Minimum Audible Movement Angles for Discriminating Upward from Downward Trajectories of Smooth Virtual Source Motion within a Sagittal Plane

David H. Benson

Music Technology Area

Schulich School of Music

McGill University

Montreal,Quebec

Submitted August 2007

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree Master of Arts

DEDICATION

To my brother, Chris,

my father, Jim,

and the memory of my mother, Phyllis.

ACKNOWLEDGEMENTS

Thanks are due to several individuals for their help with writing this thesis. Louise Frenette and Alexandre Bouenard translated the abstract, Andrea Bellissimo proofread, and Chris Macivor created several diagrams. Thanks are also due to my colleagues in the Schulich School of Music for many hours of stimulating discussion, and to my two supervisors, William L. Martens and Gary P. Scavone. Without the encouragement and guidance of these two wise and dedicated professors this work would have been all but impossible. Finally, thanks are due to the many excellent choirs and vocal ensembles I've had the privilege of singing with over the last few years. To the Canadian Chamber Choir, the MSO Chorus, J.S. Allaire's groups, the Liederwölfe vocal collective and the rest: you make it all worthwhile. Now that this thesis is finished I can stop neglecting you.

ABSTRACT

In virtual auditory display, sound source motion is typically cued through dynamic variations in two types of localization cues: binaural disparity cues and spectral cues. Generally, both types of cues contribute to the perception of sound source motion. For certain spatial trajectories, however, namely those lying on the surfaces of cones of confusion, binaural disparity cues are constant, and motion must be inferred solely on the basis of spectral cue variation. This thesis tests the effectiveness of these spectral variation cues in eliciting motion percepts. A virtual sound source was synthesized that traversed sections of a cone of confusion on a particular sagittal plane. The spatial extent of the source's trajectory was systematically varied to probe directional discrimination thresholds.

ABRÉGÉ

Dans le domaine de la spatialisation, le mouvement de la source sonore est généralement indiqué par des variations dynamiques selon deux types d'indices de localisation : des indices de disparite binaurale et des indices spectraux. En règle générale, les deux types d'indices contribuent à la perception du mouvement de la source sonore. Cela dit, dans le cas de certaines trajectoires spatiales, savoir celles qui reposent sur la surface des cônes de confusion, les indices de disparité binaurale sont constants et le mouvement ne s'induit forcément qu'à partir des variations spectrales. La présente thèse sonde l'efficacité de ces indices de variation spectrale en indiquant les perceptions du mouvement. Une source sonore virtuelle a été synthétisée et chemine sur la surface d'un cône de confusion sur un plan sagittal déterminé. L'étendue spatiale de la trajectoire de la source a été ajustée systématiquement afin de sonder les seuils critiques de discrimination de mobilité directionnelle.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

ABSTRACT iv

ABRÉGÉ v

LIST OF TABLES viii

LIST OF FIGURES ix

1 Introduction 1

1.1 Unanswered questions in spatial hearing 2
1.2 Definitions and operational terms 5
1.2.1 Motion detection vs. motion discrimination 6
1.2.2 Horizontal, median and sagittal planes 6
1.2.3 Spherical coordinate systems 7
1.3 Thesis organization 9

2 Background 10

2.1 Sound localization cues 10
2.1.1 Binaural Disparity cues 11
2.1.2 Spectral cues 14
2.2 Binaural virtual auditory display technologies 16
2.2.1 Efficient HRTF filtering 18
2.2.2 Functional and Structural HRTF models 18
2.2.3 The Multicomponent HRTF model 20
2.3 Minimum audible movement angles in sagittal planes . . . 26

3 Methodology 29

3.1 Stimulus creation 29
3.1.1 Filter cases 30
3.1.2 Source signal 35
3.1.3 Reverberation processing 36
3.1.4 Spatial trajectories of moving stimuli 36
3.1.5 Starting elevations 39
3.2 Adaptive staircase threshold tracking 40

3.2.1 Staircase step size 40
3.3 Structure of experimental sessions 43

4 Results 45

4.1 Data selection 45
4.1.1 The inter-block mean difference 47
4.1.2 Criteria for rejecting session 1 data 47
4.1.3 Problematic subjects 47
4.2 Two-way ANOVA 48

5 Discussion and Conclusion 52

5.1 Phenomenology 52
5.2 Comparison with previous results 53
5.2.1 Data selection 53
5.2.2 Previous results 54
5.2.3 Discussion of previous results 55
5.3 Individual Differences 56
5.3.1 Significance of individual differences 57
5.4 Threshold shift in the multicomponent case 57
5.4.1 Elevation dependence in the multicomponent case . 59
5.5 Conclusions 59
5.6 Future work 60

Appendix A: Principal Components Analysis and the Singular Value Decomposition 61

Appendix B: Certificate of Ethical Acceptability 63

References 66

LIST OF TABLES

Table page

3–1 The structure of each testing session. 43

4–1 Two way ANOVA for starting angles above ear-level 48

4–2 Two way ANOVA for starting angles near ear-level 49

4–3 Two way ANOVA for starting angles below ear-level. 49

LIST OF FIGURES

Figure		page
1–1	Horizontal, median and sagittal planes described in cartesian coordinates.	6
1–2	The interaural-polar spherical coordinate system.	8
2–1	The cone of confusion.	13
2–2	A 'traditional' approach to directional filtering.	22
2–3	The multicomponent model approach to directional filtering.	23
3–1	Measured HRTF angles.	30
3–2	Impulse responses of the measured case HRTF filters.	33
3–3	Impulse responses of the multicomponent HRTF model.	34
3–4	The signal processing structure used in stimulus generation.	37
3–5	The two spatial motion trajectories between which subjects were required to discriminate.	38
3–6	Examples of motion trajectories at different elevations.	41
3–7	An example staircase of subject responses showing changes in step size.	42
4–1	Large and small 'inter-block mean differences'.	46
4–2	Differences between Multicomponent Case thresholds and Measured Case thresholds.	49
4–3	Directional discrimination thresholds.	50

CHAPTER 1
Introduction

Virtual auditory display is the technique of presenting sounds to a listener in such a way that they appear to occupy distinct locations in an imaginary three-dimensional space. These displays are often used in multimodal human-computer interfaces, for instance in 3D video game interfaces where they augment the visual display in order to increase the player's sense of immersion in the virtual environment. More utilitarian applications of these technologies also exist. In architectural acoustics, virtual auditory displays can be used to preview the acoustics of concert halls before they are built. In the aerospace industry, they enable communications personnel to more effectively monitor multiple streams of speech simultaneously [4]. These streams become more intelligible if they are presented from distinct locations in auditory space.[1]

Virtual auditory displays function by digitally simulating the acoustic cues used in human sound localization. As such, they present a concrete application of spatial hearing research results devoted to identifying these cues. Such research is ongoing and specific questions about the nature of localization cues remain unanswered. Particularly underrepresented in the research literature are questions concerning cues used in the perception of moving sound sources.

[1] This phenomenon is formally known as 'spatial unmasking', e.g. [12].

This thesis investigates the perception of sound source motion in a virtual auditory display. Specifically, it asks how far a virtual source must move before a listener can reliably discriminate its direction of motion. An experiment is conducted to determine this just detectable angular distance, known as a Minimum Audible Movement Angle (MAMA), and to assess how the size of this just detectable angle of movement varies under different conditions.

This initial chapter serves to introduce and justify the work performed. It first gives a summary of relevant research results to motivate the specific questions under investigation. The research problems and experimental hypotheses are then stated. Finally, some domain-specific terminology is explained.

1.1 Unanswered questions in spatial hearing

Within the spatial hearing research community, a distinction is commonly made between two types of localization cues. The first type can be best understood in the time-domain and is related to the difference in time of arrival and overall sound pressure level of a wavefront at the two ears. These might in general be termed 'binaural disparity cues', however, the most important of these would be the interaural time delay (ITD). The second type of cues are spectral in nature and are related to the direction-dependent acoustic filtering of the head, upper body, and pinna. These might in general be termed 'spectral cues,' some of which are captured in the monaural HRTF, and others of which exist as interaural spectral differences. The categorical distinction between spectral and temporal cues provides a useful means for understanding the complex transformations underlying binaural HRTFs, and these cues have often been held to play differing roles in supporting human directional hearing: interaural

cues, dominated by the ITD, tend to indicate the lateral angle to a sound source [21], while spectral cues determine the perceived elevation [16], and disambiguate front from rear [3], at least when the listener's head is immobile.[2]

In headphone-based virtual auditory display, holding these cues constant gives rise to spatial auditory images that are stationary. By contrast, the smooth and systematic modulation of these cues can induce percepts of virtual sound source motion. Due to the finite spatial resolution of the auditory system [38], however, thresholds for the detection of cue modulation exist. Above these thresholds, clear correlations can be found between smooth changes in binaural signals and auditory image motion. Below these thresholds, the tracking of source motion is impossible.

Probing these motion discrimination thresholds has been the focus of a large body of research. Most studies in this field have created their experimental stimuli using real, as opposed to virtual, sound sources. These studies typically modulate the position of a loudspeaker by ever decreasing amounts until the direction of its motion can no longer be discerned. Through this process a minimum perceptible angular displacement for the speaker, known as a minimum audible movement angle (MAMA), is uncovered [46].

---

[2] Indeed, research in auditory neuroscience has further justified this distinction between these two types of cues. Distinct neural structures have been found that encode, respectively, time-based interaural cues and monaural spectral cues. Structures directly encoding overall interaural time delay have been long hypothesized [19] and have more recently been observed in some mammals (e.g., [36]). Similarly, other mammalian neural structures that encode notch frequencies and other complex features of the frequency spectrum have also been experimentally verified [60].

By measuring MAMAs using real sound sources, however, such studies have left unasked more subtle questions concerning thresholds for the modulation of ITD and spectral cues in isolation. These studies usually employ source motion trajectories that give rise to concomitant variation in both types of cues, and so it is not always clear which is primarily responsible for motion discrimination near the threshold. A seldom asked question, then, is how effectively source motion can be elicited by either of these cues individually. What minimum change in ITD is required to induce virtual sound source motion *when spectral cues are held constant*, or, conversely, what amount of spectral variation is sufficient *when ITD is held constant*. The latter question is the focus of this thesis.

This study aimed to determine the minimum amount of spectral variation, in the absence of ITD variation, necessary to enable directional discrimination of source trajectories in a virtual auditory display. In effect, the study measured MAMAs for spectrally induced motion of virtual sound sources.

Holding the ITD cues constant and varying only spectral cues was expected to generate source trajectories whose motion was restricted to a single sagittal plane. In this respect the present study differed from most previous investigations of the MAMA. MAMAs on sagittal planes have received little attention, even though those on horizontal planes have been extensively studied. In addition to confirming the results of previous studies, this work strove to investigate the effect on the sagittal plane MAMA of two independent variables:

- the starting elevation of the motion trajectory
- the level of spectral detail in the directional filters

To test the first variable, a moving stimulus was presented at three elevations: below ear-level, near ear-level, and above ear-level. Previously reported results with static stimuli have shown that localization acuity degrades for elevated source positions relative to ear-level positions. Accordingly, it was hypothesized that, at elevated source positions, degradation would also be observed in motion discrimination.

To evaluate the effect of varying levels of spectral detail, stimulus creation employed two different types of directional filtering. In the first case directional filters were based exactly on the measured acoustic response of a human head. The transfer functions of these filters exhibited fine detail in their frequency spectra. This case was termed the 'measured' case.

In the second case spectral detail was reduced by modeling these filters in a low-dimensional subspace. This process was termed 'multicomponent' modeling, and this case, the 'multicomponent case'. Previous studies had shown that removing spectral detail from directional filters in this way had the effect of increasing rates of confusions between frontal and rearward sound source locations [48]. Since front/rear discrimination was known to be degraded when this type of spectral smoothing was performed, it was hypothesized that similar degradation would be observed in motion discrimination.

1.2 Definitions and operational terms

This thesis makes use of several key terms that are either not in standard use, or that would benefit from precise definitions to explain their usage in the present context. In this section, motion detection and motion discrimination are disambiguated, various spatial regions of interest are described, and the inter-aural polar coordinate system is explained.

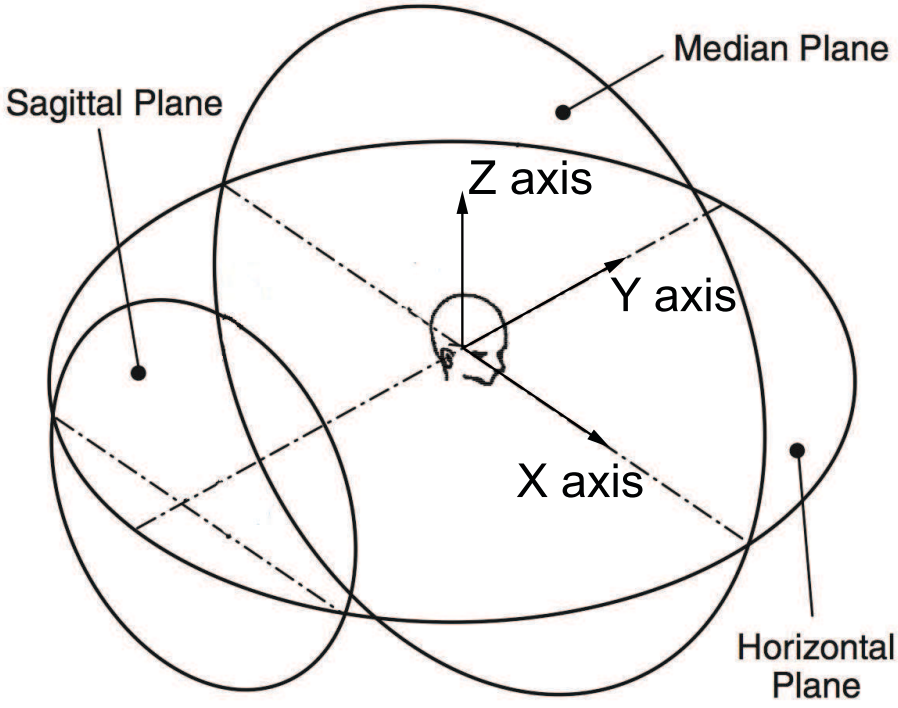

Figure 1–1: The horizontal and median planes and a sagittal plane described in cartesian coordinates. *Figure adapted from [43]*

1.2.1 Motion detection vs. motion discrimination

A distinction must be made between two types of motion thresholds: those for the detection of motion versus those for the discrimination of the direction of motion. We will refer to thresholds for motion *detection* as the limits above which listeners can report that motion has occurred but *cannot consistently report its direction.* By contrast, thresholds for motion *discrimination* indicate the limits above which listeners *can consistently report the direction* in which a source has moved. The present investigation is concerned with these latter motion *discrimination* thresholds.

1.2.2 Horizontal, median and sagittal planes

The horizontal and frontal planes and the set of sagittal planes that includes the median plane are regions of space commonly referenced in the

spatial hearing literature. These are perhaps most easily described using a three-dimensional cartesian coordinate system (Fig. 1–1). The x axis in this system includes the line that connects a listener's nose to a point directly on the back of his head. The y axis includes the line extending through the listener's head and connecting his two ears. This y axis is also known as the inter-aural axis. The x and y axes define a plane parallel to the floor. This is the horizontal plane.

The x and y axes meet at a point in the center of the head. The z axis includes a line that passes through this point and extends up though the top of the head. The y and z axes define the frontal plane, and the x and z axes then define the median sagittal plane, also known simply as the median plane.

