Spatial Audio for Bat Biosonar

Hyeon Lee

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Mechanical Engineering

Michael J. Roan, Chair
Ricardo Burdisso
Rolf Müller
Jeffrey H. Reed
Pablo Tarazaga

July 9th, 2020
Blacksburg, Virginia

Keywords: bat biosonar, spatial audio, soundfield microphone, Ambisonics, matched-filter, sound localization, echolocation, HARPEx, signal modeling

Spatial Audio for Bat Biosonar

Hyeon Lee

ABSTRACT (ACADEMIC)

Research investigating the behavioral and physiological responses of bats to echoes typically includes analysis of acoustic signals from microphones and/or microphone arrays, using time difference of arrival (TDOA) between array elements or the microphones to locate flying bats (azimuth and elevation). This has provided insight into transmission adaptations with respect to target distance, clutter, and interference. Microphones recording transmitted signals and echoes near a stationary bat provide sound pressure as a function of time but no directional information.

This dissertation introduces spatial audio techniques to bat biosonar studies as a complementary method to the current TDOA based acoustical study methods. This work proposes a couple of feasible methods based on spatial audio techniques, that both track bats in flight and pinpoint the directions of echoes received by a bat. A spatial audio/soundfield microphone array is introduced to measure sounds in the sonar frequency range (20-80 kHz) of the big brown bat (*Eptesicus fuscus*). The custom-built ultrasonic tetrahedral soundfield microphone consists of four capacitive microphones that were calibrated to match magnitude and phase responses using a transfer function approach. Ambisonics, a signal processing technique used in three-dimensional (3D) audio applications, is used for the basic processing and reproduction of the signals measured by the soundfield microphone. Ambisonics provides syntheses and decompositions of a signal containing its directional properties, using the relationship between the spherical harmonics and the directional properties.

As the first proposed method, a spatial audio decoding technique called HARPEx (High Angular Resolution Planewave Expansion) was used to build a system providing angle and elevation estimates. HARPEx can estimate the direction of arrivals (DOA) for up to two simultaneous sources since it decomposes a signal into two dominant planewaves. Experiments proved that the estimation system based on HARPEx provides

accurate DOA estimates of static or moving sources. It also reconstructed a smooth flight-path of a bat by accurately estimating its direction at each snapshot of pulse measurements in time. The performance of the system was also assessed using statistical analyses of simulations. A signal model was built to generate microphone capsule responses to a virtual source emitting an LFM signal (3 ms, two harmonics: 40-22 kHz and 80-44 kHz) at an angle of 30° in the simulations. Medians and RMSEs (root-mean-square error) of 10,000 simulations for each case represent the accuracy and precision of the estimations, respectively. Results show lower d (distance between a capsule and the soundfield microphone center) or/and higher SNR (signal-to-noise ratio) are required to achieve higher estimator performance. The Cramer-Rao lower bounds (CRLB) of the estimator are also computed with various d and SNR conditions. The CRLB which is for TDOA based methods does not cover the effects of different incident angles to the capsules and signal delays between the capsules due to a non-zero d, on the estimation system. This shows the CRLB is not a proper tool to assess the estimator performance.

For the second proposed method, the matched-filter technique is used instead of HARPEx to build another estimation system. The signal processing algorithm based on Ambisonics and the matched-filter approach reproduces a measured signal in various directions, and computes matched-filter responses of the reproduced signals in time-series. The matched-filter result points a target(s) by the highest filter response. This is a sonar-like estimation system that provides information of the target (range, direction, and velocity) using sonar fundamentals. Experiments using a loudspeaker (emitter) and an artificial or natural target (either stationary or moving) show the system provides accurate estimates of the target's direction and range. Simulations of imitating a situation where a bat emits a pulse and receives an echo from a target (30°) were also performed. The echo sound level is determined using the sonar equation. The system processed the virtual bat pulse and echo, and accurately estimated the direction, range, and velocity of the target. The simulation results also appear to recommend an echo level over -3 dB for accurate and precise estimations (below 15% RMSE for all parameters).

This work proposes two methods to track bats in flight or/and pinpoint the directions of targets using spatial audio techniques. The suggested methods provide accurate estimates of the direction, range, or/and velocity of a bat based on its pulses or of

a target based on echoes. This demonstrates these methods can be used as key tools to reconstruct bat biosonar. They would be also an independent tool or a complementary option to TDOA based methods, for bat echolocation studies. The developed methods are believed to be also useful in improving man-made sonar technology.

Spatial Audio for Bat Biosonar

Hyeon Lee

ABSTRACT (GENERAL AUDIENCE)

While bats are one of the most intriguing creatures to the general population, they are also a popular subject of study in various disciplines. Their extraordinary ability to navigate and forage irrespective of clutter using echolocation has gotten attention from many scientists and engineers. Research investigating bats typically includes analysis of acoustic signals from microphones and/or microphone arrays. Using time difference of arrival (TDOA) between the array elements or the microphones is probably the most popular method to locate flying bats (azimuth and elevation). Microphone responses to transmitted signals and echoes near a bat provide sound pressure but no directional information.

This dissertation proposes a complementary way to the current TDOA methods, that delivers directional information by introducing spatial audio techniques. This work shows a couple of feasible methods based on spatial audio techniques, that can both track bats in flight and pinpoint the directions of echoes received by a bat. An ultrasonic tetrahedral soundfield microphone is introduced as a measurement tool for sounds in the sonar frequency range (20-80 kHz) of the big brown bat (*Eptesicus fuscus*). Ambisonics, a signal processing technique used in three-dimensional (3D) audio applications, is used for the basic processing of the signals measured by the soundfield microphone. Ambisonics also reproduces a measured signal containing its directional properties.

As the first method, a spatial audio decoding technique called HARPEx (High Angular Resolution Planewave Expansion) was used to build a system providing angle and elevation estimates. HARPEx can estimate the direction of arrivals (DOA) for up to two simultaneous sound sources. Experiments proved that the estimation system based on HARPEx provides accurate DOA estimates of static or moving sources. The performance of the system was also assessed using statistical analyses of simulations. Medians and RMSEs (root-mean-square error) of 10,000 simulations for each simulation case represent

the accuracy and precision of the estimations, respectively. Results show shorter distance between a capsule and the soundfield microphone center, or/and higher SNR (signal-to-noise ratio) are required to achieve higher performance.

For the second method, the matched-filter technique is used to build another estimation system. This is a sonar-like estimation system that provides information of the target (range, direction, and velocity) using matched-filter responses and sonar fundamentals. Experiments using a loudspeaker (emitter) and an artificial or natural target (either stationary or moving) show the system provides accurate estimates of the target's direction and range. Simulations imitating a situation where a bat emits a pulse and receives an echo from a target (30°) were also performed. The system processed the virtual bat pulse and echo, and accurately estimated the direction, range, and velocity of the target.

The suggested methods provide accurate estimates of the direction, range, or/and velocity of a bat based on its pulses or of a target based on echoes. This demonstrates these methods can be used as key tools to reconstruct bat biosonar. They would be also an independent tool or a complementary option to TDOA based methods, for bat echolocation studies. The developed methods are also believed to be useful in improving sonar technology.

DEDICATION

Dedication goes to my parents, Youngchun and Younglan Lee, my brother, Hyunho Lee, and my best friend as well as another brother, Matthew Nichols, for their limitless support and love that have made every piece of my achievement possible.

ACKNOWLEDGEMENT

I first give my advisor, Dr. Michael Roan, special and sincere thanks. I still remember my smooth transition between research topics in the middle of my program thanks to his help. His guidance enabled me to become familiar with and intrigued by this new area quickly. He's been around the lab whenever I needed his insight and knowledge, and I have learned a lot by watching his lab work over the shoulder. This provided me an idea of how to be a good hands-on expert for my future career. My academic achievement made in the acoustics field during this Ph.D. program wouldn't have been possible without his support and mentorship. I will miss honest, down-to-earth, and jovial conversations with him on any topic.

I also owe thanks to our colleagues from Brown University, Dr. James Simmons and Dr. Chen Ming, for their wonderful collaboration. With his knowledge and experience, Jim helped me become familiar with the bat, a fascinating and yet still academically unknown to me at the beginning of this research. Chen conducted all experiments involving bats for my research, and she helped design the experiments. Since I know how tedious experiments with any alive animals can be, I really appreciate her time and effort.

