

ABSTRACT

Title of dissertation: **SOUND LOCALIZATION BY
ECHOLOCATING BATS**

Murat Aytekin, Doctor of Philosophy, 2007

Dissertation directed by: **Professor Cynthia F. Moss**
Neuroscience and Cognitive Science Program

Echolocating bats emit ultrasonic vocalizations and listen to echoes reflected back from objects in the path of the sound beam to build a spatial representation of their surroundings. Important to understanding the representation of space through echolocation are detailed studies of the cues used for localization, the sonar emission patterns and how this information is assembled.

This thesis includes three studies, one on the directional properties of the sonar receiver, one on the directional properties of the sonar transmitter, and a model that demonstrates the role of action in building a representation of auditory space. The general importance of this work to a broader understanding of spatial localization is discussed.

Investigations of the directional properties of the sonar receiver reveal that interaural level difference and monaural spectral notch cues are both dependent on sound source azimuth and elevation. This redundancy allows flexibility that an echolocating bat may need when coping with complex computational demands for sound localization.

Using a novel method to measure bat sonar emission patterns from freely behaving bats, I show that the sonar beam shape varies between vocalizations. Consequently, the auditory system of a bat may need to adapt its computations to accurately localize objects using changing acoustic inputs.

Extra-auditory signals that carry information about pinna position and beam shape are required for auditory localization of sound sources. The auditory system must learn associations between extra-auditory signals and acoustic spatial cues. Furthermore, the auditory system must adapt to changes in acoustic input that occur with changes in pinna position and vocalization parameters. These demands on the nervous system suggest that sound localization is achieved through the interaction of behavioral control and acoustic inputs. A sensorimotor model demonstrates how an organism can learn space through auditory-motor contingencies. The model also reveals how different aspects of sound localization, such as experience-dependent acquisition, adaptation, and extra-auditory influences, can be brought together under a comprehensive framework.

This thesis presents a foundation for understanding the representation of auditory space that builds upon acoustic cues, motor control, and learning dynamic associations between action and auditory inputs.

SOUND LOCALIZATION BY ECHOLOCATING BATS

by

Murat Aytekin

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2007

Advisory Committee:

Professor Cynthia F. Moss Chair/Advisor
Professor Timothy K. Horiuchi, Co-Advisor
Professor David D. Yager
Professor Jonathan Z. Simon
Professor Catherina Carr

© Copyright by
Murat Aytakin
2007

DEDICATION

ANNE ve BABAMA

ACKNOWLEDGMENTS

I consider myself one of the most fortunate graduate students, starting a career in science, because of the support and guidance that I received from my advisor and mentor, Dr. Cynthia F. Moss. Working with her deepened my respect and appreciation for the natural sciences, in particular, neuroethology. Her modesty, patience, commitment, and critical thinking are only some of the skills and virtues I continue to strive for as a young scientist.

Dr. Timothy K. Horuichi, my second advisor, is a truly multi-dimensional scientist. I am thankful for his stimulating discussions and guidance. Being in the mid-field of engineering and biology he is living my dream.

I thank Jonathan Z. Simon for his help and collaboration with the sensorimotor model for auditory space learning. His help made it possible to bring this work to its current state. I am looking forward to more collaboration with him in taking this work to the next level.

I thank Dr. David D. Yager, a great scientist and teacher, for being a member of my dissertation committee. Who would have guessed that I would learn to respect cockroaches?

I also thank Dr. Catherine E. Carr for being in my dissertation committee and for making herself available despite her busy schedule.

I thank the NACS faculty. I have learned a lot from them.

I thank Dr. Manjit Sahota and Dr. Elena Grassi for their collaboration with the HRTF work. I thank Dr. Bradford J. May for his comments on the manuscript that is part of this

thesis. Also to Dr. Mark Holderied for his collaboration with the HRTF and pinna position relations study.

I thank Dr. P. S. Krishnaprasad for his comments at the beginning stage of the sensori-motor work. I thank Dr. Ramani Duraiswami and Dr. Dimitry Zotkin for their discussions on the reciprocity problem of sonar emission shape measurements.

My thanks also goes to Fei Sha for providing assistance and software for conformal component analysis for manifold learning.

I thank the past and current BATLAB members – particularly Amy Kryjak, Kari Bohn, Amaya Perez, Aaron Schurger, Rose Young, Tameeka Tameeka Williams (my partner in object discrimination experiments), Ben Falk, Dr. Kaushik Ghose, Dr. Shiva Sinha and Dr. Ed Smith – for their support and friendship. Additional thanks to Kasuhik for his English editing of the thesis.

I thank Poul Higgins for his help with English editing of the earlier version of the thesis.

I thank Tim Stranges, the executive director of the Conflict Resolution Center of Montgomery County (CRCMC), for making space available in the office to write my thesis. I am also grateful to the CRCMC staff – Marcia Cartagena, Margie Henry, and Peter Meleney – for accommodating my presence.

I am thankful to a number of people from Yıldız Technical University in İstanbul, Turkey, who encouraged me to come to the United States for Ph.D work and supported my education: Prof. Metin Yücel; Dr. Atilla Ataman; Dr. Tulay Yıldırım; the faculty of Electronics and Telecommunication Department, the staff of the Department of Electrical and Electronic Engineering; and Hülya Hanım in the business office.

I thank two important people who were indispensable with their support and guidance

with my scientific ambitions: my dear teachers Artemel Çevik and late Fuat Alkan.

I am grateful for my second family in the USA: Helen and Laura for their patience and love. They embraced me as one of their own. I am forever in debt to my love, Monika, from whom I am learning to be a better person.

I thank my roommate and friend Mary Fetzko (the little bug) for her friendship, her patience with English editing, and her camera work during my defense.

Mom and Dad, you are the source of my inspiration and passion. You have shown me how to be open to what life brings and appreciate to how it can help me change for the better. Unlimited honesty, goodwill, and unconditional love for family are values I hope to carry forward. If my graduation is a success, it is a product of your accomplishment. I hope that whatever I choose to do with my life will bring you the well deserved pride and happiness.

Burcu and Burçak, my litter sisters, you are God's gift to our family. I love you both very much.

I am thankful to echolocating bats, true miracles of nature, for opening some of their secrets to me.

Lastly, whether believing in a higher power is scientific or not, Allah, The One, Love, I am thankful for the beauty of your creation that is worth spending a lifetime to appreciate and learn.

Funding

The work presented in this thesis is supported by grants from National Science Foundation (IBN-0111973 and IBN-0111973), National Institute of Health (R01-MH056366), and Comparative and Evolutionary Biology of Hearing Center grant P30-DC004664.

List of Figures

| | | |
|------|--|----|
| 2.1 | Schematic of the frame used to suspend the bat | 54 |
| 2.2 | Maximum gain of head related transfer functions | 59 |
| 2.3 | Transfer functions for S4 | 61 |
| 2.4 | Left ear DTF contour plot for sagittal planes | 63 |
| 2.5 | Primary spectral notch center frequency in frontal space | 65 |
| 2.6 | Spatial contour maps of HRTF magnitude | 67 |
| 2.7 | Frequency-dependent changes in acoustical axis orientation | 68 |
| 2.8 | Directivity index of S1, S2, and S4 HRTFs | 70 |
| 2.9 | Interaural level difference (ILD) spatial contour maps | 72 |
| 2.10 | Angle of spatial gradient vectors of ILD | 73 |
| 2.11 | Interaural time difference (ITD) spatial contour map for S1. | 74 |
| 3.1 | Experimental apparatus. | 87 |
| 3.2 | Narrowband data acquisition. | 88 |
| 3.3 | Frequency response of the Knowles microphones | 89 |
| 3.4 | Reference microphone performance | 92 |
| 3.5 | The marker-headpost | 93 |

| | | |
|------|--|-----|
| 3.6 | Head Tracking | 94 |
| 3.7 | Head pose estimation accuracy | 95 |
| 3.8 | Heterodyne tuning frequency stability | 97 |
| 3.9 | Dynamic range of a typical channel | 98 |
| 3.10 | Speaker beams obtained by the microphone array (35 to 50 kHz) | 101 |
| 3.11 | Speaker beams obtained by the microphone array (55 to 70 kHz) | 102 |
| 3.12 | Ultrasound loudspeaker directivity | 103 |
| 3.13 | Single microphone versus the microphone array | 104 |
| 3.14 | Beam-to-noise ratio for loudspeaker beams | 105 |
| 3.15 | Bat sonar beams (35 to 50 kHz) | 107 |
| 3.16 | Bat sonar beams (55 to 70 kHz) | 108 |
| 3.17 | Spatial variation of bat sonar beams | 109 |
| 3.18 | Bat sonar beamshape from raw data envelope (35 to 50 kHz) | 110 |
| 3.19 | Bat sonar beamshape from raw data envelope (55 to 70 kHz) | 111 |
| 3.20 | Reference microphone and spatial variance | 113 |
| 3.21 | Consequences of sonar beamshape variability | 121 |
| 4.1 | Total local dissimilarity for each human subject | 143 |
| 4.2 | Uniformity of the local tangent space dimensions | 146 |
| 4.3 | Mean local distances of extended-tangent vectors and HRTFs | 148 |
| 4.4 | Learned global coordinates of auditory space of two echolocating bats | 150 |
| 4.5 | Mean local distances of extended-tangent vectors and HRTF (echolocating bats) | 152 |

Contents

1 Introduction:

| | |
|---|----------|
| Sound localization and bats | 2 |
| 1.1 Echolocation | 2 |
| 1.2 Sound localization | 4 |
| 1.2.1 Computation of the vertical component of sound sources | 7 |
| 1.2.2 Are two ears necessary? | 9 |
| 1.2.3 Spatial processing by the auditory system | 15 |
| 1.2.4 Evidence for separate pathways to process vertical and horizontal components of sound location | 16 |
| 1.2.5 Compartmentalized computation of the binaural difference cues | 20 |
| 1.2.6 Computation of sound localization is experience-dependent | 26 |
| 1.2.7 Multi-modal processing of sound location | 29 |
| 1.2.8 Summary | 32 |
| 1.3 Sound localization by bats | 34 |
| 1.3.1 Bat sound localization accuracy | 36 |

| | | |
|----------|---|-----------|
| 1.3.2 | Sound location cues for echolocating bats | 39 |
| 1.3.3 | Neural mechanisms for bat sound localization | 42 |
| 1.3.4 | Summary | 43 |
| 1.4 | Outline of the dissertation | 45 |
| 2 | The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation | 48 |
| 2.1 | Introduction | 49 |
| 2.2 | Methods | 53 |
| 2.2.1 | Animal preparation | 55 |
| 2.2.2 | Data acquisition | 56 |
| 2.2.3 | Signal processing | 57 |
| 2.3 | Results | 58 |
| 2.3.1 | Acoustic gain | 58 |
| 2.3.2 | Monaural spectral properties | 62 |
| 2.3.3 | Monaural spatial features | 65 |
| 2.3.4 | Interaural level differences | 70 |
| 2.3.5 | Interaural time differences | 73 |
| 2.4 | Discussion | 74 |
| 2.4.1 | A model for sound localization | 77 |
| 2.4.2 | Summary and conclusions | 80 |
| 3 | Sonar beam patterns of the freely vocalizing echolocating bat | 81 |
| 3.1 | Introduction | 81 |

| | | |
|----------|--|------------|
| 3.2 | Methods | 86 |
| 3.2.1 | The approach | 86 |
| 3.2.2 | Apparatus | 86 |
| 3.2.3 | Data acquisition | 87 |
| 3.2.4 | Experiment and sonar beam construction | 95 |
| 3.2.5 | Stability of the measurement set up | 96 |
| 3.3 | Results | 99 |
| 3.3.1 | Beam patterns of a stable emitter | 99 |
| 3.3.2 | Beam patterns of bat sonar calls | 105 |
| 3.3.3 | Reference microphone and spatial variance | 112 |
| 3.4 | Discussion | 113 |
| 3.4.1 | Implications of variable sonar beam patterns | 117 |
| 3.5 | Conclusions and future work | 122 |
| 4 | A sensorimotor approach to sound localization | 125 |
| 4.1 | Introduction | 126 |
| 4.1.1 | The sensorimotor approach | 128 |
| 4.1.2 | Sound localization and sensorimotor approach | 131 |
| 4.1.3 | Manifold of spatial acoustic changes | 132 |
| 4.1.4 | Computation of the space coordinate system | 135 |
| 4.2 | Demonstration 1: Obtaining spatial coordinates in humans | 138 |
| 4.2.1 | Simulation of the sensory inputs | 139 |
| 4.2.2 | Determining the spatial parameters | 140 |

| | | |
|----------|---|------------|
| 4.3 | Demonstration 2: Obtaining spatial coordinates by echolocating bats | 147 |
| 4.4 | Discussion | 151 |
| 4.4.1 | Geometry of the auditory space | 153 |
| 4.4.2 | Role of sensorimotor experience in auditory spatial perception . . . | 157 |
| 4.4.3 | Multi-sensory nature of spatial hearing | 161 |
| 4.4.4 | Role of the motor state in sound localization | 162 |
| 4.4.5 | Localization of sound sources without head movements | 164 |
| 4.4.6 | Neurophysiological implications | 165 |
| 4.4.7 | Applications to robotics | 166 |
| 4.5 | Conclusion | 167 |
| 5 | General Discussion | 168 |
| | Bibliography | 175 |

İlim ilim bilmektir.
(Knowledge means knowing knowledge.)
İlim kendin bilmektir.
(Knowledge is to know oneself.)
Sen kendini bilmezsin,
(If you do not know yourself,)
Ya bu nice okumaktır.
(What is the point of studying.)

Yunus Emre (d. 1320?)

Chapter 1

Introduction:

Sound localization and bats

1.1 Echolocation

Echolocating bats emit ultrasound pulses and listen to the echoes reflected back from the objects around them to build a representation of their surroundings. As nocturnal hunters they rely mainly on their biological sonar system to navigate, detect, and capture prey. This specialized utilization of the auditory system has been the interest of many researchers. How is space represented in the bat's brain and how detailed is this representation? How is the bat's auditory system different from that of other mammals? Can other mammals, such as humans, learn to use echolocation? These are only some of the questions that have motivated many studies in the field of biosonar research.

Although the details of how space is represented in the bat brain still remain to be fully clarified, our understanding of this topic continues to grow. Among the many interesting

aspects of echolocation, perhaps the most important one is the ability of echolocating bats to determine the relative positions of echo sources in their surroundings. Without localization of echo sources, a complete spatial representation of the bat's immediate environment cannot be realized. Studies focusing on sound localization suggest that the bat's localization accuracy is close to, if not better than, that of humans (Schnitzler and Henson, 1980).

Sound localization research has been conducted on a large variety of animals using a wide range of techniques. The study of sound localization in bats has proven difficult, due to technical challenges in experimental design, a problem that is common to behavioral studies of hearing in many species of animals. Consequently, the studies directly addressing sound localization by bats are sometimes limited in their interpretation and ability to address wider questions. Because the anatomy and function of the auditory system is largely preserved in mammals, a better understanding of sound localization can be achieved by combining the knowledge gained from studying different species.

This work is aimed at expanding our understanding of how and where bat sound localization fits into the current model of spatial hearing. Using the echolocating bat as a model system this work also aims to improve our understanding of sound localization in general. In accordance with these goals, it is important to remember motivations and background of the important issues as well as the current model of mammalian sound localization, before we move onto an introduction of our current understanding of bat sound localization.

1.2 Sound localization

“Localization is the law or the rule by which the location of an auditory event is related to a specific attribute or attributes of a sound event, or of another event that is in some way correlated with the auditory event to the position of the sound source ...”

Blauert (1997)

The definition given above summarizes the motivation and the objective of sound localization research very effectively. Listeners, human or otherwise, perceive sound sources as located at a particular location external to their body. The spatial perception of the sound is elicited by the interpretation of the acoustic signals at the ears by the listener’s the auditory system. The brain computes the spatial attributes within these signals and establishes where the sound originated in space. Sound localization research focuses on what these special attributes are, how they are extracted and interpreted by a listener.

Sounds arriving at a listener are transformed by their interaction with the listener’s body, head, and external ears. This transformation is a function of the location of the sound source and causes the spectra of the two signals received at either ear to be different from each other. The direction-dependent differences between the two ears are mostly studied in two categories, interaural time differences (ITDs) and level differences (ILDs). ITD is the result of different path lengths that a sound has to travel to reach to each ear. The time difference increases with the horizontal offset of the sound position from the median plane¹. Similarly, a sound signal arriving to the far ear will be partially blocked by the

¹ The median plane passes through the midline of the head, dividing it to right and left halves.

head and weakened. Consequently, its intensity at the far ear will be smaller compared to the sound intensity at the closer ear. This level difference in the acoustic signal received at both ears will be greater for larger horizontal angles with reference to the median plane.

Lord Raleigh performed a detailed study on the operating frequency range of ITD and ILD cues in humans. Based on this influential study Raleigh proposed the duplex theory (see, references in Macpherson and Middlebrooks, 2002). According to this theory, at low frequencies ITD is the dominant cue for sound localization in azimuth while at high frequencies ILD is the more dominant cue. Although the duplex theory has been influential in understanding sound localization, it is limited to interaural difference cues. For most animals, and particularly for humans, interaural difference cues are not useful in determining the vertical component of the sound location. The left-right symmetry of the head results in so-called “cones of confusion”, conic surfaces whose main axis coincides with the interaural axis, on which ILD and ITD cues remain invariant (Blauert, 1997). Consequently, binaural difference cues for sound sources placed on a cone of confusion will be ambiguous and not sufficient for localization. The ambiguity of this position is most obvious in the median plane. Sound sources placed in the median plane cause zero ITD and ILD values regardless of what vertical position they occupy.

The insufficiency of the duplex theory to explain sound localization completely created a new surge of effort in sound localization research. One interesting result of psychoacoustical investigations revealed that while horizontal location of a narrowband or a pure tone stimulus can be localized accurately, its perceived vertical position is dependent on frequency and independent of where it is located in space (Butler and Helwig, 1983; Musicant and Butler, 1984; Middlebrooks, 1992; Blauert, 1997). These and similar results

established that the cues associated with the computation of the vertical component of the sound source requires broadband signals. This conclusion also led to the hypothesis that the auditory system computes horizontal and vertical components of sound source positions independently (Middlebrooks and Green, 1991; Hofman and Van Opstal, 2003). Subsequently, how the auditory system could determine the vertical component of the sound source location became the question.

Later studies focused on the direction-dependent effects of the pinnae on the received sound spectrum (Batteau, 1967; Shaw and Teranishi, 1968). It is now understood that the spectral cues created by the directional effects of the pinnae on sound signals are used by the auditory system to determine the source elevation (Hofman and Van Opstal, 2002; Huang and May, 1996; Middlebrooks and Green, 1991).

The direction-dependent effects of the pinna can be modeled as linear time invariant filters. An acoustic signal at the ears caused by a sound source can be replicated by the filtering of the acoustic signal at the sound source. These filters can be measured for each sound source position relative to the listener's head (Wightman and Kistler, 1989a). The replicated signals when fed directly to the ears are perceived in space (externalization) and cannot be discriminated from real sound sources positioned at the simulated location (Wightman and Kistler, 1989b). This implies that the transfer functions of these filters, also known as head-related transfer functions (HRTF), completely capture the directional effects of the pinna, the head, and the torso on sounds arriving to the ears. Therefore HRTFs contain all directional-acoustic cues for sound localization. The availability of HRTF measurements has given rise to the development of virtual sound techniques, which allowed researchers to design experiments that manipulate directional features of the sounds pos-

sible. With these techniques it is now possible to investigate the relationship between a subject's localization performance and the directional features in the HRTF, hence isolate the particular directional features that could be used by the auditory system.

1.2.1 Computation of the vertical component of sound sources

One of the most challenging issues in sound localization research is determining what spatial cues are actually being used by the auditory system. Although the answer seems to be obvious for the horizontal component of the sound source location, i.e. usage of ITD and ILD, spatial cues for the vertical component are still not clearly understood. Investigations on vertical localization has revealed that the vertical position of a sound conveyed by the spectral shape of the acoustic signals at the ears (Hofman and Van Opstal, 2002; Huang and May, 1996; Middlebrooks and Green, 1991). Acoustic-spectral cues for the vertical components of sound source locations have been mostly studied in the median plane. On the median plane both ILD and ITD are zero and therefore cannot interfere or contribute to the computation of the vertical component. Having signals practically identical at both ears for any position in the median plane raised the question of whether signals from single ear are sufficient for localization under these conditions. Some experiments have tested subjects' ability to use only one ear to localize the vertical position of a sound source and shown that it is possible to localize sound using monaural spectral features by normal listeners (Hebrank and Wright, 1974; Slattery and Middlebrooks, 1994). Because localization of the horizontal component requires binaural difference cues, these results further supported the dichotomy of computation of sound the sources' horizontal and vertical

components.

Narrowing down the search for vertical localization cues to monaural spectral features, focus was drawn towards the spectral features of the HRTF. Studying sound localization in the median plane provided an ideal experimental paradigm because it forced subjects to use the spectral cues to localize. One of the most studied robust feature is the deep spectral notches in the HRTF whose center frequency has a linear relationship to the vertical angle of the sound source in the median plane (Rice et al., 1992; Wotton, Haresign, and Simmons, 1995; Blauert, 1997). This feature has been found to be common within the species that are studied and shown to be related to the concave structure of the ear (Lopez-Poveda and Meddis, 1996). Depending on the species, there can be multiple notch patterns at different frequency intervals (Rice et al., 1992; Wotton, Haresign, and Simmons, 1995). The illusion of the vertical motion created by the insertion of a notch to the spectrum of a broadband noise stimulus and moving its center frequency systematically supported the notch hypothesis (Bloom, 1977; Watkins, 1978). Huang and May (1996) have shown that cats' sound-evoked orientation responses are accurate if the sound stimuli are provided with the frequency range where the spectral notch cues are prominent. Moreover, May and Huang (1997) have shown that spectral notches can be well represented by the auditory nerve responses suggesting that the auditory system could utilize these cues to compute vertical components of the sound source locations. The single-notch hypothesis which proposes the first notch trend in the frequency axis as the elevation cue is based on these results (Rice et al., 1992).

Despite this converging evidence, recent studies questioned the significance of the individual notch patterns for the computation of the vertical component. Langendijk and

Bronkhorst (2002) have shown that removing the spectral notch from the virtual acoustic stimuli (VAS), which is otherwise identical to the free field sound at a particular location in the median plane, did not effect the localization performance. They concluded that the elevation cues are distributed across the spectrum suggesting a robust computational scheme for localization that is not limited to a single local spectral feature like the primary spectral notch.

A different line of experiments where the HRTFs were smoothed by reconstructing them using five principle components, revealed that localization does not require the detail of the spectral shape of the HRTF (Kistler and Wightman, 1992). These reconstructed HRTFs contained the robust features like the spectral notches. Based on this line of research it can be concluded that the reliance on a single feature, e.g spectral notch, in HRTF is not significant and the computation of the vertical component of a sound source is most likely based on combination of multiple spectral features, e.g. multiple notches at different frequencies. Redundant cues for elevation result in reliable computation of the vertical component of a sound source position.

1.2.2 Are two ears necessary?

Results of the monaural localization studies, mentioned above, gave rise to a paradox. Acoustic signals received at the ears are mixed, i.e convolved, with the associated HRTF, and they are a function of sound source spectrum and location. Both qualities are merged together and it is not clear how the auditory system can extract out features that are associated solely with the position of the sound. This ambiguity is not a problem for the

computation of the horizontal component of sound source locations using ITD and ILD cues since these are only a function of horizontal position of the sound source. Subjects participate in monaural localization experiments have to identify the spectral cues associated with the vertical location of the sound source to localize. How then, can they separate the sound source spectrum and directional filtering of the ear the received signal spectrum without having prior information about the sound source?

One hypothesis is that the spectra of natural sounds do not include fast and large transients such as notches that are observed in HRTFs (Brown and May, 2003). Thus, if the auditory system takes fast changes in the spectrum to be spatial cues and uses them to determine sound source location, it will be successful as long as sound source spectra are smooth. The constraint that natural sounds have smooth spectra is not valid. For instance, echolocating bats experience non-smooth echo spectrum very often (Simmons et al., 1995). Zakarauskas and Cynader (1993) proposed that these conditions can be relaxed if the gradient of the sound source spectrum is required to be smoother than certain spectral features of the HRTF. Validity of these constraints has not been systematically investigated. However, Macpherson (1998) has shown that his subjects' localization performance with sound spectra manipulated in different ways were not well explained by the type of computational method suggested by Zakarauskas and Cynader.

Macpherson (1998) has shown that, unlike earlier findings (Hebrank and Wright, 1974; Searle et al., 1975), the auditory system is robust against the irregularities in the sound source spectrum to a limited extent. However, this robustness is not limitless. Hofman and Van Opstal (2002) showed that if the sound source spectrum includes random variations above three cycles-per-octave with amplitude root mean square value 7.5 dB on the fre-

quency axis, an illusion of position change in the median plane can occur. Macpherson and Middlebrooks (2003) investigated vertical localization performance using rippled-spectrum noise systematically to determine the limits of the auditory system's ability to discriminate spatial features from the signals received at the ears under binaural listening conditions.

Wightman and Kistler (1997) challenged conclusions that vertical localization can be achieved monaurally. They argued that many monaural localization studies used ear plugs which are not perfect insulators and leakage of acoustic signal to the blocked ear is always possible. They were also concerned that the spectral characteristics of the sound stimuli being used could be learned by the subjects and this familiarity could facilitate better localization. Theoretically, if either the sound source or its location is known, it is possible to deconvolve the unknown component from the signal received at the ears. Wightman and Kistler tested their subjects ability to localize sound in the far field using virtual sound techniques that guaranteed monaural listening conditions. In order to remove familiarity to sound stimuli as a factor, they scrambled the sound spectrum profile. They showed that subjects could localize the sound source accurately under binaural listening conditions, but not under monaural listening conditions. Their results thus invalidated earlier conclusions on the monaural localization of vertical positions of sound sources. Further evidence that challenges the monaural processing view followed.

Studies on chronic monaural listeners have shown that these subjects can utilize the spectral cues in the HRTF to localize sounds in azimuth and elevation (Van Wanrooij and Van Opstal, 2004). This ability breaks down if the sound source intensity is unpredictable (Wanrooij and Opstal, 2007), confirming the results of Wightman and Kistler (1997). As a result, evidence from these lines of study suggest that monaural localization is possible

under very limited conditions. In contrast, binaural localization is very robust at least under broadband listening conditions.

Such results indicated that the computation of both horizontal and vertical components of sound sources require acoustic signals from both ears. A number of studies investigated the relative contribution of each ear to sound localization. Hofman and Van Opstal (2003) tested how the localization of sound sources in the frontal hemisphere (within -30° and 30° in the vertical and horizontal angles) is influenced when subjects wear ear-molds that eliminates spectral cues in one or both ears. They found that both in monaural and binaural mold conditions where subjects wear the one or two ear-molds, respectively, the horizontal component is accurately computed. The perceived vertical locations for each horizontal plane were collapsed to a single location. This location was linearly dependent on the sound source azimuth under the binaural mold condition. With monaural molds, sound source azimuths again localized accurately and the perceived elevations on the side of the mold correlated to the sound source elevation, but the accuracy was lower than no-mold control conditions. Perceived elevations for stimuli presented contralateral to the mold side, were closer to the target elevation values, yet in comparison to the control conditions they showed slightly different linear relations to them. Authors concluded that spectral information is important for vertical localization and both ears contribute to computation of this component. Further analysis by the authors showed that the auditory system might be weighting the cues contributed by both ears with reference to azimuth (Hofman and Van Opstal, 2003). Wanrooij and Opstal (2007) recently showed that the binaural weighting process is most likely to take place with reference to the perceived azimuth of the sound

source positions.

Jin et al. (2004) tested sufficiency of the interaural level difference cues as well as importance of the binaural information for vertical localization. In this study using virtual acoustic techniques, they presented a sound with flat spectrum to the left ear while presenting either the normal monaural spectrum or the binaural difference spectrum associated with a particular position in space to the right ear. They found that neither the preserved interaural spectral difference nor that the preserved monaural spectral cue was sufficient to maintain accurate localization of the vertical component when a sound with flat spectrum was presented to the other ear. Subjects could, however, localize the horizontal component of the sound source.

Note that the robust computation of the horizontal component and the sensitivity of the vertical component to monaural manipulation not only points to the importance of binaural signal processing for sound localization, but also to the independence of the computations of these components by the auditory system. Recently however, this independence is also questioned. Studies on chronic (Van Wanrooij and Van Opstal, 2004) and acute (Wanrooij and Opstal, 2007) effects of the monaural localization revealed that listeners adopt a strategy that uses monaural spectral cues to localize both vertical and horizontal components of the sound source. For acute monaural conditions, the switch in strategy is immediate. Suggesting the dependence of both computations to the cues created by the frequency and direction dependent shadowing effect of the head on the incoming sound source, hence the interdependence of the azimuth and elevation computations.

A different approach also provided a similar conclusion on the interdependence of the vertical and horizontal components' computation. Hartmann and Wittenberg (1996) stud-

ied whether subjects perceive an externalized and non-diffused sound source if one ear receives interaural level differences and the other sounds with flat spectrum. In theory under these conditions, binaural cues are intact. They progressively introduce this stimulus condition each time up to a certain frequency while keeping the rest of the signal identical to the normal listening conditions. They found that maintaining the ILD cues was not sufficient to create a normal experience of externalization of the sound. These results also bring a different perspective to Jin et al. (2004) study where the investigators did not test the externalization their subjects experienced.

Binaural processing is especially important under normal listening conditions. Sound spectrum can be unpredictable and require processing schemes that are robust enough to solve the ambiguity problem. The auditory system can take advantage of a priori information about the sound source, e.g. familiarity to sound source can contribute to the localization and this is very clear under monaural listening conditions. An important conclusion based on this line of evidence is that binaural listening help disambiguate spatial cues from the sound source spectrum and it is necessary for externalization and accurate localization of the sound sources. The auditory system utilizes spectral cues for both components of the sound location and is very efficient with combining the available cues to cope with different listening conditions.

The disambiguation, however, may not necessarily manifest itself as a deconvolution process that would allow computing of the HRTF for a known source spectrum or vice versa. A recent study (Rakerd, Hartmann, and McCaskey, 1999) shows that subjects cannot identify sounds positioned randomly in the median plane, whose distinguishing properties are at the high frequencies. By contrast, subjects can localize the positions of these sound

sources, in spite of the fact that vertical spatial cues are more robust at the high frequencies where an ambiguity between the sound-source spectra and localization cues is expected. These results suggest that listeners could not extract the sound spectrum via deconvolution. Consequently, the auditory system might be evaluating the acoustic features for localization and recognition rather than constructing a complete frequency representation of the sound source or the HRTF.

1.2.3 Spatial processing by the auditory system

In parallel with psychoacoustical studies of sound localization, neurophysiological studies have been conducted to identify neural mechanisms corresponding to the proposed computational strategies. Physiological investigations of sound localization have been mainly motivated by the cue-based model of sound-source localization. The cue-based model simply postulates that the location of a sound source is computed based on the directional-acoustic cues extracted from the signals received at the ears. More specifically, most neurophysiological studies focus on a particular form of the cue-based model that assumes independent computation of interaural difference and monaural spectral cues. Efforts have been made to find neural mechanisms in the auditory brain stem that are responsible for processing these sound localization cues.

Since the horizontal and vertical components of sound source location are believed to be computed independently, it is not unreasonable to hypothesize that the neural mechanisms that are associated with these computations are separate. Moreover, this compartmentalized form of the cue-based model makes specific predictions that are well suited to

electrophysiological investigation. One difficulty faced by this line of research is that for spatial acoustic cues to influence the sound location computation, their observable representations within the auditory system are not necessary. Thus, direct confirmation of the proposed models of sound localization through physiological studies could be problematic. However, their contribution to the understanding of spatial computation by the auditory system is vital and cannot be underestimated.

This discussion will be limited to the findings that strongly suggest that the binaural difference cues and monaural spectral cues are first computed separately within the auditory system before being combined together to represent sound location in space.

1.2.4 Evidence for separate pathways to process vertical and horizontal components of sound location

There are two main pathways that are believed to be involved in sound localization (Cohen and Knudsen, 1999). The first pathway branches from the inferior colliculus (IC), projects to the external nucleus of the IC (ICx), and ends at the deep sensory layers of the superior colliculus (SC), a sensory-motor nucleus responsible for orientation behavior (Jay and Sparks, 1987). Neurons in the SC that are sensitive to auditory spatial location are arranged topographically in alignment with the visual neurons in the superficial layers of the SC (Palmer and King, 1982; Middlebrooks and Knudsen, 1984). The SC also has cells that are driven by tactile stimuli and are organized somatotopically. In addition to sensory maps, motor neurons in the SC are also organized topographically with reference to space. Local stimulations in the motor map directs the eyes, the ears and the head to corresponding

positions in space. Sensory maps and the motor map are in alignment (Stein and Meredith, 1993).

The second pathway projects from the IC to the medial geniculate nucleus of the thalamus and later projects to the auditory cortex and frontal eye field (FEF) in the forebrain. The FEF sends descending projections back to SC and to the motor nuclei (Meredith and Clemo, 1989; Perales, Winer, and Prieto, 2006). This second pathway is not as well understood as the first partly because it does not have a topographic organization making it harder to investigate.

The relative contributions of these separate pathways to sound localization are not clear (see Cohen and Knudsen, 1999). Lesion studies reveal that auditory processing below the IC is necessary for sound localization (Casseday and Neff, 1975; Jenkins and Masterton, 1982). By contrast, severing either pathway after the IC does not eliminate auditory localization (Strominger and Oesterreich, 1970; Middlebrooks, L. Xu, and Mickey, 2002). However, localization tasks requiring navigation and memory are disrupted by lesioning of the auditory cortex. An animal's ability to discriminate source locations orient and/or approach towards a sound source is impaired in the region of space contralateral to the lesion site in the auditory cortex (see references in Middlebrooks, L. Xu, and Mickey, 2002). The auditory cortex is hypothesized to integrate and refine the auditory spatial information and relay it to other systems. In addition the auditory cortex facilitates and mediates learning and adaptation of spatial cues (King et al., 2007). For example, inactivation of the primary auditory cortex (AI) slowed the learning rate of adult animals, compared to control animals under altered listening conditions, involving unilateral ear plugs (King et al., 2007).

The IC-SC pathway seems to be involved in controlling reflexive orientations to sound

sources (Middlebrooks, L. Xu, and Mickey, 2002). Lesions of the auditory cortex does not seem to affect reflexive orientation. Lesions of the SC, however, eliminate an animal's ability to integrate multisensory information for localization in the space contralateral to the lesioned side (Burnett et al., 2004). Recent investigations determined that multisensory integration in SC is mediated by the anterior ectosylvian sulcus (AES) and the rostral lateral suprasylvian cortex (rLS) (Jiang, Jiang, and Stein, 2002; Jiang et al., 2007). Lesion of either cortical site replicates the behavioral observations obtained by SC lesions. Furthermore, physiological investigations showed that multisensory SC cell responses lose their cross-modal integrative functionality as a result of the inactivation of the AES and/or rLS. The cells that are driven by a single modality, however, remain unaffected (Jiang et al., 2001; Jiang, Jiang, and Stein, 2006).

As a result of these findings it is possible to conclude that both pathways can control sound-evoked orientation behavior that requires sound localization and control different aspects of the spatial orientation tasks. Two pathways interact to facilitate multisensory integration and spatial learning.

Below IC sound location computation shows compartmentalized processing areas that are specialized for different localization cues. The first separation happens at the level of the cochlear nucleus. This area is the first site that receives projections from the cochlea in the auditory system. Dorsal and ventral subnuclei of the cochlear nucleus, DCN and VCN respectively, are thought to be initial processing locations that give rise to two pathways specialized for monaural spectral and binaural difference cues (May, 2000).

DCN receives primarily ipsilateral projections from the cochlea. A subset of neurons with type-IV response properties show sensitivity to spectral features of the monaural signal

spectrum (Young and Davis, 2002). The nonlinear behavior of these cells are believed to be well suited to code for spectral notches. DCN gives direct projections to the IC. A set of neurons that show excitatory response to spectral notches in IC, classified as type-O, are hypothesized to be related to the projection of type-IV neurons in DCN (Davis, Ramachandran, and May, 2003).

The VCN pathway consist of the superior olivary complex (SOC) and the lateral lemniscus (LL). The SOC projects to LL and IC and LL projects to the IC (Yin, 2002). The SOC is known as the first area in the auditory system that is involved in binaural processing. Although DCN also receives contralateral inhibitory inputs in addition to ipsilateral excitatory inputs these projections are believed to enhance the processing of the spectral localization cues generated by pinna (Davis, 2005). The lateral and medial superior olives (LSO and MSO) are the two subnuclei of the SOC and specialized in processing ILD and ITD respectively (Yin, 2002). Neurons in MSO receive excitatory inputs from both sides and act as coincident detectors that are sensitive to instantaneous arrival of the binaural excitatory signals. MSO cells' ipsilateral inputs are delayed for a different amount that allow signals with different ITD to arrive simultaneously to different cells. This results in activation of separate subsets of MOS cells in response to different ITDs. Thus, creating a representation of the ITD (Yin, 2002).

Neurons in the LSO are excited by the unilateral projections and inhibited by the contralateral projections. Thus effectively compare two signals that represent sound pressure levels of acoustic signals received in both ears. Because ipsilateral and contralateral projections to LSO neurons change their firing rate linearly when sound pressure changes logarithmically (May and Huang, 1997), these neurons' are sensitive to ILD.

The evidence for the dichotomy of the monaural and binaural pathways comes from studies that systematically eliminated the DCN pathway monolaterally and bilaterally (Sutherland, Masterton, and Glendenning, 1998; May, 2000). Eliminating the DCN pathway, in cats caused significant deficits in localization of the vertical component of sound source locations, while, variation of the horizontal components showed an increase. Interestingly, DCN elimination did not cause abnormal discrimination of changes in sound source location in cats (May, 2000). This last finding suggests that different processing schemes are involved in localization of a sound source and in detecting a change in its position (May, 2000).

1.2.5 Compartmentalized computation of the binaural difference cues

As mentioned earlier ITD and ILD cues are thought to be processed in separately the auditory system. MSO and LSO are the first nuclei that show correlated responses to the binaural difference cues, ITD and ILD respectively. Studies on animals with monolateral lesions of SOC showed impairment in discriminating sound source locations on the left and the right lateral positions (Casseday and Neff, 1975; Kavanagh and Kelly, 1992; van Adel and Kelly, 1998).

ITD encoding

Neurons in MSO receive excitatory inputs from the ipsilateral and contralateral VCN (Yin, 2002). Excitatory projections converging on an MSO cell match in their characteristic

frequency (CFs). Cells respond when the two excitatory inputs from the left and the right VCNs arrive simultaneously acting as coincidence detectors. The ipsilateral projections are delayed in certain amounts, such that signals arriving with different time delays to the ears can stimulate different MSO cells, thus encoding the ITD (Yin, 2002). ITDs corresponding to the ipsilateral spatial positions to the MSO are predominantly represented. Most cell's CFs are in the low frequencies where ITD acts as the primary binaural difference cue (Stevens and Newman, 1936) and auditory fibers respond in phase with the acoustic energy in their CF (Pickles, 1988). These properties make MSO an ideal location to encode ITD.

The range of ITD experience by an animal is related to its head size. A small head size result in a small range of ITDs. If MSO's function is to code ITDs, it is reasonable to expect that animals with small heads should have smaller MSO for ITD may not be as efficient for a localization cue. However, a weaker than expected correlation between relative sizes of the head and MSO has been reported (Heffner, 1997). Moreover, certain species of mammals have been found to be exceptions to the hypothesized trends. For instance, echolocating bats have a larger MSO than one could expect from their head size if MSO is assumed to be responsible for ITD processing. Moreover, depending on the species of bats, MSO shows significant variation in terms of circuitry and responses to auditory signals (Grothe and Park, 2000). Other mammals with small heads also show ITD sensitivity but most cells' best ITD is beyond the range of time delays that can be caused by sound sources. However, an ITD encoding can still be accomplished by this type ITD response if the slopes of the ITD tuning curves are considered for ITD encoding instead of best ITDs (McAlpine and Grothe, 2003).

ILD encoding

LSO receives ipsilateral excitation and contralateral inhibition. They are matched in frequency and inputs arriving from ipsilateral DCN and ipsilateral medial nucleus of trapezoid body (MNTB) are in temporal register. MNTB converts the excitation it receives from the contralateral cochlea to inhibition (Tollin, 2003). Hence, an LSO unit would receive the relevant signals with the right signs to generate a response as a function of ILD. Furthermore, the CFs of most cells in the LSO are in high frequencies where ILD is a robust cue for sound localization (Macpherson and Middlebrooks, 2002). The CF of neurons in the high frequency region of hearing is distributed fairly uniformly (Yin, 2002). Most of the cells show a firing rate pattern that has sigmoidal relationship to the ILD favoring sound positions in the ipsilateral side (Yin, 2002). The firing rate is higher for the sound source positions ipsilateral to the LSO. The LSO cells react to temporal changes in the signals almost as fast as MNTB units (Joris and Yin, 1998). These cells can follow amplitude modulations with high fidelity and can compare bilateral inputs within short time periods. One interesting finding, however, is that the inhibition lasts relatively longer than excitation. Sanes (1990) has reported that inhibition lasts between in gerbils and physiologically this inhibition is functional for only about 1 to 2ms. Recent studies utilizing VAS methods to manipulate ILD, ITD, and interaural spectrum independently revealed that the LSO units are influenced predominantly by the ILD, supporting LSO's involvement in ILD processing (Tollin and Yin, 2002).