While the terms "horizontal plane" and "median plane" refer to two distinct planes defined by $z = 0$ and $y = 0$, respectively, the term "sagittal plane" can refer to the entire set of planes parallel to the median plane. The particular sagittal plane of interest in this thesis is a sagittal plane shifted to the right of the median.

1.2.3 Spherical coordinate systems

More than one spherical coordinate system is used in the spatial hearing literature. Though vertical-polar spherical coordinates have historically been popular, this thesis uses interaural-polar (IP) spherical coordinates (Fig. 1–2). The IP coordinate system is becoming increasingly common in spatial hearing research to describe locations and trajectories in 3D space. It is convenient both because it succinctly describes cones of confusion and because it effectively captures the distinctive roles in localization of binaural disparity and spectral cues [41].

subsection check above

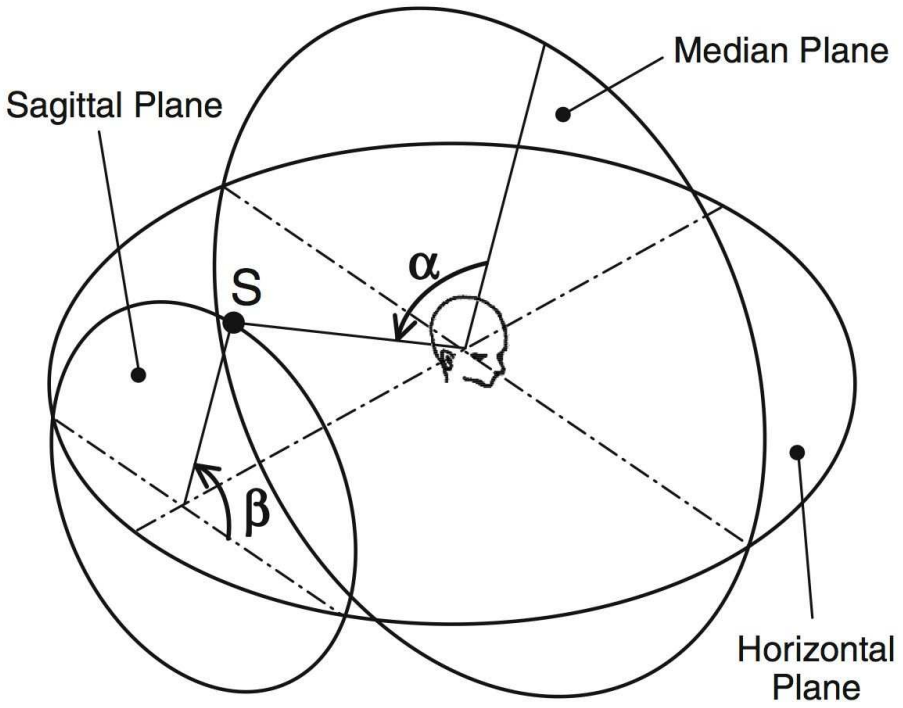

Figure 1–2: The interaural-polar spherical coordinate system. Using IP coordinates, directions are expressed using two angles: a lateral angle α and a rising angle β. The lateral angle α describes the angle between the median plane and the interaural axis. The rising angle β describes the angle around a circle centered on the interaural axis. One advantage of this system is that a cone of confusion can be defined by holding α constant and letting β and the distance from the head vary. At a constant distance from the head, such a cone reduces to a circle, as shown. We dub this a 'circle of confusion'. *Figure taken from [43]*

The interaural-polar coordinate system uses two angles to specify directions in three-dimensional space: a lateral angle and a rising angle. One advantage of the system is that cones of confusion, regions of roughly constant binaural disparity, may be defined by fixing the lateral angle and letting the rising angle and distance vary.

IP coordinates are related to the two types of acoustical localization cues discussed earlier. Specifically, binaural disparity cues are strongly correlated with perceived lateral angle (α), and spectral cues are strongly modulated by rising angle (β). In other words, for a given sagittal plane, binaural disparity cues are thought to specify a particular circle of confusion, and spectral cues are thought to influence the perceived position on that circle.

1.3 Thesis organization

Having clarified key terms and objectives, an overview of the structure of this document will now be given. The second chapter of the work provides a more extensive review of research in spatial hearing and virtual auditory display. The third chapter describes the methodology employed to measure the effect on sagittal plane MAMAs of source elevation and spectral detail of the directional filters. The fourth chapter presents the results of the investigation and the fifth interprets these results and draws conclusions.

CHAPTER 2
Background

This thesis investigates minimum audible movement angles (MAMAs) for spectrally-cued motion in virtual auditory display. As virtual auditory display is an inherently multidisciplinary topic, drawing on research results from both psychoacoustics and digital signal processing (DSP), this chapter will survey relevant results from both of these disciplines. It will begin by discussing acoustical cues used in human sound localization. It will then examine the evolution of technologies that synthesize these cues in order to generate spatial percepts. Finally, it will survey the results of several related studies that have attempted to determine MAMAs for spectrally cued source motion.

2.1 Sound localization cues

In the published research of Japanese psychoacoustician Masayuki Morimoto a categorical distinction is often made between two types of sound localization cues: binaural disparity cues [1] and spectral cues (e.g. [41, 40, 43]). The distinction is useful because the two categories of cues play differing roles in localization. The present discussion will respect Morimoto's distinction and will discuss binaural disparity cues and spectral cues in succession.

[1] Note that the 'binaural disparity' cue referred to here should not be confused with the 'binaural pinna disparity' cue as defined in [51].

2.1.1 Binaural Disparity cues

Binaural disparity cues were first identified by the British physicist Lord Rayleigh near the turn of the 19th century. Rayleigh's pioneering research in spatial hearing produced the venerable duplex theory of sound localization [47]. The duplex theory states that two different types of acoustic cues are used to determine the lateral angle to a sound source: the inter-aural time delay (ITD) and the inter-aural level difference (ILD). Both of these cues make use of differences between the signals at the two ear-drums, and so are termed binaural disparity cues.

The ITD arises because of the difference in path length to the two ears from most sound source locations. This path length difference gives rise to a time delay between the arrival of a wavefront at one ear and the other. This time delay varies from about $0\mu s$, for sound source locations on the median plane, to about $600\mu s$ for locations to the side. This time delay, the ITD, is used by the auditory system to deduce the lateral angle to a sound source.

Rayleigh is also credited with pointing out an ambiguity in the ITD that arises in the localization of pure (sinusoidal) tones. Pure tones lack the initial transient that is characteristic of most natural sounds, and, in the absence of this transient, the ITD effectively becomes a difference in sinusoidal phase between the two ears rather than a difference in time of arrival. This phase difference is a useful cue for low frequency sounds with wavelengths greater than the size of the head (below about 500 Hz), but it is problematic for higher frequency sounds for which a given phase difference can correspond to multiple angular locations.

For these higher frequency sounds, Rayleigh considered the effect of acoustic head shadowing. While low-frequency sound waves can diffract around the head and have similar intensity levels on either side, higher

frequency sound waves cannot, and they are at least partially reflected off one side of the head, arriving at the contralateral ear at a lower intensity than the ipsilateral ear. This head shadowing phenomenon explains the inter-aural level difference (ILD), the second cue in Rayleigh's duplex theory.

Stated fully, the duplex theory asserts that the ITD is used to localize low frequency sounds and the ILD is used to localize high frequency sounds, with the boundary between the two cases occuring at wavelengths on par with the size of the head.

Shortcomings of the duplex theory

The duplex theory can predict localization percepts only under certain specific conditions. Firstly, the theory only applies when ITD and ILD cues are consistent with one another, that is, when broadband signals first arrive loudly at one ear and then quietly at the other. Inconsistent cues, resulting from, say, a quieter signal at the leading ear, would rarely arise in natural listening environments but are relevant since they could be synthesized in a virtual auditory display. Inconsistent cues were explored in recent investigations, such as [21], which showed that when ITD and ILD are in conflict, judgments are usually dominated by the low-frequency ITD.

Secondly, and perhaps more grievously, the duplex theory can only explain the perception of lateral angle in a single hemifield, either in the front or the rear. The theory does not explain how differing hemifields or elevations are distinguished, at least for stationary source and listener (cf. [55]). This is because many sound source positions give rise to a nearly identical ITD and ILD, and as such cannot be distinguished on the basis of either cue alone.

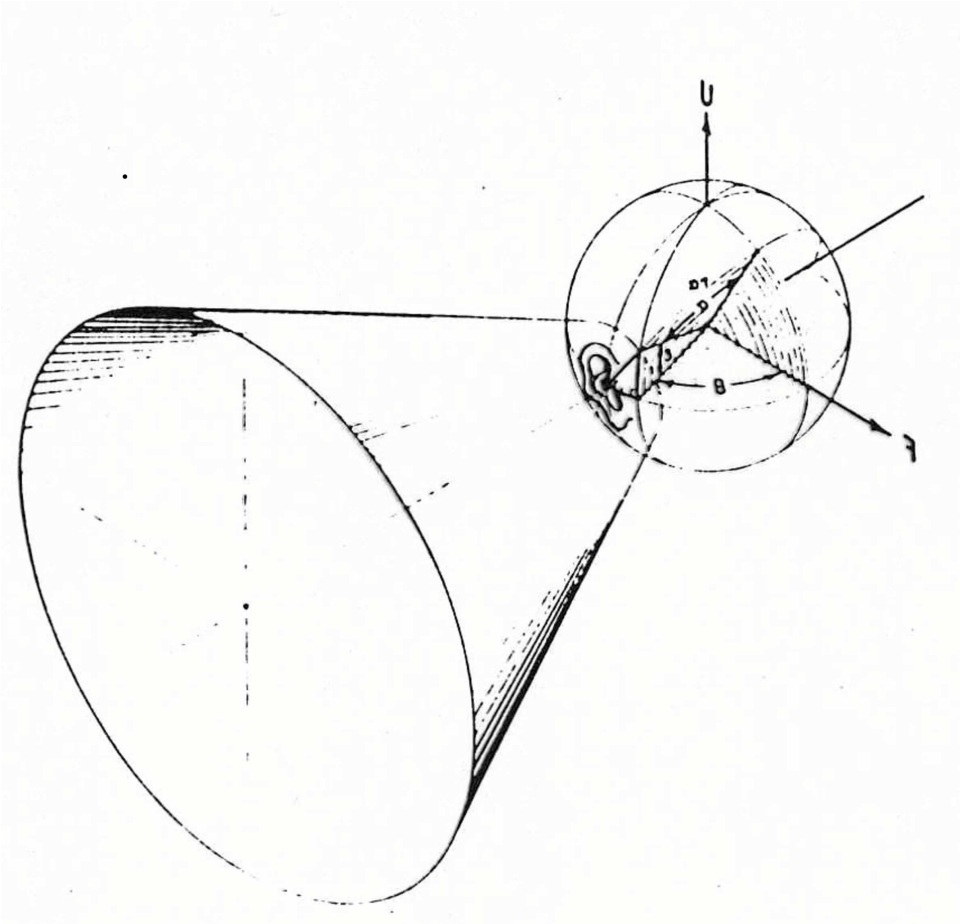

Figure 2–1: The cone of confusion. *Adapted from [39].*

Circles of confusion

The set of source positions giving rise to a nearly identical ITD and ILD form a conical region centered on the inter-aural axis and opening to the side (fig. 2–1). Such a cone is known traditionally as a 'cone of confusion' ([54] cited in [52]). If the distance to the head is fixed, as is the case with most sounds in this thesis, such a cone is reduced to a circle which we dub a 'circle of confusion' (see fig. 1–2)[2] .

2.1.2 Spectral cues

Since ITD and ILD are roughly constant on a circle of confusion, other types of cues are needed to explain discrimination between different positions around its circumference. Two types of cues are credited here, one of which is salient but difficult to synthesize, and the other which is weaker but more amenable to usage in virtual auditory display.

The more salient cues are the dynamic changes in ITD and ILD induced by head movements. These are known as dynamic cues. These changes provide unambiguous information about sound source hemifield (front or rear) [59] and elevation [45]. While these head movement cues are strong, they are only useful in virtual auditory displays when their synthesis is coupled with tracking of the listener's head movement. Of course, head tracking requires specialized hardware that is unavailable in many application contexts. Due to the practical difficulties of synthesizing

[2] Although we employ the term 'circle of confusion' for the sake of simplicity, we acknowledge that there is additional power in distinguishing it as a 'torus of confusion' (or, colloquially, a 'doughnut of confusion') in terms of human error. The just noticeable difference (JND) for the ITD and ILD can only allow sounds to be localized to within a toroidal area, rather than a circle of infinitely thin circumference [52].

head motion cues in virtual auditory display, the present discussion will focus on localization cues that function in the absence of head movement.

The secondary set of cues used to disambiguate locations on circles of confusion, used when the listener's head is immobile, are known as spectral cues. These consist of particular spectral features of binaural signals and result from the interaction of the incoming sound waves with the head, torso, and particularly the pinnae (outer ears). These parts of the anatomy act as acoustical filters that impose direction-dependent features on the spectra of incoming sounds, and these features – notches and resonances – are used to cue the IP rising angle (equivalently, the hemifield and elevation) of the source's location [16].

With regard to determining a sound source's elevation, one set of spectral cues is thought to be particularly important: the so-called 'pinna notches' [18]. These are deep nulls in the HRTF spectrum that result from reflections off the back of the concha cavity (clearly seen in the ipsilateral ear spectrum of figs. 3–2 and 3–3). Pinna geometry causes these reflections to arrive earlier in time as a sound source rises, hence causing the notch frequencies to rise proportionally with source elevation. The auditory system is sensitive to the locations of these notches in frequency and, presumably, uses them as cues to source elevation.

In summary, the acoustical cues used to localize sound sources in natural environments can be divided into two categories, at least when the listener's head is immobile. The first type of cues results from a difference in time of arrival and overall level in the binaural signals. These are termed binaural disparity cues. The second type is spectral in nature and results from the complex acoustical filtering of the head, body, and particularly the pinna. These are termed spectral cues. The two sets of cues have distinct

roles, which are particularly well illustrated using the inter-aural polar coordinate system. In this two-angle coordinate system, lateral angle is primarily influenced by binaural disparity cues, while spectral cues tend to influence rising angle [42]. In other words, at a fixed distance from the head, binaural disparity cues specify on which circle of confusion a sound source lies, and spectral cues influence the perceived location on that circle.

2.2 Binaural virtual auditory display technologies

Since all the cues used in static source localization result from the acoustical interaction of the head with an incoming sound wave, the complete set of all cues is contained in the acoustical response of the head and upper body. This acoustical response can be measured for particular sound source locations and is known as the Head-Related Transfer Function (HRTF). Virtual auditory displays exploit the fact that digitally simulating the acoustical filtering of the HRTF is sufficient to produce illusions of virtual sound sources positioned at particular locations in auditory space. That is, if the HRTF is measured and implemented as a set of digital filters, these filters can synthesize the same signals that would appear at a listener's ear drums if she were listening to a real sound source in a natural environment. If these filtered signals are then presented via headphones, illusory sound images will result, which are termed 'virtual' sound sources .

Perceptual studies have shown that this procedure is quite effective in generating spatial imagery. In 1989, Wightman and Kistler reported that these synthetic binaural signals produced localization percepts very similar to those arising in free field listening [57, 58]. They did note, however, that subjects confused locations in the front and rear hemifields more often when listening over headphones, perhaps due to the lack of dynamic head movement cues.