Dr. Rolf Müller and Ruihao Wang are other colleagues who deserve my acknowledgement. Dr. Müller allowed me to use his already well-established facilities for bat biosonar studies. Ruihao helped me access the facilities and provided technical assistance using the equipment. This helped build my academic foundation for bat biosonar research from the first stage.

I also would like to thank my doctoral advisory committee members, Dr. Ricardo Burdisso, Dr. Pablo Tarazaga, and Dr. Jeffrey Reed, for their time serving on my committee and insightful advice on my research.

My gratitude also goes to the friends with whom I spent this crucial time of my life. I wouldn't have been able to complete my journey without the time we shared together discussing personal matters, academic matters, celebrating each other's important events, and just having fun.

We were cheerful on great events and frustrated over bad ones. But most importantly, we were together, and we walked through the time together. I won't forget the time I had with all of you here. Now we have different lives at different places, but our friendship will remain forever. Since listing everyone's name is cliché to me, I won't do it here. Instead, I will express my gratitude to all of you for your friendship in person so that you know this is a message for you.

I also would like to thank Cathy Hill, Amanda McCoy, Lauren Mills, and Gail Coe from our remarkable department office crew. Since their excellent assistance and service has eased my graduate school life, I could focus on my academic work. They are dedicated staff, supporting faculty and students tirelessly so that we can each. I believe they are one of the precious assets our department possesses.

My final and greatest gratitude is for my family (parents and brothers) that always stand by me, wish me the best, and make me feel like the happiest and most amazing person in the world. 사랑!

This research was supported by the Multidisciplinary University Research Initiative (MURI) program from the Office of Naval Research (ONR) Grant No. N00014-17-1-2736.

ATTRIBUTION

Chapter 2 was coauthored with Michael Roan (my advisor), Chen Ming, James A. Simmons, Ruihao Wang, and Rolf Müller. All is coworkers in our MURI (Multidisciplinary University Research Initiative) project titled "Bio-inspired Adaptive Sonar for Classification and Guidance in Complex Environments." Every author contributed to the preparation of this manuscript.

Chapter 3 and 4 were coauthored with Michael Roan, Chen Ming, and James A. Simmons. Every author contributed to the preparation of this manuscript.

Chen Ming, Ph.D., is a post-doctoral research associate from the Department of Neuroscience at Brown University. She conducted all experiments involving bats for my research, and helped design the experiments.

James A. Simmons, Ph.D., is a professor from the Department of Neuroscience at Brown University. He provided behavioral, physiological, and acoustical insight about bats and helped design the study, as the PI of the MURI project.

Ruihao Wang, B.S., is a graduate research assistant from the Department of Mechanical Engineering at Virginia Tech. He provided technical assistance on acoustical measurements.

Rolf Müller, Ph.D., is a professor from the Department of Mechanical Engineering at Virginia Tech and a member of my doctoral advisory committee. He provided required facilities and various equipment for acoustical measurements.

Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 1

Introduction

1.1 Bats in Science

There are many fascinating creatures in the animal kingdom. The bat is one that intrigues not only scientists, but laypeople alike in various ways; this is because of their unique features, such as the wing-like long digits connected with a membrane that enable them to fly. The common question of whether the bat is a mammal or bird was risen from this observation. Although it has been verified the bat is a mammal (Altringham, 2011), it is still a popular question among people and students in biology classes. This could be because the bat is the only truly flying mammal, which challenges intuition. The bat has also been frightening to some due to their nocturnality, arguably mean personality, and sometimes hideous appearance (Holland *et al.*, 2010). It actually inspired many vampire works of fiction coinciding with the discovery of vampire bats (Rydell *et al.*, 2018). Seeing how often bats discussed in daily life and the diverse reactions bats draw, we can confidently say it is an enigmatic creature.

Reflecting the high interest in the unique features of bats, there have been ta great number of studies on bats in various aspects. Many scientists have conducted comparison studies with

diverse focuses (e.g. metabolism, anatomy, flight performance, etc.), and between bats and birds as both are capable of flying, even while they fall into different biological classes (Norberg, 1981; Nagy *et al.*, 1999; Hedenström *et al.*, 2009). Hedenstrom et al. found that bats and birds show different wingbeat kinematics because of the difference between bats' long digits with membranes and birds' feathered wings (Hedenström *et al.*, 2009). For instance, bats seem to create more convoluted wakes, which are beneficial for bats' slow flight maneuverability. Bats are also often compared with birds in migration patterns (Altringham, 2011; Krauel and McCracken, 2013). Bats' higher longevity and lighter digestion system are other distinctive features, even compared to other mammals (Herreid II, 1964; Caviedes-Vidal *et al.*, 2008; Ibáñez *et al.*, 2018).

1.2 Bat echolocation

While researchers have studied bats from countless angles for better understanding, as discussed above, bats' extraordinary ability to navigate the air and hunt prey using echolocation is one of the features paid most attention to. As bioinspired technologies have also grown, engineers from various disciplines have attempted to learn from bat biosonar, reproduce its mechanism, and apply it to man-made sonar to improve the technology. Since bats' ability to echolocate objects is literally sound-based, researchers have studied their biosonar/echolocation dynamics in acoustics.

Echolocation is used to locate an object – or objects – of interest using the echoes from the object as the word defines. Echolocating bats emit acoustic waves into the air through their mouth or nose, and receive the echoes from the object using their auditory system (including auditory cortex) to process and analyze the returned signals (echoes). This is a type of active sonar.

1.3 High-frequency of bat echolocation signals

Bats use ultrasonic waves (above the human hearing range) for echolocation. Here are several reasons why bats use high-frequency sonar calls.

First, small insects, which are main prey of insectivorous bats, reflect only high-frequency sounds due to their small size. Since high-frequency calls of bat echolocation have short wavelengths which are in the order of magnitude around the size of insects, the calls enable bats not only to detect small insects (targets), but to discriminate multiple targets from one another or targets from clutter. Higher frequency provides higher resolution of range. In other words, higher frequency echolocation signals enable bats to detect smaller range differences (e.g. ~8.6 mm with 20 kHz vs. ~4.3 mm with 40 kHz) (Simmons *et al.*, 1975; Altringham, 2011). Based on this fact, it has also been believed that small bats use higher frequency sonar calls than larger bats to hunt even smaller prey (Neuweiler, 1980; Altringham, 2011). Jakobsen et al., however, pointed out that small bats would also be capable of detecting smaller prey with lower frequency sounds than expected (Jakobsen *et al.*, 2013). They claimed small bats use higher frequencies not necessarily because of the smaller size of their prey, but because of their smaller emitter size (body size) (Jakobsen *et al.*, 2013). Higher frequencies generate narrower sonar beam-widths, ensuring a higher directivity of the sonar calls towards the target. Since small bats are smaller emitters themselves than are larger bats, they need to use higher frequency calls to achieve a certain level of sonar beam directivity that larger bats can achieve without requiring higher frequency sounds. Since researchers are still working on this matter, and are studying various bats, we expect a solid conclusion in the future.

Another benefit from the use of high-frequency sounds is bats can remove possible acoustical noises, sounds from nature, to their echolocation process (Altringham, 2011). Because

bats avoid using acoustic waves as sonar calls in the frequency range of natural sounds, they can focus solely on their higher frequency sonar calls out of the natural sound frequency range for echolocation. This is the same strategy as a wireless communication technology being assigned an exclusive operation frequency band to avoid signal jamming.

