Human Sound Localisation Cues and their Relation to Morphology

Jonathan Benjamin Alexis THORPE

Submitted to the University of York
for the Degree of Doctor of Philosophy

University of York,
Department of Electronics,
York, United Kingdom, YO10 5DD

March 2009

Abstract

Binaural soundfield reproduction has the potential to create realistic three-dimensional sound scenes using only a pair of normal headphones. Possible applications for binaural audio abound in, for example, the music, mobile communications and games industries. A problem exists, however, in that the head-related transfer functions (HRTFs) which inform our spatial perception of sound are affected by variations in human morphology, particularly in the shape of the external ear. It has been observed that HRTFs simply based on some kind of average head shape generally result in poor elevation perception, weak externalisation and spectrally distorted sound images. Hence, HRTFs are needed which accommodate these individual differences. Direct acoustic measurement and acoustic simulations based on morphological measurements are obvious means of obtaining individualised HRTFs, but both methods suffer from high cost and practical difficulties. The lack of a viable measurement method is currently hindering the widespread adoption of binaural technologies. There have been many attempts to estimate individualised HTRFs effectively and cheaply using easily obtainable morphological descriptors, but due to an inadequate understanding of the complex acoustic effects created in particular by the external ear, success has been limited. The work presented in this thesis strengthens current understanding in several ways and provides a promising route towards improved HRTF estimation. The way HRTFs vary as a function of direction is compared with localisation acuity to help pinpoint spectral features which contribute to spatial perception. 50 subjects have been scanned using magnetic resonance imaging to capture their head and pinna morphologies, and HRTFs for the same group have been measured acoustically. To make analysis of this extensive data tractable, and so reveal the mapping between the morphological and acoustic domains, a parametric method for efficiently describing head morphology has been developed. Finally, a novel technique, referred to as morphoacoustic perturbation analysis (MPA), is described. We demonstrate how MPA allows the morphological origin of a variety of HRTF spectral features to be identified.

Contents

List of Figures

List of Tables

Acknowledgements

I cannot overstate how crucial my Ph.D. supervisor Tony Tew has been to the completion of this thesis. His enthusiastic guidance and passionate commitment to research are exemplary. I am truly grateful to him for the many discussions which have shaped this work over the course of the last four years and for the way I felt both free to explore my own ideas and safe in the knowledge that care, advice, concern, encouragement and good ideas would be there when I needed them.

The research work described in this thesis would also have been completely impossible without the dedication and skills of Carl Hetherington who's title of "research associate" hardly begins to describe his participation. His coding was crucial for all acoustic simulations, mesh processing and reconstructions, shape parameterisation as well as the development and maintenance of back-up systems, literature databases and so much more. I am truly thankful for his help and advice in matters of programming especially for showing me the great power of the open source community to which I am indebted. Open source software libraries have been of huge benefit to our research and my thanks go to everyone involved in their development.

I would like to thank the senior members of CARlab, Craig Jin and Andre van Schaik for their invaluable input, ideas and technical assistance as well as CARlab students (Aengus Martin, Alan Kan and Andrew Wabnitz in particular) who were very welcoming and helpful throughout my stay in Sydney. Acoustic measurements would not have been possible without the

enthusiastic assistance of the Auditory Neuroscience Laboratory staff (Simon Carlile, Joel Cooper, Johahn Leung, Caitlin Corkhill and Virginina Best) and I am grateful for having been granted access to their state of the art facilities. The level of collaboration between CARlab and the ANL is an example to Engineering and Psychology departments worldwide. Kirsten Moffat from the Symbion Imaging Centre somehow made the MRI scanning of 25 subjects fun and her expertise was crucial in obtaining high quality MRI scans. Also, the professional support and advise from Patrick Macey (PACSYS Ltd.) in all matters relating to their BEM acoustic simulation software (PAFEC) was invaluable throughout the project and I speak for everyone in the research team when expressing my gratitude.

I would also like to thank everyone in (the) Audio lab for providing a friendly, helpful, dynamic and fun working environment, in no particular order Jude Brereton (for being invariably funny, cheerful and for making lunch for me in the final days), Alex Southern, Alastair Moore (the wise one), Ag Asri Ag Ibrahim (the badminton master), Giorgos Siamantas, Helena Daffern, Rachel van Besouw (thanks for the kitchen and for keeping the plants alive), Eva Björkner, Jez Wells, Sandra Pauletto, Matt Speed, Damian Murphy, Andy Hunt and David Howard. Other people in the Electronics department who have helped me in a number of ways include Brenda, Camilla, Joan, Dereck, Arthur, Owen and the many others too numerous to mention.

Last but certainly not least I would like to thank my family, especially my parents Simon and Michèle, for their love and support throughout my undergraduate and postgraduate studies. I would not be where I am today without their help and my debt to them is beyond measure.

Declaration

I certify that this Thesis is entirely the product of my own work unless otherwise indicated in the text, references or acknowledgements. Any errors and inaccuracies are my own and, accordingly, I take full responsibility for them.

Chapter 1

Introduction

1.1 Background

The ability of humans to accurately estimate the location of sound sources in their three-dimensional environment solely on the basis of auditory input is nothing short of astonishing considering the nature of auditory information. The neural encoding of spatial attributes observed for vision and touch for example, flows from the spatial arrangements of numerous and essentially identical sensory receptors working as a unit, allowing a two-dimensional topographic representation of space. This intrinsically spatial representation does not occur in audition. The organ of Corti (or spiral organ) contained within the cochlea encodes sonic vibrations as a function of frequency by means of 15,000-20,000 auditory nerve receptors excited by hair cells. These hair cells have different resonant frequencies which together cover the entire auditory range giving a highly detailed frequency domain picture with no evident spatial information. Our understanding of the processes which allow us to achieve auditory-based spatial awareness was for the best part of last century limited to the inter-aural time and level difference cues ITD

and ILD, respectively. They arise from the spatial separation of the two ears and the masking effect of the head, which are known to play a crucial role for the perception of sound location in the horizontal plane but do not allow elevation perception. Instantaneous three dimensional sound localisation purely using inter-aural time and level differences would require more than two ears.

Although head movements provide additional location information, it is clear that three-dimensional localisation is, to some extent, possible without them. This ability has been attributed to the filtering of incoming sound waves by the auditory periphery; namely the torso, shoulders, head and, most importantly, the outer ears (pinnae). The directionally dependent character of this filtering is often expressed as a set of head-related transfer functions (HRTFs) or, equivalently, as a set of head-related impulse responses (HRIRs). Sound source location (most importantly elevation) can be inferred by the brain using these directionally dependent spectral and temporal variations. This makes it possible for humans to construct a three-dimensional representation of their surrounding sonic environment even in static conditions. Although this requires assumptions to be made about the acoustic character of the sound source, sound localisation in natural conditions is remarkably robust. This is thought to be due to sharp spectral features observed in HRTFs (Hebrank and Wright, 1974; Langendijk and Bronkhorst, 2002). Such spectral characteristics are rarely present in the mostly broadband natural sounds produced by hunting predators and their prey which are likely to be the main driving force behind the evolution of sound localisation capabilities in animals (Manley *et al.*, 2004; Popper and Fay, 2005). An HRTF can therefore be thought of as a spectral fingerprint uniquely associated with a particular direction, which is imprinted

upon the original source spectrum providing the information required for human sound localisation processes to operate.

In principle, filtering a sound source with a pair of HRTFs (one for each ear) and playing the result through headphones will fool the brain into an artificial sensation of sound location (see Figure 1.1). Further sources in different directions can be spatialised using a similar process and the results summed to create soundfields of arbitrary complexity, a process often referred to as binaural sound rendering. This potentially makes binaural sound an extremely powerful and affordable technology with almost limitless commercial applications. There are, however, technical challenges to be overcome if the technique's full potential is to be achieved. Both the production and perception of spectral localisation cues are complex, intricate processes. Furthermore it is generally accepted that morphological disparities across individuals affect cue production mechanisms. The neural circuitry which decodes source position for a given individual is finely tuned to the acoustics which characterise their unique auditory periphery (Wenzel *et al.*, 1993). As a consequence the use of generic HRTFs based on population means for binaural sound rendering generally results in poor externalisation and elevation perception, as well as unwanted timbral colouration and distortion. All these problems are, to some extent, attributable to the unfamiliar acoustics resulting from the use of unmatched HRTFs.

1.2 Scope and Motivation

It has been shown that individuals can adapt to a set of unfamiliar HRTFs through prolonged exposure and visual feedback leading to somewhat im-

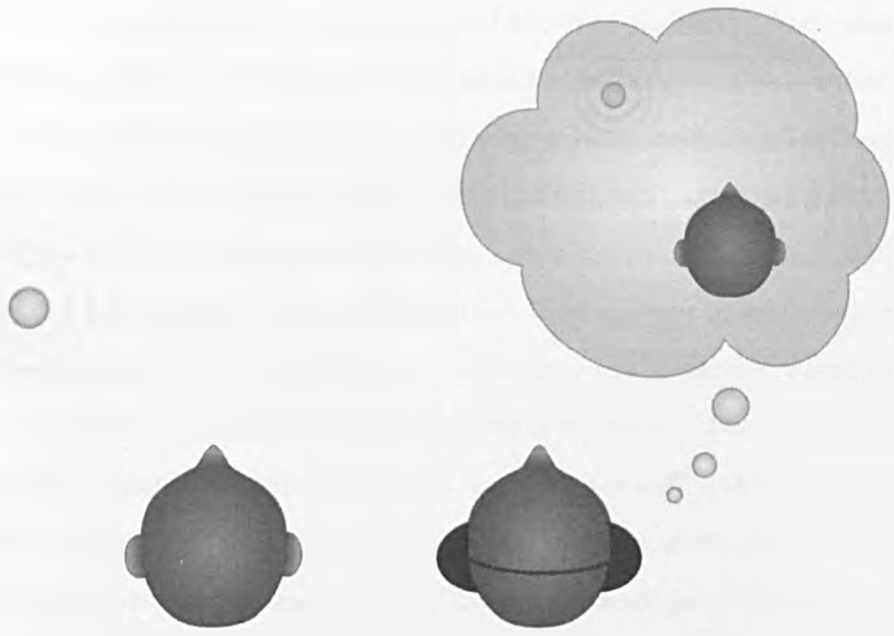


Figure 1.1: A real and a virtual sound source

proved localisation performance (Shinn-Cunningham *et al.*, 1998a,b, 2001; Zahorik, 2001; Zahorik *et al.*, 2006). The plasticity of the brain allows it to partially adapt to an unfamiliar set of HRTFs, however, the improved performance is often attributed to subjects associating spectral patterns with source locations rather than to a true sensation of source location. Binaural sound should be capable of providing realistic rendering of sound-fields without the need for extra listening effort or prior conditioning which are clearly undesirable.

Adapting the virtual auditory space to each listener by providing matching HRTFs has been shown to drastically improve localisation performance, providing a close match to free-field listening conditions (Wightman and Kistler, 1989b; Carlile and Pralong, 1995). This process is known as HRTF individualisation, also referred to as personalisation. HRTFs are most commonly

obtained by making acoustic measurements (Algazi *et al.*, 2001c; Wightman and Kistler, 1989a), but they are time consuming and require costly equipment. More recently, estimation of HRTFs by acoustic modelling has been achieved (Katz, 2001a,b; Kahana and Nelson, 2005), but numerical simulation is, again, time-consuming and involves expensive computing resources. Difficulties in obtaining accurate HRTFs have been a major obstacle to the potential commercial success of binaural technology and a number of studies have attempted to simplify the individualisation procedure. The origin and effects of the inter-aural differences, ITD and ILD, are well understood and reliable techniques have been developed to estimate them, given an appropriate set of morphological measurements. The intricate shape of the external ear, however, makes the estimation of individualised spectral HRTF features from morphology far more challenging.

The pursuit of a model which will allow the cheap, efficient and effective individualisation of spectral localisation cues from morphological descriptors has led to simplified mathematical models describing multi-path reflection patterns (Batteau, 1967; Watkins, 1978). Alternate models which break the external ear down into separate resonant cavities have also been developed (Brown and Duda, 1997, 1998; Gupta *et al.*, 2004). Individual cavities are modelled as isolated linear time-invariant systems which are then combined by superposition. Although these models accounted for some prominent HRTF features, they were conceded to be oversimplifications. Indeed, the great complexity of pinna acoustics forces assumptions and approximations which place severe limits on their validity.

Statistical approaches aiming to extract mappings between variations in morphology and variations in HRTFs have received considerable attention in

recent years (Algazi *et al.*, 2001c; Rodriguez and Ramirez, 2005). These approaches aim to bypass the mathematical difficulties involved in attempting to model external ear acoustics. Results have been disappointing, however, and it has been suggested that there is a need for "a deeper understanding of the perceptually important characteristics of the HRTF and of their dependence on detailed pinna features"(Algazi *et al.*, 2001c). Such an understanding would allow the morphological and acoustic datasets to focus description on relevant attributes, thereby maximising chances for the extraction of meaningful mappings allowing the effective individualisation of artificially reproduced auditory spaces. Filtering out superfluous data while enhancing the focus on relevant features would greatly improve the chances of extracting meaningful mappings allowing accurate HRTF estimation from morphology and, ultimately, the successful individualisation of binaural virtual auditory spaces.

1.3 Thesis Outline

The overall objective of this thesis is to deepen the current understanding of spectral localisation cues and of the acoustic mechanisms which produce them. A review of the relevant literature is presented in Chapter 2, including a general overview of the principles of sound localisation with a special emphasis on spectral cues and elevation perception. Shape description methods, including landmark measurements, are described followed by a review of the chief HRTF estimation techniques will be described highlighting possible areas for improvement. The final section describes the principles of the differential pressure synthesis (DPS) method, which efficiently estimates the acoustic effects of applying small arbitrary deformations to an

object using a pre-computed database (referred to as DPS database) cataloguing the acoustic effects of a set of orthogonal deformations. The results of performance tests carried out to determine its bounds of validity are also described.

A study of HRTF variations in the auditory space surrounding a Knowles Electronic Manikin for Acoustic Research (KEMAR) is presented in Chapter 3. It is conducted using boundary element method (BEM) based acoustic simulations for which a novel multi-resolution approach to meshing is introduced. Results are compared with previous findings. The analysis focuses especially on HRTF variations in the horizontal, frontal and median planes (see Figure 1.2) as well as a number of ipsilateral sagital planes which are parallel to the median plane. The directional spectral variations which allow source location discrimination along loci of constant ITD and ILD are investigated in detail. During the analysis, the distinction between spectral features and their spatial variation is highlighted.

Chapter 4 describes a collection of head and torso morphology data for 49 subjects and the acoustic measurement of their HRTFs. The shape description method differs significantly from those employed in the great majority of previous attempts to identify connections between morphology and HRTFs and which have revealed only relatively weak correlations. This novel approach addresses the intrinsically incomplete, sparse nature of landmark measurement based shape description, which has been commonly used and which makes the omission of relevant shape detail a likely explanation for a number of disappointing results. An accurate and comprehensive shape description technique based on MRI scans of the head and shoulders capturing the finest cross-subject morphological variations is described. Pre-processing

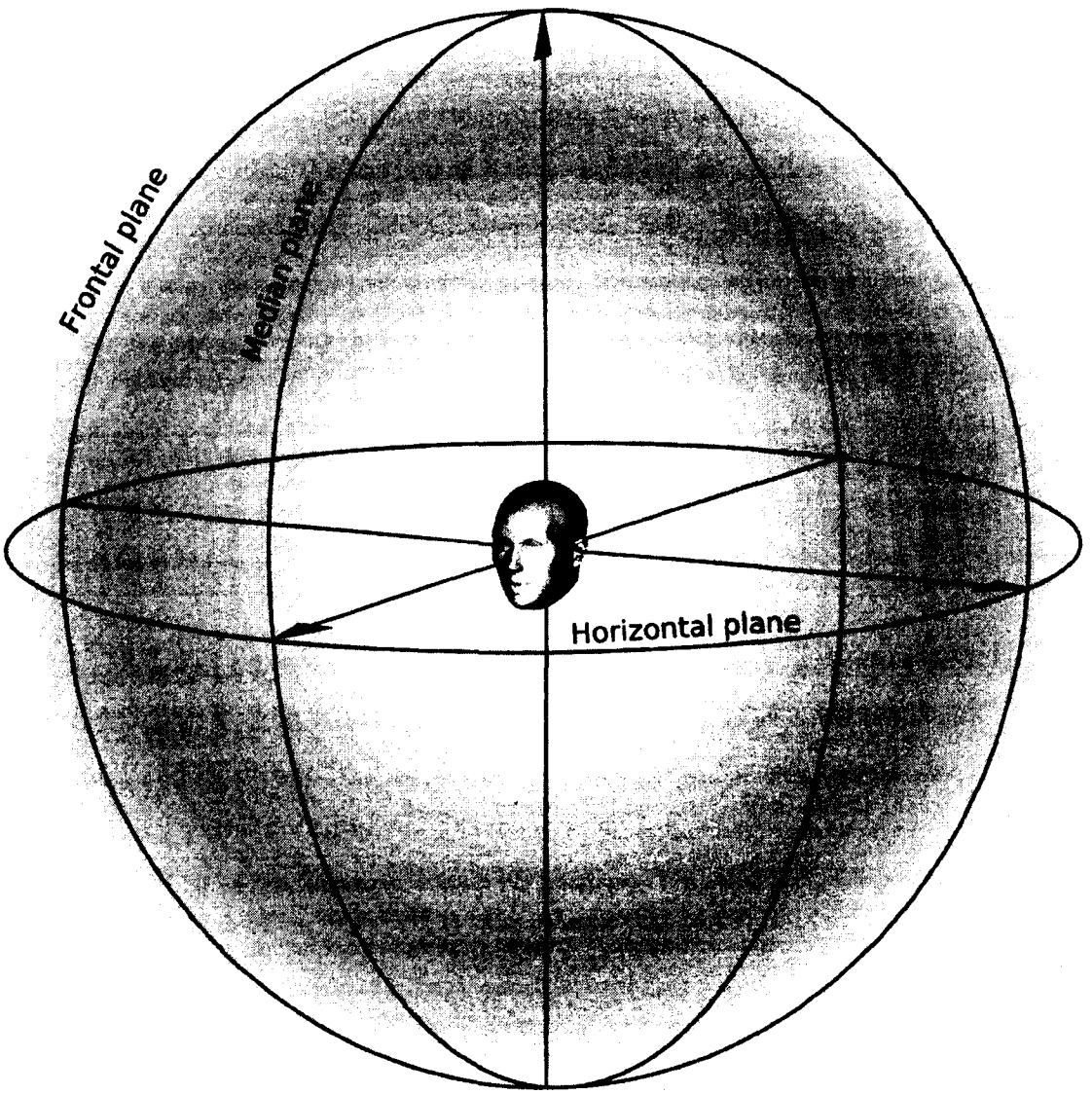


Figure 1.2: Planes in the auditory space

stages extract a clean outer surface contour of the head and pinnae from the raw MRI scan data. At this stage, the description, though complete is highly redundant. To condense the data, a parameterisation method based on principal component analysis, which exploits this redundancy, is proposed and

tested.

Chapter 5 extends work on the DPS method developed by Tao *et al.* (2003a). Although DPS performance tests conducted by Tao *et al.* yielded promising results, the method as it was could not be applied to the human pinna because of intrinsic limitations of the orthogonal shape descriptors they used. New orthogonal deformation functions which solve this problem are presented, along with details of a large DPS database compiled for the KEMAR head and ears using these functions. The results of performance tests are presented and possible applications are discussed, including the introduction of a new technique known as morphoacoustic perturbation analysis (MPA).

The novel aspects of the work described in this thesis will be summarised in Chapter 6. The possibility of combining them to achieve affordable, efficient and effective HRTF estimation is considered. Finally, possible future research paths are proposed which, it is hoped, will greatly enhance the prospect of widely accessible high performance binaural technology.

Chapter 2

Literature Review

"Inquiry is fatal to certainty."

Will Durant

The aim of this thesis is to further our understanding of the relation between the shape of the human head (especially the pinnae) and the cues used for sound localisation. This understanding is crucial for developing techniques allowing the affordable and efficient individualisation of localisation cues; key to optimised virtual auditory spaces. An overview of past research efforts which have shaped our understanding of three-dimensional human sound localisation is given after an initial clarification of the terminology used throughout the thesis. The successes and shortcomings of the current approaches to HRTF estimation will then be overviewed. Finally, the recently developed differential pressure synthesis (DPS) method, central to the final chapter in this thesis, will be described.

2.1 Terminology

This section clarifies the terminology used to define sound source location in auditory space and to describe the external ear in detail.

2.1.1 Planes in the auditory space

The perception of source distance will not be investigated in this thesis. Consequently, source location is defined uniquely in terms of direction and is constrained to the surface of a sphere around the listener. We therefore use the term 'hemisphere' in place of the more usual 'half-space' in the following plane definitions. The main planes in the auditory space were illustrated in Figure 1.2. The **horizontal** (also known as transverse) plane separates the top and bottom hemispheres, the **frontal** (also known as coronal) plane separates the front and back hemispheres and the **median** (also known as mid-sagittal) plane separates the left and right hemispheres. Planes parallel to the mid-sagittal plane but which do not contain the origin are simply referred to as sagittal planes. The horizontal, median and sagittal planes are defined by three landmark points; the centres of both eardrums E_1 and E_2 (which define the interaural axis) and the tip of the nose N. The horizontal plane contains all three points E_1, E_2 and N. The frontal plane contains E_1 and E_2 and is perpendicular to the horizontal plane. The median plane contains N and is perpendicular to both the horizontal and frontal planes.

2.1.2 Spherical coordinate systems

The origin of the coordinate system is placed half way between the points E_1 and E_2 (defined in section 2.1.1). Two different spherical coordinate

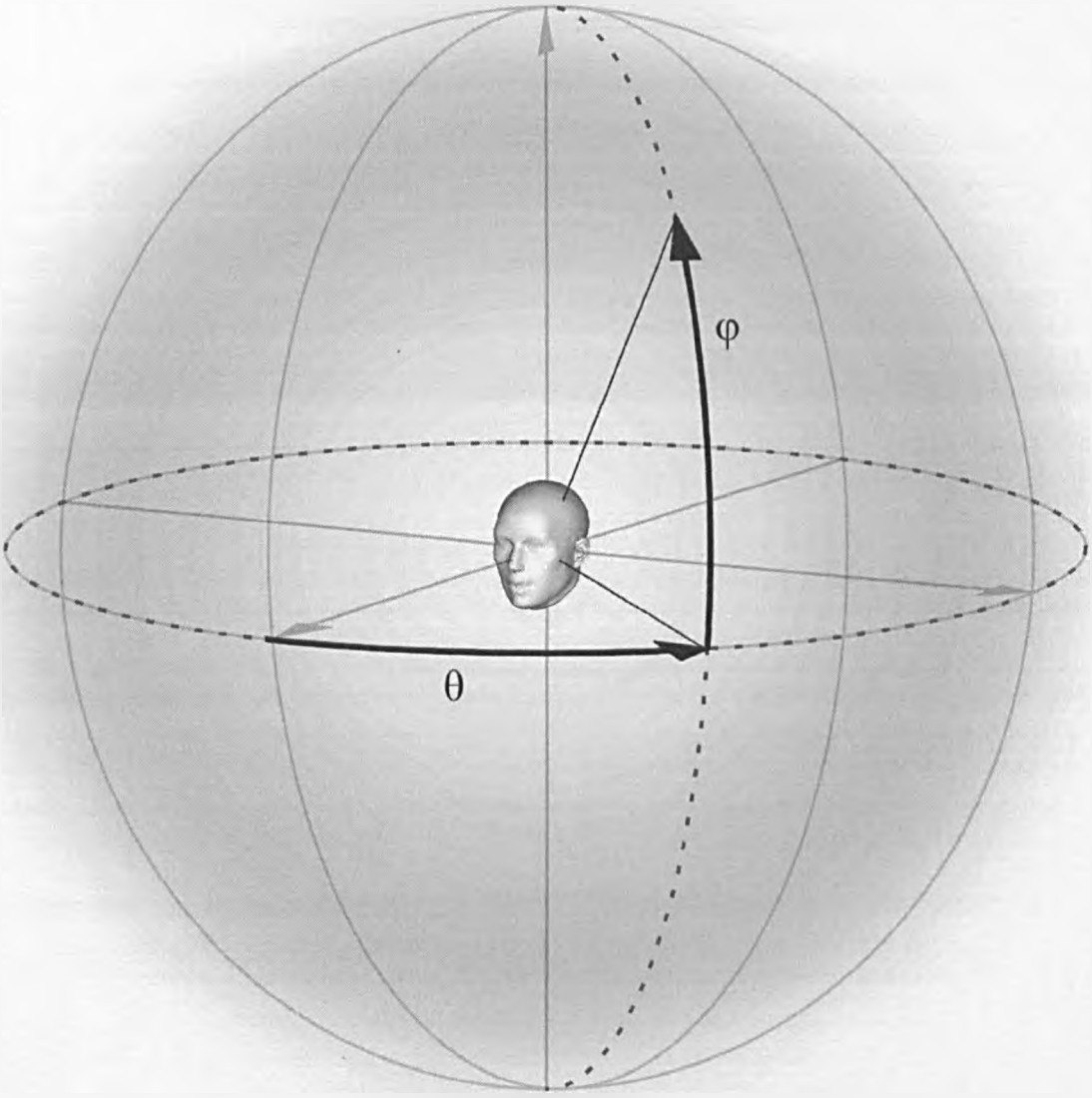


Figure 2.1: Vertical-polar spherical coordinate system. The dotted lines show the range of variation for the azimuth (θ) and elevation (φ).

systems have mainly been used to define source locations in the auditory space. The spherical vertical-polar coordinate system shown in Figure 2.1 is favoured by most researchers. In this system, azimuth (θ) ranges from $-\pi$ to

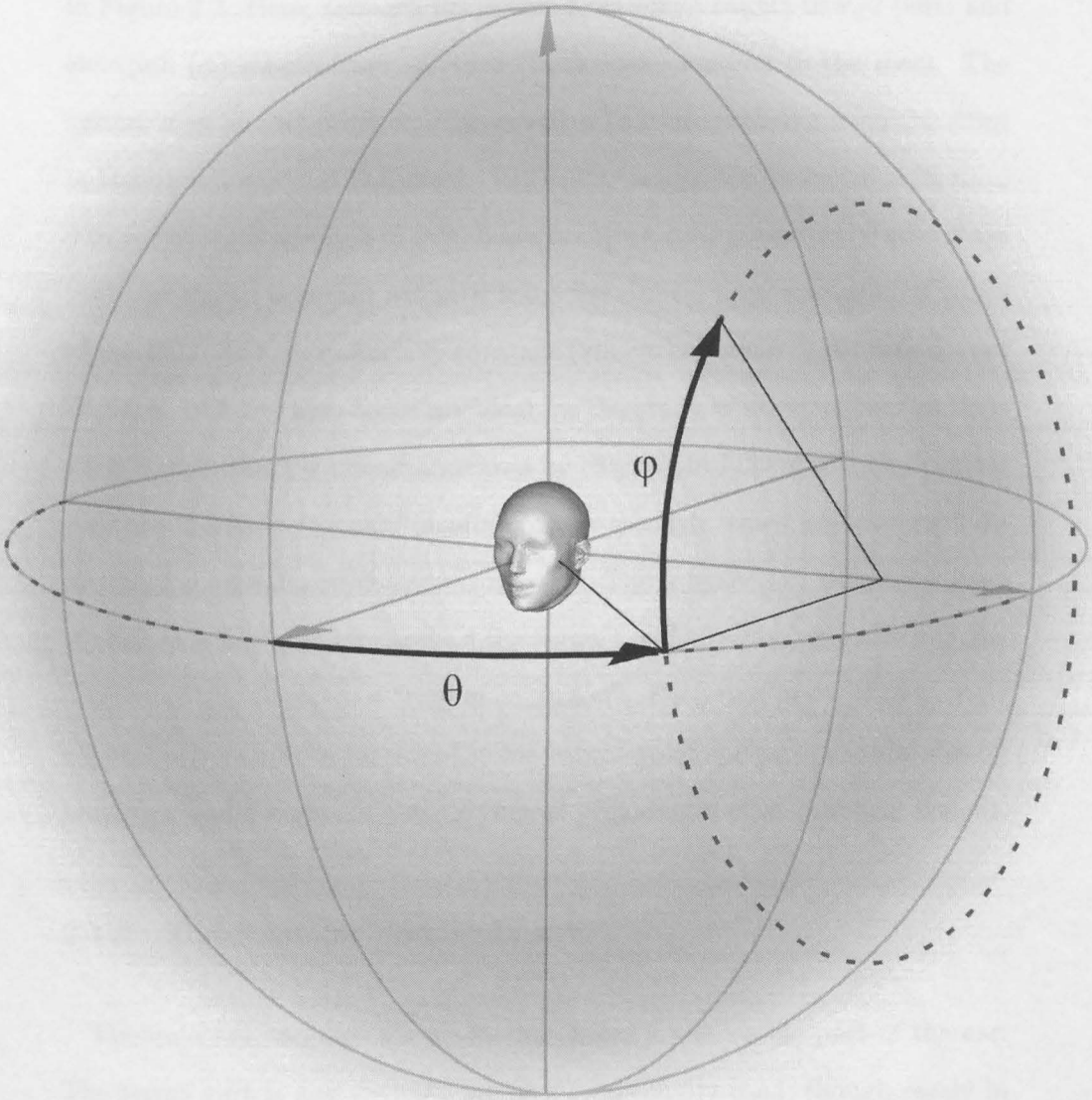


Figure 2.2: Interaural-polar spherical coordinate system. The dotted lines show the range of variation for the azimuth (θ) and elevation (φ).

π rad (both corresponding to the rear) and elevation (φ) ranges from $-\pi/2$ (bottom) to $\pi/2$ (top) rad, as shown by the dotted lines. The other coordinate system, used by Algazi *et al.* (2001c) and Brown and Duda (1998)

amongst others, is the spherical interaural-polar coordinate system shown in Figure 2.2. Here, azimuth (θ) ranges from $-\pi/2$ (right) to $\pi/2$ (left) and elevation (φ) ranges from $-\pi$ to π (both corresponding to the rear). The system may at first seem counter-intuitive (a source moving from the front to the back horizontal half-planes will see its associated elevation shift from 0 to π rad). It is adopted in this thesis, however, as it presents the advantage that surfaces of constant azimuth trace the conical approximation of zones where ITD/ILD are essentially constant (known as "cones of confusion", see Wallach, 1939). These cones are ideal for the study of spectral cues as their effect within them is uncontaminated by changes in ITD and ILD. Spectral cues are therefore the only possible means through which any observed directional source discrimination can occur. The intersection between a cone of confusion and a sphere around the listener will often be referred to in the context of this thesis as a "ring of confusion". Each ring lies within a single sagittal plane and can be traced in interaural-polar spherical coordinates by keeping r and θ constant and varying φ, providing a clear practical benefit.

2.1.3 External ear nomenclature

The external ear, also known as the pinna, is the visible part of the ear. The terms auricle and auricula are also occasionally used, though rarely in the context of binaural sound research. Its complex shape has brought about a sophisticated nomenclature. Figure 2.3 shows the nomenclature defined by Shaw (1997), which is generally accepted amongst specialists. The large, central hollow in the pinna, is referred to as the concha. It is a broad cavity, partially divided by the crus helias. Its lower part (referred to as cavum) is adjacent to the antitragus and ear canal (also known as the external acoustic meatus, or just, meatus) and its upper part (referred to as the cymba) is

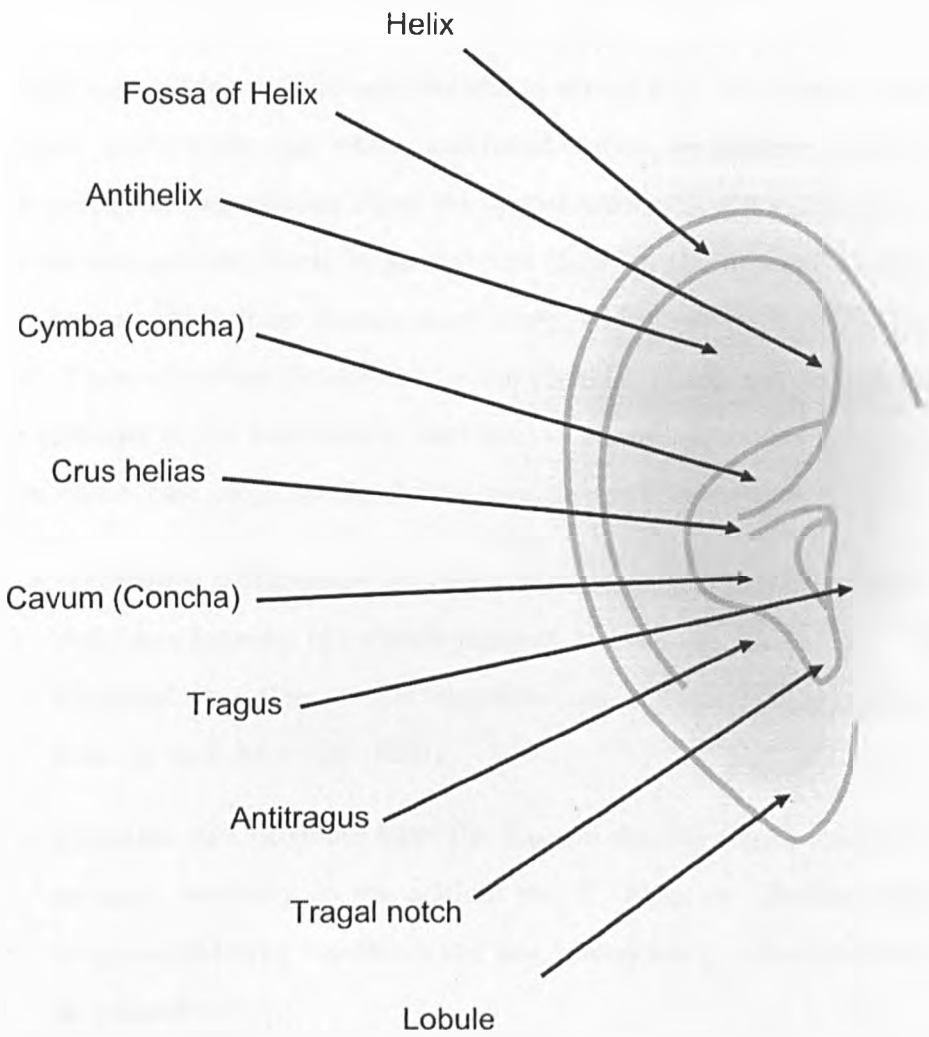


Figure 2.3: External ear nomenclature adapted from Shaw (1997)

bordered by the fossa of helix and the antihelix. These features form the bedrock of the acoustic localisation cue production mechanism, as will be described later.

2.2 Acoustic cues for sound localisation

Although it is known that acoustic effects arising from the external environment (early reflections, echoes, and reverberation, for example) can provide perceptual information about the spatial attributes of a sound source, this section and the thesis in general will focus on the acoustic localisation cues resulting from listener morphology, which require individualisation. These are primarily affected by the shoulders, head and pinnae and are captured in the individual's head-related transfer functions (HRTFs). Localisation cues are generally divided into two main categories:

- **Interaural differences** are cues whose perceptual power lies in the *difference* between the signals incident at each ear. Also known as **binaural** cues, they are the interaural time difference (**ITD**) and interaural level difference (**ILD**).

- **Spectral** cues originate from the spectral shaping imprinted by the auditory periphery on the original sound. They are effective under monaural listening conditions and are, consequently, often referred to as **monaural** cues.

There is some debate as to whether differences between spectral cues at both ears are perceptually salient (Jin *et al.*, 2004). These differences are referred to as interaural spectral differences (**ISDs**). Unlike monaural spectral cues, ISDs are robust to source spectral variation and potentially support the localisation of unfamiliar sounds (Searle *et al.*, 1975). The question of whether the increased perceptual salience of spectral cues in a binaural context is due to a reinforcement of the spectral cues at each ear or to the difference between them (ISD) is hard to settle experimentally and will

not be considered further in this thesis. Although the perceptual salience of the ISD is still uncertain, it is clear that spectral cues provide a huge amount of spatial information under monaural conditions (see Perrott and Saberi, 1990, amongst many others). It is possible that spectral cues operate both binaurally and monaurally. To avoid controversy the terms "monaural" and "binaural" will therefore be avoided when referring to spectral and ITD/ILD cues (respectively) as they both imply that spectral cues operate exclusively monaurally, which remains unproven.

2.2.1 Inter-aural time and level differences

2.2.1.1 ITD/ILD production mechanisms

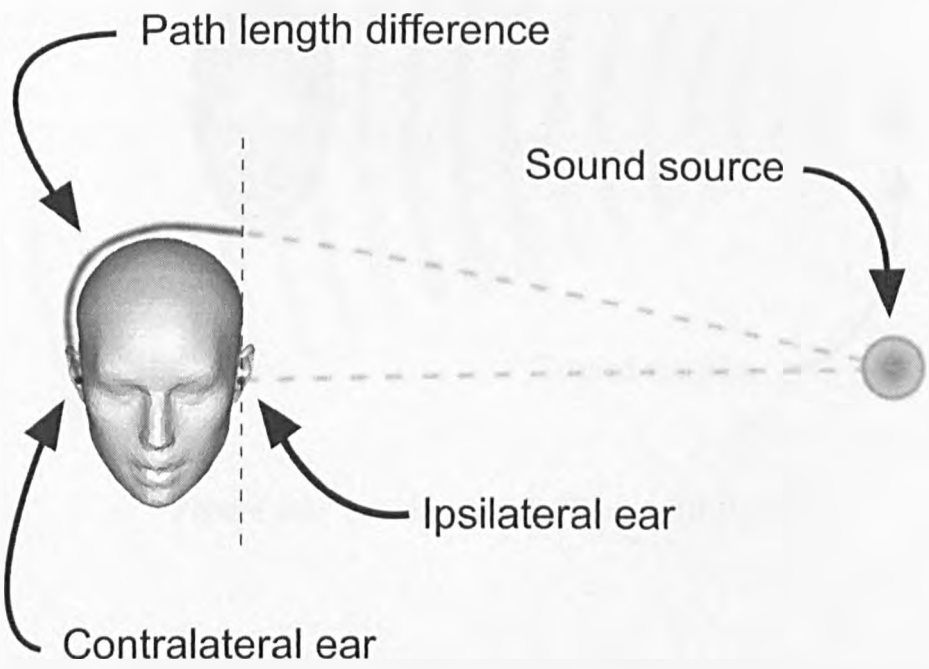


Figure 2.4: Interaural time difference (ITD).

The interaural time difference (ITD) originates from the distance between the two ears and the finite speed of sound. John Strutt, 3^{rd} Baron Rayleigh, better known as Lord Rayleigh, observed in the early 1900's that human listeners are sensitive to low frequency phase differences between the signals arriving at each ear from a common sound source. He inferred their perceptual role as localisation cues (see Rayleigh, 1907, and Figure 2.4). These phase differences are an inherent consequence of the ITD, which gives rise to the related term, interaural phase difference (IPD).

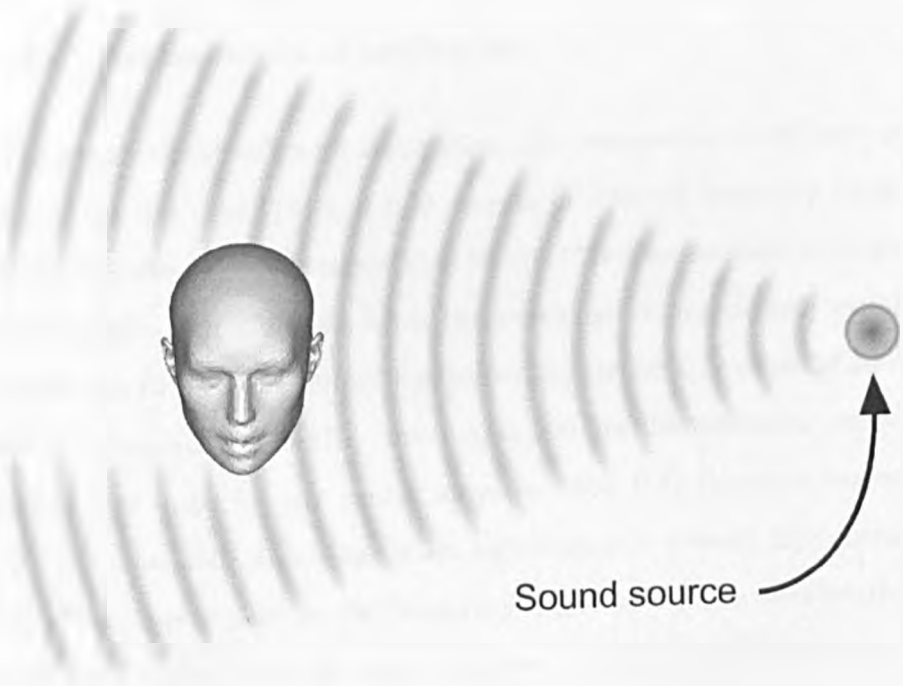


Figure 2.5: Interaural level difference (ILD).

Lord Rayleigh also observed that the acoustic shadowing effect caused by the head (see Figure 2.5) results in an attenuation of the signal received at the more distant (contralateral) ear relative to the closer (ipsilateral) ear. This effect causes the interaural level difference (ILD), also known

as interaural intensity or amplitude difference (IID and IAD, respectively). These cues are generally considered to be primary and overriding source azimuth indicators (Wightman and Kistler, 1992; Jin *et al.*, 2004). However they do not allow localisation within a cone of confusion, where they are by definition constant. The origin and effects of the ITD and ILD are well understood. Reliable techniques have been developed to extract them from HRTFs (see Kuhn, 1977, for example) as well as to estimate them given an appropriate set of morphological measurements.

2.2.1.2 Duplex theory of localisation

The duplex theory of sound localisation, also proposed by Lord Rayleigh, rests on the fact that ITD and ILD operate in different frequency ranges. The head-shadowing effect responsible for the ILD only becomes noticeable above 3-4 kHz approximately, when the wavelength is significantly smaller than the size of the head. As an illustration, maximum ILD values of 20 dB have been reported at 6 kHz, whereas at 200 Hz the difference scarcely departs from 0 dB for any source location. The ILD therefore becomes important for sounds with a significant high frequency content. By contrast, the ITD is mostly used in the frequency range where the wavelength is significantly longer than the size of the head so that no ambiguity exists as to whether the phase difference results from a lead or a lag between the signals at the two ears.

The first rigorous perceptual testing of the duplex model was performed by Stevens and Newman (1936) using pure tone stimuli. These tests showed a relatively constant localisation performance at low frequencies, followed by a sharp decrease above 2 kHz and an improvement above 4 kHz. This dip

in auditory acuity for tones between 2 kHz and 4 kHz cycles is an expected consequence of the duplex theory as this range is too high for ITD and too low for ILD to operate effectively. Other, more recent perceptual studies have supported this model to some extent (see Mills, 1958; Wightman and Kistler, 1992; Chandler and Grantham, 1992; Macpherson and Middlebrooks, 2002). It is now generally accepted, however, that although ITD is the dominant localisation cue at low-frequencies, its effects are not bound to them. At low frequencies the ITD is coded neurally through a process referred to as phase locking, whereby auditory neurons fire in phase with the acoustic pressure oscillations. This process does not occur at high frequencies, which exceed maximum possible neural firing rates. Instead, the auditory system exploits the ITD information present in the temporal envelopes of high frequency stimuli at both ears to estimate azimuth, although this process is less reliable than low frequency phase locking (Henning, 1974; Dreyer and Delgutte, 2006).

2.2.2 Spectral cues

It was initially suggested by Batteau (1967) that the localisation cues produced by the pinna are best described as time-domain transformations (multi-path reflections). However, important works presented by Shaw *et al.* (see Shaw and Teranishi, 1968; Shaw, 1974, 1997) have shifted the general understanding of these cues towards filtering operations in the frequency domain by considering their production in terms of resonances and modal behaviour inside the pinna acting to shape HRTF spectra (hence they are referred to as spectral cues). Since then, a number of studies have tended to confirm that elevation cues are spectral in nature.

Under binaural conditions, spectral cues leave azimuth perception (for which the more reliable and powerful ITD and ILD cues are very dominant) almost unaffected, but allow elevation perception which cannot be achieved using ITD/ILD cues (Roffler and Butler, 1968a; Hebrank and Wright, 1974; Watkins, 1978; Butler and Helwig, 1983; Middlebrooks, 1992; Martin *et al.*, 2006). However, under monaural listening conditions (where ITD/ILD cues are absent), spectral cues allow the location of broadband sources (elevation and azimuth) to be estimated with less than 5° error over the entire auditory space (Perrott and Saberi, 1990), a testament to the fact that the effects of spectral cues are not bound to elevation perception. The following sections will review key research efforts which have provided insight into their operation.

2.2.2.1 Spectral cue production mechanisms

Spectral cues originate from the acoustic effects in the auditory periphery; namely the torso, shoulders, head and pinnae. Together these create an acoustic filtering system with directionally dependent characteristics, which provide cues for localisation. The pinnae are especially remarkable in this respect due to the many different cavities and folds it is comprised of, which produce complex resonances and reflection patterns. Disturbing the natural filtering of the external ear by filling one of the pinna cavities has been shown to affect the accuracy of source elevation estimation and to increase front-back confusions (Musicant and Butler, 1984; Morimoto *et al.*, 2001). When this procedure was extended to both pinnae the resulting effect was greatly amplified (Humanski and Butler, 1988; Hofman and Opstal, 2003).

Shaw *et al.* carried out extensive experimental studies to investigate the acoustic characteristics of the external ear using physical measurements (Shaw and Teranishi, 1968; Shaw, 1974). Pinna resonances were analysed by rotating a sound source around an isolated pinna at the grazing incidence angle. In the interaural-polar coordinate system described in section 2.1.1, this is equivalent to $-180 < \varphi < 180$ and $\theta = 0$, with the origin shifted from the centre of the head to the pinna. The pinna was isolated by attaching a flat plate to the measured head. This reduced the diffraction of waves around the head. The pressure was measured with a probe microphone in the immediate region around the blocked meatus (ear canal). Modes were identified as amplitude maxima as the source frequency and source position were varied. Where modes were identified, further pressure amplitude and phase measurements were performed at the base of the concha and the fossa of helix. The concha was identified as a primary cue production area; a centre of modal activity. Its lower and upper parts (cavum and cymba, respectively) combine with each other as well as with the fossa of helix to produce a variety of acoustic modes (Shaw and Teranishi, 1968). Other structures extending from the concha, such as the helix, anti-helix and lobule have been shown to act collectively as a multi-delay system (Teranishi and Shaw, 1968). The production mechanisms behind spectral variations in HRTFs will be considered further in section 2.4.

2.2.2.2 Operating frequency range

In order to narrow down the spectral features within HRTFs which are potentially perceptually salient, a number of studies have investigated the localisation of narrow-band or band-limited stimuli (Blauert, 1970; Watkins, 1978; Butler, 1987; Humanski and Butler, 1988; Asano and Sone, 1990;

Middlebrooks, 1992, amongst others). Experimental results suggest that spectral cues operate over a substantial portion of the normal hearing frequency range. A reduction in source bandwidth causes a decrease in the acuity of elevation perception and an increase in front-back confusions (Roffler and Butler, 1968a; Hebrank and Wright, 1974; Butler and Belendiuk, 1977; Butler, 1986; King and Oldfiled, 1997). The absence of frequency components above 5 kHz leads to a non-elevated sound image (see Morimoto *et al.*, 2003b) indicating that spectral cues operate above this threshold. Indeed subjects perceived the sounds as originating from the horizontal plane. It has been shown that optimal median plane localisation will only be achieved for broadband signals with a frequency content extending up to 15 kHz (Hebrank and Wright, 1974; Langendijk and Bronkhorst, 2002; Best *et al.*, 2005). However, a number of studies suggest that spectral cues above 9 kHz, although present, are limited in effect and serve mainly to resolve front-back confusions (Morimoto *et al.*, 2003b; Bronkhorst, 1995). These studies showed that the bulk of the information, allowing localisation over the entire auditory space, is contained below 9 kHz.

2.2.2.3 Spectral peaks and notches

Spectral cues are ambiguous in some respect as they require the auditory system to make assumptions about the original source spectrum, which could contain features resembling spectral localisation cues causing localisation errors. However, HRTF spectral patterns have characteristically sharp, steep spectral features which rarely occur in the typically broad and smooth spectra of naturally occurring sound sources. Whether source locations are cued by specific spectral features or by overall spectral patterns is still a matter of conjecture. A number of studies have identified isolated spec-

tral peaks and notches as powerful localisation cues in and of themselves (Hebrank and Wright, 1974; Butler and Belendiuk, 1977; Bloom, 1977, for example).

As they are morphology-dependent these peaks and notches vary across individuals in shape, prominence and frequency. Despite this variability, particular spectral peaks and notches have been shown to be associated with similar perceptual effects across individuals. Both Hebrank and Wright (1974) and Langendijk and Bronkhorst (2002) have identified a one octave notch in the 5-8 kHz region and a prominent peak around 13 kHz as likely frontal cues, the latter being also associated with rear HRTFs. They also found a peak between 7 and 9 kHz combined with a high frequency cut-off at 10 kHz to be associated with high elevations. An increase in frontal elevation has been found by many studies to be associated with an increase in center frequency of the 1 octave notch from 5 to 11 kHz approximately (Hebrank and Wright, 1974; Shaw, 1974; Butler and Belendiuk, 1977; Bloom, 1977; Langendijk and Bronkhorst, 2002).

Investigations into the spectral edge sensitivity of neural circuits in the dorsal cochlear nucleus and its role in vertical sound localization were made by Reiss and Young (2005). They showed that some dorsal cochlear nucleus neurons (referred to as type IV) presented a sharp response peak when the rising spectral edge of a notch was aligned with certain frequencies. This study, amongst others, suggests that neural circuits contributing to sound localization are very sensitive to steep spectral slopes in a way that is comparable to the visual system, which relies heavily on edge and boundary processing. When spectral edges correspond to those observed in a particular individual's HRTF, these neural circuits identify the match, decoding spatial

attributes in the process.

2.2.2.4 Spectral variation with direction

The cues for localisation lie, it seems reasonable to assume, in the variation of HRTFs over the auditory space. The way in which the energy in a given frequency band varies as a function of source direction should therefore be taken into account when attempting to assess its relevance to the perception of source location. A spectral feature, no matter how prominent, cannot contribute to localisation if it does not vary with source direction. Conversely, an unremarkable spectral feature could find perceptual salience if it exhibits a marked directional variation. Analysing the variation of individual frequency bands with direction is therefore an important aspect of cue identification. It should be noted that the direction of maximum transmission for a given frequency-band does not necessarily correspond to a spectral peak in the HRTF for this direction. The term "covert peak" is therefore used to define a maximum as a function of direction for a given frequency band to avoid confusion with "spectral peak" which refers a maximum over frequency for a single direction. The covert peak area (CPA) is generally defined as the area around the source location that generates maximal sound pressure at the entrance of the ear canal for specific bands of frequency (Butler, 1987; Rogers and Butler, 1992; Butler and Musicant, 1993).

It has been argued that relatively high acoustic energy at a given frequency is likely to project the sound image toward the regions of space where the ear is most directional at that frequency (Musicant and Butler, 1984; Butler, 1987). In agreement with this suggestion Butler and Flannery

(1980) and Butler (1987) showed that the localisation of a narrow-band sound by monaural listeners is dependent on the centre frequency of the narrow-band noise rather than the actual location of the source. Also, Butler and Musicant (1993) showed that, although filtering particular frequency bands generally decreases localisation performance, the decrease is far worse when the sound originates from the covert peak area for the frequency band in question, providing further evidence for the spatially referent character of the CPA. A number of other studies have also reported that the center frequency of narrow band sounds affected perceived elevation (Roffler and Butler, 1968b; Blauert, 1970; Middlebrooks and Green, 1992).

2.3 Spatial sound perception

This section will focus on the psychophysics of spatial sound perception; the relationships between physical stimuli and their subjective correlates. It is important to give careful consideration to these relationships as they allow better informed attempts to identify the spectral variations which are salient to localisation. Firstly, perceptual tests which have given insight into the spatially varying acuity of source location perception across the auditory space will be reviewed. Secondly, studies investigating the limitations imposed by the neural encoding of sonic information are described along with techniques which have been designed to efficiently model the effects of encoding processes. Finally, studies which have attempted to assess the salience of individual spectral features through perceptual testing will be reviewed.

Source Azimuth	Horizontal localization error
0°	4.6°
15°	13.0°
30°	15.6°
45°	16.3°
60°	16.2°
75°	15.6°
90°	16.0°

Table 2.1: Horizontal localisation error as a function of azimuth, averaged for all pure tone stimuli frequencies and subjects, reported by Stevens and Newman (1936).

2.3.1 Variations in localisation acuity

2.3.1.1 Acuity of azimuth perception

Stevens and Newman (1936) used pure tones to perform the first significant study on localisation acuity for sources in the horizontal plane. Their results are presented in Table 2.1, averaged over all frequencies. Deviations from the average were small except for a dip in performance around 3 kHz, presumably due to neither ITD nor ILD being strong cues in this region (see Section 2.2.1.2).

In more recent studies, the acuity of auditory localisation has generally been measured in terms of the minimum audible angle (MAA). This term was original coined by Mills (1958) who defined it as the "smallest detectable difference between the azimuths of two identical sources of sound". This definition was later extended to include different planes of angular variations (sagittal and oblique for example). Reported MAAs vary with source direction, source spectrum and experimental setup. Most experiments performed with modern equipment generally agree on a best-case frontal horizontal MAA of approximately 1° (Mills, 1958; Grantham, 1986).

Another, similar, measure of localisation acuity is the minimum audible movement angle (MAMA), which has been investigated in a number of studies (Harris and Sergeant, 1971; Perrott and Musicant, 1977; Grantham, 1986; Perrott and Tucker, 1988; Perrott and Marlborough, 1989). A MAMA threshold is defined as the smallest angular distance a moving sound must traverse in the free field to be just discriminable either from a stationary source or from a source moving in the opposite direction. Compared with MAAs, MAMAs have the added variable of stimulus velocity. Harris and Sergeant (1971) reported that for sound sources moving very slowly in the horizontal plane, the MAMA is only slightly larger than the MAA. However, it increases with source velocity (Perrott and Musicant, 1977; Perrott and Tucker, 1988; Saberi and Perrott, 1990), giving rise to the notion of minimum integration time.

Recent psychophysical experiments measuring the MAA are in general agreement with the early studies of Stevens and Newman (1936), reporting decreasing acuity as azimuth increases from 0° to 90° (Mills, 1958, for example). Chandler and Grantham (1992) reported a similar pattern of deterioration when investigating MAMA variation with azimuth. Grantham (1986) used a stereo speaker arrangement and crosstalk cancellation to emulate source movement. He came to the same conclusion, save a very abrupt deterioration for stimuli beyond 90° azimuth, which is arguably due to the fact that the stereo speaker arrangement could not produce convincing sound images in the back hemisphere. In both MAA and MAMA studies, increased source bandwidth generally improved localisation performance. In general, investigating localisation performance through the study of MAMAs or MAAs leads to similar conclusions, prompting Perrott and Tucker (1988) to suggest that "both static and dynamic spatial discrimination func-

Source Elevation	Vertical localisation error
0°	9°
30°	10°
74°	13°
112°	22°
153°	15°

Table 2.2: Vertical localisation error in the median plane, as a function of elevation, reported by Damaske and Wagener (1969).

tions are dependent upon the same underlying mechanism".

2.3.1.2 Acuity of elevation perception

The acuity of elevation perception can be measured using sagittal MAA, which requires extending Mills' definition (see Section 2.3.1.1) to apply to elevation changes as well as azimuth changes. Minimum audible angle thresholds obtained for changes in elevation are generally two to four times larger than those observed in the horizontal plane when broad-band stimuli are employed. Perrott and Saberi (1990) reported a mean MAA threshold of 0.97° in the horizontal plane and a mean MAA threshold of 3.65° in the vertical plane. These results are consistent with earlier findings by Wettschurek (1973), Morrongiello and Rocca (1987) and Blauert (1997) who all report vertical MAAs around $3 - 4°$.

There is little experimental data on the effect of elevation on sagittal auditory resolution. Test results reported by Damaske and Wagener (1969) showed that vertical localisation error in the median plane increases with elevation (see Table 2.2). This data shows a clear decrease in performance with increasing elevation, a result also reported by Wettschurek (1973). Median plane localisation performance seems to reach a minimum around 120°

and then improves as the source nears the back of the horizontal plane.

Variations in the acuity of elevation perception around rings of confusion away from the median plane (which lie within single sagittal planes) have also been investigated. Leung and Carlile (2004) found that the variation pattern was consistent across rings of confusion. In all cases, the best performance was observed for locations in front ($\varphi = 0°$) with a minor decrease for locations behind the subjects ($\varphi = 180°$) and a large decrease in performance (by a factor of 2 to 5 depending on subjects) for locations above the subjects ($\varphi = 60°$ and $\varphi = 120°$).

2.3.2 The limits of human sound perception

The information contained in perceived auditory signals is intrinsically shaped and ultimately limited by the psychophysical and physiological aspects of their neural encoding. A number of studies have shown that spectral variations present in the sound incident on the eardrum are smoothed during neural encoding (Patterson, 1976; Patterson and Moore, 1986; Moore *et al.*, 1989, amongst others). This observation has led to hearing models based on a series of band-pass filters (Patterson *et al.*, 1992; Moore *et al.*, 1990; Glasberg and Moore, 1990). The way in which the width and shape of these band-pass filters change as a function of frequency has been determined in the case of human hearing in a number of psychophysical studies. These trace spectral excitation patterns by analysing the ability of subjects to discriminate tones within notched noise (Patterson, 1976; Moore and Glasberg, 1987). These models have been shown to provide a good, although coarse, representation of the neural representation of sonic input as it reaches the primary auditory cortex, where neural processes responsible

for sonic awareness take place. Indeed, the quantitative and predictive character of this approach allows its validity to be perceptually evaluated and excitation patterns have been used to predict the limits of partial resolution in complex tones with good success (see Moore and Glasberg, 1987, for example).

A large body of experimental evidence acquired through careful perceptual studies suggests that the width of the bandpass filters comprising hearing models increases exponentially with frequency (Patterson, 1976; Moore and Glasberg, 1987; Moore, 1989; Glasberg and Moore, 1990). This results in the normalised filters being constant in shape across the audible spectrum when plotted on a logarithmic scale. The effects of spectral smoothing therefore also remain constant when plotted logarithmically. This is consistent with the quasi-logarithmic nature of pitch and intensity perception. The output of each filter is determined by the total energy falling within its pass-band and the smoothed spectrum results from the outputs of all the filters across the audible frequency range (Glasberg and Moore, 1990). HRTFs possess a large amount of fine spectral detail which is lost when captured within the bandwidth of a single auditory filter and the perceptible spectral resolution inherently suffers thereafter. This smoothing has been shown to substantially reduce HRTF variations across space and subjects (Carlile and Pralong, 1994).

2.3.3 Gammatone filter-bank models

Patterson *et al.* (1992) have shown that the impulse response of the gamma-tone function of order 4 accurately fits the human auditory filter shapes derived by Patterson and Moore (1986). As a result, gammatone

filters are commonly used by researchers to efficiently model the neural encoding of sound in the auditory system. The corresponding time-domain function of a gamma-tone filter is

$$g(t) = at^{n-1}\cos(2\pi f_c t + \varphi)e^{-2\pi bt} \tag{2.1}$$

where n is the order of the filter, b its bandwidth, f_c its center frequency, a is amplitude and φ is phase. The equivalent rectangular bandwidth (ERB) scale (see Moore *et al.*, 1990; Shailer *et al.*, 1990) was developed to align the frequency dependent bandwidth of gammatone filters with that measured in human listeners. It is essentially a psychoacoustic measure of auditory filter bandwidth at each frequency and presents logarithmic behaviour over the audible frequency range. Glasberg and Moore (1990) defined the ERB scale by the equation

$$ERB(f) = 24.7\left(\frac{4.37f}{10^3} + 1\right) \tag{2.2}$$

An efficient implementation of gammatone filter banks is available in the Auditory Toolkit[1] developed by Slaney (1993) who defines a general, adjustable form for the ERB scale as

$$ERB(f, EarQ, minBW, n) = \left(\left(\frac{f}{EarQ}\right)^n + minBW^n\right)^{1/n} \tag{2.3}$$

where $EarQ$ defines the filter selectivity (quality or Q factor) reached asymptotically at large frequencies, $minBW$ is the required minimum bandwidth at low frequencies and n is the order. Expressed in this form, the parameters

[1]Url: `http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/`

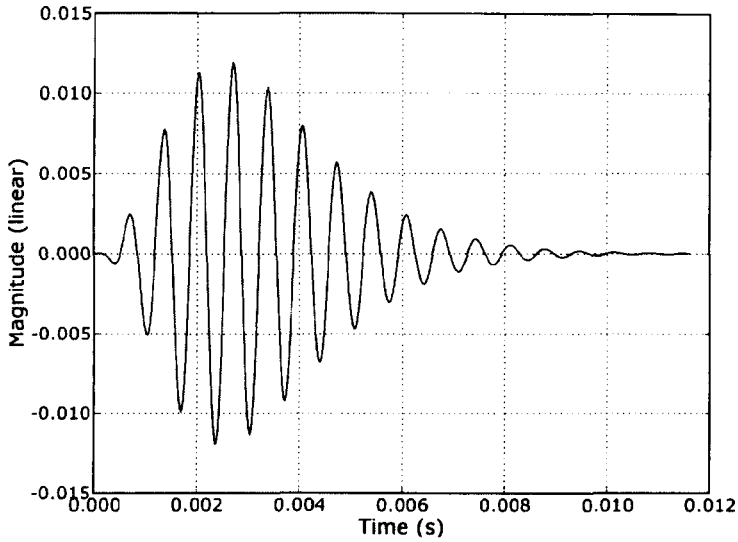
for the ERB scale proposed by Glasberg and Moore (1990) are:

$$\begin{aligned} EarQ &= 9.26449 \\ minBW &= 24.7 \\ n &= 1 \end{aligned} \tag{2.4}$$

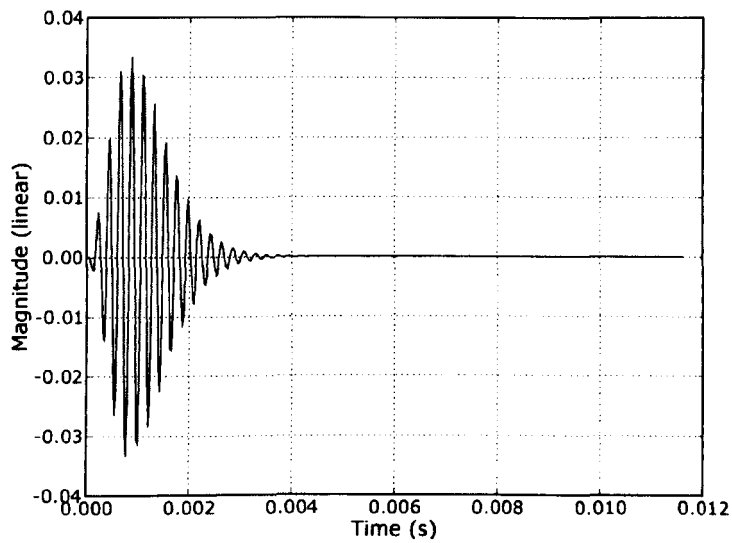
Lyon (1982) and Greenwood (1990) proposed slightly different filter banks which can be described by altering these parameters (see Slaney, 1988). Patterson *et al.* (1992) recommend setting the bandwidth (b in Equation 2.1) of the gamma-tone filters to 1.019 times the ERB. Figure 2.6 shows the time-domain filter coefficients for different gamma-tone filter center frequencies in a normalised filter bank as implemented by Slaney (1993), following the recommendations of Patterson *et al.* (1992) and setting the filter parameters defined by the set of Equations 2.4. As the center frequency of the gammatone filter increases, the impulse response duration decreases and its amplitude increases.

The frequency spacing between neighbouring filters required for accurately modeling human perception is unclear. In the cochlea there are many thousands of hair cells, but computer models can only approximate this density of channels. The filter bank implementation proposed by Slaney (1993) allows the highest and lowest frequencies, along with the desired number of channels, to be specified. From these parameters, the filter centre frequencies are calculated. The Auditory Toolkit calculates the center frequency of the nth gammatone filter in an N channel filter bank using (from Slaney, 1993)

$$f_n = -A + (f_h + A)e^{n(-\log(f_h+A)+\log(f_l+A))/N} \tag{2.5}$$



(a) 1479 Hz center frequency



(b) 4555 Hz center frequency

Figure 2.6: Gammatone filter impulse responses produced using the Auditory Toolkit implementation (Slaney, 1993). As frequency increases the length of the response decreases and its magnitude increases.

where $A = EarQ \times minBW$, f_l is the lowest frequency, f_h the highest

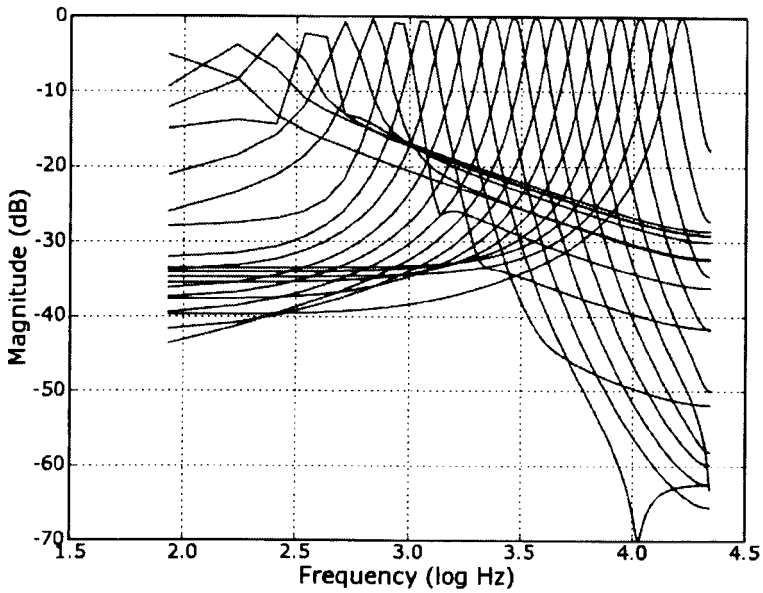


Figure 2.7: 20 channel gammatone filter bank covering the 100 Hz - 20 kHz frequency range produced using the Auditory Toolkit.

frequency. Figure 2.7 shows a bank of 20 filters produced using the Auditory Toolkit, spread over frequencies ranging from 100 Hz to 20 kHz. The filter characteristics are presented in the frequency domain, frequency points being spaced linearly over frequency. This results in a visible improvement of the filter pass-band shape definition as frequency (and consequently, filter bandwidth) increases. Figure 2.8 shows the frequency domain character of a complete 20 channel gammatone filter bank obtained by summing the impulse responses of all the individual filters. The visible ripples are a result of the low density of the filter bank. As the density increases the ripples disappear as shown in Figure 2.9 which shows the overall response of a 128 channel gammatone filter bank produced for the same frequency range using the same implementation.

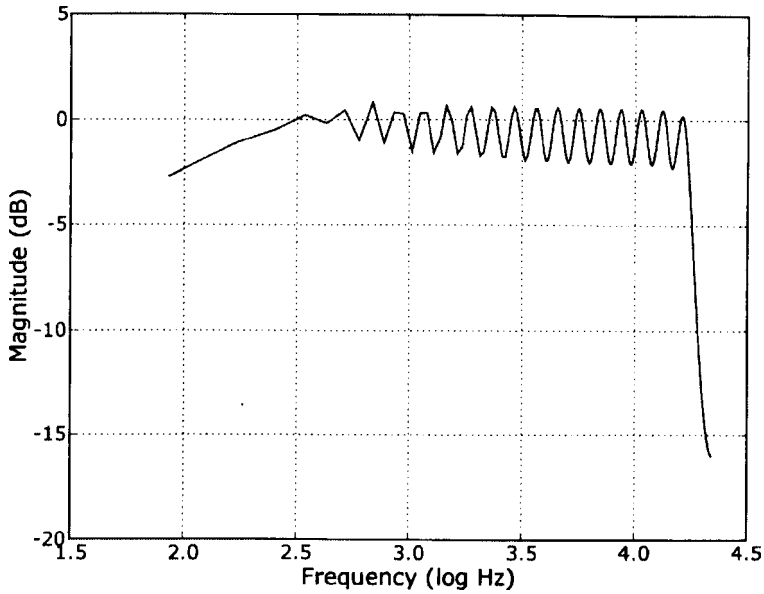


Figure 2.8: Overall response of a 20 channel gammatone filter bank covering the 100 Hz - 20 kHz frequency range produced using the Auditory Toolkit.

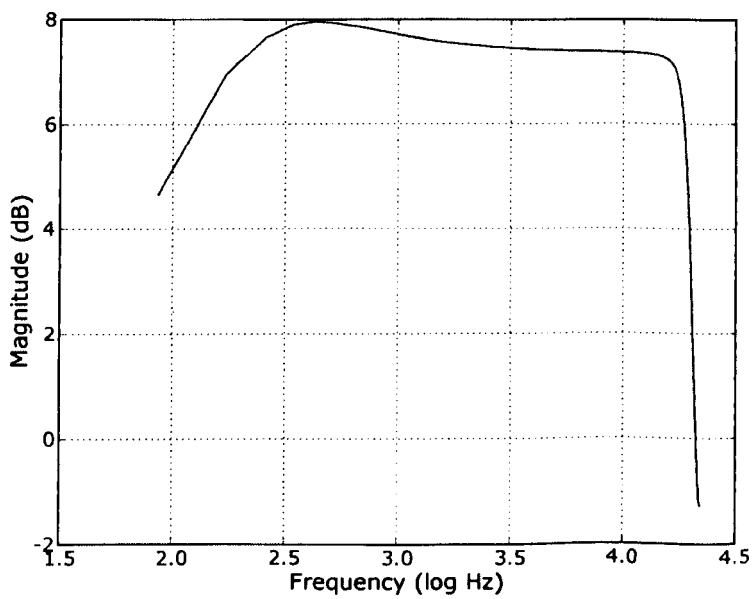


Figure 2.9: Overall response of a 128 channel gammatone filter bank covering the 100 Hz - 20 kHz frequency range produced using the Auditory Toolkit.

2.3.4 Perceptual approaches to cue identification

It has been shown that the accuracy of source elevation estimates significantly decreases when the frequency content of broadband stimuli is scrambled (Macpherson, 1996; Wightman and Kistler, 1997; Macpherson and Middlebrooks, 2003) and when the content of certain frequency bands is attenuated or removed (Hebrank and Wright, 1974; Butler and Musicant, 1993; Burlinghame and Butler, 1998). These studies confirm the spectral nature of elevation cues. Other studies have attempted to trace the effect of specific spectral features. Langendijk and Bronkhorst (2002), for example, masked selected portions of HRTF spectra ranging from a half-octave to 2 octaves and analysed the resulting change in localization performance. They noted that the masking of half-octave spectral features and narrower did not seem to affect localisation performance in conflict with Hebrank and Wright (1974) who identified spectral features as narrow as a quarter-octave as prime indicators of source elevation in the 7-9 kHz range. Langendijk and Bronkhorst (2002) reports that the effects of 1 and 2 octave spectral feature masking indicated that most up/down cues are present in the 5.7-11.3 kHz region and most front-back cues in 8-16 kHz region. Perceptual test performed by Bronkhorst (1995) seem to be compatible with these observations as the percentage of front-back confusions in low passed broadband stimuli location increased significantly when the cut-off frequency was reduced from 16 to 7 kHz comparing to a relatively small increase in up-down confusions.

Some studies have attempted to estimate the perceptual relevance of isolated peaks and notches to sound localisation by measuring detection thresholds (see Moore *et al.*, 1989, for example). These results should be interpreted very carefully as low detection thresholds are not necessarily a sign

of greater participation in localisation processes. Perceptual awareness is a result of activity in the primary auditory cortex, a very late stage of the auditory pathway, whereas the extraction of sound location information occurs as early as the dorsal cochlear nucleus. The intensive neural processing occurring between these two stages, which will be described further in Section 3.4 suggests the two processes should be considered independently. Moreover, it has been argued that a "re-equalisation" process, whereby the brain compensates to some extent for the sharp spectral features introduced by the pinna so as to minimise changes in perceived sound timbre with source location is possible (Jin, 2001). This would suggest that the intermediate neural processing may actually act to prevent the detection of spectral features used for localisation. Low detection thresholds for changes in a given spectral feature would then counter-intuitively indicate its lack of relevance to the sound localisation process.