While Wightman and Kistler's experimental procedure produced consistent localizations, it did not lend itself well to many practical applications. To begin with, the procedure used filters based on HRTFs measured for each individual user. Individual measurements are impractical in most contexts due to their length and expense. Instead, virtual auditory display designers typically measure the HRTFs of one individual and then use this one individual's cues to process sounds for all users of the display [33]. The use of such non-individualized cues is associated with 'lower quality' imagery typified by more frequent hemifield confusions (both up/down and front/rear) and images which are more often inside the head, rather than externalized [56]. Despite these disadvantages, non-individualized directional filters are expected to remain the norm in consumer applications until the advent of more economical measurement techniques.[3]

A second problem with Wightman and Kistler's technique was the complexity of their directional filters. They exhaustively recreated both the magnitude and phase responses of the measured HRTFs, which required storage and processing capacities that prohibited real-time operation on devices of modest means. Thus, while Wightman and Kistler were successful in demonstrating that binaural virtual auditory display was possible in theory, further research was needed to render the technologies feasible in consumer application contexts.

---

[3] Some of the drawbacks of non-individualized cues can also be mitigated through HRTF 'customization', where directional filters are adapted to individual users. Customization is an ongoing topic of research (e.g., [33, 62, 37, 32]).

2.2.1 Efficient HRTF filtering

Fortunately, subsequent research showed that the *exact* replication of HRTF acoustical filtering was not always necessary to create spatial imagery. Measured HRTFs can be altered in specific ways without significantly degrading spatial image quality. In particular, studies have shown that the auditory system is generally insensitive to the phase spectra of directional filters [22, 25], so long as the low frequency ITD is maintained [21]. Fine spectral details have also been shown to be unimportant in localization [2, 24]. These findings allowed measured directional filters to be simplified in ways that increased their efficiency, for example by approximating them by infinite impulse response (IIR) filters [30, 26] or warped filters [17, 15].

2.2.2 Functional and Structural HRTF models

While these techniques have increased the processing speed of directional filtering, they have not mitigated other problems in virtual auditory display. Issues of customization and spatial interpolation are better addressed through, respectively, structural and functional models of the HRTF.

Structural and functional modeling represents a paradigm shift in the search for efficient spatial sound processing techniques. Structural models use distinct filters or sets of filters to represent not particular directions, but rather different parts of the anatomy. These models are appealing because they are parameterized by anatomical measurements. Measurements such as pinna depth or head circumference are typically much easier to obtain than acoustical HRTF measurements, and allow structural models to be customized for individuals with relative ease. One elaborate structural model developed in the 1980's by Klaus Genuit was parameterized by a total of 34 measurements of the head and upper torso [13]. A model with

considerably fewer parameters was later designed by Brown and Duda [6]. However, informal subjective evaluations of this model reported poor image externalization.

Functional models, by contrast, are not driven by the demands of HRTF customization, but rather of spatial interpolation. HRTFs are commonly measured at discrete locations, yet the synthesis of moving sound sources requires smooth and continuous variation of directional cues. Functional models thus attempt to represent HRTFs by a continuous function of direction that can be evaluated at any point, and that interpolates smoothly between measured values. One functional model proposed by Evans, Angus and Tew decomposed the HRTF into a weighted sum of surface spherical harmonics [11]. This model was not particularly efficient in terms of storage or computation. An alternative functional model that appears to be quite popular is the so-called multicomponent model. This model represents the HRTF as a weighted sum of orthogonal filters usually derived from measurements. The multicomponent approach features prominently in this thesis, and so its evolution will be discussed in detail.

2.2.3 The Multicomponent HRTF model

Origins: Principal Components Analysis of the HRTF magnitude spectrum

The multicomponent HRTF model[4] was inspired by exploratory research analyzing HRTF magnitude spectra using Principal Components Analysis (PCA) (see appendix). The goals of these early studies were primarily psychoacoustic, as they sought to use PCA to identify spectral cues to sound source location. Martens' work in 1987 analyzed 35 HRTFs from the horizontal plane and identified characteristic spectral features of four hemifields (left vs. right, front vs. rear) [31]. A later paper by Kistler and Wightman expanded Martens' idea to a larger number of subjects and source positions [22]. In all, 265 HRTF angles for both ears of 10 subjects were analyzed: a total of 5300 spectra. As in Martens' results, Kistler and Wightman observed systematic variations in the component scores with source position. They noted that 90% of the variation in the 5300 spectra could be accounted for by five components. This five-component model was also validated perceptually.

These papers were useful for several reasons. Firstly, they provided insight into the structure of HRTF spectral variation with direction. Secondly, they showed that PC-based representations could be useful in HRTF data compression, greatly reducing the amount of memory required

[4] Several papers discussing this model use the term 'multichannel', rather than 'multicomponent' (e.g. [20, 49, 48]). While the term 'multichannel' is descriptive, it is also potentially confusing, since, within the audio community, 'multichannel' is strongly associated with loudspeaker reproduction. 'Multicomponent' is advantageous due to its freedom from such connotations. Furthermore, it is both descriptive and historically appropriate since the early papers that inspired the model analyzed HRTF data using Principal Components Analysis [31, 22].

to store measured HRTFs. Unlike subsequent research, however, they were not oriented towards real-time implementations of directional filtering.

Orthogonal decompositions of the Head-Related Impulse Response

Real-time spatialization has been the focus of most subsequent work applying PCA-like orthogonal decompositions to HRTF data. The first related paper with this focus, by Chen, Van Veen and Hecox, described a functional model in the time-domain [9]. This model was based on acoustic beamforming theory. While it effectively decomposed the Head-Related Impulse Response (HRIR) into a weighted sum of components, the components in this case were not orthogonal. The model was computationally cumbersome and sometimes numerically unstable.

A second approach by the same researchers extended the earlier PCA efforts of Martens and Kistler and Wightman to the complex frequency domain [10]. This paper modeled the HRTF as a weighted sum of complex basis functions, and, unlike the previous frequency-domain analyses, modeled its measured phase as well. This model was effective in capturing HRTF spectral variation, but, in operation, necessitated costly complex-valued computations.

Later works showed that analysis in the complex frequency domain was, in fact, mathematically equivalent to a more straightforward analysis in the time-domain [49]. That is, given two matricies whose columns are related by the Fourier transform, for example a matrix of HRIRs and a corresponding matrix of HRTFs, the principal components of each will also be related to each other by the Fourier transform. Thus, PCA yields equivalent results whether it is performed in the frequency domain or the time domain. These two types of decomposition are not equivalent in usefulness, however.

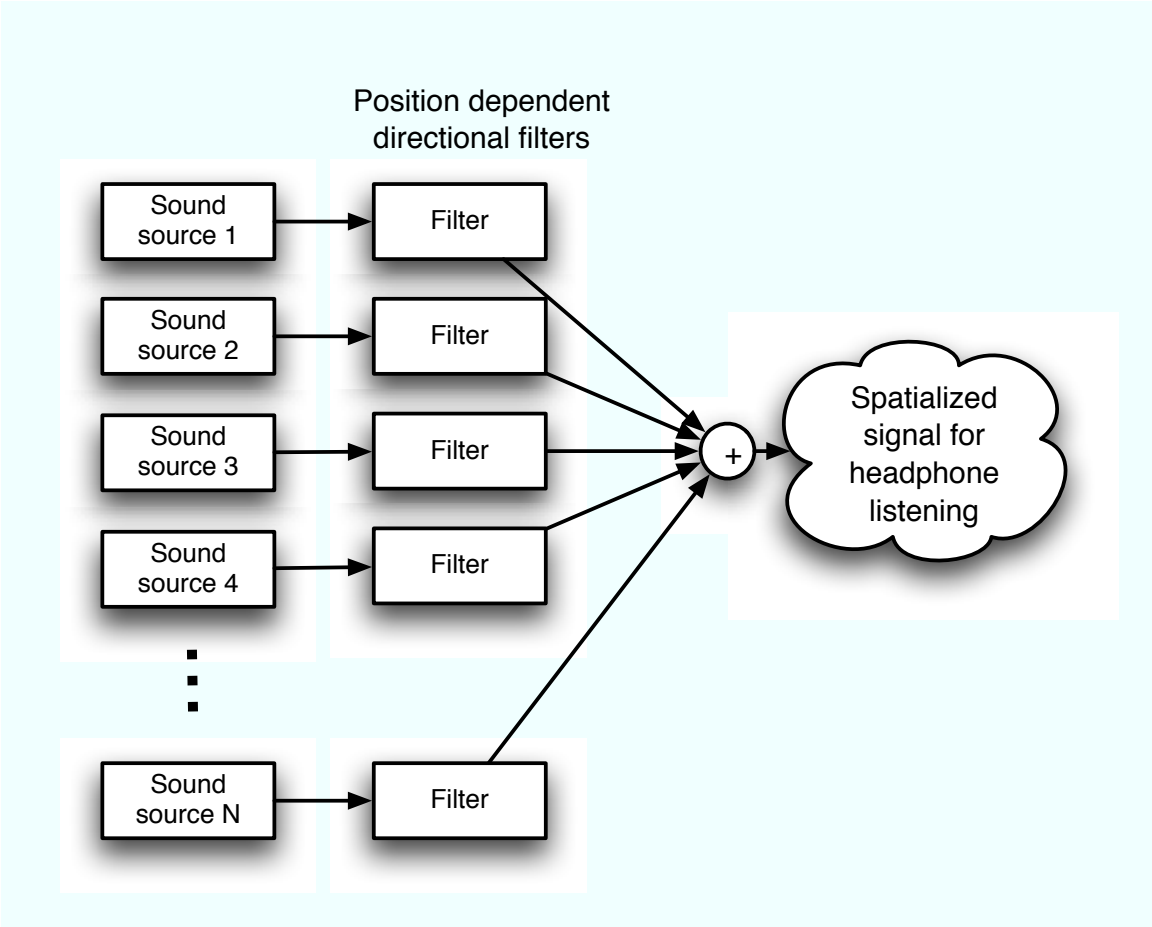

Figure 2–2: A 'traditional' approach to directional filtering for virtual auditory display. Processing for only one ear is shown.

Unlike frequency domain components, time-domain components can be implemented directly as FIR filters, resulting in extremely efficient signal processing structures for directional filtering.

The multicomponent model

The 'multicomponent model' relies on just such a time-domain decomposition. Specifically, the procedure employed in this thesis follows a patent by Abel and Foster [1]. It can be described simply as follows:

- A set of HRIRs is measured and assembled columnwise into a matrix.

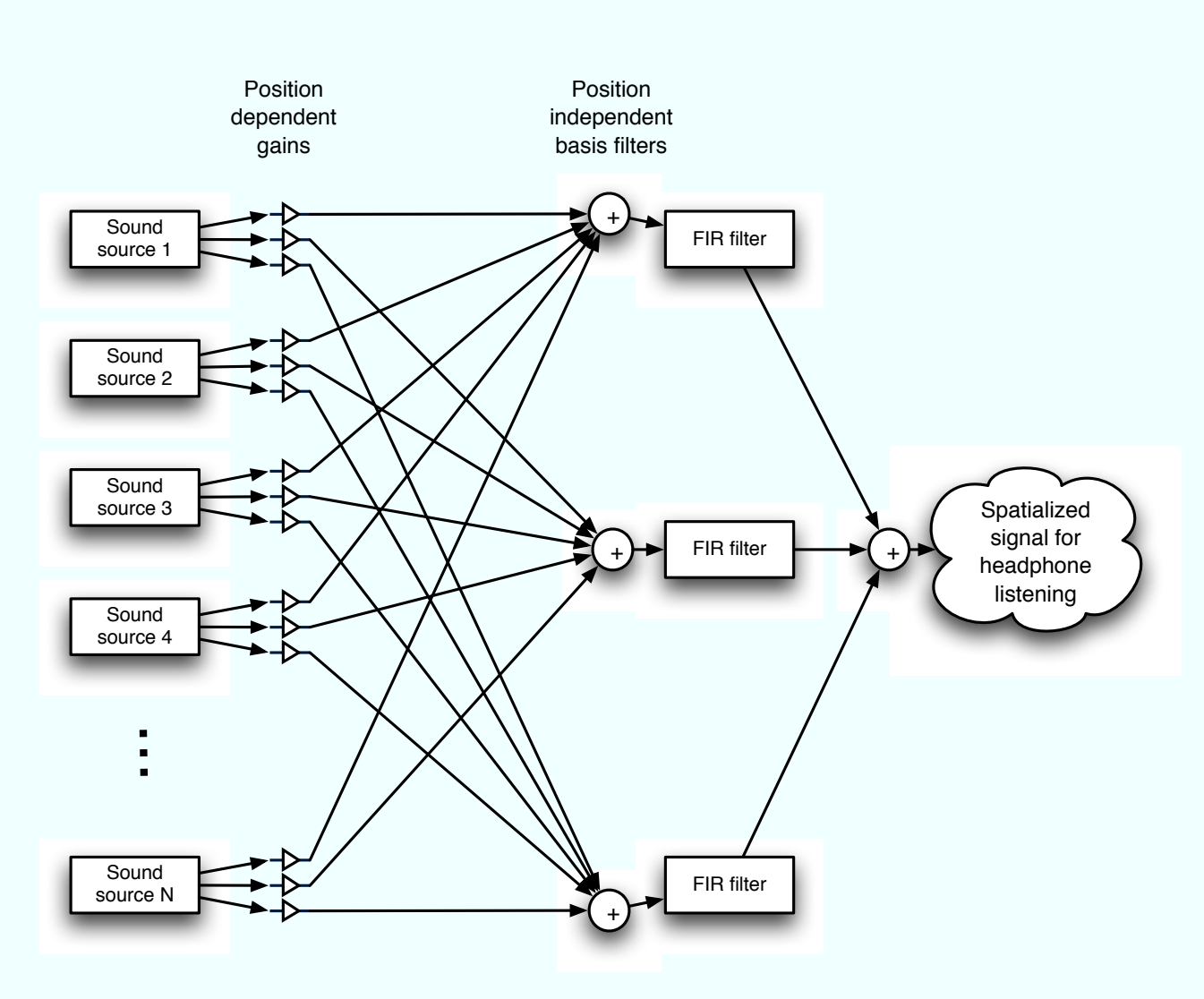

Figure 2–3: The multicomponent model approach to directional filtering. Processing for only one ear is shown.

- This matrix is then approximated by a small set of orthogonal vectors, derived from a singular value decomposition (SVD).[5]
- The small set of orthogonal vectors, referred to as 'components', are implemented as FIR filters and weighted by the direction-dependent gain values also derived from the SVD.

The appeal of the multicomponent model can be clearly seen when it is compared with 'traditional' directional filtering schemes, such as the one shown in fig. 2–2. This figure shows a number of sound sources, each being processed by its own directional filter.

Consider the run-time complexity of such a processing scheme. As complexity is dominated by the filtering operation, processing time here is roughly linear with the number of sound sources to be spatialized ($O(n)$, where n is the number of sources). This level of run-time complexity quickly leads to unmanageable computation loads when rendering scenes with many sound sources. Such scenes commonly occur, for example, in architectural acoustic auralizations.[6] In these applications, individual

[5] The SVD is a procedure closely related to PCA. The two are compared in the appendix. PCA and SVD are not the only types of matrix decompositions that have been proposed in similar contexts. For example, Larcher *et al.* experimented with independent components analysis (ICA) in an effort to reduce the number of components associated with each spatial direction and hence the amount of filtering required [27]. Variations on the SVD have also been proposed that involved weighting sections of the frequency spectrum prior to analysis [49, 48]. These weighted techniques show particular promise since they rely on perceptually informed error measures. Nonetheless, due to its simplicity, this thesis employed a straightforward SVD.