Bats can also minimize interference with each other in echolocation by using high-frequency sonar calls (Altringham, 2011). Since high-frequency sounds attenuate more easily than low-frequency ones, they are not able to travel far. As long as bats keep a sufficient distance between one another, they do not hear distracting sounds from other bats while hunting. Then, a question arises; how can bats fly together in a dense group without crashing (Gillam *et al.*, 2010; Cvikel *et al.*, 2015)? Group flight is impossible if the sonar calls of bats in the group disturb each other's echolocation. Bats actually prefer to communicate with each other using their sonar, but then how does this work? The answer from reports is that bats adjust the frequency range of their sonar calls to avoid signal jamming (Masters *et al.*, 1995; Hiryu *et al.*, 2010). Some studies also showed that bats are actually able to discriminate individuals by their echolocation calls (Yovel *et al.*, 2009; Amichai *et al.*, 2015; Hase *et al.*, 2018) beyond the level of female bats being able to recognize their own offspring (Jones *et al.*, 1991; De Fanis and Jones, 1995). Masters et al. further claimed that bats' echolocation signals not only provide individually identifiable information, but also convey cues as to age, sex, and family affiliation of the individual (Masters *et al.*, 1995). However, those studies were conducted under controlled lab conditions using only few bats. Since the complexity is limitless when analyzing a group of tens of thousands of bats (Gillam *et al.*, 2010; Cvikel *et al.*, 2015; Kloepper *et al.*, 2016), no one yet has a clear idea of how bat biosonar operates for navigation and communication in a densely populated group. As for the types of information bat sonar calls possibly carry about the individual, some studies have agreed with one

another's finding, whereas others have shown disagreement. While answers as to how bat biosonar functions in a dense group and what bat biosonar calls can broadcast about individuals' traits remain undiscovered or controversial, many researchers are working on these matters to find definite answers.

1.4 Effects of bat echolocation circumstance

There have also been studies to find the effects of various circumstances on bat echolocation.

Greif et al. observed bats attempting to drink from panels with smooth surfaces put on the ground (Greif and Siemers, 2010). The bats identified these panels as water's surface because the smooth plates have the same acoustical features. Bats' sonar calls are reflected away, rather than returning to the bat, on the water's surface. Only sound waves reflected in the normal direction of the surface turn back to the bat. This happens when bats fly over a type of stationary waterbody such as a lake or pond, therefore the bats in the study misidentified the smooth plates as water. This so-called 'acoustic mirror' phenomenon explained why bats frequently smash into buildings, a finding in their succeeding study (Greif *et al.*, 2017). Since many modern buildings have smooth surfaces made of glass or metal, which do not send echoes back, bats interpret the buildings as an open pathway. By the time bats receive some echoes traveling perpendicularly to the building surface and detect the obstacle, it is already too late to avoid it.

Weather specifically, and climate more broadly, also affect bat echolocation dynamics since atmospheric conditions including temperature, humidity, pressure, etc., changes atmospheric attenuation and the speed of sound (Griffin, 1971; Evans *et al.*, 1972; Lawrence and Simmons, 1982; Snell-Rood, 2012; Goerlitz, 2018). Acoustic waves attenuate in the air more easily at high-

frequencies (over 30 kHz) that most bat species use for their sonar (Lawrence and Simmons, 1982; Snell-Rood, 2012; Goerlitz, 2018). As ambient temperature and relative humidity increase, atmospheric attenuation increases (Griffin, 1971; Evans *et al.*, 1972; Lawrence and Simmons, 1982), while the speed of sound also increases (Goerlitz, 2018). As an adaptive response to high atmospheric attenuation due to high temperature and humidity, bats adjust their sonar calls by lowering the sonar frequencies or/and extending call duration (Snell-Rood, 2012). The wet season with high temperature and humidity from Snell-Rood's study accompanies lower atmospheric pressure that also leads to higher atmospheric attenuation (Goerlitz, 2018).

Bats' flight speed also affects their echolocation strategy. Bats lower their sonar call frequency as they fly faster, responding to the Doppler shift that changes the frequency of echoes (Schnitzler, 1973). This enables echoes to remain around the frequency bats can hear best (Schuller and Pollak, 1979; Jones and Holderied, 2007). This is also one reason larger bats use lower frequencies than smaller bats alongside longer call durations for their echolocation. Since larger bats fly faster than smaller ones, lower frequency calls are more suitable for larger bats to compensate for the frequency shift of echoes (Doppler shift) and to empower calls to reach farther for longer ranging (Jakobsen *et al.*, 2013).

1.5 Bat echolocation signal characteristics & Bat behaviors

The types of information bats can collect from echoes are probably the most important subject to study among many others to understand the core of bat echolocation. Thus, tremendous efforts have been made to scrutinize the characteristics of bat echolocation signals and their echoes. Researchers have further investigated bats' behavior associated with their echolocation features.

Bat echolocation signals vary depending on the species and the circumstance (Simmons *et al.*, 1975; Altringham, 2011). Although there are many types of bat sonar calls (Melendez *et al.*, 2006), most calls are composed of constant-frequency (CF) or/and frequency-modulated (FM) signals (Simmons *et al.*, 1975). The duration of bat CF signals varies depending on the species; either long (>10 ms) or short. In contrast, bats' FM signals are short (<10 ms) and downward frequency-sweeps (for most bats) with wide bandwidths covering approximately an octave (Simmons *et al.*, 1975; Altringham, 2011). Since these two elemental signals have different characteristics from each other, they deliver different information. FM signals characterized by wide bandwidths are advantageous on spatial localization, and detecting the range and angle of a target (Simmons *et al.*, 1975; Warnecke *et al.*, 2016). FM signals also can convey other features of the target such as shape, size, acoustic impedance, and even surface texture (Simmons *et al.*, 1975). CF signals with a long enough duration are, in contrast, advantageous on detecting relative velocity due to their Doppler intolerance (Simmons *et al.*, 1975; Warnecke *et al.*, 2016). CF signals also tell the existence of a target (Simmons *et al.*, 1975). Because of the unique properties and functions of FM and CF signals, most bats use combinations of those two signals for echolocation such as FM/short-CF, short-CF/FM, and long-CF/FM (Simmons *et al.*, 1975). Note that there are also bats that use only FM signals.

Bats choose the best call types considering various parameters for echolocation in each phase of hunting. In the first, or so-called 'search' phase (or 'detection'), most bats use lower frequency (in their echolocation frequency range), CF signal dominant (except FM bats), and/or long (duration) sonar calls (Parsons *et al.*, 1997; Denny, 2004; Altringham, 2011). Lower frequency calls are more useful searching for a target because these travel farther in range and also have wider beamwidths to cover a wider search area (Jakobsen *et al.*, 2013). Extending call

duration enhances the quality of information gathered from echoes. CF signals are, as discussed previously, more suitable for target detection in the initial stage. In the second 'approach' phase, bats shorten their calls to avoid overlapping between emitting signals and echoes while approaching to a target. Bats also emit echolocation calls more frequently to collect information at a faster rate. Finally, in the 'terminal' (or 'buzz') phase, the duration and pulse rate of sonar calls decreases and increases, respectively, even more than in the approach phase as the bat is closing to the moment of capture (Altringham, 2011).

It has also been reported that bats' hearing sensitivity varries with frequency. Their hearing is most sensitive to the sounds around the center of their echolocation frequency bandwidth which is the center of their sonar receiver operation frequency range (Simmons *et al.*, 1975). This means bats intentionally use a frequency range having the most sensitive frequency at the center for their sonar to ensure high audibility. This is why bats' primary sonar calls are around the center frequency of their operation bandwidth, which is considered the characteristic frequency of the species. Bats' hearing sensitivity also varries with direction (Simmons *et al.*, 1975). Their frequency and direction-dependent hearing sensitivity is another reason bats prefer broadband FM signals for detecting the direction of a target.

Bats' adaptive responses are observed when echoes go off the preferred frequency for their sonar receivers. Simmons reported that a bat (Rhinolophus ferrumequinum) adjusted the frequency of its emitting sonar calls as an adaptive response to a Doppler shift to have echoes around its preferred hearing frequency (83 kHz in the study) (Simmons, 1974; Simmons *et al.*, 1975). As discussed previously, bats respond adaptively to changes of their environment by adjusting sonar frequency and/or call duration. When the echoes from objects are unclear or distorted due to any external disturbances, bats generally increase the amplitude and/or duration of their echolocation

signals to boost the power of the emission signals, enhancing the signal-to-noise ratio (SNR) of echoes (Simmons *et al.*, 1975; Snell-Rood, 2012). This is probably the most basic adaptive response of bats to changes of condition that are unfavorable to their echolocation.

