* [243] suggest that asymmetrically located pairs of loudspeakers are more robust to head movement, the practicality of the findings is questionable as the suggested pairs involve high elevation values.

Example inversion filters for the loudspeaker pair where the left loudspeaker is at $\theta = 45^{\circ},\ \phi = 0^{\circ},\ r = 0.15\,\text{m}$ are shown in Fig. 6.3.

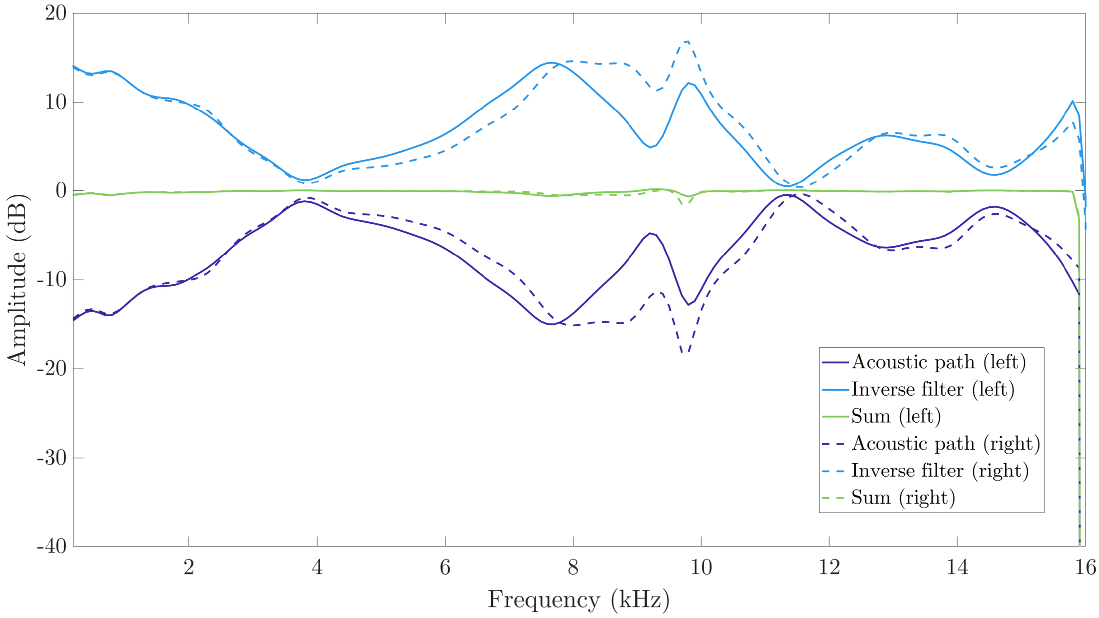


Figure 6.3: Frequency spectra of acoustic paths, ipsilateral inversion filters and their respective sums generated for the loudspeaker pair where the left is at $\theta = 45^{\circ},\ \phi = 0^{\circ},\ r = 0.15\,\text{m}$.

Using the VRS, a modified HRIR pair was produced for each of the 18 reference HRIR pairs, reproduced over each of 325,780 loudspeaker pairs, giving a total of 5,864,040. The three stages of the VRS for each reference HRIR pair reproduced over each $\mathbf{LP}^{\mathbf{KEMAR}}$ loudspeaker pair are listed below with their corresponding mathematical descriptions. The signal flow of the stages is also shown in Fig. 6.4.

1. Ipsilateral inversion: convolution of the left and right signal of the reference HRIR with the corresponding inverse filter for the intended paths;

$$h_L^N * i_{LL} \qquad\qquad h_R^N * i_{RR} \tag{6.3}$$

2. System reproduction: convolution of the pair of filtered signals with the IRs for each corresponding intended and crosstalk path;

$$\begin{aligned} &\left(h_L^N * i_{LL}\right) * a_{LL} \qquad\qquad &\left(h_L^N * i_{LL}\right) * a_{LR} \\ &\left(h_R^N * i_{RR}\right) * a_{RR} \qquad\qquad &\left(h_R^N * i_{RR}\right) * a_{RL} \end{aligned} \tag{6.4}$$

3. Summation: summation of the signals to produce a signal for each ear.

$$\begin{aligned} \hat{h}_{M_L}^N &= \left(\left(h_L^N * i_{LL}\right) * a_{LL}\right) + \left(\left(h_R^N * i_{RR}\right) * a_{RL}\right) \\ \hat{h}_{M_R}^N &= \left(\left(h_R^N * i_{RR}\right) * a_{RR}\right) + \left(\left(h_L^N * i_{LL}\right) * a_{LR}\right) \end{aligned} \tag{6.5}$$

In matrix notation, the VRS described in steps 1–3 can be written as:

$$\begin{bmatrix} \hat{H}^N_{M_L} \\ \hat{H}^N_{M_R} \end{bmatrix} = \begin{bmatrix} A_{LL} & A_{LR} \\ A_{RL} & A_{RR} \end{bmatrix} \begin{bmatrix} I_{LL} & 0 \\ 0 & I_{RR} \end{bmatrix} \begin{bmatrix} H^N_L \\ H^N_R \end{bmatrix} \tag{6.6}$$

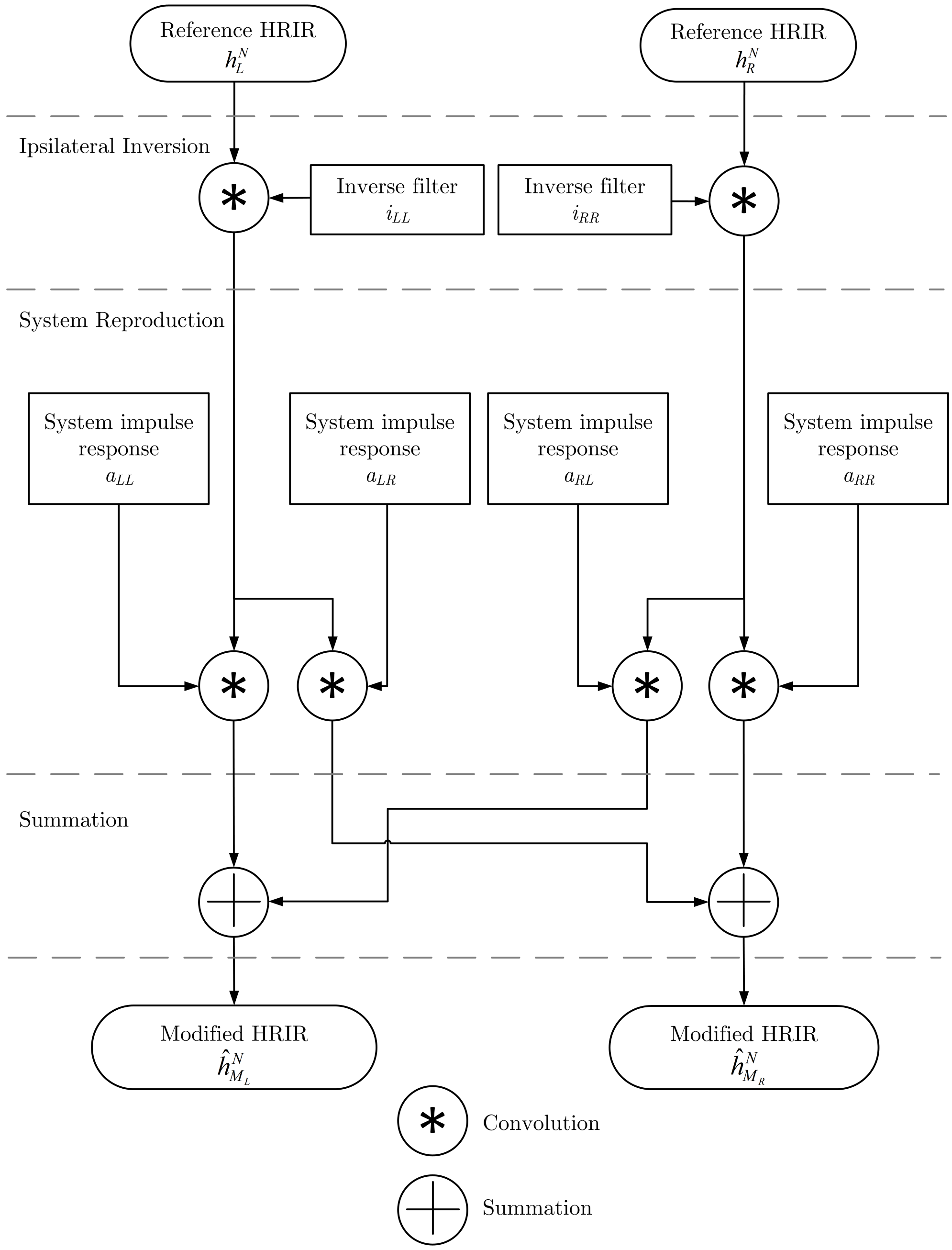


Figure 6.4: The signal flow of the virtual CTC-less binaural reproduction system (VRS) used to simulate the modified HRIRs. The three stages are indicated: ipsilateral inversion, system reproduction and summation.

6.2.3 Estimation of Binaural Cues

The interaural binaural cues of all reference HRIR pairs and modified HRIR pairs were estimated in order to compare each reference HRIR pair with the corresponding modified HRIR pairs (that is, for the same value of N). ITD was calculated below 800 Hz using the cross-correlation method (see Section 2.5.1). ISD was calculated as the unsigned spectral difference at each frequency point between 200 Hz and 16 kHz, and ILD calculated as the arithmetic mean of these values. Lower frequencies than those usually included in an ILD calculation (usually above 1.5 kHz) are used due to the importance of low-frequency content at near-field source positions [36]. Whilst not strictly the ILD calculation method described in Section 2.5.2, this method allows for inclusion of more of the variation between the HRIR pairs over frequency than simply calculating a wideband ILD value for each HRIR pair and comparing these.

6.3 Results

For each $\mathbf{LP^{KEMAR}}$ loudspeaker pair, the binaural cues delivered by each modified HRIR pair were investigated, both in absolute units (µs and dB) and perceptual units (just-noticeable difference (JND)). For example, for loudspeaker pair $M = 314$ and reference HRIR pair $N = 2$, the difference between reference HRIR pair $\mathbf{h^2}$ and modified HRIR pair $\mathbf{\hat{h}^2_{314}}$ was calculated. From this, loudspeaker pairs exhibiting deterioration from the reference HRIR pair to the modified HRIR pair below the threshold of perception, for all reference HRIRs, were determined. An indication of the region of best-performing loudspeaker pairs was also determined.

6.3.1 Differences in Binaural Cues

To estimate the impact of each $\mathbf{LP^{KEMAR}}$ loudspeaker pair on the binaural cues, the unsigned differences in ITD and ILD (ΔITD and ΔILD, respectively) were calculated between each reference HRIR pair and the corresponding modified HRIR pair. This gave 5,864,040 ($18 \times 325{,}780$) values for each of ΔITD and ΔILD. Fig. 6.5 shows the distribution properties of ΔITD, split by reference HRIR pair using the values of N in Table 6.1. The global maximum value is 750.4 µs and the global minimum value is 0 µs. The global 95 % interval is 537.1 µs. The global median value is 6.5 µs, whilst the global mean value is much larger at 98.4 µs. This indicates that the data is positively skewed, with a long tail to the right side of the median value. This is due to the differences for $N = 1,\ 2,\ 17,\ 18$ and those on the median plane ($9 \leq N \leq 16$). The differences for these values of N are generally small, in particular those on the median plane, where values reach a maximum of 29.2 µs and median and mean values are 2.5 µs and 3.6 µs, respectively. This is due to the difference between the left and right signals in median plane HRIR pairs being relatively small; hence, the loudspeaker pair does not need a large CS to be able to maintain the small difference.

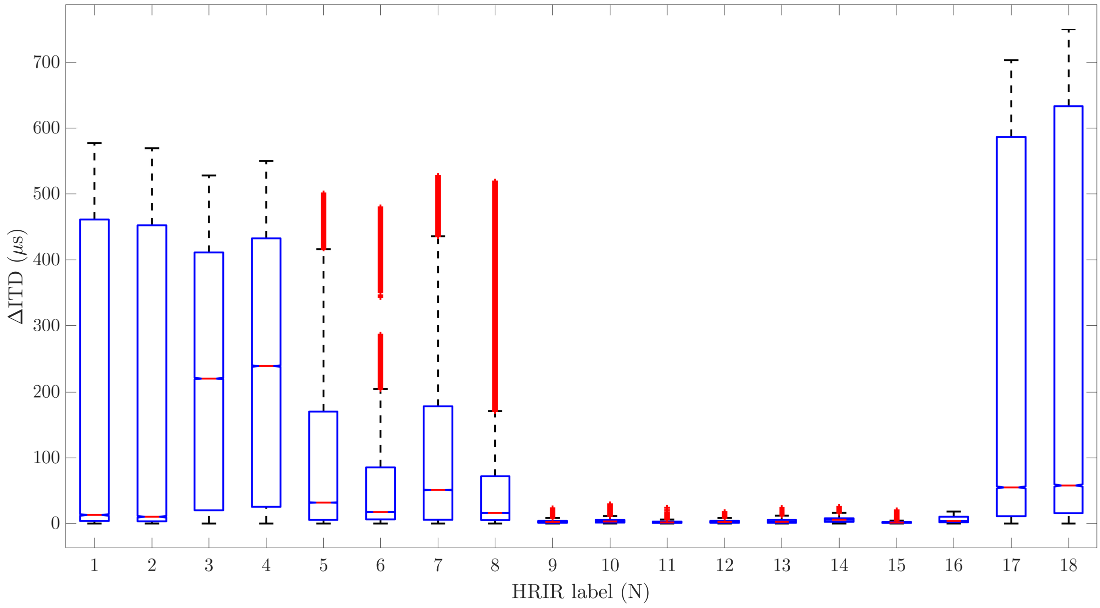


Figure 6.5: Distribution properties of ΔITD, split by reference HRIR pair, where the x-axis label corresponds to those listed in Table 6.1. The global maximum value is 750.4 µs and the minimum value is 0 µs, with median and mean values of 6.5 µs and 98.4 µs, respectively.

Fig. 6.6 shows a similar boxplot to Fig. 6.5, but for the distribution properties of ΔILD rather than ΔITD. The global maximum value is 23.5 dB and the global minimum value is 0.0 dB. The global 95 % interval is 16.5 dB. The global median value is 3.8 dB and the mean value is 5.4 dB, again indicating positively-skewed data as observed for ΔITD. The reduction in difference values for HRIR pairs $9 \leq N \leq 16$ compared to other values of N is apparent here, although it is less marked than in the ΔITD case. Here, the maximum ΔILD value across reference HRIR pairs $9 \leq N \leq 16$ is 10.2 dB, with median and mean values of 1.7 dB and 2.2 dB, respectively. As with ΔITD, this is due to the loudspeaker pair not needing a large CS to be able to maintain the small difference between the left and right signals in the median plane HRIR pairs.

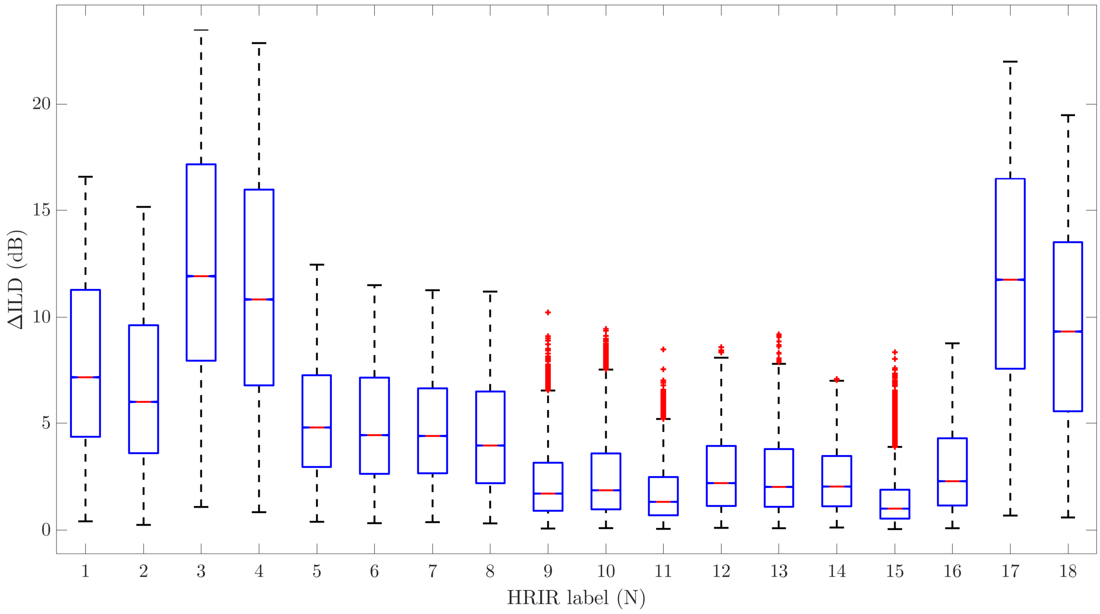


Figure 6.6: Distribution properties of ΔILD, split by reference HRIR pair, where the x-axis label corresponds to those listed in Table 6.1. The global maximum value is 23.5 dB and the minimum value is 0.0 dB, with median and mean values of 3.8 dB and 5.4 dB, respectively.

To identify the loudspeaker pairs which are capable of reproducing a range of binaural signals faithfully without a CTC stage, it is more useful to look at the performance of each loudspeaker pair across all 18 reference HRIRs, rather than the performance for each reference HRIR pair across all loudspeaker pairs. Fig. 6.7 and Fig. 6.8 show ΔITD and ΔILD for all 18 reference HRIR pairs for a single example loudspeaker pair where the left loudspeaker is at $\theta = 1°$, $\phi = 0°$, $r = 1.5$ m. The centre of the figure aligns with $\theta = 0°$, $\phi = 0°$, spot colour represents the magnitude and the numerical values correspond to N in Table 6.1.

Fig. 6.7 and Fig. 6.8 show that this loudspeaker pair reproduces binaural cues for median plane sources well, with small differences between reference and modified HRIR pairs in the region of 3.1 to 13.3 µs and 1.7 to 3.9 dB, respectively. However, at more lateral positions, differences up to 749.6 µs and 22.0 dB are evident. This pattern extends across all loudspeaker pairs, which also mirror the distribution properties shown in Fig. 6.5 and Fig. 6.6. Smaller differences are present for reference HRIR pairs in the median plane, and these differences increase with increasing lateral position of the sound source.

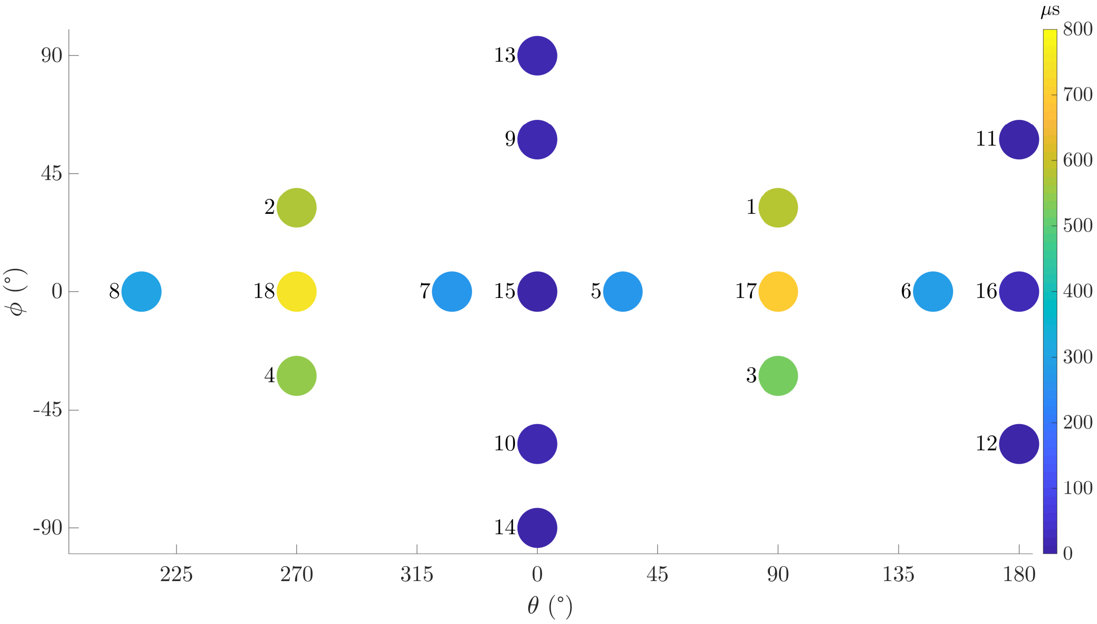


Figure 6.7: ΔITD (spot colour) between each reference and modified HRIR pair for a loudspeaker pair where the left is at $\theta = 1^{\circ}$, $\phi = 0^{\circ}$, $r = 1.5\,\mathrm{m}$. Values correspond to N in Table 6.1.

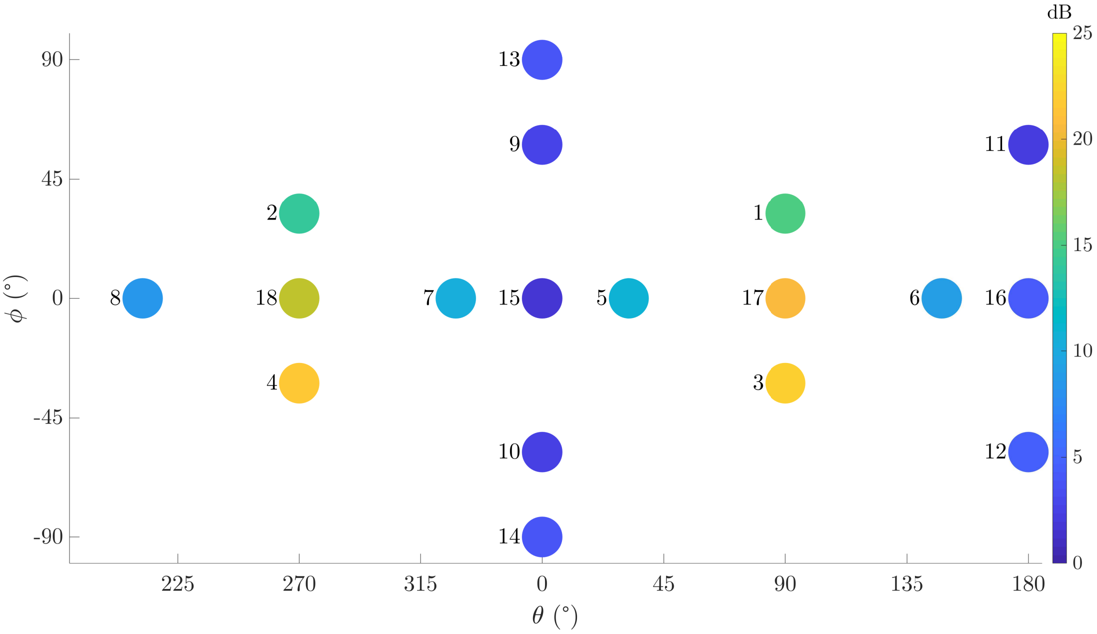


Figure 6.8: ΔILD (spot colour) between each reference and modified HRIR pair for a loudspeaker pair where the left loudspeaker is at $\theta = 1^{\circ}$, $\phi = 0^{\circ}$, $r = 1.5\,\mathrm{m}$. Values correspond to N in Table 6.1.

To describe the behaviour of each loudspeaker pair across all 18 reference HRIR pairs, the arithmetic means, standard deviations and ranges for ΔITD and ΔILD (referred to collectively as the *six statistical metrics*) were calculated, respectively.

Arithmetic mean was chosen rather than median, as it is of importance to characterise the loudspeaker pair behaviour with respect to all reference HRIRs, and explicitly include possible outliers, should they exist. For example, if 17 of the 18 differences for a loudspeaker pair were low but one was much larger, this would indicate that that loudspeaker pair was unable to reproduce binaural signals from that specific direction, which would merit further investigation. This situation would likely produce a median lower than the arithmetic mean. Therefore, the arithmetic mean describes the overall performance, and standard deviation and range together describe the consistency of the loudspeaker pair across all 18 reference directions. These results should help to identify any loudspeaker pair positions where binaural reproduction is possible without the need for CTC across all the reference HRIRs under test.

The six statistical metrics calculated for the loudspeaker pair where the left loudspeaker is at $\theta = 1°,\ \phi = 0°,\ r = 1.5\,\text{m}$ (as plotted in Fig. 6.7 and Fig. 6.8) are listed in Table 6.3.

Cue	Arithmetic Mean	Standard Deviation	Range
ΔITD	270.4 µs	275.2 µs	746.5 µs
ΔILD	9.8 dB	7.2 dB	20.3 dB

Table 6.3: Values of the six statistical metrics used in analysis, describing the characteristics of the loudspeaker pair where the left is at $\theta = 1°,\ \phi = 0°,\ r = 1.5\,\text{m}$ over all 18 reference HRIRs.

Whilst these values are in the units appropriate for ITD and ILD, they shed little light on the associated perceptual implications. As discussed in Section 2.5.6, the thresholds of perception for the spatial cues are not uniform across the sphere: the JND thresholds for changes in ITD and ILD are smallest for sounds emanating from the frontal region, and become larger as the sound source moves away in both azimuth and elevation. Therefore, the same numerical value may have a different perceptual impact, depending on the source direction. The smallest perceivable changes (and, therefore, defined as $1\,\text{JND}_{\text{ITD}}$ and $1\,\text{JND}_{\text{ILD}}$) have been widely reported as [110, 111, 115, 116, 118]:

$$1\,\text{JND}_{\text{ITD}} = 16\,\mu\text{s} \tag{6.7}$$

$$1\,\text{JND}_{\text{ILD}} = 1\,\text{dB} \tag{6.8}$$

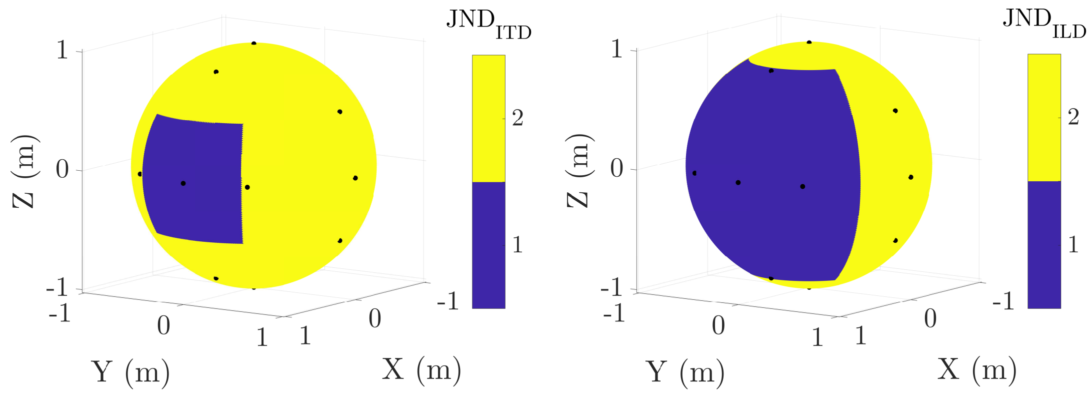
These values can be used to describe the smallest perceivable change at other directions around the sphere. For example, if the smallest perceivable change in ITD at a given direction is 80 µs, the JND can be described as $5\,\text{JND}_{\text{ITD}}$. However, the rates at which the perceivable changes increase for ITD and ILD are different. In the horizontal plane, for a sound source at $\theta = 30°$ the JND for a change in ITD is approximately twice the minimum change ($2\,\text{JND}_{\text{ITD}}$), and for a sound source at $\theta = 90°$ the threshold can be up to $5\,\text{JND}_{\text{ITD}}$ [116]. However, the JND for a change in ILD for a sound source at $\theta = 30°$ is less than $2\,\text{JND}_{\text{ILD}}$; this occurs for directions closer to $\theta = 60°$, with the threshold reaching 4 to 6 JND at $\theta = 90°$ [119, 120].

To relate the calculated ΔITD and ΔILD to their respective perceptual impact, the values in µs and dB, respectively, were mapped to the estimated JND in multiples of

1 $\mathrm{JND}_{\mathrm{ITD}}$ and 1 $\mathrm{JND}_{\mathrm{ILD}}$ as a function of source direction. Due to a lack of detailed information in the literature on how the relationships between the binaural cues and their associated perceptual thresholds vary with direction, a simplistic approach was taken. Also for simplicity, the same value was used for both horizontal and vertical directions as a worst-case measure; vertical JND thresholds in reality are always larger than horizontal JND thresholds for the same angular increase [111].

A two-level perceptual model was devised. For frontal directions within 30° of straight ahead in both azimuth ($0° \leq \theta \leq 30°$ and $330° \leq \theta \leq 359°$) and in elevation ($-30° \leq \phi \leq 30°$), the JND for ΔITD was set to 1 $\mathrm{JND}_{\mathrm{ITD}}$ (16 µs). For source directions outside of that range the JND was set to 2 $\mathrm{JND}_{\mathrm{ITD}}$ (32 µs). Similarly, for ΔILD, the JND was set to 1 $\mathrm{JND}_{\mathrm{ILD}}$ (1 dB) for frontal directions within 60° of straight ahead in both azimuth ($0° \leq \theta \leq 60°$ and $300° \leq \theta \leq 359°$) and in elevation ($-30° \leq \phi \leq 30°$). For source directions outside of this range the JND was set to 2 $\mathrm{JND}_{\mathrm{ILD}}$ (2 dB). This was then used to map numerical difference to estimated perceptual impact, creating a two-level perceptual model. As an example, a ΔITD of 50 µs at $\theta = 0°$, $\phi = 0°$ is mapped to a perceptual impact of 3.1 JND. For a source direction at $\theta = 90°$, $\phi = 0°$, the same numerical difference is mapped to 1.6 JND.

These two-level mappings are shown in Fig. 6.9. Purple indicates the regions of 1 $\mathrm{JND}_{\mathrm{ITD}}$ and 1 $\mathrm{JND}_{\mathrm{ILD}}$, and yellow indicates the regions of 2 $\mathrm{JND}_{\mathrm{ITD}}$ and 1 $\mathrm{JND}_{\mathrm{ILD}}$. The reference HRIR pair directions are indicated as black dots. The JND mappings used for each reference HRIR pair individually are listed in Table 6.4.



(a) ΔITD: 1 $\mathrm{JND}_{\mathrm{ITD}}$ within 30° of straight ahead (b) ΔILD: 1 $\mathrm{JND}_{\mathrm{ILD}}$ within 60° of straight ahead

Figure 6.9: JND thresholds used to map binaural cue differences to perceptual impact for ITD and ILD. Reference HRIR pair directions are indicated as black dots.

Source			JND threshold			
Label #	θ (°)	ϕ (°)	ΔITD (µs)	$\mathrm{JND_{ITD}}$	ΔILD (dB)	$\mathrm{JND_{ILD}}$
1	90	32	32	2	2	2
2	270	32	32	2	2	2
3	90	-32	32	2	2	2
4	270	-32	32	2	2	2
5	32	0	32	2	1	1
6	148	0	32	2	2	2
7	328	0	32	2	1	1
8	212	0	32	2	2	2
9	0	58	32	2	1	1
10	0	-58	32	2	1	1
11	180	58	32	2	2	2
12	180	-58	32	2	2	2
13	0	90	32	2	2	2
14	0	-90	32	2	2	2
15	0	0	16	1	1	1
16	180	0	32	2	2	2
17	90	0	32	2	2	2
18	270	0	32	2	2	2

Table 6.4: JND thresholds used to map the binaural cue differences to perceptual impact in both native and perceptual units for each of the 18 reference HRIR pairs.

This conservative approach imposes more stringent conditions on achieving perceptual transparency than might be typical in reality. That is, $2\,\mathrm{JND_{ITD}}$ and $2\,\mathrm{JND_{ILD}}$ are likely to represent perceptual thresholds which are lower than the actual perceivable change, particularly at the greater absolute angles of azimuth and elevation. However, since variations in JND thresholds with direction have not been reported in sufficient detail, they cannot be accurately accounted for. Preliminary work examining the impact of using a more complex perceptual model is presented in Section 8.2.2.

Fig. 6.10 and Fig. 6.11 show cue differences in JND for all 18 reference HRIR pairs for the same loudspeaker pair as shown in Fig. 6.7 and Fig. 6.8 (left loudspeaker at $\theta = 1°,\ \phi = 0°,\ r = 1.5\,\mathrm{m}$). Once again, $\theta = 0°,\ \phi = 0°$ is at the centre of the figure, spot colour represents the magnitude and the numerical values correspond to the values for N in Table 6.1. For ΔITD, directions on the median plane show differences below the threshold of perception, ranging from 0.16 to 0.57 JND. At lateral directions, the differences increase to a maximum of 23.42 JND. For ΔILD, all differences are above the threshold of perception (median plane differences have a range of 1.09 to 2.58 JND), with lateral position differences rising to a maximum of 11.00 JND. As previously discussed, statistical metrics are used to describe the overall performance of the loudspeaker pair. Table 6.5 lists the arithmetic mean, standard deviation and range metrics calculated for the binaural cues for this loudspeaker pair: all are larger than the threshold of perception.

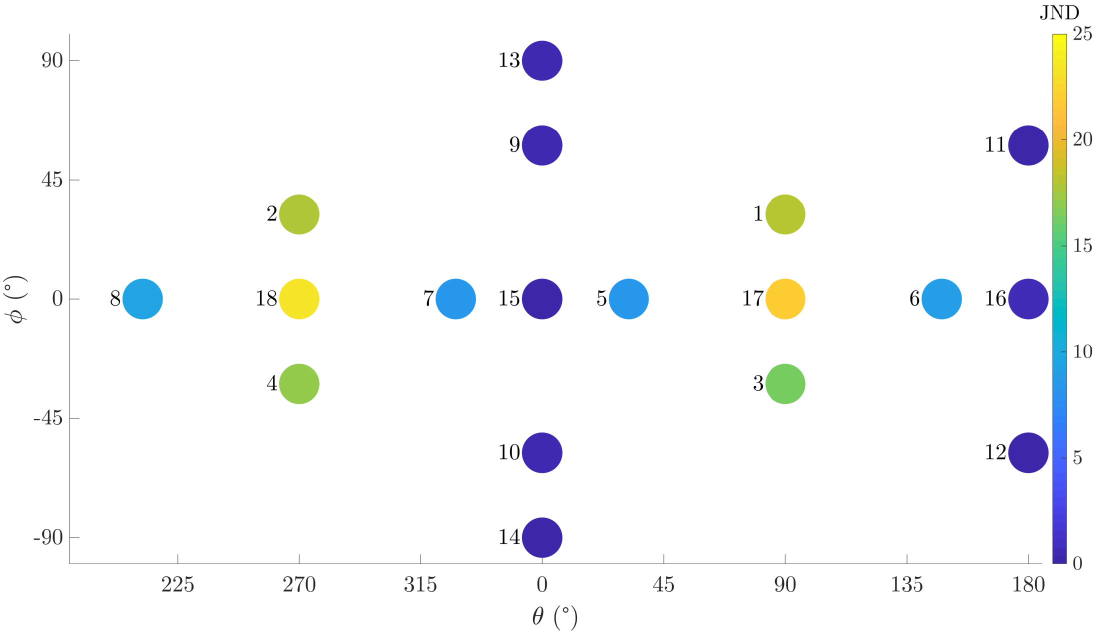


Figure 6.10: ΔITD in estimated JND (spot colour) between each reference and modified HRIR pair for a loudspeaker pair where the left is at $\theta = 1°$, $\phi = 0°$, $r = 1.5\,\mathrm{m}$. Values correspond to N in Table 6.1.

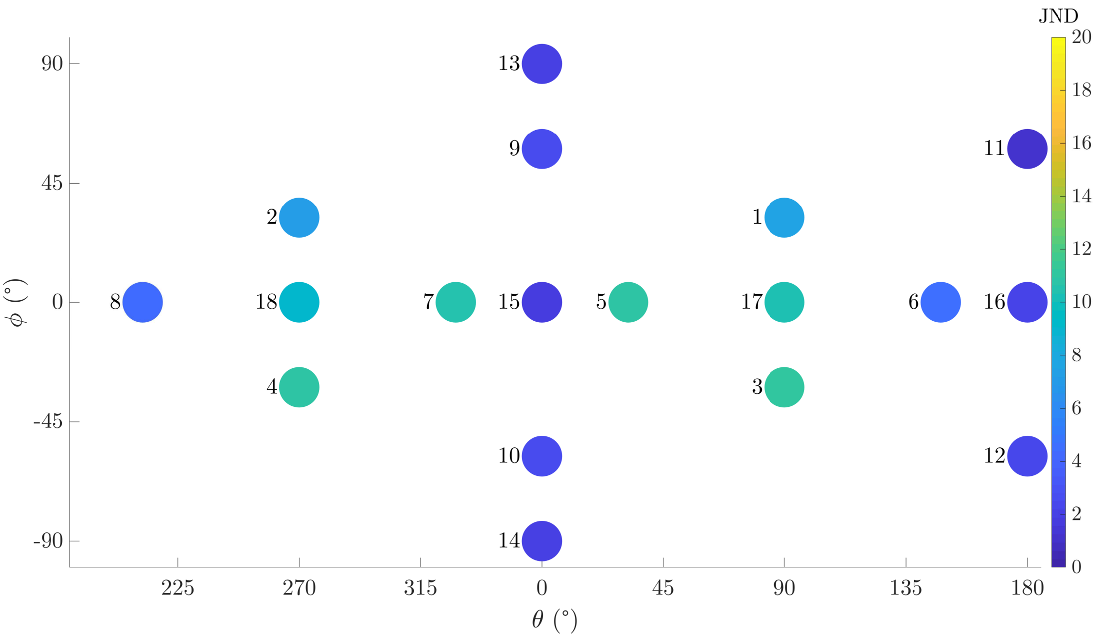


Figure 6.11: ΔILD in estimated JND (spot colour) between each reference and modified HRIR pair for a loudspeaker pair where the left is at $\theta = 1°$, $\phi = 0°$, $r = 1.5\,\mathrm{m}$. Values correspond to N in Table 6.1.

Cue	Arithmetic Mean (JND)	Standard Deviation (JND)	Range (JND)
ΔITD	8.45	8.60	23.26
ΔILD	5.68	3.89	9.91

Table 6.5: Statistical metrics in perceptual units describing the performance of a loudspeaker pair where the left is at $\theta = 1°,\ \phi = 0°,\ r = 1.5\,\mathrm{m}$ over all 18 reference HRIRs.

As a comparison, Table 6.6 lists the statistical metrics for a loudspeaker pair at the same elevation angle of $\phi = 0°$, but at azimuth angles of $\theta = 90°,\ 270°$ and a radial distance $r = 0.15\,\mathrm{m}$. The values in this case are much smaller and are below the threshold of perceptibility of 1 JND. This suggests that all reference HRIRs are reproduced using this loudspeaker pair in a perceptually transparent manner.

Cue	Arithmetic Mean (JND)	Standard Deviation (JND)	Range (JND)
ΔITD	0.19	0.19	0.85
ΔILD	0.34	0.22	0.66

Table 6.6: Statistical metrics in perceptual units describing the performance of loudspeaker pair where the left is at $\theta = 90°,\ \phi = 0°,\ r = 0.15\,\mathrm{m}$ over all 18 reference HRIRs.

These six statistical metrics can be used to express the binaural reproduction performance of every loudspeaker pair. Fig. 6.12 and Fig. 6.13 show the values of these metrics for all loudspeaker pairs at three radial distances: $r = 0.15\,\mathrm{m},\ 0.5\,\mathrm{m},\ 1.5\,\mathrm{m}$. The perceptual degradation (indicated by rising JND values) is apparent with both increasing radial distance and with varying azimuth and elevation angles.

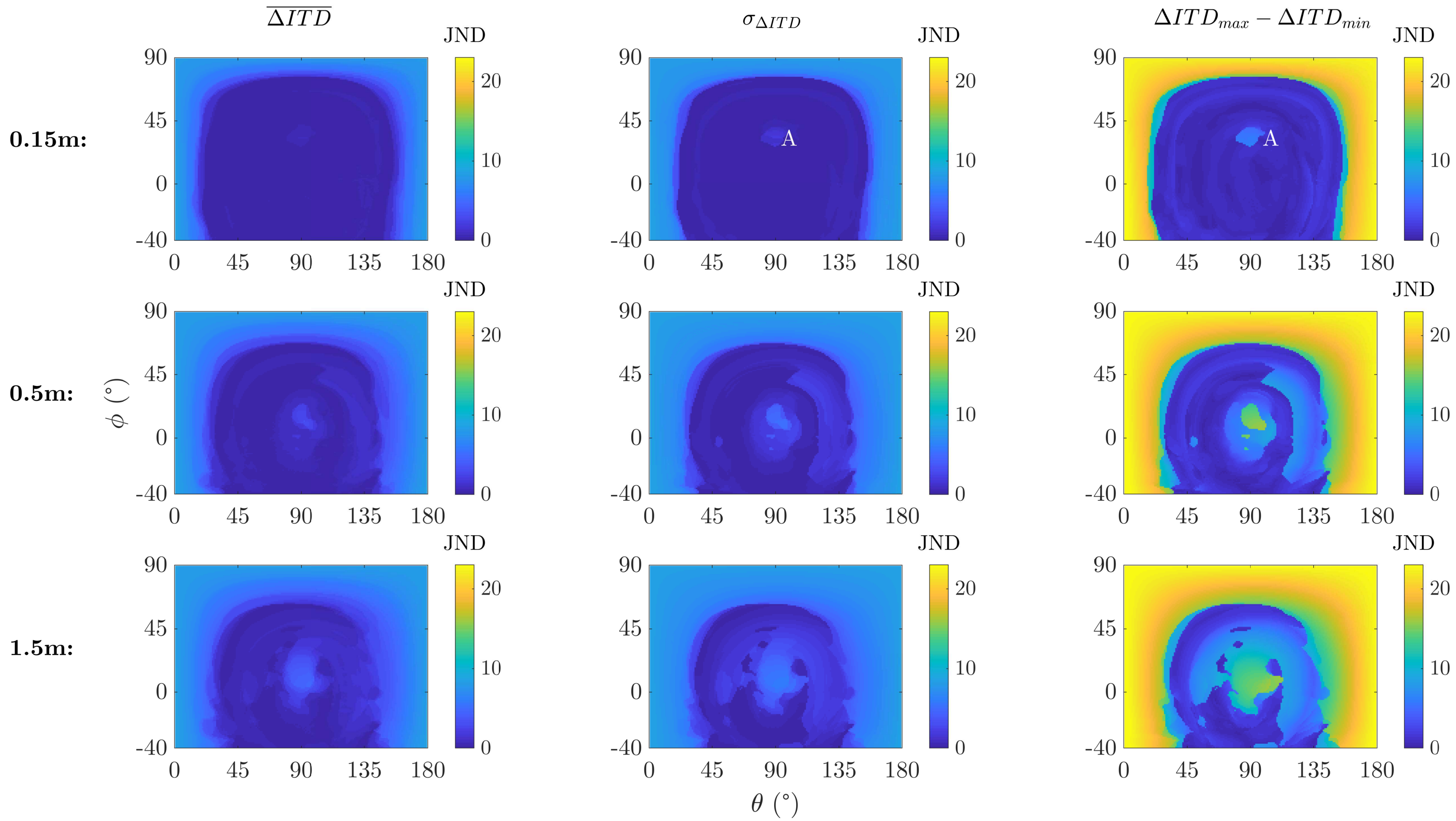


Figure 6.12: Arithmetic mean (left), standard deviation (centre) and range (right) values for ΔITD for three radial distances: $r = 0.15\,\mathrm{m}$ (top), 0.5 m (centre), 1.5 m (bottom). Perceptual deterioration generally increases as a function of radial distance, and varies markedly as a function of direction.

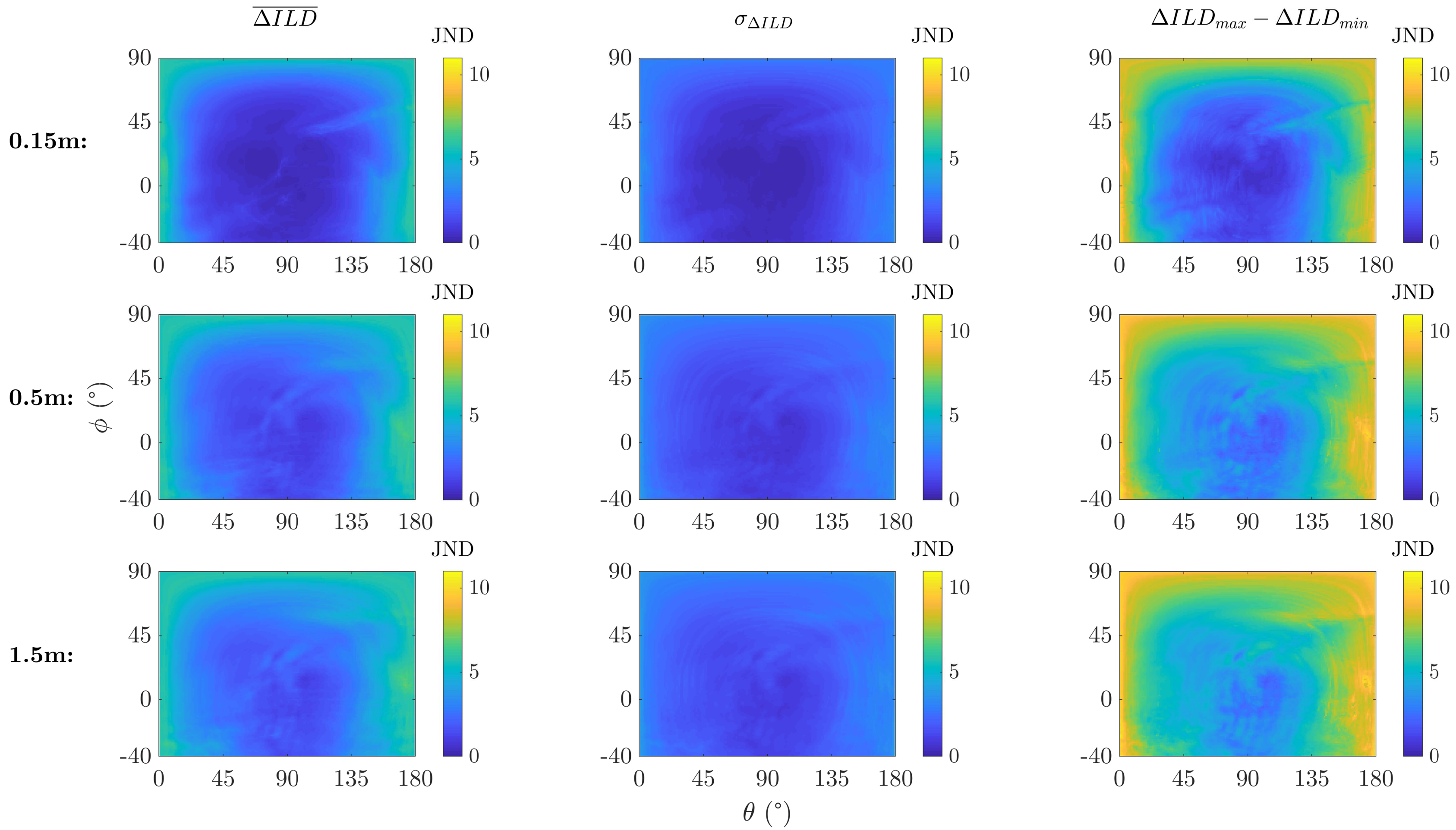


Figure 6.13: Arithmetic mean (left), standard deviation (centre) and range (right) values for ΔILD for three radial distances: $r = 0.15\,\mathrm{m}$ (top), $0.5\,\mathrm{m}$ (centre), $1.5\,\mathrm{m}$ (bottom). Perceptual deterioration generally increases as a function of radial distance, and varies as a function of direction.

The two 'range' metrics (right-most column) are generally larger than the corresponding arithmetic mean or standard deviation for any given loudspeaker pair, with maximum values of 23.29 JND and 10.55 JND for ΔITD and ΔILD, respectively. This is due to the phenomenon discussed in relation to Fig. 6.5 and Fig. 6.6, where the median plane reference HRIR pairs are reproduced with substantially lower differences than those at lateral directions, therefore producing a large range.

An anomalous region is present in the top row figures of Fig. 6.12 ($r = 0.15\,\mathrm{m}$). The small internal region of larger JND values, exemplified in the standard deviation and range plots (labelled with 'A'), is a result of the ITD estimation method. The maximum peak within the interaural cross-correlation (IACC) function with multiple local maxima suddenly switches to a different maximum and therefore creates a discontinuity in ITD value over a small angular change. This is demonstrated in Fig. 6.14, which shows the IACC functions between left and right signals for three HRIR pairs: reference HRIR pair $\mathbf{h^4}$ ($\theta = 270°$, $\phi = -32°$), and the modified HRIR pairs $\mathbf{\hat{h}^4_{11637}}$ (loudspeaker pair where the left loudspeaker is at $\theta = 90°$, $\phi = 26°$, $r = 0.15\,\mathrm{m}$) and $\mathbf{\hat{h}^4_{11638}}$ (1° higher in elevation than $\mathbf{\hat{h}^4_{11637}}$). $\mathbf{\hat{h}^4_{11638}}$ is just inside the anomalous region; $\mathbf{\hat{h}^4_{11637}}$ is just outside it.

The sample number at which the maximum value of IACC occurs (that which is used to determine the ITD) is 12,243 for $\mathbf{h^4}$, and 12,308 and 13,243 for $\mathbf{\hat{h}^4_{11637}}$ and $\mathbf{\hat{h}^4_{11638}}$, respectively. This leads to respective ΔITD values of 13.5 µs and 208.3 µs. This, therefore, accounts for the observed discontinuities in both the standard deviation and the range metrics for this loudspeaker pair. Without further investigation, which is outside the scope of this work, it is not possible to say whether this is a flaw in the ITD estimation method or whether the difference can really be perceived.

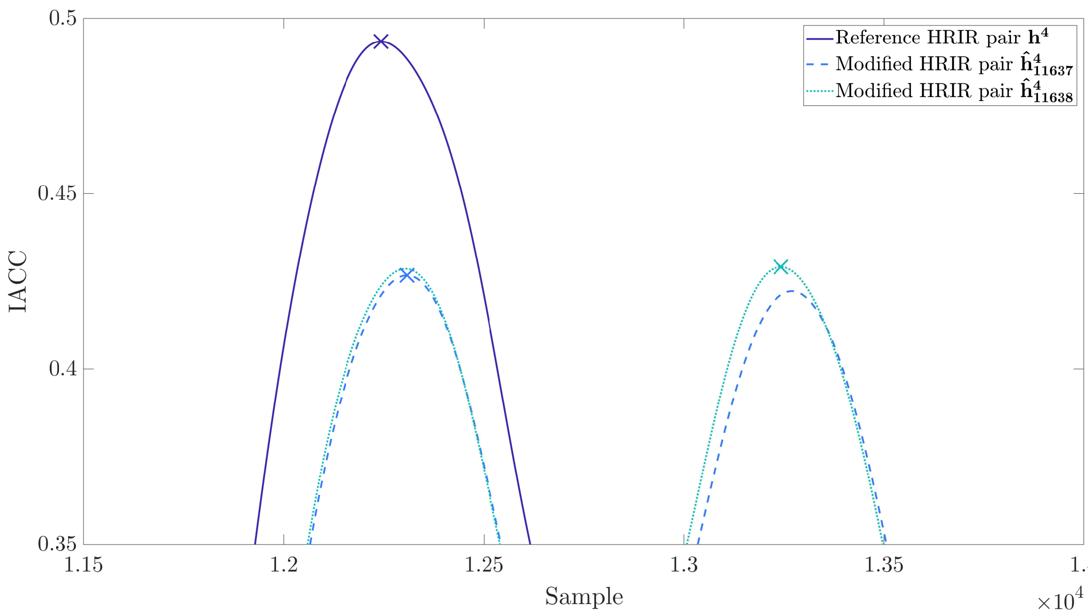
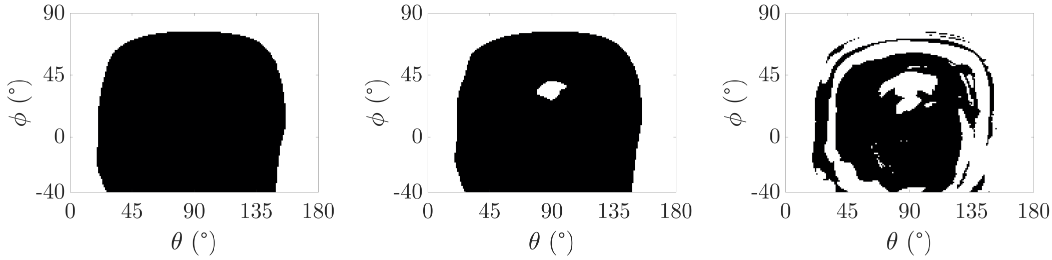


Figure 6.14: Peaks in the IACC functions between left and right signals for three HRIR pairs: reference HRIR pair $\mathbf{h^4}$, and modified HRIR pairs $\mathbf{\hat{h}^4_{11637}}$ and $\mathbf{\hat{h}^4_{11638}}$. The jump between local maxima (indicated with crosses) creates a jump in ITD value for a small angular change.

6.3.2 Regions of Perceptually-Transparent Performance

Whilst perceptual deterioration of up to 23.29 JND for ΔITD and 10.55 JND for ΔILD exist, there are regions within each plot where the degradation is below the estimated threshold of perception. Fig. 6.15 shows the regions under 1 JND (indicated in black) for arithmetic mean, standard deviation and range for ΔITD values at $r = 0.15$ m. More positions where loudspeaker pairs exhibit deterioration below this perceptual threshold exist for arithmetic mean and standard deviation than for range. The discontinuities shown in Fig. 6.15c are likely a result of the same direction-dependent phenomenon seen previously, where a change in the relative magnitudes of two IACC peaks creates an abrupt change in the calculated ITD.



(a) Arithmetic mean (b) Standard deviation (c) Range

Figure 6.15: Regions where statistical metrics for ΔITD have a value of less than the estimated threshold of perception at $r = 0.15$ m (indicated in black).

As discussed previously, these statistical metrics each describe a different aspect of the performance of the loudspeaker pair. By considering the set of loudspeaker pair positions for which these metrics simultaneously have a value of less than 1 JND, the loudspeaker pairs which reproduce ITD in a perceptually-transparent way for all reference HRIR pairs can be determined.

The set $\mathbf{LP^{KEMAR}}$ has been previously defined in Equation 6.2 as:

$$\mathbf{LP^{KEMAR}} = \{(v, w) \mid v, w \in \mathbb{R}^3\} \tag{6.9}$$

where $\mathbf{LP^{KEMAR}}$ is a set of pairs of vectors v and w, such that v and w are each a three-dimensional, real-valued column vector, related through being mirrored in the median plane. Using set notation, a function $f : \mathbf{LP^{KEMAR}} \to \mathbb{R}^3$ can be defined such that, if $lp \in \mathbf{LP^{KEMAR}}$ where lp is a pair:

$$f(lp) = (\overline{\Delta ITD},\ \sigma_{\Delta ITD},\ \Delta ITD_{max} - \Delta ITD_{min}) \tag{6.10}$$

where the function operating on each member lp in $\mathbf{LP^{KEMAR}}$ maps the pairs of vectors to the three real-valued metrics and can be thought of as analogous to the calculations which produce those metrics. A relation $\prec$ can be defined over $\mathbb{R}^3$ such that:

$$a \prec b \iff \forall i\ a_i < b_i \tag{6.11}$$

which creates a set of pairs where elements are paired together if and only if each element of a is strictly less than the corresponding element of b. This can be extended to define the set $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{ITD}}$ (where all three metrics are less than one) as follows, using the function $f(lp)$ as a in the relation $\prec$ and a comparison vector $\mathbf{1} = (1, 1, 1)$ as b in the relation $\prec$:

$$\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{ITD}} = \left\{ lp \in \mathbf{LP}^{\mathbf{KEMAR}} \mid f(lp) \prec \mathbf{1} \right\} \tag{6.12}$$

There are 39,361 $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{ITD}}$ loudspeaker pair positions across all radial distances, shown in Fig. 6.16 in black and listed in Table 6.7. 9,030 (38.8 % of all $\mathbf{LP}^{\mathbf{KEMAR}}$ positions at that radial distance) exist at the smallest radial distance $r = 0.15$ m (Fig. 6.16a). These positions occupy the majority of the region bounded by $35° < \theta < 140°$, $-40° < \phi < 60°$. The number of $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{ITD}}$ loudspeaker pair positions within $\mathbf{LP}^{\mathbf{KEMAR}}$ decreases with increasing radial distance and drops to 898 at the largest radial distance (Fig. 6.16n). A somewhat elliptical pattern of variation is apparent. As radial distance increases, a hole develops in the vicinity of the interaural axis and increases in size, whilst the region behind the interaural axis decreases in size.

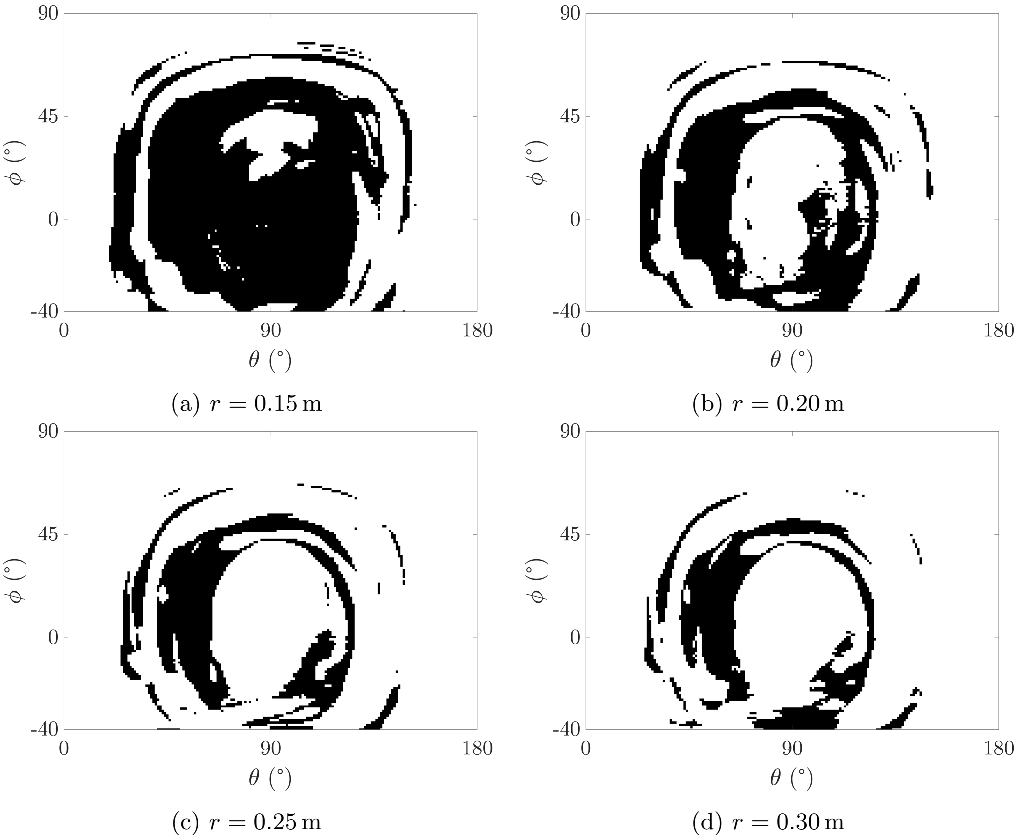


Figure 6.16: Regions of loudspeaker pair positions for which the estimated perceptual differences are less than 1 JND for ΔITD average, standard deviation and range simultaneously ($\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{ITD}}$, indicated in black) at radial distances 0.15 to 0.30 m. *Figure continued overleaf.*

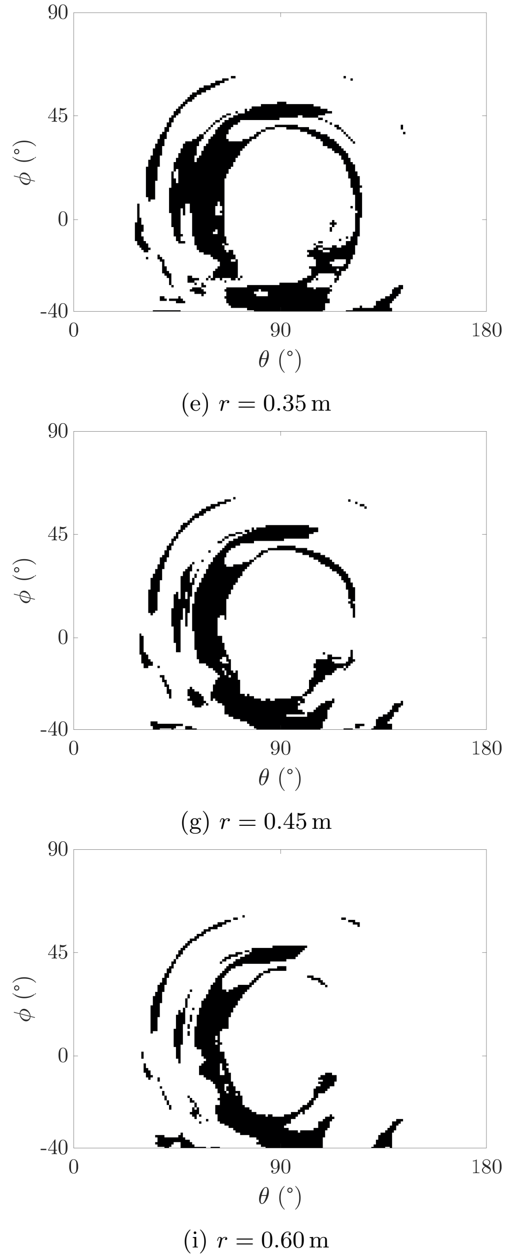


Figure 6.16: Regions of loudspeaker pair positions for which the estimated perceptual differences are less than 1 JND for ΔITD average, standard deviation and range simultaneously ($\mathbf{LP^{KEMAR}_{ITD}}$, indicated in black) at radial distances 0.35 to 0.70 m. *Figure continued overleaf.*

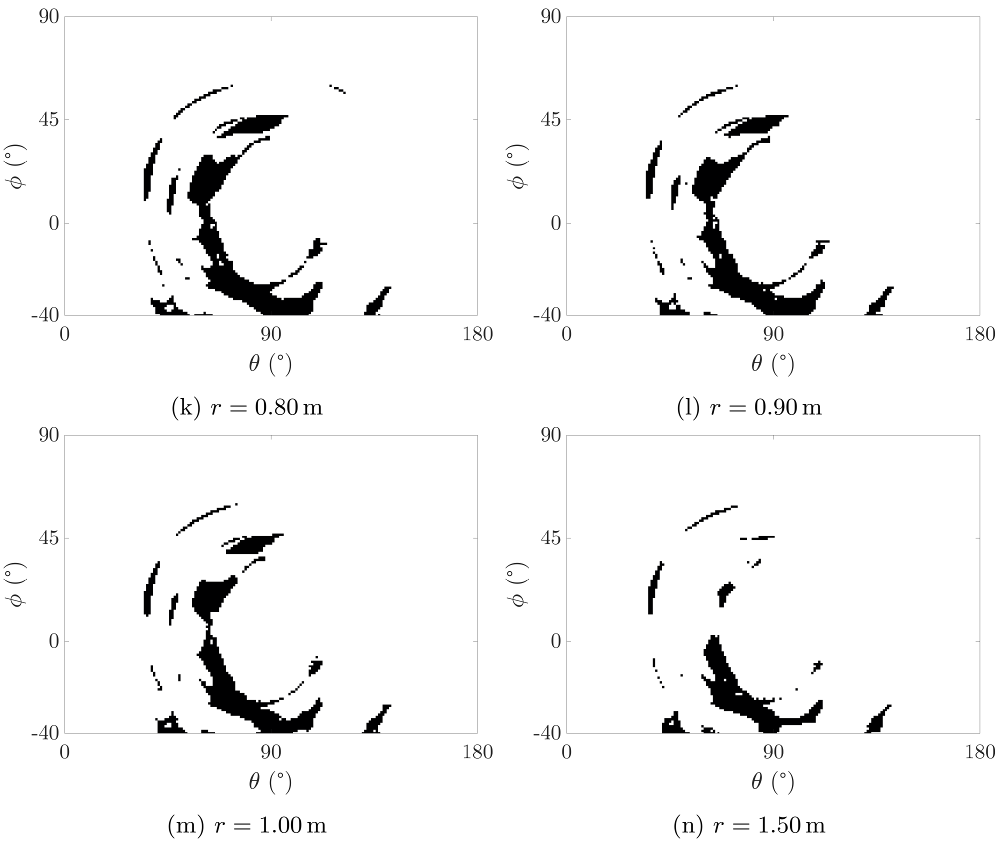


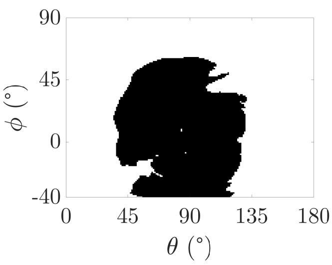
Figure 6.16: Regions of loudspeaker pair positions for which the estimated perceptual differences are less than 1 JND for ΔITD average, standard deviation and range simultaneously ($\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{ITD}}$, indicated in black) at radial distances 0.80 to 1.50 m. 39,361 such pairs exist in total, with 9,030 at $r = 0.15\,\mathrm{m}$, and reducing with increasing radial distance to 898 at $r = 1.5\,\mathrm{m}$.

r (m)	$\mathbf{LP^{KEMAR}_{ITD}}$ Number	%
0.15	9,030	38.8
0.20	5,164	22.2
0.25	3,470	14.9
0.30	2,997	12.9
0.35	2,783	12.0
0.40	2,565	11.0
0.45	2,348	10.1
0.50	2,143	9.2
0.60	1,902	8.2
0.70	1,710	7.3
0.80	1,546	6.6
0.90	1,464	6.3
1.00	1,341	5.8
1.50	898	3.9

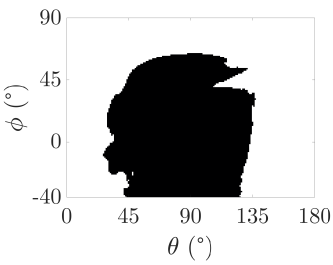
Table 6.7: Number and percentage of $\mathbf{LP^{KEMAR}_{ITD}}$ loudspeaker pairs within $\mathbf{LP^{KEMAR}}$ at each radial distance.

The number of loudspeaker pair positions in $\mathbf{LP^{KEMAR}_{ITD}}$ is, perhaps, surprisingly high. This is likely due to low ΔITD values calculated when the ITD of the reference HRIR pair is similar to that between the intended and crosstalk paths for the loudspeakers within the loudspeaker pair. In this situation, the crosstalk signal from one loudspeaker arrives at the same time as the intended signal from the other, resulting in a smaller impact on the timing information due to the relatively low amplitude.

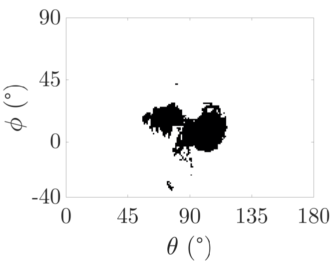
A similar analysis can be applied to the ΔILD metrics to determine the positions where loudspeaker pairs reproduce ILD in a perceptually-transparent way for all reference HRIR pairs. The regions where each individual statistical metric has a value below the estimated threshold of perception is shown in Fig. 6.17 in black. As with ΔITD, more positions meet the criterion for arithmetic mean and standard deviation than for range.



(a) Arithmetic mean



(b) Standard deviation



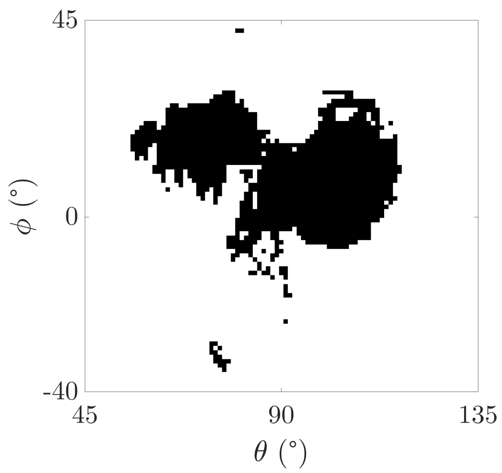
(c) Range

Figure 6.17: Regions where statistical metrics for ΔILD have a value of less than the estimated threshold of perception at $r = 0.15\,\mathrm{m}$ (indicated in black).

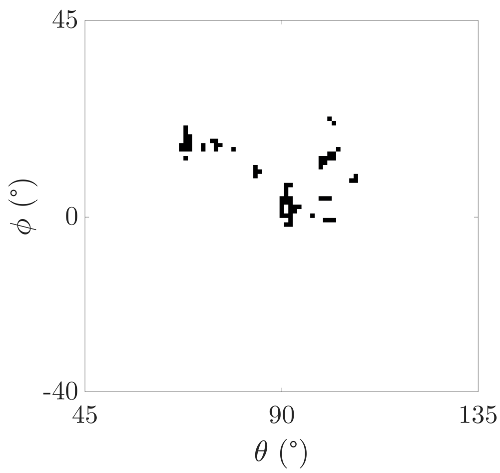
The subset of $\mathbf{LP^{KEMAR}}$ positions where the arithmetic mean, standard deviation and range of ΔILD simultaneously have a value less than 1 JND is referred to as $\mathbf{LP^{KEMAR}_{ILD}}$. An expression for $\mathbf{LP^{KEMAR}_{ILD}}$ can be derived following the steps presented for $\mathbf{LP^{KEMAR}_{ITD}}$ in Equation 6.9 to Equation 6.12:

$$\mathbf{LP^{KEMAR}_{ILD}} = \left\{ lp \in \mathbf{LP^{KEMAR}} \mid f(lp) \prec \mathbf{1} \right\} \tag{6.13}$$

There are far fewer loudspeaker pair positions in $\mathbf{LP^{KEMAR}_{ILD}}$ (shown in Fig. 6.18) than there are in $\mathbf{LP^{KEMAR}_{ITD}}$. Compared with the 9,030 $\mathbf{LP^{KEMAR}_{ITD}}$ positions at $r = 0.15\,\mathrm{m}$, only 1,360 $\mathbf{LP^{KEMAR}_{ILD}}$ positions exist at $r = 0.15\,\mathrm{m}$. These lie within the region bounded by $56° < \theta < 117°$, $-35° < \phi < 43°$, although the majority are within the region bounded by $56° < \theta < 117°$, $-18° < \phi < 29°$. Another 70 exist at the next radial distance of $r = 0.20\,\mathrm{m}$, and these are scattered within the region bounded by $67° < \theta < 107°$, $-2° < \phi < 22°$. No positions exist in $\mathbf{LP^{KEMAR}_{ILD}}$ loudspeaker pairs exist at radial distances beyond $r = 0.20\,\mathrm{m}$.