Despite the observations listed above that suggest LSO neurons are ideal ILD encoders, other properties of LSO cells reveal that the function of the LSO could be more

complicated. Only a small percentage of LSO cells show monaural level invariant responses to binaural stimuli (Irvine and Gago, 1990). Most cells' responses are influenced by the average binaural level (ABL), therefore they respond at different firing rates for the same ILD value at different sound intensities (Tsuchitani and Boudreau, 1969; Semple and Kitzes, 1993). Furthermore, some LSO cells are sensitive to ILD values beyond natural listening conditions. Although in dispute, some low frequency LSO cells seem to be driven monaurally (Yin, 2002).

ILD sensitivity has also been seen in the dorsal nucleus of the lateral lemniscus (DNLL). ILD sensitive cells from both LSO and DNLL project onto the contralateral inferior colliculus. These cells in the IC receive ipsilateral inhibition and contralateral excitation (IE cells). Because IC cells receive inputs from different nuclei, the ILD response patterns are more variable (Pollak and Park, 1995). A systematic study of the IC cell responses to binaural stimuli (Irvine and Gago, 1990) showed that approximately 72% of the IE cells showed sigmoidal response patterns to ILD and the rest of the cells showed response patterns between sigmoidal and peaked. Most cells, 77% to 89%, showed sensitivity to ABL. Most cells showed similar sensitivity patterns to broadband and tone stimuli but some showed marked differences (Irvine and Gago, 1990).

The fact that most ILD sensitive neurons observed in the auditory system are not purely a function of ILD has implications on the hypothesis that the ILD is represented in the auditory system at the level of a single neuron. Unless the processing of the ILD is limited to a small percentage of the cells observed in the IC, the ILD-dependent neural activity will vary in relation to the acoustic signal level. Alternatively, ILD and ABL sensitive neurons could give rise to a population coding scheme where ILDs are coded at particular sound

levels (Semple and Kitzes, 1987).

Encoding of monaural spectral cues

Efforts to find neural correlates of monaural spectral features are relatively more recent and have focused on cats. This is partly because it is not clearly understood what these cues are. Monaural spectral cues are hypothesized to be processed in the DCN (Young and Davis, 2002; May, 2000). DCN rate response patterns to acoustic stimuli are mostly similar to the VCN except for the type-IV cells that respond to sound nonlinearly. Type-IV cells react to broadband sound with notches in the spectrum. These cells are driven strongly by broadband noise, but are inhibited by spectral notches close to their CF (Young and Davis, 2002). The DCN projects directly to the IC. A group of cells in the IC, categorized as type-O cells, are excited by broadband stimuli with notches and appear to be related to the type-IV DCN cells (Davis, Ramachandran, and May, 2003).

Representation of auditory space

Unlike vision or somatosensation, the auditory sensory epithelium is not topographically correlated with the spatial positions of the sounds. Consequently, neither the outputs of the cochlea (auditory nerve responses) nor the neurons in the lower auditory brainstem show sensitivity to a small isolated part of space, i.e. a spatial receptive field. Although neurons with spatial receptive fields may not be necessary to perceive spatial qualities of sound sources (Kaas, 1997; Weinberg, 1997), finding neurons with spatial receptive fields is important because it helps to narrow the search for computational mechanisms of sound localization. Once such neurons are found, the computational evolution of auditory signals

at the ears can be observed and studied. For this reason search and study of neurons with spatial receptive fields has been a large part of sound localization research (Middlebrooks, L. Xu, and Mickey, 2002). What could particularly facilitate our understanding of spatial computation in the auditory system is a topographical organization of such neurons into a map of space. A topographic representation of space implies highly organized local circuitry that may make structure-function investigations easier to carry out (Weinberg, 1997).

A topographic representation of auditory space is found in the SC (Palmer and King, 1982; Middlebrooks and Knudsen, 1984) with varying degrees of elaboration in different mammals. Cells in SC increase their firing rate resulting in larger receptive fields with increasing sound intensity and some show shift in their best location although this shift is not significant in multi-unit responses (Campbell et al., 2006). SC cells show sensitivity to spectral cues (Carlile and Pralong, 1994) and ILD, but are mostly uninfluenced by ITD (Campbell et al., 2006). Insensitivity to ITD in SC is unexpected in light of the psychoacoustical data supporting its importance for sound localization (Wightman and Kistler, 1992) and not completely understood thus far.

A spatial topography for azimuth is also found in ICx, a site that sends the auditory projections to SC, but its topography is not as robust as the map in SC (Binns et al., 1992). It has been hypothesized that the ICx is involved in the development of the auditory map in SC (Thornton and Withington, 1996). Lesioning ICx in young guinea pigs (8 to 13 days old) results in larger spatial receptive fields in SC and the best azimuths of the cells do not show normal topographic order. A more selective disruption of the auditory map in SC was obtained when ICx lesions were limited to a smaller area: topographic representation of the part of space corresponding to the lesioned site of the ICx was not developed in

the SC. Other studies found that the development of the ICx map requires normal auditory experience and not dependent on the visual input (Binns, Withington, and Keating, 1995).

Both ICx and SC are part of the first sound localization pathway. Neurons with auditory spatial receptive fields have been also found in the auditory cortical areas (Middlebrooks, L. Xu, and Mickey, 2002). Unlike ICx and SC however no topographical organization for space has been found. Location sensitive neurons show a rather distributed representation of auditory space (Stecker and Middlebrooks, 2003). Cortical lesion studies strongly suggest that the cortex is essential for accurate sound localization, adaptation to changing cues, and learning (King et al., 2007). This implies that a topographic organization is not necessary for the representation of auditory space. Since SC is a multi-modal sensory motor area responsible for spatial orientation, it is possible that the topographical alignment of sensory and motor areas provides architectural and computational advantage for multi-sensory processing. This is supported by the studies suggesting that the auditory map in the SC is driven by the vision and visual map (Withington-Wray, Binns, and Keating, 1990). It is well established that the blind subjects who should have a disrupted topographical representation of space in SC, can localize sounds (Zwiers, Van Opstal, and Cruysberg, 2001b).

1.2.6 Computation of sound localization is experience-dependent

The shape of the head and the ears vary across individuals of the same species and so do the acoustic cues for sound localization. This implies that the computational processes using these cues should be customized for each individual (Middlebrooks, 1999;

Xu and Middlebrooks, 2000; Schnupp, Booth, and King, 2003; Maki and Furukawa, 2005). Individuals' head and pinna shapes change during development (Clifton et al., 1988) and could change even in adulthood, as do the sound localization cues generated by them. The auditory system is faced with the problem of maintaining sound localization capability and accuracy under such changing conditions. The auditory system's ability to adapt to changes in listening conditions has been demonstrated (Held, 1955; Hofman, Van Riswick, and Van Opstal, 1998). It has been shown that the acquisition computation of the auditory localization is also experience dependent: Infants show a U-shape performance function for sound evoked orientation responses during development (Muir, Clifton, and Clarkson, 1989; Kelly and Potash, 1986). Initial orientation responses after birth is proposed to be reflexive at first which later drops to chance levels. During this change the reflexive orientation responses slowly switch to the control of the mid and forebrain structures which mature later in the development (Muir and Hains, 2004).

Auditory deprivation in early ages can impair auditory processing in later ages. Behavioral experiments on animals raised with monaural ear plugs during development show deficits in sound evoked orientation responses under binaural listening conditions (Clements and Kelly, 1978; Kelly and Potash, 1986; King, Parsons, and Moore, 2000). For young animals that are not raised with ear plugs the orientation response deteriorates when monaural plugs are placed. The orientation performance is not significantly different from normal conditions when both ears are plugged (ear plugs do not completely block sound). These findings suggest that the computation of sound source location is shaped by the experience.

Adaptation to changed localization cues is also possible in adulthood (Hofman, Van Riswick, and Van Opstal, 1998). With ear molds placed monaurally or binaurally adult subjects'

monaural spectral cues can be changed. Interestingly, when molds are removed, subjects can use previous localization cues immediately without significant residual effect. When molds are placed again, adaptation is equally quick to previously experienced changes (Hofman, Van Riswick, and Van Opstal, 1998). Furthermore, subjects do not seem to be consciously aware of the changes they experience in the mold experiments, implying that no conscious effort was needed to adapt to new conditions.

The necessary and sufficient conditions for adaptation and underlying mechanisms are not completely understood. It has been suggested that sensorimotor (Held, 1955) and cross modal correlations (Moore and King, 2004), and behavioral tasks that demands accurate localization and attention could mediate the adaptation (Kacelnik et al., 2006).

Most investigations of the neural mechanisms for sound plasticity have focused on the SC. Monaural and binaural manipulations on the ear shape, disruption of the visual input, and rearing in omni-directional noise were some of the conditions that animals were exposed to during development. All these conditions have shown to be disrupting to the auditory map in SC. The sound-evoked orientation behavior as a result of these manipulations is also influenced. Behavioral studies for omni-directional noise rearing condition, unfortunately, has not been investigated yet. Relation between the localization performance and the quality of SC map brings up an interesting issue about the relation between subjects' localization ability and the intactness of a normal SC map. Can the normality or abnormality of the auditory map in SC be an indicator of the sound localization ability of an animal? The results summarized here and the results from blind subjects suggest that abnormal sound experience with insufficient directional quality and unpatterned visual input cause abnormal auditory map development in SC. Yet, in the absence of vision and insuf-

ficient directional acoustic cues, there is still a robust localization behavior. In the case of blind rearing, sound localization cues are still available. In the case of ear removal, they are insufficient for normal sound localization. Under both circumstances, however, the animal learns to localize sound using the directional cues available. Consequently, an abnormal SC map does not always lead to inability to localize sound.

The case of omni-directional rearing is particularly interesting and different from the other alterations of auditory rearing conditions (Withington-Wray et al., 1990). Under this condition the sound stimulus is available for the normal development of the auditory functions that are not dependent on spatial information. Since experience is an essential part of the development of spatial hearing one could expect that the auditory system has no opportunity to learn the localization cues because they are not available for acquisition.

Young and adult animals differ in the amount of plasticity they exhibit and the mechanisms that mediate them. Adult plasticity is most limited in range, whereas the plasticity during development is more flexible and sensitive to the changes in experience. Therefore, it is possible that different types of plasticity could be involved in both stages of life (Moore and King, 2004). Preliminary results suggest that adaptation is mediated by the corticocollicular influence, since severing these projections impairs adaptation (King et al., 2007).

1.2.7 Multi-modal processing of sound location

Although computation of sound location requires processing of acoustic inputs, it is also influenced by other sensory modalities. Vision is thought to be important for accurate sound

localization, for it provides high resolution spatial information that can act as a teaching signal to adjust acoustic-cue computation. This effect can be seen by comparing head orientation responses of normal and blind listeners (Zwiers, Van Opstal, and Cruysberg, 2001b; Lewald, 2002). Although, vision is not necessary for auditory spatial perception it certainly aids in making accurate orientations to auditory targets (Moore and King, 2004). A general principle that could explain this type of collaborations between multiple sensory modalities is to have a coherent spatial representation of the world where the visual objects and the sounds they generate overlap in the perceived locations from both modalities.

It is not surprising then that sound localization is also influenced by proprioceptive signals. For instance, vestibular signal (Lewald and Karnath, 2000), eye and head orientation (Alexeenko and Verderevskaya, 1976; Lewald, 1997; Lewald, Dörrscheidt, and Ehrenstein, 2000), neck and arm muscles afferent stimulations (Lackner and DiZio, 2005), and pinna orientation (Oertel and Young, 2004) all effect sound localization.

It is possible that most of these extra-auditory inputs could influence localization after the sound location is computed with reference to the head using ITD, ILD and spectral cues. Evidence suggests that these inputs contribute to the sound localization at very early stages of the auditory processing within the auditory system. For instance, experiments involving gaze orienting to a sound source by making a saccade to a different direction first, after or during the sound stimulation, revealed that head position signal is interacting with the sound localization computation at every frequency (Vliegen, Van Grootel, and Van Opstal, 2004; Vliegen, Grootel, and Opstal, 2005). Moreover, many animals with movable pinna may not be able to compute sound source location without initially considering the pinnae's state since HRTF is dependent on the pinna position (Young, Rice, and Tong,

1996). Proprioceptive inputs from the pinna muscles have found to project to the DCN (Oertel and Young, 2004) and ICx (Jain and Shore, 2006).

More evidence of multi modal influence on auditory processing came from the study that investigated the modulatory effect of gaze orientation on the IC neurons (Groh et al., 2001). This study has shown that integration of the gaze position signals creates a distributed frame of reference between eye-centered and head-centered coordinate frames for sound locations.

Multi-modality of sound localization reveals a different aspect of sound location computation. As mentioned earlier, the working model of sound localization hypothesizes location cues obtained from the acoustic signals. However, spatial auditory perception is dependent strongly on listeners' spatial relation to sound sources in their environment. Relative positions of the ears, the head with reference to the body have to be accounted for to guide sound-evoked actions. Experiments that manipulate listeners' body, head, pinna and gaze orientations (Lewald, 1997; Lewald and Karnath, 2000; Lewald and Karnath, 2002), and the correlation between vision, somatosensation and auditory inputs (Lackner and DiZio, 2000; Lackner and DiZio, 2005) clearly show that extra-auditory information is an essential part of spatial hearing without which the perception of auditory space would not be complete.

A dramatic example that depicts the importance of multi-modal interaction comes from experiments where subjects are provided with acoustic cues that are insufficient in creating a well localized and externalized perception of the sound source. If the auditory signals are manipulated to correlate with the subjects' movements, an immediate sense of externalization and localization is accomplished. Spatial sensation gained this way was utilized by the

subjects to guide their movements (Loomis, Hebert, and Cicinelli, 1990), despite the fact that no detailed spectral information was available to these subjects. In studies involving blind infants who were equipped with a sonar aid, babies quickly learned to use the distal spatial information provided by the sonar aid and used it as if it is a new exterosensor to locomote and guide themselves in space (Bower, 1989). The sonar aid did not provide localization cues that were based on HRTF but the acquisition is thought to be based on sensorimotor coherence.

1.2.8 Summary

Most of what we know about sound localization comes from a common experimental condition involving one stationary sound source and a stationary listener. Although sound localization can occur under more complicated conditions involving multiple moving sound sources and moving listeners with mobile pinna, this level of complexity makes scientific investigation mostly impractical. Since it is established that sound localization is possible under the stationary condition, it presents itself as an appropriate first step to understand the computational mechanism underlying sound localization. We can then hypothesize that what we learn about this well controlled and simpler condition can be generalized and expanded to more complicated situations.

Assuming stationary listening conditions are the most basic form of sound localization, the most reasonable approach is to explore location cues embedded in the acoustic signals received at the ears, since there is no other possible source for spatial information. The ability to localize azimuth, but not the elevation, of narrowband signals' suggests that vertical

localization requires broadband signals. In addition, horizontal localization of tones could only be possible if the auditory system is capable of determining the differences between the signals received at both ears. A tone signal could only differ in two ways, in phase and in amplitude, corresponding to the two well-known horizontal cues ITD and ILD. Since both cues for sound source in the median plane are zero and equal to constant values on the cones of confusion, they cannot be used for vertical localization, hence the dichotomy of the computation of vertical and horizontal components of sound sources.

Vertical localization, therefore, should rely on monaural spectral cues, especially on the median plane. This brings up a new condition for sound location computation. Because, sounds with complex unknown spectrum can be localized the computation should rely on location cues that are independent of sound source spectrum. This condition can only be realized if the computation uses information from both ears.

Computational mechanisms involved in sound localization should allow adaptation to changing conditions and develop through experience. Adaptation requires mechanisms by which the system can monitor the condition of the computation. This is where the extra-auditory information becomes important. Vision, proprioception and perhaps efferent motor signals should be correlated with the acoustic cues and changes in this relation can act as a signal to control adaptation.

Unfortunately, the cue-based model has not yet been generalized to address mechanisms for plasticity and influence of extra-auditory information. Next, a brief review of what we know about bat sound localization and how it fits to the general framework I developed so far will be discussed.

1.3 Sound localization by bats

The echolocation system can be thought of functionally as containing an active and a passive component. The active component is the ensonification of the immediate environment by the bat with sonar vocalizations. The passive component is the reception of sounds by passive listening. Spatial information is obtained as a result of the combination of these two components, i.e. active sensing. Thus, information obtained by the bat's sonar system requires these two components to work together. This is obvious in the example of sensing the range of a distant object which is not available without the evaluation of the timing between the outgoing vocalization and the received echoes. Similarly, recognition of objects by sonar requires the monitoring of how the outgoing sound is transformed by its interaction with the object that generated the echo.

In contrast, the role of active sensing in bat sound localization is not entirely clear since sound localization can be an independent process so long as there is a sound source in the environment, e.g. passive listening. In this sense, use for the sonar vocalizations could be limited to generating these sound sources to be localized. On the other hand, it may well be that the bat sonar system goes one step forward and takes advantage of spatial information only available through mutual interaction between active and passive components. By controlling the timing and the spatial shape of the vocalizations and adjusting its head direction and pinna position, a bat has a very flexible system to monitor space. By properly orienting its ears and adjusting its outgoing pulse timing, intensity, and time-frequency profile, a bat could increase the resolution of its localization system for a particular portion of space while making it robust to interference caused by other sound sources. An efference copy

of the motor signals or sensory feedback signals are necessary to incorporate ear positions, intended sonar emission patterns, and sonar signal timing and time-frequency profile for the processing auditory spatial information for the interaction of active and passive components.

Bats' sonar vocalizations are directional, and this directionality is frequency dependent (Hartley and Suthers, 1987; Shimozawa et al., 1974). As a result, spectral content of the outgoing ultrasound pulse will be different at different directions. Furthermore, high frequency components of sounds are absorbed more by the atmosphere. Consequently, the farther away an object is, the weaker the acoustic energy at the high frequencies in echoes received from it. It has been demonstrated that bats counteract this weakening by increasing their outgoing sonar vocalization strength by 6 dB per halving of target range (Hartley, 1992). This results in constant sound intensity at the target. The simultaneous contractions of the middle ear muscles with the sonar emissions results in decrease in sound intensity which is thought to help the compensation of the increasing echo intensity for closer target ranges (Kick and Simmons, 1984). This cumulative spatial filtering effect can influence sound localization as well as other functions of the sonar system, e.g object recognition and texture identification, that could be crucial for bats' ability to recognize its prey. By manipulating the spatial properties of the echolocation system, a bat could compensate for the loss of information at high frequencies in the regions of interest.

Unfortunately, most of these possibilities have not been directly demonstrated. From the sound localization perspective it is important to know whether bats can resolve the location of the sound within the sonar beam or do they just detect the existence of the target. Answering this question requires understanding of the spatial properties of directional

properties of the sonar emissions as a function of frequency and directional effects of the pinna, i.e. HRTF. Both of which are among the goals of my investigations. If we assume that bats are good passive localizers, then the answer to this question is related to the availability of localization cues to determine the sound location accurately. Echoes arriving from directions with limited spectral content will not be localized as reliably.

1.3.1 Bat sound localization accuracy

For many, the idea of navigating and hunting using echolocation implies a very accurate perception of the auditory space through hearing. However, it does not necessarily follow that echolocating bats have to have high or better accuracy for sound localization. The characteristics and limits of the spatial representation through biosonar is still being worked out. In particular, since bats can make vocalization at different rates, depending on the information they demand from the environment (Moss and Schnitzler, 1995), the spatial information can presumably become more detailed and accurate with the integration of the information in time. It is not clear how accurate the spatial information they can obtain is with a single sonar vocalization. Although, answers that can directly address these questions are not available yet and require complex experimental designs, it is still possible to take creative approaches to understand if bats need accurate localization of sounds. Erwin, Wilson, and Moss (2001) built a detailed computational model of the echolocation system to investigate sensory-motor control of the flight during target capture. Using data obtained from target capture experiments to calibrate different components of the model, Erwin, Wilson, and Moss predicted that one of the most important components of flight capture

behavior is sound localization and concluded that this model needs a highly accurate sound localization to be able to mimic bats flight capture behavior.

Accuracy of sound localization has been measured in big brown bats, *Eptesicus fuscus*, the subject of my investigations, and other species in horizontal and vertical planes using two alternative force choice paradigms (Peff and Simmons, 1972; Simmons et al., 1983; Lawrence and Simmons, 1982; Wotton and Simmons, 2000). Peff and Simmons (1972) reported that bats can discriminate two small spherical targets on the horizontal plane when the angle separates them is larger than (6° - 8°) for *Eptesicus fuscus*, an oral vocalizer, and (4° - 6°) for *Phyllostomus hastatus*, a nasal vocalizer. Both bat species, however, use broadband frequency modulated (FM) ultrasound pulses for echolocation. In more recent experiments, bats were required to discriminate between two sets of vertical bars or horizontally arranged beads on either side of a platform, each set with a different angular separation. By changing the angular difference between the two sets, researchers determined bats ability to resolve small angular differences. The horizontal accuracy threshold was found to be 1.5° corresponding to ILD differences of 0.3 to 1.5 dB and ITD differences of $1\mu s$ (Simmons et al., 1983). Vertical discrimination threshold is reported as 3° in the front and 4.1° and 8.3° in -40° and 40° vertical angles respectively (Wotton et al., 1996). Wotton et al. (1996) also reported that deflection of the tragus increased the threshold to $18.3^{\circ} - 20.1^{\circ}$ in -20° elevation in the lower hemisphere where spectral notch cues are thought to be operating as localization cues (Lawrence and Simmons, 1982; Wotton, Haresign, and Simmons, 1995).

In passive localization accuracy experiments on the same species for 100 ms long broadband noise and for bats' own recorded vocalizations as stimuli, minimum audible

angle (MAA) of 14° and 15° are reported, respectively for *Eptesicus fuscus* (Koay et al., 1998). In the same study, bats could localize high frequency tones but not the low frequencies (amplitude modulated or otherwise). Since ILD cues operate in high and ITD in the low frequencies, they concluded that *Eptesicus fuscus* does not use ITD cues for passive sound localization. The large MAA measured by the study can be a result of the experimental paradigm and the type of the stimuli used.

In an experiment where bats were capturing meal worms catapulted in the air, Webster and Brazier (1965) estimated that *Myotis lucifugus*, another FM-bat, orient their heads to the target with an accuracy of 5° . It is also reported that after the initial detection, bats bring the target in alignment with their head within 0.1 ms. Note that determining the initial detection time of a target is very difficult experimentally, for bats may not necessarily produce a reliable change in their echolocation behavior associated with their detection of the target immediately; therefore, estimation of the detection time inherently varies. More reliable estimates, however, could be obtained when a bat is close to the target at the time of detection since its needs to react quickly to intercept (Webster and Brazier, 1965).

A recent study by Ghose and Moss (2003) revealed that bats lock their head at least in one plane with 3° in standard deviation to the target during a prey capture flight. This result is also comparable with an earlier finding of $\pm 1.5^\circ$ head alignment to a moving tracking when perched on a platform (Masters, Moffat, and Simmons, 1985).

There are no sound localization studies that investigated sound localization in two dimensional space. However, there is good evidence to suggest that bats can localize sounds in space as accurately as humans, close to 1° of accuracy. Fuzessery et al. (1993) has reported that pallid bats, use prey-generated sounds to locate their prey (passive listening),

and can land within 1° to a target 4.9 m away from the perch after only hearing it briefly before take off.

1.3.2 Sound location cues for echolocating bats

Although, echolocation is composed of active and passive components, efforts to understand how bats might determine the location of sounds mostly assume passive-listening conditions. This allows generalization of the assumptions underlying mammalian sound localization to echolocating bats and provides a starting point. It has been largely accepted that bats most likely use ILD and monaural spectral cues for sound localization. ITD is not thought to be utilized by bats as a localization cue for its range is too small ($\pm 45\mu s$) and even smaller within the main lobe of the sonar beam (beamwidth is found to be decreasing from 60° to 20° from 20 kHz to 80 kHz for *Eptesicus fuscus* Hartley and Suthers, 1989). Furthermore, the MSO which is believed to be involved in ITD encoding in the auditory system, is predominantly monaural, although degree and type of binaurality varies across different species (Grothe and Park, 2000). However, ITD could be indirectly involved by influencing response latencies of neurons contributing to spatial computation, which is also valid for other mammals (Yin, 2002).

Measurements of the directionality of the bat ears (Jen and Chen, 1988; Obrist et al., 1993; Wotton, Haresign, and Simmons, 1995) and neural recordings of free field sounds presented from different locations (Grinnell and Grinnell, 1965) revealed that the ILD range is within 30 dB and could serve as a very robust cue for sound location. One interesting aspect of ILD in bats is that at high frequencies ILD changes with both azimuth and elevation.

This characteristic of ILD has not attracted much attention in studies on other mammalian species. Grinnell and Grinnell (1965) and more recently Fuzessery and Pollak (1984) propose that this complex behavior of ILD could serve as a cue for the computation of both horizontal and vertical components of a sound source location.

For vertical localization, spectral notches have been proposed as potential localization cues (Lawrence and Simmons, 1982; Wotton, Haresign, and Simmons, 1995). It is partially supported by the experiments that manipulated the tragus. The bat's tragus unlike many other mammals, is particularly large relative to the pinna size. It is therefore thought to play a significant role in shaping directional properties of the ear. Measurements taken from the ear with and without a tragus revealed that tragus contributes in shaping the spectral features relevant to vertical localization, in particular, elevation-dependent spectral notches (Lawrence and Simmons, 1982; Wotton, Haresign, and Simmons, 1995). This is a disputed issue and will be touched upon in chapter two. Bending of the tragus has been shown to disrupt elevation accuracy in the lower hemisphere (Wotton and Simmons, 2000). Later in a complementary study, using behavioral paradigms, Wotton et al. (1996) demonstrated that bats can detect notches embedded in the broadband noise stimuli. A more detailed analysis of the spectral notches and their previously unreported properties will be discussed in chapter-two. Although these studies offer spectral notch pattern as a cue for elevation, this cue could not be effective at the upper hemisphere for which there seem to be no clear spatial feature proposed (Wotton, Haresign, and Simmons, 1995).

Different roles are attributed to ILD cues in echolocating bats and in other mammals. This difference is concerned with addressing different types of difficulties in our understanding of sound localization. It is very likely that for most other mammals, including

human, at high frequencies, ILD also changes on the cone of confusion (Carlile, 1990). What creates the complex nature of ILD at these frequencies is the shape and size of the ear and the wavelength of the sound. At least for cats and ferrets whose ear shapes are similar to bats, the ratio of ear-head size and audible sound wavelength is similar to bats for the good part of the hearing range. As discussed previously, ILD is traditionally known as an azimuth cue, and for most frequencies it behaves so. The different look in the field of echolocation originates from efforts to explain how bats that use constant frequency (CF) signals could localize the vertical positions of the echoes. If ILD is an azimuth cue and unreliable for vertical component computation and a broadband feature like spectral notch could not be available to the bat, simplifying ILD as an azimuth cue will present a problem. While the greater horseshoe bat, *Rhinolophus ferrumequinum*, an echolocating bat that use long CF calls with short durations FM component at the end, employs alternating vertical pinna movements to localize in vertical axis, it is not thought to be a general solution adapted by the other CF bats. Fuzessery and Pollak (1984) motivated by the idea proposed earlier by Grinnell and Grinnell (1965) suggested that CF-bats could take advantage of the complex ILDs at three harmonically related frequencies (CF-bat calls consist of harmonic stacks) to triangulate the sound source position in space. Pollak and Park (1995) in their review explains in detail that this idea could be feasible based on his analysis of the binaural neural sensitivities in the IC.

Experiments on *Myotis lucifugus* with monaural earplugs, and *Plecotus auritus* and *Rhinolophus mehelyi*, with a unilaterally perforated tympanic membrane showed that monaural listening conditions decrease the bats' localization accuracy (Schnitzler and Henson, 1980). Obstacle avoidance studies on *Rhinolophus ferrumequinum* showed that monaural

ear plugs, that attenuate sound by 15 - 20 dB, resulted in reduced scores. When both ears are plugged, however, the scores returned to normal levels suggesting that interaural level comparison plays an important role for sound localization by bats.

1.3.3 Neural mechanisms for bat sound localization

Bats have been investigated intensely to uncover the physiology of hearing (Popper and Fay, 1995). Cells with spatial sensitivity are found in LSO, DNLL, IC, SC and auditory cortex. Similar to other animals LSO and DNLL are binaural nuclei receiving inhibitory and excitatory from opposite the ear and showing common EI responses. LSO sends inhibitory projections to ipsilateral and contralateral IC and so does the DNLL. Inhibition from LSO and DNLL are glycergic and GABAergic respectively. IC also receives ipsilateral inhibitory (GABAergic) projections from VNLL and INLL (Zhang et al., 1998), both of which are monaural. All these nuclei are topographically organized for frequency. Bats have unusually large VNLL and INLL which are sensitive to relatively larger frequency range and are very precise in their latency. The response latency in these areas are independent from frequency and sound intensity (Pollak and Park, 1995) and hypothesized to be involved in range detection. Projections to IC from these subnuclei creates rich and complex binaural responses in the central nucleus of the IC (ICc). Binaural neurons in ICc are also topographically organized for frequency, type of binaural response, response latency and ILD threshold for neurons with sigmoidal ILD response functions.

Most of the observations presented above are from mustached bats, that use CF-FM echolocation signals, who have a larger representation of 60 kHz frequency providing a

unique opportunity for detailed physiological investigations. One curious finding is that despite the existence of MSO, its potential function for bats is not well understood. In a mustached bat, MSO is predominantly monaural, in a Mexican free-tailed bat, on the other hand, it is mostly binaural. MSO in a big brown bat show both features (Grothe and Park, 2000). In all cases best ITDs are larger than the natural range of ITD.

Shimozawa, Sun, and Jen (1984) reported that a topographical organization of auditory space does not exist in the SC and speculated that the highly movable pinna could be the reason for it. However, other mammals with movable pinna do show the SC map suggesting the proposed reason may not be valid (Middlebrooks and Knudsen, 1984). Wong (1984) on the other hand, did find a topographical organization for azimuth extending along the rostral-caudal axis corresponding to ipsilateral-contralateral sound positions, in the SC of *myotis*. Valentine and Moss (1997) reported existence of cells that are not only sensitive to the azimuth and elevation but also range. Cells with three-dimensional and 2-dimensional receptive fields also showed the hint of topography. More recently, in addition to an auditory map a motor map, of space for the control of pinna and head orientation is also found (Valentine, Sinha, and Moss, 2002).

1.3.4 Summary

Localization of echo sources in space by bats depends on the directional characteristics of sonar emissions and HRTF. In order to achieve the required accuracy and detail for the spatial representation of space bats could control both aspects of the bat echolocation system independently. This provides opportunities to obtain and process spatial information that

are not available to passive listeners.

Echolocating bats most likely use ILD and monaural spectral cues to localize sound sources in space. Although, there has not been a systematic study on their ability to localize sound sources in different positions in space, target tracking and capture studies indicate that they can orient and land within 1° to 5° to the source. Minimum accuracy experiments revealed that bats can discriminate objects separated by on average 5° in azimuth and elevation by echolocation. Vertical discrimination of sound source positions is better in the lower hemisphere than in the upper hemisphere, and spectral notches are thought to be used as localization cues. Spectral notches are effective for elevations below the interaural plane and could be influenced by the removal of the tragus, whereas elevation cues for the upper hemisphere are not well understood.

ILD gains a special status for bat sound localization as a sound localization cue for both azimuth and elevation, contrary to the common view of ILD's role in mammalian sound localization. In particular, bats that rely on narrowband signals, where spectral cues would be of minimum use, ILD is proposed as vertical localization cue.

Most echolocating bats can move their pinna which significantly complicates the study of sound localization, because at different positions, a new set of HRTFs have to be considered. The topography observed in SC of other animals has been found to be less strong in bats and it's proposed that it might have been because of the dynamic complexity of the localization cues.

1.4 Outline of the dissertation

This dissertation is a collection of works that explore directional properties of active and passive components of the echolocation system of the big brown bat, *Eptesicus fuscus*. The second and third chapters focus on the directional properties of the bats' external ears and sonar emissions respectively. Chapter two takes a close look at the binaurally measured bat HRTFs with intact head and body and investigates spatial and frequency-dependent properties of the HRTF, in particular, sound localization cues ILD and spectral notches. In addition inferences to the separate contributions of the pinna and tragus to the directional properties of HRTF are drawn by analyzing HRTF obtained after removing tragus and pinna.

Chapter three reports on the results of a novel method developed to measure sound emission patterns from a freely behaving bat. This is the first time we have the ability to take these measurements from an unrestrained echolocating bat which can give us the opportunity to study sonar beam patterns under different conditions demanding different information from the echolocation system. The method being developed provides an opportunity to test if bats control their sonar beam patterns functionally. The analysis is mostly limited to the preliminary results obtained by the new method that lays the foundation for future experiments.

Chapter four enlarges the focus of the general cue-based model to address sound localization by echolocating bats. I develop a comprehensive approach to sound localization that incorporates adaptive and multi-modal characteristics of sound localization computation and allows a more appropriate framework to think about the sound localization problem

in a larger scale that is not limited to the static scenario. More importantly, by considering sensorimotor contingencies, a vital source to acquire spatial information processing, the proposed approach offers a way to bridge the explanatory gap in the acquisition of the sound localization cues.

Paper Title The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation

Authors Murat Aytakin, Elena Grassi, Manjit Sahota, and Cynthia F. Moss

Journal J. Acoust. Soc. Am. (2004) 116(6): 3594-3605

Chapter 2

The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation

Directional properties of the sound transformation at the ear of four intact echolocating bats, *Eptesicus fuscus*, were investigated via measurements of the head-related transfer function (HRTF). Contributions of external ear structures to directional features of the transfer functions were examined by remeasuring the HRTF in the absence of the pinna and tragus. The investigation mainly focused on the interactions between the spatial and the spectral features in the bat HRTF. The pinna provides gain and shapes these features over a large frequency band (20-90 kHz), and the tragus contributes gain and directionality at the high frequencies (60 to 90 kHz). Analysis of the spatial and spectral characteristics of the bat HRTF reveals that both interaural level differences (ILD) and monaural spectral

features are subject to changes in sound source azimuth and elevation. Consequently, localization cues for horizontal and vertical components of the sound source location interact. Availability of multiple cues about sound source azimuth and elevation should enhance information to support reliable sound localization. These findings stress the importance of the acoustic information received at the two ears for sound localization of a sonar target position in both azimuth and elevation.

2.1 Introduction

Echolocating bats, as auditory specialists, present a valuable model for the study of spatial hearing. These flying mammals produce ultrasonic vocalizations and listen to the echoes reflected from objects around them (Griffin, 1958). Their vital activities, such as spatial orientation, prey detection, localization, and capture can be accomplished entirely with the use of a biological sonar system. An important question is whether specialized sound localization capacity is required for successful orientation by echolocation in bats. If so, what specializations are evident that can help to deepen our understanding of spatial hearing?

Azimuthal discrimination experiments by Simmons et al. (1983) and horizontal tracking experiments by Masters, Moffat, and Simmons (1985) report about 1.5° accuracy in the echolocating big brown bat, *Eptesicus fuscus*. This species uses broadband frequency-modulated (FM) echolocation signals, which are well suited for spatial localization. Lawrence and Simmons (1982) measured the minimum audible angle (MAA) along the vertical axis and report discrimination by *E. fuscus* of about 3° . In a later study, Wotton and Simmons (2000) closely studied the bat's MAA in the midsagittal plane at 5 different eleva-

tions ($\pm 40^\circ$, $\pm 20^\circ$, and 0°). They reported that the bat's MAA falls between 2.9° and 4.1° at around 0° and $\pm 20^\circ$ elevation and increases to 6.7° at around -40° and 8.3° at around 40° . In the same study, when the bat's tragus was deflected, the MAA increased significantly, 18.3° - 20.1° , at around -20° elevation; however, the MAA in the upper hemisphere did not show a measurable change with the tragus manipulation. In these studies, the bats used biosonar to perform the localization tasks.

Studying bats in a passive listening localization task, Koay et al. (1998) used a conditioned avoidance procedure to measure left-right azimuthal sound source discrimination by *E. fuscus*. They reported that the bat's threshold for discrimination is 14° , comparable to a number of other mammalian species they tested. Further, they reported that this bat species could not discriminate narrow-band sounds that were positioned at 30° off the midline at low frequencies (5.6 to 8 kHz). They concluded that *E. fuscus* is not able to use interaural time difference (ITD) cues at sound frequencies where interaural phase differences would be available. Koay et al. (1998) estimated the maximum ITD experienced by the bats during the experiment to be approximately $55 \mu\text{s}$. Although the results of this study suggest that *E. fuscus* cannot use interaural phase differences for sound localization, the result does not eliminate the possibility of ITD localization using the envelope of wideband sounds.

To explore sound localization one step further, attention has been given to the directional properties of the external ears. Studies of the directionality of sound-pressure transformation at the pinna of several different bat species have been conducted (Jen and Chen, 1988; Obrist et al., 1993; Guppy and Coles, 1988). In these studies, sound pressure at the tympanic membrane was measured at different frequencies by moving a narrow-band sound source in space and recording signals at the base of the ear canals. Disturbance of the

iso-pressure contours and a decrease in the gain of the sound pressure for all frequencies tested was reported after removing the pinna and the tragus together (Obrist et al., 1993; Guppy and Coles, 1988).

Using a similar approach, Fuzessery (1996) measured acoustic pressure transformation by the intact ears of the pallid bat, *Antraous pallidus*. He reported that monaural spectral features of the HRTF of this species show parallel patterns to what has been reported in other mammals, such as humans and cats. He also showed that binaural difference cues changed with both azimuth and elevation. Grinnell and Grinnell (1965) reported a similar observation in the directional dependence of evoked potential recordings from the inferior colliculus (IC) of FM bats, *Plecotus townsendii* and *Myotis lucifugus*. They found that directional sensitivity of the neural responses changed with small changes in sound source angle in both the horizontal and vertical planes. Moreover, they reported that directional sensitivity was different at different frequencies. Based on these observations and assuming that directional responses of the left and right ICs are similar, they hypothesized that bats can use interaural level differences (ILD) obtained from several different frequencies to localize both sound source azimuth and elevation. Fuzessery and Pollak (1984) conducted cochlear microphonic recordings in mustached bats, *Pteronotus parnellii*, a species that produces signals with a combination of constant frequency and FM components. These researchers also proposed a similar role for ILD in horizontal and vertical sound localization. These important findings stand apart from the common view that interaural differences are used for azimuthal localization. Thus, the role of binaural cues for localization of sound source elevation warrants further study.

Sounds reaching the ear are transformed in a directionally dependent manner as a result

of their interaction with the body, the head, and the pinnae. The direction-dependent transformations can be modeled as a linear time-invariant systems transfer function, commonly known as the head-related transfer function (HRTF). Contrary to what the name suggests, the torso, in particular neck and shoulders, also introduce strong effects on the properties of the HRTF (Blauert, 1997).

To our knowledge, the most detailed published measurement of the HRTF in bats was reported by Wotton, Haresign, and Simmons (1995). The study focused on elevation cues: in particular, spectral notches and peaks in the HRTF of the big brown bat, *Eptesicus fuscus*. They measured the HRTF of full and half-head preparations and reported a linear relationship between center frequency (CF) of spectral notches (30 to 50 kHz) and sound source elevations (-30 to 0°). They also reported that the relation between notch CF and elevation was disrupted when the bat's tragus was deflected. Based on their findings, Wotton, Haresign, and Simmons (1995) suggested that position of the spectral notches along the frequency axis could contribute to vertical localization in bats, as is the case in other animals that have been studied, such as cats (Huang and May, 1996) and humans (Bloom, 1977).

Wotton and Simmons (2000) and Wotton et al. (1996) provided behavioral evidence for the role of the position of a spectral notch in sound source localization along the vertical plane. They trained the bats to discriminate a linear FM sound with a spectral notch at a particular frequency, and they demonstrated that spectral notches in the HRTF are perceived by the bat (Wotton et al., 1996). Bats showed a significant decrease in their performance in this behavioral task when the target sound source was presented at a particular elevation at which the filtering characteristics of the external ear produced a spectral notch at the same

frequency. In another experiment, they showed that deflection of the bat's tragus resulted in an increase in vertical MAA (Wotton and Simmons, 2000). Based on these and other MAA measurements (Simmons et al., 1983; Lawrence and Simmons, 1982; Wotton and Simmons, 2000) they suggested that there are distinct sound localization cues for azimuth and elevation, consistent with the classic dichotomy of separate cues for horizontal and vertical sound localization.

In our paper, we present the HRTF of an echolocating FM bat, *Eptesicus fuscus*. Our first goal was to confirm earlier reports by Wotton, Haresign, and Simmons (1995) and extend observations to intact preparations that preserve the head, neck, body, and wings. The measurements were taken from the entire frontal hemisphere at a high spatial and spectral resolution to capture a complete and detailed HRTF. We carried out analyses of HRTF data that enabled us to investigate the interactions between spatial and spectral features of the HRTF that were not previously reported. By comparing measured transfer functions in the absence of pinna and tragus, we examined the contribution of the external ear structures to the directional properties of the HRTF. In our analysis presented here, we emphasize the acoustic features that are believed to play a role in mammalian sound localization. This approach has led us also to consider a role for binaural cues for vertical localization by animals that hear in the ultrasonic range.