1.6 Motivation & TDOA (Time Difference of Arrival) method

As sonar technology has grown rapidly over the past few decades, sonar signal processing techniques have been developed considerably (Estrada and Starr, 2005). With the improvement of sonar signal processing techniques, the understanding of bats' echolocation mechanism has also significantly grown. Alongside the improved sonar signal processing techniques, sound measurement techniques have been improved with the revolutionary growth of digital technologies (Estrada and Starr, 2005; Hui *et al.*, 2013). This has enabled scientists to collect spatio-temporally more accurate data of bat sonar calls and echoes. However, man-made sonar is not yet able to function as well as bat biosonar. Thus, researchers still study bat biosonar for better understanding, and for developing man-made sonar with higher performance closer to that of bat biosonar. This dissertation is also an effort in the contribution to this mission.

Based on the acoustic signal measurement and processing techniques, TDOA method has been the tool biosonar researchers probably use the most (Arlettaz *et al.*, 2001; Surlykke *et al.*, 2009; Jakobsen *et al.*, 2013; Warnecke *et al.*, 2018). TDOA method requires microphone arrays for sonar measurements. Arrays of sensor elements (microphones) with various arrangements (e.g. cross, t-shaped, etc.) are used for both room (Arlettaz *et al.*, 2001; Surlykke *et al.*, 2009; Barchi *et al.*, 2013; Jakobsen *et al.*, 2013; Wheeler *et al.*, 2014; Knowles *et al.*, 2015; Warnecke *et al.*, 2018) and open field (Barclay, 1985; Jones, 1994; Waters *et al.*, 1995; Jensen and Miller, 1999; Surlykke *et al.*, 2009) measurements. Distance (spatial difference) between two elements is already known,

then the delay of recorded signals (temporal difference) between the elements in a pair is measured. The mathematical relationship between the spatial difference and the temporal difference for the pair draws a hyperbolic plot which is a range of possible sound source locations (Krizman *et al.*, 1997). Deriving hyperbolic plots from relationships of multiple pairs yields the most-overlapped range by the hyperbolic curves. The range (or spot) is the estimated location of the sound source.

Although TDOA method is a straightforward tool, it contains some limits (Jensen and Miller, 1999; Arlettaz *et al.*, 2001; Surlykke *et al.*, 2009; Jakobsen *et al.*, 2013; Warnecke *et al.*, 2018). This method has spatial constraints because the elements of a microphone array require a certain distance between the elements for accurate measurements. The structure of microphone arrays needs to change depending on the layout of the test site. Clutter can also block some microphones from sound, but not the others in the array. As for indoor measurements, the available space microphones can be installed is limited.

Consequently, we introduce hardware and software systems complementary or alternative to TDOA method in this work. These novel systems are sound intensity-based using a soundfield microphone that measures not only acoustic pressure, but also acoustic velocity (Gerzon, 1973; 1980). The acoustic vector information of particle velocity leads to reliable estimations of a target's direction (Nehorai and Paldi, 1994). Measured data is processed using Ambisonics (Gerzon, 1973; 1980) signal processing techniques in the system. Another prominent advantage of this system employing spatial audio techniques is it provides acoustical views directly from bats' perspective. The main purpose of spatial audio techniques is to provide immersive sound to listeners by realistic soundfield reproduction. By the same token, reproduction of sound fields around a bat can provide an acoustical map of its surrounding, which is the main goal of this work. This gives insights of acoustics around the bat, and how the world would look to the bat. This dissertation presents the

development of the hardware and software system using soundfield microphone and Ambisonics, and bat experiments used to validate the system. It demonstrates the possible application of this target detection system to bat biosonar studies.

1.7 Big brown bat (*Eptesicus fuscus*)

Big brown bats in Dr. Simmons group's at Brown University were used for the experiments in this work. *Eptesicus fuscus* is one of FM bat species that solely use FM signals (Simmons *et al.*, 1975). Its FM signals tend to be short in duration and broad in bandwidth (Masters *et al.*, 1995). It emits sonar calls with a basic downward sweep in generally the 50-25 kHz frequency range (typically 1-5 ms) as the first harmonic (Simmons *et al.*, 1975). The sounds contain multiple, up to three or four depending on the bandwidth and duration of the calls, noticeable FM harmonics (Simmons *et al.*, 1995). Bates et al. reported that big brown bats use these characteristic harmonics to discriminate a target from clutter (Bates *et al.*, 2011). Weak echoes of the second FM harmonic from clutter cause delays in neural response, leading bats to defocusing the clutter.

1.8 Dissertation Overview

This dissertation consists of five chapters. The current chapter has covered the fundamentals of bat echolocation, the effects of variables to bat biosonar, and the motivation of this work.

Chapter 2 explains the development of an ultrasonic tetrahedron soundfield microphone for bat echolocation signal measurements. A spatial audio decoding technique called High Angular Resolution Planewave Expansion (HARPEx) is selected alongside Ambisonics and programmed for digital signal processing and direction estimation. The soundfield microphone is the hardware,

and the programmed signal processing and estimation algorithm including Ambisonics and HARPEx is the software of the system. Experiments with artificial sources (loudspeakers) and a natural source (bat) are demonstrations of estimations by the system.

Chapter 3 is devoted to performance assessment of the direction estimation system introduced in chapter 2. A statistical study of simulations is implemented to assess the performance of the estimator. A monopole emitting a chirp at an angle is used for the simulations with two variables. Medians and root-mean-square errors (RMSE) of estimations from simulations for each case represent the accuracy and precision of the estimator performance, respectively. The Cramer-Rao lower bounds (CRLB) is computed and compared to the statistical study result.

Chapter 4 presents another signal processing method to estimate a target's direction using a matched-filter approach which is a fundamental technique in sonar (and radar) technology. Ambisonics signal processing technique reproduces measured sound fields around the measurement point. The matched-filter method analyzes the sound fields and estimates a target's direction, while also providing characteristics of the target. An upgraded soundfield microphone that can measure ultrasonic waves up to 80 kHz is introduced.

Chapter 5 summarizes the achievements and contributions of this work. Limitations of this study and suggested future work are also discussed.

References

Altringham, J. D. (**2011**). *Bats: from evolution to conservation* (Oxford University Press).

Amichai, E., Blumrosen, G., and Yovel, Y. (**2015**). "Calling louder and longer: how bats use biosonar under severe acoustic interference from other bats," Proceedings of the Royal Society B: Biological Sciences **282**, 20152064.

Arlettaz, R., Jones, G., and Racey, P. A. (**2001**). "Effect of acoustic clutter on prey detection by bats," Nature **414**, 742-745.

Barchi, J. R., Knowles, J. M., and Simmons, J. A. (**2013**). "Spatial memory and stereotypy of flight paths by big brown bats in cluttered surroundings," Journal of Experimental Biology **216**, 1053-1063.

Barclay, R. M. (**1985**). "Long-versus short-range foraging strategies of hoary (Lasiurus cinereus) and silver-haired (Lasionycteris noctivagans) bats and the consequences for prey selection," Canadian Journal of Zoology **63**, 2507-2515.

Bates, M. E., Simmons, J. A., and Zorikov, T. V. (**2011**). "Bats use echo harmonic structure to distinguish their targets from background clutter," Science **333**, 627-630.

Caviedes-Vidal, E., Karasov, W. H., Chediack, J. G., Fasulo, V., Cruz-Neto, A. P., and Otani, L. (**2008**). "Paracellular absorption: a bat breaks the mammal paradigm," PLoS One **3**, e1425.

Cvikel, N., Berg, K. E., Levin, E., Hurme, E., Borissov, I., Boonman, A., Amichai, E., and Yovel, Y. (**2015**). "Bats aggregate to improve prey search but might be impaired when their density becomes too high," Current Biology **25**, 206-211.

De Fanis, E., and Jones, G. (**1995**). "Post-natal growth, mother-infant interactions and development of vocalizations in the vespertilionid bat Plecotus auritus," Journal of Zoology **235**, 85-97.

Denny, M. (**2004**). "The physics of bat echolocation: signal processing techniques," American Journal of Physics **72**, 1465-1477.

Estrada, R. F., and Starr, E. A. (**2005**). "50 years of acoustic signal processing for detection: Coping with the digital revolution," IEEE Annals of the History of Computing **27**, 65-78.