(a) $r = 0.15\,\mathrm{m}$



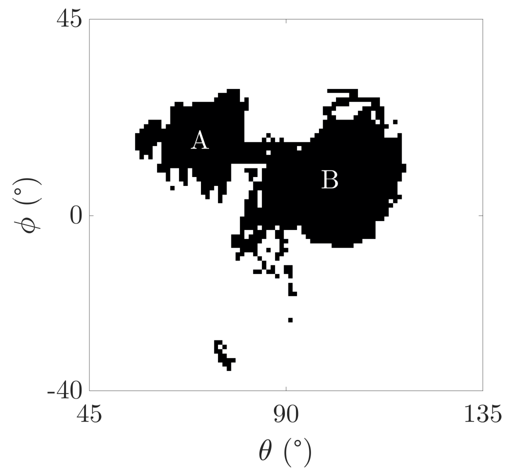
(b) $r = 0.20\,\mathrm{m}$

Figure 6.18: Regions of loudspeaker pair positions for which the estimated perceptual differences are less than 1 JND for ΔILD average, standard deviation and range simultaneously ($\mathbf{LP^{KEMAR}_{ILD}}$, indicated in black). 1,360 such positions exist at $r = 0.15\,\mathrm{m}$ and 70 such positions exist at $r = 0.20\,\mathrm{m}$.

The subsets $\mathbf{LP^{KEMAR}_{ITD}}$ and $\mathbf{LP^{KEMAR}_{ILD}}$ are clearly different, so it is important to identify any loudspeaker pair positions that they have in common. Thus, $\mathbf{LP^{KEMAR}_{CM}}$ is defined to be the intersection of $\mathbf{LP^{KEMAR}_{ITD}}$ and $\mathbf{LP^{KEMAR}_{ILD}}$:

$$\mathbf{LP^{KEMAR}_{CM}} = \mathbf{LP^{KEMAR}_{ITD}} \ \cap \ \mathbf{LP^{KEMAR}_{ILD}} \tag{6.14}$$

$\mathbf{LP^{KEMAR}_{CM}}$ contains the loudspeaker pair positions with an estimated perceptual difference between all reference HRIR pairs and the corresponding modified HRIR pairs of less than 1 JND for all six criteria simultaneously. The positions of these loudspeaker pairs are shown in Fig. 6.19.



(a) $r = 0.15\,\mathrm{m}$

(b) $r = 0.20\,\mathrm{m}$

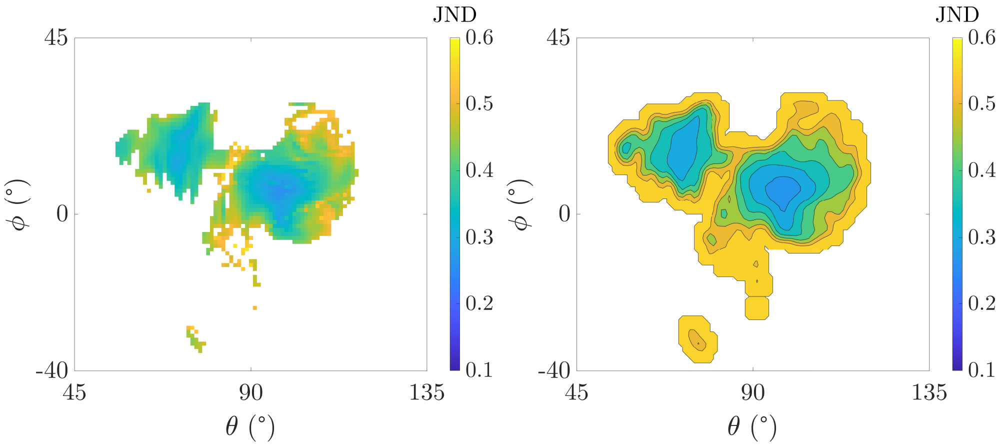
Figure 6.19: loudspeaker pair positions for which the estimated perceptual differences are less than 1 JND simultaneously for the metrics: average, standard deviation and range, for both ΔITD and ΔILD ($\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}}$, indicated in black). 1,314 such pairs exist at $r = 0.15\,\mathrm{m}$ in two almost disconnected subsets (labelled A and B). 11 pairs exist at $r = 0.20\,\mathrm{m}$.

1,314 loudspeaker pair positions exist in $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}}$ at $r = 0.15\,\mathrm{m}$ (Fig. 6.19a), in the region bounded by $56° < \theta < 117°$, $-35° < \phi < 28°$ with the majority at an elevation angle above $\phi = -7°$. Two almost disconnected subsets (indicated by A and B) are apparent, one centred about approximately $\theta = 70°$, $\phi = 15°$, and a larger section centred about $\theta = 100°$, $\phi = 10°$. At the second radial distance of $r = 0.20\,\mathrm{m}$ (Fig. 6.19b), the 11 positions are more sparsely distributed and do not form a distinct region. It can be inferred that the range metric of the ΔILD values is the dominant metric among the six statistical metrics used, as the resulting shape of $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}}$ closely matches that shown in Fig. 6.17c.

6.3.2.1 Region of best performance

All $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}}$ loudspeaker pair positions are associated with high binaural reproduction performance without the need for CTC. However, it is also of interest to determine whether a region of best performance exists within this set. As an indication of performance across all metrics, the overall arithmetic mean of the six statistical metrics (M_6) was calculated for each $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}}$ loudspeaker pair. Fig. 6.20a shows the intensity surface for M_6 for radial distance $r = 0.15\,\mathrm{m}$. The global minimum and maximum values are 0.26 JND and 0.57 JND, respectively. The pattern suggests that two regions of better performance may exist within $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}}$; one centred at approximately $\theta = 70°$, $\phi = 15°$ and one centred at approximately $\theta = 100°$, $\phi = 5°$. At the radial distance $r = 0.20\,\mathrm{m}$ (not plotted here), results are broadly similar across the 11 values with a range of 0.48 to 0.55 JND.

A contour plot divides the values into bands, and so indicates the regions of best performance more clearly. Fig. 6.20b shows the contour plot of the same data using steps of 0.05 JND, where a 2D Gaussian smoothing kernel ($\sigma = 1$) has been applied prior to plotting using the 'imgaussfilt' MATLAB function [286] to provide more even contours.



(a) Intensity surface (b) Contour with 2D Gaussian smoothing

Figure 6.20: Arithmetic mean of the six statistical metrics (M_6) for the loudspeaker pair positions in $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}}$ at $r = 0.15\,\mathrm{m}$, plotted as (a) an intensity surface and (b) a contour plot.

The single region of lowest value is at $M_6 = 0.3\,\mathrm{JND}$. The centroid for this region is at $\theta = 98°$, $\phi = 6°$, with the contour occupying the majority of the region bounded by $93° \leq \theta \leq 103°$, $1° \leq \phi \leq 10°$. $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}}$ positions within this contour are referred to as the set $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{best}}$:

$$\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{best}} = \left\{ lp \in \mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}} \mid f(tl) < 0.3 \right\} \tag{6.15}$$

where the function $f : \mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}} \to \mathbb{R}$ is defined such that, if $lp \in \mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}}$ where lp is a pair:

$$f(lp) = \overline{x} \tag{6.16}$$

and is analogous to the calculation of M_6 for that loudspeaker pair position. The next lowest contour in Fig. 6.20b is 0.35 JND. It comprises two regions with centroid values at $\theta = 72°$, $\phi = 17°$ and $\theta = 97°$, $\phi = 6°$, respectively. The latter centroid value is very close to the centroid of the contour for 0.3 JND. The two portions 'A' and 'B' and the join between them, as discussed in relation to Fig. 6.19, are also more clearly shown in this plot.

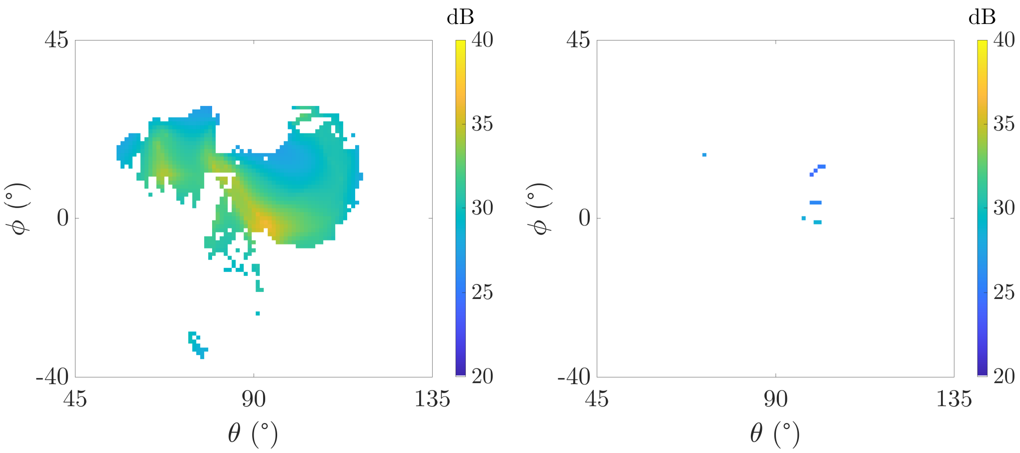
6.4 Comparisons of NCS and Performance Predictions

From the results presented above, it can be concluded that positions exist within the set $\mathbf{LP}^{\mathbf{KEMAR}}$ where loudspeaker pairs are capable of faithfully reproducing binaural signals using CTC-less binaural reproduction. This set of loudspeaker pair positions, referred to as $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}}$, predominantly exists close to the interaural axis and close to the head.

This thesis is founded on the notion that the presence of sufficient NCS for a left-right pair of loudspeakers makes satisfactory binaural reproduction possible without the additional complication of incorporating CTC. The validity of this notion can be tested by comparing regions identified as possessing high NCS (Section 5.4) with regions where loudspeaker pair placement is predicted to achieve perceptually faithful reproduction of binaural cues (Section 6.3.2). Masiero [227] suggested that CS may only be an indicator of localisation performance under certain circumstances, but those results are based on a method where the CS of a single loudspeaker pair was varied using individualised and non-individualised CTC, rather than the approach taken here of a continuous range of CS (or in this case, NCS) values obtained through varying loudspeaker placement.

6.4.1 NCS at Loudspeaker Pair Positions in $\mathbf{LP}_{\mathbf{CM}}^{\mathbf{KEMAR}}$

Fig. 6.21 shows NCS for each of the $\mathbf{LP}_{\mathbf{CM}}^{\mathbf{KEMAR}}$ loudspeaker pair positions, calculated as the arithmetic mean of the NCS values for the two $\mathbf{SP}^{\mathbf{KEMAR}}$ positions within each pair. The NCS values calculated in Chapter 5 were used for the left hemisphere $\mathbf{SP}^{\mathbf{KEMAR}}$ positions. As the right hemisphere $\mathbf{SP}^{\mathbf{KEMAR}}$ positions were time- and level-aligned as part of the VRS, therefore potentially altering their NCS, the NCS values for those loudspeaker positions were recalculated.



(a) $r = 0.15\,\mathrm{m}$ (b) $r = 0.20\,\mathrm{m}$

Figure 6.21: NCS for each $\mathbf{LP}_{\mathbf{CM}}^{\mathbf{KEMAR}}$ loudspeaker pair position. All values exceed the 20 dB separation reported as necessary for binaural reproduction [17, 228].

NCS values at the $\mathbf{LP}_{\mathbf{CM}}^{\mathbf{KEMAR}}$ loudspeaker pair positions range from 26.8 to 36.2 dB at $r = 0.15\,\mathrm{m}$, and from 24.4 to 28.6 dB at $r = 0.20\,\mathrm{m}$. The largest values at $r = 0.15\,\mathrm{m}$ appear around $\theta = 93°$, $\phi = -2°$. All NCS values at $\mathbf{LP}_{\mathbf{CM}}^{\mathbf{KEMAR}}$ positions are larger than the 20 dB minimum separation reported as necessary to maintain binaural accuracy [17, 228]. However, based on the analysis approaches in this thesis, it is difficult to quantify the benefits that substantially larger values of NCS might yield in terms of improved accuracy of binaural reproduction. Whilst this would be an interesting avenue to pursue, it is outside the scope of this work.

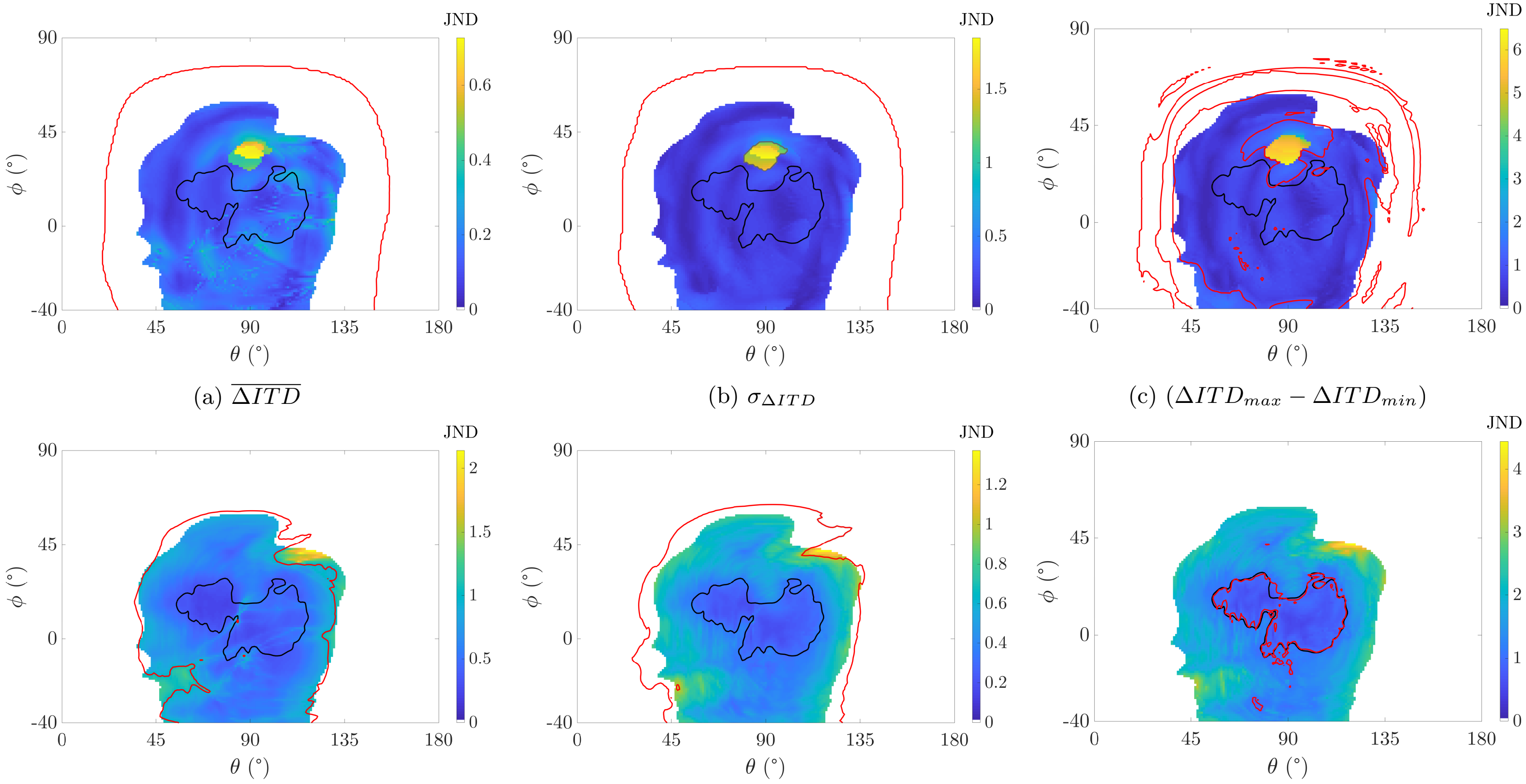
6.4.2 Relationships between $\mathbf{LP}_{\mathbf{CM}}^{\mathbf{KEMAR}}$, $\mathbf{SP}_{\mathbf{20}}^{\mathbf{KEMAR}}$ and 1JND

It is also of interest to examine the relationship between the set of loudspeaker positions determined as exhibiting high enough NCS for binaural reproduction ($\mathbf{SP}_{\mathbf{20}}^{\mathbf{KEMAR}}$, see Section 5.4) and the set $\mathbf{LP}_{\mathbf{CM}}^{\mathbf{KEMAR}}$. Some loudspeaker positions within the set $\mathbf{SP}_{\mathbf{20}}^{\mathbf{KEMAR}}$, when paired and used in the VRS, do not exist in the set $\mathbf{LP}_{\mathbf{CM}}^{\mathbf{KEMAR}}$. To examine why this might be the case, arithmetic mean, standard deviation and range values for ΔITD and ΔILD for the $\mathbf{SP}_{\mathbf{20}}^{\mathbf{KEMAR}}$ positions used in the VRS at the closest radial distance of $r = 0.15\,\mathrm{m}$ were isolated and plotted using the colour scale in Fig. 6.22. For comparison purposes, the boundary of the $\mathbf{LP}_{\mathbf{CM}}^{\mathbf{KEMAR}}$ set is indicated in black and the boundary of the 1 JND region for each statistical metric is indicated in red. These red contours indicate the outer edge of the black regions plotted in Fig. 6.15 and Fig. 6.17 for the top and bottom rows, respectively.

Comparing the colour regions ($\mathbf{SP}_{\mathbf{20}}^{\mathbf{KEMAR}}$) with the black contours ($\mathbf{LP}_{\mathbf{CM}}^{\mathbf{KEMAR}}$), a large proportion (83.4 %) of the loudspeaker pairs formed from $\mathbf{SP}_{\mathbf{20}}^{\mathbf{KEMAR}}$ positions are not in $\mathbf{LP}_{\mathbf{CM}}^{\mathbf{KEMAR}}$, that is, they have a colour value but lie outside of the black contours in Fig. 6.22. It may be that using mapping values of $2\,\mathrm{JND}_{\mathrm{ITD}}$ and $2\,\mathrm{JND}_{\mathrm{ILD}}$ at the most lateral source directions (rather than a larger, less cautious, and perhaps more accurate JND threshold), produces an excessively high estimate for the perceptual impact that the loudspeaker positions have on reproduction of these lateral sources. This would exclude loudspeaker positions from the set $\mathbf{LP}_{\mathbf{CM}}^{\mathbf{KEMAR}}$ when they may, to a human listener, have perceptually acceptable performance. The implications of this are examined further in Section 8.2.2.

Comparing the colour regions with the red contours (that is, the statistical metric values calculated for $\mathbf{SP}_{\mathbf{20}}^{\mathbf{KEMAR}}$ positions and the 1 JND region of each statistical metric), three observations can be made. First, for the ΔITD metrics (Figs. 6.22a–6.22c), the 1 JND contour is predominantly outside the region of $\mathbf{SP}_{\mathbf{20}}^{\mathbf{KEMAR}}$ positions. Second, for the arithmetic mean and standard deviation of ΔILD (Fig. 6.22d and Fig. 6.22e, respectively), the boundary of the $\mathbf{SP}_{\mathbf{20}}^{\mathbf{KEMAR}}$ region roughly aligns with the 1 JND boundary. Third, for ΔILD range (Fig. 6.22f), the 1 JND boundary is considerably within the region of $\mathbf{SP}_{\mathbf{20}}^{\mathbf{KEMAR}}$ positions, and aligns with the boundary of the set $\mathbf{LP}_{\mathbf{CM}}^{\mathbf{KEMAR}}$ almost exactly.

These observations are perhaps unsurprising, based on previous discussions. The fact that the 1 JND contour for ΔITD metrics (Figs. 6.22a–6.22c) contains more loudspeaker pair positions than exist in $\mathbf{SP}_{\mathbf{20}}^{\mathbf{KEMAR}}$ is most likely a result of the phenomenon discussed in Section 6.3.2. A low deterioration in ITD is calculated for some loudspeaker pairs due to similarity between the reference HRIR pair and the ipsilateral and contralateral acoustic paths. NCS is calculated in the frequency domain, and so the alignment between $\mathbf{SP}_{\mathbf{20}}^{\mathbf{KEMAR}}$ positions and the 1 JND contour in Fig. 6.22d and Fig. 6.22e is encouraging. The smaller number of positions within the 1 JND contour for ΔILD range (Fig. 6.22f) is most likely a result of an overly-conservative perceptual model.



(a) $\overline{\Delta ITD}$

(b) $\sigma_{\Delta ITD}$

(c) $(\Delta ITD_{max} - \Delta ITD_{min})$

(d) $\overline{\Delta ILD}$

(e) $\sigma_{\Delta ILD}$

(f) $(\Delta ILD_{max} - \Delta ILD_{min})$

Figure 6.22: Arithmetic mean (left column), standard deviation (centre column) and range (right column) values for ΔITD (top row) and ΔILD (bottom row), calculated using $\mathbf{SP_{20}^{KEMAR}}$ positions at $r = 0.15\,\text{m}$ in pairs in the VRS. Different colour limits have been used between figures so as to not lose detail at the lower values. For comparison, the boundary of the set $\mathbf{LP_{CM}^{KEMAR}}$ is indicated in black and the boundary of the 1 JND region as shown in Fig. 6.15 is indicated in red.

6.4.3 Relationship between Regions of Best Performance

Regions of largest NCS and best CTC-less binaural reproduction performance have been identified. Fig. 6.23 presents both sets of results for comparison purposes. NCS values at the loudspeaker pair positions in the set $\mathbf{LP}_{\mathbf{CM}}^{\mathbf{KEMAR}}$ (as shown in Fig. 6.21a) are shown using the dB colour scale. The contour encapsulating positions where loudspeaker pairs perform CTC-less binaural reproduction best (where $M_6 \leq 0.3\,$JND, $\mathbf{LP}_{\mathbf{best}}^{\mathbf{KEMAR}}$) is imposed on this in black. Within this contour, NCS has a range of 29.4 to 33.2 dB. Contours for the next highest performance boundary (corresponding to $M_6 \leq$0.35 JND) are indicated in red. NCS has a range of 28.5 to 35.5 dB within these two regions. The region of largest NCS value does not coincide completely with the region of best binaural performance. Nevertheless, there are similarities, most notably the existence of two distinct regions: one in front of and above the interaural axis; the other slightly behind it.

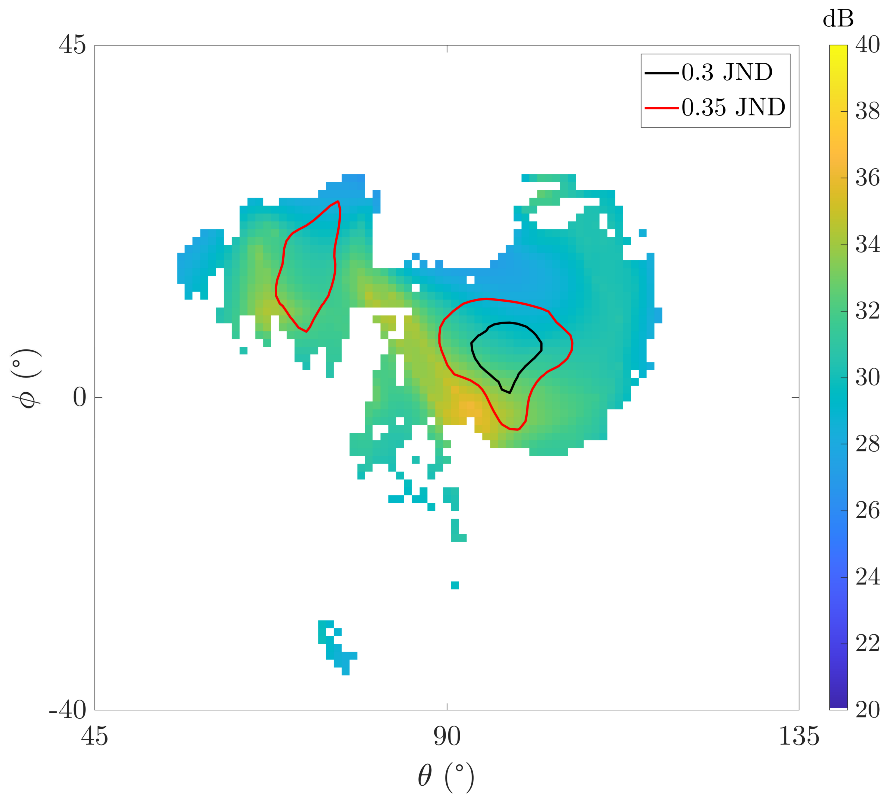


Figure 6.23: NCS for $\mathbf{LP}_{\mathbf{CM}}^{\mathbf{KEMAR}}$ loudspeaker pair positions at $r = 0.15\,$m, with the regions of best ($M_6 \leq 0.3\,$JND) and next best ($M_6 \leq 0.35\,$JND) CTC-less binaural performance bounded in black and red, respectively. The best-performing regions in the two sets of data do not coincide exactly, but they share common features.

One reason for the observed differences may be the frequency range used to calculate NCS. To enable comparison with previous literature, NCS has so far been calculated over the frequency range 200 Hz to 8 kHz. As the inclusion of higher frequencies can lead to larger values of NCS at a given loudspeaker position (as demonstrated in Section 5.5.1), and as this increase is non-linear with respect to the increase in frequency range used, it is possible that calculation of NCS including frequencies up to 16 kHz would produce a distribution of NCS which better aligns with the region of best CTC-less binaural reproduction performance. Table 6.8 lists the Spearman's rank correlation coefficients[2] calculated between NCS and the six statistical binaural metrics for each loudspeaker pair position within $\mathbf{LP_{CM}^{KEMAR}}$, plus the arithmetic mean of all six metrics (M_6), all at $r = 0.15$ m. Correlation values for NCS calculated from 200 Hz to 8 kHz $\left(NCS^{8k}\right)$ and from 200 Hz to 16 kHz $\left(NCS^{16k}\right)$ are shown. The significance value is $p < 0.01$ for all pairs. For radial distance $r = 0.20$ m, the values range from -0.23 to 0.25 ($0.46 \leq p \leq 0.88$) and -0.6 to 0.25 ($0.04 \leq p \leq 0.9$) for NCS^{8k} and NCS^{16k}, respectively.

Metric	Correlation NCS^{8k}	Correlation NCS^{16k}
$\overline{\Delta ITD}$	0.17	-0.23
$\sigma_{\Delta ITD}$	0.14	-0.30
$\Delta ITD_{max} - \Delta ITD_{min}$	0.20	-0.22
$\overline{\Delta ILD}$	-0.42	-0.63
$\sigma_{\Delta ILD}$	-0.55	-0.84
$\Delta ILD_{max} - \Delta ILD_{min}$	-0.36	-0.69
M_6	-0.14	-0.65

Table 6.8: Spearman's rank correlation coefficients between NCS, calculated using frequency ranges 200 Hz to 8 kHz (NCS^{8k}) and 200 Hz to 16 kHz (NCS^{16k}), and each of the listed statistical metrics for the $\mathbf{LP_{CM}^{KEMAR}}$ loudspeaker pair positions. The correlations are weak for NCS^{8k} but stronger for NCS^{16k}.

The values indicate weak correlation between any of the pairs when calculating NCS using the 8 kHz frequency limit, however, the correlation appears to be stronger when including frequencies up to 16 kHz. NCS calculated using this frequency range for $\mathbf{LP_{CM}^{KEMAR}}$ loudspeaker pair positions at $r = 0.15$ m is shown in Fig. 6.24. The best-performing regions, whilst again not coinciding exactly, appear to align more so than the regions in Fig. 6.23. This suggests that NCS calculated using a larger frequency range more closely aligns with perceptual performance, particularly (and, perhaps, not surprisingly) for the spectral metrics.

[2] Spearman's rank correlation was used as Kolmogorov-Smirnov tests indicated that the data was not normally distributed and, therefore, that calculation of Pearson correlation coefficients was not appropriate.

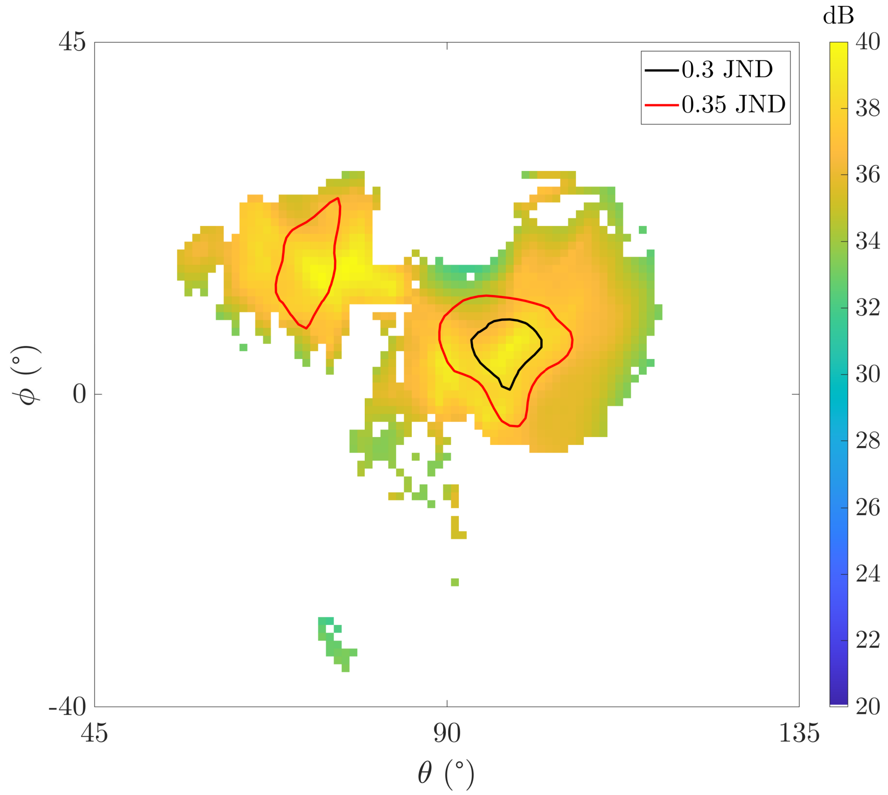


Figure 6.24: NCS^{16k} for $\mathbf{LP}_{\mathbf{CM}}^{\mathbf{KEMAR}}$ loudspeaker pair positions at $r = 0.15$ m, with the regions of best ($M_6 \leq 0.3$ JND) and next best ($M_6 \leq 0.35$ JND) CTC-less binaural performance bounded in black and red, respectively. The best-performing regions in the two sets of data appear to align more than in Fig 6.23.

There are, additionally, unavoidable differences between the idealised simulation test environment presented here and the real world. These findings are indicative of the existence of a region where CTC may not be required for binaural reproduction over a near-field loudspeaker pair. However, it is likely that real-world mechanical and acoustical influences will have a significant impact on results in ways that are not accounted for within this study. These aspects are discussed further in Section 8.2.1.

6.5 Conclusion

This chapter presents an analysis of the impact of loudspeaker position on the binaural reproduction of 18 reference HRIR pairs. A VRS based on KEMAR has been implemented to simulate 325,780 pairs of loudspeakers, with their positions referred to as the set $\mathbf{LP}^{\mathbf{KEMAR}}$. Modified HRIR pairs, corresponding to the reproduction of each reference HRIR pair through each loudspeaker pair under test, have been synthesised using the VRS. The VRS consists of three stages: ipsilateral inversion (to account for the acoustic path between each loudspeaker and the ipsilateral ear), system reproduction (reproduction over the loudspeaker pair without a CTC stage), and summation of the intended and crosstalk paths to create the overall signal at each ear. 18 HRIRs from the SADIE v1 database [123] have been used as inputs to the VRS to create 5,864,040

(325,780 × 18) modified HRIR pairs. For brevity, loudspeaker pairs and bounding angular regions are referred to using the angle(s) in the left hemisphere.

The ITD and ILD were estimated for each reference HRIR pair and for the corresponding modified HRIR pair for all $\mathbf{LP}^{\mathbf{KEMAR}}$ loudspeaker pair positions. The unsigned differences between the two sets of binaural cues (ΔITD and ΔILD, respectively) were calculated to determine the deterioration in binaural cues as a result of reproduction using each loudspeaker pair. The differences for frontal and median plane HRIR pairs are generally low, regardless of loudspeaker pair, though large differences are present (up to 750.4 µs and 23.5 dB) for more lateral and elevated reference HRIR pairs for a large number of loudspeaker pairs. As the binaural cue JND thresholds vary with source direction, the ΔITD and ΔILD values are mapped to JND to enable discussion of the perceptual impact of loudspeaker placement.

To compare the performance of all loudspeaker pairs across all 18 reference HRIR pairs, the arithmetic mean, standard deviation and range (collectively referred to as the statistical metrics) of the perceptually-mapped ΔITD and ΔILD values were calculated for each loudspeaker pair. Loudspeaker pairs where the statistical metrics for ΔITD and ΔILD have values below the threshold of perception for all reference HRIRs were determined. These are first presented independently and then in combination, to identify the positions of loudspeaker pairs where the statistical metrics for both interaural binaural cues are simultaneously below the threshold of perception.

39,361 loudspeaker pair positions exist in the subset $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{ITD}}$, for which all three ΔITD statistical metrics simultaneously have values below 1 JND. The number of qualifying loudspeaker pairs at each radial distance decreases with increasing distance, but pairs exist across all radial distances with a peak of 9,030 at $r = 0.15$ m and reducing to 898 at $r = 1.5$ m. At the radial distance of $r = 0.15$ m, $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{ITD}}$ loudspeaker pair positions exist within the region bounded by $35° < \theta < 140°,\ -40° < \phi < 60°$. With increasing radial distance, two subregions progressively decrease in size, one close to the interaural axis and the other behind the interaural axis.

1,430 loudspeaker pair positions exist in the subset $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{ILD}}$, for which all three ΔILD statistical metrics simultaneously have values below 1 JND. Such positions exist only at the two smallest radial distances, with 1,360 at $r = 0.15$ m within the region bounded by $56° < \theta < 117°,\ -35° < \phi < 43°$, and 70 at $r = 0.20$ m scattered within the region bounded by $67° < \theta < 107°,\ -2° < \phi < 22°$.

As $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{ITD}}$ and $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{ILD}}$ are different sets of loudspeaker pair positions, those common to both subsets, $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}}$, are of special interest. 1,314 $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}}$ positions exist at $r = 0.15$ m, in the region bounded by $56° < \theta < 117°,\ -35° < \phi < 28°$. The majority lie at elevation angles above $\phi = -7°$, where one large region is present. At the second radial distance of $r = 0.20$ m, the 11 positions are more sparsely distributed and do not form a distinct region.

An estimate for the possible region of best performance within the set of $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}}$ loudspeaker pair positions was determined through computation of the arithmetic mean across all six perceptually-mapped statistical metrics for each loudspeaker pair. This data was used to generate contours in the azimuth-elevation plane in steps of 0.05 JND. The global minimum is 0.26 JND, and the region contained by the lowest contour value (0.3 JND) occupies the majority of the region bounded by $93° \leq \theta \leq 103°$, $1° \leq \phi \leq 10°$.

A channel separation of 20 dB is considered the minimum value for maintaining binaural accuracy [17, 228]. Loudspeaker pair positions satisfying this criterion through NCS (without the use of CTC) were determined in Chapter 5 and placed in the set $\mathbf{SP_{20}^{KEMAR}}$. To determine whether NCS is indeed a reliable indicator of perceptual performance, the mean NCS value for the left-right pair of $\mathbf{SP^{KEMAR}}$ positions associated with each member of the set $\mathbf{LP_{CM}^{KEMAR}}$ was calculated. All NCS values for $\mathbf{LP_{CM}^{KEMAR}}$ loudspeaker pair positions are greater than 20 dB. It was found that some loudspeaker positions exist in the set $\mathbf{SP_{20}^{KEMAR}}$ which are not present in the set $\mathbf{LP_{CM}^{KEMAR}}$, and the indicated regions of best performance within $\mathbf{LP_{CM}^{KEMAR}}$ and $\mathbf{SP_{20}^{KEMAR}}$ do not precisely coincide. It may be the case that, while acoustic shadowing (the principal process for the generation of NCS) provides sufficient channel separation to preserve the binaural cues, the method of calculation (using a frequency range of 200 Hz to 8 kHz) does not map simply to perception. Increasing the frequency range used in calculation of NCS appears to align the best-performing regions more closely. In addition, the use of a conservative constant JND threshold value outside the frontal region may have indicated that the deterioration in cues for some source directions was above the threshold of perception when, in fact, it is not.

The findings presented in this chapter suggest that positions exist near the head where pairs of loudspeakers can reproduce binaural signals in a perceptually-transparent way without requiring the use of CTC. These findings are based on the use of a basic virtual reproduction system and a rudimentary perceptual model. Furthermore, they are specific to KEMAR, and the existence of suitable positions of loudspeaker pairs for other subjects is of crucial importance. The extension of the VRS to accommodate the varying morphology of the general population would allow conclusions to be drawn as to whether a binaural reproduction system without CTC might be rolled out successfully in consumer applications. The development of a VRS to use simulation data from different subjects, and the subsequent analysis to determine candidate regions for loudspeaker placement with regards to varying subject morphology, are discussed in Chapter 7.

Chapter 7

Simulation of the Perceptual Performance of Candidate Loudspeaker Pairs with respect to SYMARE subjects

All I wanna get is a little bit closer.

Tegan and Sara, *Closer*
Heartthrob (2013)

7.1 Introduction

The primary objective of this thesis is to determine whether any pairs of positions exist each side of a listener where loudspeakers will faithfully reproduce binaural signals without the need for crosstalk cancellation (CTC). To be of practical use, such pairs of positions should lead to acceptable binaural reproduction for most listeners. Using a rudimentary perceptual model, the positions of loudspeaker pairs which exhibit deterioration in interaural binaural cues of less than the threshold of perception were identified in Chapter 6 for the Knowles Electronics Manikin for Acoustic Research (KEMAR) (referred to as the set $\mathbf{LP}_{\mathbf{CM}}^{\mathbf{KEMAR}}$). It is, therefore, of interest to determine whether such sets of positions for loudspeaker pairs exist for subjects other than KEMAR. If so, it is of additional interest to know whether these sets have any loudspeaker pair positions in common.

This chapter describes the process of adapting the methods and analyses developed in Chapter 6 to investigate the existence of common loudspeaker pair positions across subjects. Although the number of subjects used in this analysis is relatively small, the existence of any common loudspeaker positions would indicate the viability of a commercial implementation of binaural reproduction system without CTC.

The virtual reproduction system (VRS) described in Section 6.2.2 was modified to allow investigation of binaural cue consistency for listeners other than KEMAR.

15 subjects from the Sydney-York Morphological and Recording of Ears (SYMARE) database [80] were chosen. This number was chosen to encompass a reasonable variation in morphological features without extending the computation time excessively. Head-related transfer functions (HRTFs) and the corresponding head-related impulse responses (HRIRs) were simulated for each SYMARE subject. The HRIRs were used in the VRS as the acoustic paths and, for each subject in turn, the deterioration in interaural binaural cues was calculated between each *reference* HRIR pair and the signal at the ears of the subject generated using each loudspeaker pair (the *modified* HRIR pair)[1].

Using the perceptual model described in Section 6.3.2, a set of positions was identified for each subject for which loudspeaker pairs exhibited deterioration below the threshold of perception between reference and modified HRIR pairs. The sets for each subject were compared to identify any loudspeaker pairs common to all subjects used in the analysis. If overlaps between subjects exist, and if the overlaps are great enough, this would suggest that a binaural reproduction system without CTC might perform well for any listener. Such a simple rendering method could form part of a commercially attractive personal spatial audio solution.

7.2 Methods

To enable the analysis of subjects other than KEMAR, further simulation of HRTFs was required. HRTFs for 15 subjects from the SYMARE database [80] were simulated. The SYMARE database contains BEM-ready meshes of 61 subjects at varying resolutions, both with and without shoulders, in addition to measured and simulated HRTF data. Subjects in the database are numbered in ascending order according to the degree of correlation between the acoustically measured HRTFs and the numerically simulated HRTFs. To ensure that simulated results bear the most resemblance to the real-world counterpart, subjects with the lowest number labels were included. The first 15 subjects were initially selected; this was considered to be a reasonable compromise between ensuring analysis of enough subjects to provide substantial variation in morphological features, and keeping computational requirements and processing time to within acceptable limits. However, during initial testing it was determined that the mesh for subject #5 comprised too many faces to be simulated using the available computational resources. It was not possible to reduce the number of faces without compromising mesh quality, and so subject #5 was removed from the analysis and a further subject (subject #16) added. This left the 15 SYMARE subjects shown in Fig. 7.1: #1 to #4 and #6 to #16. While the SYMARE database contains 48 male and 13 female subjects, the sex of the specific 15 used here is not available.

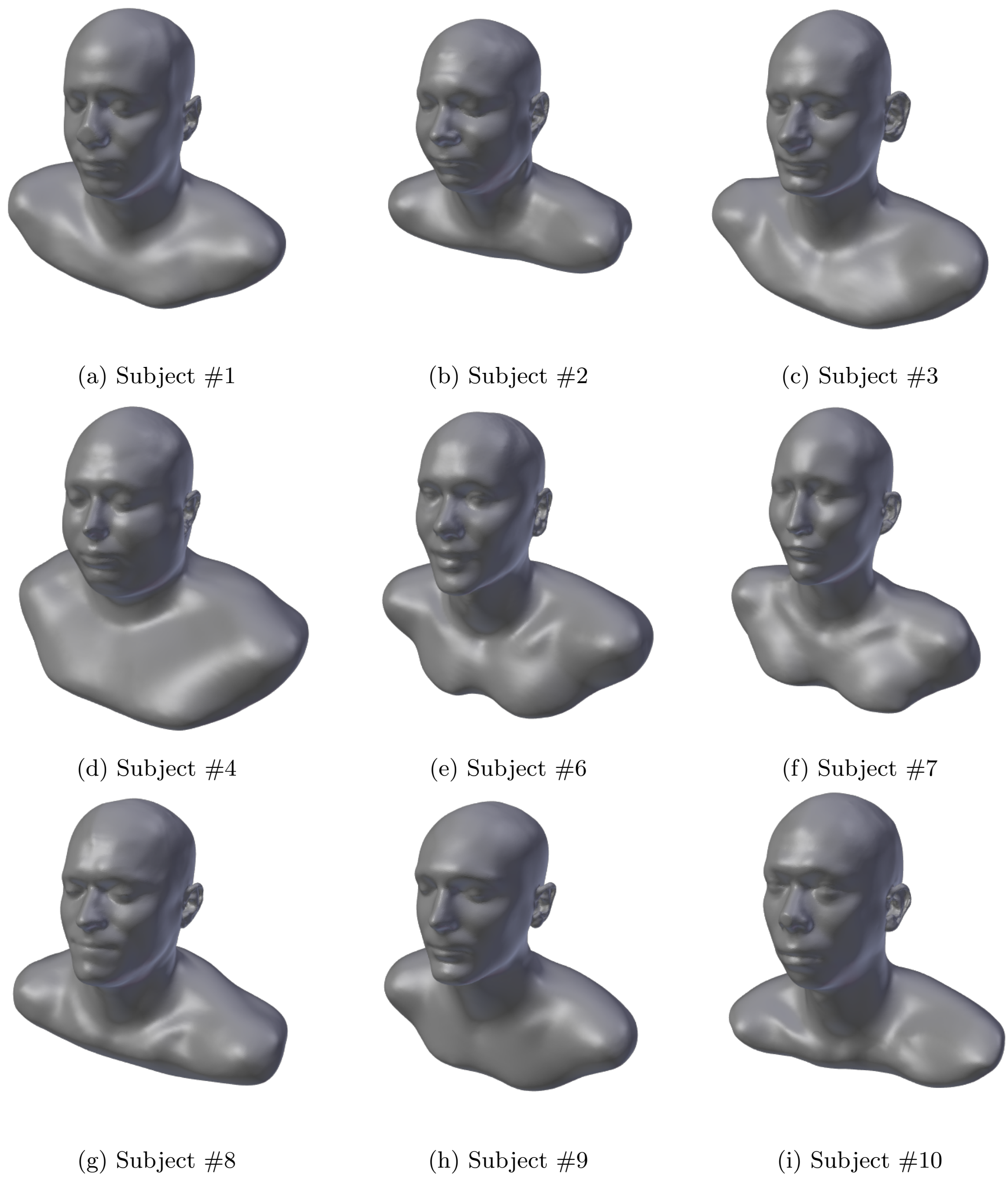
The methods described in Section 6.2 were then run for each SYMARE subject to: a) generate the modified HRIR pairs using a VRS for each loudspeaker pair (where the set of loudspeaker pairs is referred to as the set $\mathbf{LP^{SYMARE}}$), and b) estimate the binaural cues associated with each modified HRIR pair. Analysis using the processes described in Section 6.3 determined three sets of loudspeaker pairs for each SYMARE subject, where the integer S refers to the index of the SYMARE subject:

- $\mathbf{LP^{S}_{ITD}}$: loudspeaker pairs which exhibit deterioration in interaural time difference

[1]More detailed descriptions of the terms *reference* and *modified* HRIR pair are given in Section 6.1.

(ITD) of less than the threshold of perception, defined using the same derivation as that presented for Equation 6.12;

- $\mathbf{LP^{S}_{ILD}}$: loudspeaker pairs which exhibit deterioration in interaural level difference (ILD) of less than the threshold of perception, defined in the same way as Equation 6.13;
- $\mathbf{LP^{S}_{CM}}$: loudspeaker pairs which exhibit deterioration in both ITD and ILD simultaneously of less than the threshold of perception, defined in the same way as Equation 6.14.

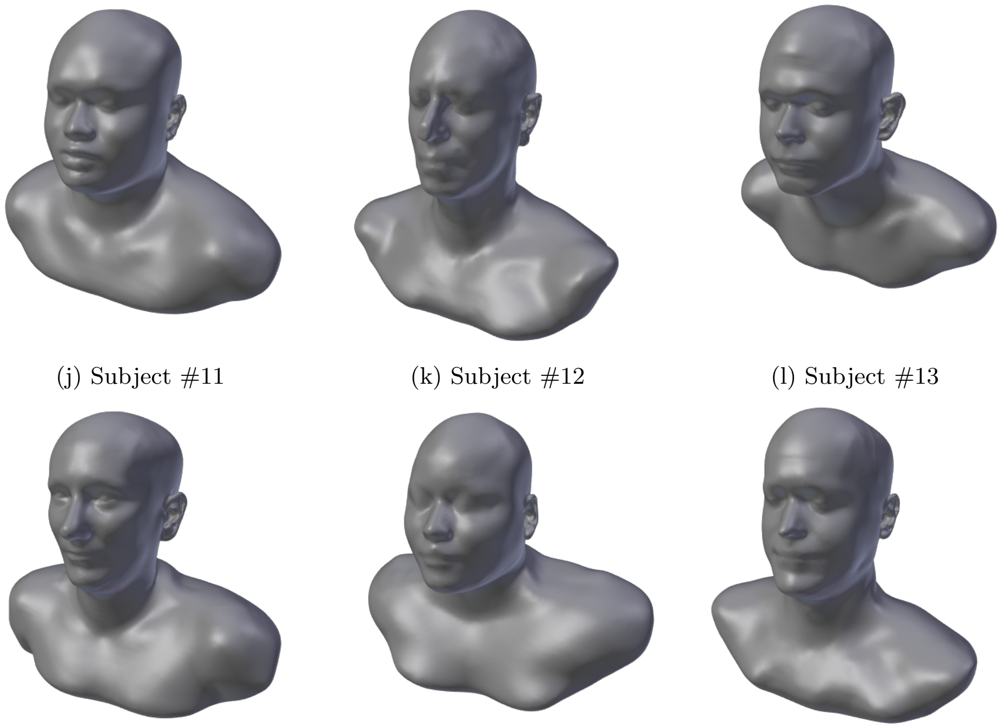


(a) Subject #1 (b) Subject #2 (c) Subject #3

(d) Subject #4 (e) Subject #6 (f) Subject #7

(g) Subject #8 (h) Subject #9 (i) Subject #10

Figure 7.1: The first 9 of the 15 subjects from the SYMARE database [80] used in this work. Note the omission of subject #5 (see main text). *Figure continued overleaf.*



(j) Subject #11 (k) Subject #12 (l) Subject #13

(m) Subject #14 (n) Subject #15 (o) Subject #16

Figure 7.1: The remaining 6 of the 15 subjects from the SYMARE database [80] used in this work. The variation in the morphology of the subjects is apparent.

7.2.1 Simulation of Near-Field HRTFs

HRTFs of the 15 SYMARE subjects were simulated using the ‘HeadTorsoAndEars’ meshes (valid to 16 kHz) and the multi-level fast multipole boundary element method (FM-BEM), as implemented in the open source software Mesh2HRTF [179–181]. As discussed in Section 2.6.2, the BEM requires knowledge of the speed of sound and the density of the medium. To maintain consistency with the simulated HRTFs obtained for KEMAR in Chapter 5, default values were used for the speed of sound and the density of air ($c = 343.0\,\mathrm{m\,s^{-1}}$ and $\rho = 1.21\,\mathrm{kg\,m^{-3}}$, respectively).

The analysis undertaken in Section 6.3 indicated that positions within the set $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}}$ existed only at radial distances of $r < 0.25\,\mathrm{m}$, and at angular positions within 50° of the interaural axis in azimuth (θ) and within 40° of the interaural axis in elevation (ϕ). Therefore, to reduce computational requirements, a reduced set of source positions was defined to include only this most relevant region. The bounds and increment of the source positions within this set, referred to as $\mathbf{SP}^{\mathbf{SYMARE}}$, are listed in Table 7.1 and plotted in full in Fig. 7.2. There are 14,342 source positions at $r = 0.15\,\mathrm{m}$ and 4,182 at $r = 0.20\,\mathrm{m}$, giving a total of 18,524 source positions. This can be written using set notation as:

$$\mathbf{SP}^{\mathbf{SYMARE}} = \{v \mid v \in \mathbb{R}^3\} \tag{7.1}$$

in that $\mathbf{SP}^{\mathbf{SYMARE}}$ is a set with members v, such that each member v is a three-dimensional, real-valued column vector.

r (m)	θ bounds (°)	ϕ bounds (°)	Increment (°)
0.15	40 to 140 220 to 320	-30 to 40	1
0.20	60 to 110 250 to 300	-20 to 20	1

Table 7.1: The 18,524 source positions in the set $\mathbf{SP}^{\mathbf{SYMARE}}$, organised as four partial radial shells, used in the simulations of HRTFs of the SYMARE subjects. There are 14,342 source positions at $r = 0.15\,\text{m}$ and 4,182 source positions at $r = 0.20\,\text{m}$.

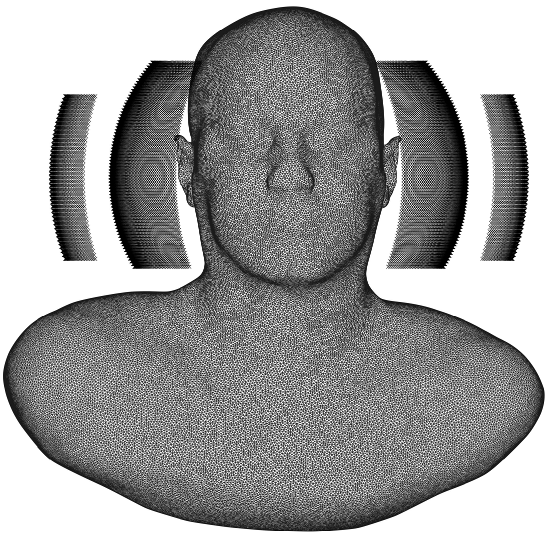


Figure 7.2: The 18,524 source positions in the set $\mathbf{SP}^{\mathbf{SYMARE}}$ used in the simulations of HRTFs of the SYMARE subjects, plotted with the mesh of subject #1.

As in previous chapters, θ increases from zero in the frontal direction counter-clockwise such that $\theta = 90°$ is on the left and $\theta = 270°$ is on the right. ϕ is measured with respect to the horizontal plane containing the interaural axis, where $\phi = 90°$ describes a source directly above the head and $\phi = -90°$ lies directly below. Radial distance r is measured relative to the origin (the mid-point of the interaural axis). The evaluation grid of 18,524 source positions was prepared using Delauney triangulation in MATLAB [270, 271] and the pre-processing steps were performed in Blender [272] to satisfy the requirements of Mesh2HRTF. The vibrating element in each ear of each mesh was defined to be the element through which the interaural axis passed.

The 18,524 HRTF pairs for each SYMARE subject mesh were computed at 159 frequencies in 100 Hz increments from 200 Hz to 16 kHz. The simulations ran consecutively on an Ubuntu 16.04.2 machine (the same machine as discussed in Chapter 3 and Chapter 5) with central processing units (CPUs) clocked at 2.60 GHz sharing 256 GB of random-access memory (RAM).

The number of CPUs used in each simulation was adjusted based on the number of faces in each mesh to ensure that enough RAM was available for all CPU processes to successfully complete. The average number of faces per mesh is 191,874, with the meshes ranging in size from 139,830 to 219,486 faces. Simulations using meshes with more faces require more RAM for each CPU process, and so it is possible for the situation to occur where too many processes are running at once for them all to access the RAM they require. This causes one or more of the CPU processes to be stopped by the machine running the software, and results in a failed simulation.

Initial testing indicated that 16 CPU processes was appropriate for all subjects other than subject #4, which required a reduction to 12. Use of more than 16 CPU processes may have been possible for some of the meshes with fewer faces. However, it was deemed more sensible to let the simulations run with a high likelihood of success, rather than to experiment with higher numbers of CPU processes and risk the simulations failing due to a lack of RAM resources. The longest duration of any process within each simulation (and, therefore, the actual simulation run time, since the processes run concurrently) is listed in Table 7.2, along with the cumulative simulation duration and the average duration of a CPU process.

Subject #	Faces	Processes	Max time (hours)	Cumulative time (hours)	Average per process (hours)
1	184,450	16	8.4	126.5	7.9
2	139,830	16	5.3	78.8	4.9
3	202,226	16	9.5	143.8	9.0
4	219,486	12	14.7	164.6	13.7
6	209,316	16	10.3	155.0	9.7
7	174,188	16	7.7	116.4	7.3
8	182,162	16	8.4	126.2	7.9
9	181,460	16	8.3	125.6	7.9
10	182,362	16	8.4	128.0	8.0
11	192,616	16	9.1	137.7	8.6
12	181,884	16	8.3	126.2	7.9
13	187,014	16	8.9	133.9	8.4
14	189,340	16	9.1	137.7	8.6
15	208,070	16	10.2	153.4	9.6
16	176,286	16	7.9	119.9	7.5

Table 7.2: Number of faces in the mesh for each SYMARE subject, number of CPU processes used and simulation times for each simulation. As processes run in parallel, the actual duration is the same as the longest individual process time.

7.2.2 Virtual Reproduction System

In Chapter 6, a VRS was used to examine the impact of loudspeaker position on binaural reproduction in the absence of a CTC stage (referred to as CTC-less binaural reproduction). The VRS facilitated the generation of a *modified* HRIR pair ($\hat{\mathbf{h}}^{\mathbf{N}}_{\mathbf{M}}$) as a result of reproducing each of 18 *reference* HRIR pairs ($\mathbf{h}^{\mathbf{N}}$, listed in Table 6.1), using each of 325,780 pairs of loudspeaker positions (giving a total of 5,864,000 modified HRIR pairs). The VRS comprises three stages: ipsilateral inversion (inverse filtering of the acoustic path between each loudspeaker and its ipsilateral ear), system reproduction (reproduction of the filtered reference HRIR pair over the CTC-less loudspeaker pair under test), and summation of the acoustic paths to create the signals at each ear. These stages are shown in Fig. 6.4. The binaural cues in the signals at the ears (the modified HRIR pair) could then be compared to those in the corresponding reference HRIR pair to determine the amount of deterioration introduced by that loudspeaker pair. In Section 6.3.1 and Section 6.3.2, the deterioration with respect to perceptual impact was estimated and the region of loudspeaker pair positions found which exhibited a deterioration smaller than 1 JND, the threshold of perception.

The VRS described in Section 6.2.2 was used as a basis for the VRS used in this chapter to examine the impact of a varying morphology on reproduction performance for each loudspeaker pair under test. The two main differences between the two versions of the VRS are: a) that the system impulse responses (IRs) to describe the acoustic paths between each loudspeaker position and the ipsilateral and contralateral ears (intended and crosstalk paths, respectively) were taken from the $\mathbf{SP^{SYMARE}}$ simulation results from each SYMARE head, rather than from those simulated with respect to KEMAR ($\mathbf{SP^{KEMAR}}$), and b) that a reduced set of loudspeaker pairs was used. Specifically, a set of 9,262 loudspeaker pairs, referred to as the set $\mathbf{LP^{SYMARE}}$, was defined by mirroring $\mathbf{SP^{SYMARE}}$ positions in the median plane. Using set notation, $\mathbf{LP^{SYMARE}}$ can be defined as:

$$\mathbf{LP^{SYMARE}} = \{(v, w) \mid v, w \in \mathbb{R}^3\} \tag{7.2}$$

which states that $\mathbf{LP^{SYMARE}}$ is a set of pairs of vectors v and w, such that v and w are each a three-dimensional, real-valued column vector, related through being mirrored in the median plane. Using the convention established in Chapter 6, each member of the set $\mathbf{LP^{SYMARE}}$ is referred to using the angle of the left hemisphere $\mathbf{SP^{SYMARE}}$ position. Regions of loudspeaker pair positions will be discussed and plotted with reference to the left hemisphere.

As described in Chapter 6, ipsilateral inversion using inverse filters was required to equalise the intended path for each loudspeaker. The inverse filters were generated from the corresponding SYMARE HRIRs using the implementation of the least-mean-square regularisation method outlined by Kirkeby *et al.* [269], as discussed in Chapter 3 and Chapter 6. Individualised system inversion was used to ensure that only the loudspeaker position had an impact on the binaural reproduction, as opposed to the difference between the subject and generic inverse filters (such as those generated from KEMAR HRIRs). The impact of using generic inverse filters, whilst perhaps relevant to a commercial implementation, is outside the scope of the work presented here. Discussion on further work relating to the role of system inversion is contained in Section 8.3.5. Fig. 7.3 plots example inversion filters for SYMARE subject #1, for the loudspeaker pair where the left loudspeaker is at $\theta = 45°,\ \phi = 0°,\ r = 0.15\,\mathrm{m}$. 20 dB of regularisation was used to achieve numerical accuracy. However, this example demonstrates that even 20 dB is not enough to ensure complete accuracy under all circumstances.

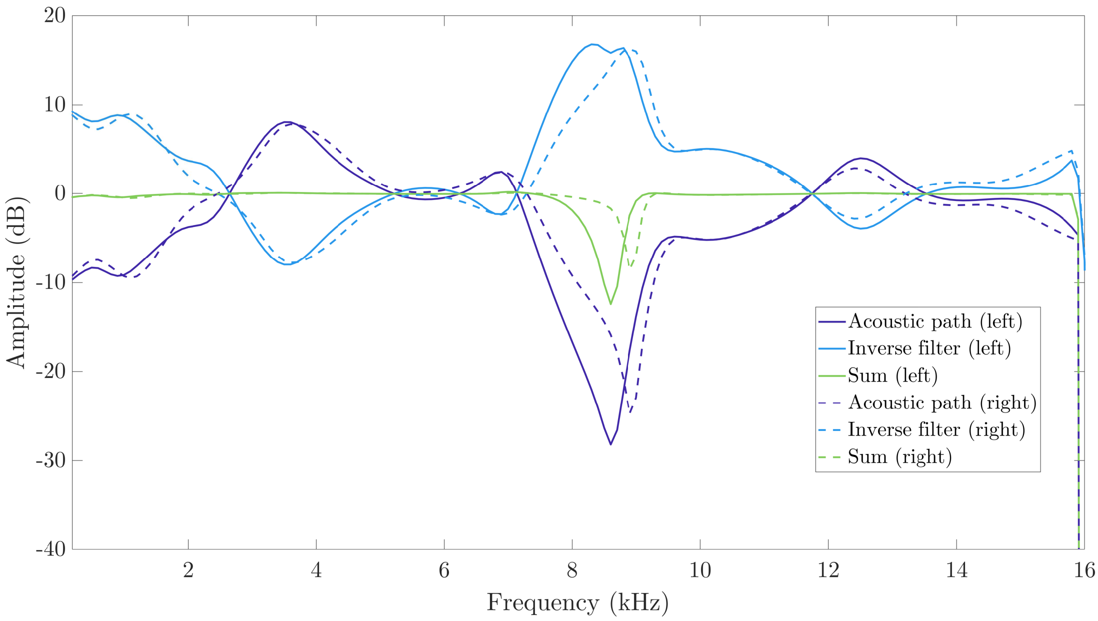


Figure 7.3: Frequency spectra of acoustic paths, ipsilateral inversion filters and their respective sums, generated for SYMARE subject #1 for the loudspeaker pair where the left is at $\theta = 45^{\circ},\ \phi = 0^{\circ},\ r = 0.15\,\mathrm{m}$.

The same 18 reference HRIR pairs as in Chapter 6 (denoted by $\mathbf{h^N}$, where the left and right signals are h_L^N and h_R^N, respectively) were used as inputs to the VRS to allow comparison across all subjects. This emulates the most common commercial condition of 'generic listening', in which the HRIR pair used to create the virtual binaural sound source was not captured from the subject under test. Using individualised HRIRs would likely create a more accurate listening experience when compared to the reference, but the goal in this work is commercial viability as opposed to absolute accuracy, and so the generic listening condition was used.

Using the same terminology as in Chapter 6, 9,262 modified HRIR pairs $\mathbf{\hat{h}_M^N}$ were produced for each of the 18 reference HRIR pairs. This gives a a total of 166,716 modified HRIR pairs per SYMARE subject.

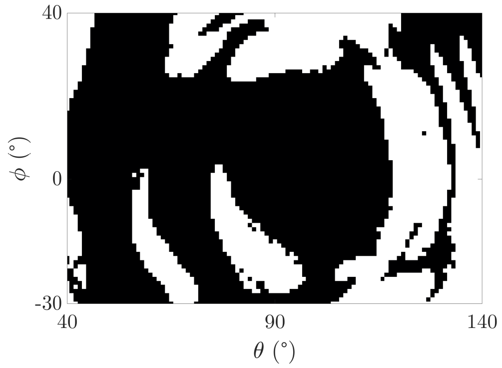
7.2.3 Calculation of Binaural Cue Deterioration

Using the methods described in Section 6.2.3 and Section 6.3.1, the deterioration in interaural binaural cues between the reference HRIR pairs and each corresponding modified HRIR pair was calculated for each SYMARE subject. This deterioration was characterised using the six statistical metrics defined previously (the arithmetic means, standard deviations and ranges for ΔITD and ΔILD for each loudspeaker pair: $\overline{\Delta ITD}$, $\sigma_{\Delta ITD}$, $(\Delta ITD_{max} - \Delta ITD_{min})$, $\overline{\Delta ILD}$, $\sigma_{\Delta ILD}$, $(\Delta ILD_{max} - \Delta ILD_{min})$) and related to perceptual impact through mapping of the delta values to just-noticeable difference (JND) threshold values as a function of reference HRIR pair direction.

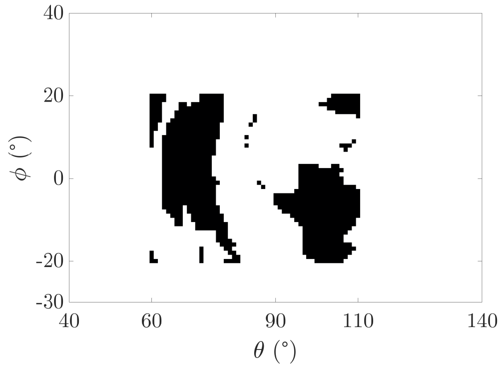
The perceptually-mapped deterioration values enabled the identification of pairs of positions where loudspeaker binaural rendering would create an estimated additional perceptual impact of under 1 JND for ΔITD and ΔILD separately and for all six criteria simultaneously ($\mathbf{LP^S_{ITD}}$, $\mathbf{LP^S_{ILD}}$ and $\mathbf{LP^S_{CM}}$ respectively, where the integer S refers to the index of the SYMARE subject). The processes for this analysis are described in Section 6.3.2. The intersection of the 15 $\mathbf{LP^S_{CM}}$ sets was then found in order to identify the region with a perceptual impact smaller than the threshold of perception common to all SYMARE subjects (referred to as $\mathbf{LP^{all}_{CM}}$).

7.3 Results

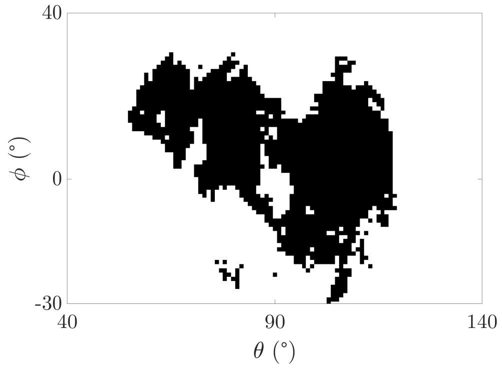
Using the methods outlined in Section 6.3.2, the $\mathbf{LP^S_{ITD}}$ and $\mathbf{LP^S_{ILD}}$ sets were determined for each of the SYMARE subjects used. As an example for subject #1, these regions of loudspeaker pair positions are indicated in black in Fig. 7.4a to Fig. 7.4d, separated into the two radial distances analysed. The intersection of these two regions, $\mathbf{LP^S_{CM}}$, is indicated in black in Fig. 7.4e and Fig. 7.4f, also separated into the two radial distances analysed. For visual comparison, the same azimuth and elevation axes limits are used across the two plots, and the smaller angular limits of the simulated region at $r = 0.20\,\text{m}$ have been indicated on the axes. The regions for all subjects analysed are included in Appendix F. The number of loudspeaker pairs in each set, separated into radial distances, are listed in Table 7.3.



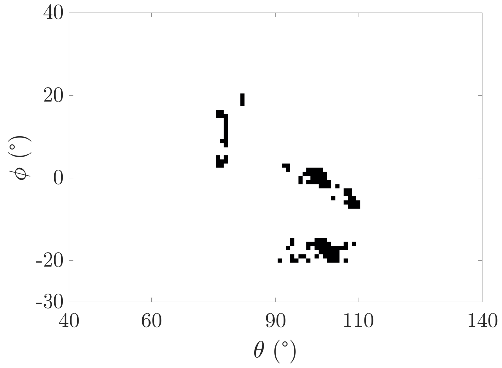
(a) $\mathbf{LP^1_{ITD}}$: $r = 0.15\,\mathrm{m}$



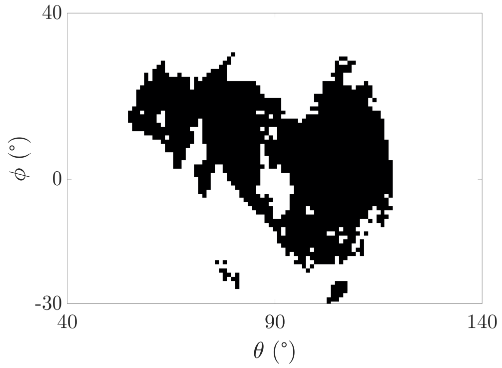
(b) $\mathbf{LP^1_{ITD}}$: $r = 0.20\,\mathrm{m}$



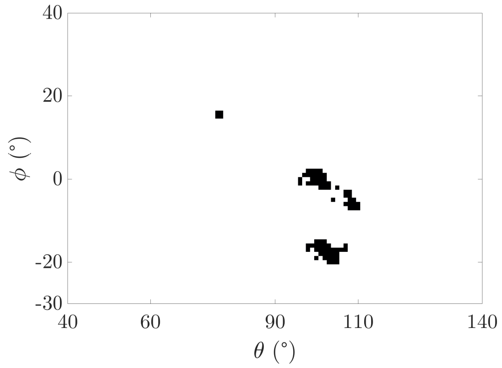
(c) $\mathbf{LP^1_{ILD}}$: $r = 0.15\,\mathrm{m}$



(d) $\mathbf{LP^1_{ILD}}$: $r = 0.20\,\mathrm{m}$



(e) $\mathbf{LP^1_{CM}}$: $r = 0.15\,\mathrm{m}$



(f) $\mathbf{LP^1_{CM}}$: $r = 0.20\,\mathrm{m}$

Figure 7.4: Regions of loudspeaker pair positions, indicated in black, with deterioration below 1 JND for: ITD cues ($\mathbf{LP^S_{ITD}}$, top row); ILD cues ($\mathbf{LP^S_{ILD}}$, middle row) and both cues at once ($\mathbf{LP^S_{CM}}$, bottom row) for SYMARE subject #1.