Some results are compatible with this suggestion. Moore *et al.* (1989) note that the detection of changes in peak height or notch depth is generally less good at 8 kHz (at the heart of the range of cue operation) than at 1 kHz. Also, thresholds for the detection of peaks and notches were shown to increase markedly with decreasing bandwidth, which implies that detection thresholds are higher when the spectral slope is steeper. It should also be noted that investigating the detection thresholds and perceptual effects of single spectral peaks or notches in a flat-spectrum background would only be satisfactory under the assumption that their combined perceptual effect can be inferred from those measured in isolation. There are no grounds for such an assumption. Intense sound location sensation over the entire auditory space cannot be achieved using only the most prominent peaks and notches, although these are enough to achieve some sense of sound direction, as shown

by Blauert (1970), Hebrank and Wright (1974), Bloom (1977) and Butler and Helwig (1983). More subtle HRTF features are necessary to achieve observed localisation acuity. Kistler and Wightman (1992) described the location-dependent components of HRTFs (referred to as directional transfer functions or DTF) using a principal component analysis. They showed that the smoothing of spectral detail resulting from component truncation causes an increase in the number of front-back confusions and errors in elevation estimation, which is a clear indication that finer HRTF structures play a role in the perception of sound location. All these factors make it difficult to justify discarding subtle spectral features as localisation cues based on perceptual tests.

2.4 HRTF Estimation

The need for affordable, efficient and effective HRTF estimation is the central motivation for the work reported in this thesis. This section reviews previous research efforts carried out with this objective in mind. Approaches to shape description are considered, followed by an overview of mathematical models, structural decompositions and statistical analyses. Finally, studies which have made use of acoustic simulations to calculate HRTFs will be summarised.

2.4.1 Head/Pinna Shape Description

2.4.1.1 Landmark measurements

However significant shape variations are, a clear underlying morphological structure allows a large number of landmarks to be identified across

individuals. Studies investigating the relationships between morphology and localisation cues have often relied on simple landmark measurements as shape descriptors. Algazi *et al.* (2001c) compiled morphological landmark measurements for a group of subjects along with corresponding HRTF measurements to create the CIPIC database. The landmark measurements contained in this database are shown in Figure 2.10.

Although these measurements give a measure of variation across individuals, the landmarks chosen are inevitably somewhat arbitrary, largely based on well-informed guess work. Even a large set of landmark measurements does not guarantee that all the relevant shape detail will be recorded. This is particularly crucial in the case of the external ear. As a consequence, problems have been encountered when attempting to correlate shape and acoustic variations as is further described in Section 2.4.4.

2.4.1.2 Shape parameterisation

To address the need for a more comprehensive shape description technique, Hetherington *et al.* investigated the application of the elliptic Fourier transform (EFT) for parameterising the human head, including the pinnae (see Hetherington and Tew, 2003; Hetherington *et al.*, 2003). The EFT is an established mathematical process which has been used in a number of other fields (Kuhl and Giardina, 1982; Park and Lee, 1987; Stevenson *et al.*, 1987; le Minor and Schmittbuhl, 1999; Wu and Sheu, 2001; Cheong, 2001).

The first step in the EFT process is to express the three-dimensional surface of interest as a set of two-dimensional contours (or slices). Hetherington's slices are obtained from the intersection of the surface with a plane

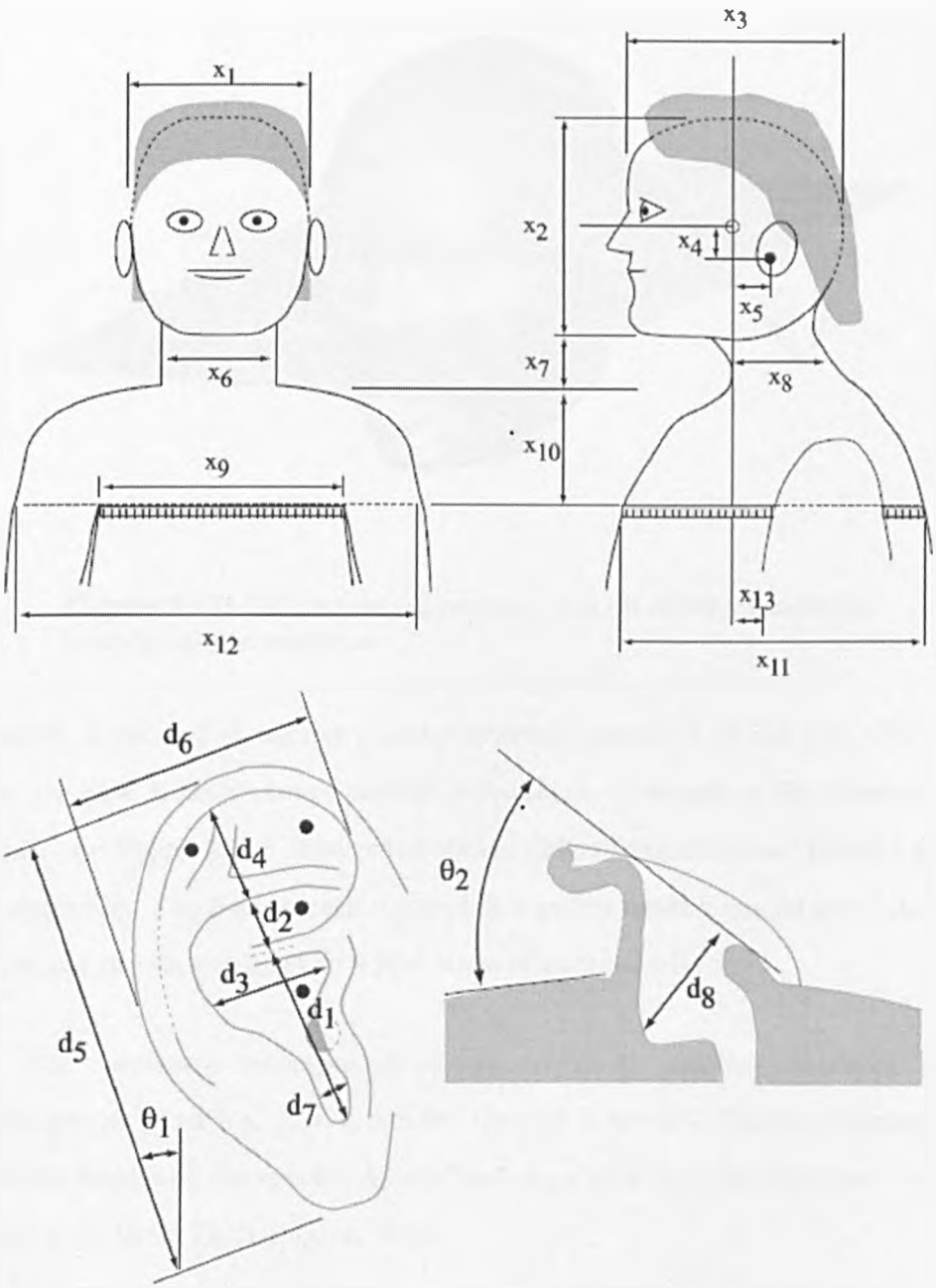


Figure 2.10: Landmark measurements used by Algazi *et al.* (2001c) for the CIPIC database. Height, seated height, head circumference and shoulder circumference were also measured; a total of 27 measurements.

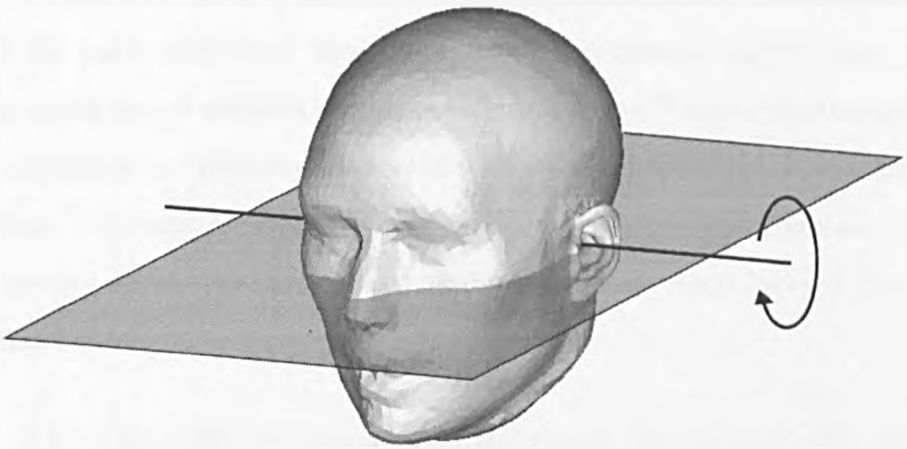


Figure 2.11: Slicing process resulting in a set of two dimensional mesh/plane intersections

which is rotated at regular angular intervals around a slicing axis which, in the case of head/pinnae parameterisation, corresponds to the interaural axis (see Figure 2.11). The radial slicing differs from the usual parallel arrangement. The S slices each containing T points equally spaced along their contour are then subject to a first stage of parameterisation.

The parametric x-component of slice s, $f_x[s,t]$, and the parametric y-component of slice s, $f_y[s,t]$, are fed through a forward Fourier transform, which results in the spectra $A_x[s,n]$ and $A_y[s,n]$ defined in Equations 2.6a and 2.6b (from Hetherington, 2004).

$$A_x[s,n] = \sum_{t=0}^{T-1} f_x[s,t] \mathrm{e}^{-j2\pi nt/T} \tag{2.6a}$$

$$A_y[s,n] = \sum_{t=0}^{T-1} f_y[s,t] \mathrm{e}^{-j2\pi nt/T} \tag{2.6b}$$

where n is the first EFT stage spectral component index and t is the index of the point along each slice. The spectral coefficients $A_x[s,n]$ are then arranged into N complex signals of length S. The n^{th} signal represents the variation of the coefficient associated with the n^{th} spectral component across slices. The same re-arrangement is applied to the $A_y[s,n]$ coefficients. The resulting signals are then separately fed through a second forward Fourier transform.

$A_x[s,n]$ and $A_y[s,n]$ originate from real signals and so exhibit Hermitian symmetry. Consequently, the second transform stage only needs to be performed for spectral components under the Nyquist frequency ($N/2$). The outputs of the second stage, $B_x[m,n]$ and $B_y[m,n]$ are (from Hetherington, 2004).

$$B_x[m,n] = \sum_{s=0}^{S-1} A_x[m,n]\mathrm{e}^{-j2\pi ms/S} \tag{2.7a}$$

$$B_y[m,n] = \sum_{s=0}^{S-1} A_y[m,n]\mathrm{e}^{-j2\pi ms/S} \tag{2.7b}$$

where m is the second EFT stage spectral component index. $B_x[m,n]$ and $B_y[m,n]$ are the final EFT parameters. The strength of the EFT lies in the concentration of energy in the lower components of the first and second stage transforms. This allows parameter sets to be truncated while retaining most of the shape information (investigated by Hetherington, 2004).

2.4.2 Structural approaches

One approach for simplifying the analysis of acoustic processes responsible for generating spatial HRTF variations is to treat the components of the auditory periphery as independent acoustic entities. The entire system is then described as a combination of filters, each of which accounts for the contribution of a different morphological structure. This structural approach was originally proposed by Genuit (1984). Each filter block is adjusted according to the morphological specificities of the structure it represents in order to achieve individualisation.

Algazi *et al.* (2001b) justify investigating the acoustic effects of an isolated pinna-less spherical approximation of the head as a structural component. They suggest that these effects could be superimposed on those of isolated pinnae arguing that "a spherical head model can be used to estimate the sound field in the vicinity of the pinna" and that "a pinna model can then be used to estimate the sound field that enters the ear canal" (Algazi *et al.*, 2001b). They conclude that "the HRTF might be represented by a cascade of two filters, one for the head and the other for the pinna" (Algazi *et al.*, 2001b). Their work shows that the acoustic effects of the head, torso and pinnae can be successfully isolated and recombined. Structural models described by Brown and Duda (1997), Brown and Duda (1998) and Gupta *et al.* (2004), amongst others, attempt to break the problem down further by applying the same principle to different acoustic systems within the pinna. However, this approach has failed to model measured acoustic phenomena effectively. Decomposing the HRTF into a set of isolated anatomical structures discounts the acoustic interactions between these structures. Although this approximation appears to be valid for the head, shoulders and torso, it

has proved inadequate when applied to the many different resonance cavities observed within the pinna, where acoustic effects are heavily inter-related and overall behaviour cannot be inferred from a superposition of isolated sub-systems.

2.4.3 Mathematical models

Acoustic theory allows sound fields around relatively simple three-dimensional objects, like spheres and ellipsoids, to be modelled analytically. Duda and Martens (1998) used mathematical modeling in the context of HRTF estimation to describe the sound field around a spherical approximation of a head. In later work they refined the model by using a combination of a sphere and an ellipsoid to approximate the low frequency acoustic effects of the head and torso (Avendano *et al.*, 1999; Algazi *et al.*, 2001a, 2002; Algazi and Duda, 2002), the so-called snowman model. Although these models succeeded in their stated objective, they offered little hope of extensions that would incorporate the higher frequency pinna effects.

Modeling external ear acoustics has been attempted using a quite different approach describing them in terms of multi-path pinna reflection patterns (see Batteau, 1967; Watkins, 1978). These models have provided valuable insights. Reflections from the posterior wall of the concha have, for example, been shown to be at least partly responsible for the production of the elevation-dependent spectral notch in the median plane (see Shaw and Teranishi, 1968; Rodgers, 1981). However, these models have been criticised as oversimplifications and were unsuccessful in modeling the more detailed aspects of external ear acoustics. For example, Lopez-Poveda and Meddis (1996) note that similar elevation-dependent notches observed for

incident sounds from the frontal plane (Shaw and Teranishi, 1968; Bloom, 1977; Carlile and Pralong, 1994) cannot be attributed to concha reflection processes, as the concha posterior wall is parallel to the direction of incidence.

Lopez-Poveda and Meddis (1996) presented an approximate physical model of the human concha in the form of a spiral-shaped cylinder. The model includes diffraction, reflection, and interference effects in an attempt to account for some of the less well understood spectral features observed in measured HRTFs. They pay particular attention to the characteristic elevation-dependent notch and the role of the concha. Although this model offered some insight there were still significant discrepancies between modelled responses and measured ones. These were attributed to the combined acoustic effects of other external ear regions, such as the helix, tragus and ear canal. Repeated attempts to analyse pinna acoustics by isolating its constituent parts have all proved unsuccessful suggesting that the external ear should be studied as a whole if its acoustic properties are to be better understood.

2.4.4 Statistical analyses

Statistical analyses of the relationships between morphological descriptors and localisation cues have been the focus of a number of research efforts aiming to efficiently derive pinna acoustics from shape descriptors without the need for complex mathematical modeling. These studies attempt to extract relationships between inter-subject shape and HRTF (more specifically, localisation cue) variations. Algazi *et al.* (2001c) created a large database of morphological landmark measurements (see Section 2.4.1.1) and analysed correlations between these measurements and the centre frequency of

the pinna notch in corresponding HRTFs. They identified the cavum concha height, pinna rotation angle, pinna flare angle and fossa height as best (although still relatively poor) predictors. Rodriguez and Ramirez (2005) investigated this morphoacoustic database further, extracting a set of pinna-related transfer functions (PRTFs) from the HRTFs and modelled them using principal component analysis (PCA). They then calculated correlations between these components and the original landmark measurements. They also generated new morphological parameters, derived by algebraically combining the original ones, in an attempt to strengthen correlations. In both cases the results were disappointing; correlation coefficients remaining low. It is commonly acknowledged that a deeper understanding of cue production mechanisms is required (Brown and Duda, 1998; Algazi *et al.*, 2001c).

2.4.5 Perceptual evaluation of estimation techniques

Zotkin *et al.* (2003) investigated the perceptual effects of modifying generic HRTFs using the results reported in some of the aforementioned models. The incorporation of low frequency head, shoulder and torso effects consistent with the subjects' morphology was shown to improve elevation perception significantly. This provided support for the validity of the model proposed by Algazi *et al.* and testified to the importance of shoulder reflections and torso shading for the acuity of elevation perception for sources outside and inside the torso shadow, respectively. However, the effects of replacing generic HRTFs with those for which the CIPIC landmark measurements most closely fitted the listener's own morphology were disappointing. Although the acuity of elevation perception was marginally improved in some cases, neutral and even negative effects were also observed.

2.4.6 Acoustic simulations

2.4.6.1 The boundary element method

The boundary element method (BEM) is a common engineering tool, a detailed introduction to which is provided by Brebbia and Dominguez (1992). It provides a numerical solution to a boundary integral equation formulation of a problem. In acoustics it can be used to produce a frequency-dependent description of the acoustic field around an object. The surface of the object is discretised into a number of planar elements (the boundary elements) referred to as *patches*, *polygons* or just *polys* which define the connexions between *points* (also known as *vertices*) forming a *mesh* description of the object. Each BEM simulation calculates surface pressures on the mesh created by a point source at a single frequency. Far-field pressures can then be calculated from these results. HRTF calculations using the BEM are made on the assumption that only the outer surface of the auditory periphery needs to be modelled and that internal effects are negligible.

Acoustic simulations performed using the BEM have already been used to investigate the acoustics of the human head and pinnae. Weinrich (1984) was the first to attempt to model the acoustics of the external ear. The low computational power at his disposal however, only allowed a mesh containing 20 patches, which produced a very coarse approximation to the shape of the concha. Later, improvements in computational power allowed Katz to perform acoustic simulations on a denser mesh up to 5 kHz (Katz, 1998, 2001a,b), which still fell short of covering the range of frequencies over which pinna acoustics are known to produce localisation cues. More recently, results obtained by performing BEM-based simulations on high-resolution

meshes have shown good agreement with acoustic measurements across the audible frequency range (Walsh *et al.*, 2003; Otani and Ise, 2006; Kahana and Nelson, 2005, 2007).

Although the BEM is a very powerful tool which allows HRTFs to be calculated accurately, software implementations come at very high cost and full HRTF sets take hours to calculate even with powerful computers. Simulations also require a mesh of the outer surface, which is expensive, time-consuming and technically challenging to produce in the case of the human pinna. These practicalities prohibit the widespread use of BEM algorithms, which would make fully spatialised sound more accessible.

2.4.6.2 Capturing a surface description of the human head and pinnae

One of the main challenges to be overcome in acoustic simulation is the difficulty of obtaining an accurate computer model of the object of interest. The complex shape of the external ear makes a complete shape capture demanding in terms of the hardware needed. Katz (2001a) obtained a raw mesh using a Cyberware 3-D laser scanner. He concedes that laser scanning presented a major disadvantage in that, "due to the linear design of the system, the data were restricted to line-of-sight data in a radial direction about the head" (Katz, 2001a). As a result, areas such as the space behind the ears and detailed folds of the ears were solid-filled.

Otani and Ise (2006) analysed the effects of such approximations by comparing the BEM results of meshes acquired using a similar laser scanner and a micro-CT (computerised tomography) scanner. They observe that the

frontal HRTF calculated for the meshes obtained from CT and laser scans were consistent below 4 kHz. However, significant discrepancies were observed in the frequency band where spectral cues operate, of which a shift in the main pinna notch from 9 kHz to 10 kHz is the most notable. The authors attribute the observed differences to the inaccurate shape description resulting from the laser scanner they used and conclude that an accurate ear model is crucial for BEM-based HRTF calculations.

Kahana and Nelson (2005) obtained raw meshes using another laser scanner, the Cyberware Mini Model. This scanner allowed finer detail to be achieved behind the ear and in the pinna folds. The authors also succeeded in capturing some of the ear canal geometry. Walsh *et al.* (2003) included the ear canal in their mesh using CT scans as their starting point. A mesh of the ear canal was reconstructed by joining a number of cross-sectional slices shaped to fit its surface. The resulting mesh was then "scaled, rotated, and joined to the original head mesh by cutting a small opening on the ear, attaching the canal, and redefining the local connectivities" (Walsh *et al.*, 2003).

2.4.6.3 Re-meshing and patch size issues

A mesh used for BEM-based acoustic simulations is required to satisfy a number of conditions if the calculations are to yield valid results. The mesh topology is crucial. Katz (2001a) describes the meshing process as a balance between maximum edge length, mesh accuracy and homogeneous discretisation. The mesh resolution should allow relevant shape detail to be described and the maximum patch edge length used to describe the object places an upper limit on the frequency for which the results are valid. Katz

(2001a) proposed a maximum edge length of $\lambda/6$, where λ is the acoustic wavelength in air of the maximum frequency to be simulated. He suggests however, that a maximum edge length of $\lambda/4$ may, in the limit, lead to results of satisfactory accuracy. Models fulfilling this condition have been shown to yield results agreeing with acoustic measurements (see Kahana and Nelson, 2005; Otani and Ise, 2006, and Section 2.4.6.5). A rigorous study of the effects of coarsening the mesh beyond the currently agreed limit remains to be performed however.

The advantages of increasing the number of patches are counterbalanced by significant drawbacks. The computational requirements (mainly time and memory) rise approximately proportionally to the cube of the patch-count in the case of large meshes. A compromise which achieves a satisfactory balance between speed and accuracy has to be found. To optimise performance the vertices should be spread homogeneously on the surface, which minimises the number of patches needed to satisfy the constraint on edge length. To achieve this most researchers have created a dense mesh and then applied various decimation algorithms to improve performance (see Katz, 2001a; Kahana and Nelson, 2005; Otani and Ise, 2006, for example).

2.4.6.4 The reciprocity principle

A dramatic improvement in the efficiency of the simulations can be achieved using the Helmholtz principle of reciprocity. In a linear time-invariant acoustic scene, the principle states that the pressure at an arbitrary point A caused by a source at point B is equal to the pressure which would be measured at point B if the source were placed at point A (see Morse and Ingard, 1968, for example).

Instead of separately modeling the acoustic effects of sources around a head mesh at every location for which HRTFs are required, it is possible to place a source in the ear canal and to calculate HRTFs in the reverse direction. Reducing the problem to a single source at the ear greatly accelerates calculations, because most of the computation required for BEM simulations is used in determining the pressures created by a source on the mesh patches. Once these are known, there is relatively little computational overhead involved in calculating large numbers of far-field pressures. Hence, the computationally intensive calculation of patch pressures needs to be performed only once and the fast far-field pressure calculations can be performed for a large number of points in a relatively short time. Failing to exploit the principle of reciprocity in this way would require patch pressures to be calculated for every single source location, dramatically increasing computation times. The principle can also be used in the context of acoustic measurements as described by Zotkin *et al.* (2006).

2.4.6.5 Results and agreement with acoustic measurements

Kahana and Nelson (2005) performed a BEM-based study of KEMAR ear acoustics widely considered to be the state of the art. They related their results to the acoustic measurements reported by Shaw and Teranishi (1968), described in Section 2.2.2.1. They extracted the modes of the pinna using a procedure similar to that proposed by Shaw *et al.*, but through numerical modeling. The plate used to isolate the pinna by Shaw and Teranishi (1968) was replaced by an infinite baffle. The simulations were conducted by placing an ideal monopole source 1 mm away from the blocked meatus and using the reciprocity principle to calculate pressure variations 1 metre away. This section describes the results reported by Kahana *et al.* in detail as they

Resonance frequency	Wave incidence angle (φ)
4.0 kHz	16°
7.2 kHz	60°
9.5 kHz	94°
11.6 kHz	0°
14.8 kHz	4°
18 kHz	−16°

Table 2.3: Pinna resonances and corresponding source elevation angle causing maximum excitation for the DB-65 KEMAR ears, reported by Kahana and Nelson (2005).

will be related to the results obtained during a similar exercise, which has been carried out during the course of the research reported in this thesis (see Chapter 3).

Figure 2.12 shows the normalised response of DB-65 baffled pinnae for a source at grazing incidence angle with $0° \leq \varphi \leq 180°$ reported by Kahana and Nelson (2005), emulating the acoustic measurements performed by Shaw and Teranishi (1968). The resonance frequencies noted by the authors are shown in Table 2.3 along with the source elevation giving rise to maximum excitation at grazing incidences (0° azimuth). Kahana *et al.* observed that the excitation patterns for resonances below 10 kHz were highly similar to those reported by Shaw and Teranishi (1968).

Along with far-field excitation patterns which were reported for the entire ipsilateral hemisphere, Kahana and Nelson (2005) also investigated the pressure on the surface of the pinnae for different modes using the singular value decomposition (SVD) technique. This utilises the Green function matrix of field and source points, a commonly used approach for solving boundary element problems in the context of sound radiation and scattering analysis (Photiadis, 1995; Borgiotti, 1990). This formulation is useful because it

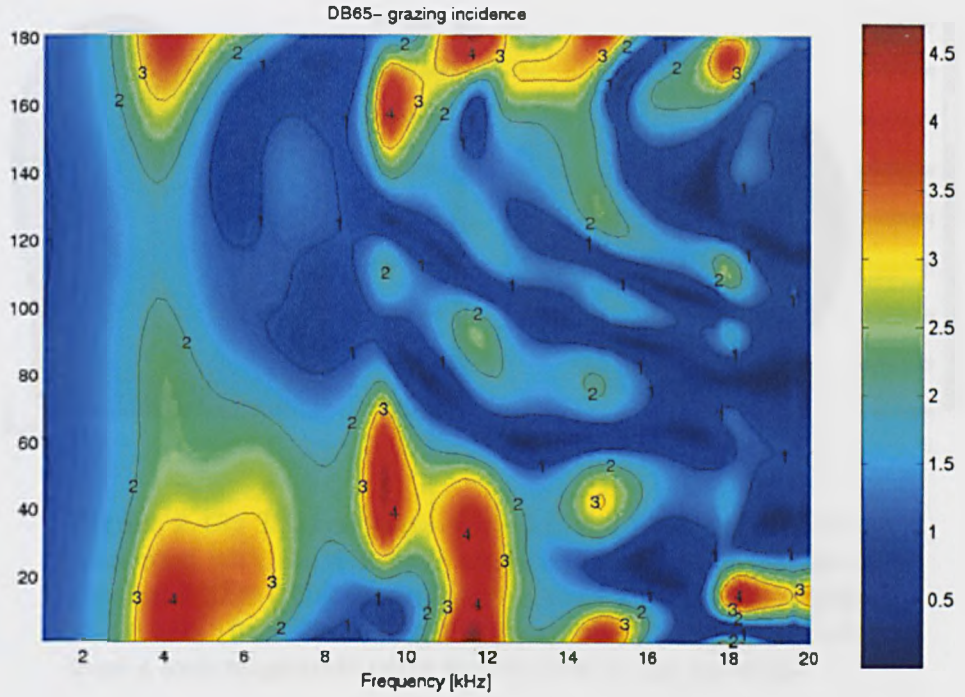


Figure 2.12: The normalised response of the DB-65 baffled pinnae for a grazing incidence angle (from Kahana and Nelson, 2005). Values for each source angle were obtained by dividing the response at the entrance to the blocked ear canal with the response at the same location on the baffle but without the pinna.

allows far-field acoustics to be investigated concurrently with surface pressures. The two sets of data are partitioned into the right and left singular vectors, respectively, during the singular value decomposition process. Kahana *et al.* chose to display the surface pressures by assigning colours to each vertex representing absolute pressure. The colour scheme restricted phase information to positive or negative. They note that a more accurate and intuitive presentation would be to plot the magnitude and phase of the surface pressures separately, but explained that their choice was driven by a desire for consistency with the works of Shaw and Teranishi (1968).

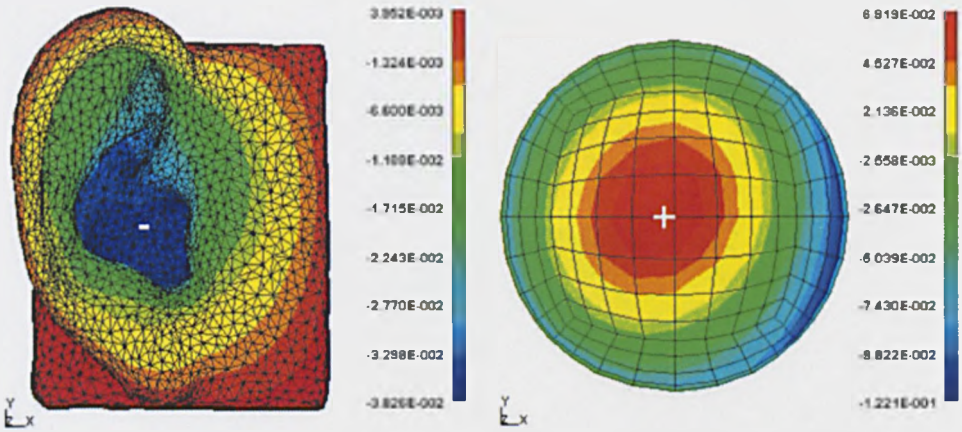


Figure 2.13: The real parts of the left and right singular vectors associated with the dominant singular values of the numerically generated 3389×209 Green function matrix for the DB-65 pinna at 4.0 kHz (from Kahana and Nelson, 2005). The values result from a unit magnitude point source close to the ear canal.

A number of pinna modes were identified by Kahana and Nelson (2005). A 4.0 kHz mode, described as a concha monopole, is shown in Figure 2.13. It is slightly more pronounced for sources originating from the front, where there is a clear line of sight to the concha, than from the back, where the source is occluded by the pinna. Overall, the far-field excitation distribution is smooth, with a maximum in the far ipsilateral area. This closely matches Shaw's "unidirectional resonance of the concha" at 4.2 kHz (Shaw and Teranishi, 1968). The surface pressure distribution for this resonance was similar for all five pinnae tested by Kahana *et al*; significant variation was observed only with respect to the phase.

The 6.8 kHz mode (see Figure 2.14), weak compared with the others, according to Kahana and Nelson (2005) due to a "vertical dipole pattern" involving an "oscillatory flow between the cavum concha, cymba concha and the antihelix". This seems to correspond to a resonance identified by Shaw

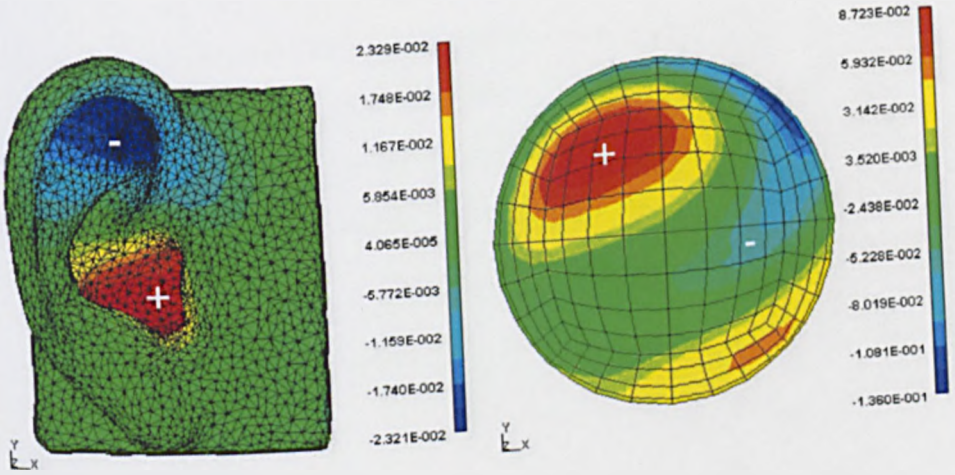


Figure 2.14: The real parts of the left and right singular vectors associated with the dominant singular values of the numerically generated 3389×209 Green function matrix for the DB-65 pinna at 6.8 kHz (from Kahana and Nelson, 2005). The values result from a unit magnitude point source close to the ear canal.

and Teranishi (1968) at 7.1 kHz, which was described as a vertical mode and exhibits a similar far-field excitation pattern for the grazing source locations which were measured.

A 9.5 kHz mode, according to Kahana and Nelson (2005) due to another "vertical mode" involving "oscillatory flow between the cavum concha, cymba concha and the antihelix, as before but with different signs in the cymba concha and the antihelix". Although the highest level of excitation was observed for source elevations around $\varphi = 94°$ grazing source incidences (see Table 2.3), it is also high for source elevations around $\varphi = -35°$, a characteristic also observed in the case of acoustic measurements (Shaw and Teranishi, 1968).

At higher frequencies, surface pressure and far-field excitation patterns become more complex. Cross-subject similarities become less pronounced and

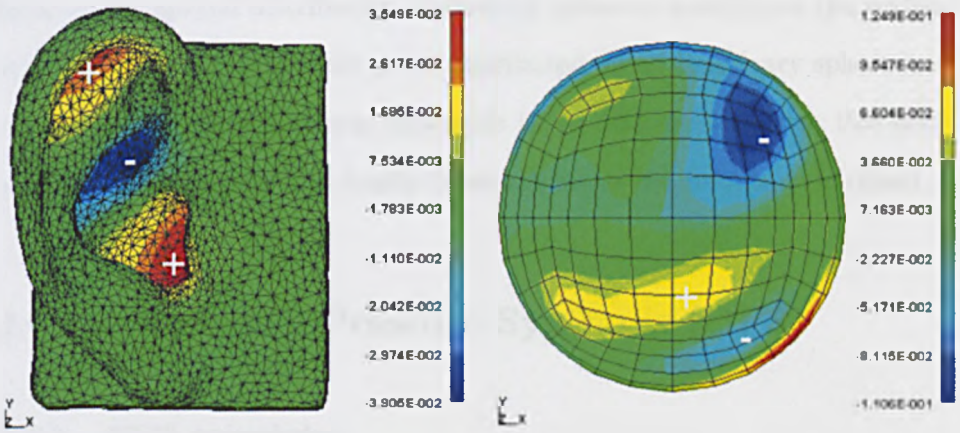


Figure 2.15: The real parts of the left and right singular vectors associated with the dominant singular values of the numerically generated 3389×209 Green function matrix for the DB-65 pinna at 9.5 kHz (from Kahana and Nelson, 2005). The values result from a unit magnitude point source close to the ear canal.

the agreement between simulated and measured acoustics deteriorates. This was also reported by Otani and Ise (2006) who noted that HRTFs calculated with a micro-CT scanned ear model agree well with actual measurements only up to 10 kHz.

Accurate pressure values over the whole surface of the simulated object can provide an insight into the acoustic mechanisms which cause modal behaviour. This data, all but impossible to obtain using traditional acoustic measurements, allowed Kahana and Nelson (2005) to confirm the modal factors underlying some of the acoustic effects within the pinna as suggested by Shaw and Teranishi (1968). The surface pressure variations in the area surrounding the pinna are negligible in all cases, and most of the resonant behaviour appears in the cavum concha, cymba concha, fossa of helix and antihelix, confirming their status as primary cue production regions. Kahana and Nelson (2005) made the interesting observation that a resemblance

between the spatial distribution of absolute pressure maxima on the surface of the pinnae and at far-field point distributed on an imaginary sphere surrounding the KEMAR ears is noticeable below 10 kHz. However, this similarity becomes weaker and finally breaks down as the frequency is raised.

2.5 Differential Pressure Synthesis (DPS)

2.5.1 DPS principles

It seems intuitively obvious that the calculated acoustic field around an object could somehow be used as a basis for estimating the field around a slightly deformed version of that object. Traditional simulation techniques such as the BEM, however, cannot exploit this similarity to shorten the simulation time and require the entire calculation process to be repeated. The differential pressure synthesis (DPS) method introduced by Tao *et al.* (see Tao *et al.*, 2003a,b) offers a solution to this problem. This solution requires expressing the small objects deformation as a weighted sum of a set of orthogonal shape deformations. The pressure changes associated with each orthogonal deformation are pre-calculated and stored into a database. Using this database, the acoustic effect of the original deformation is estimated by summing the accordingly weighted pressure changes associated with each orthogonal shape deformation.

The linear superposition of pressure changes caused by orthogonal shape perturbations amounts to a first order approximation of the complex nonlinear relationships linking shape perturbations with their associated pressure changes. Consequently, this pressure estimate is only valid for small deformations where the magnitude of an orthogonal deformation and the

resulting pressure change are linearly related. Within this region DPS allows the acoustic effect of an arbitrary deformation to be estimated without the need for further acoustic simulation. Instead, it can be expressed as a summation of the orthogonal deformations for which the acoustic effects have been pre-computed. The technique was developped in order to rapidly estimate the acoustic effects of arbitrary, small head shape variations.

2.5.2 Mathematical description

2.5.2.1 Two-dimensional DPS

Tao *et al.* (2003a) describe DPS in two dimensions using orthogonal radial harmonic deformations. Any singled-valued, differentiable radial function $r(\theta)$ defined over $0 \leq \theta < 2\pi$ can be described using radial harmonics, which in effect are the terms of a radial Fourier series (the polar angle θ replacing the usual Cartesian variable x or t). In discretised form this can be expressed as (from Tao *et al.*, 2003a):

$$r[n] = a_0 + \sum_{m=0}^{N-1} a_m \cos\left(\frac{2\pi}{N} mn\right) + \sum_{m=0}^{N-1} b_m \sin\left(\frac{2\pi}{N} mn\right) \qquad (2.8)$$

where

$$a_m = \frac{1}{N} \sum_{n=0}^{N-1} r[n] \cos\left(\frac{2\pi}{N} mn\right) \qquad (2.9a)$$

$$b_m = \frac{1}{N} \sum_{n=0}^{N-1} r[n] \sin\left(\frac{2\pi}{N} mn\right) \qquad (2.9b)$$

The bedrock of DPS is the assumption that given the acoustic pressure around an object can be determined from its shape, $r[n]$, it can also be determined by the Fourier series coefficients which define $r[n]$. Assuming the pressure p near the surface of an object is differentiable with respect to a_m and b_m, the pressure difference dp caused by an infinitesimal arbitrary deformation can be represented as a linear superposition of the pressure differences caused by its orthogonal constituent deformations. To ease notation, Tao *et al.* define the variable

$$e_m^\sigma = \begin{cases} a_m & \text{for } \sigma = 0 \\ b_m & \text{for } \sigma = 1 \end{cases} \tag{2.10}$$

Using this variable the pressure difference dp can be expressed as

$$dp = \sum_{\sigma=0}^{1} \sum_{m=0}^{N-1} \frac{\partial p}{\partial e_m^\sigma} de_m^\sigma \tag{2.11}$$

Within the deformation range where the relationships between orthogonal deformation magnitudes e_m^σ and acoustic pressure p can be considered linear, the difference in pressure Δp caused by an arbitrary summation of orthogonal deformations is approximately

$$\Delta p \approx \sum_{\sigma=0}^{1} \sum_{m=0}^{N-1} \frac{\partial p}{\partial e_m^\sigma} \Delta e_m^\sigma \tag{2.12}$$

which has the form of a first-order multidimensional Taylor series.

2.5.2.2 Three-dimensional DPS

Tao *et al.* (2003a) extended DPS principles to three dimensions using surface spherical harmonics (SSHs) to define a surface $r(\theta, \varphi)$ on which the pressure $p(\theta, \varphi)$ is to be estimated:

$$r(\theta,\varphi) = \sum_{n=0}^{\infty} a_{n0} P_n^0(\cos\theta) \quad \cdots$$
$$\cdots \quad + \sum_{n=0}^{\infty} \sum_{m=1}^{n} P_n^m(\cos\theta) \left[a_{nm}\cos(m\varphi) + b_{nm} sin(m\varphi)\right] \quad (2.13)$$

where a_{nm} and b_{nm} are the SSH coefficients and $P_n^m(\cos\theta)$ is the Legendre polynomial of degree n and order m. When a finite order N is used to represent a shape, incorporating the first summation into the second and isolating the $(0,0)$th order ($n = 0$, $m = 0$), Equation 2.13 becomes

$$r(\theta,\varphi) = a_{00} + \sum_{n=0}^{\infty} \sum_{m=1}^{n} P_n^m(\cos\theta) \left[a_{nm}\cos(m\varphi) + b_{nm} sin(m\varphi)\right] \quad (2.14)$$

As in the two-dimensional case, Tao *et al.* use the variable

$$e_{nm}^{\sigma} = \begin{cases} a_{nm} & \text{for } \sigma = 0 \\ b_{nm} & \text{for } \sigma = 1 \end{cases} \quad (2.15)$$

to simplify notation, allowing the pressure difference dp caused by an infinitesimal arbitrary deformation to be expressed as the following summa-

tion

$$dp = \sum_{\sigma=0}^{1} \sum_{n=0}^{N} \sum_{m=0}^{n} \frac{\partial p}{\partial e_{nm}^{\sigma}} de_{nm}^{\sigma} \tag{2.16}$$

Within the deformation range where the relationships between deformation magnitude e_{nm}^{σ} and acoustic pressure p can be considered linear, the difference in pressure Δp caused by an arbitrary deformation is approximately

$$\Delta p \approx \sum_{\sigma=0}^{1} \sum_{n=0}^{N} \sum_{m=0}^{n} \frac{\partial p}{\partial e_{nm}^{\sigma}} \Delta e_{nm}^{\sigma} \tag{2.17}$$

which, like Equation 2.12 has the form of a first-order multidimensional Taylor series.

2.5.3 Applications of DPS in HRTF estimation

2.5.3.1 Effects of head shape simplification

As described in Section 2.4.3, a number of studies have resorted to simplifying the shape of the head and torso to facilitate the estimation of their low frequency acoustic effects. Perhaps the most widely documented of these was the use of spherical and ellipsoidal approximations by Avendano *et al.* (1999), Algazi *et al.* (2001a), Algazi *et al.* (2002) and Algazi and Duda (2002). Tao *et al.* (2003b) rigorously examined the effects of shape simplifications by investigating the effects of truncations applied to the SSH

series describing the KEMAR[2] head. The pinnae were not included in the original model as their shape cannot be described using SSHs (see Section 2.5.3.3). This constrained the frequency range of validity to below 3 kHz. As expected, the shape representation deteriorated when the degree/order truncation threshold was lowered. Tao *et al.* noted that below 3 kHz the truncation of spherical harmonics of degrees 11 and above induced an RMS pressure error of no more than 5% and that no further improvement in pressure accuracy was achieved by increasing the truncation threshold beyond spherical harmonics of degree 14. They concluded that the fine detail of the head (excluding pinnae), specifically the nose and other facial features, do not significantly contribute to the production of localisation cues.

2.5.3.2 Evaluation of DPS performance

Tao *et al.* (2003a) tested the validity of DPS estimation in the two-dimensional case using a circular template of radius $a_0 = 0.1$ m. The template boundary pressures caused by a nearby source were computed using the BEM along with the effects of radial harmonic deformations, which were stored in a 2D-DPS database. The circular template was then deformed using randomly generated coefficients for the first to seventh order radial harmonics $(a_1, ..., a_7)$ such that the maximum shape deformation was 10% of the radius a_0. The boundary pressures for the deformed template were computed directly using the BEM and results were compared to DPS estimates. Any observed error was attributed to the DPS estimation process. The results showed good agreement at low frequencies, but errors increased with frequency. The range of deformation amplitudes for which linearity may be

[2]Knowles Electronic Manikin for Acoustic Research (Url: http://www.gras.dk/00012/00330/)

Frequency	% Error
250 Hz	0.9
500 Hz	2.2
1 kHz	6.9
2 kHz	18.9
3 kHz	34.6

Table 2.4: DPS estimation error, as a percentage of the actual pressure difference caused by deforming a spherical template into a pinna-less KEMAR head, determined through direct BEM simulation, reported by Tao *et al.* (2003a).

assumed decreases for higher order radial harmonic deformations and higher audio frequencies.

Tao *et al.* (2003a) also tested DPS performance in the three-dimensional case by deforming a spherical template into a pinnaless KEMAR head (shown in Figure 2.16) for frequencies ranging from 250 Hz to 3 kHz. The errors reported at different frequencies are shown in Table 2.4 as a percentage of the actual pressure difference determined using the BEM directly. While the estimation inaccuracies revealed in this study were significant at higher frequencies, results were, again, promising. In particular, the authors drew attention to the large shape perturbations required to deform a sphere into a KEMAR head which were considerably greater than the variations that might be expected across human heads.

2.5.3.3 The limitations of surface spherical harmonics

The use of SSHs inherently limits the range of shapes to which the DPS database compiled by Tao *et al.* can be applied. Although spherical harmonics are a very elegant form of orthogonal shape description, their applicability is constrained to surfaces which can be described by a single-valued

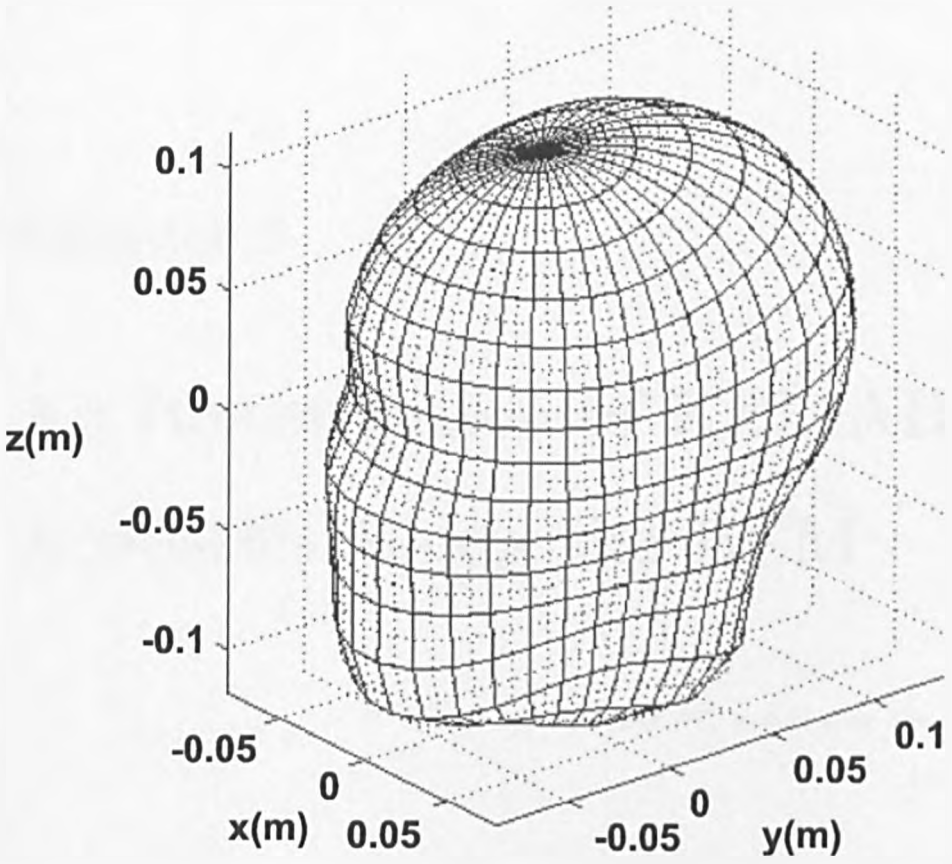


Figure 2.16: Pinna-less KEMAR head expressed using surface spherical harmonics (from Tao *et al.* (2003a))

spherical polar function. Because of this, pinna deformations cannot be accommodated. Given the importance of pinna acoustics in the spatial cue production process, alternative basis functions which allow orthogonal pinna deformations are required. It will then become possible to estimate the effects of morphological variations on high frequency pinna spectral cues, a crucial step for effective DPS-based HRTF individualisation from morphological data.

Chapter 3

An Investigation of KEMAR Acoustics Using the BEM

"Pentiums melt in your PC, not in your hand."

Anon

The acoustic origin of ITD and ILD localisation cues are well understood and reliable techniques have been developed to estimate them given an adequate set of morphological measurements (Kuhn, 1977; Algazi *et al.*, 2001b, amongst others). The individualisation of spectral cues, however, has proved a far more challenging task. The acoustic mechanisms from which they originate, as well as the intricacies of their perceptual operation, remain, despite numerous research efforts, only partially understood. Section 2.2.2 gives an overview of these research efforts. Several studies have shown that a sense of sound source location can, in some cases, be achieved through a coarse approximation of macroscopic spectral features, namely prominent HRTF peaks and notches (Hebrank and Wright, 1974; Bloom, 1977; Butler and Helwig, 1983). Unsimplified spectral cues, however, allow the localisation of

broadband sources (elevation and azimuth) to within 5° error (Perrott and Saberi, 1990) over the entire auditory space. This suggests that the auditory system extracts some source location information from the finer spectral detail which should, therefore, be preserved in order to achieve full localisation acuity.

Whether specific HRTF variations with direction have an impact on localisation, no matter how slight, is difficult to establish experimentally. Variations are easily distorted by shortcomings in transducer design and imperfections in the sound recording and reproduction processes. These errors overwhelm the more subtle HRTF features which may be contributing to a highly accurate and realistic sound image. The fact that the extraction of source location from sonic input occurs, at least partly, far earlier in the auditory pathway than the stages responsible for sonic awareness (see Section 3.4) complicates matters further. A fuller knowledge of the spectral variations which occur in HRTFs over the entire auditory space is crucial to our understanding of the more illusive and sophisticated aspects of spectral cue operation.

A study of spatial HRTF variations for a KEMAR manikin using data acquired through BEM simulations is presented in this chapter. Rings of confusion receive particular attention since, by definition, source localisation observed within them cannot result from changes in interaural time and level differences and must be entirely due to spectral variation. The steps involved in acquiring the raw shape data, preparing a satisfactory mesh model and performing the BEM simulations are described. The spectral variations determined using the acoustic simulations are analysed and discussed primarily in relation to localisation processes. Where possible, our

findings are compared with previously published results and observations.

3.1 Acoustic simulations

Aural acoustic measurements require the introduction of a microphone, generally in the vicinity of the ear canal. Its presence inevitably results in sound field distortions, leading to measurement errors. Although Pralong and Carlile (1994) have shown that these perturbations can be mostly restricted to ultrasonic frequencies, subtle, lower frequency secondary effects are essentially inevitable. The acoustic transparency of simulation point sources and far-field pressure measurement points eliminate such errors and minimise further inaccuracies resulting from background noise, transduction imperfections, slight environmental echoicity and involuntary subject movements. Acoustic simulations also offer a level of control and flexibility over the system being investigated, the position of the sound sources and microphones as well as the excitation signals which is unachievable through acoustic measurements. These benefits are invaluable in the context of HRTF analysis.

Section 2.4.6 reviews prior research efforts into the acoustics of the auditory periphery using acoustic simulations. Several results obtained with BEM-based simulations have been validated against real acoustic measurements. The maximum frequency of validity has gradually risen in step with growing computing capabilities and decreasing patch size (Walsh *et al.*, 2003; Otani and Ise, 2006; Kahana and Nelson, 2005, 2007). However, a comparison of measurements performed by a number of institutions, reported by Katz and Bergault (2007), shows that slight differences in measurement

setup and methods can cause large discrepancies between the results of different groups. This must lead to doubts concerning their validity. Although it is hard to affirm whether simulated results or their measured counterparts best represent the physical reality, it is clear that consistency across measurements and frequencies is far better achieved using numerical simulations than traditional means.

All acoustic simulations conducted in this research were performed using the PAFEC-FE software package[1] which applies the boundary element method, amongst other techniques, to vibration and acoustic problems. Simulations exploited the reciprocity principle (see Section 2.4.6.4), allowing the far field excitation to be calculated at a practically unlimited number of points with very little overhead by placing the source close to the ear canal. In a further step to reduce simulation time, the KEMAR head is assumed to be perfectly symmetrical as this characteristic allows the calculation of surface pressures to be split into two problems each involving half as many patches. Although this process doubles the number of simulations, the benefits of the patch-count reduction far outweigh this drawback in terms of computational demand in the case of large meshes.

3.2 Obtaining a KEMAR mesh description

The KEMAR's head, torso and small ears (DB-60, DB-61) were used as the basis for creating the template mesh. This decision was taken because KEMAR is a recognised industry-standard for anthropomorphic testing in the fields of telecommunications, noise abatement, sound recording

[1]The PAFEC-FE software package is distributed and maintained by PACSYS Limited (Url: `http://www.vibroacoustics.co.uk/`

and sound-quality evaluation amongst others[2]. The manikin is designed to simulate acoustic effects caused by the auditory periphery (such as diffraction and reflection) and aims to represent the acoustical and morphological mean for the adult human population.

A CT scanner[3] was used to obtain an adequate mesh description of the small KEMAR ears allowing their shape to be captured completely. The KEMAR head and torso could not however be scanned using CT or MRI technology because of its inadequate constituent materials. Instead, they were captured optically, using the Polhemus FasTRACK laser scanner[4]. As mentioned in Section 2.4.6.2 the data obtained via laser scanning is restricted to the line of sight from the cameras. Although this is a significant problem in the case of the pinnae, most of the head and torso surfaces could be captured, with only small gaps under the arms.

The ear mesh obtained using CT and the head and torso mesh created by the laser scanner were combined using 3ds Max[5]. This software package contains powerful mesh editing tools, which allow vertices and polygons to be created and edited manually. As such it was ideal to remove scanning imperfections and smoothly combine the ear meshes with the head mesh.

Given the need for a sufficiently low patch count to enable BEM simulations to be possible using available computing resources, it was decided not to include the torso in the final mesh. Although torso effects have been shown to be perceptually significant, especially for elevation perception, they

[2]Url: `http://www.gras.dk/00012/00330/`

[3]The CT scanner was a GE LightSpeed Plus located at Yorkshire Clinic Imaging Centre, UK.

[4]`http://www.polhemus.com/?page=Motion_Fastrak`

[5]3ds Max is an Autodesk product, more information on which can be found at `http://usa.autodesk.com/adsk/servlet/index?siteID=123112&id=5659302`

can be easily re-incorporated using one of the structural models described in Section 2.4.2. Since the BEM requires a closed surface, the neck was sealed as smoothly as possible to reduce unwanted acoustic effects, also using 3ds Max. Once the raw mesh had been produced, it had to be cleared of unwanted elements and re-meshed in order to allow acoustic simulations.

3.3 A multi-resolution approach to meshing for the BEM

3.3.1 Motivation

Meshing the acquired surface description so that it is suitable for BEM calculations is an important step which requires careful consideration if full head/pinnae simulations are to be made possible. It is widely accepted that the mesh topology determines the upper frequency limit of validity. The general consensus is that in order to guarantee the validity of acoustic BEM simulation results, a mesh should have a maximum edge length of $\lambda/6$, where λ is the wavelength of the acoustic source, rising to $\lambda/4$ in the limit (see Section 2.4.6.3). The shorter the maximum edge length, the higher the frequency for which valid simulations can be performed. However, the computational cost of running simulations increases sharply with the number of patches. A balance must be reached to satisfy these competing constraints. Ideally, the meshing resolution should allow valid simulations up to the required analysis frequency without any superfluous definition in order to minimise computational cost. Simulating a mesh of a whole head, following this guideline, requires a maximum edge length of

$$\frac{\lambda}{6} = \frac{c}{6f} \tag{3.1}$$

where $c = 343$ m/s is the speed of sound and $f = 20$ kHz is the maximum simulation frequency. This gives

$$\frac{\lambda}{6} = \frac{343}{6 \times 20,000} = 0.0028\text{m} = 2.8\text{mm} \tag{3.2}$$

Such a maximum edge length over the entire head/pinnae surface area makes acoustic simulations unachievable on generally available computer systems. Kahana and Nelson (2005) succeeded in modelling several baffled pinnae up to 20 kHz and heads with pinnae up to 10 kHz. Given the extent to which BEM simulations are used in the course of the research presented in this thesis, particularly the creation of a DPS database described in Chapter 5, and the constraints on computational resources, it is important to investigate whether, in some cases, the generally accepted edge length limit can be exceeded without significantly affecting the accuracy of the required acoustic simulations.

In practice, mesh resolution requirements are rather more complicated to establish than a straightforward application of the $\lambda/6$ rule might suggest. Many factors influence the accuracy and the upper frequency limit. They include patch shape, the rate of change of acoustic pressure across the surface, surface integrity and the relative pressures on patches. A more conclusive means of assessing accuracy is therefore to conduct a convergence test, in which the shape is re-meshed using successively shorter patch edges until the change in computed pressures from one mesh to the next is con-

sidered small enough at a particular frequency that the pressures can be said to have converged to the true solution. In the case of BEM simulations applied to HRTF calculations, a number of facts should be taken into consideration. For example, the sharp folds in the pinna require mesh edge lengths shorter than $\lambda/6$ to preserve its shape sufficiently. Conversely, shape variations over the remainder of the head are relatively smooth, with the possible exception of the nose, which has been shown to be of little relevance to the production of localisation cues (see Section 2.5.3.1) and its shape can be described using far larger patches introducing only very small shape errors. Also, at frequencies above about 5 kHz ($\lambda/6 = 11.4$ mm), the HRTF is almost entirely determined by the shape of the pinnae.

On these grounds, the effect of increasing the maximum edge length of a mesh over the surface of the head, excluding the pinna, was investigated. Considering the fairly small proportion of head surface occupied by the pinna regions, this approach, if it proves to be justifiable, has the potential to considerably reduce the number of patches required for valid acoustic simulations. It would consequently result in very significant time savings. This section presents a multi-resolution meshing algorithm, along with a location and frequency-dependent investigation of the performance of a multi-resolution KEMAR head mesh.

3.3.2 Procedure

The meshing algorithm used in the context of this research was designed and implemented by Carl Hetherington[6]. It acts upon a high-resolution mesh, selecting a subset of the mesh points which are fed into a point-

[6]Research associate (University of York, Audio lab)

cloud reconstructor originally developed by Lin *et al.* (2004). The point-selection algorithm is driven by the constraint of an optimum (target) edge length $L_0(\mathbf{p})$ at some point $\mathbf{p} = [p_x\ p_y\ p_z]$ in the mesh. The points of the high-resolution source mesh are iterated through, and only those which are further than the $L_0(\mathbf{p})$ threshold from any already-used point are added to the output point list.

The target edge length $L_0(\mathbf{p})$ is the only control variable for the algorithm. An optimum edge length L_a is defined for the lower-resolution areas of the mesh. A set of N "detail ellipsoids" within which the mesh resolution is to be magnified are then defined as, E_1, E_2, . . . E_N. Each of these is described by a centre point $\mathbf{c}_1$, $\mathbf{c}_2$, . . . $\mathbf{c}_N$ with $\mathbf{c}_m = [c_{mx}\ c_{my}\ c_{mz}]$ and a semi-principal axes $\mathbf{a}_1$, $\mathbf{a}_2$, . . . $\mathbf{a}_N$ with $\mathbf{a}_m = [a_{mx}\ a_{my}\ a_{mz}]$. The mth ellipsoid E_m is defined mathematically as:

$$\frac{(x - c_{mx})^2}{a_{mx}^2} + \frac{(y - c_{my})^2}{a_{my}^2} + \frac{(z - c_{mz})^2}{a_{mz}^2} = 1 \tag{3.3}$$

and all the $\mathbf{p}$ points inside E_m satisfy:

$$0 \leq \frac{(p_x - c_{mx})^2}{a_{mx}^2} + \frac{(p_y - c_{my})^2}{a_{my}^2} + \frac{(p_z - c_{mz})^2}{a_{mz}^2} < 1 \tag{3.4}$$

Each detail ellipsoid E_m has an associated "detail factor" F_m. For a given point $\mathbf{p}$, the optimum edge length $L_0(\mathbf{p})$ is computed using

$$L_0(\mathbf{p}) = \frac{L_a}{S(\mathbf{p})} \tag{3.5}$$

where $S(\mathbf{p})$ is determined by the detail ellipsoids. Given the N detail ellipsoids, we find the subset M ellipsoids which contain the point $\mathbf{p}$. For each such ellipsoid, a fade factor G_m, representing a measure of how close the point $\mathbf{p}$ is to the centre of the ellipsoid, is computed as follows

$$G_m(\mathbf{p}) = \cos\left(\frac{\pi}{2}\left[\frac{(p_x - c_{mx})^2}{a_{mx}^2} + \frac{(p_y - c_{my})^2}{a_{my}^2} + \frac{(p_z - c_{mz})^2}{a_{mz}^2}\right]\right) \tag{3.6}$$

Since $\mathbf{p}$ is inside the ellipsoid, and taking into account Equation 3.4, the value of G_m will fade between 1, at the centre of the ellipsoid, to 0 at the edges. This allows a smooth edge length transition throughout the ellipsoid, avoiding an unwanted abrupt change at the boundary. $S(\mathbf{p})$ is then computed using

$$S(\mathbf{p}) = 1 + \sum_{m=1}^{N} F_m G_m(\mathbf{p}) \tag{3.7}$$

3.3.3 KEMAR simulation model validation

A set of 524,288 points was used as starting point for the creation of the KEMAR head mesh representation. This data was supplied to the multi-resolution algorithm. The high point count serves to improve the regular (equilateral) character of the three-sided polygons comprising the output mesh. It also helps to avoid meshing imperfections which would result in invalid BEM simulations. A high level of detail is kept in the pinna regions to preserve its shape and the target edge length is increased smoothly to

the substantially larger value employed over the remainder of the head. As mentioned in Section 3.1, the KEMAR head is assumed to be perfectly symmetrical. Under this assumption, only half the KEMAR mesh is required for simulation. The final simulation mesh is, therefore, a representation of the right half of the KEMAR head, with the DB-61 KEMAR pinna, and consists of 3,816 polygons.

The multi-resolution algorithm achieved the desired results such that a KEMAR mesh comprising small patches for the pinnae and larger patches for the remainder of the head was generated (see Figure 3.1). A close-up view of the external ear and the transition area is shown in Figure 3.2. Simulations were carried out at 222 Hz frequency intervals over the 0-14 kHz range. In order to assess the performance of the multi-resolution mesh, BEM simulations were carried out for a high-resolution mesh for comparison purposes, with maximum edge length inferior to $\lambda/4$ at 15 kHz over the entire surface of the head. The resulting mesh was comprised of slightly over 17,000 polygons. These simulations were very time consuming (several hundred times more than the multi-resolution mesh) and so were carried out for a limited number of frequencies. Simulations were also carried out for a subdivided version of the multi-resolution mesh containing 13,879 polygons. Unlike the high-resolution mesh, the subdivided mesh describes exactly the same shape as the multi-resolution mesh, but has a much smaller maximum edge length. In each case the point source is placed 2.12 mm away from the mesh and HRTFs are calculated by using the reciprocity principle (see Section 2.4.6.4).

Given the extremely high computational requirements for performing BEM simulations of the high-resolution mesh (around 16 CPU-hours per fre-

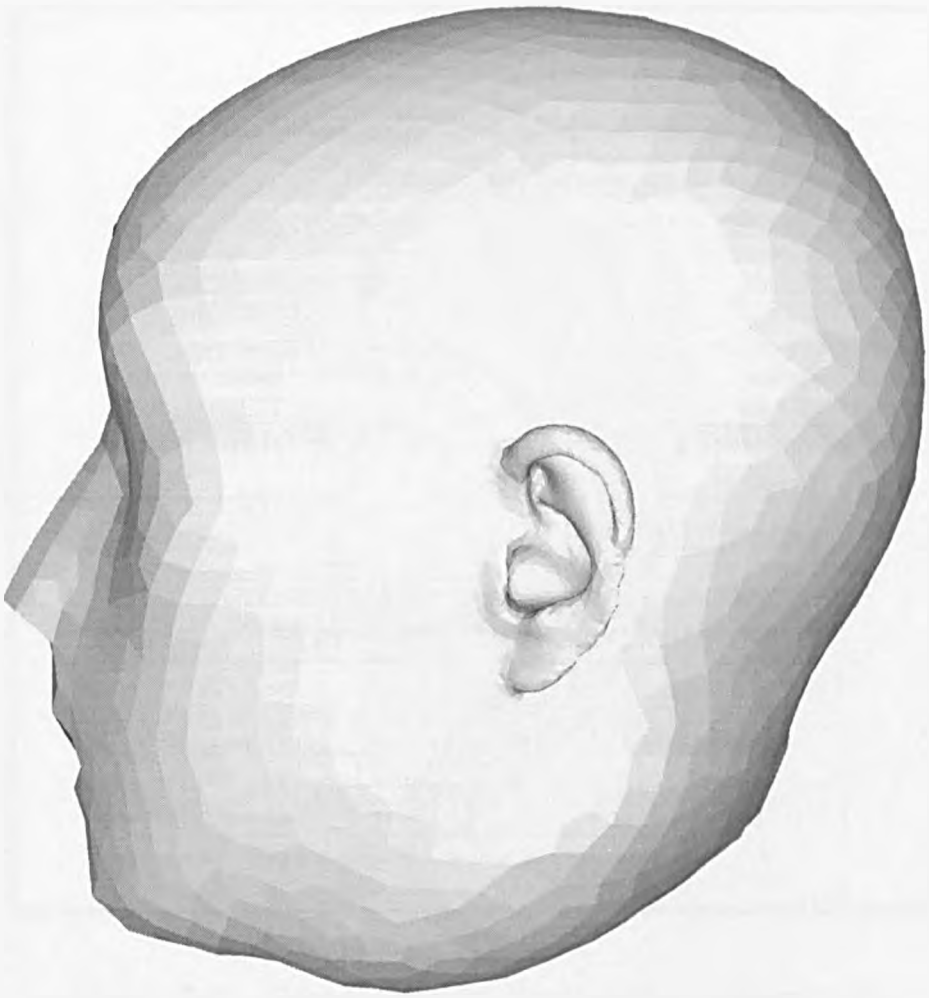


Figure 3.1: Multi-resolution KEMAR right-half head mesh with a DB-61 pinna. The mesh contains 3,816 polygons, with high-resolution meshing in the pinna area and lower resolution meshing over the remainder of the head. Maximum edge length is kept below $\lambda/6$ at 15 kHz in the pinna region but is significantly increased over the remainder of the head, dramatically reducing patch count.

quency), only a small range of high frequencies (10.0-14.8 kHz) were computed for performance comparison. For similar reasons, the simulation frequency range for the subdivided multi-resolution mesh were constrained to

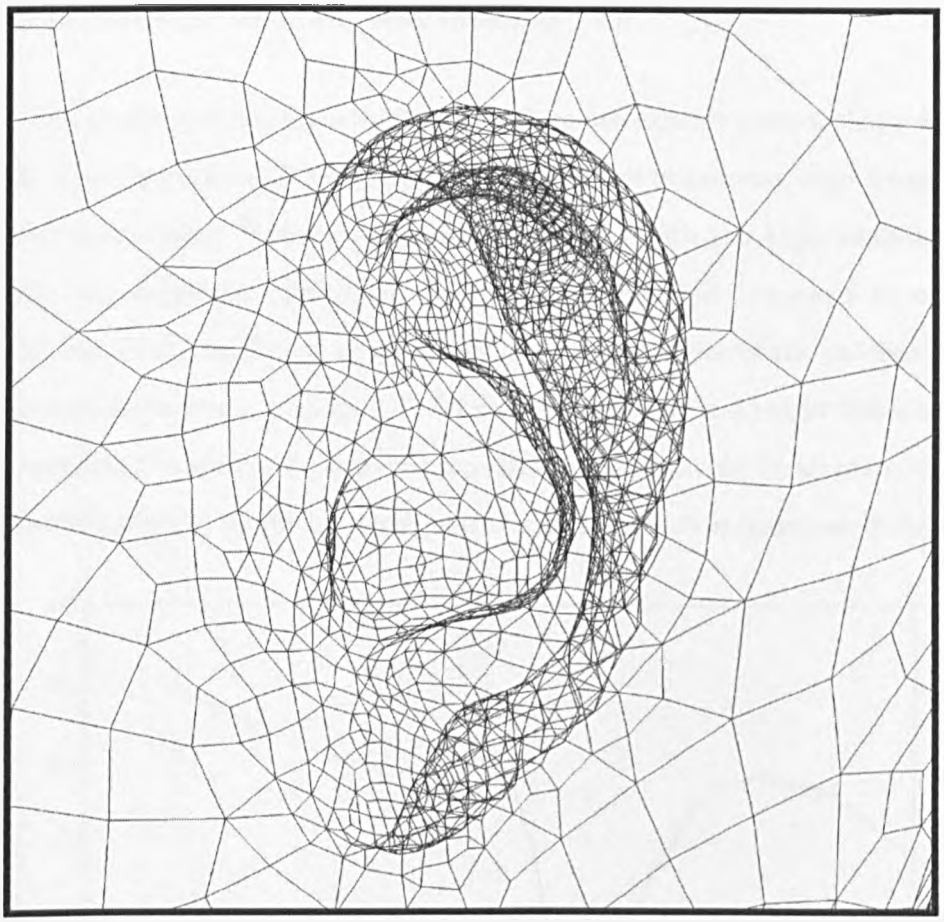


Figure 3.2: Close-up view of the transition between high-resolution meshing (in the pinna region) to low-resolution meshing (over the remainder of the head). The multi-resolution meshing algorithm (see Section 3.3.2) allows a smooth transition with intermediary target edge lengths.

the 7.8-14.8 kHz frequency range. Figure 3.3 shows the performance comparison in the case of the left (ipsilateral) HRTF. The subdivided multi-resolution mesh, is in near perfect agreement with the high-resolution mesh over the entire range of simulated frequencies. The multi-resolution mesh results, however, show some deviations from those obtained with both higher patch count meshes. These deviations appear over the entire simulated fre-

quency range but are small, never exceeding 1 dB.

The subdivided multi-resolution mesh does not express greater shape detail than the non-subdivided one, it only reduces maximum edge length. After subdivision, however, the slight deviations with the high-resolution mesh are essentially eliminated. This indicates that the deviations do not originate from insufficient shape detail, but from the size of the patches in low-resolution areas. Figures 3.4, 3.5 and A.1 (Appendix A) show the same comparison for the front, back and top HRTFs, respectively. In all cases, the subdivided multi-resolution mesh and the high-resolution mesh are in near

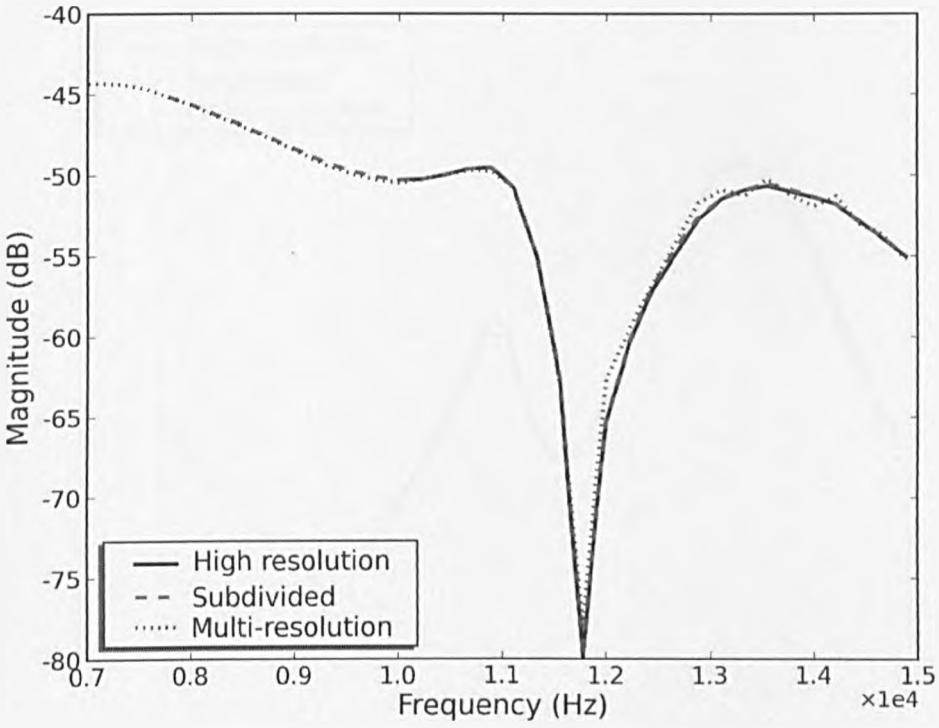


Figure 3.3: BEM simulation results for the left (ipsilateral) HRTF in the case of the multi-resolution mesh (3,816 polygons), a subdivided version (13,879 polygons) and a high-resolution mesh, with maximum edge length inferior to $\lambda/4$ at 15 kHz.

perfect agreement. The multi-resolution mesh results show, again, slight deviations from those obtained using the higher patch count meshes. These deviations remain mostly constrained to below 1 dB, occasionally slightly higher in the case of the back HRTF, in the 11.7-12.5 kHz range (see Figure 3.5). A very sharp notch in this frequency range is the main cause of the higher deviation values.