[6] Auralization is defined as "the process of rendering audible, by physical or mathematical modeling, the soundfield of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modeled space" [23].

soundwave reflections off the walls of the room are often modeled as additional sound sources, a practice which quickly leads to heavy computational loads (e.g. [29]).

The multicomponent model, by contrast, is shown schematically in fig. 2–3. Here a fixed number of filters is used irrespective of the number of sound sources. Three filters are shown here, similar to the model used in this thesis, each of which is based upon one of the components derived through the SVD. Due to the fixed number of filters, this model has a processing time that is roughly constant with increasing numbers of sound sources ($O(1)$). This processing advantage makes the multicomponent model more suitable for rendering complex scenes.

The cost of this increase in processing efficiency is a reduction of the spectral detail of the directional filters. This reduction of detail can be seen by comparing fig. 3–2, which shows a set of directional filters based exactly on measured HRIRs, with fig. 3–3, which shows the impulse responses generated by the model. Broad spectral features are preserved, but much detail is lost. This loss of detail is associated with increases in front-back confusions of static sound sources as compared with measured individualized directional filter conditions [48].

In summary, the multicomponent model belongs to a family of functional HRTF models. These models are thought to interpolate smoothly between measured HRTF directions, and, in the case of the multicomponent model, offer significant computational advantages as well. Due to the spectral smoothing inherent in the model, however, the quality of spatial imagery produced is thought to be poorer than that achieved with measured HRTF filters. When used to synthesize static sound sources, the model has been shown to increase rates of front/back confusions.

2.3 Minimum audible movement angles in sagittal planes

The goal of the present work is to examine the performance of the multicomponent model not with static sound sources, but rather with moving sources. Specifically, this work seeks to evaluate how well the model allows a sound source's direction of motion to be discerned when the motion is cued solely by variations in spectral cues. Spectral variations of this sort should create the perception of sound sources moving smoothly around a circle of confusion. The final section of this chapter, then, will review investigations of the minimum audible movement angle (MAMA) for such spectrally cued sound source motion.

Several papers have investigated MAMAs on horizontal planes [8, 46, 53, 14, 50], but, to the best knowledge of the author, only two have also reported thresholds for smoothly changing elevations on sagittal planes [14, 50]. In particular, these two studies both restricted source motion to the median sagittal plane, a region that gives rise to a constant ITD of zero. As there is no ITD variation on this plane, all motion judgements are based solely on spectral cues.

Saberi and Perrott measured sagittal plane MAMAs for 3 individuals [50]. The sound source they employed was a train of broadband pulses emitted from individual elements of a loudspeaker array. Source motion was synthesized by rapidly re-routing the signal to closely spaced adjacent elements. The sound source's velocity was varied along with its extent of motion and, at an optimal velocity of 7-11 degrees per second, an average sagittal plane MAMA of about 11 degrees was reported.

Instead of using loudspeakers, Grantham, Hornsby and Erpenbeck presented motion cues over headphones [14]. Stimuli were generated from the binaural response to a noise source of a slowly rotating KEMAR

mannequin. A pool of 20 subjects was initially recruited, but some 15 were rejected due to poor localization acuity. For the remaining 5 subjects, an average MAMA of 15.3 degrees was observed using a wideband noise stimulus.

The studies of Saberi and Perrott [50], and Grantham *et al.* [14] leave two key MAMA-related issues unresolved. These issues concern motion discrimination at varying elevations and motion discrimination of virtual sound sources synthesized with smoothed spectral cues.

Firstly, the existing literature provides no experimental data about elevation dependence of vertical MAMAs. The papers discussed above averaged together measurements from various elevations, making no attempt to determine if motion was more readily discriminable in some elevation ranges as compared to others. Studies have shown that static source localization is highly elevation-dependent – poorer at elevated positions than at ear-level [5] – but equivalent studies have not yet been carried out to measure the elevation dependence of the sagittal plane MAMA.

Secondly, the two studies provide no data about the discrimination of motion of virtual sources synthesized using smoothed spectral cues. In Saberi and Perrott's study, subjects listened to physical sources in a free-field condition. In Grantham *et al.*'s study, subjects listened to virtual sources created using the time-varying acoustical filtering of a rotating KEMAR mannequin. In both cases, it can reasonably be assumed that the spectral cues in the signals arriving at the subjects' eardrums were richly detailed. Since smoothed spectral cues are often used in application contexts for the sake of computational efficiency, it would be useful to determine whether the smoothing of spectral cues has a measurable effect on the MAMA.

The experiment reported in this thesis attempted to address these issues. The experimental methodology is described in the next chapter.

CHAPTER 3
Methodology

This thesis aimed to address the two issues left unresolved in the existing literature, as identified in the previous chapter: the effect of elevation and of directional filter spectral detail on the sagittal plane MAMA. This chapter describes in detail the experimental procedure used to assess the impact of these two variables.

The experiment reported in this thesis measured motion discrimination thresholds in six different stimulus cases. These six cases resulted from all possible permutations of the two independent variables: the 'filter case', which had two possible values, and the spatial trajectory 'starting angle', which had three. In each of these six cases, the size of the motion trajectory was varied using an adaptive staircase paradigm to track the directional discrimination threshold. Testing was completed by six subjects in two one-hour sessions spread over two days.

This chapter is divided into three sections. In the first, the techniques used in stimulus creation are described. In the second, the adaptive staircase paradigm that controlled the order of stimulus presentation is presented. Finally, the fine structure of each experimental session is shown.

3.1 Stimulus creation

This section describes how the sound stimuli used in the experiment were synthesized. The two 'filter cases' are explained, and a description of the source signal used to excite the directional filters is given. The addition of artificial reverberation is discussed, and the spatial trajectories traversed by the virtual sources are then illustrated.

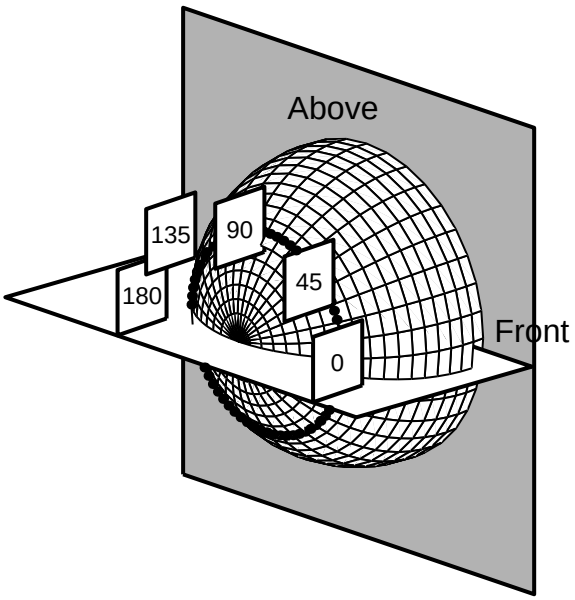

Figure 3–1: Measured HRTF angles. This figure shows a circle of confusion at an IP-lateral angle of 50°. Black dots indicate IP rising angle increments of 5°. Note that only the 37 measurements to the rear of the Subject were used to synthesize stimuli in the present experiment (i.e., IP rising angles of 90° to 270°).

3.1.1 Filter cases

To evaluate the effect of filter spectral detail on the MAMA, two different spatial processing schemes were used to generate experimental stimuli. In the first case, directional filtering was accomplished using filters based exactly on the measured HRTFs of one of the subjects. This was referred to as the 'measured HRTF case'. In the second case, directional filtering was accomplished using the multicomponent HRTF model described in Section 2.2.3. This was referred to as the 'multicomponent model case'.

Measured HRTF case

In the measured HRTF case, directional filters were based exactly on the measured HRTFs of experimental Subject 1. This meant that Subject 1 effectively listened to 'individualized' directional cues (his own), while the five other subjects listened to 'non-individualized' directional cues. These

filters had been used in several previous studies and were believed to be effective in generating useful variations in spatial imagery (e.g. [32]). HRTF measurements were taken inside a 16'-by-16'-by-10' anechoic chamber with the Subject 1 seated. Golay codes [1] were presented via a small loudspeaker, and blocked meatus responses were captured using an Etymotic Research ER-7C probe microphone [61]. The process resulted in a set of head-related impulse responses (HRIRs) from which 128-tap finite impulse-response (FIR) filters were designed (Fig. 3–2).

During measurement, the loudspeaker traversed a complete circle of confusion 1.5 m from the listener's head at an IP azimuth angle of 50° (Fig. 3–1). Measurements were taken at 5 degree increments in IP rising angle, resulting in a set of 72 measured angles in total.

The circle of confusion on which measurements were taken was shifted to the right of the median plane. This caused the wavefront of the measurement signal to arrive at one ear before the other, creating a natural ITD on the order of 500 μs. However, in the present experiment, this ITD was removed. The onsets of the ipsilateral and contralateral ear responses were time aligned, creating an ITD of zero. The resulting combination of ITD and spectral cues was thus unnatural and did not correspond to any physically possible sound source location. Nonetheless, the auditory images created by the filters were informally reported to be similar for all

[1] *?* Golay codes are pairs of signals whose numerical properties are convenient for acoustical measurement. Namely, the sum of their autocorrelations is exactly zero at every time lag except for the zeroth time lag. While the impulse also shares this autocorrelation property, Golay codes are often preferable to impulses due to the greater signal to noise ratios they achieve.

subjects. A more detailed discussion of the nature of these auditory images is presented in the phenomenology section of chapter 5.

Multicomponent model case

In the second case, spatial stimuli were generated using the 'multicomponent model' to accomplish directional filtering. As stated in Section 2.2.3, this model functions by approximating a matrix of HRIRs with a small set of orthogonal vectors, or 'components' [1, 27]. These vectors are derived from a singular value decomposition of the matrix of impulse responses (see appendix).

The gains in processing power associated with the model come at the cost of reducing spectral detail, effectively 'smoothing' the spectra of the measured filters. The smoothing effects of the model can be seen by comparing Fig. 3–3, the model output, with the measured filters in Fig. 3–2.

A free parameter in the model is the number of SVD-derived components that are retained. This value represents a trade-off between fidelity of HRTF reconstruction and computational efficiency. The present experiment retained three components to model the ipsilateral ear filter. This number was chosen somewhat arbitrarily, but was ultimately selected because it created a model that was 'reasonably similar', visually and aurally, to the original data.

Though three components were used to model the ipsilateral ear filter, only a single component was retained to model the filter representing the contralateral ear. This unequal distribution of modeling effort was an attempt to simulate application conditions. The multicomponent model is typically applied to large sets of HRIRs measured over a nearly complete sphere of incidence angles. Within these datasets, HRIRs from the side of the head closest to the sound source tend to have more energy than

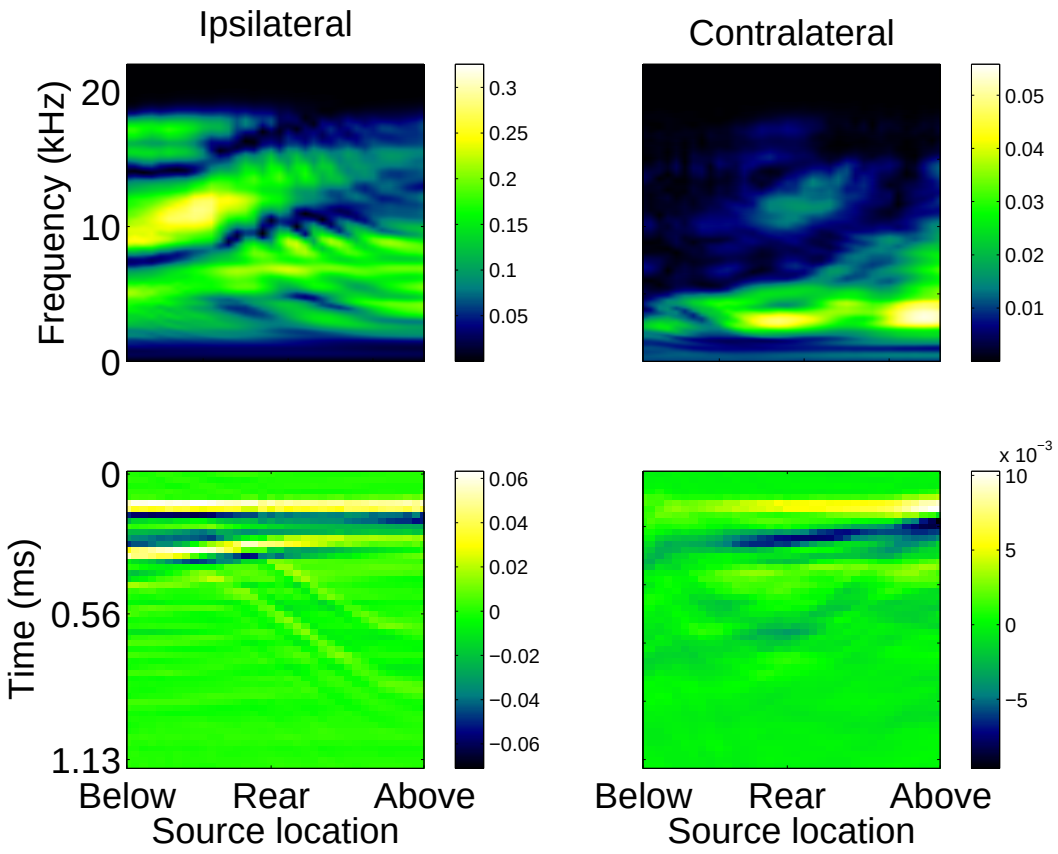

Figure 3–2: Impulse responses of the measured case HRTF filters in the time and frequency domains. Filters representing the ear nearest the sound source (ipsilateral) are shown at left, and those representing the further (contralateral) ear at right. Each figure shows filters representing 37 IP rising angles, spanning from the bottom of the circle of confusion at left (below), to a point behind and at ear-level in the center (rear), to a point on top of the circle at right (above). These angles correspond to the rear hemifield of the circle of confusion in Fig. 3–1. In the ipsilateral spectrum, note especially the "interference pattern" resembling ripples emanating from the top right corner of the image. These are thought to result from a delayed reflection off the shoulder which arrives progressively later in time as the source rises in elevation.