Subject #	$\mathbf{LP^S_{ITD}}$		$\mathbf{LP^S_{ILD}}$		$\mathbf{LP^S_{CM}}$	
	$r = 0.15\,\mathrm{m}$	$r = 0.20\,\mathrm{m}$	$r = 0.15\,\mathrm{m}$	$r = 0.20\,\mathrm{m}$	$r = 0.15\,\mathrm{m}$	$r = 0.20\,\mathrm{m}$
1	4,543 (63.4 %)	831 (39.7 %)	1,792 (25.0 %)	110 (5.3 %)	1,682 (23.5 %)	76 (3.6 %)
2	3,078 (42.9 %)	302 (14.4 %)	985 (13.7 %)	104 (5.0 %)	621 (8.7 %)	47 (2.2 %)
3	3,938 (54.9 %)	610 (29.2 %)	1,556 (21.7 %)	222 (10.6 %)	1,198 (16.7 %)	36 (1.7 %)
4	5,631 (78.5 %)	1,284 (61.4 %)	2,339 (32.6 %)	860 (41.1 %)	2,112 (29.5 %)	599 (28.6 %)
6	6,091 (84.9 %)	1,621 (77.5 %)	2,433 (33.9 %)	633 (30.3 %)	2,313 (32.3 %)	553 (26.4 %)
7	4,389 (61.2 %)	795 (38.0 %)	1,115 (15.5 %)	30 (1.4 %)	690 (9.6 %)	5 (0.2 %)
8	4,543 (63.4 %)	1,086 (51.9 %)	1,356 (18.9 %)	396 (18.9 %)	1,255 (17.5 %)	255 (12.2 %)
9	5,682 (79.2 %)	1,445 (69.1 %)	2,001 (27.9 %)	346 (16.5 %)	1,873 (26.1 %)	291 (13.9 %)
10	3,366 (46.9 %)	671 (32.1 %)	896 (12.5 %)	47 (2.2 %)	659 (9.2 %)	24 (1.1 %)
11	5,858 (81.7 %)	1,494 (71.4 %)	2,272 (31.7 %)	942 (45.1 %)	2,166 (30.2 %)	268 (12.8 %)
12	3,355 (46.8 %)	620 (29.7 %)	533 (7.4 %)	21 (1.0 %)	337 (4.7 %)	3 (0.1 %)
13	4,947 (69.0 %)	1,120 (53.6 %)	1,903 (26.5 %)	697 (33.3 %)	1,722 (24.0 %)	416 (19.9 %)
14	2,134 (29.8 %)	40 (1.9 %)	82 (1.1 %)	0 (0.0 %)	21 (0.3 %)	0 (0.0 %)
15	5,768 (80.4 %)	1,385 (66.2 %)	2,046 (28.5 %)	452 (21.6 %)	1,882 (26.2 %)	293 (14.0 %)
16	5,128 (71.5 %)	1,221 (58.4 %)	2,641 (36.8 %)	772 (36.9 %)	2,350 (32.8 %)	458 (21.9 %)

Table 7.3: Number and percentage of $\mathbf{LP^S_{ITD}}$, $\mathbf{LP^S_{ILD}}$ and $\mathbf{LP^S_{CM}}$ loudspeaker pair positions at each radial distance for each SYMARE subject used in the analysis. Percentages are calculated within each radial distance, therefore are out of 7,171 and 2,091 for $r = 0.15\,\mathrm{m}$ and $r = 0.20\,\mathrm{m}$, respectively.

As previously found in determining optimal positions of loudspeaker pairs with respect to KEMAR (Section 6.3.2), for each SYMARE subject there are more positions in the set $\mathbf{LP^{S}_{ITD}}$ than in the set $\mathbf{LP^{S}_{ILD}}$. Within each set, there are more positions at $r = 0.15\,\mathrm{m}$ than at $r = 0.20\,\mathrm{m}$.

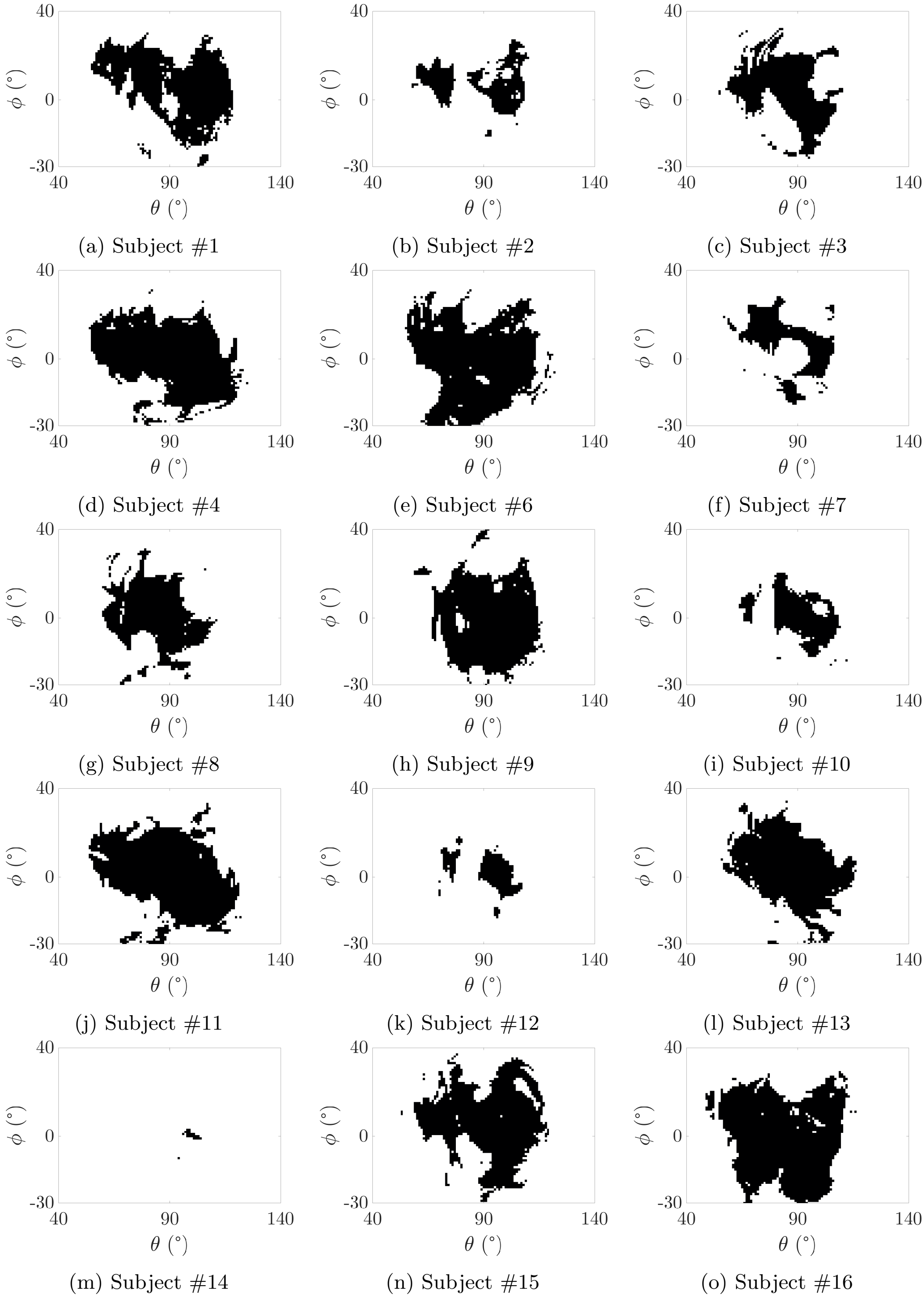
The patterns of positions at $r = 0.15\,\mathrm{m}$ are similar to those found with respect to KEMAR. As with $\mathbf{LP^{KEMAR}_{ITD}}$, the loudspeaker pair positions in the $\mathbf{LP^{S}_{ITD}}$ sets (subplots (a) in Figs. F.1–F.15) generally occupy a large proportion of the region in the vicinity of the interaural axis, with similar elliptical patterns of holes which vary in location for each SYMARE subject. The positions at $r = 0.15\,\mathrm{m}$ in the $\mathbf{LP^{S}_{ILD}}$ sets (subplots (c) in Figs. F.1–F.15) also follow a similar pattern to the set $\mathbf{LP^{KEMAR}_{ILD}}$, with positions in the vicinity of the interaural axis, as well as behind and below, and in front and above.

A higher degree of variation is present across SYMARE subjects for the regions at $r = 0.20\,\mathrm{m}$. For the $\mathbf{LP^{S}_{ITD}}$ sets (subplots (b) in Figs. F.1–F.15), a hole in the vicinity of the interaural axis is present for all SYMARE subjects. However, the extents of the hole in azimuth and elevation, as well as the presence of other holes, varies greatly between subjects. For the $\mathbf{LP^{S}_{ILD}}$ sets (subplots (d) in Figs. F.1–F.15), there is little in common between the number and position of loudspeaker pairs.

In the set $\mathbf{LP^{KEMAR}_{ITD}}$, loudspeaker pair positions are present within all radial distances analysed, up to and including a maximum of $r = 1.50\,\mathrm{m}$. As the regions determined with respect to SYMARE subjects appear to follow similar patterns to those determined with respect to KEMAR, it is feasible that more positions would be included in the $\mathbf{LP^{S}_{ITD}}$ sets if a larger number of directions and radial distances had been analysed. However, the positions of loudspeaker pairs under test were limited to the $\mathbf{SP^{SYMARE}}$ source positions. Therefore, conclusions can only be drawn for the loudspeaker pair positions within the regions bounded by $40° \leq \theta \leq 140°$, $-30° \leq \phi \leq 40°$ for radial distance $r = 0.15\,\mathrm{m}$, and $60° \leq \theta \leq 110°$, $-20° \leq \phi \leq 20°$ for radial distance $r = 0.20\,\mathrm{m}$, as defined in Table 7.1.

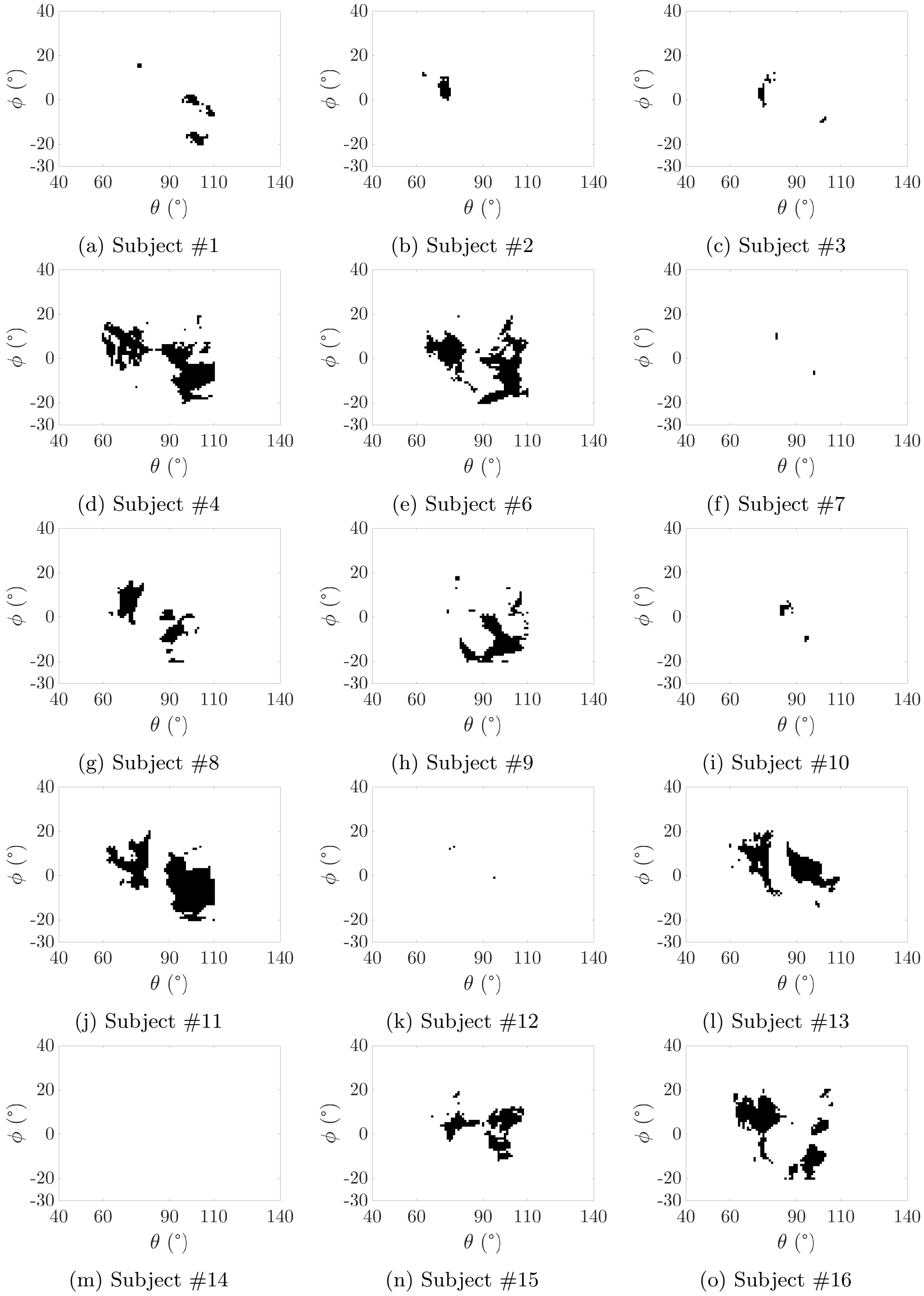
Since the $\mathbf{LP^{S}_{ITD}}$ and $\mathbf{LP^{S}_{ILD}}$ sets follow similar patterns of distribution to those of sets $\mathbf{LP^{KEMAR}_{ITD}}$ and $\mathbf{LP^{KEMAR}_{ITD}}$ for $r = 0.15\,\mathrm{m}$, it follows that $\mathbf{LP^{S}_{CM}}$ does too. For ease of visual comparison across all SYMARE subjects at once, Fig. 7.5 and Fig. 7.6 show the $\mathbf{LP^{S}_{CM}}$ regions for all SYMARE subjects included in this study at $r = 0.15\,\mathrm{m}$ and $r = 0.20\,\mathrm{m}$, respectively. For each SYMARE subject, larger regions exist at $r = 0.15\,\mathrm{m}$ than at $r = 0.20\,\mathrm{m}$. The size of the region varies across subjects, however, as does the spatial location. As with the set $\mathbf{LP^{KEMAR}_{CM}}$, the inclusion of positions in $\mathbf{LP^{S}_{CM}}$ is predominantly defined by the corresponding $\mathbf{LP^{S}_{ILD}}$ set.

The $\mathbf{LP^{S}_{CM}}$ loudspeaker pair positions at $r = 0.15\,\mathrm{m}$ (Fig 7.5) generally demonstrate a similar pattern of distribution to the region determined with respect to KEMAR, with positions in the vicinity of the interaural axis and with coverage in two general areas: in front and above the interaural axis, and behind and below the interaural axis. The regions for some subjects (for example, subject #9 (Fig. 7.5h) and subject #16 (Fig. 7.5o) exhibit less of this diagonal shape and have greater front-back symmetry. As a result of the distribution of loudspeaker pair positions in the $\mathbf{LP^{S}_{ILD}}$ sets, the regions for $\mathbf{LP^{S}_{CM}}$ at $r = 0.20\,\mathrm{m}$ (Fig. 7.6) vary much more across the subjects, with some showing few (for example, only 3 for subject #12 (Fig. 7.6k)) or even no loudspeaker positions at all (such as for subject # 14 (Fig 7.6m)).



(a) Subject #1 (b) Subject #2 (c) Subject #3
(d) Subject #4 (e) Subject #6 (f) Subject #7
(g) Subject #8 (h) Subject #9 (i) Subject #10
(j) Subject #11 (k) Subject #12 (l) Subject #13
(m) Subject #14 (n) Subject #15 (o) Subject #16

Figure 7.5: Regions of loudspeaker pair positions at $r = 0.15\,\mathrm{m}$ with deterioration below 1 JND for both ITD and ILD at once ($\mathbf{LP^{S}_{CM}}$) for the 15 SYMARE subjects included in this study, indicated in black.



(a) Subject #1 (b) Subject #2 (c) Subject #3

(d) Subject #4 (e) Subject #6 (f) Subject #7

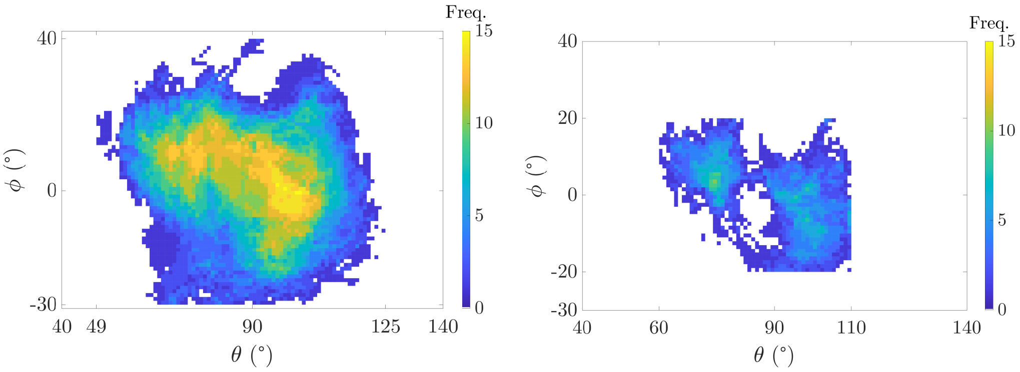
(g) Subject #8 (h) Subject #9 (i) Subject #10

(j) Subject #11 (k) Subject #12 (l) Subject #13

(m) Subject #14 (n) Subject #15 (o) Subject #16

Figure 7.6: Regions of loudspeaker pair positions at $r = 0.20\,\mathrm{m}$ with deterioration below 1 JND for both ITD and ILD at once ($\mathbf{LP^{S}_{CM}}$) for the 15 SYMARE subjects included in this study, indicated in black.

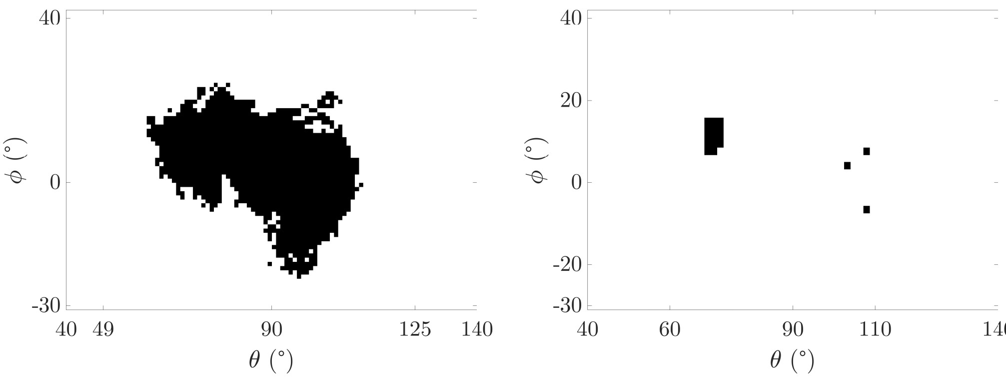
Fig. 7.7 shows the frequency of occurrence of each unique loudspeaker pair position within all 15 $\mathbf{LP}^{\mathbf{S}}_{\mathbf{CM}}$ sets at the two radial distances. Colour indicates the frequency of occurrence between one (dark blue) and 15 (yellow), White indicates zero occurrences. For $r = 0.15\,\mathrm{m}$, the 3,526 unique positions lie within the region bounded by $49° \leq \theta \leq 125°,\ -30° \leq \phi \leq 40°$. For $r = 0.20\,\mathrm{m}$, the 1,268 unique positions lie within the region bounded by the simulation limits of $60° \leq \theta \leq 110°,\ -20° \leq \phi \leq 20°$.



(a) $r = 0.15\,\mathrm{m}$ (b) $r = 0.20\,\mathrm{m}$

Figure 7.7: Frequency of occurrence (indicated by colour) of all loudspeaker pair positions within all $\mathbf{LP}^{\mathbf{S}}_{\mathbf{CM}}$ sets, where white indicates zero occurrences.

Despite the variation in the determined regions for each SYMARE subject, some loudspeaker pair positions are common to a number of $\mathbf{LP}^{\mathbf{S}}_{\mathbf{CM}}$ sets. Fig. 7.8 shows the positions which exist in the $\mathbf{LP}^{\mathbf{S}}_{\mathbf{CM}}$ sets for more than half of the subjects (that is, with a frequency of occurrence greater than seven out of fifteen), indicated in black. 1,264 such positions exist at $r = 0.15\,\mathrm{m}$ (35.4 % of all loudspeaker pair positions at that radial distance) in the region bounded by $60° \leq \theta \leq 112°,\ -23° \leq \phi \leq 24°$. 17 positions exist at $r = 0.20\,\mathrm{m}$ (1.3 % of all loudspeaker pair positions at that radial distance) with one region centred at $\theta = 70°,\ \phi = 12°$ and another three disconnected points.



(a) $r = 0.15\,\mathrm{m}$ (b) $r = 0.20\,\mathrm{m}$

Figure 7.8: Regions of loudspeaker pair positions which exist in more than half of all $\mathbf{LP}^{\mathbf{S}}_{\mathbf{CM}}$ sets. 1,264 positions exist at $r = 0.15\,\mathrm{m}$ and form one distinct region with a small number of disconnected points. 17 positions exist at $r = 0.20\,\mathrm{m}$.

Table 7.4 lists the number of loudspeaker pair positions for each frequency of occurrence. At radial distance $r = 0.15\,\mathrm{m}$, a small number of positions exist in all 15 $\mathbf{LP^S_{CM}}$ sets. This subset is referred to as the set $\mathbf{LP^{all}_{CM}}$, which is the intersection of all $\mathbf{LP^S_{CM}}$ sets (see Equation 7.3).

Frequency of occurrence within all $\mathbf{LP^S_{CM}}$ sets	Number of loudspeaker pair positions $r = 0.15\,\mathrm{m}$	$r = 0.20\,\mathrm{m}$
1	577	370
2	415	245
3	357	199
4	246	204
5	190	108
6	215	91
7	262	34
8	212	11
9	184	5
10	190	1
11	310	0
12	174	0
13	125	0
14	65	0
15	4	0
Total:	3,526	1,268

Table 7.4: Number of loudspeaker pair positions at each frequency of occurrence across all $\mathbf{LP^S_{CM}}$ sets. 3,526 unique positions exist with a frequency of occurrence greater than zero out of fifteen.

$$\mathbf{LP^{all}_{CM}} = \mathbf{LP^{1}_{CM}} \cap \mathbf{LP^{2}_{CM}} \cap \mathbf{LP^{3}_{CM}} \cap \mathbf{LP^{4}_{CM}} \cap \mathbf{LP^{6}_{CM}} \cap \mathbf{LP^{7}_{CM}} \cap \mathbf{LP^{8}_{CM}} \cap \mathbf{LP^{9}_{CM}} \cap \mathbf{LP^{10}_{CM}} \cap \mathbf{LP^{11}_{CM}} \cap \mathbf{LP^{12}_{CM}} \cap \mathbf{LP^{13}_{CM}} \cap \mathbf{LP^{14}_{CM}} \cap \mathbf{LP^{15}_{CM}} \cap \mathbf{LP^{16}_{CM}} \quad (7.3)$$

$\mathbf{LP^{all}_{CM}}$ is shown in Fig. 7.9 and contains 4 loudspeaker pair positions (0.1 % of all positions at that radial distance) at the following angles:

1. $\theta = 96°,\ \phi = 1°$

2. $\theta = 98°,\ \phi = 0°$

3. $\theta = 101°,\ \phi = -1°$

4. $\theta = 102°,\ \phi = -1°$

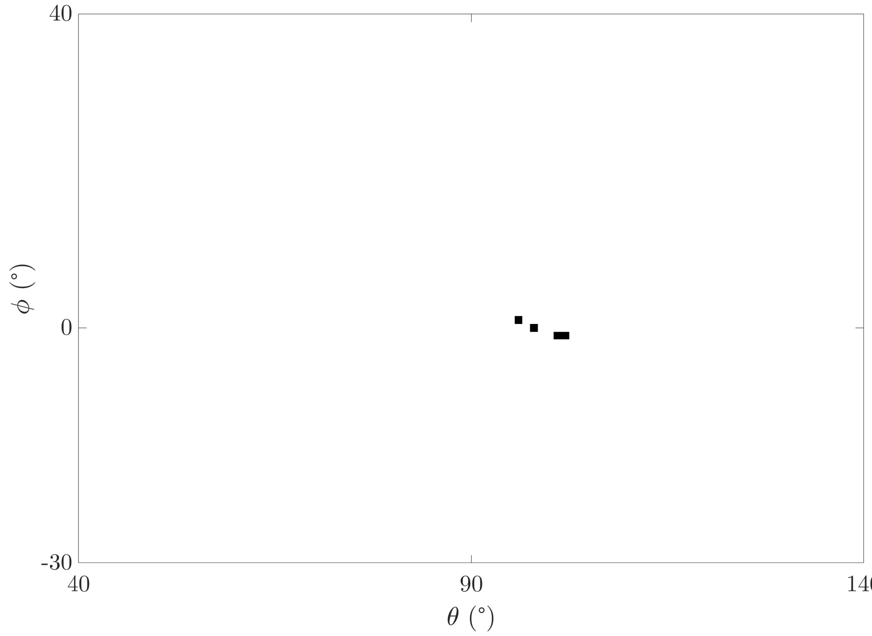


Figure 7.9: Loudspeaker pair positions which exist in $\mathbf{LP^{S}_{CM}}$ sets for all SYMARE subjects ($\mathbf{LP^{all}_{CM}}$). Such positions (of which there are four) only exist at $r = 0.15\,\mathrm{m}$.

The existence of common positions across all $\mathbf{LP^{S}_{CM}}$ sets suggests that loudspeaker pairs in these positions would produce perceptually-transparent binaural signals for all subjects. This small subset is of limited practical use, however, as it would be impractical to create a real-world acoustic loudspeaker reproduction system using such precise loudspeaker placement. As this subset is the intersection of all $\mathbf{LP^{S}_{CM}}$ sets, the size of this region is determined by the smallest region within any of the SYMARE subjects. In this case, it is subject #14 which has the smallest $\mathbf{LP^{S}_{CM}}$ set (Fig. F.13e), with 21 loudspeaker pair positions due to its small $\mathbf{LP^{S}_{ILD}}$ region (Fig. F.13c).

As a thought experiment, Fig. 7.10 shows the loudspeaker pair positions which exist in all $\mathbf{LP^{S}_{CM}}$ sets other than that of subject #14 ($\mathbf{LP^{14}_{CM}}$). There are 60 such positions: 59 positions lie in a roughly diagonal region bounded by $94° \leq \theta \leq 103°,\ -6° \leq \phi \leq 6°$ and one disconnected point at $\theta = 73°,\ \phi = 12°$. The increase in number of candidate loudspeaker pair positions through removal of one subject suggests that inclusion and exclusion of loudspeaker pair positions in this way may be overly sensitive with regards to so-called anomalous sets. However, such anomalies evidently occur within the population and so perhaps should not be excluded. It may also be the case that more of these so-called anomalous sets may be present if more subjects were to be included in this analysis.

It is feasible that an increase in the number of subjects used in analysis may also reduce further the number of candidate loudspeaker pair positions. Performing analysis on the first seven subjects results in 153 candidate loudspeaker pair positions, a larger number than the four positions indicated with 15 subjects. The inclusion or exclusion of subject #14 has a substantial impact on the number of loudspeaker pair positions; it is likely that including more subjects results in a greater proportion of the loudspeaker pair positions being excluded from the set $\mathbf{LP^{all}_{CM}}$, as they do not appear in all $\mathbf{LP^{S}_{CM}}$

sets due to inter-subject variation.

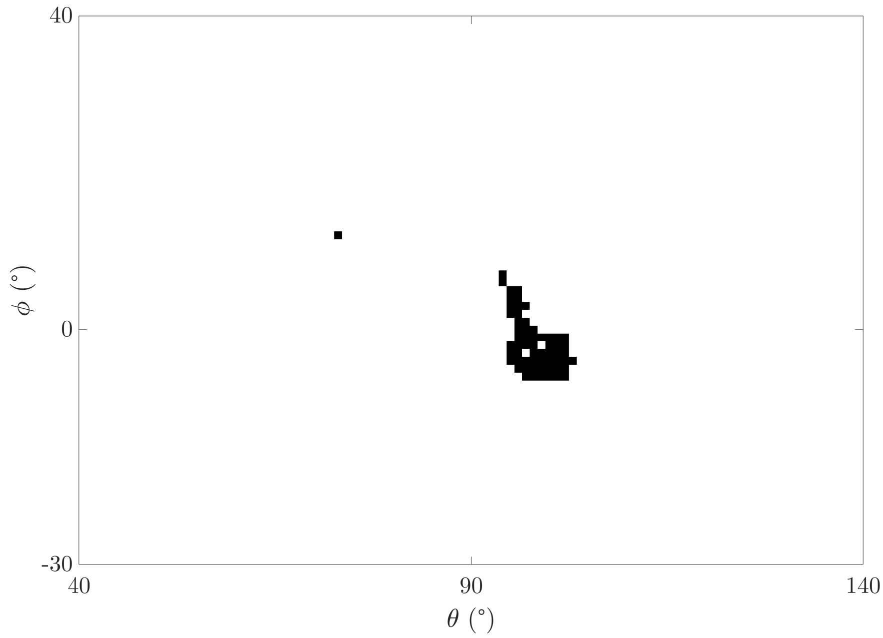


Figure 7.10: Loudspeaker pair positions which exist in $\mathbf{LP^{S}_{CM}}$ sets for all SYMARE subjects, excluding subject #14. Such pairs (of which there are 60) only exist at $r = 0.15\,\mathrm{m}$.

As discussed in Section 6.4.2 and Section 8.2.2, the perceptual model used in this work to relate deterioration in binaural cues to perceptual impact is likely over-cautious. That is, it errs on the side of over-estimating the perceptual impact of lateral sources. It may be the case that, as in Chapter 6, a larger, less conservative (and perhaps more accurate) JND threshold for lateral sources would result in more loudspeaker pair positions within the set $\mathbf{LP^{all}_{CM}}$.

It is evident from the variation across the $\mathbf{LP^{S}_{CM}}$ sets that some form of relationship exists between the morphology of the subject and the region in which loudspeaker pairs may faithfully reproduce binaural signals without CTC. Some initial work regarding this is presented in Section 8.3.3. However, substantial investigation of this relationship is outside the scope of the work presented here.

7.4 Conclusion

This chapter expands on the methods and processes described in Chapter 6 to determine pairs of positions where loudspeakers exhibit deterioration in interaural binaural cues below the threshold of perception for a range of subjects.

The VRS described in Section 6.2.2 was modified to use HRIRs simulated for 15 subjects from the SYMARE database [80]. 18,524 HRTF source positions were simulated (referred to as the set $\mathbf{SP^{SYMARE}}$). This is fewer source positions than were simulated for analysis with respect to KEMAR ($\mathbf{SP^{KEMAR}}$), as the results in Chapter 6 indicate that few loudspeaker positions of interest exist for radial distances greater than $r = 0.20\,\mathrm{m}$ or for angular directions greater than 50° from the interaural axis.

Accordingly, to reduce computational effort these source positions were not included in the simulations.

As in Section 6.2.2, pairs of loudspeaker positions were formed from the simulated source positions through mirroring in the median plane. This gave 9,262 positions, referred to as the set $\mathbf{LP^{SYMARE}}$. The 18 HRIR pairs from the SADIE database v1 [123] were again used as inputs to the modified VRS to create 166,716 (9,262 $\times$ 18) modified HRIR pairs for each SYMARE subject (where a modified HRIR pair is the result at the ears of reproducing the reference HRIR pair using the simulated loudspeaker pair under test).

Using the methods described in Section 6.2.3 and Section 6.3.1, the deterioration in interaural binaural cues between each reference HRIR pair and the corresponding modified HRIR pair was calculated for each loudspeaker pair position and for each SYMARE subject. The deterioration was then mapped to an estimate of its perceptual impact and the positions at which loudspeaker pairs exhibit deterioration below the threshold of perception were determined. Sets of positions were identified for each SYMARE subject and referred to as the $\mathbf{LP^{S}_{ITD}}$ and $\mathbf{LP^{S}_{ILD}}$ sets, where S is the index of the relevant SYMARE subject.

For all SYMARE subjects, more candidate loudspeaker pair positions exist in the set $\mathbf{LP^{S}_{ITD}}$ than in the set $\mathbf{LP^{S}_{ILD}}$. Within each set, more candidate positions exist at $r = 0.15\,\mathrm{m}$ than at $r = 0.20\,\mathrm{m}$. The distribution patterns of loudspeaker pair positions at $r = 0.15\,\mathrm{m}$ are generally similar to those found with respect to KEMAR ($\mathbf{LP^{KEMAR}_{ITD}}$ and $\mathbf{LP^{KEMAR}_{ILD}}$). The patterns at $r = 0.20\,\mathrm{m}$ vary more across SYMARE subjects, particularly the $\mathbf{LP^{S}_{ILD}}$ sets, where there is little commonality across subjects.

For each SYMARE subject, the intersection of the $\mathbf{LP^{S}_{ITD}}$ and $\mathbf{LP^{S}_{ILD}}$ sets was calculated to determine the positions where loudspeaker pairs would provide faithful binaural reproduction across both interaural cues. These sets are referred to as $\mathbf{LP^{S}_{CM}}$. More positions exist at $r = 0.15\,\mathrm{m}$ than at $r = 0.20\,\mathrm{m}$ for all SYMARE subjects. The $\mathbf{LP^{S}_{CM}}$ sets generally demonstrate similar position distributions to each other and to the set determined with respect to KEMAR $\left(\mathbf{LP^{KEMAR}_{CM}}\right)$. Loudspeaker pair positions exist in the vicinity of the interaural axis and cover two general regions: in front and above the interaural axis, and behind and below. Many positions exist in more than one $\mathbf{LP^{S}_{CM}}$ set: for example, 1,264 positions exist in seven of the 15 $\mathbf{LP^{S}_{CM}}$ sets.

The principal objective of this research is to determine whether loudspeaker pair positions exist where satisfactory CTC-less binaural reproduction can be achieved independent of a listener's unique morphology. Thus, ideally, such positions should exist in all 15 $\mathbf{LP^{S}_{CM}}$ sets. Four such positions were determined, between $98^\circ \leq \theta \leq 102^\circ$ and $-1^\circ \leq \phi \leq 1^\circ$. This set is referred to as the set $\mathbf{LP^{all}_{CM}}$. The small number is predominantly the result of a small $\mathbf{LP^{S}_{CM}}$ region related to one specific SYMARE subject. However, as discussed in Chapter 6, the perceptual model used in this analysis is rudimentary and conservative, and it is possible that use of more accurate mapping of binaural cue deterioration to perceptual impact would result in the inclusion of more loudspeaker pair positions in the set $\mathbf{LP^{all}_{CM}}$.

In reporting these findings, further questions and issues have been identified, some of which have been discussed in the relevant chapters throughout this thesis. Further considerations are reflected upon and explored more deeply in Chapter 8, and some recommendations for further work are proposed.

Chapter 8

Additional Discussion

Take this passion, turn it into action.

Tegan and Sara, *Stop Desire*
Love You to Death (2016)

8.1 Introduction

The work presented in this thesis suggests that it is possible to achieve perceptually transparent binaural reproduction using loudspeakers and without the need for crosstalk cancellation (CTC) by placing the loudspeakers close to the head. This conclusion has been reached using several methods. Firstly, the boundary element method (BEM) has been used to simulate a large number of head-related transfer functions (HRTFs) for the Knowles Electronics Manikin for Acoustic Research (KEMAR) and for subjects from the Sydney-York Morphological and Recording of Ears (SYMARE) database [80]. These HRTFs were used to implement a virtual reproduction system (VRS) which simulates the acoustic paths for a two-loudspeaker binaural reproduction system without CTC (referred to as *CTC-less* binaural reproduction). The VRS allowed for the calculation of natural channel separation (NCS) for each loudspeaker position, and made it possible to examine the deterioration in binaural cues introduced by different placements of the loudspeaker pairs. The set of loudspeaker pair positions for which the deterioration in performance is below the estimated threshold of perception was then determined.

During presentation of the methods and findings, however, further questions and issues were identified. Two such issues have been discussed in the relevant chapters and so will not be discussed further in this chapter. The impact of the range of frequencies used for calculating NCS for a given source direction was discussed in Section 5.5. The variation of NCS with this range was shown to be complex, highlighting the need for using the same range as in the literature if comparisons are intended to be made. Comparisons between NCS and other performance predictors were presented in Section 6.4, with results indicating that the relationship between NCS and binaural performance is also complex.

This chapter reflects on broader considerations of the presented work, including the implications of using simulation-based environments and a rudimentary perceptual model. Some opportunities for further work are also presented.

8.2 Limitations

As with any experimental work, limitations of the methods and conclusions must be considered. With respect to the work presented in this thesis, the two major aspects for consideration are the limitations of the procedures used, and the limitations of the perceptual model.

8.2.1 Procedural Limitations

This work develops a proof of concept: it is a starting point from which further research can progress. The procedures used in the simulations and experimental work described unavoidably restrict the validity of the results obtained.

The first limitations to be considered are those imposed by the use of simulation-based environments. Use of the BEM enabled a large number of HRTFs to be acquired for many subjects. The VRS enabled analysis of the difference in interaural binaural cues contained within each HRIR pair before and after reproduction. However, simulated environments are not fully representative of the real-world acoustic environment. Some of the attractions of simulation are that the loudspeakers are ideal point sources, no other objects are present in the environment, and there is no subject movement. Whilst these aspects are useful, they represent an ideal version of the real environment. The future development of an acoustic prototype system to examine the impact of the differences between simulation and acoustic environments is discussed in Section 8.3.4.

The interaural binaural cues for reference and modified HRIR pairs[1] were quantified using established methods: cross-correlation for interaural time difference (ITD) and arithmetic mean over frequency for interaural level difference (ILD). These methods are referred to as *estimation* methods, as they can only estimate the value of the perceptual cue based on a model of the behaviour of the auditory system. The alignment of ITD estimations methods with perceptual results has been explored by Andreopoulou *et al.* [65]. Although ITD estimation methods generally produce comparable values and vary with azimuth in a similar manner to each other, they also tend to display characteristic artefacts [65]. It is not clear, for example, whether the large jumps in ITD for a small angular change, as shown in Section 6.3.1, are a true representation of the perception of a listener, or whether this is a flaw with the estimation method. To avoid such outcomes, analysis would need to move away from using estimated interaural binaural cues as the difference metrics, and perhaps move towards localisation error estimation using binaural models (such as those described in the Auditory Model Toolbox [287, 288]) or using human listeners in a listening test (once an acoustic implementation has been constructed). Such analysis methods were considered for this work, and although they would have provided an excellent grounding for future investigations, there was insufficient time to include them in this initial proof of principle work.

Some other factors which may have affected the results can be grouped under the term *quantisation*. These include the use of a finite number of radial distances at which

[1] More detailed descriptions of the terms *reference* and *modified* HRIR pair are given in Section 6.1.

HRTFs were simulated, and the finite numbers of SYMARE subjects and reference HRIRs used in the analysis.

Increasing the number of test conditions would characterise the performance of the loudspeaker pairs in more detail. That is, using more radial distances, reference HRIR pairs and SYMARE subjects would give a more complete description of the performance of the candidate loudspeaker pairs. However, the numbers used in this work were constrained by the computational resources available. The aim was to provide a detailed enough description for a proof of concept: anything more than a small increase in the resolution of these three variables would have had an adverse impact on the computation and analysis time.

8.2.2 Perceptual Limitations

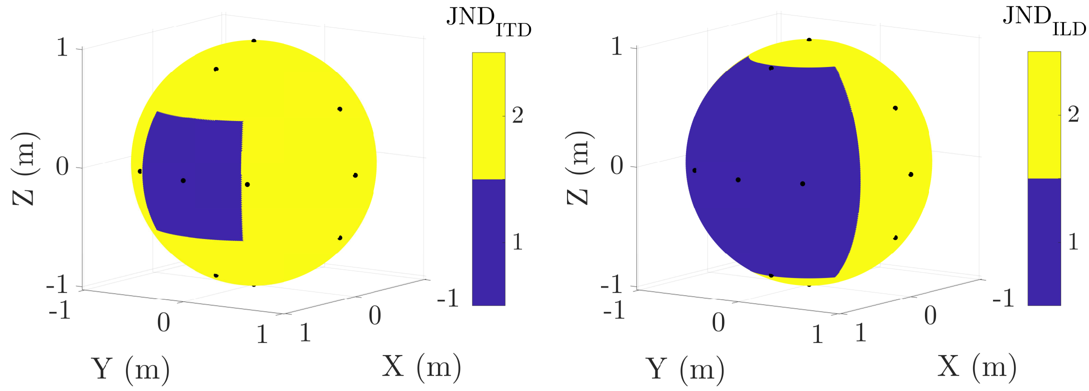
Perhaps the main limitation of the work is the rudimentary perceptual model used to map numerical differences to perceptual impact in Chapter 6 and Chapter 7. The localisation accuracy of the human auditory system varies with sound source direction, with the area of highest accuracy for both ITD and ILD being directly ahead on the horizontal plane [110, 111]. As a result of this directional variation, the perceptual impact of numerical changes in the binaural cues will vary as a function of the direction of the sound source. The magnitude of a change in a cue which can just be perceived (the just-noticeable difference, or JND) varies with source direction. These changes can be quantified and compared with a reference - in this case, the numerical change which corresponds to 1 JND in the region of best performance. These 1 JND reference values for ITD and ILD were defined in Equation 6.7 and Equation 6.8 as $1\,\mathrm{JND}_{\mathrm{ITD}}$ and $1\,\mathrm{JND}_{\mathrm{ILD}}$, respectively, where:

$$1\,\mathrm{JND}_{\mathrm{ITD}} = 16\,\mu\mathrm{s} \tag{8.1}$$

$$1\,\mathrm{JND}_{\mathrm{ILD}} = 1\,\mathrm{dB} \tag{8.2}$$

As described in Section 6.3, a two-level perceptual model was implemented to map numerical differences in binaural cues to JND in multiples of $1\,\mathrm{JND}_{\mathrm{ITD}}$ and $1\,\mathrm{JND}_{\mathrm{ILD}}$. For example, if the smallest perceivable change in ITD at a given direction is $32\,\mu\mathrm{s}$, the JND can be described as having a value of $2\,\mathrm{JND}_{\mathrm{ITD}}$. While literature does exist regarding the variation of localisation accuracy with source direction, there is a lack of detail on how the relationships between binaural cues and their associated perceptual thresholds vary with direction, particularly for an arbitrary change in either azimuth or elevation. The model implemented for this work was, as far as possible, implemented using the angles and JND thresholds reported in the literature. However, this left regions of uncertainty where experimental values are unavailable.

The two-level model used in this work defines two distinct regions of perceptual mapping, shown in Fig. 6.9 and reproduced in Fig 8.1 for convenience. Within the frontal region (indicated in purple), the lowest JND thresholds of $1\,\mathrm{JND}_{\mathrm{ITD}}$ and $1\,\mathrm{JND}_{\mathrm{ILD}}$ ($16\,\mu\mathrm{s}$ and $1\,\mathrm{dB}$, respectively) are used to map numerical difference to perceptual impact in JND. A second region (indicated in yellow) was defined beyond a designated angle for azimuth and elevation, as informed by the literature. Within this region, JND thresholds of $2\,\mathrm{JND}_{\mathrm{ITD}}$ and $2\,\mathrm{JND}_{\mathrm{ILD}}$ ($32\,\mu\mathrm{s}$ and $2\,\mathrm{dB}$, respectively) are used. The boundary angles in both azimuth and elevation were set at 30° for ITD and 60° for ILD, with the same angle used for elevation as for azimuth as a worst-case measure, since elevation localisation accuracy is always worse than azimuth accuracy [111]. As an example, a ΔITD of 50 µs at $\theta = 0°$, $\phi = 0°$ is mapped to a perceptual impact of 3.1 JND. For a source direction at $\theta = 90°$, $\phi = 0°$, the same numerical difference is mapped to 1.6 JND.



(a) ΔITD: 1 JND$_{\text{ITD}}$ within 30° of straight ahead (b) ΔILD: 1 JND$_{\text{ILD}}$ within 60° of straight ahead

Figure 8.1: JND thresholds used to map binaural cue differences to perceptual impact for ITD and ILD using the two-level perceptual model. Reference HRIR pair directions are indicated as black dots.

As this perceptual model is overly simplistic, it is of interest to examine the impact of using a more sophisticated model. This would incorporate more levels of JND threshold and extend the range of JND variation, since localisation accuracy for sound sources at the more lateral directions can deteriorate by 5 to 6 times compared with sources in the frontal region [116, 119, 120]. Some initial work on this has been conducted, and is described below.

Although it is likely that the increase in JND with direction is nonlinear, for simplicity here, regions were defined using a linear stepped increase in JND between a region of highest accuracy directly ahead (denoted by 1 JND$_{\text{min}}$) and a region of lowest accuracy at $\theta = 90°$ (denoted by 6 JND$_{\text{min}}$). This gives six regions of perceptual mapping values with boundaries at 18° increments, as listed in Table 8.1. The JND thresholds were mirrored in the median plane to create the thresholds for the front right quadrant. The same step intervals were used for azimuth and elevation, and also for both ITD and ILD (hence the use of the same value JND$_{\text{min}}$ for both cues). The regions of the sphere which these levels correspond to are shown in Fig. 8.2, where colour indicates JND between 1 and 6 JND$_{\text{min}}$, and the reference HRIR pair directions are indicated as black dots. The values used to map numerical differences at each reference HRIR pair direction to perceptual impact using this six-level model are listed in Table 8.2.

Angle region (°)	JND threshold		
	JND	ΔITD (µs)	ΔILD (dB)
0 to 17	$1\,\mathrm{JND_{min}}$	16	1
18 to 35	$2\,\mathrm{JND_{min}}$	32	2
36 to 53	$3\,\mathrm{JND_{min}}$	48	3
54 to 71	$4\,\mathrm{JND_{min}}$	64	4
72 to 89	$5\,\mathrm{JND_{min}}$	80	5
≥90	$6\,\mathrm{JND_{min}}$	96	6

Table 8.1: Multiples of $\mathrm{JND_{min}}$ for the front left quadrant used in the six-level perceptual model. The $\mathrm{JND_{min}}$ multipliers are mirrored in the median plane to create the thresholds for the front right quadrant.

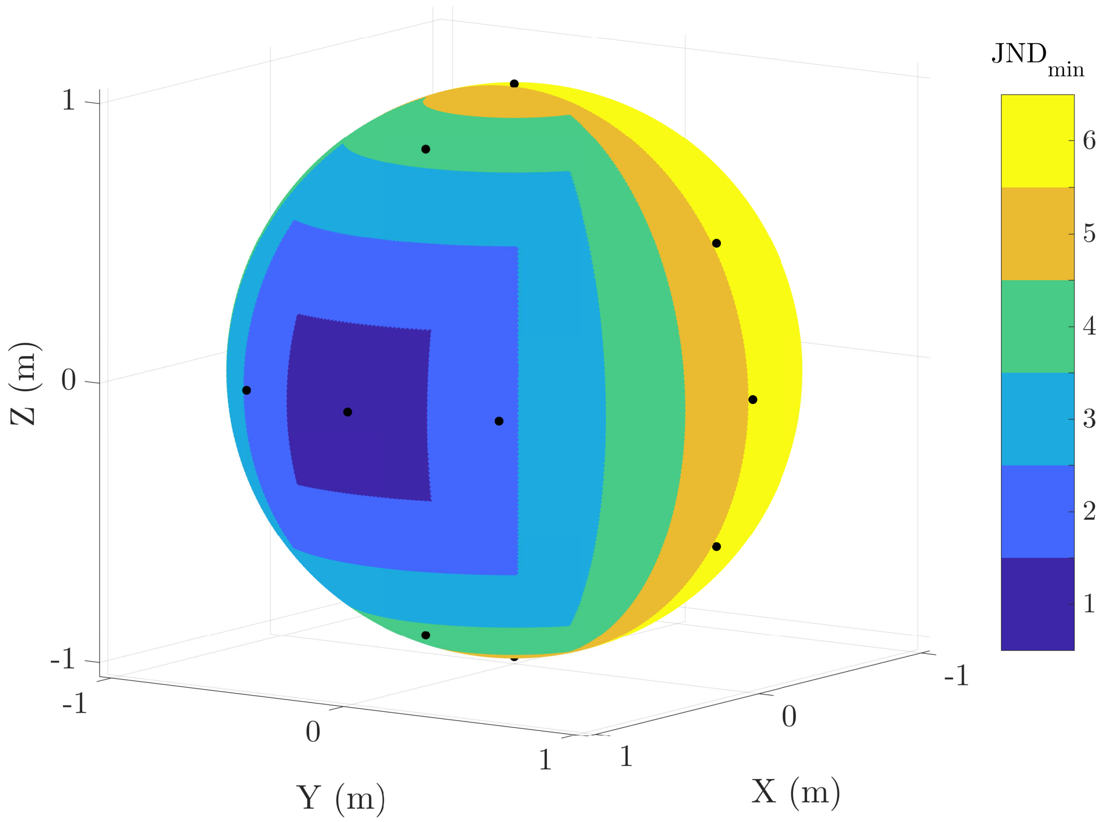


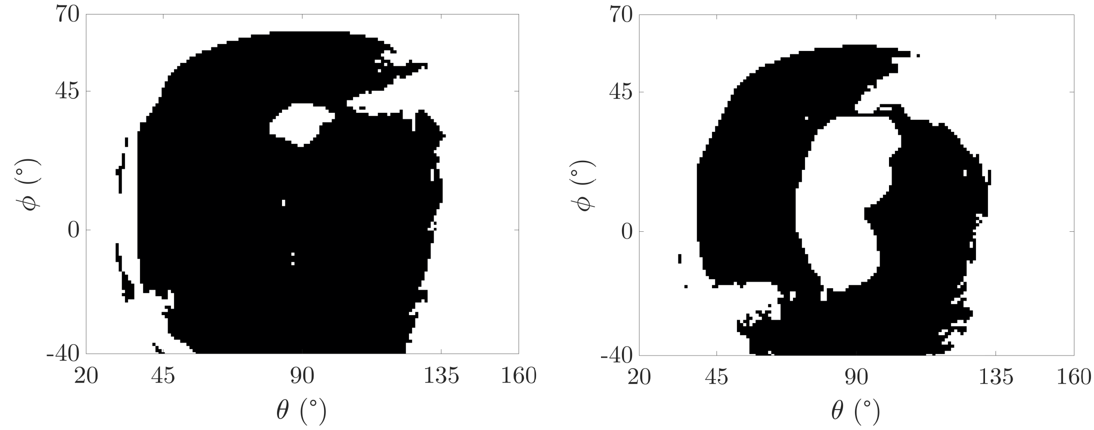
Figure 8.2: JND thresholds (in multiples of $\mathrm{JND_{min}}$) used to map binaural cue differences to perceptual impact for ITD and ILD using the six-level perceptual threshold model. Reference HRIR pair directions are indicated as black dots.

Source			JND threshold			
Label #	θ (°)	ϕ (°)	ΔITD (µs)	$\mathrm{JND_{min}}$	ΔILD (dB)	$\mathrm{JND_{min}}$
1	90	32	96	6	6	6
2	270	32	96	6	6	6
3	90	-32	96	6	6	6
4	270	-32	96	6	6	6
5	32	0	32	2	2	2
6	148	0	96	6	6	6
7	328	0	32	2	2	2
8	212	0	96	6	6	6
9	0	58	32	2	2	2
10	0	-58	64	4	4	4
11	180	58	96	6	6	6
12	180	-58	96	6	6	6
13	0	90	96	6	6	6
14	0	-90	96	6	6	6
15	0	0	16	1	1	1
16	180	0	96	6	6	6
17	90	0	96	6	6	6
18	270	0	96	6	6	6

Table 8.2: JND thresholds used in the six-level perceptual threshold model to map the binaural cue differences to perceptual impact for each of the 18 reference HRIR pairs.

The numerical differences between each reference and corresponding modified HRIR pair calculated in Section 6.3.1 were mapped to perceptual impact using the six-level model rather than the two-level model. The loudspeaker pairs which exhibit a perceptual difference below 1 JND for all the six statistical metrics used in this work ($\overline{\Delta ITD}$, $\sigma_{\Delta ITD}$, $(\Delta ITD_{max} - \Delta ITD_{min})$, $\overline{\Delta ILD}$, $\sigma_{\Delta ILD}$, $(\Delta ILD_{max} - \Delta ILD_{min})$) were then identified. The positions of these loudspeaker pairs, referred to as the set $\mathbf{LP^{KEMAR}_{CM_6}}$, are indicated in black in Fig. 8.3 at the two radial distances $r = 0.15\,\mathrm{m}$ and $r = 0.20\,\mathrm{m}$.

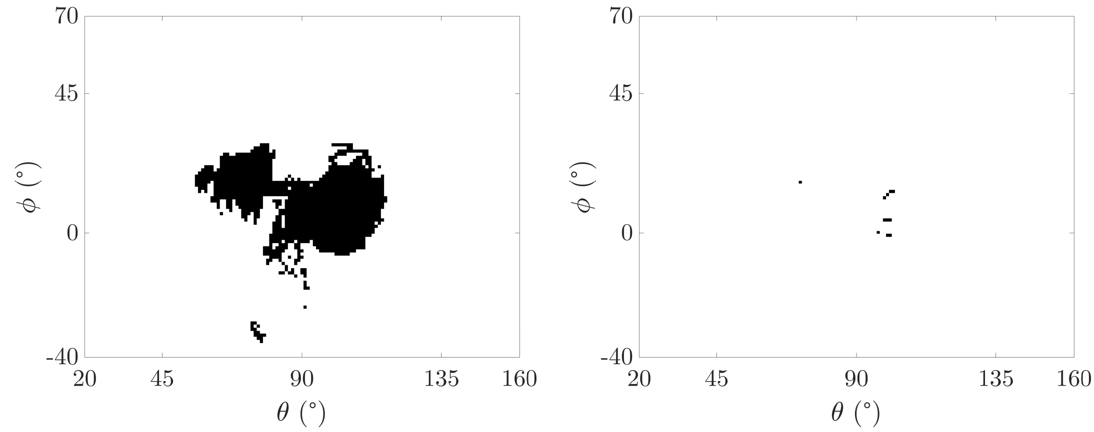
Many more loudspeaker pair positions exist in the set $\mathbf{LP^{KEMAR}_{CM_6}}$ than in the set $\mathbf{LP^{KEMAR}_{CM}}$ (that is, those positions identified using the mappings from the two-level perceptual model, shown in Fig. 6.19 and reproduced in Fig. 8.4 for convenience). At $r = 0.15\,\mathrm{m}$, 8,622 loudspeaker pair positions exist in $\mathbf{LP^{KEMAR}_{CM_6}}$, in the region bounded by $30° < \theta < 136°$, $-40° < \phi < 64°$, whereas in $\mathbf{LP^{KEMAR}_{CM}}$ at $r = 0.15\,\mathrm{m}$ there exist 1,314 positions in the region bounded by $56° < \theta < 117°$, $-35° < \phi < 43°$. Additionally, positions were identified at more radial distances when using the six-level model. The number and percentage of $\mathbf{LP^{KEMAR}_{CM_6}}$ loudspeaker pair positions in the set $\mathbf{LP^{KEMAR}}$ at each radial distance is listed in Table 8.3.



(a) $r = 0.15\,\text{m}$

(b) $r = 0.20\,\text{m}$

Figure 8.3: Regions of loudspeaker pair positions in the set $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM_6}}$ (indicated in black) at the two smallest radial distances. 8,622 such positions exist at $r = 0.15\,\text{m}$. 5,767 positions exist at $r = 0.20\,\text{m}$. The axes limits used in Fig. 6.19 have been included here to enable visual comparison.



(a) $r = 0.15\,\text{m}$

(b) $r = 0.20\,\text{m}$

Figure 8.4: Regions of loudspeaker pair positions in the set $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}}$ (indicated in black), using the same axes limits as in Fig. 8.3. 1,314 such pairs exist at $r = 0.15\,\text{m}$. 11 pairs exist at $r = 0.20\,\text{m}$.

r (m)	$\mathbf{LP^{KEMAR}_{CM_6}}$ Number	%
0.15	8,622	37.1
0.20	5,767	24.8
0.25	4,300	18.5
0.30	3,179	13.7
0.35	1,950	8.4
0.40	1,402	6.0
0.45	1,019	4.4
0.50	737	3.2
0.60	411	1.8
0.70	279	1.2
0.80	189	0.8
0.90	151	0.6
1.00	121	0.5
1.50	60	0.3

Table 8.3: Number and percentage of $\mathbf{LP^{KEMAR}_{CM_6}}$ loudspeaker pair positions within $\mathbf{LP^{KEMAR}}$ at each radial distance.

These preliminary findings indicate that a more sophisticated, less conservative perceptual model identifies a greater number of loudspeaker pairs capable of reproducing satisfactory binaural reproduction without using CTC. It is, therefore, conceivable that a real-world implementation of such a system would perform better than indicated here due to the conservative choices made in the models in this work. Although the six-level model is more sophisticated, it too has limitations. Without evidence to justify which JND values to apply in the mapping from numerical difference to perceptual impact, doubts must remain about its correctness. Additionally, no attempt is made to account for variation in localisation accuracy for sound sources behind the interaural axis: in the six-level model, $6\,\mathrm{JND_{min}}$ is used for all source directions on and behind the interaural axis.

8.3 Further Work

Several directions for work building on that presented here have been identified: acoustic validation of the full torso BEM-suitable mesh of KEMAR; further examination of the relationship between NCS and perceptual performance; investigation into the relationship between candidate loudspeaker pair positions and subject morphology; building a physical acoustic prototype of the simulated system, and further examination of the role of ipsilateral inversion. These will be considered below in more detail.

8.3.1 Validation of the Full Torso Mesh

In Chapter 3, the acoustic validation of a 3D mesh of KEMAR through comparison of acoustically-measured and simulated HRTFs is described. Due to a lack of computational resources available at the time of writing, a head-and-neck version of the mesh (referred to as the Neck-Extended Easily Computable KEMAR, or NEECK) was used for the acoustic validation. The validity of the full torso mesh is inferred from this result, but not explicitly determined. An obvious further step is to simulate HRTFs for the *full-torso-noEC-20* mesh and compare these with accurate acoustically-measured HRTFs, such as those in the SADIE database v2 [74]. It is anticipated that, whilst HRTFs for the full torso will differ from those for the NEECK (as the NEECK lacks shoulders and has an elongated neck), the relationship between simulated and acoustically-measured full torso HRTFs will be similar to the relationship demonstrated between simulated and acoustically-measured NEECK HRTFs. This assertion is based on the fact that the two sets of simulated HRTFs are derived from the same numerically-validated mesh.

8.3.2 Examination of the Relationship between NCS and Perceptual Performance

NCS is the inherent acoustic isolation between the two ears, and occurs primarily as a result of acoustic shadowing introduced by the head (see Section 2.7.2.2). Chapter 5 describes the calculation of NCS for 660,240 loudspeaker positions and the subsequent identification of loudspeaker positions which exhibit NCS of greater than 20 dB. This value has been suggested as the minimum channel separation required for binaural robustness [17, 228]. However, the results presented in Chapter 6 indicate that NCS values corresponding to binaurally-transparent loudspeaker positions range from 24.4 to 36.2 dB, and that some loudspeaker positions with NCS greater than 20 dB are not included in the binaurally-transparent set. It would, therefore, be of interest to explore further the relationship between NCS (and channel separation (CS) more generally) and perceptual performance. In particular, it would be useful to study the potential impact of larger NCS on binaural reproduction, and whether high values of NCS deliver perceptually-superior binaural sound reproduction compared with loudspeaker positions with NCS closer to the accepted 20 dB threshold.

The use of NCS as a predictor of subject localisation performance is also of interest. In comparing the localisation ability of subjects with the achieved CS for various CTC implementations calculated over various frequency ranges, Masiero [227] concluded that CS is only useful as a predictor of localisation performance for certain conditions (specifically, in evaluating lateral errors in mismatched CTC systems using a high-frequency CS). In this work, CS was achieved using various CTC implementations, and so only a limited number of CS values to localisation performance could be compared to subject localisation performance. When calculated using a frequency range of 0.3 to 8 kHz, CS for the mismatched CTC systems across the eight subjects ranged from, on average, 12.3 dB to 16.6 dB. For the matched CTC system, CS was, on average, 68.4 dB [227] (Table 6.5). It would be of interest to extend this type of comparison between CS and subject performance using the more continuous NCS values calculated in this work.

8.3.3 Investigation of the Relationship between Candidate Loudspeaker Pairs and Subject Morphology

Variations in the shape and angular extent of the regions of loudspeaker pair positions meeting the requirements for binaural rendering without CTC were identified in

Chapter 7 for 15 SYMARE subjects. It can be inferred from this that some form of relationship exists between the morphology of a subject and the candidate loudspeaker pair positions (referred to as the $\mathbf{LP^S_{CM}}$ sets, where S refers to the index of the SYMARE subject). To examine this relationship further, two sets of descriptors are required: one to describe the distribution of candidate loudspeaker pair positions in some way, and another to describe morphological aspects of the subject.

A preliminary study was conducted to determine if a simple relationship exists between the number of loudspeaker pair positions in each of the $\mathbf{LP^S_{CM}}$ sets and a selection of parameters relating to the subjects' morphology. The Spearman's rank correlation coefficient[2] was calculated between the number of loudspeaker pair positions in each $\mathbf{LP^S_{CM}}$ set and the following parameters:

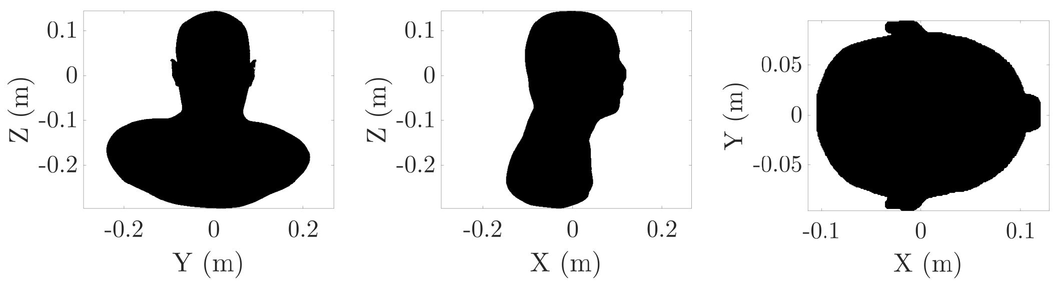
- Volume contained within the mesh (m^3), calculated using the 'meshVolume' function from the geom3d toolbox [289];

- Head width (m), calculated as the distance between the coordinates of the two simulation receiver positions in the occluded ear canals;

- Area of the projection of the mesh in each of the three principal anatomical planes (m^2) (demonstrated for subject #1 in Fig. 8.5 in black);

- Length of the contour around the upper portion of the head (m) when viewed in both the median and frontal planes (demonstrated for subject #1 in Fig. 8.6a and Fig. 8.6b in red);

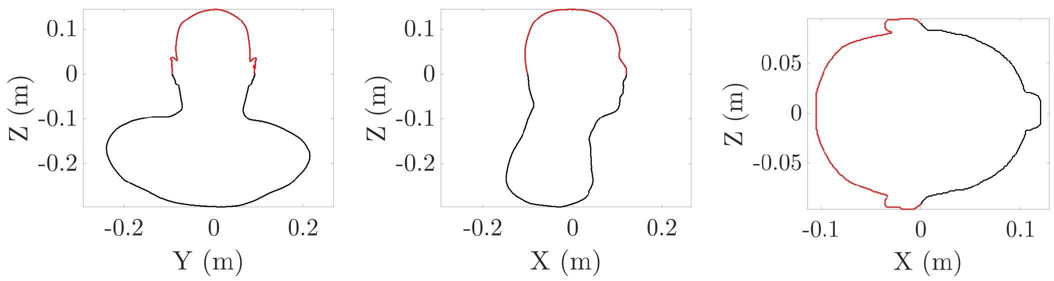
- Length along the contour of the rear portion of the head (m) when viewed in the horizontal plane (demonstrated for subject #1 in Fig. 8.6c in red).



(a) Frontal plane (b) Median plane. (c) Horizontal plane

Figure 8.5: SYMARE subject #1 with mesh surface projected onto each of the three anatomical planes (indicated in black) to allow calculation of the projections' areas.

[2]Kolmogorov-Smirnov test results indicated that the data was not normally distributed and, therefore, that calculation of Pearson correlation coefficients was not appropriate.



(a) Frontal plane (b) Median plane. (c) Horizontal plane

Figure 8.6: SYMARE subject #1 with mesh outline projected onto each of the three anatomical planes to allow calculation of the length of the upper and rear portions of the contour (indicated in red).

Of the eight parameters used to describe the morphology of each subject mesh, only the volume of the meshes displayed a correlation coefficient of greater than 0.5 and a significance value below $p = 0.05$. This suggests that the relationship between 2D projections of a subject and the number of candidate loudspeaker pair positions is weak. It is possible that the number of loudspeaker pair positions in each $\mathbf{LP}^{\mathbf{S}}_{\mathbf{CM}}$ set is not a particularly useful metric. A more spatial description of the positions may prove more useful, such as: the mean distances to the origin, the interaural axis or the ear; the solid angles or the spatial distributions of pair positions; or indeed some mixture of these and other spatial parameters.

It is also possible that the morphological parameters chosen here do not capture sufficient information about the subjects' individuality. Additional morphological dimensions could include those discussed in the CIPIC HRTF database [24]. It is, perhaps, more likely to be the smaller-scale morphological features of the head and pinnae that are of greater importance, particularly those which may increase acoustic shadowing, such as pinna size and ear flare angle, rather than the overall size of the subject mesh.

8.3.4 Development of a Prototype

The development of a physical prototype is a logical progression of the work presented in this thesis. This would not only provide confirmation of the performance of the candidate loudspeaker pairs identified here, but it would also permit examination of the impact of aspects which cannot be considered using a simulated environment. For example, an acoustic CTC-less binaural reproduction system would have physical and mechanical factors that differ from those in the simulated environment, including a loudspeaker enclosure, some sort of mounting system and driver directivity. Additionally, no reflective or absorptive objects are present in the simulated environment other than the mesh of the subject of interest, which is not representative of real listening environments.

Implementation of loudspeakers close to the head would likely involve a headrest of some form, either between the loudspeakers or as an enclosure, with the headrest constructed of absorptive materials. As discussed in Section 2.7.2.3, consideration of the construction materials is perhaps more important within the near-field than for loudspeakers located in the far-field. It is anticipated that the presence of a headrest would increase NCS, and that loudspeakers with increased directivity would reduce crosstalk to the contralateral side. As a result, an acoustic implementation of the VRS may reveal a greater range of satisfactory loudspeaker pair positions than those identified in this work using simulation. Additionally, a physical implementation would allow for perceptual listening tests to replace the perceptual model in the simulation.

8.3.5 Examination of the Role of Ipsilateral Inversion

The work presented in Chapter 6 and Chapter 7 indicates that informed placement of loudspeaker pairs allows for the removal of CTC. This avoids some of the common downsides associated with CTC, including spectral colouration and errors introduced by inverse filtering (see Section 2.7.2.3). However,the VRS used in the simulation work actually includes ipsilateral inversion to maintain binaural fidelity (see Section 6.2.2).

The inverse filters in the VRS were calculated using a high regulation value (in-band regularisation of 20 dB, out-of-band regularisation of −12 dB) and a wide frequency range (50 Hz to 16 kHz) to ensure numerical accuracy. However, should these parameters be used to generate filters for use in an acoustic implementation of the VRS, downsides associated with inverse filtering (such as undue strain on the loudspeakers [16]) may be introduced. A lower regularisation value and a reduced frequency range would usually be more appropriate for acoustic ipsilateral inversion. This reduces the numerical accuracy of the filter in exchange for more stable behaviour. Additionally, the use of a single filter for each ear is likely to become problematic in the presence of listener movement. Dynamic system inversion (where the filters are updated based on tracked listener movement) could be implemented to address this, but its complexity may make it impractical in a commercial implementation. Generic system inversion, where a single set of inverse filters is used for every listener, may also be worthy of further investigation.

Alternatively, the impact of entirely removing the ipsilateral inversion stage could be examined. It may be that loudspeaker positions exist which are capable of satisfactory binaural reproduction without compensating for the ipsilateral acoustic path. Such loudspeaker positions would have to exhibit a sufficiently flat frequency response (comprising both the loudspeaker frequency response and the acoustic path to the ear) to be of negligible perceptual impact on the reproduced binaural signals.

8.4 Conclusion

This chapter discusses the main procedural and perceptual limitations of the work presented in this thesis, and proposes a number of directions for future work.

The implications of using simulated environments and interaural binaural cue estimation methods are considered. The results presented in earlier chapters are inevitably different from what would occur in a real-world acoustic reproduction system, due to important differences between simulated and acoustic environments.

The limitations of the two-level perceptual model used to map numerical differences in binaural cues to estimates of their perceptual impact are discussed. An experimental six-level perceptual model was implemented to investigate the impact of a less conservative model. This model identified a larger number of loudspeaker pair positions for which interaural binaural cue reproduction differences are below the threshold of perception.

These positions lie within a larger angular range and at greater radial distances. However, without evidence to justify which JND values to use in the mapping (unfortunately, to the author's knowledge, such data does not exist), such a model is difficult to implement well in a simulated environment. It is, therefore, likely that an acoustic implementation could perform better than is indicated by either the two-level or six-level simulated systems employed in this work.

Several directions for future work have also been identified. It would be of use to the wider research community for the BEM-suitable full torso mesh of KEMAR to be validated acoustically in addition to numerically. However, this depends on the availability of computational resources to perform the required HRTF simulations. The relationship between NCS and perceptual performance has been addressed somewhat in the literature (for example, in [227]), but could be expanded upon to look at perceptual performance over a more continuous range of NCS values, and in particular to address the question of whether larger NCS leads to significantly more accurate and perceptually improved binaural reproduction. Limited evidence of a relationship between subject morphology and the loudspeaker pair positions identified for CTC-less binaural reproduction has been found. Further understanding of the impact of specific morphological features on binaural reproduction would inform binaural loudspeaker system design. Finally, the advantages of implementing a physical embodiment of the loudspeaker setup within a headrest are discussed, in addition to the examination of the importance of ipsilateral inversion.

This thesis has explored the feasibility of a near-field binaural loudspeaker system which avoids the general requirement for CTC through suitable placement of the loudspeakers. Chapter 9 provides a summary of the work presented and revisits the hypothesis introduced in Chapter 1.

Chapter 9

Conclusions

It's not the time for us to give in.

Tegan and Sara, *I'll Be Back Someday*
Hey, I'm Just Like You (2019)

This thesis explores the feasibility of a near-field binaural loudspeaker system which avoids the general requirement for crosstalk cancellation (CTC) through suitable placement of the loudspeakers. This chapter draws the thesis to a close, summarising the work presented and revisiting the hypothesis introduced in Chapter 1. The contributions to the wider field are detailed, after which the thesis is concluded with some final thoughts.

9.1 Thesis Summary

Chapter 1 presents the historical and cultural context for the thesis, and forms the hypothesis statement and objectives for the work. Chapter 2, the literature review, presents an overview of acoustic principles relevant to psychoacoustics and binaural technology in order to give context to the subject matter of this thesis. The chapter covers: sound wave behaviour; how the human auditory system detects, perceives and localises sound; various techniques for the capture of the acoustic localisation cues, and methods of reproduction of those cues using headphones and loudspeakers. Since this work focuses on the acoustic simulation of localisation cues and the removal of crosstalk cancellation (CTC) from loudspeaker reproduction, emphasis is given to previous simulation work using the boundary element method (BEM) and near-field loudspeaker systems with and without CTC.