Figure 3.6 shows the performance comparison for the right (contralateral) HRTF. Results obtained using the subdivided multi-resolution mesh are in good agreement with those obtained using high-resolution mesh up to

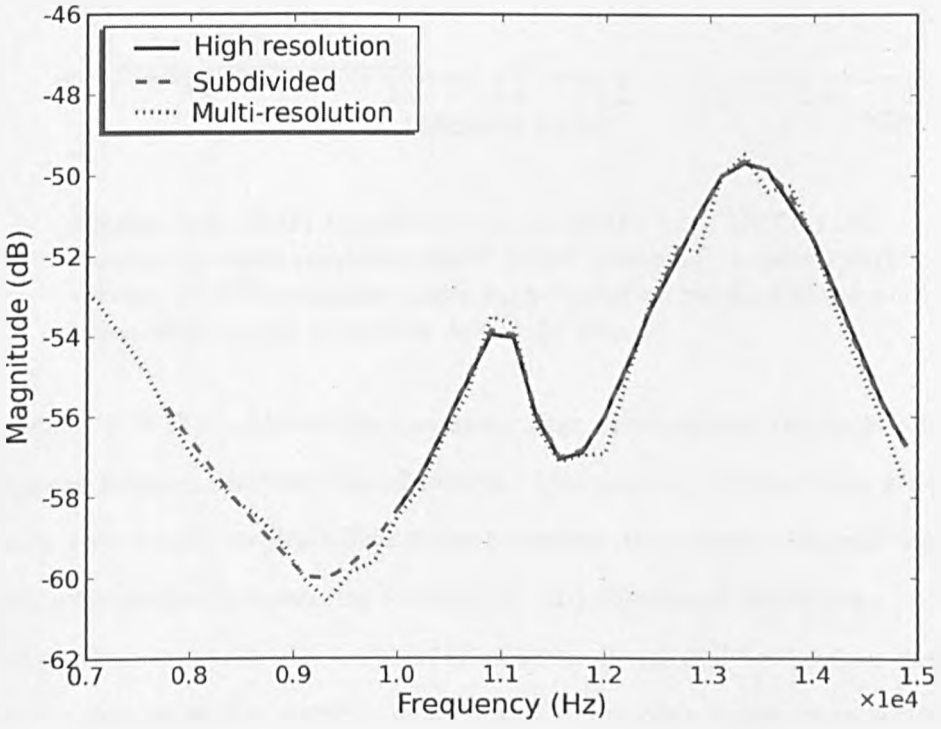


Figure 3.4: BEM simulation results for the front HRTF in the case of the multi-resolution mesh (3,816 polygons), a subdivided version (13,879 polygons) and a high-resolution mesh, with maximum edge length inferior to $\lambda/4$ at 15 kHz.

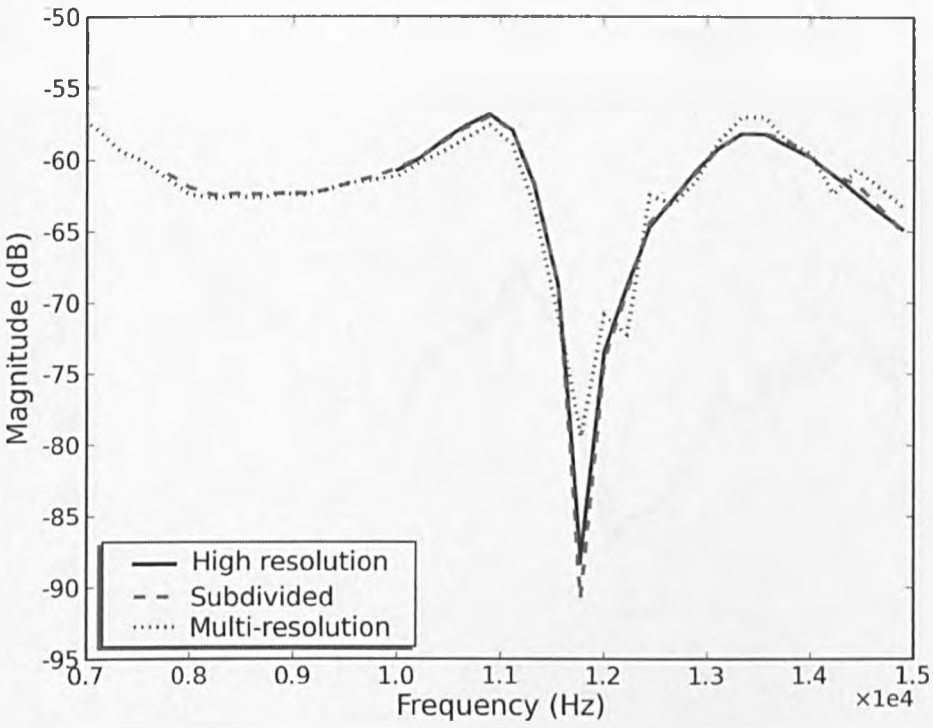


Figure 3.5: BEM simulation results for the back HRTF in the case of the multi-resolution mesh (3,816 polygons), a subdivided version (13,879 polygons) and a high-resolution mesh, with maximum edge length inferior to $\lambda/4$ at 15 kHz.

around 11.8 kHz. Above this frequency large discrepancies (up to 5-6 dB) appear between the two sets of results. The multi-resolution mesh generates very erratic results which strongly deviate from those obtained using the high-resolution mesh (up to about 10 dB). The effect of the large polygons which comprise the low-resolution areas of the multi-resolution mesh is far greater in the contralateral region, where they cause unacceptable distortions, than in the ipsilateral region.

In summary, the variable patch size employed in creating the database reduces the patch count, which allows pressures to be computed in a practical period of time at the expense of making the maximum frequency of analysis

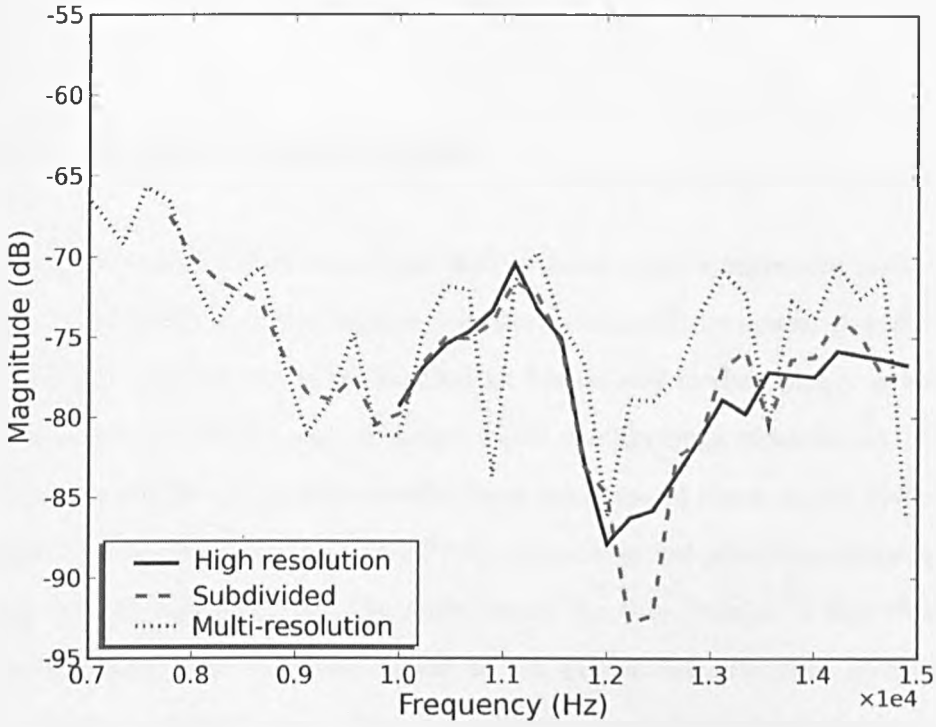


Figure 3.6: BEM simulation results for the right (contralateral) HRTF in the case of the multi-resolution mesh (3,816 polygons), a subdivided version (13,879 polygons) and a high-resolution mesh, with maximum edge length inferior to $\lambda/4$ at 15 kHz.

direction dependent. This dependence can be explained by a higher relative participation of the large head patches for sources in the contralateral region. A possible alternative to the multi-resolution meshing approach is to model a high-resolution pinna in isolation and use a low-resolution pinnaless head to compute low-frequency effects. Comparing the high-frequency acoustic character of a lone high-resolution pinna and that of a high-resolution head/pinna mesh would give some insight as to the role of the head at high frequencies, which must be relatively small if this approach is to be valid. However, this approach requires "combining" the low-frequency effects of a pinnaless head with the high-frequency effects of an isolated pinna which is

a non-trivial problem that must be solved.

3.4 Results and analysis

This Section presents an analysis of directional acoustic variations induced by the KEMAR head and pinnae over the entire auditory space. Attention is focused particularly on the horizontal, frontal and median planes as well as a number of lateral sagital planes which contain rings of confusion. Although a number of previous studies have made use of the auditory models described in Sections 2.3.2 and 2.3.3 to pre-process acoustic data, these are not used in this analysis. The main reason for this decision is that these models have been developed based on the perceptual awareness of tones masked by notched noise. Sound localisation and discrimination are two different psycho-acoustic tasks and it is clear that "at some point, the auditory processing that underlies localisation and spectral discrimination must be different" (Jin, 2001). It has been shown that auditory information is consciously perceived only when received and processed by cortical areas. Perceptual awareness in general is known to be the result of processes within the cerebral cortex. The primary auditory cortex is the first region of the cerebral cortex to receive auditory input and before reaching this late stage of the auditory pathway, neurally encoded signals have been relayed and processed by a number of stages in the deep auditory system, such as the superior olivary complex, the lateral lemniscus, the inferior colliculus and the medial geniculate nucleus (Berhrend *et al.*, 2003, 2004; Branoner *et al.*, 2005; McLaughlin *et al.*, 2008; Tollin *et al.*, 2008, amongst others, have given some insight into the processes which take place within these neural systems).

Conversely, it has been shown that at least some source location information is extracted very early in the auditory pathway, and is then relayed to the cerebral cortex as such, separate from actual sonic information (Chase and Young, 2005, 2008). Neural circuitry performs localisation-related spectral processing as early as the dorsal cochlear nucleus, where type IV cells have been shown to be particularly suited to detecting the spectral edges observed in HRTFs (Reiss and Young, 2005). While the information received by the dorsal cochlear nucleus has been been encoded in the basilar membrane, it has not gone through the later stages of the auditory pathway and the associated neural processing, which is inevitably applied to perceived sonic information. Disregarding the effects of intermediate stages of the auditory pathway, which pre-condition sonic information for the extraction of high level sonic attributes (such as timbre, pitch, rhythm, melody and words) in the auditory cortex seems an unnecessary risk, as a substantial amount of the original spectral and temporal detail could be lost during these processes. The very fact that the extraction of sound location information occurs, partly, early in the auditory pathway suggests the need to bypass later neural processing. These reservations are supported by reports that damage to the primary auditory cortex induces a loss of sonic awareness while the ability to react reflexively to sounds, in many instances, remains because of subcortical processing. The choice not to apply auditory filter models is therefore made on the basis that sound localisation and sonic awareness are two processes that are, at least in part, dissociated and that although spectral detail may be imperceptible in the context of studies described in Section 2.3.2, it could play a direct role in sound localisation.

3.4.1 Surface pressures and far-field pressures

Section 2.4.6.5 describes work by Kahana and Nelson (2005) relating acoustic simulation results to physical measurements performed by Shaw *et al.* (Shaw and Teranishi, 1968; Shaw, 1974). By placing a point source close to the ear canal, Kahana *et al.* used the reciprocity principle (see Section 2.4.6.4) to calculate the far field pressures for a large number of points with every simulation. They described the far field excitation patterns as a function of source position and frequency. A similar exercise, based on BEM-based simulations of the multi-resolution mesh described in Section 3.3.3, containing 3,816 polygons, is presented. Results are compared to those reported by Kahana and Nelson (2005). It should be noted that the mesh used here describes the entire KEMAR head, unlike those used by Kahana *et al.*, which described isolated pinnae.

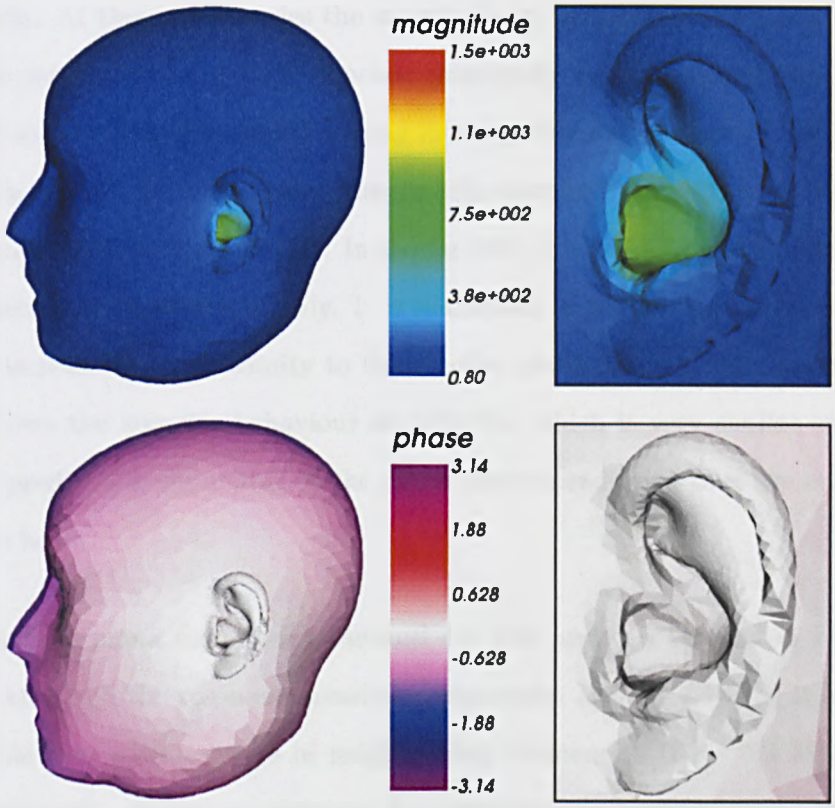
Kahana *et al.* also described the pressures generated by the source at the ear drum on the surface of the pinna meshes. To obtain similar data, Shaw *et al.* measured the acoustic pressure at different points in the pinna to investigate its modal behaviour, the transducers causing inevitable acoustic disturbances. During BEM-based acoustic simulations, however, surface pressures over the entire object mesh are inherently calculated, making a more detailed and accurate study of external ear acoustics possible. This exercise will also be conducted using the multi-resolution mesh described in Section 3.3.3 for comparison.

The best solution for displaying complex pressures on three-dimensional meshes is not necessarily obvious. Colour is necessary as the shading generated during the rendering would distort any grayscale colour-map. Kahana

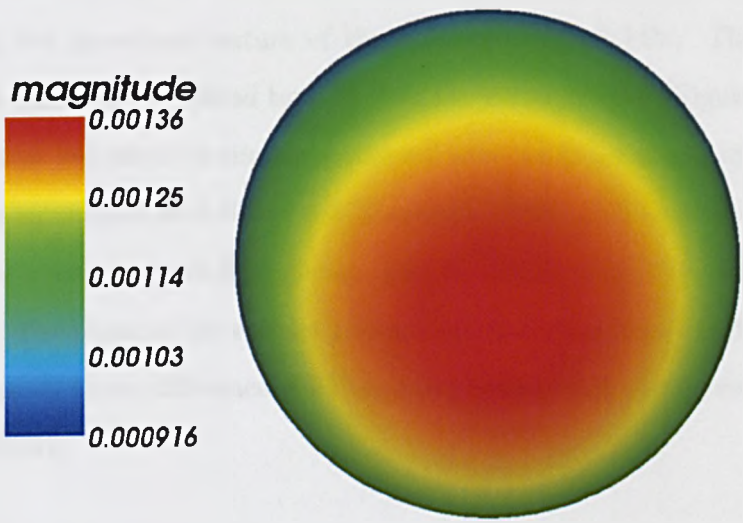
and Nelson (2005) assigned colours to each vertex corresponding to absolute pressure. Negative and positive values indicated positive and negative pressure phase. Using this method, two pressures with identical magnitude will be plotted as different colours if their respective phases are 1° and −1° but will be plotted as the same colour if the respective phases are 1° and 179°. This is undesirable. Kahana *et al.* noted this, but used the method to maintain consistency with the work of Shaw and Teranishi (1968). They suggested that a more accurate way of representing the results would be to plot real and imaginary parts or magnitude and phase separately. The latter approach has been adopted here as it reveals modal behaviour more clearly. The surface pressure magnitudes and phases are plotted separately on the mesh representation of the KEMAR head used for acoustic simulation.

Far-field pressures are calculated for points spread on the ipsilateral region of a 1 m radius imaginary sphere, centred on the origin (the centre of the KEMAR head). The far-field pressure magnitude is plotted on the surface of this sphere, with the view point on the ipsilateral side of the horizontal plane ($\theta, = -90°, \varphi = 0°$). The display orientation is the same as the one used for surface pressure plots. Phase is omitted in the far-field plots because it does not carry relevant information (except for well known ITD/IPD effects) unlike in the case of surface pressures, where it reveals the character of the modes. All pressures result from an omni-directional point source of unit magnitude placed 2.12 mm from the blocked ear-canal. The external ear nomenclature defined in Section 2.1.3 will be extensively used to describe acoustic phenomena throughout the analysis.

Figure 3.7 shows the acoustic behaviour of the KEMAR head and pinnae at 444 Hz. A similar result is observed for all frequencies up to around



(a) Surface pressures.



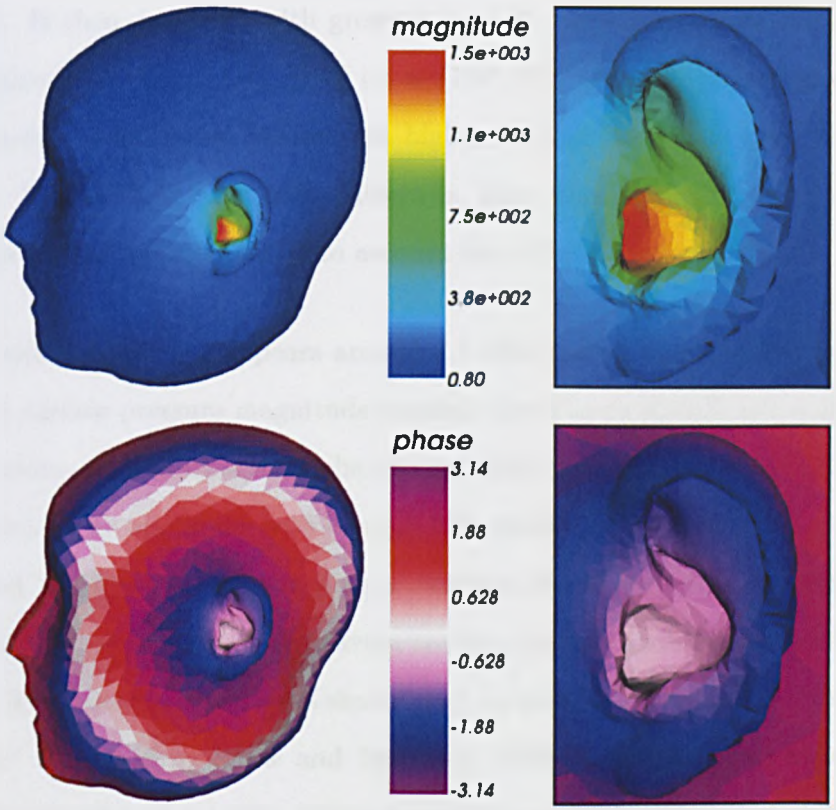
(b) Far field pressures.

Figure 3.7: KEMAR acoustics at 444 Hz. Pressure and phase are relative to a unit pressure point source close to the ear canal.

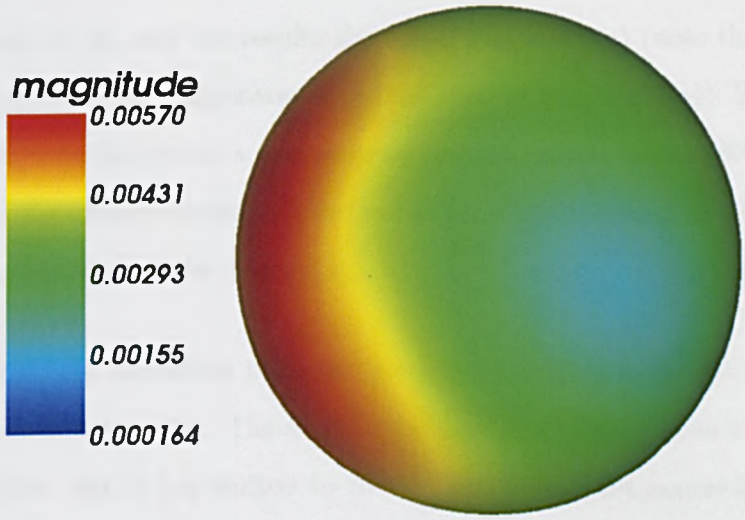
3.5 kHz. At these frequencies the magnitude of the pressure on the surface of the mesh (Figure 3.7(a)) depends principally on its distance from the point source. The wavelength is long in comparison to the size of the head and the surface pressure phase changes only marginally, also as a function of distance from the point source. In the far field (Figure 3.7(b)), the pressure magnitude changes very slowly. It is maximum in the far ipsilateral region and decreases with proximity to the median plane. Figure A.2 (Appendix A) shows the acoustic behaviour at 1333 Hz, which is very similar except for a predictable tightening of the phase pattern radiating over the surface of the head.

The first pinna mode peaks around 4.4 kHz and can be seen in Figure 3.8. Although the resonance reaches a maximum around 4.4 kHz, it is noticeable over a wide range of neighbouring frequencies (3.8 - 5.3 kHz approximately). Acoustic behaviour for neighbouring frequencies are shown in Appendix A (Figures A.4 (3778 Hz), A.5 (4000 Hz) and A.7 (5111 Hz)) showing the prominent nature of the resonance at 4.4 kHz. The surface pressure maximum is spread broadly over the entire concha (Figure 3.8(a)). This modal behaviour is similar to the 4.0 kHz mode identified by Kahana *et al.*, in agreement with Shaw and Teranishi (1968), although the different magnitude scale may, at first glance, lead to believe otherwise (see Section 2.4.6.5). The phase of the surface pressure starts to vary significantly within the pinna. A phase difference of 90° appears between the cavum concha and the antihelix.

The far field excitation pattern (Figure 3.8(b)) has also changed significantly compared to that observed at lower frequencies. A large covert peak area is now located in front of the listener, over a broad range of eleva-



(a) Surface pressures.



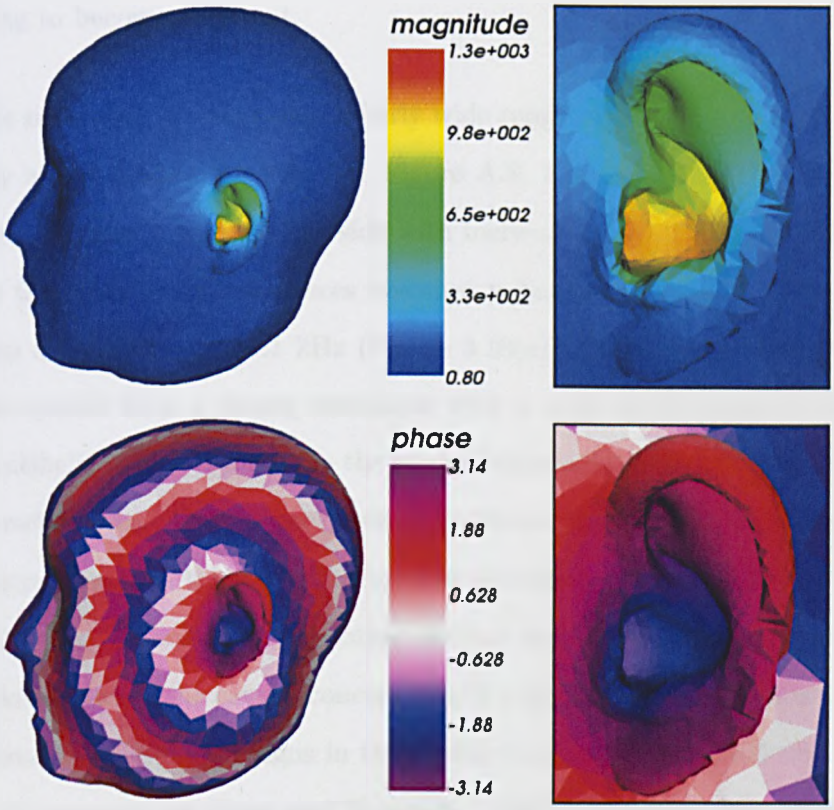
(b) Far field pressures.

Figure 3.8: KEMAR acoustics at 4222 Hz. Pressure and phase are relative to a unit pressure point source close to the ear canal.

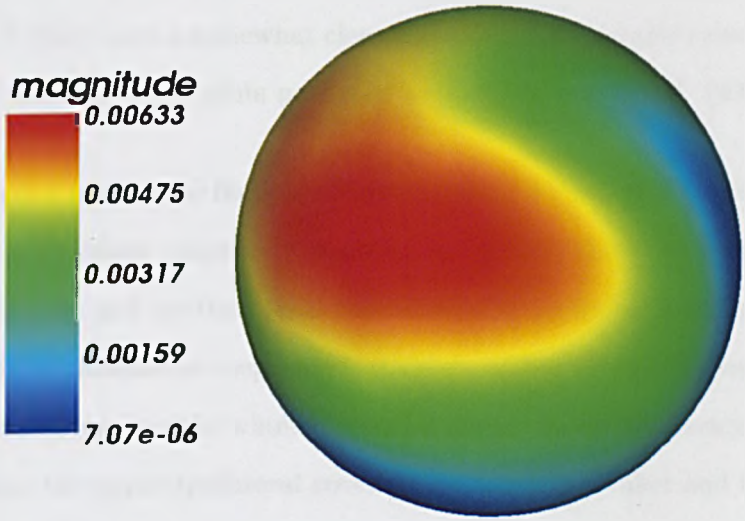
tions. It then decreases with growing azimuth, with a region of minimum magnitude in the vicinity of $(\theta, \varphi) = (120°, 0°)$. The elevation-dependent character of excitation patterns at higher frequencies is becoming noticeable. The far-field excitation pattern is, also, similar to that reported by Kahana *et al.* when taking into account the different colour scales.

Another resonance appears around 7.1 kHz and is shown in Figure 3.9. Local surface pressure magnitude maxima (see Figure 3.9(a)) are visible in the cavum concha, cymba concha and antihelix (in order of decreasing magnitude). It is consistent with the 6.8 kHz mode identified by Kahana and Nelson (2005) who describe it as a "vertical dipole pattern" involving an "oscillatory flow between the cavum concha, cymba concha and the antihelix". This was identified by Kahana *et al.* as corresponding to the "vertical mode" identified by Shaw and Teranishi (1968) whose acoustic measurements placed it at 7.1 kHz. Kahana and Nelson (2005) noted that the resonance is weak in comparison with others. This observation is in agreement with Shaw *et al.* and the results shown on Figure 3.9(a) (note the reduced magnitude scale in comparison with that used in Figure 3.8(a)). The phase variation over the pinna is now very significant. A full flip of 180° appears between the cavum concha and the antihelix, which is now actively involved in the production of the resonance.

The far field excitation patterns (see Figure 3.9(b)) have once again increased in directionality. The main covert peak is still noticeable in the front hemisphere, but it has shifted to the the ipsilateral side somewhat and is spread over a smaller range of elevations above the horizontal plane. An area of low pressure magnitudes has appeared in the median region of the back and low hemispheres due to the effect of pinna shadowing, which is



(a) Surface pressures.



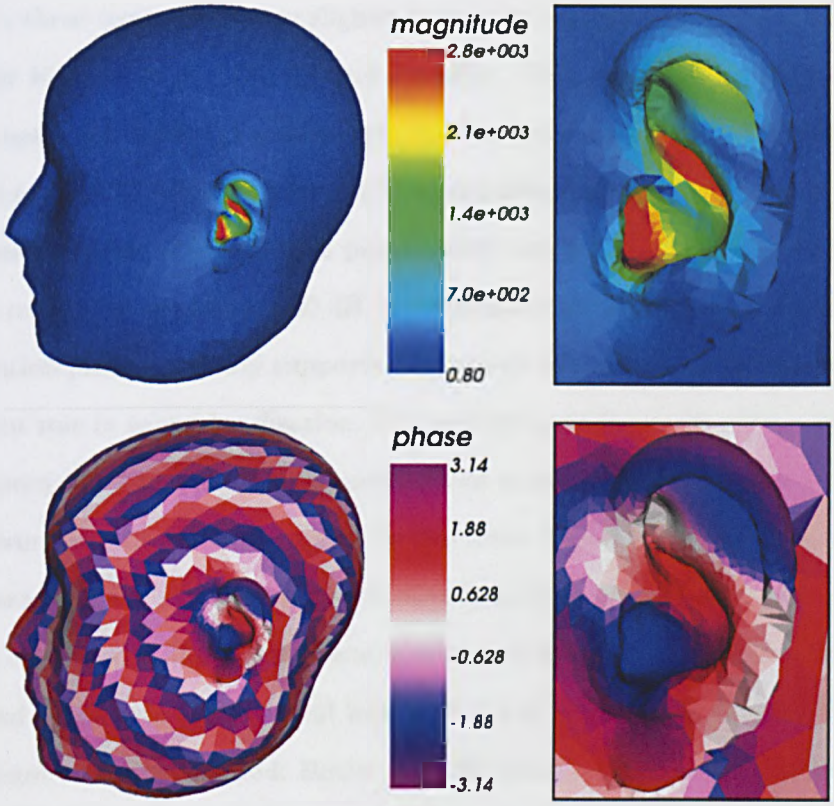
(b) Far field pressures.

Figure 3.9: KEMAR acoustics at 7111 Hz. Pressure and phase are relative to a unit pressure point source close to the ear canal.

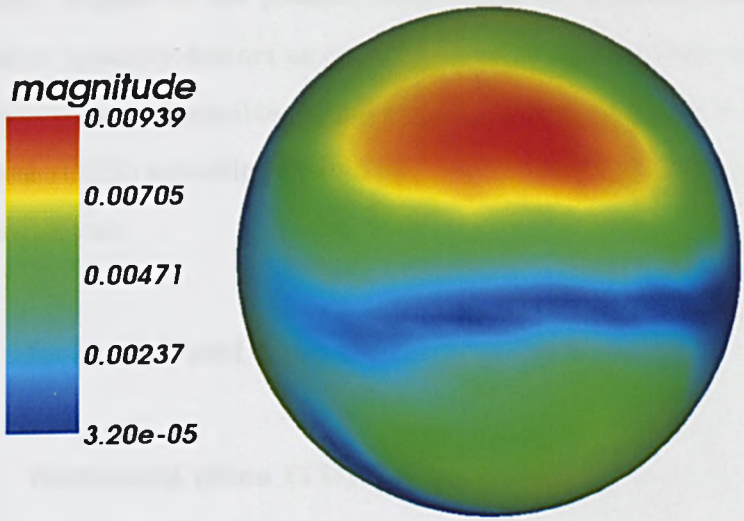
starting to become apparent.

This resonance is visible over a fairly wide range of frequencies and is still clearly active around 8.5 kHz (see Figure A.8, Appendix A). The far-field covert peak slowly shifts to the side with increasing frequency. The brightest of the three main resonances becomes noticeable around 9.5 kHz and reaches a peak around 11.1 kHz (Figure 3.10(a)). The cavum concha and cavum cymba form a strong resonance with a node at the crux of helias. The antihelix is also involved in the modal behaviour, although less so than the concha areas. The surface pressure patterns for this resonance seem to be compatible with the 9.5 kHz resonance identified by Kahana and Nelson (2005). They described it as another vertical mode also involving "oscillatory flow between the cavum concha, cymba concha and the antihelix", but this time "with different signs in the cymba concha and the antihelix" and was also reported by Shaw and Teranishi (1968) at 9.6 kHz. The separate magnitude and phase plots for the 7.1 kHz (Figure 3.9(a)) and 11.1 kHz (Figure 3.10(a)) give a somewhat clearer picture of the acoustic mechanisms at work than the single plots produced by Kahana and Nelson (2005).

The covert peak in the far field (Figure 3.10(b)) is very strong (note, again, the change in colour range in comparison to previous plots) in the upper ipsilateral region and another, local maximum appears in the lower ipsilateral region. Both maxima are separated by a strip of low far-field pressure magnitude extending over the whole horizontal plane. As the frequency is raised to 12 kHz, the upper ipsilateral covert peak becomes weaker and the lower ipsilateral covert peak becomes stronger (see Figures A.12, A.13, Appendix A).



(a) Surface pressures



(b) Far field pressures

Figure 3.10: KEMAR acoustics at 11111 Hz. Pressure and phase are relative to a unit pressure point source close to the ear canal.

The three main modes are slightly higher in frequency than those identified by Kahana *et al.* (4.0, 6.8 and 9.5 kHz). This can be explained by the fact that small KEMAR pinnae were used in our simulations (DB-61 as opposed to DB-65), which tends to shift modal behaviour up in frequency. The increase in pressure observed in pinna cavity resonances, which can in some cases reach and even exceed 30 dB, is remarkable and the spatial variation in excitation patterns clearly supports suggestions that covert peaks play a significant role in sound localisation. The perceptual salience of a given HRTF frequency band may lie just as much, if not more, with its contrast (higher *or* lower magnitude) with respect to the same frequency band in other directions than with its contrast with respect to other frequency bands in the same direction. Although the term covert peak is well accepted and its perceptual salience in terms sound localisation has been proven (Butler, 1987; Musicant and Butler, 1984; Butler and Musicant, 1993), spatial variations away from the covert peak area are significant and may also be perceptually salient. Regions of low pressure magnitudes, for example, are in some cases just as spatially distinct as covert peak areas making their salience as localisation cues very possible. These regions could be referred to as covert notch areas (CNA) extending the naming convention proposed by the aforementioned studies.

3.4.2 Horizontal and frontal HRTF variations

3.4.2.1 Horizontal plane ITD/ILD variations

The analyses of directional spectral variations in this section, as well as the following Section 3.4.3, investigate their "static" and "dynamic" character.

The former refers to the spectral features which potentially allow the location of a static sound source to be estimated, while the latter refers to the rate of spectral variation as a function of azimuth or elevation. This rate of spectral variation is compared to published experimental data describing changes in minimum audible angles (MAA) and minimum audible movement angles (MAMA) across the auditory space (see Section 2.3.1).

Source direction in the horizontal and frontal planes (see Section 2.1.1) is cued primarily by the ITD and ILD (see Section 2.2.1). The power of these cues stems mostly from their largely unambiguous information content and their operation over a broad range of frequencies (see Section 2.2.1.2). The perception of changes in azimuth is consistently better under binaural listening conditions than monaural ones. Given the ITD and ILD are overriding azimuth cues (Wightman and Kistler, 1992; Jin *et al.*, 2004) and that they invariably improve localisation acuity in the horizontal plane, it is likely that they are responsible for changes in horizontal localisation acuity with source direction (described in Section 2.3.1.1).

Different studies have reported quite different localisation acuity depending on experimental setup and source spectrum (Stevens and Newman, 1936; Mills, 1958; Oldfield and Parker, 1984a,b, 1986; Chandler and Grantham, 1992; Grantham, 1986). All agree, however, that the acuity of azimuth perception is best in front and deteriorates as sources move away from the median plane (as θ is increased from 0° to 90°). This effect occurs for low and high frequency pure tones as well as broadband sounds (Stevens and Newman, 1936). For sounds with no high frequency content, the ITD is the only accepted cue available for localisation. In order to explain this change in localisation acuity the concept of a rate of change of cue with direction is

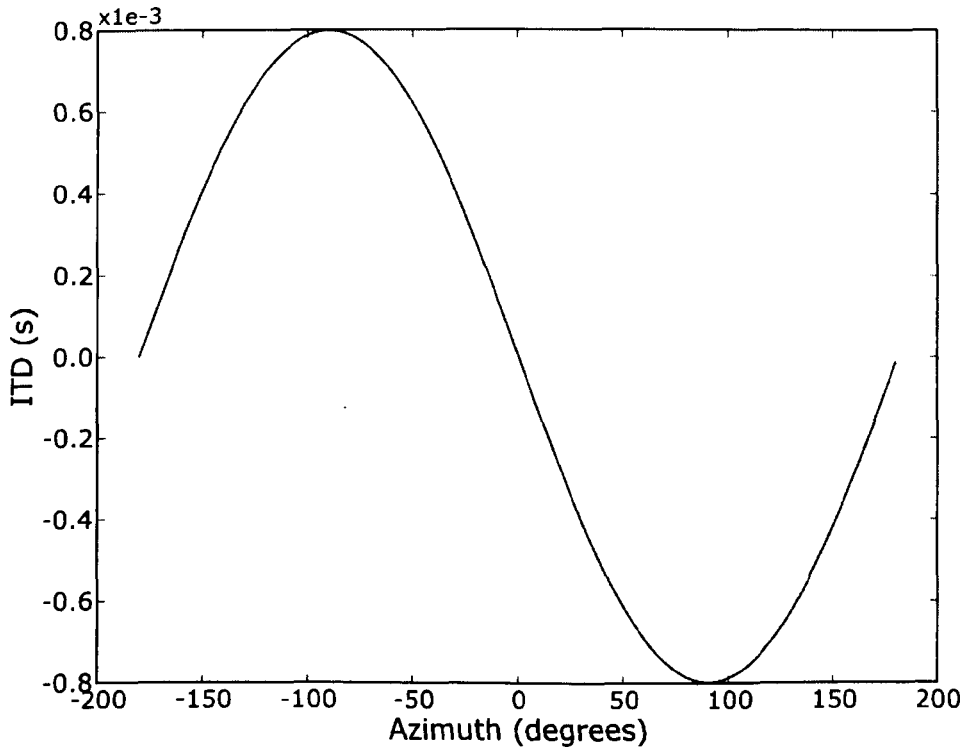


Figure 3.11: ITD variations with azimuth in the horizontal plane. The ITD is calculated as the phase difference at 200 Hz.

introduced. Figure 3.11 shows the ITD calculated for the KEMAR head as a function of azimuth in the horizontal plane. The ITD is calculated using the phase difference at 200 Hz. A vertical-polar spherical coordinate system (see Section 2.1.2) is used as it allows the entire horizontal plane to be swept with values $\varphi = 0$ and $-180° \leq \theta < 180$. The ITD traces a near-sinusoidal cycle peaking at $\pm$ 0.8 ms for a source on either side of the head ($\theta \approx \mp 90°$).

Figure 3.12 shows the absolute rate of change with azimuth (ARCA) of ITD variations. ITD-ARCA is maximum for sources in front ($\theta \approx 0°$) and at the back ($\theta \approx \mp 180°$). It decreases as the source moves away from the median plane to reach minima on either side of the head ($\theta \approx \mp 90°$). This pattern bears a strong similarity to variations in horizontal localisation

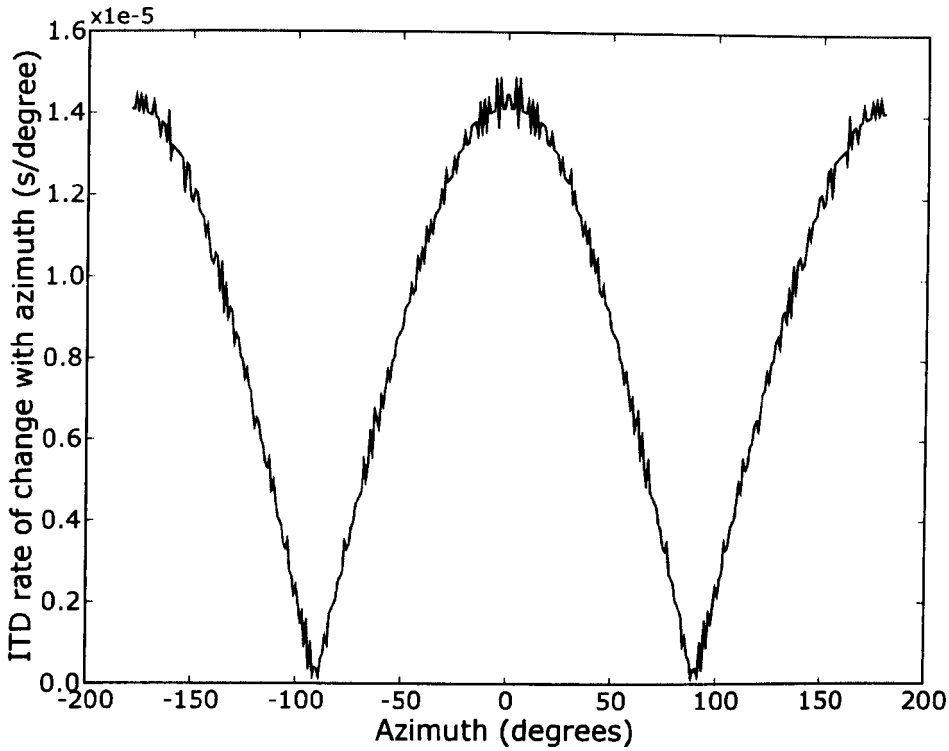


Figure 3.12: Absolute rate of ITD change with azimuth (ITD-ARCA) in the horizontal plane (in seconds per degree). The ITD is calculated as the phase difference at 200 Hz.

acuity described in Section 2.3.1.1. The disparity in reported localisation acuity measures makes a rigorous numerical assessment of the correlation difficult. It seems reasonable to suggest, however, that localisation acuity in the horizontal plane can be roughly determined from the local ITD-ARCA. A comprehensive study of MAA/MAMA variations in the horizontal plane is needed to confirm this suggestion. However, it has not yet, to the author's knowledge, been performed. Mathematically, the proposed relationship between ITD-ARCA and MAA can be expressed as

$$\mathrm{MAA}_{\theta}^{-1} \approx \frac{1}{p}\frac{d\,\mathrm{ITD}}{d\,\theta} \tag{3.8}$$

where MAA_θ is the minimum audible horizontal angle at azimuth θ, MAA_θ^{-1} is a measure of localisation acuity and p is the just noticeable ITD difference. Assuming an MAA at $\theta = 0°$ of around 1° (see Section 2.3.1.1), the ITD-ARCA for $\theta = 0$ in Figure 3.12 would suggest a value of approximately

$$p \approx 1.45 \times 10^{-5} \tag{3.9}$$

This relationship breaks down around $\theta = \pm 90°$, where observed MAAs are smaller than it predicts since the ITD-ARCA is essentially zero in these directions (see Figure 3.12). Perrott and Saberi (1990) showed that subjects can discriminate sources less than $\theta = 5°$ apart over the entire auditory space. A commonly reported 4-5 fold increase in MAA as source azimuth increases from $\theta = 0°$ to $\theta = 90°$ is compatible with this estimated MAA limit.

The change in ITD between $\theta = 0°$ and $\theta = 1°$ (a commonly reported just noticeable difference) is 14.4 μs (see Table 3.1 for ITD change calculations). This is compatible with ITD thresholds which are known to be in some cases as low as 10-20 μs (see Skottun *et al.*, 2001). By contrast, the change in ITD between $\theta = -90°$ and $\theta = -85°$ is 3 μs, a far smaller change. The change in ITD between $\theta = -90°$ and $\theta = -80°$ (far in excess of any modern agreed maximum MAA) is 11.5 μs, still smaller than the ITD change witnessed for a 1° change at $\theta = 0°$. Also, an increase in the bandwidth of the sound has been shown to drastically improve these MAA values (Butler, 1986; Chandler and Grantham, 1992). All these facts suggest that other localisation cues contribute to observed azimuth acuity in the lateral regions of the horizontal plane.

Azimuth change	ITD change
0° to 1°	$0 - 14.4 = 14.4\ \mu s$
−90° to −85°	$798.9 - 795.9 = 3\ \mu s$
−90° to −80°	$798.9 - 787.4 = 11.5\ \mu s$

Table 3.1: Azimuth change and associated ITD change

The relationship between localisation acuity and the rate of cue change is more difficult to assert in the case of the ILD as the operating range overlaps with that of spectral cues. There is evidence, however, that a similar process occurs as a drop in accuracy is observed for pure high frequency tones as well as low frequency ones (Stevens and Newman, 1936). The processes by which spectral cues are understood to operate (see Section 2.2.2) require the spectral content of sound sources to be spread in frequency and so cannot be involved in the case of pure tone stimuli. Using the same reasoning as for ITD and pure low frequency tones, the deterioration in horizontal localisation acuity as sound sources move from the front ($\theta = 0°$) to lateral areas ($\theta = \pm 90°$), in the case of pure high frequency tones, is plausibly attributable to the varying absolute rate of change of ILD with azimuth (ILD-ARCA).

Figure 3.13 shows variations in ILD with azimuth in the horizontal plane. ILD is calculated as the average magnitude difference across all simulated frequencies (0-14 kHz). This data should be accompanied with a note of caution as the multi-resolution mesh used for BEM simulations has been shown to generate somewhat erroneous results in contralateral areas at high frequencies (see Section 3.3.3). Figure 3.14 shows the ILD-ARCA. Although the variations generally follow a similar pattern to ITD variations there are a number of significant differences. The ILD-ARCA observed for sources around $\theta = 0$ is higher than would be produced by a sinusoidal approxima-

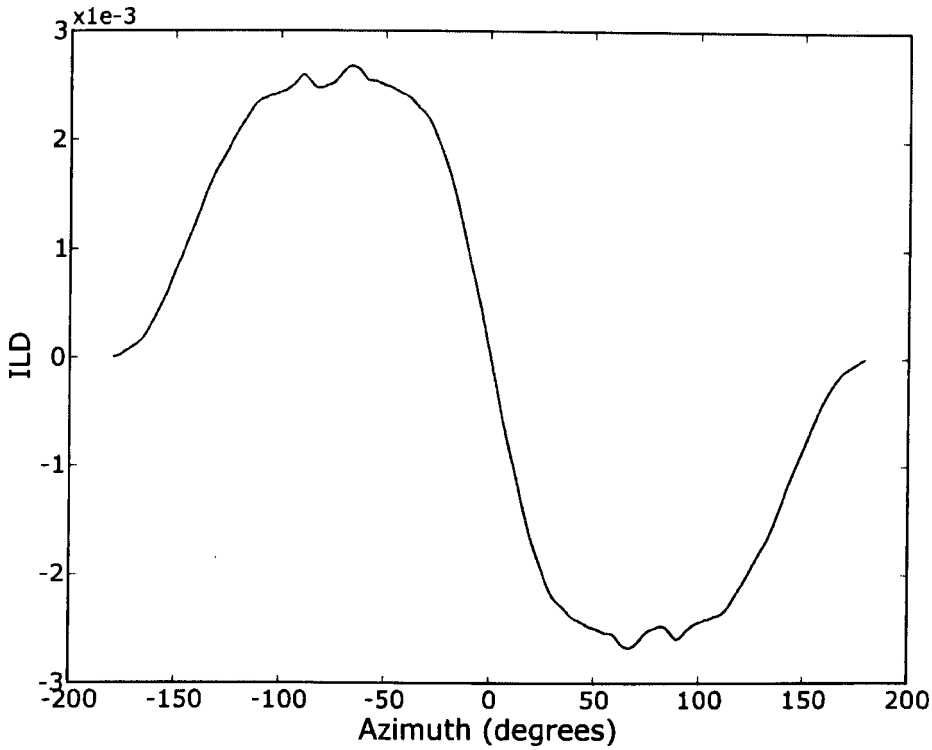


Figure 3.13: ILD variations with azimuth in the horizontal plane, calculated as mean linear magnitude difference across all simulated frequencies (0-14 kHz).

tion, which was highly accurate in the case of the ITD. It is at its highest in the horizontal plane. ILD changes in the $40° < |\theta| < 120°$ range are small, but include relatively rapid variations. These variations are most noticeable around $|\theta| = 90°$ and are attributable to creeping waves diffracted around the surface of the head, which creates a "bright spot" at the shaded contralateral ear. This effect is well known and has recently been shown to have an adverse effect on ILD based sound localisation (Rakerd *et al.*, 2008). Another significant difference appears around $\theta = 180°$ where the ILD variation decreases to a minimum, whereas maximum ITD variation is observed. This can be explained by the shadowing effect of the pinnae blocking high frequencies for sources at the rear, reducing the effects of source movements

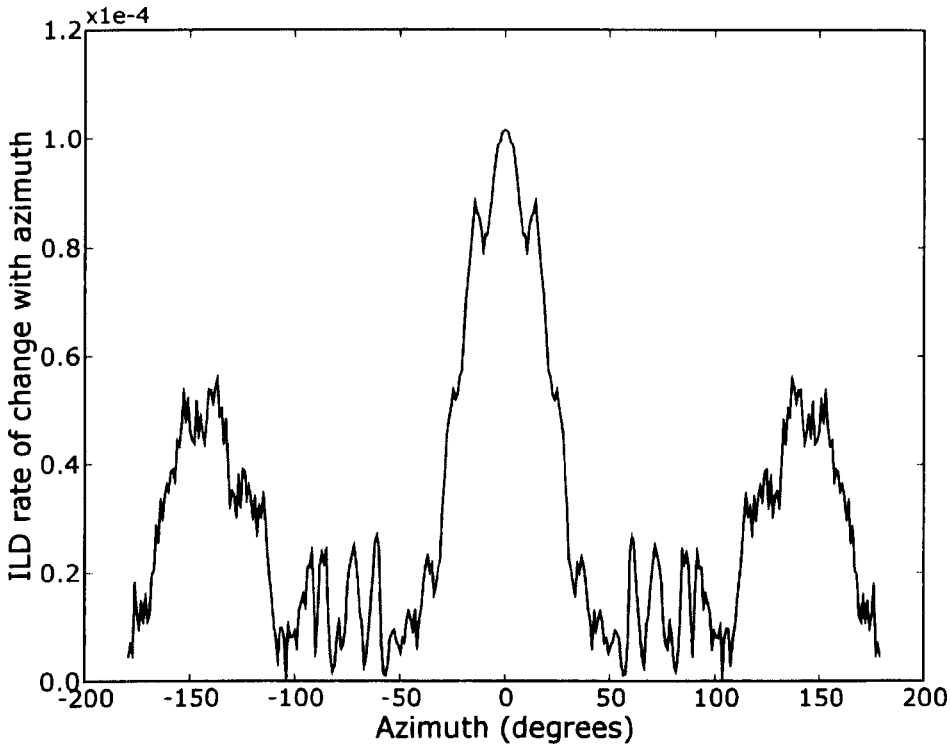


Figure 3.14: Absolute rate of ILD change with azimuth (ITD-ARCA), in the horizontal plane, in magnitude units per degree. The ILD is calculated as mean linear magnitude difference across all simulated frequencies (0-14 kHz).

on the overall ILD.

3.4.2.2 Horizontal plane spectral variations

ILD variations in the lateral regions ($|\theta| \approx 90°$) have been shown by Rakerd *et al.* (2008) not to improve horizontal localisation acuity. However, spectral variation with azimuth, shown in Figure 3.15, contain potentially powerful localisation information over a broad range of frequencies which could explain the increase in localisation performance with source bandwidth reported by a number of studies (Stevens and Newman, 1936; Butler,

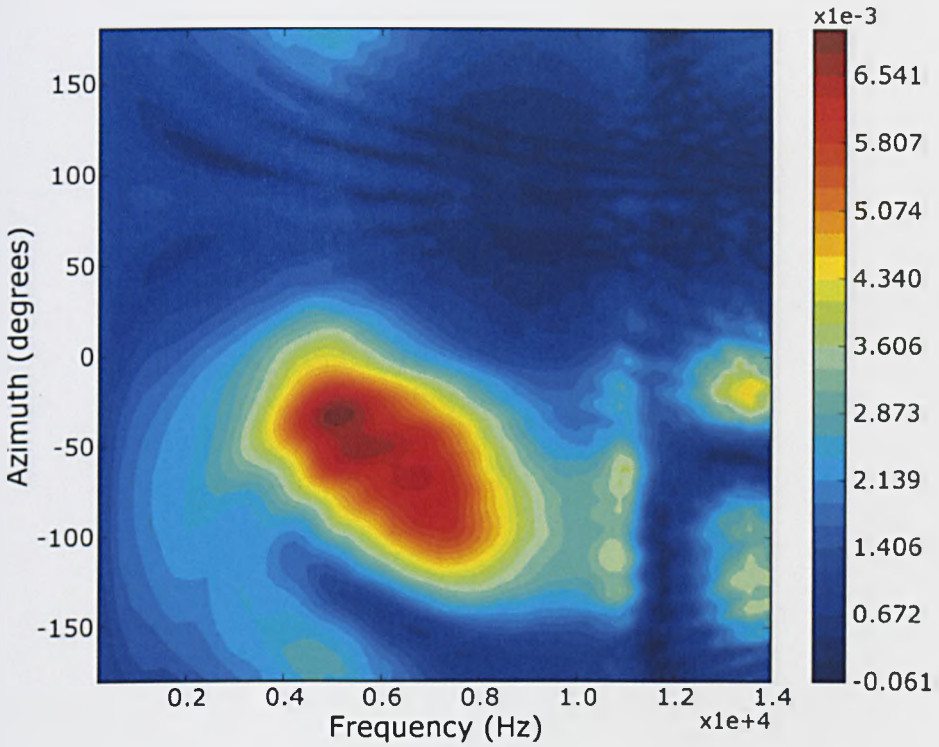


Figure 3.15: Spectral variations with azimuth in the horizontal plane up to 14 kHz in linear magnitude units (relative to a unit source placed close to the ear canal).

1986; Chandler and Grantham, 1992). Note that the ipsilateral hemisphere corresponds to $\theta < 0$ and that contralateral simulation data ($\theta > 0$) is prone to simulation errors at higher frequencies due to the multi-resolution meshing approach used (see Section 3.3.3).

As would be expected pressures are overall higher in the ipsilateral area than in the contralateral ear. Sharp variations in magnitude in the $-10° < \theta < 10°$ is particularly noticeable in the 4-8 kHz range. A strong falling notch edge from 11 kHz to 12 kHz is visible in the entire ipsilateral area and is particularly marked at the back $-150° < \theta < -90°$ as the 12 kHz notch becomes deeper making it a good static localisation cue candidate.

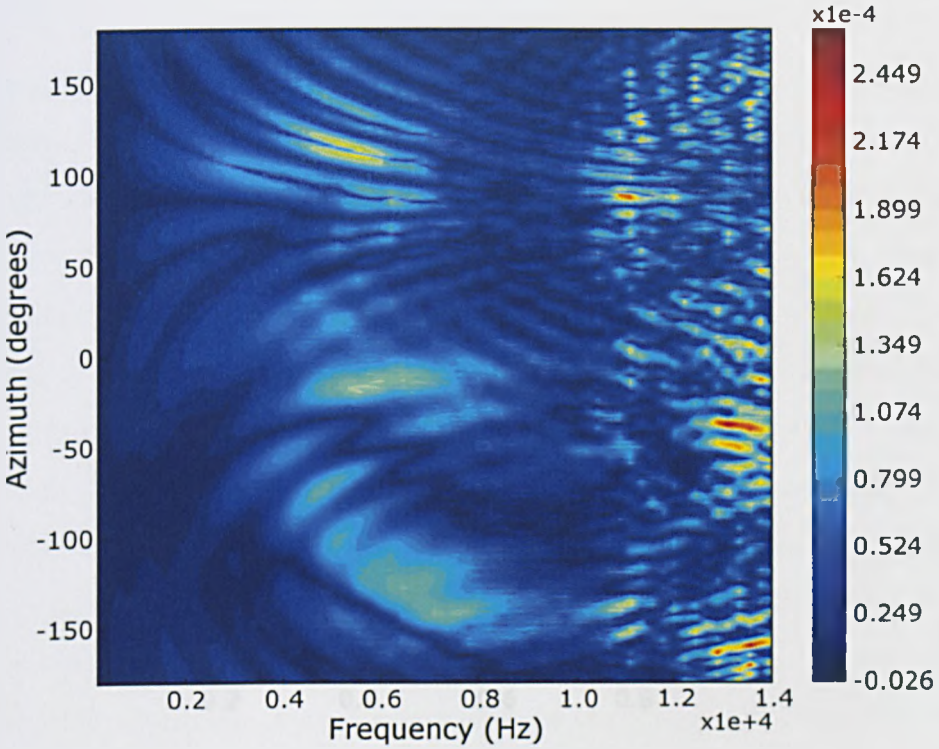


Figure 3.16: Absolute rate of spectral change with azimuth (HRTF-ARCA) in the horizontal plane up to 14 kHz in magnitude units (relative to a unit source placed close to the ear canal) per degree.

The spectral edges on either side of the major peak, in the centre of the plot (around 5 kHz and $\theta = -40°$), seem to be strong dynamic localisation cue candidates. A frequency shift occurs for the ascending spectral edge from around 8 kHz at $\theta = -140°$ to around 4 kHz at $\theta = -60°$, while the same frequency shift for the spectrally descending edge from $-60°$ to $0°$ is also visible. These changes have been reported by Lopez-Poveda and Meddis (1996), Mehrgardt and Mellert (1977) and Carlile and Pralong (1994). The shifts are relatively slow and unlikely to compete with ITD/ILD changes for small azimuths but could be central in explaining unexpected azimuth acuity in the far ipslilateral region (around $\theta = -90°$). Figure 3.16 shows the

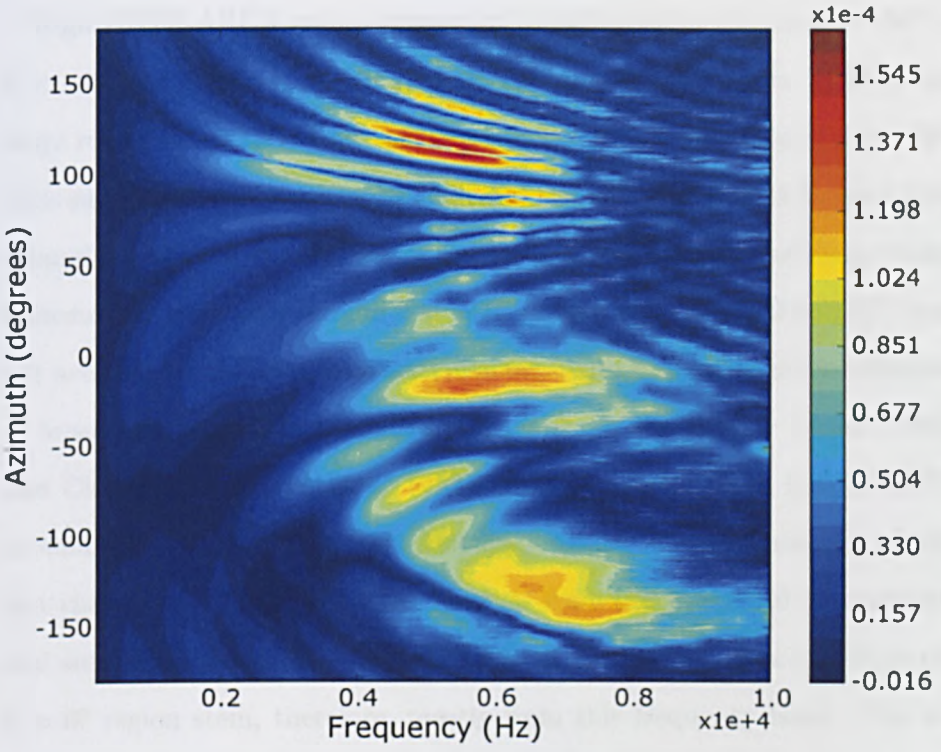


Figure 3.17: Absolute rate of spectral change with azimuth (HRTF-ARCA) in the horizontal plane up to 10 kHz in magnitude units (relative to a unit source placed close to the ear canal) per degree.

absolute rate of change with azimuth of the HRTF spectrum (HRTF-ARCA) in the horizontal plane. In this and subsequent plots of spectral variation rates, high levels of spectral variation occur in the 11-12 kHz band as a result of the complex pinna resonances. These high HRTF-ARCA values tend to occupy most of the colour range, drowning out lower frequency data. In order to better visualise the HRTF-ARCA below 10 kHz, a plot focussing on the 0-10 kHz frequency range with an appropriately rescaled colour bar, is shown in Figure 3.17.

High HRTF-ARCA values appear in broad areas in the range $-140° < \theta < -0°$ and 4 kHz $< f <$ 8 kHz. It varies erratically above 11 kHz, but stays relatively small in that spectral region for azimuths around $\theta = -90°$. This suggests that a consistently high and stable HRTF-ARCA below 9 kHz, around $-140° < \theta < -0°$, works alongside ITD and ILD providing crucial information, especially around $\theta = -90°$ (where both the ITD and ILD cannot account for observed localisation acuity) and improving the localisation of broadband sounds in the horizontal plane, as observed by Butler (1986) and Chandler and Grantham (1992) amongst others. High HRTF-ARCA around $\theta = 0°$, in the 4-8 kHz frequency range, is a manifestation of the fast change in ILD as the source changes from the ipsilateral to contralateral side of the head. The high ILD-ARCA values (see Figure 3.14) in the $\theta = 0°$ region stem, therefore, mostly from this frequency band. The absence of spectral variation under 3 kHz is a clear indicator of the small low-frequency spectral information content in terms of sound localisation (as suggested by Hebrank and Wright, 1974; Langendijk and Bronkhorst, 2002; Best *et al.*, 2005, amongst many others).

3.4.2.3 Frontal plane ITD/ILD variations

In order to plot frontal plane (see Section 2.1.1) variations a vertical polar spherical coordinate system (see Section 2.1.2) is used, with a slight modification. In this case, it is convenient to set $\theta = -90$ and let the elevation sweep the entire plane starting and ending below the modeled head ($-90° < \varphi < 270°$) in order to allow a continuous plot (see Figure 3.19).

Figure 3.18 shows ITD variations calculated as the phase difference at 200 Hz. The variation is, once again, very smooth and the maximum rates

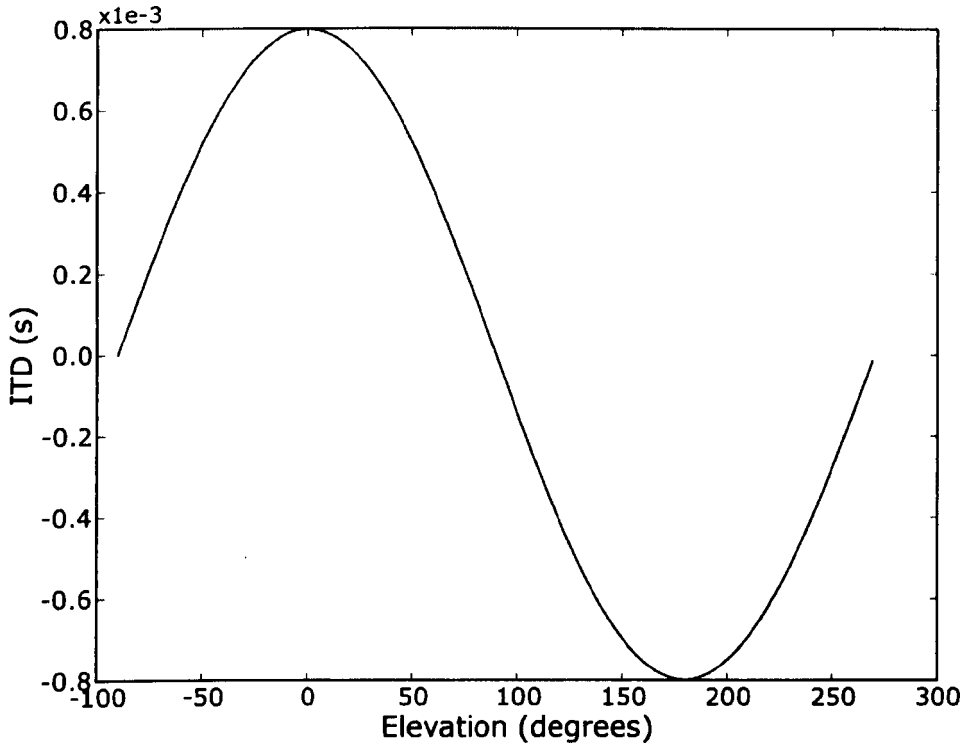


Figure 3.18: ITD variations with elevation in the frontal plane. The ITD is calculated as the phase difference at 200 Hz.

of change with elevation occur, unsurprisingly, below and above the head ($\varphi = -90°$ and $\varphi = 90°$) with minimum rates on the sides ($\varphi = 0°$ and $\varphi = -180°$, translating to $(\theta, \varphi) = (-90°, 0°)$ and $(\theta, \varphi) = (90°, 0°)$ in normal vertical polar spherical coordinates).

There is very little published experimental work investigating minimum audible angles directly above and below listeners ($\varphi = -90°$ and $\varphi = 90°$) for sources moving perpendicularly to the median plane. However, the ITD absolute rate of change with elevation (ITD-ARCE) for these movements are essentially identical to the ITD-ARCA (see Section 3.4.2.1) in the frontal ($\theta = 0°$, $\varphi = 0°$) region, which is a maximum in the entire auditory space. Since the ITD is the only cue operating at low frequencies, it seems likely

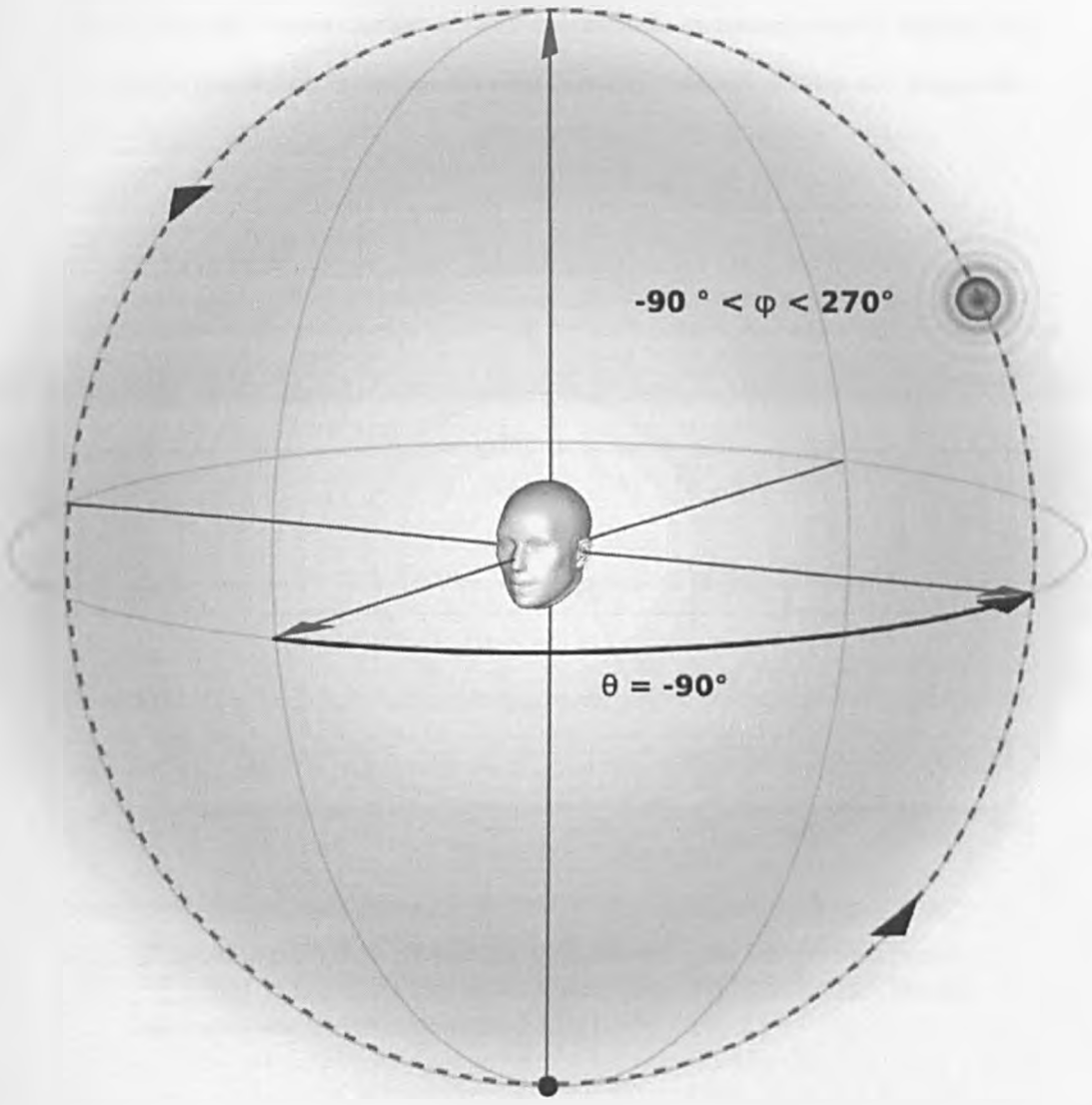


Figure 3.19: Altered vertical spherical polar coordinate system used for frontal plane plots. Azimuth is set to $\theta = -90°$. The black dot shows the starting point ($\varphi = -90°$), which is also the ending point ($\varphi = 270°$).

that, for low frequency sound sources, the MAA perpendicularly to the median plane directly above and below the listener will be very similar

to those observed for θ variations in the frontal region (around 1°). This reasoning can be extended to any movements perpendicularly exiting the median plane for, in all cases, the absolute ITD variation rates are maximum.

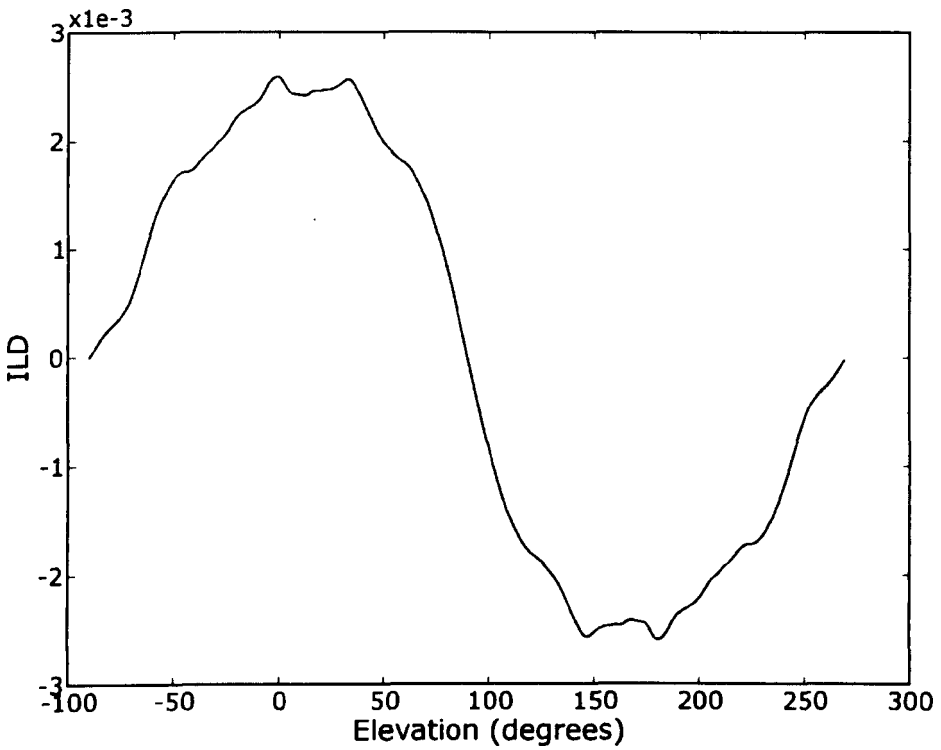


Figure 3.20: ILD variations with elevation in the frontal plane. The ILD is calculated as mean linear magnitude difference across all simulated frequencies (0-14 kHz).

Figure 3.20 shows ILD variations over the frontal plane using the same modified vertical-polar spherical coordinate system (see Figure 3.19) for $\theta = -90°$ and $-90° < \varphi < 270°$. As in the case of the ITD, the similarity with horizontal plane variations is clear, although the relationship between elevation and ILD is slightly more turbulent in the left and right directions ($\varphi = 0°$ and $\varphi = 180°$). The effects of creeping waves is again visible for

these directions, where they create a "bright spot", causing relatively rapid fluctuation. The local slope of these fluctuations is not far from the highest levels observed for sources moving perpendicularly away from the median plane, however, the turbulent nature of the variations causes them to disturb localisation processes rather than improve them (as shown by Rakerd *et al.*, 2008).

The absolute rates of change of ITD and ILD with elevation (ITD-ARCE and ILD-ARCE) are plotted in Figures B.1 and B.2 (Appendix B). The ITD-ARCE in the frontal plane is essentially identical to the ITD-ARCA in the horizontal plane (see Figure 3.12). The ILD-ARCE in the frontal plane, however, is significantly different from the ILD-ARCA in the horizontal plane (see Figure 3.14). This difference stems from pinna shadowing effects, which differ strongly for movements in the frontal and horizontal planes. This difference is only significant at higher frequencies, hence significantly different ILD variations and similar ITD variations.