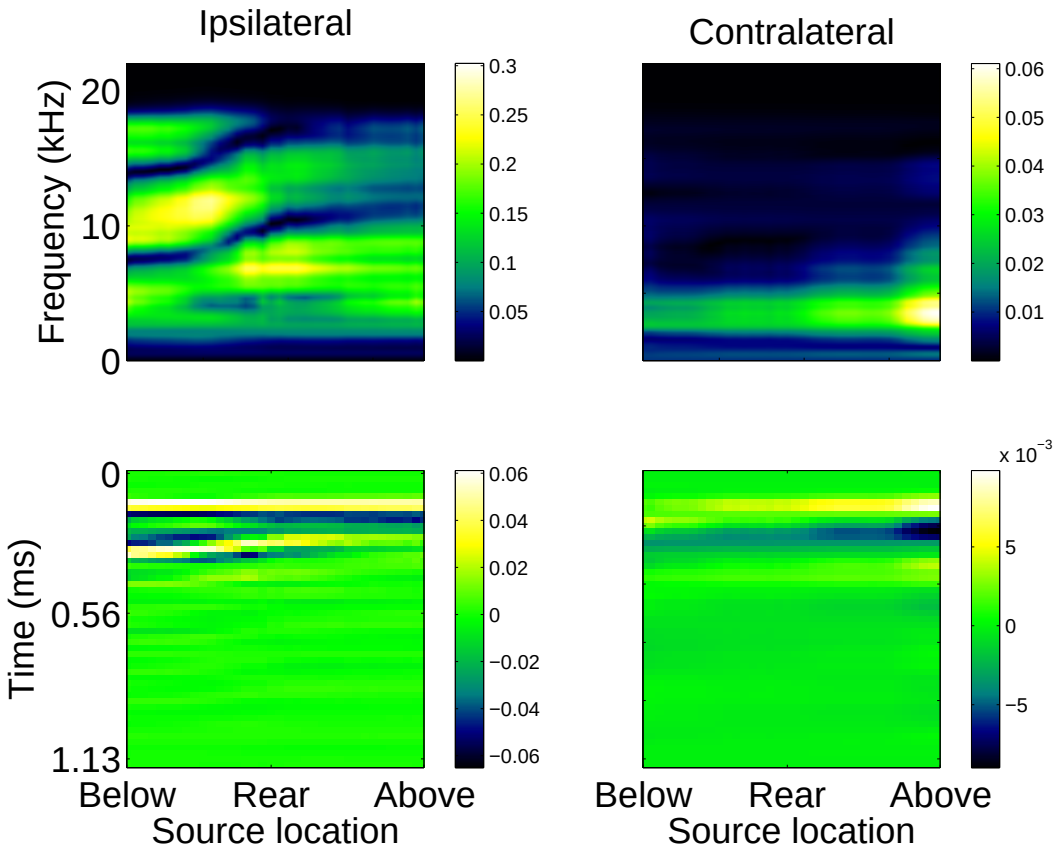

Figure 3–3: Impulse responses of the multicomponent HRTF model: time and frequency domains. The layout is identical to Fig. 3–2 and shows filters representing the ipsilateral and contralateral ears. In the ipsilateral spectrum, note the lack of detail as compared with the measured case filters. Conspicuously absent from the multicomponent model is the "interference pattern" resulting from interactions with the shoulder that was present in the measured case, as indicated in the previous figure (Fig. 3–2).

those on the far side. Since the matrix decompositions used to derive the components attempt to minimize the amplitude of error between the original and modeled data, most of the modeling effort is devoted to the ipsilateral HRIRs which are of high amplitude. In consequence, contralateral HRIRs are modeled less accurately.

Unlike application contexts, however, the present experiment did not attempt to model an entire sphere of incidence angles with a single decomposition. Rather, it modeled two simpler datasets (the ipsilateral and contralateral ear filters) with two separate decompositions. Had an equal number of components been used to model both datasets (both ear filters), the modeled contralateral ear response would have been much more accurate than would ever be possible in an application context.

Thus, to approximate the low fidelity associated with contralateral responses in typical applications, the quality of these filters was deliberately degraded by modeling them with a smaller number of components (three components for the ipsilateral ear vs. one component for the contralateral ear).

3.1.2 Source signal

To generate the experimental stimuli, a source signal was input to the measured HRTF or multicomponent model directional filters. This source signal was a close-miked recording of a bowed double bass, playing the note A2 (a fundamental frequency of approximately 110 Hz), taken from the McGill Master Samples [44].

Rationale for selecting a musical source signal

This musical stimulus was selected instead of a noise stimulus for three reasons. First, it was expected that a familiar harmonic musical source would be more likely to form a stable auditory image, in accordance with

the principles of spectral fusion [35]. Noise sources have been anecdotally reported to segregate into distinct auditory objects in similar situations, with each one potentially following a different path of motion through space. This segregation was to be avoided for fear that it would confuse subjects.

Secondly, the use of a recorded musical sound increased the study's ecological validity. A bowed double bass could potentially appear in an application context.

Thirdly, it was expected that a source with natural spectral-temporal variation (due to vibrato, etc.) would be gentler on the listener's ears and slower to induce auditory fatigue. Further, the rich spectrum and quasi-periodic nature of the sound were expected to increase the audibility of spectral details in the directional filters much in the same way that voice source jitter is thought to enhance the perception of vowel formants.

In summary, then, this musical source signal was preferable to a noise source because it was better able to test the questions under investigation, as well as providing results of more practical interest.

3.1.3 Reverberation processing

Some low level reverberation was also added to the stimulus signals to aid in image externalization [34], as diagrammed in Fig. 3–4.

3.1.4 Spatial trajectories of moving stimuli

Since the experiment was focused on the perception of sound source motion, auditory images were required to move smoothly through auditory space. To accomplished this, it was necessary to approximate values of the HRTF in between the measured angles. In the 'measured HRTF' case this was accomplished by linearly interpolating the coefficients of the directional filters between the two nearest measured angles. In the 'multicomponent

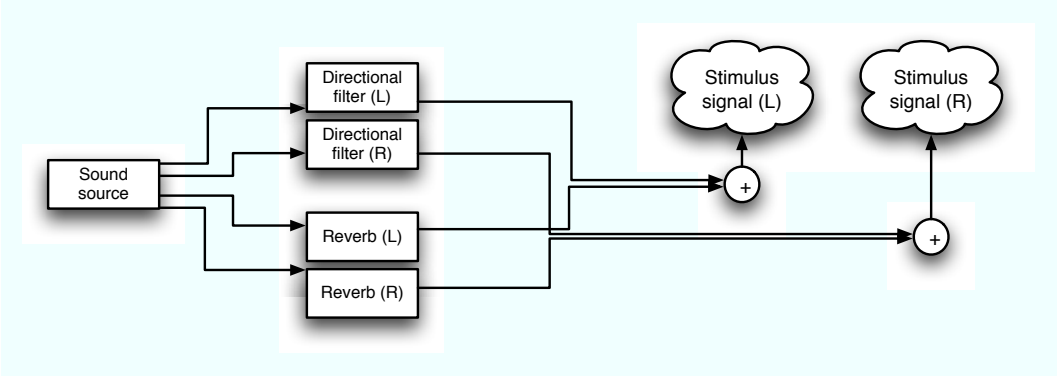

Figure 3–4: The signal processing structure used to generate stimuli. Note the parallel nature of the reverberation processing: reverberation was added in parallel, rather than in series, with directional filtering. As such, the reverberation signals were not themselves processed by the directional filters. The 'directional filter' blocks in the diagram contain either 'measured case' filters or 'multicomponent model' filters as illustrated in Figs. 2–2 and 2–3, respectively.

model' case, linear interpolation was performed on the basis filter weights.[2] Filter coefficients and basis filter weights were updated at each output sample.

Sound sources moved along one of two types of spatial trajectories (Fig. 3–5). The first type resembled a sine curve and was known as an 'up first' trajectory. The second type resembled a sine curve with a 180° phase shift and was known as a 'down first' trajectory. In both cases, the virtual sources moved above and below a central starting elevation by an angular distance termed here the movement angle. The size of this movement angle was varied from trial to trial throughout the experiment.

[2] Note that the interpolation process was much more efficient in the multicomponent model case since only 3 basis filter weights needed to be interpolated. By contrast, in the measured HRTF filter case, interpolation was performed on the 128 coefficients of the measured FIR filters.

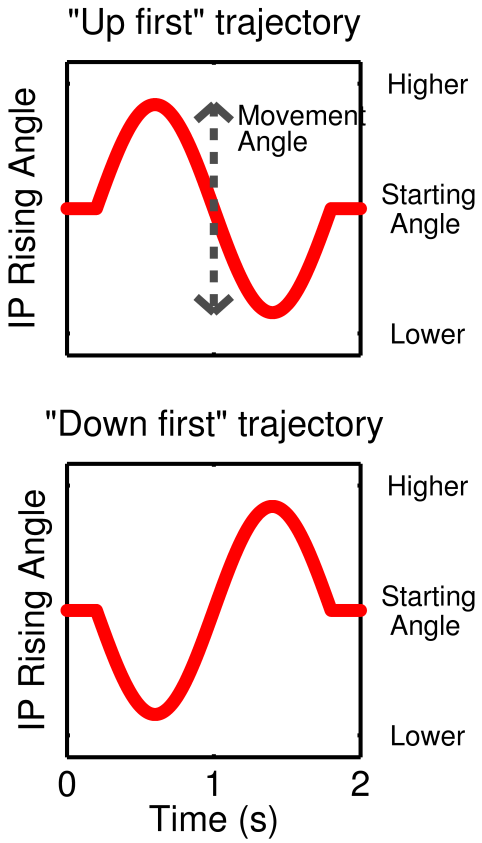

Figure 3–5: The two spatial motion trajectories between which subjects were required to discriminate.

Varying the movement angle while maintaining a constant trajectory shape and duration led to a concomitant variation in source velocity. Velocities ranged from about $72°/s$ at the largest movement angle (of 90°) to about $1.6°/s$ at the smallest angle (of 2°). These values represent the average velocity during the middle section of the trajectory, when the source moved from one extreme to the other. This section lasted for about $0.8s$.

Sinusoidal trajectories were used in the present study because they allowed the key directional attribute of the stimulus (its 'up first' or 'down first' shape) to be varied independently of its average elevation. That is, from a fixed starting angle, a source could move initially upward or initially downward without experiencing any net motion in either direction by the end of its trajectory. As the source always returned to its starting position after visiting the upper and lower extremes, a net motion of zero was maintained.

Net motion would have occurred, by contrast, if a source with a unidirectional trajectory had moved up or down from a fixed starting angle without returning to its starting position. This net motion might have provided an unwanted localization cue, since the up and down unidirectional trajectories would have had higher and lower average elevations, respectively. Subjects might then have based their 'motion' judgments not on motion cues, per se, but rather on the perceived average elevation of the sound source in relation to a known starting angle.

3.1.5 Starting elevations

All stimulus motion revolved around one of three starting elevations: above ear-level (135 degrees IP rising angle), ear-level (180 degrees), or below ear-level (225 degrees) (Fig. 3–6). These base elevations were all behind the subject and to the right.

Rearward locations were chosen for pragmatic reasons. Chiefly, they avoided the issue of front/back confusions, a type of localization error common in binaural listening [56]. Also, it may be argued that reliable control over spatial auditory cues is more valuable behind the listener than in front, since the motion of rearward virtual objects cannot be reinforced by visual imagery.

3.2 Adaptive staircase threshold tracking

The experiment employed an adaptive staircase paradigm, meaning that the movement angle in a given trial depended on the subject's responses in previous trials. At the beginning of each staircase, subjects were presented with a stimulus that was given a movement angle of 50°. The stimulus' trajectory ('up first' or 'down first') was chosen randomly. Subjects were asked, in a two-alternative forced-choice task, to report on which motion trajectory they heard. A response corresponding to the trajectory used in stimulus creation was deemed 'correct'. After giving a response, the subject was immediately presented with the next stimulus.

When a subject gave three correct responses in a row, the size of the movement angle on the following trial was reduced. Conversely, if a single incorrect response was given, the movement angle increased. This scheme, a three-down one-up transformed adaptive staircase, tracks the movement angle at which trajectories are correctly identified 79.4% of the time [28].

Note that subjects were never given explicit feedback about the correctness of their responses.

3.2.1 Staircase step size

The amount by which the movement angle changed at each step up or down, known as the 'step size', was reduced gradually throughout each block 3–7. Specifically, the step size depended on the number of turnarounds

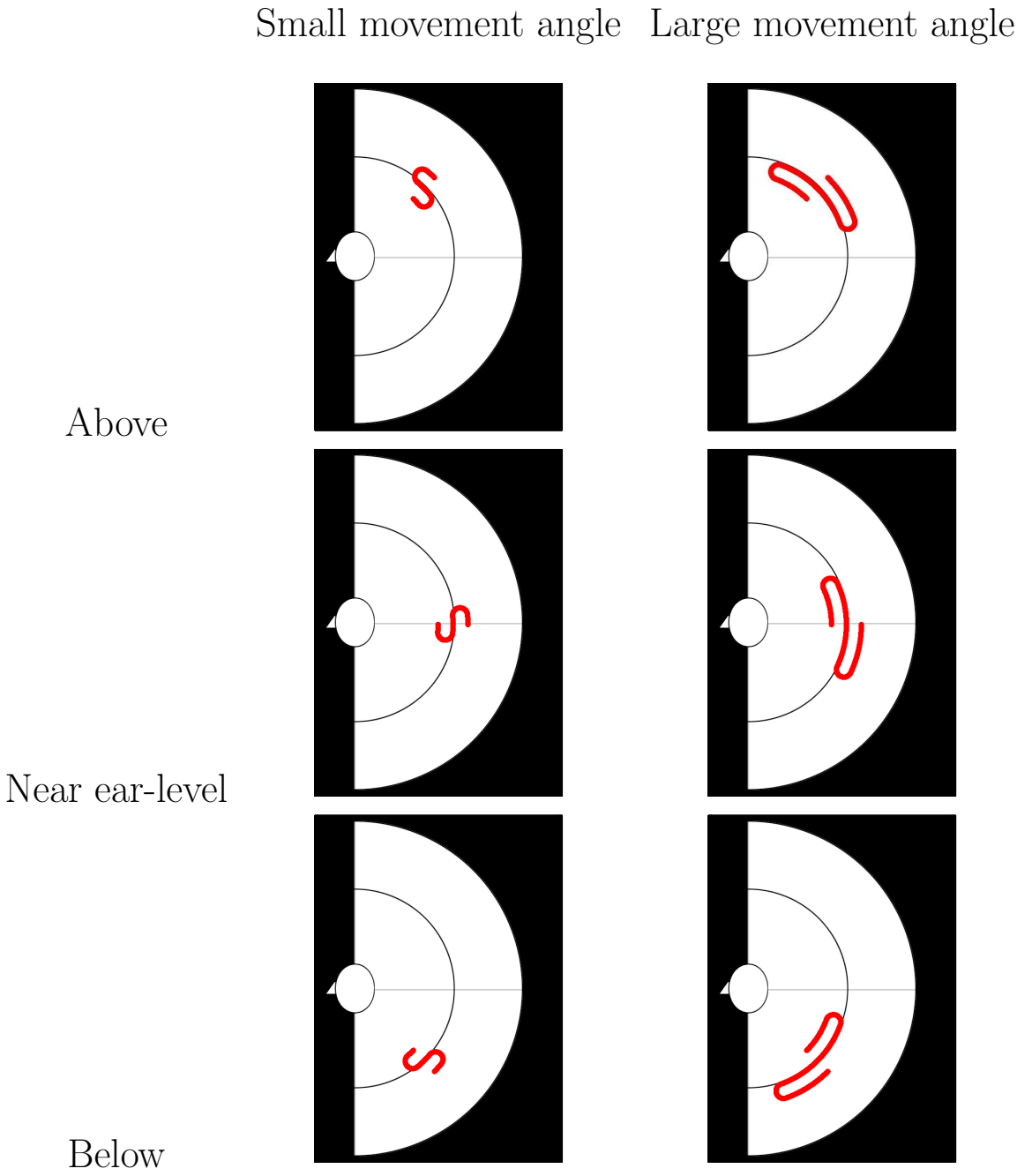

Figure 3–6: Examples of motion trajectories at different elevations. The semi circular line in the figure represents the rear hemifield of the circle of confusion on which sources moved. The curved trajectory lines have been 'jittered' and offset from the circle to show their temporal evolution. Small movement angles have been drawn as 'down first' trajectories, and large movement angles have been drawn as 'up first' trajectories, although, in the experiment, both trajectories were equally likely to occur at any given movement angle.