Evans, L., Bass, H., and Sutherland, L. (**1972**). "Atmospheric absorption of sound: Theoretical predictions," The Journal of the Acoustical Society of America **51**, 1565-1575.

Gerzon, M. A. (**1973**). "Periphony: With-height sound reproduction," Journal of the Audio Engineering Society **21**, 2-10.

Gerzon, M. A. (**1980**). "Practical periphony: The reproduction of full-sphere sound," in *Audio Engineering Society Convention 65* (Audio Engineering Society).

Gillam, E. H., Hristov, N. I., Kunz, T. H., and McCracken, G. F. (**2010**). "Echolocation behavior of Brazilian free-tailed bats during dense emergence flights," Journal of Mammalogy **91**, 967-975.

Goerlitz, H. R. (**2018**). "Weather conditions determine attenuation and speed of sound: Environmental limitations for monitoring and analyzing bat echolocation," Ecology and evolution **8**, 5090-5100.

Greif, S., and Siemers, B. M. (**2010**). "Innate recognition of water bodies in echolocating bats," Nature communications **1**, 1-6.

Greif, S., Zsebők, S., Schmieder, D., and Siemers, B. M. (**2017**). "Acoustic mirrors as sensory traps for bats," Science **357**, 1045-1047.

Griffin, D. R. (**1971**). "The importance of atmospheric attenuation for the echolocation of bats (Chiroptera)," Animal Behaviour **19**, 55-61.

Hase, K., Kadoya, Y., Maitani, Y., Miyamoto, T., Kobayasi, K. I., and Hiryu, S. (**2018**). "Bats enhance their call identities to solve the cocktail party problem," Communications biology **1**, 1-8.

Hedenström, A., Johansson, L. C., and Spedding, G. R. (**2009**). "Bird or bat: comparing airframe design and flight performance," Bioinspiration & Biomimetics **4**, 015001.

Herreid II, C. F. (**1964**). "Bat longevity and metabolic rate," Experimental gerontology **1**, 1-9.

Hiryu, S., Bates, M. E., Simmons, J. A., and Riquimaroux, H. (**2010**). "FM echolocating bats shift frequencies to avoid broadcast–echo ambiguity in clutter," Proceedings of the National Academy of Sciences **107**, 7048-7053.

Holland, R. A., Borissov, I., and Siemers, B. M. (**2010**). "A nocturnal mammal, the greater mouse-eared bat, calibrates a magnetic compass by the sun," Proceedings of the National Academy of Sciences **107**, 6941-6945.

Hui, A., Kursell, J., and Jackson, M. W. (**2013**). "Music, Sound, and the Laboratory from 1750 to 1980," Osiris **28**, 1-11.

Ibáñez, C., Novella-Fernandez, R., Alonso, P., and Agirre-Mendi, P. T. (**2018**). "SHORT NOTE New longevity record for the Mediterranean horseshoe bat (Rhinolophus euryale," Journal of Bat Research & Conservafion Volume **11**, 1.

Jakobsen, L., Ratcliffe, J. M., and Surlykke, A. (**2013**). "Convergent acoustic field of view in echolocating bats," Nature **493**, 93-96.

Jensen, M. E., and Miller, L. A. (**1999**). "Echolocation signals of the bat Eptesicus serotinus recorded using a vertical microphone array: effect of flight altitude on searching signals," Behavioral Ecology and Sociobiology **47**, 60-69.

Jones, G. (**1994**). "Scaling of wingbeat and echolocation pulse emission rates in bats: why are aerial insectivorous bats so small?," Functional Ecology, 450-457.

Jones, G., and Holderied, M. W. (**2007**). "Bat echolocation calls: adaptation and convergent evolution," Proceedings of the Royal Society B: Biological Sciences **274**, 905-912.

Jones, G., Hughes, P., and Rayner, J. (**1991**). "The development of vocalizations in Pipistrellus pipistrellus (Chiroptera: Vespertilionidae) during post-natal growth and the maintenance of individual vocal signatures," Journal of Zoology **225**, 71-84.

Kloepper, L. N., Linnenschmidt, M., Blowers, Z., Branstetter, B., Ralston, J., and Simmons, J. A. (**2016**). "Estimating colony sizes of emerging bats using acoustic recordings," Royal Society open science **3**, 160022.

Knowles, J. M., Barchi, J. R., Gaudette, J. E., and Simmons, J. A. (**2015**). "Effective biosonar echo-to-clutter rejection ratio in a complex dynamic scene," The Journal of the Acoustical Society of America **138**, 1090-1101.

Krauel, J. J., and McCracken, G. F. (**2013**). "Recent advances in bat migration research," in *Bat evolution, ecology, and conservation* (Springer), pp. 293-313.

Krizman, K. J., Biedka, T. E., and Rappaport, T. S. (**1997**). "Wireless position location: fundamentals, implementation strategies, and sources of error," in *1997 IEEE 47th Vehicular Technology Conference. Technology in Motion* (IEEE), pp. 919-923.

Lawrence, B. D., and Simmons, J. A. (**1982**). "Measurements of atmospheric attenuation at ultrasonic frequencies and the significance for echolocation by bats," The Journal of the Acoustical Society of America **71**, 585-590.

Masters, W. M., Raver, K. A., and Kazial, K. A. (**1995**). "Sonar signals of big brown bats, Eptesicus fuscus, contain information about individual identity, age and family affiliation," Animal Behaviour **50**, 1243-1260.

Melendez, K. V., Jones, D. L., and Feng, A. S. (**2006**). "Classification of communication signals of the little brown bat," The journal of the Acoustical Society of America **120**, 1095-1102.

Nagy, K. A., Girard, I. A., and Brown, T. K. (**1999**). "Energetics of free-ranging mammals, reptiles, and birds," Annual review of nutrition **19**, 247-277.

Nehorai, A., and Paldi, E. (**1994**). "Acoustic vector-sensor array processing," IEEE Transactions on signal processing **42**, 2481-2491.

Neuweiler, G. (**1980**). "How bats detect flying insects," PhT **33**, 34-40.

Norberg, U. M. (**1981**). "Allometry of bat wings and legs and comparison with birds wings," Philosophical Transactions of the Royal Society of London. B, Biological Sciences **292**, 359-398.

Parsons, S., Thorpe, C. W., and Dawson, S. M. (**1997**). "Echolocation calls of the long-tailed bat: a quantitative analysis of types of calls," Journal of mammalogy **78**, 964-976.

Rydell, J., Eklöf, J., and Riccucci, M. (**2018**). "Cimetière du Père-Lachaise. Bats and vampires in French Romanticism," Journal of Bat Research & Conservafion Volume **11**, 1.

Schnitzler, H.-U. (**1973**). "Control of Doppler shift compensation in the greater horseshoe bat, Rhinolophus ferrumequinum," Journal of comparative physiology **82**, 79-92.

Schuller, G., and Pollak, G. (**1979**). "Disproportionate frequency representation in the inferior colliculus of Doppler-compensating greater horseshoe bats: evidence for an acoustic fovea," Journal of Comparative Physiology **132**, 47-54.

Simmons, J. A. (**1974**). "Response of the Doppler echolocation system in the bat, Rhinolophus ferrumequinum," The Journal of the Acoustical Society of America **56**, 672-682.

Simmons, J. A., Howell, D. J., and Suga, N. (**1975**). "Information content of bat sonar echoes: recent research on echolocation in bats identifies some of the kinds of information conveyed by echoes of their sonar sounds," American Scientist **63**, 204-215.

Simmons, J. A., Saillant, P. A., Wotton, J. M., Haresign, T., Ferragamo, M. J., and Moss, C. F. (**1995**). "Composition of biosonar images for target recognition by echolocating bats," Neural Networks **8**, 1239-1261.

Snell-Rood, E. C. (**2012**). "The effect of climate on acoustic signals: does atmospheric sound absorption matter for bird song and bat echolocation?," The Journal of the Acoustical Society of America **131**, 1650-1658.

Surlykke, A., Boel Pedersen, S., and Jakobsen, L. (**2009**). "Echolocating bats emit a highly directional sonar sound beam in the field," Proceedings of the Royal Society B: Biological Sciences **276**, 853-860.