2.2 Methods

We constructed an apparatus and developed the methods to permit measurements of the head related transfer function (HRTF) of the intact bat, in the frontal hemisphere. A 1-m-

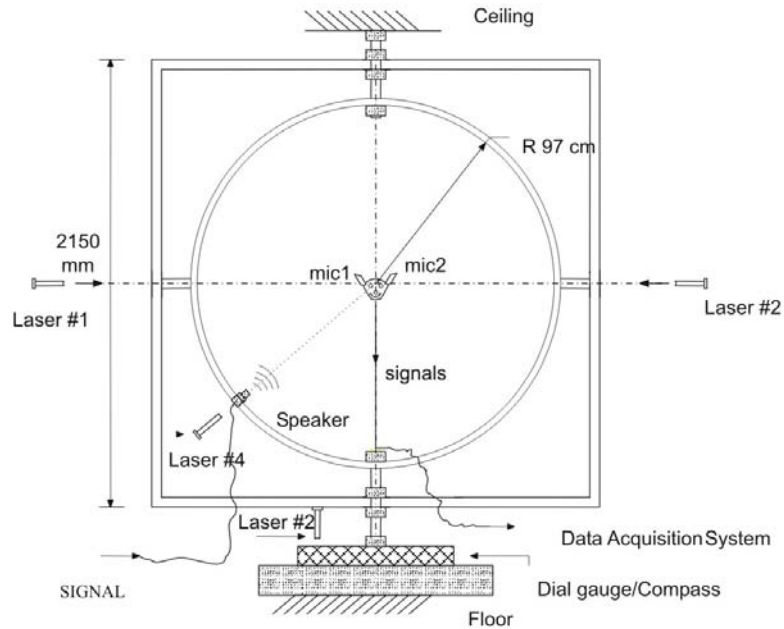


Figure 2.1: Schematic of the frame used to suspend the bat, with microphones implanted in the ear canals, and to deliver acoustic signals from changing azimuths and elevations.

radius hoop was used to carry out the acoustic measurements (Figure 2.1). The hoop was mounted upright and fixed to a rectangular frame, which was attached to the ceiling and floor of a large room (7×6 m) lined with acoustic foam. This device could rotate around its vertical axis and could be locked on a reference dial at any selected azimuth angle with 1° resolution. A custom 1.6-cm-diameter ultrasound loudspeaker (design by Lee Miller, Odense, Denmark, frequency response flat ± 3 dB between 20 and 100 kHz) was attached to a sliding fixture that could be locked at any position on the hoop. The hoop was marked in 5° steps, indicating elevation coordinates.

2.2.1 Animal preparation

Measurements were taken from four intact *Eptesicus fuscus* specimens (S1, S2, S3, S4). Each bat was perfused with 4% paraformaldehyde, followed by a saline flush. A microphone with a 1.5-mm-diameter metal probe (FG 3453-Knowles Electronics) was inserted into each of the bat's ear canals after rupturing the tympanic membrane. Lithium batteries (1.5 V) were used to power the microphones. The microphones were glued in place (Loctite 411 with Zipkicker accelerator) and sealed tightly around the ear canal to avoid any leak of sound pressure. Cross-talk between the microphones was measured before each experiment, and it was no greater than -60 dB. The bat was suspended horizontally with its wings spread, as in flight, and the pinnae were preserved in an upright position.

The bat's head was positioned at the center of the hoop pointing at the 0° azimuth and elevation. Photographs of the bat were taken to make necessary coordinate transformations to align the bat's eye-nostril axis to 0° azimuth and 0° elevation. The coordinate transformations that were applied to S1, S2, S3, and S4 were 40°, 30°, 0°, and 22° downward tilt, respectively. Two laser pointers were installed at positions -90° and 90° azimuth, each aimed at the tip of the corresponding tragus. These laser beams were permanently illuminated to monitor the position of the bat and to ensure the correct position of the frame in the calibration stage. Another laser pointer was used to adjust the directionality of the speaker with respect to the bat's head any time the elevation of the speaker was changed.

To investigate the contribution of the pinna and the tragus to the directional properties of the HRTF, a separate set of measurements for tragus-absent and pinna-absent ears was conducted. Tragus-absent measurements were taken from S3 and S4 by keeping the right

ear intact (control ear) and deflecting the tragus of the left ear. Pinna-absent measurements were taken from S2, S3, and S4 by cutting away the pinna and returning the tragus of the left ear to its natural position. Measurements of the HRTF from the intact right ear allowed comparisons across different sets of measurements to control for any time-related changes that may have occurred in the condition of the bat. During data collection from S3, an unintended change in the orientation of the bat after the first set of measurements prevented a detailed comparison of the HRTF features between the intact and manipulated ears.

2.2.2 Data acquisition

The data were recorded for sounds broadcast from a loudspeaker at a distance of 84 cm from the center of the bat's head. Microphone recordings from 685 different speaker positions on the frontal hemisphere were taken. Spatial positions were sampled from a double-polar coordinate system with 5° resolution in both dimensions. Speaker positions were specified by their azimuth and elevation values. Azimuth values changed from -90° to 90° from left to right of the midline, while elevation values changed from -90° to 90° from below to above the horizon.

Computer-generated signals, consisting of 2-ms-duration downward linear frequency-modulated (FM) sweeps, were broadcast at a rate of 33 Hz. The duration and interpulse interval of the signals were selected to prevent overlap of the broadcast signals with echoes from the hoop and surrounding walls. The signals recorded with the implanted microphones were amplified and bandpass filtered (10-100 kHz) using a commercial filter unit

(Stewart). The signals were sampled through a data acquisition system (IOtech Wavebook/512) connected to a Pentium III PC computer. The outputs from the microphones were monitored at all times on a two-channel oscilloscope (LeCroy).

2.2.3 Signal processing

Fifty signals per channel (left and right ears) were recorded, with a sampling rate of 500 kHz for each speaker position. To improve SNR, these signals were time aligned and averaged, making sure that the relative time delay between channels was unchanged. Then, HRTFs for the left and the right ears at each source direction were calculated as the ratio of the Fourier transform (2048-point FFT) of the averaged signals to the Fourier transform of the averaged signals obtained from free-field sound recordings. Free-field sound recordings were taken from the speaker position 0° azimuth and 0° elevation with both microphones placed at the center of the coordinate frame before they were implanted in the bat's ears.

The magnitude of the HRTF was smoothed in logarithmic frequency domain using a rectangular sliding window. The width of the sliding window was chosen to obtain a quality factor of 20. In this case, the quality factor is defined as the proportion of the window's center frequency to its width. This quality factor was chosen based on measured tuning curves of cells in the inferior colliculus of *E. fuscus* (Ferragamo, Haresign, and Simmons, 1998; Covey and Casseday, 1995) to achieve biologically realistic frequency resolution.

For some analyses, the HRTF was divided into direction-dependent and direction-independent linear time-invariant systems. The magnitude of the direction-independent component (here referred to as the common transfer function, CTF) was calculated as the

spatial average of the magnitude function across the entire set of HRTF measured. The phase of this component was computed, based on a minimum-phase assumption (Middlebrooks and Green, 1990). The direction-dependent component, also called direction-dependent transfer function (DTF), was calculated by dividing the HRTF by the CTF.

Spatial interpolation was applied to the measured HRTF at each frequency using spherical harmonic expansion (Wahba, 1981). This process allowed us to obtain HRTF for the directions that were not sampled during the measurements. After interpolation, iso-level contours were obtained using built-in MATLAB functions.

2.3 Results

2.3.1 Acoustic gain

Acoustic gain can be defined as the ratio of sound intensity at the ear canal to sound intensity in free field at a particular frequency. Thus, acoustic gain is the magnitude of the HRTF at that frequency. The direction of the maximum gain of HRTF (or DTF) across azimuth and elevation, at a particular frequency, is commonly known as the acoustic axis.

Figure 2.2a depicts the maximum gain (MG) provided by the HRTFs of the left (solid) and the right (dashed) ear of S1 as a function of frequency, across all the directions. Each curve shows distinct peaks and troughs, the acoustical modes of the external ears. A plateau at the lower frequencies spans a frequency region between 10 and 30 kHz. A peak can be seen at the higher end of the spectral range. The peak point of this mode is around 57 and 67 kHz for the right and the left ears, respectively. A trough in between the plateau and

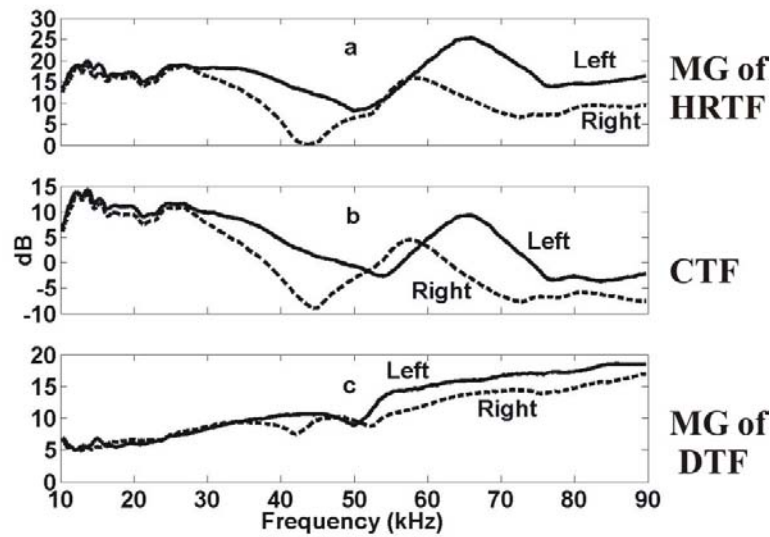


Figure 2.2: (a) Maximum gain (MG) of head related transfer functions (HRTF). (b) Common transfer functions (CTF). (c) MG of directional transfer functions (DTF). Functions for the left and the right ears of S1 given as solid and dashed lines, respectively.

the peak occurs at 43 kHz for the right ear and 50 kHz for the left. Across most of the frequency range, except the trough, the gain is positive and reaches a maximum of 25 dB for the left ear.

Asymmetry between the left and the right ears

Another interesting feature of the data is that the left ear's and the right ear's MG of HRTF and DTF are not identical, suggesting asymmetry between the ears. This asymmetry may be a result of structural and/or orientation differences between ears. Orientation differences between the two pinnae could be natural or might have been created during preparation of the animals for acoustic measurements. However, the asymmetry might also be an artifact caused by the placement of the microphones in the ear canals. Many HRTF studies have

investigated effects of microphone position in the ear canal on the directional changes in HRTF (Middlebrooks, Makous, and Green, 1989; Keller, Hartung, and Takahashi, 1998; Carlile, 1990). These studies have reported that transformation in the ear canal is independent of sound direction. Thus, microphone placement in the ear canal likely does not contribute to directional changes in the HRTF. To evaluate unintended effects of microphone placement in both ears that may potentially produce asymmetry, we examine the common transfer function (CTF) and the MG of the direction-dependent component of the HRTF, i.e. the direction-dependent transfer function (DTF). Effects of the ear canal and its interactions with the microphones on the HRTF are accounted for in CTF (Figure 2.2b). Any structural differences between the ears, before the ear canal opening, would still be present in DTF (for a technical discussion, see (Bloom, 1977; Hammershøi and Møller, 1996)). However, some asymmetry in the MG of DTF still persists between the two ears (see figure 2.2c). In particular, above 35 kHz the left ear shows a higher gain than the right ear.

The local troughs in both DTFs (Figure 2.2c) are the result of differences in center frequencies of the troughs in MG functions and CTFs, shown in figures 2.2(a) and (b), respectively. The center frequencies of the troughs for both ears shift to a lower frequency in the CTF. We hypothesize that the asymmetry between the left and the right ear DTF is a result of structural and/or orientation differences between the two ears.

Pinna and tragus manipulations

Figure 2.3 illustrates the left and right MG of HRTF, CTF and MG of DTF of S4. The left ear of S4 was manipulated after the intact ear measurement, while the right remained

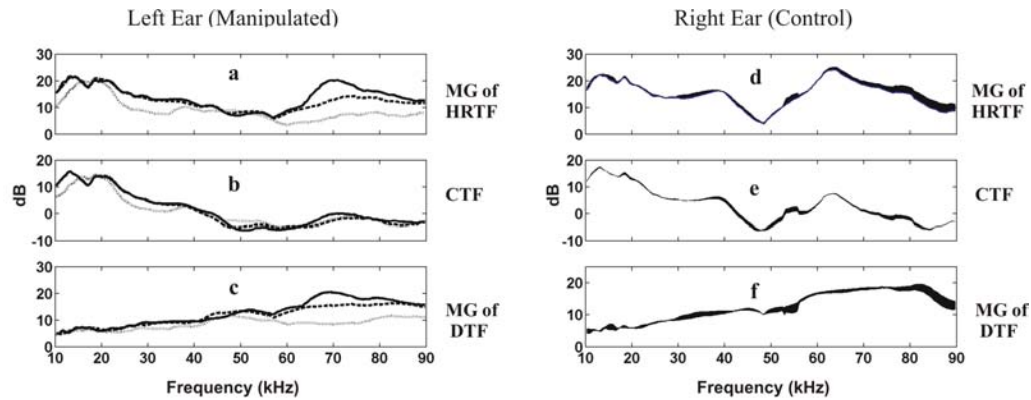


Figure 2.3: Transfer functions for S4: Intact (solid), tragus-removed (dashed), and pinna-removed (dotted). The right ear remained intact as a control. Characteristic functions for the right ear depicted as lines with varying thickness, representing the range of change of the functions across different measurements. The characteristic functions, from top to bottom, are maximum gain (MG) of HRTF, CTF, and MG of DTF.

untouched to serve as a control. Figures 2.3a and (d) show MG of HRTFs of the left and right ears, respectively.

As expected, the right ear characteristic functions overlap across different measurement sessions, ensuring that no significant decay of the preparation occurred (Figures 2.3d, e, and f). Shaded areas represent the range of change of characteristic functions along the frequency axis across three sets of measurements. Notice also that CTFs of the left ear show only small variations across the different manipulations. In contrast, the MG of the left ear HRTF for the intact and the tragus-absent cases differ above 57 kHz, where tragus removal causes a drop in the gain (Figures 2.3a, b, dashed line). Notice that this effect persists in the maximum gain of the DTF of the left ear (Figure 2.3c). This observation suggests that the presence of the tragus affects the acoustical mode, causing an improvement in gain and directionality and a small change in the frequency of the resonant mode.

Following removal of the pinna on the left ear, the maximum gain drops up to 6 dB between 20 and 35 kHz and up to 14 dB at frequencies above 57 kHz. The maximum gain of the DTF, after the pinna was removed, shows a very shallow increase with frequency (Figure 2.3c, dotted line).

MG of HRTFs obtained from S4 (Figures 2.3a and d) and S1 (Figure 2.2a) were slightly different from each other, not only across subjects but also across ears in the same subject. These differences may be a result of natural variations in ear shape and/or differences in pinna orientation and microphone placement in the subjects' ear canals. However, MG of DTFs show a similar pattern across subjects.

2.3.2 Monaural spectral properties

The HRTF includes direction-dependent spectral features that can be used for the computation of sound localization by the auditory system. Spectral notches are the most systematic of these features, and are believed to play an important role in vertical sound localization (Wotton, Haresign, and Simmons, 1995). Figure 2.3.2 shows the contour plots of the DTF magnitudes measured from the left ear of S2, as a function of frequency, on three sagittal planes, 23° ipsilateral (top), midsagittal 0° (middle), and 23° contralateral (bottom).

Position of spectral notches along the frequency axis changes with elevation

DTFs at different sagittal planes share a similar profile. Relatively higher intensities are observed for the planes that are more ipsilateral (negative azimuths in this case) as a result

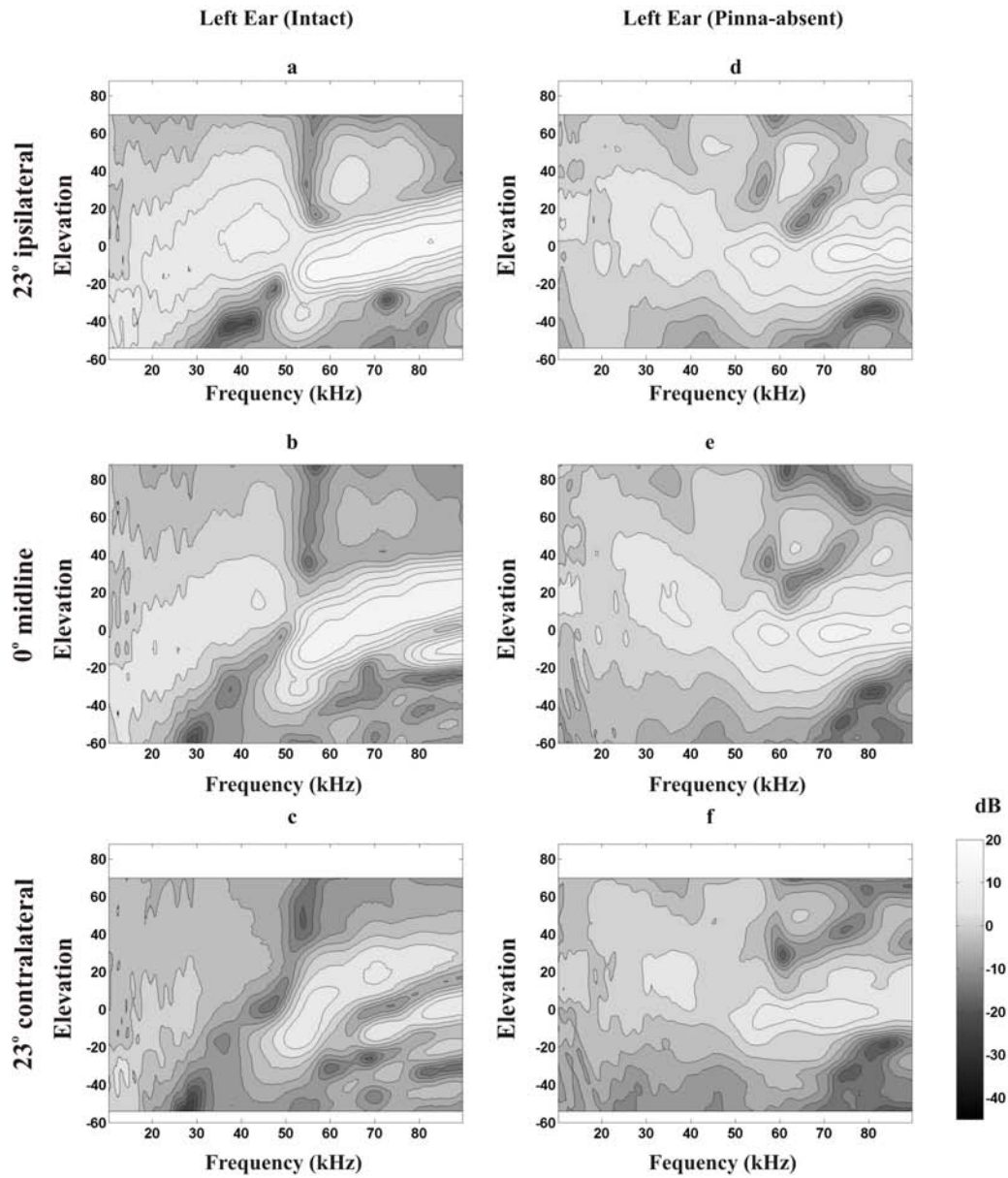


Figure 2.4: Left ear DTF contour plot for sagittal planes at 23° ipsilateral (top row), 0° (middle row), and 23° contralateral (bottom row). Intact (a), (b), (c) and pinna-absent (d), (e), (f) cases for S2. Contour lines are plotted with 3-dB separation.

of head shadow (Figure 2.3.2a). A peak covering almost the entire frontal hemisphere below 55 kHz is interrupted by a primary spectral notch that shifts from 30 to 55 kHz with increasing elevation (Figures 2.3.2a, b and c) and remains fixed at approximately 55 kHz at higher elevations. This notch changes roughly linearly for elevations -60° to 20° (Figures 2.3.2b and c). A secondary notch appears between 55 and 90 kHz over the elevation range of -60° to 0° (Figure 2.3.2b) and -60° to 20° (Figure 2.3.2c), respectively. A third notch trend is observed between -50° and -20° elevations along the lower border of a second peak for frequencies above 55 kHz (Figures 2.3.2a and b).

The pinna is essential for the elevation-dependent nature of the spectral notches

Figures 2.3.2d, (e), and (f), illustrate the changes in DTF as a result of pinna removal in the same subject. As can be seen, primary and secondary notch trends are no longer present for the pinna-absent ear. Starting at 70 kHz, a linear notch trend, less visible in the intact ear measurements (Figures 2.3.2a, b, c), shows similar elevation-dependent behavior as the primary and secondary notch trends in the lower hemisphere. Moreover, at high elevation and frequency, a different group of spectral notches surrounds a peak in this region. With the removal of the pinna, less directional dependence of spectral profile is observed.

Spectral notches are azimuth- and elevation dependent

The frequency where a notch reaches its lowest point is defined as the notch center frequency. In Figure 2.5 the spatial changes in the frequency of the primary notch, obtained from the HRTF of S1, for both ears are represented as a function of azimuth and elevation. The contour lines in the figure represent directions that share the same primary notch center

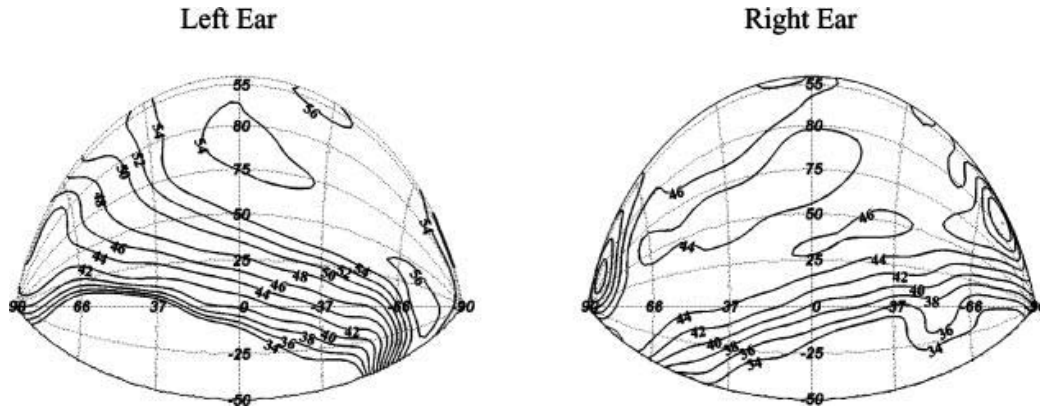


Figure 2.5: Primary spectral notch center frequency as a function of elevation and azimuth, for the intact left and right ears of S1. Iso-frequency contour lines were obtained using spherical harmonic interpolation.

frequency in the HRTF. To distinguish the primary notch, from other notches, a clustering algorithm in the three-dimensional domain of azimuth, elevation, and frequency was applied. This algorithm groups points that are close together in this 3D space. Interpolation between the obtained points was achieved using spherical harmonic decomposition. Iso-frequency contour lines of the center frequencies for both ears extend from ipsilateral low elevations to contralateral high elevations. The center frequencies monotonically increase with elevation in the medial-frontal space in both ears.

2.3.3 Monaural spatial features

Spatial patterns of HRTF consist of directional lobes separated by deep spatial nulls that are co-localized with the spectral notches

In Figure 2.6, intact and pinna-absent DTF magnitude contour plots at selected frequencies are compared. The figures are based on data collected from the left ear of S2. Contour maps

of the intact ear differ from the ones corresponding to the pinna-absent manipulation. In the intact case, a large lobe covers almost the entire ipsilateral and part of the contralateral frontal hemisphere for frequencies below 40 kHz. Above 40 kHz, two sidelobes, in addition to the main lobe, appear. Deep spatial nulls separate the main lobes and the sidelobes.

In the pinna-absent case, the main lobe for all frequencies covers a much larger spatial area in the frontal hemisphere. At lower frequencies (32 kHz), the main lobe occupies a large portion of the contralateral side. The acoustic axis falls between 0° and -34° of azimuth, around the 0° elevation plane, for almost the entire frequency region. In general, for the pinna-absent ear not only is the spatial selectivity low (larger lobes) but also the gain across frequency. This observation is in agreement with the effect of the pinna removal on MG of HRTF and DTF in Figure 2.3.

It is interesting to note the relation between the spatial nulls in the HRTF magnitude at a particular frequency, and the presence of spectral notches at that frequency. Figure 2.6 (left column) illustrates this relationship via overlaid spatial contour maps and spatial position of the spectral notches, white dots, at given frequencies. Note that spectral notches are co-localized with spatial nulls. This is a result of a rapid change in spatial location of the deep nulls in a small frequency range.

Direction of the acoustic axis changes with frequency

Another important observation is that, in the intact case, the acoustic axis (maximum point of the lobe) moves to higher elevations and lower azimuths as frequency increases (Figure

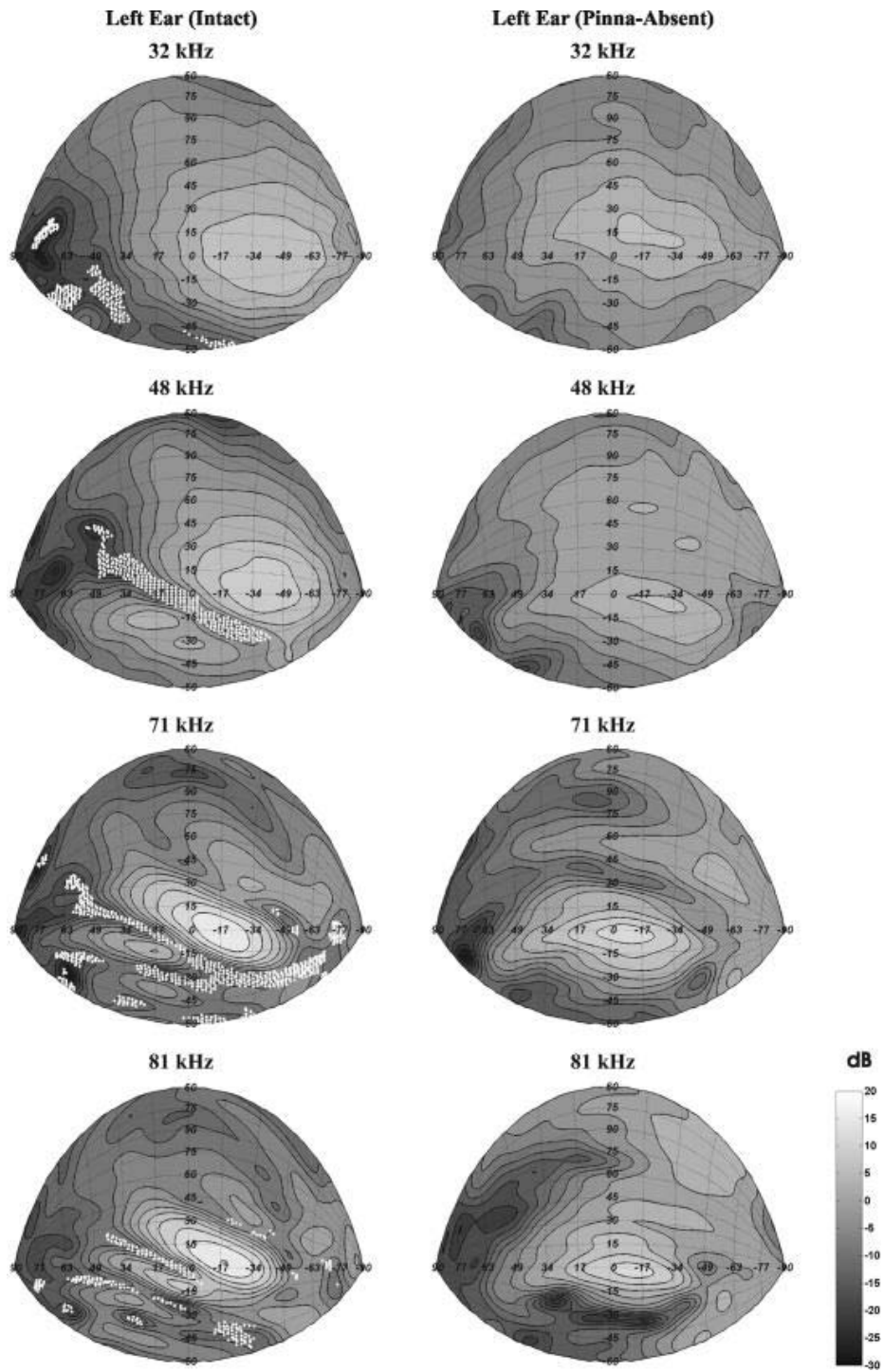


Figure 2.6: Spatial contour map of HRTF magnitude at selected constant frequencies (32, 48, 71, 81 kHz) for intact and pinna-absent cases of S2. Contour lines are plotted with 3-dB separation. Overlaid on the left ear plots are white dots representing the existence of spectral notches in the corresponding HRTF, at that particular frequency.

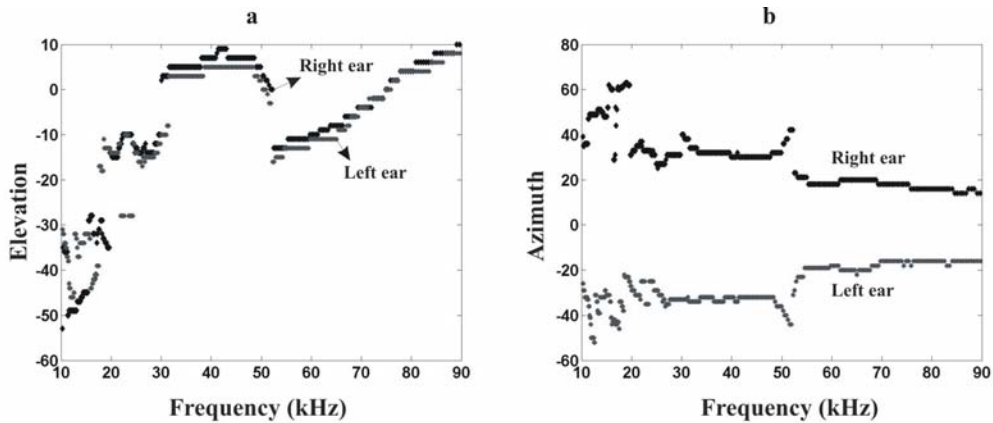


Figure 2.7: Frequency-dependent changes in acoustical axis elevation (a), and azimuth (b) for the left and the right ear of S2.

2.7). Figures 2.7a and (b) show the elevation and the azimuth of the acoustic axis as a function of frequency obtained from the HRTF of S2. Acoustic axis elevation shows a linear increase from -50° at 10 kHz to 10° at 42 kHz. A second linear trend between 50 and 90 kHz also shows an increase in acoustic axis elevation from -15° to 10° . At the frequency region (40 to 50 kHz) where the first linear trend is interrupted a second spatial lobe becomes prominent, as can be seen in Figure 2.6 (48 kHz). Note that behavior of the acoustic axis is very similar in both ears.

The azimuth of the acoustic axis shifts towards 0° azimuth (from $\pm 40^\circ$ to $\pm 0^\circ$) as frequency increases from 20 to 90 kHz. From Figure 2.7 one can conclude that external ears provide more sensitivity to the sounds arriving from directly in front of the bat at frequencies above 50 kHz.

The pinna and the tragus contribute to the directionality of the HRTF

One can observe in Figure 2.6 that the main lobes and the sidelobes become more directional with sound frequency. We quantified the frequency-dependent increase in directionality using a directivity index (Beranek, 1988). The directivity index is defined as the log ratio of the acoustic power delivered by a directional receiver on the acoustic axis to that of an omnidirectional with the same sensitivity in a diffused sound field. The directionality of the spatial maps of the left ears of S1, S2, and S4, as a function of frequency, are given in Figure 2.8 for manipulated (left) and control (right) ears. For all subjects spatial directionality of intact ears increases monotonically with frequency (Figure 2.8, solid lines), indicating that acoustic energy is collected from a restricted directional range as the frequency increases. Thus, this finding is consistent with the observation that the HRTF becomes more spatially selective with sound frequency. Using linear regression, an increase in the directionality was computed for the intact left and right ears, respectively: 0.148 and 0.121 dB/kHz for S1, 0.13 and 0.135 dB/kHz for S2, and 0.150 and 0.139 dB/kHz for S4. Note that the slopes were slightly different between the left and the right ears of S1 and S4, consistent with the asymmetry between the left and right MGF described earlier (Figure 2.2).

The effects of the manipulation of the left ear of S2 (pinna-absent) and S4 (pinna-absent and tragus-absent) on directionality of the HRTF as a function of frequency can also be seen in Figure 2.8 (dashed lines for tragus-absent and dotted lines for pinna-absent cases). For both S2 and S4 (Figures 2.8b and c) directionality still increases monotonically, but

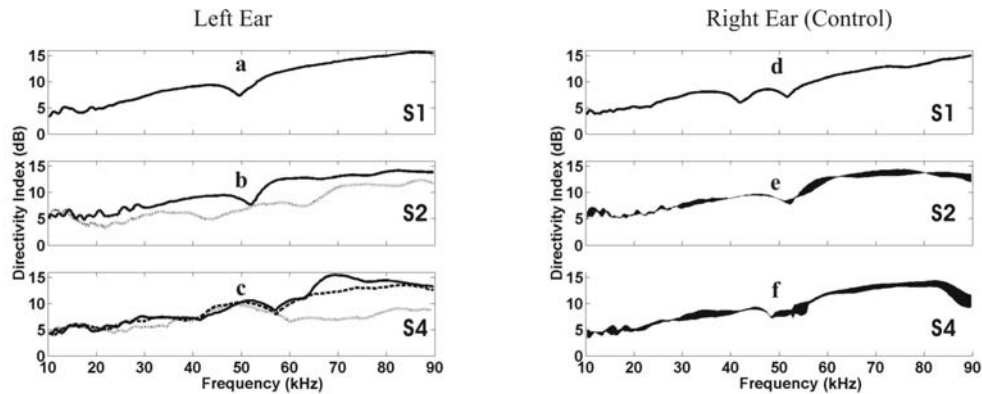


Figure 2.8: Directivity index of S1, S2, and S4 for left (a), (b), (c) and right (d), (e), (f) ear. Pinna and tragus manipulations were only applied to the left ears of S2 and S4. Data are shown from the right ear as a control. Intact (solid), pinna-absent (dotted), and tragus-absent (dashed) cases are illustrated. Directivity index for the right ear depicted as lines with varying thickness representing the range of change of the functions across (different) measurements.

removal of the pinna decreases directionality. The decrease in directionality is small below 30 kHz and more apparent above 57 kHz for S4; however, for S2 it occurs over a much larger frequency range. Removal of the tragus also causes a decrease in the directionality over 57 kHz for S4.

2.3.4 Interaural level differences

Spatial characteristics of the ILD becomes complex with sound frequency

Spatial contour maps of interaural level differences (ILD) at different frequencies are displayed in Figure 2.9 for S2. For frequencies below 40 kHz, ILD changes monotonically from the right to the left, as a function of azimuth (Figure 9, 38 kHz). In contrast, at higher frequencies this monotonic change no longer holds for the frontal hemisphere. In fact, at high frequencies negative ILDs (Carlile, 1990) were observed: negative ILD refers to measurements for which sound pressure at the contralateral ear is larger than the ipsilateral

ear. The contour plot for 48 kHz (Figure 2.9) shows that the monotonic ILD trend along the azimuth was disrupted with a negative ILD pattern around the center (0° azimuth and elevation.). Above this frequency, negative and positive ILD patterns alternate along the vertical axis around the midsagittal plane. The occurrence of this alternating ILD pattern increases with sound frequency. Note that large ILD values also appear at directions around the mid-sagittal axis.

ILD changes on the cones of confusion

Figure 2.10 shows the distribution of the angle of the spatial gradient vector of ILD with respect to the vector perpendicular to the cone of confusion surface, for each HRTF measurement location, as a function of frequency, for S4. A gradient vector's norm and orientation indicate the amount and the relative direction of composite local changes of the ILD at a given sound source direction and frequency. Having all the local gradient vectors orienting perpendicular to the surface of the cone of confusion would indicate that ILD change is independent from sound source elevation. Distributions are obtained as normalized histograms of the angles at each frequency. Normalization was undertaken by weighting each angle by a factor. The weight for each position was obtained as the ratio of the local gradient vector norm to the spatial integral of the norm of all the gradient vectors. This way the local ILD change at a particular position contributed to the distribution in proportion to the overall change in ILD on the frontal hemisphere. The histogram should show a peak around 90° if ILD changes on the cone of confusion, and it should peak around 0° if it does

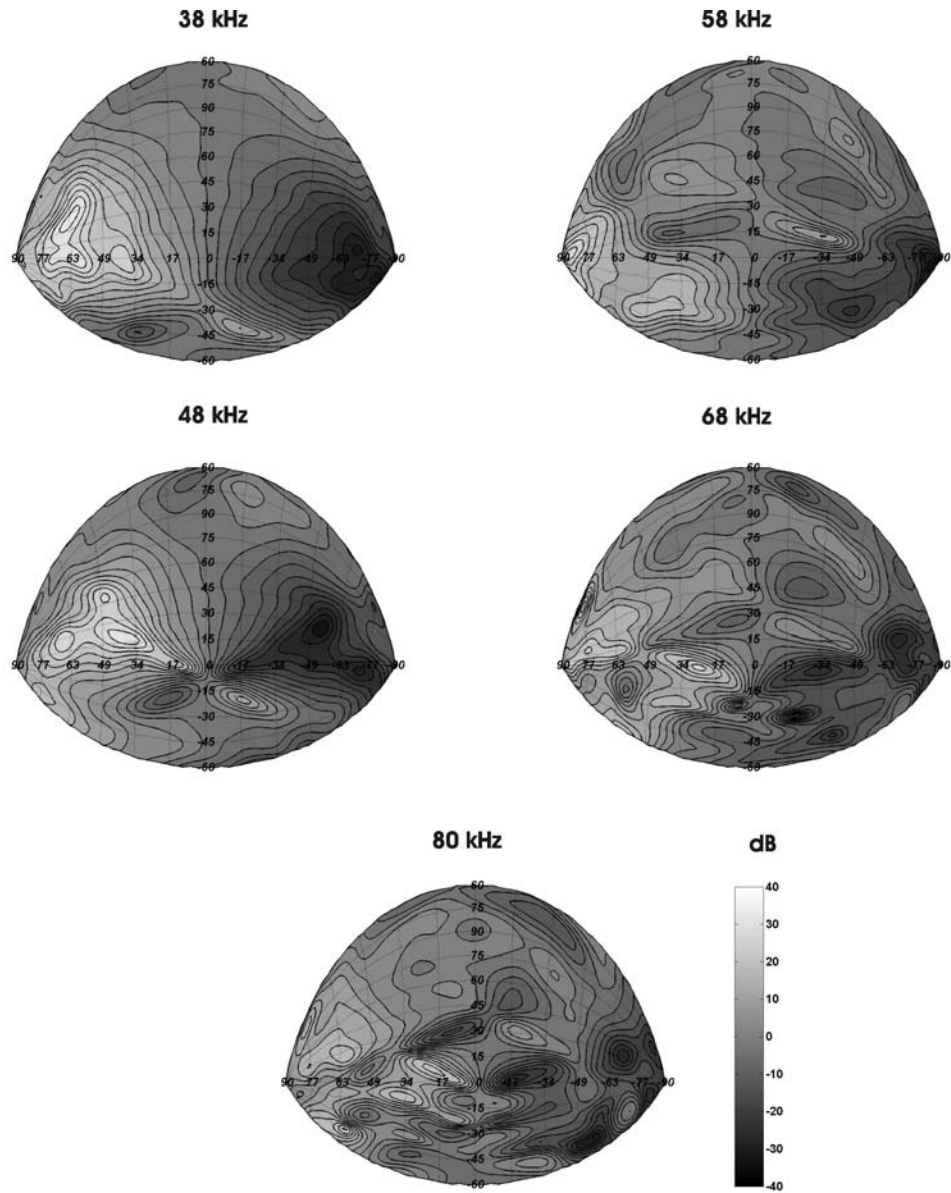


Figure 2.9: Interaural level difference (ILD) spatial contour maps for 38, 48, 58, 68, and 80 kHz of S2. Contour lines are plotted with 3-dB separation.

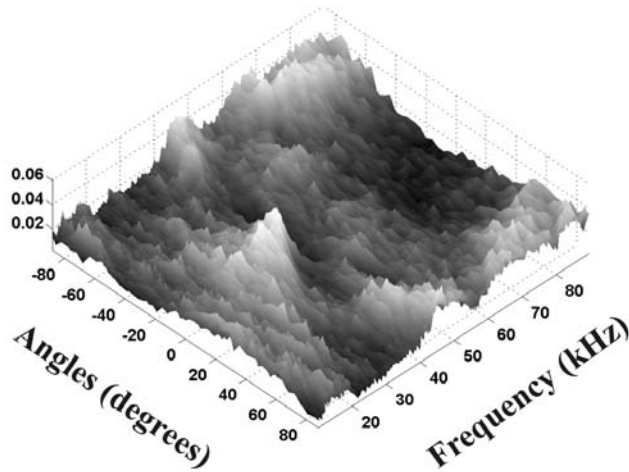


Figure 2.10: Angle of spatial gradient vectors of ILD to the vector perpendicular to cone of confusion plane at each frequency for S4.

not. The distributions show a peak around 0° , below 40 kHz, extending to 20° on each side, suggesting a constant ILD along the cone of confusion. Above 40 kHz, histograms show peaks around 90° , suggesting that local change of ILD is dominant on the cones of confusion.