The first portion of work presented in the thesis details the creation of 3D mesh models for use within the BEM. Chapter 3 describes the workflow developed to create a 3D mesh model of the Knowles Electronics Manikin for Acoustic Research (KEMAR). During initial investigations, existing meshes of KEMAR available in the literature were deemed not to meet the resolution requirements for the intended work. Therefore, a mesh of higher resolution was created, following a workflow adapted from that described by Jin *et al.* [80]. To ensure topological consistency and acoustic accuracy with the physical manikin, numerical and acoustical validation were performed, and the mesh was shown to be consistent with the original. Chapter 4 describes the development of a simplified, but still topologically accurate, 3D mesh of KEMAR comprising only the head and shoulders. This allows for acoustic simulation using the BEM, as the full torso mesh required more computational resources than were available at the time of writing.

The second portion of work presented in the thesis details the simulation of localisation cues and the implementation of a virtual loudspeaker system. Chapter 5 describes the simulation of head-related transfer functions (HRTFs) of KEMAR for a large number of loudspeaker directions and distances, and the subsequent calculation of natural channel separation (NCS) for each loudspeaker position. Pairs of loudspeaker positions were found which exhibit NCS above the threshold reported in the literature as required for robust binaural reproduction. Chapter 6 describes the creation of a virtual test environment to assess the impact of loudspeaker placement on binaural reproduction. The virtual environment enables simulation of the acoustic paths within a CTC-less loudspeaker reproduction system, and, therefore, allows examination of the deterioration in interaural binaural cues brought about by different loudspeaker positions. Loudspeaker pairs in the vicinity of the interaural axis at close radial distances were identified as introducing less deterioration in binaural cues than the estimated threshold of perception. Chapter 7 discusses the adaptation of the virtual reproduction system to investigate the existence of positions where loudspeaker pairs do not require CTC for a variety of head and ear shapes besides those of KEMAR. Loudspeaker pairs in a few positions were found to perform satisfactorily across all subjects tested.

The final portion of the thesis reflects on broader considerations of the work presented. Chapter 8 discusses limitations of the conclusions, for example, due to the use of simulated environments and a conservative perceptual model. Some initial work on the development of a model more akin to human perception is described. Preliminary results suggest that, in a real-world implementation of such a loudspeaker system, there may be a greater number of acceptable loudspeaker pair positions than indicated by the simulation results. Chapter 8 also proposes some directions for future research. These include: investigation of the relationship between NCS and perceptual performance; exploration of the relationship between loudspeaker performance and subject morphology; the development of an acoustic prototype to further validate the findings; and examination of the role of ipsilateral inversion.

9.2 Contributions

The work presented in this thesis has resulted in several novel contributions to the field, as listed in Chapter 1 and repeated here for convenience:

- Seven high-resolution 3D mesh models of KEMAR, numerically and acoustically validated for use with the BEM and valid up to audio frequencies not previously presented in the literature;
- NCS values for 655,216 loudspeaker positions over 15 radial distances, within both the near-field and the far-field;
- Identification of candidate pairs of loudspeaker positions capable of reproducing satisfactory binaural audio without requiring CTC across multiple subjects,

determined using a simulated environment.

3D Mesh Models of KEMAR: The maximum valid frequency when using the BEM for acoustic simulation is dependent on the resolution of the mesh (see Section 2.6.2.3). However, the meshes described in the literature are either of a lower maximum valid frequency than desired for this work, or they are not available for general use. The work described in Chapter 3 has resulted in an accurate mesh model of KEMAR valid to 20 kHz, from which derivative meshes can be created. In this way, a number of versions of the mesh have also been created: one without ear canals valid to 23 kHz, one without ear canals valid to 16 kHz, two head-and-neck derivatives referred to as the Neck-Extended Easily Computable KEMAR (NEECK), valid to 16 kHz and 20 kHz, and two head-and-shoulder variants valid to 17 kHz and 20 kHz.

NCS Values: NCS has been calculated in the literature for some loudspeaker locations (see Section 2.7.2.2). However, these discussions are often limited to a small set of radial distances. The results presented in Chapter 5 provide a more complete picture, with results for loudspeaker locations in 1° increments for both azimuth and elevation, at 15 radial distances in both the near-field and far-field.

Candidate Loudspeaker Positions: A small number of previous studies have investigated reproduction of binaural audio over CTC-less near-field loudspeakers (see Section 2.7.2.3). However, in general these studies have used only a small number of loudspeaker positions, and have been limited to examining the performance of that specific position rather than the performance of all possible loudspeaker positions, which is needed for determining if an optimal position exists. The results presented in Chapter 6 and Chapter 7 go some way to expanding this discussion. Loudspeaker pairs which do not require CTC to reproduce satisfactory binaural audio have been identified for KEMAR and for 15 human subjects. A small number of loudspeaker pairs which perform satisfactorily for all human subjects have also been identified. It must be noted, however, that the significance of these findings is inevitably restricted by limitations of the simulated environment and of the perceptual models employed in this work.

9.3 Restatement of Hypothesis

As originally stated in Section 1.1, the hypothesis tested within this thesis is as follows:

It is possible to negate the general requirement for crosstalk cancellation when reproducing binaural audio using loudspeakers through suitable placement of the loudspeakers close to the head.

The results presented in this thesis broadly support this hypothesis, although they are affected by the procedures used, specifically by the use of simulated environments and the conservative perceptual model. Since the hypothesis does not specify the environment within which it should be tested, however, it is correct to state that the hypothesis has been validated, provided that the conditions in which it has been tested are clearly stated. The conclusions of this thesis cannot comprehensively argue that CTC-less binaural reproduction is possible in all real-world situations using loudspeakers placed close to the head. Nevertheless, they provide an encouraging and substantive basis upon which further work may build to test a refined hypothesis in more realistic conditions.

9.4 Closing Remarks

This thesis has provided valuable insight into the binaural reproduction performance of loudspeaker pairs in the absence of CTC. The use of the BEM and a virtual reproduction system has enabled examination of a dense set of loudspeaker locations, and allowed application of the procedures to a range of subjects in order to generalise the findings more broadly. It is hoped that this research can provide a foundation for the development of simplified commercial near-field binaural loudspeaker systems.

Appendix A

Derivation of Density as a Function of Temperature

This appendix derives the equation for calculating the density of a medium as a function of temperature from first principles (adapted from [33]), as discussed in Section 2.2 and used in Section 3.6.4.

First, consider the ideal gas law for n_{mol} moles of gas:

$$P_0 V_{gas} = n_{mol} \bar{R} T_K \tag{A.1}$$

where P_0, V_{gas} and T_K are the standard atmospheric pressure (101,325 Pa), the volume of gas (m^3) and the absolute temperature (K), and $\bar{R}$ is the ideal gas constant ($8.3145\,\mathrm{J\,K^{-1}\,mol^{-1}}$) [33].

When formulated for a particular gas, Equation A.1 becomes:

$$P_0 V_{gas} = m_V \hat{R} T_K \tag{A.2}$$

where m_V is the mass contained in V_{gas} in kg, and $\hat{R}$ is the molar-weight-specific gas constant. Density in $\mathrm{kg\,m^{-3}}$ can be calculated using [30]:

$$\rho = \frac{m_V}{V_{gas}} \tag{A.3}$$

Therefore, Equation A.3 can be rewritten as:

$$P_0 = \rho \hat{R} T_K \tag{A.4}$$

and so the density of a particular gas as a function of temperature can be calculated using:

$$\rho = \frac{P_0}{\hat{R} T_K} \tag{A.5}$$

To calculate using a temperature value in Celsius, T_K can be replaced by $T_C + 273.15$.

Appendix B

Mathematics of Mesh Analysis Techniques

This appendix describes the mathematical steps used for numerical validation of the full torso KEMAR mesh in Section 3.5. For prerequisite mathematical knowledge, the reader is directed towards dedicated works on the subject such as [290] (Part II, Programme 6).

B.1 Calculation of Angles between Edges

This calculation can be treated as an example of calculating the angle between two vectors, which is commonly derived from the definition of the dot product of vector **a** and **b**:

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}|\,|\mathbf{b}| \cos\theta_V \tag{B.1}$$

which gives the angle between the vectors, θ_V, as:

$$\theta_V = \cos^{-1}\left(\frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}|\,|\mathbf{b}|}\right) \tag{B.2}$$

This, however, has two problems. High computational power is required for the dot product and for the two norm calculations for the vector magnitudes, and the inverse cosine becomes less accurate with smaller angles [291]. In cases where this calculation will be run many times, an alternative formulation can be used, such as the method described by Simon [292]:

$$\theta_V = \tan^{-1}\left(\frac{|\mathbf{a} \times \mathbf{b}|}{\mathbf{a} \cdot \mathbf{b}}\right) \tag{B.3}$$

This formulation uses only one vector norm calculation; the other is replaced with a cross product, which can be quicker to calculate. More importantly, using inverse tangent avoids the accuracy problems associated with inverse cosine.

B.2 Comparison of the Local Topology of Two Meshes

The algorithm described in Listing 1 to compare the topology of two meshes through calculation of perpendicular distances between faces along the face normal is based on several 3D geometrical principles applied consecutively:

1. The calculation of face centroid and centred normal vector;
2. The calculation of the intersection point between a line and a plane;
3. Determining whether a point on a plane lies within a specified triangle.

B.2.1 Calculation of the Face Centroid and the Centred Normal Vector

The coordinates of the face centroid $\mathbf{V_C}$ are calculated by taking the mean of each trio of x, y and z coordinates of the three vertices of the face:

$$\begin{aligned} V_{C_x} &= \frac{1}{3}(V_{1_x} + V_{2_x} + V_{3_x}) \\ V_{C_y} &= \frac{1}{3}(V_{1_y} + V_{2_y} + V_{3_y}) \\ V_{C_z} &= \frac{1}{3}(V_{1_z} + V_{2_z} + V_{3_z}) \end{aligned} \tag{B.4}$$

where $\mathbf{V_1}$–$\mathbf{V_3}$ are the three vertices of the face. The face normal is calculated by taking the cross product of two edges. The direction of the normal is not of importance in this case, so which edges are used does not matter.

$$\mathbf{v} = \mathbf{a} \times \mathbf{b} \tag{B.5}$$

where $\mathbf{a}$ and $\mathbf{b}$ are edge vectors and $\mathbf{v}$ is the face normal. The centred face normal vector is then created by locating the face normal at the centroid.

B.2.2 Calculation of the Intersection Point between an Infinite Plane and an Infinite Line

The intersection point in 3D space between an infinite line and an infinite plane can be calculated using the method described by Weisstein [293]. An infinite plane defined by the 3D points $\mathbf{V_{P1}}$, $\mathbf{V_{P2}}$ and $\mathbf{V_{P3}}$ and the line passing through points $\mathbf{V_{L1}}$ and $\mathbf{V_{L2}}$ intersect at a point $\mathbf{I}$ which can be found by solving the following set of simultaneous equations:

$$0 = \begin{vmatrix} I_x & I_y & I_z & 1 \\ V_{P1_x} & V_{P1_y} & V_{P1_z} & 1 \\ V_{P2_x} & V_{P2_y} & V_{P2_z} & 1 \\ V_{P3_x} & V_{P3_y} & V_{P3_z} & 1 \end{vmatrix} \tag{B.6}$$

$$I_x = V_{L1_x} + (V_{L2_x} - V_{L1_x})\alpha \tag{B.7}$$

$$I_y = V_{L1_y} + (V_{L2_y} - V_{L1_y})\alpha \tag{B.8}$$

$$I_z = V_{L1_z} + (V_{L2_z} - V_{L1_z})\alpha \tag{B.9}$$

This gives the constant α in terms of the known coordinates as:

$$\alpha = -\frac{\begin{vmatrix} 1 & 1 & 1 & 1 \\ V_{P1_x} & V_{P2_x} & V_{P3_x} & V_{L1_x} \\ V_{P1_y} & V_{P2_y} & V_{P3_y} & V_{L1_y} \\ V_{P1_z} & V_{P2_z} & V_{P3_z} & V_{L1_z} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 & 0 \\ V_{P1_x} & V_{P2_x} & V_{P3_x} & (V_{L2_x} - V_{L1_x}) \\ V_{P1_y} & V_{P2_y} & V_{P3_y} & (V_{L2_y} - V_{L1_y}) \\ V_{P1_z} & V_{P2_z} & V_{P3_z} & (V_{L2_z} - V_{L1_z}) \end{vmatrix}} \tag{B.10}$$

which can then be found from the coordinate values and substituted into Equations B.7–B.9 to find the point of intersection $\mathbf{I}$.

In this work, the coordinates of the plane $\mathbf{V_{P1}}$, $\mathbf{V_{P2}}$ and $\mathbf{V_{P3}}$ are the three vertices of a triangular face. However, the calculation assumes that the plane and line are infinite, so an intersection point could exist which is outside the face of interest. A further step is required to ensure that intersections only between the coordinates (that is, within the face) are included.

B.2.3 Testing the Validity of the Intersection Point

The test for whether a point of intersection lies within a designated face is taken from [294], using the 'barycentric' system of describing triangles.

Consider a triangle with vertices $\mathbf{V_1}$, $\mathbf{V_2}$, $\mathbf{V_3}$, then consider a weight attached to each vertex. Each weight has 100 % contribution from that vertex, and 0 % contribution from the other vertices. A position anywhere within the triangle can be thought of as a contribution of the three weights. The barycentre $\mathbf{V_B}$ of the triangle is the point where the weights are balanced; each weight is contributing equally at that point (this is equal to the centroid as described in Section B.2.1). Assigning each weight to be equal to 1, this can be formulated as:

$$\mathbf{V_B} = \frac{1}{3}\mathbf{V_1} + \frac{1}{3}\mathbf{V_2} + \frac{1}{3}\mathbf{V_3} \tag{B.11}$$

From this, the coordinates for the barycentre $\mathbf{V_B}$ can be read as $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. The barycentric axes of a triangle are those which extend from a triangle edge, where the weight for the opposite vertex is 0, through the barycentre and to the opposite vertex, where the weight for that vertex is 1. This is shown in Fig. B.1, where $\mathbf{V_B}$ is the barycentre of a triangle with vertices $\mathbf{V_1},\ \mathbf{V_2},\ \mathbf{V_3}$ and the barycentric axes are indicated by dotted lines.

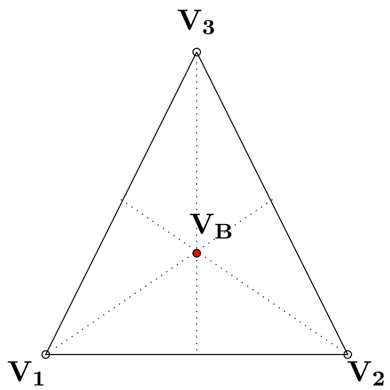


Figure B.1: Barycentric axes (dotted lines) and barycentre $\mathbf{V_B}$ of a triangle with vertices $\mathbf{V_1},\ \mathbf{V_2},\ \mathbf{V_3}$.

For each barycentric axis, the values can only range from 0 (on an edge) to 1 (at the vertex reached after passing through the barycentre). It therefore follows that a value less than zero or greater than one is outside the triangle of interest. For any point of interest $\mathbf{V_Q}$, a barycentric equation can be formed:

$$\mathbf{V_Q} = s\mathbf{V_1} + r\mathbf{V_2} + t\mathbf{V_3} \tag{B.12}$$

where s, r and t are the barycentric coefficients which must sum to 1:

$$\begin{aligned} 1 &= s + r + t \\ &\therefore \\ s &= 1 - r - t \end{aligned} \tag{B.13}$$

Substituting this into Equation B.12 gives

$$\begin{aligned} \mathbf{V_Q} &= (1 - r - t)\mathbf{V_1} + r\mathbf{V_2} + t\mathbf{V_3} \\ &= \mathbf{V_1} - r\mathbf{V_1} - t\mathbf{V_1} + r\mathbf{V_2} + t\mathbf{V_3} \\ &= \mathbf{V_1} + r(\mathbf{V_2} - \mathbf{V_1}) + t(\mathbf{V_3} - \mathbf{V_1}) \end{aligned} \tag{B.14}$$

Equation B.14 defines the relationship between point of interest $\mathbf{V_Q}$ and the triangle with vertices $\mathbf{V_1},\ \mathbf{V_2},\ \mathbf{V_3}$ in terms of coordinates. To allow calculation using vector-based maths, this can be reformulated in terms of vectors by defining $\mathbf{V_1}$ as the local origin, and thus the following relationships:

$$\boldsymbol{u} = \mathbf{V_2} - \mathbf{V_1} \tag{B.15}$$

$$\boldsymbol{v} = \mathbf{V_3} - \mathbf{V_1} \tag{B.16}$$

$$\boldsymbol{w} = \mathbf{V_Q} - \mathbf{V_1} \tag{B.17}$$

Rearranging Equation B.14 and substituting in these relationships gives:

$$\boldsymbol{w} = r\boldsymbol{u} + t\boldsymbol{v} \tag{B.18}$$

Using vector relationships, r and t can then be calculated from:

$$r = \frac{\boldsymbol{v} \times \boldsymbol{w}}{\boldsymbol{v} \times \boldsymbol{u}} \tag{B.19}$$

$$t = \frac{\boldsymbol{u} \times \boldsymbol{w}}{\boldsymbol{u} \times \boldsymbol{v}} \tag{B.20}$$

The dot product is used to determine if r or t are greater than zero. If either r or t are less than zero, the dot product between the respective cross products for each will return a negative value. If this is the case, the intersection is not valid.

$$sign(r) = (\boldsymbol{v} \times \boldsymbol{w}) \cdot (\boldsymbol{v} \times \boldsymbol{u}) \tag{B.21}$$

$$sign(t) = (\boldsymbol{u} \times \boldsymbol{w}) \cdot (\boldsymbol{u} \times \boldsymbol{v}) \tag{B.22}$$

If r and t are greater than zero, their values are calculated and the magnitudes summed to ensure $\|r\| + \|t\| < 1$. If this is not the case, the intersection is not valid.

$$\|r\| = \left\| \frac{\boldsymbol{v} \times \boldsymbol{w}}{\boldsymbol{v} \times \boldsymbol{u}} \right\| \tag{B.23}$$

$$\|t\| = \left\| \frac{\boldsymbol{u} \times \boldsymbol{w}}{\boldsymbol{u} \times \boldsymbol{v}} \right\| \tag{B.24}$$

$$\mathsf{sum} = \|r\| + \|t\| \tag{B.25}$$

If all three conditions are met, the point of interest falls within the triangle.

Appendix C

NEECK HRTF Measurement and Simulation Apparatus

Section 3.6 describes the measurement and simulation of head-related transfer functions (HRTFs) for the purpose of acoustic validation of a 3D mesh model of the Knowles Electronics Manikin for Acoustic Research (KEMAR). This appendix provides technical details of the measurement rig and of the physical and mesh versions of the mounting unit used to hold the head-only section of KEMAR, which, when combined, is referred to as the Neck-Extended Easily Computable KEMAR (NEECK).

C.1 HRTF Measurement Rig

A lightweight loudspeaker rig was designed using SketchUp [295] and constructed to allow the acoustic measurement of HRTFs at five elevation angles at a radial distance of $r = 1.20$ m. The SketchUp design is shown in Fig. C.1. Five lengths of 20x20 mm aluminium extrusion[1] were cut, two at 382 mm and three at 685 mm, and arranged at 150° to each other (shorter bar at each end). These were affixed to one another using a custom-designed bracket on either side which were laser cut from 6 mm ply (shown in Fig. C.2). The lengths and angles were designed such that, when fixed together, loudspeakers mounted at the centre point of each bar were positioned at 30° increments at a radial distance of $r = 1.20$ m. The vertical portions in Fig. C.1, extending upward from the rig, indicate the points on the frame where the rig was suspended from the ceiling of the anechoic chamber, such that the horizontal plane loudspeaker was 1.3 m above floor level (approximately 30 cm from the tip of the acoutic wedges).

[1]For an example of aluminium extrusion, see [296].

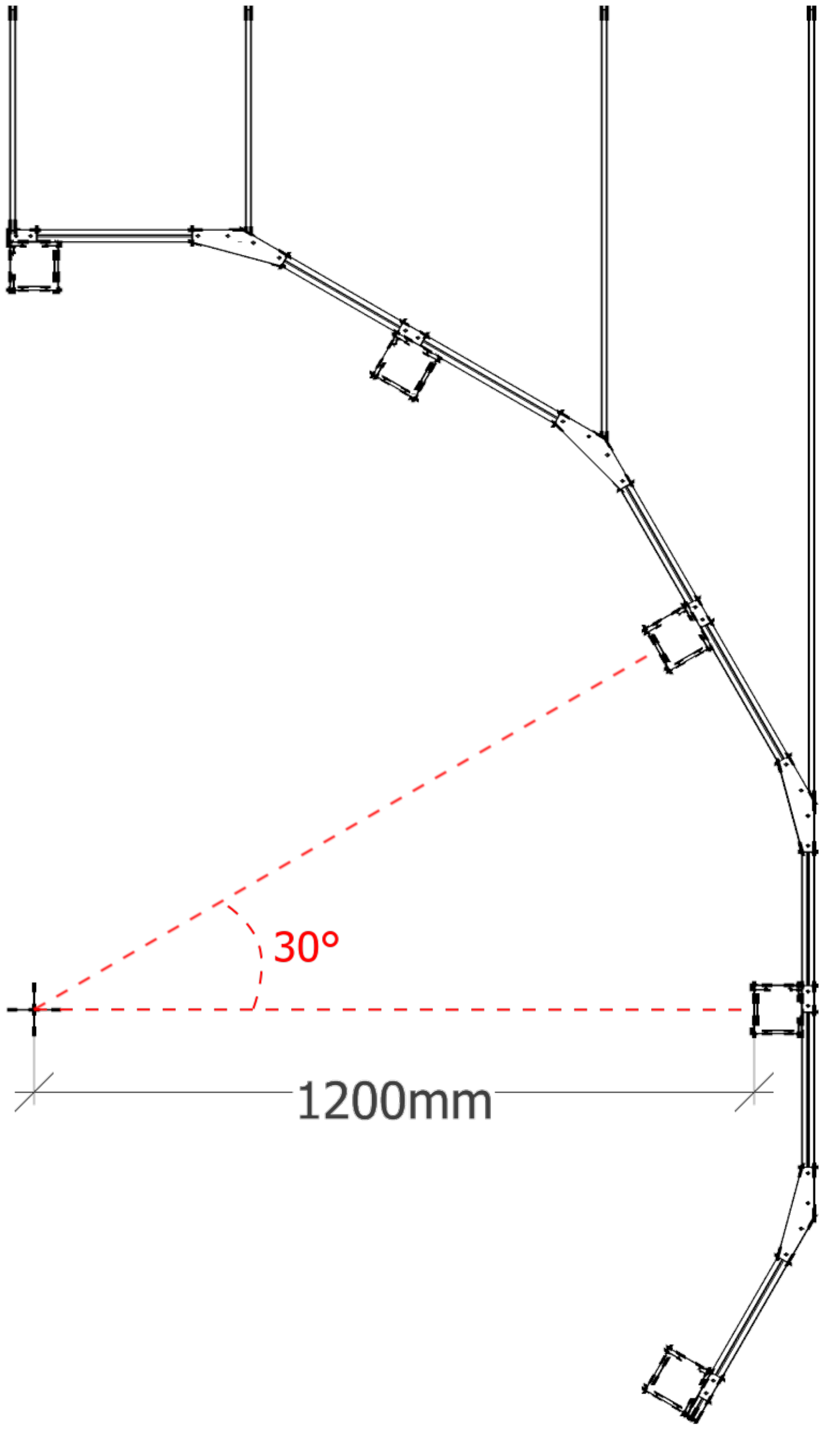


Figure C.1: Design of the acoustic HRTF measurement rig. The vertical portions that can be seen extending upward from the rig indicate the points on the frame where the rig was suspended from the ceiling of the anechoic chamber.

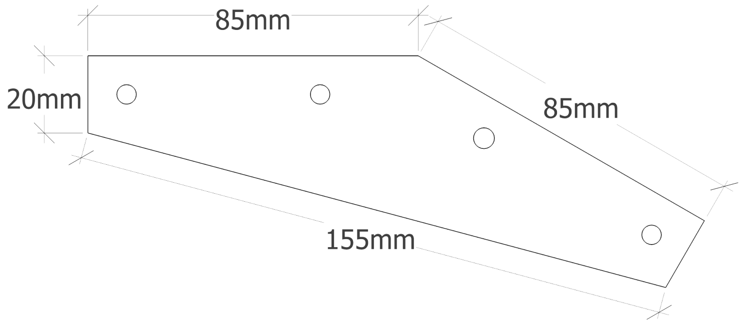


Figure C.2: Design of the 6 mm laser ply bracket designed to hold lengths of aluminium extrusion at 150°.

A cube was designed in SketchUp to mount each loudspeaker driver, the net for which is shown in Fig. C.3. The nets for five cubes were laser cut from the same 6 mm ply as the brackets, formed into cubes (Fig. C.4) and a Tectonic Elements BMR driver (TEBM46C20N-4B) [267] mounted on the front of each. The connections were soldered to post terminals attached to the back piece of each cube, the air gap stuffed with mineral wool, and the pieces glued together to form the loudspeaker enclosures (Fig. C.5). The protruding pieces at the back of the loudspeaker enclosures were designed to snugly attach to the aluminium extrusion using slot nuts[2]. The five loudspeaker cubes were mounted to each length of aluminium extrusion and the angle and radial distance of each was confirmed. The exposed portions of the rig were then weighted to reduce low-frequency emissions and covered with acoustic foam to reduce the impact of reflections and cable vibration. The complete rig is shown in Fig. 3.18b.

[2]For an example of slot nuts, see [297].

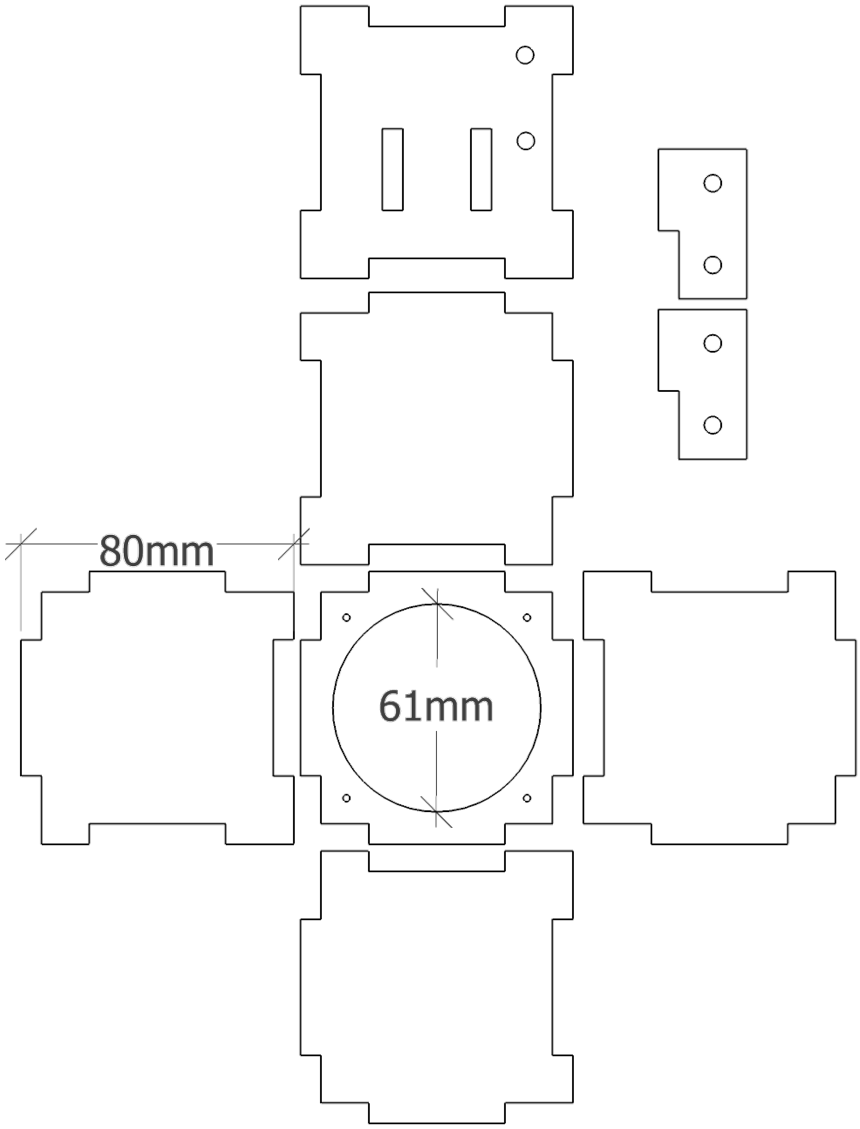


Figure C.3: Design of the net for the 80 mm cube loudspeaker enclosure.

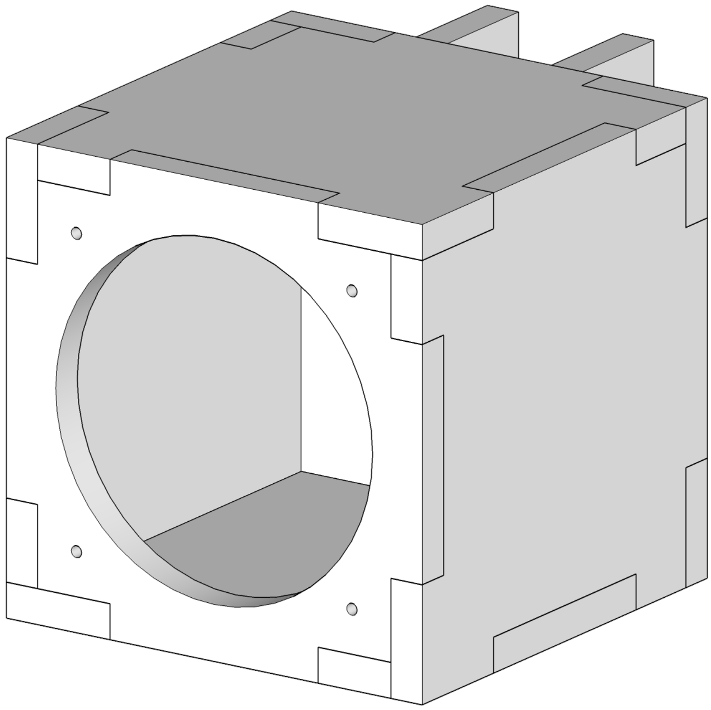


Figure C.4: Design of the cube loudspeaker enclosure before mounting the loudspeaker driver.

Figure C.5: Complete loudspeaker unit, including the Tectonic Elements BMR loudspeaker driver.

C.2 NEECK Mounting Unit

The NEECK mounting unit (shown in Fig. 3.16 and reproduced in Fig. C.6 for reference) was constructed to allow acoustic measurement of HRTFs of a comparable subject to that which could be simulated using computational resources available at the AudioLab. The mounting unit comprised several 5 mm acrylic rings which could be screwed together using M3 threaded rods and nuts. These were then screwed to the KEMAR microphone stand mounting plate (shown in Fig C.7) with the cables passing through a gap in the rings, and slotted over the ribbed portion on the underside of the head-only KEMAR. Laser cut acrylic was chosen, as it is both lightweight and stable.

(a) Head section of KEMAR

(b) Mounting unit.

(c) Head section of KEMAR mounted on a microphone stand.

Figure C.6: Mounting unit created to hold the head only portion of KEMAR.

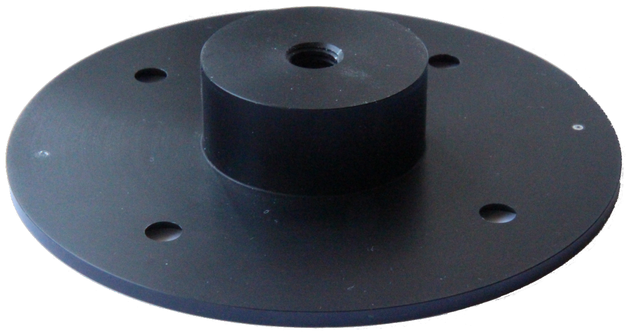


Figure C.7: Microphone mounting stand from the KEMAR kit.

Five types of rings were used (shown in Fig. C.8). Rings D and E are the top-most and bottom-most rings, respectively, and both have small counterbored spaces to accommodate the M3 nuts required for screwing the set of rings together. Ring B has a 5 mm gap in the rear side to allow the cables from the KEMAR microphones to pass through the unit. The majority of the unit comprises ring types A and C.

The rings had an exterior diameter of 114 mm to match the outer diameter of the neck of KEMAR and ensure a smooth transition between the head of KEMAR and the mounting unit. Two inner diameters were used. Rings A, B and D had an inner diameter of 90 mm to fit over the ribbed portion of the KEMAR head. Rings C and E had an inner diameter of 40 mm to fit over the microphone stand mounting plate. 4 mm holes allowed the M3 threaded rod to pass through all rings, and 7 mm holes in ring types C and E allowed the microphone stand mounting plate to be screwed to the underneath using M5 screws. In total, 18 rings were cut (12 of A, 2 of B and C, and one each of D and E) and were constructed in the following order from top to bottom: D, 12 x A, 2 x B, 2 x C, E, with E rotated such that the counterbored spaces for the nuts are on the underside. This construction is shown in Fig. C.9, with views from above and below shown in Fig. C.10.

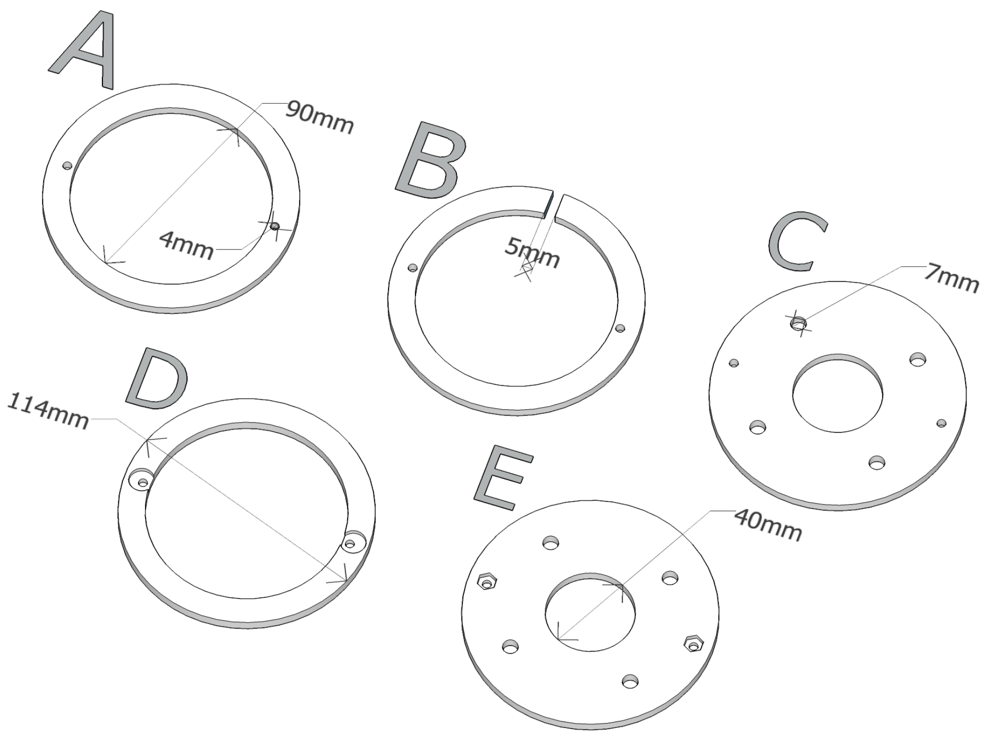
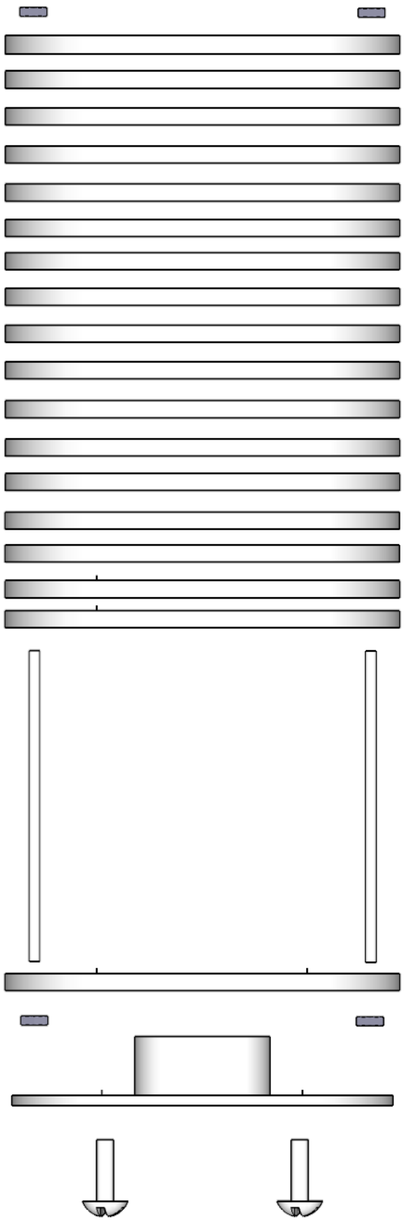
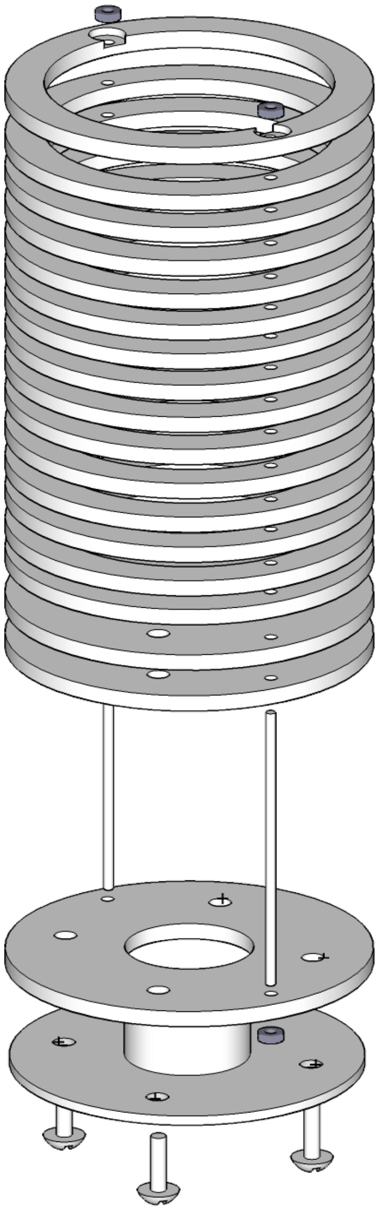


Figure C.8: Design of the five types of ring used in construction of the NEECK.

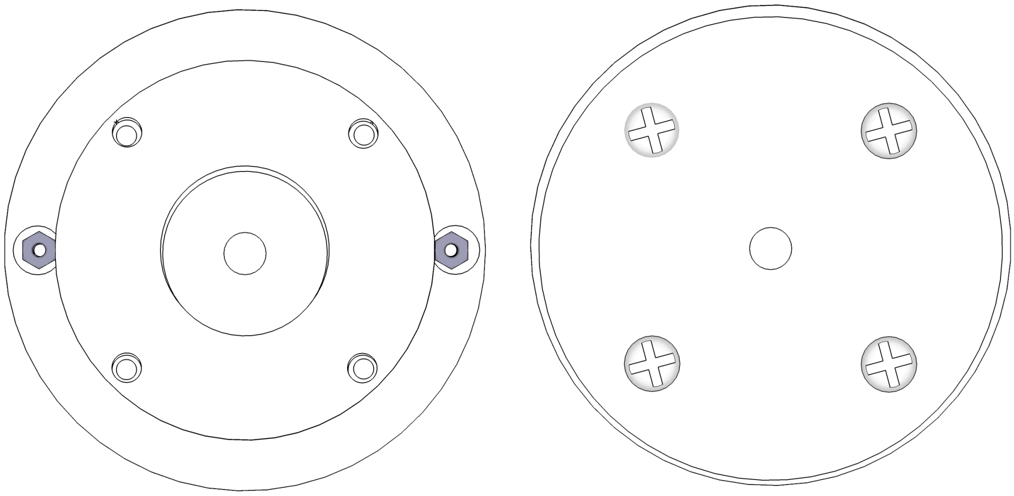


(a) Front view



(b) Isometric view

Figure C.9: Exploded view of the components of the NEECK mounting unit.



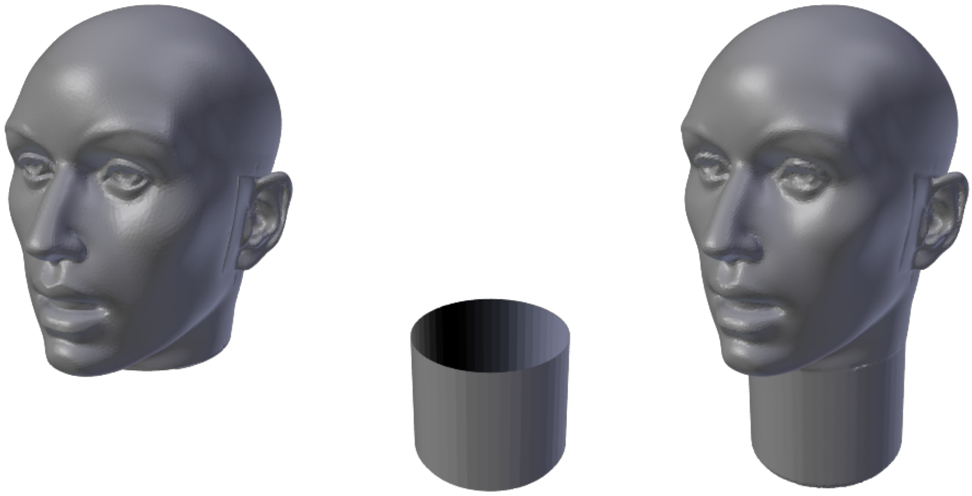
(a) Above (b) Below

Figure C.10: Views of the constructed NEECK mounting unit.

C.3 NEECK Mesh

In order to simulate HRTFs of the physical NEECK, a mesh representation was required (shown in Fig. 3.17 and reproduced in Fig. C.6 for reference). The head portion of the *original-polygonal* mesh was sliced at the collar using the ‘Trim with XY-plane’ tool in Geomagic Wrap (Fig. C.11a). An angle of 7° was required to match the line of the collar. However, this angle was rotated to 0° in both the physical and mesh NEECK as the stabilising base plate shown in Fig. C.6 is secured horizontally onto the microphone stand.

Two open cylinders matching the dimensions of the bolted rings and the mounting plate were created with 50 segments in Autodesk 3DS Max and attached using the ‘Edit Geometry → Attach’ tool. Four additional hemispheres of 6 mm radius were created to mimic the screws. These were aligned and joined using the ‘Fill Hole → Bridge’ and ‘Fill Hole → Complete’ tools in Geomagic Wrap to create a mesh of the mounting unit (Fig. C.11b). The open top of the mounting unit mesh was aligned with the open underside of the sliced head mesh, maintaining the origin at the centre of the head, and the two were joined to form a single mesh (Fig. C.11c).



(a) Mesh of head section of KEMAR

(b) Mesh of mounting unit.

(c) Mesh representation of complete NEECK unit.

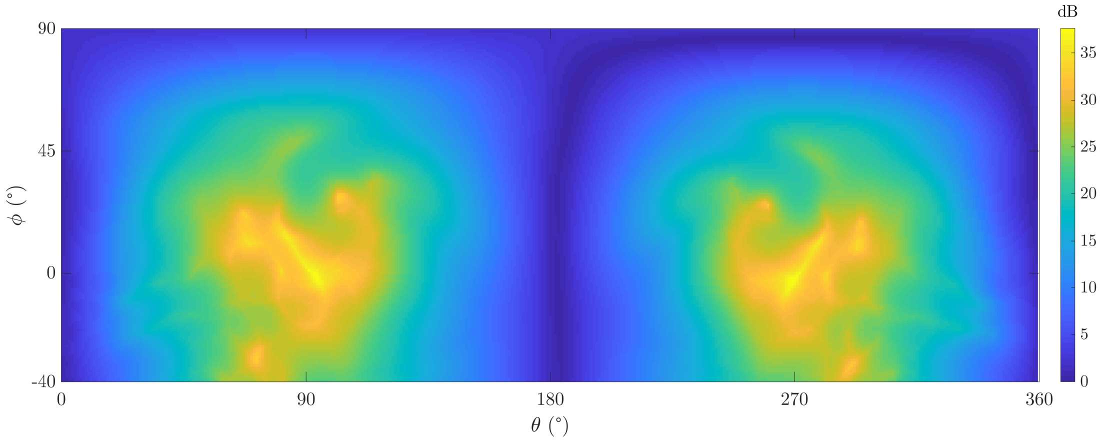
Figure C.11: Mesh representation of the NEECK.

Appendix D

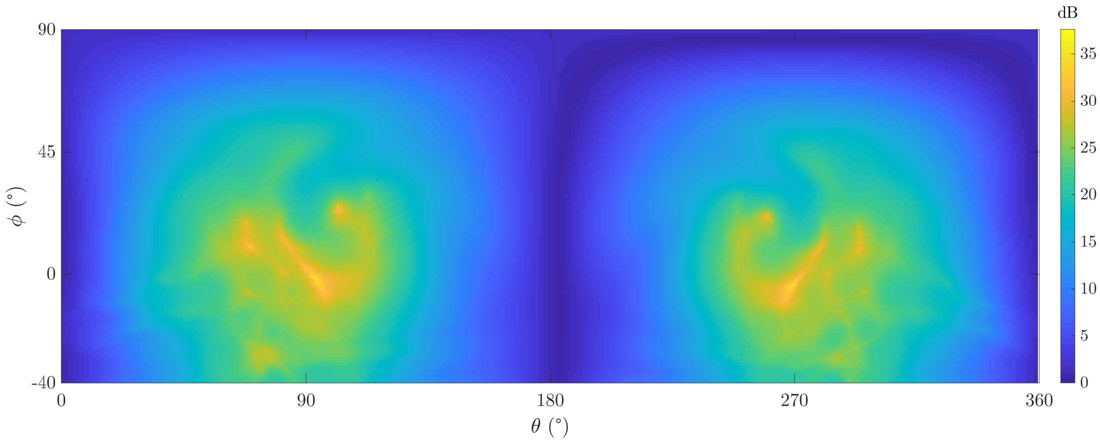
Additional Figures: NCS (larger)

This appendix contains larger versions of the figures plotted in Chapter 5:

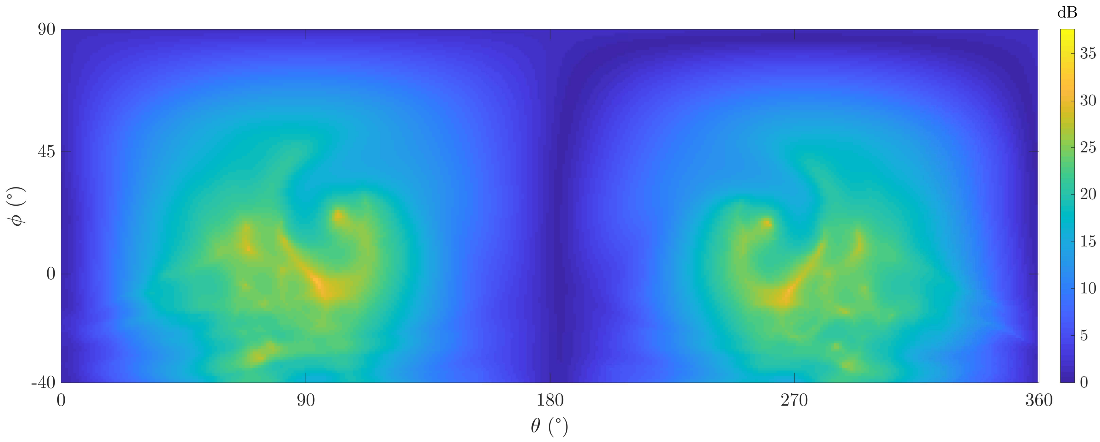
- Variation of natural channel separation (NCS) as a function of direction in heatmap form (discussed in Section 5.3.1);
- Variation of NCS as a function of direction in contour form (discussed in Section 5.3.1);
- Variation of normalised NCS as a function of direction in heatmap form (discussed in Section 5.3.2);
- Regions of loudspeakers which exhibit NCS of greater than 20 dB (referred to as the set $\mathbf{SP_{20}^{KEMAR}}$, discussed in Section 5.4).



(a) $r = 0.15\,\mathrm{m}$

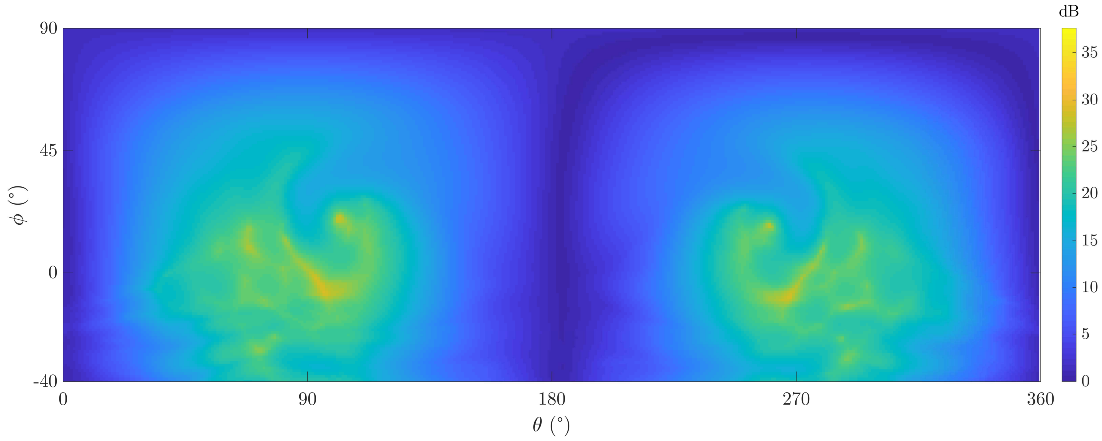


(b) $r = 0.20\,\mathrm{m}$

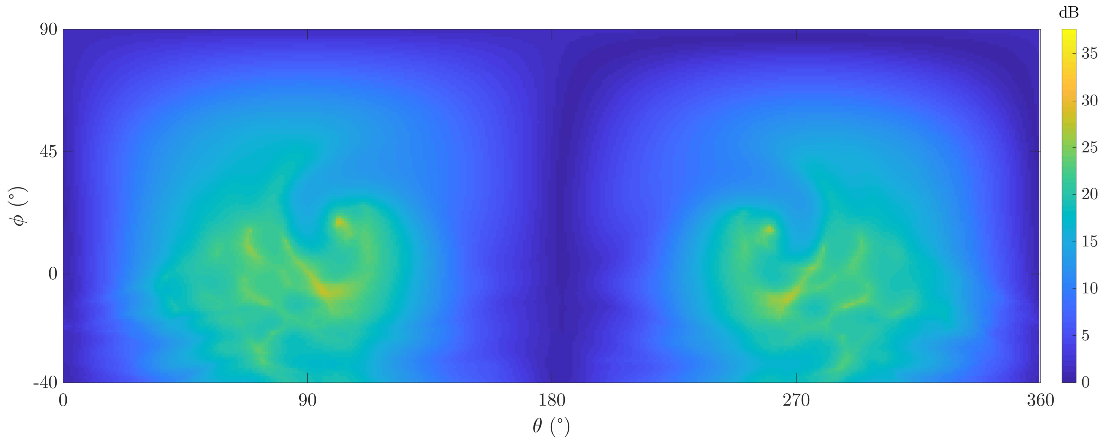


(c) $r = 0.25\,\mathrm{m}$

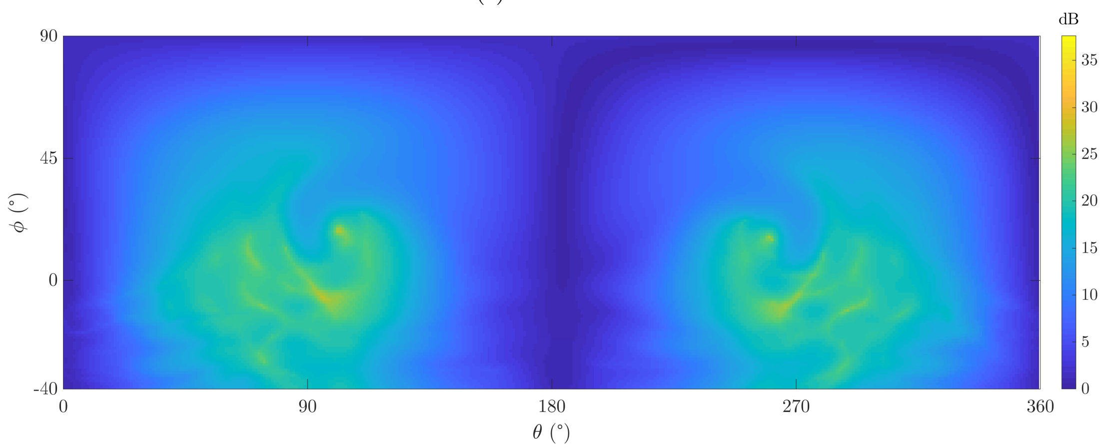
Figure D.1: NCS plotted as intensity surfaces for domes within $\mathbf{SP^{KEMAR}}$ for radial distances 0.15 m to 0.25 m. *Figure continued overleaf.*



(d) $r = 0.30\,\text{m}$

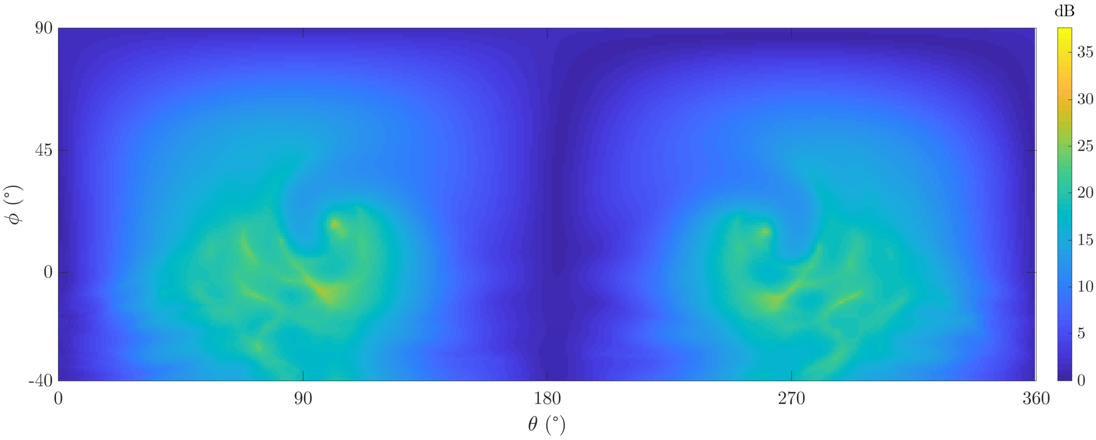


(e) $r = 0.35\,\text{m}$

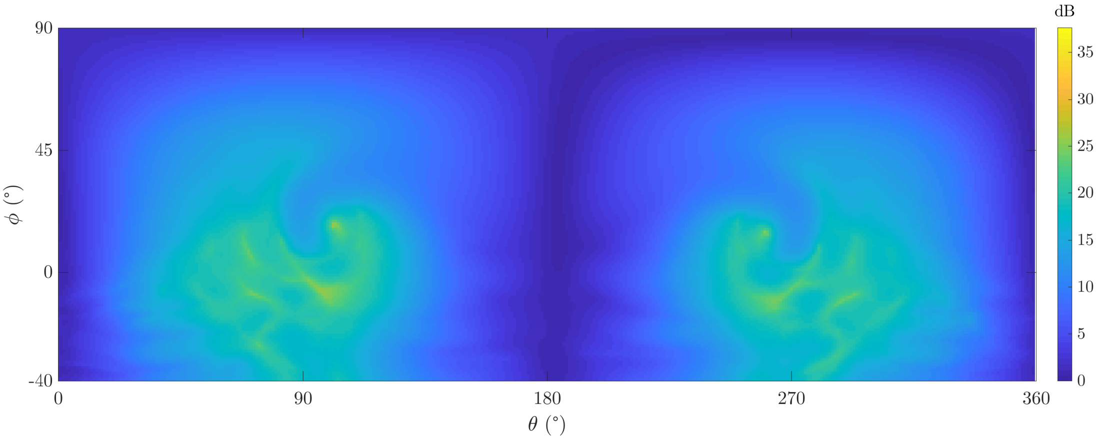


(f) $r = 0.40\,\text{m}$

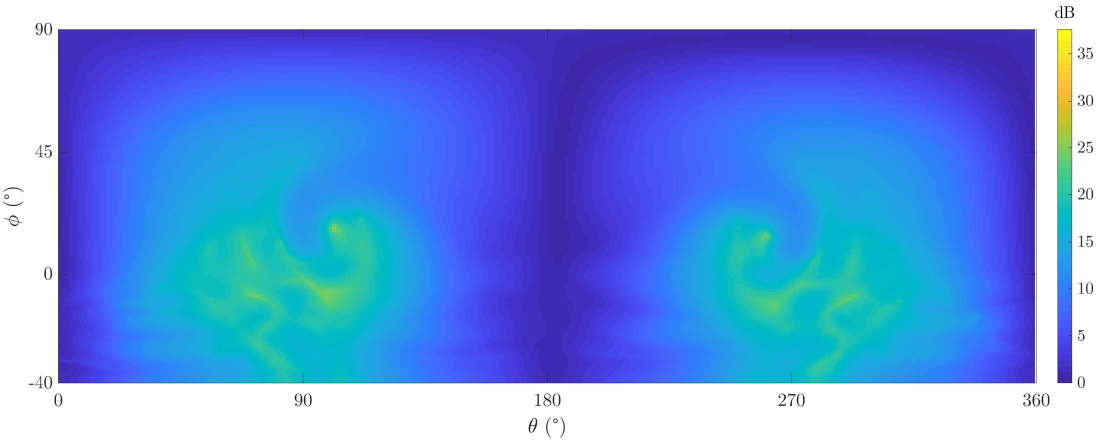
Figure D.1: NCS plotted as intensity surfaces for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 0.30 m to 0.40 m. *Figure continued overleaf.*



(g) $r = 0.45\,\mathrm{m}$

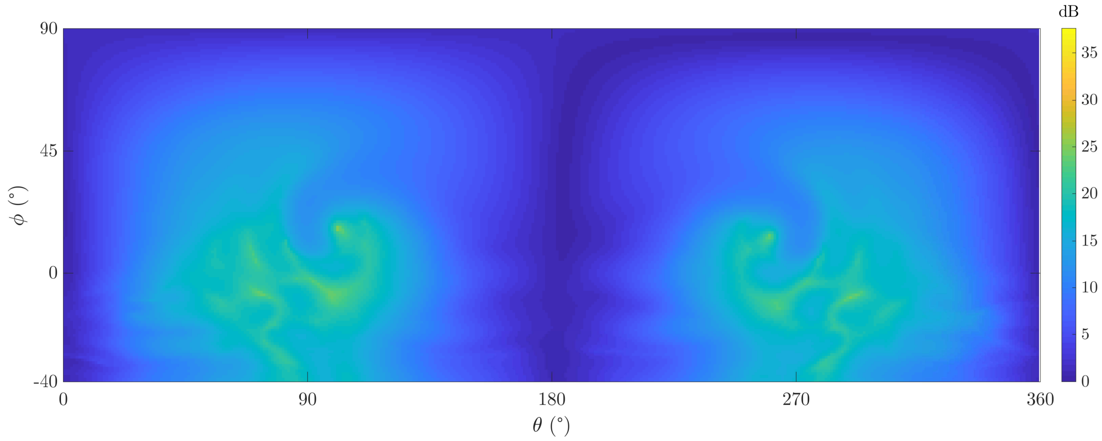


(h) $r = 0.50\,\mathrm{m}$

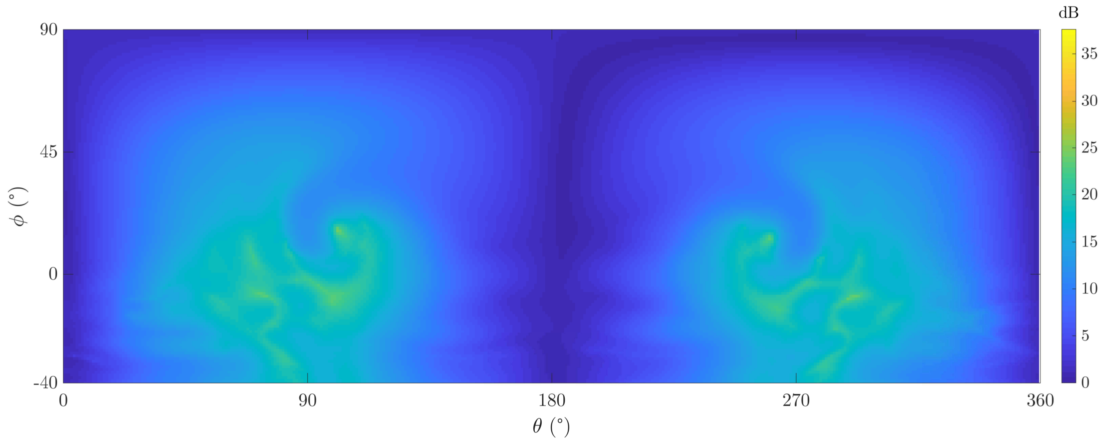


(i) $r = 0.60\,\mathrm{m}$

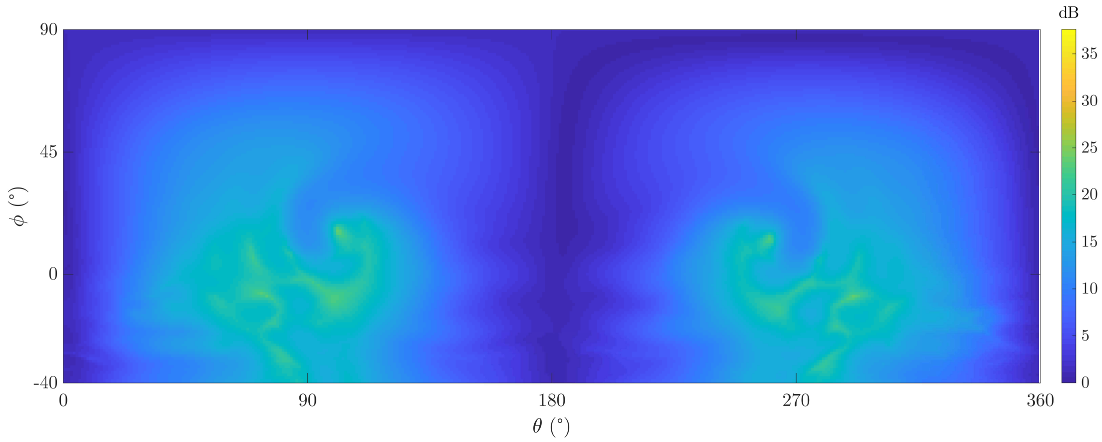
Figure D.1: NCS plotted as intensity surfaces for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 0.45 m to 0.60 m. *Figure continued overleaf.*



(j) $r = 0.70\,\mathrm{m}$

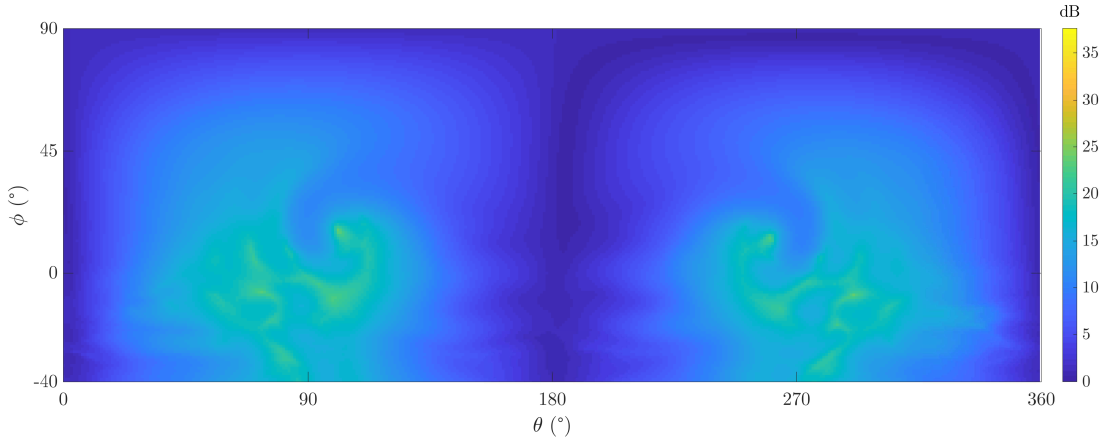


(k) $r = 0.80\,\mathrm{m}$

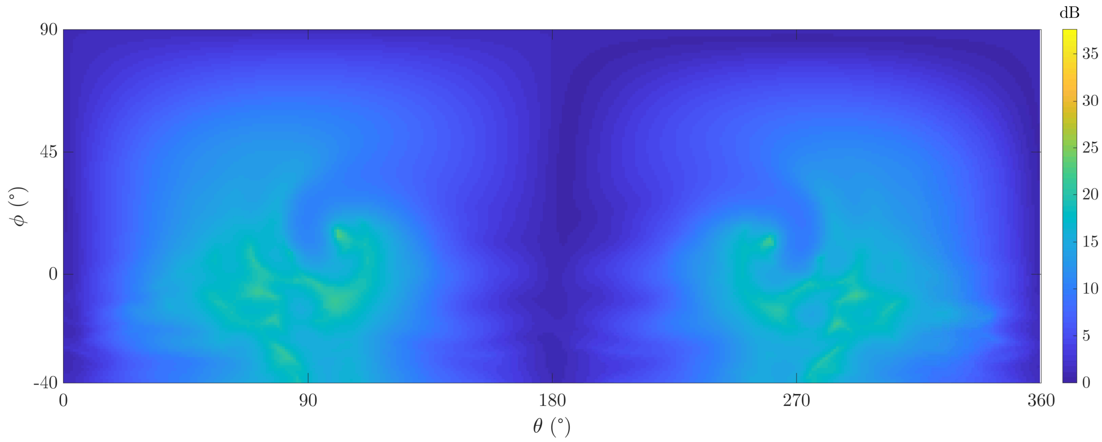


(l) $r = 0.90\,\mathrm{m}$

Figure D.1: NCS plotted as intensity surfaces for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 0.70 m to 0.90 m. *Figure continued overleaf.*

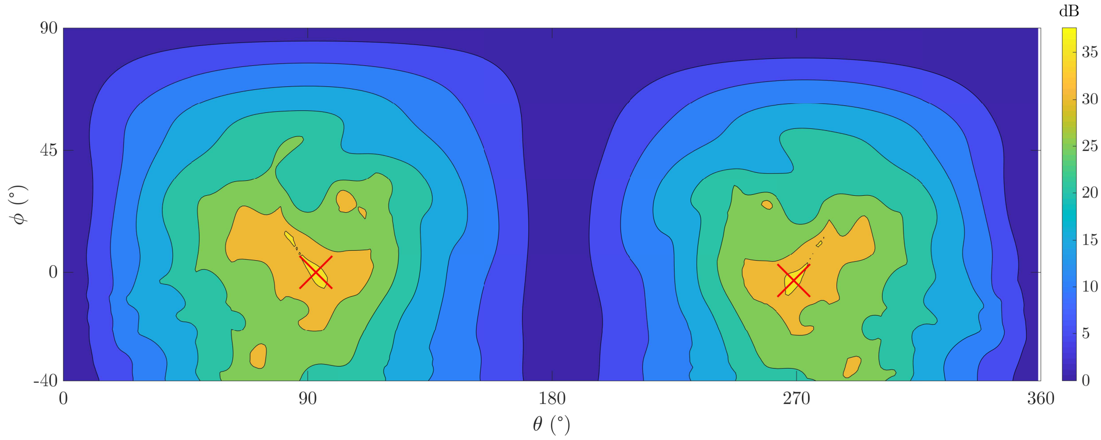


(m) $r = 1.00\,\mathrm{m}$

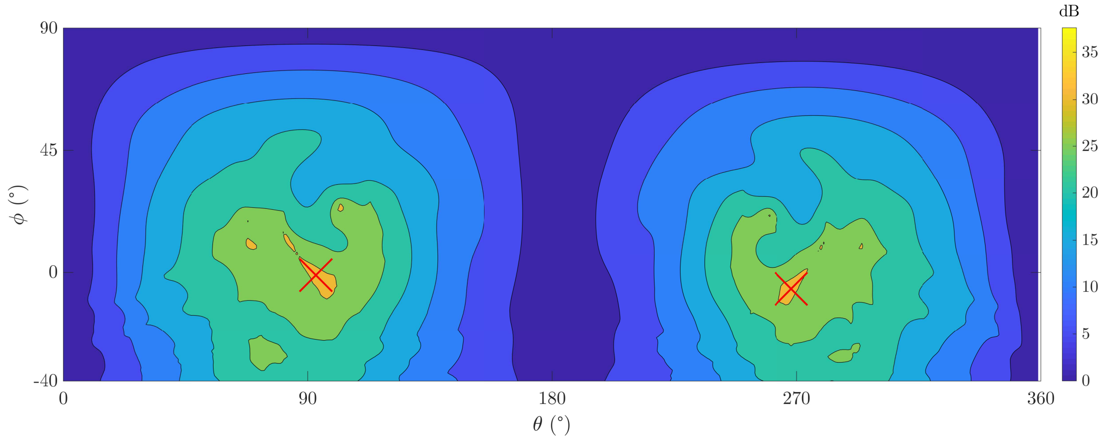


(n) $r = 1.50\,\mathrm{m}$

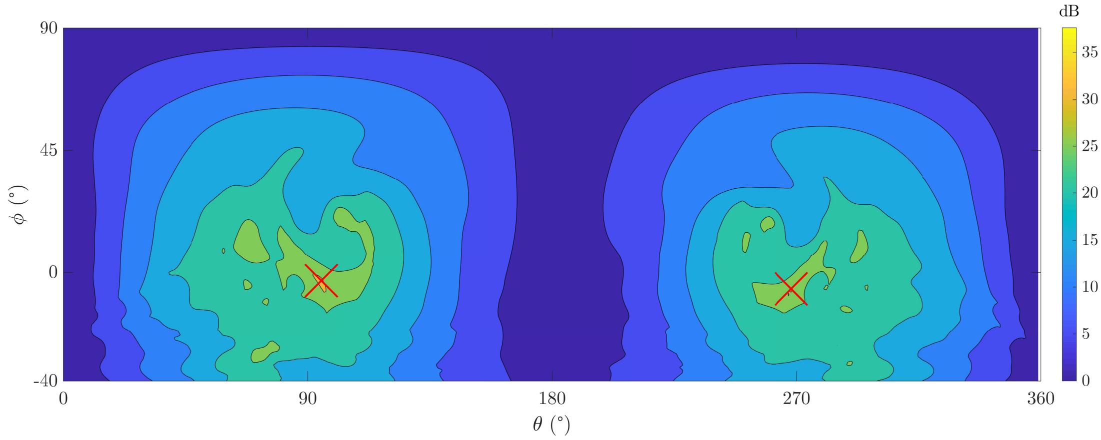
Figure D.1: NCS plotted as intensity surfaces for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 1.00 m and 1.50 m.