3.4.2.4 Frontal plane spectral variations

Figure 3.21 shows the spectral variation as a function of elevation, again using the modified vertical-polar spherical coordinate shown in Figure 3.19, for $\theta = -90°$ and $-90° < \varphi < 270°$. The activity seems to be concentrated in the ipsilateral side ($-90° < \varphi < 90°$), more specifically around $-20° < \varphi < 50°$. The 11.1 kHz pinna resonance identified in Section 3.4.1 creates a very strong covert peak around $10° < \varphi < 40°$ and the 7.1 kHz pinna resonance also produces a covert peak around $-20° < \varphi < 20°$. The roll-off for this covert peak starts around 6 kHz at $\varphi = -40°$ and smoothly increases to around 8 kHz as the source reaches the horizontal plane. A

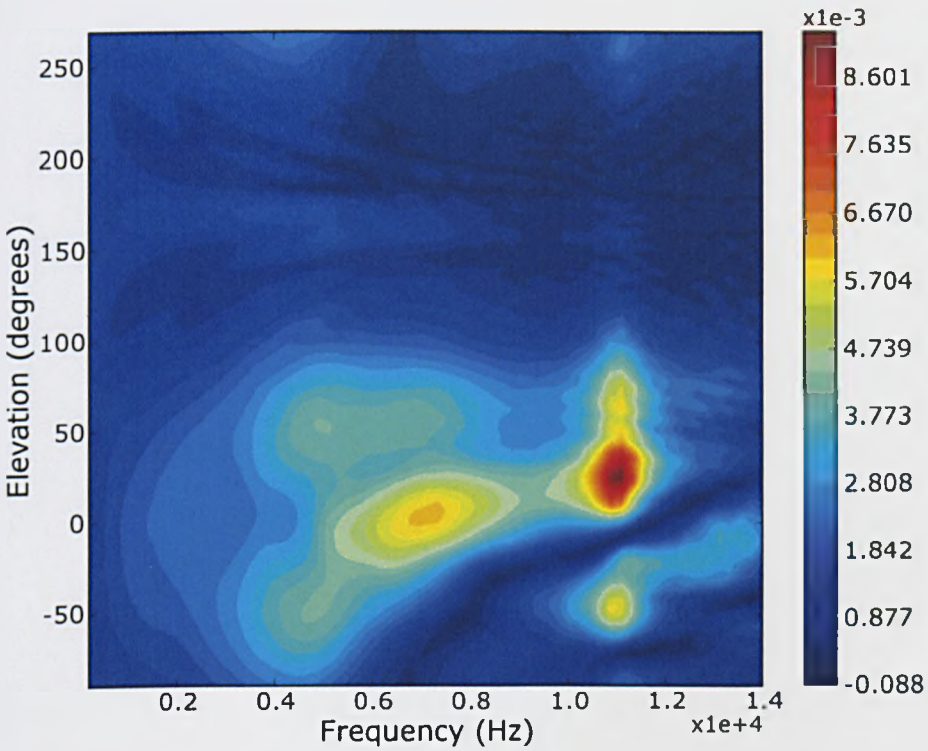


Figure 3.21: Spectral variations with elevation in the frontal plane, up to 14 kHz, in linear magnitude units relative to a unit source placed close to the ear canal.

similar increase was noted by Carlile and Pralong (1994) at slightly lower frequencies. The absence of spectral variation under 3 kHz and in the contralateral side suggests, as in the case of the horizontal plane, a small spectral information content for localisation.

Figure 3.22 shows the absolute spectral rate of change with respect to elevation (HRTF-ARCE) in the frontal plane. The strong 11.1 kHz covert peak (see Figure 3.21) creates a very prominent HRTF-ARCE peak in its surroundings ($-10° < \varphi < 20°$ in the 10-12 kHz range). The values in this region are so large they relegate the rest of the plot to the lower third of the colour map. Variation rates in the 0-10 kHz frequency band can be far

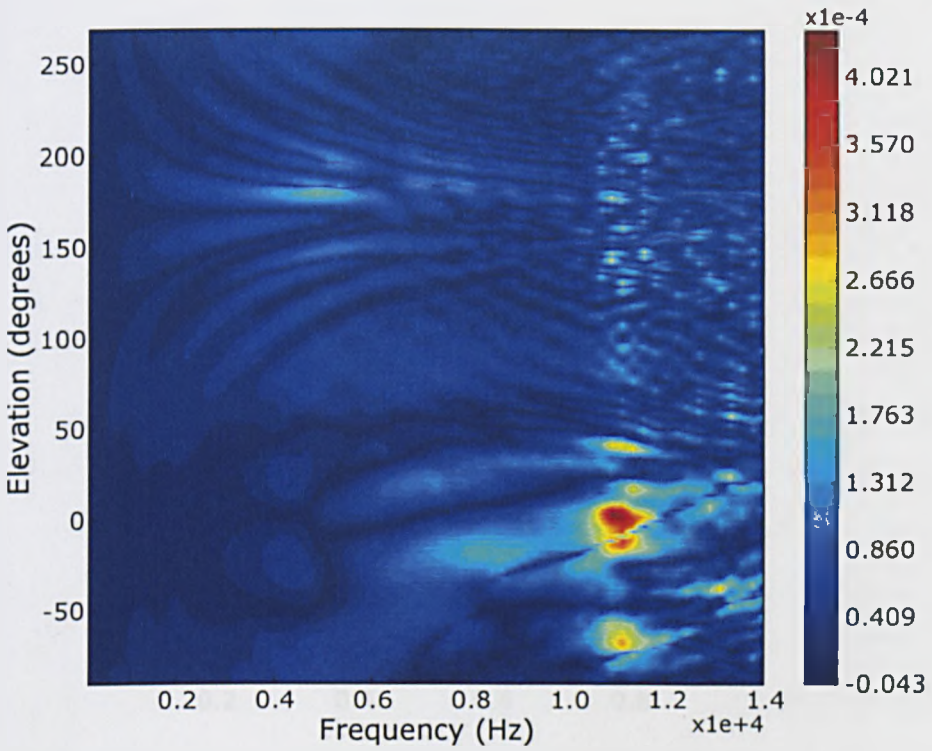


Figure 3.22: Absolute rate of spectral change with elevation, in the frontal plane, up to 14 kHz, in magnitude units (relative to a unit source placed close to the ear canal) per degree.

better interpreted in Figure 3.23 where the frequency range has once again been limited to under 10 kHz, as explained in Section 3.4.2.2. Although the 7.1 kHz resonance creates a clearly visible peak in Figure 3.21, the HRTF-ARCE values it generates are fainter than those observed around 8-9 kHz. This is caused by the larger spatial spread (hence a slower change with elevation) in the case of the 7.1 kHz covert peak. Indeed, although the 8-9 kHz covert peak area ($0° < \varphi < 30°$) is fainter, its proximity to a deep spectral notch ($-30° < \varphi < -10°$) generates relatively high HRTF-ARCE values in that frequency band.

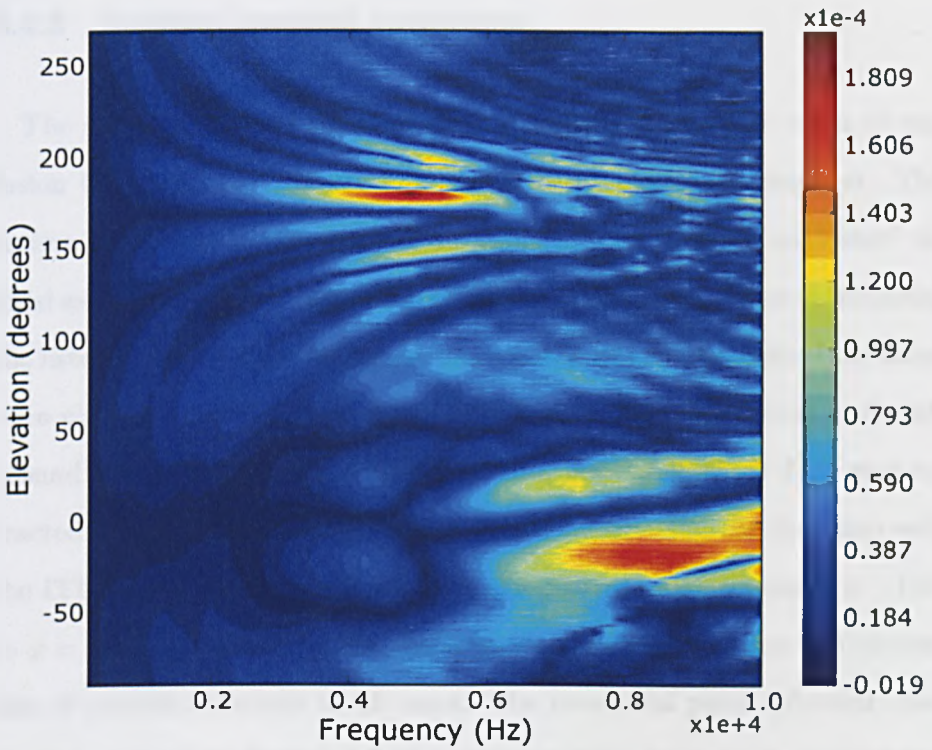


Figure 3.23: Absolute rate of spectral change with elevation in the frontal plane, up to 10 kHz, in magnitude units (relative to a unit source placed close to the ear canal) per degree.

As in the case of the horizontal plane, dynamic spectral cues seem in the ipsilateral region (around $\varphi = 0°$) provide information which allows better localisation acuity than that expected from weak ITD/ILD variations. It is interesting to note that the high spectral variation rates which allow good localisation acuity fall mostly within lateral regions where ITD and ILD variations provide insufficient information. This could be a coincidence, but it may be that evolutionary trends have shaped the pinna and its acoustic characteristics in order to rectify the shortcomings of ITD/ILD cues in these regions.

3.4.3 Sagittal spectral variations

The study of spectral variations is particularly relevant on cones of confusion (Wallach, 1939) where ITD and ILD are essentially constant. This section focuses on HRTF spectral variations around "rings of confusion" defined as the intersection between cones of confusion and a sphere surrounding the listener. In this case the sphere has a radius of 1 m. Acoustic simulations were carried out for a dense set of HRTF directions distributed uniformly around the KEMAR head. Using this data, lines of constant ITD were extracted. Each ring of confusion was traced by tracking the direction with the ITD closest to the target for each elevation φ ranging from $\varphi = -180°$ to $\varphi = 180°$ in sequence. This gave a very good approximation to a circular line of constant azimuth in all cases. The interaural polar spherical coordinate system (see Section 2.1.2) is very practical here, as it allows whole rings of confusion to be swept by varying the elevation, the azimuth remaining constant. In each case an approximate value for this azimuth is given. In the context of this analysis, the terms "covert peak" and "covert notch" will be used to describe local as well as global magnitude maxima and minima, respectively, at a given frequency, within the ring of confusion. This is as opposed to defining such terms in a spatially unconstrained sense over the entire auditory space.

3.4.3.1 Ring of confusion variations for ITD = 0.0ms

Spectral variations with elevation

Figure 3.24 shows spectral variations around the ITD = 0.0 ms ring of confusion (median plane) across the whole range of simulated frequencies. The entire ring of confusion can be swept by varying φ between $-180°$ and

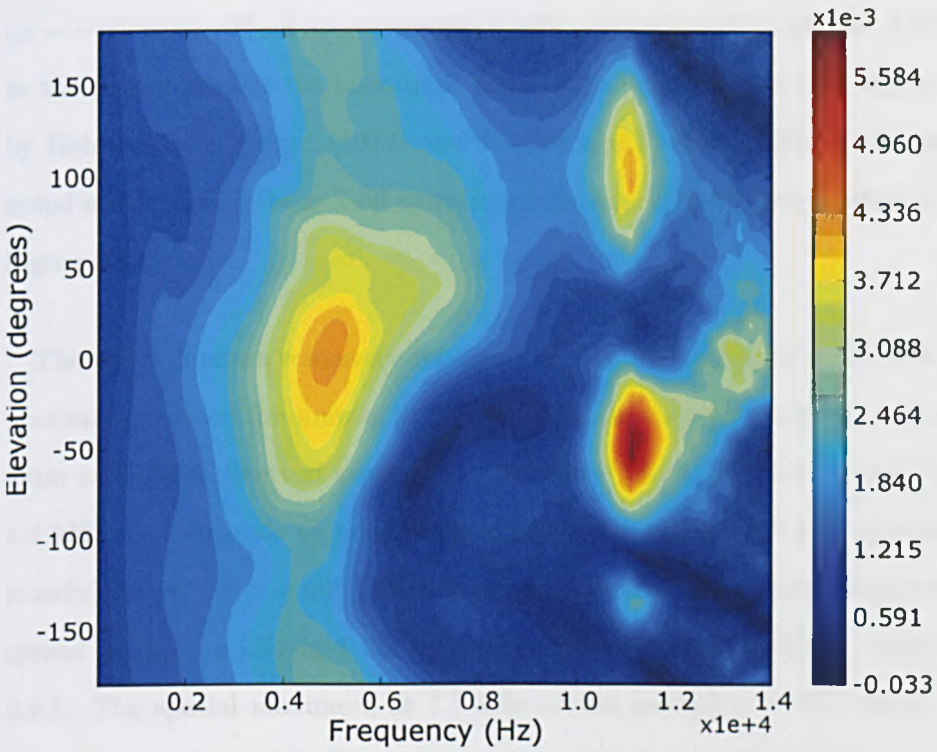


Figure 3.24: Spectral variations with elevation in the median plane (ITD = 0.0 ms), up to 14 kHz, in linear magnitude units relative to a unit source placed close to the ear canal.

180°, with a stable azimuth, $\theta \approx 0°$. Simulation results will be compared to those reported by Kahana and Nelson (2005), who measured the pinna-related transfer functions (PRTFs) for an isolated pinna, for sound sources at grazing incidence. This amounts to a slightly translated version of the median plane investigation presented here, where the origin is placed at the center of the head as opposed to at the ear itself.

As in the case of the horizontal and frontal planes, spectral variation is very small under 3 kHz. The 4.4 kHz pinna resonance described in Section 3.4.1 produces a strong covert peak in the $-20° < \varphi < 20°$ region, although its effects are noticeable over the entire elevation range. At lower elevations

($\varphi = -60°$) the roll off starts around 4 kHz and increases to around 5 kHz as the source reaches the horizontal plane. This increase was also reported by Hebrank and Wright (1974) and Carlile and Pralong (1994), who also noted a decrease in the roll-off slope in agreement with the results shown in Figure 3.24.

The 4.2 kHz pinna resonance reported by Kahana and Nelson (2005) was maximally excited for sources at $\varphi = 16°$, which is consistent with a maximum excitation observed around $\varphi = 10°$ here. The bright area around the 4.4 kHz peak extends to neighbouring frequencies (3.8 - 7.5 kHz approximately) for $40° < \varphi < 60°$. This is presumably due to the large frequency spread of the 4.4 kHz and 7.1 kHz pinna resonances identified in Section 3.4.1. The spatial maximum at 7.1 kHz occurs around $\varphi = 50°$, which is compatible with the 7.2 kHz maximum identified at $\varphi = 60°$ by Kahana and Nelson (2005). The covert peak is fainter and and more localised than in the case of the 4.4 kHz pinna resonance as would be expected from the far field plots shown in Section 3.4.1 (Figure 3.9).

The most noticeable peaks appear in a narrow higher frequency band (10.5 - 11.8 kHz approximately) resulting from the pinna resonance identified at 11.1 kHz in Section 3.4.1. Two covert peaks are clearly visible, a very strong one in the $-70 < \varphi < -20°$ range and a fainter one in the $80 < \varphi < 120°$ range. The spatial distribution of these covert peaks can be related to the covert peaks observed at $\varphi = 94°$ and $\varphi = -40°$ by Kahana *et al.* around 9.5 kHz. The frequency shift is attributable to the smaller size of KEMAR ear used in our simulations. A third 11.1 kHz peak, fainter still, can be seen around $\varphi = -140°$. The major 11.1 kHz covert peak around $-70 < \varphi < -20°$ extends into higher frequencies with growing elevation

and, although it becomes weaker, it remains prominent. It then regains some strength around 14 kHz, peaking around $\varphi = 0°$ (in agreement with the 14.8 kHz covert peak at $\varphi = 4°$ reported by Kahana and Nelson, 2005). This shift causes a smooth monotonic variation in peak roll-off frequency (11 kHz to 14 kHz, approximately) as elevation rises from $\varphi = -90°$ to $\varphi = 0°$. Changing roll-off frequencies for a notch around this frequency have also been reported in previous studies (Hebrank and Wright, 1974; Butler and Belendiuk, 1977; Carlile and Pralong, 1994).

Hebrank and Wright (1974) and Langendijk and Bronkhorst (2002) identified a one octave notch in the 4-8 kHz region combined with a prominent peak around 13 kHz as a strong frontal cue valid across subjects. This is compatible with our results (see Figure 3.24), although the main notch is shifted upwards in frequency somewhat (5-10 kHz). Hebrank and Wright (1974) and Langendijk and Bronkhorst (2002) found a similar spectral pattern cued sources coming from the back except for the notable absence of the 13 kHz peak. Again, this is compatible with our results. The 13 kHz peak is clearly visible for (and around) $\varphi = 0°$ but is essentially non-existent elsewhere, especially $\varphi = 180°$. This spatial distribution makes the 13 kHz peak a prime static localisation cue candidate. The similarity of spectral patterns for source locations at $\varphi = 0°$ and $\varphi = 180°$ explains frequent front-back confusions and supports suggestions that the key to resolving these confusions lies above 10 kHz (see Bronkhorst, 1995; Langendijk and Bronkhorst, 2002, amongst others). The increased spectral energy between 7 and 9 kHz found by Hebrank and Wright (1974) and Langendijk and Bronkhorst (2002) to be symptomatic of high elevations across subjects is also visible here between 8 and 10 kHz. As before, the upward shift in frequency can be explained by the relatively small KEMAR ears. This is a prime example of a covert peak

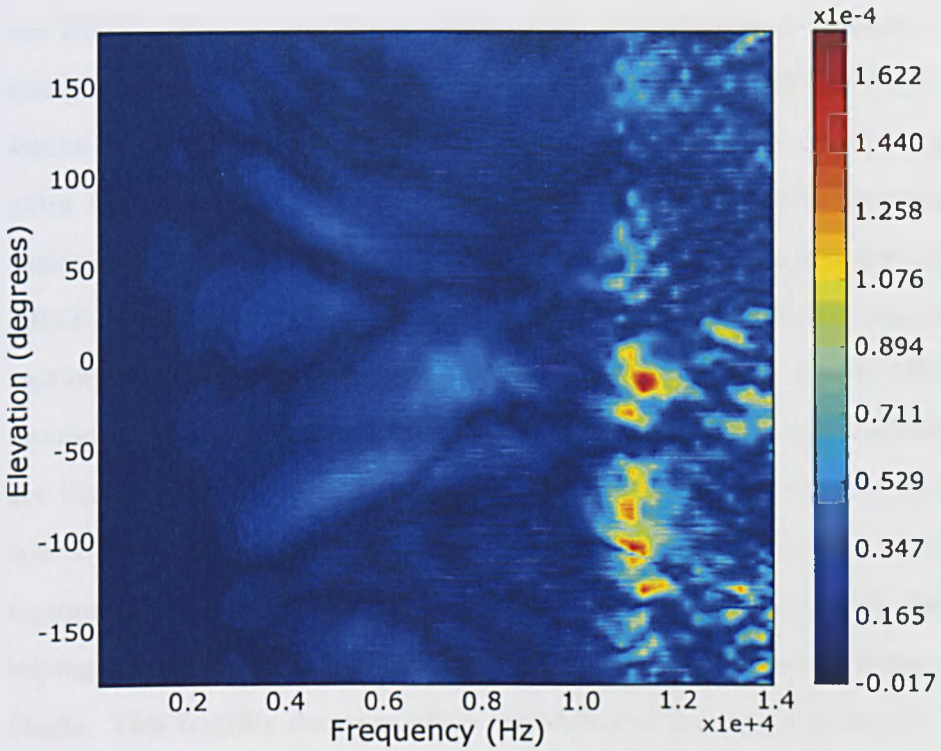


Figure 3.25: Absolute rate of spectral change with elevation (HRTF-ARCE) in the median plane (ITD = 0.0 ms), up to 14 kHz, in magnitude units (relative to a unit source placed close to the ear canal) per degree.

which is by no means a spectral peak. In fact, for sources around $\varphi = 90°$, the 8-10 kHz frequency range is the heart of a spectral notch. The high frequency roll-off above 10 kHz also identified by Hebrank and Wright (1974) and Langendijk and Bronkhorst (2002) at high elevation is also visible from 11 kHz. This roll-off is even more pronounced for lower elevations and at the back.

Rate of spectral variation with elevation

Figure 3.25 shows the absolute spectral rate of change with elevation (HRTF-ARCE) in the median plane. Note the important difference between

the HRTF-ARCE in the frontal plane, where the elevation is measured in altered vertical polar coordinates (see Figure 3.19) and the HRTF-ARCE in sagital planes, where the elevation is measured in interaural polar coordinates (see Section 2.1.2). The excitation of the 11.1 kHz pinna resonance varies very rapidly with source elevations around $\varphi = 0°$ generating high HRTF-ARCE values. This strong variation could contribute to the perception of frontal vertical source movements. However, the HRTF-ARCE around $\varphi = 0°$ for frequencies neighbouring the 11.1 kHz pinna resonance are highly volatile. This suggests that any localisation process relying on spectral variations within this spectral region for the discrimination of contiguous sound sources would be relatively fragile in comparison to those relying on the broader, smoother variations observed in lower frequency bands. This fragility does not affect the ability of the peaks generated by the 11.1 kHz pinna resonance (see Figure 3.24) to act as static localisation cues, however. As in Sections 3.4.2.2 and 3.4.2.4 the plot is also shown with its frequency range constrained to 0-10 kHz and its colour map re-adjusted in Figure 3.26, in order to better visualise the HRTF-ARCE for frequencies below 10 kHz.

The higher HRTF-ARCE values in the $-20° < \varphi < 20°$ area, for the 0-10 kHz frequency range (see Figure 3.26), occur between 6.5 kHz and 8 kHz, in the main pinna notch region. The important role of this notch in sound location perception finds support in a number of studies (Hebrank and Wright, 1974; Shaw, 1974; Butler and Belendiuk, 1977; Bloom, 1977; Langendijk and Bronkhorst, 2002). However, numerical variations with elevation are principally restricted to the falling (lower-frequency) edge of the notch. This suggests that the perceptual salience of the notch stems mostly from that region. The rising notch edge (9-10 kHz) seems to generate very

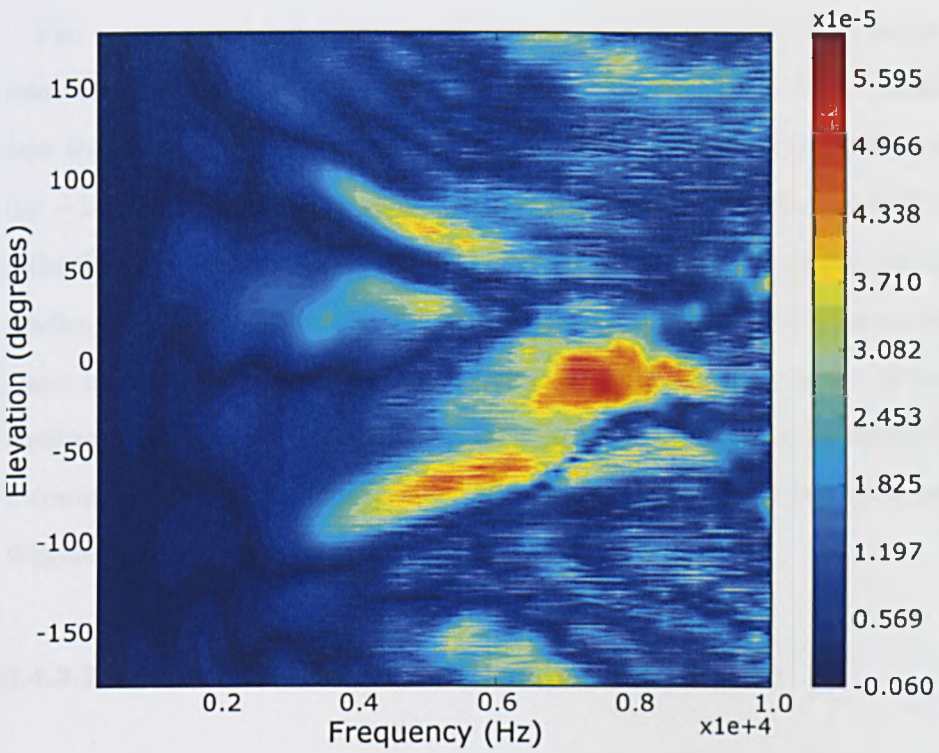


Figure 3.26: Absolute rate of spectral change with elevation in the median plane, up to 10 kHz, in magnitude units (relative to a unit source placed close to the ear canal) per degree.

little numerical variation over the entire ring of confusion.

As source location moves away from the horizontal plane, both upwards and downwards, the higher HRTF-ARCE values shift down in frequency in a surprisingly symmetrical fashion. Most of the variation for the $50° < |\varphi| < 100°$ directions occurs in the 4-6 kHz frequency range. The variation is, however, less pronounced than that observed closer to the horizontal plane, especially in the upper region ($50° < \varphi < 100°$). This slowing rate of variation may help to explain the decrease in vertical localisation performance for increasing source elevation observed by Damaske and Wagener (1969), Wettschurek (1973) and Leung and Carlile (2004) amongst others.

The symmetrical behaviour continues as sources become more remote from the median plane. A region of extremely low HRTF-ARCE values over the entire frequency range is visible for both the $110° < \varphi < 140°$ and the $-140° < \varphi < -110°$ regions. Again, this correlates with observed localisation performance, a minimum being noted in these directions by the studies mentioned previously. As source position approaches the horizontal plane in the back hemisphere ($\varphi = 180°$), spectral variation picks up momentum in the 5-8 kHz region. This increase in spectral variation probably accounts for the rise in localisation performance reported by Damaske and Wagener (1969) and Leung and Carlile (2004).

3.4.3.2 Ring of confusion variations for ITD = 0.2ms

Figure 3.27 shows spectral variations around the ITD = 0.2 ms ring of confusion. As for the previous case, the entire ring of confusion can be swept by varying φ between $-180°$ and $180°$, this time with a fairly stable azimuth, $\theta \approx -14.3°$. The visual resemblance with the ITD = 0.0 ms ring of confusion is quite striking. However, a closer inspection reveals a number of differences. The relative magnitudes of covert peaks is slightly altered. The 4.2 kHz covert peak ($-30 < \varphi < 30°$), both the main 11.1 kHz covert peaks ($80 < \varphi < 120°$ and $-70 < \varphi < -20°$) and the 13 kHz peak ($-10 < \varphi < 10°$) are now all on a par with each other in terms of magnitude. Also, the 4.4 kHz covert peak spreads further in elevation and the effects of the 7.1 kHz pinna resonance (visible across the neighbouring frequency band, see Section 3.4.1) are significantly more pronounced for $60° < \varphi < 110°$ area creating a plateau. The faint 11.1 kHz covert peak identified around $\varphi = -140°$ in the median (ITD = 0.0 ms) ring of confusion is still visible and its energy has markedly increased relative to the two other

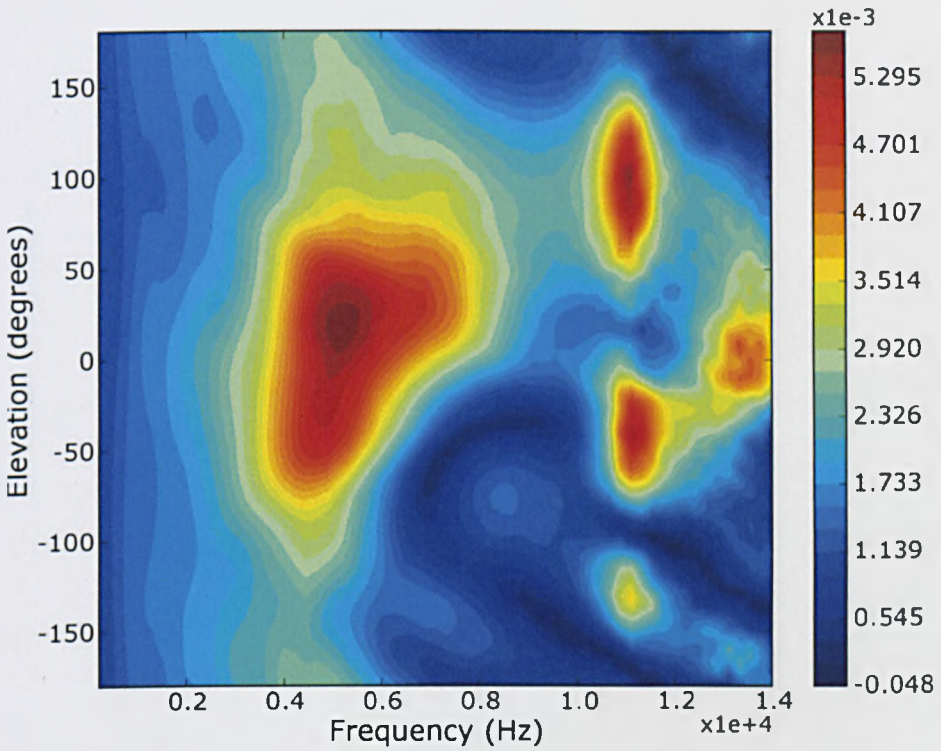


Figure 3.27: Spectral variations with elevation at the ipsilateral ear around the 0.2 ms ring of confusion, up to 14 kHz, in magnitude units (relative to a unit source placed close to the ear canal).

covert peaks at that frequency.

As a result of these differences the HRTF-ARCE (plotted in Figure 3.28 and again, with reduced frequency range and re-adjusted colour map, in Figure 3.29) is slightly altered. Overall, the HRTF-ARCE seems to be higher in the 4-10 kHz range as a result of the increased energy in the 4.4 kHz and 7.1 kHz covert peaks compared to the median plane ring of confusion. However, the greater spread of the 4.4 kHz covert peak across elevations results in a slight HRTF-ARCE decrease for $90 < \varphi < 110°$. The growth in the faint 11.1 kHz peak around $\varphi = -130°$ has also resulted in locally

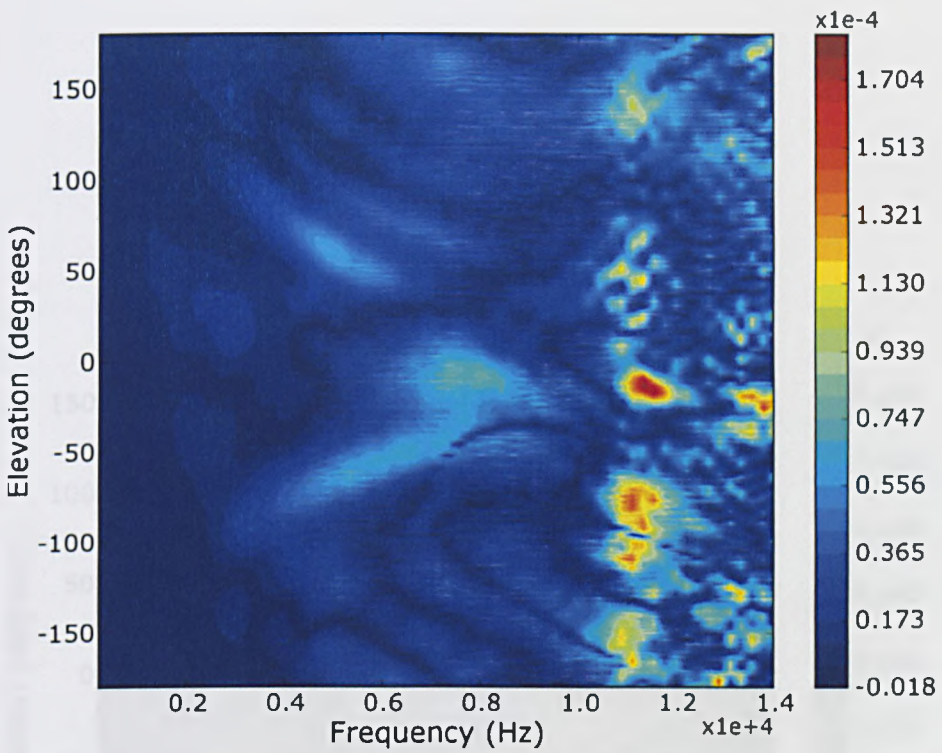


Figure 3.28: Absolute rate of spectral change with elevation (HRTF-ARCE) at the ipsilateral ear around the 0.2 ms confusion ring, up to 14 kHz, in magnitude units (relative to a unit source placed close to the ear canal) per degree.

increased HRTF-ARCE values. Spectral fluctuations in the $110° < \varphi < 140°$ and $-140° < \varphi < 110°$ have slightly increased from the extremely low levels observed in the median plane. This suggests that a marginal improvement in localisation performance is possible in these areas although to the author's knowledge, this has not been reported.

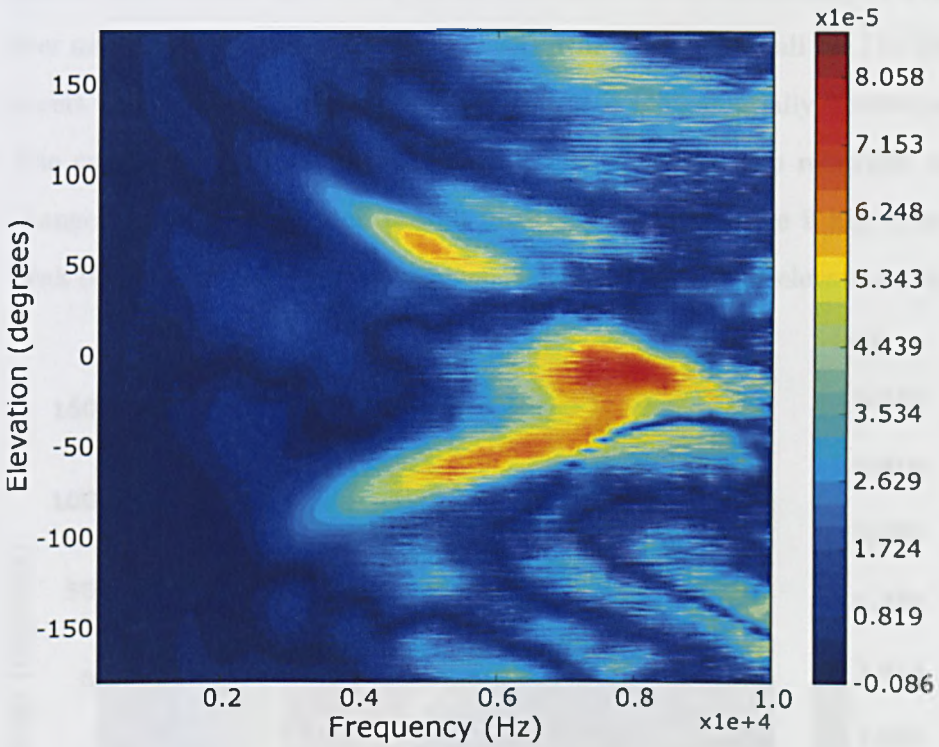


Figure 3.29: Rate of spectral change with elevation at the ipsilateral ear around the 0.2 ms confusion ring, up to 10 kHz, in magnitude units (relative to a unit source placed close to the ear canal) per degree.

3.4.3.3 Ring of confusion variations for ITD = 0.4ms

Figure 3.30 shows spectral variations around the ITD = 0.4 ms ring of confusion, which exhibits only small deviations from $\theta = -29.8°$. Its resemblance to the two previous rings of confusion stands out as a prominent characteristic, although further small differences can be identified. In a further magnitude growth, the 4.4 kHz covert peak now exceeds all the 11.1 kHz covert peaks. Save this magnitude change, it is left essentially unchanged. The connected 7.1 kHz covert peak ($0° < \varphi < 50°$) is also relatively unchanged apart from a small overall increase in energy. The 9 kHz covert peak ($50° < \varphi < 100°$) gains prominence over the rest of the elevation range

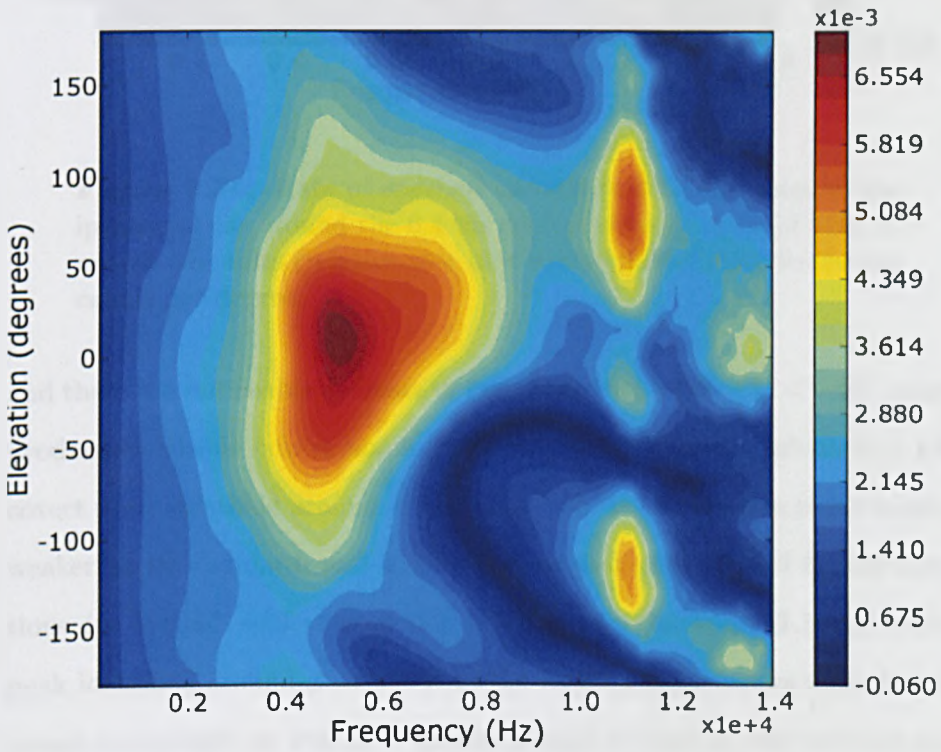


Figure 3.30: Spectral variations with elevation at the ipsilateral ear around the 0.4 ms confusion ring, up to 14 kHz, in magnitude units (relative to a unit source placed close to the ear canal).

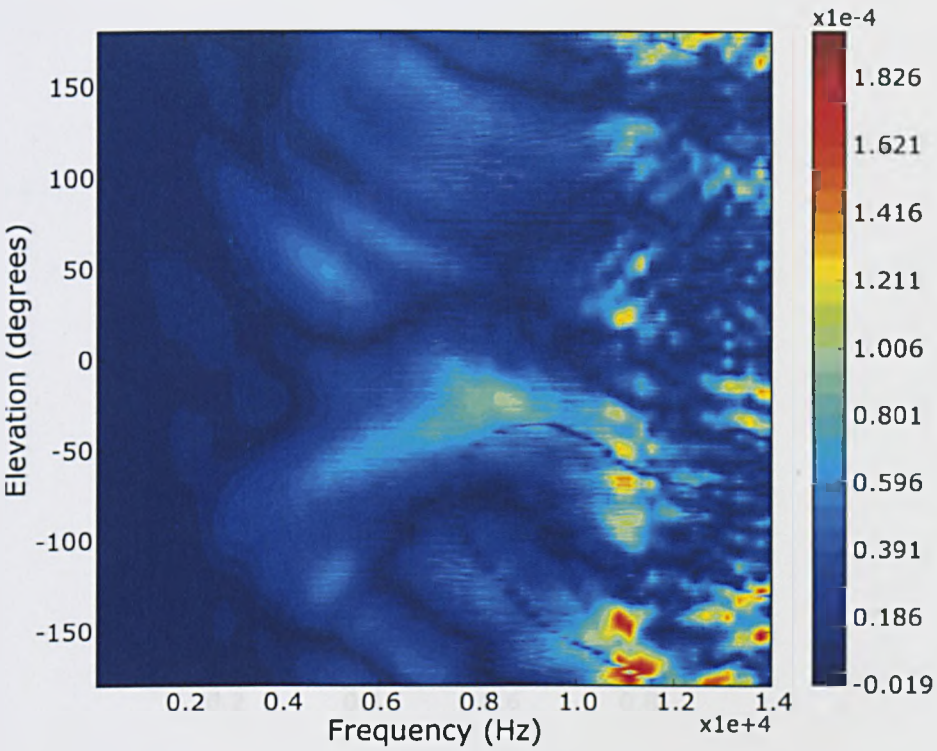


Figure 3.31: Rate of spectral variations with elevation at the ipsilateral ear around the 0.4 ms confusion ring, up to 14 kHz, in magnitude units (relative to a unit source placed close to the ear canal) per degree.

and there is a noticeable decrease in energy around $-60° < \varphi < -20°$, which produces a visible covert notch. The very strong median plane 11.1 kHz covert peak identified around $-70 < \varphi < -20°$ in the median plane is much weaker at this azimuth and it shows a marked shift toward higher elevations (to around $-30 < \varphi < 0°$). Conversely, the faint 11.1 kHz covert peak identified around $\varphi = -140°$ in the median plane shows a marked increase in strength as well as a spatial spread so that in now extends over $-140° < \varphi < -90°$. Deep covert notches are now discernible on both sides of this peak.

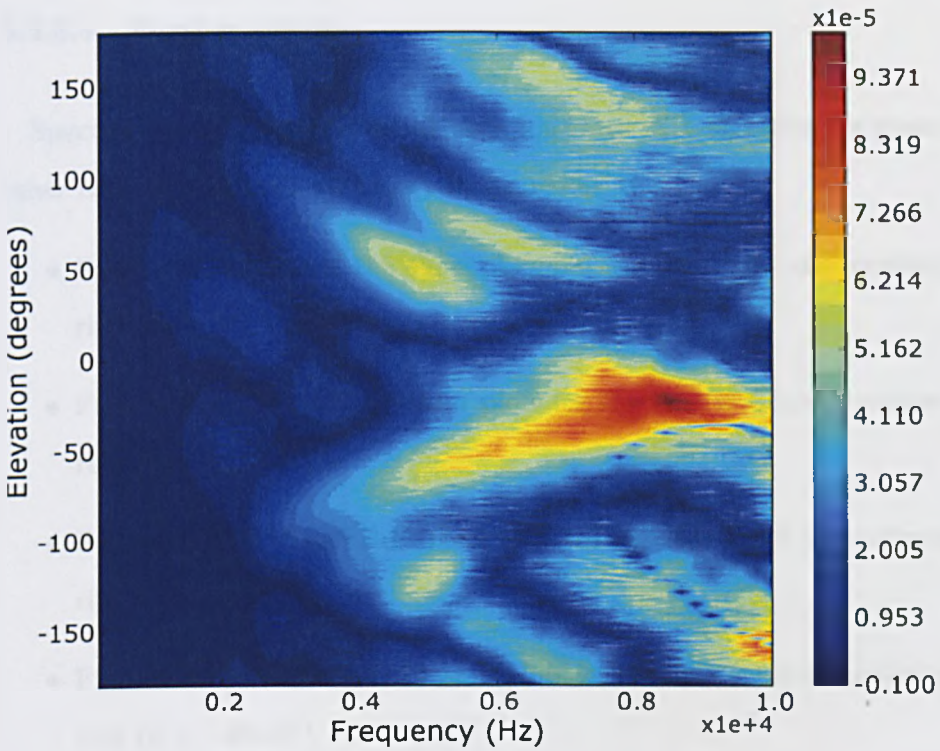


Figure 3.32: Rate of spectral variations with elevation at the ipsilateral ear around the 0.4 ms confusion ring, up to 10 kHz, in magnitude units (relative to a unit source placed close to the ear canal) per degree.

These differences, again, result in altered HRTF-ARCE values. These are plotted in Figure 3.31 and, with reduced frequency range and re-adjusted colour map, in Figure 3.32). The 9 kHz covert peak ($60° < \varphi < 110°$) is not as bright as those associated with the resonant pinna frequencies (4.4, 7.1 and 11.1 kHz, see Figure 3.30), however, its adjacency to a narrow, deep covert notch ($-50° < \varphi < 0°$) creates high HRTF-ARCE values, which form the prominent features in the plot shown in Figure 3.32. The HRTF-ARCE in the 11.1 kHz frequency band are much reduced around $\varphi = 0°$; the highest rates are now observed around $\varphi = 170°$. A marked difference in comparison to the median plane.

3.4.3.4 Further results

Spectral variations and spectral variation rates with elevation for several other rings of confusion are shown in Appendix B:

- Figures B.3, B.4 and B.5 show plots for the ITD = 0.1 ms confusion ring ($\theta \approx -7.27°$).
- Figures B.6, B.7 and B.8 show plots for the ITD = 0.3 ms confusion ring ($\theta \approx -21.68°$).
- Figures B.9, B.10 and B.11 show plots for the ITD = 0.5 ms confusion ring ($\theta \approx -38.37°$).
- Figures B.12, B.13 and B.14 show plots for the ITD = 0.6 ms confusion ring ($\theta \approx -48.43°$).

3.4.4 Discussion

3.4.4.1 Simulation results

Simulation results are in good agreement with previously published studies over the entire simulated frequency range, for the ipsilateral region. This gives further support to the multi-resolution meshing technique described in Section 3.3. The gain, in terms of required simulation time and memory is extremely significant and is vital to the work described in Chapter 5.

3.4.4.2 Similarity across rings of confusion

The similarities across cones of confusion are quite striking in the 0-10 kHz frequency band. This offers some support to suggestions that spectral cues

for elevation (as defined using interaural polar coordinates, see Section 2.1.2) are common for all azimuths (see Morimoto *et al.*, 2003a, amongst others). Further support comes from the fact that localisation accuracy around different rings of confusion displays a similar variation pattern (Leung and Carlile, 2004; Morimoto *et al.*, 2003a). Martin *et al.* (2004), by contrast, found that accuracy improved dramatically when a low frequency sound was played in the contralateral ear, giving ITD information. These results seem, at first, to be incompatible with those of Morimoto *et al.* (2003a). However, the changes in elevation for the vertical polar coordinates (see Figure 2.1, Section 2.1.2) used by Martin *et al.* inherently introduce ITD and ILD variations unlike changes in elevation for the interaural polar coordinates used by Morimoto *et al.* (2003a), Leung and Carlile (2004) and ourselves (see Figure 2.2 , Section 2.1.2). The additional cue variations with elevation help to explain the differing conclusions.

Above 10 kHz, however, our results indicate that significant changes occur across rings of confusion in the case of the KEMAR head and pinnae. The changes are particularly noticeable for the spatial arrangement of the local covert peaks created by the 11.1 kHz pinna resonance. These covert peaks shift in elevation and vary in amplitude across different rings of confusion. Although the results obtained in this study do not take into account torso and shoulder effects, these would be unlikely to counter the cross-azimuthal variations generated by the head and pinnae and, therefore, would not alter these conclusions. It has been reported by Morimoto *et al.* (2003b) and Bronkhorst (1995) that the bulk of localisation information is contained below 10 kHz, however, it would be dangerous to disregard higher frequency variations across rings of confusion, even though taking them into account would complicate HRTF estimation. Indeed, there is evidence that optimal

localisation performance is achieved only when the spectrum of broadband sound sources extends up to 15 kHz (Hebrank and Wright, 1974; Langendijk and Bronkhorst, 2002; Best *et al.*, 2005), which warns against simply extrapolating high-frequency median plane elevation cues across all azimuths.

3.4.4.3 Cue variation rates and localisation acuity

Repeated observations that the rate of cue change and the acuity of localisation seem to be correlated are perhaps the most important aspect of the data analysis. Again, the effects of the torso and shoulders are not taken into account, however, it seems reasonable to suggest that the localisation information they generate is mostly static, as opposed to dynamic. By this, it is meant that they are not responsible for large spatial cue variation rates like those generated by the pinnae, but rather, create localisation cues which change slowly and smoothly across the auditory space. As such, these effects are of little relevance to the following discussion.

ITD/ILD variations in the horizontal plane

The absolute rate of change in ITD/ILD with azimuth, or ITD-ARCA/ILD-ARCA (see Section 3.4.2.1), for example, could explain the decrease in horizontal localisation acuity as azimuth rises from $\theta = 0°$ to $\theta = 90°$. Although different studies agree over this decrease, numerical acuity measures differ significantly. This makes a numerical verification of the theory difficult. A comprehensive study of changes in azimuth acuity over the entire horizontal plane (see Section 2.3.1) for broadband and pure tone (both high and low frequency) is needed to rigorously test the theory. Such a study would be preferably conducted by detecting MAA/MAMA thresholds, as distortions due to motor control factors, near-inevitable when using techniques such as

source pointing, especially at the back, are completely eliminated.

Spectral variations in rings of confusion

The absolute rate of spectral change also seems to affect localisation acuity. The acuity of vertical localisation in the median plane, for example, seems to be highly correlated with the absolute rate of spectral change with elevation, or HRTF-ARCE (see Section 3.4.3.1). With general consensus, studies have reported maximum elevation acuity in front, progressively deteriorating with rising elevation, reaching a minimum around $\varphi = 120°$, and then improving as the source approaches the back of the horizontal plane ($\varphi = 180°$). Again, different studies agree on the general pattern of acuity variations but the numerical measures differ significantly. A significant cross-subject variation has also been reported. Leung and Carlile (2004), for example, reported that elevation acuity as source position moved from in front to above the listener, decreased by a factor of 2 to 5 depending on subjects (see Section 2.3.1.2).

In the case of the KEMAR head, the maximum HRTF-ARCE values (see Figures 3.25 and 3.26) are shown for a number of elevations over the 0-10 kHz frequency range, in Table 3.2. These values strongly support the suggestion that elevation acuity varies as a function of maximum HRTF-ARCE. The highest values appear in front, deteriorate with growing elevation to reach a minimum at $\varphi = 120°$, then rise as the source elevation nears 180°. Maximum values are spread between 4.9 kHz and 8.1 kHz, which suggests that dynamic localisation cues are particularly potent in this frequency range.

As described Section 3.4.3.1, HRTF-ARCE values are significantly higher above 10 kHz, than below. This results mainly from the 11.1 kHz pinna res-

Elevation (φ)	Maximum HRTF-ARCE (units/degree)	Frequency (kHz)
0°	5.6×10^{-5}	7.0
30°	3.7×10^{-5}	4.9
60°	3.7×10^{-5}	5.6
90°	3.0×10^{-5}	4.0
120°	2.1×10^{-5}	8.1
180°	3.8×10^{-5}	6.8

Table 3.2: The maximum absolute rate of spectral variation with elevation (HRTF-ARCE) in the median plane, below 10 kHz, as a function of elevation (see Figure 3.26). Maximum HRTF-ARCE values are in magnitude units (relative to a unit source placed close to the ear canal) per degree. The frequency at which the maximum occurs is shown in each case.

onance, whose excitation patterns are more complex than those observed for lower frequency resonances. HRTF-ARCE values vary erratically as a result, both within and across rings of confusion. This suggests that psychophysical processes exploiting them would not allow robust dynamic localisation. Such robustness would result, instead, from the smoother, broader regions of high HRTF-ARCE values observed at lower frequencies (4-9 kHz). This suggestion finds further support in the fact that similar patterns of localisation acuity, as well as similar spectral variations are observed across rings of confusion, below 10 kHz (see Section 3.4.4.2). Indeed, the spectral differences which appear across confusion rings, above 10 kHz, do not appear to alter localisation acuity patterns.

Maximum HRTF-ARCE values in the 10-14 kHz frequency range (shown for different elevations in Table 3.3) seem, however, to follow the same trend as that observed below 10 kHz. As in the previous case, highest HRTF-ARCE values are observed in front, deteriorate with growing elevation to reach a minimum around $\varphi = 120°$ then rise again as source elevation approaches $\varphi = 180°$. The maximum HRTF-ARCE values occur around

Elevation (φ)	Maximum HRTF-ARCE (units/degree)	Frequency (kHz)
0°	1.4×10^{-4}	11.1
30°	1.0×10^{-4}	11.1
60°	0.8×10^{-4}	11.1
90°	0.7×10^{-4}	11.1
120°	0.7×10^{-5}	13.8
180°	1.0×10^{-5}	13.8

Table 3.3: The maximum absolute rate of spectral variation with elevation (HRTF-ARCE) in the median plane, in the 10-14 kHz frequency range, as a function of elevation (see Figure 3.25). Maximum HRTF-ARCE values are in magnitude units (relative to a unit source placed close to the ear canal) per degree. The frequency at which the maximum occurs is shown in each case.

the 11.1 kHz pinna resonance, except for the back hemisphere directions ($\varphi = 120°$ and $\varphi = 180°$), where they occur around 13.8 kHz.

Spectral variations in lateral areas

A possible link between localisation acuity and absolute rate of spectral change is also visible in the far ipsilateral area ($\theta = -90°$, $\varphi = 0°$). Saberi *et al.* (1991) showed that the vertical (φ variations) MAAs and horizontal (θ variations) MAAs in that area ($\theta - 90°, \varphi = 0°$) are essentially equivalent. Accordingly, maximum absolute rates of spectral variation (in the 0-10 kHz range) are numerically very similar for θ and φ variations. ITD and ILD cues in either case are unlikely to lie behind observed acuity (see Sections 3.4.2.1 and 3.4.2.3).

In the case of θ variations, maximum HRTF-ARCA are around 1.3×10^{-4} magnitude units per degree and occur in the 4-6 kHz frequency range. In the case of φ variations, maximum HRTF-ARCE are around 1.2×10^{-4} magnitude units per degree and occur in the 8-10 kHz range. The close

numerical similarity of absolute spectral variation rates with direction in the horizontal and vertical cases, although they do not occur in the same frequency bands, could lie behind the similar localisation acuities observed.

It is important to note that, while maximum absolute spectral variation rates in lateral areas are similar under 10 kHz for θ and φ variations, a large discrepancy appears at 11.1 kHz, for the same direction. A strong local covert peak is present at $(\theta, \varphi) = (-90°, 20°)$. It rapidly falls as φ descends from 20° to 0°, which produces high spectral variation rates. These high variation rates are not observed as sources approach $(\theta, \varphi) = (-90°, 0°)$ along the horizontal plane. MAAs, however, are similar in both cases. This suggests that in lateral areas, as in the case of the median plane described earlier, spectral variation around 11.1 kHz do not act as a dynamic localisation cues.

When related to reported spatial variations in localisation acuity, our results seem to indicate that smooth, broad regions of spectral variations observed below 10 kHz are systematically preferred to stronger and more unstable higher frequency spectral variation as robust and accurate dynamic localisation cues.

3.4.4.4 The role of pinna resonances

It is clear that pinna resonances play a very significant role in generating high HRTF-ARCE values around any confusion ring. Depending on their direction of incidence, sound waves stimulate varying levels of resonance, which leads to spatially fluctuating spectral excitation patterns. The frequency of spectral peaks observed in HRTFs seems to be directly inferable

from pinna resonances. Estimating the frequency of the pinna resonances for a given individual seems, in any case, to be an important aspect of HRTF estimation. The extent to which the far-field excitation patterns of these resonances can be extrapolated across subjects is a matter for investigation. Such extrapolations, if they are justified, would constitute an important tool in facilitating HRTF estimation.

Although spectral peaks play an important role in generating spatial HRTF fluctuations, the role of spectral minima in enhancing these fluctuations should not be overlooked. A role for spectral notches in localisation has been widely proposed (Hebrank and Wright, 1974; Butler and Belendiuk, 1977; Bloom, 1977), and in a number of cases, the highest HRTF-ARCE values were observed not in the vicinity of strong covert peaks but where spatially adjacent covert peaks and covert notches in a given frequency band combine to enhance these values. It seems reasonable to suggest that both the notches produced by pinna reflections and the peaks produced by pinna resonances (see Batteau, 1967; Shaw and Teranishi, 1968; Rodgers, 1981; Kahana and Nelson, 2005, and Section 3.4.1) co-operate to generate the information required for source location discrimination within cones of confusion.

Chapter 4

A Morphoacoustic Database

"Facts are the air of scientists.
Without them you can never fly."
Linus Pauling

"Where is the knowledge that is lost in information?
Where is the wisdom that is lost in knowledge?"
T.S.Eliot

This chapter describes the acquisition of shape and acoustic data for a large number of subjects, compiled into what is referred to as a "morphoacoustic database". HRTF data was data obtained through anechoic acoustic measurements using Golay codes for 393 positions around each of 49 subjects. For these same subjects, shape capture was performed using MRI scanning, which resulted in comprehensive descriptions. A novel shape parameterisation technique, which aims to compress the large amounts of gathered shape data to a more manageable size, is presented. This parameterisation technique is expected to facilitate and accelerate later analyses

and its efficiency is tested on the acquired dataset.

4.1 Morphology capture and data pre-processing

Section 2.4.1 gave an overview of the general approach to shape description used for the study of relationships between morphology (particularly of the head, pinna, shoulder and torso) and auditory localisation cues. The weak statistical links reported in these studies suggest an alternative approach involving more complete shape description is required. The landmark measurement approach, commonly adopted in this type of study (Jin *et al.*, 2000; Algazi *et al.*, 2001c), lacks a reliable basis on which to justify a selection of features to be measured. Because of this they tend to aim for a well distributed set of measures over the entire pinna structure. The measurements used for the CIPIC database (see Algazi *et al.*, 2001c), for example, give little attention to areas identified as having primary cue production roles such as the cavum concha, cymba concha, fossa of helix and antihelix (see Sections 2.2.2.1, 2.4.6.5 and 3.4.1). Each of those complex three dimensional structures is generally described using one or two linear measurements, which leaves ample scope for relevant shape variations across individuals to be overlooked.

The objective of the shape-capture procedure in the context of this study was to acquire a *complete* shape description for the surface of the head, ears and upper torso. With this in mind, the choice of MRI scanning was made, as it allows not only a complete external ear description to be obtained but also captures internal morphology such as the deep ear canal, nasal passages and other features which are of potential interest for binaural sound

research. Original MRI trials pioneered at the York Neuroimaging Centre[1] demonstrated the technical challenges faced when attempting to reveal the fine detail of the pinnae, an unusual application for this kind of technology. Difficulties in scanning the pinna lead to the consideration alternate capturing methods.

The constituent material of the human pinna is chiefly responsible for poor imaging results. A possible solution to this problem is the creation of ear replicas, which can be scanned more effectively. Test ear molds were produced using alginate, a recognised industry standard for safely capturing and reproducing detailed human body parts[2]. A plaster of Paris cast was then produced from the alginate molds. Good results were obtained, however, the procedure is delicate and requires extensive training. Also, minor shape distortions are essentially inevitable and the casts are fragile particularly in the regions of thin, delicate pinna cartilage.

Another possible technique for shape-capture is stereo-photography. This technique simulates human binocular vision to capture three-dimensional images. The hardware needed is expensive, however, and shape-capture suffers from occlusion problems similar to those encountered through laser scanning. During the course of the project, it became apparent that recent developments in technology and careful adjustments to the scanning parameters did, in fact, allow MRI-based shape capture of the head and pinnae. This option was ultimately considered favourable.

[1] `https://www.ynic.york.ac.uk/`
[2] `http://www.smooth-on.com/index.php?cPath=1234_1240`

4.1.1 MRI Scanning

Two Philips 3T Achieva MRI scanners located in Sydney, Australia[3] were employed. Each was equipped with a Quasar dual gradient system set at 40 mT/m with a slew rate of 200 mT/m/ms. Two MRI scans were performed on each of the 49 subjects in the present study.

The first of these was a high definition scan optimised for obtaining a detailed pinna description with sub-millimetre resolution. This scan delivered 280 axial 2D cross-sectional images for the whole head[4]. The image slices had a spatial separation of 870 μm and the scan lasted for 10 minutes. The second scan was performed at a lower resolution, but over a larger area. It was aimed at providing shape information for the shoulders and torso. This scan delivered 100 coronal 2D cross-sectional images[5] at 3 mm intervals and lasted 3 minutes. The extent of the shoulder/torso morphology which could be captured depended on subject's size and ranged from just below the shoulder-line down to about 15 cm below. Originally, the subjects were simply asked to minimise involuntary movements as much as possible. During later scans, however, the head was immobilised to reduce gross movement. However, changes in facial expression, blinking, coughing and other involuntary movements remained potential sources of image artefacts.

[3]The scanners were located in the Symbion imaging center (Prince of Wales Medical Research Institute, Randwick) and St Vincents' Hospital (Darlinghurst)

[4]Axial images are parallel to the horizontal planes defined in Section 2.1.1

[5]Coronal images are parallel to the frontal plane defined in Section 2.1.1

Figure 4.1: MRI DICOM image example

4.1.2 Subject mesh model extraction

The raw data resulting from each scan was stored as a DICOM[6] description. It consisted of a set of two-dimensional gray-scale image files. This data was converted into three-dimensional mesh through a process referred to as surface rendering. The rendering was performed using the OsiriX[7] freeware package set to the highest possible resolution, which resulted in a very dense mesh. OsiriX allows regions of specific pixel brightness ranges to be extracted and rendered. Although this is a useful feature, the brightness range observed on the surface of the head and pinnae overlaps greatly with that of internal detail (see Figure 4.1). The rendering therefore creates a large quantity of unwanted polygons inside the head. Several further processing steps are necessary to obtain a clean and complete mesh of the outer surface only:

[6]DICOM stands for digital imaging and communications in medicine. It is a standard for handling, storing and transmitting medical imaging information.

[7]Url: `http://www.osirix-viewer.com/`

- Using the Autodesk 3ds Max software package[8], holes in the mesh, mostly present in the neck, nostrils and ear canals, were filled and scanning imperfections were eliminated as far as possible.

- A set of 256 two-dimensional images were obtained from the cross-section of subject meshes and a plane, which was rotated around the interaural axis. This slicing is similar to that applied prior to EFT parameterisation (see Section 2.4.1.2 and Figure 2.11). This was done using the VTK C++ library[9]. Each image contained a closed contour delineating the outer surface, but also contained unwanted mesh elements (Figure 4.2, step 1).

- The inner and outer parts of the main closed contour were filled with different colours using the Cairo C++ 2D graphics library[10] (Figure 4.2, step 2).

- The closed contour was traced using edge extraction (Figure 4.2, step 3).

- Each contour was converted to a set of 2048 points and the plane in which these lay was rotated about the slicing axis to restore it to its original orientation in three-dimensional space.

- The point sets for each slice were fed through a multi-resolution point cloud reconstructor algorithm (see Section 3.3.2) to produce a clean surface mesh. The mesh resolution was set to approximately 0.8 mm in the pinna regions and 8 mm elsewhere. The high resolution was necessary for the pinna to avoid unwanted polygons bridging surfaces in regions of high surface curvature.

[8]Url: `http://usa.autodesk.com/adsk/servlet/index?siteID=123112\&id=5659302`
[9]Url: `http://www.vtk.org/`
[10]Url: `http://www.cairographics.org/`

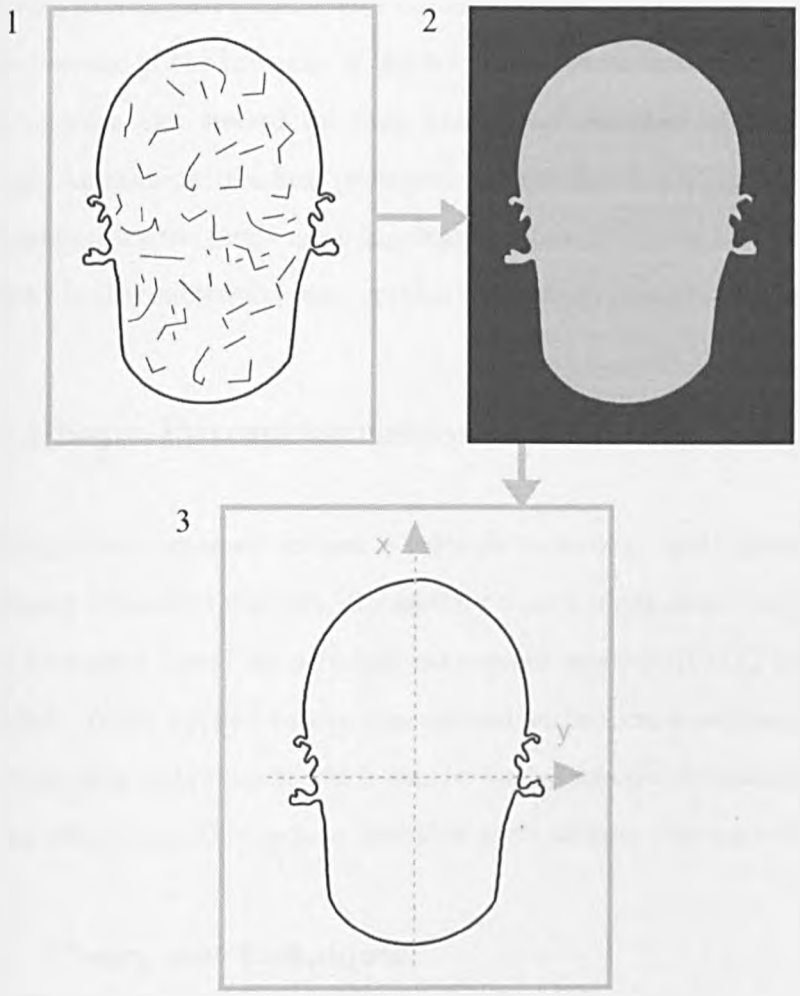


Figure 4.2: Slice "cleaning" process. Step 1 shows a typical cross-section of a sealed subject mesh, step 2 shows the coloured inner and outer contour regions and step 3 shows the surface contour obtained through edge extraction.

The slicing process (Figure 2.11) can lead to the formation of inner "islands" (separate closed contours). Although these islands can describe valid surface contours, they are discarded by subsequent processing steps (colouring and edge extraction), leaving a gap in the description. They can, however, be avoided by placing the rotation axis of the slicing plane with care.

A rotation axis aligned exactly with the inter-aural axis produced unwanted islands because of the presence of the tragus. However, a slight translation of the rotation axis toward the back and top of the head eliminates this problem. An example of a final processed mesh is shown in Figure 4.3 and a close-up exposing the pinna mesh topology is shown in Figure 4.4. The steps described in this subsection were applied to all 49 subjects in the database.

4.2 Shape Parameterisation

Although this processed dataset is valuable in its own right, its sheer size complicates statistical analysis. To address this, a novel shape parameterisation technique based on principal component analysis (PCA) has been developed. When applied to two-dimensional surface cross sections it provides high data compression which can be traded against acceptable levels of shape distortion. This section describes and validates the method.

4.2.1 Theory and Techniques

4.2.1.1 The EKLT and its relation to the EFT

The proposed parameterisation procedure is inspired by the works of Hetherington *et al.* who investigated the potential of the elliptic Fourier transform (EFT) as a parameterisation technique for the human head, including the pinnae (Hetherington and Tew, 2003; Hetherington *et al.*, 2003). The EFT expresses a three-dimensional surface as a set of two-dimensional slices obtained from the intersection of the surface with a plane which is rotated at regular angular intervals around a slicing axis. A two-dimensional Fourier transform is then applied to the parametric slice components (see Section

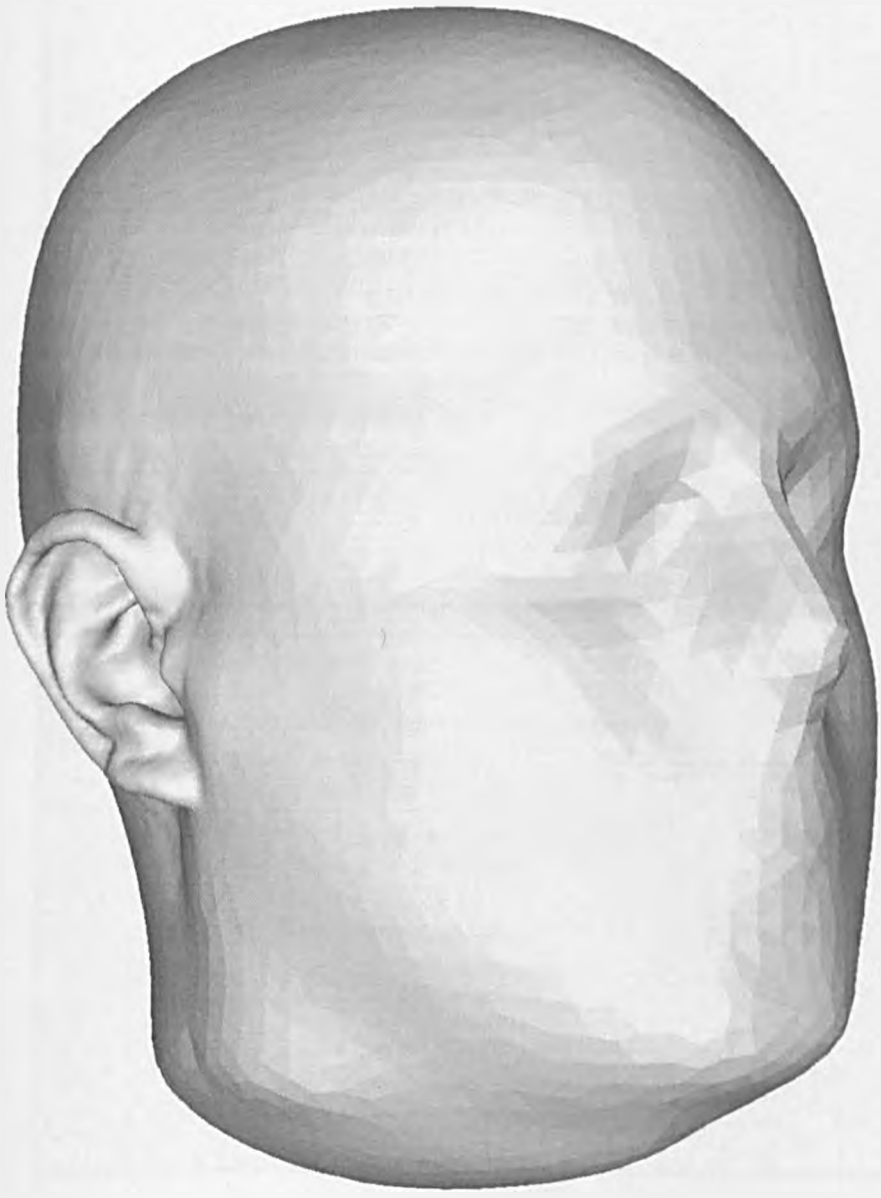


Figure 4.3: Example head reconstruction

2.4.1.2).

The large size of the morphological dataset acquired during this study and the redundant nature of the data allows the efficiency of the EFT parameter-

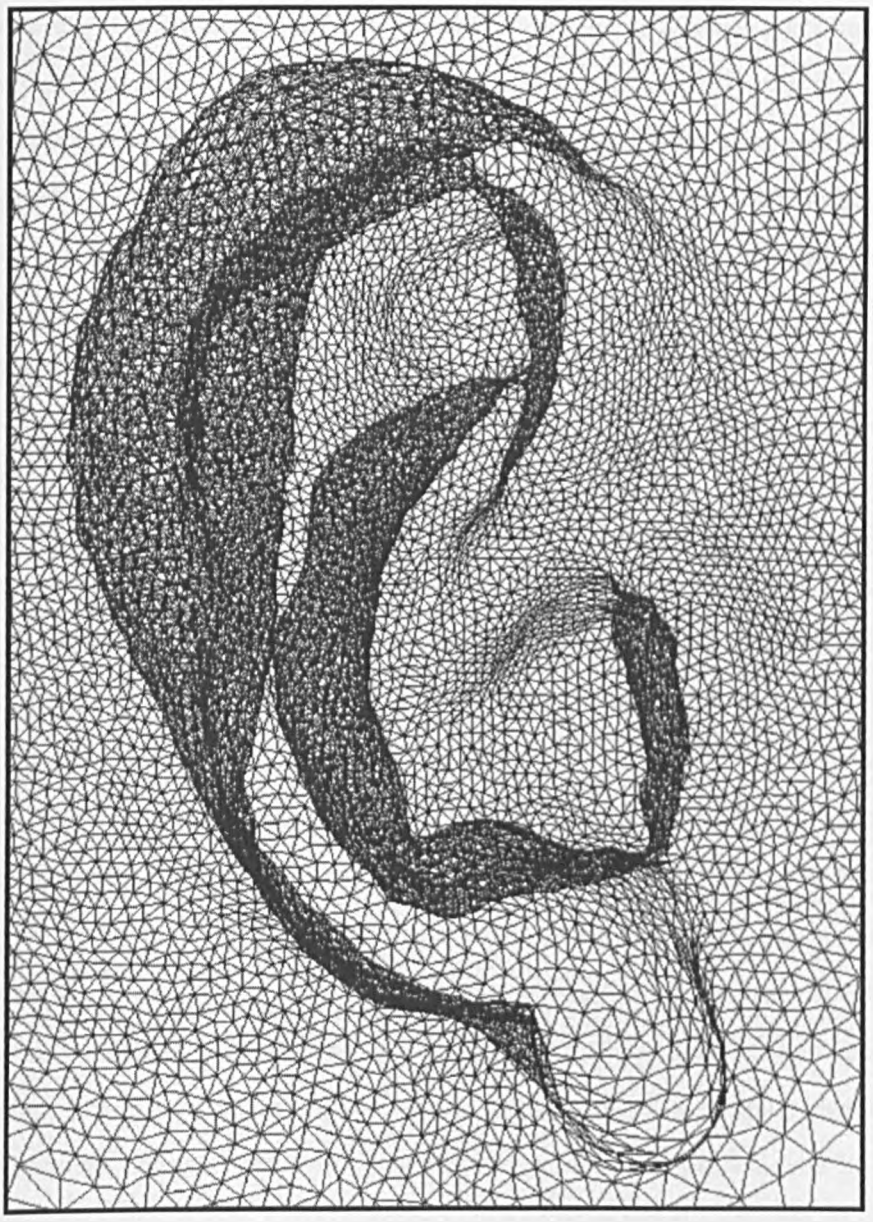


Figure 4.4: Example ear mesh reconstruction

isation to be significantly bettered by employing a discrete Karhunen-Loève transform (or KLT, see Loève, 1978) in both EFT stages instead of the DFT. In the context of the EFT, the purpose of the DFT is to concentrate energy into the lower order components so that truncation can be effected with

minimal loss of data. The KLT represents a dataset as a linear combination of orthogonal functions in a manner very much comparable to the DFT. In contrast to the DFT, however, the expansion basis functions are not sinusoidal, but are optimised to suit the original data. This allows the dataset to be synthesised with some specified accuracy using the fewest possible basis functions. Hence, we adopt the name elliptic Karhunen-Loève transform (EKLT) for the proposed method.