2.3.5 Interaural time differences

Figure 2.11 illustrates the interaural time differences (ITD) obtained from S1. ITD was calculated based on the spectrogram correlation of the impulse responses (Middlebrooks and Green, 1990). For this process a 100-channel, second-order Butterworth bandpass filter bank, consisting of filters with 1/3-octave bandwidth and envelope detectors, was used. The cross correlation obtained from outputs of the right-and the left ear frequency channel pairs was summed across frequency. The time lag of the maximum of this summed

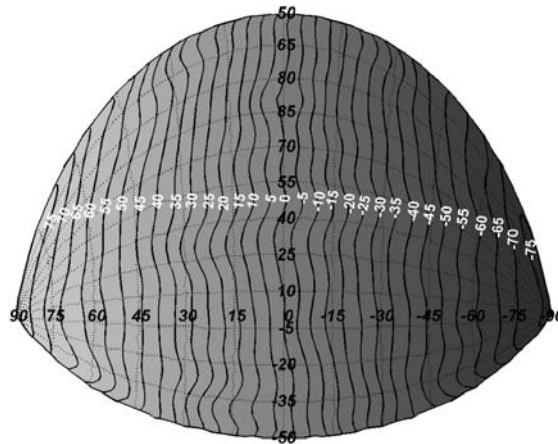


Figure 2.11: Interaural time difference (ITD) spatial contour map for S1.

cross correlation was taken as the ITD for that sound source location. ITD changed approximately between -75 and $75 \mu\text{s}$ as the sound source moved from right to left for all the subjects.

2.4 Discussion

Accurate sound localization is essential for bats when they negotiate complex acoustic environments. Extraordinary sensitivity to echo delay, used by the bat for target range discrimination, has been reported (for a review, see (Moss and Schnitzler, 1995)). However, less is known about the acoustic cues used by bats to localize sound source direction. Several behavioral studies reported accurate azimuthal and elevational sound localization performance in echolocating bats (Simmons et al., 1983; Masters, Moffat, and Simmons, 1985; Wotton and Simmons, 2000). The bats' accuracy in sound localization depends on robust

localization cues created by the directional properties of the ears and/or specializations of acoustic information processing that take place in the auditory system. Studying the HRTF is an important step in our understanding of sound localization, as localization cues available to all mammals are captured entirely by the HRTF. Investigating bat sound localization from the HRTF is limited to a situation where the pinnae are immobile during passive listening. *E. fuscus* has mobile pinnae, and thus, HRTF-based analysis cannot capture the role of the pinna mobility in sound localization. Yet, our results can still apply to the localization of brief sounds when the pinnae may be immobile during the sound reception. Bats have been shown to localize brief sounds with high accuracy (Fuzessery et al., 1993). Thus, our results are likely to be relevant to the understanding of sound localization in echolocating bats.

Our analysis of the echolocating bat HRTF suggests that the pinna plays a role in shaping its directional features. Comparisons between the intact-and the pinna-absent ears showed that the pinna improves the directionality and the gain of the HRTF. The pinna also affects the directional properties of the spectral notches.

Spectral notches systematically changed with sound direction in the bat HRTF. There were three prominent notch trends, which were between 30 and 50 kHz, 55 and 90 kHz for sound source positions on the lower hemisphere, and 55 and 90 kHz for sound source positions on the upper hemisphere. A closer look at the spatial distribution of these notches showed that iso-frequency lines of notch center frequencies extend diagonally from contralateral at low elevations to ipsilateral at high elevations, exhibiting both elevation and azimuth dependence. Iso-frequency lines of notch center frequencies from both ears result in a grid-like organization of these lines, as observed earlier in cats by Rice et al. (1992).

Another salient feature of the HRTF is the presence of spatial nulls. These nulls were found in between the main lobe and the sidelobes of the HRTF. A comparison between the spatial positions of the spectral notches and the spatial nulls indicates that the two are overlapping (Figure 2.6, left column). In addition to our observations, many earlier studies on the directional properties of the bat external ear also mention that the acoustical axis moves in the frontal hemisphere and increases in directionality with sound frequency (Jen and Chen, 1988; Obrist et al., 1993; Guppy and Coles, 1988). In relation to these characteristics of the HRTF, the spatial nulls also move with the acoustical axis. These changes in spatial nulls within a small frequency region are related to a deep and sharp notch in the HRTF. Ear manipulations in this study revealed that the pinna contributes to this phenomenon.

We observed that the ILD also shows both azimuth and elevation dependence. An azimuth-dominant change in ILD appears at low frequencies, between 10 and 40 kHz. For frequencies above 40 kHz, the ILD spatial distribution was more complex and sensitive to the elevation of the sound source.

Wotton, Haresign, and Simmons (1995) reported that the absence of the tragus causes disturbances in the primary notch trend of the bat HRTF. In a related behavioral study these authors reported tragus deflection produced a dramatic increase in the bat's vertical MAA in the frontal-lower hemisphere, where they observed the primary notch trend of the HRTF from the intact ear. They hypothesized that bats use the primary notch trend as a cue for localization of the sound source elevation.

We found that the tragus contributes to the gain and the directionality of the HRTF at frequencies between 70 and 90 kHz. Aside from these effects, we did not observe distur-

bances in the vertical dependence of spectral notches in the frequency range of 30-50 kHz following tragus removal, as reported in the Wotton, Haresign, and Simmons (1995) study.

ITD based on the measured HRTF changed between -75 and 75 μ s. Although this range of change is relatively small, we believe that ITD could be a reliable cue for sound localization during echolocation. The range difference discrimination threshold of *E. fuscus* is 30-60 μ s; (Moss and Schnitzler, 1995) thus, it may be possible for a bat to detect echo arrival differences from the echo delay difference between the two ears using FM sweeps.

2.4.1 A model for sound localization

Earlier theoretical and experimental studies of sound localization have asserted that mammals localize sound using binaural difference cues (ILD and ITD) for azimuth and monaural spectral cues, such as spectral notches, for elevation (Fay and Popper, 2000) in the frontal (lower) hemisphere. However, neither the monaural cues that are investigated here nor the ILD in our data behaves in a way that is consistent with this localization scheme. By contrast, both the ILD and the spectral notches were dependent on sound source azimuth and elevation.

Our observations are in agreement with some earlier reports. Grinnell and Grinnell (1965) and Fuzessery and Pollak (1984) suggested a role for ILD in vertical localization in other bat species. In addition to bats, earlier studies report elevation-dependent ILD changes in humans (Middlebrooks, Makous, and Green, 1989) and other mammals such as ferret (Carlile, 1990). However, these reports did not lead to further studies of the role of ILD on

vertical sound localization in bats and other animals.

Wotton, Haresign, and Simmons (1995) briefly mentioned a weak relation between the primary notch trend and sound source azimuth. They mentioned a possible role for the spectral notch in binaural sound localization, as was suggested in cats; (Rice et al., 1992) however, they focused on the role of the spectral notch in vertical sound localization. They reported that elevation-dependent notch trends are limited to the lower hemisphere, and suggested that this observation can explain the vertical MAA difference between high and low elevations along the vertical plane. Moreover, disturbance of the primary notch trend and the increase in the MAA at lower elevations in bats with deflected tragus were considered to provide further support for the role of the spectral notch in vertical sound localization. Combining their behavioral and HRTF data, Wotton, Haresign, and Simmons (1995) suggested separate sound localization cues for vertical and horizontal components of the sound source position.

We observed that the pinna contributes more to the direction-dependent spectral patterns than does the tragus. In addition, we showed that spectral notches are not limited to the lower hemisphere. Spectral notches mainly surround the acoustic energy peaks in the HRTF spatial patterns, and they can appear for sound sources at high elevations. We found that there is a coupling between the spatial nulls and the spectral notches and most important, the spatial nulls give rise to complex ILD patterns. Consequently, elevation dependence of the ILD is related to the directional properties of the spectral notches. Thus, ILD and spectral notches should not be viewed as entirely separate cues that could be used exclusively for horizontal and vertical sound localization, respectively.

We propose that both the ILD and the spectral notches contribute to localization of

sound source azimuth and elevation, probably in cooperation with other spectral features of the HRTF. The fact that both cues can provide horizontal and vertical information about sound source position improves the robustness of computation for sound localization. This new scheme can replace the simple dichotomy of acoustic cues for sound localization in azimuth and elevation.

Several recent psychoacoustical findings showed that human subjects' vertical localization performance was not in accordance with the single-notch theory (Macpherson, 1998). The single-notch theory predicts that subjects should perceive a sound source at a vertical position that depends on the frequency of the spectral notch embedded in the sound source. This prediction is based on the relationship between the sound source elevation and the notch frequency in the HRTF. Langendijk and Bronkhorst (2002) using virtual acoustic techniques, observed that human subjects' localization acuity did not change when the spectral notch cue was removed from HRTFs.

Our data suggest a model of sound localization in which the absence of the primary notch cue should not cause loss of acuity in vertical localization outside the midsagittal plane, since the ILD cue is still available. The fact that in our model the spectral features in the HRTF from both ears contribute to the computation of sound source localization is also in agreement with the study by Hofman and Van Opstal (2003) In this study it was shown that spectral cues from left and right ears are weighted to construct an elevation percept around the midsagittal plane.

We suggest that computation of sound location in the bat auditory system should be based on information received from both ears. Localization cues that were investigated here are not distinctly separate for sound source azimuth or elevation, but rather can contribute

together to the estimation of sound source location.

2.4.2 Summary and conclusions

A spatial-spectral analysis of the HRTF of *E. fuscus* is reported. HRTFs obtained from intact, pinna-absent, and tragus-absent ears are compared to understand contributions of external ear structures to the directional transformation of the sounds at the ear canals.

We made key observations that may have important consequences to our understanding of bat sound localization. We found that both ILD and spectral notches, believed to be important localization cues, are dependent on sound source azimuth and elevation

These findings lead us to a localization model for *E. fuscus*, for which both of these cues contribute to localization of sound source direction. This model provides robustness in sound localization by employing multiple cues for sound source azimuth and elevation.

Chapter 3

Sonar beam patterns of the freely vocalizing echolocating bat

3.1 Introduction

Spatial information obtained via the bat's sonar system is influenced by the directional properties of the ears and the vocalizations. Bat sonar emissions are directional (Simmons, 1969) and the size of the ensonified region around the bat can have important implications for the inner workings of the sonar system. For instance, a narrow sonar emission beam focuses acoustic energy. It provides for a robust reception of the echoes on the beam's main axis and reduces interference from echo sources that are outside of it. On the other hand, a narrow sonar emission beam limits spatial coverage and, consequently, detection of prey. Conversely, a broad sonar emission beam allows detection of obstacles in a larger area of space, while being susceptible to problems caused by low signal-to-noise ratio and interference. A broad beam dissipates the emitted acoustic energy resulting in weaker

echoes and a shorter detection range. Thus, in order to detect targets at larger distances, bats emitting wider beams need to produce higher intensity sonar vocalizations. This trade off between beam width and signal quality can be managed by actively controlling the beam's timing, direction, spatial coverage, and frequency bandwidth.

A wideband sonar system, with enough directional complexity, can achieve the benefits of both narrow and broad beams. Different beamshapes across the frequency can allow construction of spatial filters by controlling the spectra of the sonar vocalizations. Similarly, with a flexible sound generator a bat could even steer its beam mechanically (Hartley, 1990). Moreover, since the ear position can be controlled by many different bat species, it is possible to change the spatial properties of the entire echolocation system by taking advantage of the dynamic properties of emission and reception components. This can also be achieved by filtering the acoustic signals at the ears. Monitoring sonar emission patterns from behaving bats is an important step in understanding how bats control their sonar emission beams.

Simmons measured sonar emission patterns of *Eptesicus fuscus*, an oral emitter, and *Chilonycteris rubiginosa*, a nasal emitter at 30 and 28 kHz. Bats were trained to vocalize towards a target during flight. and microphones placed at different horizontal angles (0°, 22.5°, 43°, and 90°) with respect to a target were used to measure the sonar beam. This study reported that the sonar beam of both bats decreases from maximum to half intensity within 22.5° azimuth. Sonar beam intensity of *Chilonycteris* weakened faster with azimuth, above 22.5°, than that of *Eptesicus*. These and similar studies on different bat species have shown that the sonar emission beams are oriented directly in front of the bat and they are highly directional.

Sonar beamshapes change with frequency. It is, therefore, important to understand the spatial properties of the sonar emission patterns as a function of frequency for bats that use wideband sonar vocalizations. Moreover, depending on the complexity of the emitter, e.g. shape of mouth or nasal structures, beamshapes can be spatially complex and may not be symmetric along the median plane (Schnitzler and Grinnell, 1977; Hartley and Suthers, 1987; Hartley and Suthers, 1989). To capture such details, a dense sampling of the bat's sonar beam is necessary. Many recent studies of bat sonar emission patterns employed methods that fixed the bat's head in space (Shimozawa et al., 1974; Hartley and Suthers, 1987; Hartley and Suthers, 1989). Methods that limit head movements are particularly difficult for some species like, *Eptesicus fuscus*, since they do not voluntarily vocalize when restrained. Thus, more invasive techniques, involving electrical stimulations of the certain brain areas to elicit vocalizations has been necessary (Shimozawa et al., 1974; Hartley and Suthers, 1989). By stabilizing the head and controlling sound production, investigators can sample the beam at different points in space, using a microphone, at any desired resolution.

In agreement with theory of acoustic radiation (Beranek, 1988), bat sound emissions are directional with higher directionality at higher frequencies (Hartley and Suthers, 1989). At all frequencies, the mainlobe is oriented to the front and its vertical orientation changes systematically with frequency (Shimozawa et al., 1974; Schnitzler and Grinnell, 1977; Hartley and Suthers, 1987; Hartley and Suthers, 1989). A sidelobe directed to lower elevations has also been observed in different species (both nasal and oral vocalizers) (Schnitzler and Grinnell, 1977; Hartley and Suthers, 1987; Hartley and Suthers, 1989). The vertically oriented sidelobe increases sensitivity of the echolocation system for lower elevations in

the frontal hemisphere. Combined with the directional characteristics of the ears the vertical sidelobe seems compensate for the relative insensitivity of the ears for lower elevations (Schnitzler and Grinnell, 1977). Hartley and Suthers (1989) pointed out that directional characteristics of the sonar beamshapes of *Eptesicus fuscus* show similarities to simple a piston radiator. Unlike a piston radiator sonar beams of these bats do not show sidelobes.

One important function of the echolocation system is to determine the relative position of objects in the environment. Bats can localize the direction of echoes reflected from objects using interaural level difference (ILD) and monaural spectral cues. For echolocating bats, ILD could be a useful cue for determining both the elevation and azimuth of echo source locations (Grinnell and Grinnell, 1965; Fuzessery and Pollak, 1984; Aytakin et al., 2004) especially for constant frequency vocalizers like mustached bats. Although, the ILD is not effected by the sonar emission, the sonar beam pattern can influence the availability of ILD cues by amplifying and attenuating sound intensity arriving from different directions (Fuzessery, Hartley, and Wenstrup, 1992). Wotton, Jenison, and Hartley (1997) reported that the spectral notch cues sharpen when head related transfer functions (HRTF) are combined with sonar emission patterns for *Eptesicus fuscus*. It has been hypothesized that sharper notch cues can improve vertical localization of echo sources (Wotton, Haresign, and Simmons, 1995).

Studies investigating the combined effect of the emission and reception spatial characteristics, showed that the effective beamshape of the echolocation system is highly directional (Shimozawa et al., 1974; Grinnell and Schnitzler, 1977; Fuzessery, Hartley, and Wenstrup, 1992), providing a homogeneous region around the main lobe (Fuzessery, Hartley, and Wenstrup, 1992).

Different bat species employ different methods to elicit sonar vocalizations. Those that produce sonar vocalizations nasally, have elaborate nose structures. It has been hypothesized that sonar emission patterns generated by these bats can have advantages over patterns emitted orally. Emission patterns in different species of nasal emitters has been studied (Grinnell and Schnitzler, 1977; Hartley and Suthers, 1987). Based on the similarities between sonar beams patterns of oral and nasal emitters, Grinnell and Schnitzler (1977) reported that there are no advantages for the nasally-emitted sonar beamshapes

The complex structure of nasal apparatus is important for shaping of the sonar beam. Some speculate that nasal emitters have the ability to steer their emission beam by manipulating this structure, which can be done voluntarily (Schnitzler and Grinnell, 1977; Hartley and Suthers, 1987). Schnitzler and Grinnell (1977) found that directionality of the sonar beam in the horseshoe bat increases with increasing sound intensity. Whether horseshoe bats can control the directionality of their emission beamshape remains unclear.

For oral emitters, variation of the beam was also reported in an earlier study by Shimozawa et al. (1974). This variation was within $\pm 12\%$ of the mean emission beamshape and was associated with the dynamics of mouth opening during vocalizations. The peak intensity did not always correspond to the maximum mouth opening but the aperture of mouth remained between 30° and 60° during vocalizations. In addition to a different sound generation mechanism, The Schnitzler and Grinnell study was different from Shimozawa et al. for their bats were generating the sonar signals on their own.

The goal of our study is to record complete sonar beam patterns from freely vocalizing *Eptesicus fuscus*. Unlike earlier studies on this species by Hartley and Suthers (1989) this approach eliminates the need to restrain the animal's head and stimulate its brainstem to

generate sonar vocalizations. The approach taken here relies on the assumption that the bat's sonar beam does not change. Stability of sonar emission patterns was also assumed in the earlier reports. With this assumption it is possible to combine spatial measurements taken from different vocalizations to construct the sonar beamshape. Results obtained in this study show that this assumption is not met by freely vocalizing bats. Supporting evidence and consequences of this outcome will be discussed and potential improvements to the new technique will be discussed.

3.2 Methods

3.2.1 The approach

The method developed here aims to reconstruct the sonar beam from a freely vocalizing echolocating bat sitting on a platform. Using an array of microphones, 22 points of the beam are sampled with each sonar vocalization. Spatial samples from multiple vocalizations are combined using head position estimations, obtained from a high speed camera, to reconstruct the beam.

3.2.2 Apparatus

The sound recording apparatus is a hemispherical metal structure with a radius of 50 cm (Figure 3.1). The axis which connects the hemisphere's polar point (0° azimuth and 0° elevation) with the center, extends parallel to the room floor. Eighteen metal rods ($1/8''$ radius) with 20° of separation meeting at the polar point comprise the hemisphere. On



Figure 3.1: Experimental apparatus.

each arm there are small mounts for the microphones. These mounts can slide along the arms and can be fixed at a desired angular position. A small platform for the bat is placed at the center and oriented toward the polar point. Large surfaces and electrical cables are covered with sound-absorbent foam (Sonex-1) to suppress echoes.

3.2.3 Data acquisition

To sample the sonar beam at multiple spatial locations simultaneously a microphone array with 22 electret microphones (Knowles FG-3236) is used. Eighteen of these microphones are placed at positions 30° off from center-pole axis. Another microphone is placed at the center position and two more 60° off center on the horizontal plane on the opposite sides

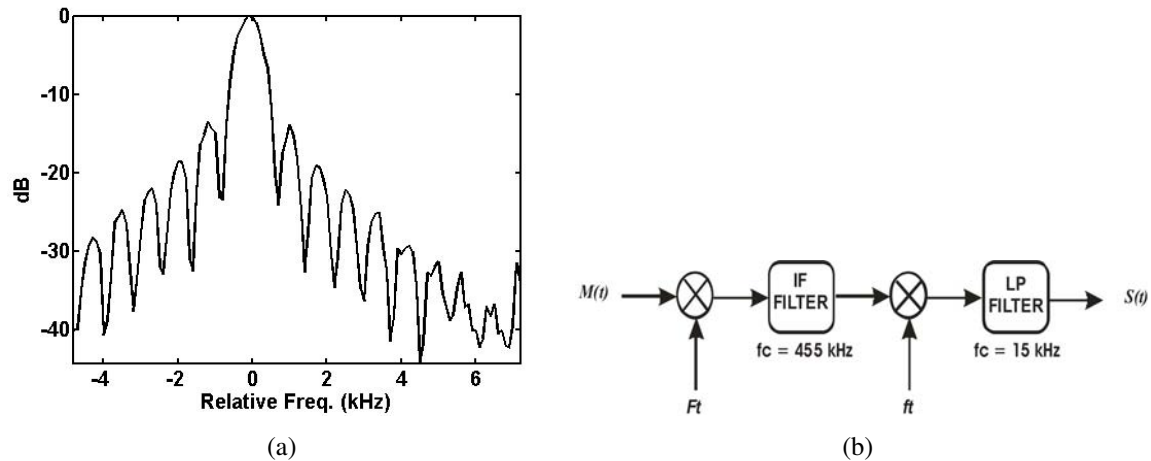


Figure 3.2: Narrowband data acquisition. Wideband sonar signal received by a microphone is amplified and fed to a super-heterodyne circuit that converts the acoustic energy at the desired frequency (1 kHz bandwidth) to a lower frequency signal centered at 5 kHz. **(a)** Effective frequency response of a channel. Center frequency represented as 0 corresponds to tuning frequency. **(b)** A block diagram of the super-heterodyne principle. Wideband signal is first shifted to higher frequency such that the desired frequency is centered at 455kHz. This is achieved by mixing the microphone output with a square wave signal whose frequency is adjusted to tune the circuit to the desired frequency, ($f_{des} = f_{adj} - 455kHz$). A narrowband filter centered at 455 kHz isolates the desired frequency region from frequency up-shifted signal. Another mixing operation with 460 kHz square wave mixing signal down shifts the output of the filter to 5 kHz.

of the center-pole axis. Additionally, a microphone is mounted on the marker-head post to monitor outgoing sound intensity.

Each microphone channel is designed to relay narrowband signals with 1 kHz bandwidth, so that 22 or more channels can be acquired with one data acquisition board (NI PCI-6071E). For this purpose signals received by the microphones are fed to custom build super-heterodyne circuits (see Figure 3.2(b) for a diagram of the super-heterodyne principle) that isolates and transforms a narrow frequency band of a broadband signal to a low frequency signal centered at 5 kHz. This principle allows acquisition of the desired signal at a lower sampling rate. With heterodyning, the microphone array can be tuned to any frequency within the sonar range of the bat; 20 kHz to 100 kHz.

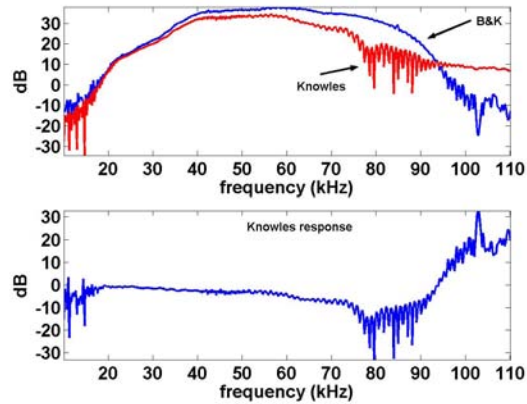


Figure 3.3: Frequency response of the Knowles and B&K microphones. A wideband sound pulse is presented 100 times and recorded via both microphones simultaneously. The top figure depicts the averaged sound spectra for both microphones. The frequency response of the Knowles microphone is obtained as the ratio of the signal spectrum obtained via the Knowles and the B&K microphones.

The particular type of microphones we use are not designed specifically for ultrasound frequencies. We used frequency matching filters to improve sensitivity at ultrasound frequencies before heterodyning. Using 1/8" B&K microphone that has flat frequency response (within ± 1 dB), as a reference, frequency response of the microphone-filter system is obtained (figure 3.3). Notice that with the matching filter, each channel can detect acoustic signals reliably up to 80 kHz.

Microphone array calibration

Determining the orientation of the microphones is an important because they can be positioned arbitrarily at any direction. With large number of microphones, a fast but effective method is preferable. An ultrasound loudspeaker, mounted on a computer controlled pan-and tilt mechanism is placed on the platform. The loudspeaker scans orientations that

constitute a uniform grid with 5° resolution within $\pm 45^\circ$ azimuth and $\pm 30^\circ$ elevation range. Linear frequency-modulated (LFM) pulses with 2 ms durations are broadcast at each speaker orientation. For each microphone the direction that provides the maximum output is determined as that microphone's position with reference to the pan-and-tilt coordinate system. In order to increase estimation accuracy spatial interpolation is used.

Each microphone channel can have different frequency and gain properties. To combine the spatial data from multiple microphones, differences between individual microphones need to be minimized. Towards this purpose, each channel's frequency response for a given tuning frequency is measured. After the loudspeaker is oriented towards a microphone, tone pulses with 1.5 ms duration are presented with 100 Hz steps within the frequency range of ± 10 kHz around the tuning frequency. The frequency response of a given channel is obtained by measuring the signal energy at the output of each channel. Channel differences are minimized by multiplying the channel outputs with the appropriate weights obtained from the channel responses obtained at the tuning frequency.

Reference microphone

Bats can change intensity and spectral-temporal characteristics of their sonar calls. Such variations between sonar vocalizations, if not dealt with, will introduce errors to the beam reconstruction process since the overall shape of the beam is obtained from spatial data collected from multiple vocalizations. Consequently, different snapshots (signals captured by the microphone array from a single sonar vocalization) may vary from each other even if the head position remains fixed in space. The nature of this additional source of variance is non-spatial, in other words, two snapshots from the same head pose will only differ

by a constant gain factor. In an ideal measurement scenario, identical sonar vocalizations would be recorded from the bat. Since this is not practical, we chose to place a reference microphone on the bat's head to monitor the changes in the intensity of the outgoing sonar vocalizations to remove the non-spatial variance during the computations. This microphone provides a direction-independent reference to follow the sound intensity across different vocalizations. Once the signals recorded with the microphone array are corrected by factoring out the intensity differences between vocalizations, the relationship between any two microphone signals should be purely spatial.

In figure 3.4 signal level at the reference microphone output with reference to intensity of the 2ms long LFM sound pulses generated by the loudspeaker is given. The microphone, glued on the marker-headpost, was placed on the loudspeaker. The signal level at the reference microphone output changed linearly with the normalized signal amplitude that was fed to the loudspeaker. The linear relation indicates that the reference microphone is reliable enough to be used for the compensation of the intensity differences between the emitted sounds.

Head tracking

Head tracking is an essential part of this study. During the beam pattern measurements the bat is unrestrained and free to move on the platform. An accurate estimation of the head orientation is essential to determine the relative directions where microphones sample the beam. This is accomplished by placing a light weight apparatus, the marker-headpost, with

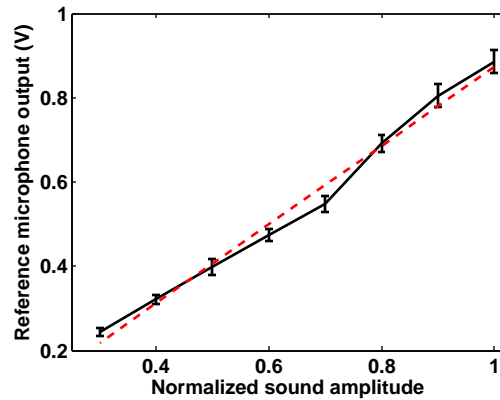


Figure 3.4: Reference microphone performance. The reference microphone placed on the marker-headpost reliably tracks the intensity of the sound emitted by the loudspeaker.

five infrared reflective markers on the head (figure 3.5). At the end of each measurement session the marker-headpost is removed for the comfort of the bat. Sonar beam measurements for different frequencies are made in different days. Therefore, a stable coordinate system attached to bat's head is important to obtain day-to-day consistency. To achieve this we glued a threaded metal post on the skull such that it stands perpendicular to the skull surface. The marker-head post is mounted on the metal post before each measurement session. Because the marker-headpost is designed to fit in one way with the post, the day-to-day consistency of the head-centered coordinate system is assured.

Video images of the markers are obtained via a fixed camera that monitors the bat during a recording session. Images from this camera are later used to predict the orientation of the bat's head. Because the relative spatial relationships of marker points are known, it is possible to compare camera images of the markers in two different poses and compute the relative transformation in terms of rotation and translation parameters (all 6 of them). To

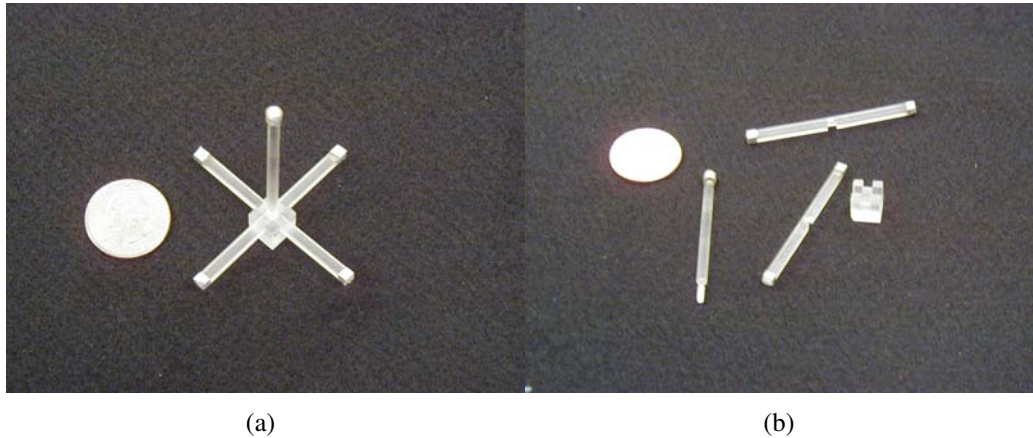
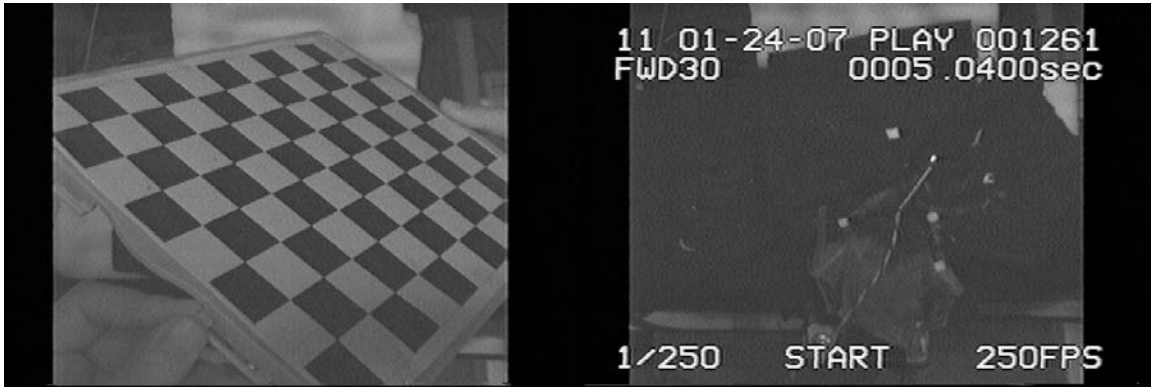


Figure 3.5: Marker-headpost. **(a)** The headpost consists of five equal length arms with infrared reflector tapes at the end of each arm. **(b)** The marker-headpost can be disassembled.

accomplish this with one camera, the intrinsic and extrinsic parameters of the camera also need to be known. We have used a freely available software, a camera calibration toolbox for MATLAB¹, to calibrate the camera. The software uses different poses of checkerboard pattern with known dimensions (see figure 3.6(a) for a sample frame) to estimate the camera parameters.

The head pose estimation process is done offline, using video data obtained in synchrony with the microphone array recordings. Analysis involves extracting infrared reflective marker image positions by a custom developed software in MATLAB. The software estimates marker point locations based on the contrast difference with the background (see figure 3.6(b)). Once the image positions are obtained an algorithm proposed by Dementhon and Davis (1995) is implemented to estimate the head pose parameters.

¹MATLAB camera calibration toolbox is available at http://www.vision.caltech.edu/bouguetj/calib_doc



(a)

(b)

Figure 3.6: Head Tracking. **(a)** A calibration frame. Checkerboard pattern with known dimensions is monitored with the camera at different orientation to obtain camera calibration parameters. **(b)** A video frame during pose estimation. Five infrared-reflecting markers' coordinates in the video image is used to determine the head pose. The relative position of the markers, with reference to each other, need to be known a priori for the head pose computation.

The accuracy of the head pose estimation is tested using a pan-and-tilt mechanism mounted an ultrasound loudspeaker to predetermined poses. The marker-headpost is placed on the loudspeaker and images of the markers are taken at predetermined orientations of the loudspeaker using the calibrated camera. Employing the camera tracking software of-line, the marker-headpost orientations are estimated and compared with the orientation parameters used to control the pan-and-tilt mechanism. Figure 3.7 shows azimuth and elevation coordinates of the different poses on the pan-and-tilt reference frame (in black) and the estimated pose parameters for the marker-headpost after being transformed to the same coordinate system (in red).

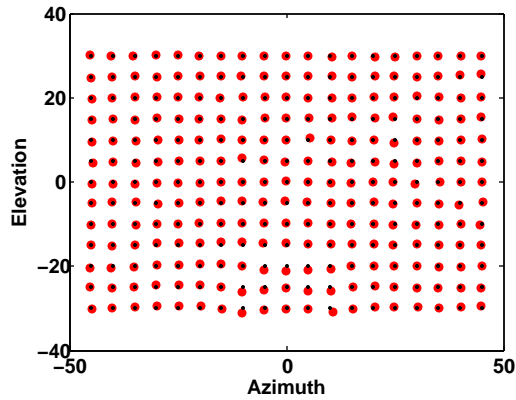


Figure 3.7: Head pose estimation using video images obtained with the infrared camera. The black dots are speaker orientations in azimuth and elevation coordinates given to the pan-and-tilt mechanism. The red dots represent the estimated orientation parameters estimated with the camera-based head pose estimation method. Since both sets of parameters are in different coordinate frames (one in head pan-and-tilt reference frame, the other is in camera centered reference frame) estimated head pose parameters are converted to pan-and-tilt reference frame.

3.2.4 Experiment and sonar beam construction

An echolocating bat, *Eptesicus fuscus*, is trained to remain on the platform and encouraged to make sonar vocalizations towards a tethered mealworm moving in front of it. Bat vocalizations are captured in 8.7 sec long trials. Simultaneously a high-speed video camera (250 frame/sec) positioned above the head records head movements. Beam pattern measurements are made at frequencies of 35, 40, 45, 50, 55, 60, 65, 70 kHz. Measurements for different frequencies are made on separate days.

In order to reconstruct a beam pattern, vocalizations from multiple trials at a given frequency are pooled (100 vocalizations on average, per frequency). For each trial, timing of the vocalizations are obtained using the reference microphone channel output. Using the time index of the vocalizations, a fixed-length window whose length is tailored to individual vocalizations, is used to extract the signals from the output of each channel. A Fourier

transform is applied to each window across all the channel outputs for a given vocalization and the frequency component at 5 kHz is determined. These values are divided to the similar values obtained at the reference microphone to predict relative sonar beam amplitudes at the sampled position in space.

Amplitudes obtained in this fashion, across different vocalizations and trials, are later pooled to obtain enough spatial coverage and density to reconstruct the beam. Spatial smoothing and interpolation are implemented using spatial filtering and spherical harmonic decomposition (SHD), to obtain the estimated sonar beam.

3.2.5 Stability of the measurement set up

In order to test the reliability of the measurements we employed the pan-and-tilt mechanism to move a ultrasound loudspeaker at the center of the measurement setup. Ability to control the speaker orientation provides us with an opportunity to compare spatial variations of beam measurements for both bat and the loudspeaker. A loudspeaker beam is stable, thus beamshape measurements from the loudspeaker will provide an estimate of the measurement reliability of the measurement setup. The pan-and-tilt setup also provides a test bed for the camera tracking method and allows for estimation of accuracy of the head tracking method. A difference in measured spatial variations between bat and loudspeaker beam can then be attributed to the variability of the bat's beams patterns.

Frequency response stability

An important design characteristic we demand from the measurement system is stable frequency tuning. The tuning properties of the heterodyne system can be influenced by temper-

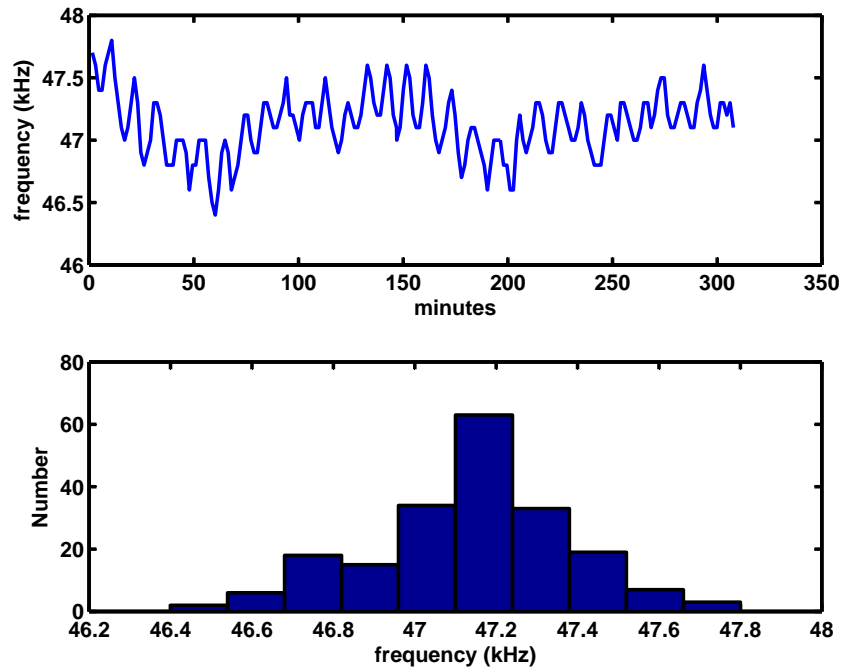
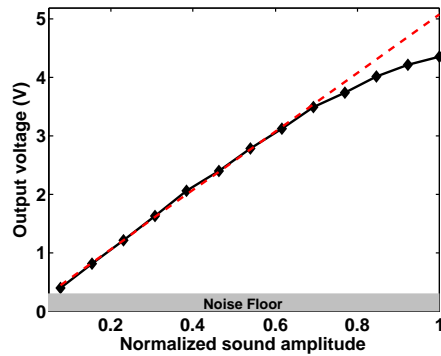


Figure 3.8: Tuning frequency stability with time. **(a)** Tuning frequency change with time. **(b)** Histogram of the measured tuning frequencies. Tuning frequency within 200 Hz around the preset value (47.2 kHz)

ature changes and other internal and external effects to circuitry. In order to determine variation in the frequency tuning we took measurements of tuning frequency at every 2-3 minutes for approximately five hours. Tuning frequency is estimated from the frequency response of an arbitrarily chosen channel. Figure 3.8 shows the result of these measurements. The system is manually tuned to approximately 47 kHz and, on average, maintained that frequency throughout the measurement session. A distribution of the frequencies are given in figure 3.8.



(a)

Figure 3.9: Dynamic range of a typical channel for frequency modulated pulses. The normalized amplitude of the LFM pulses are given on the x-axis and corresponding channel output voltages on the y-axis. The dynamic linear range (from 0.07 to 0.7 normalized signal amplitude) within which the channel output can reliably represent sonar signal amplitude is approximately 40 dB.

Dynamic range of microphone array

The data acquisition system must have a large enough range to reliably monitor sonar signals. By presenting 2 ms long LFM pulses with different amplitudes to one of the array microphones we monitored the signal level at the channel output. Figure 3.9 depicts the result of these measurements. The input signal amplitude is normalized. Since the gain of the individual channels can be set independently, the rule of thumb is to adjust the dynamic range to a level that fits to individual bats' vocalization intensity range. The channel output is linearly related to the input amplitude between 0 and 0.7 normalized input levels corresponding 0 V and 3.5 V respectively. The nonlinearity above 3.5 V is caused by the amplifier and super-heterodyne circuits. From figure 3.9 the linear region is between 0.007 and 0.7 corresponding to approximately 40 dB range of acoustic pressure. The lower limit of this range in practice is determined by the noise conditions at time of measurements.

3.3 Results

3.3.1 Beam patterns of a stable emitter

The proposed method involves processing of the acoustic information obtained from multiple microphones, and monitoring the bat's head position during vocalizations. Consequently, reconstructing a sonar beam with this technique involves accurate estimation of head positions, microphone positions with reference to the bat's head as well as their frequency response and acoustic energy generated by the bat at the location of the microphones. An accurate sonar beam reconstruction requires each component this method to be accurate and stable. We followed a step by step procedure where, at each step, a new component of the method was introduced and evaluated. During these test we used an ultrasound loudspeaker. The orientation of the loudspeaker and the acoustic signals it generates were controlled via a computer.

Beam reconstruction with the microphone array

The loudspeaker oriented to angles within $\pm 45^\circ$ azimuth and $\pm 30^\circ$ elevation range with 5° steps in each dimension. The acoustic signal energy measured at each microphone output is assigned to corresponding azimuth and elevation angles with reference to the speaker orientation. Acoustic signals generated by the loudspeaker were 2 ms long LFM pulses. Speaker beam patterns were obtained for frequencies between 35 to 70 kHz separated by 5 kHz intervals. The beamshape is estimated by fitting a two dimensional surface to data at

each frequency using spatial interpolation and smoothing (figure 3.9 and 3.11).

The beams obtained with the microphone array show increased directionality with frequency (figure 3.12) as expected. Furthermore, the main lobe axes across the different frequencies were found to be very small within 3° . Constancy of the beam axis across the frequency range is a feature of the circular loudspeaker we used.

To determine whether the loudspeaker beam obtained from the microphone array is a reasonable estimate of the loudspeaker beam we compared the reconstructed beam with the beams that are obtained with a single microphone positioned at the center, at approximately 0° azimuth and elevation. In order to compare the two sets of beam pattern estimates correlation of the beamshapes from a single and array of microphones are obtained, at a given frequency. We calculated the spatial variability in a beam obtained by the microphone array that could be explained by a beamshape estimated by a single microphone, at the same frequency, using regression analysis. The match between these beam estimates was defined as the the square of the correlation coefficient obtained.

Results obtained from the microphone array and the center microphone show 75% similarity on average (figure 3.13). The loudspeaker beam patterns obtained with the microphone array are reliable in terms of matching a beam pattern that obtained by a single microphone. However, there is still a moderate amount of mismatch (25%).