(a) $r = 0.15\,\mathrm{m}$

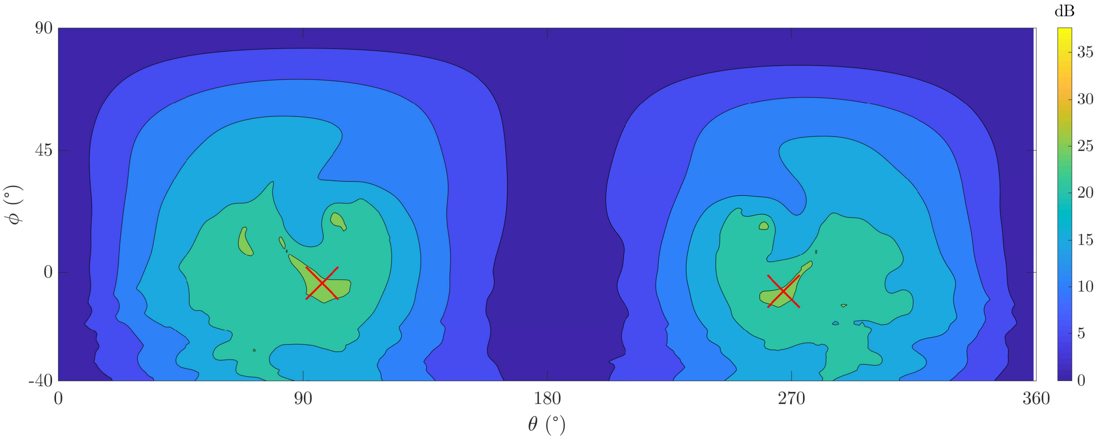


(b) $r = 0.20\,\mathrm{m}$

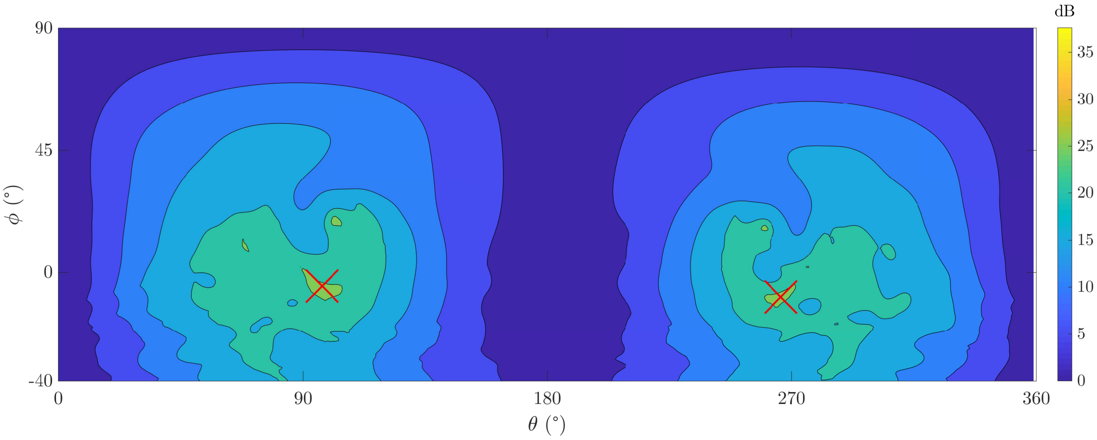


(c) $r = 0.25\,\mathrm{m}$

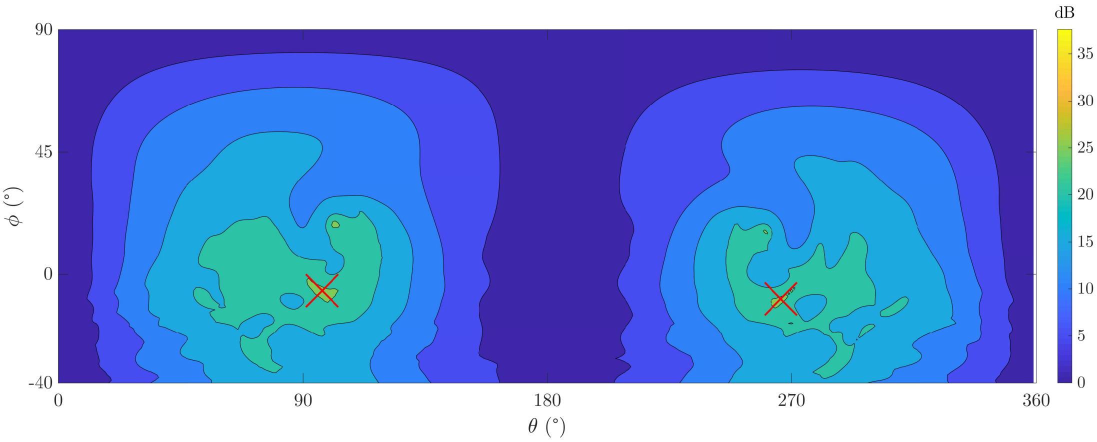
Figure D.2: NCS plotted as contour plots (increments of 5 dB) for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 0.15 m to 0.25 m. The centroid of each region with the largest threshold value is indicated with a red cross. *Figure continued overleaf.*



(d) $r = 0.30\,\mathrm{m}$

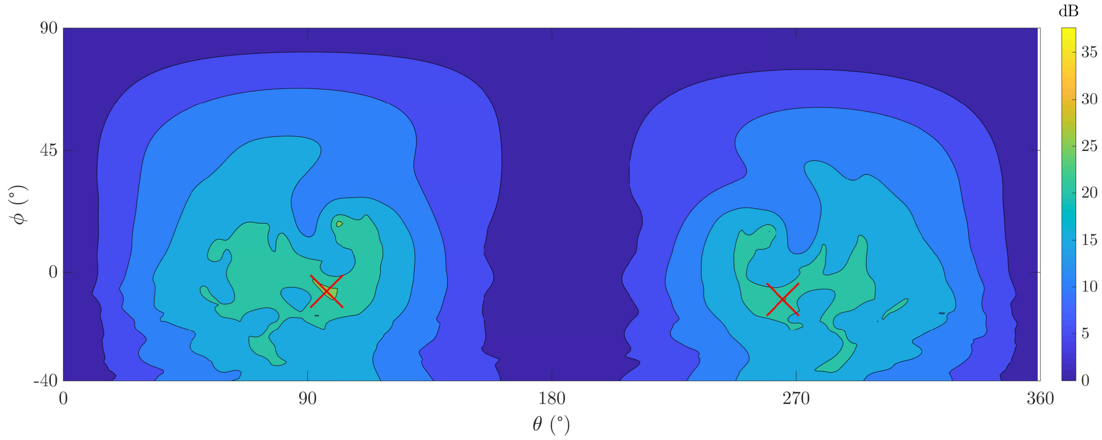


(e) $r = 0.35\,\mathrm{m}$

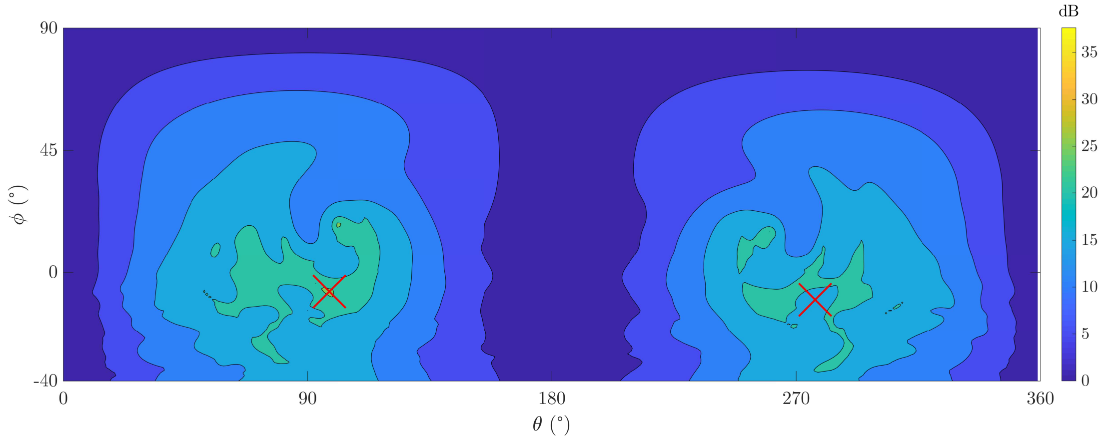


(f) $r = 0.40\,\mathrm{m}$

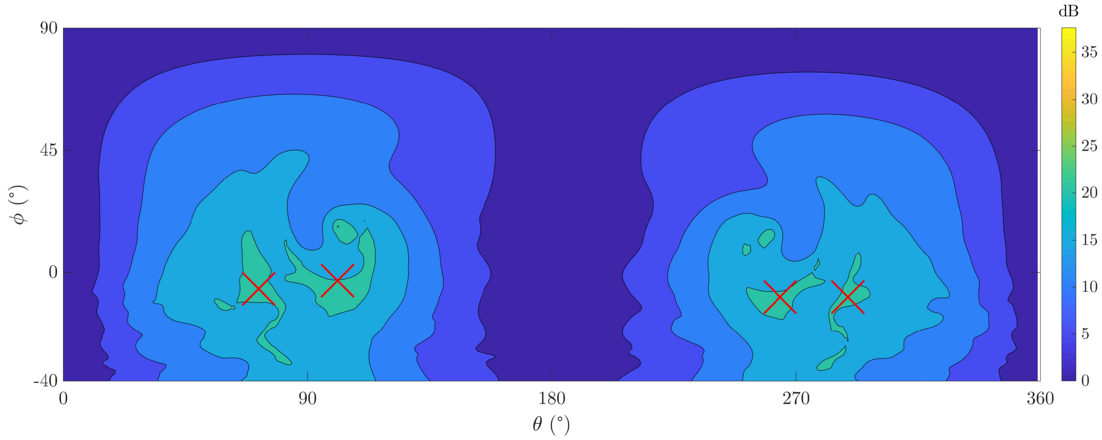
Figure D.2: NCS plotted as contour plots (increments of 5 dB) for domes within $\mathbf{SP^{KEMAR}}$ for radial distances 0.30 m to 0.40 m. The centroid of each region with the largest threshold value is indicated with a red cross. *Figure continued overleaf.*



(g) $r = 0.45\,\mathrm{m}$

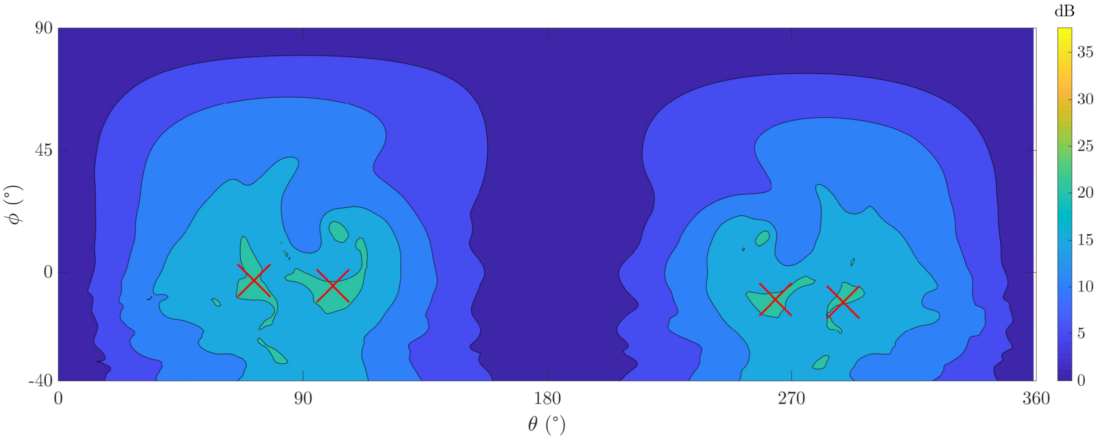


(h) $r = 0.50\,\mathrm{m}$

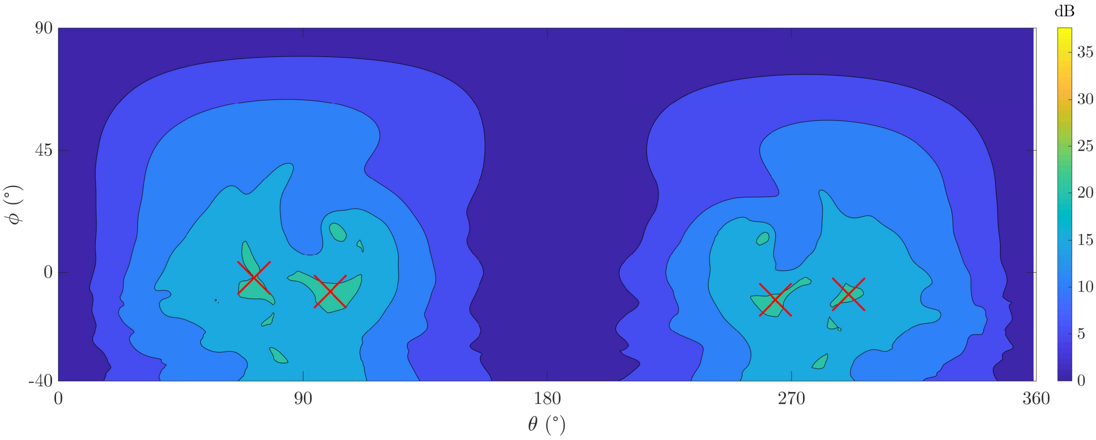


(i) $r = 0.60\,\mathrm{m}$

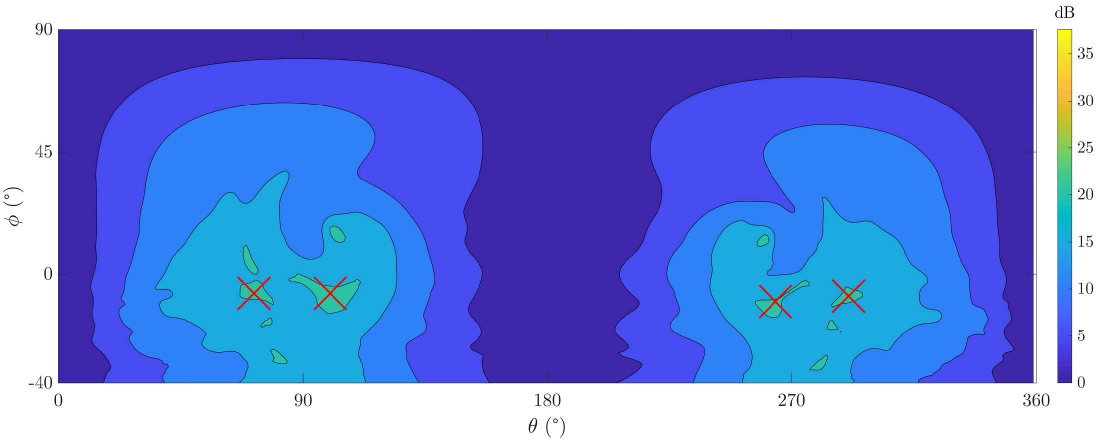
Figure D.2: NCS plotted as contour plots (increments of 5 dB) for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 0.45 m to 0.60 m. The centroid of each region with the largest threshold value is indicated with a red cross. Where two separate regions exist, two centroids are indicated. *Figure continued overleaf.*



(j) $r = 0.70\,\mathrm{m}$

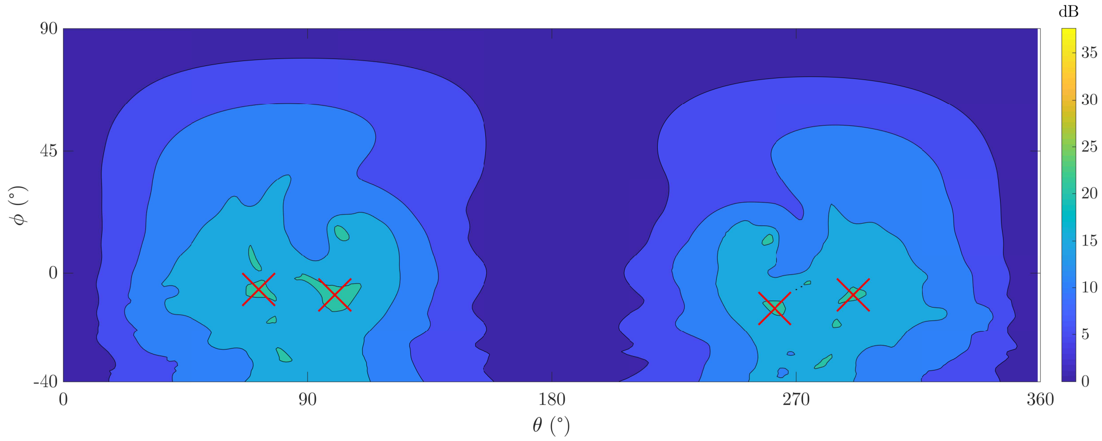


(k) $r = 0.80\,\mathrm{m}$

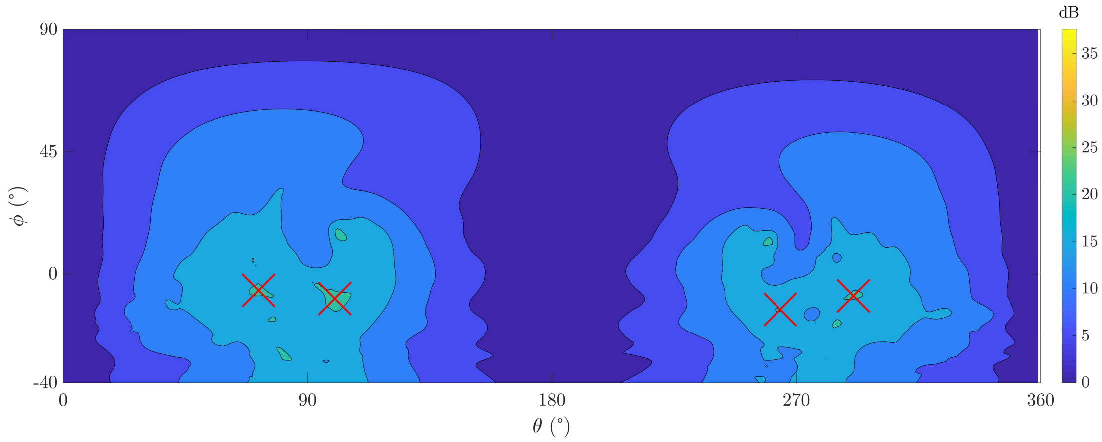


(l) $r = 0.90\,\mathrm{m}$

Figure D.2: NCS plotted as contour plots (increments of 5 dB) for domes within $\mathbf{SP^{KEMAR}}$ for radial distances 0.70 m to 0.90 m. The centroid of each region with the largest threshold value is indicated with a red cross. Where two separate regions exist, two centroids are indicated. *Figure continued overleaf.*

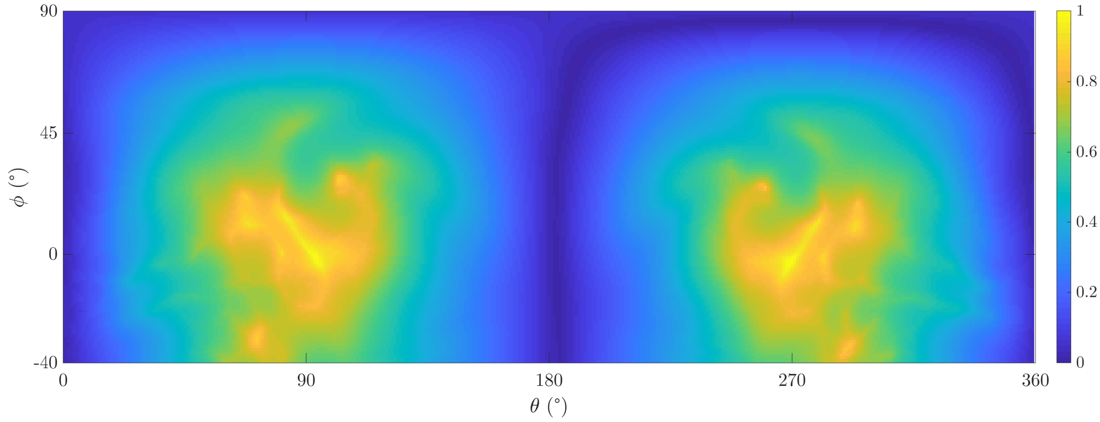


(m) $r = 1.00\,\mathrm{m}$

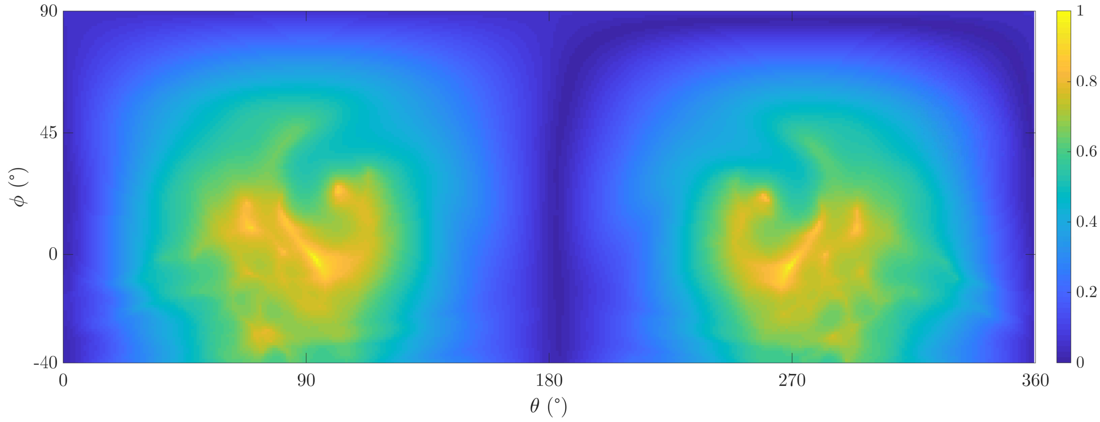


(n) $r = 1.50\,\mathrm{m}$

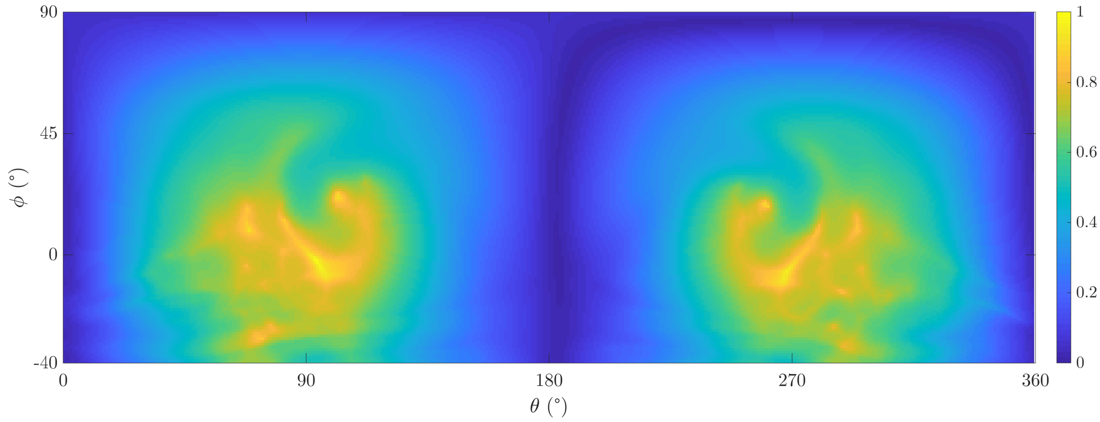
Figure D.2: NCS plotted as contour plots (increments of 5 dB) for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 1.00 m and 1.50 m. The centroid of each region with the largest threshold value is indicated with a red cross. Where two separate regions exist, two centroids are indicated.



(a) $r = 0.15\,\mathrm{m}$

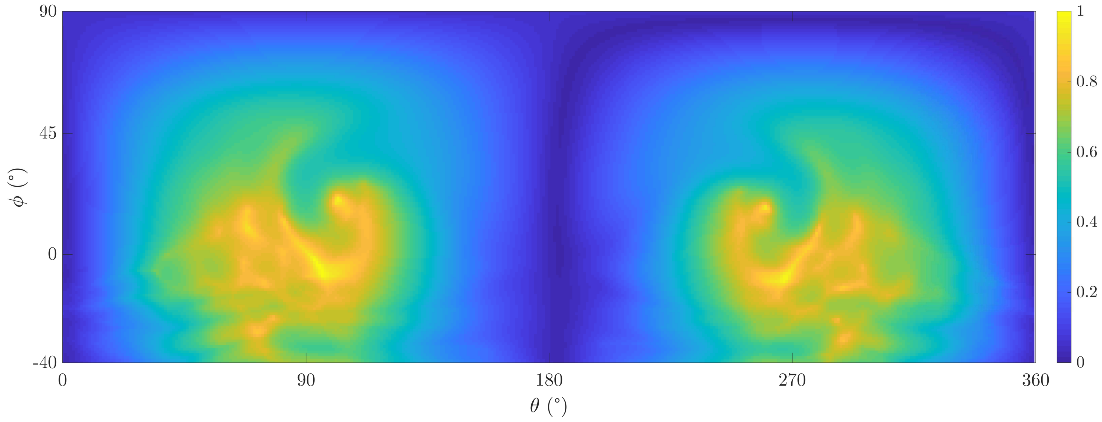


(b) $r = 0.20\,\mathrm{m}$

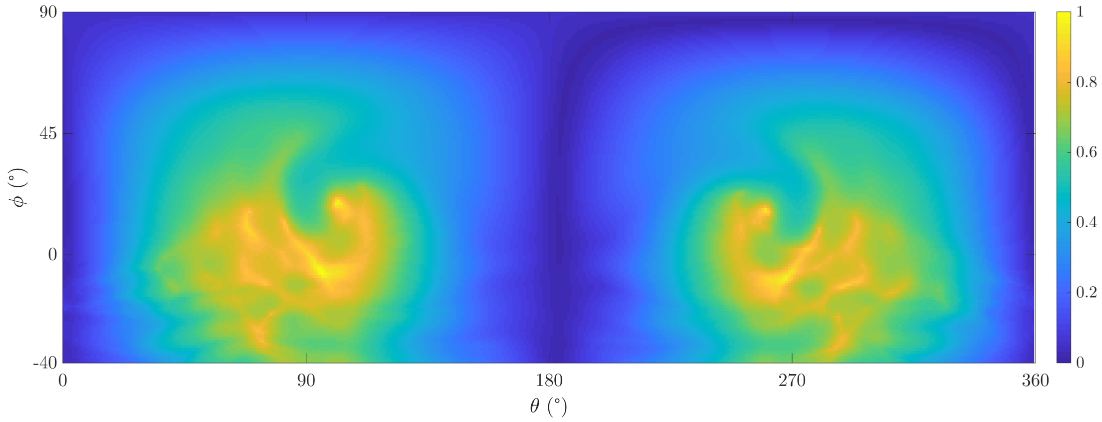


(c) $r = 0.25\,\mathrm{m}$

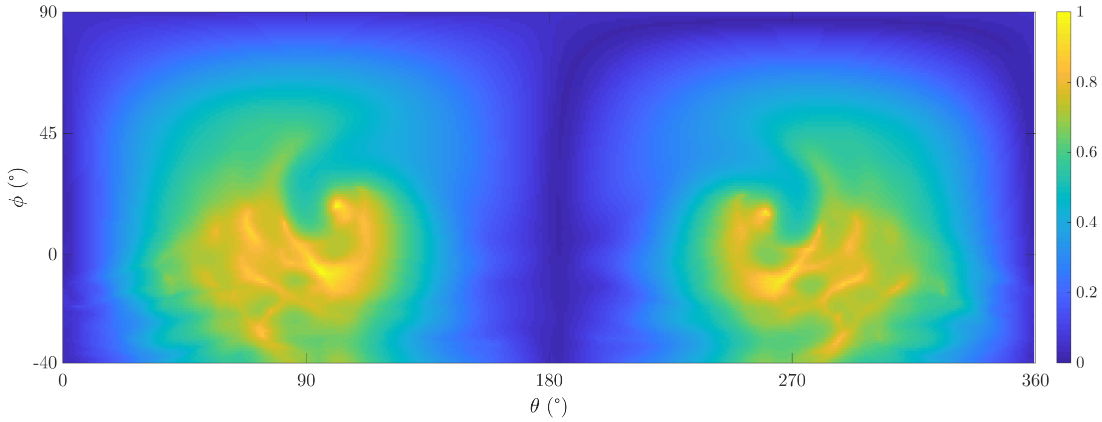
Figure D.3: NCS values normalised with respect to the corresponding maximum value, plotted as intensity surfaces for domes within $\mathbf{SP^{KEMAR}}$ for radial distances 0.15 m to 0.25 m. *Figure continued overleaf.*



(d) $r = 0.30\,\mathrm{m}$

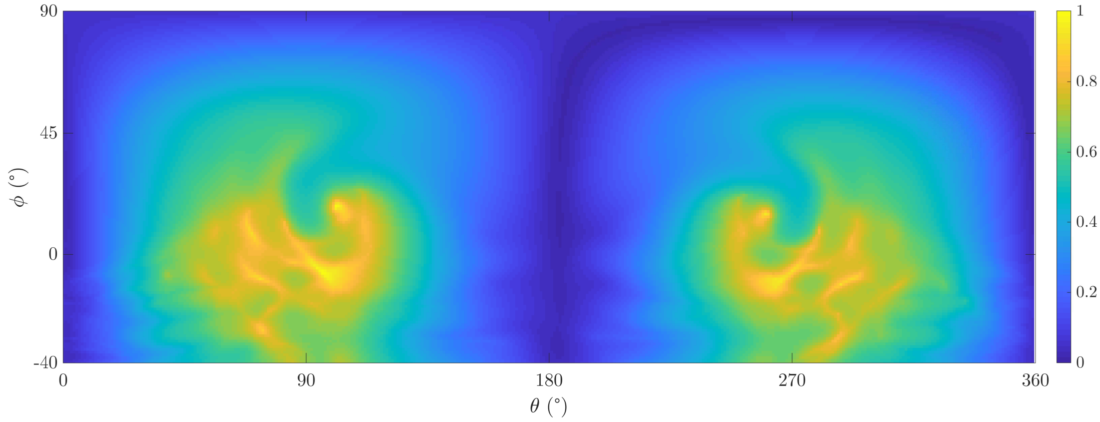


(e) $r = 0.35\,\mathrm{m}$

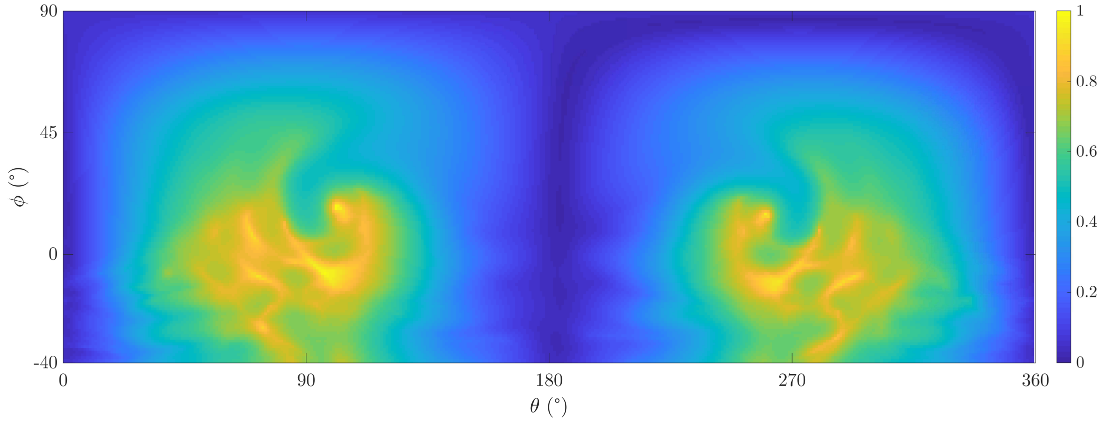


(f) $r = 0.40\,\mathrm{m}$

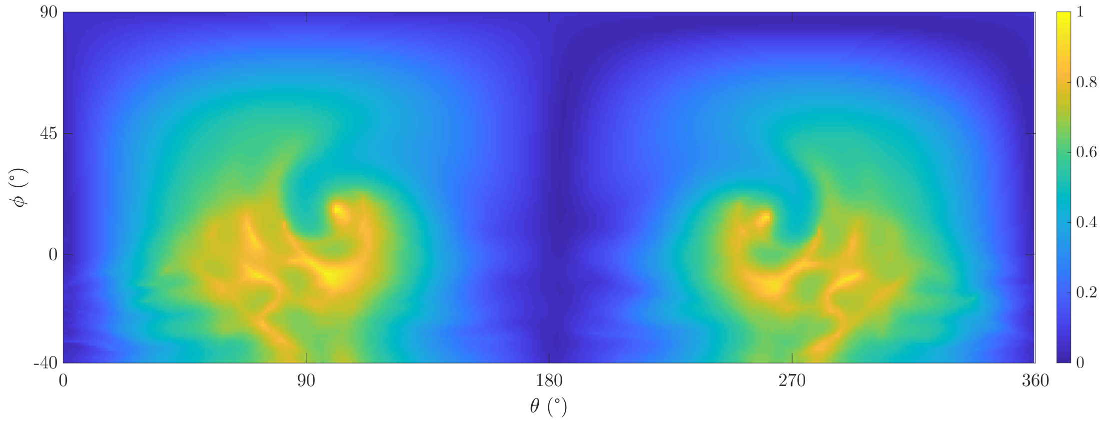
Figure D.3: NCS values normalised with respect to the corresponding maximum value, plotted as intensity surfaces for domes within $\mathbf{SP^{KEMAR}}$ for radial distances $0.30\,\mathrm{m}$ to $0.40\,\mathrm{m}$. *Figure continued overleaf.*



(g) $r = 0.45\,\mathrm{m}$

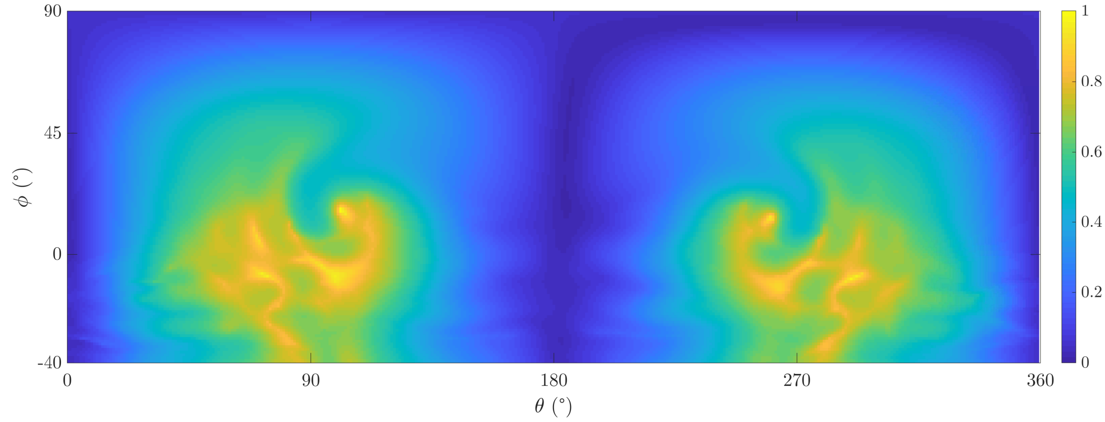


(h) $r = 0.50\,\mathrm{m}$

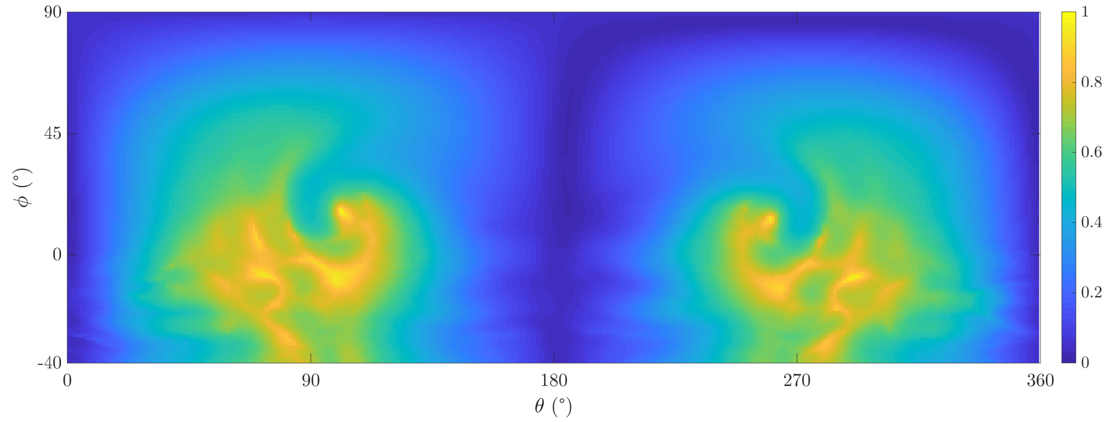


(i) $r = 0.60\,\mathrm{m}$

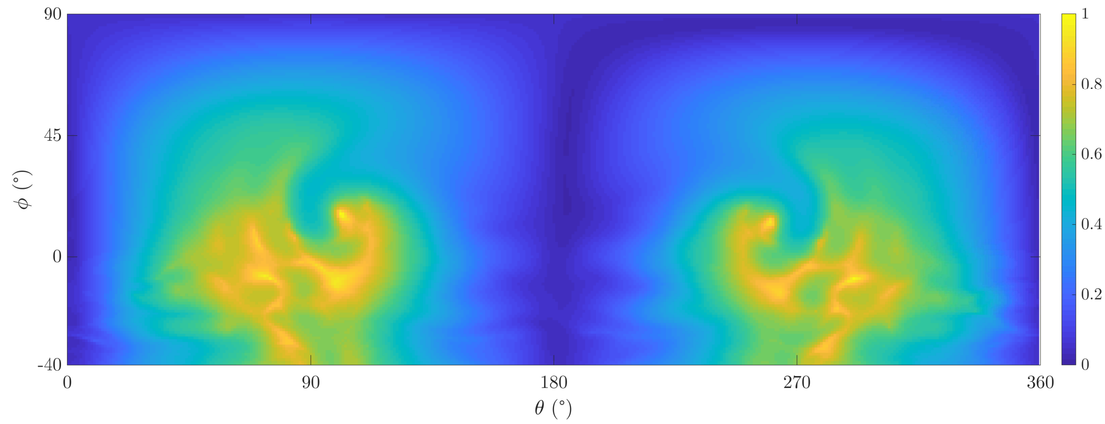
Figure D.3: NCS values normalised with respect to the corresponding maximum value, plotted as intensity surfaces for domes within $\mathbf{SP^{KEMAR}}$ for radial distances 0.45 m to 0.60 m. *Figure continued overleaf.*



(j) $r = 0.70\,\mathrm{m}$

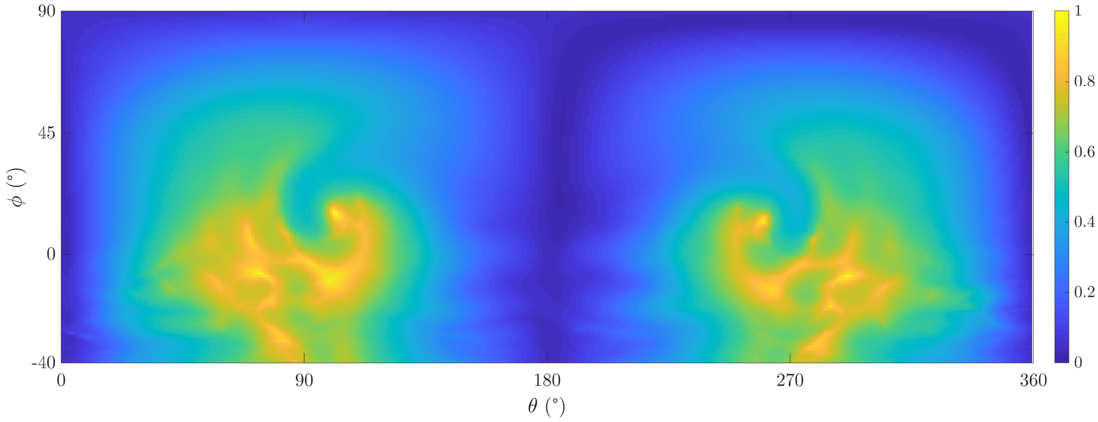


(k) $r = 0.80\,\mathrm{m}$

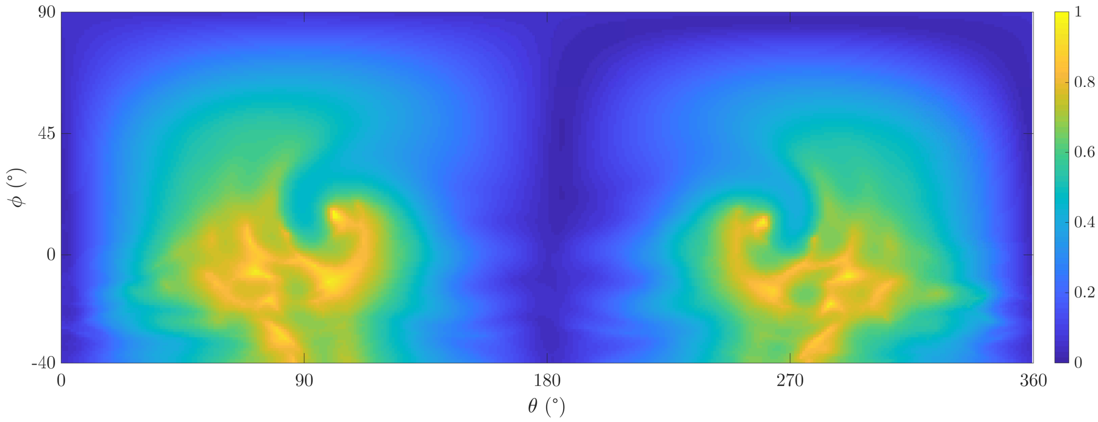


(l) $r = 0.90\,\mathrm{m}$

Figure D.3: NCS values normalised with respect to the corresponding maximum value, plotted as intensity surfaces for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 0.70 m to 0.90 m. *Figure continued overleaf.*

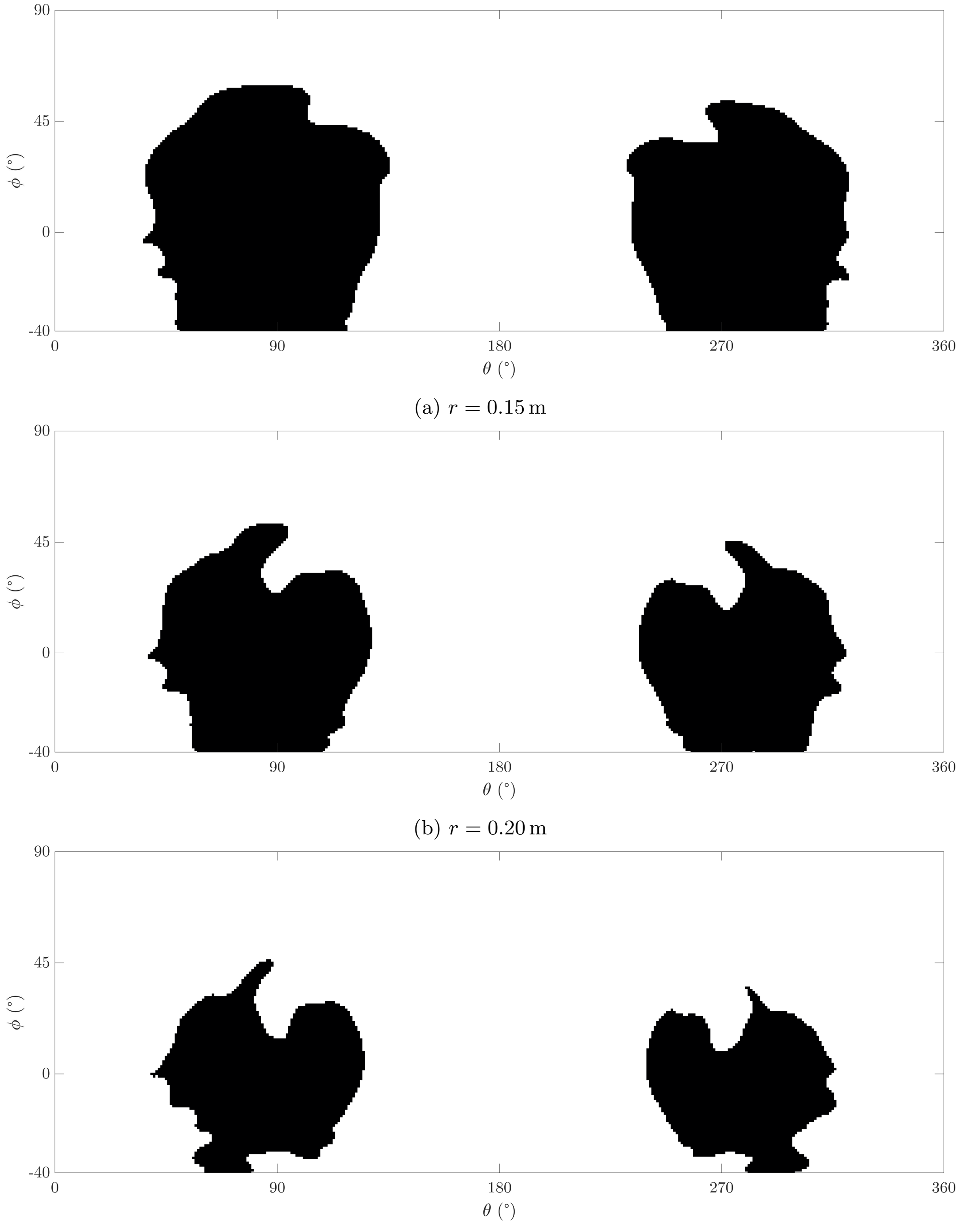


(m) $r = 1.00\,\mathrm{m}$



(n) $r = 1.50\,\mathrm{m}$

Figure D.3: NCS values normalised with respect to the corresponding maximum value, plotted as intensity surfaces for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 1.00 m and 1.50 m.

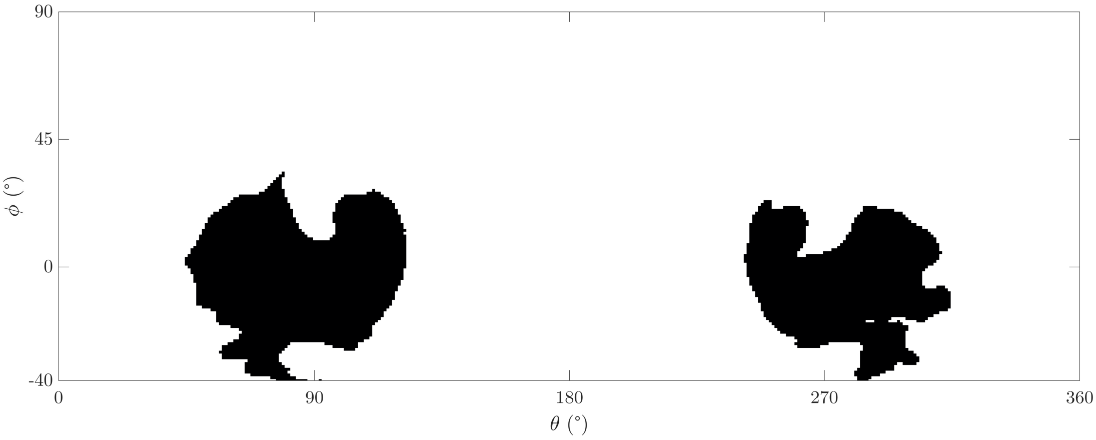


(a) $r = 0.15\,\mathrm{m}$

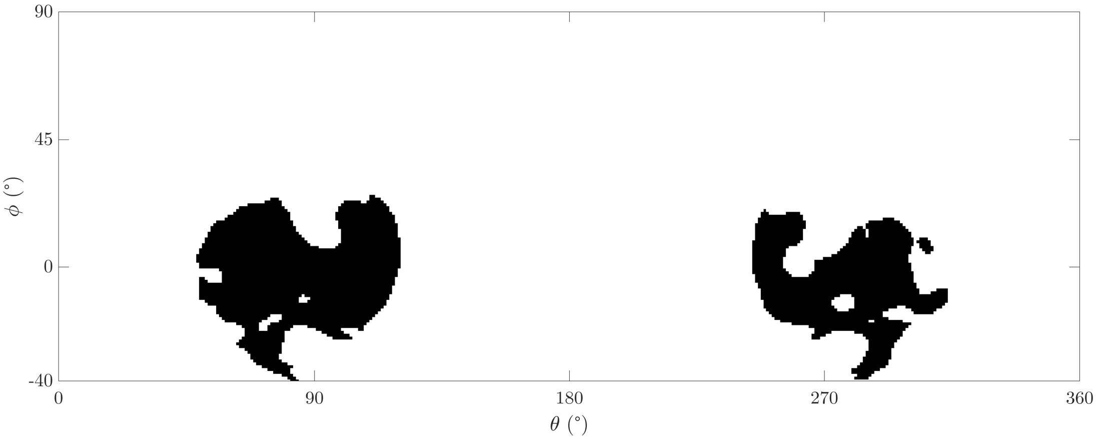
(b) $r = 0.20\,\mathrm{m}$

(c) $r = 0.25\,\mathrm{m}$

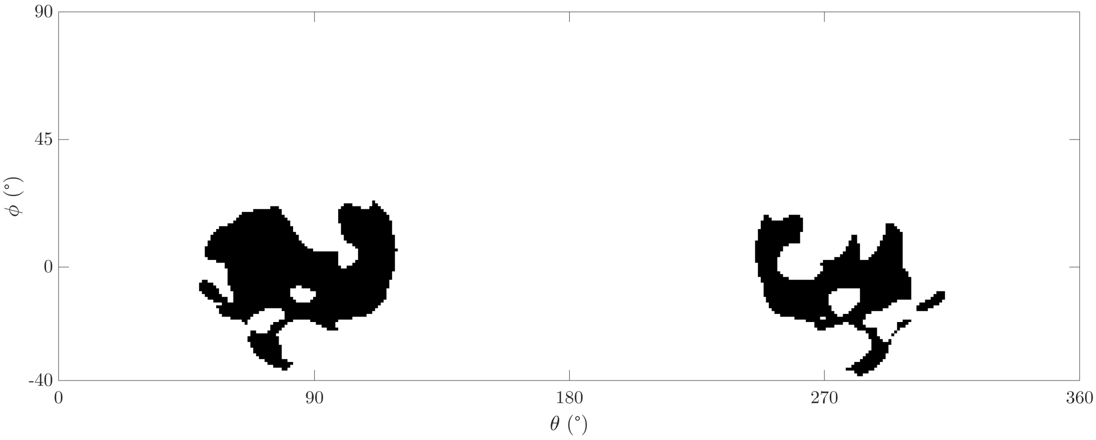
Figure D.4: Regions of $\mathbf{SP}_{\mathbf{20}}^{\mathbf{KEMAR}}$ positions (indicated in black) for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 0.15 m to 0.25 m. *Figure continued overleaf.*



(d) $r = 0.30\,\mathrm{m}$

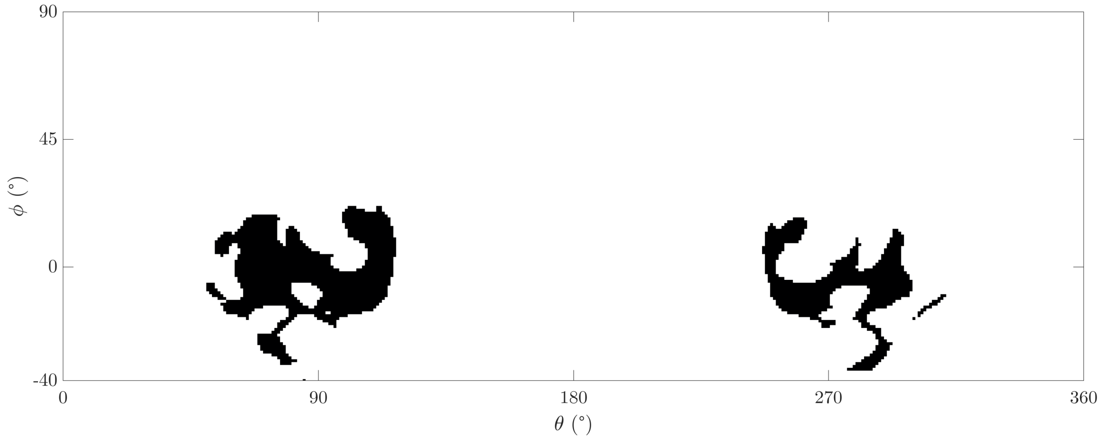


(e) $r = 0.35\,\mathrm{m}$

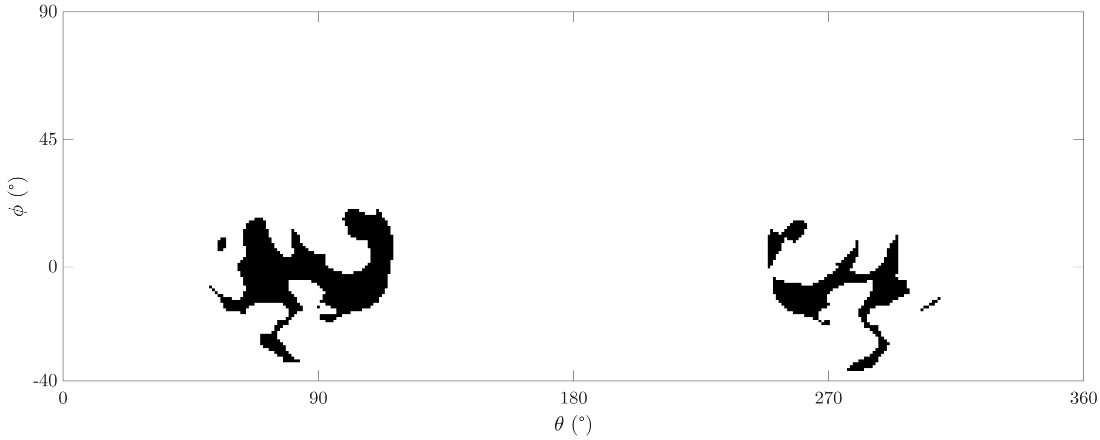


(f) $r = 0.40\,\mathrm{m}$

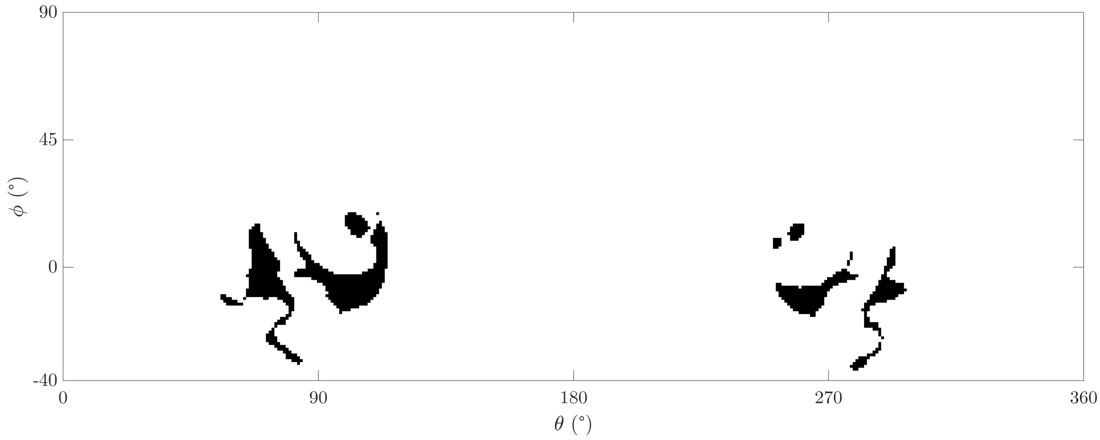
Figure D.4: Regions of $\mathbf{SP}_{\mathbf{20}}^{\mathbf{KEMAR}}$ positions (indicated in black) for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 0.30 m to 0.40 m. *Figure continued overleaf.*



(g) $r = 0.45\,\mathrm{m}$

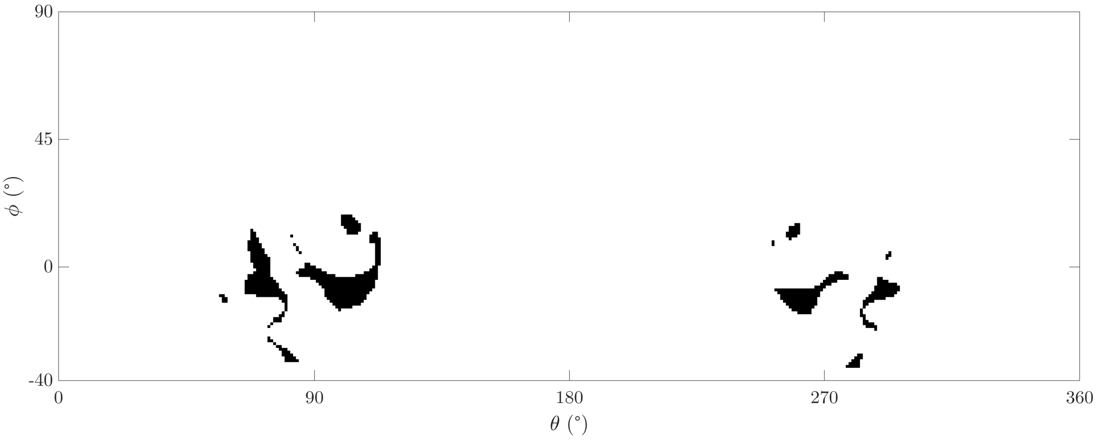


(h) $r = 0.50\,\mathrm{m}$

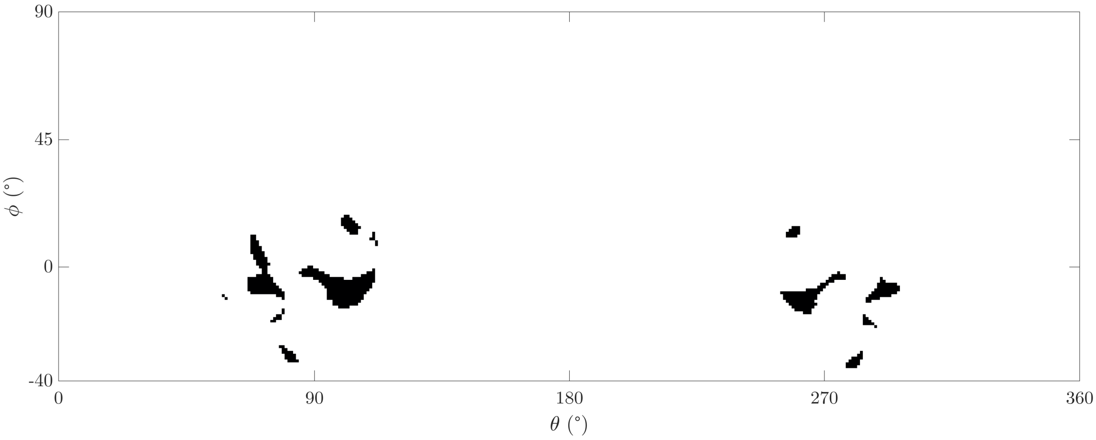


(i) $r = 0.60\,\mathrm{m}$

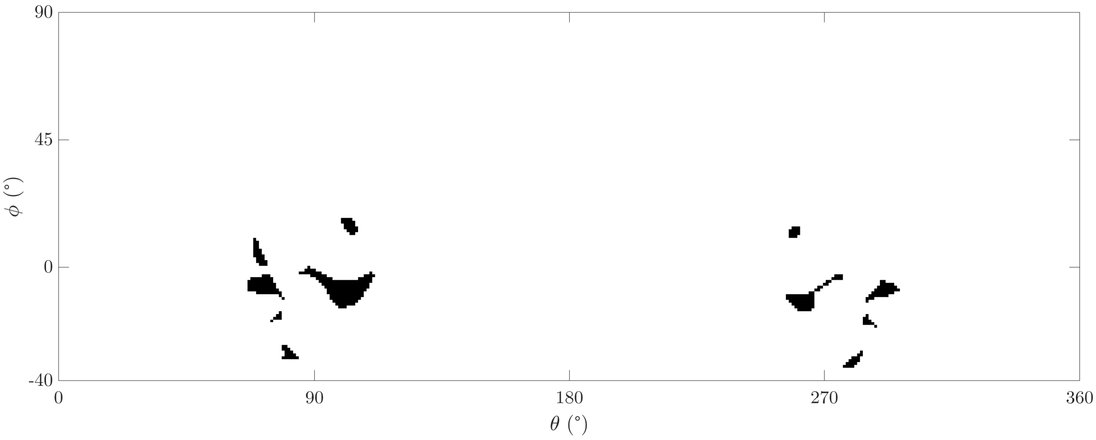
Figure D.4: Regions of $\mathbf{SP}_{\mathbf{20}}^{\mathbf{KEMAR}}$ positions (indicated in black) for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 0.45 m to 0.60 m. *Figure continued overleaf.*



(j) $r = 0.70\,\mathrm{m}$

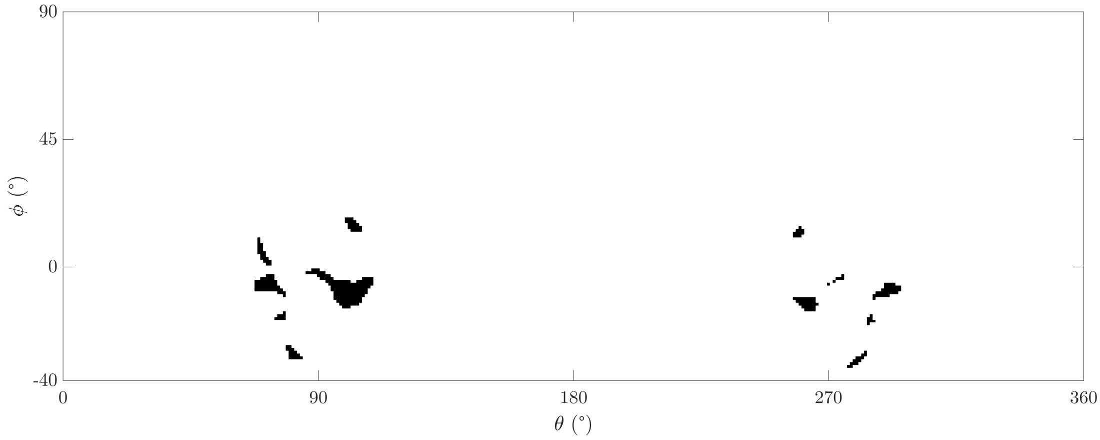


(k) $r = 0.80\,\mathrm{m}$

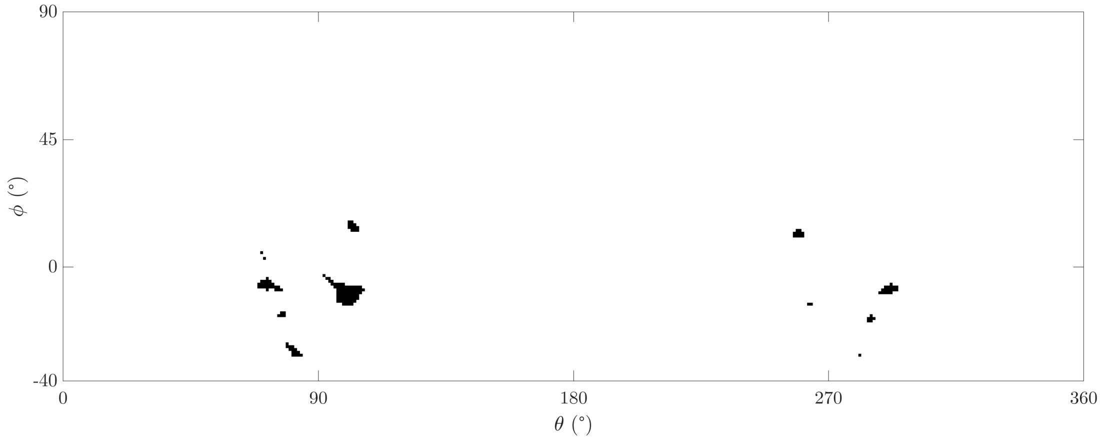


(l) $r = 0.90\,\mathrm{m}$

Figure D.4: Regions of $\mathbf{SP}_{\mathbf{20}}^{\mathbf{KEMAR}}$ positions (indicated in black) for domes within $\mathbf{SP}^{\mathbf{KEMAR}}$ for radial distances 0.70 m to 0.90 m. *Figure continued overleaf.*



(m) $r = 1.00\,\mathrm{m}$



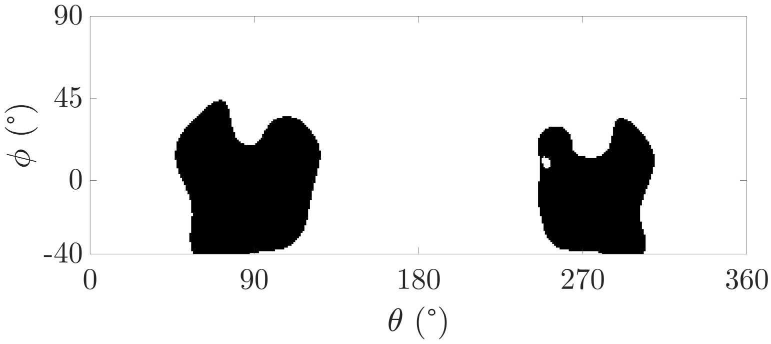
(n) $r = 1.50\,\mathrm{m}$

Figure D.4: Regions of $\mathbf{SP_{20}^{KEMAR}}$ positions (indicated in black) for domes within $\mathbf{SP^{KEMAR}}$ for radial distances 1.00 m and 1.50 m.

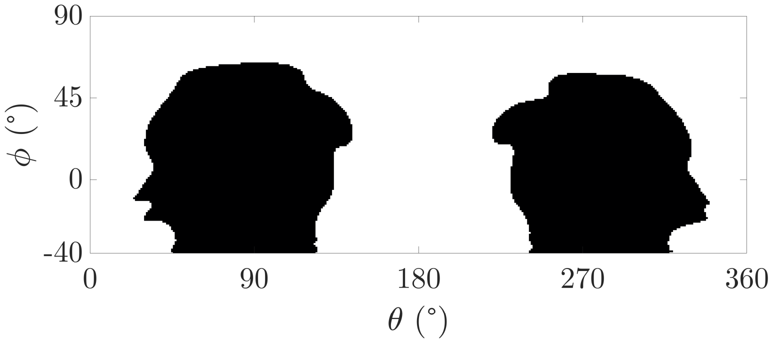
Appendix E

Additional Figures: Source Positions with Sufficient NCS calculated in Frequency Bands

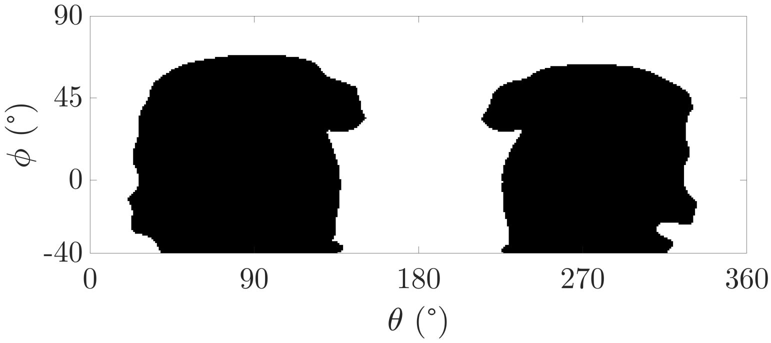
This appendix contains the plots of regions of source positions with sufficient NCS for binaural reproduction for each frequency band f at each of the 14 radial distances used in the analysis discussed in Chapter 5.