The pre-conditioning of the subjects' morphological data through the use of radial slicing posseses two major advantages. The first is that it inherently concentrates shape definition on the region close to the slicing axis, namely the external ear. This property is highly beneficial due to the key role of this region in the production of directional and cross-subject HRTF variations. The other advantage is that the data can be easily geometrically registered across subjects before the EKLT is applied. This creates a correspondence between the morphological parameters across subjects. Although achieving this correspondence when using landmark based shape description (used by Jin *et al.*, 2000; Algazi *et al.*, 2001c, amongst others) is relatively straightforward, complete shape description (using points spread on the surface contour) complicates matters a great deal. Slicing provides a simple yet powerful solution, since data alignment can be achieved by using the inter-aural axis as the slicing axis and aligning slice 0 with the tip of the nose.

4.2.1.2 Mathematical description of the EKLT

The application of the KLT to exploratory data analysis is commonly referred to as principal component analysis (PCA). PCA is an orthogonal

linear transformation which describes a dataset with a new set of optimised bases. Each basis for the dataset is obtained in turn to account for the maximum possible residual variance present after having computed and subtracted the previous bases, with the constraint that it is orthogonal to each of them. The theoretical and practical aspects of PCA are discussed at length in the multivariate statistics literature (Jolliffe, 2002; Johnson and Wichern, 2002; Everitt and Dunn, 2001, for example). The principal components of a set of n observations of m variables, all within $\mathbb{R}$, described by a matrix $\mathbf{X}$ (of dimensions $m \times n$) can be obtained by a singular value decomposition (SVD), expressing $\mathbf{X}$ as

$$\mathbf{X} = \mathbf{W}\mathbf{\Sigma}\mathbf{V}^T \tag{4.1}$$

where $\mathbf{V}^T$ the transpose of $\mathbf{V}$, a unitary matrix over $\mathbb{R}$ of dimensions $n \times n$. $\mathbf{V}$ contains a set of orthonormal "input" basis vectors for $\mathbf{X}$. $\mathbf{\Sigma}$ contains the singular values which scale the normalised input vectors contained in $\mathbf{V}$. $\mathbf{W}$ is a unitary matrix over $\mathbb{R}$ of dimensions $m \times m$ and contains a set of "output" basis vectors for $\mathbf{X}$, the principal components. The original data matrix $\mathbf{X}$ is then rotated (with the matrix operator $\mathbf{W}^T$), giving the matrix $\mathbf{Y}$, as follows:

$$\mathbf{Y} = \mathbf{W}^T\mathbf{X} \tag{4.2}$$

$\mathbf{Y}$ is a matrix of column vectors, where each vector is the projection of the corresponding original data vector from $\mathbf{X}$ onto the basis vectors (principal components) contained in the columns of $\mathbf{W}$. The elements of the

column vectors contained in $\mathbf{Y}$ are referred to as weights for the principal components contained in $\mathbf{W}$. The number of principal components required to reconstruct the original observations depends on the redundancy in the original data and the target accuracy of the reconstruction.

Prior to applying the EKLT, the original data consisting of S slices each described by a sequence of T xy-coordinates is conditioned to improve PCA performance. The highly (though clearly not completely) symmetrical nature of the human head results in a high degree of correlation between the left and right halves of each slice. The original slices are therefore split into left and right half-slice observations and right half-slices are flipped to be aligned with left half-slice observations (see step 2 in Figure 4.5). The point order of the right half-slices has to be reversed for the similarity between the right and left half-slices to be exploited (the start points are indicated by black squares on this and subsequent diagrams). This process halves the number of variables for each observation and doubles the number of observations. Another performance improvement can be achieved by exploiting the similarity between the top and bottom quarter slices, treating each as a separate observation (step 3 in Figure 4.5). The bottom quarter slices are flipped to be aligned with the top ones and, again, the point order is reversed so that the starting point for all slices is the interaural axis (see Figure 4.7).

As all the observations must be described by the same number of variables for a PCA analysis to be applied and noting that the length of contour portions OA and OB (see Figure 4.7) will generally be slightly different, the sampling interval between O and A is adjusted so that $T/4$ samples lie between them (as a full slice contains T samples). The same process is

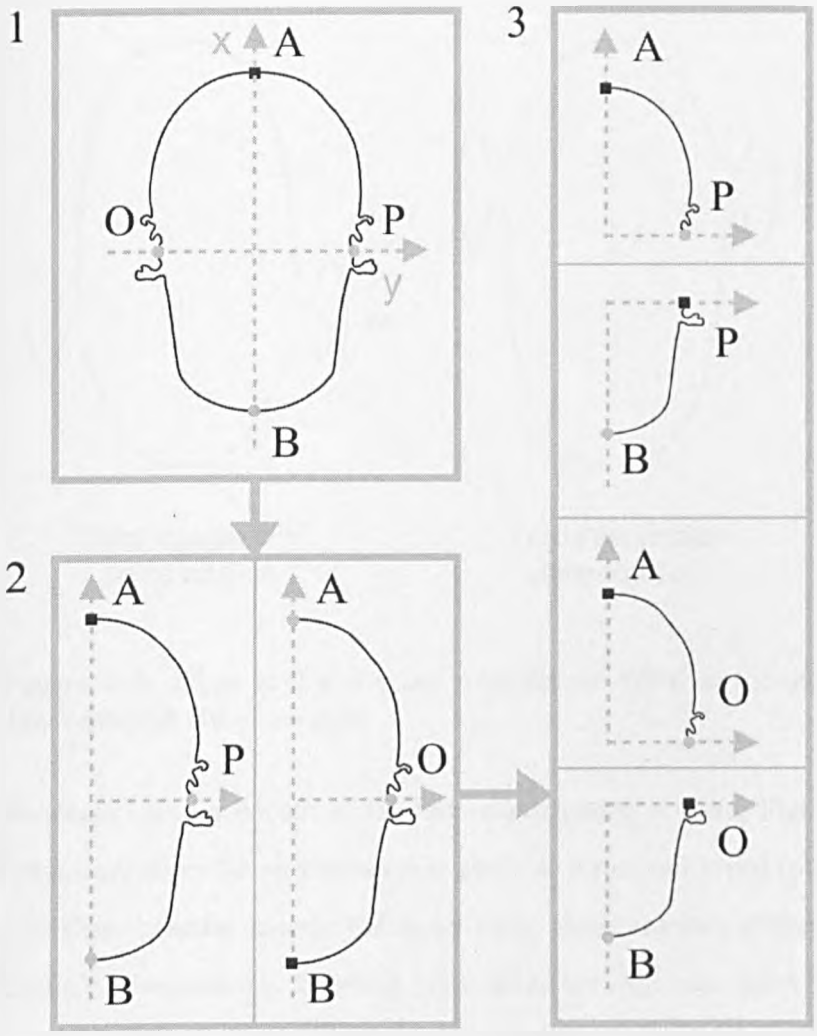


Figure 4.5: Slice conditioning for first EKLT stage

applied to obtain $T/4$ equally spaced samples along the contour portions OB, PA and PB. This step presents the significant advantage of improving the alignment of pinna features from one slice to the next and across subjects. The re-sampling ensures that the interaural axis occurs at the same sample number on each slice.

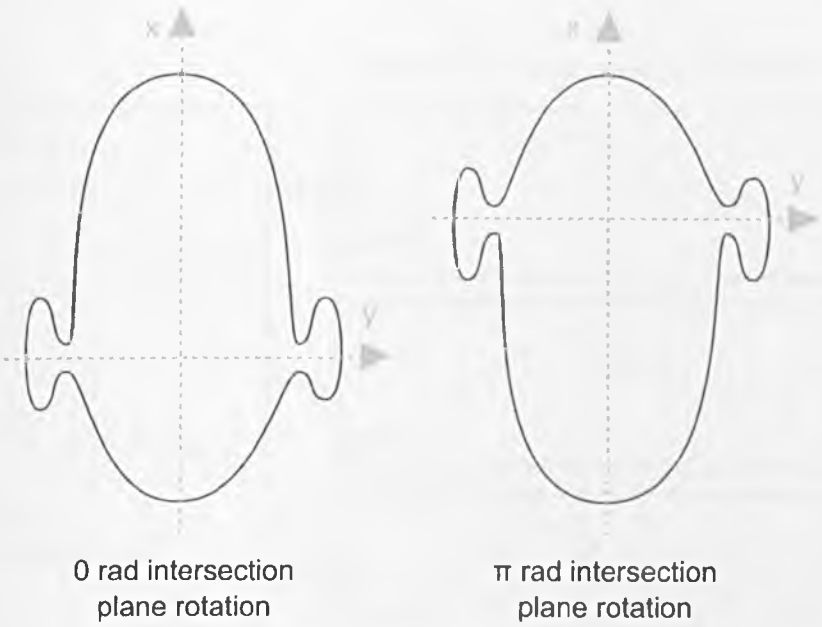


Figure 4.6: Slices at 0 and π rad rotation are different though they represent the same data

To illustrate another benefit of the data conditioning process, Figure 4.6 shows (dummy) slices for an intersection plane at 0 rad and π rad rotation. Although they describe exactly the same data, these are very different as full slice PCA observations. However, once slices are split into quarter slice observations, which are then aligned, the similarity between slices around 0 rad and π rad is exploited and further energy is concentrated into lower order PCA components.

For the first stage of the EKLT, each observation in a data matrix $\mathbf{X}_1$ is assembled by concatenating the x and y co-ordinates for a given quarter slice (see Figure 4.7). Each quarter slice contains $T/4$ points, each with x and y coordinates giving a total of $2(T/4) = T/2$ variables per quarter slice observation. If there are N subjects and S full slices per subject, there are

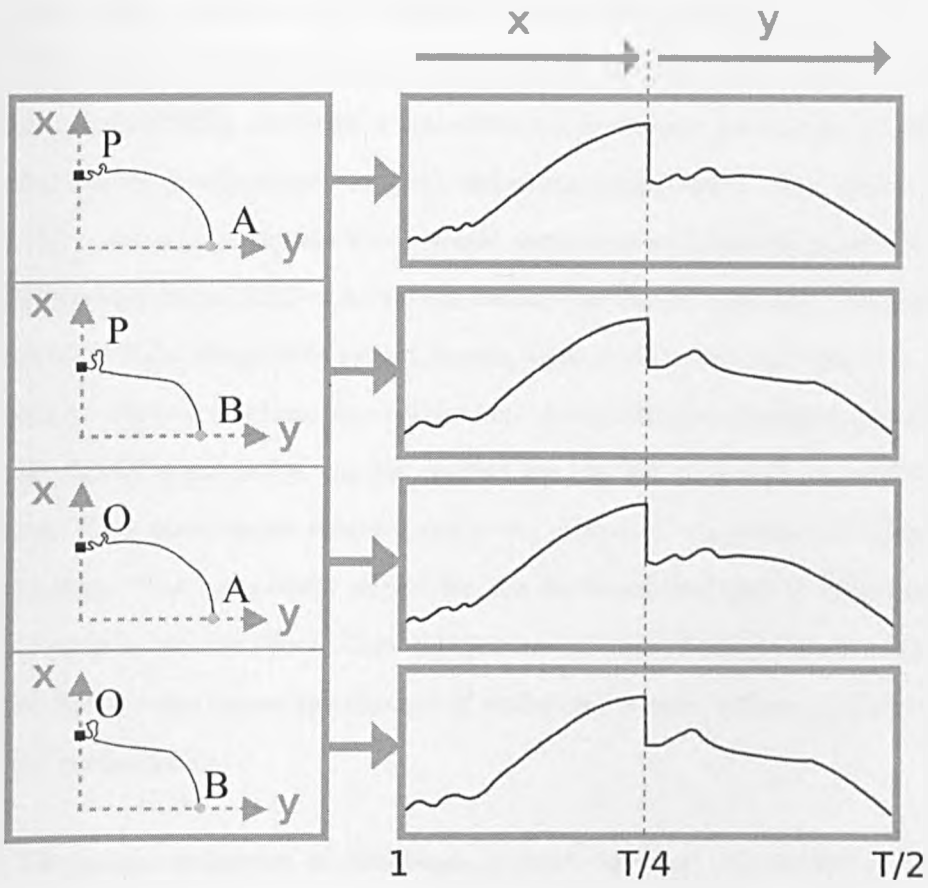


Figure 4.7: The four quarter slices resulting from each original slice are aligned so as to exploit their similarity during PCA (the black square shows the starting point in each case). For each quarter slice comprised of $T/4$ points, the x and y coordinates are concatenated to form a first stage EKLT observation containing $T/2$ variables.

$4SN$ quarter slice observations. $\mathbf{X}_1$ therefore has the dimensions $T/2 \times 4SN$.

The PCA of $\mathbf{X}_1$ gives the weights matrix $\mathbf{Y}_1$ according to

$$\mathbf{Y}_1 = \mathbf{W}_1^T \mathbf{X}_1 \tag{4.3}$$

where the columns of $\mathbf{W}_1$ contain the principal components of $\mathbf{X}_1$. $\mathbf{Y}_1$ contains the results from the first EKLT stage. As for the EFT, the second stage of the EKLT performs a transform on first stage parameter (in this case, the PCA component weight) variations across slices. The objective of the process is to express these weight variations as concisely as possible. Each second stage EKLT observation should describe the variation of a given first stage PCA component weight, across slices, for the four quarter slices, in order to allow a full-head reconstruction. Alternatively, a separate second stage EKLT observation can be created for the left and right side of the head. Each observation would contain the cross-slice variations of a given first stage PCA component weight for the corresponding (left or right) top and bottom quarter slices. This doubles the number of second stage EKLT observations and halves the number of variables for each of them, improving PCA performance.

The parameterisation of half-heads presents another, significant advantage looking ahead to the extraction of morphoacoustic mappings. It seems reasonable to suggest that the HRTF for a given ear can be calculated accurately by approximating the opposite side of the head to a perfect symmetrical reflection of the side on which the ear lies about the median plane. If this is the case, information describing the shape of the opposite side of the head is irrelevant to the estimation of a monaural HRTF for the ear in question. Discarding this information would not, however, be an option should whole heads be parameterised using the EKLT. This could hinder the extraction of mappings between morphology and monaural HRTFs and, as a consequence, half-head parameterisation is preferred.

In order to achieve half-head parameterisation, $\mathbf{Y}_1$ is re-arranged to form the second stage observations matrix $\mathbf{X}_2$. Since there are $2N$ half-head observations per component, if the first C_1 most significant first stage EKLT principal components are sufficient to reconstruct $\mathbf{X}_1$ with satisfactory accuracy, there will be $2NC_1$ second stage EKLT observations in all. Each observation describes the variation, across slices, of the weight for a given first stage EKLT principal component for the top and bottom quarter slices which comprise a given half-head. This equates to a total of $2S$ variables per observation. $\mathbf{X}_2$ will therefore have dimensions $(2S) \times (2NC_1)$. Fig 4.8 gives a visual representation of the rearrangement. Performing a PCA on $\mathbf{X}_2$ gives the weights matrix $\mathbf{Y}_2$ according to

$$\mathbf{Y}_2 = \mathbf{W}_2^T \mathbf{X}_2 \tag{4.4}$$

where the columns of $\mathbf{W}_2$ contain the principal components of $\mathbf{X}_2$. Assuming the first C_2 components contained in $\mathbf{W}_2$ are sufficient to reconstruct $\mathbf{X}_2$ with satisfactory accuracy, then the entire shape data for a given half-head can be reconstructed with C_1C_2 parameters.

4.2.2 Parameterisation Performance

The EKLT parameterisation method was tested for a morphological database of 49 subjects. The head and pinnae of each subject were described by a set of 128 slices each containing 512 points giving a total of 65,536 points per subject. Each point is described by x and y coordinates, therefore the morphology of the head and pinnae for a single subject is described by 131,072 real values. As mentioned previously, shape data for the right and left sides of the head were treated as separate observations, the right side being flipped so as to be aligned with the left side. This resulted in 98 half

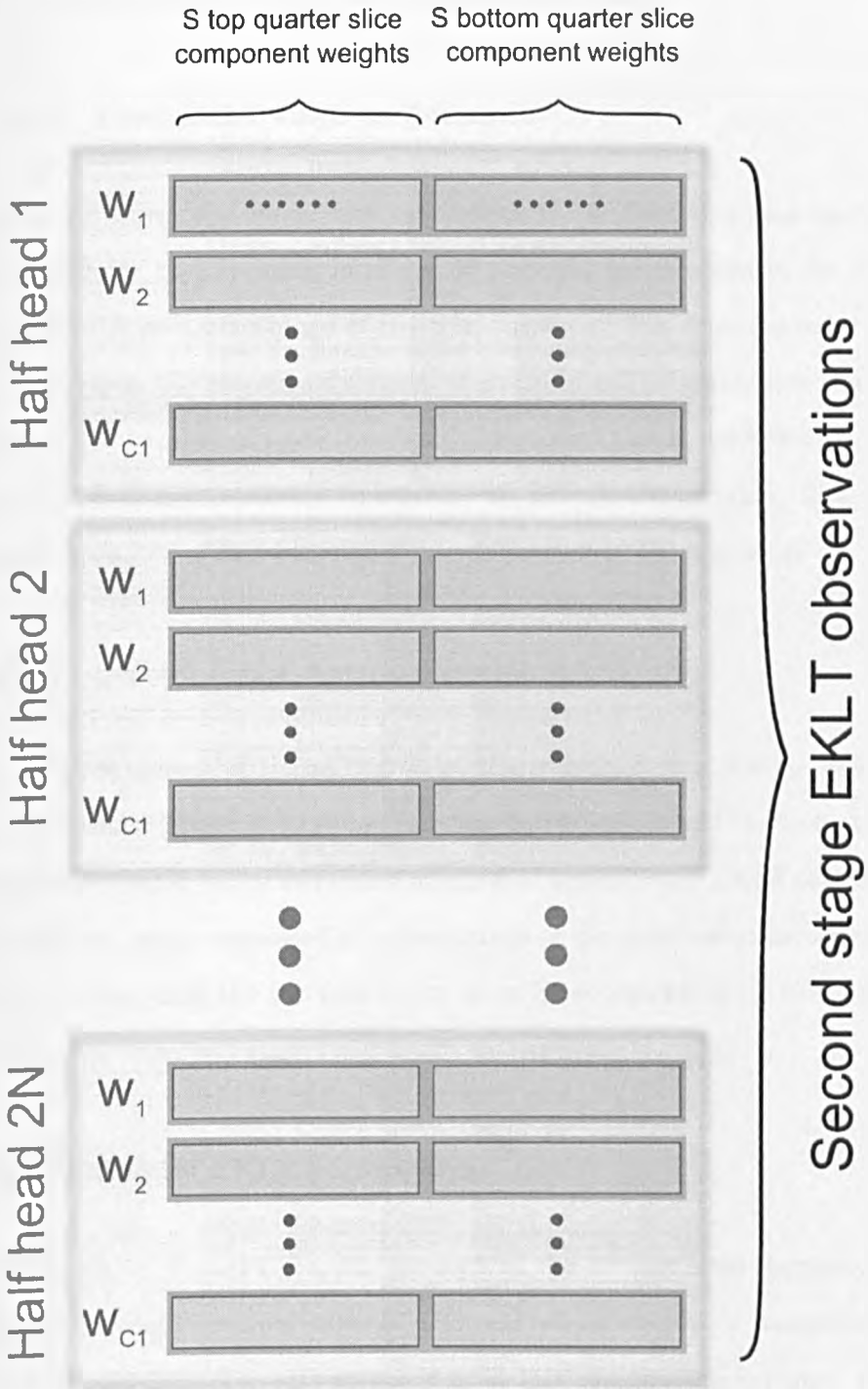


Figure 4.8: Second EKLT stage observations

head observations each described by 65,536 real values.

4.2.2.1 First EKLT stage performance

Fig. 4.9 shows the cumulative variance of the original slice observations accounted for by increasing numbers of principal components in the first stage EKLT, as a percentage of the total variance. The required accuracy depends upon the perceptual impact of errors in HRTF estimation due to limiting the number of basis functions. At present this is unknown. If, for example, it proves necessary to account for 99% of the variance, then the required number of first stage EKLT basis functions, C_1, equals 10.

4.2.2.2 Second EKLT stage performance

The performance of the second stage is determined in a similar way to the first stage. Figure 4.10 shows the cumulative variance of the first stage weights accounted for by increasing numbers of second stage EKLT principal components, again expressed as a percentage of the total variance. Assuming, as before, that the variance which must be accounted for is 99%, then the number of second stage basis functions required, C_2, is 12.

4.2.2.3 Overall EKLT performance

Using $C_1 = 10$ and $C_2 = 12$, for each of the 98 half-head descriptions reduces the data required from $65,536$ real values to 120, a compression factor of over 546. The loss resulting from this compression is shown as a point reconstruction error distribution in Figure 4.11. The majority of points in the original data set are reconstructed with less than 1 mm error with

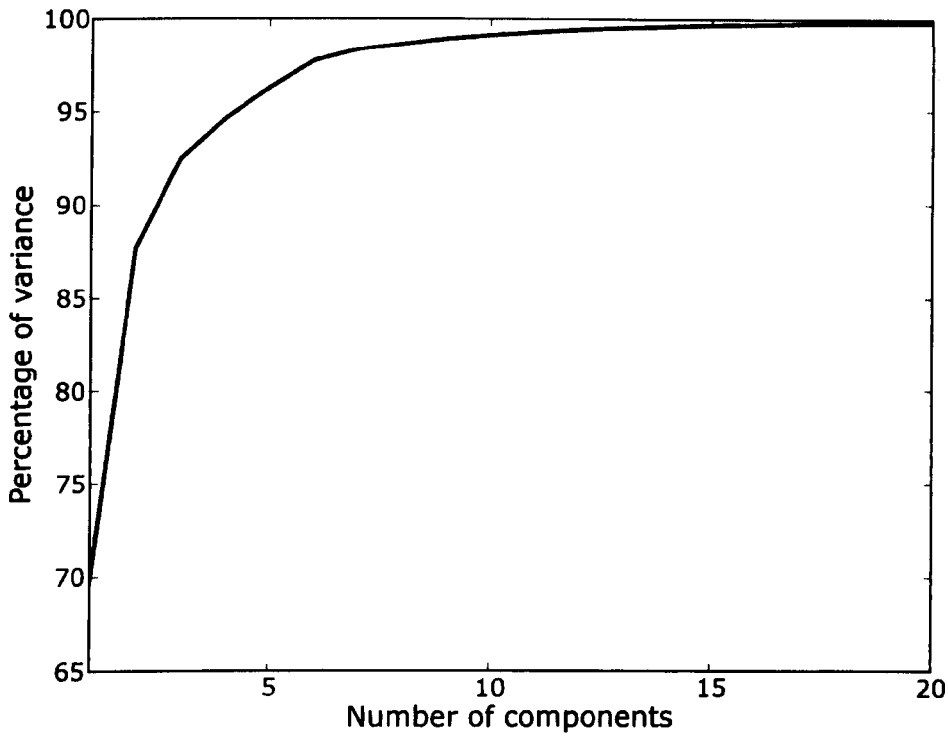


Figure 4.9: Cumulative variance of the original slice observations accounted for by increasing numbers of principal components in the first stage EKLT, as a percentage of the total variance.

fewer than 1% errors greater than 4 mm. The shape distortion caused by parameter truncation is small. Unlike EFT parameterisation this truncation does not result in the inherent elimination of pinna shape detail as most of that information is present in the lower order EKLT bases. The main source of errors are mesh imperfections and small anomalies, present in the original slice observations. Whether the proposed values for C_1 and C_2 are sufficient for the purposes of creating a perceptually valid morphology-to-HRTF mapping is a matter for investigation. Even if a higher number of components is necessary, the parameterisation method is still likely to

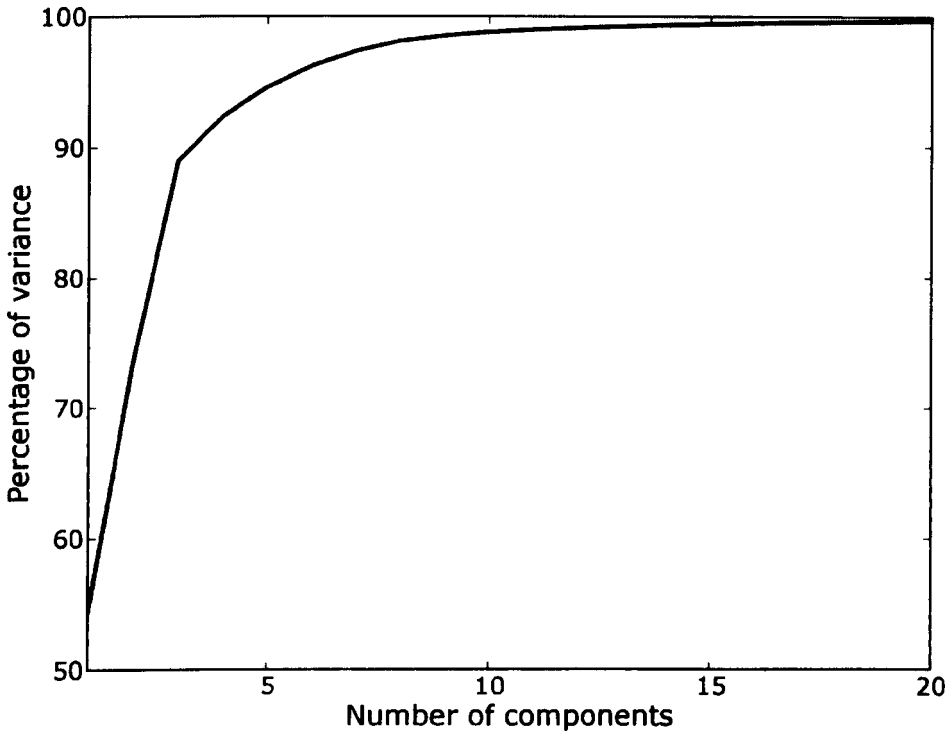


Figure 4.10: Cumulative variance of the first stage EKLT principal component weights accounted for by increasing numbers of principal components in the second stage EKLT, as a percentage of the total variance.

produce an extremely concise representation of the original data.

4.3 HRTF measurements and data parameterisation

4.3.1 Measurement procedure

The shape capture of the subjects' head, pinnae and torso was accompanied by corresponding HRTF measurements to form a morphoacoustic

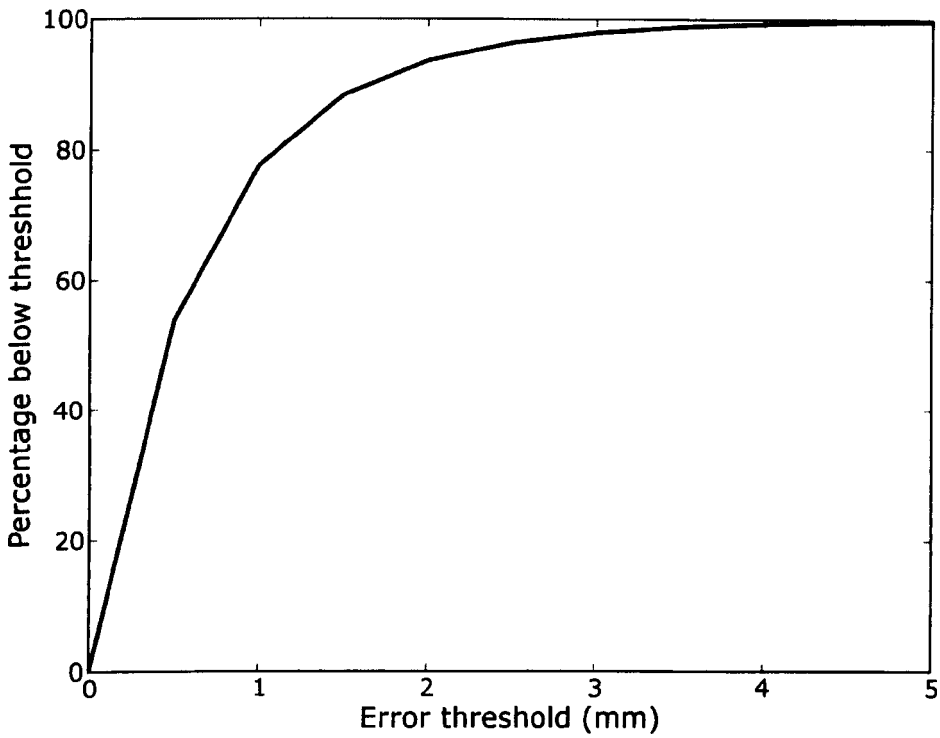


Figure 4.11: EKLT point reconstruction error distribution

database. The measurements were performed at the Auditory Neuroscience Laboratory[11] (ANL), located in the School of Medical Sciences[12], Discipline of Physiology[13] at the University of Sydney[14]. The Laboratory contains a 64 m^3 anechoic chamber with an insertion loss of better than 30 dBs for sound frequencies greater than 100 Hz rising rapidly to greater than 60 dBs above 500 Hz[15]. The absorption of the chamber is over 99% for all frequencies down to 200 Hz. The subject is located in the middle of the chamber at the center of an imaginary sphere of 1 m radius. A fully automated robotic

[11]Url: `http://www.physiol.usyd.edu.au/research/labs/auditory/`
[12]Url: `http://www.medsci.usyd.edu.au/`
[13]Url: `http://www.physiol.usyd.edu.au/`
[14]Url: `http://www.usyd.edu.au/`
[15]Url: `http://www.physiol.usyd.edu.au/research/labs/auditory/facilities.htm`

arm allows a small speaker to be placed anywhere on the surface of the sphere, down to an elevation of $-40°$ with placement errors under $0.1°$. The HRTF measurement procedure[16], developped by the ANL staff (Carlile and Pralong, 1995; Pralong and Carlile, 1994) is briefly summarised below.

A pair of AuSIM[17] in-ear microphones were fitted securely inside the ear canals of the subject using foam rings[18]. Microphone output signals were fed through a Sound Devices MP-1 Portable Microphone Preamp[19]. Subjects were then carefully placed in the center of the imaginary sphere traced by the speaker. This was achieved using laser beams to align the head correctly, so that the two ear drums and the tip of the nose lay in the horizontal plane and that the mid-point of the interaural axis corresponded to the origin of the measurement coordinate system.

After a calibration process design to reduce measurement errors, HRTFs were measured using complementary series[20], also known as Golay codes (see Golay, 1961), at regular angular intervals for 393 positions around the listener. During the procedure the frontal HRTF ($\theta = 0$, $\varphi = 0$) was re-checked several times in order to detect and eliminate unwanted changes (in microphone placement, for example). A head-tracker monitored involuntary subject movements and triggered a halt to the procedure if head position drifted beyond a tolerance threshold from the reference position. A set of coloured LEDs informed the subject of the unwanted drift in position and indicated the direction of the required correction. Once the head position

[16]Url: `http://www.physiol.usyd.edu.au/~simonc/hrtf_rec.htm`

[17]`http://www.ausim3d.com/about/index.html`

[18]The foam rings are produced by Etymotic Research, Inc. `http://www.etymotic.com/ha/ha-acc.aspx`

[19]`http://www.sounddevices.com/products/mp1master.htm`

[20]The series are 1024 samples long, recorded at 80kHz. They are repeated 16 times then averaged to improve SNR

was back within tolerance limits, the procedure resumed automatically.

4.3.2 HRTF Parameterisation

The acoustic data processing is beyond the scope of this thesis. The extensive nature of the data gathering exercise has delayed this crucial part of the project. HRTF data needs to be paramaterised in a way which will allow the relevant information they contain to be expressed comprehensively and concisely. For each ear, HRTF data consists of a large number of transfer functions, each associated with a different sound source position. A PCA analysis of HRTF data for all positions and subjects is a possible first step towards data compression. The spatial variation of principal component weights would then need be parameterised in some way for the purpose of further dimensionality reduction.

A number of studies have shown that PCA can be used as a powerful HRTF dimensionality reduction tool (Martens, 1987; Kistler and Wightman, 1992; Middlebrooks and Green, 1992; Chen *et al.*, 1995). Other, more recent studies, however, suggest that alternative methods such as Isomap and locally linear embedding (LLE), which adapt bases according to local neighbourhood information, perform better when attempting to encode the spatially varying character of HRTFs (Kapralos and Mekuz, 2007; Kapralos *et al.*, 2008). An alternative HRTF description technique proposed by (Grindlay and Vasilescu, 2007), based on a multi-linear (tensor) framework has also compared favourably with PCA.

A generally active area of current research, the development a comprehensive, yet concise description of HRTF variations across space and individ-

uals is key to the extraction of meaningful morphoacoustic mappings. We are currently working in close collaboration with Craig Jin and Andre van Schaik, researchers at the Computer & Audio Research Laboratory (CARlab[21]) and associates in the EPSRC funded project which gave rise to this thesis, in an attempt to improve the expression of HRTF data as much as current knowledge allows.

4.4 Discussion

The complete capture of head and pinna morphology using MRI, combined with a novel outer surface extraction procedure, provides a rich dataset. This dataset, in itself, is a valuable resource and will allow the exploration and analysis of shape variations across human subjects with a very high degree of precision. In addition, the EKLT shape parameterisation method enables a complete morphological description of the head and pinnae to be expressed using only a few hundred numbers. This method can be applied to greatly accelerate any statistical analysis.

In addition to the morphological dataset, corresponding HRTF measurements have been made for a large number of positions in state of the art facilities. These measurements constitute an extremely high-dimensional dataset. Even though an HRTF for a given position can be expressed using only the first few most significant principal components, HRTFs have been measured for 393 positions. Assuming 5 principal components are required to completely reconstruct a single monaural HRTF, a complete HRTF dataset for a given individual would need $393 \times 2 \times 5$ or 3730 val-

[21] `http://www.ee.usyd.edu.au/research/allresearch/?group=carlab`

ues; a complete mismatch with the size of the corresponding, parameterised morphological data.

A technique allowing spatial HRTF variation to be expressed concisely is required. As described in Section 4.3.2, this is a highly active field of research and a number of different parameterisation techniques have given good results. A performance comparison will be needed to determine the best option for our dataset. Once the HRTF data is adequately compressed, the extraction of mappings between the morphological and acoustic datasets is expected to result in a significant improvement over previous studies (see Section 2.4.4), while the data compression will avoid unreasonable computing requirements.

The complete description of head and pinnae morphology presented in this chapter greatly reduces the chance of omitting shape detail which is relevant to the production of localisation cues. However, the method inevitably provides excessive shape detail in areas that are irrelevant to this process. This is expected to produce unwanted effects. Namely, the extraction of morphoacoustic mappings needed for effective HRTF estimation will be complicated by large amounts of superfluous data. Also, the required shape capture imposes unrealistic demands on a morphology measurement system, which should be generally accessible. Adapting the method to focus morphological description on areas responsible for the cue production mechanism, retaining accurate surface description only where necessary, seems a viable route towards solving these problems.

Chapter 5

Acoustic Effects of Shape Variations

"We learn more by looking for the answer to a question and not finding it than we do from learning the answer itself."

Lloyd Alexander

"Nothing is a waste of time if you use the experience wisely."

Auguste Rodin

The steady growth in computing power has permitted acoustic simulations using the boundary element method (BEM) to be performed with correspondingly decreasing patch size. This has led to the BEM being validated with good accuracy against real acoustic measurements for ever-increasing frequencies (Walsh *et al.*, 2003; Otani and Ise, 2006; Kahana and Nelson, 2005, 2007). However, software implementations are costly and valid simulations of the human head and pinnae in the upper audible frequency range

still require extremely powerful computing platforms. This has so far prevented the widespread use of acoustic simulations for HRTF estimation. The differential pressure synthesis (DPS) method proposed by Tao *et al.* (2003a) (see Section 2.5) was originally developed to accelerate the estimation of individualised HRTFs by exploiting underlying morphological similarities across individuals.

The DPS estimation process uses pre-computed acoustic fields around a template head shape. Each field is associated with an orthogonal shape deformation to the template. Arbitrary small deformations may be synthesised as a weighted sum of this orthogonal set and the associated acoustic effects estimated by superposition (see Section 2.5). The range of deformations over which a sufficiently linear mapping exists between shape changes and acoustic pressure changes (a requirement for the application of the DPS principle) was found to be too small compared with the natural variations observed across individuals (Tao *et al.*, 2003a). However, DPS remains a powerful tool for investigating the acoustic effects of perturbing the shape of the template head mesh and it is in this role, referred to as *morphoacoustic perturbation analysis* (MPA), that it is applied in this chapter.

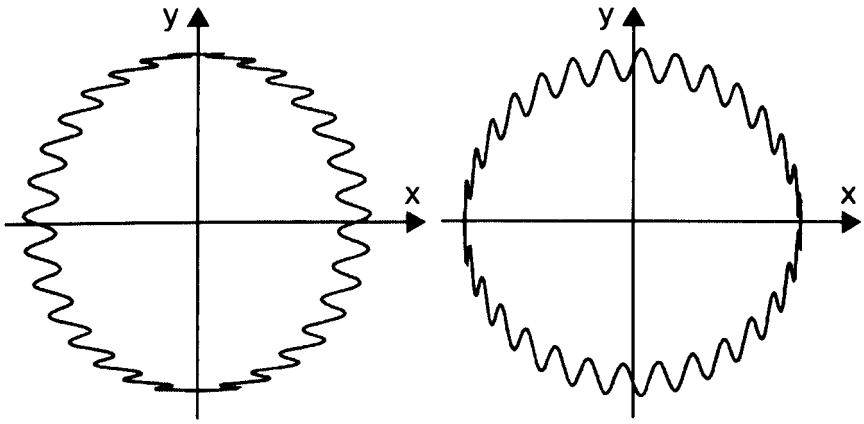
As explained in Section 2.5.3.3, the surface spherical harmonic (SSH) deformations used by Tao *et al.* could not be used to describe the external ear region. Their work was thus constrained to investigating the validity of DPS principles for simplified pinna-less heads and the lower frequency bands where such a simplification is acceptable. This chapter describes new orthogonal deformations which include the pinna. The compilation of a DPS database for a KEMAR head using these deformations is described and an analysis of its performance is presented. This analysis focuses particularly

on the higher frequencies where spectral cues are known to operate.

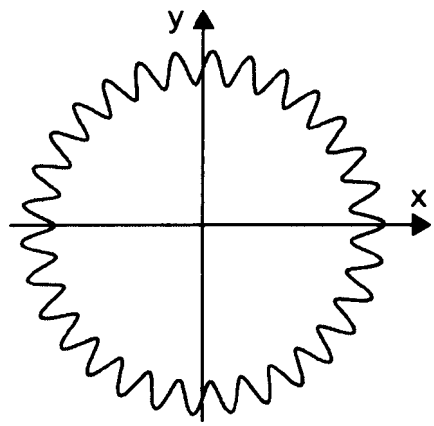
5.1 Orthogonal deformation of the human head and pinnae

The DPS technique relies on a pre-computed database which describes the acoustic effects of applying a set of orthogonal shape deformations to a template shape. The orthogonality of the shape deformations ensures that, by summing them in the appropriate proportions, an *arbitrary* shape deformation can be applied to the template. Limitations arise, however, due to the range and nature of the orthogonal deformations for which the database has been compiled. In the same way, a Fourier series can represent an "arbitrary" signal but requires this signal to follow certain rules, i.e. being single valued and differentiable.

Hetherington *et al.* used the elliptic Fourier transform (EFT) shape parameterisation technique to express head and pinna morphology (Hetherington and Tew, 2003; Hetherington *et al.*, 2003). The technique is described in detail in Section 2.4.1.2. It requires a slice description of the object of interest to be obtained (see Figure 2.11), then applies a transform on the x- and y-component slice signals. This transformation results in a set of EFT parameters, which can then be perturbed giving rise to orthogonal deformations. EFT parameter perturbations fulfil two requirements for constructing a DPS database similar to that used by Tao *et al.* (2003a); namely, they are orthogonal and allow the head and pinnae to be described entirely. However, potential problems were identified.



(a) EFT x-component deformation (b) EFT y-component deformation



(c) Contour harmonic deformation

Figure 5.1: The effect of EFT parameter perturbations on one of the slices of a sphere description in the case of x-component and y-component EFT parameters. In the case of x-component deformations (a) the oscillations are spatially compressed when the contour slope is near horizontal and in the case of y-component deformations (b), the effect occurs when the contour slope is near vertical. A sinusoidal deformation applied perpendicularly to the contour of the slice (c) referred to as a contour harmonic deformation, is a possible solution to the problem.

Individual EFT parameter perturbations do not result in smooth, evenly spatially distributed deformations like those introduced by spherical harmonic deformations. In fact, the spatial frequency of these deformations

changes erratically over the surface. For each slice, a single EFT parameter perturbation results in a sinusoidal signal being superimposed on either its parametric x- or y-component signals. Consequently, the spatial frequency of the deformations changes depending on the slope of the parameterised slice contour. This is most obvious when the contour slope is horizontal in the case of x-component perturbations, shown in Figure 5.1(a), or vertical in the case of y-component perturbations, shown in Figure 5.1(b).

The problem is accentuated in the case of head slices because the pinna contour slope often changes very rapidly. Rapid spatial variation markedly reduces the range for which a linear approximation of the relationship between deformation amplitude and associated acoustic effects is valid; a requirement for DPS estimation. Also, in regions of high spatial variation, spatial aliasing resulting from the finite mesh resolution is a significant problem. The next section presents a solution, based on applying oscillatory deformations perpendicularly to the slice contour, as shown in Figure 5.1(c).

5.1.1 Elliptic surface harmonic deformations

Elliptic surface harmonic deformations are a variation on EFT parameter perturbations better suited to the compilation of a DPS database. The first step towards generating them is identical to that used in EFT parameterisation and consists in obtaining a slice description of the head mesh (see Section 2.4.1.2, Figure 2.11). This process results in S slices, each containing P points uniformly spread along its contour. Next, the points comprising the head surface slice description are mapped to points on a rectangular, two-dimensional surface. Specifically, the pth point along the contour of the sth slice defines a corresponding point at

$$(x, y, z) = (s, p, 0) \tag{5.1}$$

in three-dimensional Cartesian space. The sliced representation of the head shape (including the pinnae) is therefore mapped to a set of points spread on a rectangle of dimensions $S \times P$ on the xy plane of a three-dimensional Cartesian coordinate system. Each point in the slice description maps to a corresponding integer (x, y) pair obeying $0 \leqslant x < S$ and $0 \leqslant y < P$. A deformation applied to this flat surface by altering the z coordinate associated with each (x, y) pair can be described in terms of a discrete two-dimensional inverse Fourier transform. Equating z to $f[x, y]$, this may be expressed as

$$f[x, y] = \sum_{u=0}^{S-1} \sum_{v=0}^{P-1} F[u, v] e^{2\pi j(xu/S + yv/P)} \tag{5.2}$$

where

$$F[u, v] = \frac{1}{SP} \sum_{x=0}^{S-1} \sum_{y=0}^{P-1} f[x, y] e^{-2\pi j(xu/S + yv/P)} \tag{5.3}$$

The deformation applied to the flat surface can then be mapped back to the slice description. Denoting $d_{s,p}$ the deformation to the pth point on the sth slice, applied perpendicularly to the local slice contour. The mapping is performed, simply, by setting

$$d_{s,p} = f(s, p) \tag{5.4}$$

As the slice description is cyclic, or more specifically, two-fold periodic in nature, any deformation applied to the xy plane representation should also exhibit this property in order to avoid a discontinuity in the slice description. When this is the case, the transformation described in Equation 5.3 can be performed without the need for a windowing function. The exponential form of the discrete two-dimensional inverse Fourier transform (Equation 5.2) can be separated into a sine and a cosine part:

$$\begin{aligned} f[x,y] &= \sum_{u=0}^{S-1}\sum_{v=0}^{P-1} F[u,v]\cos\left(2\pi(xu/S+yv/P)\right) \quad \cdots \\ &+ \sum_{u=0}^{S-1}\sum_{v=0}^{P-1} jF[u,v]\sin\left(2\pi(xu/S+yv/P)\right) \end{aligned} \tag{5.5}$$

Since the deformation $f[x,y]$ is real, $F[u,v]$ displays two-dimensional Hermitian symmetry. That is,

$$\begin{aligned} & F[S-u,P-v]\ \cos 2\pi\left(x(S-u)/S\ +\ y(P-v)/P\right) \\ &= F[u,v]^*\ \cos\left(2\pi(xu/S\ +\ yv/P)\right) \end{aligned} \tag{5.6}$$

and

$$\begin{aligned} & F[S-u,P-v]\ \sin\left(2\pi(x(S-u)/S\ +\ y(P-v)/P)\right) \\ &= -\ F[u,v]^*\ \sin\left(2\pi(xu/S\ +\ yv/P)\right) \end{aligned} \tag{5.7}$$

Using these identities, Equation 5.5 becomes

$$f[x,y] = \sum_{u=0}^{S/2-1} \sum_{v=0}^{P/2-1} (F[u,v] + F[u,v]^*) \cos(2\pi(xu/S + yv/P)) \quad \cdots$$

$$+ \sum_{u=0}^{S/2-1} \sum_{v=0}^{P/2-1} j(F[u,v] - F[u,v]^*) \sin(2\pi(xu/S + yv/P)) \qquad (5.8)$$

Substituting

$$A_{u,v} = F[u,v] + F[u,v]^* = 2|F[u,v]| \cos \angle F[u,v] \qquad (5.9)$$

$$B_{u,v} = j(F[u,v] - F[u,v]^*) = -2|F[u,v]| \sin \angle F[u,v] \qquad (5.10)$$

Equation 5.8 may be written as

$$f[x,y] = \sum_{u=0}^{S/2-1} \sum_{v=0}^{P/2-1} A_{u,v} \cos(2\pi(xu/S + yv/P)) \quad \cdots$$

$$+ \sum_{u=0}^{S/2-1} \sum_{v=0}^{P/2-1} B_{u,v} \sin(2\pi(xu/S + yv/P)) \qquad (5.11)$$

where $A_{u,v}$ and $B_{u,v}$ are real coefficients. Once it is mapped back to the slice representation (see Equation 5.4), the deformation produced by perturbing a single one of these coefficients is referred to as an elliptic surface harmonic deformation of slice-harmonic v and cross-harmonic u. A DPS database is compiled using these deformations. This entails systematically deforming a template shape for all (u, v) pairs and calculating the acoustic effect of each deformation using the BEM (see Section 2.5). Following the same notation

conventions as Tao *et al.* (2003a), a new variable

$$E^{\sigma}_{u,v} = \begin{cases} A_{u,v} & \text{for } \sigma = 0 \\ B_{u,v} & \text{for } \sigma = 1 \end{cases} \tag{5.12}$$

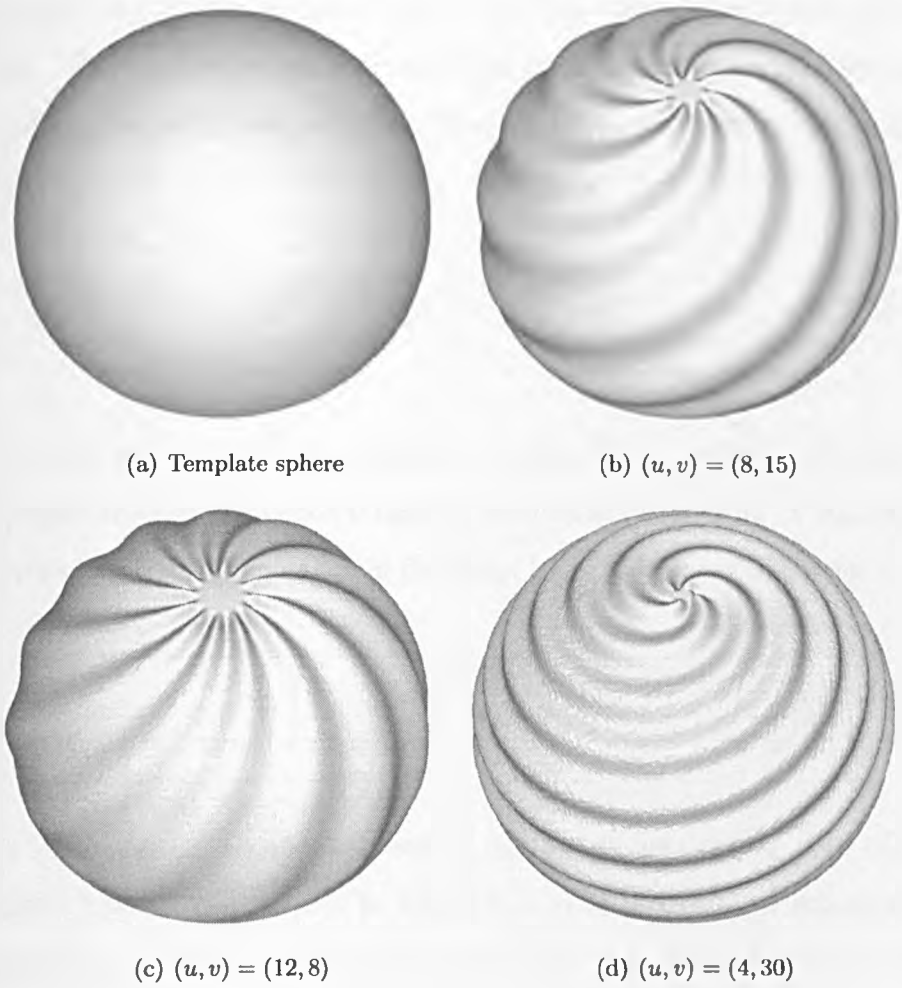
is introduced to simplify notation allowing the pressure difference dp caused by an infinitesimal arbitrary deformation to be expressed as the following summation

$$dp = \sum_{\sigma=0}^{1} \sum_{u=0}^{S/2-1} \sum_{v=0}^{P/2-1} \frac{\partial p}{\partial E^{\sigma}_{u,v}} dE^{\sigma}_{u,v} \tag{5.13}$$

As in the cases described in Section 2.5.2, within the deformation range where the relationships between deformation amplitude $E^{\sigma}_{u,v}$ and acoustic pressure p is substantially linear, the difference in pressure Δp can be approximated by

$$\Delta p = \sum_{\sigma=0}^{1} \sum_{u=0}^{S/2-1} \sum_{v=0}^{P/2-1} \frac{\partial p}{\partial E^{\sigma}_{u,v}} \Delta E^{\sigma}_{u,v} \tag{5.14}$$

The greater the shape detail required in the analysis, the higher the upper limits for u and v required in the DPS database. The effects of applying to a template sphere a selection of the deformations obtained by changing the $A_{u,v}$ coefficients in Equation 5.11, are shown in Figure 5.2. As the slices converge toward the slicing axis in the central pinna region, the distance between them shrinks and, consequently, cross-slice oscillations become spatially tighter. To avoid spatial aliasing in the area around the crus helias and the cymba concha, spatial filtering is applied to eliminate surface



(a) Template sphere

(b) $(u, v) = (8, 15)$

(c) $(u, v) = (12, 8)$

(d) $(u, v) = (4, 30)$

Figure 5.2: A spherical template (a) to which have been applied example elliptic surface harmonic deformations (b), (c), (d). In each case, the deformations are obtained by perturbing the coefficient $A_{u,v}$ in Equation 5.11, to which the u and v values shown have been assigned. $B_{u,v}$ changes result in similar deformations, with the pattern rotated around the slicing axis. The value of v, in effect, controls the intensity of a visible spiralling effect. Note that as u increases, the area where anti-aliasing spatial filtering is effective grows.

oscillations whose spatial frequency cannot be supported by the mesh. The spatial frequency of cross-slice oscillations is a function of the cross-harmonic u and the distance from the slicing axis. For the pth point on the sth slice,

this distance is noted as $a_{s,p}$. Given a minimum cross-oscillation wavelength λ_{min}, determined by the maximum edge length of the mesh, the deformation $d_{s,p}$ applied to this point (see Equation 5.4) is zeroed when $a_{s,p} < a_{min}$, where

$$a_{min} = \frac{u\lambda_{min}}{2\pi} \tag{5.15}$$

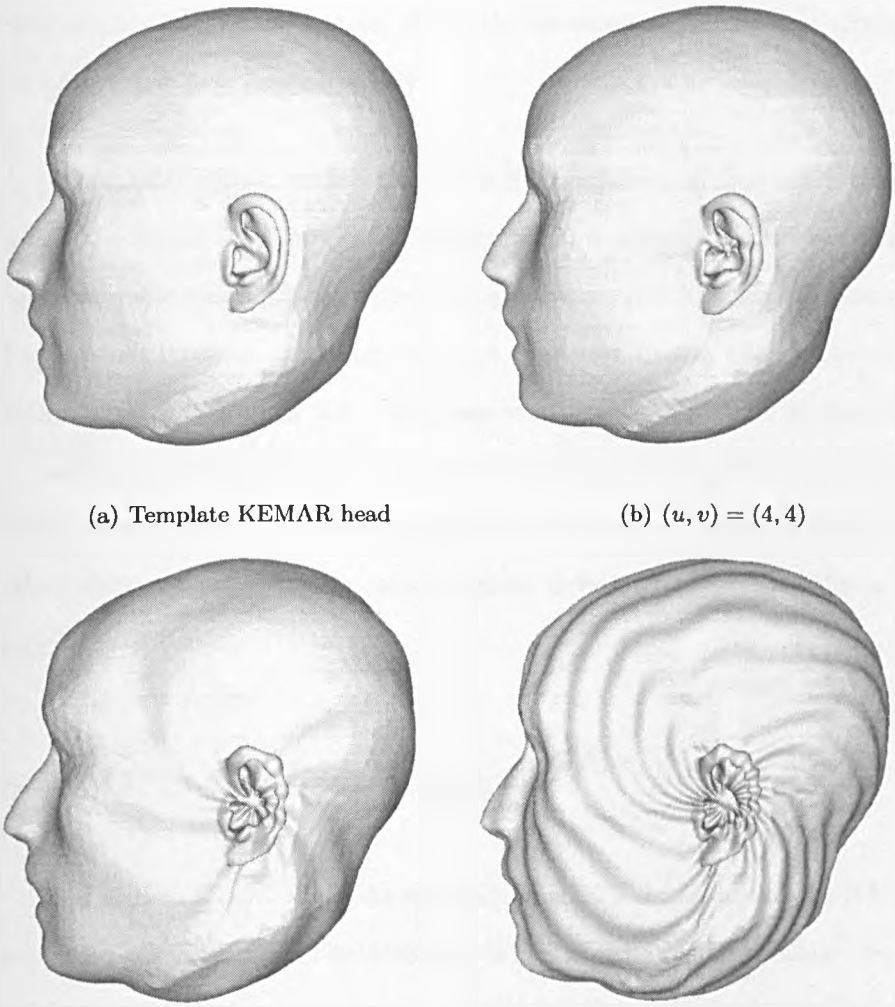
In order to avoid an abrupt change at distance a_{min} from the slicing axis, a cosine windowing function is applied, starting at distance $a_f > a_{min}$ from the slicing axis. When $a_{s,p}$ is in the range $[a_{min}, a_f]$, $d_{s,p}$ is scaled by

$$\frac{1}{2}\left(1 - \cos\left(\frac{(a_{s,p} - a_{min})\pi}{a_{min} - a_f}\right)\right) \tag{5.16}$$

When $a_{s,p} > a_f$, the deformation $d_{s,p}$ is left unchanged. The effects of this filtering can be seen in Figure 5.2. The proportional relationship between a_{min} and u is noticeable, indeed, the area where filtering occurs grows with increasing u.

5.1.2 Deformations in practice

To employ elliptic harmonic surface deformations for the creation of a DPS database, a slice description of the object of interest is required. In this case, the KEMAR head and pinnae were chosen, because of their widespread use in acoustics research. The slice description is obtained from the KEMAR mesh model obtained as described in Section 3.2, using radial slicing around a slightly shifted version of the interaural axis (see Figure 2.11). As in the



(a) Template KEMAR head

(b) $(u, v) = (4, 4)$

(c) $(u, v) = (8, 12)$

(d) $(u, v) = (12, 30)$

Figure 5.3: The KEMAR head template (a) to which have been applied example elliptic surface harmonic deformations with amplitude 2 mm (b), (c), (d). As in Figure 5.2, the deformations were obtained by perturbing the coefficient $A_{u,v}$ in Equation 5.11, to which the u and v values shown have been assigned. $B_{u,v}$ changes result in similar deformations, with a pattern rotated around the slicing axis. The low u, v values in (b) produce subtle effects, which are visible in the pinna area but hardly noticeable away from the slicing axis.

case of EFT (and EKLT) parameterisation, each slice is required to consist of a single continuous contour if elliptic harmonic surface deformations are to be applied (see Section 4.1.2).

Examples of elliptic surface harmonic deformations applied to the KEMAR mesh are shown in Figure 5.3. Deformation amplitudes have been exaggerated to make their effects more clearly visible. For high (u, v) values (see Figure 5.3(d)), the spiralling patterns observed in the case of sphere deformations (see Figure 5.2), are somewhat irregular. This is due to the variation in contour length from one slice to the next, which can be relatively sudden due to occasionally rapid changes in the shape of pinna. The effect does not occur in the case of sphere deformations, where the contour length is constant.

5.2 DPS database creation

This section describes the compilation of a DPS database for the KEMAR head and small pinnae. The database is a catalogue of the acoustic pressure changes resulting from applying a large number of individual elliptic surface harmonic deformations (ESHDs) to the template KEMAR shape. The main motivation for this effort is to accelerate the exploration of the acoustic effects of making small shape changes to this template. Indeed, using the principles of DPS described in Section 2.5, the database allows the acoustic effects of arbitrary micro-deformations to the template to be computed relatively quickly, without the need for further BEM simulation. To achieve this, DPS estimation requires a lengthy and expensive database compilation process, but this only needs to be performed once. A full BEM acoustic sim-

ulation is performed for each orthogonal deformation comprising the DPS database (see Sections 2.4.6 and 3.1). Since acoustic simulations are lengthy and a large number of deformations is required to build a database of sufficient size, many factors had to be considered carefully before creating it. These are discussed in turn.

5.2.1 Mesh resolution and topology

The multi-resolution KEMAR head mesh described in Section 3.3.3 formed the starting point for creating the DPS database. It is comprised of 3,816 polygons. The time savings provided by multi-resolution meshing are particularly desirable in the case of DPS database building, given the large number of simulations required and the strong effect of patch count on BEM simulation time. Results obtained using the multi-resolution mesh were validated in Chapter 3, for the 10.0-14.8 kHz frequency range, against a high-resolution mesh for which edge length did not exceed a quarter wavelength at 15 kHz. The multi-resolution mesh pressures deviated from the high-resolution mesh by no more than 1 dB for sources in the median plane (see Figures 3.4 and A.1), the only exception being in the back HRTF (Figure 3.5), where a sharp notch caused slightly greater disagreement in a very restricted frequency band. The performance of the multi-resolution mesh was shown to improve as sources enter the ipsilateral region (see Figure 3.6). Further empirical support for the mesh's performance comes from the simulation results discussed in Section 3.4, which compare favourably with published real and virtual measurements (Shaw and Teranishi, 1968; Kahana and Nelson, 2005; Hebrank and Wright, 1974, amongst others).

5.2.2 Deformation amplitude

Two main factors were taken into consideration when deciding on the amplitude of elliptic surface harmonic deformations used to compile the DPS database. Firstly, the acoustic effects of deformations should clear computation noise levels generated by the solving of BEM equations. However, these acoustic effects must be kept linear as much as possible, which places an upper limit to deformation amplitude. These issues were considered in detail.

5.2.2.1 BEM computation noise

The lower limit for the deformation amplitude is dictated by the computation noise resulting from the BEM. A series of tests were conducted to ascertain the most appropriate amplitude to use. The pressure change was computed, using a mesh similar to the template, for a series of small deformations with slowly increasing amplitude. This investigation revealed computation noise most clearly for very small incremental deformation amplitudes, whilst pressure changes due to the deformation dominate at larger amplitudes. This analysis was conducted for a number of HRTF directions and frequencies. As a result of this investigation, it was decided that each ESHD should be applied to the template slice set with a peak amplitude of 0.3 mm, in order for associated acoustic effects to adequately clear computation noise levels. After the database building process, updates to the BEM-based simulation software allowed computation noise levels to be lowered significantly. Although re-building the DPS database using a smaller orthogonal deformation amplitude would be preferable, time constraints did not allow this to be done within the scope of this project. The following sub-

section describes non-linear acoustic effects which could be mostly avoided by lowering the amplitude of orthogonal deformations.

5.2.2.2 Investigating non-linear behaviour

Determining the range where the acoustic effects of ESHDs can be considered linear is of crucial importance as it ultimately determines the maximum magnitude of the arbitrary deformations for which DPS estimations are valid (see Section 2.5). This section describes a number of tests which were carried out in order to investigate the extent of this range for a number of different (u, v) pairs and simulation frequencies, for HRTFs in the front and left (ipsilateral) positions. Linearity was assessed by performing a number of BEM-based acoustic simulations at small linearly-spaced deformation amplitude intervals, ranging from -0.5 mm to 0.5 mm. The acoustic pressure magnitude is, as before and for the remainder of the chapter, relative to a unit magnitude sound source. In each case a best-fit linear approximation to the curve, obtained through linear regression, is shown alongside the data. The slope (m) and intersect (b) of the best linear fit, for a set of frequency and pressure magnitude pairs (f_1, p_1), (f_2, p_2) . . . (f_N, p_N) are calculated as

$$m = \frac{n\sum_{n=1}^{N} f_n p_n - \sum_{n=1}^{N} f_n \sum_{n=1}^{N} p_n}{n\sum_{n=1}^{N} f_n^2 - \left(\sum_{n=1}^{N} f_n\right)^2} \tag{5.17}$$

and

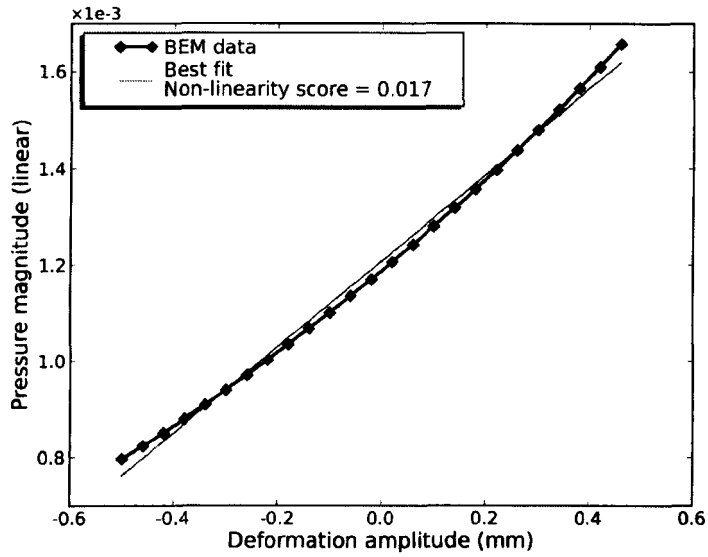
$$b = \frac{\sum_{n=1}^{N} p_n - m\sum_{n=1}^{N} f_n}{n} \tag{5.18}$$

This method is known as ordinary least squares and minimizes the sum of the squares of errors. From this, a measure of non-linearity is calculated by summing the absolute pressure differences between the linear pressure predictions (l_n) and the actual pressures (p_n), and dividing the result by the range of pressure change over the deformation amplitude range ($p_{max} - p_{min}$), as follows

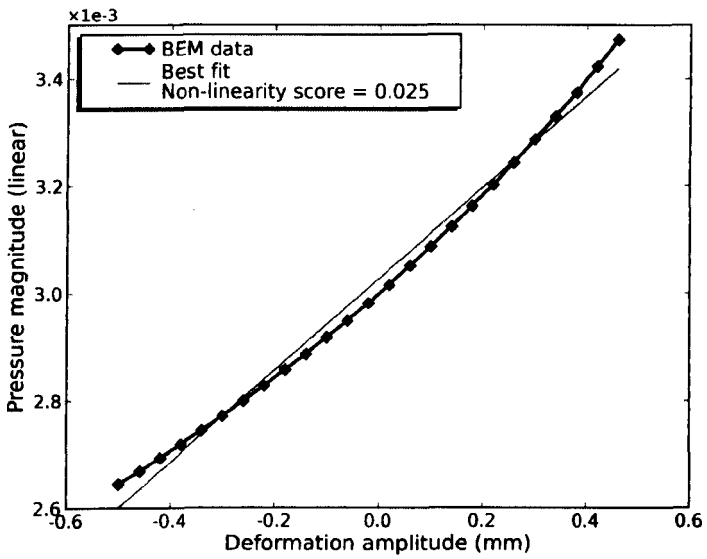
$$\text{Non-linearity score} = \frac{\sum_{n=1}^{N} |p_n - l_n|}{p_{max} - p_{min}} \tag{5.19}$$

The results were obtained using the updated version of the BEM-based simulation software, hence the near absence of computation noise described in Section 5.2.2.1. For low u and v values pressure variation is approximately linear over the full range of deformation amplitudes, for all tested frequencies and both for a frontal and ipsilateral sound source. For example, Figure 5.4 shows the linearity plot for the $\cos(2\pi(2x/S + 2y/P))$ ESHD at 10 kHz, where a linear approximation of the curve is relatively accurate over the entire amplitude range (low non-linearity scores of 0.017 for the frontal HRTF and 0.025 for the left HRTF). Non-linearity scores are similar at lower frequencies for the same u and v values (see Figures C.1 and C.2 in Appendix C). It should be noted that the slope of the curve, changes radically with frequency.

For higher frequency slice-harmonics (v), an approximately linear relationship is visible throughout the ± 0.5 mm range of deformation amplitudes investigated. Figure 5.5 shows the linearity plot for the $\cos 2\pi(2x/S + 15y/P)$ ESHD at 10 kHz, with a low non-linearity score. For constant frequency,



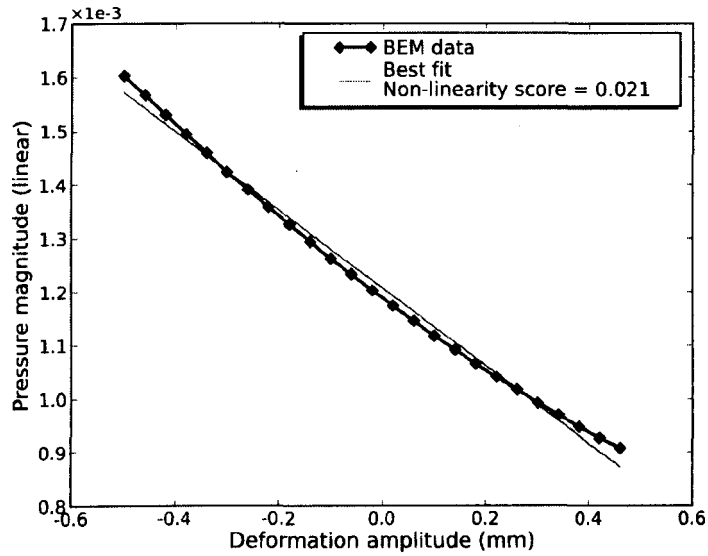
(a)



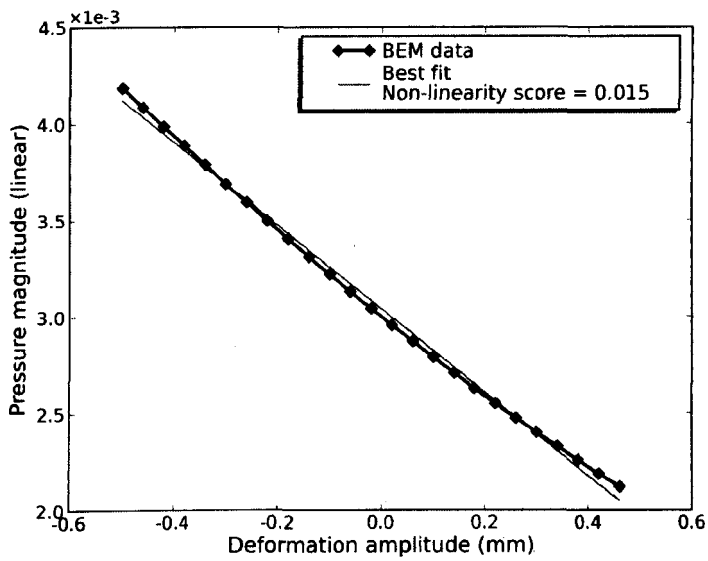
(b)

Figure 5.4: Deformation peak amplitude vs pressure plots for the $\cos 2\pi(2x/S + 2y/P)$ ESHD at 10 kHz, for the front (a) and left (b) positions. A linear approximation to the curve is accurate over the entire range in both cases.

the slope of the curve changes with the slice harmonic v (see Figures 5.4 and 5.5). The quality of the linear approximation stagnates as frequency



(a)



(b)

Figure 5.5: Deformation peak amplitude vs pressure plot for the $\cos 2\pi(2x/S + 15y/P)$ ESHD at 10 kHz, for the front (a) and left (b). Here, again, a linear approximation to the curve is accurate over the entire range in both cases. Note that, in each case, the slope of the curve is opposite to that observed for the $\cos 2\pi(2x/S + 2y/P)$ ESHD at the same frequency (see Figure 5.4).

decreases (see Figure C.3, Appendix C).

Deformations with higher cross-harmonics (u) generate strongly characteristic acoustic behaviour at high frequencies. Figure 5.6 shows the linearity plot for the $\cos 2\pi(15x/S + 2y/P)$ ESHD at 10 kHz. A quadratic approximation of the relationship between deformation amplitude and pressure seems to be more appropriate than a linear one in this case. The parabolic appearance is slightly less marked in the case of the left source and the frequency minimum is shifted slightly towards negative deformation amplitudes (see Figure 5.6(b)). This behaviour also occurs at 5 kHz (see Figure C.6), although it is less pronounced. At 200 Hz, the plot is more linear in appearance over the full range of deformations (see Figure C.5), although the non-linearity score is still superior to that observed for the $\cos 2\pi(2x/S + 2y/P)$ and $\cos 2\pi(2x/S + 15y/P)$ ESHD.

For ESHDs with both high cross-harmonic (u) and high slice-harmonic (v) values, the same approximately quadratic relationship is observed at high frequencies (see Figure 5.7). It is not as marked as in the previous case (Figure 5.6), especially for the left source, but this may be peculiar to the particular values of u and v, which have been chosen somewhat arbitrarily. A parabolic appearance is also observed at 5 kHz (Figure C.8). At lower frequencies, any parabolic relationship is probably broader and so is left relatively unrevealed by the limited range of deformations investigated (see Figure C.7). Linearity scores for all tested (u, v) pairs, frequencies and positions are shown in Tables 5.1 and 5.2.

The observed parabolic relationships pose a particular challenge when deciding on a suitable deformation amplitude to use for generating the DPS

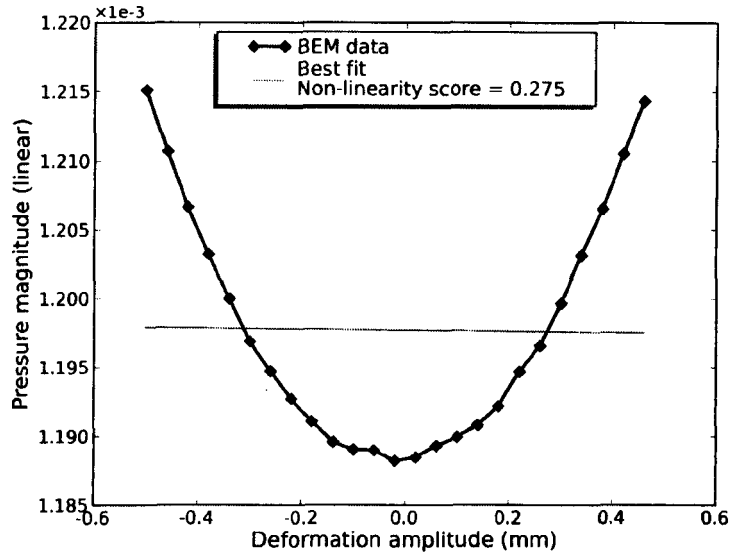
(u, v)	Frequency (kHz)	Position	Non-linearity	Range
(2,2)	0.2	Front	0.009	9.73×10^{-08}
(2,2)	0.2	Left	0.008	1.23×10^{-06}
(2,2)	5.0	Front	0.029	2.91×10^{-05}
(2,2)	5.0	Left	0.006	1.57×10^{-04}
(2,2)	10.0	Front	0.017	8.60×10^{-04}
(2,2)	10.0	Left	0.006	8.27×10^{-04}
(2,15)	0.2	Front	0.018	2.57×10^{-07}
(2,15)	0.2	Left	0.023	4.54×10^{-07}
(2,15)	1.0	Front	0.025	2.78×10^{-06}
(2,15)	1.0	Left	0.032	2.81×10^{-06}
(2,15)	5.0	Front	0.001	2.10×10^{-04}
(2,15)	5.0	Left	0.009	7.08×10^{-04}
(2,15)	10.0	Front	0.021	6.94×10^{-04}
(2,15)	10.0	Left	0.015	2.06×10^{-03}

Table 5.1: Non-linearity scores and ranges of pressure magnitude perturbation caused by ESHDs with (u, v) pairs $(2, 2)$ and $(2, 15)$ and with amplitude ranging ± 5 mm. The non-linearity score, calculated as described in Equation 5.2.2.2, is shown for different frequencies, for front and left sound sources. It stays below 0.04 and is generally much lower. The pressure magnitude perturbation range generated by ESHDs rises by around four orders of magnitude as frequency increases from 200 Hz to 20 kHz.