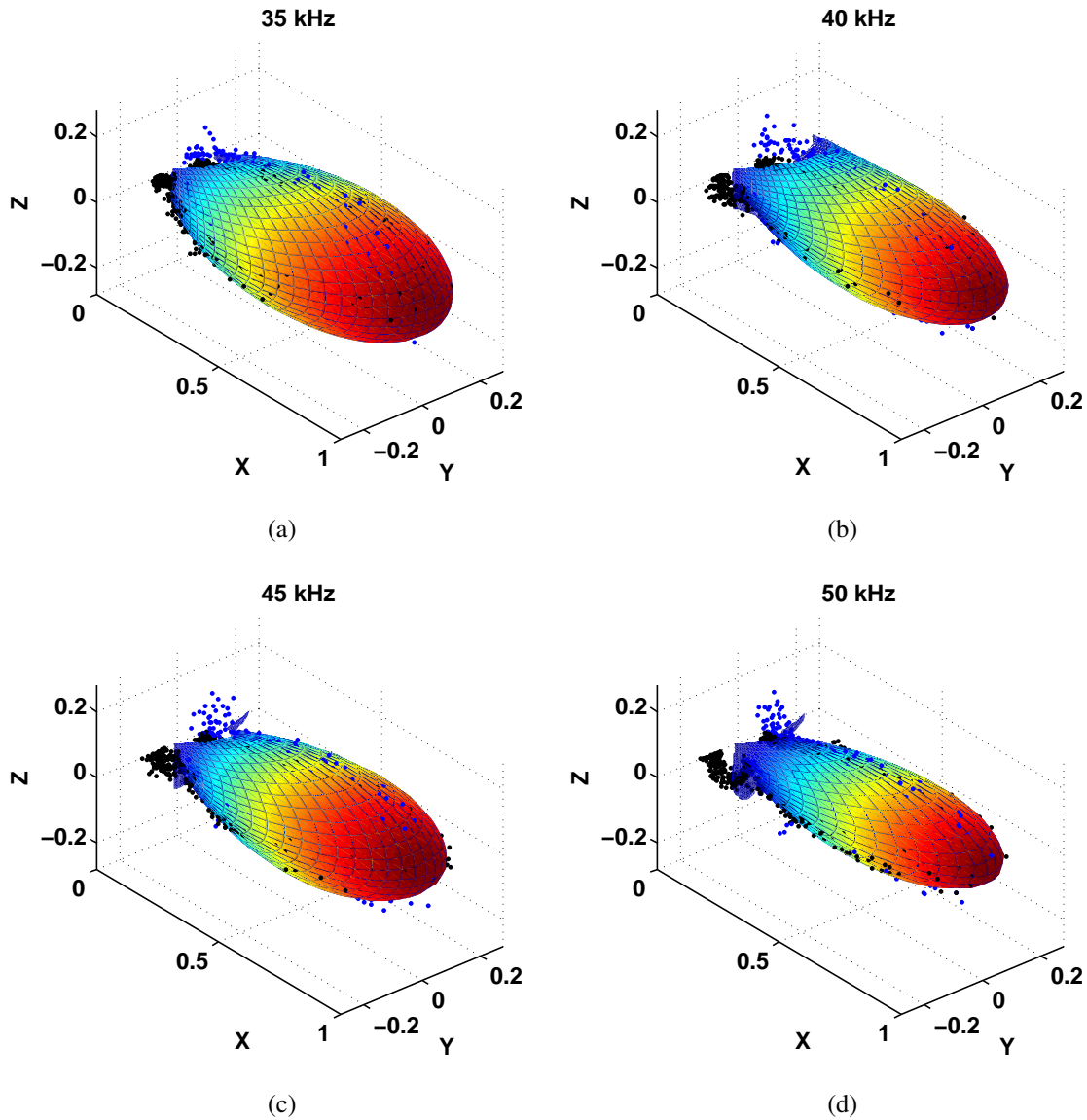


Figure 3.10: Speaker beams obtained by the microphone array. The loudspeaker was mounted on a pan-and-tilt mechanism and oriented towards positions that are 5° apart, on a grid spanning an area from -45° to 45° azimuth and -30° to 30° elevation. **(a to d)** normalized beams obtained for frequencies 35 to 50 with 5 kHz steps respectively. Blue and black dots are raw data points in the neighborhood of azimuth and elevation planes (within 3°) respectively.

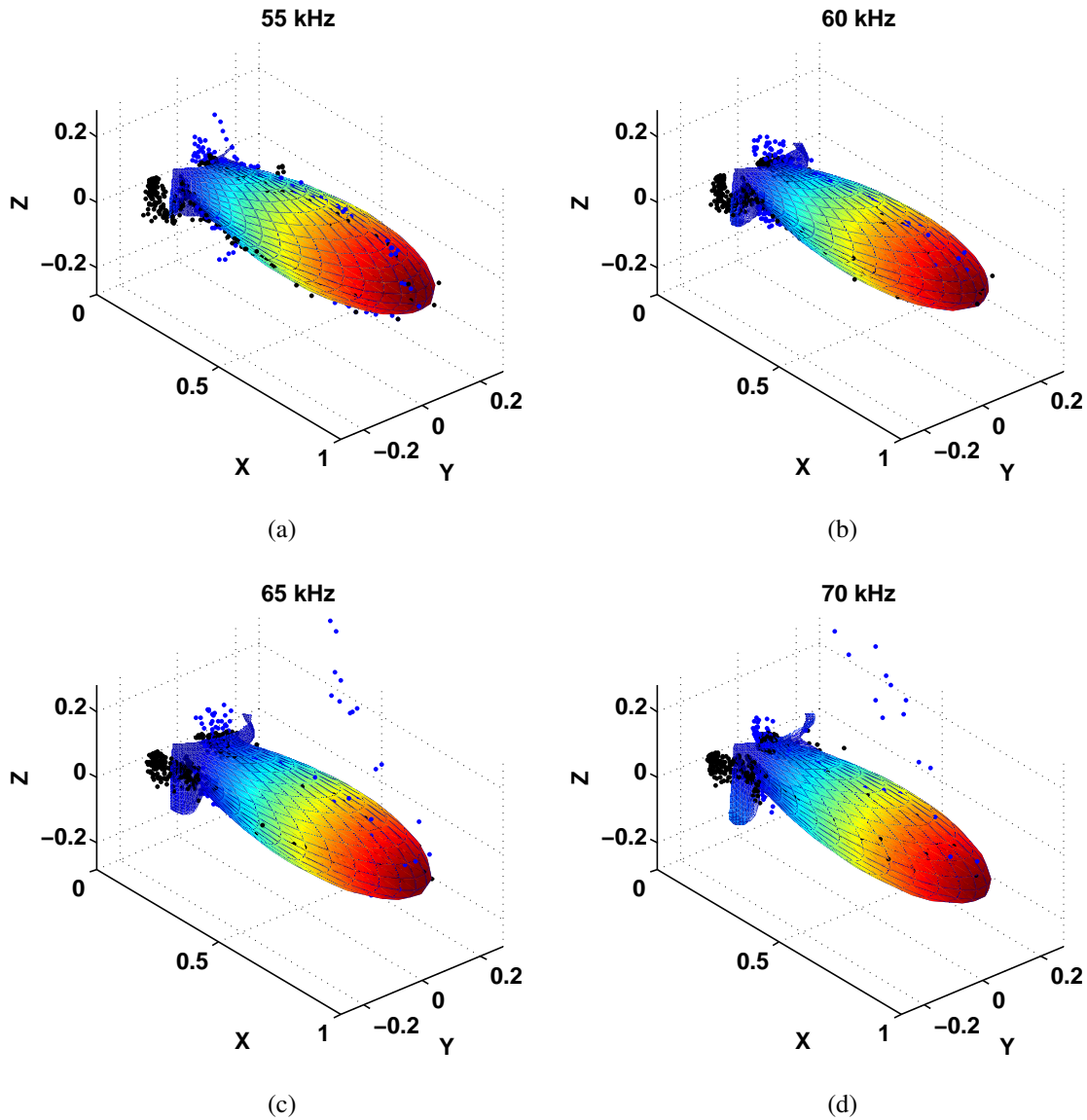


Figure 3.11: Speaker beams obtained by the microphone array. The loudspeaker was mounted on a pan-and-tilt mechanism and oriented towards positions that are 5° apart, on a grid spanning an area from -45° to 45° azimuth and -30° to 30° elevation. **(a to d)** normalized beams obtained for frequencies 55 to 70 with 5 kHz steps respectively. Blue and black dots are raw data points in the neighborhood of azimuth and elevation planes (within 3°) respectively.

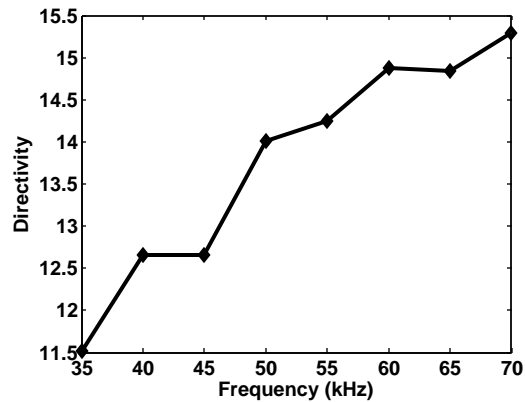


Figure 3.12: Ultrasound loudspeaker directivity as a function of frequency. Directivity measured in the spatial region limited to $\pm 45^\circ$ azimuth and $\pm 30^\circ$ elevation range. Directivity increases with frequency.

To better understand the nature of this mismatch we investigated the residual error between the estimated beam surface and the raw data points obtained from single and multiple microphones. We used a measure that compares the variance of the residual error with the variance of the surface obtained. We refer to this measure as beam-to-noise ratio (BNR) and computed it by taking the ratio of the two variances. BNR is large when the residual error is small and vice versa.

Figure 3.14 shows the BNR values obtained for both sets of speaker beams. In parallel with the similarity measurements, the beam patterns obtained from a single microphone give larger BNR values than those obtained from the microphone array. BNR values associated with the microphone array measurements indicate that the measurement noise is not large enough to prevent consistent estimates of the beam pattern shape. The cause for the larger variance and mismatch associated with the microphone array could be related to the differences between the microphones caused by imperfect compensations and errors

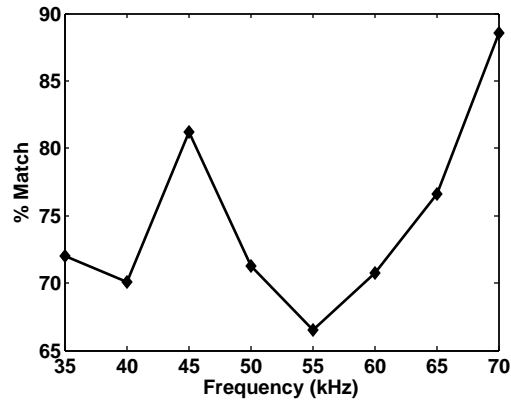


Figure 3.13: Comparison of speaker beams obtained from a single microphone and the microphone array for frequencies between 35 and 70 kHz in 5 kHz steps. A correlation coefficient r is computed using linear correlation between the two sets of beam samples at each frequency to determine similarity. Percent match of the beams from the microphone array to the single speaker beams is represented by r^2 .

related to the estimation of the the microphones positions on the setup.

Speaker beams obtained from the microphone array for known and estimated speaker orientations

In order to test the reliability of the method, we compare speaker beams obtained from the microphone array for known speaker orientations (static condition) with the beams that are reconstructed from a speaker that is in motion for 8 seconds while presenting 2 ms long LFM signals with regular intervals (dynamic condition). The orientation of the speaker in the second case is determined by video data obtained with high speed camera. Vocalization times are estimated by analyzing the reference microphone channel output.

Speaker orientations in both static and dynamic conditions are selected on a curve that makes a figure-eight in azimuth and elevation coordinates. Mismatch variance of 0.0387

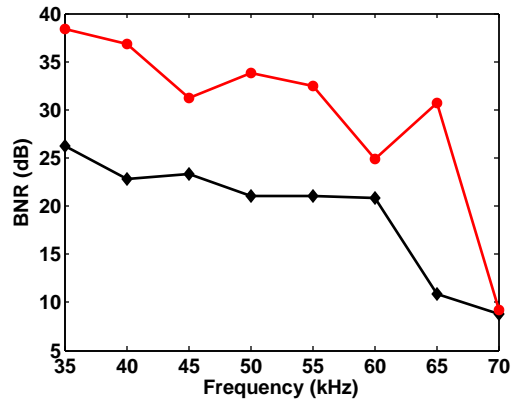


Figure 3.14: Beam-to-noise ratio (BNR) for estimated speaker beams using the microphone array (black line) and a single microphone (red line) across frequencies. For both conditions, BNR values are well above zero suggesting the consistency of measurements. BNR values obtained from a single microphone are moderately lower suggesting that usage of multiple microphones introduced additional variance to the beam reconstruction.

($r = 0.991$) is obtained when correlation between both beams are computed. The match of the two beam patterns suggests that addition of the camera tracking does not degrade the reliability of the beam reconstruction process.

The step by step analysis of the measurement system with a ultrasound loudspeaker revealed that the proposed measurement method is effective and provide reasonable and consistent estimates of beam patterns. Although using multiple microphones increases measurement error, beam patterns estimated via the microphone array are stable and robust. Spatial variations measured as BNR values show that variation increased with sound frequency.

3.3.2 Beam patterns of bat sonar calls

Having established that the measurement setup is capable of generating meaningful beam patterns from ultrasound emitters we tested the new method on an echolocating bat. Figures

3.15 and 3.16 show the reconstructed beam patterns obtained for frequencies between 35 kHz to 70 kHz with 5 kHz steps. These figures depict raw data points and the surface that is fitted to the data. The surfaces are obtained through local spatial smoothing and interpolation. For spatial smoothing data points that are within 8° neighborhood of each other are determined. In each local area the median acoustic intensity value is assigned to the spatial location at the center of the neighborhood. Then a two dimensional surface is fitted to the median values using SHD.

Unlike what we have observed from the loudspeaker tests, there is a larger spatial variation in the raw data obtained from the bat sonar beams. More specifically, the bat sonar beam showed large variations for a given relative orientation. This result is also reflected in BNR values obtained at each frequency (figure 3.17(a)). In figure 3.17(a), in addition to BNR values obtained from bat sonar vocalizations, BNR obtained from the loudspeaker is redrawn for comparison. Spatial variations of the bat sonar emission patterns are clearly lower. BNR did not increase with frequency for bats.

In figure 3.17(b) a histogram of the residual deviations of the raw data from the estimated surface are given at 55 kHz. The distribution of the deviations were similar in other frequencies. A range of 30 dB deviation around the fitted surface with a standard deviation of 6.26 dB is observed.

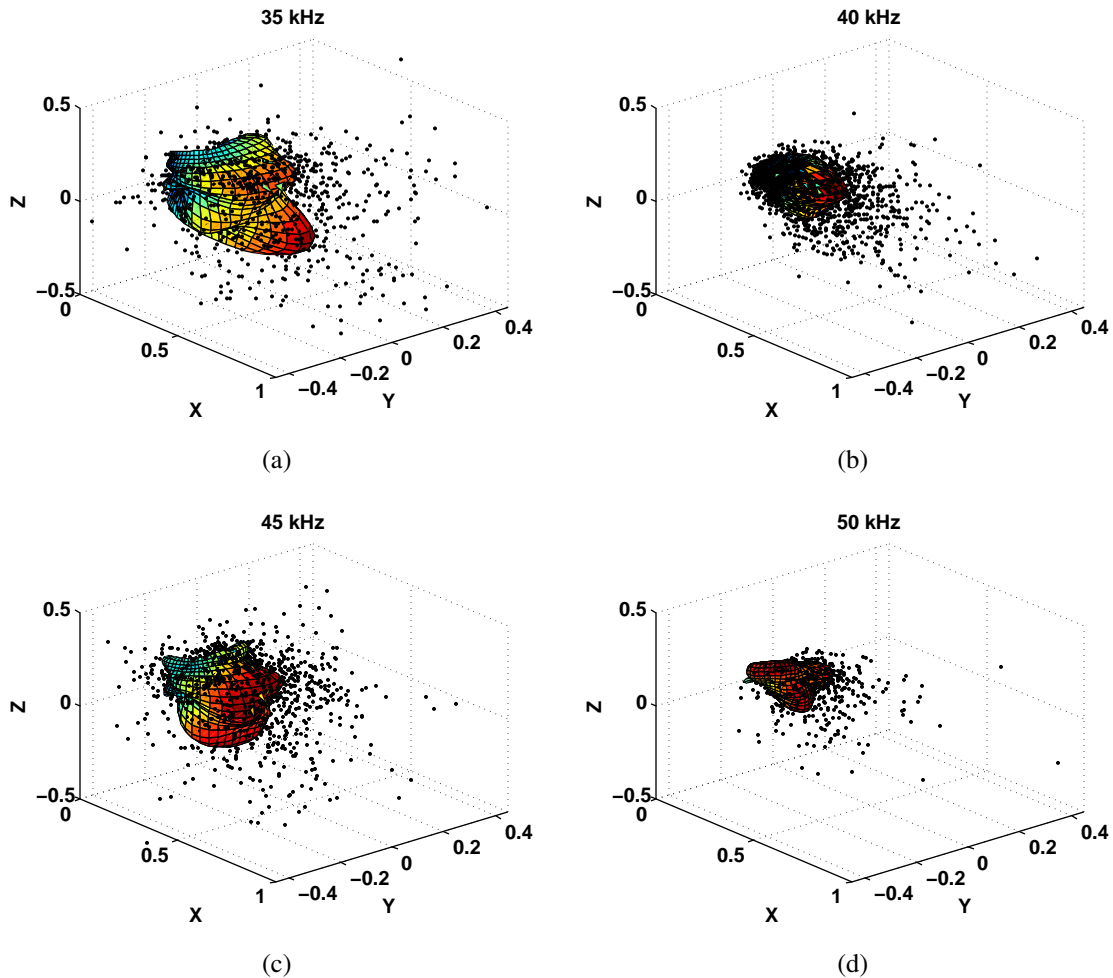


Figure 3.15: Bat sonar beams obtained by the microphone array. The bat placed on the platform at the center of the microphone array freely vocalizes while its head orientation is monitored by high speed infrared camera. **(a to d)** sonar beams are normalized with reference to the raw data obtained for frequencies of 35 to 50 kHz with 5 kHz steps. Black dots represent the raw data points obtained at a particular direction. The distances of these points from the origin represent the intensity of the sonar beam in that direction for that vocalization. Surfaces fitted to the raw data represent an estimate of overall shape of the sonar beam under the assumption that the beam is constant across different vocalizations.

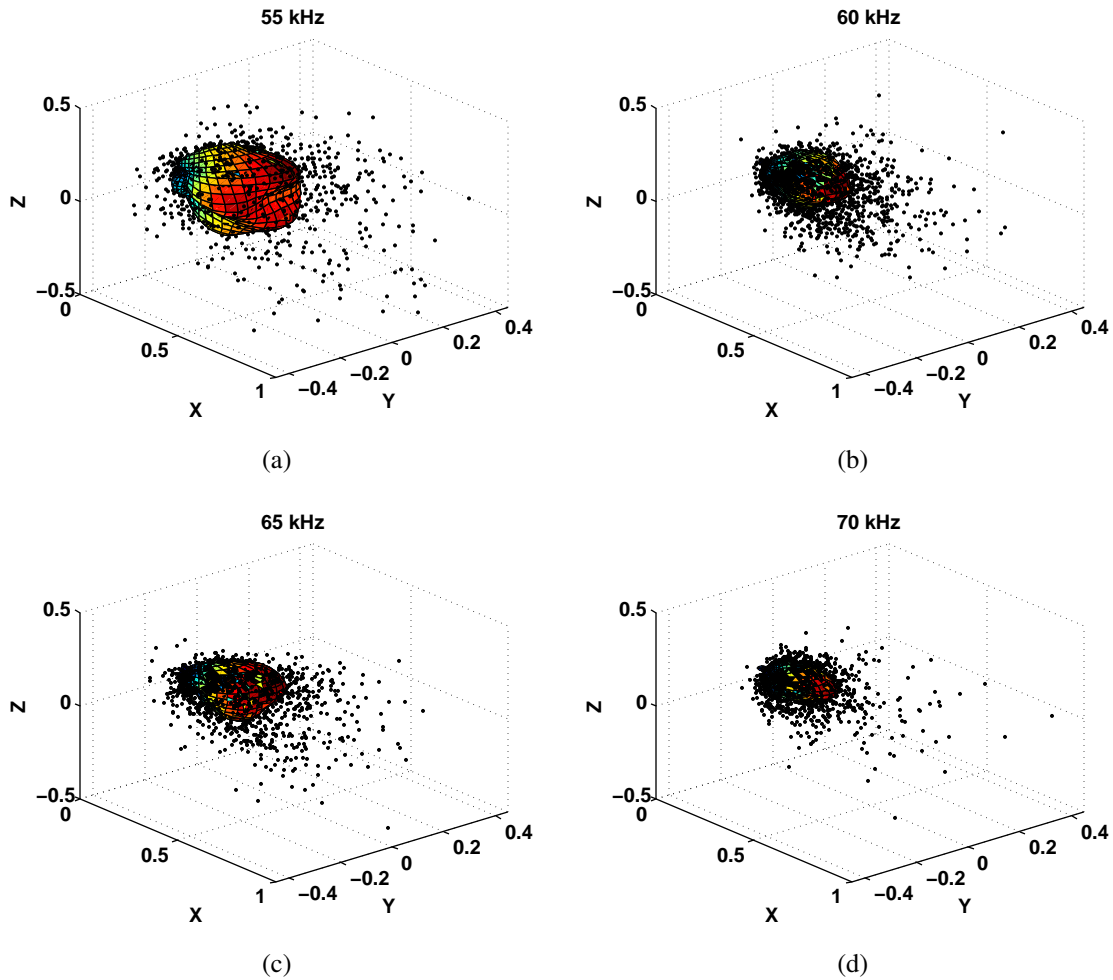


Figure 3.16: Bat sonar beams obtained by the microphone array. The bat placed on the platform at the center of the microphone array freely vocalizes while its head orientation is monitored by high speed infrared camera. **(a to d)** sonar beams are normalized with reference to the raw data obtained for frequencies of 55 to 70 kHz with 5 kHz steps. Black dots represent the raw data points obtained at a particular direction. The distances of these points from the origin represent the intensity of the sonar beam in that direction for that vocalization. Surfaces fitted to the raw data represent an estimate of overall shape of the sonar beam under the assumption that the beam is constant across different vocalizations.

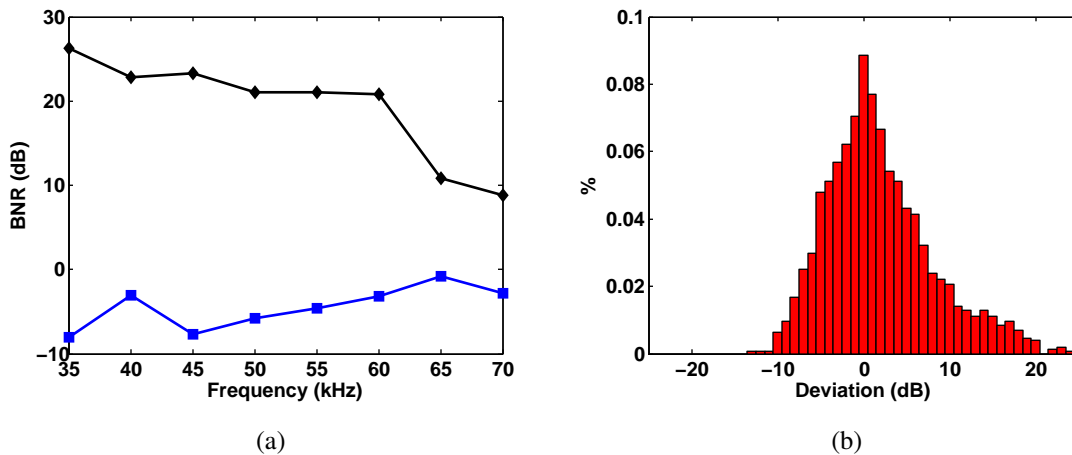


Figure 3.17: Spatial variation of bat sonar beams. **(a)** The variation around the estimated beam surface is larger than the spatial variation associated with the beam surface (blue). BNR values as a function of frequency for the loudspeaker beam obtained via microphone array is also given for comparison (black). **(b)** Histogram of the dB deviations from the estimated surface intensity and the raw intensity data across the sampled spatial orientations for 55 kHz.

Although the sonar beamshapes obtained from the bat may not be reliable in their directional features, the overall distribution of the raw data points cover a limited spatial area and show directionality. In order to capture the gross directional characteristics of the beams, we computed the spatial envelope of the raw data. In figures 3.20 and 3.19, we show contour representation of the beam amplitude in the frontal hemisphere space of the bat. Although, bat sonar beams seem to favor the frontal directions, frequency dependent narrowing of the beam expected from a fixed-aperture sound emitter is not obvious in these results.

In light of these observations bats may not qualify as stable ultrasound emitters. However, it is also possible that the larger variation might be a result of the experimental condi-

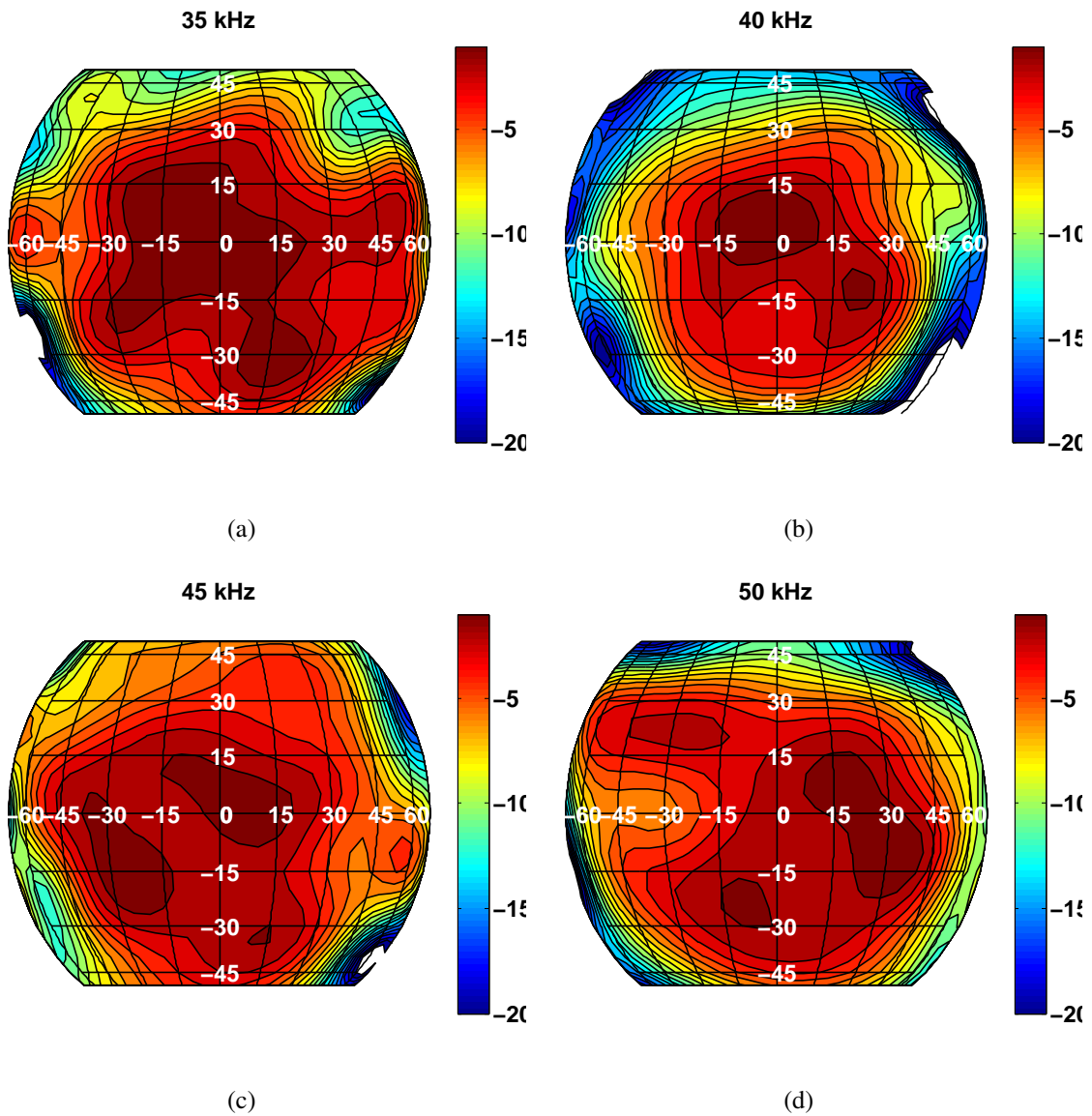


Figure 3.18: Bat sonar beams obtained as the envelope of the raw data points. **(a to d)** Estimated beam patterns for frequencies of 35 to 50 with 5 kHz steps respectively. Contour lines are drawn at every 2dB. Beam amplitudes are normalized with reference to the maximum value.

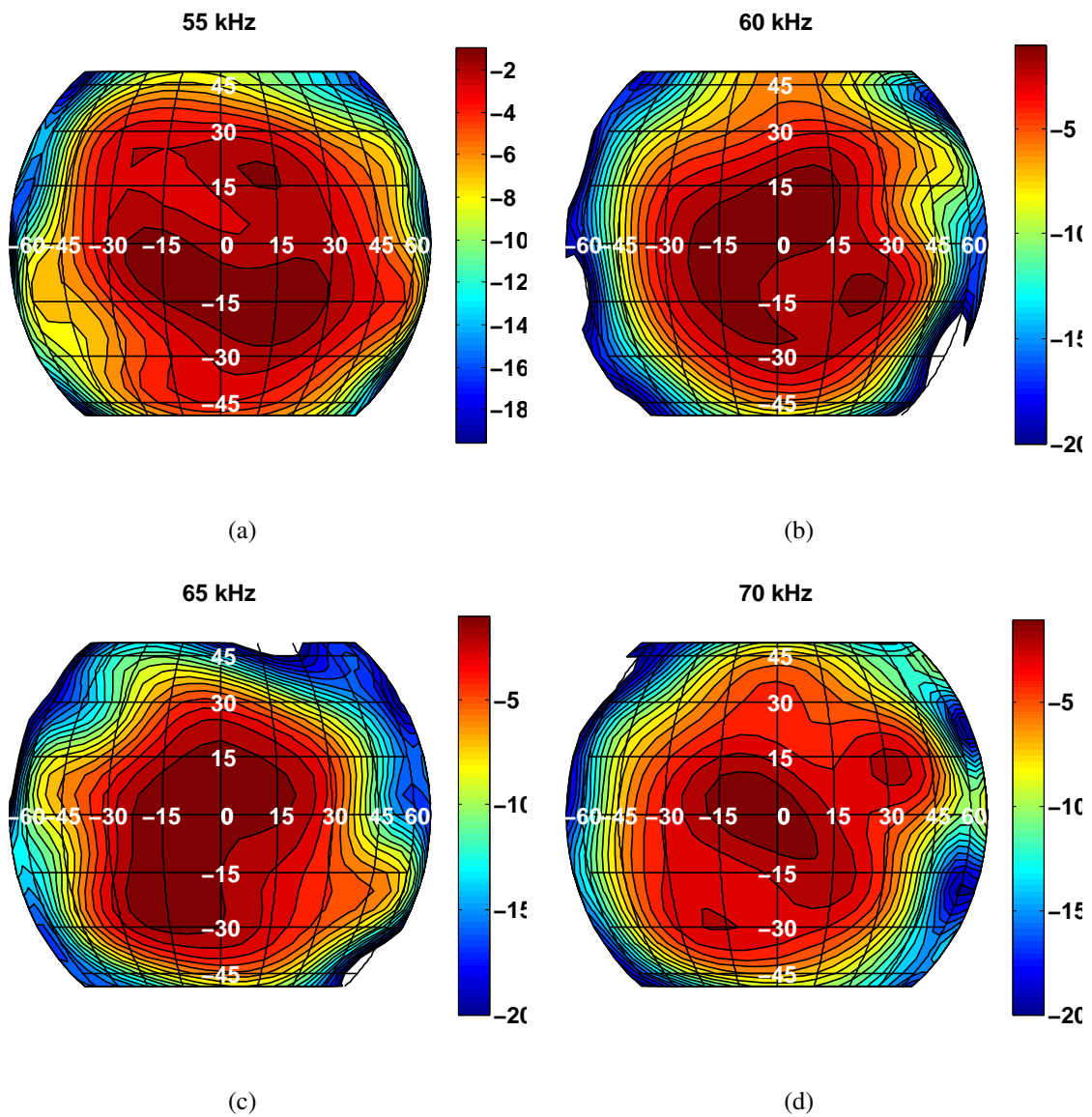


Figure 3.19: Bat sonar beams obtained as the envelope of the raw data points. **(a to d)** Estimated beam patterns for frequencies of 55 to 70 with 5 kHz steps respectively. Contour lines are drawn at every 2dB. Beam amplitudes are normalized with reference to the maximum value.

tions. The sonar vocalizations generated by the bat were generally louder than the speaker generated sounds that could result in stronger echoes from the measurement setup. These echoes can interfere with the measurement. Being free to move the bat, can position itself in such a way that the outgoing sonar emissions could interact with the platform it sits on. This interference would cause unintended changes in the measured beamshape. Last but not least, inaccurate intensity estimates of the outgoing sonar vocalizations from the reference microphone signal output could introduce variation between snapshots.

Sonar vocalizations can be emitted at different phases of mouth opening. As a result, the sonar beam pattern will change as a function of mouth opening. The effects of these type of structural changes can be larger in near field that it is in the far field. The reference microphone, being in the near field, could be influenced by these changes while effect on the microphones in the far field might be much smaller.

3.3.3 Reference microphone and spatial variance

Assume two points on the beam that are sampled simultaneously by two sonar vocalizations. If the reference microphone is the dominant contributor of the spatial variance, the ratios of the beam intensities between the two points obtained for each vocalization should be identical. We isolated such pairs of points that differ less than 4° between vocalizations. For each point pair ratios of the estimated beam intensities were calculated in dB units. We determined the similarity of two matching intensity ratios by subtracting them from one another. The result of this subtraction should be equal to zero, if the sonar beamshape is constant. The standard deviation of the subtractions across the measured frequencies was

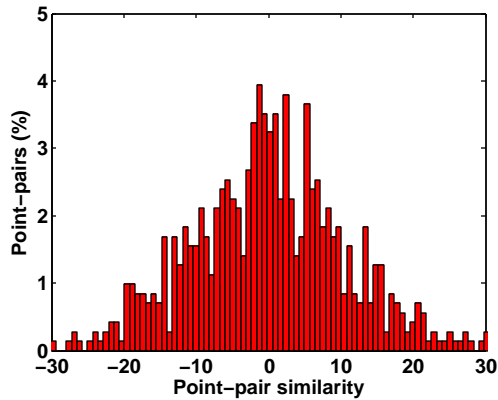


Figure 3.20: Similarity of intensity ratios of sonar beam point-pairs between two vocalizations at 50 kHz. If the sonar beamshape is fixed then similarity values should give a small variation around 0. For 50 kHz similarity values this variation is 11dB. Variation is 10.57 dB on average across the measured frequencies.

on average 10.12 dB. This result is comparable to the variation of measured amplitude values around the estimated sonar beam surface. A distribution of the similarity values for 50 kHz is given in figure 3.20.

Based on this result it can be concluded that the reference microphone cannot be the cause for the spatial variation in the sonar beam measurements that are observed from the bat.

3.4 Discussion

Studying spatial properties of bat sonar beams is important in understanding spatial hearing by echolocating bats and, particularly, spatial localization of echoes. In general, beam patterns of sound emitters become more directional with increasing sound frequency. Bats seem to be no exception to this rule: Previous studies on bat sonar emission patterns show

that directionality of the sonar beam increases with frequency (Schnitzler and Grinnell, 1977; Hartley and Suthers, 1989). This results in a low pass filtering effect on echoes from objects that are off with axis to the beam. The filtering effect can influence the sound localization process, since an accurate sound localization requires wideband acoustic inputs. The absence of higher frequencies can result in less accurate estimates of the sound location. Hence, a bat's ability to localize echo sources may not be uniform in the frontal hemisphere and it is expected to become more accurate for spatial locations near the main axis of the sonar beam.

Most of the previous attempts to measure sonar emission patterns especially for *Eptesicus fuscus*, were made under conditions that prevented the bats' natural echolocation behavior. Sonar vocalizations were produced by these bats as a result of brain stimulations. Limitations on the bats behavior and unnatural production of sonar signals raise the question of whether sonar beams obtained under these conditions reflect what a bat naturally generates. Furthermore, studies on other bat species revealed that sonar emission patterns may not be constant (Schnitzler and Grinnell, 1977). Can echolocating bats alter their sonar beam pattern to control and manipulate the spatial information they receive from their environment (see proposition made by Hartley for instance)?

The answer to this question requires an approach that permits sonar beam pattern measurements from freely behaving bats. The exploratory study presented here is a first step in investigating the feasibility of measuring sonar emission patterns of bats engaging in natural echolocation behavior. The main difficulty with this type of attempt is to know at which relative positions, with reference to bat, a measurement is taken. This requires knowing bat's head position at the time of vocalizations. A second demand is to obtain the

beamshape with high spatial resolution. Obtaining a detailed shape of the bat's sonar beam from each individual sonar vocalization would be ideal but technically very demanding. For this reason we have taken an intermediate step between this ideal method and methods that immobilize subjects during the measurements.

We have assumed that bat sonar beams are constant. If this is true then it is possible to rely on spatial information obtained from different vocalizations to build a complete picture of the sonar beam pattern. In order to give the bat mobility and still be able to determine the head position in space we employed a high speed camera tracking method. With each vocalization 21 new samples from the sonar beam were obtained. Spatial samples were later brought together to estimate the sonar emission pattern.

The implementation of this alternative measurement method required developing practical solutions to a number of technical challenges. First, in order to record from a large number of microphones simultaneously, we sacrificed measurement bandwidth. By limiting the bandwidth, it was possible to use a smaller sampling rate and support more microphones with one data acquisition system. Second, it was important to use small size and reasonably priced microphones. The size of the microphone affects its directional properties. A microphone with smaller dimensions is less directional and is less influenced by sound source location. Third, a method that is minimally invasive and easy to implement, and accurate for head pose estimation was necessary.

Less restriction on bat's mobility resulted in more complex technical solutions. To evaluate the feasibility of the resulting measurement system, we tested its performance while increasing its complexity by adding each essential component of the method one at a time. The result of these test from a stable sound emitter revealed that the measurement setup

performs as intended.

The results we have obtained from echolocating bats however, were not as successful. At every frequency tested, the sonar beam measurements were highly variable. A negative result like this raises more questions than it answers. One interesting explanation for this failure could be that the assumption about the constancy of the sonar beamshape of a bat is not valid and bat sonar emissions varies in shape. On the other hand it is also possible that failure is caused by a variety of reasons relating to the experimental conditions.

Although it is possible, to achieve an exhaustive investigation of each potential explanation for the large spatial variations, we believe that we have eliminated the most important ones. We show that the measurement system is capable of obtaining the beam patterns of emitters with constant beam. On the other hand there could be reasons that can not be replicated by the loudspeaker experiments. For instance, the bat's vocalization could interact with the sonar vocalizations because the bat's head was not always far from the edges of the platform. Furthermore, bat sonar vocalizations were in general louder and could potentially create larger echoes from the setup that can interfere with the acoustic signals measured. Lastly, it is possible that the reference microphone output was not reflecting the outgoing vocalization strength reliably. Being close to the mouth it could be subject to changes caused by different mouth openings at different vocalizations that are more prominent in the near field than they are in the far field. In these conditions, although, the far field signals received by the array microphones records from a near constant beamshape, the variation in the reference microphone may cause the variability in our results.

3.4.1 Implications of variable sonar beam patterns

If the sonar beamshape of echolocating bats do vary however, it raises interesting questions: What is the source of this variation? Can it be controlled by the bat. What do sonar beam patterns of individual vocalizations look like? What are the implications to spatial hearing, particularly to sound localization?

Echolocating bats that generate sonar signals from elaborate nasal structures could change their sonar emission patterns voluntarily by moving different parts of these structures (Schnitzler and Grinnell, 1977; Hartley and Suthers, 1987; Hartley, 1990). However, a bat that emits sonar signals orally, the only obvious way to introduce change to a beam pattern could be by controlling how much they open their mouths before and during the vocalization. A previous result by Shimozawa et al. (1974) suggests that bats are vocalizing at and around the peak opening of the mouth and the opening remains within the range of 30° and 60° . This resulted in variation within $\pm 12\%$ of the mean amplitude value obtained for the beam at a particular position. This variation is negligible compared to variations we have observed in our study. It is possible, however, Shimozawa et al. study does not reflect the variability that can be obtained from freely behaving bats due to artificial methods they used to elicit vocalizations. In our experiments the bat was not trained for a particular task, hence, there was no constraint on the type of vocalization it generates.

Bats could use different mouth openings during vocalizations for purposes that are not directly associated with beam shaping. For instance, bats could open their mouths to vocalize louder or to change the frequency content and structure of the outgoing sonar signals. Mouth acting as a resonance cavity could boost and suppresses different frequencies at

different openings.

The voluntary manipulation of sonar beamshapes could increase functionality of the bat's echolocation system. A bat searching for a target at close range may prefer to decrease its sonar beams directionality, covering a larger frontal space. In a different scenario that requires monitoring a target at a particular direction while eliminating interfering echoes coming from other locations, a bat, this time, could prefer a highly directional beam. Whether bats control their emission pattern can be investigated with behavioral tasks that could demand a bat to seek different type of information. Estimating the beamshapes from individual vocalization should be the preferred method for these investigations.