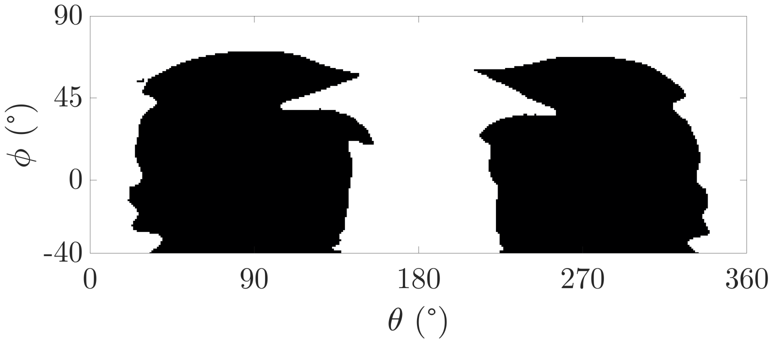
(a) 200 Hz to 4 kHz



(b) 4 to 8 kHz

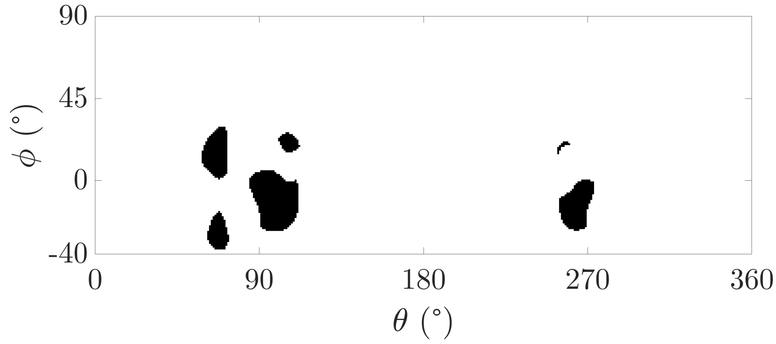


(c) 8 to 12 kHz

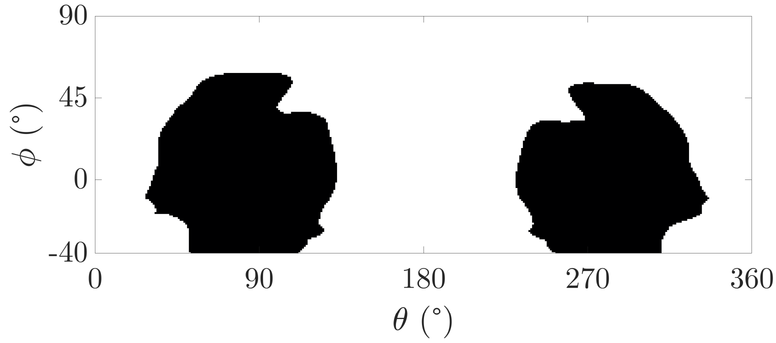


(d) 12 to 16 kHz

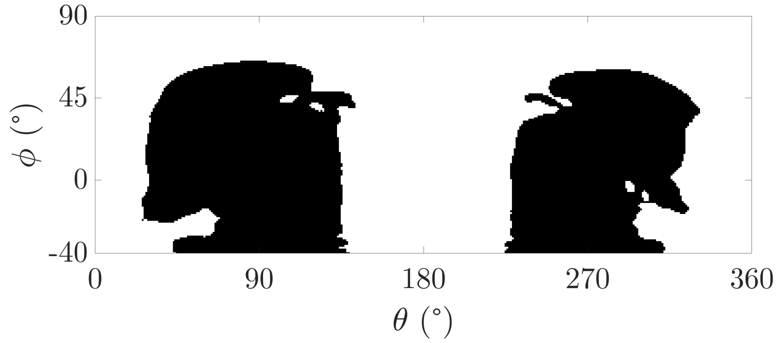
Figure E.1: Regions of $\mathbf{SP}^{\mathbf{KEMAR}}_{\mathbf{20_f}}$ positions (indicated in black) within $\mathbf{SP}^{\mathbf{KEMAR}}$ at $r = 0.15\,\mathrm{m}$ for the four frequency bands used in analysis.



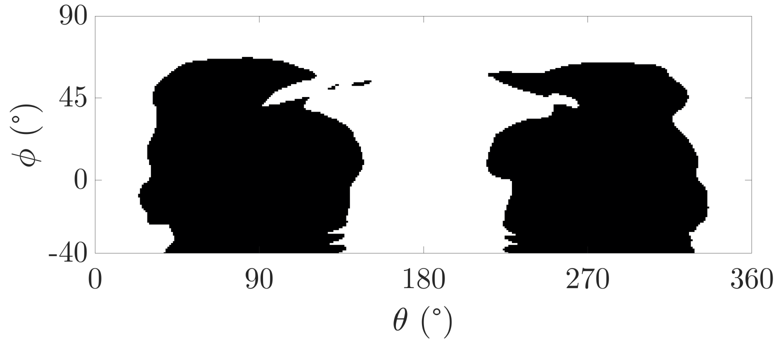
(a) 200 Hz to 4 kHz



(b) 4 to 8 kHz

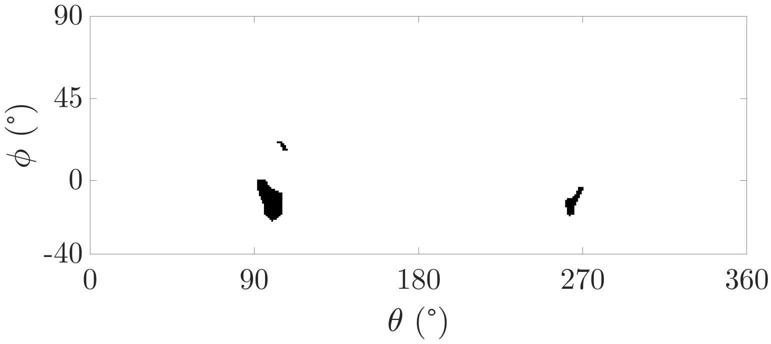


(c) 8 to 12 kHz

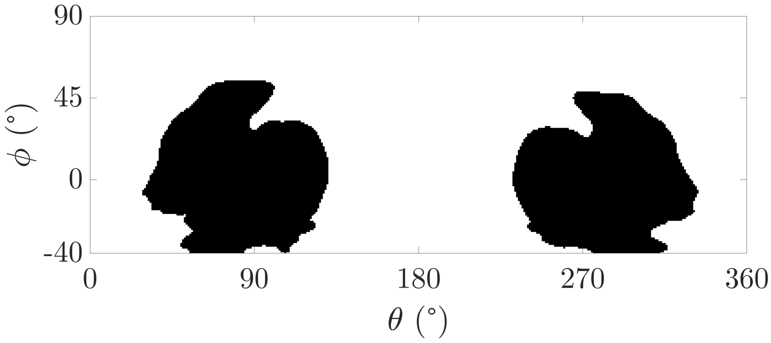


(d) 12 to 16 kHz

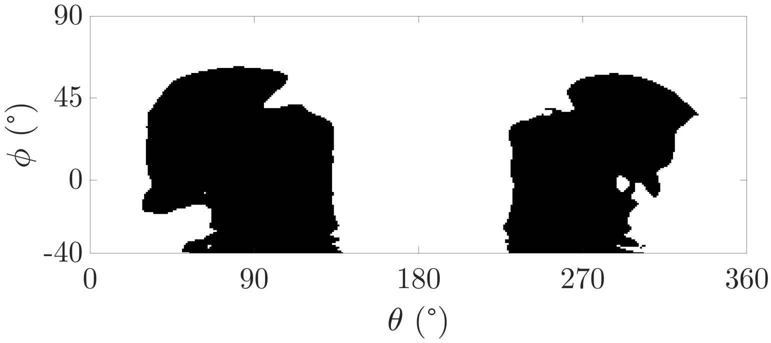
Figure E.2: Regions of $\mathbf{SP^{KEMAR}_{20_f}}$ positions (indicated in black) within $\mathbf{SP^{KEMAR}}$ at $r = 0.20\,\mathrm{m}$ for the four frequency bands used in analysis.



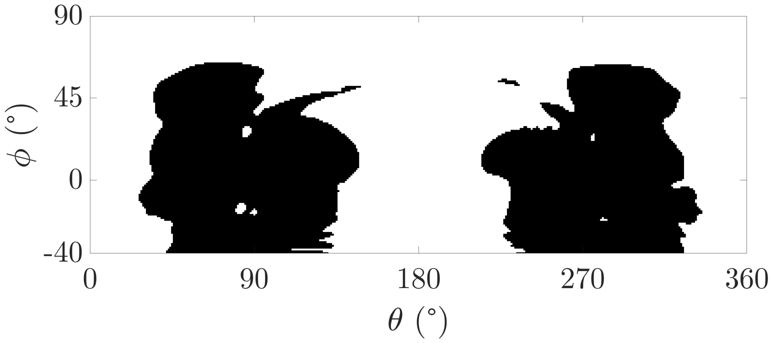
(a) 200 Hz to 4 kHz



(b) 4 to 8 kHz

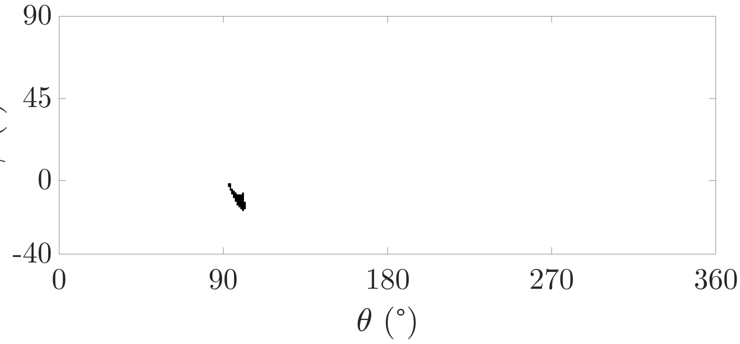


(c) 8 to 12 kHz

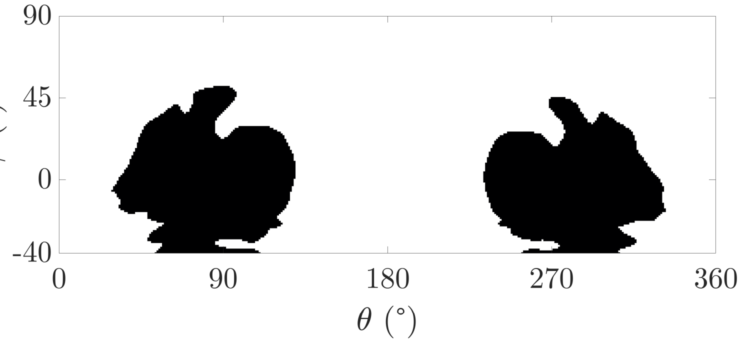


(d) 12 to 16 kHz

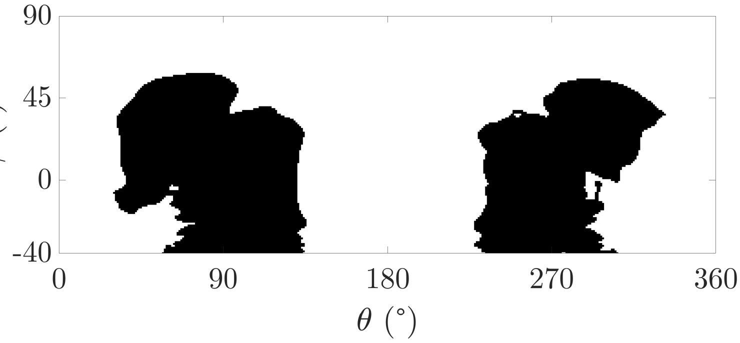
Figure E.3: Regions of $\mathbf{SP}^{\mathbf{KEMAR}}_{\mathbf{20_f}}$ positions (indicated in black) within $\mathbf{SP}^{\mathbf{KEMAR}}$ at $r = 0.25\,\mathrm{m}$ for the four frequency bands used in analysis.



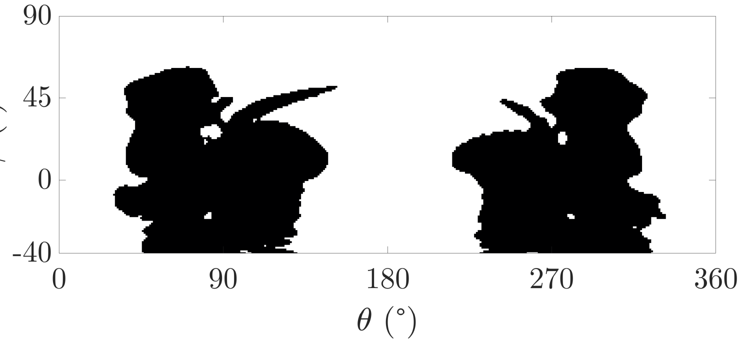
(a) 200 Hz to 4 kHz



(b) 4 to 8 kHz

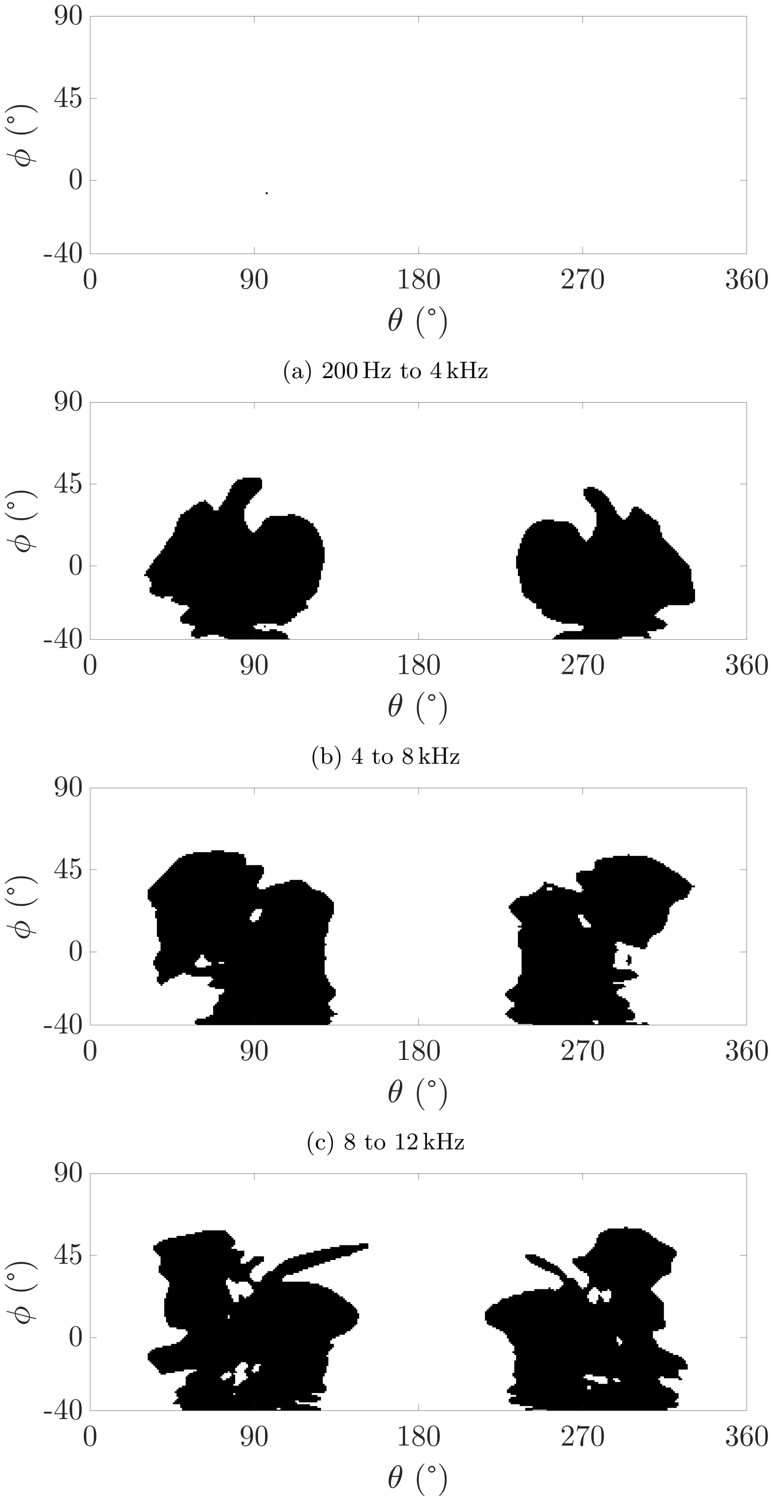


(c) 8 to 12 kHz



(d) 12 to 16 kHz

Figure E.4: Regions of $\mathbf{SP}^{\mathbf{KEMAR}}_{\mathbf{20_f}}$ positions (indicated in black) within $\mathbf{SP}^{\mathbf{KEMAR}}$ at $r = 0.30\,\mathrm{m}$ for the four frequency bands used in analysis.



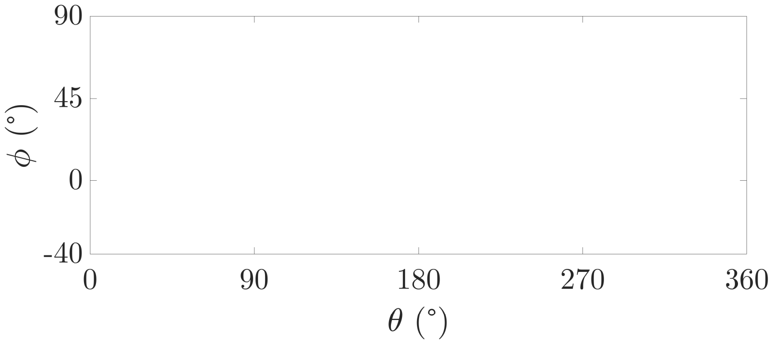
(a) 200 Hz to 4 kHz

(b) 4 to 8 kHz

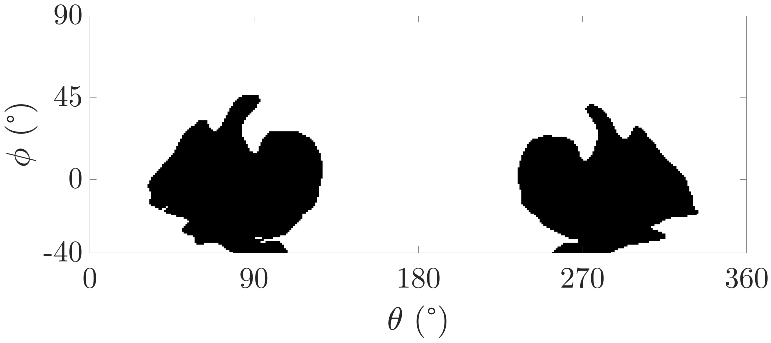
(c) 8 to 12 kHz

(d) 12 to 16 kHz

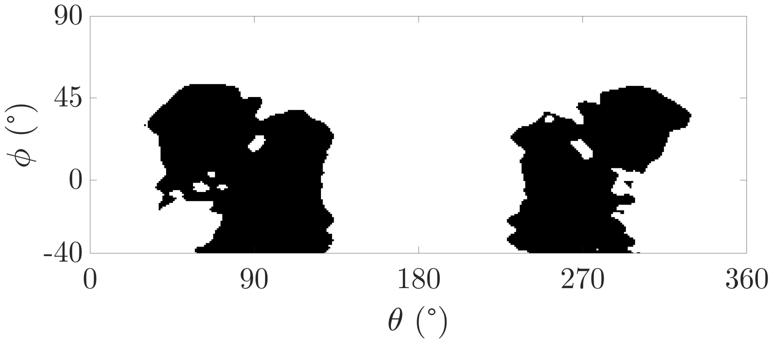
Figure E.5: Regions of $\mathbf{SP^{KEMAR}_{20_f}}$ positions (indicated in black) within $\mathbf{SP^{KEMAR}}$ at $r = 0.35\,\mathrm{m}$ for the four frequency bands used in analysis.



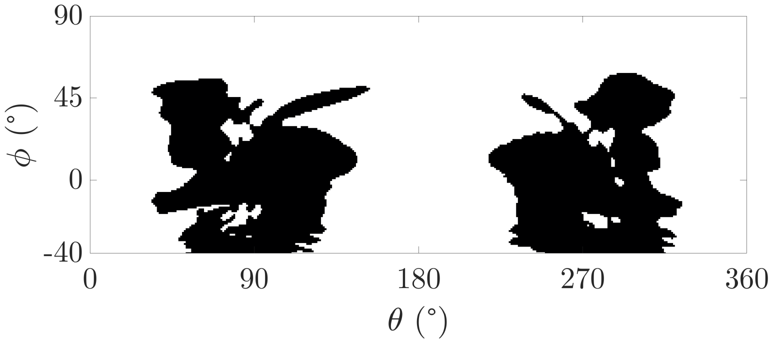
(a) 200 Hz to 4 kHz



(b) 4 to 8 kHz



(c) 8 to 12 kHz



(d) 12 to 16 kHz

Figure E.6: Regions of $\mathbf{SP}^{\mathbf{KEMAR}}_{\mathbf{20_f}}$ positions (indicated in black) within $\mathbf{SP}^{\mathbf{KEMAR}}$ at $r = 0.40\,\mathrm{m}$ for the four frequency bands used in analysis.

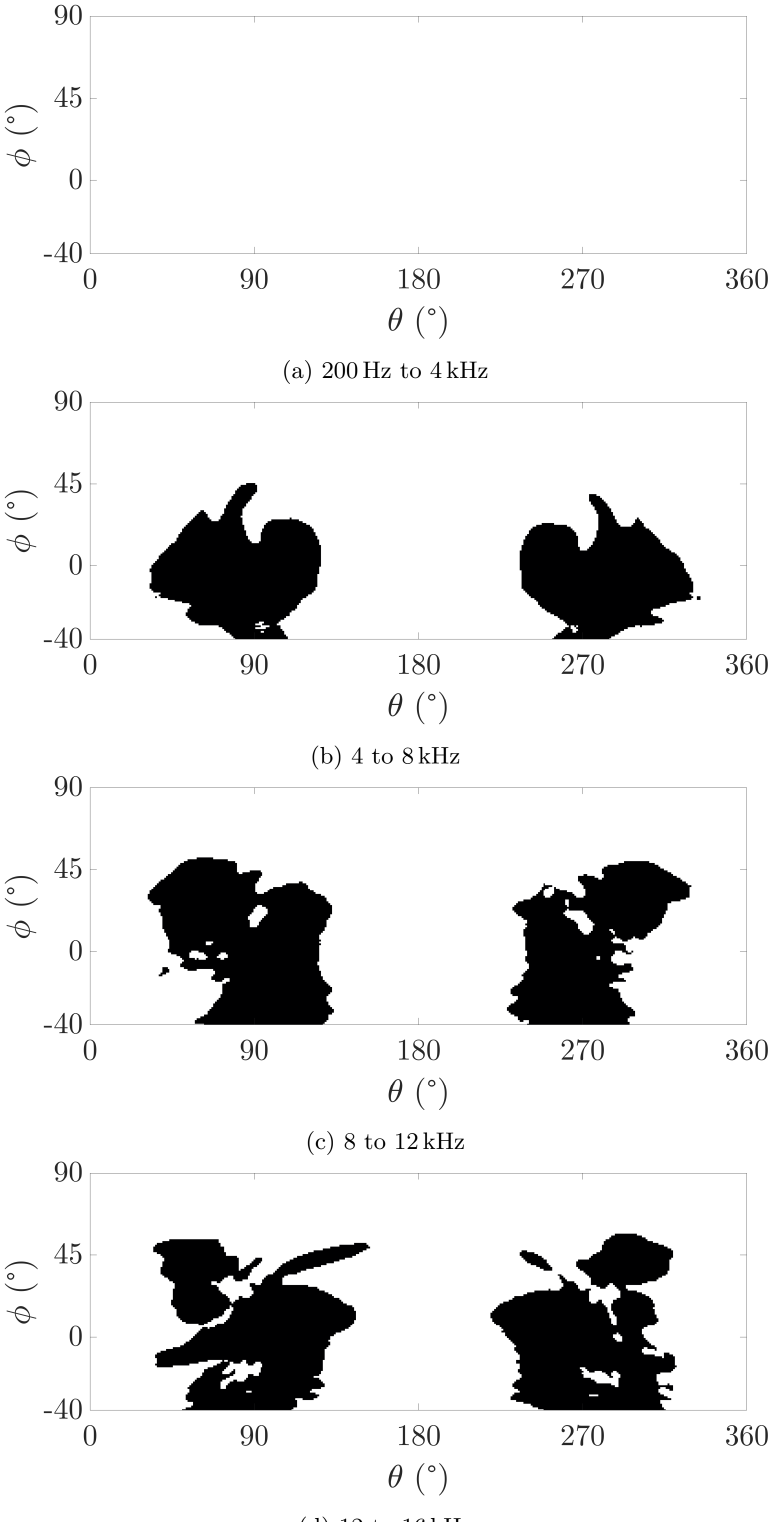
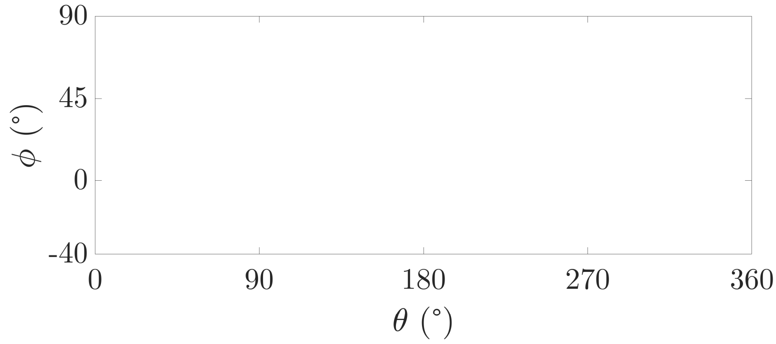
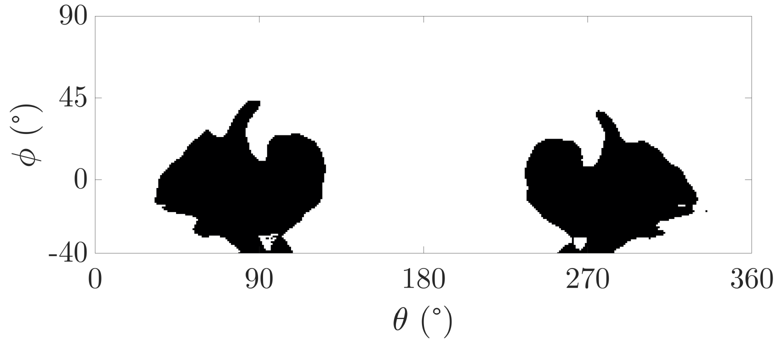


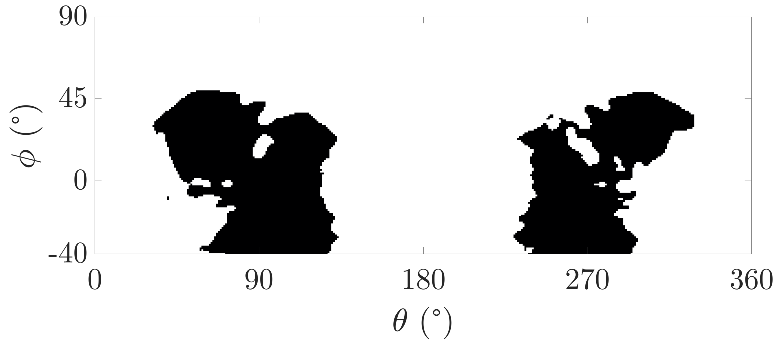
Figure E.7: Regions of $\mathbf{SP^{KEMAR}_{20_f}}$ positions (indicated in black) within $\mathbf{SP^{KEMAR}}$ at $r = 0.45\,\mathrm{m}$ for the four frequency bands used in analysis.



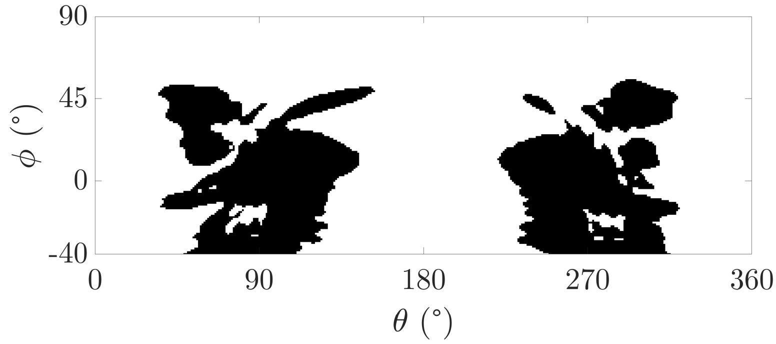
(a) 200 Hz to 4 kHz



(b) 4 to 8 kHz

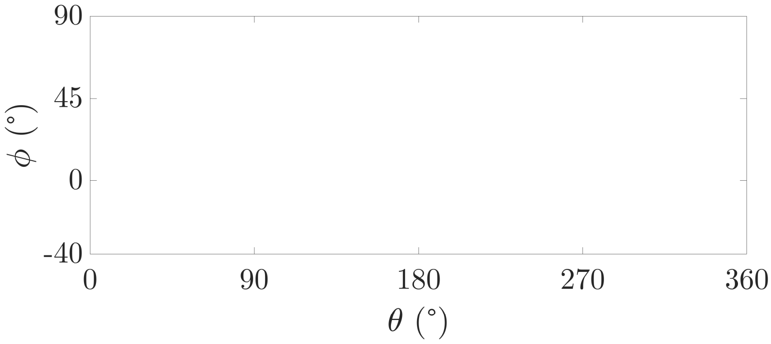


(c) 8 to 12 kHz

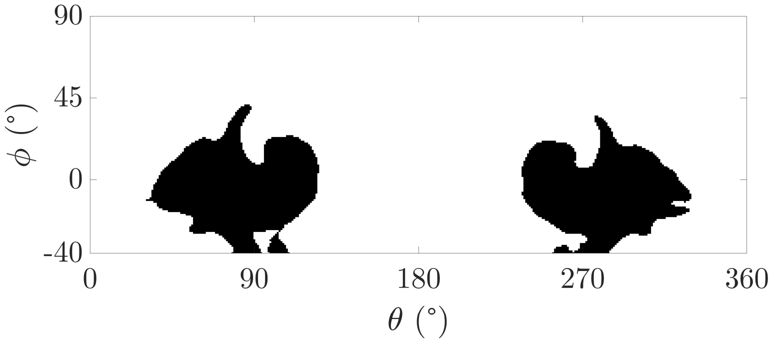


(d) 12 to 16 kHz

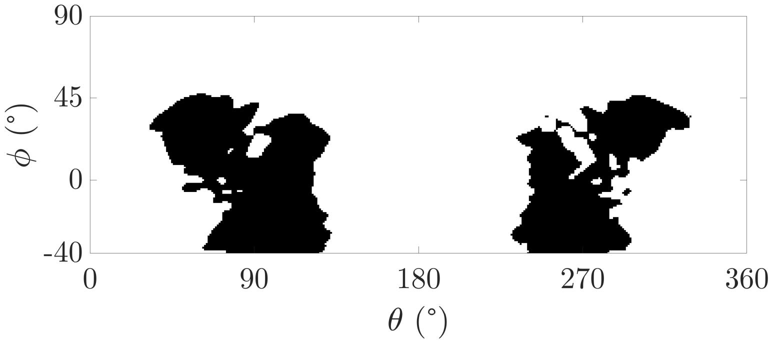
Figure E.8: Regions of $\mathbf{SP}^{\mathbf{KEMAR}}_{\mathbf{20_f}}$ positions (indicated in black) within $\mathbf{SP}^{\mathbf{KEMAR}}$ at $r = 0.50\,\mathrm{m}$ for the four frequency bands used in analysis.



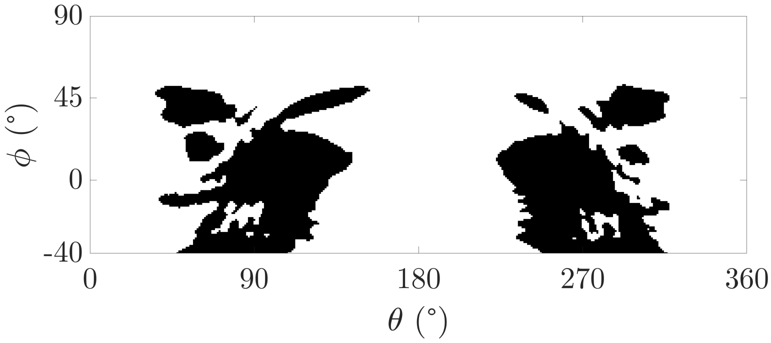
(a) 200 Hz to 4 kHz



(b) 4 to 8 kHz

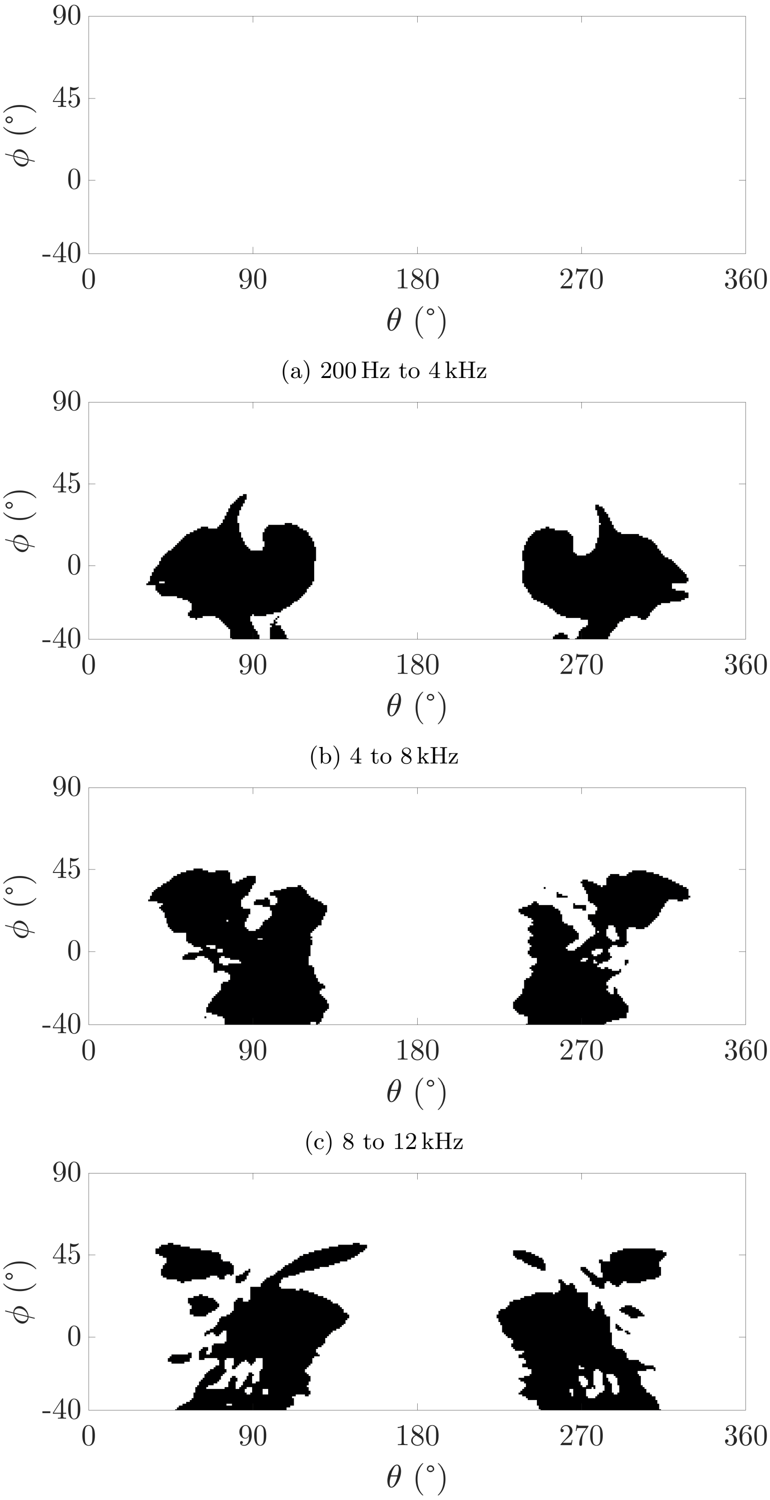


(c) 8 to 12 kHz



(d) 12 to 16 kHz

Figure E.9: Regions of $\mathbf{SP_{20_f}^{KEMAR}}$ positions (indicated in black) within $\mathbf{SP^{KEMAR}}$ at $r = 0.60\,\mathrm{m}$ for the four frequency bands used in analysis.



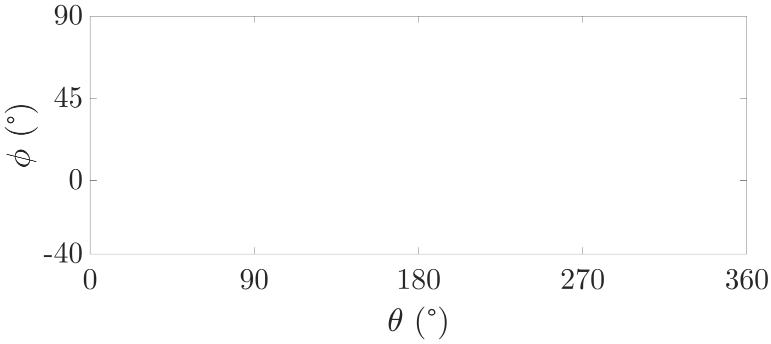
(a) 200 Hz to 4 kHz

(b) 4 to 8 kHz

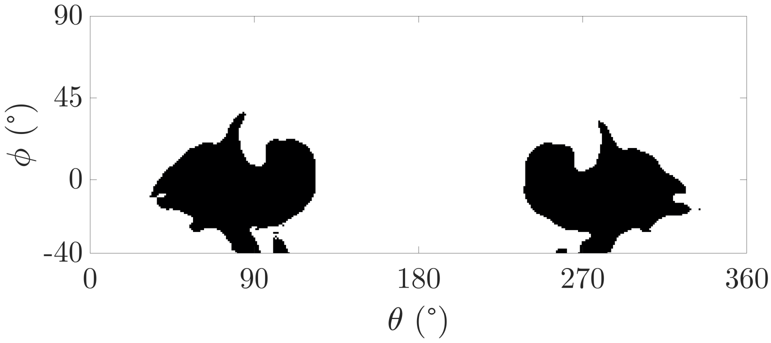
(c) 8 to 12 kHz

(d) 12 to 16 kHz

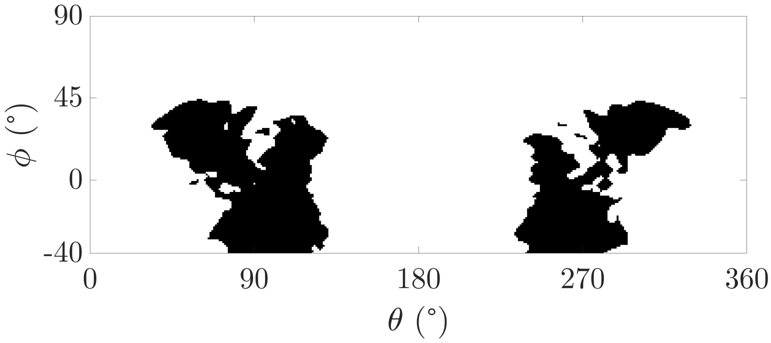
Figure E.10: Regions of $\mathbf{SP^{KEMAR}_{20_f}}$ positions (indicated in black) within $\mathbf{SP^{KEMAR}}$ at $r =$ 0.70 m for the four frequency bands used in analysis.



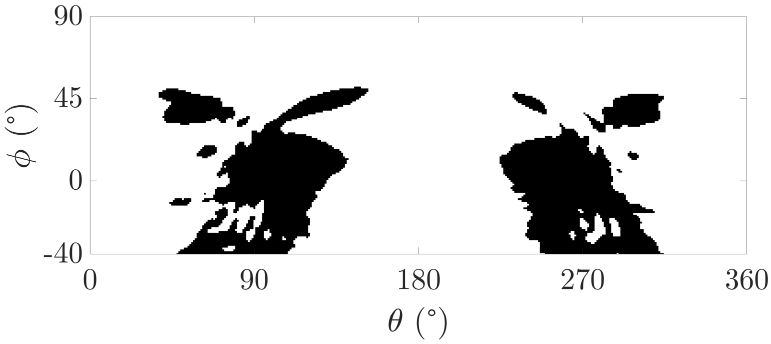
(a) 200 Hz to 4 kHz



(b) 4 to 8 kHz

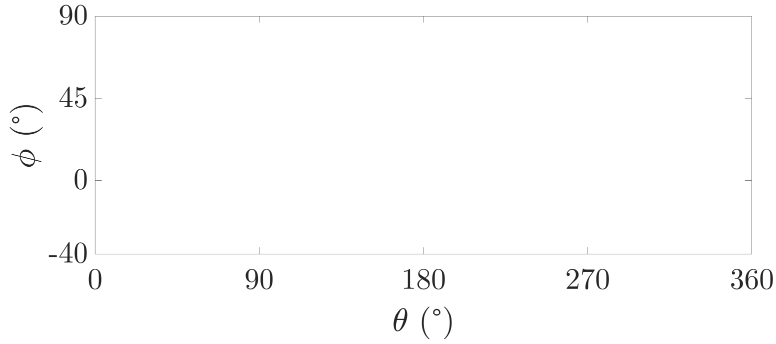


(c) 8 to 12 kHz

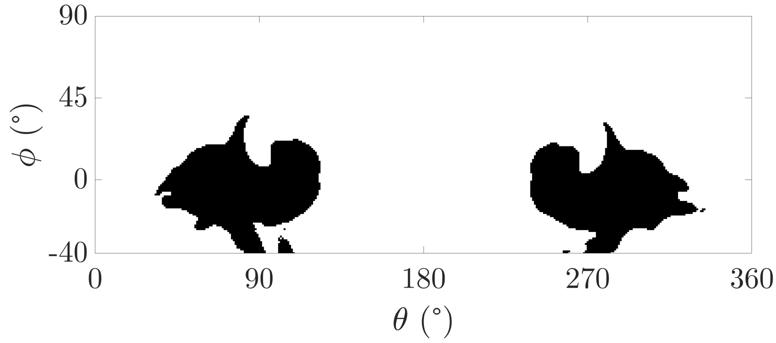


(d) 12 to 16 kHz

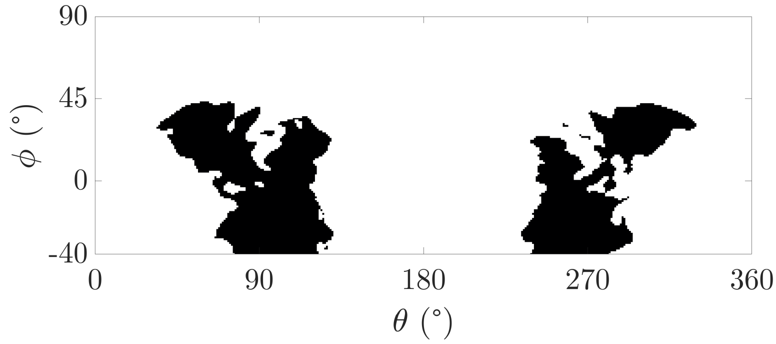
Figure E.11: Regions of $\mathbf{SP}^{\mathbf{KEMAR}}_{\mathbf{20_f}}$ positions (indicated in black) within $\mathbf{SP}^{\mathbf{KEMAR}}$ at $r = 0.80\,\mathrm{m}$ for the four frequency bands used in analysis.



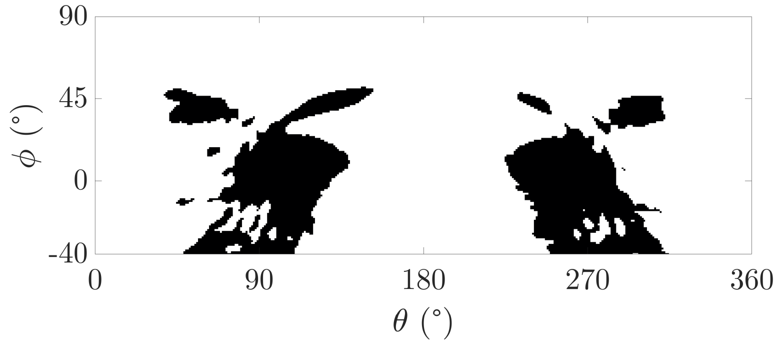
(a) 200 Hz to 4 kHz



(b) 4 to 8 kHz

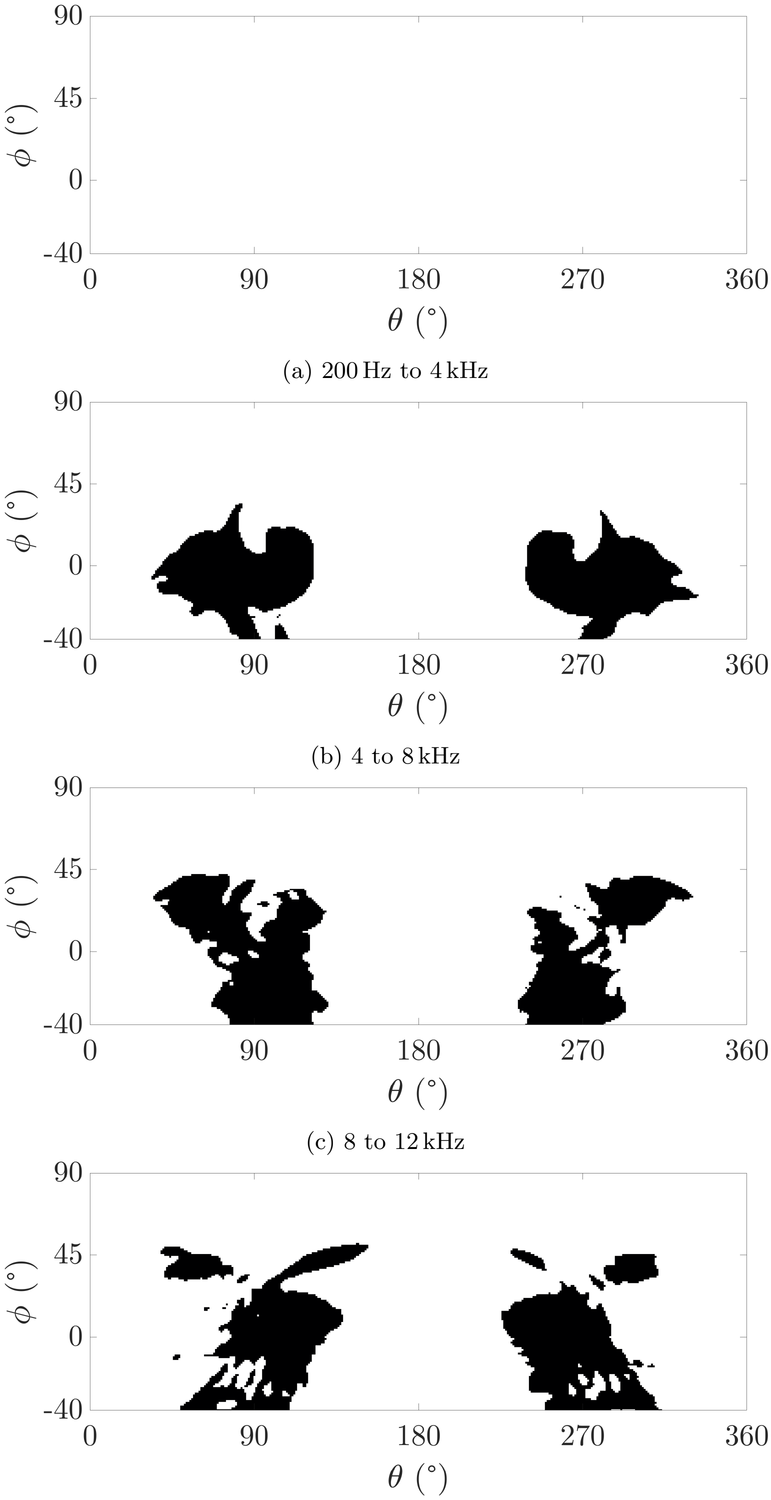


(c) 8 to 12 kHz



(d) 12 to 16 kHz

Figure E.12: Regions of $\mathbf{SP}^{\mathbf{KEMAR}}_{\mathbf{20_f}}$ positions (indicated in black) within $\mathbf{SP}^{\mathbf{KEMAR}}$ at $r = 0.90\,\mathrm{m}$ for the four frequency bands used in analysis.



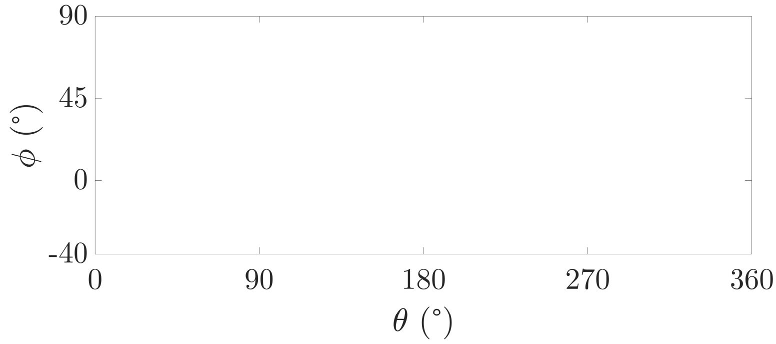
(a) 200 Hz to 4 kHz

(b) 4 to 8 kHz

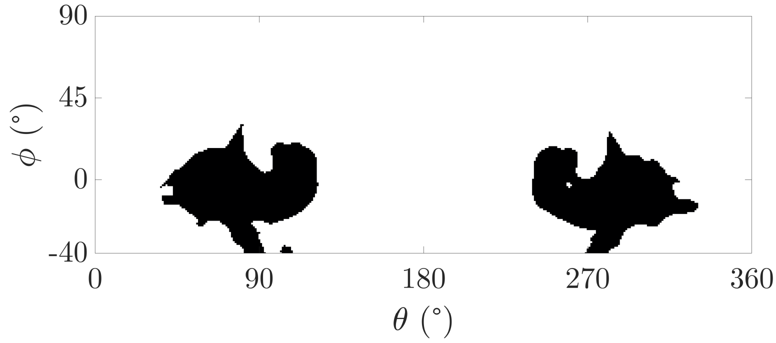
(c) 8 to 12 kHz

(d) 12 to 16 kHz

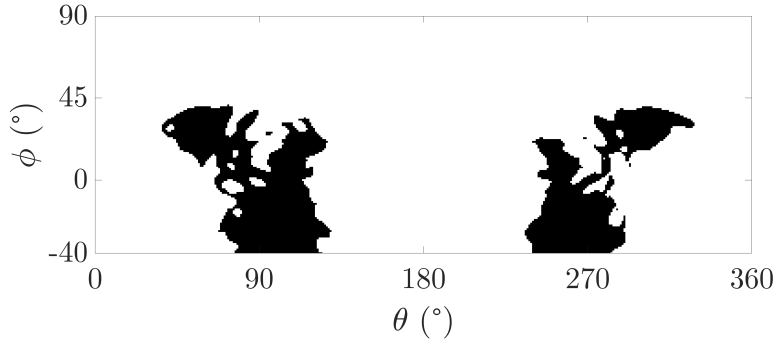
Figure E.13: Regions of $\mathbf{SP^{KEMAR}_{20_f}}$ positions (indicated in black) within $\mathbf{SP^{KEMAR}}$ at $r =$ 1.00 m for the four frequency bands used in analysis.



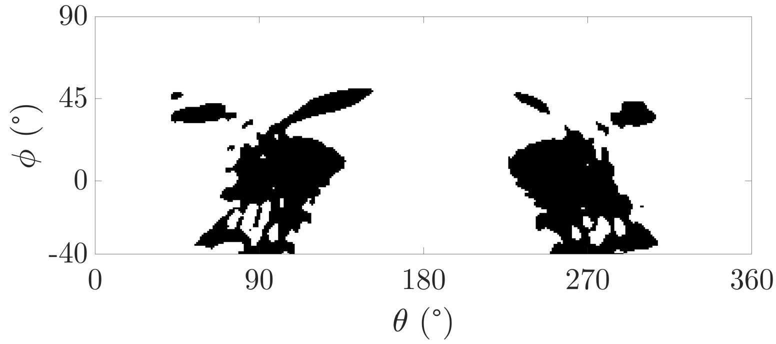
(a) 200 Hz to 4 kHz



(b) 4 to 8 kHz



(c) 8 to 12 kHz



(d) 12 to 16 kHz

Figure E.14: Regions of $\mathbf{SP^{KEMAR}_{20_f}}$ positions (indicated in black) within $\mathbf{SP^{KEMAR}}$ at $r = 1.50\,\mathrm{m}$ for the four frequency bands used in analysis.

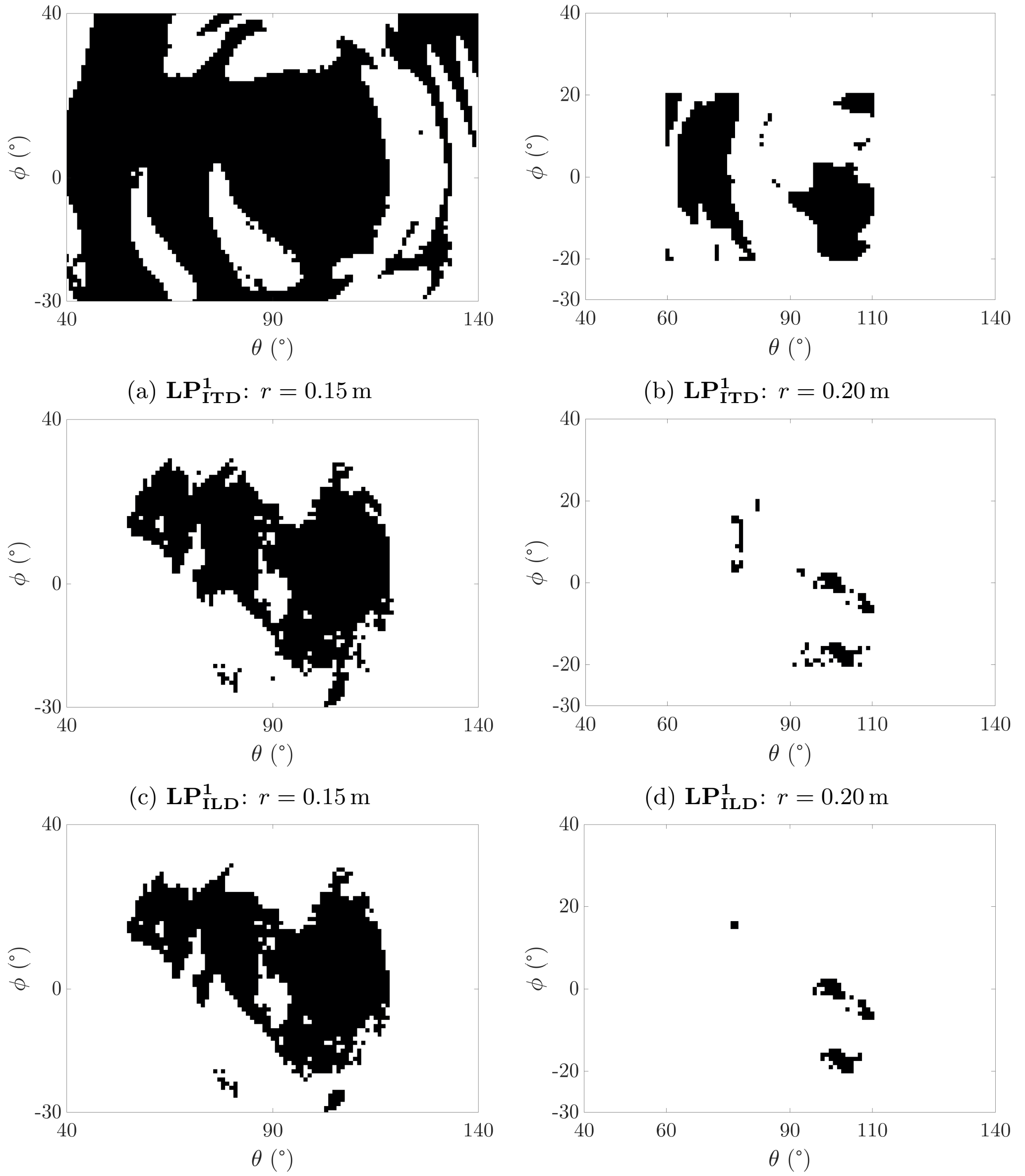
Appendix F

Additional Figures: Candidate Loudspeaker Pairs for SYMARE Subjects

This appendix contains the plots of six regions of loudspeaker pair positions for each of the 15 subjects from the SYMARE database [80] used in the analysis discussed in Chapter 7:

- The set of loudspeaker pair positions with deterioration below 1 JND for ITD cues ($\mathbf{LP^S_{ITD}}$) at $r = 0.15\,\mathrm{m}$ and $r = 0.20\,\mathrm{m}$
- The set of loudspeaker pair positions with deterioration below 1 JND for ILD cues ($\mathbf{LP^S_{ILD}}$) at $r = 0.15\,\mathrm{m}$ and $r = 0.20\,\mathrm{m}$
- The set of loudspeaker pair positions with deterioration below 1 JND for both sets of cues ($\mathbf{LP^S_{CM}}$) at $r = 0.15\,\mathrm{m}$ and $r = 0.20\,\mathrm{m}$

The integer S refers to the index of the SYMARE subject.

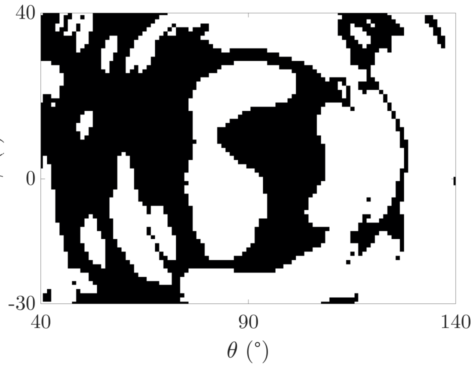


(a) $\mathbf{LP^1_{ITD}}$: $r = 0.15\,\mathrm{m}$ (b) $\mathbf{LP^1_{ITD}}$: $r = 0.20\,\mathrm{m}$

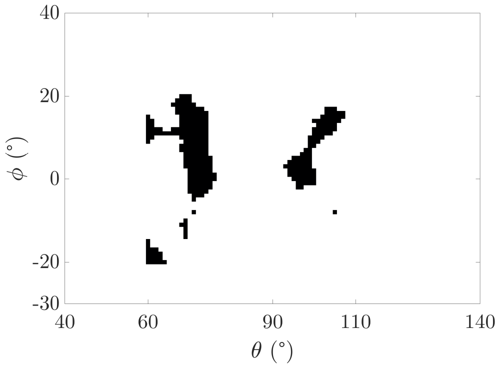
(c) $\mathbf{LP^1_{ILD}}$: $r = 0.15\,\mathrm{m}$ (d) $\mathbf{LP^1_{ILD}}$: $r = 0.20\,\mathrm{m}$

(e) $\mathbf{LP^1_{CM}}$: $r = 0.15\,\mathrm{m}$ (f) $\mathbf{LP^1_{CM}}$: $r = 0.20\,\mathrm{m}$

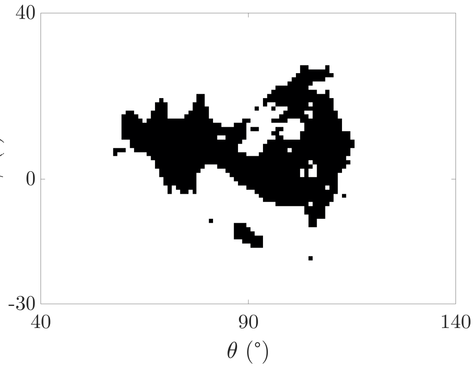
Figure F.1: Regions of loudspeaker pair positions, indicated in black, with deterioration below 1 JND for: ITD cues ($\mathbf{LP^S_{ITD}}$, top row); ILD cues ($\mathbf{LP^S_{ILD}}$, middle row) and both types of cue at once ($\mathbf{LP^S_{CM}}$, bottom row) for SYMARE subject #1.



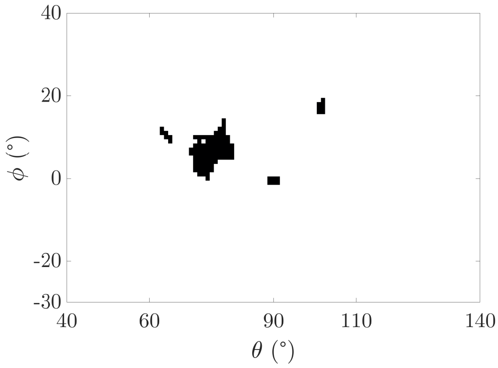
(a) $\mathbf{LP^2_{ITD}}$: $r = 0.15\,\mathrm{m}$



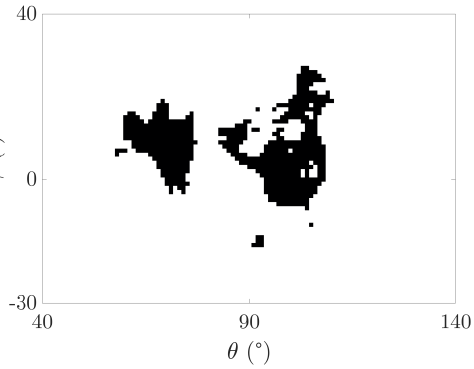
(b) $\mathbf{LP^2_{ITD}}$: $r = 0.20\,\mathrm{m}$



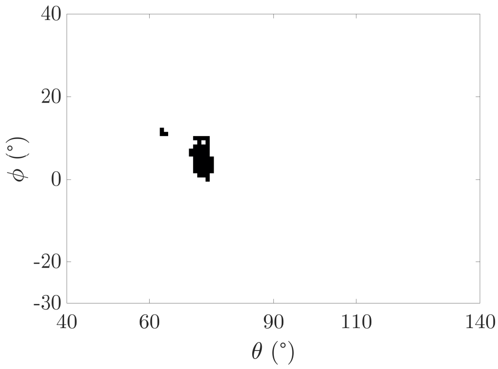
(c) $\mathbf{LP^2_{ILD}}$: $r = 0.15\,\mathrm{m}$



(d) $\mathbf{LP^2_{ILD}}$: $r = 0.20\,\mathrm{m}$

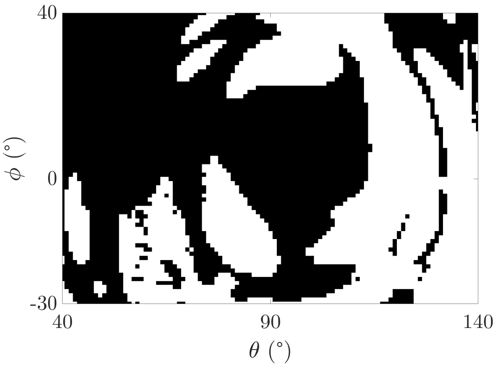


(e) $\mathbf{LP^2_{CM}}$: $r = 0.15\,\mathrm{m}$

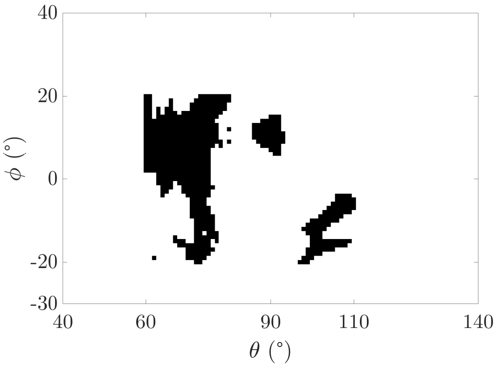


(f) $\mathbf{LP^2_{CM}}$: $r = 0.20\,\mathrm{m}$

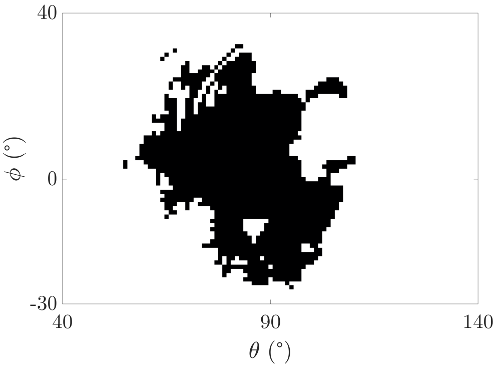
Figure F.2: Regions of loudspeaker pair positions, indicated in black, with deterioration below 1 JND for: ITD cues ($\mathbf{LP^S_{ITD}}$, top row); ILD cues ($\mathbf{LP^S_{ILD}}$, middle row) and both types of cue at once ($\mathbf{LP^S_{CM}}$, bottom row) for SYMARE subject #2.



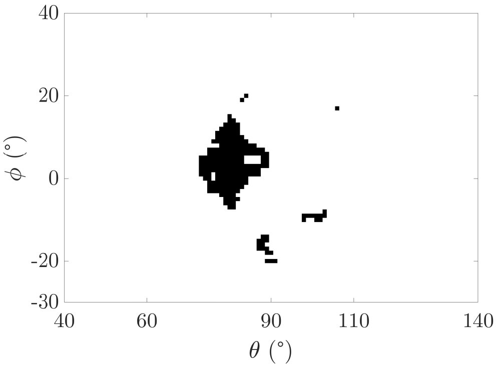
(a) $\mathbf{LP^3_{ITD}}$: $r = 0.15\,\mathrm{m}$



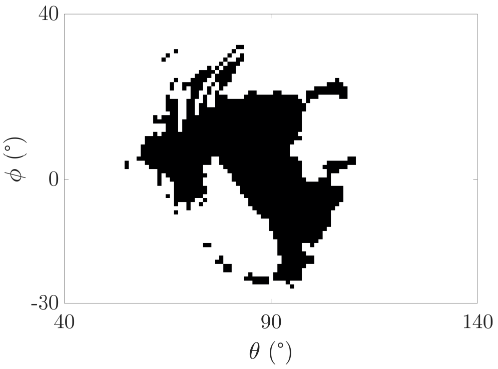
(b) $\mathbf{LP^3_{ITD}}$: $r = 0.20\,\mathrm{m}$



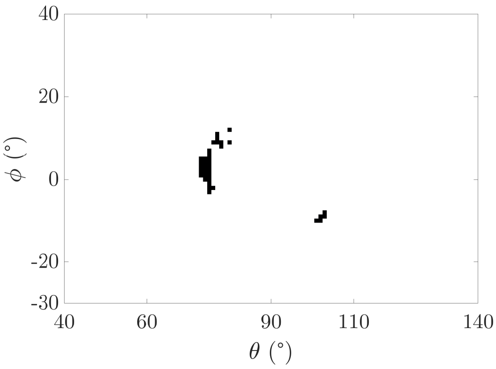
(c) $\mathbf{LP^3_{ILD}}$: $r = 0.15\,\mathrm{m}$



(d) $\mathbf{LP^3_{ILD}}$: $r = 0.20\,\mathrm{m}$

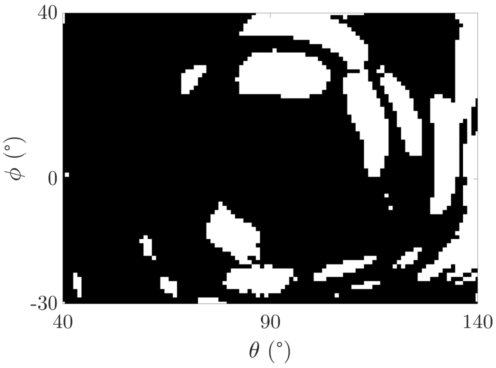


(e) $\mathbf{LP^3_{CM}}$: $r = 0.15\,\mathrm{m}$

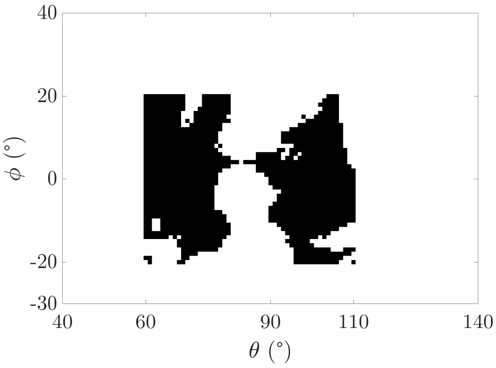


(f) $\mathbf{LP^3_{CM}}$: $r = 0.20\,\mathrm{m}$

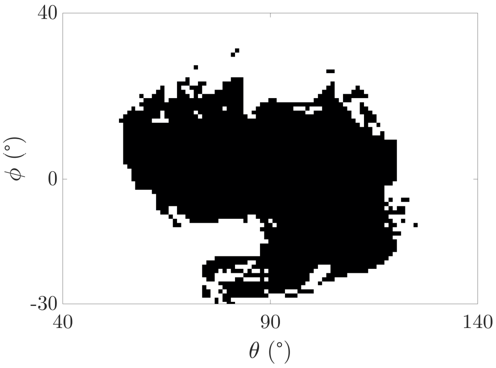
Figure F.3: Regions of loudspeaker pair positions, indicated in black, with deterioration below 1 JND for: ITD cues ($\mathbf{LP^S_{ITD}}$, top row); ILD cues ($\mathbf{LP^S_{ILD}}$, middle row) and both types of cue at once ($\mathbf{LP^S_{CM}}$, bottom row) for SYMARE subject #3.