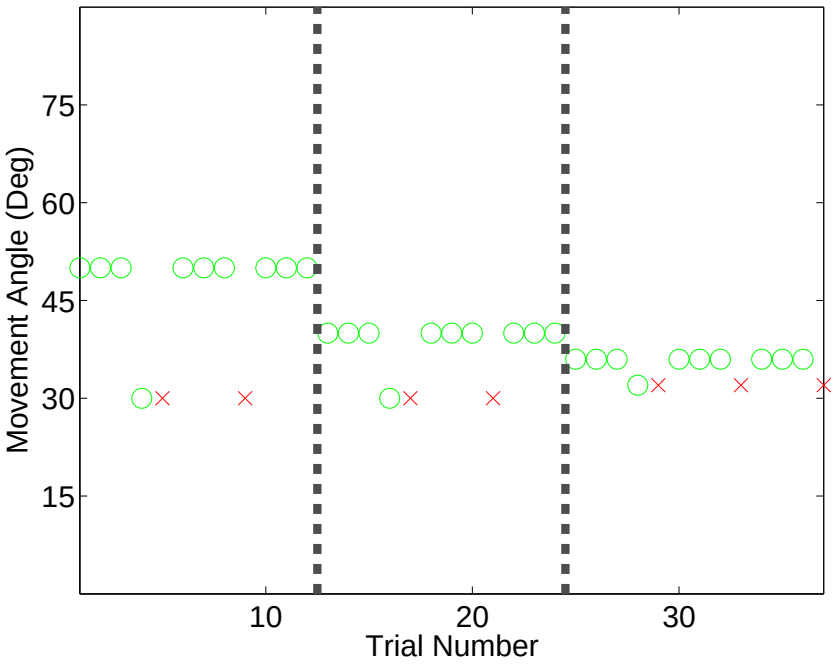

Figure 3–7: An example staircase of subject responses showing changes in step size. Circles indicate correct responses while crosses indicate incorrect responses. Note that the stimulus Movement Angle increased following incorrect responses and decreased following three consecutive correct responses. Local maxima and minima in the history Movement Angles are referred to in the text as 'turnarounds'. After each group of four turnarounds a dashed line appears. At each dashed line the staircase 'step size' is reduced.

Session Structure

Starting angle 1	Training task
	Block 1 (measured or multicomponent case)
	Block 2 (measured or multicomponent case)
Starting angle 2	Training task
	Block 1 (measured or multicomponent case)
	Block 2 (measured or multicomponent case)
Starting angle 3	Training task
	Block 1 (measured or multicomponent case)
	Block 2 (measured or multicomponent case)

Table 3–1: The structure of each testing session. The order of presentation of starting angles and filter cases was randomized between subjects.

(local maxima and minima) in the history of movement angles. Until four turnarounds had been obtained the step size stayed at its initial value of 20 degrees. It then dropped to 10 degrees for four turnarounds, after which it fell to its final value of 4 degrees. Once four more turnarounds had been obtained at this final step size, the staircase was terminated. At each step size reduction, a new movement angle was calculated as the mean of the turnarounds at the previous step size. This formula for sound source variation had been anecdotally reported to converge quickly [28], and allowed the staircase to first find a 'ballpark estimate' of the subject's threshold, and then hone in on a more fine-grained measure.

3.3 Structure of experimental sessions

Subjects completed two testing sessions on two separate days, each one consisting of three training tasks plus six experimental blocks. Each experimental block contained two interleaved staircases that were run simultaneously, in the sense that movement angles were alternately taken

from staircase 1 or staircase 2. Staircases were interleaved in this way because they made it difficult for subjects to learn the pattern of stimulus variation. The structure of each session is shown in table 3–1.

Training tasks preceded each group of blocks. These served to familiarize subjects with the experimental procedure and stimuli. Each training task contained a single staircase, rather than an interleaved staircase, to save time. Also, they began at relatively large movement angles (90°) to initially provide subjects with unambiguous stimuli that would build their confidence in their ability to recognize upward from downward trajectories. Training blocks contained stimuli processed exclusively by the measured HRTF case filters.

The order of presentation of filter cases and starting elevations was randomized for each subject.

CHAPTER 4
Results

This chapter discusses the analysis of the experimental data and presents the results of the investigation. Before presenting the findings, however, the reliability of the collected data is considered. The subject responses in each experimental block are gauged for consistency on the basis of an objective criterion. This criterion provides grounds for rejecting some of the data, namely all the results from one Subject (Subject 6) and the results from all subjects given in the first experimental session. Following this pruning of the results, the remaining data are then submitted to an analysis of variance (ANOVA).

Experimental data consisted of staircase tracks showing subjects' correct or incorrect responses at various movement angles. Note that each experimental block contained two interleaved staircases. Since each subject completed 12 blocks in total, and each block contained 2 staircases, 24 individual staircases were created by each subject.

Staircases varied in length, but each contained an equal number of local maxima and minima, or 'turnarounds' (see Section 3.2.1). Only the last six turnarounds in each staircase were analyzed. The mean value of these final six turnarounds was considered the 'threshold estimate' for that particular staircase.

4.1 Data selection

The first step in the data analysis process was to determine the reliability of the subject responses. For this purpose the means of the two interleaved staircases in each block were compared. Each pair of these

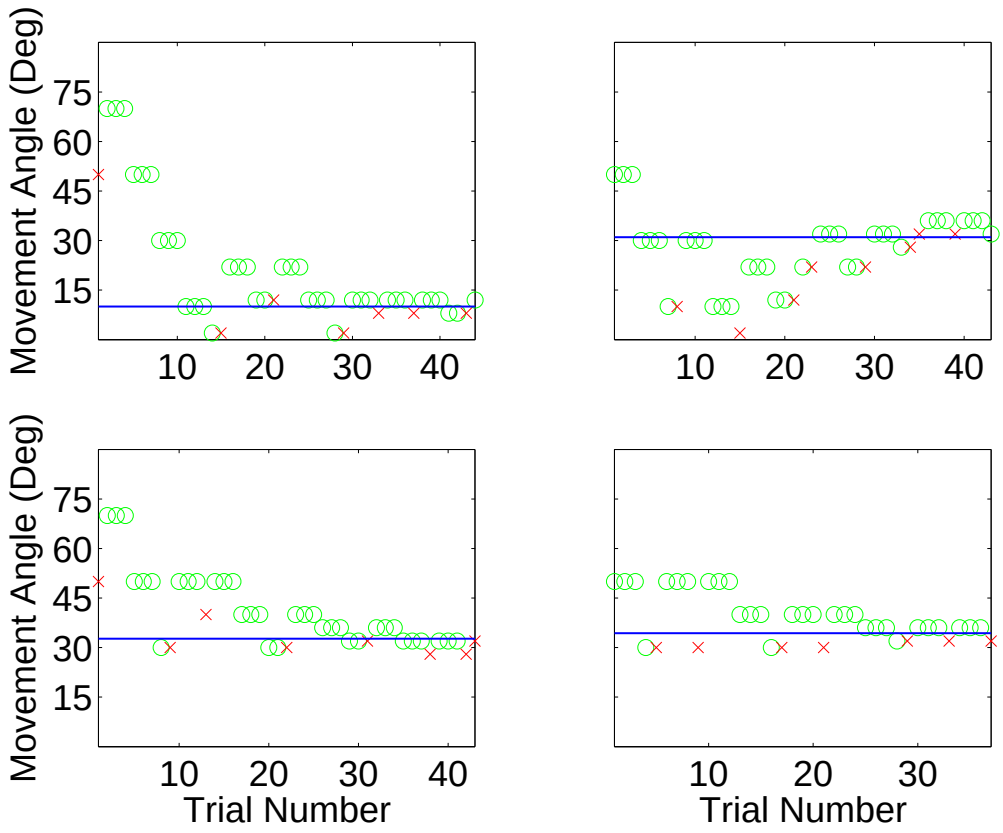

Figure 4–1: Large and small 'inter-block mean differences'. Each row in this figure shows two staircases that were interleaved in one experimental block. The horizontal line indicates the threshold estimate for that particular staircase. The top row of the figure shows a block with a large 'inter-block mean difference'. The bottom row shows a block where the 'inter-block mean difference' is small because the two staircases have converged to nearly the same value.

interleaved staircases tested the exact same experimental conditions, i.e., the same Filter Case and the same Starting Angle. The effect of interleaving the staircases, then, was to perform two independent tests of the same conditions simultaneously. If subjects were focused on the task and gave thoughtful responses, both staircases would be expected to converge to nearly the same value, yielding two similar threshold estimates. Blocks in which the two threshold estimates diverged significantly were treated as suspect, and indicated that the subject had given inconsistent responses, perhaps due to a lack of focus, motivation or attention.

4.1.1 The inter-block mean difference

A metric was devised to gauge the consistency of the responses in each block. This metric was known as the 'inter-block mean difference' and was calculated as the absolute value of the difference between the two threshold estimates in each pair of interleaved staircases. Examples of pairs of staircases with large and small inter-block mean differences are shown in Figure 4–1.

4.1.2 Criteria for rejecting session 1 data

Comparing the average inter-block mean differences across the two experimental sessions suggested that, overall, more reliable responses were given in the second session. This was likely because subjects were better practiced at the task. The average inter-block mean difference across all subjects in the first session was 5.5 degrees, but this fell to 4 degrees in the second session. Thus, the first session data were deemed less reliable and discarded. All analysis was perfomed on the data from the second session.

4.1.3 Problematic subjects

The inter-block mean difference was also used to gauge the consistency of each subject's responses. All subjects except Subject 6 were able to keep

Source	SS	df	MS	F	p
Filter Case Factor	870.01	1	870.01	63.42	< 0.01
Subject	1730.61	3	576.872	42.05	< 0.01
Interaction	982.61	3	327.538	23.88	< 0.01
Error	1207.25	88	13.719		
Total	4790.49	95			

Table 4–1: Two way ANOVA for starting angles above ear-level

inter-block mean differences below about 16 degrees. Subject 6 gave one inter-block mean difference of 48 degrees, and so was deemed unreliable. Subject 6 was excluded from further analysis.

Subject 5 was also problematic in that his discrimination was exceptionally poor. On several occasions he incorrectly identified the stimulus with the largest available movement angle (of 90°). At this point the adaptive staircase should have presented him with a still larger movement angle but was unable to do so, since stimuli with larger movement angles had not yet been synthesized. Instead, the maximum stimulus level was simply maintained. This failure in the algorithm's performance should be considered a type of 'ceiling effect' and should be interpreted to mean that Subject 5's thresholds may be larger than those reported.

Subject 5's thresholds were significantly different from those of the other subjects, at the 0.05 level, as calculated by the MATLAB *multcompare* routine using the Tukey-Kramer option. This Subject's thresholds are included in figures 4–2 and 4–3 but were omitted from the ANOVA.

4.2 Two-way ANOVA

The results of a two-way ANOVA were calculated for all staircase results at each of three starting angles (Tables 4–1, 4–2 and 4–3). The last six turnarounds of the two interleaved staircases in each block were combined and used as input.

The two-way ANOVA had the following factors:

Source	SS	df	MS	F	p
Filter Case Factor	442.04	1	442.042	31.48	< 0.01
Subject	2876.37	3	958.792	68.27	< 0.01
Interaction	227.37	3	75.792	5.4	< 0.01
Error	1235.83	88	14.044		
Total	4781.63	95			

Table 4–2: Two way ANOVA for starting angles near ear-level

Source	SS	df	MS	F	p
HRTF Factor	30.38	1	30.375	2.59	$= 0.111$
Subject	141.42	3	47.139	4.02	< 0.01
Interaction	354.37	3	118.125	10.08	< 0.01
Error	1031.17	88	11.718		
Total	1557.33	95			

Table 4–3: Two way ANOVA for starting angles below ear-level.

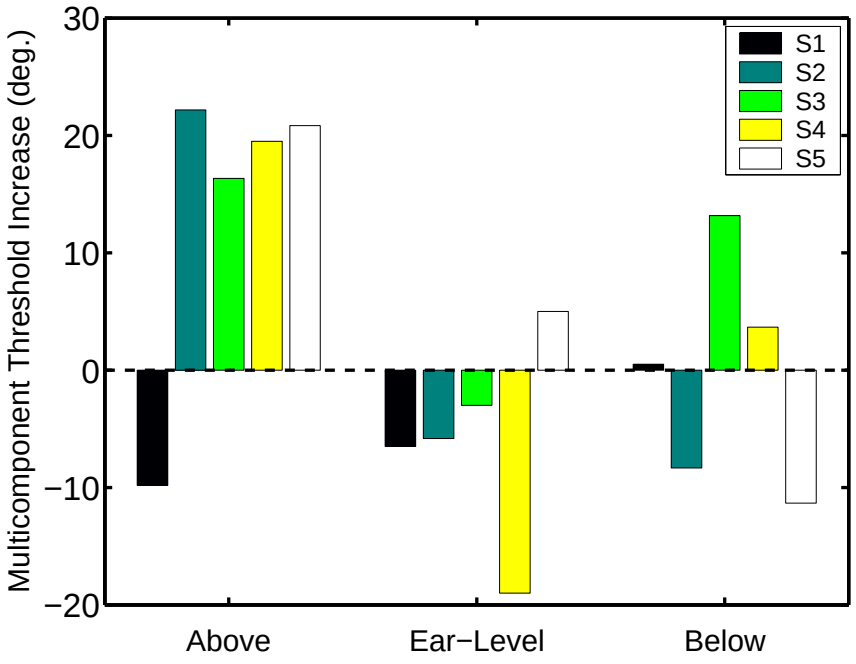

Figure 4–2: Differences between Multicomponent Case thresholds and Measured Case thresholds for three elevations. Note that a negative value in this figure indicates a case in which the multicomponent model gave more acute directional discrimination than did the measured case.