Warnecke, M., Falk, B., and Moss, C. F. (**2018**). "Echolocation and flight behavior of the bat Hipposideros armiger terasensis in a structured corridor," The Journal of the Acoustical Society of America **144**, 806-813.

Warnecke, M., Lee, W.-J., Krishnan, A., and Moss, C. F. (**2016**). "Dynamic echo information guides flight in the big brown bat," Frontiers in behavioral neuroscience **10**, 81.

Waters, D. A., Rydell, J., and Jones, G. (**1995**). "Echolocation call design and limits on prey size: a case study using the aerial-hawking bat Nyctalus leisleri," Behavioral Ecology and Sociobiology **37**, 321-328.

Wheeler, A., Kloepper, L., Matsuo, I., and Simmons, J. A. (**2014**). "Temporal dynamics of echolocation during clutter navigation in Eptesicus fuscus," The Journal of the Acoustical Society of America **135**, 2208-2208.

Yovel, Y., Melcon, M. L., Franz, M. O., Denzinger, A., and Schnitzler, H.-U. (**2009**). "The voice of bats: how greater mouse-eared bats recognize individuals based on their echolocation calls," PLoS Comput Biol **5**, e1000400.

Chapter 2

High-Frequency Soundfield Microphone for the Analysis of Bat Biosonar

Hyeon Lee[1*], Michael J. Roan[1], Chen Ming[2], James A. Simmons[2], Ruihao Wang[1], Rolf Müller[1]

[1]Department of Mechanical Engineering, Virginia Tech, 460 Old Turner Street, Blacksburg, VA 24061, USA

[2]Department of Neuroscience, Brown University, 185 Meeting Street, Providence, RI 02912, USA

The Journal of the Acoustical Society of America 146, no. 6 (2019): 4525-4533.

Abstract

Numerous bat species emit wideband frequency-modulated (FM) signals for echolocation to hunt prey and avoid obstacles. Research investigating the behavioral and physiological responses of bats to echoes typically includes analysis of acoustic signals from microphones and/or microphone arrays, using time difference of arrival (TDOA) between array elements or the microphones to locate flying bats (azimuth and elevation). This has provided insight into transmission adaptations such as pulse duration and duty cycle with respect to target distance, clutter, and interferers. Microphones recording transmitted signals and echoes near a stationary bat provide sound pressure as a function of time but no directional information. In this work, we propose a spatial audio/soundfield microphone array to both track bats in flight and pinpoint the directions of echoes received by a bat. We introduce an ultrasonic (20-80 kHz) tetrahedral soundfield microphone to capture bat sounds up to 80 kHz. A spatial audio decoding technique called HARPEx (High Angular Resolution Planewave Expansion) supplies angle and elevation estimates, either for a flying bat based on the bat pulses or for targets based on echoes. Experiments using the soundfield microphone and HARPEx show that the approach accurately estimates the sound direction of arrival in both scenarios.

Keywords: soundfield microphone; echolocation; HARPEx; angle estimation; Ambisonics; spatial audio; biosonar.

2.1 Introduction

Numerous bat species emit wideband frequency-modulated (FM) signals for echolocation in the context of hunting prey and obstacle avoidance. Research investigating the behavioral and physiological response of bats to echo returns typically includes analysis of acoustic signals from microphones and/or microphone arrays. This has given insight into transmission adaptations such as pulse duration and duty cycle with respect to target distance, clutter, and interferers. In scenarios where the bat is in flight, microphone arrays are a widely used method to locate bats in flight. Typically, the time difference of arrival (TDOA) between array elements is used to estimate the azimuth and elevation of a bat (Arlettaz *et al.*, 2001; Surlykke *et al.*, 2008; Jakobsen *et al.*, 2013; Warnecke *et al.*, 2018). This method has been used to record flying bats in rooms (Arlettaz *et al.*, 2001; Surlykke *et al.*, 2008; Barchi *et al.*, 2013; Jakobsen *et al.*, 2013; Wheeler *et al.*, 2014; Knowles *et al.*, 2015; Warnecke *et al.*, 2018) as well as in open spaces (Barclay, 1985; Jones, 1994; Waters *et al.*, 1995; Jensen and Miller, 1999; Surlykke *et al.*, 2008). The layout of the test site or the interests of the researchers dictate the placement of the microphone array (Jensen and Miller, 1999; Arlettaz *et al.*, 2001; Surlykke *et al.*, 2008; Jakobsen *et al.*, 2013; Warnecke *et al.*, 2018). Arrangements of microphones within the array have included cross and t-shaped arrays with array element numbers ranging from 6-12 microphones (Fawcett *et al.*, 2015).

In this work, we introduce a hardware and software system complementary to the larger microphone arrays that use TDOA measurements, as described above. Our system relies on the soundfield microphone and the related Ambisonics processing (Gerzon, 1973; 1980). Use of a single soundfield microphone can provide estimates of the direction of a target. Combining the direction estimates from two soundfield microphones can yield estimates of the complete position of a sound source, e.g., direction and distance. To demonstrate the target direction finding

capability, we developed a 1st order soundfield microphone capable of localizing sounds from 20 kHz to 80 kHz. Commercially available soundfield microphones can only operate up to 20 kHz. This microphone produces four channels of data that are converted to B-format (Gerzon, 1973; 1980). The B-format signal contains the spatial and temporal properties of the received signal. To analyze the received signal, we chose the HARPEx (High Angular Resolution Planewave Expansion) algorithm (Barrett and Berge, 2010; Berge and Barrett, 2010). The outputs from HARPEx are the directions and amplitudes of two plane waves. These are calculated in each time-frequency bin. We use the HARPEx algorithm to process overlapping time blocks of time-series data (the B-format signal) and produce a movie-like display that provides the direction of arrival of sources, their relative magnitudes, and an indication of their frequency content all as a function of time. We propose the use of this tool to spatio-temporally analyze the echoes that the bat receives in parallel with high-speed video of the bat. In this way, the bat's physical adaptations (ear shape, head tilt, etc.) can be matched to the soundfield around the bat's head.

The layout of the paper is as follows. The operating principles of 1st order soundfield microphones and Ambisonics are introduced first. Second, we provide a review of the decomposition part of HARPEx, and construction details of the ultrasonic soundfield microphone. And last, we provide experimental results confirming the operation of the system.

2.2 Soundfield microphones and Ambisonics

Gerzon introduced the soundfield microphone/Ambisonics for audio applications in the early 1970's (Gerzon, 1973; 1980). In the audio engineering realm, soundfield microphones are used to capture 3D audio, and then reproduce the sound using a high-density loudspeaker array.

The 1st order soundfield microphone is the most straight-forward (Fig. 2.1). It consists of four cardioid capsules mounted on the faces of an equilateral tetrahedron. This arrangement places the capsule elements at four points on the surface of a sphere. The four capsules each have a cardioid response pattern. Simple combinations of the individual capsule signals create figure-of-eight microphone receive patterns that align with the Cartesian axes (Eqs. (2.1a)-(2.1d)). The omnidirectional signal W (sum of all capsules) and the three figure-of-eight (X, Y, Z) outputs are the four-channel B-format signals. Additionally, the B-format signal derived from this arrangement is also the decomposition of the soundfield into the first four spherical harmonics (Fig. 2.1). The four microphone capsules are labeled LFU (Left Front Up), RFD (Right Front Down), LBD (Left Back Down), and RBU (Right Back Up) as shown in Fig. 2.2(a). (Arteaga, 2015; Ortolani, 2015) provide a detailed discussion of Ambisonics. The spherical harmonics are calculated in this paper using the following convention:

$$W = LFU + RFD + LBD + RBU, \qquad (2.1a)$$

$$X = LFU + RFD - LBD - RBU, \qquad (2.1b)$$

$$Y = LFU - RFD + LBD - RBU, \qquad (2.1c)$$

$$Z = LFU - RFD - LBD + RBU. \qquad (2.1d)$$

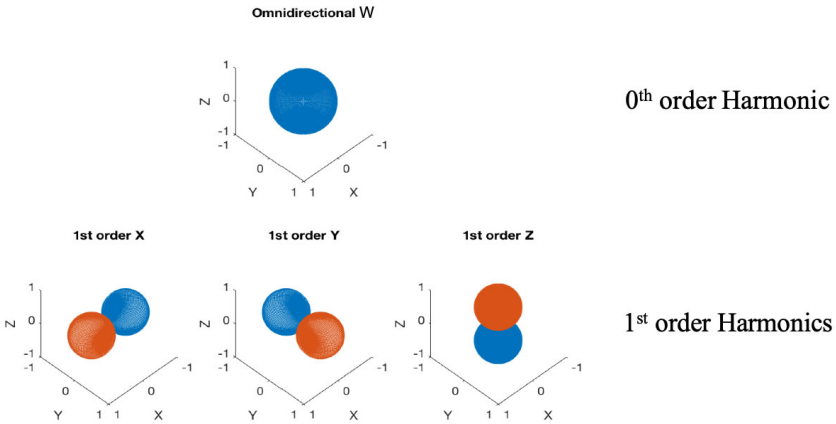


Figure 2.1. First order spherical harmonics.