Variability of the sonar beamshape and sound localization

Potential variability of the sonar beam could be particularly informative in our quest to understand how bats localize sounds. A frequency dependent directionality of the beam, as discussed earlier, will influence the frequency content of the arriving echo. Especially, complex beam patterns with spatial notches between two side lobes can result in spectral notches. Such features induced by the sonar beam pattern could interfere with the computation of sound location. Conversely, binaural difference cues particularly ILD-spectrum will not be effected by the shape of the sonar beam. However, its detectability will be influenced because of the low-pass effect. Hartley and Suthers (1989) reported that no significant sidelobes are observed in sonar beam patterns of *Eptesicus fuscus*, which could help to prevent appearance of the notch type features in the echo spectrum induced by the sonar beamshape.

If bats have constant sonar beamshape, it is possible that they can learn the localiza-

tion cues generated by the combination of directional filtering effect of the external ears and sonar emission pattern. This implies that there could be two sets of localization cues for passive listening and active listening (echolocation) for bats. As an alternative, since the effects of the sonar beam results in identical spectral features for both ears, the bat's auditory system could rather be specialized for the spectral features that are not identical between the two ears, at least during echolocation.

Another possibility could be that bats can change the weight they give to localization cues in active and passive modes. Since ILD is not effected by the sonar beam a bat could rely more on this cue during echolocation. Especially for CF-FM bats this could be an ideal strategy. Use of ILD cues for azimuth and elevation localization is definitely makes this a likely possibility.

If the beamshape is variable however, sound localization relying on monaural localization cues are going to be problematic for they are subject to change. If the bat's auditory system can somehow monitor the sonar beam changes then localization cues with reference to certain beamshapes can still be available for computation. This is analogous to change in sound localization cues as a result of pinna movements. A proprioceptive input and/or a signal indicating what type of beam will be generated can keep the sound localization process adapted to the changes. On the other hand, if sonar beamshape changes are not integrated to the sound localization computation, localization should rely on cues that are independent of the sonar beam patterns. In this condition, bats could get away with the difficulty by using ILD only localization schemes. However, previous studies suggest that monaural spectral cues are influential in sound localization during echolocation (Lawrence and Simmons, 1982; Wotton and Simmons, 2000) implying bats either have no significant

change in their beamshape or they can adapt to the changes in the sonar emission patterns.

In figure 3.21 we have put potential outcomes of fixed and variable beamshapes for sound localization discussed so far. Depending on controllability and the complexity of the beamshapes, computation of the sound localization has to be implemented differently for passive listening and echolocation to maintain or improve localization accuracy. These outcomes are highlighted with the star symbol in figure 3.21. Only if the beamshape is simple with one lobe, can the sound localization be independent of the mode of listening. However, this can result in less accurate localization due to lowpass filtering effects of the sonar beam.

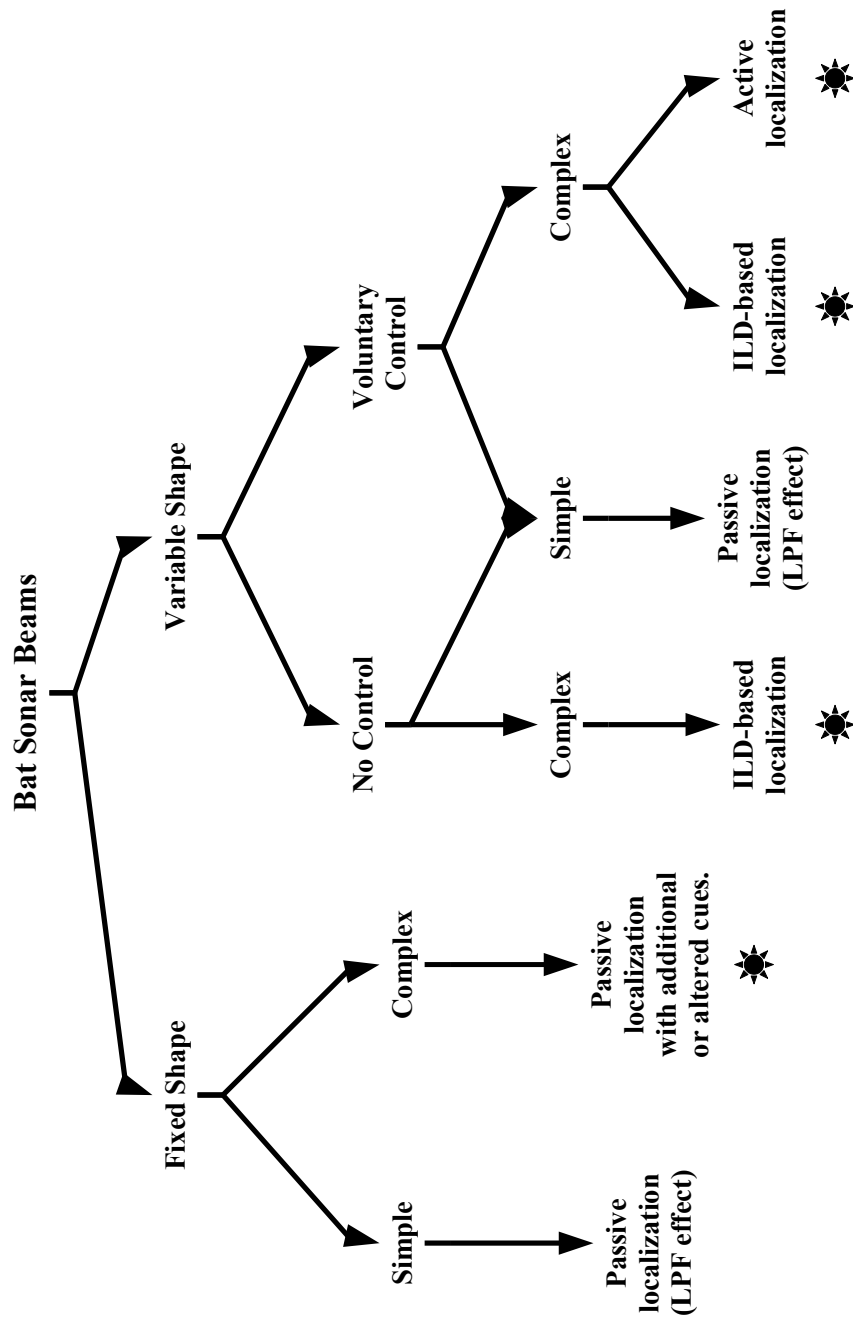


Figure 3.21: Consequences of sonar beamshape variability. Sound localization computation can vary depending on the complexity of the sonar beamshape and, if variable, bats' ability of control it. For most of the cases (emphasized with) bats need to adapt computation of sound location to altered acoustic cues or fall back to ILD-based localization. A beamshape with a single lobe is considered simple and result in computation of sound localization based on the same cues in both passive and active listening (echolocation) conditions. It is expected that during echolocation localization accuracy decreases for the off beam axis locations due to the lowpass filter (LPF) effect.

For all possible outcomes limiting computation of sound localization to ILD cues is an option. This would imply sacrificing from localization accuracy that can be attained under passive listening conditions. The idea of an extra-auditory signal to the bat's auditory system, expected with controllable complex beamshapes, brings up a difficulty with our current thinking of how sound localization is computed. Under the complex conditions involving changing pinna positions and sonar beam patterns, acoustic inputs are not sufficient to compute locations of the echo sources.

Without studying sonar beam patterns for each vocalization it is not possible to determine the relevance of these possibilities, yet the complexity of the sound localization problem with these hypothesized conditions motivates us to pursue more detailed measurements of the bat sonar beam patterns.

3.5 Conclusions and future work

In this study we were able to show that it is possible to utilize techniques that allow monitoring the head position of a freely moving bat in three-dimensions and determine at what relative positions a measurement of the acoustic signal is taken. Furthermore we have successfully shown that having used the proposed measurement system we can obtain beam patterns of sound emitters with constant beam pattern.

What we were unable to do, however, is to obtain sonar beamshapes from bat vocalizations. This failure was related to the need to use spatial data from multiple vocalizations to build the beam patterns. This could be a result of natural variation in the sonar beamshapes and/or variation created by the experimental conditions. Whatever the causes there may

be a solution that could circumvent them by increasing the number of microphones and reconstructing the sonar beamshapes for individual sonar vocalizations.

The difficulty with the proposed solution is related to the practicality of recording with a large number of microphones simultaneously. If however, signals at the output of the super-heterodyne process, could be further decreased in bandwidth by envelope detection a larger number of channels can be supported by one data acquisition (DAQ) system, while keeping the cost of the system in reasonable range.

Further improvements can be made with the head position monitoring with miniature gyroscopes and accelerometers. Gyroscopes and accelerometers can be used in determining orientation of a bat's head and body. With the improvements in technology of micro-electro-mechanical systems (MEMS) these type of sensors are becoming smaller in size and lighter in weight to be placed on a bat's head. With the addition miniature wireless data transmission module that can be carried by a bat in flight and a microphone array with numerous microphones, sonar beamshapes can be monitored during flight. Using this methodology utilization of sonar emissions by bats while they are engaging in to sonar guided obstacle avoidance and prey capture behaviors can be investigated.

Paper Title A sensory motor approach to sound localization

Authors Murat Aytekin, Cynthia F. Moss and Jonathan Z. Simon

Journal Submitted to Neural Computation. (2007)

Chapter 4

A sensorimotor approach to sound localization

**We can't solve problems by using the same kind of thinking we
used when we created them.**

Albert Einstein

Sound localization is known to be a complex phenomenon, combining multisensory information processing, experience-dependent plasticity and movement. Here we present a sensorimotor model that addresses the question of how an organism could learn to localize sound sources without any *a priori* neural representation of its head related transfer function (HRTF) or prior experience with auditory spatial information. We demonstrate quantitatively that the experience of the sensory consequences of its voluntary motor actions allows an organism to learn the spatial location of any sound source. Using examples from humans and echolocating bats, our model shows that a naive organism can learn the auditory space based solely on acoustic inputs and their relation to motor states.

4.1 Introduction

It is largely accepted that the relative position of a sound source is determined by binaural acoustical cues, e.g. interaural level and time differences (ILD and ITD) and monaural spectral features, embedded in the acoustic signals received at the ears. Recent advances in our understanding of sound localization, however, suggest that sound localization is not purely an acoustic phenomenon, an inherent assumption of any solely cue-based model. Studies report that aside from the acoustic information received at the ears, input to other sensory modalities can also affect a subject's estimate of sound position in space. Vision, for instance, can influence and guide calibration of sound localization (Knudsen and Knudsen, 1985; Knudsen and Knudsen, 1989). Proprioceptive senses, e.g. position and the motion of the head, as well as the perceived direction of gravitational forces, and gaze direction (Lewald and Ehrenstein, 1998; Goossens, 1999; Lewald, Dörrscheidt, and Ehrenstein, 2000; Lewald and Karnath, 2000; DiZio et al., 2001; Getzmann, 2002; Prieur et al., 2005; Sparks, 2005), also play essential roles in sound localization.

There is strong evidence that normal development and maintenance of the ability to localize sound requires auditory experience (Knudsen, 1982; Wilmington, Gray, and Jahrsdoerfer, 1994; King, Parsons, and Moore, 2000), a finding also proposed for the visiomotor system (Held and Hein, 1963; Hein, Held, and Gower, 1970). The auditory system has the capability to adapt to changes that occur in the directional characteristics of the external ears during development (Moore and Irvine, 1979; Clifton et al., 1988) and in adulthood (Hofman, Van Riswick, and Van Opstal, 1998; Van Wanrooij and Van Opstal, 2005; Kacelnik et al., 2006). Subjects can adapt to artificial changes to specific sound localiza-

tion cues (Held, 1955; Loomis, Hebert, and Cicinelli, 1990; Javer and Schwarz, 1995) and changes introduced to the information content of acoustic input, e.g. blind infants can learn to use a sonar device to reach objects in their environments (Bower, 1989). The notion that the auditory system has the capability to learn to localize sound suggests that it should not rely on an innate, pre-existing representation of how to interpret spatial acoustic cues. This line of evidence further suggests that the auditory system is plastic enough to acquire the spatial information, and perform the computations that are needed to determine it, through experience.

Here we demonstrate an approach that aims to provide a comprehensive computational scheme that incorporates the adaptive and multi-sensory nature of the computation of spatial acoustic information by the nervous system. The approach addresses the question of how a naive nervous system might develop the ability to compute the spatial locations of sound sources. This approach to the sound localization problem is complementary to, but fundamentally different from, those that rely solely on acoustic cues for sound localization. Most acoustic cue-based approaches are limited to specific computational strategies that use exclusively the head related transfer functions (HRTFs), i.e. the direction-specific acoustic filtering by the pinnae and the head (Colburn and Kulkarni, 2005). Standard modeling approaches rely on acoustic information alone, and so cannot explain the effects of motor actions and other sensory modalities on the computation. Moreover, unlike the proposed model, by taking the outside observer's point of view they are not concerned with how the auditory system acquires the knowledge of the spatial coordinates to utilize these acoustic cues. We propose a sensorimotor approach (Poincare, 1929; O'Regan and Noë, 2001; Philipona, O'Regan, and Nadal, 2003; Philipona et al., 2004) to the general problem of

sound localization, with an emphasis on questions of development and learning that allow spatial information to be acquired and refined by a mobile agent. Thus, the sensorimotor approach does not require *a priori* representation of space by the system.

4.1.1 The sensorimotor approach

How can a naive animal acquire the sense of space, i.e. how does the brain acquire enough information about its surroundings to localize a sound source? If experience is necessary for the development of spatial hearing, i.e. sound localization, one might infer that this also involves learning to interpret spatial information embedded in the acoustic signals it receives. This would require learned associations between certain features of the acoustic input received at the ears and the spatial information carried by sounds. Although behavioral evidence suggests that a reflexive orientation to sound sources by neonatal animals and human infants is hardwired (Kelly and Potash, 1986; Muir, Clifton, and Clarkson, 1989), this reflex disappears and is later replaced by spatial cognition that develops through experience (Clifton, 1992; Campos et al., 2000). The purpose of this early orientation reflex, controlled by the lower brain stem, may then provide an initial state for this learning process (Metta, 2000; Muir and Hains, 2004). Without the knowledge about associations between acoustic cues and spatial information, a naive nervous system would be unable to interpret spatial characteristics of the acoustic inputs (Poincare, 1929; Philipona, O'Regan, and Nadal, 2003).

What gives rise to auditory spatial perception? Poincare and recently Fermuller and Aloimonos (1994) and O'Regan and Noë (2001) argue that if an organism has no means

to generate movements and sense them (proprioception) the perception of space cannot develop. With voluntary movements, the nervous system learns sensorimotor contingencies (O'Regan and Noë, 2001), which, in turn, reveal the spatial properties of the acoustic inputs. To clarify this point we adapt an example given by Poincare for visual perception of space. Consider two sound sources located at a given position with reference to a listener's head at different times. Assume that both sounds are displaced 30° to the right with reference to the listener. These relative displacements will induce different acoustic inputs at the subject's ears. Despite the differences in the acoustic sensations at these two instances in time, the nervous system will interpret them both as sound sources that underwent the same spatial displacement. There is something in common between the two sets of acoustic inputs that allows the brain to calculate the change in spatial position of the two sound sources. The acoustic changes associated with the 30° displacement of each sound source are different, but the argument is that these changes are interpreted by the auditory system as comparable because they can be compensated by the same set of motor actions. By compensation it is implied that sensory inputs are brought back to their initial conditions, e.g. 30° head movement to the right to recover the initial positions of the sound sources. As a realization of Poincare's insight, it follows that in order to perceive the space acoustically, self-generated movements are required. These movements sample acoustic space by changing the positions of the ears, and, crucially, by conveying the corresponding proprioceptive sensation.

Acoustic inputs received at the ears vary with the dynamics of a sound source (external changes) and according to the motor state of the body (internal changes). External changes may be due to non-stationary acoustic properties or a spatial displacement of a

sound source. External and internal changes can be distinguished by the nervous system, since only internal changes are also sensed proprioceptively. Note that spatial external changes can be mimicked by certain motor actions that move the ears and head rigidly (rotation and translation). However, this cannot be used to mimic external changes due to dynamic acoustic properties of the sound source. Stated another way, spatial-external changes can be compensated by a certain set of changes imposed on internal states by motor actions. It is this notion of compensation that leads to the commonality between the internal and external changes, which then give rise to the entity of space encoded by the nervous system (Poincare, 1929; Vuillemin, 1972; Philipona, O'Regan, and Nadal, 2003). Thus, the compensation is a direct consequence of the organism-environment interaction. The brain can distinguish the body and its exterior and, using this dichotomy, can learn and explore space through observation of the body's actions and the resulting sensory consequences. Sensorimotor laws (contingencies) should be invariant under changes in the way the acoustic information is encoded by the nervous system and changes in the sensor structure, provided that these changes do not hinder the notion of compensation (Philipona, O'Regan, and Nadal, 2003; Philipona et al., 2004). In other words, spatial displacements of sound sources can be compensated by the same physical displacements of the sensors, regardless of the shape of the ear, head or body, or the way the information is represented in the central nervous system. Changes in the sound-source properties and of body state can be thought of as transformations acting on the acoustic environment-body system. Explorations of those transformations that lead to compensable actions yield the mathematical foundation that is necessary to formalize a mechanistic way to explore the acoustic space.

4.1.2 Sound localization and sensorimotor approach

We envision the problem of learning auditory space as the organism's acquisition of the spatial relations between the points that can be occupied by sound sources in space. Our goal here is to show that using movements of the listener, i.e. compensable motor actions, it is possible to identify points in space and represent them with coordinates, and hence their geometric relations in the auditory space. The learning of the auditory space could then allow the system to determine acoustic and/or audio-motor features associated with a particular point in space, i.e. sound localization. In order to achieve this task we make three assumptions. First, we assume that an organism can distinguish the differences between the exteroceptive and proprioceptive inputs. Based on the classification and separability of sensory inputs in these two groups, an organism can identify the dimension of space through its interaction with the environment (Poincare, 2001; Philipona, O'Regan, and Nadal, 2003). Thus, we also assume that the dimension of auditory space is known. Since only two parameters are necessary to identify a sound-source location in auditory space in the far field, auditory space is assumed to be two-dimensional. Lastly, we assume that the organism can distinguish the motor actions that can induce spatial displacement (spatial motor actions). Philipona, O'Regan, and Nadal (2003) demonstrate how a naive system can identify these special movements.

Our initial method will be limited to the learning of points and their spatial neighborhood relations in auditory space by the organism, without concern for the detailed metric properties. This approach is substantially simpler in its goals (and weaker in accomplishments) than that applied by Philipona et al. (2004) in the visual system. For our purposes

we need a way to identify points in the space from the organism's perspective.

4.1.3 Manifold of spatial acoustic changes

One can view the set of all possible states of a sound source, its time, frequency and spatial properties as a manifold, E , a topological space that is locally diffeomorphic to a linear space. Similarly, all the motor states (including, e.g., head position) can be thought of as elements of a manifold, M and all the acoustic sensory inputs received at the ears also constitute a manifold, S . Since the acoustic sensory input is determined by both the sound source and the current motor state, there is a functional relationship between the manifolds E , M and S .

$$S = f(E, M)$$

If an organism makes all the possible head movements when the state of the sound sources is e_o , the resulting set of sensory inputs, $f(e_o, U)$, will be an embedding, where U represents a subset of motor states in M . For sound sources in the far-field these embeddings will be two-dimensional, since the relative position change of a sound source caused by head movements can be represented with two parameters. These embeddings can be referred as the *orbits* of a sound-source state. Sensory inputs corresponding to the different sound sources with the same relative spatial positions, will lie on different orbits: two sound sources with different spectra will result in different sets of acoustic inputs during similar relative head motions. The orbits are smooth surfaces, i.e. $f(e_i, U)$ and $f(e_j, U)$ where e_i and e_j represent two different sound sources, consist of sets of sensory inputs that

correspond to the environment-organism states. There can be as many orbits as there are different sound sources. As discussed earlier, the entity of space emerges as a result of observations that identical displacements of any two sound sources can be compensated by the same motor outputs. Exteroceptive and proprioceptive sensory inputs associated with the same displacement, however, could be different for different motor states in general.

Exteroceptive sensory changes that are compensated for by a particular motor action, e.g. a head rotation, may not be unique to particular spatial locations, and so alone are insufficient to give rise to the concept of a point in space. Yet listeners, when they localize the spatial position of a sound source, do perceive it as a unique spatial point. Thus, there must be something in common (an invariant property of a spatial point) among these different sensory inputs so that the organism can identify their spatial parameters as characterizing a single point. From the organism's point of view, the locations in space of two potential sound sources' positions would be identical if both could be brought to a reference sensorimotor state with the same set of motor actions, e.g. head movements. The set of potentially identifiable points (by movements of a head with immobile ears) results in a two-dimensional representation of space, since there are only two independent parameters, e.g. azimuth and elevation, available to identify the direction of a sound source in space, with reference to the body.

If there were only a single orbit in the sensory input space (only one kind of sound), the solution to the auditory space learning problem would have been described as fitting a global coordinate system to this surface with two parameters. In this case each sound-source location could have been represented by the global parameters of its corresponding point on the orbit, and any arbitrary point on this surface could have been picked as the

reference sensory input or the origin of the global coordinate system.

The practical case of multiple sound sources, i.e. multiple orbits, however, poses a more challenging problem in terms of finding the parameters that represent the spatial positions of the sound sources. We can determine each orbit and fit a global coordinate system to each one of them, but as a result we would obtain as many global coordinate systems as the number of orbits in the sensory input space. A critical task then is, for each spatial source position, to find the points on the orbits that correspond to that same spatial position and assign a single parametric representation. In general, there might not be additional information available to register these separate global coordinate systems with reference to each other. An extra sensory input, e.g. vision, favoring a particular set of organism-environment states (W, m_o) such that W represent the subset of sound-source states with the same position (0° azimuth and 0° elevation for instance), could be sufficient as a reference. We argue, however, that under circumstances special to the auditory system, such an extra input is not necessary. In fact, we show below that the directional acoustic properties of the external ear make it possible to solve the localization problem without the need for another exteroceptive sense.

Directional properties of the acoustic inputs at the ears

Directional features of the acoustic signal received at the ears can be completely captured by linear-time-invariant (LTI) system models of the transformation of the sound caused by its interaction with the head and the pinnae. The transfer functions of these LTI models are the HRTFs (Blauert, 1997). Acoustic inputs received at the ears caused by a sound source in space can be mathematically depicted as follows:

$$S_{left}(f) = A(f) \cdot H_{left}(f, \theta, \phi)$$

$$S_{right}(f) = A(f) \cdot H_{right}(f, \theta, \phi)$$

The H_i s represent the HRTF of left and the right ears for a given source at position, θ azimuth and ϕ elevation, respectively, for frequency f . $A(f)$ is the sound source's frequency representation (spectrum), and the $S_i(f)$ results are the spectra of the acoustic inputs as measured at the left and the right ears. Now we will show that the orbits that can be obtained from such a sensory system have common features (invariants) for the sensory inputs that are generated by the sound sources with the same relative positions. The existence of such features allows a definition for a point in space.

4.1.4 Computation of the space coordinate system

Head movements induce changes to the acoustic inputs produced by a sound source located at a fixed point in space. These changes can be thought of as transformations acting on the points on the orbit of this sound source. For the sake of simplicity we assume that head movements are fast enough that during its motion there is no change in the state of the acoustic environment, and we assume that the head motion occurs only when there is a single sound source. Without loss of generalization (though see discussion below) we may limit the organism's spatial motor actions to infinitesimal head movements around a fixed head orientation (0° azimuth and 0° elevation for instance) when a measurement from the environment is taken. Each measurement is assumed independent of all others. Each head movement induces a vector tangent to a point, $f(e_i, m_o)$, on the sensory input manifold, S , representing the change in the sensory input. Each tangent vector is attached to the

observed sensory input before the head movement.

In order for the nervous system to represent spatial points independently of their sound source, based on audio-motor contingencies, there must be a unique, i.e. sound-source-invariant, feature of spatial displacements that is present for all sets of exteroceptive sensory input changes associated with particular motor state change. Note that each member of such a set arises from one point on a sound-source orbit. Only by utilizing source-invariant features is it possible to bring points on different orbits associated with the same relative spatial position into a single representation of that point in space.

How this feature manifests itself depends on the neural representation of the acoustic inputs. For instance, an organism that encodes changes in acoustic energy in logarithmic units (dB) could realize that some changes in the sensory inputs associated with its head movements are identical. In the case of acoustic energy differences encoded linearly in amplitude, a head movement would generate external changes that are linear functions of the acoustic inputs. These linear relationships would be constant for sound sources at the same relative position, (e.g $\Delta S = S(f) \frac{\partial H(\theta, \phi, f)}{\partial \theta}$), which would qualify as the necessary sound-source-invariant feature. Thus, different coding schemes require different computational solutions for the system to identify the invariance property, some of which might be easier to implement by the nervous system. For each coding scheme an appropriate metric on the neural signals is needed to determine similarity of the neural representations of any two acoustic signals. In pure logarithmic coding, this metric takes its simplest form, in terms of the level of computation required, since similar inputs result in similar changes of the neural signals. But the specific neural coding of acoustic sensory input employed is not critical: it can be shown that under any sufficiently complete neural coding scheme a

diffeomorphism¹ exists between the submanifolds of E and S .

In order to obtain a full coordinate representation of auditory space, no two sound-source-invariant features can be identical. Failure of this uniqueness would naturally result in (well-established) psychoacoustical spatial ambiguities.

From sound-source-invariant features to coordinate representation of spatial locations

Consider the sensory input changes caused by a set of infinitesimal head movements obtained for a given sound-source location. If we assume logarithmic coding for the sake of simplicity, the tangent vectors on the orbits obtained by the same head movements for different sound sources will be identical, i.e. a sound-source-invariant feature, and will change smoothly with sound-source location. If we further assume that the set of tangent vectors from a spatial location is unique to that location, then the set of tangent vector sets constitutes a two-dimensional manifold: the manifold of spatial points. Since this manifold is independent of the sound source's internal characteristics, and dependent only on the relative position of the sound sources, one may then attempt to assign to it a global coordinate system, corresponding to the relative sound locations. The manifold can be represented as a two-dimensional embedding in an N -dimensional space, where N is the number of frequencies discriminable by the auditory system, and a manifold learning algorithm can be employed to obtain a coordinate system.

¹A diffeomorphism is a map between manifolds which is differentiable and has a differentiable inverse.

4.2 Demonstration 1: Obtaining spatial coordinates in humans

We implemented the proposed computational scheme for simplified human subjects. We show that by using voluntary head movements and observing their sensory consequences, it is possible to compute the (two-dimensional) global parameters that represent the direction of the far field sound sources in space. We use 45 human HRTFs from publicly available CIPIC database (Algazi et al., 2001), in magnitude only (i.e. no ITD information is used). The simulated humans take measurements from the environment by making small (1°) head movements from a resting state where the nose points at 0° azimuth and 0° elevation. For each measurement the head makes the following three head movements; rightward, upward, and tilting downward to the right. Each head movement results in a vector that is on the tangent space at $f(e_i, m_o)$, the sensory input caused by a particular sound source. Sensory inputs are assumed to be represented in logarithmic (dB) coding. An extended-tangent vector is then produced for each measured sound source by concatenating the three measured tangent vectors. Each tangent vector is sound-source invariant and the set of them, extended-tangent vector, qualifies as a location-dependent feature, and the set of location-dependent features constitutes a manifold isomorphic to auditory space. The particular choice of head movements is not important as long as they generate a set of independent vector fields on the sensory input space. Any arbitrary vector field caused by an arbitrary head movement can be written as the linear combination of any independent vector field set. The number of independent vectors is limited by the dimension of the manifold. We have chosen three head movements to guarantee that at each spatial point there are at least two

independent tangent vectors that can be obtained. For example, at the north pole right to left turns of the head cannot introduce any external changes but the other two head movements can still generate two independent tangent vectors.

4.2.1 Simulation of the sensory inputs

Simulated sound sources are represented as vectors that are comprised of the Fourier Transform magnitudes of sounds in the frequency interval of 1 kHz to 20 kHz, in steps of 200 Hz, resulting in 95-dimensional real vectors. These vectors are generated from a uniform random process and have units of dB. The HRTF phase information, normally available to humans up to 3 kHz (Stevens and Newman, 1936), is ignored only to simplify the analysis.

We have selected our source positions on the hemisphere that extends between two polar points located at -45° elevation, 0° azimuth and 45° elevation, 180° azimuth (elevations below -45° are not available for CIPIC data). Each sound spectrum is filtered with their corresponding HRTF to compute the acoustic signals at the ear canal (represented as 190-dimensional vectors). The HRTFs are smoothed spectrally using a Gaussian smoothing filter with a constant quality factor ($Q = 24$) and interpolated spatially using spherical harmonics (Evans, Angus, and Tew, 1998). The spatial positions of the sound sources are chosen using uniform spiral sampling (Saff and Kuijlaars, 1997) with the spiral axis extending from one ear to the other (total of 1748 points). This sampling approaches a uniform distribution on the sphere when the number of points approaches infinity. The tangent vectors are computed as the difference of the spectra of the simulated sound at the ears before and after the three head movements. Later these vectors are concatenated to build

a 570-dimensional extended-tangent vector. Each extended-tangent vector is used to represent a spatial point in the space since they are independent of the sound-source spectrum under the logarithmic coding. If the extended-vectors are unique for each sound-source location in (far-field) space then they should lie on a two-dimensional embedding (manifold) in 570-dimensional space.

4.2.2 Determining the spatial parameters

Assuming the manifold of extended-tangent vectors exists, we can use a manifold learning algorithm, e.g adaptive-LTSA (local tangent space alignment), to assign global parameters to its points. There is a one-to-one association between these parameters and the sound-source locations. Adaptive-LTSA is a nonlinear manifold learning algorithm with three steps. First, using an adaptive nearest-neighbors method (Wang, Zhang, and Zha, 2005) the algorithm finds the local neighborhood around each point. Then for every point local coordinates of its neighboring points on the local tangent space are computed. In the third step these local coordinates are aligned to construct a low dimensional global representation of the points on the manifold.

The algorithm can capture local isometry only if the manifold is locally isometric to a Euclidean parameter space, but a hemisphere is not: one cannot flatten a hemisphere to a plane without distorting the pairwise distances between the points. Thus if the algorithm can successfully capture the topological relations, the pairwise distances between any two points in the maps will be distorted. In addition, like other well known spectral manifold learning methods (e.g. locally linear embedding (Roweis and Saul, 2000), Laplacian eigen-

maps (Belkin and Niyogi, 2003)), LTSA (Saul et al., 2006) cannot guarantee the recovery of the global parameters obtained from the alignment step (Zha and Zhang, 2005). To remedy the recovery problem we implement the conformal component analysis (CCA) method proposed by Sha and Saul (2005) on the results obtained from LTSA. CCA computes the low dimensional embedding by trying to preserve the angles in between the extended-tangent vector as a result of minimizing total local dissimilarity, and hence generates a conformal map of the points on the manifold. CCA can also be used to predict intrinsic dimension of the locally isometric manifolds. Local dissimilarity, $D_i(s_i)$, is a measure based on the similarity (similarity of the corresponding angles) of the triangles in the neighborhood of an extended-tangent vector, $\{x_i\}$, and its image, $\{z_i\}$, in the low dimensional embedding obtained via manifold learning methods.

$$D_i(s_i) = \sum_{jj'} \eta_{ij} \eta_{ij'} \left(\|z_j - z_{j'}\|^2 - s_i \|x_j - x_{j'}\|^2 \right)^2$$

Here x_i and z_i represents the extended-tangent vectors and their low dimensional images respectively. $\eta_{ij} = 1$ if x_j is a member of x_i 's neighborhood and $\eta_{ij} = 0$ otherwise. s_i is the constant representing the scaling as a result of the conformal mapping at the neighborhood of x_i . The total local dissimilarity is obtained as $\sum_i D_i(s_i)$. This measure allows one to quantify and compare the two-dimensional embeddings obtained from this method using local dissimilarity measure (Sha and Saul, 2005). Although, this manifold learning method is not necessarily the best suited for obtaining the metric structure on the manifold, embeddings obtained from LTSA are sufficient to show that the neighborhood relationships of spatial points can be learned using auditory-motor contingencies.

In Figure (4.1) total local dissimilarity values are given for each subject. Forty three out of 45 subjects gave similar dissimilarity values. The manifold learning method failed to give reliable low dimensional embeddings for the remaining two subjects (subject 008 and subject 126) with the highest values of total local dissimilarity. Examples of global parameter maps are also provided in Figure (4.1). These maps correspond to subjects with 15th (subject 154), 30th (subject 050) and 43rd (subject 131) highest dissimilarity values. Points on these maps are color coded to indicate the corresponding azimuth and elevation values of the sound-source locations. The stereographic projection of the spatial distribution of the sampled azimuth and elevation values can be seen in the insets on the left corner of the figure.

An ideal map of the global parameters of the sound-source locations should maintain the pairwise neighborhood relationship between the positions of the sound sources, hence showing a smooth change in relation to their spatial parameters. Notice that the relationship between the sound-source directions in space is preserved in the global parameter maps. For many subjects the two-dimensional extended-tangent manifold requires three-dimensions from the CCA results to globally describe the parameter space. This is expected since the manifolds studied here are not locally isometric. In these cases the points in the lower dimensional representation of the manifold manifest as a two-dimensional surface embedded in three-dimensional coordinate system. For the global parameter maps given in Figure (4.1) a two-dimensional projection was possible. We found that the density of the points in the global parameter maps increases at and around the north pole for all subjects. With increasing dissimilarity, value deformations are observed in the same region (subjects 050 and 131). The overall shapes of the global parameter maps, however, are not partic-

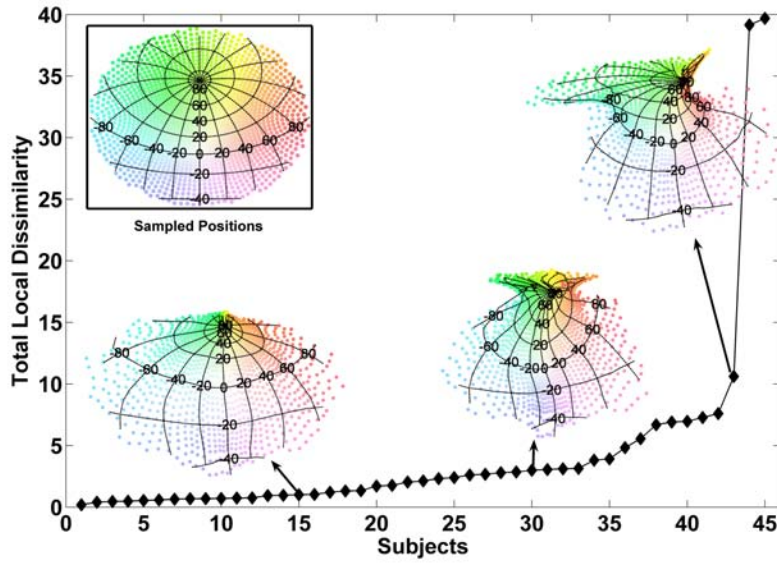


Figure 4.1: Total local dissimilarity for each subject. Spatial points are sampled from the hemisphere of interest in stereographic projection depicted in the inset. Positions are color coded to allow comparisons with the global parameter maps obtained from the manifold learning method. Global maps of three subjects corresponding to 15th (subject 154), 30th (subject 050) and 43rd (subject 147) largest total local dissimilarity values are also shown (arrows). Manifold learning step failed to capture geometric organization of the spatial points for the subjects with two outlying local dissimilarity values, the (subject 008) and (subject 126).

ularly important as long as a unique pair of parameters is obtained for each sound-source direction (though it is interesting that the two dimensions correspond roughly to the directions of elevation and azimuth). Smoothness of the global parameters provides the ability to interpolate, i.e. to predict any inexperienced sound-source direction on this hemisphere.

Deformations are always observed at high elevations ($> 60^\circ$). We have investigated potential reasons for these effects. Accurate determination of the local tangent space is dependent on the local curvature, local sampling density, noise level and the regularity of the Jacobian matrix (Zhang and Zha, 2004). The estimated pairwise distances between any two

neighbor points are less accurate in the high curvature regions and thus, uniform sampling of the manifold can introduce a bias in the alignment of the local coordinate system when a global minimization scheme is employed. The process of global alignment of the local coordinates uses a weighting procedure inversely proportional to the local curvature estimates which minimizes any potential bias that could be introduced by the high curvatures on the manifold (Wang, Zhang, and Zha, 2005). The local curvature values estimated by LTSA, as well as the condition of the Jacobian matrix, were examined in detail.

Spatial distribution of the local curvatures showed higher curvature values at and around the north pole, coinciding with the previously mentioned high density regions. Mean curvature value above 60° elevation across all subjects is found to be 1.07 ± 0.27 and 0.51 ± 0.04 below this elevation. Although these curvature estimations are only approximations it is clear that at higher elevations extended-tangent vector manifolds show higher local curvature values.

The Jacobian matrix represents the best linear approximation to the local tangent space at a given extended-tangent vector. Since the manifold of the extended-tangent vectors should be two-dimensional, the rank of the Jacobian matrix should be equal to two. Non-uniformity of the local tangent space dimensions on the manifold can result in non-optimum solutions by the LTSA (Zhang and Zha, 2004). In order to investigate the condition of the Jacobian matrix we compared the singular values of the matrices comprised of the local coordinates of the points in each neighborhood obtained in the first step of the manifold learning algorithm. These matrices and the corresponding Jacobian matrix should span the same local linear subspace and thus have the same rank. Singular values of a (noisy) local coordinate matrix are expected to have two large values and $K - 2$ smaller values for a

two-dimensional local linear space. Here K represents the number of points in the neighborhood. We investigated singular values obtained at each point's neighborhood for subjects with the smallest (subject 147) and largest (subject 131) dissimilarity values (Figure (4.2)). In Figures (4.2(a)) and (4.2(b)) we have shown the ratios of the singular values $\frac{\sigma_2}{\sigma_1}$ and $\frac{\sigma_3}{\sigma_2}$, depicted as '+' and '•', for each neighborhood, where σ_i is the i^{th} largest singular value ($\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_K$). For both subjects the majority of the points show well separated singular value ratios, confirming that first two singular values, σ_1 and σ_2 , are comparable and larger than the rest of the singular values. However, this separation for subject 131 is not as robust as it is in the subject 147, particularly for the points corresponding to the most peripheral left and right sound-source locations, as well as points near the north pole. Closer inspection reveals that σ_1 is moderately larger ($\frac{\sigma_2}{\sigma_1} < 0.7$, $\frac{\sigma_3}{\sigma_2} > 0.5$) than the rest of the singular values, suggesting that the dimension of the local linear spaces is less than 2. These points are highlighted in Figures (4.2(a)) and (4.2(b)) by marking corresponding values by 'o' and '◇' respectively. Based on these observations, deformations observed in the global coordinate maps could be caused by the irregularity of the local tangent space dimensions in the extended-tangent vector set.

In addition to the condition of the Jacobian matrix and local curvatures, local geodesic distances can also be influential on the global parameter maps. Because the global maps are only conformal representations, the local distances between the global parameters are not directly interpretable. Thus, we cannot directly conclude that points corresponding to the higher density region are simply closer to each other (which would imply greater dif-

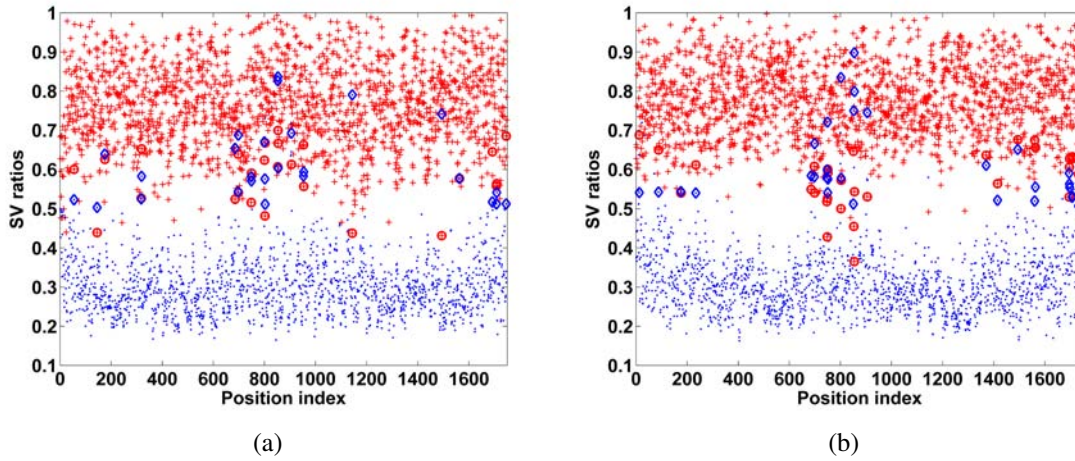


Figure 4.2: Singular value ratios $\frac{\sigma_2}{\sigma_1}$ and $\frac{\sigma_3}{\sigma_2}$, where $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_K$. a: subject 147 (lowest ranked total local dissimilarity) and b: subject 131 (43rd ranked total local dissimilarity). Ratios obtained within the neighborhood of each point (total 1748 points uniformly distributed on the hemisphere) are depicted in '+' and '.' respectively. Moderate separation of the two sets of values implies full rank Jacobian matrix for 2-dimensional local tangent space. The ratios with potential rank deficiency problem are stressed with 'o' (for $\frac{\sigma_2}{\sigma_1}$) and 'o' (for $\frac{\sigma_3}{\sigma_2}$). A local neighborhood is determined as problematic if the largest eigenvalue is moderately larger than the rest of them (if $\frac{\sigma_2}{\sigma_1} < 0.7$, $\frac{\sigma_3}{\sigma_2} > 0.5$).

difficulty in distinguishing source locations near to each other). Assuming that the uniform spatial sampling is dense enough, however, it is possible to compare the local distances of the extended-tangent vectors and determine if this region contains points that are relatively closer to each other (i.e. more difficult to distinguish neighboring source locations). Spatial distributions of the mean local distances of each point to their local neighbors (neighborhoods are obtained using K-nearest neighborhood method where K=8) are given in Figures (4.3(a)) and (4.3(b)). Mean local distances for both subjects show a steady decrease with increasing elevation confirming that at poles the extended-tangent vectors are relatively close to each other. In Figures (4.3(c)) and (4.3(d)) we have given the mean local distance maps of the underlying HRTF for each of the subjects. Notice that for both subjects, local HRTF distances are low above 40° elevation. Minimum local distance values are obtained

at 90° elevation. In order to determine the robustness of these results with increased sampling density we doubled the number of points and repeated the analysis, and obtained quantitatively similar results.