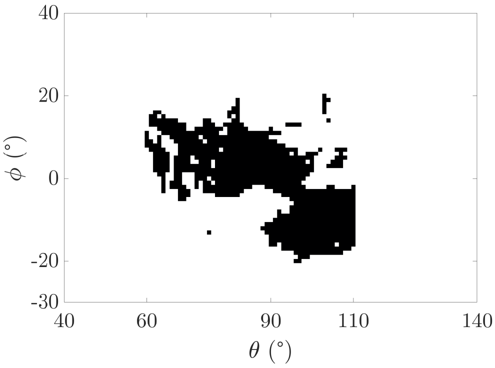
(a) $\mathbf{LP^4_{ITD}}$: $r = 0.15\,\mathrm{m}$



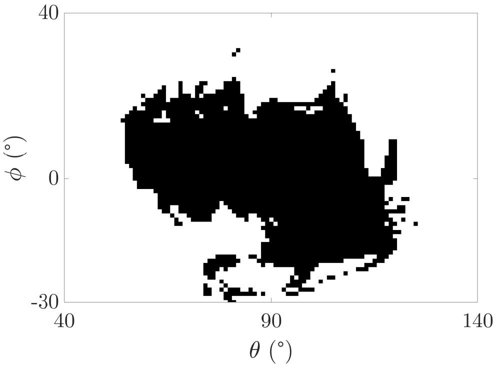
(b) $\mathbf{LP^4_{ITD}}$: $r = 0.20\,\mathrm{m}$



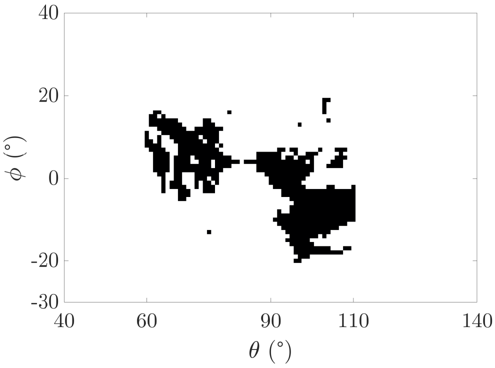
(c) $\mathbf{LP^4_{ILD}}$: $r = 0.15\,\mathrm{m}$



(d) $\mathbf{LP^4_{ILD}}$: $r = 0.20\,\mathrm{m}$

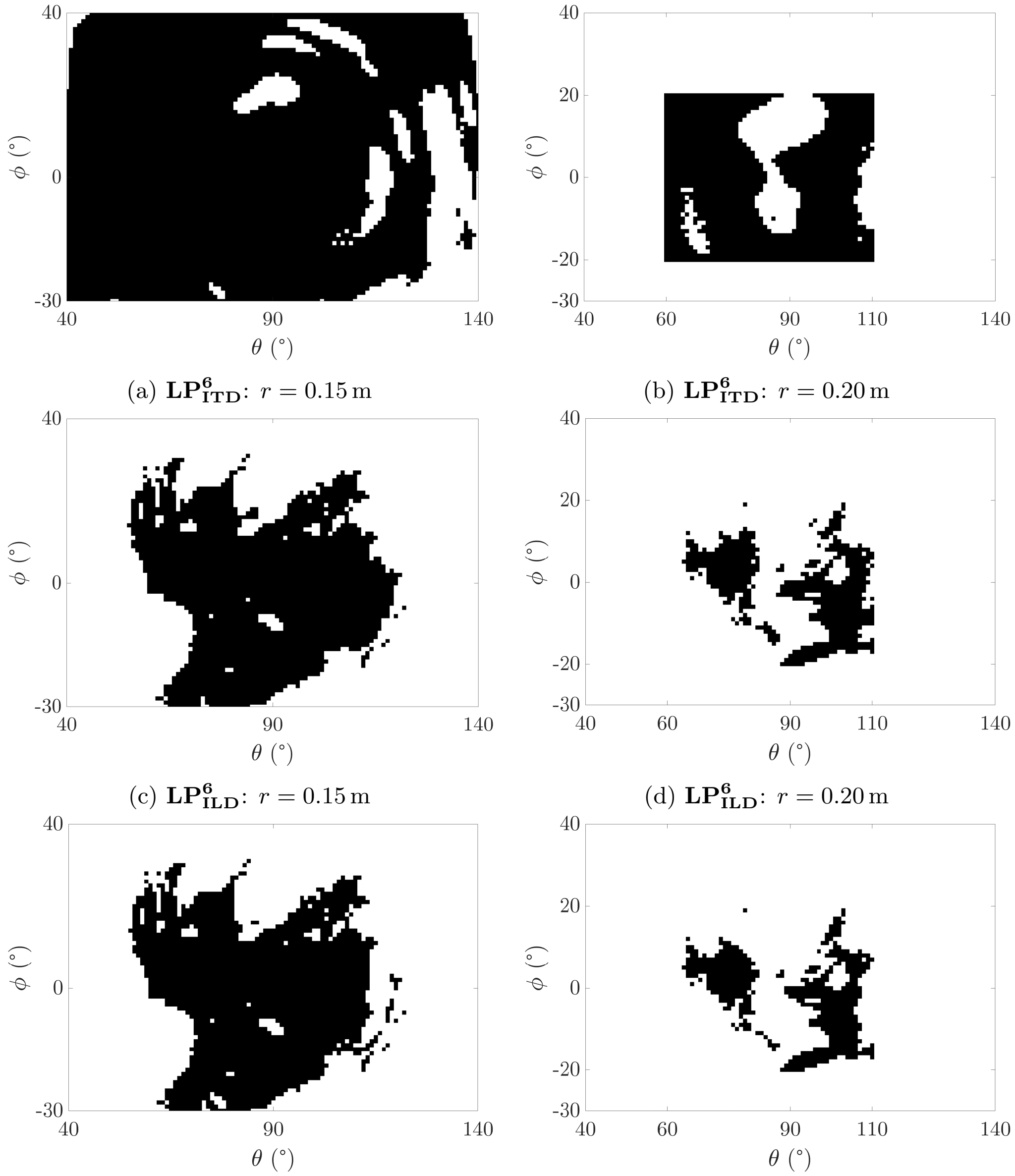


(e) $\mathbf{LP^4_{CM}}$: $r = 0.15\,\mathrm{m}$



(f) $\mathbf{LP^4_{CM}}$: $r = 0.20\,\mathrm{m}$

Figure F.4: Regions of loudspeaker pair positions, indicated in black, with deterioration below 1 JND for: ITD cues ($\mathbf{LP^S_{ITD}}$, top row); ILD cues ($\mathbf{LP^S_{ILD}}$, middle row) and both types of cue at once ($\mathbf{LP^S_{CM}}$, bottom row) for SYMARE subject #4.

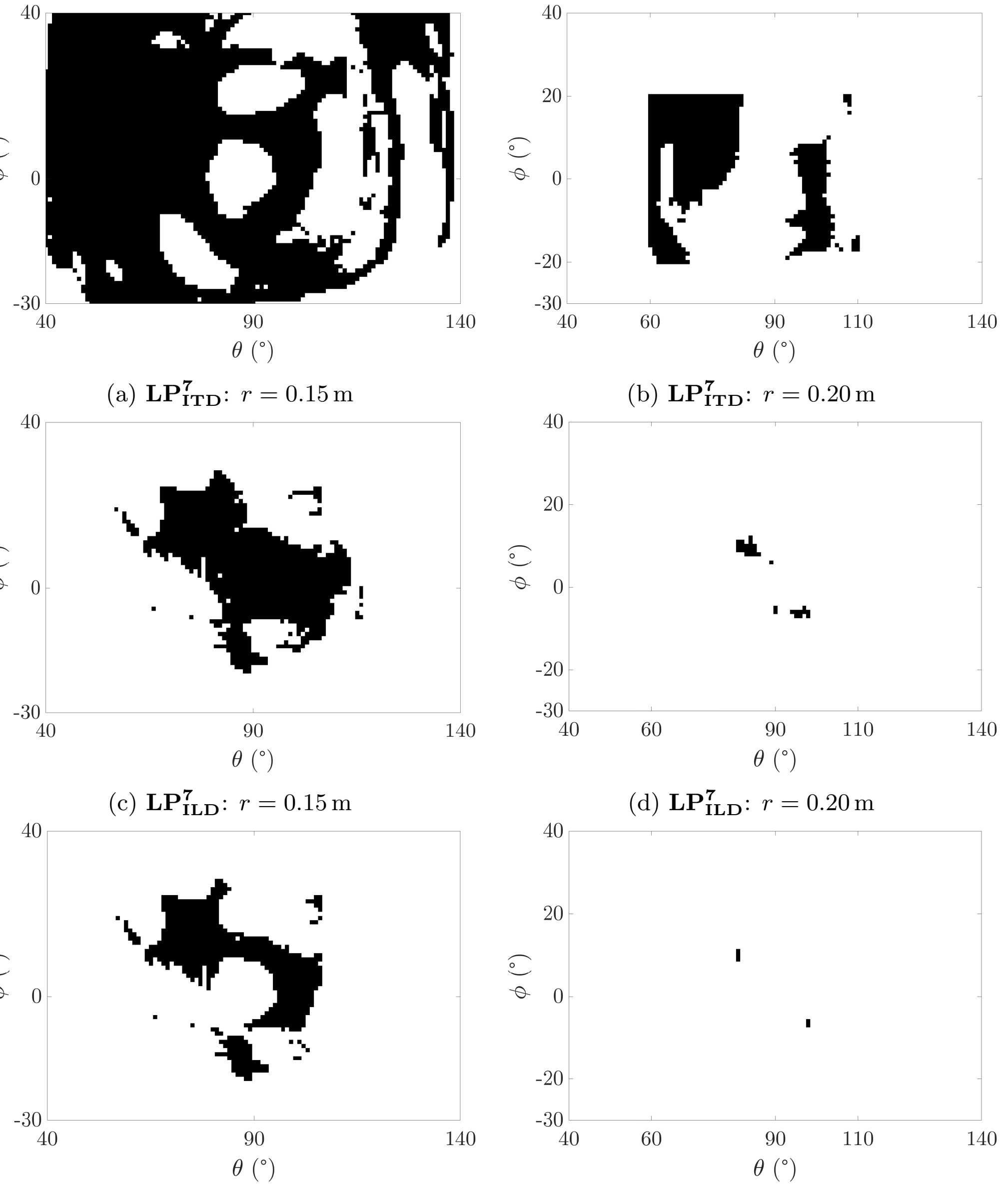


(a) $\mathbf{LP^{6}_{ITD}}$: $r = 0.15\,\mathrm{m}$ (b) $\mathbf{LP^{6}_{ITD}}$: $r = 0.20\,\mathrm{m}$

(c) $\mathbf{LP^{6}_{ILD}}$: $r = 0.15\,\mathrm{m}$ (d) $\mathbf{LP^{6}_{ILD}}$: $r = 0.20\,\mathrm{m}$

(e) $\mathbf{LP^{6}_{CM}}$: $r = 0.15\,\mathrm{m}$ (f) $\mathbf{LP^{6}_{CM}}$: $r = 0.20\,\mathrm{m}$

Figure F.5: Regions of loudspeaker pair positions, indicated in black, with deterioration below 1 JND for: ITD cues ($\mathbf{LP^{S}_{ITD}}$, top row); ILD cues ($\mathbf{LP^{S}_{ILD}}$, middle row) and both types of cue at once ($\mathbf{LP^{S}_{CM}}$, bottom row) for SYMARE subject #6.



(a) $\mathbf{LP^{7}_{ITD}}$: $r = 0.15\,\mathrm{m}$

(b) $\mathbf{LP^{7}_{ITD}}$: $r = 0.20\,\mathrm{m}$

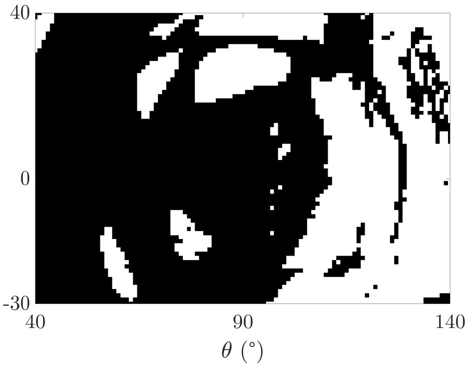
(c) $\mathbf{LP^{7}_{ILD}}$: $r = 0.15\,\mathrm{m}$

(d) $\mathbf{LP^{7}_{ILD}}$: $r = 0.20\,\mathrm{m}$

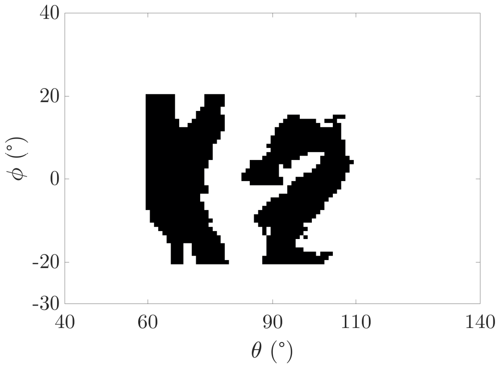
(e) $\mathbf{LP^{7}_{CM}}$: $r = 0.15\,\mathrm{m}$

(f) $\mathbf{LP^{7}_{CM}}$: $r = 0.20\,\mathrm{m}$

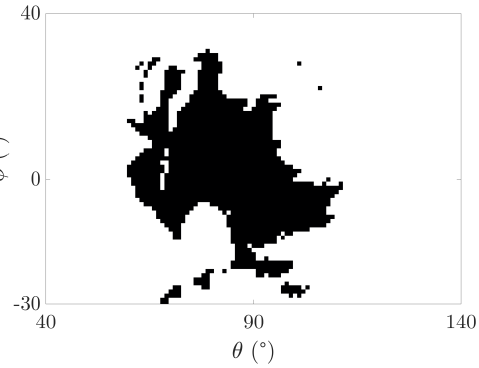
Figure F.6: Regions of loudspeaker pair positions, indicated in black, with deterioration below 1 JND for: ITD cues ($\mathbf{LP^{S}_{ITD}}$, top row); ILD cues ($\mathbf{LP^{S}_{ILD}}$, middle row) and both types of cue at once ($\mathbf{LP^{S}_{CM}}$, bottom row) for SYMARE subject #7.



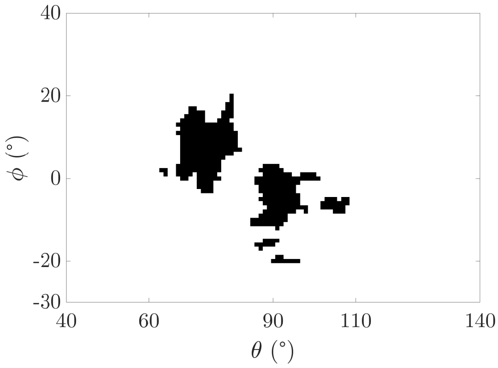
(a) $\mathbf{LP^{8}_{ITD}}$: $r = 0.15\,\mathrm{m}$



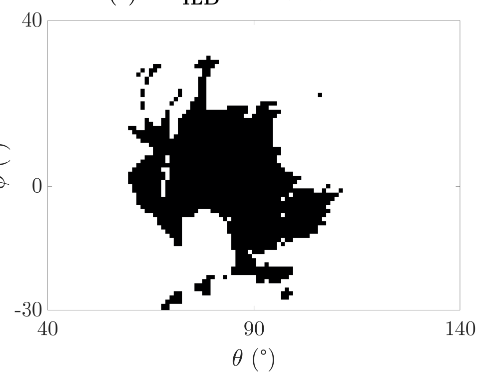
(b) $\mathbf{LP^{8}_{ITD}}$: $r = 0.20\,\mathrm{m}$



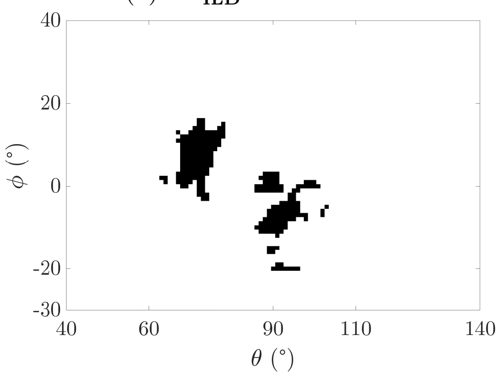
(c) $\mathbf{LP^{8}_{ILD}}$: $r = 0.15\,\mathrm{m}$



(d) $\mathbf{LP^{8}_{ILD}}$: $r = 0.20\,\mathrm{m}$

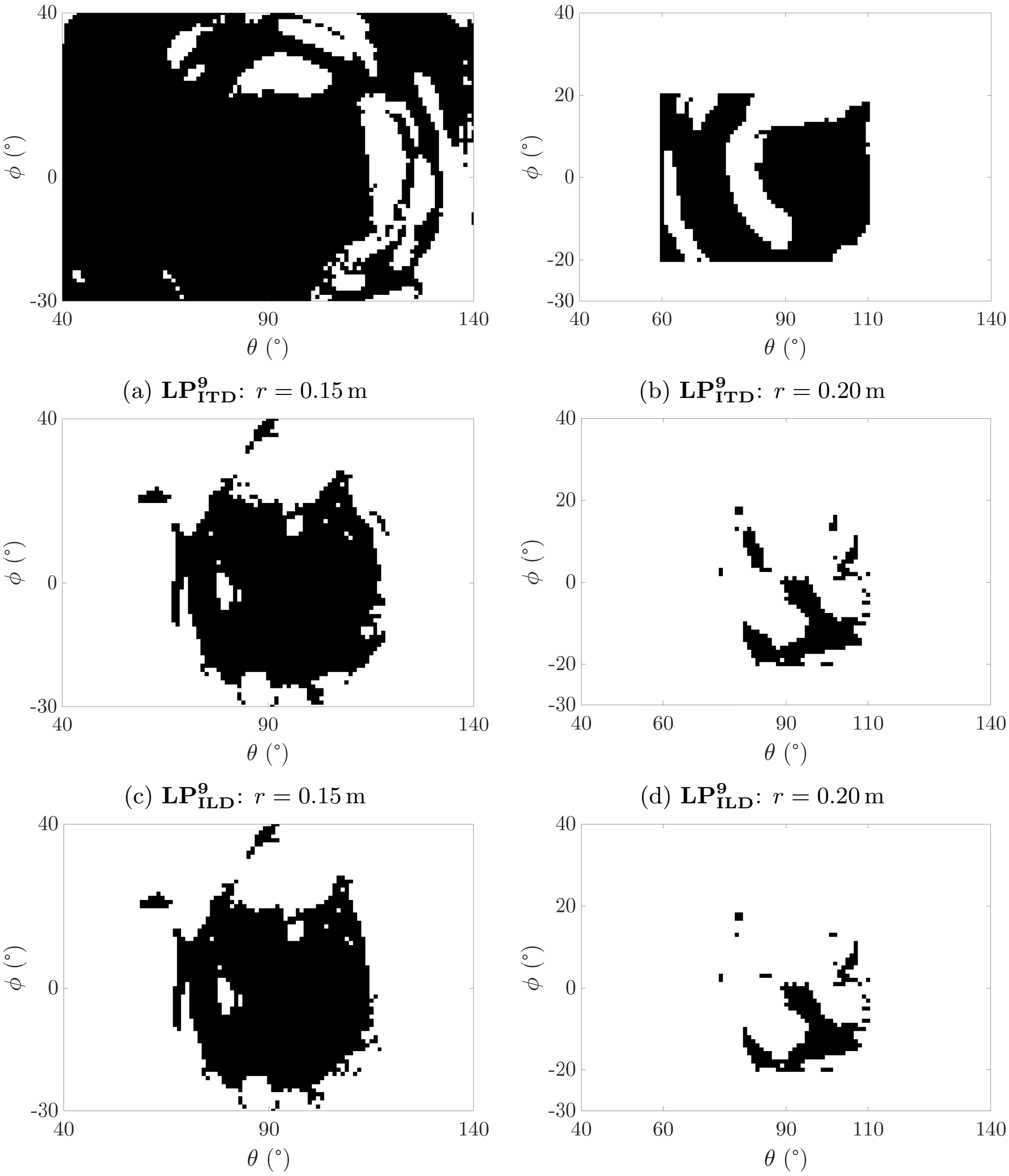


(e) $\mathbf{LP^{8}_{CM}}$: $r = 0.15\,\mathrm{m}$



(f) $\mathbf{LP^{8}_{CM}}$: $r = 0.20\,\mathrm{m}$

Figure F.7: Regions of loudspeaker pair positions, indicated in black, with deterioration below 1 JND for: ITD cues ($\mathbf{LP^{S}_{ITD}}$, top row); ILD cues ($\mathbf{LP^{S}_{ILD}}$, middle row) and both types of cue at once ($\mathbf{LP^{S}_{CM}}$, bottom row) for SYMARE subject #8.

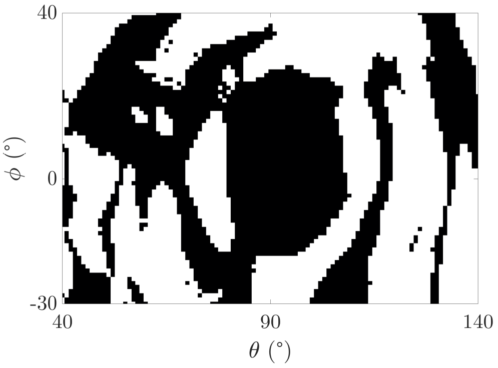


(a) $\mathbf{LP^{9}_{ITD}}$: $r = 0.15\,\mathrm{m}$ (b) $\mathbf{LP^{9}_{ITD}}$: $r = 0.20\,\mathrm{m}$

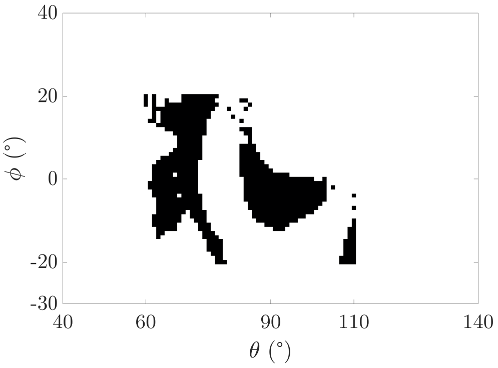
(c) $\mathbf{LP^{9}_{ILD}}$: $r = 0.15\,\mathrm{m}$ (d) $\mathbf{LP^{9}_{ILD}}$: $r = 0.20\,\mathrm{m}$

(e) $\mathbf{LP^{9}_{CM}}$: $r = 0.15\,\mathrm{m}$ (f) $\mathbf{LP^{9}_{CM}}$: $r = 0.20\,\mathrm{m}$

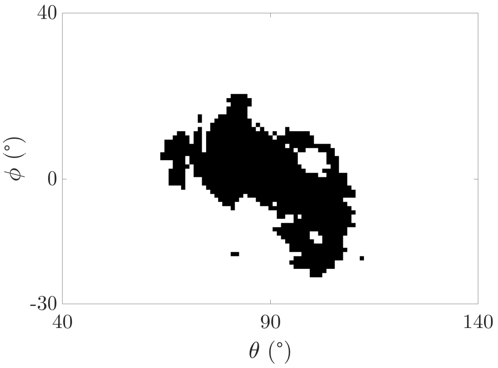
Figure F.8: Regions of loudspeaker pair positions, indicated in black, with deterioration below 1 JND for: ITD cues ($\mathbf{LP^{S}_{ITD}}$, top row); ILD cues ($\mathbf{LP^{S}_{ILD}}$, middle row) and both types of cue at once ($\mathbf{LP^{S}_{CM}}$, bottom row) for SYMARE subject #9.



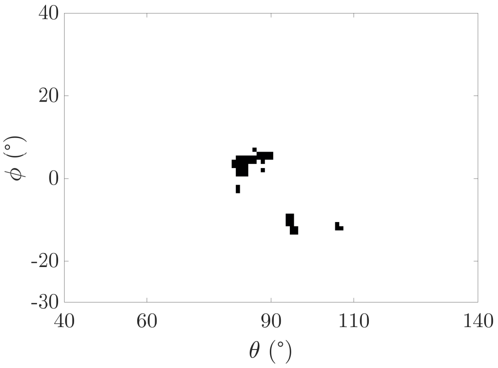
(a) $\mathbf{LP^{10}_{ITD}}$: $r = 0.15\,\mathrm{m}$



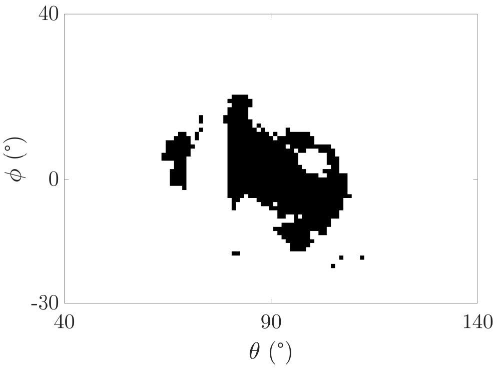
(b) $\mathbf{LP^{10}_{ITD}}$: $r = 0.20\,\mathrm{m}$



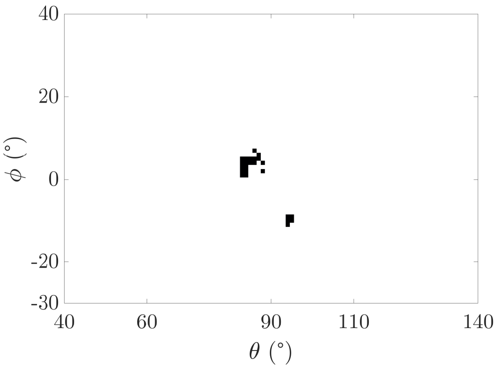
(c) $\mathbf{LP^{10}_{ILD}}$: $r = 0.15\,\mathrm{m}$



(d) $\mathbf{LP^{10}_{ILD}}$: $r = 0.20\,\mathrm{m}$

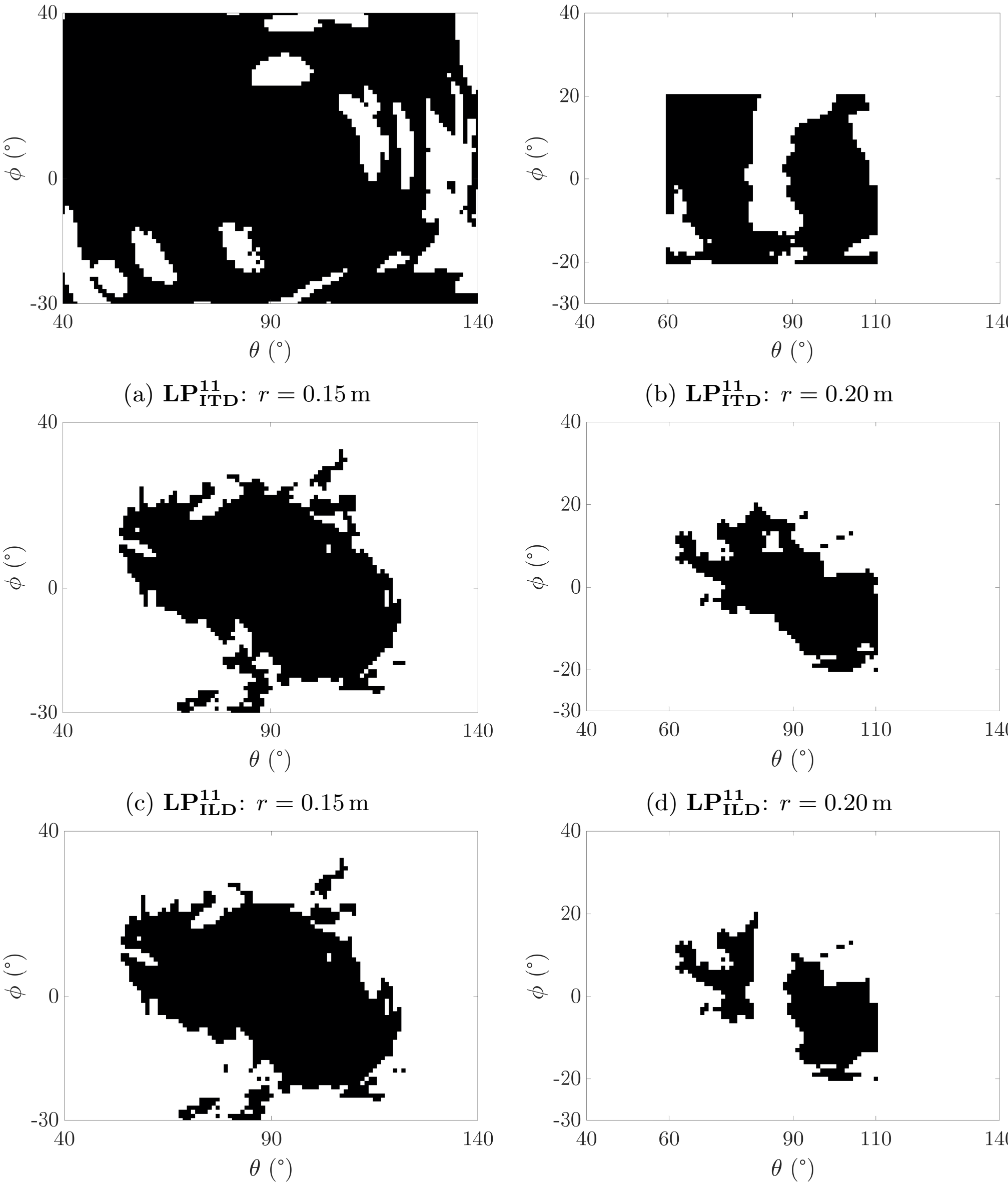


(e) $\mathbf{LP^{10}_{CM}}$: $r = 0.15\,\mathrm{m}$



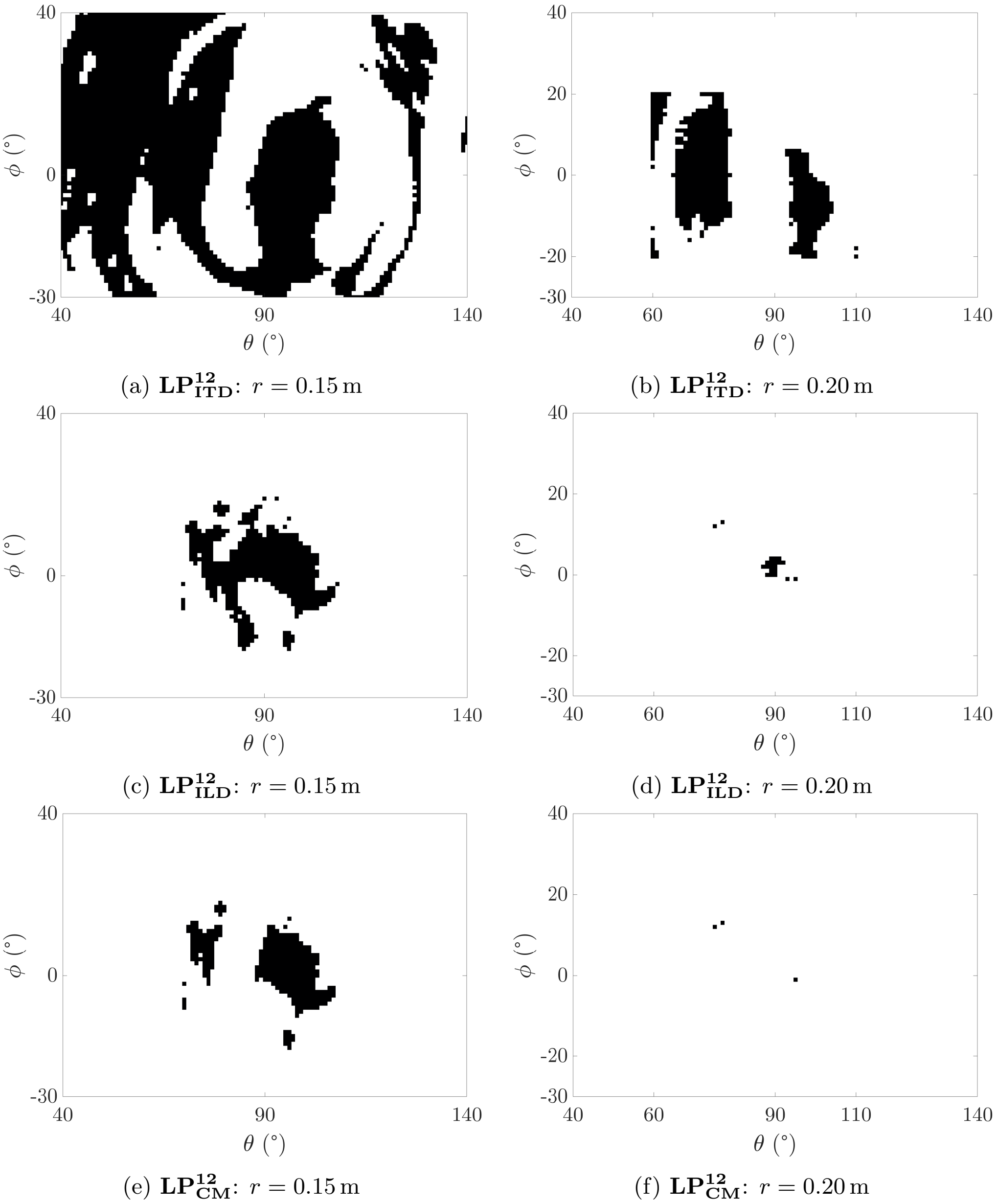
(f) $\mathbf{LP^{10}_{CM}}$: $r = 0.20\,\mathrm{m}$

Figure F.9: Regions of loudspeaker pair positions, indicated in black, with deterioration below 1 JND for: ITD cues ($\mathbf{LP^{S}_{ITD}}$, top row); ILD cues ($\mathbf{LP^{S}_{ILD}}$, middle row) and both types of cue at once ($\mathbf{LP^{S}_{CM}}$, bottom row) for SYMARE subject #10.



(a) $\mathbf{LP^{11}_{ITD}}$: $r = 0.15\,\mathrm{m}$
(b) $\mathbf{LP^{11}_{ITD}}$: $r = 0.20\,\mathrm{m}$
(c) $\mathbf{LP^{11}_{ILD}}$: $r = 0.15\,\mathrm{m}$
(d) $\mathbf{LP^{11}_{ILD}}$: $r = 0.20\,\mathrm{m}$
(e) $\mathbf{LP^{11}_{CM}}$: $r = 0.15\,\mathrm{m}$
(f) $\mathbf{LP^{11}_{CM}}$: $r = 0.20\,\mathrm{m}$

Figure F.10: Regions of loudspeaker pair positions, indicated in black, with deterioration below 1 JND for: ITD cues ($\mathbf{LP^{S}_{ITD}}$, top row); ILD cues ($\mathbf{LP^{S}_{ILD}}$, middle row) and both types of cue at once ($\mathbf{LP^{S}_{CM}}$, bottom row) for SYMARE subject #11.



(a) $\mathbf{LP^{12}_{ITD}}$: $r = 0.15\,\mathrm{m}$

(b) $\mathbf{LP^{12}_{ITD}}$: $r = 0.20\,\mathrm{m}$

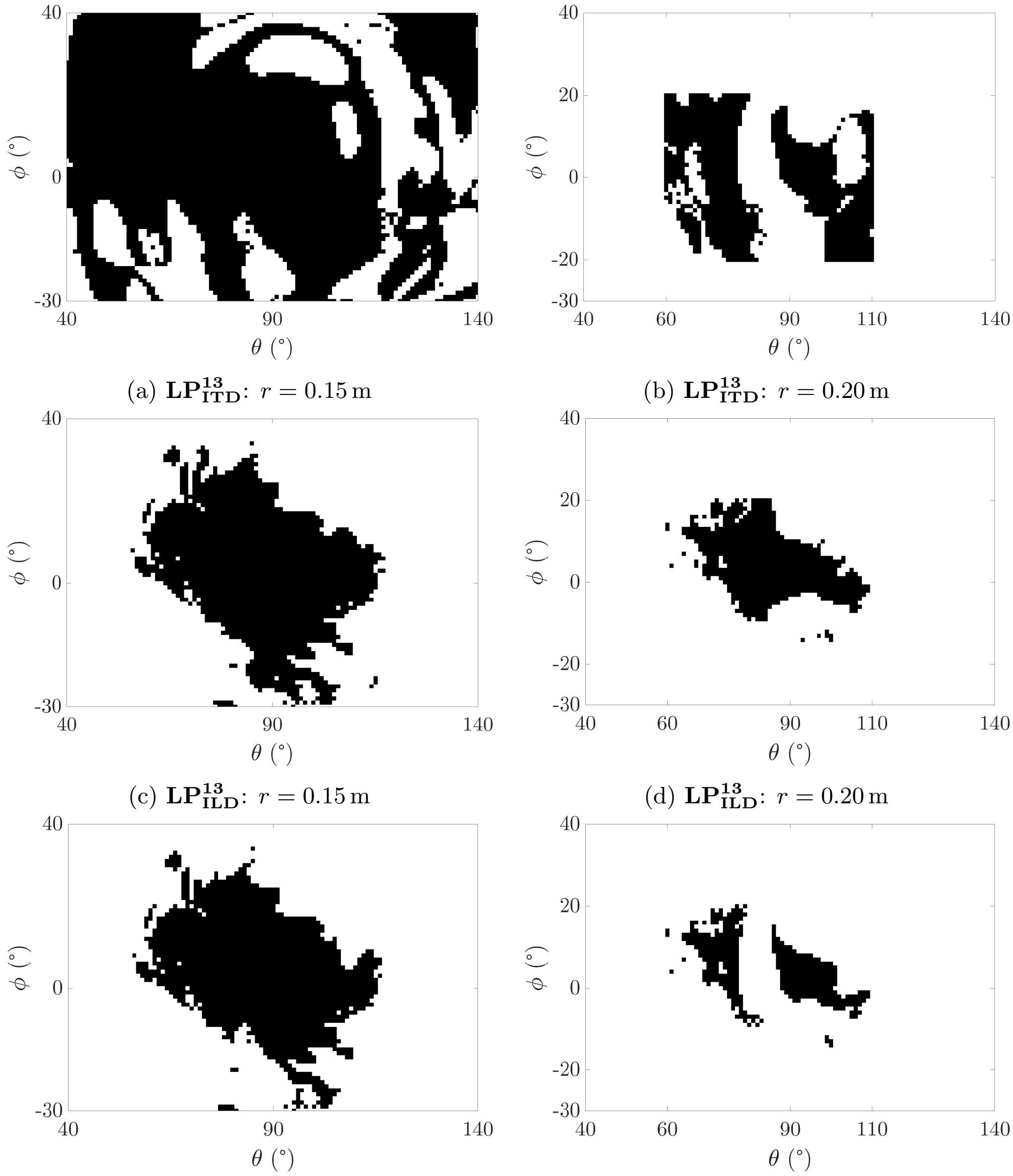
(c) $\mathbf{LP^{12}_{ILD}}$: $r = 0.15\,\mathrm{m}$

(d) $\mathbf{LP^{12}_{ILD}}$: $r = 0.20\,\mathrm{m}$

(e) $\mathbf{LP^{12}_{CM}}$: $r = 0.15\,\mathrm{m}$

(f) $\mathbf{LP^{12}_{CM}}$: $r = 0.20\,\mathrm{m}$

Figure F.11: Regions of loudspeaker pair positions, indicated in black, with deterioration below 1 JND for: ITD cues ($\mathbf{LP^{S}_{ITD}}$, top row); ILD cues ($\mathbf{LP^{S}_{ILD}}$, middle row) and both types of cue at once ($\mathbf{LP^{S}_{CM}}$, bottom row) for SYMARE subject #12.



(a) $\mathbf{LP^{13}_{ITD}}$: $r = 0.15\,\mathrm{m}$

(b) $\mathbf{LP^{13}_{ITD}}$: $r = 0.20\,\mathrm{m}$

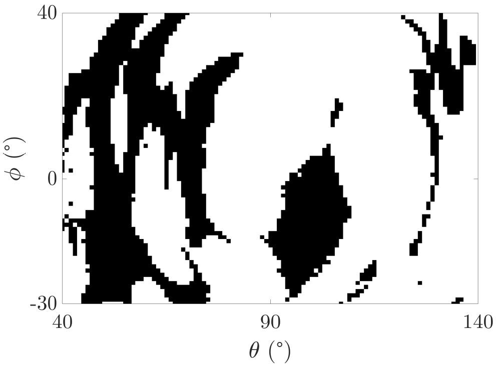
(c) $\mathbf{LP^{13}_{ILD}}$: $r = 0.15\,\mathrm{m}$

(d) $\mathbf{LP^{13}_{ILD}}$: $r = 0.20\,\mathrm{m}$

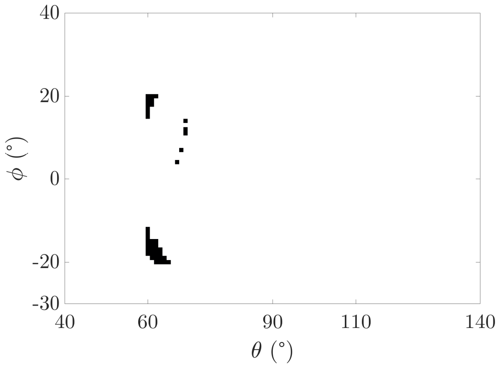
(e) $\mathbf{LP^{13}_{CM}}$: $r = 0.15\,\mathrm{m}$

(f) $\mathbf{LP^{13}_{CM}}$: $r = 0.20\,\mathrm{m}$

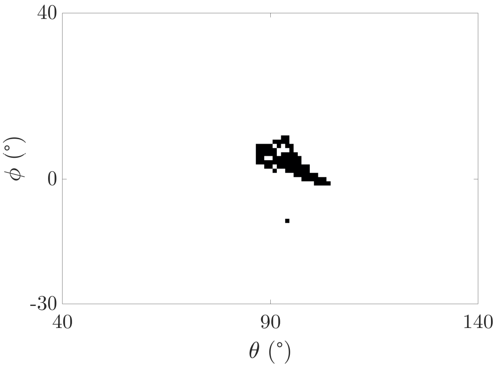
Figure F.12: Regions of loudspeaker pair positions, indicated in black, with deterioration below 1 JND for: ITD cues ($\mathbf{LP^{S}_{ITD}}$, top row); ILD cues ($\mathbf{LP^{S}_{ILD}}$, middle row) and both types of cue at once ($\mathbf{LP^{S}_{CM}}$, bottom row) for SYMARE subject #13.



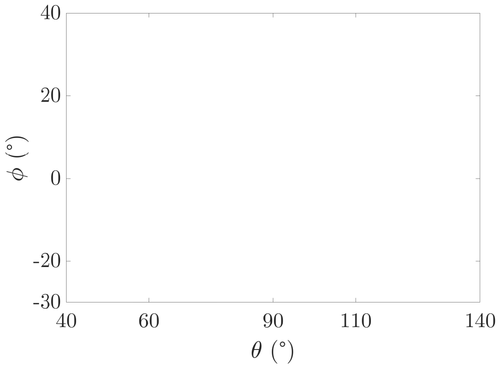
(a) $\mathbf{LP^{14}_{ITD}}$: $r = 0.15\,\mathrm{m}$



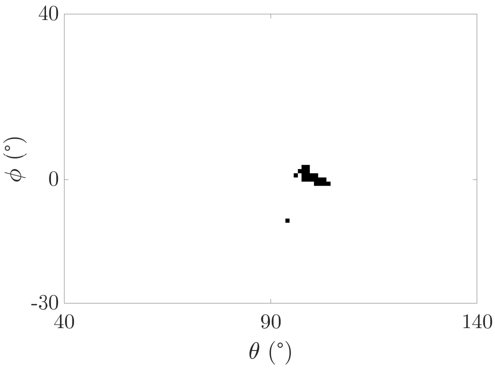
(b) $\mathbf{LP^{14}_{ITD}}$: $r = 0.20\,\mathrm{m}$



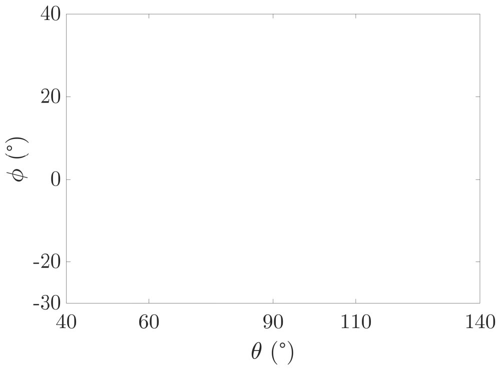
(c) $\mathbf{LP^{14}_{ILD}}$: $r = 0.15\,\mathrm{m}$



(d) $\mathbf{LP^{14}_{ILD}}$: $r = 0.20\,\mathrm{m}$

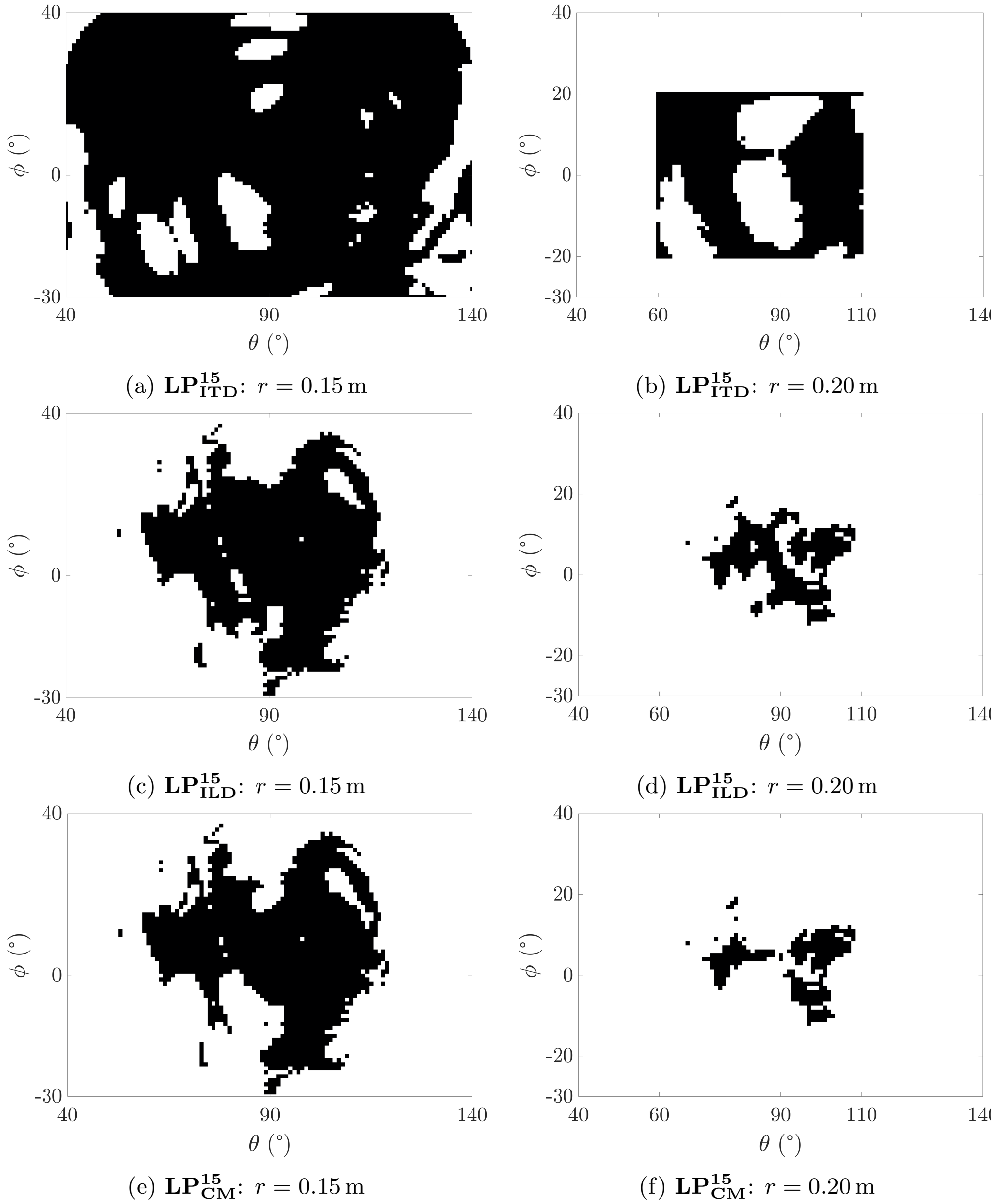


(e) $\mathbf{LP^{14}_{CM}}$: $r = 0.15\,\mathrm{m}$



(f) $\mathbf{LP^{14}_{CM}}$: $r = 0.20\,\mathrm{m}$

Figure F.13: Regions of loudspeaker pair positions, indicated in black, with deterioration below 1 JND for: ITD cues ($\mathbf{LP^{S}_{ITD}}$, top row); ILD cues ($\mathbf{LP^{S}_{ILD}}$, middle row) and both types of cue at once ($\mathbf{LP^{S}_{CM}}$, bottom row) for SYMARE subject #14.



(a) $\mathbf{LP^{15}_{ITD}}$: $r = 0.15\,\mathrm{m}$

(b) $\mathbf{LP^{15}_{ITD}}$: $r = 0.20\,\mathrm{m}$

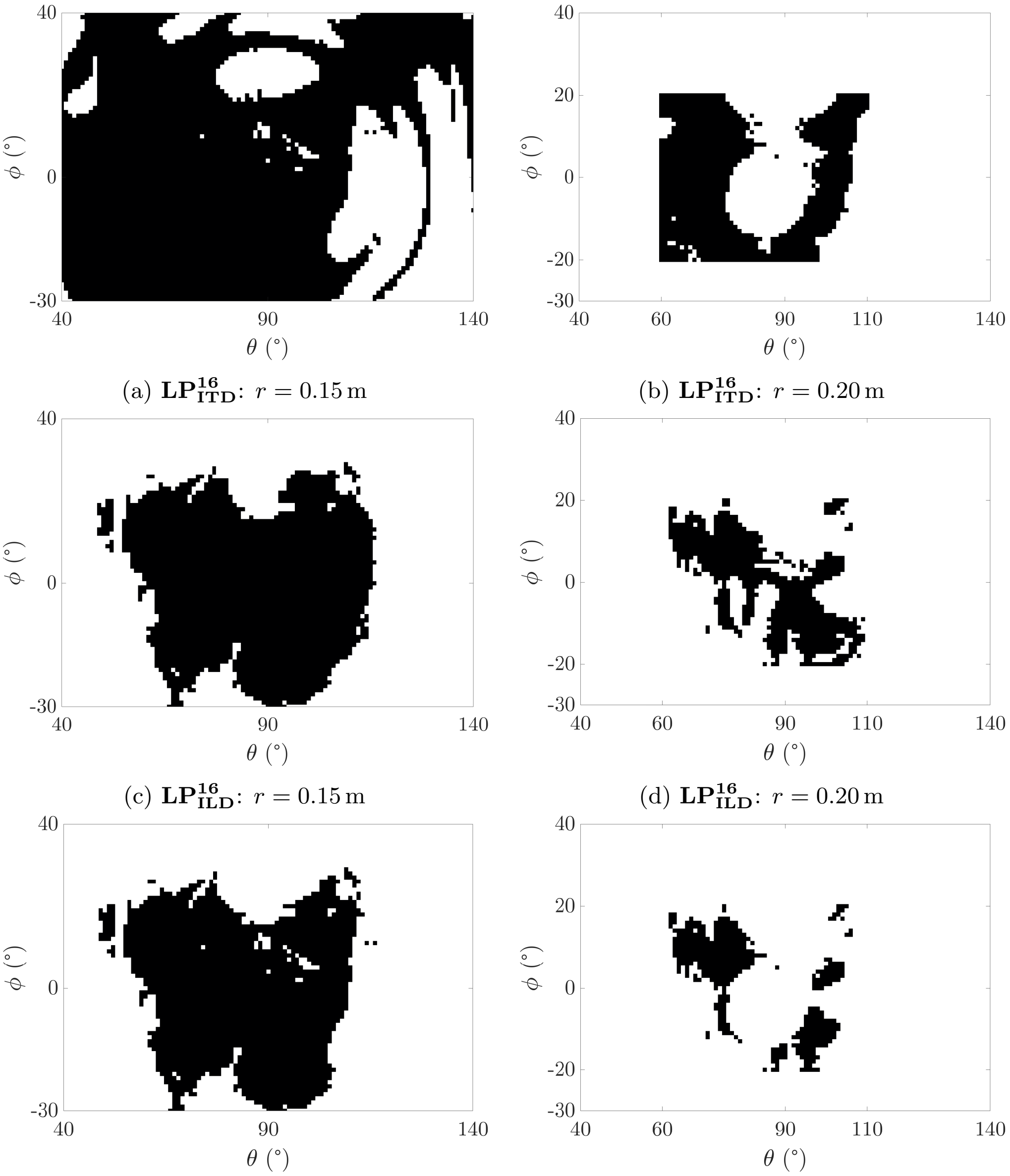
(c) $\mathbf{LP^{15}_{ILD}}$: $r = 0.15\,\mathrm{m}$

(d) $\mathbf{LP^{15}_{ILD}}$: $r = 0.20\,\mathrm{m}$

(e) $\mathbf{LP^{15}_{CM}}$: $r = 0.15\,\mathrm{m}$

(f) $\mathbf{LP^{15}_{CM}}$: $r = 0.20\,\mathrm{m}$

Figure F.14: Regions of loudspeaker pair positions, indicated in black, with deterioration below 1 JND for: ITD cues ($\mathbf{LP^{S}_{ITD}}$, top row); ILD cues ($\mathbf{LP^{S}_{ILD}}$, middle row) and both types of cue at once ($\mathbf{LP^{S}_{CM}}$, bottom row) for SYMARE subject #15.



(a) $\mathbf{LP^{16}_{ITD}}$: $r = 0.15\,\mathrm{m}$ (b) $\mathbf{LP^{16}_{ITD}}$: $r = 0.20\,\mathrm{m}$

(c) $\mathbf{LP^{16}_{ILD}}$: $r = 0.15\,\mathrm{m}$ (d) $\mathbf{LP^{16}_{ILD}}$: $r = 0.20\,\mathrm{m}$

(e) $\mathbf{LP^{16}_{CM}}$: $r = 0.15\,\mathrm{m}$ (f) $\mathbf{LP^{16}_{CM}}$: $r = 0.20\,\mathrm{m}$

Figure F.15: Regions of loudspeaker pair positions, indicated in black, with deterioration below 1 JND for: ITD cues ($\mathbf{LP^{S}_{ITD}}$, top row); ILD cues ($\mathbf{LP^{S}_{ILD}}$, middle row) and both types of cue at once ($\mathbf{LP^{S}_{CM}}$, bottom row) for SYMARE subject #16.

Appendix G

Accompanying Data

The folders of accompanying data have been separated by relevant chapter, then divided to meet upload size requirements. Each chapter folder contains a README.txt to describe the contents in more detail. For convenience, these README files have been additionally included as a separate folder. The data folders are organised as follows:

AccompanyingData_1.zip (10 kB)

- chapter3_KEMARmesh_README.txt
- chapter4_halfTorsoMesh_README.txt
- chapter5_NCSvariation_README.txt
- chapter6_candidatePositions-KEMAR_1_README.txt
- chapter6_candidatePositions-KEMAR_2_README.txt
- chapter6_candidatePositions-KEMAR_3_README.txt
- chapter6_candidatePositions-KEMAR_4_README.txt
- chapter7_candidatePositions-SYMARE_1_README.txt
- chapter7_candidatePositions-SYMARE_2_README.txt

AccompanyingData_2.zip (7.75 GB)

- chapter3_KEMARmesh (273 MB)
  - README.txt
  - Meshes
    - full-torso-20.ply
    - full-torso-noEC-16.ply
    - full-torso-noEC-20.ply
    - original-polygonal.ply
    - remeshed-2mm.ply
    - NEECK-16.ply
    - NEECK-20.ply
  - measuredData.mat
  - simulatedData.mat
- chapter4_halfTorsoMesh (50.5 MB)
  - README.txt
  - half-torso-17.ply
  - half-torso-20.ply
  - head-and-shoulders-sliced.ply
  - joined.ply
  - rounded-portion.ply
- chapter5_NCSvariation (7.44 GB)
  - README.txt
  - info.mat
  - simulatedData_radius0.10m.mat
  - simulatedData_radiusX.XXm.mat (x14)

AccompanyingData_3.zip (28.2 GB)

- chapter6_candidatePositions-KEMAR_1
  - README.txt
  - info.mat
  - modifiedHRIRpairs_azAelB.mat (x5)

AccompanyingData_4.zip (28.2 GB)

- chapter6_candidatePositions-KEMAR_2
  - README.txt
  - info.mat
  - modifiedHRIRpairs_azAelB.mat (x5)

AccompanyingData_5.zip (28.2 GB)

- chapter6_candidatePositions-KEMAR_3
  - README.txt
  - info.mat
  - modifiedHRIRpairs_azAelB.mat (x5)

AccompanyingData_6.zip (16.9 GB)

- chapter6_candidatePositions-KEMAR_4
  - README.txt
  - info.mat
  - modifiedHRIRpairs_azAelB.mat (x3)

AccompanyingData_7.zip (27.9 GB)
- chapter7_candidatePositions-SYMARE_1
  - README.txt
  - SubjectXX (x10)
    - simulatedData.mat
    - modifiedHRIRs.mat

AccompanyingData_8.zip (18.6 GB)
- chapter7_candidatePositions-SYMARE_2
  - README.txt
  - SubjectXX (x6)
    - simulatedData.mat
    - modifiedHRIRs.mat

List of Acronyms

$(\Delta ILD_{max} - \Delta ILD_{min})$	Range of ΔILD values
$(\Delta ITD_{max} - \Delta ITD_{min})$	Range of ΔITD values
ΔILD	Unsigned difference between two ILD values
ΔISD	Unsigned difference between two ISD values
ΔITD	Unsigned difference between two ITD values
ΔNCS	Signed difference between two NCS values
$\overline{\Delta ILD}$	Arithmetic mean of ΔILD values
$\overline{\Delta ITD}$	Arithmetic mean of ΔITD values
$\sigma_{\Delta ILD}$	Standard deviation of ΔILD values
$\sigma_{\Delta ITD}$	Standard deviation of ΔITD values
2D	Two-dimensional
3D	Three-dimensional
BEM	Boundary element method
CAD	Computer-aided design
CPU	Central Processing Unit
CS	Channel separation
CSI	Channel separation index
CT	Computerised tomography
CTC	Crosstalk cancellation
DIR	Directional impulse response
DTF	Directional transfer function
DVF	Distance variation function
FDTD	Finite-difference time-domain
FM-BEM	Fast multipole boundary element method
GPU	Graphics Processing Unit
HATS	Head-and-torso simulator
HpTF	Headphone to ear canal transfer function
HRIR	Head-related impulse response
HRTF	Head-related transfer function
IACC	Interaural cross-correlation
IFEM	Infinite-finite element method
ILD	Interaural level difference

IPD	Interaural phase difference
IR	Impulse response
ISD	Interaural spectral difference
ISD_f	Interaural spectral difference at a given frequency point f
ITD	Interaural time difference
JND	Just-noticeable difference
KEMAR	Knowles Electronics Manikin for Acoustic Research
LTI	Linear-time invariant
MLS	Maximal-length sequence
MRI	Magnetic resonance imaging
NCS	Natural channel separation
NEECK	Neck-Extended Easily Computable KEMAR
OSD	Optimal source distribution
PLY	Polygon File Format
RAM	Random-access memory
RMS	Root mean square
SADIE	Spatial Audio Domestic Interactive Entertainment (database)
SD_f	Unsigned mean spectral difference within a frequency band f
SPL	Sound pressure level
STEP	Standard for the Exchange of Product model data
SYMARE	Sydney-York Morphological and Recording of Ears (database)
TSS	Technical Support Services
UWVF	Ultra-weak variational formulation
VRS	Virtual reproduction system
WFS	Wave field synthesis

List of Symbols

β	Number of faces comprising the boundary B
Γ	Angle between the incident wave and a ray to the point on the sphere surface (°)
Δ	Amount of change in a value
$\delta(\mathbf{r}' - \mathbf{r})$	Dirac delta function
ζ	Ear (left or right)
η	Number of edges per wavelength in a BEM simulation
Θ	Starting phase of a sinusoidal wave (°)
θ	Azimuth angle (°)
θ_V	Angle between two vectors (°)
κ	Adiabatic exponent (for air, $\kappa = 1.4$)
λ	Wavelength (m)
ρ	Density ($\mathrm{kg\,m^{-3}}$)
ρ_0	Equilibrium density ($\mathrm{kg\,m^{-3}}$)
ρ_{air}	Density of air ($\mathrm{kg\,m^{-3}}$)
τ	Time shift between two signals (s)
ϕ	Elevation angle (°)
χ	Golden ratio $\left(\frac{1+\sqrt{5}}{2}\right)$
ψ_L	Phase at the left ear (°)
ψ_R	Phase at the right ear (°)
ω	Angular frequency (kc, $2\pi f$)
∇^2	Laplacian operator
$*$	Convolution operator
$\mathbb{R}$	The set of all real numbers
A	Measure of amplitude of a sinusoidal wave
A_f	Magnitude of a HRTF at frequency point f (dB)
a_{LL}	Time domain ipsilateral (direct) acoustic path from left loudspeaker to left ear
a_{RR}	Time domain ipsilateral (direct) acoustic path from right loudspeaker to right ear
a_{LR}	Time domain contralateral (crosstalk) acoustic path from left loudspeaker to right ear

a_{RL}	Time domain contralateral (crosstalk) acoustic path from right loudspeaker to left ear
a	Radius of the head (m)
B	Boundary used in the BEM
$B(\alpha)$	Bézier curve as a function of α
B_L, B_R	Fourier transforms of b_L and b_R
$b_{n,N}(\alpha)$	Bernstein basis polynomials of degree N
b_L, b_R	Time domain binaural signals presented to the left and right ears
C_{11}, C_{12}, C_{21}, C_{22}	Crosstalk cancellation filters
c	Speed of sound ($\mathrm{m\,s^{-1}}$)
$\mathbf{CP}_n$	Bézier curve control points
D_f	Magnitude of DTF at frequency point f (dB)
d	Path difference between ipsilateral and contralateral signals (m)
d_{max}	Length of longest edge in mesh (m)
E_L, E_R	Fourier transforms of e_L and e_R
e_L, e_R	Time domain signals at the left and right ears
f_{max}	Maximum valid frequency (Hz)
f	Frequency (Hz)
$G(\mathbf{r}', \mathbf{r}, f)$	Free-space Green function of a point source
H_L, H_R	Left and right of a HRTF pair for a given direction (dB)
$h_l(ka)$	Spherical Hankel function of the second kind of order l
h_L, h_R	Left and right of a HRIR pair for a given direction
h_L^N, h_R^N	Left and right of a reference HRIR pair (Chapter 6 and Chapter 7)
$\hat{h}_{M_L}^N$, $\hat{h}_{M_R}^N$	Left and right of a modified HRIR pair (Chapter 6 and Chapter 7)
i_{LL}, i_{RR}	Inverse filters for a_{LL} and a_{RR}
$\mathbf{I}$	Intersection point of an infinite plane defined by the 3D points P_1, P_2 and P_3 and the line passing through points L_1 and L_2
K	Bulk modulus of a material
k	Wave number $\left(\frac{2\pi f}{c}\right)$
$L\prime$, $R\prime$	Left and right signals to be reproduced using headphones or loudspeakers
M_6	Arithmetic mean of the six binaural statistical metrics for a loudspeaker pair: $\overline{\Delta ITD}$, $\sigma_{\Delta ITD}$, $\Delta ITD_{max} - \Delta ITD_{min}$, $\overline{\Delta ILD}$, $\sigma_{\Delta ILD}$, $\Delta ILD_{max} - \Delta ILD_{min}$
m_V	Mass contained in volume V_{gas} (kg)
m	Single-channel time domain sound source
N_{iter}	Number of iterations in a FM-BEM calculation
n''	Outward normal direction
n_{mol}	Amount of gas (mol)
$P_l(\cos\Gamma)$	Legendre polynomial of degree l
P_0	One standard atmospheric pressure (101,325 Pa)

P_w	Pressure magnitude of incident wave
p	Pressure; sound pressure (Pa)
p_o	Reference sound pressure value of 20 µPa
Q	Volume velocity ($\mathrm{m^3\,s^{-1}}$)
$\dot{Q}$	Partial differentiation of Q with respect to time
Q_0	Intensity of a point source
$\hat{R}$	Molar-weight-specific gas constant
$\hat{R}_{air}$	Molar-weight-specific gas constant for dry air ($287\,\mathrm{J\,kg\,m^{-1}\,K^{-1}}$)
$\bar{R}$	Ideal Gas Constant ($8.3145\,\mathrm{J\,K^{-1}\,mol^{-1}}$)
r	Radial distance (m)
S	Index of the SYMARE subject
T	Period of a wave (s)
T_C	Temperature (°C)
T_K	Absolute temperature (K)
W	Weber fraction
V	Volume within which the pressure is calculated using the BEM
V_{gas}	Volume of gas ($\mathrm{m^3}$)
$\mathbf{V_B}$	Barycentre of a triangle
$\mathbf{V_C}$	Face centroid
$\mathbf{V_{L1}}$, $\mathbf{V_{L2}}$	Coordinates defining a line
$\mathbf{V_{P1}}$, $\mathbf{V_{P2}}$, $\mathbf{V_{P3}}$	Coordinates defining an infinite plane
$\mathbf{V_Q}$	Point of interest within a triangular face
$\mathbf{V_1}$, $\mathbf{V_2}$,$\mathbf{V_3}$	Coordinates of a triangular face
v_p	Particle velocity ($\mathrm{m\,s^{-1}}$)
$v_{n'}$	Normal velocity of the medium
Y	Acoustic admittance (S)
z_0	Characteristic impedance ($\mathrm{kg\,m^{-3}\,s^{-1}}$)

Sets:

$\mathbf{SP}^{\mathbf{NEECK}}_{\mathbf{sparse}}$	Source positions defined for acoustic measurement of HRTFs of the NEECK (Chapter 3)
$\mathbf{SP}^{\mathbf{NEECK}}_{\mathbf{dense}}$	Source positions defined for simulation of HRTFs of the NEECK (Chapter 3)
$\mathbf{SP}^{\mathbf{KEMAR}}$	Source positions defined for simulation of HRTFs of the half-torso mesh of KEMAR (Chapter 5)
$\mathbf{SP}^{\mathbf{KEMAR}}_{\mathbf{20}}$	Source positions within $\mathbf{SP}^{\mathbf{KEMAR}}$ which have NCS greater than 20 dB (Chapter 5)
$\mathbf{SP}^{\mathbf{KEMAR}}_{\mathbf{20_f}}$	Source positions within $\mathbf{SP}^{\mathbf{KEMAR}}$ which have NCS greater than 20 dB when calculated in frequency band f (Chapter 5)
$\mathbf{SP}^{\mathbf{SYMARE}}$	Source positions defined for simulation of HRTFs of the 15 SYMARE subjects (Chapter 7)
$\mathbf{LP}^{\mathbf{KEMAR}}$	Loudspeaker pairs formed from pairs of $\mathbf{SP}^{\mathbf{KEMAR}}$ positions mirrored in the x-axis (Chapter 6)
$\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{ITD}}$	Loudspeaker pairs within $\mathbf{LP}^{\mathbf{KEMAR}}$ which exhibit deterioration in ITD below the estimated threshold of perception (Chapter 6)
$\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{ILD}}$	Loudspeaker pairs within $\mathbf{LP}^{\mathbf{KEMAR}}$ which exhibit deterioration in ILD below the estimated threshold of perception (Chapter 6)
$\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}}$	Intersection of the $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{ITD}}$ and $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{ILD}}$ sets, therefore, loudspeaker pairs which exhibit deterioration in both ITD and ILD below the estimated threshold of perception (Chapter 6)
$\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{best}}$	Loudspeaker pairs within $\mathbf{LP}^{\mathbf{KEMAR}}_{\mathbf{CM}}$ which have an M_6 value of less than 0.3 (Chapter 6)
$\mathbf{LP}^{\mathbf{SYMARE}}$	Loudspeaker pairs formed from pairs of $\mathbf{SP}^{\mathbf{SYMARE}}$ positions mirrored in the x-axis (Chapter 7)
$\mathbf{LP}^{\mathbf{S}}_{\mathbf{ITD}}$	Loudspeaker pairs within $\mathbf{LP}^{\mathbf{SYMARE}}$ which exhibit deterioration in ITD below the estimated threshold of perception for SYMARE subject S (Chapter 7)
$\mathbf{LP}^{\mathbf{S}}_{\mathbf{ILD}}$	Loudspeaker pairs within $\mathbf{LP}^{\mathbf{SYMARE}}$ which exhibit deterioration in ILD below the estimated threshold of perception for SYMARE subject S (Chapter 7)
$\mathbf{LP}^{\mathbf{S}}_{\mathbf{CM}}$	Intersection of the $\mathbf{LP}^{\mathbf{S}}_{\mathbf{ITD}}$ and $\mathbf{LP}^{\mathbf{S}}_{\mathbf{ILD}}$ sets, therefore, loudspeaker pairs which exhibit deterioration in both ITD and ILD below the estimated threshold of perception for SYMARE subject S (Chapter 7)
$\mathbf{LP}^{\mathbf{all}}_{\mathbf{CM}}$	Intersection of all $\mathbf{LP}^{\mathbf{S}}_{\mathbf{CM}}$ sets, therefore, loudspeaker pairs which exhibit deterioration in both ITD and ILD below the estimated threshold of perception for all SYMARE subjects (Chapter 7)

$\mathbf{LP^{KEMAR}_{CM_6}}$ Loudspeaker pairs within $\mathbf{SP^{KEMAR}}$ which exhibit deterioration in both ITD and ILD below the estimated threshold of perception, calculated using the six-level perceptual model (Chapter 8)

References

[1] F. Rumsey, "Spatial Audio Psychoacoustics," in *Spatial Audio*, 1st ed., Focal Press, 2001, ch. 2, pp. 21–51, ISBN: 978-0-24-051623-3.

[2] D. Fisher, *1880-1884*, 2008. [Online]. Available: http://www.terramedia.co.uk/Chronomedia/years/1880-1884.htm (Accessed Jul. 29, 2019).

[3] S. Paul, "Binaural Recording Technology: A Historical Review and Possible Future Developments," *Acta Acustica united with Acustica*, vol. 95, no. 5, pp. 767–788, 2009. DOI: 10.3813/AAA.918208.

[4] M. Perry, *The History (and Future) of 3D Audio*, 2016. [Online]. Available: https://hookeaudio.com/blog/binaural-3d-audio/history-and-future-3d-audio/ (Accessed Jul. 29, 2019).

[5] A. Kall, *A Brief History of Binaural Audio*, 2011. [Online]. Available: http://www.kallbinauralaudio.com/a-brief-history-of-binaural-audio/ (Accessed Jul. 29, 2019).

[6] S. Krebs, *Dummy Head research in Göttingen and Oldenburg*, 2016. [Online]. Available: https://binauralrecording.wordpress.com/2016/04/05/dummy-head-research-in-gottingen-and-oldenburg/ (Accessed Jul. 29, 2019).

[7] Discogs, *Audiostax*. [Online]. Available: https://www.discogs.com/label/155436-Audiostax (Accessed Jul. 29, 2019).

[8] BBC R&D, *Binaural Sound*. [Online]. Available: https://www.bbc.co.uk/rd/projects/binaural-broadcasting (Accessed Jul. 29, 2019).

[9] J. Ramos, *Hellblade is an audio nightmare*, 2017. [Online]. Available: https://www.polygon.com/2017/8/9/16120082/hellblade-binaural-audio-psychosis (Accessed Jul. 29, 2019).

[10] S. Spagnol, G. Wersényi, M. Bujacz, O. Bălan, M. Herrera Martínez, A. Moldoveanu, and R. Unnthorsson, "Current Use and Future Perspectives of Spatial Audio Technologies in Electronic Travel Aids," *Wireless Communications and Mobile Computing*, vol. 2018, 2018. DOI: 10.1155/2018/3918284.

[11] BBC Media Applications Technologies Limited, *Home - A VR Spacewalk*, 2017. [Online]. Available: https://store.steampowered.com/app/512270/Home__A_VR_Spacewalk/ (Accessed Jul. 29, 2019).

[12] G. De Gasperis, A. Cordisco, and F. Cucchiara, "Immersive Virtual Reality As A Resource For Unaccessible Heritage Sites," *IOP Conference Series: Materials Science and Engineering*, vol. 364, no. 1, 2018. DOI: `10.1088/1757-899x/364/1/012035`.

[13] D. Henderson, *VR starts to prove its worth as a training tool*, 2019. [Online]. Available: `https://www.technative.io/vr-starts-to-prove-its-worth-as-a-training-tool/` (Accessed Jul. 29, 2019).

[14] A. H. Moore, A. I. Tew, and R. Nicol, "Headphone Transparification: A Novel Method for Investigating the Externalization of Binaural Sounds," in *AES 123rd Convention*, 2007. [Online]. Available: `http://www.aes.org/e-lib/browse.cfm?elib=14224` (Accessed Jul. 17, 2019).