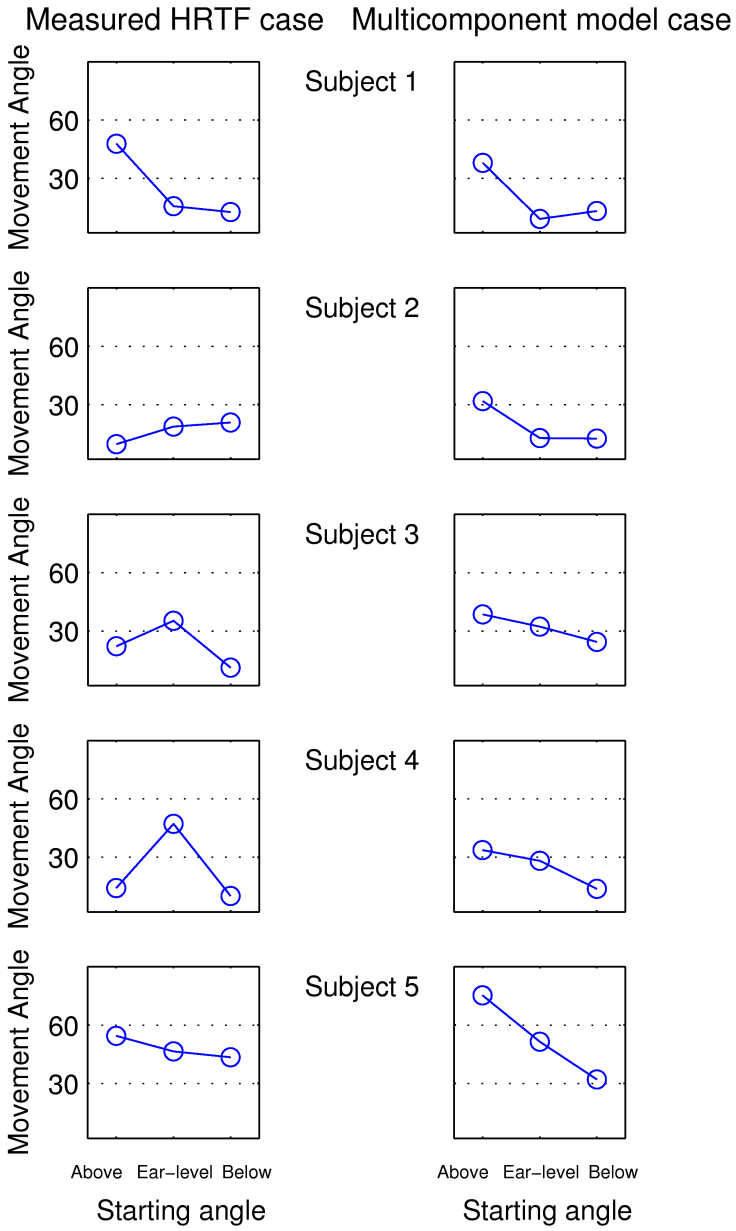

Figure 4–3: Directional discrimination thresholds plotted on starting elevation with Filter Case Factor as the parameter. Thresholds are measured in terms of IP rising angles on a circle of confusion at an IP lateral angle of 50°.

- A Filter Case Factor with two levels (measured case versus multicomponent case)
- A Subject Factor (separating the results from each of the four listeners included in the analysis)

The results show that, at the 0.01 level, threshold differences due to the Filter Case were statistically significant for starting angles above and near ear-level. Only for sources moving below ear-level was this effect not significant. The effect of Filter Case on threshold values is shown in Fig. 4–2.

It was also true that differences due to the Subject were always statistically significant, as were the interactions between Subject and Filter Case. This means that for different Subjects, the two Filter Cases had different effects on their thresholds.

Doing multiple comparisons between means calculated for the whole group of four listeners is problematic because of the significant interaction that is always present in the data. Therefore, the means must be examined for each listener in a case-by-case fashion. Subject means for the two Filter Cases are shown in Fig. 4–3.

CHAPTER 5
Discussion and Conclusion

In this final chapter the results of the investigation are interpreted. First, the phenomenology of the experiment is discussed. Verbal reports from Subjects are considered to establish whether or not Subjects actually experienced auditory motion when presented with the experimental stimuli. Following this discussion of subjective experiences, the results of the present experiment are compared with those of previous studies. Agreement between current and previous studies increases our confidence in the experiment's methodology and hence our confidence in its results. Individual differences in the measured thresholds of the Subjects are then considered. Possible justifications are given for the wide range of collected responses. Finally, two conclusions concerning the effect of multicomponent modeling on motion discrimination thresholds are drawn. These results are summarized and suggestions for future work are presented.

5.1 Phenomenology

No part of the official methodology of this experiment required subject to report freely on their subjective experiences. Subjects were not systematically asked to describe the nature of the auditory imagery they experienced; they were merely asked to select between upward first or downward first motion trajectories. Nonetheless, some insight into the experiment's phenomenology was gleaned through casual conversations with the Subjects following the experiment.

In post experiment conversations, all subjects agreed that they experienced motion of virtual sound sources. There was widespread consensus

about finer details of the perceptual experience as well. When the Movement Angle of the stimulus was large, the presence of motion was unambiguous. Its direction was not always evident, however. When Movement Angles became sufficiently small, timbral variation was experienced rather than motion. Subjects also agreed on the general location of the moving sound sources. Virtual sources were heard to move up and down in the rear right hemifield. Perceived motion was not restricted to the surface of a saggital plane, however, as is implied in Fig. 3–6. Rather, auditory objects appeared closer to the median plane at some points in the trajectory and further from the median plane at others. These phenomenological descriptions were confirmed by the author, who also served as a subject in the experiment.

5.2 Comparison with previous results

5.2.1 Data selection

To make meaningful comparisons with the results of previous studies, the data presented in the previous chapter must be reinterpreted in several ways. First, it is necessary to convert the reported threshold values from IP rising angles to vertical polar elevation angles, since this latter angular measure is employed in the relevant literature.

Secondly, it is necessary in certain cases to examine the data selectively, considering only those results obtained under conditions similar to those in the study with which comparisons are being made. For example, to compare with one relevant study (i.e., [50]), it is necessary to consider only the results from Subject 1, since only this subject listened through 'individualized' directional filters. [1] To compare with the results of another

[1] For a discussion of the significance of individualized versus non-individualized filters see section 2.2.

relevant study ([14]) it is necessary to consider only those Subjects who performed best on the task and exclude those who performed poorly since the comparison study was similarly selective.

5.2.2 Previous results

One related study of vertical plane MAMAs is that of Saberi and Perrott [50]. Subjects in this study listened in free field conditions, meaning that they heard real sound sources through their own ears. The spatial cues heard by the subjects thus corresponded to their own external ear anatomy. In this respect, the free field listening in [50] is comparable to individualized cue listening in the present study. In this condition, the authors reported a MAMA of 11 degrees for elevations near ear-level. In the present study, the only Subject to listen through individualized spectral cues was Subject 1. This Subject obtained a near ear-level MAMA of 10 degrees, 1 degree lower than Saberi and Perrott's threshold.

In another related study Grantham *et al.* used non-individualized cues and reported a MAMA of 15.3 degrees [14]. This threshold value was obtained not by averaging the results from all their subjects, but rather from an average of the best 5 localizers drawn from an initial group of 20. The authors discarded results from 75% of their subjects, citing a desire to report only on individuals who were proficient at their particular experimental task. The reported results for Grantham *et al.* were highly selective (three out of four subjects were discarded), and thus, to make a meaningful comparison, we must also disregard some poorly performing Subjects. In the present study, 5 subjects listened under non-individualized conditions. Of these 5, the two best discriminators (Subjects 2 and 3) obtained an average measured HRTF case MAMA of 16.8 degrees. This was slightly higher than Grantham *et al.*'s threshold of 15.3 degrees.

5.2.3 Discussion of previous results

These previous studies differed from the current study in a number of ways. First, they measured thresholds in the frontal hemifield on the median sagittal plane, rather than in the rear hemifield on a sagittal plane slightly offset from the median. Secondly, they used noise bursts rather than musical signals as stimuli. Thirdly, their stimuli moved along unidirectional trajectories rather than sinusoidal trajectories. Despite these differences in methodology, however, the motion discrimination thresholds measured in the current study are on par with previously reported results. This fact helps to validate the experimental design used in this thesis.

An additional difference between the current study and previous studies concerns the 'congruency' of the cues in the directional filters. In previous studies, spatial cues were 'congruent', in that they contained combinations of ITD and spectral cues that could have arisen in natural listening conditions. The spatial cues in the current study, by contrast, were 'incongruent', since the combination of ITD and spectral cues did not correspond to any physically possible sound source location, and could not have occurred in any measured HRTF. Despite this incongruence, however, the binaural signals presented in the current study appear to have preserved whatever cues are used in motion discrimination. This suggests that spectral cues for source motion on sagittal planes are resilient to changes in ITD, and may be equally effective in cueing motion even when presented in combination with unnatural ITDs.[2]

[2] This suggestion is consistent with the results of Morimoto and Aokata who showed that, for static sound sources, spectral cues measured on the median plane could be used to create auditory images on any sagittal plane, provided that they were presented with an appropriate ITD cue [43].

5.3 Individual Differences

While certain trends seem to emerge from plots of the experimental results, it is also true that generalizations are problematic due to the striking differences between thresholds reported for different Subjects in different conditions. These differences can be attributed to at least three factors, namely the idiosyncratic nature of individual spatial cues, differing aptitudes for motion discrimination amongst the subjects, and imperfections in the experimental methodology.

Firstly, some individual differences might be due to the degree of similarity or dissimilarly between a Subject's own spatial cues and those cues present in the directional filters. While Subjects 2-5 all listened through the same set of 'non-individualized' filters, it is not necessarily the cases that the cues in these filters were equally foreign to all subjects. Some Subjects may have had internal cues more similar to the filter cues than others; i.e., the filters may have varied in their degree of 'non-individualiztion' for different Subjects. As a result, some of the variation in results may be attributable to differences in the level of familiarity of the cues presented.

While a lack of 'individualization' of directional cues can account for instances in which Subjects had higher thresholds than Subject 1 (for whom the filters were individualized), it does not account for those instances in which Subject 2-5 recorded *lower* thresholds. Such results can be seen in Subjects 2, 3 and 4 for Starting Angles above ear level. In these cases the Subjects appeared better able to use Subject 1's spatial cues than Subject 1 himself! Here individual differences may be due simply to different aptitudes for auditory motion perception. It may be the case that Subjects 2, 3 and 4 are more skilled than Subject 1 at detecting motion in the upper

hemifield, and so can perform better at motion discrimination tasks despite the penalty associated with non-individualized directional cues.

A final source of individual differences in the results is simply experimental noise. The experimental task was long and monotonous, and it is possible that some subjects may have lost concentration and given unintentional responses at certain points. The data are highly susceptible to this sort of contamination since each reported threshold is drawn from the convergence values of only two staircase tracks. It is expected that this experimental noise would be averaged out if the experiment were repeated with a larger subject pool or a larger number of trials.

5.3.1 Significance of individual differences

These justifications for individual differences in responses have consequences for the interpretation of the experimental results. Specifically, they provide a means of explaining any outlying data points that may appear inconsistent with otherwise reasonable conclusions. For example, one conclusion discussed below is that multicomponent modeling improves motion discrimination near ear level. This conclusion is supported by data from four subjects, but is contradicted by data from the fifth. In this case, it is possible that Subject 5's results can be explained by experimental noise. The fifth subject may have given inconsistent responses due to fatigue and this may serve to explain why his results do not agree with those of the rest of the group. Such justifications for individual differences should be kept in mind in the discussions of conclusions below.

5.4 Threshold shift in the multicomponent case

Visual inspection of figures 4–2 and 4–3 suggests that motion discrimination was not uniformly degraded in the multicomponent model case. In fact, in several conditions, the multicomponent model actually appears

to have lowered thresholds and allowed for better motion discrimination. This effect is most pronounced near ear-level, where lower thresholds were observed for four out of the five subjects.

The reason for the improved motion discrimination in the multicomponent model case is likely related to its smoothing of the HRTF spectra. While this smoothing process may indeed damage some localization cues, it also appears to remove other spectral details that may be extraneous to motion detection. In turn, the removal of these extraneous details may uncover salient motion cues that were previously hidden, such as spectral notch migration.

The 'uncovering' of such cues can be clearly seen by comparing the ipsilateral spectra of the multicomponent and measured case filters at angles near ear-level (Figs. 3–2 and 3–3, top left panel, middle). In this region, two deep spectral notches exist whose frequencies rise monotonically with source elevation. In the multicomponent model case (Fig. 3–3) the frequency migration of these notches is clear and unambigous. In the measured HRTF case (Fig. 3–2) an 'interference pattern' is superimposed on these notches that eliminates the strict monotonicity of frequency migration. These notches are well known elevation cues [7], and it is perhaps the smoother migration of these notches in the multicomponent case that enabled better motion discrimination for four out of the five subjects.[3]

[3] Indeed, while the salience of these pinna notch cues is well documented in the literature, it can also be demonstrated with simple models of the HRTF based on spectral peaks and notches. Such a model is available online as part of an HRTF sonification project initiated by Densil Cabrera of the University of Sydney. The interested reader is referred to the project website: http://wwwpeople.arch.usyd.edu.au/~densil/sos/hrtf/index.htm

5.4.1 Elevation dependence in the multicomponent case

Again examining the plots of individual Subject thresholds in the previous chapter (Fig. 4–3), it also appears that directional discrimination is more strongly elevation dependent in the multicomponent model case. In this filter case, all subjects had more difficulty discriminating motion above ear-level than near ear-level. This type of elevation dependence was not generally present in the measured HRTF case.

This effect raises the question of how multicomponent modeling could simultaneously improve directional discrimination near ear-level but impair it above ear-level. Again, the answer may lie in the multicomponent model's treatment of pinna notch migration. Comparing the ipsilateral ear spectra for the two filter cases at elevations above ear-level (Figs. 3–2 and 3–3, top left panel, left third), it appears that the multicomponent model distorts the smooth migration of pinna notches. Whereas in the measured HRTF case the lower notch continues to rise smoothly as the sound source rises, in the multicomponent model case the upward migration of this notch is much less evident. The presence of this reduction in notch migration above ear-level is somewhat counterintuitive given that the model enhances notch migration near ear-level. Nonetheless, this distortion in the filter spectra seems a reasonable explanation for the higher above ear-level motion detection thresholds in the multicomponent model case.

5.5 Conclusions

This study measured minimum audible movement angles for spectrally induced motion of virtual sound sources on a sagittal plane. The obtained motion discrimination thresholds were consistent with published results, despite the fact that the ITD had been removed from the measured transfer functions. This suggests that spectral motion cues measured in one sagittal

plane may be equally valid in others. A key contribution of the study was the suggestion that multicomponent directional filters do not necessarily worsen motion discriminations thresholds, and in fact may improve motion discrimination for sources near ear-level. These filters also appear to give rise to a strong elevation dependence for vertical motion discrimination.

5.6 Future work

The most interesting result of this research is the suggestion that, for elevations near ear-level, directional filters smoothed by the multicomponent model may enable better vertical motion discrimination than filters based exactly on measured HRTFs. This suggestion runs counter to a prevalent assumption in the spatial hearing community that filters based exactly on measured HRTFs will create optimal directional percepts in spatial auditory display. This present study suggests, however, that for the particular task of motion discrimination within a sagittal plane, simplified models of the HRTF may enable better performance. One potential goal of future work would be to validate this hypothesis with a larger subject pool.

Appendix A: Principal Components Analysis and the Singular Value Decomposition

Principal Components Analysis (PCA) is a statistical technique applied to multivariate data. Such data are often expressed as matrices of size $m \times n$, where each column contains a set of m variables and each of the n rows contains one observation of this set of variables. Matrix H (Eq. 5.1) shows such a dataset.

$$H = \begin{bmatrix} x_{11} & \cdots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{1m} & \cdots & x_{nm} \end{bmatrix} \tag{5.1}$$

In cases where large numbers of variables are observed (i.e., for large m), PCA can be used to approximate the data using a smaller and more manageable set of variables. It is particularly effective when linear dependence is present in the variables, that is, when some variables can be expressed or closely approximated by linear combinations of others.

PCA is accomplished through a linear transformation that maps the data onto a new coordinate system. The basis vectors in this new coordinate system are ordered, and have the property that the first basis vector lies in the direction of greatest variance in the data, the second lies in the direction of second greatest variance, and so on.

This prioritizing of basis vectors is useful for finding an efficient representation of the data. When the most significant basis vectors are retained and the least significant ones discarded, a close approximation to the original dataset is created that uses a smaller number of variables. Since these basis vectors are orthogonal, PCA is an attempt to reexpress a dataset using a small number of orthogonal components.