Based on these results we conclude that the high density region in the global coordinate maps for all subjects have not only higher local curvature values but also consist of extended-tangent vectors that are similar to each other. The similarity of the HRTFs corresponding to the same spatial positions suggests that acoustic inputs vary less per degree in these regions. These results predict that subjects should experience more difficulty resolving sound-source locations in these regions (depending on the level of the minimum detectable change in the acoustic input level at different frequencies). Notice that similarity of the HRTF is based on a metric defined on the neural representations of the acoustic inputs which is a result of ear shape and neural coding and not related to the metric associated with the acoustic-space.

4.3 Demonstration 2: Obtaining spatial coordinates by echolocating bats

We also test our computational scheme with a different species, an echolocating bat (*Eptesicus fuscus*), to stress that the sensorimotor approach provides a general approach that can capture spatial relations between points in auditory space for other animals that use a very different frequency range for auditory localization. Echolocating bats produce ultrasound

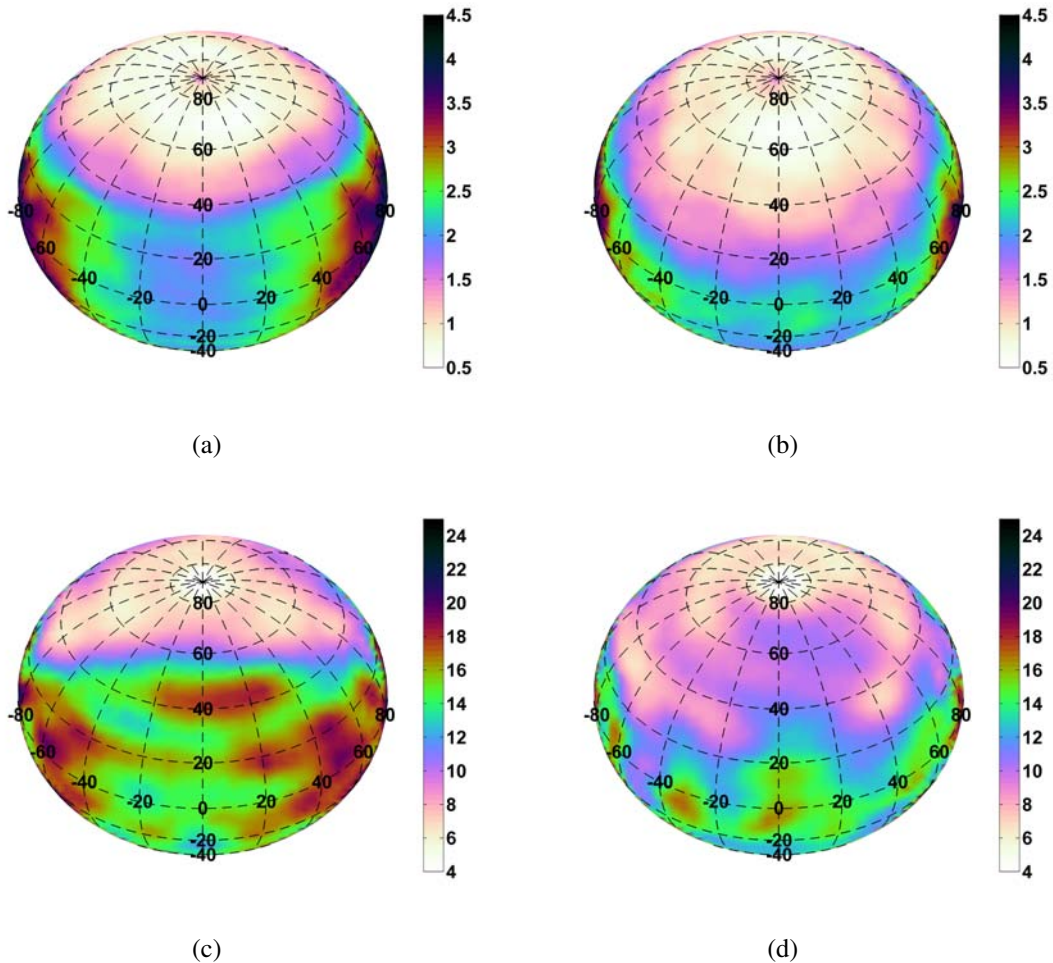


Figure 4.3: a, b: Mean local distances of learned extended-tangent vectors; c, d: Mean local distances of underlying head related transfer functions (HRTFs). Subject 147 in a and c; subject 131 in b and d. Mean distances are determined within each local neighborhood of uniformly distributed 1748 spatial positions on the hemisphere (a and b) with K-nearest neighborhood criteria ($K=8$). The local mean distance for each subject decreases with elevation and reaches its minimum value around the north pole. This property was common across all the subjects. Subject 147 (lowest total local dissimilarity) shows larger local distances below 60° elevation in comparison to subject 131 (moderately high total local dissimilarity) for both type of local distances.

vocalizations and listen to the echoes reflected from the objects around them to monitor their environment. Bats rely on active spatial hearing (biosonar) to detect, track and hunt insects (Griffin, 1958). Thus, accurate sound localization is very important for these animals' survival.

The HRTFs of three bats were measured in the frontal hemisphere (Aytekin et al., 2004). Each HTRF spans 180° azimuth and 180° elevation, centered at 40° , 30° and 22° elevation for subjects EF1, EF2 and EF3 respectively. The HRTFs were measured from 20 kHz to 100 kHz, in 68 logarithmic steps, and smoothed with a rectangular window with a constant quality factor of 20. Phase information (i.e. ITD information) was discarded and magnitude is represented in dB units.

As in the human simulations, the sound-source locations were selected using uniform-spiral sampling. The simulated bat performed the same head movements as the human subjects (in particular, without pinna movement) to generate the extended-tangent vectors. Using adaptive-LTSA algorithm the global parameters for sound-source directions are obtained. In Figure (4.4) global parameters of the echolocating bats (subjects EF2, EF3) are shown. The color coding in the global parameter maps represents the sound-source locations with reference to the head. In the inset of each figure azimuth and elevation coordinates of these sound-source locations are given. For all three global coordinate maps the local relationships between the sound-source directions in the head reference frame are preserved, showing that the sensorimotor approach can be applied successfully to echolocating bats.

The bat global parameter maps (Figures (4.4(a)) and (4.4(b))) also show similar characteristics to the ones we obtained for humans. Sound positions corresponding to high

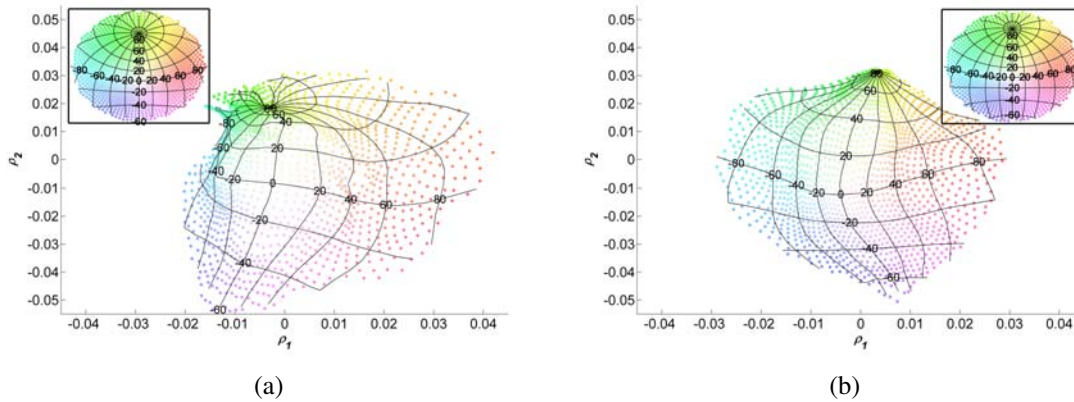


Figure 4.4: Global coordinates of two echolocating bats. a: subject EF2 ; b: subject EF3. Similar to human subjects, global maps show increased density near the north pole. Sampled spatial positions (1707 points uniformly sampled on the hemisphere) are given in the insets of each subfigure. Global parameters obtained from both subjects preserved the topology of the sound-source locations given in the insets of each figure.

elevations show a denser distribution than the mid-elevation regions. Similar to the human, local curvature values for bats are also higher at and around the north pole. Mean local curvatures below 60° elevation were 0.40 ± 0.13 , for subject EF1, 0.42 ± 0.19 for subject EF2 and 0.40 ± 0.10 for subject EF3 and above 60° elevation were, 0.59 ± 0.16 , 0.82 ± 0.37 , and 0.60 ± 0.27 respectively. All these values were lower than those of the human subjects.

Figure (4.5) shows the spatial distribution of the local distances of both the extended-tangent vectors (Figures (4.5(a)) and (4.5(b))) and the underlying HRTFs (Figures (4.5(c)) and (4.5(d))). For both sets of local distance distributions, the spatial areas corresponding to the positions at and around the north pole give the minimum local distances. Notice also that larger distances are observed at the mid-elevation ranges, where the acoustic axes of the HRTFs are observed (Aytekin et al., 2004). Sound locations that give larger local distances generally correspond to the HRTF regions with the largest gain. The direction

and frequency dependent nature of the acoustic axes contribute to the larger distance values in the mid-elevation region. This effect is especially clear for subject EF2's HRTF local distance distribution (Figure (4.5(d))).

Based on these observations we conclude that denser distributions for global parameters should be expected at high elevations. This effect might result in lower perceptual resolution of sound-source locations for the subjects at these sound-source locations.

4.4 Discussion

We have demonstrated that the acquisition of the ability to localize sounds can be achieved based on sensorimotor interactions. Unlike standard sound localization models, our model provides an unsupervised computational process, based on the sensorimotor experience of a mobile listener. The viability of the model is shown via simulations using HRTFs of the human and the echolocating bat. The spatial neighborhood relationships between sound-source locations were successfully learned for both species' HRTFs. We found that the learned global parameters used to identify sound location reflect important features of the HRTF, e.g. relative indistinguishability of the HRTF at high elevations resulted in smaller local distances between the learned global parameters corresponding to those points.

This sensorimotor approach is based on three important assumptions. We first allow the assumption that organisms are initially naive to the spatial properties of sound sources. Secondly, we limit the external sensory information to auditory signals only. Thirdly, we postulate an interaction between the auditory system and organism's motor state, i.e. pro-

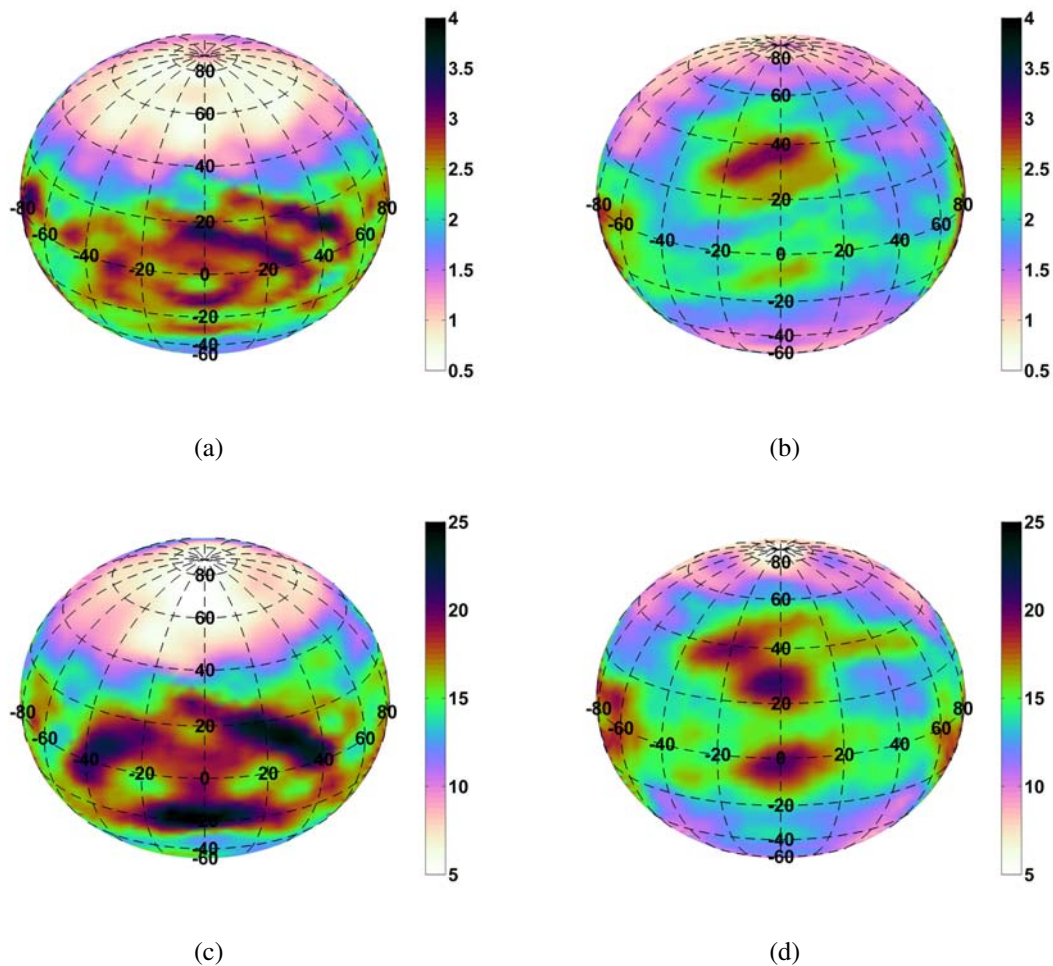


Figure 4.5: Mean local distances for two echolocating bats. a, b: Mean local distances of learned extended-tangent vectors; c, d: Mean local distances of underlying HRTFs. Subject EF2 in a and c; subject EF3 in b and d. Mean distances are determined within each local neighborhood of uniformly distributed 1707 spatial positions on the hemisphere (a and b) with K-nearest neighborhood criteria ($K=8$). The local mean distance for each subject decreases with elevation and reaches its minimum value near the north pole, similar to human subjects, and near the south pole (data for which was unavailable for human subjects).

prioception and motor actions. The first two assumptions may be viewed as a worst case scenario for sound localization, ignoring any potential mechanisms of sound localization that might be hard-wired in the brain (e.g. as a set of initial conditions to be later modified by plasticity). But they do not significantly constrain the approach. The third assumption, in contrast, is crucial to the proposed computational scheme.

With an organism's observations of the sensory consequences of its self-generated motions, there is sufficient information to capture spatial properties of sounds. We stress that the solution provided is not to find a way to match acoustic inputs to corresponding spatial parameters, but rather to show how the animal could learn that acoustic inputs have spatial properties.

4.4.1 Geometry of the auditory space

In this work we have shown that for a given motor state, m_o (0° azimuth and 0° elevation for instance) it could be possible for the nervous system to learn the manifold of sound-source locations using small head movements. However, we have not shown whether the nervous system could identify the manifolds obtained at different motor states, m_k as different images of a single entity, auditory space. In other words, it remains to be determined how the system learns the equivalence of points on two different manifolds, in terms of representing the same relative location in auditory space.

What identifies a spatial point is the unique relationship between the acoustic-input changes and the known set of head movements that generate them. Hence, both sensory and motor inputs are necessary for the learning of the spatial points. In general, identical

head movements at different starting positions, e.g. all the 30° leftward head turns, result in different proprioceptive inputs. Furthermore, acoustic input changes associated with these head movements could also be different at different initial head positions. As a result, spatial correspondence of the points representing sound-source locations on a given manifold, to the points on other manifolds obtained for different internal states is not obvious to the organism.

In order to complete the representation of auditory space, the nervous system must also be able to maintain the identification of the spatial positions of sound sources despite head movements. To illustrate this, take two points, A and B , on the manifold at m_o , where the location corresponding to B is 1° to the left of the spatial location corresponding to A . A 1° leftward head movement maps these points, to \hat{A} and \hat{B} respectively, on a second manifold at a new motor state m_k . We have not yet addressed how the organism might establish the equality of the identical spatial locations corresponding to B and \hat{A} , since they are on different manifolds. This can be accomplished if two conditions are fulfilled. First, the organism needs to be able to generate the same head movements at any motor state. Then the nervous system may identify a mapping between two different manifolds via that movement (e.g. the movement that changes the motor state from m_o to m_k), allowing the pairs A and \hat{A} , and, B and \hat{B} , to be compared to each other. But this knowledge is not sufficient to conclude the equality of the B and \hat{A} .

The second condition is that the sound-source-invariant features must be independent of motor state changes for at least one subset of motor actions, i.e. rigid movements. These movements should leave the related compensable external sensory input changes invariant. For instance, movements that rotate the body without changing relative positions of its

parts would protect these invariant features. Similarly, head movements also qualify if their effect on HRTF is negligible. (As a counter-example, pinna movements change the HRTF (Rice et al., 1992; Young, Rice, and Tong, 1996), and hence the extended-tangent vectors, giving different sound-source-invariant features at different pinna orientations.)

With these two conditions it is then possible to unify the manifolds obtained at different motor states as one entity, auditory space. The same head movements at different motor states should compensate the same external changes or transformations of points on the learned representation of space. An equivalence class of different proprioceptive inputs with identical spatial displacements can be labeled as transformations, e.g. 1° leftward head movement, that results in spatial displacement of the sound-source location. In other words, the organism can now generate similar head movements at different motor states voluntarily. Vuillemin (1972) argues that construction of these transformations require continuity of space as an idealization. Space continuity implies that a spatial displacement of sound-source location caused by organism's motion can be iterated infinitely many times.

In addition, since the organism could establish the equality of the spatial locations corresponding to \hat{A} and B , it is now possible to study the metric properties of auditory space. When the organism has a unified representation of relative spatial points in auditory space through rigid movements, the distance between any two points on the unified representation of space has a spatial meaning. The equality of the distance of any two points can be established if the transformations realized by the rigid head movements are commutative, meaning the order in which any two transformation is applied does not alter the mapping. A subgroup of rigid movements that commute will result in metric preserving mappings of

points in auditory space, i.e. an isometry. The question of how an organism can learn these spatial movements has recently been addressed by Philipona et al. (2004) (though only for the local group properties), who provide a mathematical foundation to study the metric properties of sensory space from the perspective of the organism based on sensorimotor contingencies.

On the neural representations of the invariant features

We have shown that in order for the organism to identify spatial points it has to access the auditory-motor features that are independent of sound-source spectrum. The nature of these invariant features is dependent on the neural representations of the acoustic inputs. Logarithmic coding makes them readily identifiable, since in an ideal logarithmic coding scheme the spatial displacements result in equal changes in the representations, independent of the sound spectra. Although there is an approximate linear relationship between spectral logarithmic magnitude changes and an auditory nerve discharge rates, a complete representation of the magnitude of the sound spectra requires combining the rate information from different sets of auditory nerve fibers, each of whose dynamic range is limited to different sound intensity levels (May and Huang, 1997). Hence, same amount of spectral magnitude change of an acoustic signal at different intensities may not be represented in similar ways by the auditory nerve responses. This complicates the computation of invariant features, for it cannot simply be determined by the identical neural response changes.

At this point we do not have a firm idea on how these invariant features might be determined empirically by an organism. We might hypothesize the existence of a neural mechanism that processes the relationships between the relevant movements through pro-

proprioception or efferent motor command signals and their sensory consequences, and hence represent the invariant features. Evidence for the existence of neural processes of this kind, also known as internal models, has been provided in the cerebellum (Imamizu et al., 2003; Kawato et al., 2003). Internal models are thought to be involved in estimating sensory consequences of motor actions (Blakemore, Frith, and Wolpert, 2001). A cerebellum-like circuitry has also been shown, for instance, in the dorsal cochlear nucleus (DCN), a low level auditory processing area thought to be involved in sound localization (Oertel and Young, 2004). The DCN receives somatosensory, proprioceptive signals in addition to auditory inputs (see references in (Oertel and Young, 2004)), all necessary components of an internal model. Oertel and Young (2004) have, for instance, proposed that cerebellum-like architecture in DCN could function to correct sound localization cues in relation to head and pinna movements in a similar way. If internal-model-like structures do exist in the auditory system, these structures could capture the spatial invariant properties. For now, however, these possibilities remain as hypotheses.

4.4.2 Role of sensorimotor experience in auditory spatial perception

Evidence is accumulating for the importance of sensory experience and the role of voluntary movements on the development of the exteroceptive senses, such as vision and hearing (Grubb and Thompson, 2004; Wexler and van Boxtel, 2005). Experiments on vision and hearing show that active movement is essential for the development of sensory perception and to adapt to changes that might occur after its acquisition (Held, 1955; Held and Hein, 1963; Hein, Held, and Gower, 1970; Muir and Hains, 2004). Recent par-

allel findings in human infants stress the importance of self-generated actions in the development of spatial perception (Campos et al., 2000). Studies investigating the effect of signal properties on sensory information processing also reveal that normal development of the neural circuitry in the visual and auditory systems depends on the properties of the sensory inputs (White, Coppola, and Fitzpatrick, 2001; Chang and Merzenich, 2003). The maturation of the auditory space map in the superior colliculus (SC), a sensorimotor nucleus involved in orientation behavior, is selectively affected in guinea pigs raised in an omni-directional noise environment (Withington-Wray et al., 1990). Organization of primary auditory cortex in rat is shaped by salient acoustic inputs (Chang and Merzenich, 2003). Alteration of auditory spectral-spatial features also disrupts the development of the topographic representation of acoustic space in the SC (Schnupp, King, and Carlile, 1998; King, Parsons, and Moore, 2000). This body of evidence suggests that normal sensory development requires exposure to relevant sensory inputs.

Experience-dependent plasticity and adaptation is not limited to early postnatal development. Adult humans and other animals have been shown to adapt to altered acoustic experience. Adult humans listening to sounds in the environment through modified HRTFs can re-acquire the ability to localize sound with altered cues (Hofman, Van Riswick, and Van Opstal, 1998). More recently, Kacelnik et al. (2006) have shown that adult ferrets that are subjected to altered acoustic spatial cues can re-learn to localize sounds only when they are required to use the new cues in a behaviorally relevant task. In both studies subjects were allowed the opportunity to experience and learn the altered auditory-motor relations, however, the potential role of the sensorimotor learning was not systematically studied.

Plasticity in the spatial properties of the auditory SC maps has been demonstrated dur-

ing adulthood and facilitated by behaviors that require spatial hearing, for ferret (King et al., 2001) and barn owl (Bergan et al., 2005). Thus a computational theory of sound localization should include mechanisms that can re-calibrate after changes in sensory experience. Since the sensorimotor approach is inherently experience driven, it can easily capture this observed plasticity.

More evidence comes from sensory substitution experiments performed with blind human subjects (Bach-y-Rita and Kercel, 2003; Bach-y-Rita, 2004). There, providing spatial information through a different sensory modality (tactile stimulation) was enough to allow subjects to perceive spatial information, and the effect diminished if subjects were not allowed to interact with the environment. This phenomenon can be explained by the fact that sensorimotor contingencies were similar between the visual and substituting tactile inputs (O'Regan and Noë, 2001).

An interesting example also comes from studies of blind human infants equipped with sonar aids. These sonar devices emit ultrasound pulses and use received echo parameters to modulate a tone signal that is played to the ears (the direction information of the echo is encoded by creating an intensity difference in the tone signals at the two ears, and the distance information is represented by the frequency of the tone). These studies have shown that the infants use the device as a new sensory organ to monitor their environment (Bower, 1989). What makes these sonar studies especially interesting is that the incoming acoustic information does not include most of the natural spatial cues that are used by humans to localize sound sources. Thus, specialized auditory circuits that process particular spatial cues (e.g. ITD or ILD) could not have contributed to the localization of the echo sources. Moreover, neural circuits that could interpret such artificial information could not have

been innately hardwired as such. Yet the subjects were able to interpret and use the spatial information that was available through these devices. This again can be explained by the theory asserting that the brain monitors sensorimotor contingencies to interpret space, rather than relying on innately placed circuitry that is specifically designed to serve a particular function, such as ILD and ITD processing.

Using a virtual sound-source localization experiment in which subjects could interact with the acoustic environment, Loomis, Hebert, and Cicinelli (1990) showed that in the absence of a spectral-shaping effect of the pinnae with limited and unnatural spatial cues, subjects can still externalize and localize sound sources. These findings suggest that spatial perception may not be purely innate, requiring voluntary actions to develop and maintain it.

We cannot ignore, however, innate (non-naive) components of exteroceptive sensory processing. Organisms might plausibly use genetically wired information to lay the foundations of the computations we have been examining, and it is the experience gained by active monitoring of the environment that shapes, tunes and calibrates these structures to generate meaningful interpretation of the sensory signals (Clifton, 1992; Muir and Hains, 2004). Behavioral studies on young animals demonstrate that immediately after birth, or coinciding with the onset of hearing, they show the ability to orient towards sound sources (Kelly and Potash, 1986). This behavior is slow and not as accurate compared to that of adults. Moreover, during development, accuracy of orientation to sound sources shows a U-shape function, such that accuracy of the orientation behavior initially decreases with age, e.g. 2-3 in human infants (Muir, Clifton, and Clarkson, 1989) and 24-27 days in gerbils (Kelly and Potash, 1986), and then slowly increases in accuracy and finally reaches adult

levels. It has been suggested that the initial acoustic orienting response might be diminished or disappear as the development of the forebrain progresses. Thus, later emergence of the orientation towards a sound-source location observed in the young animals might reflect localization behavior controlled by the midbrain and forebrain structures (Muir and Hains, 2004). Muir and Hains (2004) proposed that orienting to sound sources in early infancy is an example of a reflex that disappears later in development. The advantage of such reflexes in the development and learning of more complex behavior in robots has been proposed recently by Metta (2000).

4.4.3 Multi-sensory nature of spatial hearing

The examples presented in this study are limited to hearing as the only exteroceptive sense. For this restriction the HRTF must satisfy certain conditions that allow the organism to identify points in space without the need for a reference provided by another exteroceptive sense. This reference is not (though it might have been) essential for the organism to define a point as the set of acoustic inputs from which the same motor actions generates the reference sensory input. We have argued that a point in space can be identified by the organism based on the observation that different acoustic inputs originated from the same relative spatial location show similar changes to motor actions. We employ this assumption because there is considerable evidence provided by studies on blind subjects suggesting that sound localization can develop without the help of another distal exteroceptive sense such as vision (Ashmead et al., 1998; Zwiers, Van Opstal, and Cruysberg, 2001b; Lewald, 2002). However, we fully recognize the potential importance of vision, when available. The

influence of vision over the auditory system has been demonstrated in barn owl (Knudsen and Knudsen, 1989) and in human; particularly under noisy conditions vision is thought to be involved in the calibration of sound-source elevation cues (Zwiers, Van Opstal, and Cruysberg, 2001a). But the fact that sound localization can develop in the absence of vision suggest that visual influence is not required and so a computational model of sound localization should not require supplemental sensory information for calibration (Kacelnik et al., 2006).

4.4.4 Role of the motor state in sound localization

One of the important aspects of the sensorimotor approach is the organism's ability to monitor its motor states and associate them with its changing acoustic inputs. This requires the motor system, proprioception and the auditory system to interact with each other. Recent studies provide evidence suggesting that these relations do exist. Influence of proprioception on sound localization has been shown in relation to eye (Lewald, 1997) and head positions (Lewald, Dörrscheidt, and Ehrenstein, 2000), direction of gravity (DiZio et al., 2001) and whether or not the head is free to move (Tollin et al., 2005). Computation of sound localization has also been shown to be influenced by the vestibular system (Lewald and Karnath, 2000). Vliegen, Van Grootel, and Van Opstal (2004) proposed that head position signals interact with the processing of the acoustic spatial cues by way of modulating each frequency channel in a frequency specific manner. Electrophysiological findings by Kanold and Young (2001) in cats have shown that ear movements influence neurons DCN. This body of evidence suggests that computation of sound localization does not solely de-

pend on the acoustic inputs.

Reference frame of sound location

We have shown that, for a given motor state, an organism can capture a set of global parameters that represents the spatial locations of sound sources. These parameters can be different for different motor states, e.g. for an animal with mobile pinnae. A change in pinna position induces changes in the HRTF that may not be accounted for by a simple rotation of the HRTF before the position change (Young, Rice, and Tong, 1996). The motor state of the pinnae will determine the operating HRTF at every instant. With our method one can produce a family of global parameters associated with the different pinna states. These parameters represent sound locations in a pinna-related reference system. However, representations of the sound locations in different reference frames that are attached to the head, body or the exterior space are important for the behavior of an organism. One psychoacoustical study demonstrates that the sound-source locations are represented in a body-centered reference frame (Goossens, 1999). In order for the auditory system to use a body-centered reference frame, proprioceptive information from the pinna and the head should be taken into consideration by the system. Vliegen, Van Grootel, and Van Opstal (2004) suggest that the auditory system processes dynamically varying acoustic information caused by self-generated head movements, in such a way that a stable representation of the sound source location is constructed. From an animal's point of view the environment surrounding the animal is stable as it moves. This requires the ability to distinguish sensory input changes caused by self-generated movements from those that are the result of changes in the environment. This can be achieved by using proprioceptive information, plus the abil-

ity to predict sensory consequences of the organism's actions, i.e. sensorimotor expertise. It has been shown that human subjects represent visual information in an allocentric reference frame if the sensory consequences of their actions are predictable. When the sensory consequences of the movements are not predictable, or the movements are involuntary, the representation is shown to be in egocentric reference frame (eye-centered) (Wexler, 2003). Our computational scheme can be extended to create a body-centered or allocentric representation of sound source location: as mentioned earlier, different global parameters obtained at different motor states are related to each other by a coordinate transformation (Vuillemin, 1972).

4.4.5 Localization of sound sources without head movements

We have shown that it is possible for a naive organism to obtain the spatial parameters of sound source directions using voluntary head movements as a tool to explore the sensory input space. In fact, knowledge of the auditory consequences of voluntary movements has been shown to be very effective to estimate both azimuth and elevation of a sound source even for a spherical head with no pinnae (Handzel and Krishnaprasad, 2002). However, it is well known from common psychoacoustic studies that localization of a sound source does not require head movements. Rather, a subject can localize acoustic signals based on the acoustic information received at the ears. How then can the sensorimotor approach account for the sound localization without motion? We assert that sensorimotor early experience, i.e. during development, is necessary for accurate sound localization.

Sensorimotor interactions give the organism the means to identify and parameterize

the points in space, without requiring prior knowledge of the spatial parameters or the organism-environment geometry. We posit that it is through head movements that the organism learns this geometry and the necessary parameters. With this knowledge it is then possible to further investigate the properties of the acoustic signals, to determine a relationship between these inputs and their corresponding spatial parameters in the absence of head movements. We have also discussed that invariant features, obtained from sensorimotor contingencies and independent of sound spectra, should be unique to a particular location for it to be localized unambiguously. Thus, all that the organism needs is to discover the invariant feature associated with each point in space, and recognize it in the acoustic signals received at the ears. The access to these features is possible through motor actions. Once the organism is able to learn these features, it is possible to explore a function that maps the acoustic inputs to the internal coordinates of the spatial positions. Note that because of the smoothness of the orbits the invariant feature will also change smoothly across space. This property then will allow interpolation at the unexperienced acoustic sensory inputs.

4.4.6 Neurophysiological implications

Most neurophysiological studies investigating spatial properties of auditory neurons are limited to cases in which animals are prevented from moving their bodies, heads and ears. Involvement of anesthesia may also limit capturing the auditory system's normal function. For well-studied animals like bats and cats, localization cues are subject to change in relation to ear position. One would expect to see the effects of different motor states on the processing of spatial information, which may not be accessible when animals are

limited in their ability to move. Thus, it would be informative to study the effect of movement of the head and ears on the auditory nuclei that are thought to be involved in spatial information processing. The proposition of the distributed effect of head motion across frequency channels suggests that this can happen in different parts of the auditory system (Vliegen, Van Grootel, and Van Opstal, 2004). There is also evidence of somatosensory influence on the DCN in cats (Kanold and Young, 2001) and proprioceptive influence on the auditory cortex (Alexeenko and Verderevskaya, 1976). Effects of eye position on auditory processing has been reported in the inferior colliculus (IC) (Groh et al., 2001; Zwiers, Versnel, and Opstal, 2004) and superior colliculus (SC) (Jay and Sparks, 1984) in monkey . These findings suggest that auditory system receives information from many sensory modalities which may contribute to spatial processing.

4.4.7 Applications to robotics

Sensorimotor theory offers important applications in the field of robotics and design of intelligent systems. Even if the tabula rasa assumption of may not be completely valid for young living organisms, they are typically valid for artificial systems designed to interact with the surroundings. The importance of autonomous behavior in robots has been long-recognized: a system cannot be programmed to handle every possible scenario of organism-environment interaction, so it is important that robots have the ability to learn and adapt. Presently, when designing sensors, it is critically important that the sensory information not deviate from the tolerable limits of its parameters, since the robot's interpretation of its sensory inputs is dependent on how reliable they are. The sensorimotor

approach allows flexibility to the design: instead of hard-coding the properties of the sensors and the interpretation of the information provided by them, flexible algorithms can be designed, using sensorimotor principles, to allow a robot to calibrate its sensors and choose the information that is useful for a given task. This provides the freedom to the designer determining how much of the hard-coding should go in the system.

4.5 Conclusion

In this report we have proposed a computational method for the learning of the auditory space using sensorimotor theory (O'Regan and Noë, 2001; Poincare, 1929), an unexplored issue of the problem of sound localization. We have argued that a computational theory of sound localization should be able to explain the experience-dependent nature of the computation as well as its dependence on other sensory inputs. This computational method provides a framework under which integration of experience-dependent plasticity and multisensory information processing aspects of sound localization can be achieved. By way of examples from humans and bats we have shown that a naive organism can learn to localize sound based solely on dynamic acoustic inputs and their relation to motor state.

Chapter 5

General Discussion

Bats need to localize directions of the echoes accurately to be able intercept their prey or avoid obstacles during flight (Erwin, Wilson, and Moss, 2001). Studying directional properties of the echolocation system is essential to improve our understanding of the computations underlying bat sound localization. Sound localization, during passive listening and under the static listening conditions, relies on the directional filtering of the sounds by the pinna (Wightman and Kistler, 1989b). During echolocation acoustic signals are also influenced by the sonar beam. Bats could take advantage of this additional directional effect to improve their localization ability (Hartley, 1990; Wotton, Jenison, and Hartley, 1997). This work focuses on these two important components of the sound localization process in echolocating bat, *Eptesicus fuscus*.

Studying the head-related transfer functions (HRTF) is an important step in our understanding of sound localization, as localization cues available to all mammals are captured entirely by the HRTF. Chapter-two focused on the HRTF obtained from an echolocating bat, *Eptesicus fuscus*. Attention was given on the spatial properties of the binaural and

monaural sound localization cues as well as influence of the external ear to the directional properties of the HRTF.

It is largely accepted that the binaural-difference and the monaural-spectral cues are used to localize sound source azimuth and elevation, respectively. However, neither the monaural cues nor the interaural level differences (ILD) measured from bats behaved in a way that is consistent with this dichotomy. Analysis of the frequency-dependent behavior of ILD showed that below 40 kHz ILD is predominantly dependent on azimuth. Above this frequency, ILD is a complex function of both azimuth and elevation. Based on these results, I conclude that ILD can be a localization cue for both components of the sound location.

Importance of ILD cue for bats have been received attention in the past as well (Grinnell and Grinnell, 1965; Fuzessery and Pollak, 1984). Especially for CF-FM bats for whom echolocation signals are limited to constant frequency and its harmonics, ILD as a general localization cue can be highly valuable, for there is no spectral cue available for vertical localization. For most CF-FM bats, ILD cues are a non-monotonic function of sound source azimuth. Therefore, ILDs from multiple frequencies, e.g. three for mustached bat (Fuzessery and Pollak, 1984), can be used to triangulate the sound source position.

Using similar ideas combined with the ILD behavior of *Eptesicus fuscus*, an ILD based sound localization can also be possible (Aytekin and Moss, 2004). The computational strategy for ILD should allow independent contributions from individual frequencies. ILD obtained from tone signals are sufficient because horizontal sound localization supports such independent contribution (Blauert, 1997). With each frequency voting for certain spatial locations with a particular ILD observed, sound location can be decided in favor

of the spatial position that receives the most votes. Since below and above 45 kHz ILDs are biased towards azimuth, these votes can be collected separately for both frequency regions. After normalized or weighted they can be combined through a multiplication process analogous to one found in ICx of a barn owl (Peña and Konishi, 2001).

The spectral notches showed a systematic change with sound direction. The iso-frequency lines of notch center frequencies laid diagonally from contralateral at low elevations to ipsilateral at high elevations, exhibiting both elevation and azimuth dependence. When the iso-frequency lines from both ears are overlaid in space, a coordinate system for the sound source location in the frontal hemisphere is obtained. Hence, the notch center frequencies that are detected at the left and right ears can be mapped to the sound source positions in space. This grid structure is more robust at frequencies below 40 kHz, complementing the ILD cues at the low frequencies where ILD is predominantly azimuth dependent. One caveat using spectral notches as localization cues for bats could be that the spectrum of an echo signal can contain spectral notches (Simmons et al., 1995) which can create ambiguity in identifying the spatial notch cue. A sound source with a notch in its spectrum will be detected by both ears. By contrast, most of the time spectral notches caused by directional pinna filtering will be different at each ear. Consequently, it can be possible to separate the spatial notch cue from the notches in the sound source spectrum.

The fact that both cues provide azimuth and elevation information improves the robustness of sound localization computation. For instance, signals received at the ears that have no notch cue because of the gaps in the sound source spectrum, ILD can still be a useful cue for sound localization. Whereas, if received sound deprived from frequencies above 40 kHz spectral notch cue can provide the missing elevation information that ILD cannot

provide.

Results obtained in chapter two are not free from caveats. One that stands out is that the observation was limited to a situation where the pinnae are immobile during passive listening. Thus, HRTF-based analysis cannot capture the role of the pinna mobility in sound localization. Although this does not undermine the importance of the findings, it certainly does not provide a complete picture of the spatial information available to a bat during passive listening. If bats can localize sounds at different pinna positions, the bat auditory system has to somehow adapt its computations to the change in the localization cues. This implies that the auditory system has to receive input indicating the new position of the pinna,

Identifying spatial acoustic cues requires testing the hypotheses built in chapter 2. However, behavioral experiments with *Eptesicus fuscus* have their own challenges. Since these bats can move their pinna it is extremely challenging, if not impossible, to establish static listening conditions. Static listening conditions are helpful for building a direct relationship between the perceived location of space and the HRTF features. Yet, it is still possible to investigate what frequency regions are relevant to bats for azimuth and elevation localization and whether the localization performance differs between passive and active listening conditions.

Bats can evaluate sound in active (echolocation) and passive modes. During echolocation sounds received at the ears convey information about the layout of the bat's immediate environment. The echoes arriving from off-center directions will be lowpass filtered because the sonar beam gets more directional with increasing frequency. The lowpass signals could result in a larger spatial ambiguity for their location. Hence, the spatial information

received by a bat may be blurred on the periphery and sharper at the directions near the axis of the sonar beam. Moreover, sonar beams can be complex in shape with side lobes and deep spatial nulls in between them as in the case of HRTFs. If sonar beamshapes show such complexity, they will influence the frequency spectrum of the echoes by inducing direction-dependent spectral features. Thus, the auditory system of the bat could take advantage of these additional cues for the localization of the echoes. Note that this would imply an additional or separate computational organization, as opposed to using the same computational strategies for passive listening.

I took a unique approach to measure sonar beams from freely behaving and vocalizing bats. The measurement method was designed to build the sonar beamshape from multiple bat vocalizations, using a camera to estimate the bat's head orientation while simultaneously capturing sonar vocalizations at 22 different points in space at the chosen frequency. The efficiency of the method is demonstrated using an ultrasound loudspeaker that could be moved in space simulating a bat. The data obtained from the bat, however, showed a large spatial variation. The difference between the spatial variations obtained from the loudspeaker and the bat could be a result of sonar beam variations which can be explained by the natural variation in the bat sonar beams.

Although, the effects of the beamshape on the localization cues are yet to be determined, what we know so far from the previous investigations is that the sonar beam can influence the spectral cues (Wotton, Jenison, and Hartley, 1997). If the beamshape can vary the sound localization computation has to compensate for resulting changes in the spatial cues. A signal that conveys information about the sonar beam generated can be used for this compensation.

By studying directional properties of both reception and emission component of the sonar system, my goal was to get a complete description of the spatial information that is available to the bat's auditory system to localize echo sources. If I could determine the acoustic features that can be used potentially by the bat, the next step is to investigate how these features are extracted and used to represent space. This type of approach in general can be named as a black-box model. The idea behind the black-box approach is to understand the processes that maps a computational system's inputs to its corresponding outputs by studying the interdependence of the two. Input of the sound location computation is the acoustic signals that carry spatial information. The output is the position of the sound source with reference to the body (or head or some external reference point). The research efforts have been focusing on understanding what acoustic features are available to the black-box and how the black-box makes the mapping to spatial parameters of the input signals. The approach that is taken by the many investigations in sound localization boils down to understanding the relationship between input and output of the sound localization process.