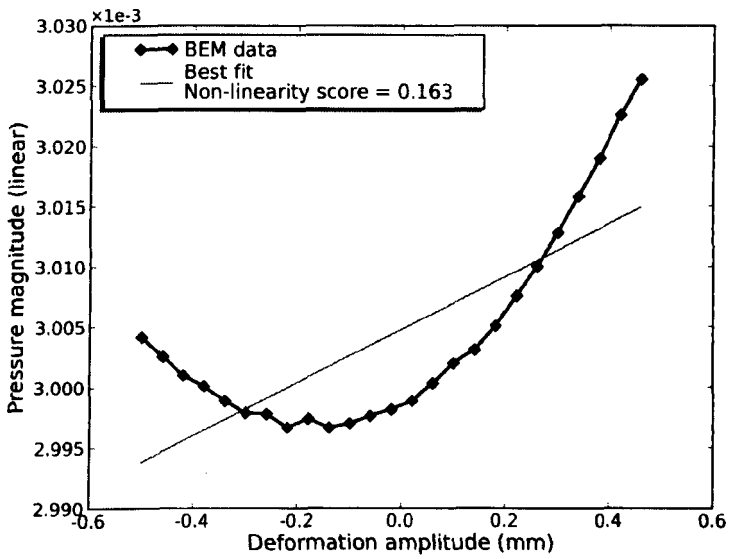
database and it is worth considering its origin. In general, a non-linear relationship between the amplitude of an ESHD applied to a template and the resulting pressure change is to be expected. For example, the resonant characteristics of cavities in the template shape may be modified by the introduction of a particular harmonic. It is observed that the higher cross-harmonic (u) values generate the most rapid deviation from linearity over the range of deformations tested. These often take the form of an approximate parabola (see Figure 5.6, for example). Observing that deformations of opposite amplitude are, in fact, equal amplitude deformations rotated about the slicing axis provides some explanation for this behaviour. Given an ESHD of cross-harmonic u applied to a spherical template (see Figure 5.2),

(u, v)	Frequency (kHz)	Position	Non-linearity	Range
(15,2)	0.2	Front	0.043	2.59×10^{-07}
(15,2)	0.2	Left	0.015	1.60×10^{-06}
(15,2)	1.0	Front	0.199	7.09×10^{-07}
(15,2)	1.0	Left	0.008	1.77×10^{-06}
(15,2)	5.0	Front	0.118	1.17×10^{-05}
(15,2)	5.0	Left	0.243	5.66×10^{-06}
(15,2)	10.0	Front	0.275	2.69×10^{-05}
(15,2)	10.0	Left	0.164	2.88×10^{-05}
(15,15)	0.2	Front	0.091	2.55×10^{-07}
(15,15)	0.2	Left	0.037	1.93×10^{-06}
(15,15)	1.0	Front	0.194	9.17×10^{-07}
(15,15)	1.0	Left	0.047	1.60×10^{-06}
(15,15)	5.0	Front	0.165	1.29×10^{-05}
(15,15)	5.0	Left	0.170	1.18×10^{-05}
(15,15)	10.0	Front	0.124	5.21×10^{-05}
(15,15)	10.0	Left	0.068	9.67×10^{-05}

Table 5.2: Non-linearity scores and ranges of pressure magnitude perturbation caused by ESHDs with (u, v) pairs $(15, 2)$ and $(15, 15)$ and with amplitude ranging ± 5 mm. The non-linearity score, calculated as described in Equation 5.2.2.2, is shown for different frequencies, for front and left sound sources. It is, generally, far greater than in the case of lower u values (see Table 5.2). The pressure magnitude perturbation range generated by ESHDs is similar to those with low u values for low frequencies but rises by only two orders of magnitude as frequency increases from 200 Hz to 20 kHz. It never exceeds 10^{-04}, more than twenty times less than that observed for ESHDs with $(u, v) = (2, 15)$ at 10 kHz.

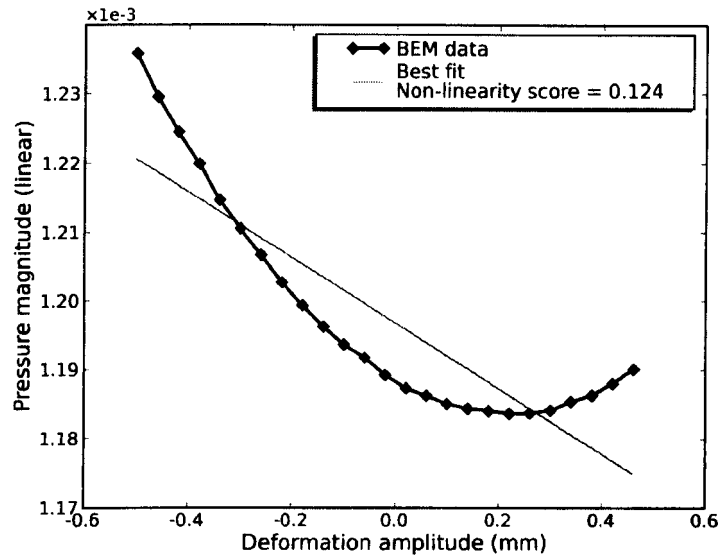


(a)

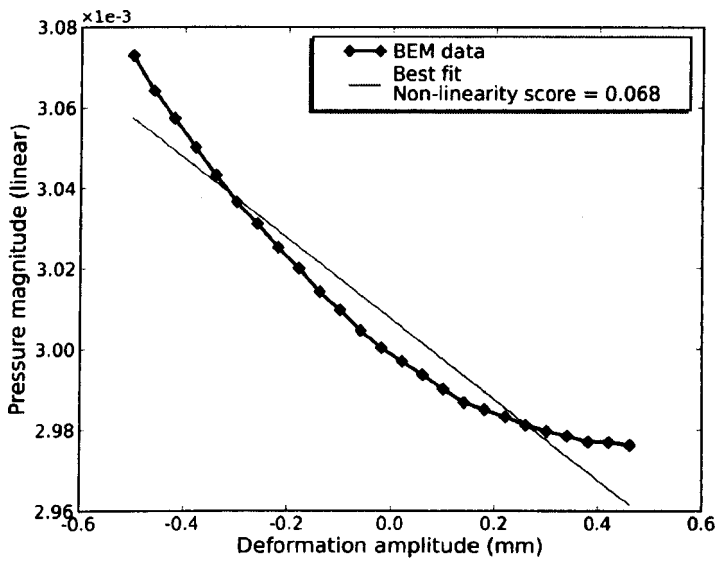


(b)

Figure 5.6: Deformation peak amplitude vs pressure plot for the $\cos 2\pi(15x/S + 2y/P)$ ESHD at 10 kHz, for the front (a) and left (b). The linear approximation breaks down for high cross-harmonic (u) deformations, instead a roughly quadratic relationship between deformation amplitude and acoustic pressure appears. For the left position the quadratic behaviour is less marked and the pressure minimum is slightly shifted towards negative deformation amplitudes.



(a)



(b)

Figure 5.7: Deformation peak amplitude vs pressure plot for the $\cos 2\pi(15x/S + 15y/P)$ ESHD at 10 kHz, for the front (a) and left (b). The quadratic relationship appears, again, although it is not as pronounced as in the case of the $\cos 2\pi(15x/S + 2y/P)$ ESHD (see Figure 5.6), particularly for the left position. The pressure minimum is, in both cases, shifted towards positive deformation amplitudes.

a deformation of opposite amplitude can be obtained through a π/u rad rotation about the slicing axis. As the cross-slice harmonic increases, the rotation becomes smaller, eventually leading to a similar effect in the case of positive and negative deformation amplitudes.

This reasoning can also be applied in the case of the KEMAR head, although the relation isn't as clear. As Figure 5.3(d) shows, the higher cross-slice harmonics generate ripples which are of considerably shorter wavelength than most of the features within the pinna. If the half-wavelength of the ripple is shorter than the features of the pinna which vary under rotation about the slicing axis, the pressure changes due to positive and negative deformations (one shifted by half a wavelength in comparison to the other) will be very similar. This provides some explanation for the approximately even appearance of observed pressure changes. Offsets in the value of deformation for which the pressure change exhibits even symmetry (see Figures 5.6(b) and C.6(a) for example) are likely due to the fact that pinna morphology is not stationary under radial rotation.

Overall, the ESHDs which impose over-riding constraints on DPS validity are the ones with high cross-harmonic (u) values, which create highly non-linear effects, introduced even for small (0.1 - 0.5 mm) deformations, especially at higher frequencies. The decision to use ESHDs with amplitude 0.3 mm for the creation of the DPS database was made because of computation noise present in the previous version of the BEM-based acoustic simulation software, which was judged unacceptable (see Section 5.2.2.1). Noise levels were substantially reduced prior to the linearity tests presented in this section, by software upgrades. Taking into account the non-linear effects which have been described, the DPS database would ideally be gener-

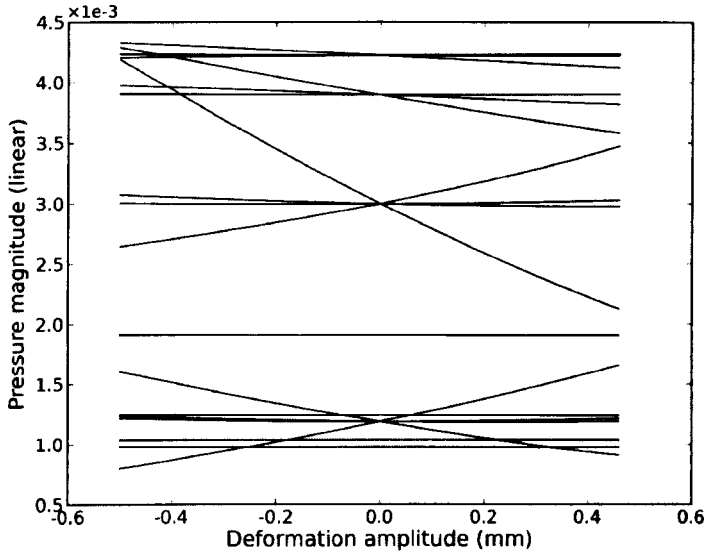


Figure 5.8: All linearity plots in this section and Appendix C, plotted on the same pressure magnitude scale. The non-linear acoustic effects observed, in particular for high cross-harmonic elliptical surface harmonic deformations, are small in comparison to others (see Figure 5.4 and 5.5, for example).

ated using an ESHD amplitude which reveals very little non-linear behaviour (no bigger than 0.1 mm). A database re-build was, however, impossible due to time constraints. The effects of non-linear acoustic effects on DPS estimation are investigated in Section 5.3. It is important to note that the acoustic effects of high cross-harmonic (u) elliptical surface harmonic deformations, although they are non-linear, are relatively small in comparison to those observed for lower cross-harmonics. Indeed, when plotting all the graphs in this section and the supplementary graphs in Appendix C on the same magnitude scale, non-linear effects appear relatively subtle (see Figure 5.8). This observation serves to reduce the DPS inaccuracies expected to result from non-linearities.

5.2.3 Mesh deformation, topology and symmetry

Once orthogonal deformations were applied to a slice set, changes in slice contour coordinates were transferred to the corresponding points in the simulation mesh. The original mesh topology was preserved in each case to avoid introducing undesirable pressure variations in addition to the effects of the deformation. In Chapter 3, the fact that the KEMAR head was perfectly symmetrical about the median plane allowed improved simulation performance. Indeed, BEM theory permits this property to be exploited to allow significant savings in computation time (see Section 3.1). After the application of orthogonal deformations, however, the resulting shape is generally not symmetrical. The benefits of symmetry can, nevertheless, be exploited by assuming, as was suggested in Section 4.2.1.2, that the HRTF for a given ear can be calculated accurately by approximating the opposite side of the head to a perfect symmetrical reflection of the side on which the ear lies about the median plane. In an effort to reduce simulation time, the acoustic effect of orthogonal deformations is calculated based on this assumption. To restore symmetry, the right side of the head is, in effect, copied and reflected about the median plane. In a narrow region, near the median plane, the deformation is smoothly reduced to zero using a cosine windowing function. This precautionary step removes potential discontinuities in the slope of the surface at this boundary.

5.2.4 Simulation frequencies

A BEM-based acoustic simulation only allows the acoustic pressures caused by a sound source to be calculated at a single frequency. For each orthogonal deformation contained in the DPS database, simulations must be carried

out for a relatively large number of frequencies in order for associated spectral changes to be adequately described. Because of time constraints, the frequencies contained in the DPS database had to be selected carefully. The choice of frequencies was made so that the resulting data would allow impulse responses at a sampling frequency of 32 kHz and of a length of 4.5 ms to be reconstructed. To achieve this goal, the required frequency spacing is of 222.2 Hz. Frequencies spreading over the principal range of pinna spectral cue activity were prioritised. Studies described in Section 2.2.2.2 provide estimates for this range and based on these, 21 frequencies linearly spaced between 7778 Hz and 12222 Hz were prioritised. Time constraints prevented further frequency-domain expansion of the database, which will be pursued in further work. The remainder of this chapter will describe DPS performance in this frequency range.

5.2.5 Database dimensions

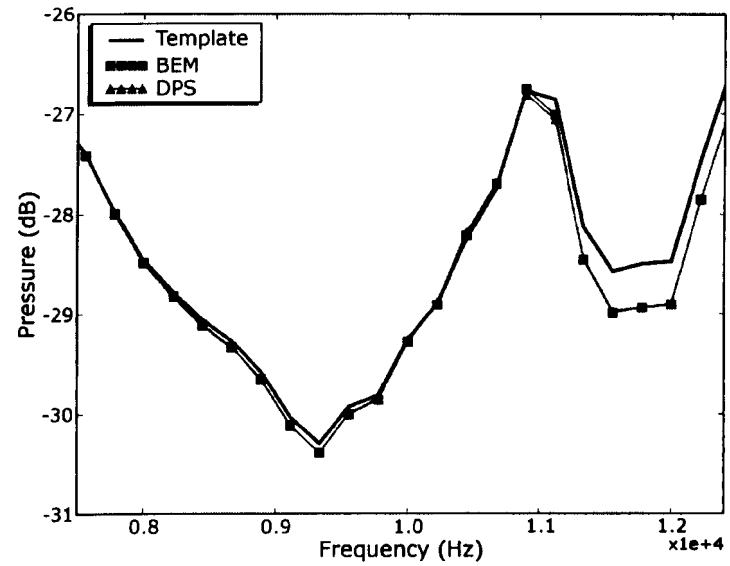
The maximum cross-harmonic (u) and slice-harmonic (v) of the ESHDs (see Section 5.1.1) for which the acoustic effects are pre-computed determines the maximum spatial resolution of possible template deformations resulting from a weighted summation. Constraints on time and computing power again limited this range. Results reported in this thesis are calculated using a DPS database describing the acoustic effect of ESHDs for all (u, v) pairs obeying $0 \leq u < 20$ and $0 \leq v < 20$. For each (u, v) pair, the acoustic effects of both $\cos(2\pi(xu/S + yv/P))$ and $\sin(2\pi(xu/S + yv/P))$ deformations (see Equation 5.11) were computed. The current DPS database therefore describes the acoustic effect of $20 \times 20 \times 2 = 800$ ESHDs to the template, for each of the 21 frequencies employed (see Section 5.2.4)

5.3 Accuracy of DPS estimation

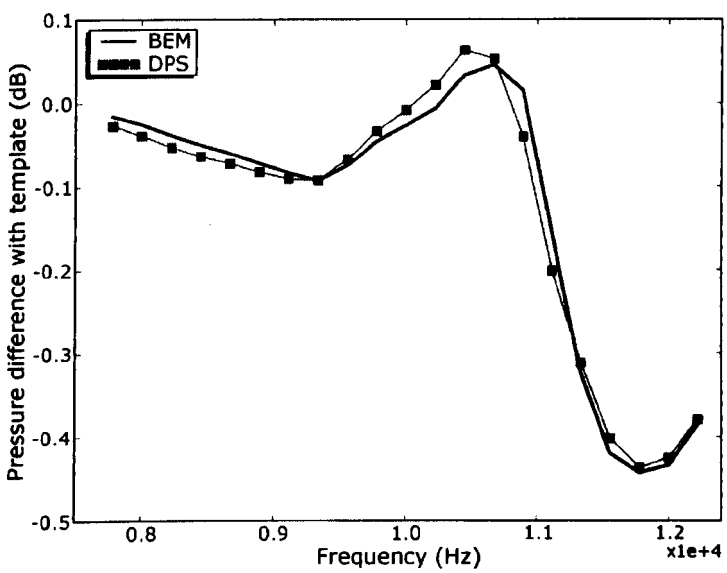
The validity of DPS estimations was investigated over the frequency range for which the database has been compiled (7778 - 12222 Hz). DPS estimation is based on the superposition of acoustic effects catalogued in the database, as described in Section 2.5. The quality of the estimation was assessed by comparing the actual acoustic effects of a summation of ESHDs, calculated directly using BEM simulations, with the corresponding DPS estimate. The more linear the relationships between deformation amplitudes and resulting pressure changes are, the better the DPS estimate is expected to be. The DPS estimate error for a given frequency point can be quantified as a percentage of the magnitude range of spectral change caused by the weighted deformation summation over the estimation range. This error measure can be expressed as

$$100 \frac{e(f) - \Delta p(f)}{\Delta p_{max} - \Delta p_{min}} \tag{5.20}$$

where $e(f)$ is the DPS estimate of the pressure magnitude change for frequency f, $\Delta p(f)$ is the actual pressure magnitude change computed through BEM, and $p_{max} - p_{min}$ is the range of pressure change caused by the total deformation over the DPS database frequency range, with all values in dBs. This is a useful measure, given that the range of pressure magnitude change varies significantly, as will be shown later. The mean and maximum DPS error percentages over all database frequencies will be given for the quality assessment of each DPS estimation. Performance was assessed for several different (u, v) ranges and two different ESHD weighting schemes. In the first scheme, ESHDs were scaled equally, so that the maximum total deformation

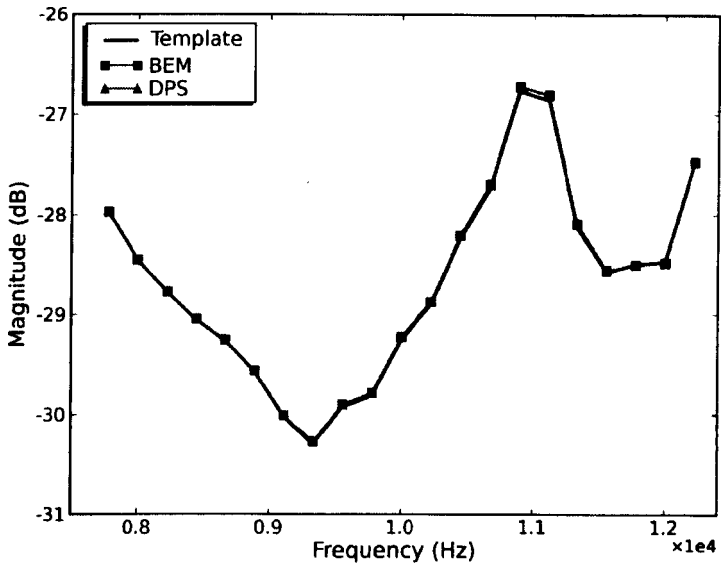


(a)

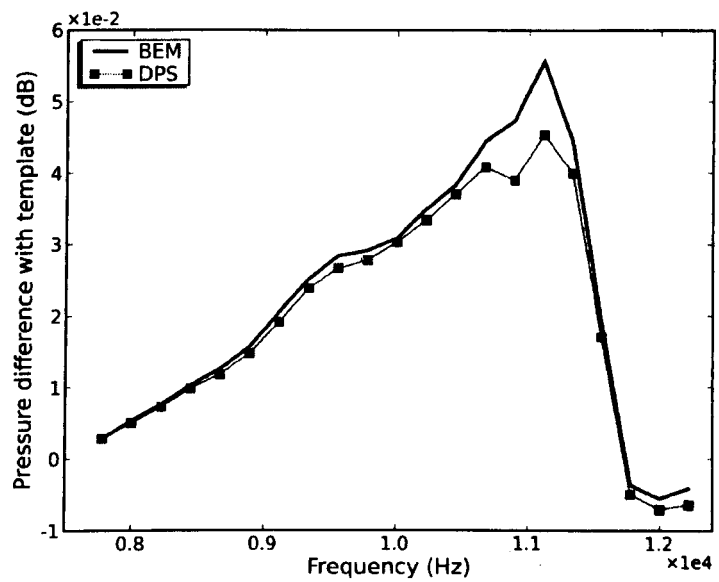


(b)

Figure 5.9: (a) The pressure at the entrance of the right ear canal of the template head generated by a frontal source is plotted along with pressures generated by a summation of all ESHDs obeying $0 \leq u < 5$ and $0 \leq v < 5$, as calculated using BEM simulation and DPS estimation. ESHD peak amplitude is 0.006 mm so that the total deformation does not exceed 0.3 mm. (b) The pressure change generated by the deformation, calculated using BEM and DPS.



(a)



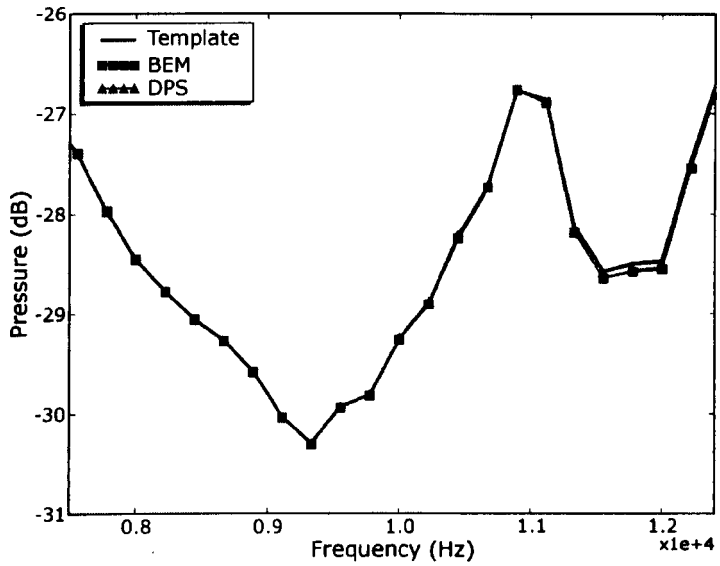
(b)

Figure 5.10: (a) The pressure at the entrance of the right ear canal of the template head generated by a frontal source is plotted along with pressures generated by a summation of all ESHDs obeying $0 \leq u < 5$ and $0 \leq v < 5$, as calculated using BEM simulation and DPS estimation. ESHD peak amplitude is attributed randomly, with a uniform probability distribution over the ± 0.006 mm amplitude range. (b) The pressure change generated by the deformation, calculated using BEM and DPS.

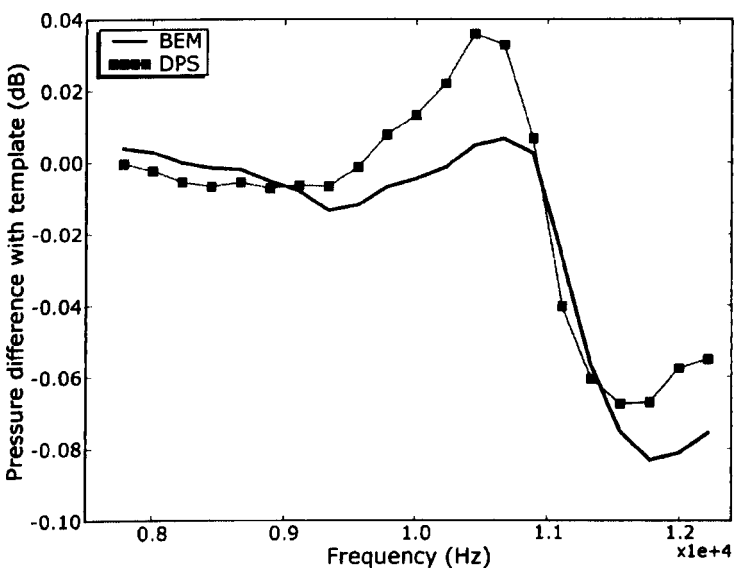
amplitude was no higher than that of the individual ESHDs used to compile the DPS database (0.3 mm). That is, each individual ESHD was given a peak amplitude of $0.3/N$ mm where N is the number of summed ESHDs. In the second scheme, deformations are attributed random peak amplitudes, with uniform probability distribution over the $\pm 0.3/N$ mm range.

Figure 5.9 shows DPS performance for a summation of all ESHDs obeying $0 \leq u < 5$ and $0 \leq v < 5$, which amounts to $5 \times 5 \times 2 = 50$ deformations, with equal weights. Figure 5.9(b) shows the pressure change caused by deformation of the template, as calculated using direct BEM simulation and by DPS estimation. The mean DPS percentage error is 3.29 % and the maximum DPS percentage error is 11.24 % (with DPS percentage error as defined in Equation 5.20). A similar performance is observed when the peak amplitude of each ESHD is attributed randomly as shown in Figure 5.10. Mean and maximum DPS percentage errors are shown in Table 5.3. The amount of pressure change generated is far smaller in the case of randomly attributed ESHD weights (around 10 times smaller). This is to be expected as the average deformation weight is far higher in the case of the equal weighting scheme than the random one.

Extending the harmonic range of deformations, Figure 5.11 shows DPS performance for a summation over all ESHDs obeying $0 \leq u < 20$ and $0 \leq v < 20$, a total of $20 \times 20 \times 2 = 800$ deformations, with equal peak magnitudes. The increased range of summed harmonics reduces the associated acoustic effect. This is due to the scaling method adopted to ensure the overall deformation stays bound to within 0.3 mm of the template. Indeed, this causes the average deformation from the template to reduce, with relatively large deformation observed over smaller, more localised areas. The

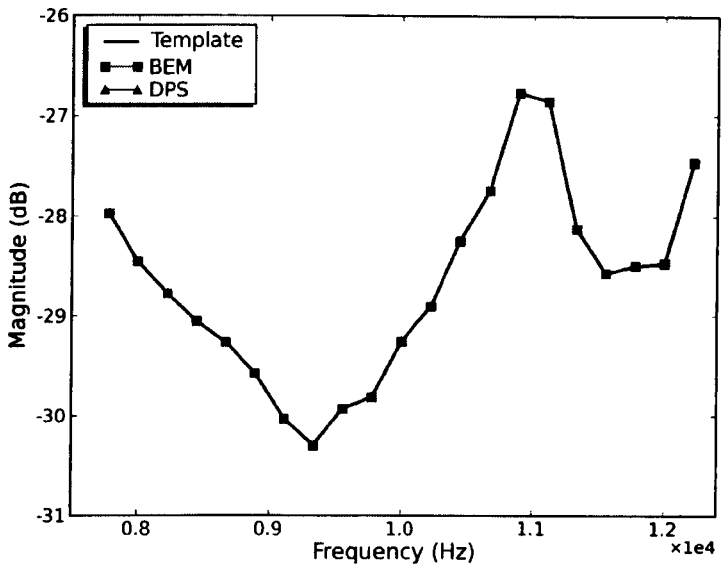


(a)

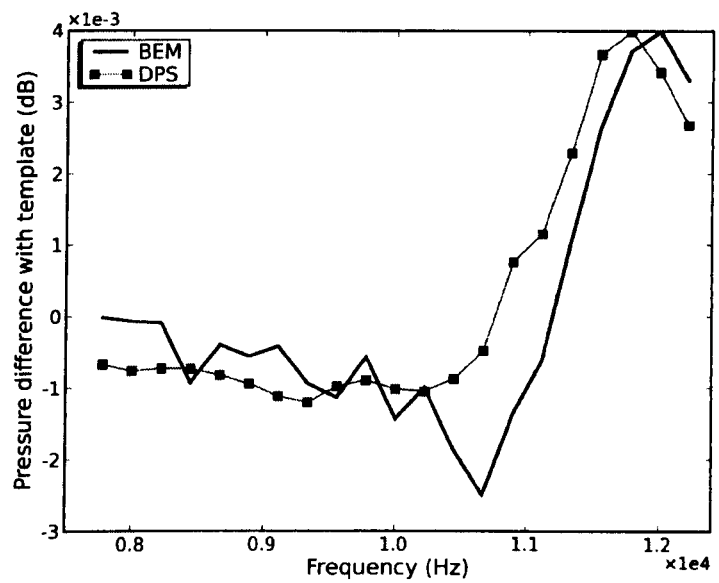


(b)

Figure 5.11: (a) The pressure at the entrance of the left ear canal of the template head generated by a frontal source is plotted along with pressures generated by a summation of all ESHDs obeying $0 \leq u < 20$ and $0 \leq v < 20$, as calculated using BEM simulation and DPS estimation. ESHD peak amplitude is 3.75×10^{-4} mm so that the total deformation does not exceed 0.3 mm. (b) The pressure change generated by the deformation, calculated using BEM and DPS.



(a)



(b)

Figure 5.12: (a) The pressure at the entrance of the left ear canal of the template head generated by a frontal source is plotted along with pressures generated by a summation of all ESHDs obeying $0 \leq u < 20$ and $0 \leq v < 20$, as calculated using BEM simulation and DPS estimation. ESHD peak amplitude is attributed randomly, with a uniform probability distribution over the $\pm 3.75 \times 10^{-4}$ mm amplitude range. (b) The pressure change generated by the deformation, calculated using BEM and DPS.

acoustic effect of the equally weighted deformation summation is similar to that observed for $(u, v) = (5, 5)$ (see Figure 5.9(a)). In both cases, ESHD within (u, v) range (5,5) are added in the same proportions, which results in similar pressure changes. Indeed, these ESHDs have been shown to have a larger effect on acoustic pressure than those with higher u values, for the particular frequencies chosen in Section 5.2.2.2 (see Tables 5.1 and 5.2).

The difference between direct BEM calculation and DPS estimate shown in Figure 5.11(b) is significantly more pronounced than in the previous case (see Figure 5.9(b)). The shape of the actual computed effect is reproduced to some extent, however, proportionally to the deviation from the template, the estimation performance has deteriorated. The enhanced discrepancy between direct BEM calculations and DPS estimates results from the non-linear behaviour described in Section 5.2.2.2. The mean DPS estimate percentage error (percentage error is as defined in Equation 5.20), is 13.12 % with a maximum percentage error of 34.63 % around 10.5 kHz. This demonstrates a marked increase compared to the case were the (u, v) range of ESHDs was constrained to (5,5).

When applying the random weighting scheme, DPS estimation performance is again very similar to those obtained with the equal weighting scheme (see Figure 5.12). The mean DPS estimate percentage error, is 11.42 % with a maximum percentage error of 33.11 % around 10.5 kHz; in fact, a slight improvement on the equal deformation amplitude case. Again the range of spectral variation generated by the deformation summation is smaller in the case of random weighting by an order of magnitude, approximately. DPS performance for intermediate ESHD (u, v) ranges can be found in Table 5.3, with plots in Appendix C (Figures C.9, C.10, C.11 and C.12).

(u, v) range	Weighting scheme	Mean % error	Maximum % error
(5,5)	Equal	3.29	11.24
(5,5)	Random	3.58	16.73
(10,10)	Equal	7.66	17.48
(10,10)	Random	2.15	9.47
(15,15)	Equal	9.53	24.56
(15,15)	Random	9.35	31.1
(20,20)	Equal	13.12	34.63
(20,20)	Random	11.42	33.11

Table 5.3: Mean and maximum DPS estimation percentage error (as defined in Equation 5.20) for different (u, v) ranges and weighting schemes.

In the case of the equal weighting scheme, DPS estimation errors increase steadily with the (u, v) range of the ESHD summation, which is expected as the non-linearities become more marked for higher u values (see Section 5.2.2.2).

This steady increase in DPS estimation error does not occur in the case of the random weighting scheme. Indeed, performance is significantly better for a (u, v) range of (10,10) than for a (u, v) range of (5,5). This is unexpected but can be explained. Although non-linear acoustic effects for ESHDs with $u = 2$ were shown to be limited for test frequencies, in Section 5.2.2.2, they are likely to be significant, especially with growing u values. The random character of weight distribution could therefore have, by chance, revealed more non-linear behaviour in the case where the (u, v) range is restricted to (5,5), by giving more weight to EHSDs which produce higher levels of non-linearity. For the broader (u, v) ranges (15,15) and (20,20), however, performance falls markedly, as expected from the highly non-linear effects described for high u values in Section 5.2.2.2.

5.4 Applying DPS to morphoacoustic perturbation analysis

The linearity checks described in Section 5.2.2.2 prove conclusively that DPS is unsuitable for estimating the acoustic effect of deforming the template KEMAR head to fit any individual morphology, at least in the form described in this chapter. Specifically, deforming the template KEMAR head into that of any given individual would require deformation amplitudes extending far beyond the sub-millimetre range for which the resulting pressure change can be considered linear. Indeed, the ESHD amplitude of 0.3 mm used to create the DPS database generated, in some instances, unacceptably non-linear acoustic effects which result in DPS estimation errors (see Section 5.3).

However, although HRTF estimation across individuals is impossible, DPS is a potentially powerful tool for performing morphoacoustic perturbation analyses. Indeed, the relationship between infinitesimal template deformations and acoustic pressure changes provides an elegant means for probing the morphological regions responsible for the production of spectral features in HRTFs. The probing process is referred to as morphoacoustic perturbation analysis (MPA). The technique will be described in detail in this section and a test case will be used to demonstrate some of its potential.

5.4.1 Motivations

Many studies have attempted to extract mappings between sets of HRTFs and corresponding sets of landmark-based morphological measurements (see Section 2.4.1.1). The objective of such studies is to bypass the complexities

of obtaining HRTFs through acoustic measurement or of modelling complex pinna acoustics mathematically. Their overall aim is HRTF estimation from easily obtainable shape descriptors (see Section 2.4.4). However, in all cases, the pairs of datasets were found to be only weakly correlated. Enhancing the description of morphological areas involved in the localisation cue production, while discarding irrelevant shape data is crucial if correlation is to be significantly improved.

The parameterised complete shape description of the head and pinnae proposed in Chapter 4 offers the advantage of intrinsically eliminating the scope for omitting relevant shape detail. The main drawback, however, is the large amount of irrelevant information contained in the parameters. Superfluous information risks swamping relevant data with a resulting blurring of statistical mappings between localisation cues and their morphological origin. Additionally, their acquisition requires a complete shape capture, which is likely to be expensive and impractical. A method for identifying the morphological regions which are salient to HRTF-feature production is required so that shape capture and description can focus on them, while avoiding irrelevant information.

A number of studies have demonstrated the importance of the pinna for the production of localisation cues by altering its shape (filling some of its cavities) and observing a loss in elevation perception acuity and an increase in front-back confusions (Gardner and Gardner, 1974; Hofman and Opstal, 2003; Humanski and Butler, 1988; Morimoto *et al.*, 2001; Musicant and Butler, 1984, see Section 2.2.2.1). This approach can be generalised by analysing the impact of a small local deformation on an acoustic feature. Indeed, this impact can be taken as a measure of the relevance of the deformed area

to the production of the acoustic feature. This approach is potentially insightful, but an exhaustive study along these lines is impractical. Similar methods based on acoustic simulation are not viable due to the huge search space involved. Acoustic simulations, on the other hand, do allow complete control over the applied deformations and the essentially ideal transducers and acoustic noise levels provide distinct advantages over acoustic measurements.

5.4.2 MPA principles

Studying the effects of small local deformations on acoustic localisation cues seems an intuitive approach, if an exhaustive exploration of the search space is realistic. However, acoustic simulations are computationally demanding and the sheer number of simulations required to cover a sufficient range of deformation sizes, locations, magnitudes and shapes, as well as investigating their combined effects makes the process prohibitively lengthy. DPS offers an elegant solution to this problem.

ESHDs are derived from the basis functions for the 2-D Fourier transform, which are orthogonal. Consequently, a micro-deformation of practically arbitrary shape can be applied to the template head and ears using a weighted sum of ESHDs. DPS provides a rapid means of estimating acoustic effect of such template micro-deformations without the need for further BEM simulations, assuming a perfectly linear relationship between the amplitude of each ESHD and its associated pressure change (see Section 2.5). MPA employs the DPS database in the reverse direction. Orthogonal shape deformations are weighted according to their effect on a given target spectral feature, and in so doing can identify the morphological origin of this feature.

Indeed, the weighted ESHDs reinforce, according to DPS principles, to create a strong deformation in morphological regions which affect the feature of interest and tend to cancel out elsewhere. Local deformation magnitude can then be taken as a measure of relevance to the cue production mechanism. The level of relevance to the production of a chosen target feature for a given morphological area will often be referred to as temperature (with high temperature equating to high relevence).

The key step in MPA is that every ESHD is assigned a score according to how strongly it contributes to the required acoustic changes in the frequency range of the feature. The procedure through which this score is established is not set and depends on the required result. The rate of change of the HRTF magnitude spectrum (defined over all computed frequencies) with respect to the amplitude of a given ESHD (denoted by $D^{\sigma}_{u,v}(f)$) is expressed as

$$D^{\sigma}_{u,v}(f) = \frac{\partial p(f)}{\partial E^{\sigma}_{u,v}} \tag{5.21}$$

where $E^{\sigma}_{u,v}$ is the deformation amplitude as described in Section 5.1.1. This is an important function in any ESHD weighting scheme. If the objective, for example, is maximum perturbation to the target feature, the score (or weight) of the corresponding ESHD, denoted $w^{\sigma}_{u,v}$, would be $D^{\sigma}_{u,v}(f)$ integrated over the (f_{min}, f_{max}) range. In discreet notation

$$w^{\sigma}_{u,v} = \sum_{n\,=\,n_{min}}^{n_{max}} D^{\sigma}_{u,v}[n] \tag{5.22}$$

where n_{min} and n_{max} are the frequency domain samples corresponding

to f_{min} and f_{max}. Once again, this weighting scheme aims to define a weighted ESHD summation which reinforces the overall spectral change in the (f_{min}, f_{max}) frequency range as much as possible. Alternative measures can be developed depending on the desired result. If the objective, for example, is to shift the center frequency of a spectral notch, ESHD score could be defined so as to favour an increase spectral energy in the falling edge and decrease spectral energy in the ascending edge of the notch.

5.5 Demonstration of MPA

The DPS database currently describes the pressure changes associated with 800 ESHDs (see Section 5.2.5). In each case the deformation amplitude is 0.3 mm. This amplitude was shown in Section 5.2.2.2 to be, in some cases, beyond the region where acoustic effects can be considered to change linearly. Ideally, the DPS database would be reconstructed, using lower ESHD amplitudes in order to reduce non-linearities. This was, however, impossible due to time restrictions. Nevertheless, the general validity of the database was established in Section 5.3 and the database was therefore considered suitable for demonstrating the principle of MPA.

5.5.1 Target spectral feature

In order to illustrate the potential MPA, a target spectral feature was selected. The chosen feature is the secondary spectral notch observed around 11.5 kHz for a source located in front ($\theta = 0$, $\varphi = 0$) of the KEMAR head (see Figure 5.13). The impact of each orthogonal deformation on the target feature is calculated using Equation 5.22, with $f_{min} = 11.11$ kHz and

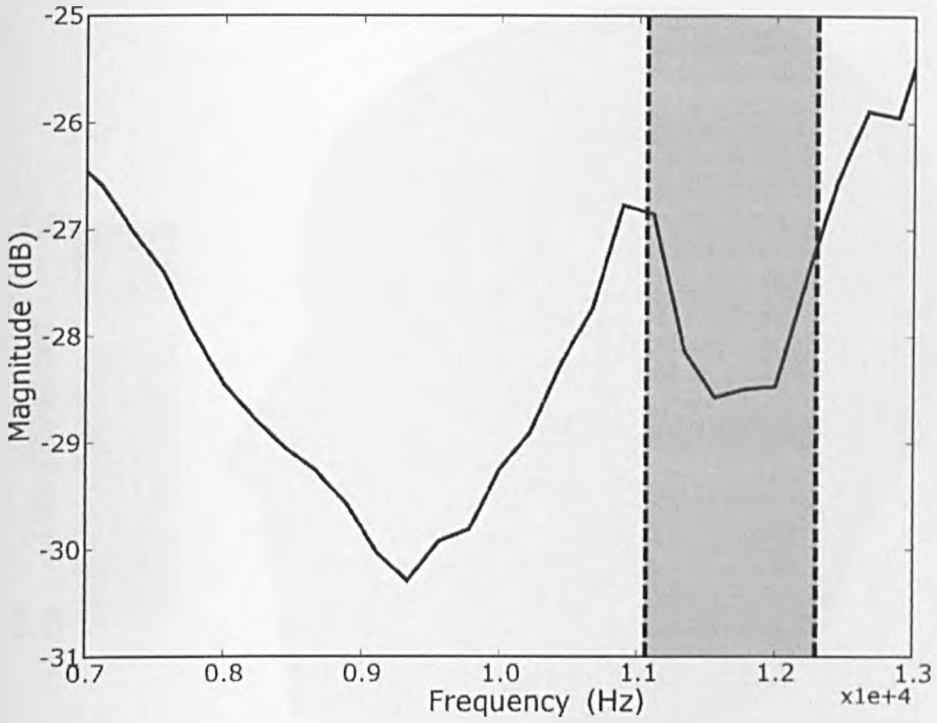


Figure 5.13: Highlighted target spectral feature on the frontal HRTF within the 11.11-12.22 kHz frequency range.

$f_{max} = 12.22$ kHz.

5.5.2 MPA results

Figures 5.14 and 5.15 show the temperature map produced using MPA for the target feature described in Section 5.5.1. This temperature is proportional to the local deformation generated by the sum of ESHDs, each weighted according to the score defined in Equation 5.22. It is taken as a measure of relevance to the production of the target feature. Areas of high temperature (highly involved in the production of the target notch) are concentrated in the inner pinna region. The temperature over the remainder of the head surface is constrained to the lower quarter of the scale, which

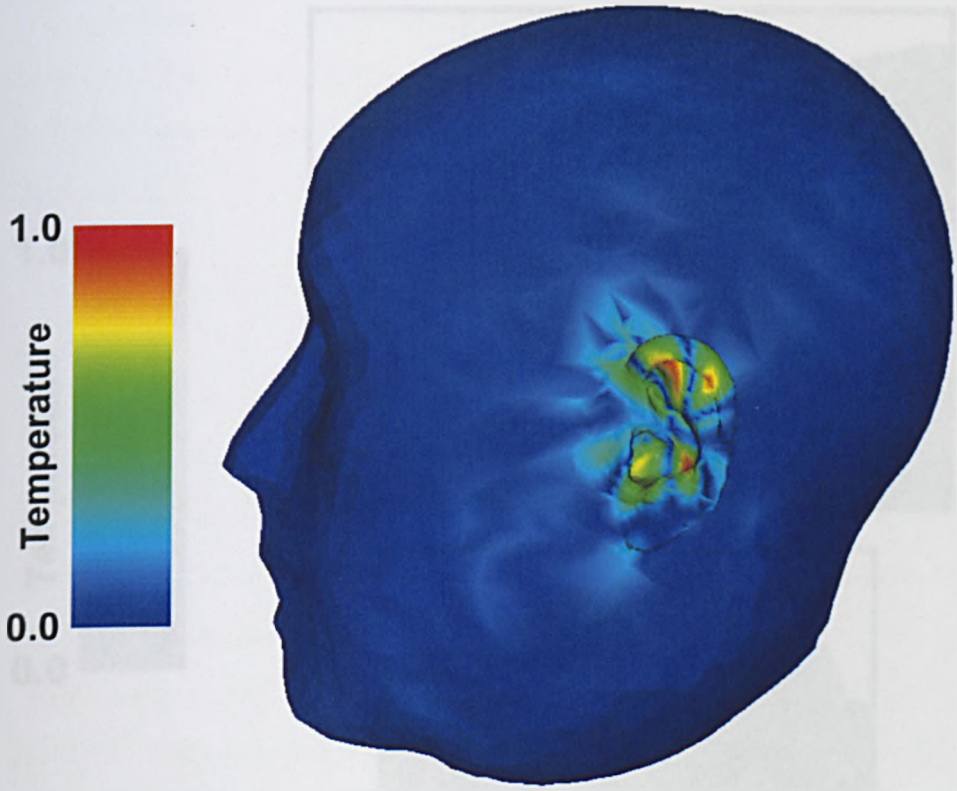


Figure 5.14: Head temperature map (arbitrary units). The temperature gives a measure of participation to the generation of the target feature (see Figure 5.13). Hot areas are concentrated within the pinna while temperature over the remainder of the head is constrained to the lower quarter of the scale.

signifies very little or no contribution to the production of the target notch. This is to be expected, as the head is known to have little involvement in the production of the target feature in comparison to the pinnae.

Figure 5.15 focuses on the temperature distribution within the pinna. MPA has picked out three distinct high-temperature areas. One in the cavum concha, the other two within the helix fold, spilling into the antihelix. This distribution is reminiscent of the surface pressure generated by a point source close to the ear canal around the 11.1 kHz resonance (see Figure

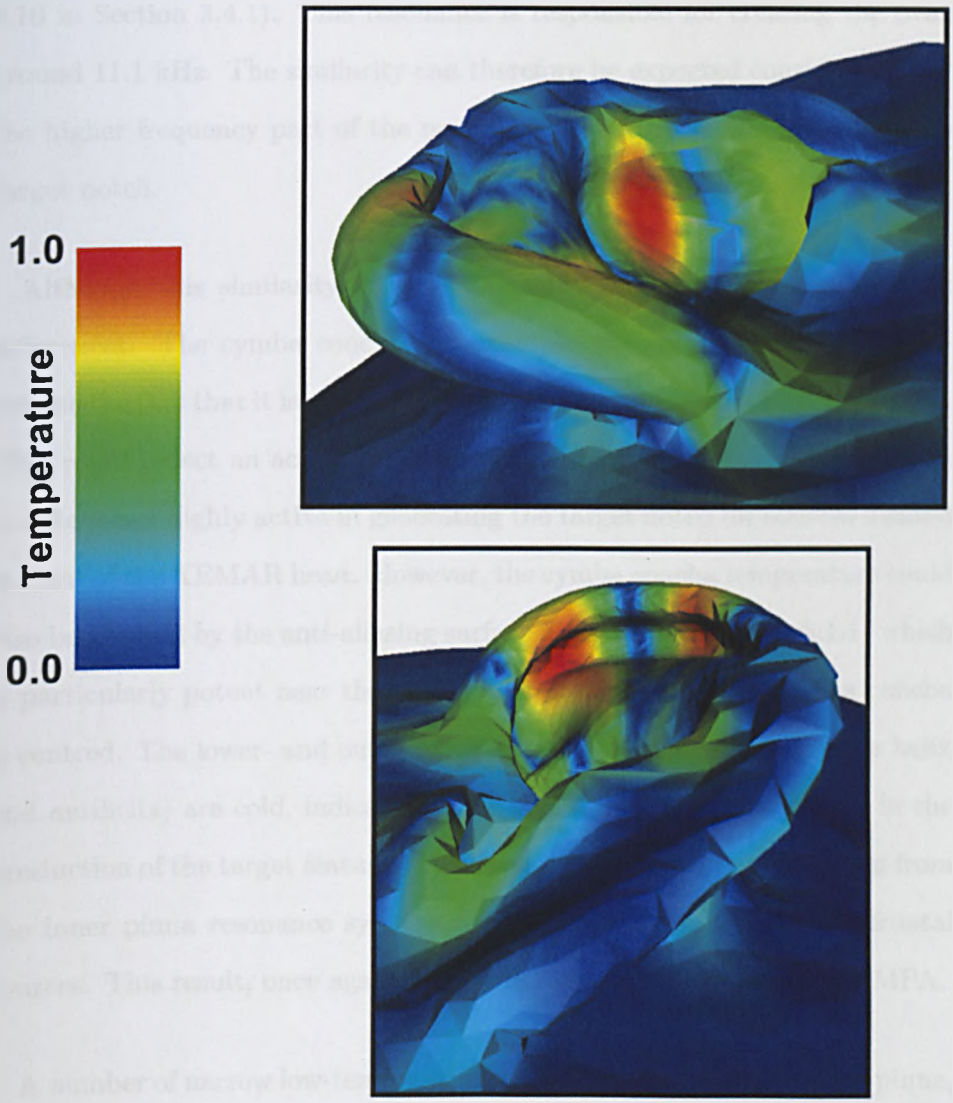


Figure 5.15: Pinna temperature map (arbitrary units). The temperature gives a measure of participation to the generation of the target feature (see Figure 5.13). For clarity, two views are given to avoid areas being left invisible because of occlusion. Three areas, in the cavum concha and the helix fold, are indicated as particularly relevant to the production of the target spectral feature (shown as high-temperature, red areas). These areas are interleaved with narrow regions of low-temperature, which signifies little or no identified contribution.

3.10 in Section 3.4.1). This resonance is responsible for creating the peak around 11.1 kHz. The similarity can therefore be expected considering that the higher frequency part of the peak overlaps with the falling edge of the target notch.

Although this similarity is very noticeable there are also a number of differences. The cymba concha, for example, shows medium temperature despite the fact that it is a major actor in the creation of the pinna resonance. This could reflect an actual tendency which would suggest that the cymba concha is not highly active in generating the target notch for sources located in front of the KEMAR head. However, the cymba concha temperature could also be affected by the anti-aliasing surface filtering (see Section 5.1.1) which is particularly potent near the slicing axis around which the cymba concha is centred. The lower- and outer-pinna areas (lobule, antitragus, lower helix and antihelix) are cold, indicating little or no identified involvement in the production of the target feature. This is compatible with their isolation from the inner pinna resonance systems and known reflection paths for frontal sources. This result, once again, supports the validity of DPS-based MPA.

A number of narrow low-temperature regions also appear within the pinna, radiating from the slicing axis. This striped pattern is likely to be, at least partly, generated by artefacts resulting from the harmonic nature of the ESHDs. Indeed, it appears to be a manifestation of the Gibb's phenomenon, due to an abrupt ESHD series truncation. The highly non-linear relationship between deformation amplitude and acoustic pressure observed for high order ESHDs (see Section 5.2.2.2) is also likely to have played a role in generating this effect.

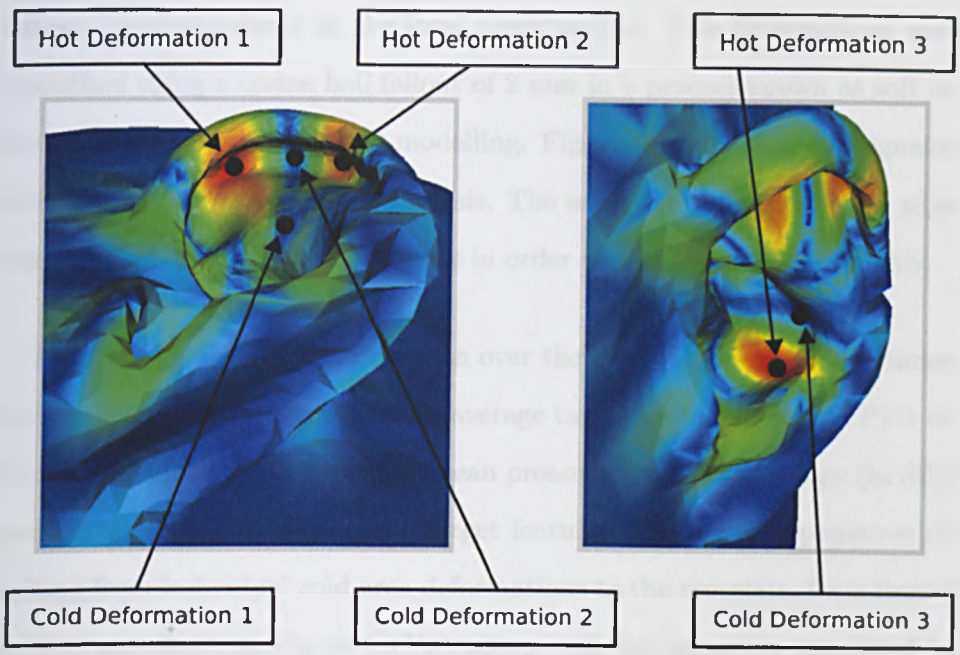


Figure 5.16: Location of the low- and high-temperature area deformations. In each case, the deformation is applied to the (single) vertex closest to the center of the dot, perpendicularly to the surface of the mesh. Soft selection is applied with a fallout range of 2 mm creating a smooth "bump" instead of the sharp notch, which would otherwise be obtained.

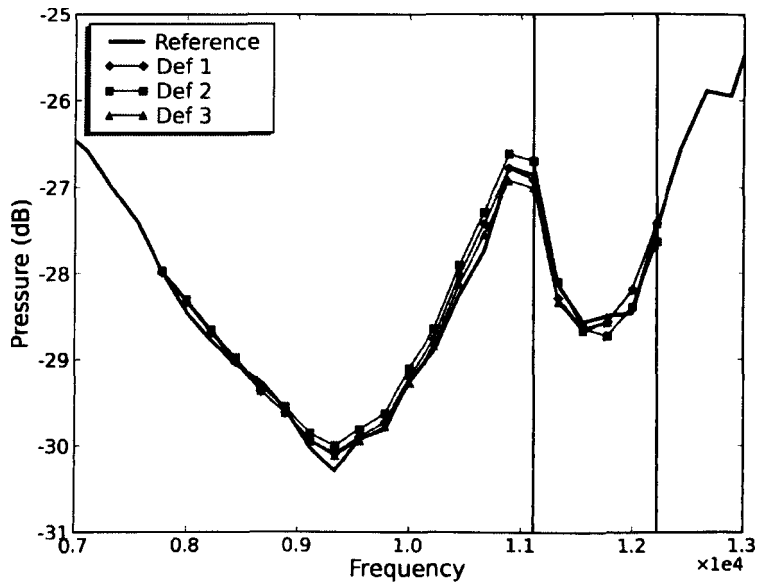
5.5.3 MPA validation

In order to assess the validity of the temperature map produced using MPA on the target notch (see Sections 5.5.1 and 5.5.2), local deformations are applied to areas identified as having high involvement and low involvement to the generation of the target feature (described for brievety as hot and cold, respectively). The effect of these local deformations was then computed using the BEM. It was anticipated that deformations in cold areas would have little effect on the notch, while deformations in hot areas would have a larger effect. Each deformation was performed by applying a 1 mm translation to the mesh vertex closest to the required deformation lo-

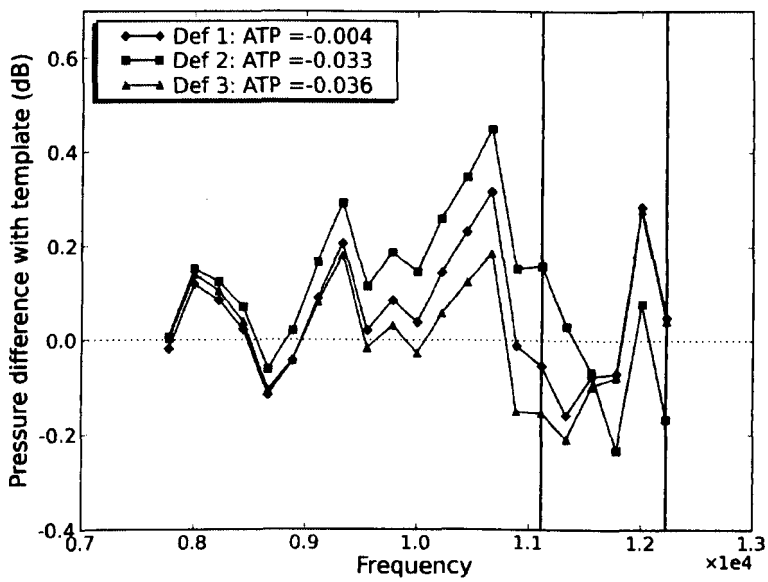
cation, perpendicularly to the local mesh surface. The deformations were smoothed using a cosine bell fallout of 2 mm in a process known as soft selection, commonly used in 3D modelling. Figure 5.16 shows the designated cold and hot spots chosen for analysis. The selection process targeted adjacent opposite temperature extremes in order to test the map thoroughly.

The net effect of each deformation over the frequency range of the target feature is expressed in terms of the average target perturbation (ATP) it inflicts. This measure is simply the mean pressure magnitude change (in dBs) over all frequencies covering the target feature. The acoustic pressures resulting from individual cold area deformations to the template, for a frontal source, are shown in Figure 5.17(a), along with the pressures generated by the undeformed reference. Figure 5.17(b) shows the change in pressure associated with each of the applied deformations. The target feature is again the notch in the 11.1-12.2 kHz range (see Section 5.5.1). Deformations in the cold areas cause some variation in pressure over the target range, which is an unwanted effect. The magnitude of the average target perturbation (see ATP in Figure 5.17(b) legend), however, is very low. Indeed, it never exceeds 0.04 dBs in magnitude.

The effects of individual hot area deformations on the frontal HRTF are shown, along with the reference, in Figure 5.18(a). Figure 5.18(b) shows the difference with the reference. These deformations consistently cause higher levels of variations than those observed for cold area deformations. Most of the variation in the target area exceeds 0.2 dB, sometimes nearing the 0.4 dB mark. The magnitude of the average target perturbation is, in all cases, far higher than that resulting from cold area deformations. Indeed, it exceeds 0.14 dBs in all cases (see ATP in Figure 5.18(b) legend). This result gives

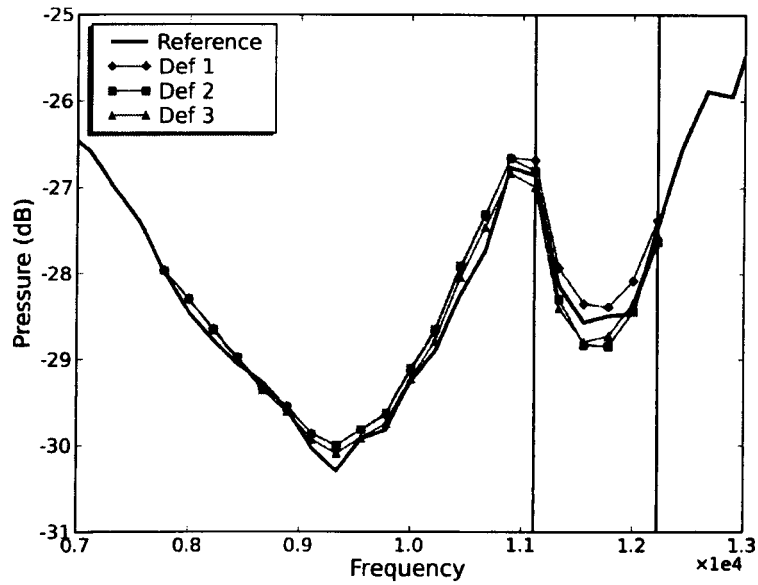


(a)

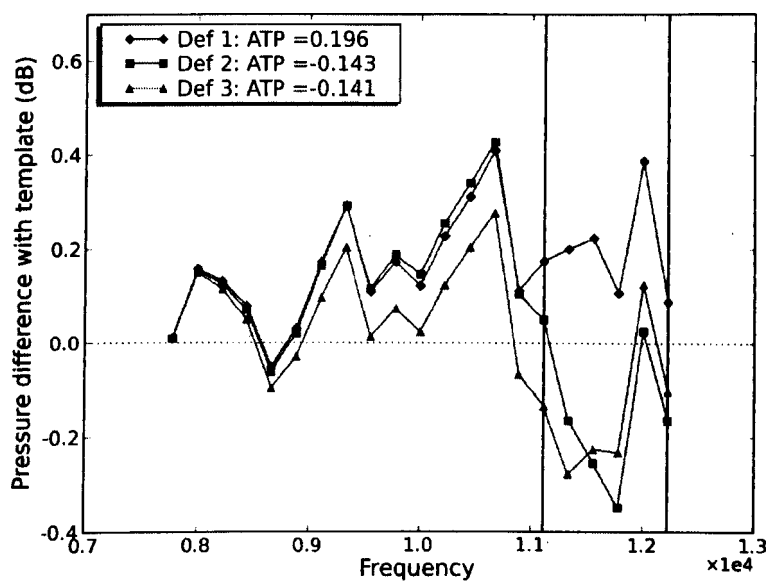


(b)

Figure 5.17: Acoustic pressures generated by deformations in the cold areas (described in Figure 5.16) on the frontal HRTF, alongside template pressures. The pressure changes generated by the deformations are shown in (b). Some variation appears within the target notch frequency range (11.1-12.2 kHz). The average target perturbation, shown in the legend, is very low. Indeed, in all cases, ATP magnitude does not exceed 0.04 dBs

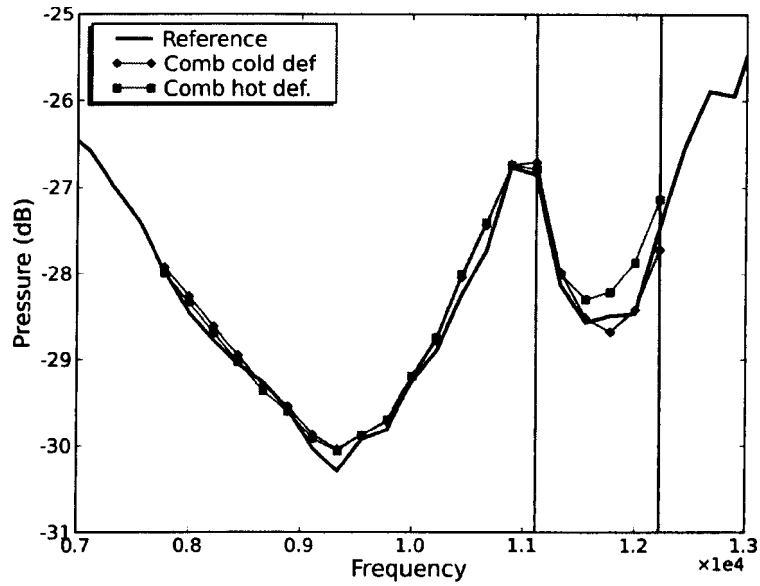


(a)

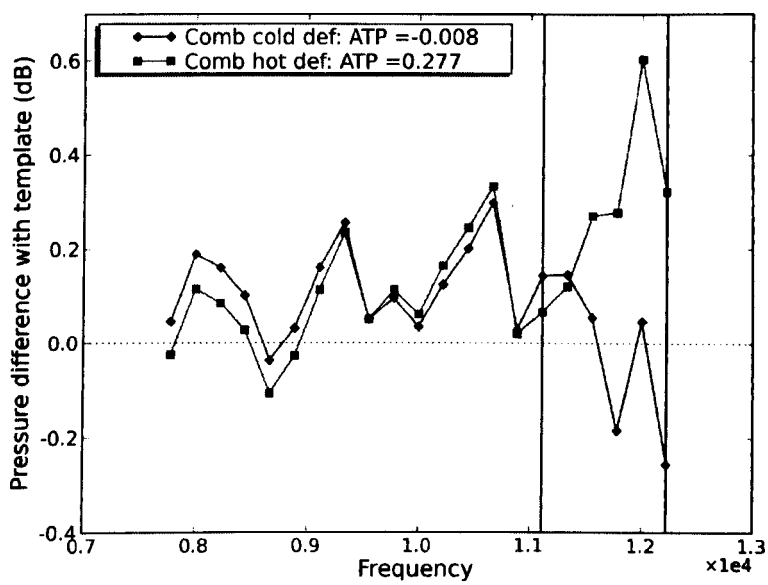


(b)

Figure 5.18: Acoustic pressures generated by deformations in the hot areas (described in Figure 5.16) on the frontal HRTF, alongside template pressures. The pressure changes generated by the deformations are shown in (b). The variation within the target notch frequency range (11.1-12.2 kHz) is consistently higher than for blue area deformation (see Figure 5.17). This translates to far larger average target perturbations (ATP, shown in the legend), exceeding 0.14 dB in all cases.



(a)



(b)

Figure 5.19: Acoustic pressures generated by combined cold area deformations and combined hot area deformations (described in Figure 5.16) on the frontal HRTF, alongside template pressures. The deformation amplitude is 1 mm in all cases. The pressure changes generated by the combined deformations are shown in (b). ATP is far higher (by a factor of over 30) in the case of combined hot area deformations than in the case of combined cold area deformations.

further validation to the temperature map generated through MPA. Indeed, although the effects of deformations in the thin cold areas (see Figure 5.17) suggest they are, in reality, slightly warmer, all deformations in the hot areas have, as expected, a more significant effect on the target notch.

Another test was carried out, this time comparing the effects of applying all hot area deformations simultaneously with that of applying all cold area deformations simultaneously (see Figure 5.16). The amplitude of all the deformations was, as before, 1 mm. Figure 5.19(a) shows the pressures generated in both cases, along with the reference. The pressure changes are shown in 5.19(b). The hot area deformation combination generates a large perturbation in the frequency range of the target notch, reaching a maximum around 0.6 dB. Conversely, the cold area deformation combination has a small effect in the target frequency range. The magnitude of the average target perturbation (see ATP in Figure 5.19(b) legend) is more than 30 times superior in the case of combined hot area deformations than in the case of cold area deformations, which strongly support the temperature distribution generated through MPA. Interestingly, the falling edge of the notch (11.1-11.5 kHz range) has changed very little in either case indicating that morphological variations have a weaker effect on it than on the rising edge of the notch (11.5-12.2 kHz range).

Overall, the results obtained using DPS-based MPA are very promising. Morphological regions identified as most significant to the generation of the chosen target notch (see Figure 5.13) are in the cymba concha and antihelix. As expected, the surface of the head is shown to be relatively uninvolved. This is in agreement with the current understanding of pinna acoustics, which attributes the generation of steep high-frequency spectral slopes to

the inner pinna region (see Section 2.2.2). Further validation has been provided by comparing the acoustic effects of small local deformations applied to areas identified by the MPA as hot and cold (all in the inner pinna region). Indeed, the perturbation inflicted on the target spectral feature is invariably more significant in the case of hot area deformations. When combined, local deformations in hot areas reinforce to generate a large perturbation of the target spectral feature, while the effect of a combination of cold area deformations stays small. Further testing is needed to assess the performance of MPA for different HRTF positions and target features, however, the potential of the technique as a powerful tool for investigating the morphological origin of acoustic features in HRTFs has been demonstrated.

Chapter 6

Conclusion and Future Work

6.1 Summary of contributions

The work reported in this thesis covers a broad range of issues. Firstly, it provides fresh insights into how spectral localisation cues, encapsulated in HRTFs, vary as a function of source direction. The approach presented for localisation cue analysis employs acoustic simulations as a powerful tool for studying HRTF features and, in particular, their spatially dynamic behaviour. Secondly, a large database of accurate and complete human head and pinnae surface meshes has been compiled, along with corresponding HRTF measurements. To augment the shape capture, a novel parameterisation method is described, which provides a major improvement in the efficiency with which the complex shape of the head and pinnae can be expressed. Finally, differential pressure synthesis (DPS) principles, first introduced by Tao *et al.* (2003a), have been applied to an innovative tool for revealing the morphological origin of acoustic localisation cues using morphoacoustic perturbation analysis (MPA). Both DPS and MPA are extensively described and validated. Each of these contributions is now considered

in turn in greater detail.

6.1.1 Simulation-based localisation cue studies

As a recognised standard in many fields relating to sound and acoustics, the KEMAR manikin was chosen as a test case for an in-depth study of directional HRTF variations. Although the constituent materials of the KEMAR manikin prevented a uni-modal shape capture, a raw mesh description was obtained by combining the results of CT and LASER scans. A novel multi-resolution meshing approach has been proposed, with performance tests revealing an error no greater than 1 dB over the ipsilateral area, when compared with a high resolution mesh (of maximum edge-length inferior to a quarter wavelength). This approach has proved useful in reducing the high computational requirements of acoustic simulations and the resulting acceleration allowed otherwise unachievable objectives to be fulfilled. Results have shown a high degree of compatibility with previously published results (Shaw and Teranishi, 1968; Hebrank and Wright, 1974; Carlile and Pralong, 1994; Kahana and Nelson, 2005). This agreement further validates the multi-resolution meshing approach, indicating that the commonly agreed $\lambda/6$ maximum edge length (with λ equating to wavelength at the maximum simulation frequency) can be broken in carefully controlled circumstances without significantly altering simulation results.

The level of accuracy offered by BEM-based acoustic simulations allowed a very detailed investigation of HRTF variations across the auditory space. The link between localisation acuity and numerical rates of ITD/ILD/spectral variation is considered an important observation, which has been made repeatedly throughout the analysis. A decrease in the acu-

ity of azimuth perception as a source in the horizontal plane moves away from the median plane, for example, can be explained by the decrease in the rate of change in ITD and ILD as azimuth grows from 0° to 90°. The principle extends to spectral variations and vertical localisation acuity in the different sagittal planes which have been investigated. This relationship was suggested by Leung and Carlile (2004), who also noted that, for some subjects, location-dependent cue variations correlated well with variations in perceptual resolution on cones of confusion. This link has been shown to be largely valid in the case of acoustic variations generated by the KEMAR head and generally accepted patterns of localisation acuity variation.

The similarity of HRTF variations around different rings of confusion is very noticeable below 10 kHz. This observation is compatible with the results of Morimoto *et al.* (2001), Morimoto *et al.* (2003a) and Leung and Carlile (2004), which suggest that elevation (φ in interaural-polar coordinates, see Section 2.1.2) is coded by a set of spectral cues which are similar for any azimuth (θ), excluding the region neighbouring $\theta = 90°$, where the decreasing radius of rings of confusion becomes an obvious issue. This being said, significant differences are noticeable above 10 kHz, where the most pronounced pinna resonance lies, casting some doubt over the correctness of the aforementioned suggestions in the upper audible frequency range. Similarities across rings of confusion under 10 kHz may allow elevation cues in any sagittal plane to be extrapolated from median plane spectral variations with some success. However, many studies have shown that frequencies above 10 kHz influence localisation (Hebrank and Wright, 1974; Bronkhorst, 1995; Langendijk and Bronkhorst, 2002; Best *et al.*, 2005). The marked disparities above 10 kHz suggest that extrapolating median plane elevation cues for all azimuths, over the entire audible frequency range, will result in unwanted

effects. Indeed, timbral distortions and a reduction in externalisation sensations, although they may be subtle, are likely consequences of ignoring or disturbing high-frequency spectral variations across sagittal planes. Another indication that sub-optimal localisation will result from such an approach is the spectral variation in the horizontal plane, which allows surprisingly robust monaural azimuth perception (Perrott and Saberi, 1990), a testament to the importance of cross-azimuth spectral HRTF changes.

The detailed picture of pressure variations on the surface of the KEMAR head/pinna mesh, provided by the BEM, compares favourably with that obtained in a similar study undertaken by Kahana and Nelson (2005). The pinna resonances which have been identified have a dramatic influence on HRTF spectra. Indeed, they lock the major spectral peaks to resonant frequencies. The strength of each peak changes with source direction, creating strong candidates for localisation cues. Although the resonance frequencies obtained for the KEMAR head are generally compatible with those reported by Kahana *et al.*, an upwards shift in frequency is noticeable for most of them. This is presumably due to the slightly smaller KEMAR pinna used in our simulations (DB-61). Obtaining these resonance frequencies for any given individual would prove very valuable in the context of HRTF estimation. Conversely, the frequency of spectral notches is not set by pinna resonances but can be predicted, to some extent, by mathematical models (Batteau, 1967; Watkins, 1978; Lopez-Poveda and Meddis, 1996). An HRTF estimation model combining the estimated effects of pinna resonances and reflections on spectral peaks and notches, respectively, could prove very powerful and has not yet, to the author's knowledge, been proposed.