A standard technique for obtaining principal components relies on the singular value decomposition (SVD). Whereas PCA is a data analysis tool, SVD is a general matrix decomposition technique. It factors a matrix into a product of three other matricies

$$H = U\Sigma V^T$$

where H is any matrix, U and V are unitary (their rows and columns are linearly independent and have length 1) and Σ is diagonal (zero everywhere except on the main diagonal). V^T is the transpose of V.

If, as above, H is interpreted as a multivariable dataset and subjected to a singular value decomposition, U, Σ and V will contain information related to the PCA. Specifically, the principal components of H will be found in the rows of U. The variances they explain will be found in Σ (on its diagonal, $\sigma_{1\ldots n}$). The rows of product ΣV^T are sometimes referred to as principal component 'scores'.

To reexpress H as data-reduced $H^{'}$ using only its first l principal components, then, requires simple manipulations. First, the elements on the diagonal of Σ are arranged in decreasing order, as in Eq. 5.2.

$$diag(\Sigma) = \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \tag{5.2}$$

The rows and columns of U and V must be similarly rearranged. Then, the data reduced $H^{'}$ can be obtained by retaining only the first l rows of U, the first l elements in Σ, and the first l columns of V^T, as shown in eq. 5.3

$$H^{'}_{m\times n} = U_{l\times m}\Sigma_{m\times n}V^H_{n\times l} \tag{5.3}$$

Appendix B: Certificate of Ethical Acceptability

McGill

Research Ethics Board Office
McGill University
845 Sherbrooke Street West
James Administration Bldg., rm 429
Montreal, QC H3A 2T5

Tel: (514) 398-6831
Fax: (514) 398-4853
Ethics website: www.mcgill.ca/rgo/ethics/human

Research Ethics Board II
Certificate of Ethical Acceptability of Research Involving Humans

REB File #: 56-0905

Project Title: The effects of modifying room acoustics on the perception of reproduced sound

Applicant's Name: Dr. William L. Martens **Department**: Music

Status: Faculty

Granting Agency and Title (if applicable): Canadian Foundation for Innovation

This project was reviewed on Sept. 20, 2005 by

Expedited Review ✓
Full Review ____

Blaine Ditto, Ph.D.
Chair, REB II

Approval Period: Sept. 20, 2005 **to** Sept. 19, 2006

This project was reviewed and approved in accordance with the requirements of the McGill University Policy on the Ethical Conduct of Research Involving Human Subjects and with the Tri-Council Policy Statement on the Ethical Conduct of Research Involving Human Subjects.

---

* All research involving human subjects requires review on an annual basis. A Request for Renewal form should be submitted at least one month before the above expiry date.
* When a project has been completed or terminated a Final Report form must be submitted.
* Should any modification or other unanticipated development occur before the next required review, the REB must be informed and any modification can't be initiated until approval is received.

McGill University

ETHICS REVIEW
RENEWAL REQUEST/FINAL REPORT

REB File #: 56-0905
Project Title: The Effects of Modifying Room Acoustics on the Perception of Reproduced Sound

Principal Investigator: William L. Martens
Department/Phone/Email: Music / 398-4535 (ext. 089795) / wlm@music.mcgill.ca
Faculty Supervisor (for student PI):

1. Were there any significant changes made to this research project that have any ethical implications? ___Yes _X_No
   If yes, describe these changes and append any relevant documents that have been revised.

2. Are there any ethical concerns that arose during the course of this research? ___ Yes _X_ No. If yes, please describe.

3. Have any subjects experienced any adverse events in connection with this research project? ___ Yes _X_ No
   If yes, please describe.

4. _X_ This is a request for renewal of ethics approval.

5. ____ This project is no longer active and ethics approval is no longer required.

6. List all current funding sources for this project and the corresponding project titles **if not exactly the same** as the project title above. Indicate the Principal Investigator of the award if not yourself.

Principal Investigator Signature: _____ Date: Sept 6, 2006

Faculty Supervisor Signature: _____ Date: _____
(for student PI)

For Administrative Use **REB:** ___AGR ___EDU ___REB-I ___REB-II

___ The closing report of this terminated project has been reviewed and accepted

___ The continuing review for this project has been reviewed and approved

___ Expedited Review ___ Full Review

Signature of REB Chair or designate: _____ Date: Sept 6. 2006

Approval Period: Sept. 20, 2006 to Sept 19, 2007

Submit to Lynda McNeil, Research Ethics Officer, James Administration Bldg., rm 419, fax: 398-4644 tel:398-6831

(version October 2002)

References

[1] J. S. Abel and S. H. Foster. Method and apparatus for efficient presentation of high-quality three dimensional audio. U.S. Patent 5,596,644, 1997.

[2] F. Asano, Y. Suzuki, and T. Sone. Role of spectral cues in median plane localization. *The Journal of the Acoustical Society of America*, 88(1):159–168, 1990.

[3] D. R. Begault. *3-D sound for virtual reality and multimedia*. AP Professional, 1994.

[4] D. R. Begault and T. Erbe. Multichannel spatial auditory display for speech communications. *J. Audio Eng. Soc.*, 42(10):819–26, 1994.

[5] J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization (Revised Edition)*. MIT Press, Cambridge, Massachusetts, 1997.

[6] C. Phillip Brown and Richard O. Duda. A structural model for binaural sound synthesis. *IEEE Transactions on speech and audio processing*, 6(5):476–488, 1998.

[7] R.A. Butler and K. Belendiuk. Spectral cues utilized in the localization of sound in the median sagittal plane. *J. Acoust. Soc. Amer.*, 61(5):1264–1269, 1977.

[8] D.W. Chandler and D.W. Grantham. Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity. *J. Acoust. Soc. Amer.*, 91(3):1624–1636, 1992.

[9] J. Chen, B.Ṽan Veen, and K. Hecox. External ear transfer function modeling: A beamforming approach. *J. Acoust. Soc. Amer.*, 92(4):1933–1943, 1992.

[10] J. Chen, B.Ṽan Veen, and K. Hecox. A spatial feature extraction and regularization model for the head related transfer function. *J. Acoust. Soc. Amer.*, 97(1):439–452, 1992.

[11] M. J. Evans, James A. S. Angus, and Anthony I. Tew. Analyzing head-related transfer functions using surface spherical harmonics. *J. Acoust. Soc. Amer.*, 104(4):2400–2411, 1998.

[12] R.L. Freyman, K.S. Helfer, D.D. McCall, and R.K. Clifton. The role of perceived spatial separation in the unmasking of speech. *J. Acoust. Soc. Amer.*, 106:3578, 1999.

[13] K. Genuit et al. Method and apparatus for simulating outer ear free field transfer function, June 9 1987. US Patent 4,672,569.

[14] D.W. Grantham, B.W.Y. Hornsby, and E.A. Erpenbeck. Auditory spatial resolution in horizontal, vertical, and diagonal planes. *J. Acoust. Soc. Amer.*, 114:1009, 2003.

[15] A. Harma, M. Karjalainen, L. Savioja, V. Valmaki, U. Laine, and J. Huopaniemi. Frequency-warped signal processing for audio applications. *J. Audio Eng. Soc.*, 48(11):1011–1031, 2000.

[16] J. Hebrank and D. Wright. Spectral cues used in the localization of sound sources on the median plane. *J. Acoust. Soc. Amer.*, 56(6):1829–1834, 1974.

[17] J. Huopaniemi, N. Zacharov, and M. Karjalainen. Objective and subjective evaluation of head-related transfer function filter design. *J. Audio Eng. Soc.*, 47(4):218–239, 1999.

[18] K. Iida, M. Yairi, and M. Morimoto. Role of pinna cavities in median plane localization. *J. Acoust. Soc. Am*, 103:2844, 1998.

[19] L. Jeffress. A place theory of sound localization. *Journal of Comparative and Physiological Psychology*, 41(1):35–39, 1948.

[20] J.M. Jot, V. Larcher, and J.M. Pernaux. A comparative study of 3-d audio encoding and rendering techniques. In *16th international conference*, Rovaneimi, 1999. Audio Engineering Society.

[21] D. J. Kistler and F. L. Wightman. The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Amer.*, 91(3):1648–1661, 1992.

[22] D. J. Kistler and F. L. Wightman. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Amer.*, 91(3):1637–1647, 1992.

[23] M. Kleiner, B.I. Dalenback, and P. Svensson. Auralization- an overview. *J. Audio Eng. Soc.*, 41(11):861–875, 1993.

[24] A. Kulkarni and H. S. Colburn. Role of spectral detail in sound source localization. *Nature*, 396:747–749, 1998.

[25] A. Kulkarni, S. K. Isabelle, and H. S. Colburn. Sensitivity of human subjects to head-related transfer-function phase spectra. *J. Acoust. Soc. Am*, 105(5):2821–2840, 1999.

[26] Abhijit Kulkarni and H. S. Colburn. Infinite-impulse-response models of the head-related transfer function. *J. Acoust. Soc. Am*, 115(4):1714–1728, 2004.

[27] V. Larcher, J. M. Jot, J. Guyard, and O. Warusfel. Study and comparison of efficient methods for 3D audio spatialization based on linear decomposition of HRTF data. In 108$^{\text{th}}$ *Convention of the Audio Engineering Society*, Paris, France, 2000.

[28] H. Levitt. Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Amer.*, 49(2):467–477, 1971.

[29] T. Lokki. *Physically-based Auralization: Design, Implementation, and Evaluation*. Helsinki University of Technology, 2002.

[30] J. Mackenzie, J. Huopaniemi, V. Valimaki, and I. Kale. Low-order modeling of head-related transfer functions usingbalanced model truncation. *Signal Processing Letters, IEEE*, 4(2):39–41, 1997.

[31] W. L. Martens. Principal components analysis and resynthesis of spectral cues to perceived direction. In *Proc. Int. Computer Music Conf.*, Champaine-Urbana, IL, Sept. 1987.

[32] W. L. Martens. Rapid psychophysical calibration using bisection scaling for individualized control of source elevation in auditory display. In *Proc. Int. Conf. on Auditory Display*, pages 199–206, Kyoto, Japan, 2002.

[33] W. L. Martens. Perceptual evaluation of filters controlling source direction: Customized and generalized HRTFs for binaural synthesis. *Acoustical Science and Technology*, 24(5):220–232, 2003.

[34] W. L. Martens. Binaural synthesis of indirect sound for positioning sources in small virtual acoustic environments: Effective yet unobtrusive global reverberation. In *Proc. of The 9$^{\text{th}}$ Western Pacific Acoustics Conference*, Seoul, Korea, June 2006.

[35] S. McAdams. Spectral fusion and the creation of auditory images. *Music, Mind, and Brain: The Neuropsychology of Music*, pages 279–298, 1983.

[36] D. McAlpine, D. Jiang, and A. Palmer. A neural code for low-frequency sound localization in mammals. *Nature Neuroscience*, 4:396–401, 2001.

[37] J.C. Middlebrooks, E.A. Macpherson, and Z.A. Onsan. Psychophysical customization of directional transfer functions for virtual sound localization. *J. Acoust. Soc. Amer.*, 108:3088, 2000.

[38] A. W. Mills. On the Minimum Audible Angle. *J. Acoust. Soc. Amer.*, 30(4):237–246, 1958.

[39] A. W. Mills. Auditory localization. In J. V. Tobias, editor, *Foundations of Modern Auditory Theory*, volume 2, pages 301–348. Academic Press, New York, 1972.

[40] M. Morimoto. The contribution of two ears to the perception of vertical angle in sagittal planes. *J. Acoust. Soc. Amer.*, 109(4):1596–1603, 2001.

[41] M. Morimoto and H. Aokata. Localization cues of sound sources in the upper hemisphere. **J**. **A***coust.* **S***oc.* **J***apan (E)*, 5(3):165–173, 1984.

[42] M. Morimoto and H. Aokata. Localization cues of sound sources in the upper hemisphere. **J**. **A***coust.* **S***oc.* **J***apan (E)*, 5(3):165–173, 1984.

[43] M. Morimoto, K. Iida, and M. Itoh. Upper hemisphere sound localization using head-related transfer functions in the median plane and interaural differences. *Acoustical Science and Technology*, 24(5):267–275, 2003.

[44] F. Opolko and J. Wapnick. *McGill University Master Samples: MUMS.* McGill University, Faculty of Music, 1989.

[45] S. Perrett and W. Noble. The effect of head rotations on vertical plane sound localization. *J. Acoust. Soc. Amer.*, 102:2325, 1997.

[46] D. R. Perrott and A. D. Musicant. Minimum auditory movement angle: Binaural localization of moving sound sources. *J. Acoust. Soc. Amer.*, 62(6):1463–1466, 1977.

[47] L. Rayleigh. On our perception of sound direction. *Philos. Mag*, 13:214–232, 1907.

[48] E. Rio, G. Vandernoot, and O. Warusfel. Perceptual evaluation of weighted multi-channel binaural format. In 6$^{\text{th}}$ *Int. Conference on Digital Audio Effects (DAFX-03)*, London, UK, 2003.

[49] E. Rio and O. Warusfel. Optimization of multi-channel binaural formats based on statistical analysis. In *Proc. of the Forum Acusticum*, Seville, Spain, 2002.

[50] K. Saberi and D.R. Perrott. Minimum audible movement angles as a function of sound source trajectory. *J. Acoust. Soc. Amer.*, 88(6):2639–2644, 1990.

[51] C. Searle, L. Braida, D. Cuddy, and M. Davis. Binaural pinna disparity: another auditory localization cue. *J. Acoust. Soc. Amer.*, 57(2):448–55, 1975.

[52] B.G. Shinn-Cunningham, S. Santarelli, and N. Kopco. Tori of confusion: Binaural localization cues for sources within reach of a listener. *J. Acoust. Soc. Amer.*, 107(3):1627–1636, 2000.

[53] T. Z. Strybel, C. L. Manligas, and D. R. Perrott. Minimum audible movement angle as a function of the azimuth and elevation of the source. *Hum. Factors*, 34(3):267–275, 1992.

[54] E. M. von Hornbostel and M. Wertheimer. Uber die wahrnehmung der schallrichtung (on the perception of the direction of sound). *Sitzungsber. K. Preuss. Akad. Wiss.*, pages 388 – 396, 1920.

[55] H. Wallach. The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology*, 27(4):339–368, 1940.

[56] E. Wenzel, M. Arruda, D.J. Kistler, and F.L. Wightman. Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Amer.*, 94(1):111–123, 1993.

[57] F. L. Wightman and D. J. Kistler. Headphone simulation of free-field listening. I: Stimulus synthesis. *J. Acoust. Soc. Am.*, 85:858–867, 1989.

[58] F. L. Wightman and D. J. Kistler. Headphone simulation of free-field listening. II: Psychophysical validation. *J. Acoust. Soc. Am.*, 85:868–878, 1989.

[59] F. L. Wightman and D. J. Kistler. Resolution of front–back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. Amer.*, 105:2841, 1999.

[60] J. Yu and E. Young. Linear and nonlinear pathways of spectral information transmission in the cochlear nucleus. *Proc. Natl. Acad. Sci. USA*, 97(22):11780–11786, 2000.

[61] B. Zhou, D.M. Green, and J.C. Middlebrooks. Characterization of external ear impulse responses using Golay codes. *J. Acoust. Soc. Amer.*, 92(2):1169–1171, 1992.

[62] D. Zotkin, J. Hwang, R. Duraiswaini, and L. Davis. HRTF personalization using anthropometric measurements. *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, pages 157–160, 2003.