With reference to studies described in chapters two and three, there were two separate cases in which I mentioned the need for extra-auditory influence on the sound location computation. The auditory system needs extra-auditory signals to incorporate ear positions and the outgoing beamshape to computation of sound locations. How do extra-auditory signals interact with the computation of sound location and how can this interaction be established? The bats and other animals with similar conditions have to somehow learn how the directional cues change with ear movements. How do animals associate acoustic-spatial cues to locations in space? To take this even further, how can they even learn to

distinguish acoustic features in sounds from the non-acoustic ones?

Chapter four takes on these questions by considering the worst case scenario which assumes an auditory system that is initially naive to the spatial nature of sounds. Under these conditions, we postulate that in addition to acoustic inputs, proprioceptive inputs are also necessary to build a representation of space and learn to localize sounds. The sensorimotor contingencies are shown as the key component for the nervous system to learn properties of auditory space. The concepts that are developed in this chapter are not limited to the auditory system but rather can be generalized to learning of space via other sensory modalities.

The sensorimotor approach also helps explain how the black-box itself can learn its proper outputs. This approach also allowed for a way to bring different aspects of sound localization, such as experience-dependent acquisition and adaptation and multi-modal influence together.

Bibliography

Alexeenko, N. Y. and N. N. Verderevskaya (1976). Proprioceptive effects on evoked responses to sounds in the cat auditory cortex. *Exp Brain Res* 26(5): 495–508.

Algazi, V. R., R. O. Duda, R. P. Morrison, and D. M. Thompson (2001). The cipic hrtf database. In *Proceedings of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 99–102, Mohonk Mountain House, New Paltz, NY.

Ashmead, D. H., R. S. Wall, K. A. Ebinger, S. B. Eaton, M. M. Snook-Hill, and X. Yang (1998). Spatial hearing in children with visual disabilities. *Perception* 27(1): 105–122.

Aytekin, M., E. Grassi, M. Sahota, and C.F. Moss (2004). The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation. *J Acoust Soc Am* 116(6): 3594–3605.

Aytekin, M. and C.F. Moss (2004). Interaural level difference based sound localization by bats. In *Society for Neuroscience, 2004 Annual Meeting*.

Bach-y-Rita, Paul (2004). Tactile sensory substitution studies. *Ann N Y Acad Sci* 1013: 83–91.

Bach-y-Rita, Paul and Stephen W Kercel (2003). Sensory substitution and the human-machine interface. *Trends Cogn Sci* 7(12): 541–546.

Batteau, D. W. (1967). The role of the pinna in human localization. *Proc R Soc Lond B Biol Sci* 168(11): 158–80. pdf file exists.

Belkin, Mikhail and Partha Niyogi (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* 15(6): 1373–1396.

Beranek, L. L. (1988). *Acoustics*. the American Institute of Physics, Woodbury, NY. Published for the Acoustical Society of America by.

Bergan, Joseph F., Peter Ro, Daniel Ro, and Eric I. Knudsen (2005). Hunting increases adaptive auditory map plasticity in adult barn owls. *J Neurosci* 25(42): 9816–9820.

Binns, K. E., S. Grant, D. J. Withington, and M. J. Keating (1992). A topographic representation of auditory space in the external nucleus of the inferior colliculus of the guinea-pig. *Brain Res* 589(2): 231–242.

Binns, K. E., D. J. Withington, and M. J. Keating (1995). The developmental emergence of the representation of auditory azimuth in the external nucleus of the inferior colliculus of the guinea-pig: the effects of visual and auditory deprivation. *Brain Res Dev Brain Res* 85(1): 14–24.

Blakemore, S. J., C. D. Frith, and D. M. Wolpert (2001). The cerebellum is involved in predicting the sensory consequences of action. *Neuroreport* 12(9): 1879–1884.

Blauert, J. (1997). *Spatial Hearing*. MIT Press, Cambridge, MA, Cambridge, MA.

Bloom, P. J (1977). Creating source elevation illusions by spectral manipulations. *J. Audio Eng. Soc.* 25: 560–565.

Bower, T. G. R. (1989). *Rational Infant: learning in infancy*. A series of books in psychology. W. H. Freeman and Company, New York, University of Texas, Dallas.

Brown, C. H. and B. J. May (2003). *Sound Source Localization*, chapter Comparative Mammalian Sound Localization, pp. 124–177. Springer Handbook of Auditory Research. Springer-Verlag, New York.

Burnett, L. R., B. E. Stein, D. Chaponis, and M. T. Wallace (2004). Superior colliculus lesions preferentially disrupt multisensory orientation. *Neuroscience* 124(3): 535–547.

Butler, R. A. and C. C. Helwig (1983). The spatial attributes of stimulus frequency in the median sagittal plane and their role in sound localization. *Am J Otolaryngol* 4(3): 165–173.

Campbell, Robert A A, Timothy P Doubell, Fernando R Nodal, Jan W H Schnupp, and Andrew J King (2006). Interaural timing cues do not contribute to the map of space in the ferret superior colliculus: a virtual acoustic space study. *J Neurophysiol* 95(1): 242–254.

Campos, J., D. Anderson, M. Barbu-Roth, E. Hubbard, M. Hertenstein, and D. Witherington (2000). Travel broadens the mind. *Infancy* 1: 149–219.

Carlile, S. (1990). The auditory periphery of the ferret. i: Directional response properties and the pattern of interaural level differences. *J Acoust Soc Am* 88(5): 2180–2195.

Carlile, S and D Pralong (1994). The location-dependent nature of perceptually salient features of the human head-related transfer functions. *J Acoust Soc Am* 95(6): 3445–59.

Casseday, J. H. and W. D. Neff (1975). Auditory localization: role of auditory pathways in brain stem of the cat. *J Neurophysiol* 38(4): 842–858.

Chang, Edward F. and Michael M. Merzenich (2003). Environmental noise retards auditory cortical development. *Science* 300(5618): 498–502.

Clements, M. and J. B. Kelly (1978). Auditory spatial responses of young guinea pigs (*Cavia porcellus*) during and after ear blocking. *J Comp Physiol Psychol* 92(1): 34–44.

Clifton, R. K. (1992). In Werner, A. and E. W. Rubel, editors, *Developmental psychoacoustics*, chapter The development of spatial hearing in human infants, pp. 135–157. American Psychological Association, Washington, D.C.

Clifton, R. K., J. Gwiazda, J. A. Bauer, M. G. Clarkson, and R. M. Held (1988). Growth in head size during infancy: Implications for sound localization. *Developmental Psychology* 24(4): 477–83.

Cohen, Y. E. and E. I. Knudsen (1999). Maps versus clusters: different representations of auditory space in the midbrain and forebrain. *Trends Neurosci* 22(3): 128–135.

Colburn, H. Steven and Abhijit Kulkarni (2005). *Models of sound localization*, Vol. 25 of *Handbook of auditory research*, chapter Models of sound localization, pp. 272–316. Springer, New York.

Covey, E. and J. H. Casseday (1995). *Hearing by Bats*, chapter The lower brainstem auditory pathways, pp. 235–295. Springer, New York, New York.

Davis, Kevin (2005). Contralateral Effects and Binaural Interactions in Dorsal Cochlear Nucleus. *J Assoc Res Otolaryngol* pp. 1–17.

Davis, Kevin A, Ramnarayan Ramachandran, and Bradford J May (2003). Auditory processing of spectral cues for sound localization in the inferior colliculus. *J Assoc Res Otolaryngol* 4(2): 148–163.

Dementhon, Daniel F. and Larry S. Davis (1995). Model-based object pose in 25 lines of code. *Int. J. Comput. Vision* 15(1-2): 123–141.

DiZio, P., R. Held, J. R. Lackner, B. Shinn-Cunningham, and N. Durlach (2001). Gravitoinertial force magnitude and direction influence head-centric auditory localization. *J Neurophysiol* 85(6): 2455–60.

Erwin, H. R., W. W. Wilson, and C. F. Moss (2001). A computational sensorimotor model of bat echolocation. *J Acoust Soc Am* 110(2): 1176–1187.

Evans, Michael J., James A. S. Angus, and Anthony I. Tew (1998). Analyzing head-related transfer function measurements using surface spherical harmonics. *J. Acoust. Soc. Am.* 104: 2400–2411.

Fay, R. R. and A. N. Popper (2000). Evolution of hearing in vertebrates: the inner ears and processing. *Hear Res* 149(1-2): 1–10.

Fermuller, Cornelia and Yiannis Aloimonos (1994). Vision and action. Technical report CAR-TR-722,CS-TR-3305, Computer Vision Laboratory Center for Automation Research, University of Maryland, University of Maryland, College Park,20742-3275.

Ferragamo, M. J., T. Haresign, and J. A. Simmons (1998). Frequency tuning, latencies, and responses to frequency-modulated sweeps in the inferior colliculus of the echolocating bat, *ptesicus fuscus*. *J Comp Physiol [A]* 182(1): 65–79.

Fuzessery, Z. M. (1996). Monaural and binaural spectral cues created by the external ears of the pallid bat. *Hear Res* 95(1-2): 1–17.

Fuzessery, Z. M., P. Buitenhoff, B. Andrews, and J. M. Kennedy (1993). Passive sound localization of prey by the pallid bat (*antrozous p. pallidus*). *J Comp Physiol [A]* 171(6): 767–777.

Fuzessery, Z. M. and G. D. Pollak (1984). Neural mechanisms of sound localization in an echolocating bat. *Science* 225(4663): 725–728.

Fuzessery, Zoltan M., David J. Hartley, and Jeffrey J. Wenstrup (1992). Spatial processing within the mustache bat echolocation system: possible mechanisms for optimization. *J Comp Physiol A* 170: 57–71.

Getzmann, Stephan (2002). The effect of eye position and background noise on vertical sound localization. *Hear Res* 169(1-2): 130–9.

Ghose, Kaushik and Cynthia F Moss (2003). The sonar beam pattern of a flying bat as it tracks tethered insects. *J Acoust Soc Am* 114(2): 1120–1131.

Goossens, H. H.; van Opstal, A. J. (1999). Influence of head position on the spatial representation of acoustic targets. *J Neurophysiol* 81(6): 2720–2736.

Griffin, Donald (1958). *Listening in the dark*. Yale University Press.

Grinnell, A. D. and V. S. Grinnell (1965). Neural correlates of vertical localization by echo-locating bats. *J Physiol* 181(4): 830–851.

Grinnell, A. D. and H. U. Schnitzler (1977). Directional sensitivity of echolocation in the horseshoe bat, *rhinolophus ferrumequinum*. ii. behavioral directionality of hearing. *J. Comp. Physiol.* 116: 63–76.

Groh, J. M., A. S. Trause, A. M. Underhill, K. R. Clark, and S. Inati (2001). Eye position influences auditory responses in primate inferior colliculus. *Neuron* 29(2): 509–18.

Grothe, B. and T. J. Park (2000). Structure and function of the bat superior olivary complex. *Microsc Res Tech* 51(4): 382–402.

Grubb, Matthew S and Ian D Thompson (2004). The influence of early experience on the development of sensory systems. *Curr Opin Neurobiol* 14(4): 503–512.

Guppy, A. and R. B. Coles (1988). Acoustical and neural aspects of hearing in the Australian gleaning bats, *macroderma gigas* and *nyctophilus gouldi*. *J Comp Physiol [A]* 162(5): 653–668.

Hammershøi, D. and H. Møller (1996). Sound transmission to and within the human ear canal. *J Acoust Soc Am* 100(1): 408–427.

Handzel, Amir A. and P. S. Krishnaprasad (2002). Biomimetic sound-source localization. *IEEE sensors journal* 2(6): 607–616.

Hartley, D. J. (1990). Phased-array beam scanning as a possible aid to horizontal localization in horseshoe bats. *J Acoust Soc Am* 88(6): 2889–2891.

Hartley, D. J. (1992). Stabilization of perceived echo amplitudes in echolocating bats. ii. the acoustic behavior of the big brown bat, *ptesicus fuscus*, when tracking moving prey. *J Acoust Soc Am* 91(2): 1133–1149.

Hartley, D. J. and R. A. Suthers (1987). The sound emission pattern and the acoustical role of the noseleaf in the echolocating bat, *Carollia perspicillata*. *J Acoust Soc Am* 82(6): 1892–1900.

Hartley, D. J. and R. A. Suthers (1989). The sound emission pattern of the echolocating bat, *Eptesicus fuscus*. *J Acoust Soc Am* 85(3): 1348–1351.

Hartmann, W. M. and A. Wittenberg (1996). On the externalization of sound images. *J Acoust Soc Am* 99(6): 3678–3688.

Hebrank, J and D Wright (1974). Are two ears necessary for localization of sound sources on the median plane? *J Acoust Soc Am* 56(3): 935–8.

Heffner, R. S. (1997). Comparative study of sound localization and its anatomical correlates in mammals. *Acta Otolaryngol Suppl* 532: 46–53.

Hein, A., R. Held, and E. C. Gower (1970). Development and segmentation of visually controlled movement by selective exposure during rearing. *J Comp Physiol Psychol* 73(2): 181–187.

Held, R. and A. Hein (1963). Movement-produced stimulation in the development of visually guided behavior. *J Comp Physiol Psychol*. 56(5): 872–876.

Held, Richard (1955). Shifts in binaural localization after prolong exposures to atypical combinations of stimuli. *Am J Psychol* 68: 526–548.

Hofman, P. M., J. G. Van Riswick, and A. J. Van Opstal (1998). Relearning sound localization with new ears. *Nat Neurosci* 1(5): 417–421.

Hofman, Paul M and A. John Van Opstal (2002). Bayesian reconstruction of sound localization cues from responses to random spectra. *Biol Cybern* 86(4): 305–316.

Hofman, Paul M and A. John Van Opstal (2003). Binaural weighting of pinna cues in human sound localization. *Exp Brain Res* 148(4): 458–470.

Huang, A. Y. and B. J. May (1996). Sound orientation behavior in cats. II. mid-frequency spectral cues for sound localization. *J Acoust Soc Am* 100(2 Pt 1): 1070–80.

Imamizu, Hiroshi, Tomoe Kuroda, Satoru Miyauchi, Toshinori Yoshioka, and Mitsuo Kawato (2003). Modular organization of internal models of tools in the human cerebellum. *Proc Natl Acad Sci U S A* 100(9): 5461–5466.

Irvine, D. R. and G. Gago (1990). Binaural interaction in high-frequency neurons in inferior colliculus of the cat: effects of variations in sound pressure level on sensitivity to interaural intensity differences. *J Neurophysiol* 63(3): 570–591.

Jain, Roshini and Susan Shore (2006). External inferior colliculus integrates trigeminal and acoustic information: unit responses to trigeminal nucleus and acoustic stimulation in the guinea pig. *Neurosci Lett* 395(1): 71–75.

Javer, A. R. and D. W. Schwarz (1995). Plasticity in human directional hearing. *J Otolaryngol* 24(2): 111–117.

Jay, M. F. and D. L. Sparks (1984). Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature* 309(5966): 345–347.

Jay, M. F. and D. L. Sparks (1987). Sensorimotor integration in the primate superior colliculus. i. motor convergence. *J Neurophysiol* 57(1): 22–34.

Jen, P. H. and D. M. Chen (1988). Directionality of sound pressure transformation at the pinna of echolocating bats. *Hear Res* 34(2): 101–117.

Jenkins, W. M. and R. B. Masterton (1982). Sound localization: effects of unilateral lesions in central auditory system. *J Neurophysiol* 47(6): 987–1016.

Jiang, W., M. T. Wallace, H. Jiang, J. W. Vaughan, and B. E. Stein (2001). Two cortical areas mediate multisensory integration in superior colliculus neurons. *J Neurophysiol* 85(2): 506–522.

Jiang, Wan, Huai Jiang, Benjamin A Rowland, and Barry E Stein (2007). Multi-sensory orientation behavior is disrupted by neonatal cortical ablation. *J Neurophysiol* 97(1): 557–562.

Jiang, Wan, Huai Jiang, and Barry E Stein (2002). Two corticotectal areas facilitate multisensory orientation behavior. *J Cogn Neurosci* 14(8): 1240–1255.

Jiang, Wan, Huai Jiang, and Barry E Stein (2006). Neonatal cortical ablation disrupts multisensory development in superior colliculus. *J Neurophysiol* 95(3): 1380–1396.

Jin, Craig, Anna Corderoy, Simon Carlile, and André Schaik (2004). Contrasting monaural and interaural spectral cues for human sound localization. *J Acoust Soc Am* 115(6): 3124–3141.

Joris, P. X. and T. C. Yin (1998). Envelope coding in the lateral superior olive. iii. comparison with afferent pathways. *J Neurophysiol* 79(1): 253–269.

Kaas, J. H. (1997). Topographic maps are fundamental to sensory processing. *Brain Res Bull* 44(2): 107–112.

Kacelnik, Oliver, Fernando R Nodal, Carl H Parsons, and Andrew J King (2006). Training-induced plasticity of auditory localization in adult mammals. *PLoS Biol* 4(4): e71.

Kanold, P. O. and E. D. Young (2001). Proprioceptive information from the pinna provides somatosensory input to cat dorsal cochlear nucleus. *J Neurosci* 21(19): 7848–7858.

Kavanagh, G. L. and J. B. Kelly (1992). Midline and lateral field sound localization in the ferret (*Mustela putorius*): contribution of the superior olivary complex. *J Neurophysiol* 67(6): 1643–1658.

Kawato, Mitsuo, Tomoe Kuroda, Hiroshi Imamizu, Eri Nakano, Satoru Miyauchi, and Toshinori Yoshioka (2003). Internal forward models in the cerebellum: fMRI study on grip force and load force coupling. *Prog Brain Res* 142: 171–188.

Keller, CH, K Hartung, and TT Takahashi (1998). Head-related transfer functions of the barn owl: measurement and neural responses. *Hear Res* 118(1-2): 13–34.

Kelly, J. B. and M. Potash (1986). Directional responses to sounds in young gerbils (*Meriones unguiculatus*). *J Comp Psychol* 100(1): 37–45.

Kick, S. A. and J. A. Simmons (1984). Automatic gain control in the bat's sonar receiver and the neuroethology of echolocation. *J Neurosci* 4(11): 2725–2737.

King, A. J., O. Kacelnik, T. D. Mrsic-Flogel, J. W. Schnupp, C. H. Parsons, and D. R. Moore (2001). How plastic is spatial hearing? *Audiol Neurootol* 6(4): 182–186.

King, A. J., C. H. Parsons, and D. R. Moore (2000). Plasticity in the neural coding of auditory space in the mammalian brain. *Proc Natl Acad Sci U S A* 97(22): 11821–11828.

King, Andrew J, Victoria M Bajo, Jennifer K Bizley, Robert A A Campbell, Fernando R Nodal, Andreas L Schulz, and Jan W H Schnupp (2007). Physiological and behavioral studies of spatial coding in the auditory cortex. *Hear Res* .

- Kistler, D. J. and F. L. Wightman (1992). A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J Acoust Soc Am* 91(3): 1637–47.
- Knudsen, E. I. (1982). Early auditory experience modifies sound localization in barn owls. *Nature* 295: 238–240.
- Knudsen, E. I. and P. F. Knudsen (1985). Vision guides the adjustment of auditory localization in young barn owls. *Science* 230(4725): 545–548.
- Knudsen, E. I. and P. F. Knudsen (1989). Vision calibrates sound localization in developing barn owls. *J Neurosci* 9(9): 3306–13.
- Koay, G., D. Kearns, H. E. Heffner, and R. S. Heffner (1998). Passive sound-localization ability of the big brown bat (*Eptesicus fuscus*). *Hear Res* 119(1-2): 37–48.
- Lackner, J. R. and P. A. DiZio (2000). Aspects of body self-calibration. *Trends Cogn Sci* 4(7): 279–288.
- Lackner, James R and Paul DiZio (2005). Vestibular, proprioceptive, and haptic contributions to spatial orientation. *Annu Rev Psychol* 56: 115–147.
- Langendijk, Erno H. A. and Adelbert W. Bronkhorst (2002). Contribution of spectral cues to human sound localization. *J Acoust Soc Am* 112(4): 1583–96.
- Lawrence, B. D. and J. A. Simmons (1982). Echolocation in bats: the external ear and perception of the vertical positions of targets. *Science* 218(4571): 481–483.

Lewald, J (1997). Eye-position effects in directional hearing. *Behav Brain Res* 87(1): 35–48.

Lewald, J, GJ Dörrscheidt, and WH Ehrenstein (2000). Sound localization with eccentric head position. *Behav Brain Res* 108(2): 105–25.

Lewald, J. and W. H. Ehrenstein (1998). Influence of head-to-trunk position on sound lateralization. *Exp Brain Res* 121(3): 230–8.

Lewald, J. and HO. Karnath (2000). Vestibular influence on human auditory space perception. *J Neurophysiol* 84(2): 1107–1111.

Lewald, Jörg (2002). Vertical sound localization in blind humans. *Neuropsychologia* 40(12): 1868–1872.

Lewald, Jörg and Hans-Otto Karnath (2002). The effect of whole-body tilt on sound lateralization. *Eur J Neurosci* 16(4): 761–6.

Loomis, J. M., C. Hebert, and J. G. Cicinelli (1990). Active localization of virtual sounds. *J Acoust Soc Am* 88(4): 1757–64.

Lopez-Poveda, E. A. and Ray Meddis (1996). A physical model of sound diffraction and reflections in the human concha. *J. Acoust. Soc. Am.* 100: 3248–3259.

Macpherson, Ewan A. (1998). Spectral Cue Processing in the Auditory Localization of Sounds with Wideband Non-flat Spectra. Doctoral dissertation, Department of Psychology, University of Wisconsin-Madison.

Macpherson, Ewan A. and John C. Middlebrooks (2002). Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. *J Acoust Soc Am* 111(5 Pt 1): 2219–36.

Macpherson, Ewan A. and John C. Middlebrooks (2003). Vertical-plane sound localization probed with ripple-spectrum noise. *J Acoust Soc Am* 114(1): 430–45.

Maki, Katuhiro and Shigeto Furukawa (2005). Reducing individual differences in the external-ear transfer functions of the mongolian gerbil. *J Acoust Soc Am* 118(4): 2392–2404.

Masters, W. M., A. J. Moffat, and J. A. Simmons (1985). Sonar tracking of horizontally moving targets by the big brown bat *ptesicus fuscus*. *Science* 228(4705): 1331–1333.

May, B. J. (2000). Role of the dorsal cochlear nucleus in the sound localization behavior of cats. *Hear Res* 148(1-2): 74–87.

May, B. J. and A. Y. Huang (1997). Spectral cues for sound localization in cats: a model for discharge rate representations in the auditory nerve. *J Acoust Soc Am* 101(5 Pt 1): 2705–2719.

McAlpine, David and Benedikt Grothe (2003). Sound localization and delay lines—do mammals fit the model? *Trends Neurosci* 26(7): 347–350.

Meredith, M. A. and H. R. Clemo (1989). Auditory cortical projection from the anterior ectosylvian sulcus (field aes) to the superior colliculus in the cat: an anatomical and electrophysiological study. *J Comp Neurol* 289(4): 687–707.

Metta, Giorgio (2000). Babyrobot: A study on sensori-motor development. Doctoral dissertation, University of Genova, LIRA-Lab, DIST.

Middlebrooks, J. C. (1992). Narrow-band sound localization related to external ear acoustics. *J Acoust Soc Am* 92(5): 2607–24.

Middlebrooks, J. C. (1999). Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *J Acoust Soc Am* 106(3 Pt 1): 1493–1510.

Middlebrooks, J. C. and D. M. Green (1990). Directional dependence of interaural envelope delays. *J Acoust Soc Am* 87(5): 2149–2162.

Middlebrooks, J. C. and D. M. Green (1991). Sound localization by human listeners. *Annu Rev Psychol* 42: 135–59.

Middlebrooks, J. C. and E. I. Knudsen (1984). A neural code for auditory space in the cat's superior colliculus. *J Neurosci* 4(10): 2621–2634.

Middlebrooks, J. C., S. Furukawa L. Xu, and B. J. Mickey (2002). *Integrative functions in the mammalian auditory pathway*, chapter Location signaling by cortical neurons, pp. 160–206. Springer-Verlag, New York.

Middlebrooks, J. C., J. C. Makous, and D. M. Green (1989). Directional sensitivity of sound-pressure levels in the human ear canal. *J Acoust Soc Am* 86(1): 89–108.

Moore, D. R. and D. R. Irvine (1979). A developmental study of the sound pressure transformation by the head of the cat. *Acta Otolaryngol* 87(5-6): 434–440.

Moore, D. R. and A. J. King (2004). *Plasticity of the auditory system*, chapter Plasticity of binaural systems, pp. 96–172. Springer Handbook of Auditory Research. Springer, New York.

Moss, C. F. and H.-U. Schnitzler (1995). *Hearing by Bats*, chapter Behavioral studies of auditory information processing, pp. 87–145. Springer, New York.

Muir, D. W., R. K. Clifton, and M. G. Clarkson (1989). The development of a human auditory localization response: a U-shaped function. *Can J Psychol* 43(2): 199–216.

Muir, Darwin and Sylvia Hains (2004). The u-shaped developmental function for auditory localization. *Journal of cognition and development* 5(1): 123–130.

Musicant, A. D. and R. A. Butler (1984). The psychophysical basis of monaural localization. *Hear Res* 14(2): 185–190.

Obrist, M. K., M. B. Fenton, J. L. Eger, and P. A. Schlegel (1993). What ears do for bats: a comparative study of pinna sound pressure transformation in chiroptera. *J Exp Biol* 180: 119–152.

Oertel, Donata and Eric D Young (2004). What's a cerebellar circuit doing in the auditory system? *Trends Neurosci* 27(2): 104–110.

O'Regan, JK and A Noë (2001). A sensorimotor account of vision and visual consciousness. *Behav Brain Sci* 24(5): 939–73; discussion 973–1031.

Palmer, A. R. and A. J. King (1982). The representation of auditory space in the mammalian superior colliculus. *Nature* 299(5880): 248–249.

- Peña, J. L. and M. Konishi (2001). Auditory spatial receptive fields created by multiplication. *Science* 292(5515): 249–252.
- Peff, T. C. and James A. Simmons (1972). Horizontal-angle resolution by echolocating bats. *J. Acoust. Soc. Amer.* 51(6B): 2063–2065.
- Perales, Mercedes, Jeffery A Winer, and Jorge J Prieto (2006). Focal projections of cat auditory cortex to the pontine nuclei. *J Comp Neurol* 497(6): 959–980.
- Philipona, D., J. K. O’Regan, and J-P. Nadal (2003). Is there something out there? Inferring space from sensorimotor dependencies. *Neural Comput* 15(9): 2029–2049.
- Philipona, David, J. Kevin O’Regan, Jean-Pierre Nadal, and Olivier J.-M.D. Coenen (2004). Perception of the structure of the physical world using multimodal unknown sensors and effectors. *Advances in Neural Information Processing Systems* 15.
- Pickles, James O. (1988). *An introduction to the physiology of hearing*. Academic Press, 2nd edition.
- Poincare, Henri (1929). *The foundations of science; Science and hypothesis, The value of science, Science and method*. The Science Press, New York. George B. Halsted, trans. of *La valeur de la science*. 1905.
- Poincare, Henri (2001). *The Value of Science, Essential Writings of Henir Poincare*. Modern Library Science Series. The Modern Library, New York, New York, NY. Translation.
- Pollak, G. D. and T. J. Park (1995). *Hearing by Bats*, chapter The inferior colliculus, pp. 296–367. Springer, New York, Berlin-Heidelberg.

Popper, A. N. and R. R. Fay, editors (1995). *Hearing by bats*. Springer Handbook of Auditory Research. Springer-Verlag.

Prieur, J-M, C Bourdin, J-L Vercher, F Sarès, J Blouin, and GM Gauthier (2005). Accuracy of spatial localization depending on head posture in a perturbed gravito-inertial force field. *Exp Brain Res* 161(4): 432–40.

Rakerd, B., WM. Hartmann, and TL. McCaskey (1999). Identification and localization of sound sources in the median sagittal plane. *J Acoust Soc Am* 106(5): 2812–2820.

Rice, JJ, BJ May, GA Spirou, and ED Young (1992). Pinna-based spectral cues for sound localization in cat. *Hear Res* 58(2): 132–52.

Roweis, S. T. and L. K. Saul (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500): 2323–2326.

Saff, E. B. and A. B. J. Kuijlaars (1997). Distributing many points on a sphere. *The Mathematical Intelligencer* 19(1): 5–11.

Sanes, D. H. (1990). An in vitro analysis of sound localization mechanisms in the gerbil lateral superior olive. *J Neurosci* 10(11): 3494–3506.

Saul, L. K., K. Q. Weinberger, J. H. Ham, F. Sha, and D. D. Lee (2006). *Semisupervised Learning*, chapter Spectral methods for dimensionality reduction. MIT Press, Cambridge, MA.

Schnitzler, H. U. and O. W. Henson (1980). Performance of airborne animal sonar systems: I. microchiroptera. In Busnel, R. G. and J. E. Fish, editors, *Animal Sonar Systems*, Vol. 28

of *NATO Advanced Study institutes: Series A. Life Sciences*, pp. 109–181. Plenum, New York.

Schnitzler, H. U. and A. D. Grinnell (1977). Directional sensitivity of echolocation in the horseshoe bat, *Rhinolophus ferrumequinum*. i. directionality of sound emission. *J. Comp. Physiol.* 116: 51–61.

Schnupp, Jan W H, John Booth, and Andrew J King (2003). Modeling individual differences in ferret external ear transfer functions. *J Acoust Soc Am* 113(4 Pt 1): 2021–2030.

Schnupp, JW., AJ. King, and S. Carlile (1998). Altered spectral localization cues disrupt the development of the auditory space map in the superior colliculus of the ferret. *J Neurophysiol* 79(2): 1053–1069.

Searle, C. L., L. D. Braida, D. R. Cuddy, and M. F. Davis (1975). Binaural pinna disparity: another auditory localization cue. *J Acoust Soc Am* 57(2): 448–455.

Semple, M. N. and L. M. Kitzes (1987). Binaural processing of sound pressure level in the inferior colliculus. *J Neurophysiol* 57(4): 1130–1147.

Semple, M. N. and L. M. Kitzes (1993). Binaural processing of sound pressure level in cat primary auditory cortex: evidence for a representation based on absolute levels rather than interaural level differences. *J Neurophysiol* 69(2): 449–461.

Sha, Fei and Lawrence K. Saul (2005). Analysis and extension of spectral methods for nonlinear dimensionality reduction. In *ICML '05: Proceedings of the 22nd international conference on Machine learning*, pp. 784–791, New York, NY, USA. ACM Press.

Shaw, E. A. and R. Teranishi (1968). Sound pressure generated in an external-ear replica and real human ears by a nearby point source. *J Acoust Soc Am* 44(1): 240–249.

Shimozawa, T., N. Suga, P. Hendler, and S. Schuetze (1974). Directional sensitivity of echolocation system in bats producing frequency-modulated signals. *J Exp Biol* 60(1): 53–69.

Shimozawa, T., X. Sun, and P. H. Jen (1984). Auditory space representation in the superior colliculus of the big brown bat, *ptesicus fuscus*. *Brain Res* 311(2): 289–296.

Simmons, J. A., M.J. Ferragamo, P. A. Saillant, T. Haresign, J. M. Wotton, S. P. Dear, and D. N. Lee (1995). *Hearing by Bats*, chapter Auditory dimensions of acoustic images in echolocation, pp. 146–190. Springer, New York.

Simmons, J. A., S. A. Kick, C. Hale, C. Bard, and B. Escudie (1983). Acuity of horizontal angle discrimination by the echolocating bats. *J. Comp. Physiol. [A]* 153: 321–330.

Simmons, James A. (1969). Acoustic radiation patterns for the echolocating bats *chilonycteris rubiginosa* and *ptesicus fuscus*. *Journal of the Acoustical Society of America* 46(4B): 1054–1056.

Slattery, W. H. and J. C. Middlebrooks (1994). Monaural sound localization: acute versus chronic unilateral impairment. *Hear Res* 75(1-2): 38–46.

Sparks, David L (2005). An argument for using ethologically "natural" behaviors as estimates of unobservable sensory processes. Focus on "Sound localization performance in the cat: the effect of restraining the head". *J Neurophysiol* 93(3): 1136–7.

Stecker, G. Christopher and John C Middlebrooks (2003). Distributed coding of sound locations in the auditory cortex. *Biol Cybern* 89(5): 341–349.

Stein, B. E. and M. A. Meredith (1993). *The merging of the senses*. MIT Press, Cambridge, MA.

Stevens, S. S. and E. B. Newman (1936). The localization of actual sources of sound. *Am J Psychol* 48(2): 297–306.

Strominger, N. L. and R. E. Oesterreich (1970). Localization of sound after section of the brachium of the inferior colliculus. *J Comp Neurol* 138(1): 1–18.

Sutherland, D. P., R. B. Masterton, and K. K. Glendenning (1998). Role of acoustic striae in hearing: reflexive responses to elevated sound-sources. *Behav Brain Res* 97(1-2): 1–12.

Thornton, S. K. and D. J. Withington (1996). The role of the external nucleus of the inferior colliculus in the construction of the superior collicular auditory space map in the guinea-pig. *Neurosci Res* 25(3): 239–246.

Tollin, D. J., L. C. Populin, J. M. Moore, J. L. Ruhland, and T. C. T. Yin (2005). Sound-localization performance in the cat: the effect of restraining the head. *J Neurophysiol* 93(3): 1223–1234.

Tollin, Daniel J (2003). The lateral superior olive: a functional role in sound source localization. *Neuroscientist* 9(2): 127–143.

Tollin, Daniel J and Tom C T Yin (2002). The coding of spatial location by single units in the lateral superior olive of the cat. ii. the determinants of spatial receptive fields in azimuth. *J Neurosci* 22(4): 1468–1479.

Tsuchitani, C. and J. C. Boudreau (1969). Stimulus level of dichotically presented tones and cat superior olive s-segment cell dcharge. *J Acoust Soc Am* 46(4): 979–988.

Valentine, D. E. and C. F. Moss (1997). Spatially selective auditory responses in the superior colliculus of the echolocating bat. *J Neurosci* 17(5): 1720–1733.

Valentine, Doreen E, Shiva R Sinha, and Cynthia F Moss (2002). Orienting responses and vocalizations produced by microstimulation in the superior colliculus of the echolocating bat, *ptesicus fuscus*. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 188(2): 89–108.

van Adel, B. A. and J. B. Kelly (1998). Kainic acid lesions of the superior olivary complex: effects on sound localization by the albino rat. *Behav Neurosci* 112(2): 432–446.

Van Wanrooij, Marc M and A John Van Opstal (2004). Contribution of head shadow and pinna cues to chronic monaural sound localization. *J Neurosci* 24(17): 4163–71.

Van Wanrooij, Marc M. and A. John Van Opstal (2005). Relearning sound localization with a new ear. *J Neurosci* 25(22): 5413–24.

Vliegen, Joyce, Tom J Van Grootel, and A. John Van Opstal (2005). Gaze orienting in dynamic visual double steps. *J Neurophysiol* 94(6): 4300–4313.

- Vliegen, Joyce, Tom J Van Grootel, and A. John Van Opstal (2004). Dynamic sound localization during rapid eye-head gaze shifts. *J Neurosci* 24(42): 9291–9302.
- Vuillemin, Jules (1972). Poincaré’s philosophy of space. *Synthese* 24(1-2): 161–179.
- Wahba, G. (1981). Spline interpolation and smoothing on the sphere. *SIAM J. Sci. Stat. Comput.* 2(1): 5–16.
- Wang, Jing, Zhenyue Zhang, and Hongyuan Zha (2005). Adaptive manifold learning. In Saul, Lawrence K., Yair Weiss, and Léon Bottou, editors, *Advances in Neural Information Processing Systems 17*, pp. 1473–1480. MIT Press, Cambridge, MA.
- Wanrooij, Marc M Van and A. John Van Opstal (2007). Sound localization under perturbed binaural hearing. *J Neurophysiol* 97(1): 715–726.
- Watkins, A. J. (1978). Psychoacoustical aspects of synthesized vertical locale cues. *J Acoust Soc Am* 63(4): 1152–1165.
- Webster, Frederick Atilla and Oliver G. Brazier (1965). Experimental studies on target detection, evaluation and interception by echolocating bats. Technical report TDR No. AMRL-TR-65-172, Aerospace Medical Division, USAF Systems Command, Tucson, AZ.
- Weinberg, R. J. (1997). Are topographic maps fundamental to sensory processing? *Brain Res Bull* 44(2): 113–116.
- Wexler, Mark (2003). Voluntary head movement and allocentric perception of space. *Psychol Sci* 14(4): 340–346.

Wexler, Mark and Jeroen J A Boxtel (2005). Depth perception by the active observer. *Trends Cogn Sci* 9(9): 431–438.

White, L. E., D. M. Coppola, and D. Fitzpatrick (2001). The contribution of sensory experience to the maturation of orientation selectivity in ferret visual cortex. *Nature* 411(6841): 1049–1052.

Wightman, F. L. and D. J. Kistler (1989a). Headphone simulation of free-field listening. I: Stimulus synthesis. *J Acoust Soc Am* 85(2): 858–867.

Wightman, F. L. and D. J. Kistler (1989b). Headphone simulation of free-field listening. II: Psychophysical validation. *J Acoust Soc Am* 85(2): 868–878.

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

Wightman, F. L. and D. J. Kistler (1997). Monaural sound localization revisited. *J Acoust Soc Am* 101(2): 1050–63.

Wilmington, D., L. Gray, and R. Jahrsdoerfer (1994). Binaural processing after corrected congenital unilateral conductive hearing loss. *Hear Res* 74(1-2): 99–114.

Withington-Wray, D. J., K. E. Binns, S. S. Dhanjal, S. G. Brickley, and M. J. Keating (1990). The maturation of the superior collicular map of auditory space in the guinea pig is disrupted by developmental auditory deprivation. *Eur J Neurosci* 2(8): 693–703.

Withington-Wray, D. J., K. E. Binns, and M. J. Keating (1990). The maturation of the superior collicular map of auditory space in the guinea pig is disrupted by developmental visual deprivation. *Eur J Neurosci* 2(8): 682–692.

Wong, D. (1984). Spatial tuning of auditory neurons in the superior colliculus of the echolocating bat, myotis lucifugus. *Hear Res* 16(3): 261–270.

Wotton, J. M., T. Haresign, M. J. Ferragamo, and J. A. Simmons (1996). Sound source elevation and external ear cues influence the discrimination of spectral notches by the big brown bat, *ptesicus fuscus*. *J Acoust Soc Am* 100(3): 1764–1776.

Wotton, J. M., T. Haresign, and J. A. Simmons (1995). Spatially dependent acoustic cues generated by the external ear of the big brown bat, *ptesicus fuscus*. *J Acoust Soc Am* 98(3): 1423–1445.

Wotton, J. M., R. L. Jenison, and D. J. Hartley (1997). The combination of echolocation emission and ear reception enhances directional spectral cues of the big brown bat, *ptesicus fuscus*. *J Acoust Soc Am* 101(3): 1723–1733.

Wotton, J. M. and J. A. Simmons (2000). Spectral cues and perception of the vertical position of targets by the big brown bat, *ptesicus fuscus*. *J Acoust Soc Am* 107(2): 1034–1041.

Xu, L. and J. C. Middlebrooks (2000). Individual differences in external-ear transfer functions of cats. *J Acoust Soc Am* 107(3): 1451–1459.

Yin, T. C. (2002). *Integrative Functions in the Mammalian Auditory Pathway*, chapter Neural mechanisms of encoding binaural localization cues in the auditory brain stem, pp. 99–159. Springer-Verlag, Heidelberg, Germany.

Young, E. D. and K. A. Davis (2002). *Integrative functions in the mammalian auditory pathway*, chapter Circuitry and function of the dorsal cochlear nucleus, pp. 160–206. Springer-Verlag, New York.

Young, E. D., J. J. Rice, and S. C. Tong (1996). Effects of pinna position on head-related transfer functions in the cat. *J Acoust Soc Am* 99(5): 3064–3076.

Zakarauskas, Pierre and Max S. Cynader (1993). A computational theory of spectral cue localization. *J. Acoust. Soc. Am.* 94: 1323.

Zha, Hongyuan and Zhenyue Zhang (2005). Spectral analysis of alignment in manifold learning. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, Vol. 5, pp. 1069–1072.

Zhang, D. X., L. Li, J. B. Kelly, and S. H. Wu (1998). Gabaergic projections from the lateral lemniscus to the inferior colliculus of the rat. *Hear Res* 117(1-2): 1–12.

Zhang, Zhenyue and Hongyuan Zha (2004). Principal manifolds and nonlinear dimensionality reduction via tangent space alignment zhenyue zhang, hongyuan zha. *SIAM Journal on Scientific Computing* 26(1): 313–338. SIAM: Society for Industrial and Applied Mathematics.

Zwiers, M. P., A. J. Van Opstal, and J. R. Cruysberg (2001a). A spatial hearing deficit in early-blind humans. *J Neurosci* 21(9): RC142: 1–RC142: 5.

Zwiers, M. P., A. J. Van Opstal, and J. R. Cruysberg (2001b). Two-dimensional sound-localization behavior of early-blind humans. *Exp Brain Res* 140(2): 206–22.

Zwiers, Marcel P, Huib Versnel, and A. John Van Opstal (2004). Involvement of monkey inferior colliculus in spatial hearing. *J Neurosci* 24(17): 4145–4156.