6.1.2 Morphoacoustic data collection and parameterisation

A database containing shape and acoustic information for 49 subjects has been compiled. This provides a unique dataset on which to perform future analyses. The MRI based shape capture has proved comprehensive and accurate. Freely available software allowed raw three-dimensional meshes to be obtained from the original DICOM images. Clean three-dimensional models describing head/pinna surfaces were extracted from the raw data by eliminating unwanted residual elements contained inside the surface using a novel, image-based, technique. The shape database therefore contains highly detailed and accurate surface shape information for 49 individuals and is, to the author's knowledge, the only one of its kind. The nature of the shape description virtually eliminates the risk of relevant detail being omitted. The inner ear-canal and nasal passages are omitted although they, arguably, play a minor a role in the localisation cue generation process. This data is present in the original MRI scans, however, and its extraction would be possible, if necessary, although technically challenging.

The comprehensive nature of the morphological description comes at the cost of information volume. Each mesh requires vertices numbering in the tens of thousands to allow the captured detail to be fully expressed. To address this problem a parameterisation technique combining aspects of the elliptic Fourier transform (EFT) with principal component analysis (PCA), referred to as the elliptic Karhunen-Loève transform (EKLT) has been proposed and tested. This technique exploits the redundancy arising from underlying morphological similarities across individuals, allowing the data to be hugely compressed with arbitrarily small distortion. The slice-based approach at the core of the technique inherently increases shape resolution

close to the interaural axis, a major advantage given the prominent role of the pinnae in the production of spectral cues. The manageable size of the parameterised morphological dataset will facilitate and greatly accelerate future analysis of a wide range of subtle cross subject morphological variations.

Acoustic data describing HRTF variations across the auditory space was gathered in parallel with morphological data in a well-established measurement environment at the Auditory Neuroscience Laboratory, in Sydney. The information volume, as in the case of the shape data, is unmanageable in its raw form. A parameterisation scheme is planned, whereby the data dimensionality will be greatly reduced. The PCA analysis of variations across space and individuals has already been performed in a number of studies (see Jin *et al.*, 2000; Rodriguez and Ramirez, 2005, for example) with good results. Other methods such as Isomap and locally linear embedding (LLE) have, however, been shown to perform better when attempting to encode the spatially varying character of HRTFs (Kapralos and Mekuz, 2007; Kapralos *et al.*, 2008). Further research is being pursued to improve the expression of the morphological and acoustic data-sets. The resulting database will provide a strong, concise basis for future research efforts investigating the mappings between variations in morphology and variations in HRTFs across individuals.

6.1.3 DPS and morphoacoustic perturbation analysis

Chapter 5 extends the work of Tao *et al.* (2003a) on differential pressure synthesis (DPS). In order to investigate the performance of DPS, applied to the KEMAR head, orthogonal deformations were adapted so as to allow

arbitrary micro-deformation of the external ear, unachievable with the surface spherical harmonics (SHHs) used by Tao *et al.* (2003a). The proposed orthogonal deformations are a variation on EFT parameter perturbation which, although they are also orthogonal, produce deformations with erratically varying spatial frequency. Each orthogonal deformation, referred to as an elliptic surface harmonic deformation, is defined by its associated slice-harmonic and cross-harmonic. A database relating these orthogonal deformations to their acoustic effects was compiled for 800 deformations allowing DPS principles to be tested. For deformations with low cross-harmonic coefficients, the linear approximation of the relationship between deformation amplitude and acoustic pressure is valid for a wide range of slice-harmonics, deformation amplitudes and frequencies. In the case of higher cross-harmonic coefficients, however, the range where linear approximation is acceptable is considerably reduced, especially at higher frequencies.

The limited range of deformation amplitudes over which a valid linear mapping to associated acoustic effects can be made prevents their use in HRTF estimation across the range of morphological variations observed in the general population, which had been the original motivation of the DPS method. However, the application of DPS to a novel technique referred to as morphoacoustic perturbation analysis (MPA) has been demonstrated. In general, MPA allows the morphological origin of an acoustic feature for any given point in auditory space to be identified. Section 5.4 describes how DPS can be applied to allow efficient and effective MPA. This process was tested by targeting a spectral notch present in the frontal KEMAR HRTF. The orthogonal deformations were summed so that associated acoustic pressure changes reinforced to perturb the target notch. The total deformation then gives a measure of the involvement of each morphological region in the

production of the target notch. The results showed great promise. The distribution of the summed deformation (also referred to as temperature map), which gives a measure of the contribution of each morphological area to the production of the target feature, is what one would expect given the current understanding of pinna acoustics. The concha and antihelix are identified as primary production regions, with the rest of the pinna being significantly cooler. The remainder of the head is identified as irrelevant. A number of minor imperfections due to the nature of the orthogonal deformations, non-linearities and an incomplete DPS database were noted.

6.2 Future Work and discussion

6.2.1 Further data gathering and validations

The value of the work reported in this thesis could be significantly enhanced by further data gathering exercises. Available computing resources, for example, have restricted the validity of acoustic simulations to the ipsilateral hemisphere and to frequencies falling short of covering the entire audible range. Although contralateral spectral cues are known to be less perceptually salient than their ipsilateral counterparts, their effects have been shown to be significant (Humanski and Butler, 1988; Jin *et al.*, 2004) and the presented study of spectral variation would benefit from being extended to include them. The results obtained through BEM-based simulations using the proposed multi-resolution approach to meshing have been shown to exhibit only slight deviations (under 1 dB) in comparison with those obtained with a high-resolution mesh (with maximum edge length inferior to a quarter wavelength), in the ipsilateral area, up to 15 kHz. Further tests

will be needed to determine how, if at all, multi-resolution meshing can be applied to reduce simulation time with valid results up to 20 kHz, over the entire auditory space.

A link between numerical rates of localisation cue change and localisation acuity has been mentioned repeatedly in the analysis of simulation results. Although published localisation acuity data agree to some extent on localisation acuity variation patterns, discrepancies between published results make an empirical verification of the theory difficult. A large scale study, conducted for several subjects, investigating both vertical and horizontal minimum audible angles (MAAs) for a large number of positions in the horizontal, frontal, median and other sagittal planes, for pure tones at different frequencies, as well as band-pass filtered sounds, is needed for a truly rigorous verification. The results of such a study, combined with a highly detailed analysis of HRTF variations for each subject, obtained possibly through acoustic simulation, would provide a strong basis for developing a deep understanding of the perceptual mechanisms underlying localisation.

Also, increasing the number of individuals contained in the morphoacoustic database would be beneficial for the development of an effective cue individualisation model. Indeed, a wider range of morphoacoustic variations present in the database can only improve the performance of statistical models aiming to estimate acoustic characteristics from morphological data. Although acoustic measurements have been performed for all the subjects in the database, the detailed mesh models obtained for each of them potentially allow the equivalent acoustic simulations to be performed, with all the inherent advantages that they potentially provide. The elimination of unwanted HRTF measurement variations would be a particularly desirable

asset. For example, variations in meatus size inherently alter microphone positioning and orientation and occasionally prevent measurements altogether, when the meatus is too narrow to allow microphone placement. Undesired effects resulting from small involuntary subject movements, resulting from coughing, swallowing or sneezing, and the effects of background noise and transducer imperfections, would also be eliminated.

The surface pressure calculations and potentially unlimited far field pressure calculations are an additional drive for pursuing simulation-based studies using human subjects. Pinna resonance frequencies, which are key HRTF descriptors, could be easily identified for each subject. A systematic comparison of simulated results and measured results would also be very informative. However, conducting acoustic simulations on all subject meshes will require the cleaning, smoothing and re-meshing algorithms to be perfected in order for BEM requirements to be fulfilled. Indeed, the complete elimination of microscopic mesh imperfections, such as overlapping or unwanted polygons and mesh holes, for example, is a problem which still requires attention. Should these problems be overcome, the resulting highly accurate acoustic data, combined with the currently available shape data would allow morphoacoustic mappings to be explored with unprecedented precision and rigour.

6.2.2 Further MPA studies

A great body of literature has provided valuable insight into the acoustic origin of spectral cues, however, some effects are still poorly understood. MPA potentially allows this understanding to be deepened and the promise it has shown during the preliminary studies reported in this thesis justifies

further work to refine and optimise it for our applications. Its adaptation to other fields involving the design of objects based on interactions with surrounding fluids is also an interesting possibility.

The database relating orthogonal shape deformations to their acoustic effects would benefit from being extended to higher orders, allowing finer detail to be probed and reducing artefacts resulting from their harmonic nature. It is possible, however, that the proposed orthogonal deformations will be refined or even superseded. MPA is extremely flexible, as it does not constrain the nature of the orthogonal shape deformations and the cost functions used to quantify their individual effects on target acoustic features. There is, therefore, ample scope for further improvements, although convincing results have already been obtained. Also, a significant acceleration in computation would be achieved by restricting MPA to the pinna area without including the head whose effects are already relatively well understood.

6.2.3 Perceptually based shape description

The comprehensive nature of the shape description captured using MRI and covered in Chapter 4, solves the problem of incomplete shape data evident in the literature. However, it presents fresh problems such as the introduction of superfluous surface shape data, which is acoustically irrelevant. Mixing large amounts of irrelevant detail with relevant shape description can only result in adverse effects on acoustic-related analyses. The ability to filter out unnecessary data would greatly benefit any shape parameterisation method used for the purpose of HRTF estimation. The regions of the head and pinnae which require a detailed description for viable HRTF

estimation from morphology must therefore be identified.

MPA introduced in Chapter 5 is a strong candidate for identifying the areas that shape description should focus on. The work carried out in Chapter 3 added to the large body of literature investigating the perceptual aspects of localisation (reviewed in Section 2.2.2) can serve to indicate the HRTF features on which MPA should be performed. One of the great assets of MPA is its ability to identify morphological areas responsible for the generation of specific HRTF spectral features for sound sources in a specific directions. Given that morphological involvement in cue production varies with HRTF direction, it is possible that extracting separate morphoacoustic mappings for each source directions, using a shape description focused on the cue producing areas particular to each of them, would improve results.

Concentrating shape description on wanted areas can be achieved, in practical terms, by increasing the sample point density in the morphological slice description for these regions, prior to feeding the slices through the EKLT parameterisation method (see Section 4.2). Currently, points are spread equally along the contour of each slice. A variable point spacing reflecting relevance to the cue production mechanism is one possible route towards filtering out irrelevant shape detail. This would enhance prospects for the extraction of clear, meaningful and powerful mappings allowing localisation cues to be efficiently and effectively individualised using parameterised shape information, our ultimate objective. The extraction itself could be performed through linear regression which has already been used by Jin *et al.* (2000) to extract mappings between PCA representations of sets of 20 landmark measurements and corresponding directional transfer function (DTF) sets with promising results. This approach will be enhanced by the

data and techniques presented in this thesis.

6.2.4 Affordable shape capture

Once a satisfactory model for the individualisation of localisation cues based on morphological descriptors is achieved, HRTF estimation should no longer require comprehensive shape capture with accuracy comparable to that of MRI scanning over the full head, which remains prohibitively expensive. Ideally, shape capture should be an affordable process; a set of pictures obtained using ordinary digital cameras, for example. In order for this to be possible, it is necessary to establish the set of pictures which would most usefully describe morphology for the purpose of HRTF estimation. The database of reconstructed subject meshes allows a large number of images to be obtained quickly, easily and accurately, from any angle, with any lighting, using automated rendering routines. Extensive investigations into the feasibility of EKLT parameter estimation (or possibly, direct cue individualisation) from this parameterised image data can be conducted much more easily than using real subjects. A global morphological temperature map, calculated through MPA over all identified localisation cues and directions could inform the choice of pictures which should be taken.

6.2.5 Incorporation of torso effects and environmental cues

The research reported in this thesis focuses on the effect of the head and pinna and discounts the effect of the torso and shoulders completely, although these have been demonstrated to play a significant role in cue production (Algazi *et al.*, 2002). Ignoring their effect would have significant repercussions on the effectiveness of binaural systems (especially when

reproducing elevated sound images). Their effects can, however, be incorporated into the final estimation procedure through previously proposed mathematical models (Avendano *et al.*, 1999; Algazi *et al.*, 2001a, 2002; Algazi and Duda, 2002) which have, to some extent, been perceptually validated (Zotkin *et al.*, 2003). The shoulder and torso shape for the database subjects is present in the MRI scans and is easily extractable. Limited computing power prohibited their acoustic simulation, although this barrier will become surmountable in the foreseeable future. A thorough, simulation-based, investigation of the error introduced by ellipsoidal approximations of the head and torso would provide useful insight into the validity of structural/mathematical model-based predictions over the entire auditory space and audible frequency range and will determine whether the scope for improvement justifies further study.

Spatial cues emanating from the acoustic environment such as reflections and reverberation are also known to contribute to auditory spatial perception, localisation and externalisation (Shinn-Cunningham, 2000; Bronkhorst and Houtgast, 1999; Schoolmaster *et al.*, 2001). These effects, because they are dependent on the listener's environment rather than his morphology, have not been considered in this thesis. Their inclusion is, however, crucial to the achievement of realistic sound-field rendering. Structural approaches allowing these effects to be incorporated into an HRTF model have been proposed (Brown and Duda, 1997, for example). These models make the assumption that cues originating from the immediate auditory periphery (pinnae, head, torso and shoulders) and those originating from the surrounding environment can be generated separately, then combined to achieve enhanced spatialisation. Whether the term HRTF refers uniquely to the response of the auditory periphery or whether environmental effects should be

incorporated is a matter of definition but the fact that both are crucial is indisputable.

Appendix A

Supplementary KEMAR Acoustics plots

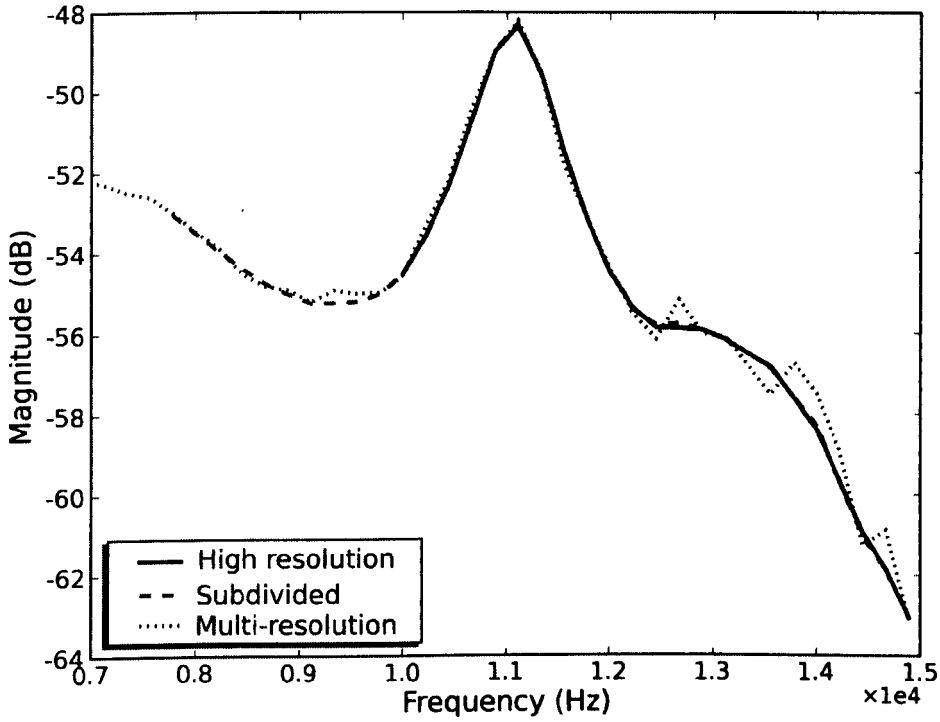
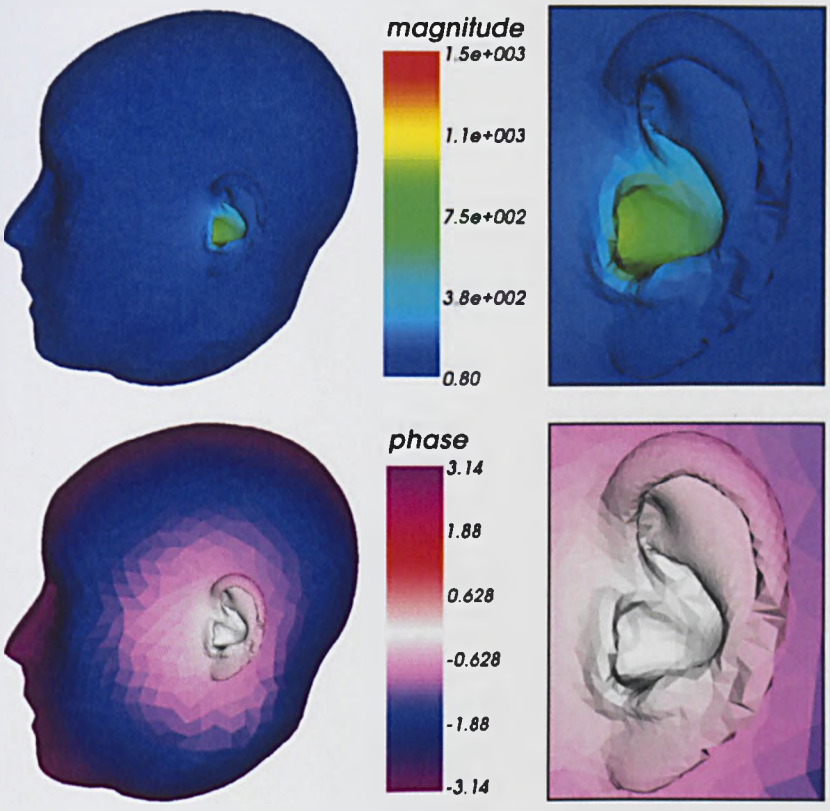
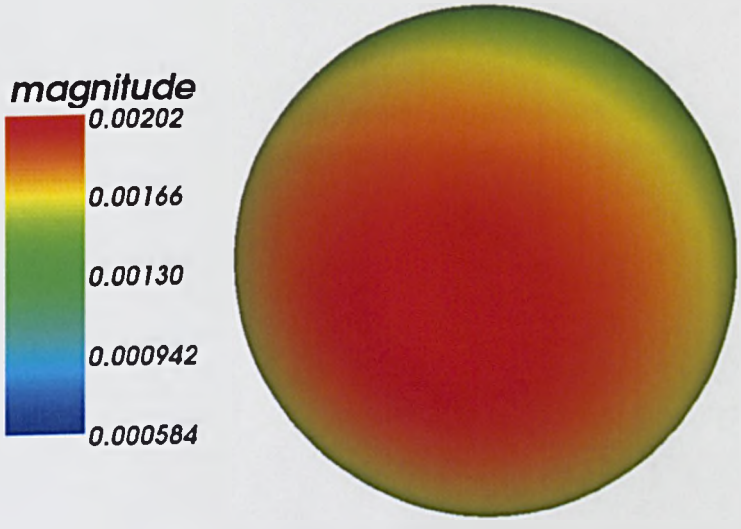


Figure A.1: BEM simulation results for the top HRTF in the case of the multi-resolution mesh (3,816 polygons), a subdivided version (13,879 polygons) and a high-resolution mesh, with maximum edge length inferior to $\lambda/4$ at 15 kHz.

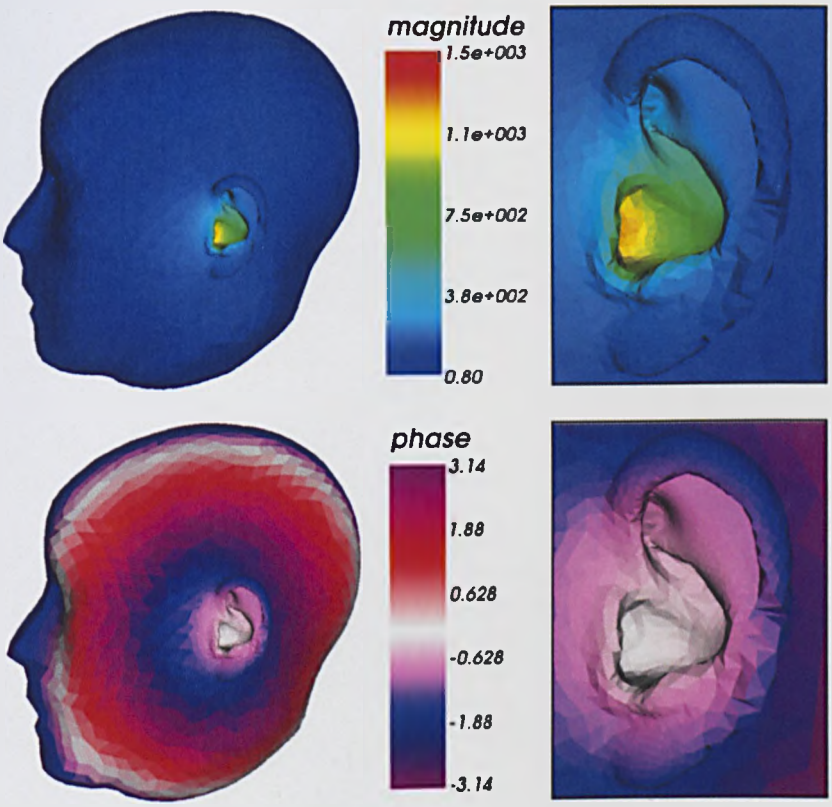


(a) Surface pressures.

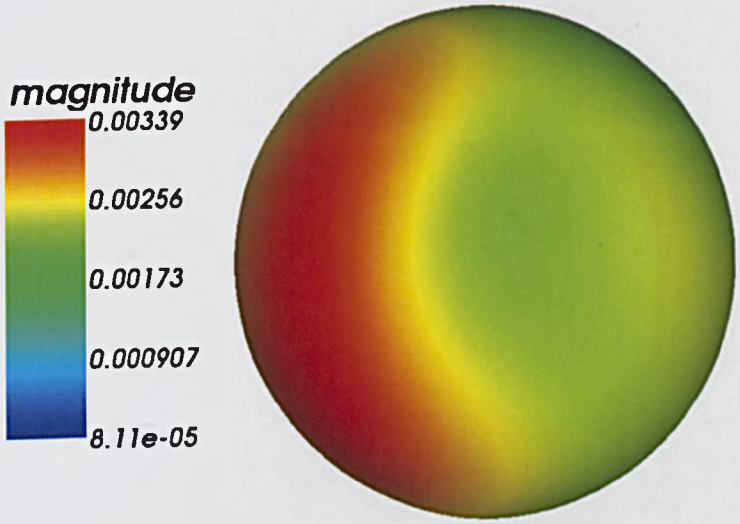


(b) Far field pressures.

Figure A.2: KEMAR acoustics at 1333 Hz. Pressure and phase are relative to a unit pressure point source close to the ear canal.

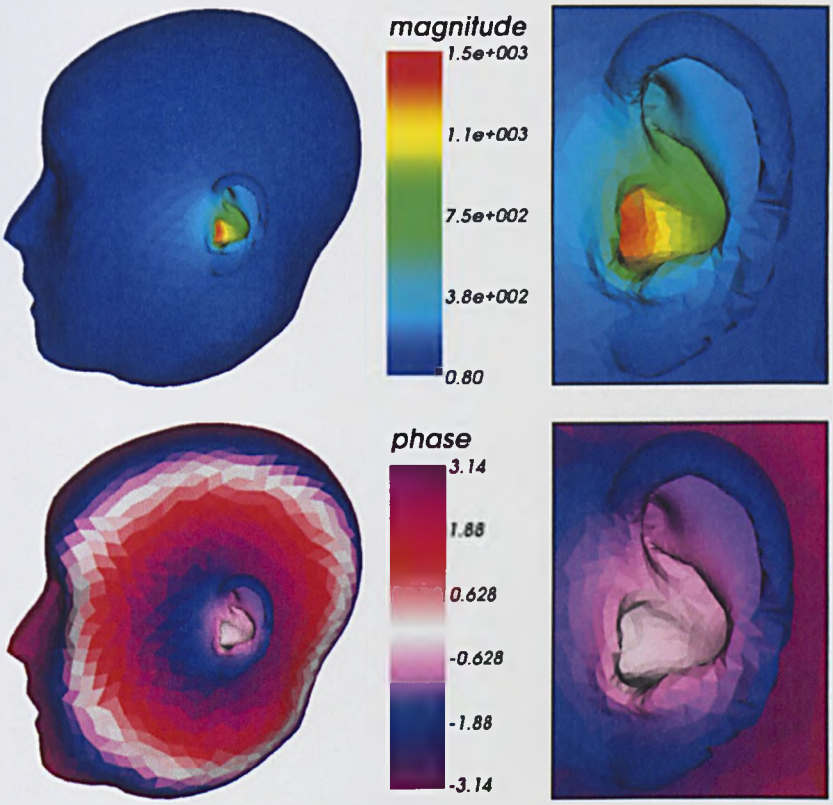


(a) Surface pressures.

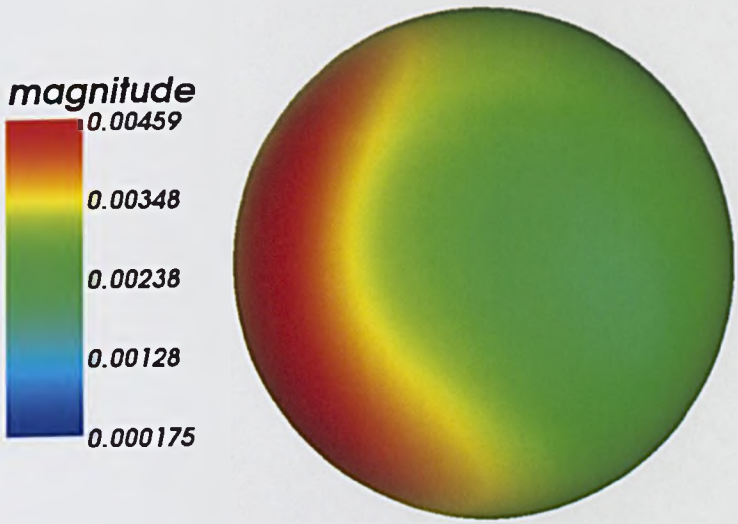


(b) Far field pressures.

Figure A.3: KEMAR acoustics at 3111 Hz. Pressure and phase are relative to a unit pressure point source close to the ear canal.

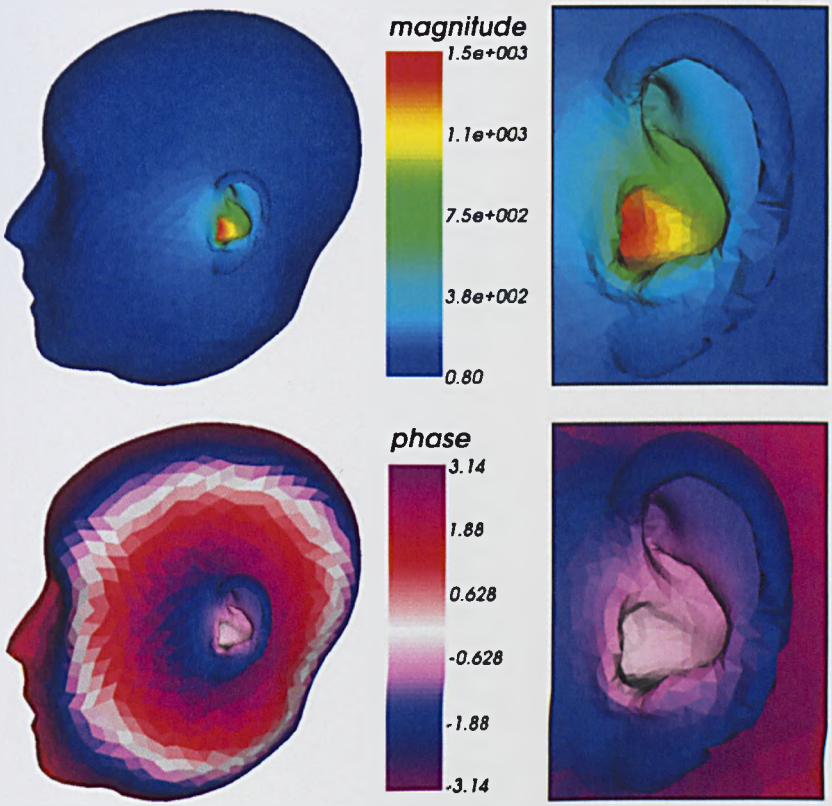


(a) Surface pressures.

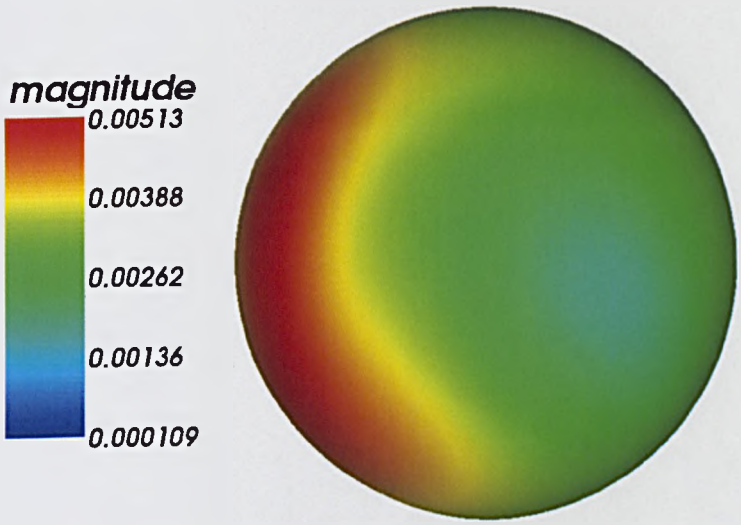


(b) Far field pressures.

Figure A.4: KEMAR acoustics at 3778 Hz. Pressure and phase are relative to a unit pressure point source close to the ear canal.

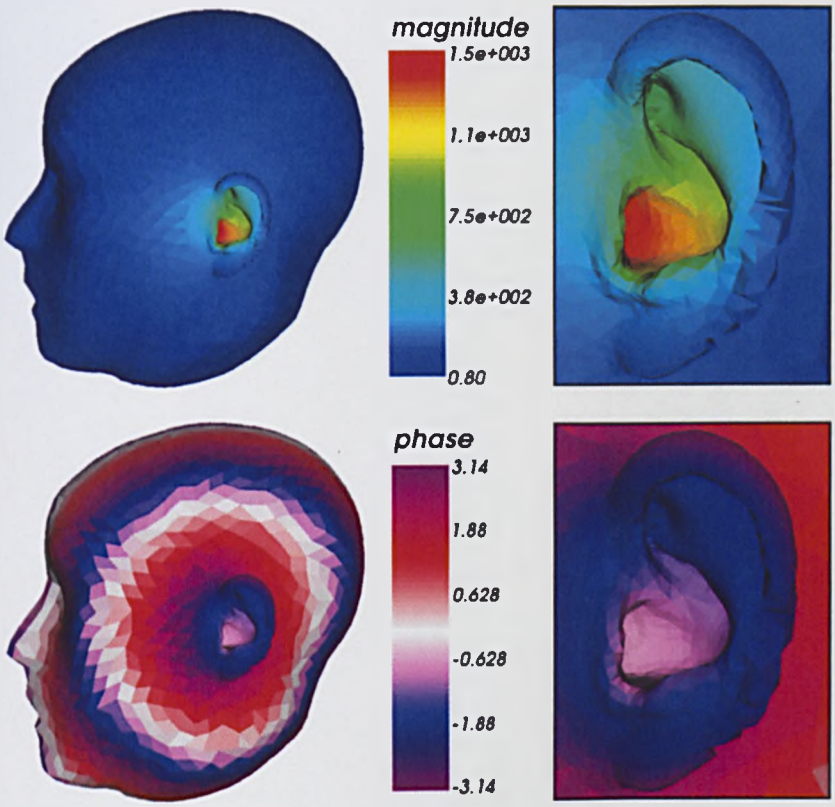


(a) Surface pressures.

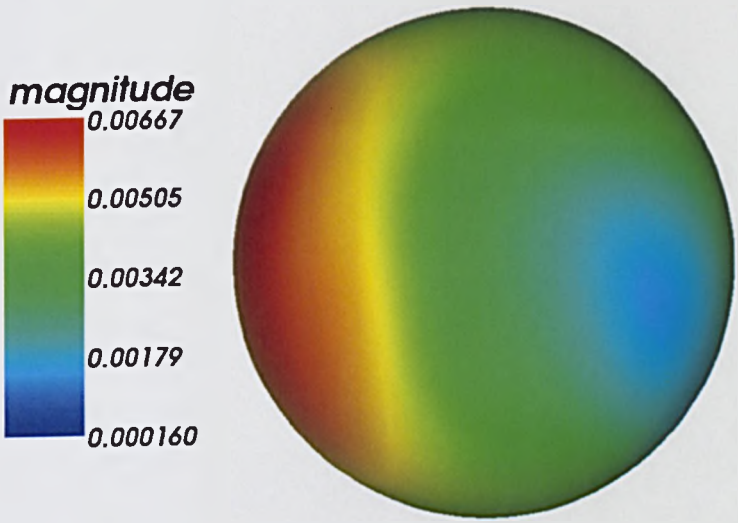


(b) Far field pressures.

Figure A.5: KEMAR acoustics at 4000 Hz. Pressure and phase are relative to a unit pressure point source close to the ear canal.

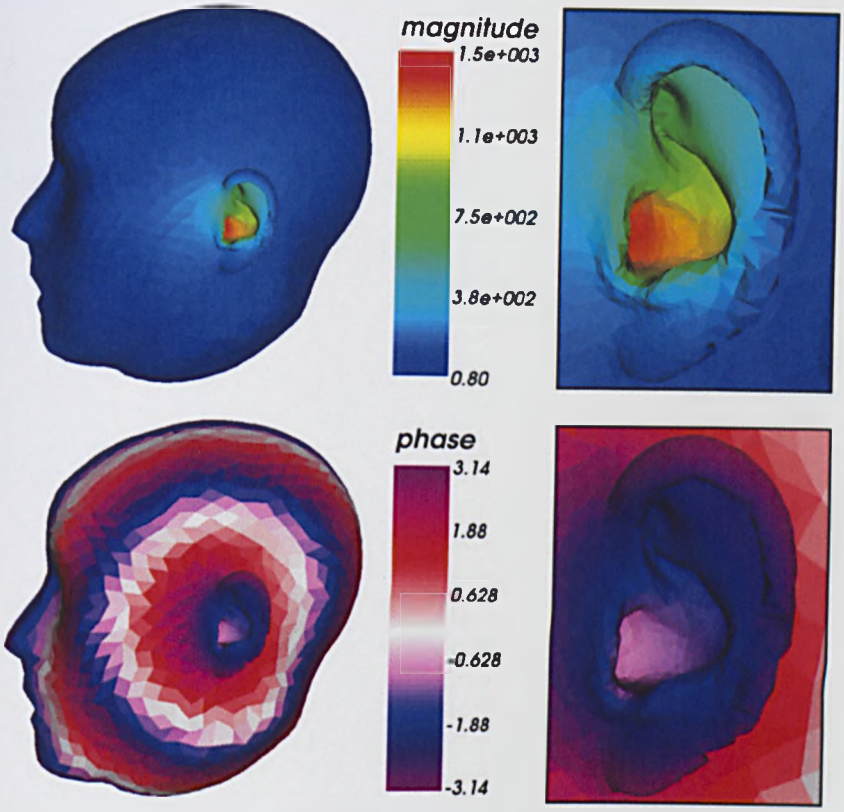


(a) Surface pressures.

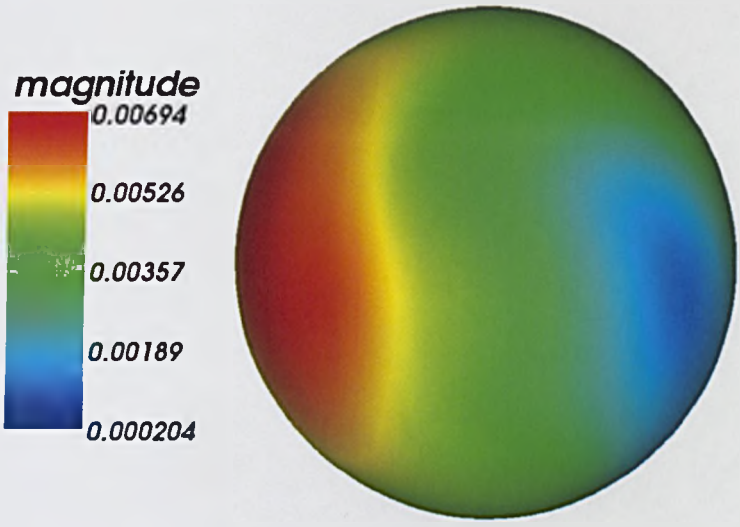


(b) Far field pressures.

Figure A.6: KEMAR acoustics at 4667 Hz. Pressure and phase are relative to a unit pressure point source close to the ear canal.

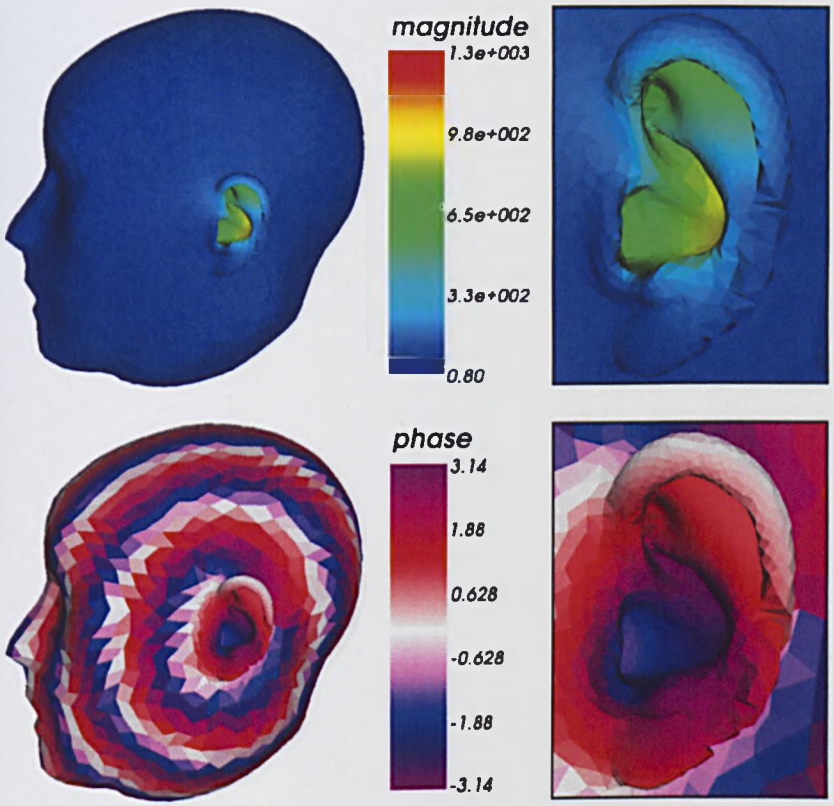


(a) Surface pressures.

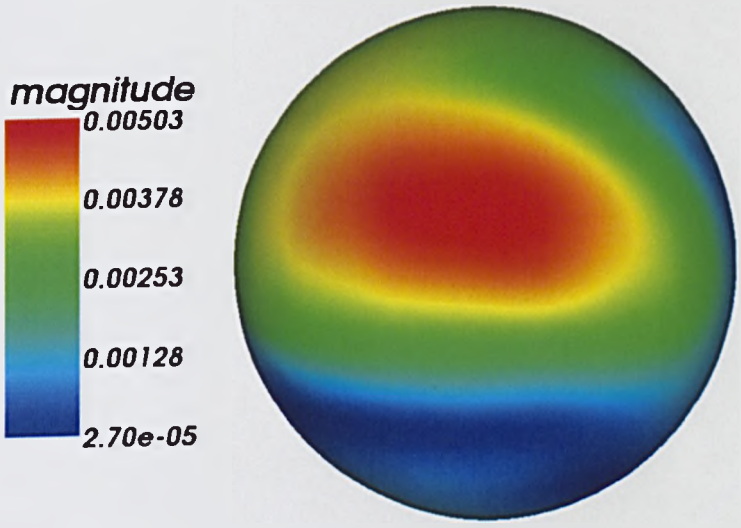


(b) Far field pressures.

Figure A.7: KEMAR acoustics at 5111 Hz. Pressure and phase are relative to a unit pressure point source close to the ear canal.

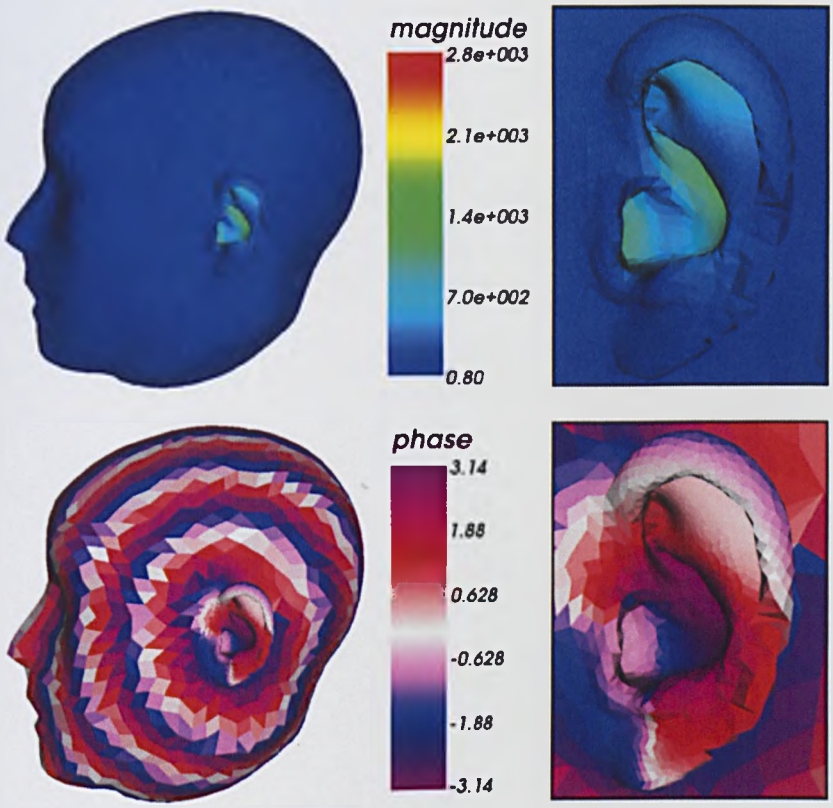


(a) Surface pressures.

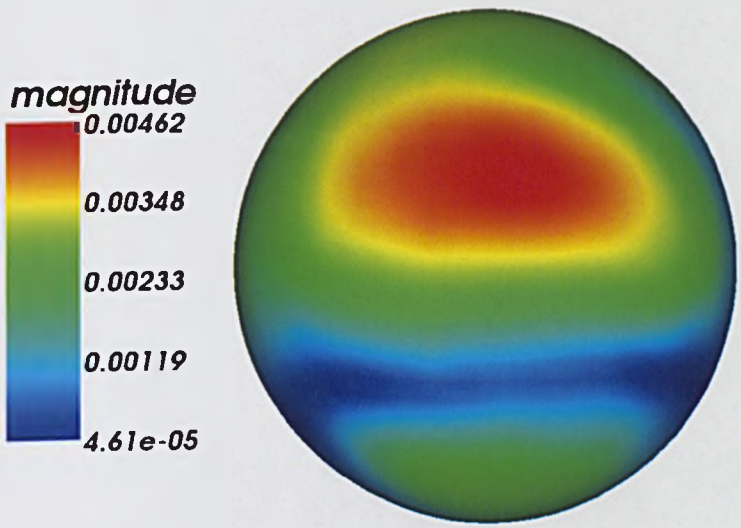


(b) Far field pressures.

Figure A.8: KEMAR acoustics at 8444 Hz. Pressure and phase are relative to a unit pressure point source close to the ear canal.

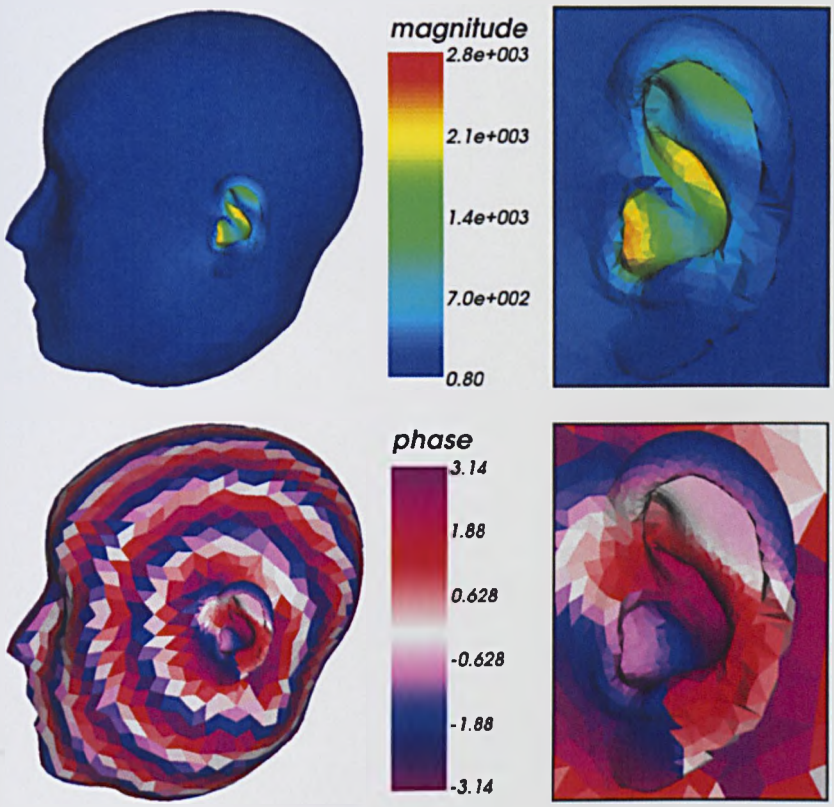


(a) Surface pressures.

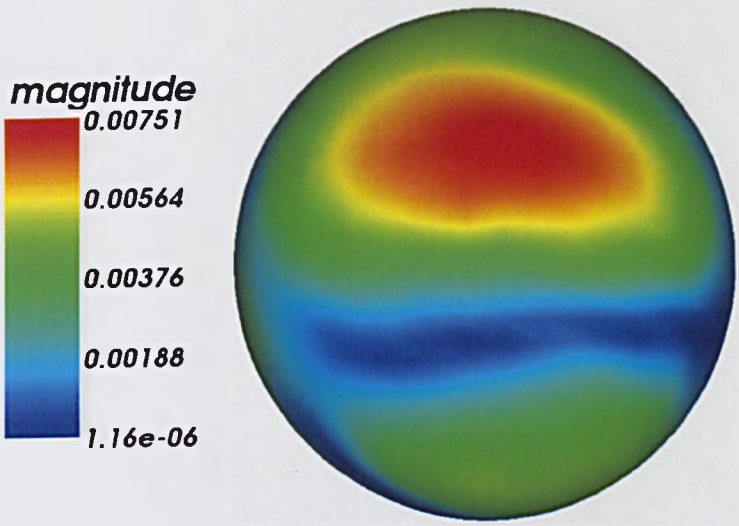


(b) Far field pressures.

Figure A.9: KEMAR acoustics at 9778 Hz. Pressure and phase are relative to a unit pressure point source close to the ear canal.

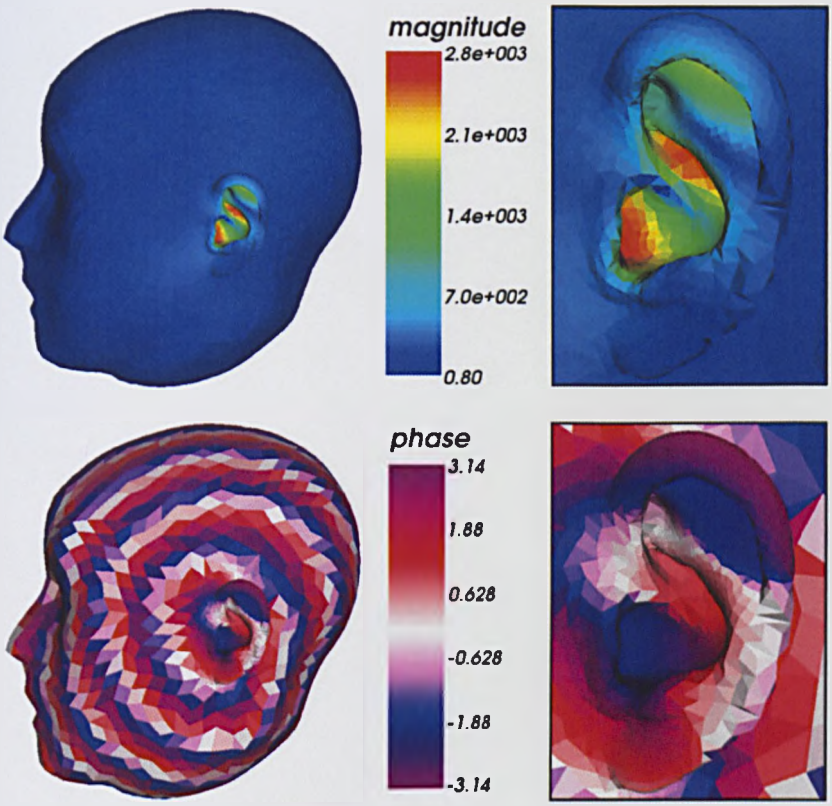


(a) Surface pressures.

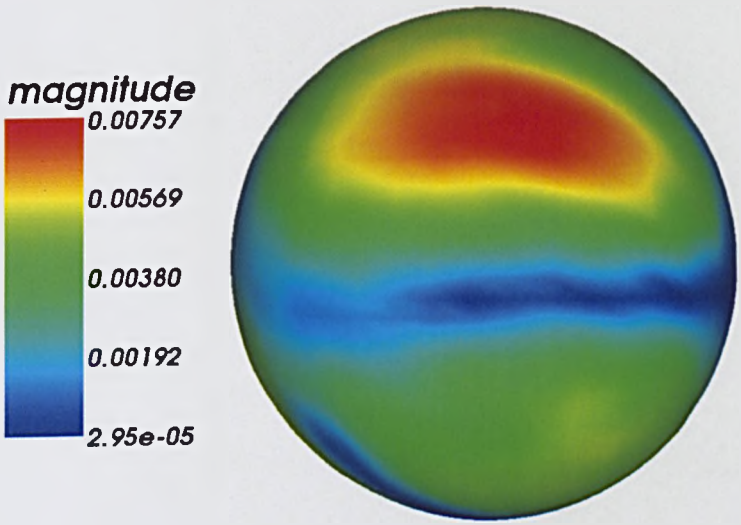


(b) Far field pressures.

Figure A.10: KEMAR acoustics at 10667 Hz. Pressure and phase are relative to a unit pressure point source close to the ear canal.

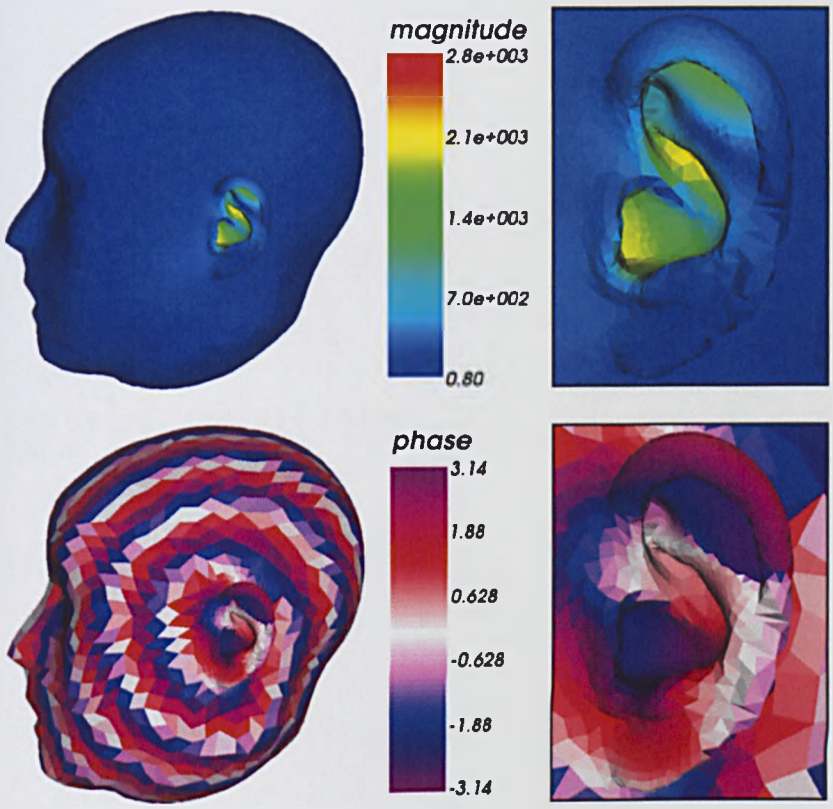


(a) Surface pressures.

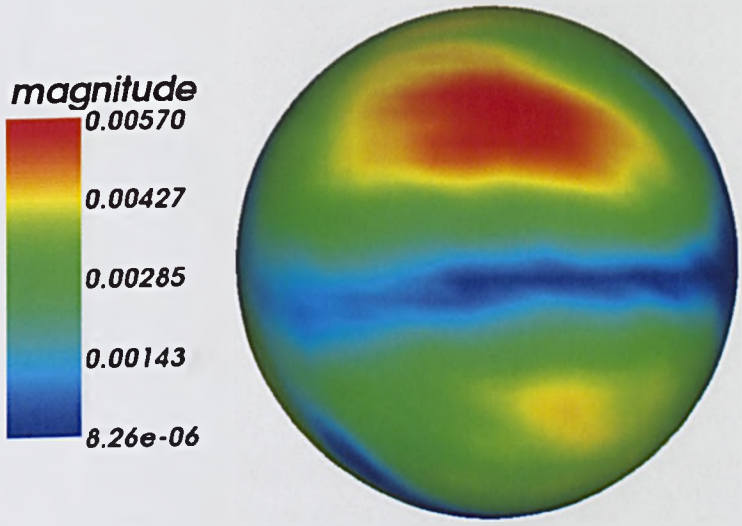


(b) Far field pressures.

Figure A.11: KEMAR acoustics at 11333 Hz. Pressure and phase are relative to a unit pressure point source close to the ear canal.

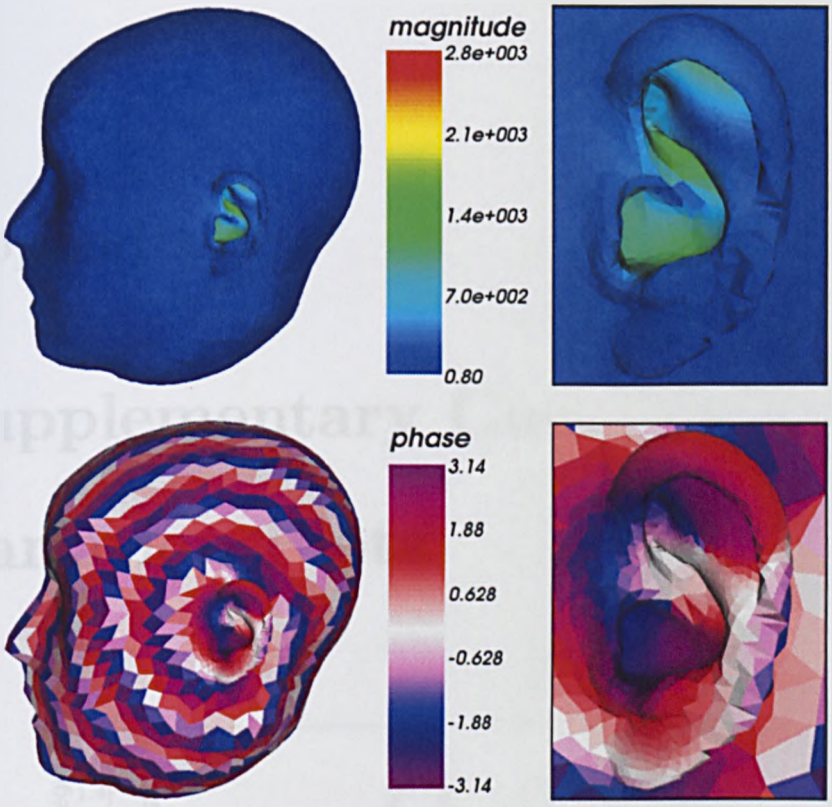


(a) Surface pressures.

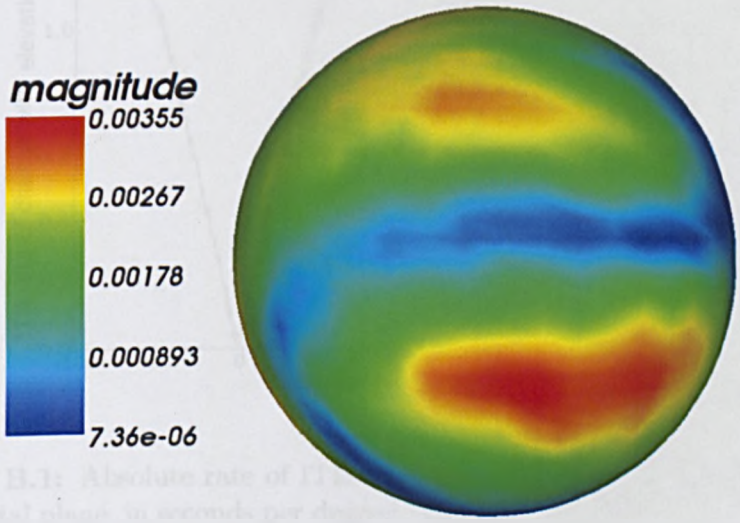


(b) Far field pressures.

Figure A.12: KEMAR acoustics at 11556 Hz. Pressure and phase are relative to a unit pressure point source close to the ear canal.



(a) Surface pressures.



(b) Far field pressures.

Figure A.13: KEMAR acoustics at 12222 Hz. Pressure and phase are relative to a unit pressure point source close to the ear canal.

Appendix B

Supplementary Cue Variation Plots

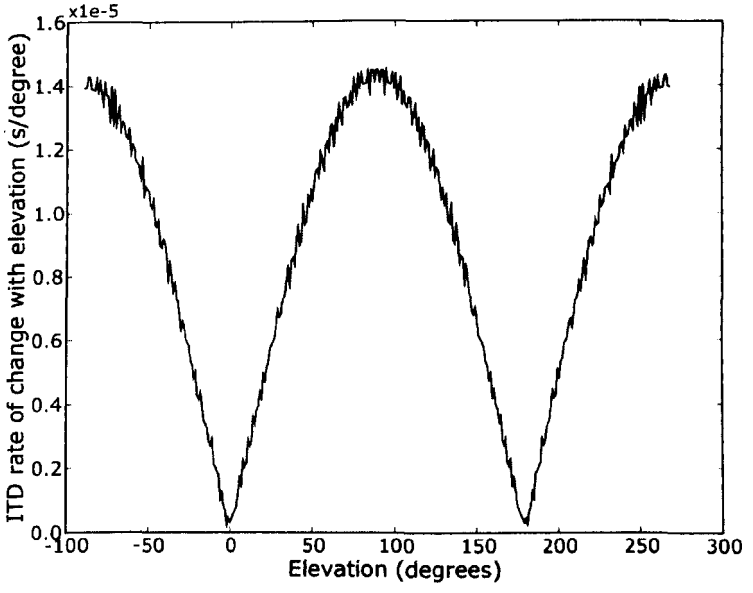


Figure B.1: Absolute rate of ITD change with elevation (ITD-ARCE) in the frontal plane, in seconds per degree. The ITD is calculated as the phase difference at 200 Hz.

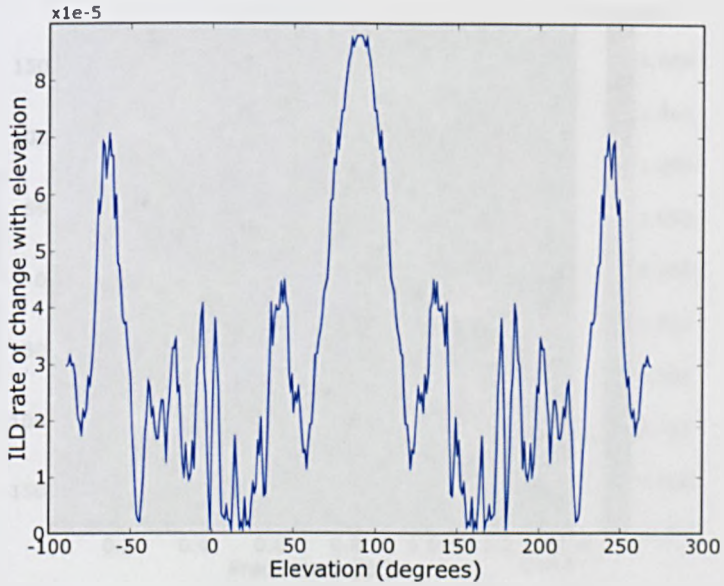


Figure B.2: Absolute rate of ILD change with elevation (ILD-ARCE) in the frontal plane, in magnitude units per degree. The ILD is calculated as mean linear magnitude difference across all simulated frequencies (0-14 kHz).

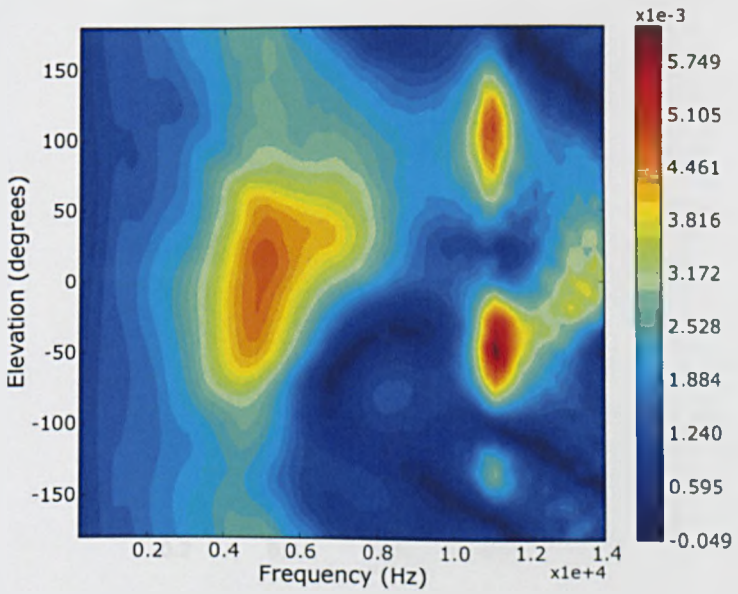


Figure B.3: Spectral variations with elevation at the ipsilateral ear around the 0.1 ms confusion ring, up to 14 kHz, in magnitude units (relative to a unit source placed close to the ear canal).

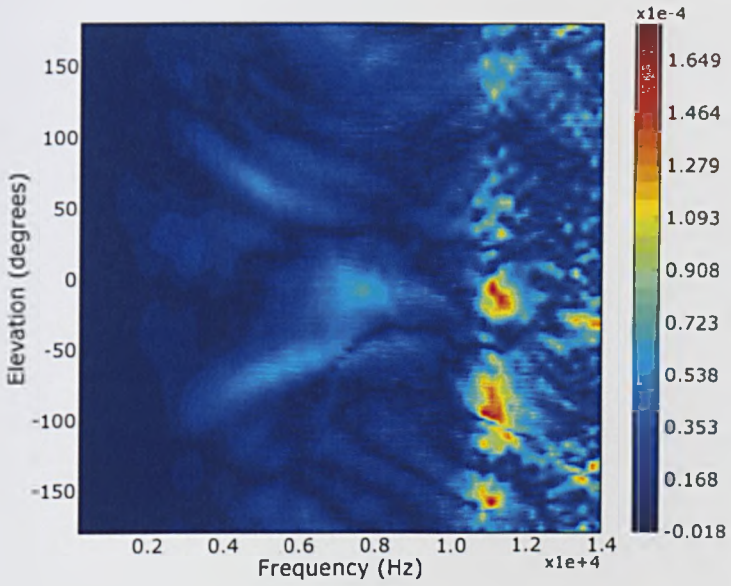


Figure B.4: Absolute rate of spectral change with elevation at the ipsilateral ear around the 0.1 ms confusion ring, up to 14 kHz, in magnitude units (relative to a unit source placed close to the ear canal) per degree.

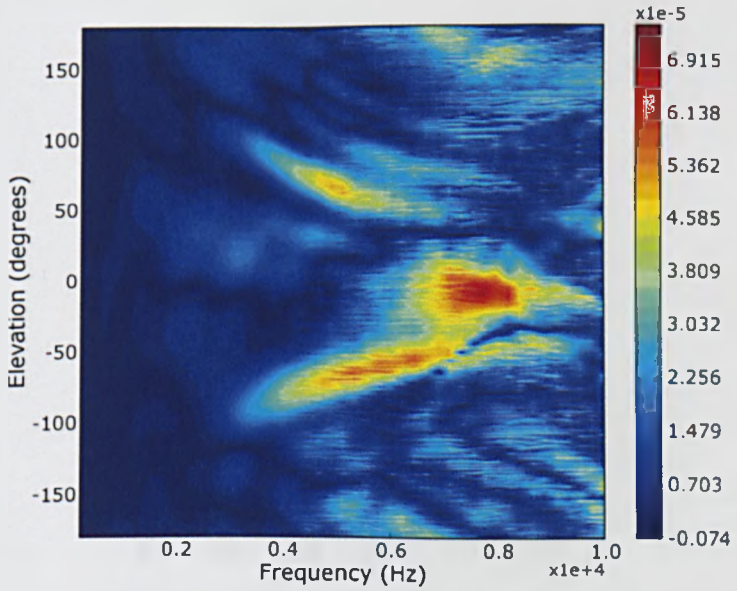


Figure B.5: Absolute rate of spectral change with elevation at the ipsilateral ear around the 0.1 ms confusion ring, up to 10 kHz, in magnitude units (relative to a unit source placed close to the ear canal) per degree.

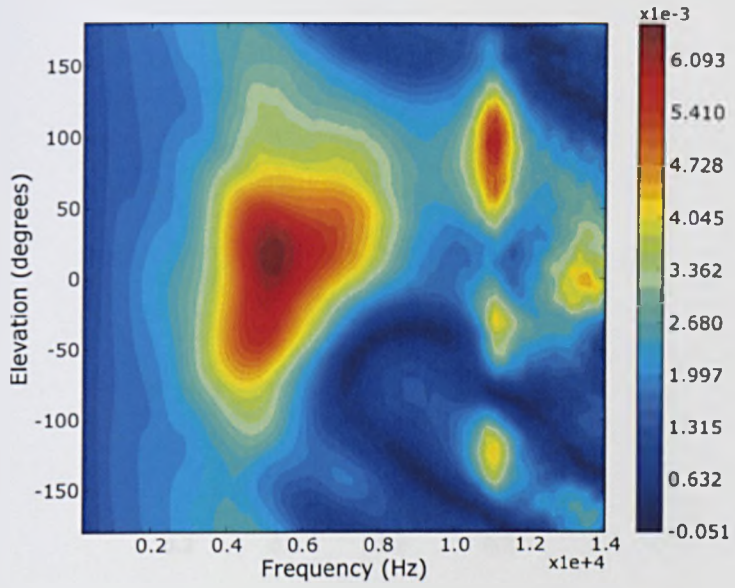


Figure B.6: Spectral variations with elevation at the ipsilateral ear around the 0.3 ms confusion ring, up to 14 kHz, in magnitude units (relative to a unit source placed close to the ear canal).

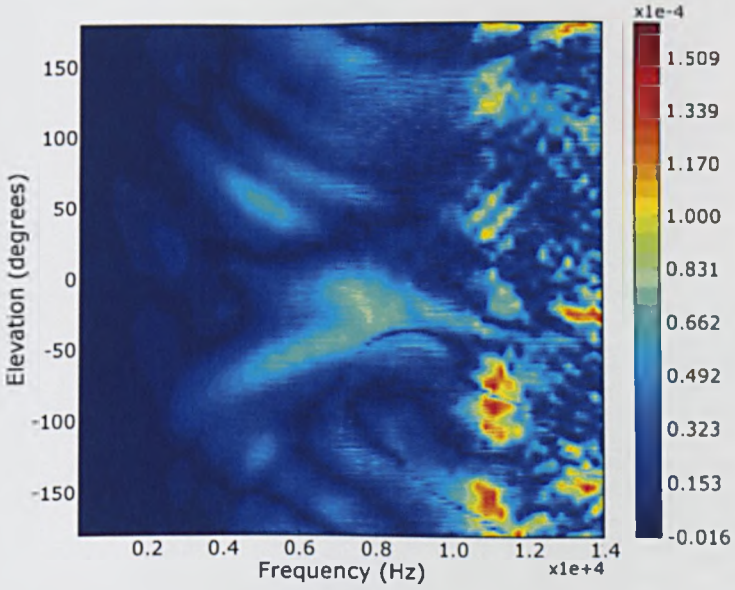


Figure B.7: Absolute rate of spectral change with elevation at the ipsilateral ear around the 0.3 ms confusion ring, up to 14 kHz, in magnitude units (relative to a unit source placed close to the ear canal) per degree.

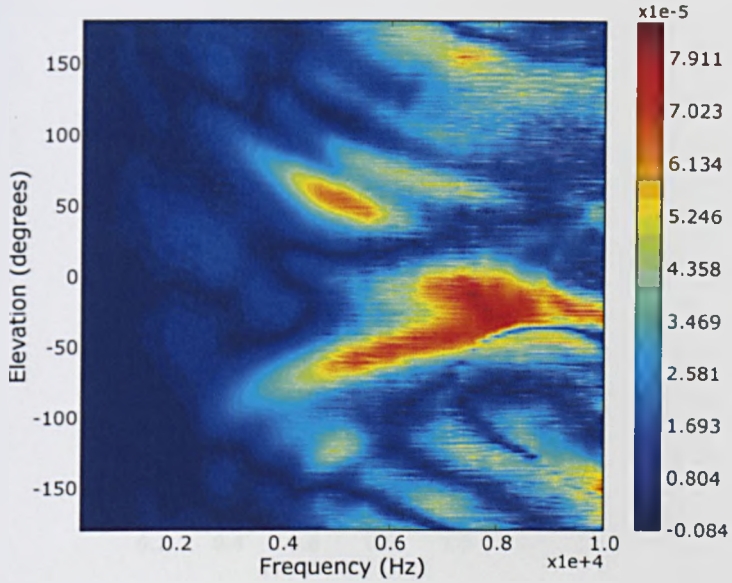


Figure B.8: Absolute rate of spectral change with elevation at the ipsilateral ear around the 0.3 ms confusion ring, up to 10 kHz, in magnitude units (relative to a unit source placed close to the ear canal) per degree.

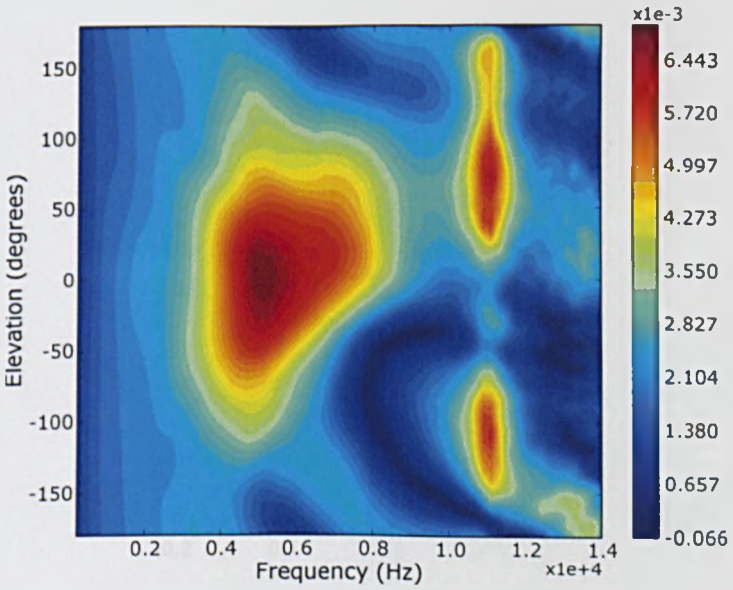


Figure B.9: Spectral variations with elevation at the ipsilateral ear around the 0.5 ms confusion ring, up to 14 kHz, in magnitude units (relative to a unit source placed close to the ear canal).

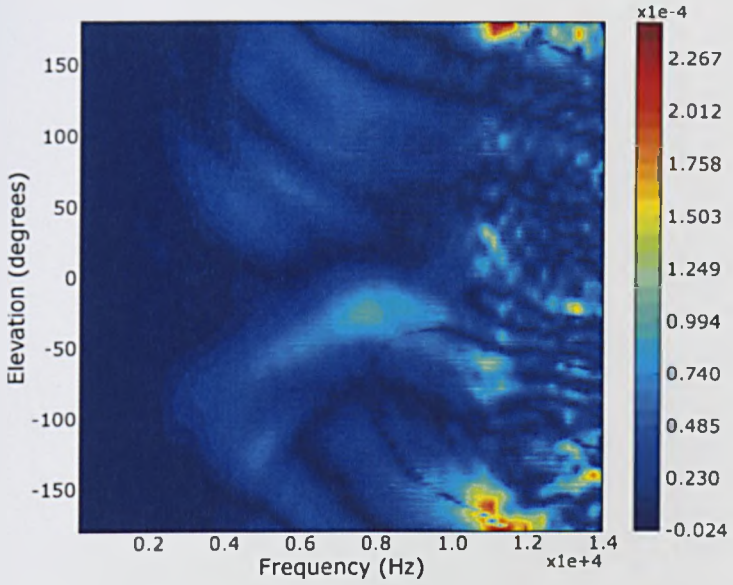


Figure B.10: Absolute rate of spectral change with elevation at the ipsilateral ear around the 0.5 ms confusion ring, up to 14 kHz, in magnitude units (relative to a unit source placed close to the ear canal) per degree.