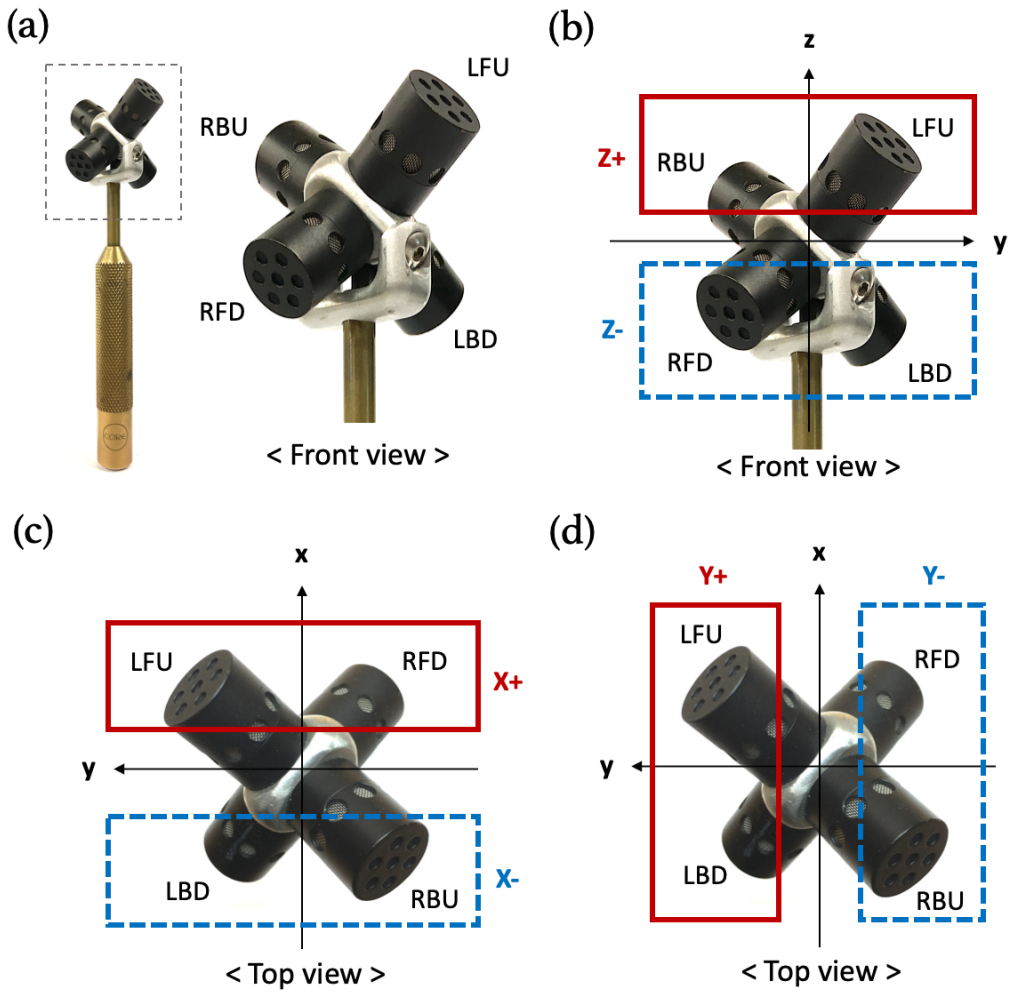


Figure 2.2. Audio signals in the first order Ambisonics A-format are acquired via a simple process of addition and subtraction of original signals recorded through a tetrahedral soundfield microphone such as (a) composed of four microphone capsules labeled LFU (left front up), RFD (right front down), LBD (left back down), and RBU (right back up), respectively. The model is a commercial soundfield microphone (Core Sound TetraMic). (b) Audio signal in z-axis component, Z = LFU – RFD – LBD + RBU, (c) Audio signal in x-axis component, X = LFU + RFD – LBD – RBU, (d) Audio signal in y-axis component, Y = LFU – RFD + LBD – RBU.

The relationship between the 1st order harmonics and the directional properties of a signal are provided in Eq. (2.2), where θ is azimuth and φ is elevation.

$$W(t) = \frac{s(t)}{\sqrt{2}}, \quad (2.2a)$$

$$X(t) = s(t)\cos\theta cos\varphi, \quad (2.2b)$$

$$Y(t) = s(t)sin\theta cos\varphi, \quad (2.2c)$$

$$Z(t) = s(t)\sin\varphi. \quad (2.2d)$$

Equations (2.2a)-(2.2d) show that we can create the 1st order Ambisonic B-format channels (W, X, Y, and Z) for a time-series, *s(t)*. The azimuth and elevation are chosen where we would like to place the signal and calculate the B-format using Eqs. (2.2b)-(2.2d). Conversely, given W, X, Y, and Z, we can calculate the azimuth and elevation at each time sample. These are the main synthesis/decomposition principles of Ambisonics. Another observation from Eqs. (2.2a)-(2.2d) is that this approach can analyze only one source at a given time instant. In order to achieve better resolution, we decided to use the HARPEx algorithm discussed below.

2.3 Analysis technique (direction-of-arrival estimation using HARPEx)

2.3.1 Principle of HARPEx

HARPEx is a spatial-sound decoding technique introduced by Berge and Barrett (Barrett and Berge, 2010; Berge and Barrett, 2010) to improve the angular resolution of Ambisonics-based spatial sound systems. The technique decomposes a B-format signal into two dominant planewaves containing their directional information and amplitudes at each time and frequency. In cases where two simultaneous signals arrive at the soundfield microphone from different directions, HARPEx provides angle and elevation estimates to both sources. This works even if the two signals have the same frequency content. For our application, this is important as it is possible that the echoes triggered by a bat's transmission can arrive simultaneously from several directions, e.g., from a target and clutter.

To implement HARPEx, the B-format signal is converted into the frequency domain using the DFT. Equation (2.3) gives the decomposition of the frequency domain B-format signal now stacked into a matrix Ψ.

$$\begin{matrix} \Psi & & V & & A \end{matrix}$$
$$\begin{bmatrix} \sqrt{2}w_r & \sqrt{2}w_i \\ x_r & x_i \\ y_r & y_i \\ z_r & z_i \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ x_1 & x_2 \\ y_1 & y_2 \\ z_1 & z_2 \end{bmatrix} \begin{bmatrix} a_{1r} & a_{1i} \\ a_{2r} & a_{2i} \end{bmatrix}, \qquad (2.1)$$

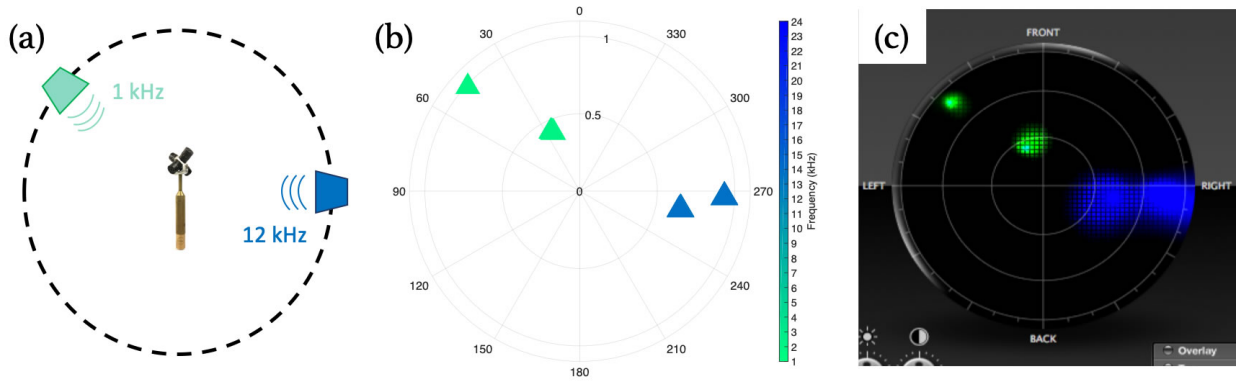
where subscript *r* and *i* denote real and imaginary parts of the signal in the frequency domain, respectively. *V* is a matrix containing unit vectors pointing in the direction of 2 arriving plane waves and *A* is a matrix containing the source amplitudes. We calculate V and A, using a QR decomposition for each time sample and frequency (Barrett and Berge, 2010).