[15] P. A. Nelson and J. F. W. Rose, "The time domain response of some systems for sound reproduction," *Journal of Sound and Vibration*, vol. 296, no. 3, pp. 461–493, 2006. DOI: `10.1016/j.jsv.2005.12.051`.

[16] E. Y. Choueiri, "Optimal Crosstalk Cancellation for Binaural Audio with Two Loudspeakers," *Princeton University*, 2014. [Online]. Available: `https://www.princeton.edu/3D3A/Publications/BACCHPaperV4d.pdf` (Accessed May 10, 2019).

[17] J. Ahrens, M. R. P. Thomas, and I. Tashev, "Gentle acoustic crosstalk cancelation using the spectral division method and Ambiophonics," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2013. DOI: `10.1109/WASPAA.2013.6701827`.

[18] B. Masiero and J. Fels, "Perceptually robust headphone equalization for binaural reproduction," in *AES 130th Convention*, 2011. [Online]. Available: `http://www.aes.org/e-lib/browse.cfm?elib=15855` (Accessed Jun. 12, 2019).

[19] A. Lundkvist, A. Nykänen, and R. Johnsson, "3D-Sound in Car Compartments Based on Loudspeaker Reproduction Using Crosstalk Cancellation," in *AES 130th Convention*, 2011. [Online]. Available: `http://www.aes.org/e-lib/online/browse.cfm?elib=15802` (Accessed May 25, 2019).

[20] O. Puomio, R. Gonzalez Diaz, and T. Lokki, "Augmented Audio with Behind-the-Head Listening Device," in *AES International Conference on Immersive and Interactive Audio*, 2019. [Online]. Available: `http://www.aes.org/e-lib/browse.cfm?elib=20420` (Accessed May 14, 2019).

[21] F. Orduña-Bustamante, J. J. López, and A. González, "Prediction and measurement of acoustic crosstalk cancellation robustness," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2001. DOI: `10.1109/ICASSP.2001.940376`.

[22] M. Jones, S. J. Elliott, T. Takeuchi, and J. Beer, "Virtual audio reproduced in a headrest," in *International Congress on Acoustics*, 2007. [Online]. Available: `http://www.sea-acustica.es/WEB_ICA_07/fchrs/papers/ppa-08-011.pdf` (Accessed Jun. 12, 2019).

[23] S. Elliott, C. House, J. Cheer, and M. Simon-Galvez, "Crosstalk Cancellation and Equalisation for Headrest Sound Reproduction," in *AES International Conference on Sound Field Control*, 2016. [Online]. Available: `http://www.aes.org/e-lib/browse.cfm?elib=18304` (Accessed Jun. 12, 2019).

[24] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF Database," in *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, 2001. DOI: `10.1109/ASPAA.2001.969552`.

[25] C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 5, pp. 476–488, 1998. DOI: `10.1109/89.709673`.

[26] M. Vorländer, "Fundamentals of Acoustics," in *Auralization*, 1st ed., Springer-Verlag Berlin Heidelberg, 2008, ch. 1, pp. 7–22, ISBN: 978-3-540-48830-9.

[27] H. Kuttruff, "Free Sound Propagation in Gases," in *Room Acoustics*, Applied Science Publishers Ltd, 1973, ch. 1, pp. 1–20, ISBN: 978-0-85-334573-2.

[28] F. A. Everest, "Sound Waves in the Free Field," in *Master Handbook of Acoustics*, 4th ed., McGraw-Hill, 2001, ch. 4, pp. 83–89, ISBN: 978-0-07-136097-5.

[29] D. Howard and J. Angus, "Introduction to hearing," in *Acoustics and Psychoacoustics*, 2nd ed., Focal Press, 2001, ch. 2, pp. 65–108, ISBN: 978-0-24-051609-7.

[30] D. Howard and J. Angus, "Introduction to sound," in *Acoustics and Psychoacoustics*, 2nd ed., Focal Press, 2001, ch. 1, pp. 1–64, ISBN: 978-0-24-051609-7.

[31] International Electrotechnical Commission, *Acoustics and electroacoustics / General terms / 801-21-20*. [Online]. Available: `http://www.electropedia.org/iev/iev.nsf/display?openform&ievref=801-21-20` (Accessed Apr. 26, 2019).

[32] W. A. Yost, "Auditory Sensitivity," in *Fundamentals of Hearing: An Introduction*, 4th ed., Academic Press, 2000, ch. 10, pp. 149–164, ISBN: 978-0-12-775695-0.

[33] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, "Appendix 9: A little thermodynamics and the perfect gas," in *Fundamentals of Acoustics*, 4th, Wiley, 2000, ch. A9, pp. 522–525.

[34] International Electrotechnical Commission, *Acoustics and electroacoustics / General terms / 801-22-07*. [Online]. Available: `http://www.electropedia.org/iev/iev.nsf/display?openform&ievref=801-22-07` (Accessed Apr. 26, 2019).

[35] H. Kuttruff, "Some facts on sound waves, sources and hearing," in *Room Acoustics*, 5th ed., CRC Press, 2009, ch. 1, pp. 7–34, ISBN: 978-0-41-548021-5.

[36] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1465–1479, 1999. DOI: `10.1121/1.427180`.

[37] F. A. Everest, "Reflection of Sound," in *Master Handbook of Acoustics*, 4th ed., McGraw-Hill, 2001, ch. 10, pp. 235–244, ISBN: 978-0-07-136097-5.

[38] S. B. Shelley, “Diffuse Boundary Modelling in the Digital Waveguide Mesh,” PhD, Department of Electronics, University of York, 2007. [Online]. Available: `http://www-users.york.ac.uk/~dtm3/Download/SBS_Thesis.pdf` (Accessed Sep. 2, 2019).

[39] M. Vorländer, “Sound Propagation,” in *Auralization*, 1st ed., Springer-Verlag Berlin Heidelberg, 2008, ch. 3, pp. 35–52, ISBN: 978-3-540-48830-9.

[40] V. Pulkki, “Sound, Hearing and Perception,” in *Sensory Evaluation of Sound*, N. Zacharov, Ed., Taylor & Francis Group, 2019, ch. 3, pp. 21–58, ISBN: 978-1-49-875136-0.

[41] J. E. Hawkins, “Human ear,” in *Encyclopædia Britannica*, Encyclopædia Britannica Inc., 2018. [Online]. Available: `https://www.britannica.com/science/ear/Transmission-of-sound-waves-through-the-outer-and-middle-ear` (Accessed Apr. 26, 2019).

[42] H. Fletcher, “Auditory patterns,” *Reviews of Modern Physics*, vol. 12, no. 1, pp. 47–65, 1940. DOI: `10.1103/revmodphys.12.47`.

[43] C. J. Plack, “Loudness and Intensity Coding,” in *The Sense of Hearing*, 2nd ed., Routledge, 2005, ch. 6, pp. 100–117, ISBN: 978-1-84-872515-7.

[44] W. A. Yost, “Loudness and Pitch,” in *Fundamentals of Hearing: An Introduction*, 4th ed., Academic Press, 2000, ch. 13, pp. 193–208, ISBN: 978-0-12-775695-0.

[45] B. Scharf, “Critical Bands,” in *Foundations of Modern Auditory Theory*, J. V. Tobias, Ed., Elsevier Inc., 1970, ch. 5, pp. 157–202, ISBN: 978-0-12-691901-1.

[46] International Organization for Standardization, *ISO 226: 2003 - Acoustics – Normal equal-loudness-level contours*, 2003. [Online]. Available: `https://www.iso.org/standard/34222.html` (Accessed Jun. 10, 2019).

[47] S. Buus, M. Florentine, and C. R. Mason, “Tuning Curves at High Frequencies and their Relation to the Absolute Threshold Curve,” in *Auditory Frequency Selectivity*, B. C. J. Moore and R. D. Patterson, Eds., Springer US, 1986, pp. 341–350, ISBN: 978-1-4613-2247-4. DOI: `10.1007/978-1-4613-2247-4_37`.

[48] I. Yasin and C. J. Plack, “Psychophysical tuning curves at very high frequencies,” *The Journal of the Acoustical Society of America*, vol. 118, no. 4, pp. 2498–2506, 2005. DOI: `10.1121/1.2035594`.

[49] K. Ashihara, “Hearing thresholds for pure tones above 16 kHz,” *The Journal of the Acoustical Society of America*, vol. 122, no. 3, EL52–EL57, 2007. DOI: `10.1121/1.2761883`.

[50] B. C. J. Moore, “Pitch Perception,” in *An Introduction to Psychology of Hearing*, 6th ed., Emerald, 2012, ch. 6, pp. 203–244, ISBN: 978-1-78052-038-4.

[51] E. Zwicker, “Scaling,” in *Handbook of Sensory Physiology Vol. 5/2: Auditory System*, W. D. Keidel and W. D. Neff, Eds., Springer-Verlag Berlin Heidelberg, 1975, ch. 9, pp. 401–448, ISBN: 3-540-07000-1.

[52] C. J. Plack and R. P. Carlyon, “Loudness Perception and Intensity Coding,” in *Hearing*, B. C. J. Moore, Ed., Academic Press, 1995, ch. 4, pp. 123–161, ISBN: 0-12-505626-5.

[53] A. J. M. Houtsma, "Pitch Perception," in *Hearing*, B. C. J. Moore, Ed., Academic Press, 1995, ch. 8, pp. 267–297, ISBN: 0-12-505626-5.

[54] C. C. Wier, W. Jesteadt, and D. M. Green, "Frequency discrimination as a function of frequency and sensation level," *The Journal of the Acoustical Society of America*, vol. 61, no. 1, pp. 178–184, 1977. DOI: 10.1121/1.381251.

[55] B. C. J. Moore, "Frequency difference limens for short-duration tones," *The Journal of the Acoustical Society of America*, vol. 54, no. 3, pp. 610–619, 1973. DOI: 10.1121/1.1913640.

[56] W. Jesteadt, C. C. Wier, and D. M. Green, "Intensity discrimination as a function of frequency and sensation level," *The Journal of the Acoustical Society of America*, vol. 61, no. 1, pp. 169–177, 1977. DOI: 10.1121/1.381278.

[57] C. Jin, A. Corderoy, S. Carlile, and A. van Schaik, "Contrasting monaural and interaural spectral cues for human sound localization," *The Journal of the Acoustical Society of America*, vol. 115, no. 6, pp. 3124–3141, 2004. DOI: 10.1121/1.1736649.

[58] C. L. Searle, L. D. Braida, D. R. Cuddy, and M. F. Davis, "Binaural pinna disparity: another auditory localization cue," *The Journal of the Acoustical Society of America*, vol. 57, no. 2, pp. 448–455, 1975. DOI: 10.1121/1.380442.

[59] B. C. J. Moore, "Space Perception," in *An Introduction to Psychology of Hearing*, 6th ed., Emerald, 2012, ch. 7, pp. 245–281, ISBN: 978-1-78052-038-4.

[60] R. S. Woodworth, *Experimental psychology*. Holt, Rinehart and Winston, 1954, ISBN: 0030074401.

[61] R. S. Woodworth and H. Schlosberg, *Experimental psychology (Revised Edition)*. Holt, Rinehart and Winston, 1954.

[62] R. V. L. Hartley and T. C. Fry, "The Binaural Location of Pure Tones," *Physical Review*, vol. 18, no. 6, pp. 431–442, 1921. DOI: 10.1103/PhysRev.18.431.

[63] J. Blauert, "Spatial Hearing with One Sound Source," in *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, 1997, ch. 2, pp. 36–200, ISBN: 978-0-58-523825-8.

[64] B. Xie, "Spatial Hearing and Virtual Auditory Display," in *Head Related Transfer Function and Virtual Auditory Display*, 2nd ed., J. Ross Publishing, 2013, ch. 1, pp. 1–42, ISBN: 978-1-60-427741-8.

[65] A. Andreopoulou and B. F. G. Katz, "Identification of perceptually relevant methods of inter-aural time difference estimation," *The Journal of the Acoustical Society of America*, vol. 142, no. 2, pp. 588–598, 2017. DOI: 10.1121/1.4996457.

[66] G. F. Kuhn, "Model for the interaural time differences in the azimuthal plane," *The Journal of the Acoustical Society of America*, vol. 62, no. 1, pp. 157–167, 1977. DOI: 10.1121/1.381498.

[67] V. R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a spherical-head model from anthropometry," *Journal of the Audio Engineering Society*, vol. 49, no. 6, pp. 472–479, 2001. [Online]. Available: `http://www.aes.org/e-lib/browse.cfm?elib=10188` (Accessed May 6, 2019).

[68] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction.," *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1637–1647, 1992. DOI: `10.1121/1.402444`.

[69] J. C. Middlebrooks and D. M. Green, "Directional Dependence of Interaural Envelope Delays," *The Journal of the Acoustical Society of America*, vol. 87, no. 5, pp. 2149–2162, 1990. DOI: `10.1121/1.399183`.

[70] J.-M. Jot, V. Larcher, and O. Warusfel, "Digital signal processing issues in the context of binaural and transaural stereophony," in *AES 98th Convention*, 1995. [Online]. Available: `http://www.aes.org/e-lib/browse.cfm?elib=7786` (Accessed Apr. 2, 2019).

[71] P. Minnaar, J. Plogsties, S. K. Olesen, F. Christensen, and H. Møller, "The Interaural Time Difference in Binaural Synthesis," in *AES 108th Convention*, 2000. [Online]. Available: `http://www.aes.org/e-lib/browse.cfm?elib=9205` (Accessed Jun. 22, 2019).

[72] S. Kuwada, R. Batra, and D. C. Fitzpatrick, "Neural Processing of Binaural Temporal Cues," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. Gilkey and T. R. Anderson, Eds., Mahwah, N.J. : Lawrence Erlbaum Associates, 1997, ch. 20, pp. 399–45, ISBN: 978-0-80-581654-9. [Online]. Available: `https://www.taylorfrancis.com/books/9781315806341/chapters/10.4324/9781315806341-31` (Accessed Apr. 3, 2019).

[73] B. F. G. Katz and M. Noisternig, "A comparative study of interaural time delay estimation methods," *The Journal of the Acoustical Society of America*, vol. 135, no. 6, pp. 3530–3540, 2014. DOI: `10.1121/1.4875714`.

[74] C. Armstrong, L. Thresh, D. Murphy, and G. Kearney, "A Perceptual Evaluation of Individual and Non-Individual HRTFs: A Case Study of the SADIE II Database," *Applied Sciences*, vol. 8, no. 11, 2018. DOI: `10.3390/app8112029`.

[75] B. Xie, "Primary Features of HRTFs," in *Head Related Transfer Function and Virtual Auditory Display*, 2nd ed., J. Ross Publishing, 2013, ch. 3, pp. 73–108, ISBN: 978-1-60-427741-8.

[76] H. Wallach, "On Sound Localization," *The Journal of the Acoustical Society of America*, vol. 10, no. 1, pp. 270–274, 1939. DOI: `10.1121/1.1902063`.

[77] F. L. Wightman and D. J. Kistler, "Resolution of front-back ambiguity in spatial hearing by listener and source movement," *The Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2841–2853, 1999. DOI: `10.1121/1.426899`.

[78] Lord Rayleigh, "XII. On our perception of sound direction," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 13, no. 74, pp. 214–232, 1907. DOI: `10.1080/14786440709463595`.

[79] V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," *The Journal of the Acoustical Society of America*, vol. 109, no. 3, pp. 1110–1122, 2001. DOI: 10.1121/1.1349185.

[80] C. Jin, P. Guillon, N. Epain, R. Zolfaghari, A. van Schaik, A. I. Tew, C. T. Hetherington, and J. B. Thorpe, "Creating the Sydney York Morphological and Acoustic Recordings of Ears Database," *IEEE Transactions on Multimedia*, vol. 16, no. 1, pp. 37–46, 2014. DOI: 10.1109/TMM.2013.2282134.

[81] V. R. Algazi, R. O. Duda, R. P. Morrison, and D. M. Thompson, "Structural composition and decomposition of HRTFs," in *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, 2001. DOI: 10.1109/ASPAA.2001.969553.

[82] S. G. Rodriguez and M. A. Ramirez, "Extracting and Modeling Approximated Pinna-Related Transfer Functions from HRTF Data," in *International Conference on Auditory Display*, 2005. [Online]. Available: http://www.icad.org/Proceedings/2005/RodriguezArjona2005WRONGPAPER.pdf (Accessed Jun. 13, 2019).

[83] V. R. Algazi, R. O. Duda, and D. M. Thompson, "The Use of Head-and-Torso Models for Improved Spatial Sound Synthesis," in *AES 113rd Convention*, 2002. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=11294 (Accessed Jun. 18, 2019).

[84] H. Takemoto, P. Mokhtari, H. Kato, R. Nishimura, and K. Iida, "Mechanism for generating peaks and notches of head-related transfer functions in the median plane," *The Journal of the Acoustical Society of America*, vol. 132, no. 6, pp. 3832–3841, 2012. DOI: 10.1121/1.4765083.

[85] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Visualization of acoustic pressure and velocity patterns with phase information in the pinna cavities at normal modes," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2014. DOI: 10.1109/ICASSP.2014.6855203.

[86] J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *The Journal of the Acoustical Society of America*, vol. 56, no. 6, pp. 1829–1834, 1974. DOI: 10.1121/1.1903520.

[87] K. Iida, M. Itoh, A. Itagaki, and M. Morimoto, "Median plane localization using a parametric model of the head-related transfer function based on spectral cues," *Applied Acoustics*, vol. 68, no. 8, pp. 835–850, 2007. DOI: 10.1016/j.apacoust.2006.07.016.

[88] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Pinna sensitivity patterns reveal reflecting and diffracting surfaces that generate the first spectral notch in the front median plane," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011. DOI: 10.1109/ICASSP.2011.5946969.

[89] D. R. Begault, “Challenges to the Successful Implementation of 3-D Sound,” *Journal of the Audio Engineering Society*, vol. 39, no. 11, pp. 864–870, 1991. [Online]. Available: `http://www.aes.org/e-lib/browse.cfm?elib=10281` (Accessed Jun. 12, 2019).

[90] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, “Head-Related Transfer Functions of Human Subjects,” *Journal of the Audio Engineering Society*, vol. 43, no. 5, pp. 300–321, 1995. [Online]. Available: `http://www.aes.org/e-lib/online/browse.cfm?elib=7949` (Accessed Jun. 14, 2019).

[91] B. Xie, “Binaural Reproduction through Loudspeakers,” in *Head Related Transfer Function and Virtual Auditory Display*, 2nd ed., J. Ross Publishing, 2013, ch. 9, pp. 283–326, ISBN: 978-1-60-427741-8.

[92] I. Nambu, M. Washizu, S. Morioka, Y. Hasegawa, W. Sakuma, S. Yano, H. Hokari, and Y. Wada, “Reinforcement-Learning-Based Personalization of Head-Related Transfer Functions,” *Journal of the Audio Engineering Society*, vol. 66, no. 5, pp. 317–328, 2018. DOI: `10.17743/jaes.2018.0014`.

[93] G. Andéol, S. Savel, and A. Guillaume, “Perceptual factors contribute more than acoustic factors to sound localization abilities with virtual sources,” *Frontiers in Neuroscience*, vol. 8, 2014. DOI: `10.3389/fnins.2014.00451`.

[94] F. Klein and S. Werner, “Auditory Adaptation to Non-Individual HRTF Cues in Binaural Audio Reproduction,” *Journal of the Audio Engineering Society*, vol. 64, no. 1/2, pp. 45–54, 2016. DOI: `10.17743/jaes.2015.0092`.

[95] P. Stitt, L. Picinali, and B. F. G. Katz, “Auditory Accommodation to Poorly Matched Non-Individual Spectral Localization Cues Through Active Learning,” *Nature Scientific Reports*, vol. 9, no. 1, 2019. DOI: `10.1038/s41598-018-37873-0`.

[96] Neumann, *Dummy Head: KU100*. [Online]. Available: `https://en-de.neumann.com/ku-100` (Accessed Feb. 7, 2019).

[97] GRAS Sound and Vibration, *Head & Torso simulators*. [Online]. Available: `http://www.gras.dk/products/head-torso-simulators-kemar` (Accessed Aug. 31, 2016).

[98] M. D. Burkhard and R. M. Sachs, “Anthropometric Manikin for Acoustic Research,” *The Journal of the Acoustical Society of America*, vol. 58, no. 1, pp. 214–222, 1975. DOI: `10.1121/1.380648`.

[99] K. Young, A. I. Tew, and G. Kearney, “Boundary Element Method Modelling of KEMAR for Binaural Rendering: Mesh Production and Validation,” in *Interactive Audio Systems Symposium*, 2016. [Online]. Available: `http://www.york.ac.uk/sadie-project//IASS2016/IASS_Papers/IASS_2016_paper_4.pdf` (Accessed Jan. 8, 2017).

[100] A. J. Kolarik, B. C. J. Moore, P. Zahorik, S. Cirstea, and S. Pardhan, “Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss,” *Attention, Perception, and Psychophysics*, vol. 78, no. 2, pp. 373–395, 2016. DOI: `10.3758/s13414-015-1015-1`.

[101] A. Kan, C. Jin, and A. van Schaik, "A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function," *The Journal of the Acoustical Society of America*, vol. 125, no. 4, pp. 2233–2242, 2009. DOI: 10.1121/1.3081395.

[102] D. S. Brungart, "Near-Field Virtual Audio Displays," *Presence: Teleoperators and Virtual Environments*, vol. 11, no. 1, pp. 93–106, 2002. DOI: 10.1162/105474602317343686.

[103] K. Young, C. Armstrong, A. I. Tew, D. T. Murphy, and G. Kearney, "A Numerical Study into Perceptually-Weighted Spectral Differences between Differently-Spaced HRTFs," in *AES International Conference on Immersive and Interactive Audio*, 2019. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=20386 (Accessed Jun. 12, 2019).

[104] G.-Z. Yu, D. Rao, and B. Xie, "Characteristics of Near-Field Head-Related Transfer Function for KEMAR," in *AES 40th International Conference: Spatial Audio - Sense the Sound of Space*, 2010. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15545 (Accessed May 11, 2019).

[105] S. Spagnol, "On distance dependence of pinna spectral patterns in head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 137, no. 1, EL58–EL64, 2015. DOI: 10.1121/1.4903919.

[106] M. Otani, T. Hirahara, and S. Ise, "Numerical study on source-distance dependency of head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 125, no. 5, pp. 3253–3261, 2009. DOI: 10.1121/1.3111860.

[107] D. S. Brungart, "Auditory parallax effects in the HRTF for nearby sources," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1999. DOI: 10.1109/ASPAA.1999.810877.

[108] T. Letowski and S. Letowski, "Localization Error: Accuracy and Precision of Auditory Localization," in *Advances in Sound Localization*, P. Strumillo, Ed., InTech, 2012, ch. 4, pp. 55–78, ISBN: 978-953-307-224-1.

[109] J. M. Moore, D. J. Tollin, and T. C. T. Yin, "Can measures of sound localization acuity be related to the precision of absolute location estimates?" *Hearing Research*, vol. 238, no. 1-2, pp. 94–109, 2008. DOI: 10.1016/j.heares.2007.11.006.

[110] A. W. Mills, "On the Minimum Audible Angle," *The Journal of the Acoustical Society of America*, vol. 30, no. 4, pp. 237–246, 1958. DOI: 10.1121/1.1909553.

[111] J. C. Makous and J. C. Middlebrooks, "Two-dimensional sound localization by human listeners," *The Journal of the Acoustical Society of America*, vol. 87, no. 5, pp. 2188–2200, 1990. DOI: 10.1121/1.399186.

[112] R. Preibisch-Effenberger, "Die Schallokalisationsfähigkeit des Menschen und ihre audiometrische Verwendbarkeit zur klinischen Diagnostik [The human faculty of soubd localization ability and its audiometric application to clinical diagnostics]," Dissertation, Technische Universität Dresden, 1966.

[113] B. G. Haustein and W. Schirmer, “Messeinrichtung zur untersuchung des richtungslokalisationsvermoegens [A measuring apparatus for the investigation of the faculty of directional localization],” *Hochfrequenztechnik und Elektroakustik*, vol. 79, pp. 96–101, 1970.

[114] S. R. Oldfield and S. P. A. Parker, “Acuity of Sound Localisation: A Topography of Auditory Space. I. Normal Hearing Conditions,” *Perception*, vol. 13, no. 5, pp. 581–600, 1984. DOI: 10.1068/p130581.

[115] M. Aussal, F. Alouges, and B. F. G. Katz, “ITD Interpolation and Personalization for Binaural Synthesis Using Spherical Harmonics,” in *AES UK Conference: Spatial Audio in Today’s 3D World*, 2012. [Online]. Available: http://www.cmapx.polytechnique.fr/~aussal/publis/AES2012_ITSpher.pdf (Accessed Jul. 2, 2019).

[116] L. S. R. Simon, A. Andreopoulou, and B. F. G. Katz, “Investigation of perceptual interaural time difference evaluation protocols in a binaural context,” *Acta Acustica united with Acustica*, vol. 102, no. 1, pp. 129–140, 2016. DOI: 10.3813/AAA.918930.

[117] R. G. Klumpp and H. R. Eady, “Some Measurements of Interaural Time Difference Thresholds,” *The Journal of the Acoustical Society of America*, vol. 28, no. 5, pp. 859–860, 1956. DOI: 10.1121/1.1908493.

[118] A. W. Mills, “Lateralization of High Frequency Tones,” *The Journal of the Acoustical Society of America*, vol. 32, no. 1, pp. 132–134, 1960. DOI: 10.1121/1.1907864.

[119] W. A. Yost and R. H. Dye, “Discrimination of interaural differences of level as a function of frequency,” *The Journal of the Acoustical Society of America*, vol. 83, no. 5, pp. 1846–1851, 1988. DOI: 10.1121/1.396520.

[120] W. Tu, R. Hu, H. Wang, and W. Chen, “Measurement and analysis of just noticeable difference of interaural level difference cue,” in *International Conference on Multimedia Technology*, 2010. DOI: 10.1109/ICMULT.2010.5630980.

[121] B. Xie, “HRTF Measurements,” in *Head Related Transfer Function and Virtual Auditory Display*, 2nd ed., J. Ross Publishing, 2013, ch. 2, pp. 43–73, ISBN: 978-1-60-427741-8.

[122] K. A. Stroud, “Introduction to the Fourier Transform,” in *Advanced Engineering Mathematics*, 4th ed., Palgrave Macmillan, 2003, ch. 7, pp. 231–271, ISBN: 978-1-40-390312-9.

[123] G. Kearney and T. Doyle, “A HRTF database for virtual loudspeaker rendering,” in *AES 139th Convention*, 2015. [Online]. Available: http://www.aes.org/e-lib/online/browse.cfm?elib=17980 (Accessed Jun. 15, 2019).

[124] C. Armstrong, A. Chadwick, L. Thresh, M. Damian, and G. Kearney, “Simultaneous HRTF Measurement of Multiple Source Configurations Utilizing Semi-Permanent Structural Mounts,” in *AES 143rd Convention*, 2017. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=19311 (Accessed May 10, 2019).

[125] D. N. Zotkin, R. Duraiswami, E. Grassi, and N. A. Gumerov, "Fast head-related transfer function measurement via reciprocity," *The Journal of the Acoustical Society of America*, vol. 120, no. 4, pp. 2202–2215, 2006. DOI: 10.1121/1.2207578.

[126] N. Matsunaga and T. Hirahara, "Fast near-field HRTF measurements using reciprocal method," in *International Congress on Acoustics*, 2010. [Online]. Available: http://www.acoustics.asn.au/conference_proceedings/ICA2010/cdrom-ICA2010/papers/p480.pdf (Accessed Jun. 2, 2019).

[127] D. Hammershøi and H. Møller, "Sound transmission to and within the human ear canal," *The Journal of the Acoustical Society of America*, vol. 100, no. 1, pp. 408–427, 1996. DOI: 10.1121/1.415856.

[128] B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," *The Journal of the Acoustical Society of America*, vol. 110, no. 5, pp. 2440–2448, 2001. DOI: 10.1121/1.1412440.

[129] H. Møller, "Fundamentals of binaural technology," *Applied Acoustics*, vol. 36, no. 3-4, pp. 171–218, 1992. DOI: 10.1016/0003-682X(92)90046-U.

[130] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. I: Stimulus synthesis," *The Journal of the Acoustical Society of America*, vol. 85, no. 2, pp. 858–867, 1989. DOI: 10.1121/1.397557.

[131] W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," *The Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3907–3908, 1995. DOI: 10.1121/1.412407.

[132] K. A. J. Riederer, "Repeatability analysis of head-related transfer function measurements," in *AES 105th Convention*, 1998. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=8334 (Accessed Jun. 12, 2019).

[133] J. Blauert, M. Brueggen, A. W. Bronkhorst, R. Drullman, G. Reynaud, L. Pellieux, W. Krebber, and R. Sottek, "The AUDIS catalog of human HRTFs," *The Journal of the Acoustical Society of America*, vol. 103, no. 5, pp. 3082–3082, 1998. DOI: 10.1121/1.422910.

[134] B. P. Bovbjerg, F. Christensen, P. Minnaar, and X. Chen, "Measuring the head-related transfer functions of an artificial head with a high directional resolution," in *AES 109th Convention*, 2000. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=9074 (Accessed Jun. 22, 2019).

[135] T. Nishino, N. Inoue, K. Takeda, and F. Itakura, "Estimation of HRTFs on the horizontal plane using physical features," *Applied Acoustics*, vol. 68, no. 8, pp. 897–908, 2007. DOI: 10.1016/j.apacoust.2006.12.010.

[136] S. Takane, D. Arai, T. Miyajima, K. Watanabe, Y. Suzuki, and T. Sone, "A database of Head-Related Transfer Functions in whole directions on upper hemisphere," *Acoustical Science and Technology*, vol. 23, no. 3, pp. 160–162, 2002. DOI: 10.1250/ast.23.160.

[137] IRCAM, *Listen HRTF Database*, 2003. [Online]. Available: http://recherche.ircam.fr/equipes/salles/listen/index.html (Accessed Apr. 21, 2016).

[138] E. Grassi, J. Tulsi, and S. Shamma, "Measurement of head-related transfer functions based on the empirical transfer function estimate," in *International Conference on Auditory Display*, 2003. [Online]. Available: http://www.icad.org/websiteV2.0/Conferences/ICAD2003/paper/29Grassi.pdf (Accessed Jun. 21, 2019).

[139] B. Xie, X. Zhong, D. Rao, and Z. Liang, "Head-related transfer function database and its analyses," *Science in China Series G: Physics, Mechanics and Astronomy*, vol. 50, no. 3, pp. 267–280, 2007. DOI: 10.1007/s11433-007-0018-x.

[140] P. Majdak, M. J. Goupell, and B. Laback, "3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training," *Attention, Perception, & Psychophysics*, vol. 72, no. 2, pp. 454–469, 2010. DOI: 10.3758/APP.72.2.454.

[141] G. Yu, Y. Liu, B. Xie, and Q. Zhong, "Fast measurement system and super high directional resolution head-related transfer function database," *The Journal of the Acoustical Society of America*, vol. 131, no. 4, 2012. DOI: 10.1121/1.4708353.

[142] N. Gupta, A. Barreto, M. Joshi, and J. C. Agudelo, "HRTF database at FIU DSP Lab," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010. DOI: 10.1109/ICASSP.2010.5496084.

[143] B. Bernschütz, "A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100," in *AIA-DAGA*, 2013. [Online]. Available: http://www.audiogroup.web.fh-koeln.de/FILES/AIA-DAGA2013_HRIRs.pdf (Accessed Jun. 1, 2019).

[144] G. Kearney, M. Gorzel, H. Rice, and F. Boland, "Distance Perception in Interactive Virtual Acoustic Environments using First and Higher Order Ambisonic Sound Fields," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 61–71, 2012. DOI: 10.3813/AAA.918492.

[145] Z. Chen and D. Mao, "Near-Field Variation of Loudness with Distance," *Acoustics Australia*, vol. 47, pp. 175–184, 2019. DOI: 10.1007/s40857-019-00158-1.

[146] G. Yu, R. Wu, Y. Liu, and B. Xie, "Near-field head-related transfer-function measurement and database of human subjects," *The Journal of the Acoustical Society of America*, vol. 143, no. 3, EL194–EL198, 2018. DOI: 10.1121/1.5027019.

[147] T. Qu, Z. Xiao, M. Gong, Y. Huang, X. Li, and X. Wu, "Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1124–1132, 2009. DOI: 10.1109/TASL.2009.2020532.

[148] P. M. Morse and K. U. Ingard, *Theoretical Acoustics*. McGraw-Hill, 1968, ISBN: 0070433305.

[149] B. Xie, “Calculation of HRTFs,” in *Head Related Transfer Function and Virtual Auditory Display*, 2nd ed., J. Ross Publishing, 2013, ch. 4, pp. 109–134, ISBN: 978-1-60-427741-8.

[150] X.-L. Zhong, F.-c. Zhang, and B.-S. Xie, “On the spatial symmetry of head-related transfer functions,” *Applied Acoustics*, vol. 74, no. 6, pp. 856–864, 2013. DOI: 10.1016/j.apacoust.2013.01.004.

[151] H. Bahu and D. Romblom, “Optimization and prediction of the spherical and ellipsoidal ITD model parameters using offset ears,” in *AES International Conference on Spatial Reproduction - Aesthetics and Science*, 2018. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=19599 (Accessed Jun. 16, 2019).

[152] R. O. Duda and W. L. Martens, “Range dependence of the response of a spherical head model,” *The Journal of the Acoustical Society of America*, vol. 104, no. 5, pp. 3048–3058, 1998. DOI: 10.1121/1.423886.

[153] V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, and Z. Tang, “Approximating the head-related transfer function using simple geometric models of the head and torso,” *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2053–2064, 2002. DOI: 10.1121/1.1508780.

[154] B. Xie, “Multipole Re-expansions for Calculating the HRTFs of the Snowman Model,” in *Head Related Transfer Function and Virtual Auditory Display*, 2nd ed., J. Ross Publishing, 2013, ch. Appendix B, pp. 437–440, ISBN: 978-1-60-427741-8.

[155] N. A. Gumerov, R. Duraiswami, and Z. Tang, “Numerical study of the influence of the torso on the HRTF,” in *IEEE International Conference on Acoustics Speech, and Signal Processing*, 2002. DOI: 10.1109/icassp.2002.5745015.

[156] N. A. Gumerov and R. Duraiswami, “Multiple Scattering from N spheres using Multipole Reexpansion,” UMIACS-TR-2001-72, University of Maryland Institute for Advanced Computer Studies, Tech. Rep., 2001. [Online]. Available: https://drum.lib.umd.edu/handle/1903/1156 (Accessed Jun. 18, 2019).

[157] N. A. Gumerov and R. Duraiswami, “Fast, Exact, and Stable Computation of Multipole Translation and Rotation Coefficients for the 3-D Helmholtz Equation,” UMIACS-TR-2001-44, University of Maryland Institute for Advanced Computer Studies, Tech. Rep., 2001. [Online]. Available: http://drum.lib.umd.edu/handle/1903/1141 (Accessed Jun. 18, 2019).

[158] S. Kirkup, “Introduction,” in *The Boundary Element Method in Acoustics*, 2nd ed., Integrated Sound Software, 2007, ch. 1, pp. 1–20. [Online]. Available: https://www.researchgate.net/publication/261760562 (Accessed Nov. 15, 2016).

[159] S. Kirkup, “The Boundary Element Method in Acoustics: A Survey,” *Applied Sciences*, vol. 9, no. 8, 2019. DOI: 10.3390/app9081642.

[160] Y. Kahana and P. A. Nelson, “Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models,” *Journal of Sound and Vibration*, vol. 300, no. 3-5, pp. 552–579, 2007. DOI: 10.1016/j.jsv.2006.06.079.

[161] Y. Kahana, P. A. Nelson, M. Petyt, and S. Choi, “Numerical Modelling of the Transfer Functions of a Dummy-Head and of the External Ear,” in *AES 16th International Conference: Spatial Sound Reproduction*, 1999. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=8025 (Accessed Jun. 20, 2019).

[162] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, “Comparison of Simulated and Measured HRTFs: FDTD Simulation Using MRI Head Data,” in *AES 123rd Convention*, 2007. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14298 (Accessed Jun. 12, 2019).

[163] O. Kirkeby, E. T. Seppälä, A. Kärkkäinen, L. Kärkkäinen, and T. Huttunen, “Some Effects of The Torso on Head-Related Transfer Functions,” in *AES 122nd Convention*, 2007. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14015 (Accessed Jun. 20, 2019).

[164] J. Turku, M. Vilermo, E. Seppälä, A. Kärkkäinen, L. Kärkkäinen, O. Kirkeby, and M. Pölönen, “Perceptual evaluation of numerically simulated head-related transfer function,” in *AES 124th Convention*, 2008. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14619 (Accessed Jun. 20, 2019).

[165] H. Ziegelwanger, P. Majdak, and W. Kreuzer, “Non-Uniform Sampling of Geometry for the Numeric Simulation of Head-Related Transfer Functions,” in *International Congress on Sound and Vibration*, 2014. [Online]. Available: https://pdfs.semanticscholar.org/5252/903a45b95813990fbd0a7a8af9c87b82d386.pdf (Accessed Mar. 20, 2017).

[166] N. A. Gumerov, A. E. O’Donovan, R. Duraiswami, and D. N. Zotkin, “Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation,” *The Journal of the Acoustical Society of America*, vol. 127, no. 1, pp. 370–386, 2010. DOI: 10.1121/1.3257598.

[167] T. Huttunen, K. Tuppurainen, A. Vanne, P. Ylä-Oijala, S. Järvenpää, A. Kärkkäinen, and L. Kärkkäinen, “Simulation of the head-related transfer functions using cloud computing,” in *International Congress on Acoustics*, 2013. DOI: 10.1121/1.4800138.

[168] F. Rugeles Ospina, M. Emerit, and B. F. G. Katz, “The three-dimensional morphological database for spatial hearing research of the BiLi project,” in *International Congress on Acoustics*, 2015. DOI: 10.1121/2.0000050.

[169] J. Fels, P. Buthmann, and M. Vorländer, “Head-related transfer functions of children,” *Acta Acustica united with Acustica*, vol. 90, no. 5, pp. 918–927, 2004. [Online]. Available: https://www.ingentaconnect.com/content/dav/aaua/2004/00000090/00000005/art00012 (Accessed Jun. 20, 2019).

[170] M. Dellepiane, N. Pietroni, N. Tsingos, M. Asselot, and R. Scopigno, “Reconstructing head models from photographs for individualized 3D-audio processing,” *Computer Graphics Forum*, vol. 27, no. 7, pp. 1719–1727, 2008. DOI: 10.1111/j.1467-8659.2008.01316.x.

[171] T. Huttunen, A. Vanne, S. Harder, R. R. Paulsen, S. King, L. Perry-Smith, and L. Kärkkäinen, "Rapid generation of personalized HRTFs," in *AES 55th International Conference: Spatial Audio*, 2014. [Online]. Available: `http://www.aes.org/e-lib/browse.cfm?elib=17365` (Accessed Dec. 7, 2016).

[172] L. Bonacina, A. Canalini, F. Antonacci, M. Marcon, A. Sarti, and S. Tubaro, "A low-cost solution to 3D pinna modeling for HRTF prediction," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016. DOI: `10.1109/ICASSP.2016.7471685`.

[173] A. Andreopoulou, D. R. Begault, and B. F. G. Katz, "Inter-Laboratory Round Robin HRTF Measurement Comparison," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 895–906, 2015. DOI: `10.1109/JSTSP.2015.2400417`.

[174] R. Sridhar, J. G. Tylka, and E. Y. Choueiri, "A database of head-related transfer function and morphological measurements," in *AES 143rd Convention*, 2017. [Online]. Available: `http://www.aes.org/e-lib/browse.cfm?elib=19308` (Accessed Dec. 1, 2017).

[175] S. Marburg, "Six Boundary Elements Per Wavelength: Is That Enough?" *Journal of Computational Acoustics*, vol. 10, no. 1, pp. 25–51, 2002. DOI: `10.1142/S0218396X02001401`.

[176] S. N. Makarov and M. Ochmann, "An iterative solver of the Helmholtz integral equation for high-frequency acoustic scattering," *The Journal of the Acoustical Society of America*, vol. 103, no. 2, pp. 742–750, 1998. DOI: `10.1121/1.421238`.

[177] N. A. Gumerov and R. Duraiswami, "A broadband fast multipole accelerated boundary element method for the three dimensional Helmholtz equation," *The Journal of the Acoustical Society of America*, vol. 125, no. 1, pp. 191–205, 2009. DOI: `10.1121/1.3021297`.

[178] W. Kreuzer, P. Majdak, and Z. Chen, "Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range," *The Journal of the Acoustical Society of America*, vol. 126, no. 3, pp. 1280–1290, 2009. DOI: `10.1121/1.3177264`.

[179] Acoustics Research Institute, *Mesh2HRTF*, version 0.1.2. [Online]. Available: `https://mesh2hrtf.sourceforge.net` (Accessed Nov. 23, 2016).

[180] H. Ziegelwanger, W. Kreuzer, and P. Majdak, "Mesh2HRTF: An Open-Source Software Package for the Numerical Calculation of Head-Related Transfer Functions," in *International Congress on Sound and Vibration*, 2015. DOI: `10.13140/RG.2.1.1707.1128`.

[181] H. Ziegelwanger, P. Majdak, and W. Kreuzer, "Numerical Calculation of Listener-Specific Head-Related Transfer Functions and Sound Localization: Microphone Model and Mesh Discretization," *The Journal of the Acoustical Society of America*, vol. 138, no. 1, pp. 208–222, 2015. DOI: `10.1121/1.4922518`.

[182] N. Tsingos, W. Jiang, and I. Williams, "Using Programmable Graphics Hardware for Acoustics and Audio Rendering," *Journal of the Audio Engineering Society*, vol. 59, no. 9, pp. 628–646, 2011. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15979 (Accessed Sep. 8, 2019).

[183] S. Hamada, "GPU-accelerated indirect boundary element method for voxel model analyses with fast multipole method," *Computer Physics Communications*, vol. 182, no. 5, pp. 1162–1168, 2011. DOI: 10.1016/j.cpc.2011.01.020.

[184] T. Takahashi and T. Hamada, "GPU-accelerated boundary element method for Helmholtz' equation in three dimensions," *International Journal for Numerical Methods in Engineering*, vol. 80, no. 10, pp. 1295–1321, 2009. DOI: 10.1002/nme.2661.

[185] Z. Fan, T. Arce, C. Lu, K. Zhang, T. W. Wu, and K. McMullen, "Computation of Head-related Transfer Functions Using Graphics Processing Units and a Perceptual Validation of the Computed HRTFs against Measured HRTFs," in *AES International Conference on Headphone Technology*, 2019. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=20511 (Accessed Sep. 8, 2019).

[186] G. Haase, M. Schanz, and S. Vafai, "Computing boundary element method's matrices on GPU," in *International Conference on Large-Scale Scientific Computing*, 2011. DOI: 10.1007/978-3-642-29843-1_39.

[187] Y. Kahana, P. A. Nelson, M. Petyt, and S. Choi, "Boundary Element Simulation of HRTFs and Sound Fields Produced by Virtual Acoustic Imaging Systems," in *AES 105th Convention*, 1998. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=8363 (Accessed Jun. 20, 2019).

[188] Y. Kahana and P. A. Nelson, "Numerical modelling of the spatial acoustic response of the human pinna," *Journal of Sound and Vibration*, vol. 292, no. 1-2, pp. 148–178, 2006. DOI: 10.1016/j.jsv.2005.07.048.

[189] T. Walsh, L. Demkowicz, and R. Charles, "Boundary element modeling of the external human auditory system," *The Journal of the Acoustical Society of America*, vol. 115, no. 3, pp. 1033–1043, 2004. DOI: 10.1121/1.1643360.

[190] B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements," *The Journal of the Acoustical Society of America*, vol. 110, no. 5, pp. 2449–2455, 2001. DOI: 10.1121/1.1412441.

[191] B. F. G. Katz, "Acoustic absorption measurement of human hair and skin within the audible frequency range," *The Journal of the Acoustical Society of America*, vol. 108, no. 5, pp. 2238–2242, 2000. DOI: 10.1121/1.1314319.

[192] F. Brinkmann, A. Lindau, S. Weinzierl, S. van de Par, M. Müller-Trapet, R. Opdam, and M. Vorländer, "A High Resolution and Full-Spherical Head-Related Transfer Function Database for Different Head-Above-Torso Orientations," *Journal of the Audio Engineering Society*, vol. 65, no. 10, pp. 841–848, 2017. DOI: 10.17743/jaes.2017.0033.

[193] Y. Rui, G. Yu, B. Xie, and Y. Liu, "Calculation of Individualized Near-Field Head-Related Transfer Function Database Using Boundary Element Method," in *AES 134th Convention*, 2013. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16801 (Accessed Apr. 26, 2019).

[194] C. D. Salvador, S. Sakamoto, J. Treviño, and Y. Suzuki, "Dataset of near-distance head-related transfer functions calculated using the boundary element method," in *AES International Conference: Spatial Reproduction - Aesthetics and Science*, 2018. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=19602 (Accessed Oct. 3, 2018).

[195] A. I. Tew, C. T. Hetherington, and J. B. A. Thorpe, "Morphoacoustic Perturbation Analysis: Principles and Validation," in *Joint Meeting of the 11th Congres Francais d'Acoustique and the 2012 Annual Meeting of the Institute of Acoustics*, 2012. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00811131 (Accessed Jun. 15, 2019).

[196] R. Zolfaghari, N. Epain, C. T. Jin, J. Glaunés, and A. Tew, "Large Deformation Diffeomorphic Metric Mapping and Fast-Multipole Boundary Element Method provide new insights for Binaural acoustics," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2014. DOI: 10.1109/ICASSP.2014.6854123.

[197] S. Kaneko, T. Suenaga, M. Fujiwara, K. Kumehara, F. Shirakihara, and S. Sekine, "Ear Shape Modeling for 3D Audio and Acoustic Virtual Reality: The Shape-Based Average HRTF," in *AES 61st International Conference: Audio for Games*, 2016. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=18091 (Accessed Jun. 15, 2019).

[198] P. J. B. Jackson and N. Desiraju, "Use of 3D Head Shape for Personalized Binaural Audio," in *AES 49th International Conference: Audio for Games*, 2013. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16661 (Accessed Jun. 15, 2019).

[199] S. Ghorbal, T. Auclair, C. Soladié, and R. Séguier, "Pinna morphological parameters influencing HRTF sets," in *International Conference on Digital Audio Effects*, 2017. [Online]. Available: http://www.dafx17.eca.ed.ac.uk/papers/DAFx17_paper_15.pdf (Accessed Jun. 21, 2019).

[200] K. Genuit, "Ein Modell zur Beschreibung von Außenohrübertragungseigenschaften [A model for describing outer ear transmission characteristics]," PhD, RWTH Aachen University, 1984.

[201] P. Satarzadeh, V. R. Algazi, and R. O. Duda, "Physical and filter pinna models based on anthropometry," in *AES 122nd Convention*, 2007. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14083 (Accessed Jun. 18, 2019).

[202] D. Schönstein and B. F. G. Katz, "HRTF selection for binaural synthesis from a database using morphological parameters," in *International Congress on Acoustics*, 2010. [Online]. Available: http://www.acoustics.asn.au/conference_proceedings/ICA2010/cdrom-ICA2010/papers/p266 (Accessed May 15, 2019).

[203] D. Y. N. Zotkin, J. Hwang, R. Duraiswaini, and L. S. Davis, "HRTF personalization using anthropometric measurements," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2003. DOI: 10.1109/ASPAA.2003.1285855.

[204] B. U. Seeber and H. Fastl, "Subjective selection of non-individual head-related transfer functions," in *International Conference on Auditory Display*, 2003. [Online]. Available: https://smartech.gatech.edu/handle/1853/50488 (Accessed Jun. 15, 2019).

[205] A. Lindau, J. Estrella, and S. Weinzierl, "Individualization of dynamic binaural synthesis by real time manipulation of the ITD," in *AES 128th Convention*, 2010. [Online]. Available: http://www.aes.org/e-lib/online/browse.cfm?elib=15385 (Accessed Mar. 11, 2017).

[206] E. A. Torres-Gallegos, F. Orduña-Bustamante, and F. Arámbula-Cosío, "Personalization of head-related transfer functions (HRTF) based on automatic photo-anthropometry and inference from a database," *Applied Acoustics*, vol. 97, pp. 84–95, 2015. DOI: 10.1016/j.apacoust.2015.04.009.

[207] K. Yamamoto and T. Igarashi, "Fully perceptual-based 3D spatial sound individualization with an adaptive variational autoencoder," *ACM Transactions on Graphics*, vol. 36, no. 6, 2017. DOI: 10.1145/3130800.3130838.

[208] S.-N. Yao, T. Collins, and C. Liang, "Head-Related Transfer Function Selection Using Neural Networks," *Archives of Acoustics*, vol. 42, no. 3, pp. 365–373, 2017. DOI: 10.1515/aoa-2017-0038.

[209] A. Kan, C. Jin, and A. van Schaik, "Psychoacoustic Evaluation of a New Method for Simulating Near-field Virtual Auditory Space," in *AES 120th Convention*, 2006. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=13605 (Accessed Jun. 12, 2019).

[210] B. Xie, "Binaural Reproduction through Headphones," in *Head Related Transfer Function and Virtual Auditory Display*, 2nd ed., J. Ross Publishing, 2013, ch. 8, pp. 265–325, ISBN: 978-1-60-427741-8.

[211] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen, "Transfer characteristics of headphones measured on human ears," *Journal of the Audio Engineering Society*, vol. 43, no. 4, pp. 203–217, 1995. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=7954 (Accessed Jun. 11, 2019).

[212] A. Lindau and F. Brinkmann, "Perceptual evaluation of individual headphone compensation in binaural synthesis based on non-individual recordings," *Journal of the Audio Engineering Society*, vol. 60, no. 1/2, pp. 54–62, 2012. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16166 (Accessed Mar. 7, 2019).

[213] D. Pralong and S. Carlile, "The role of individualized headphone calibration for the generation of high fidelity virtual auditory space," *The Journal of the Acoustical Society of America*, vol. 100, no. 6, pp. 3785–3793, 1996. DOI: 10.1121/1.417337.

[214] A. Kulkarni and H. S. Colburn, "Variability in the characterization of the headphone transfer-function," *The Journal of the Acoustical Society of America*, vol. 107, no. 2, pp. 1071–1074, 2000. DOI: 10.1121/1.428571.

[215] T. McKenzie, D. T. Murphy, and G. Kearney, "Interaural Level Difference Optimization of Binaural Ambisonic Rendering," *Applied Sciences*, vol. 9, no. 6, 2019. DOI: 10.3390/app9061226.

[216] Z. Schärer and A. Lindau, "Evaluation of Equalization Methods for Binaural Signals," in *AES 126th Convention*, 2009. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14917 (Accessed Jun. 6, 2019).

[217] B. S. Atal and M. R. Schroeder, "Apparent sound source translator," pat. US3236949A, 1966. [Online]. Available: https://patents.google.com/patent/US3236949 (Accessed May 15, 2019).

[218] D. H. Cooper and J. L. Bauck, "Prospects for Transaural Recording," *Journal of the Audio Engineering Society*, vol. 37, no. 1/2, pp. 3–19, 1989. [Online]. Available: http://www.aes.org/e-lib/online/browse.cfm?elib=6108 (Accessed Jun. 13, 2019).

[219] O. Kirkeby, P. A. Nelson, and H. Hamada, "The 'Stereo Dipole': A Virtual Source Imaging System Using Two Closely Spaced Loudspeakers," *Journal of the Audio Engineering Society*, vol. 46, no. 5, pp. 387–395, 1998. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=12148 (Accessed Apr. 12, 2019).

[220] J. Zheng, T. Zhu, J. Lu, and X. Qiu, "A Linear Robust Binaural Sound Reproduction System with Optimal Source Distribution Strategy," *Journal of the Audio Engineering Society*, vol. 63, no. 9, pp. 725–735, 2015. DOI: 10.17743/jaes.2015.0066.

[221] J. J. Lopez and A. Gonzalez, "Experimental evaluation of cross-talk cancellation regarding loudspeakers' angle of listening," *IEEE Signal Processing Letters*, vol. 8, no. 1, pp. 13–15, 2001. DOI: 10.1109/97.889637.

[222] M. Kohnen, J. Stienen, L. Aspöck, and M. Vorländer, "Dynamic CTC with and without compensation of early reflections," in *AES International Conference on Sound Field Control*, 2016. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=18330 (Accessed Jul. 22, 2016).

[223] M. R. Bai and C.-C. Lee, "Objective and subjective analysis of effects of listening angle on crosstalk cancellation in spatial sound reproduction," *The Journal of the Acoustical Society of America*, vol. 120, no. 4, pp. 1976–1989, 2006. DOI: 10.1121/1.2257986.

[224] Y. Lacouture Parodi and P. Rubak, "Analysis of design parameters for crosstalk cancellation filters applied to different loudspeaker configurations," *Journal of the Audio Engineering Society*, vol. 59, no. 5, pp. 304–320, 2011. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15931 (Accessed Apr. 2, 2016).

[225] A. H. Moore, A. I. Tew, and R. Nicol, "An Initial Validation of Individualized Crosstalk Cancellation Filters for Binaural Perceptual Experiments," *Journal of the Audio Engineering Society*, vol. 58, no. 1/2, pp. 36–45, 2010. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15240 (Accessed May 14, 2019).

[226] M. A. Akeroyd, J. Chambers, D. Bullock, A. R. Palmer, A. Q. Summerfield, P. A. Nelson, and S. Gatehouse, "The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics," *The Journal of the Acoustical Society of America*, vol. 121, no. 2, pp. 1056–1069, 2007. DOI: 10.1121/1.2404625.

[227] B. S. Masiero, "Individualized binaural technology: measurement, equalization and perceptual evaluation," PhD, RWTH Aachen University, 2012. [Online]. Available: http://publications.rwth-aachen.de/record/197486 (Accessed Jun. 22, 2019).

[228] Y. Lacouture Parodi and P. Rubak, "A Subjective Evaluation of the Minimum Channel Separation for Reproducing Binaural Signals over Loudspeakers," *Journal of the Audio Engineering Society*, vol. 59, no. 7/8, pp. 487–497, 2011. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15974 (Accessed Jan. 15, 2016).

[229] M. R. Bai and C.-C. Lee, "Development and implementation of cross-talk cancellation system in spatial audio reproduction based on subband filtering," *Journal of Sound and Vibration*, vol. 290, no. 3-5, pp. 1269–1289, 2006. DOI: 10.1016/j.jsv.2005.05.016.

[230] T. Weissgerber, K. Laumann, G. Theile, and H. Fastl, "Headphone Reproduction via Loudspeakers using Inverse HRTF-Filters," in *NAG/DAGA*, 2009. [Online]. Available: https://mediatum.ub.tum.de/doc/980062/file.pdf (Accessed Jul. 2, 2019).

[231] N. Prodi and S. Velecka, "The evaluation of binaural playback systems for virtual sound fields," *Applied Acoustics*, vol. 64, no. 2, pp. 147–161, 2003. DOI: 10.1016/S0003-682X(02)00072-5.

[232] G. Kearney, "The perception of auditory height in individualised and non-individualised dynamic cross-talk cancellation," in *AES International Conference on Sound Field Control*, 2016. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=18323 (Accessed Sep. 8, 2019).

[233] A. Mouchtaris, P. Reveliotis, and C. Kyriakakis, "Inverse filter design for immersive audio rendering over loudspeakers," *IEEE Transactions on Multimedia*, vol. 2, no. 2, pp. 77–87, 2000. DOI: 10.1109/6046.845012.

[234] T. Lentz, "Dynamic Crosstalk Cancellation for Binaural Synthesis in Virtual Reality Environments," English, *Journal of the Audio Engineering Society*, vol. 54, no. 4, pp. 283–294, 2006. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=13677 (Accessed Apr. 23, 2019).

[235] Y. Kim, P. Mannerheim, and P. A. Nelson, “Filter Update Techniques for Adaptive Virtual Acoustic Imaging,” in *AES 120th Convention*, 2006. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=13519 (Accessed Apr. 12, 2019).

[236] D. Menzel, H. Wittek, G. Theile, and H. Fastl, “The binaural sky: A virtual headphone for binaural room synthesis,” in *International Tonmeister Symposium*, 2005. [Online]. Available: https://hauptmikrofon.de/HW/BinauralSky_Hohenkammer.pdf (Accessed May 10, 2019).

[237] K. Laumann, G. Theile, and H. Fastl, “A virtual headphone based on wave field synthesis (abstract),” *The Journal of the Acoustical Society of America*, vol. 123, no. 5, 2008. DOI: 10.1121/1.2934425.

[238] K. Laumann, G. Theile, and H. Fastl, “Binaural Sky - Examination of Different Array Topologies,” in *NAG/DAGA*, 2009. [Online]. Available: https://hauptmikrofon.de/theile/2009-1_Binaural-Sky-dieffernt%20arrays_DAGA-2009.pdf (Accessed May 10, 2019).

[239] M. Guldenschuh and A. Sontacchi, “Transaural Stereo In a Beamforming Approach,” in *International Conference on Digital Audio Effects*, 2009. [Online]. Available: http://dafx09.como.polimi.it/proceedings/papers/paper_58.pdf (Accessed May 10, 2019).

[240] D. B. Ward and G. W. Elko, “Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation,” *IEEE Signal Processing Letters*, vol. 6, no. 5, pp. 106–108, 1999. DOI: 10.1109/97.755428.

[241] T. Takeuchi and P. A. Nelson, “Optimal source distribution for binaural synthesis over loudspeakers,” *The Journal of the Acoustical Society of America*, vol. 112, no. 6, pp. 2786–2797, 2002. DOI: 10.1121/1.1513363.

[242] M. Yairi, T. Takeuchi, and K. Holland, “Binaural reproduction capability for multiple off-axis listeners based on the optimal source distribution principle,” in *AES International Conference on Immersive and Interactive Audio*, 2019. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=20424 (Accessed Jun. 24, 2019).

[243] C. Han, T. Okamoto, Y. Iwaya, and Y. Suzuki, “Loudspeaker distributions suitable for crosstalk cancellers robust to head rotation,” *Acoustical Science and Technology*, vol. 33, no. 4, pp. 266–269, 2012. DOI: 10.1250/ast.33.266.

[244] M. Cohen, “The Internet Chair,” *International Journal of Human-Computer Interaction*, vol. 15, no. 2, pp. 297–311, 2003. DOI: 10.1207/S15327590IJHC1502_7.

[245] M. Cohen, O. N. Newton Fernando, T. Nagai, and K. Shimizu, “Back-Seat Driver: Spatial Sound for Vehicular Way-Finding and Situation Awareness,” in *Japan-China Joint Workshop on Frontier of Computer Science and Technology*, 2006. DOI: 10.1109/FCST.2006.1.

[246] M. Cohen and J. Villegas, "GABRIEL: Geo-Aware Broadcasting for In-Vehicle Entertainment and Localizability," in *AES 40th International Conference: Spatial Audio - Sense the Sound of Space*, 2010. [Online]. Available: `http://www.aes.org/e-lib/browse.cfm?elib=15556` (Accessed Apr. 13, 2019).

[247] M. Cohen, private communication, Feb. 2016.

[248] HUMU, *Humu*. [Online]. Available: `https://humu.fi/tilaa-oma-humu/humu-augmented-audio-cushion` (Accessed Sep. 11, 2019).

[249] C. T. Hetherington, "HRTF estimation by shape parameterization of the human head and pinnae," PhD, Department of Electronics, University of York, 2004.

[250] A. Reichinger, P. Majdak, R. Sablatnig, and S. Maierhofer, "Evaluation of methods for optical 3-D scanning of human pinnas," in *International Conference on 3D Vision*, 2013. DOI: `10.1109/3DV.2013.58`.

[251] International Organization for Standardization, *ISO 10303-21:2016 - Industrial automation systems and integration – Product data representation and exchange – Part 21: Implementation methods: Clear text encoding of the exchange structure*, 2016. [Online]. Available: `https://www.iso.org/standard/63141.html` (Accessed Jun. 10, 2019).

[252] P. Schmiechen, "Travelling Wave Speed Coincidence," PhD, Imperial College of Science, Technology and Medicine, University of London, 1997.

[253] H. Ziegelwanger, W. Kreuzer, and P. Majdak, "A-priori mesh grading for the numerical calculation of the head-related transfer functions," *Applied Acoustics*, vol. 114, no. 1, pp. 99–110, 2016. DOI: `10.1016/j.apacoust.2016.07.005`.

[254] A. E. Johnson and M. Hebert, "Control of polygonal mesh resolution for 3-D computer vision," *Graphical Models and Image Processing*, vol. 60, no. 4, pp. 261–285, 1998. DOI: `10.1006/gmip.1998.0474`.

[255] S. Valette, J.-M. Chassery, and R. Prost, "Generic Remeshing of 3D Triangular Meshes with Metric-Dependent Discrete Voronoi Diagrams," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 2, pp. 369–381, 2008. DOI: `10.1109/TVCG.2007.70430`.

[256] S. Valette, J.-M. Chassery, and R. Prost, *ACVD: Surface Mesh Coarsening and Resampling*, version 1.1. [Online]. Available: `https://www.creatis.insa-lyon.fr/site7/en/acvd` (Accessed Aug. 15, 2016).

[257] Geomagic, *Geomagic Wrap*, version 2015.0.0. [Online]. Available: `http://www.geomagic.com/en/products/wrap/overview` (Accessed Aug. 15, 2016).

[258] MeshLab, *MeshLab*, version 2016.12. [Online]. Available: `http://www.meshlab.net/` (Accessed Aug. 15, 2016).

[259] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia, "MeshLab: an Open-Source Mesh Processing Tool," in *Eurographics Italian Chapter Conference*, 2008. DOI: `10.2312/LocalChapterEvents/ItalChap/ItalianChapConf2008/129-136`.

[260] S. Kirkup, "Boundary Representation," in *The Boundary Element Method in Acoustics*, 2nd ed., Integrated Sound Software, 2007, ch. 2, pp. 21–30. [Online]. Available: https://www.researchgate.net/publication/261760562 (Accessed Nov. 15, 2016).

[261] PACSYS Ltd, *PAFEC-FE*, version 8.8. [Online]. Available: http://www.pafec.info/pafec/ (Accessed Aug. 24, 2016).

[262] Autodesk, *Autodesk Inventor*. [Online]. Available: http://www.autodesk.co.uk/products/inventor/overview (Accessed Oct. 28, 2016).

[263] Ansol LLC, *Coustyx*. [Online]. Available: http://ansol.us/Products/Coustyx/ (Accessed Jul. 10, 2019).

[264] Autodesk, *3DS Max*, version 2016. [Online]. Available: https://www.autodesk.co.uk/products/3ds-max/overview (Accessed Nov. 28, 2018).

[265] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique," in *AES 108th Convention*, 2000. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=10211 (Accessed Jun. 12, 2017).

[266] Cockos Incorporated, *Reaper*, version 5.01. [Online]. Available: https://www.reaper.fm/index.php (Accessed Dec. 5, 2018).

[267] RS Online, *Tectonic Elements Round Speaker Driver*. [Online]. Available: http://uk.rs-online.com/web/p/speaker-drivers/8765241/ (Accessed Nov. 11, 2016).

[268] H. Anton, "Real Numbers, Sets, and Inequalities (A Review)," in *Calculus with Analytic Geometry*, 1st ed., John Wiley & Sons, 1980, ch. 1.1, pp. 1–17, ISBN: 0-471-03248-4.

[269] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduña-Bustamante, "Fast Deconvolution of Multichannel Systems Using Regularization," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 189–195, 1998. DOI: 10.1109/89.661479.

[270] MathWorks, *delauney*. [Online]. Available: https://uk.mathworks.com/help/matlab/ref/delaunay.html (Accessed Feb. 13, 2019).

[271] MathWorks, *MATLAB*, version R2016b. [Online]. Available: https://uk.mathworks.com/products/new_products/release2016b.html?s_tid=srchtitle (Accessed Oct. 28, 2016).

[272] Blender Foundation, *Blender*, version 2.78c. [Online]. Available: https://www.blender.org/download/ (Accessed Nov. 23, 2016).

[273] MathWorks, *boxplot*. [Online]. Available: https://uk.mathworks.com/help/stats/boxplot.html (Accessed Jul. 3, 2019).

[274] F. Rugeles Ospina, M. Emerit, and J. Daniel, "A fast measurement of high spatial resolution head-related transfer functions for the BiLi Project," in *International Conference on Spatial Audio*, 2015.

[275] X. Zhong, X. Xu, and B. Xie, "Auditory Consistency of Head-Related Transfer Functions of Kemar From Different Databases," in *International Congress on Sound and Vibration*, 2016. [Online]. Available: http://www.iiav.org/archives_icsv_last/2016_icsv23/content/papers/papers/full_paper_111_20160512031834279.pdf (Accessed Jun. 7, 2017).

[276] F. Brinkmann, A. Lindau, M. Müller-Trapet, M. Vorländer, and S. Weinzierl, "Cross-Validation of Measured and Modeled Head-Related Transfer Functions," in *DAGA*, 2015. [Online]. Available: https://www2.ak.tu-berlin.de/~akgroup/ak_pub/2015/Brinkmann_2015_HRTF_cross_validation_DAGA.pdf (Accessed May 28, 2019).

[277] C. Armstrong, T. McKenzie, D. Murphy, and G. Kearney, "A Perceptual Spectral Difference Model for Binaural Signals," in *AES 145th Convention*, 2018. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=19722 (Accessed May 10, 2019).

[278] H. Ziegelwanger, A. Reichinger, and P. Majdak, "Calculation of listener-specific head-related transfer functions: Effect of mesh quality," in *International Congress on Acoustics*, 2013. DOI: 10.1121/1.4799868.

[279] G. Farin, "The Bernstein Form of a Bézier Curve," in *Curves and Surfaces for Computer-Aided Geometric Design*, 3rd ed., Academic Press, 1993, ch. 4, pp. 41–64, ISBN: 978-0-12-249052-1.

[280] D. M. Green, G. Kidd, and K. N. Stevens, "High-frequency audiometric assessment of a young adult population," *The Journal of the Acoustical Society of America*, vol. 81, no. 2, pp. 485–494, 1987. DOI: 10.1121/1.394914.

[281] P. G. Stelmachowicz, K. A. Beauchaine, A. Kalberer, and W. Jesteadt, "Normative thresholds in the 8- to 20-kHz range as a function of age," *The Journal of the Acoustical Society of America*, vol. 86, no. 4, pp. 1384–1391, 1989. DOI: 10.1121/1.398698.

[282] MathWorks, *regionprops*. [Online]. Available: https://uk.mathworks.com/help/images/ref/regionprops.html (Accessed Mar. 5, 2019).

[283] MathWorks, *griddata*. [Online]. Available: https://uk.mathworks.com/help/matlab/ref/griddata.html (Accessed Mar. 18, 2019).

[284] E. W. Weisstein, *Icosahedral Group*. [Online]. Available: http://mathworld.wolfram.com/IcosahedralGroup.html (Accessed Aug. 9, 2019).

[285] MathWorks, *MATLAB*, version R2017b. [Online]. Available: https://uk.mathworks.com/products/new_products/release2017b.html (Accessed Oct. 28, 2016).

[286] MathWorks, *imgaussfilt*. [Online]. Available: https://uk.mathworks.com/help/images/ref/imgaussfilt.html (Accessed Apr. 3, 2019).

[287] P. L. Søndergaard and P. Majdak, "The Auditory Modeling Toolbox," in *The Technology of Binaural Listening*, Springer Berlin Heidelberg, 2013, pp. 33–56. DOI: 10.1007/978-3-642-37762-4_2.

[288] Acoustics Research Institute, *The Auditory Modeling Toolbox*. [Online]. Available: http://amtoolbox.sourceforge.net/ (Accessed Jul. 9, 2019).

[289] D. Legland, *Geom3d*, 2018. [Online]. Available: https://uk.mathworks.com/matlabcentral/fileexchange/24484-geom3d (Accessed Jun. 3, 2019).

[290] K. A. Stroud, *Engineering Mathematics*, 6th ed. Palgrave Macmillan, 2007, ISBN: 978-1-4039-4246-3.

[291] J. Blackburne, *Angle between two 3D vectors*, 2012. [Online]. Available: http://johnblackburne.blogspot.co.uk/2012/05/angle-between-two-3d-vectors.html (Accessed Nov. 23, 2016).

[292] J. Simon, *Angle between two vectors in 3d - MATLAB Answers*, 2011. [Online]. Available: http://uk.mathworks.com/matlabcentral/answers/16243-angle-between-two-vectors-in-3d (Accessed Nov. 23, 2016).

[293] E. W. Weisstein, *Line-Plane Intersection*. [Online]. Available: http://mathworld.wolfram.com/Line-PlaneIntersection.html (Accessed Nov. 24, 2016).

[294] R. Nourai, *Barycentric Coordinates and Point in Triangle Tests*, 2011. [Online]. Available: https://blogs.msdn.microsoft.com/rezanour/2011/08/07/barycentric-coordinates-and-point-in-triangle-tests/ (Accessed Nov. 24, 2016).

[295] Trimble Inc., *SketchUp Make*. [Online]. Available: https://www.sketchup.com (Accessed Jul. 10, 2019).

[296] Valuframe, *Valuframe - The Economy Aluminium Profile System: Series 5 aluminium profiles*. [Online]. Available: http://www.valuframe.co.uk/Series-5-Aluminium-Profiles.html (Accessed May 28, 2019).

[297] Valuframe, *Valuframe - The Economy Aluminium Profile System: Slot nuts*. [Online]. Available: http://www.valuframe.co.uk/Slot-Nuts.html (Accessed May 28, 2019).

Figure Permissions

The following figures have been reproduced with the following permissions:

Fig. 2.8: Reproduced from [40] with permission of Taylor and Francis Group LLC (Books) US in the format Thesis/Dissertation via Copyright Clearance Center.

Fig 2.9: Reproduced from [29] with permission of FOCAL in the format Thesis/Dissertation via Copyright Clearance Center.

Fig. 2.22: Reproduced from [90] with permission from the Audio Engineering Society and the author.

Fig. 2.23: Reproduced from [99] with permission from the author.

Fig. 2.24: Reproduced from [74] under CC-BY-4.0 Creative Commons license.

Fig. 2.32: Reproduced from [21] with permission © 2001 IEEE.

Fig. 2.34a: Reproduced from [245] with permission © 2006 IEEE.

Fig. 2.34b: Reproduced from [23] with permission from the Audio Engineering Society and the author.

Fig. 2.35: Reproduced from [20] with permission from the Audio Engineering Society and the author.

Fig. 3.1: Reproduced from [99] with permission from the author.