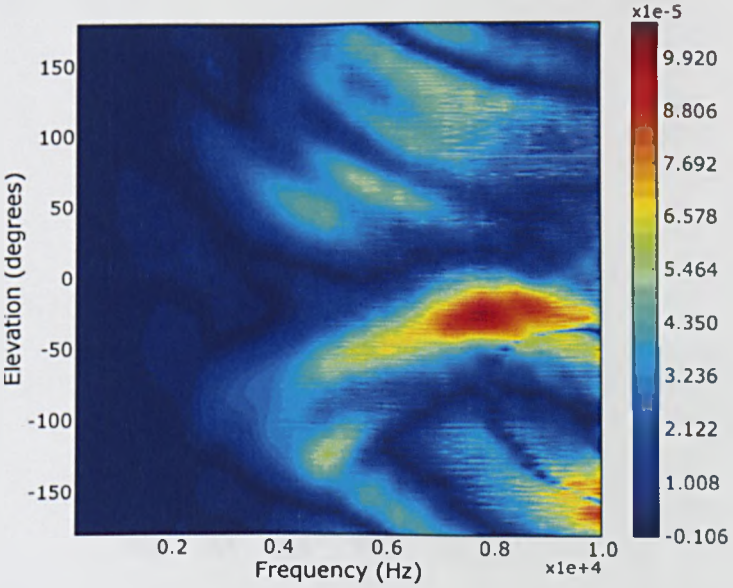


Figure B.11: Absolute rate of spectral change with elevation at the ipsilateral ear around the 0.5 ms confusion ring, up to 10 kHz, in magnitude units (relative to a unit source placed close to the ear canal) per degree.

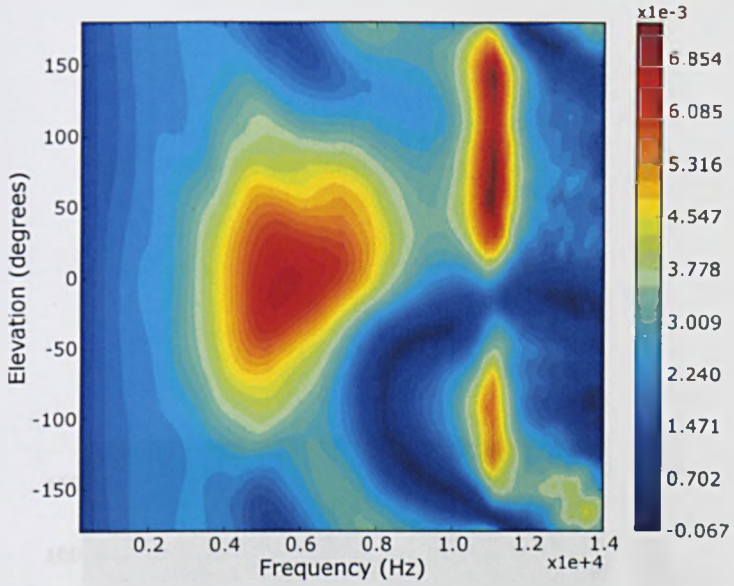


Figure B.12: Spectral variations with elevation at the ipsilateral ear around the 0.6 ms confusion ring, up to 14 kHz, in magnitude units (relative to a unit source placed close to the ear canal).

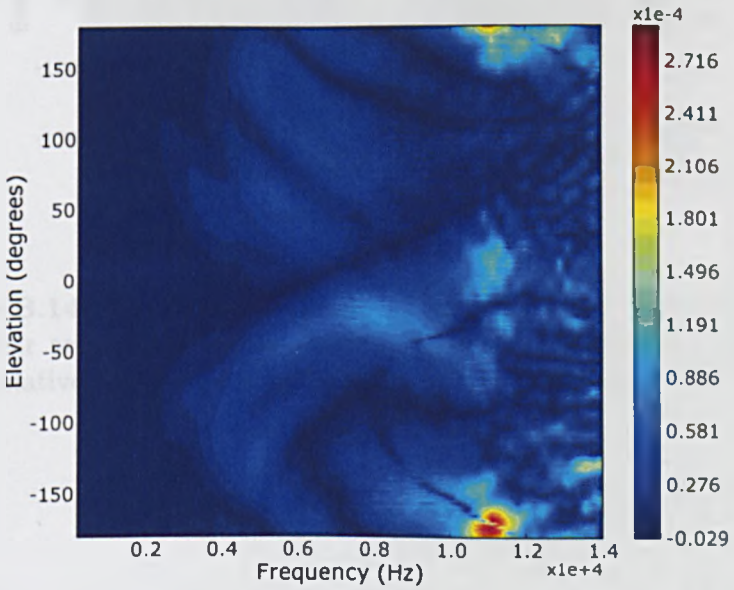


Figure B.13: Absolute rate of spectral change with elevation at the ipsilateral ear around the 0.6 ms confusion ring, up to 14 kHz, in magnitude units (relative to a unit source placed close to the ear canal) per degree.

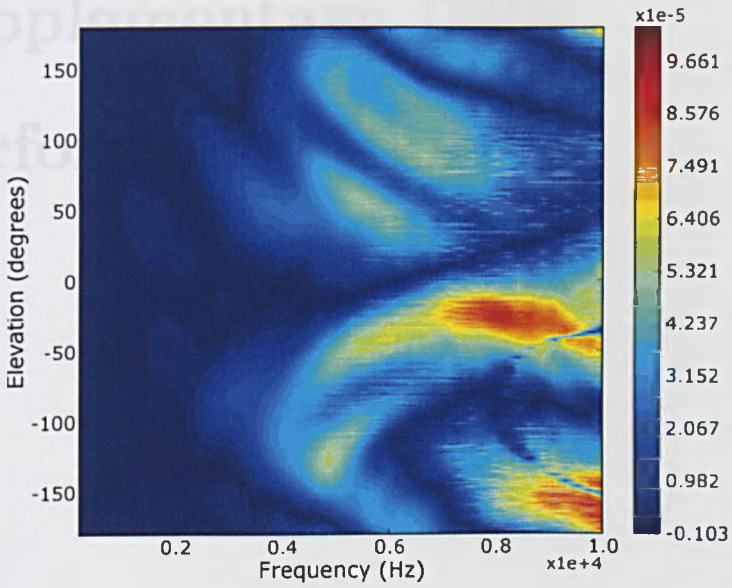
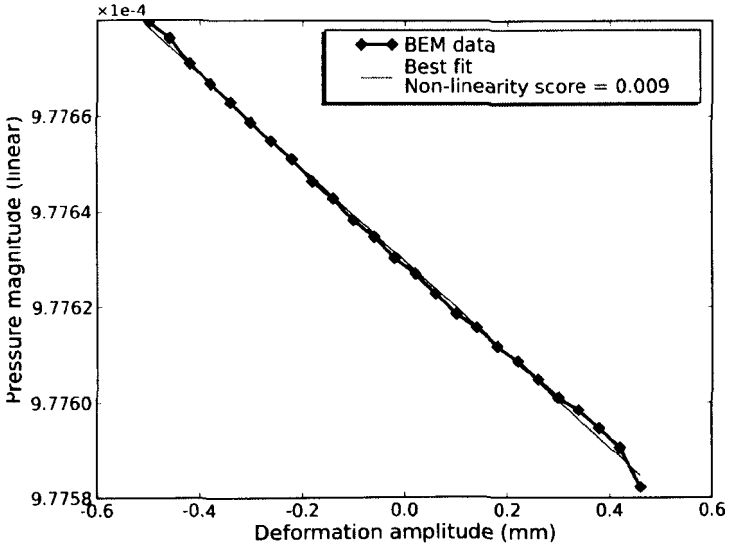


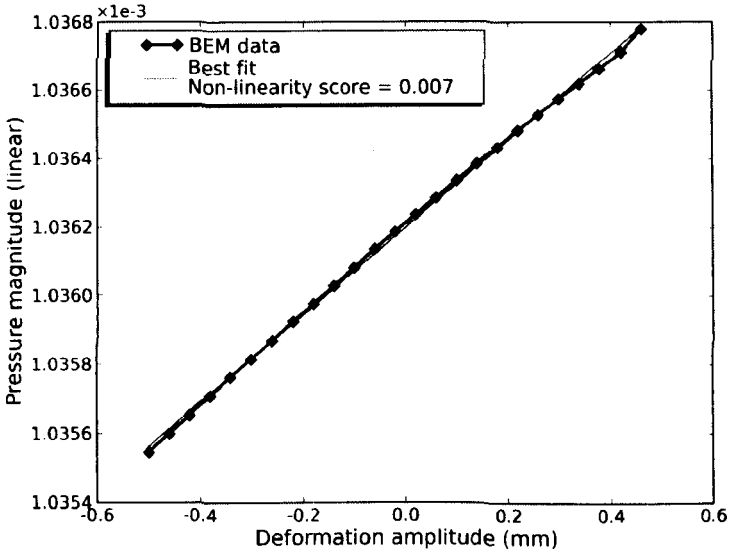
Figure B.14: Absolute rate of spectral change with elevation at the ipsilateral ear around the 0.6 ms confusion ring, up to 10 kHz, in magnitude units (relative to a unit source placed close to the ear canal) per degree.

Appendix C

Supplementary DPS Performance Plots

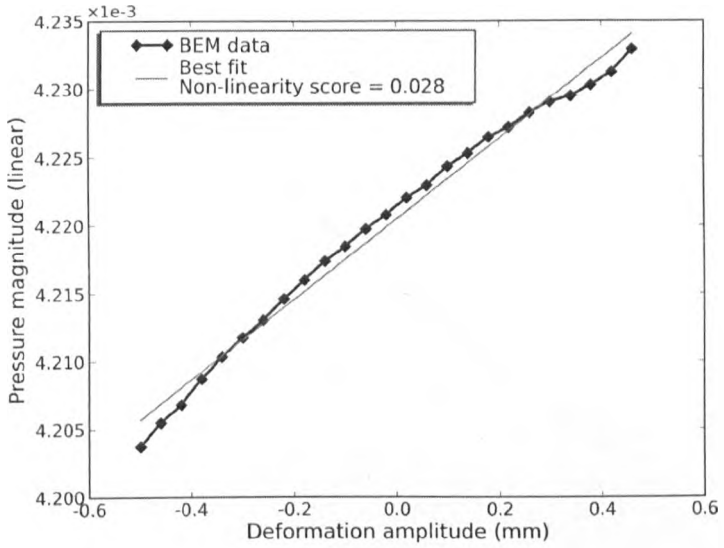


(a)

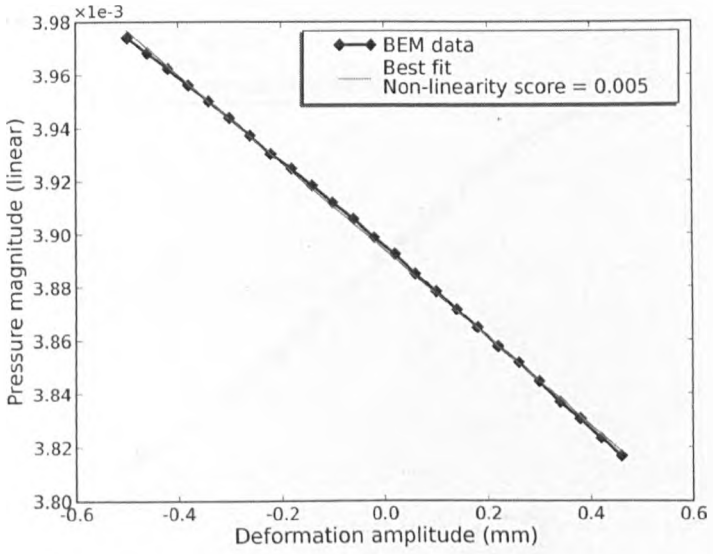


(b)

Figure C.1: Deformation peak amplitude vs pressure plot for the $\cos 2\pi(2x/S + 2y/P)$ elliptic surface harmonic deformation at 200 Hz, for the front (a) and left (b).

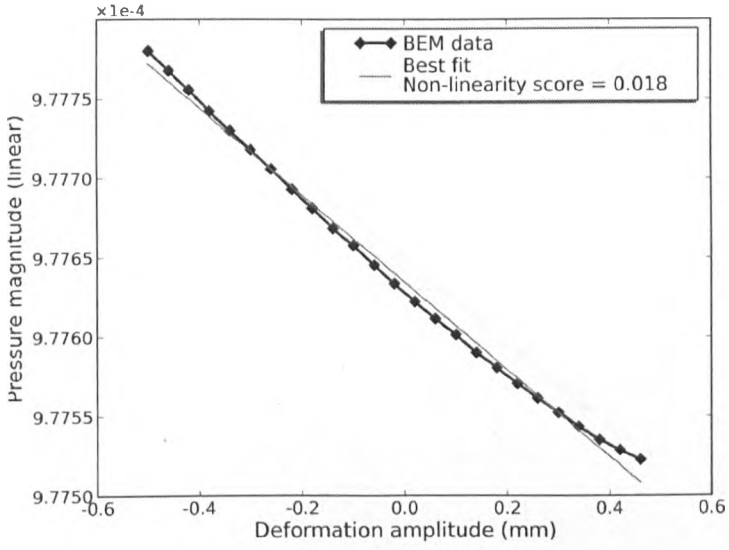


(a)

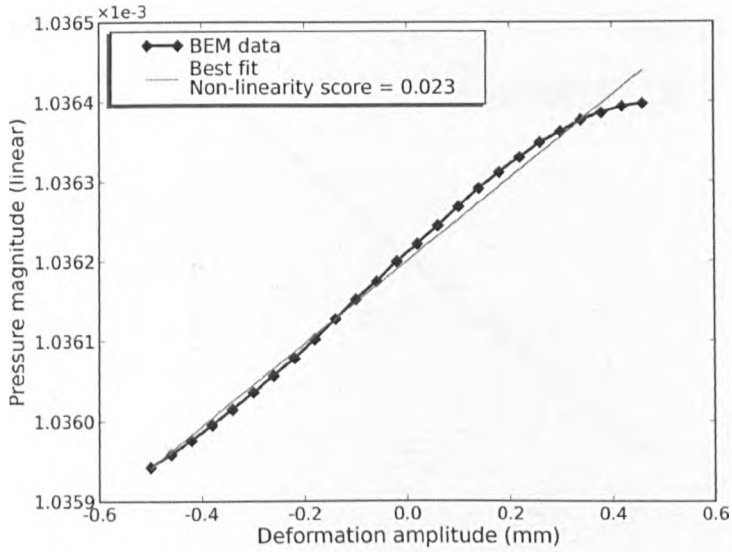


(b)

Figure C.2: Deformation peak amplitude vs pressure plot for the $\cos 2\pi(2x/S + 2y/P)$ elliptic surface harmonic deformation at 5 kHz, for the front (a) and left (b).

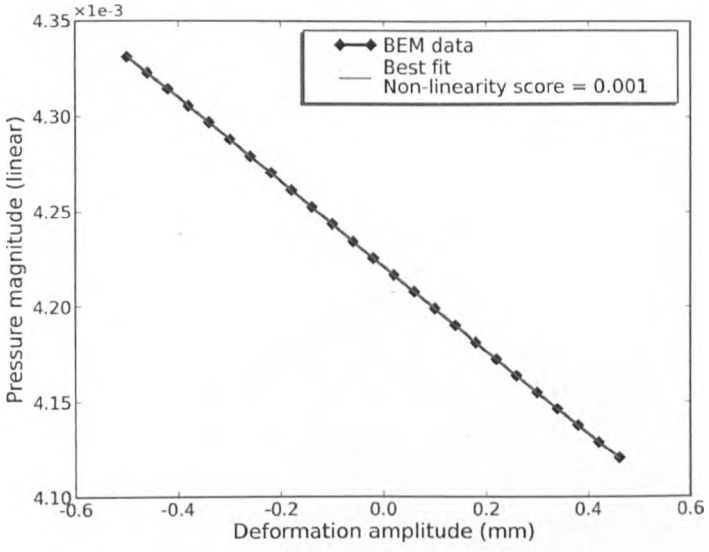


(a)

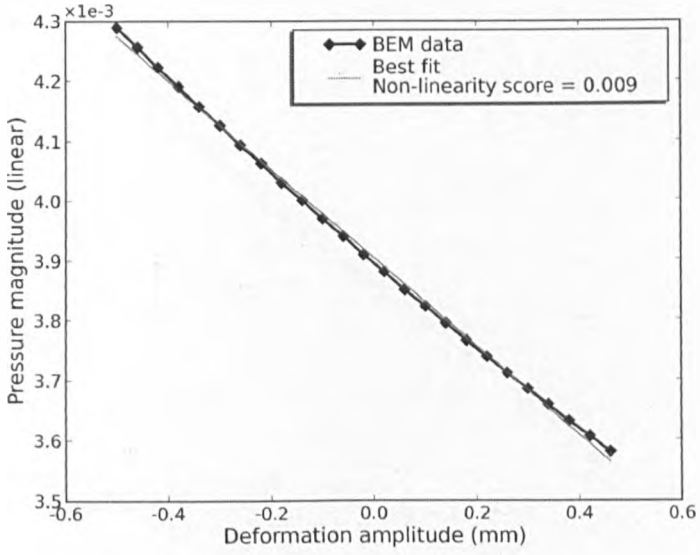


(b)

Figure C.3: Deformation peak amplitude vs pressure plot for the $\cos 2\pi(2x/S + 15y/P)$ elliptic surface harmonic deformation at 200 Hz, for the front (a) and left (b).

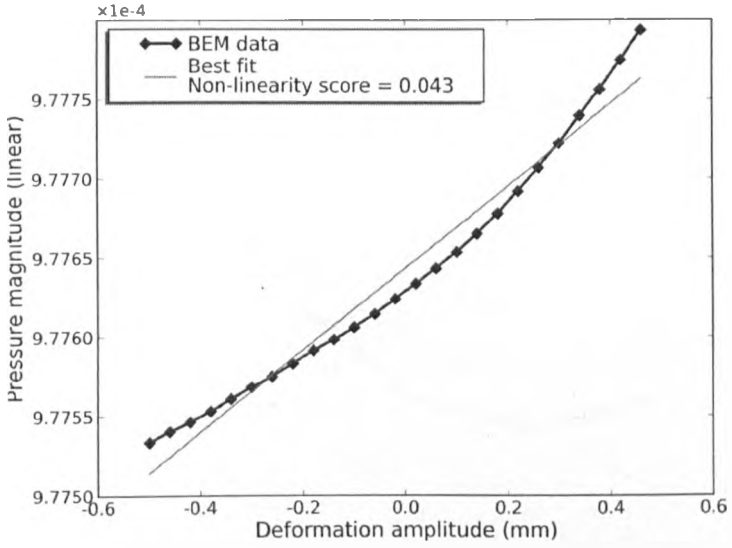


(a)

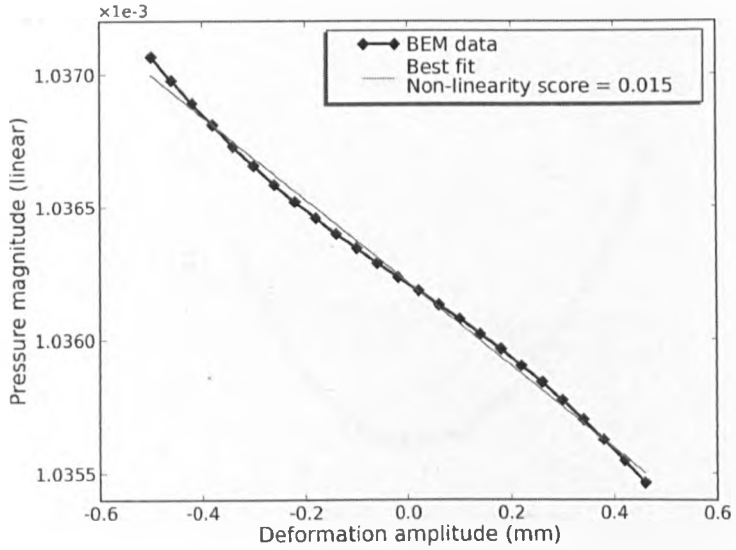


(b)

Figure C.4: Deformation peak amplitude vs pressure plot for the $\cos 2\pi(2x/S + 15y/P)$ elliptic surface harmonic deformation at 5 kHz, for the front (a) and left (b).

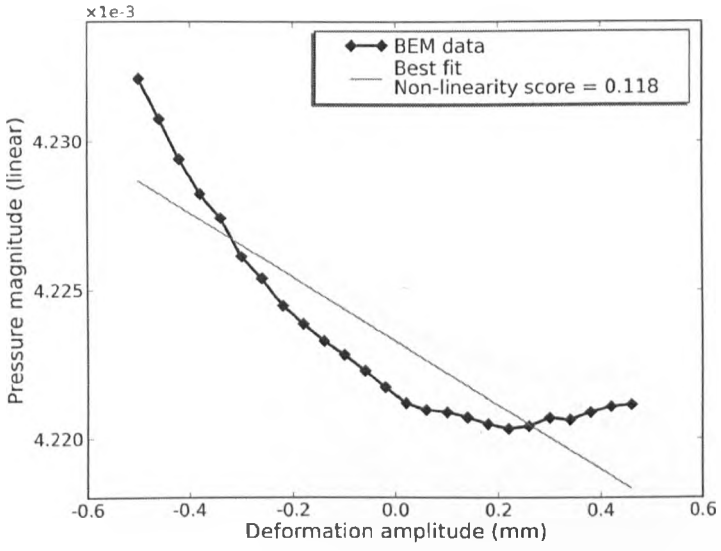


(a)

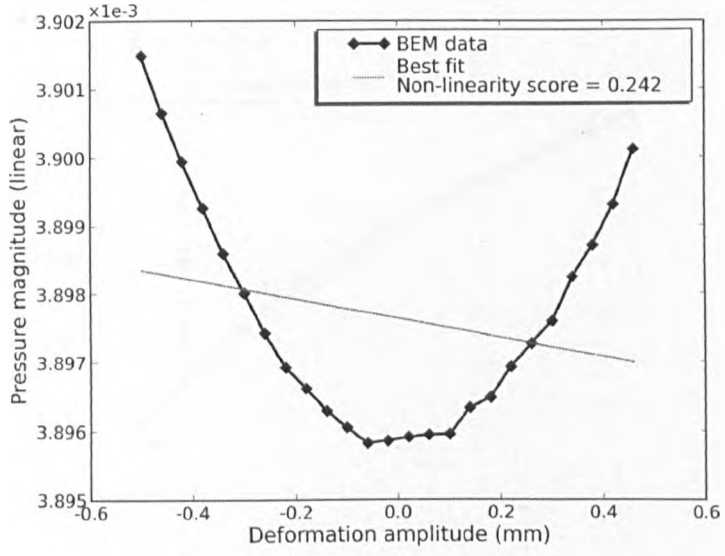


(b)

Figure C.5: Deformation peak amplitude vs pressure plot for the $\cos 2\pi(15x/S + 2y/P)$ elliptic surface harmonic deformation at 200 Hz, for the front (a) and left (b).

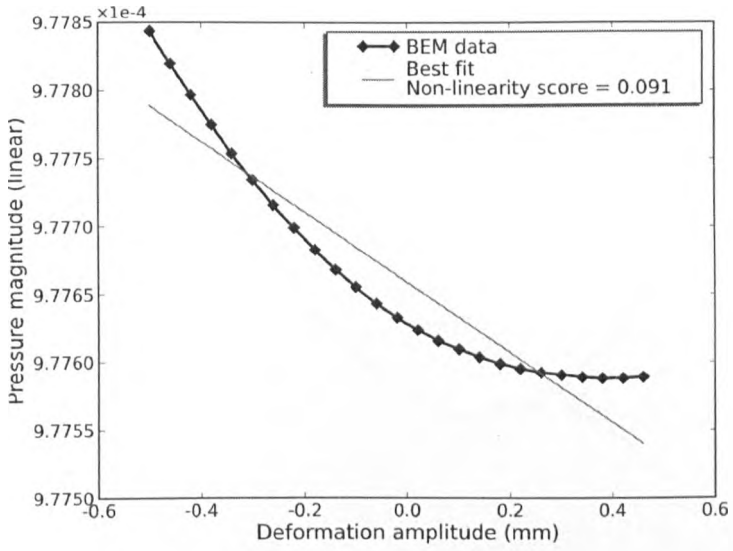


(a)

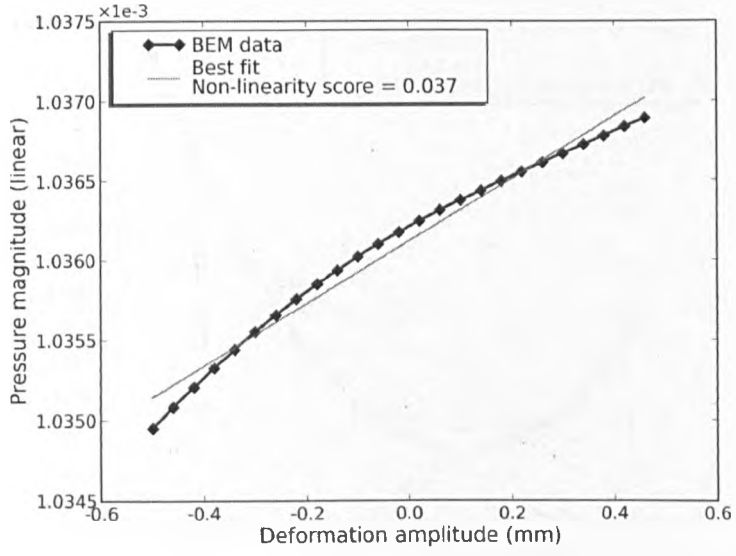


(b)

Figure C.6: Deformation peak amplitude vs pressure plot for the $\cos 2\pi(15x/S + 2y/P)$ elliptic surface harmonic deformation at 5 kHz, for the front (a) and left (b).

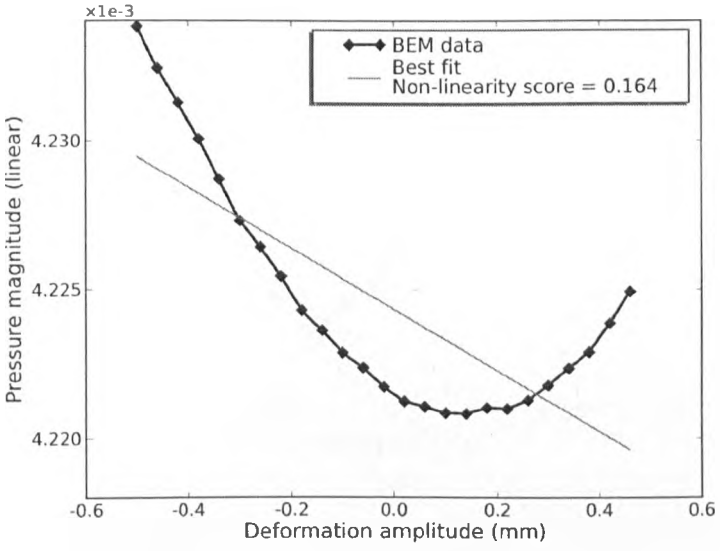


(a)

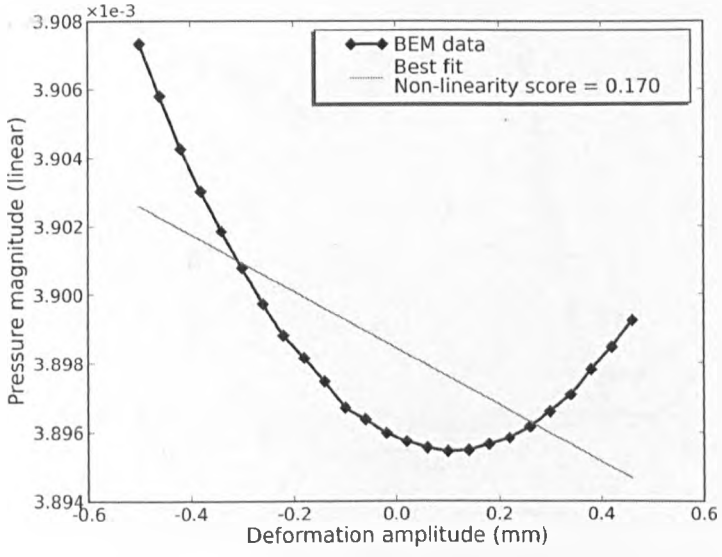


(b)

Figure C.7: Deformation peak amplitude vs pressure plot for the $\cos 2\pi(15x/S + 15y/P)$ elliptic surface harmonic deformation at 200 Hz, for the front (a) and left (b).

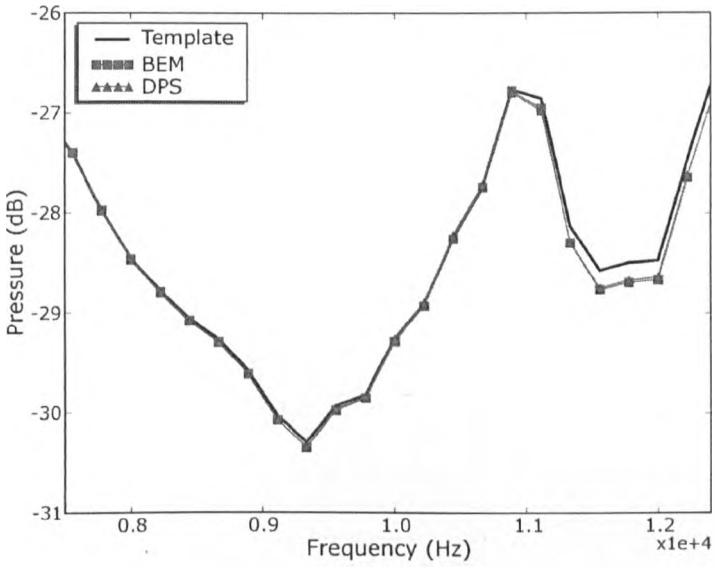


(a)

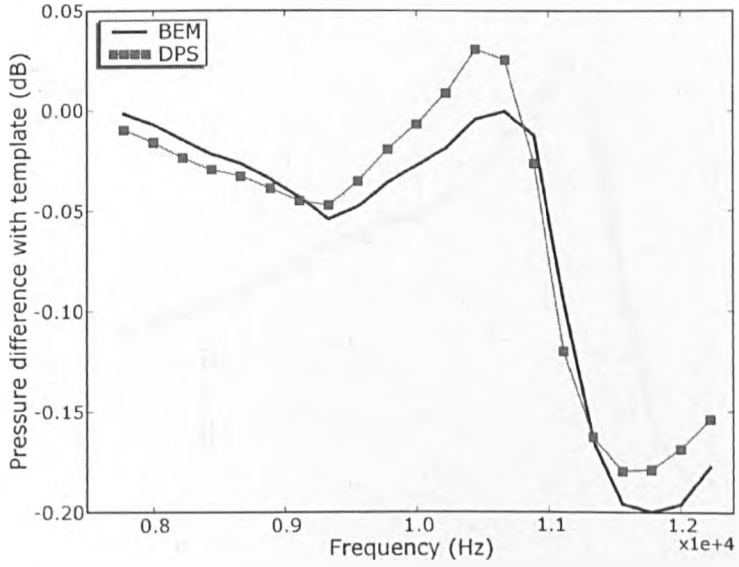


(b)

Figure C.8: Deformation peak amplitude vs pressure plot for the $\cos 2\pi(15x/S + 15y/P)$ elliptic surface harmonic deformation at 5 kHz, for the front (a) and left (b).

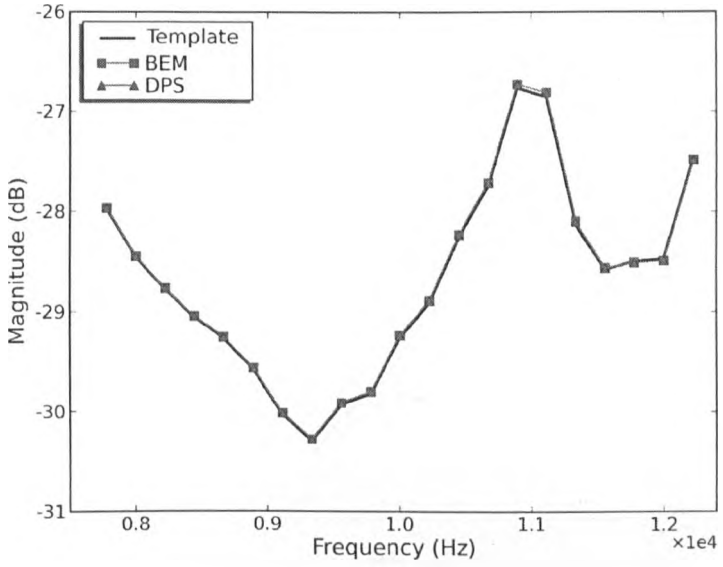


(a)

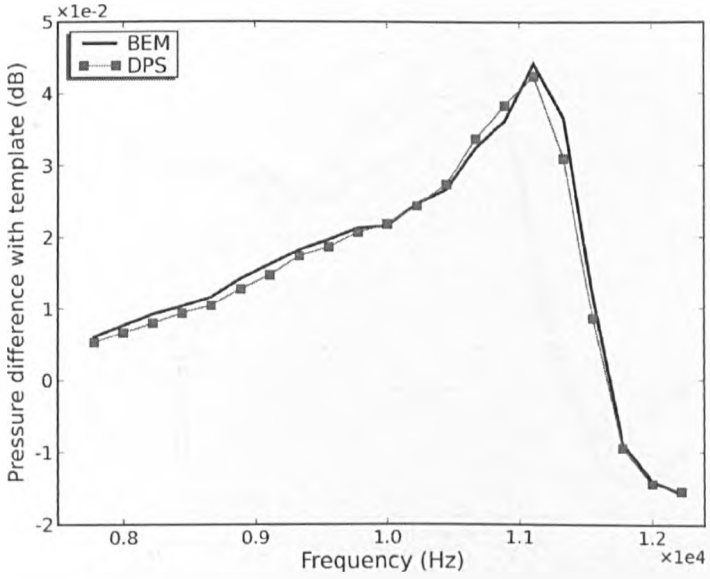


(b)

Figure C.9: (a) The pressure at the entrance of the right ear canal of the template head generated by a frontal source is plotted along with pressures generated by a summation of all ESHDs obeying $0 \leq u < 10$ and $0 \leq v < 10$, as calculated using BEM simulation and DPS estimation. ESHD peak amplitude is 3×10^{-3} mm so that the total deformation does not exceed 0.3 mm. (b) The pressure change generated by the deformation, calculated using BEM and DPS.

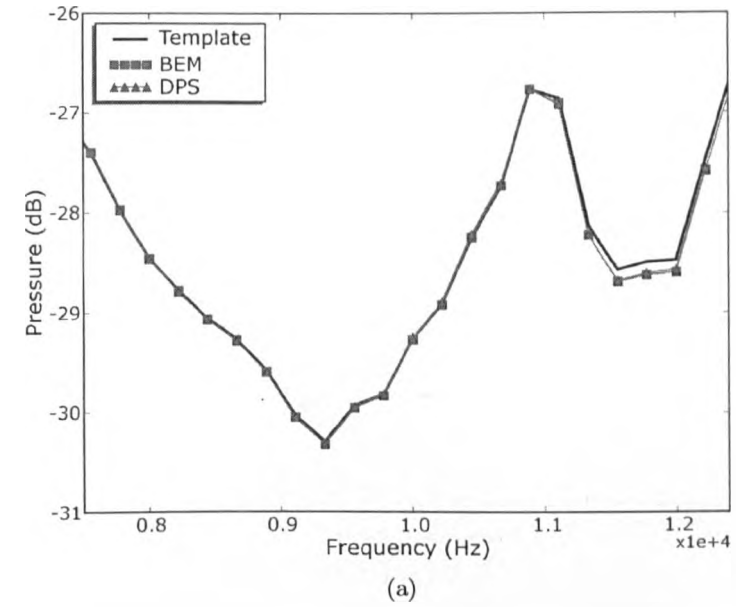


(a)

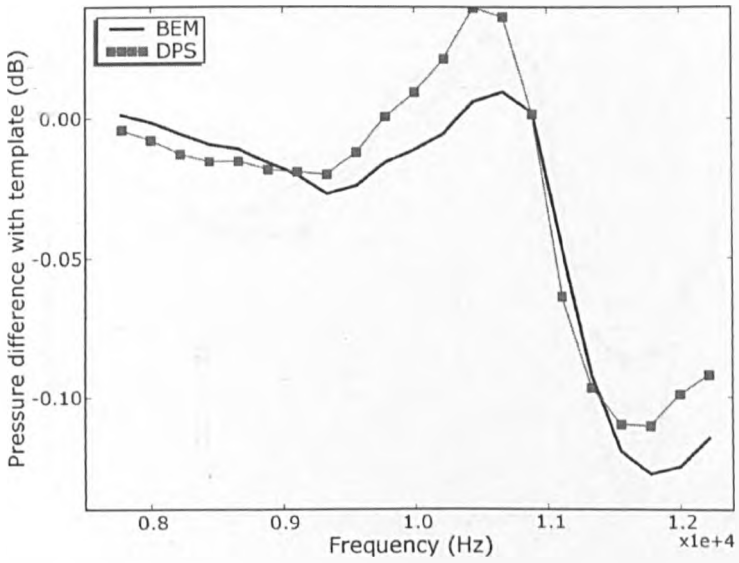


(b)

Figure C.10: (a) The pressure at the entrance of the right ear canal of the template head generated by a frontal source is plotted along with pressures generated by a summation of all ESHDs obeying $0 \leq u < 10$ and $0 \leq v < 10$, as calculated using BEM simulation and DPS estimation. ESHD peak amplitude is attributed randomly, with a uniform probability distribution over the $\pm 3 \times 10^{-3}$ mm amplitude range. (b) The pressure change generated by the deformation, calculated using BEM and DPS.

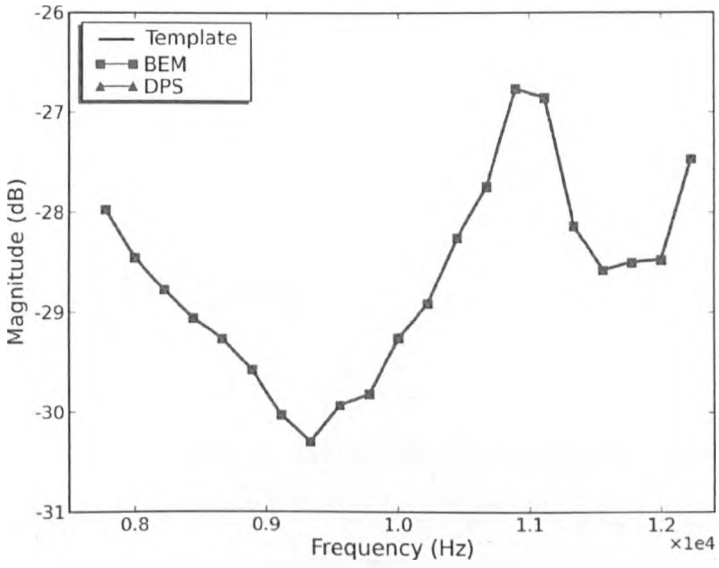


(a)

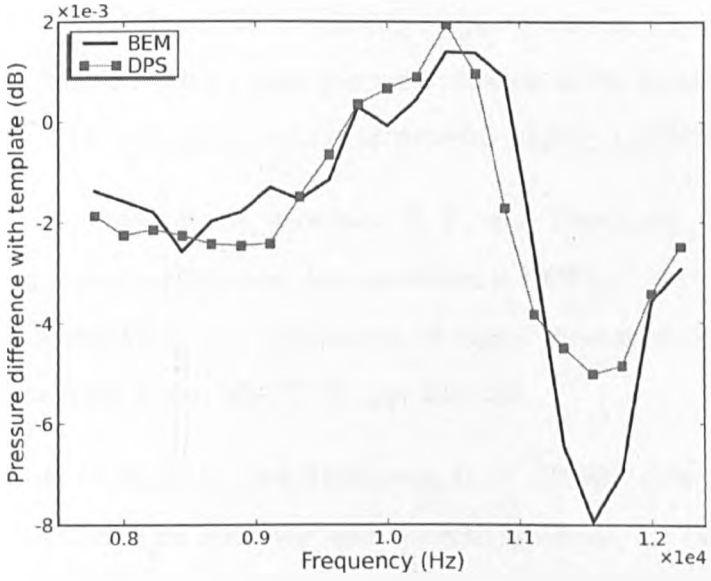


(b)

Figure C.11: (a) The pressure at the entrance of the right ear canal of the template head generated by a frontal source is plotted along with pressures generated by a summation of all ESHDs obeying $0 \leq u < 15$ and $0 \leq v < 15$, as calculated using BEM simulation and DPS estimation. ESHD peak amplitude is 6.67×10^{-4} mm so that the total deformation does not exceed 0.3 mm. (b) The pressure change generated by the deformation, calculated using BEM and DPS.



(a)



(b)

Figure C.12: (a) The pressure at the entrance of the right ear canal of the template head generated by a frontal source is plotted along with pressures generated by a summation of all ESHDs obeying $0 \leq u < 15$ and $0 \leq v < 15$, as calculated using BEM simulation and DPS estimation. ESHD peak amplitude is attributed randomly, with a uniform probability distribution over the $\pm 6.67 \times 10^{-4}$ mm amplitude range. (b) The pressure change generated by the deformation, calculated using BEM and DPS.

Bibliography

Algazi, V. R., Avendano, C., and Duda, R. O. (**2001**a), "Estimation of a spherical head model from anthropometry," *Journal of the Audio Engineering Society* **49**(6), pp. 472–478.

Algazi, V. R. and Duda, R. O. (**2002**), "Approximating the head-related transfer function using simple geometric models of the head and torso," *Journal of the Acoustical Society of America* **112**(5), pp. 2053–2064.

Algazi, V. R., Duda, R. O., Morrison, R. P., and Thompson, D. (**2001**b), "Structural composition and decomposition of HRTFs," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New Paltz, NY, USA), pp. 103–106.

Algazi, V. R., Duda, R. O., and Thompson, D. M. (**2002**), "The use of head-and-torso models for improved spatial sound synthesis," in *Proceedings of the AES 113th Convention* (Los Angeles, CA, USA), p. 5712.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**c), "The CIPIC HRTF database," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New Paltz, NY, USA), pp. 99–103.

Asano, F. and Sone, T. (**1990**), "Role of spectral cues in median plane

localization," *Journal of the Acoustical Society of America* **88**(1), pp. 159–168.

Avendano, C., Algazi, V. R., and Duda, R. O. (**1999**), "A head-and-torso model for low-frequency binaural elevation effects," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New Paltz, NY, USA), pp. 179–182.

Batteau, D. W. (**1967**), "The role of the pinna in human localization," in *Proceedings of the Royal Society of London, Series B, Biological Sciences*, vol. 168, pp. 158–180.

Berhrend, O., Dickson, B., Clarke, E., Jin, C., and Carlile, S. (**2003**), "Virtual acoustic stimulation in the inferior colliculus of the guinea pig: Validation and applications," in *Proceedings of the Australian Neuroscience Society* (Adelaide, SA, Australia), vol. 14, p. 268.

Berhrend, O., Dickson, B., Clarke, E., Jin, C., and Carlile, S. (**2004**), "Neural responses to free field and virtual acoustic stimulation in the inferior colliculus of the guinea pig," *Journal of Neurophysiology* **92**, pp. 3014–3029.

Best, V. A., Carlile, S., Jin, C., and Van Schaik, A. (**2005**), "The role of high frequencies in speech localization," *Journal of the Acoustical Society of America* **118**(1), pp. 353–363.

Blauert, J. (**1970**), "Sound localization in the median plane (Frequency function of sound localization in median plane measured psychoacoustically at both ears with narrow band signals)," *Acustica* **22**(4), pp. 205–213.

Blauert, J. (**1997**), *Spatial Hearing: The Psychophysics of Human Sound*

Localization, Revised Edition (The MIT Press, Massachusetts Institute of Technology).

Bloom, P. J. (**1977**), "Determination of monaural sensitivity changes due to the pinna by use of minimum-audible-field measurements in the lateral vertical plane," *Journal of the Acoustical Society of America* **61**(3), pp. 820–828.

Borgiotti, G. V. (**1990**), "The power radiated by a vibrating body in an acoustic field and its determination from boundary measurements," *Journal of the Acoustical Society of America* **88**, pp. 1152–1159.

Branoner, F., Zhivkov, Z., Ziehm, U., and Berhrend, O. (**January 2005**), "Neural responses to water surface waves in the midbrain of the African clawed frog," in *Proceedings of the Australian Neuroscience Society* (Perth, WA, Australia), vol. 96, p. 173.

Brebbia, C. A. and Dominguez, J. (**1992**), *Boundary Elements: An Introductory Course* (WITPRESS Computational Mechanics Publications).

Bronkhorst, A. W. (**1995**), "Localization of real and virtual sound sources," *Journal of the Acoustical Society of America* **98**(5), pp. 2542–2553.

Bronkhorst, A. W. and Houtgast, T. (**1999**), "Auditory distance perception in rooms," *Nature* **397**(6719), pp. 517–520.

Brown, C. P. and Duda, R. O. (**1997**), "An efficient HRTF model for 3-D sound," in *Proceedings of the Workshop on Applications of Signal Processing to Audio and Acoustics* (New Paltz, NY, USA).

Brown, C. P. and Duda, R. O. (**1998**), "A structural model for binaural sound synthesis," *IEEE Transactions on Speech and Audio processing* **6**(5), pp. 476–488.

Burlinghame, J. A. and Butler, R. A. (**1998**), "The effects of attenuation of frequency segments on binaural localization of sound," *Perception & Psychophysics* **60**, pp. 1374–1383.

Butler, R. A. (**1986**), "The bandwidth effect on monaural and binaural localization." *Hearing Research* **21**(1), pp. 67–73.

Butler, R. A. (**1987**), "An analysis of the monaural displacement of sounds in space," *Perception & Psychophysics* **41**, pp. 1–7.

Butler, R. A. and Belendiuk, K. (**1977**), "Spectral cues utilised in the localization of sound in the median sagittal plane," *Journal of the Acoustical Society of America* **61**(5), pp. 1264–1269.

Butler, R. A. and Flannery, R. (**1980**), "The spatial attributes of stimulus frequency and their role in monaural localization of sound in the horizontal plane," *Perception & Psychophysics* **28**, pp. 449–457.

Butler, R. A. and Helwig, C. C. (**1983**), "The spatial attributes of stimulus frequency in the median sagittal plane and their role in sound localization," *American Journal of Otolaryngology* **4**, pp. 165–173.

Butler, R. A. and Musicant, A. D. (**1993**), "Binaural localization: Influence of stimulus frequency and the linkage to covert peak areas." *Hearing Research* **67**(1–2), pp. 220–229.

Carlile, S. and Pralong, D. (**1994**), "The location-dependent nature of perceptually salient features of the human head-related transfer function," *Journal of the Acoustical Society of America* **95**(6), pp. 3445–3459.

Carlile, S. and Pralong, D. (**1995**), "Recording the head related transfer functions (HRTF): the essential link in high fidelity three dimensional

audio displays," in *Proceedings of the 3rd International meeting of the TTCP Subgroup U* (Department of Defence, Australia).

Chandler, D. W. and Grantham, D. W. (**1992**), "Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity," *Journal of the Acoustical Society of America* **91**(3), pp. 1624–1636.

Chase, S. M. and Young, E. D. (**2005**), "Limited segregation of different types of sound localization information among classes of units in the inferior colliculus," *Journal of Neuroscience* **25**(33), pp. 7575–7585.

Chase, S. M. and Young, E. D. (**2008**), "Cues for sound localization are encoded in multiple aspects of spike trains in the inferior colliculus," *Journal of Neurophysiology* **99**(4), pp. 1672–1682.

Chen, J., Van Veen, B. D., and Hecox, K. E. (**1995**), "A spatial feature extraction and regularization model for the head-related transfer function," *Journal of the Acoustical Society of America* **97**(1), pp. 439–452.

Cheong, H. B. (**2001**), "Double Fourier series on a sphere: Applications to elliptic and vorticity equations," *Journal of Computational Physics* **157**(1), pp. 327–349.

Damaske, P. and Wagener, B. (**1969**), "Richtungshörversuche über einen nachgebildeten Kopt," *Acustica* **21**, pp. 30–35.

Dreyer, A. and Delgutte, B. (**2006**), "Phase Locking of Auditory-Nerve Fibers to the Envelopes of High-Frequency Sounds: Implications for Sound Localization," *Journal of Neurophysiology* **96**(5), pp. 2327–2341.

Duda, R. O. and Martens, W. L. (**1998**), "Range-dependence of the response

of a spherical head model," *Journal of the Acoustical Society of America* **104**(5), pp. 3048–3058.

Everitt, B. S. and Dunn, G. (**2001**), *Applied Multivariate Data Analysis, Second Edition* (Arnold).

Gardner, M. B. and Gardner, R. S. (**1974**), "Problem of localization in the median plane: Effect of pinna cavity occlusion," *Journal of the Acoustical Society of America* **53**, pp. 400–408.

Genuit, K. (**1984**), "Ein Modell zur Beschreibung von Auflenohrübertragungseigenschaften," Ph.D. thesis, Rheinisch-Westfälischen Technichen Hoch-schule, Aachen, Germany.

Glasberg, B. R. and Moore, B. C. J. (**1990**), "Derivation of auditory filter shapes from notched-noise data," *Hearing Research* **44**, pp. 99–122.

Golay, M. (**1961**), "Complementary series," in *IRE Transactions on Information Theory*, vol. 7, pp. 82–87.

Grantham, D. W. (**1986**), "Detection and discrimination of simulated motion of auditory targets in the horizontal plane," *Journal of the Acoustical Society of America* **79**(6), pp. 1939–1949.

Greenwood, D. D. (**1990**), "A cochlear frequency-position function for several species – 29 years later," *Journal of the Acoustical Society of America* **87**(6), pp. 2592–2605.

Grindlay, G. and Vasilescu, M. A. O. (**2007**), "A multilinear (tensor) framework for HRTF analysis and synthesis," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (Honolulu, HI, USA), vol. 1, pp. 161–164.

Gupta, N., Barreto, A., and Coudhury, M. (**2004**), "Modeling Head-Related Transfer Functions Based on Pinna Anthropometry," in *Proceedings of the Second LACCEI International Latin American and Caribbean Conference for Engineering and Technology* (Miami, FL, USA).

Harris, J. D. and Sergeant, R. L. (**1971**), "Monaural/binaural minimum audible angles for a moving sound source," *Journal of Speech and Hearing Research* **14**, pp. 618–629.

Hebrank, J. and Wright, D. (**1974**), "Spectral cues used in the localization of sound sources on the median plane," *Journal of the Acoustical Society of America* **56**(6), pp. 1829–1834.

Henning, G. B. (**1974**), "Detectability of interaural delay in high-frequency complex waveforms," *Journal of the Acoustical Society of America* **55**, pp. 84–90.

Hetherington, C. T. (**2004**), "HRTF Estimation by Shape Parameterisation of the Human Head and Pinnae," Ph.D. thesis, University of York, UK.

Hetherington, C. T. and Tew, A. I. (**2003**), "Parameterizing human pinna shape for the estimation of head-related transfer functions," in *AES 114th Convention, Preprint 5753* (Amsterdam, Netherlands).

Hetherington, C. T., Tew, A. I., and Tao, Y. (**2003**), "Three-dimensional elliptic fourier methods for the parameterization of human pinna shape," in *IEEE International Conference on Acoustics, Speech and Signal Processing, Preprint V-612* (Hong-Kong, China).

Hofman, P. M. and Opstal, J. (**2003**), "Binaural weighting of pinna cues in human sound Binaural weighting of pinna cues in human sound localization," *Experimental Brain Research* **148**, pp. 458–470.

Humanski, R. A. and Butler, R. A. (**1988**), "The contribution of the near and far ear toward localization of sound in the sagittal plane," *Journal of the Acoustical Society of America* **83**(6), pp. 2300–2310.

Jin, C. (**2001**), "Spectral analysis and resolving spatial ambiguities in human sound localization," Ph.D. thesis, School of Electrical and Information Engineering, University of Sydney, NSW, Australia.

Jin, C., Corderoy, A., Carlile, S., and Van Schaik, A. (**2004**), "Contrasting monaural and interaural spectral cues for human sound localization," *Journal of the Acoustical Society of America* **115**(6), pp. 3124–3141.

Jin, C., Leong, P., Leung, J., Corderoy, A., and Carlile, S. (**2000**), "Enabling individualized virtual auditory space using morphological measurements," in *Proceeding of the First IEEE Pacific-Rim Conference on Multimedia (International Symposium on Multimedia Information)* (Sydney, NSW, Australia), pp. 235–238.

Johnson, R. A. and Wichern, D. W. (**2002**), *Applied Multivariate Statistical Analysis, Fifth Edition* (Pearson Education International).

Jolliffe, I. T. (**2002**), *Principal Component Analysis, Second Edition* (Springer).

Kahana, Y. and Nelson, P. A. (**2005**), "Numerical modelling of the spatial acoustic response of the human pinna," *Journal of Sound and Vibration* **292**, pp. 148–178.

Kahana, Y. and Nelson, P. A. (**2007**), "Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models," *Journal of Sound and Vibration* **300**, pp. 552–579.

Kapralos, B. and Mekuz, N. (**2007**), "Application of dimensionality reduction techniques to HRTFs for interactive virtual environments," in *Proceedings of the International Conference on Advances in Computer Entertainment Technology* (Salzburg, Austria), pp. 256–257.

Kapralos, B., Mekuz, N., Kopinska, A., and Khattak, S. (**2008**), "Dimensionality reduced HRTFs: a comparative study," in *Proceedings of the International Conference on Advances in Computer Entertainment Technology* (Yokohama, Japan), pp. 59–62.

Katz, B. F. G. (**1998**), "Measurement and calculation of individual head-related transfer functions using a boundary element model including the measurement and effect of skin and hair impedance," Ph.D. thesis, Pennsylvania State University, PA, USA.

Katz, B. F. G. (**2001**a), "Boundary element method calculation of individual head related transfer function. I. Rigid model calculation," *Journal of the Acoustical Society of America* **110**(5), pp. 2440–2448.

Katz, B. F. G. (**2001**b), "Boundary element method calculation of individual head related transfer function. II. Impedance effects and comparison to real measurements," *Journal of the Acoustical Society of America* **110**(5), pp. 2449–2455.

Katz, B. F. G. and Bergault, D. R. (**2007**), "Round robin comparison of HRTF measurement systems: preliminary results," in *Proceedings of the 19th International Congress on Acoustics* (Madrid, Spain).

King, A. J. and Oldfiled, S. R. (**1997**), "The impact of signal bandwidth on auditory localization: Implications for the design of three-dimensional audio displays," *Human Factors* **39**, pp. 287–295.

Kistler, D. J. and Wightman, F. L. (**1992**), "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *Journal of the Acoustical Society of America* **91**(3), pp. 1637–1647.

Kuhl, F. P. and Giardina, C. R. (**1982**), "Elliptic Fourier features of a closed contour," *Computer Graphics and Image Processing* **18**, pp. 236–258.

Kuhn, G. F. (**1977**), "Model for the interaural time differences in the azimuthal plane," *Journal of the Acoustical Society of America* **62**(1), pp. 157–167.

Langendijk, E. H. A. and Bronkhorst, A. W. (**2002**), "Contribution of spectral cues to human sound localization," *Journal of the Acoustical Society of America* **112**(4), pp. 1583–1596.

le Minor, J. M. and Schmittbuhl, M. (**1999**), "Importance of elliptic Fourier methods for morphometry of complex outlines: application to the distal human femur," *Surgical and Radiologic Anatomy* **21**(6), pp. 387–391.

Leung, J. and Carlile, S. (**2004**), "Spatial resolution along a cone of confusion," in *Proceedings of the ARO Midwinter Meeting Conference* (Daytona Beach, FL, USA), vol. 27, p. 208.

Lin, H. W., Tai, C. L., and Wang, G. J. (**2004**), "A mesh reconstruction algorithm driven by an intrinsic property of a point cloud," *Computer-Aided Design* **36**, pp. 1–9.

Loève, M. (**1978**), *Probability theory. Vol. II, 4^{th} ed., Graduate Texts in Mathematics*, vol. 46 (Springer-Verlag).

Lopez-Poveda, E. A. and Meddis, R. (**1996**), "A physical model of sound

diffraction and reflections in the human concha," *Journal of the Acoustical Society of America* **100**, pp. 3248–3259.

Lyon, R. F. (**1982**), "A computational model of filtering, detection and compression in the cochlea," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (Paris, France).

Macpherson, E. A. (**1996**), "Effects of source spectrum irregularity and uncertainty on sound localization," *Journal of the Acoustical Society of America* **99**, p. 2515.

Macpherson, E. A. and Middlebrooks, J. C. (**2002**), "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," *Journal of the Acoustical Society of America* **111**(5), pp. 2219–2236.

Macpherson, E. A. and Middlebrooks, J. C. (**2003**), "Vertical-plane sound localization probed with ripple-spectrum noise," *Journal of the Acoustical Society of America* **114**(16), pp. 430–445.

Manley, G. A., Popper, A. N., and Fay, R. R. (**2004**), *Evolution of the Vertebrate Auditory System* (Springer-Verlag, New York).

Martens, W. L. (**1987**), "Principal components analysis and resynthesis of spectral cues to perceived direction," in *Proceedings of International Computer Music Conference* (San Francisco, CA, USA), pp. 274–281.

Martin, R. L., McAnally, K. I., Watt, T., and Flanagan, P. (**2006**), "The Effect of Spectral Variation on Sound Localisation," Tech. Rep. DSTO-RR-0308, Defence Science and Technology Organisation, Australia.

Martin, R. L., Paterson, M., and McAnally, K. I. (**2004**), "Utility of monaural spectral cues is enhanced in the presence of cues to sound-source lateral

angle," *Journal of the Association for Research in Otolaryngology* **5**, pp. 80–89.

McLaughlin, M., Chabwine, J. N., Van der Heijden, M., and Joris, P. X. (**2008**), "Comparison of bandwidths in the inferior colliculus and the auditory nerve. II: Measurement using a temporally manipulated stimulus," *Journal of Neurophysiology* **100**(4), pp. 2312–2327.

Mehrgardt, S. and Mellert, V. (**1977**), "Transformation characteristics of the external human ear," *Journal of the Acoustical Society of America* **61**(6), pp. 1567–1576.

Middlebrooks, J. C. (**1992**), "Narrow-band sound localization related to external ear acoustics," *Journal of the Acoustical Society of America* **92**, pp. 2607–2624.

Middlebrooks, J. C. and Green, D. M. (**1992**), "Observations on a principal components analysis of head-related transfer functions," *Journal of the Acoustical Society of America* **92**, pp. 597–599.

Mills, A. W. (**1958**), "On the Minimum Audible Angle," *Journal of the Acoustical Society of America* **30**(4), pp. 237–246.

Moore, B. C. J. (**1989**), *An Introduction to the Psychology of Hearing* (Academic, London).

Moore, B. C. J. and Glasberg, B. R. (**1987**), "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns," *Hearing Research* **28**, pp. 209–225.

Moore, B. C. J., Oldfiled, S. R., and Gary, J. D. (**1989**), "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," *Journal of the Acoustical Society of America* **85**(2), pp. 820–836.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1990**), "Auditory filter shapes at low center frequencies," *Journal of the Acoustical Society of America* **88**(1), pp. 132–140.

Morimoto, M., Iida, K., and Itoh, M. (**2003**a), "Upper hemisphere sound localization using head related transfer functions in the median plane and interaural differences," *Acoustical Science and Technology* **24**, pp. 267–275.

Morimoto, M., Motoki, Y., Kazuhiro, I., and Motokuni, I. (**2001**), "The contribution of two ears to the perception of vertical angle in sagittal planes," *Journal of the Acoustical Society of America* **109**(4), pp. 1596–1603.

Morimoto, M., Motoki, Y., Kazuhiro, I., and Motokuni, I. (**2003**b), "The role of low frequency components in median plane localization," *Acoustical Science and Technology* **24**(2), pp. 76–82.

Morrongiello, B. A. and Rocca, P. T. (**1987**), "Infants' localization of sounds in the median vertical plane: Estimates of minimum audible angle," *Journal of Experimental Child Psychology* **43**, pp. 181–193.

Morse, P. M. and Ingard, K. U. (**1968**), *Theoretical Acoustics* (Princeton University Press, New Jersey).

Musicant, A. D. and Butler, A. R. (**1984**), "The influence of pinnae-based spectral cues on sound localization," *Journal of the Acoustical Society of America* **75**(4), pp. 1195–1200.

Oldfield, S. R. and Parker, S. P. A. (**1984**a), "Acuity of sound localisation: a topography of auditory space. I. Normal hearing conditions," *Perception* **13**(5), pp. 581–600.

Oldfield, S. R. and Parker, S. P. A. (**1984**b), "Acuity of sound localization: A topography of auditory space. II. Pinna cues absent," *Perception* **13**, pp. 601–617.

Oldfield, S. R. and Parker, S. P. A. (**1986**), "Acuity of sound localization: A topography of auditory space. III. Monaural hearing conditions," *Perception* **15**, pp. 61–81.

Otani, M. and Ise, S. (**2006**), "A fast calculation system specialized for the HRTF based on the BEM," *Journal of the Acoustical Society of America* **119**, pp. 2589–2598.

Park, K. S. and Lee, N. S. (**1987**), "A three-dimensional Fourier descriptor for human body representation/reconstruction from serial cross sections," *Computers and Biomedical Research* **20**, pp. 125–140.

Patterson, R. D. (**1976**), "Auditory filter shapes derived with noise stimuli," *Journal of the Acoustical Society of America* **59**, pp. 640–654.

Patterson, R. D. and Moore, B. C. (**1986**), *Frequency Selectivity in Hearing* (Academic Press Ltd., London), chap. Auditory filters and excitation patterns as representations of frequency resolution, pp. 123–177.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1992**), *Auditory physiology and perception* (Pergamon, Oxford), chap. Complex sounds and auditory images, pp. 429–446.

Perrott, D. R. and Marlborough, K. (**1989**), "Minimum audible movement angle: Marking the end points of the path traveled by a moving sound source," *Journal of the Acoustical Society of America* **85**(4).

Perrott, D. R. and Musicant, A. D. (**1977**), "Minimum auditory movement

angle: Binaural localization of moving sound sources," *Journal of the Acoustical Society of America* **62**, pp. 1463–1466.

Perrott, D. R. and Saberi, K. (**1990**), "Minimum audible angle threshold for sources varying in both elevation and azimuth," *Journal of the Acoustical Society of America* **87**, pp. 1728–1731.

Perrott, D. R. and Tucker, J. (**1988**), "Minimum audible movement angle as a function of signal frequency and the velocity of the source," *Journal of the Acoustical Society of America* **83**(4).

Photiadis, D. M. (**1995**), "The relationship of singular value decomposition to wave-vector filtering in sound radiation problems," *Journal of the Acoustical Society of America* **98**, pp. 1570–1580.

Popper, A. N. and Fay, R. R. (**2005**), *Sound Source Localization* (Springer).

Pralong, D. and Carlile, S. (**1994**), "Measuring the human head-related transfer functions: A novel method construction and calibration of a miniature 'in ear' recording system," *Journal of the Acoustical Society of America* **95**, pp. 3435–3444.

Rakerd, B., Ryan, Z., Macaulay, E., and Hartmann, W. M. (**2008**), "The acoustical bright spot and the erroneous localization of tones by human listeners," in *Proceedings of the Acoustics'08 Conference* (Paris, France), p. 3294.

Rayleigh, L. (**1907**), "On our perception of sound direction," *Philosophical Magazine* **13**, pp. 214–232.

Reiss, A. J. and Young, E. D. (**2005**), "Spectral Edge Sensitivity in Neural Circuits of the Dorsal Cochlear Nucleus," *Journal of Neuroscience* **25**(14), pp. 3680–3691.

Rodgers, C. A. (**1981**), "Pinna Transformations and Sound Reproduction," *Journal of the Audio Engineering Society* **29**(4), pp. 226–234.

Rodriguez, S. and Ramirez, M. (**2005**), "Linear relationships between spectral characteristics and anthropometry of the external ear," in *Proceedings of the International Conference on Auditory Display* (Limerick, Ireland), pp. 336–339.

Roffler, S. K. and Butler, R. A. (**1968**a), "Factors that influence the localization of sound in the vertical plane," *Journal of the Acoustical Society of America* **43**, pp. 1255–1259.

Roffler, S. K. and Butler, R. A. (**1968**b), "Localization of tonal stimuli in the vertical plane," *Journal of the Acoustical Society of America* **43**, pp. 1260–1266.

Rogers, M. E. and Butler, R. A. (**1992**), "The linkage between stimulus frequency and covert peak areas as it relates to monaural localization," *Perception & Psychophysics* **52**(5), pp. 536–546.

Saberi, K., Dostal, L., Sadralodabai, T., and Perrott, D. R. (**1991**), "Minimum audible angles for horizontal, vertical and oblique orientations: Lateral and dorsal planes," *Acustica* **75**.

Saberi, K. and Perrott, D. R. (**1990**), "Minimum audible movement angles as a function of sound source trajectory," *Journal of the Acoustical Society of America* **88**(6), pp. 2639–2644.

Schoolmaster, M., Kopčo, N., and Shinn-Cunningham, B. G. (**2001**), "Auditory distance perception in fixed and varying simulated acoustic environments," *Journal of the Acoustical Society of America* **115**(5), pp. 2459–2459.

Searle, C. L., Braida, L. D., Cuddy, D. R., and Davis, M. F. (**1975**), "Binaural pinna disparity: Another auditory localization cue," *Journal of the Acoustical Society of America* **57**, pp. 448–455.

Shailer, M. J., Moore, R. W., Glasberg, B. R., and Watson, N. (**1990**), "Auditory filter shapes at 8 and 10 kHz," *Journal of the Acoustical Society of America* **88**(1), pp. 131–148.

Shaw, E. A. G. (**1974**), "Transformation of the sound pressure level from the free field to the eardrum in the horizontal plane," *Journal of the Acoustical Society of America* **56**, pp. 1848–1861.

Shaw, E. A. G. (**1997**), *Binaural and Spatial Hearing in Real and Virtual Environments* (Lawrence Erlbaum Associates, Mahwah, NJ, USA), chap. Acoustical features of the human external ear, pp. 25–48.

Shaw, E. A. G. and Teranishi, R. (**1968**), "Sound pressure generated in an external ear replica and real human ears by a nearby point source," *Journal of the Acoustical Society of America* **44**, pp. 240–249.

Shinn-Cunningham, B. G. (**2000**), "Distance cues for virtual auditory space," in *Proceedings of the First IEEE Pacific-Rim Conference on Multimedia* (Sydney, NSW, Australia), pp. 227–230.

Shinn-Cunningham, B. G., Durlach, N. I., and Held, R. M. (**1998a**), "Adapting to supernormal auditory localization cues. I. Bias and resolution," *Journal of the Acoustical Society of America* **103**, pp. 3656–3666.

Shinn-Cunningham, B. G., Durlach, N. I., and Held, R. M. (**1998b**), "Adapting to supernormal auditory localization cues. II. Constraints on adaptation of mean response," *Journal of the Acoustical Society of America* **103**, pp. 3667–3676.

Shinn-Cunningham, B. G., Streeter, T., and Gyss, J. (**2001**), "Perceptual Plasticity in spatial auditory displays," in *Proceedings of the International Conference on Auditory Display* (Helsinki, Finland), pp. 181–184.

Skottun, B. C., Shackleton, T. M., Arnott, R. H., and Palmer, A. R. (**2001**), "The ability of inferior colliculus neurons to signal differences in interaural delay," in *Proceedings of the (US) National Academy of Sciences*, vol. 98, pp. 14050–14054.

Slaney, M. (**1988**), "Lyon's Cochlear Model," Tech. rep., Apple Computer, Inc, `http://rvl4.ecn.purdue.edu/~malcolm/apple/tr13/LyonsCochlea.pdf`.

Slaney, M. (**1993**), "An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank," Tech. rep., Apple Computer, Inc, `https://eprints.kfupm.edu.sa/24374/1/24374.pdf`.

Stevens, S. S. and Newman, E. B. (**1936**), "The Localization of Actual Sources of Sound," *American Journal of Psychology* **48**(2), pp. 297–306.

Stevenson, R. G., Lestrel, P. E., and Read, D. W. (**1987**), "A method for analyzing complex two-dimensional shapes — elliptic Fourier functions," *American Journal of Physical Anthropology* **72**(2), pp. 257–258.

Tao, Y., Tew, A. I., and Porter, S. J. (**2003**a), "The differential pressure synthesis method for efficient acoustic pressure estimation," *Journal of the Audio Engineering Society* **51**(7), pp. 647–656.

Tao, Y., Tew, A. I., and Porter, S. J. (**2003**b), "A study on head-shape simplification using spherical harmonics for HRTF computation at low frequencies," *Journal of the Audio Engineering Society* **51**(9), pp. 799–805.

Teranishi, R. and Shaw, E. A. G. (**1968**), "External-ear acoustic models with simple geometry," *Journal of the Acoustical Society of America* **44**(1), p. 257.

Tollin, D. J., Koka, K., and J., T. J. (**2008**), "Interaural Level Difference Discrimination Thresholds for Single Neurons in the Lateral Superior Olive," *Journal of Neurophysiology* **28**(19), pp. 4848–4860.

Wallach, H. (**1939**), "On sound localization," *Journal of the Acoustical Society of America* **10**, pp. 270–274.

Walsh, T., Demkowicz, L., and Charles, R. (**2003**), "Boundary element modeling of the external human auditory system," *Journal of the Acoustical Society of America* **115**(3), pp. 1033–1043.

Watkins, A. J. (**1978**), "Psychoacoustical aspects of synthesized vertical locale cues," *Journal of the Acoustical Society of America* **63**(4), pp. 1152–1165.

Weinrich, S. G. (**1984**), "Sound field calculations around the human head," Tech. Rep. 37, The Acoustics Laboratory, Technical University of Denmark.

Wenzel, E. M., Kistler, D. J., and Wightman, F. L. (**1993**), "Localization Using Non-individualized Head-Related Transfer Functions," *Journal of the Acoustical Society of America* **94**, pp. 111–123.

Wettschurek, R. G. (**1973**), "Die absoluten Unterschiedswellen der Richtungswahrnehmung in der Medianebene beim natürlichen Hören sowie beim Hören über ein Kunstkopf-Übertragungssystem," *Acustica* **28**, pp. 197–208.

Wightman, F. L. and Kistler, D. J. (**1989**a), "Headphone simulation of free-field listening. I: Stimulus synthesis," *Journal of the Acoustical Society of America* **85**, pp. 858–867.

Wightman, F. L. and Kistler, D. J. (**1989**b), "Headphone simulation of free-field listening II : Psychophysical validation," *Journal of the Acoustical Society of America* **85**, pp. 868–878.

Wightman, F. L. and Kistler, D. J. (**1992**), "The dominant role of low-frequency interaural time differences in sound localization," *Journal of the Acoustical Society of America* **91**(3), pp. 1648–1661.

Wightman, F. L. and Kistler, D. J. (**1997**), "Monaural sound localization revisited," *Journal of the Acoustical Society of America* **101**, pp. 1050–1063.

Wu, M. and Sheu, H. (**2001**), "Representation of 3D surfaces by two-variable Fourier descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(8), pp. 858–836.

Zahorik, P. (**2001**), "Localization accuracy in 3-d sound displays: The role of visual-feedback training," in *Proceedings of the Advanced Displays and Interactive Displays Federal Laboratory Consortium* (College Park, MD, USA), pp. 17–22.

Zahorik, P., Bangayan, P., Sundareswaran, V., Wang, K., and Tam, C. (**2006**), "Perceptual recalibration in human sound localization: Learning to remediate front-back reversals," *Journal of the Acoustical Society of America* **120**(1), pp. 343–359.

Zotkin, D. N., Duraiswami, R., Grassi, E., and Gumerov, N. A. (**2006**),

"Fast head related transfer function measurement via reciprocity," *Journal of the Acoustical Society of America* **120**(4), pp. 2202–2215.

Zotkin, D. N., Hwang, J., Duraiswami, R., and Davis, L. S. (**2003**), "HRTF personalization using anthropomentric measurements," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New Paltz, NY, USA), pp. 157–160.