2.3.2 Validation of the estimation system

The HARPEx algorithm discussed above was implemented in MATLAB. In order to validate the implementation, several controlled experiments were conducted in an anechoic chamber. We conducted both static and moving source experiments.

Figure 2.3 (a) gives the experimental arrangement for the static tests. This is the top-down view of the anechoic chamber (located in the ASPIRe lab at Virginia Tech). A commercial soundfield microphone (Core Sound TetraMic from Fig. 2.2) captured audio signals transmitted from two loudspeakers (JBL LSR 308). Azimuths of the loudspeakers were 45° (azimuth) from the front (0 degree) of the soundfield microphone, and 270° (90° to clockwise) from the microphone front (Fig. 2.3(a)). The speakers were both located 1.5 m away from the microphone. They transmitted tones at 1 kHz and at 12 kHz respectively. A 4-channel audio recorder (Roland R-44E field recorder) captured the microphone signals at a sampling rate of 48 kHz. The audio

recording was processed using HARPEx in 100 ms time blocks with 90% overlap. Figure 2.3(b) shows that HARPEx provided accurate direction estimates to the two simultaneous sources (green and blue triangles). Figures 2.3(b) and 2.3(c) are top-down views of a hemisphere with marker symbols (estimated sound source directions). Symbols that are closer to the center are above or below the horizontal. Sources near the edge are nearly in the horizontal plane. In Fig. 2.3(b), triangles indicate signals arriving from below the horizontal. In Fig. 2.3(c), cross hatching indicates the same. The size of the symbols indicates the relative intensity of the sound to each other. The frequency of each sound source was also correct based on the color of the symbols. To test the result further, we processed the signal using a commercially available plugin of the HARPEx algorithm. The estimates from our implementation agreed closely with this commercial version (Fig. 2.3(c)). Although there were only two sources present, HARPEx indicates the presence of 4 sources. This is due to a reflected path arriving almost directly from beneath the array.



Experimental setup HARPEx result Commercial HARPEx

Figure 2.3. (a) Two loudspeakers transmitting sound at 1 kHz and at 12 kHz, respectively, were simultaneously recorded by a commercial soundfield microphone in an anechoic chamber. (b) The direction estimation system using the HARPEx algorithm accurately detected the direction of each sound including its frequency. (c) The direction estimation result showed an agreement with the result derived by the commercial HARPEx software.

The next step in algorithm verification was to run the software on signals from moving sources. For this purpose, synthetic signals were generated using an audio VST plugin called AMBIx (Fig. 2.4). This plugin generates B-format signals (Eq. (2.2) above) for two moving sources. For this test, the sound sources (yellow spheres in Fig. 2.4) were both the same white noise signal. The sources moved counterclockwise with a speed of 20 degree/s. Each one departed from approximately a 130° and 270° (90° to clockwise) angle, respectively, without any elevation. The view in Fig. 2.4 is from above the space. **Error! Reference source not found.** gives the positions of the synthetic sound sources (Figs. 2.4 (a), (c), (e), (g)) and positions of the sources produced by HARPEx (Figs. 2.4 (b), (d), (f), (h)) at four different snapshots in time. Again, HARPEx gave accurate results of the source positions as a function of time. It is important to note that both sources transmitted the exact same noise sequence and both arrived at the microphone at the exact same time. Only the angle of arrival was different. These results are for frequencies below 20 kHz. The next step was to develop a soundfield microphone that operates in the bat's frequency range (10-100kHz) and test our implementation of the HARPEx algorithm at these much higher frequencies.

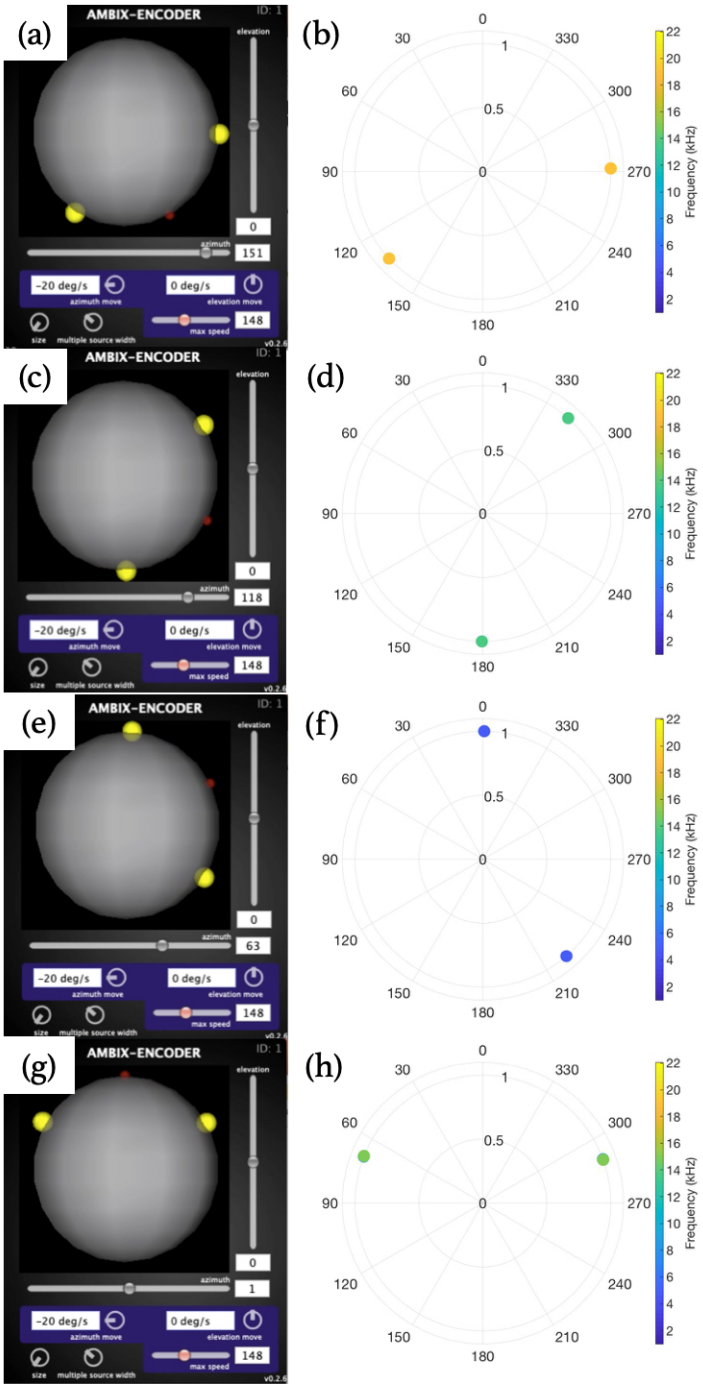


Figure 2.4. Two artificially generated moving sound sources (white noise orbiting the center counterclockwise) were analyzed by the direction estimation system using the HARPEx algorithm. The estimation system accurately tracked the directions of the dynamic sound sources retaining their speed and the angle gap between them. Figures (a)-(h) display the pairs of sound sources and the direction estimation result at four different points in time.

2.4 Ultrasonic soundfield microphone development

2.4.1 Building process and features of the microphone

There is no commercially available ultrasonic soundfield microphone that operates in the bat's frequency range of 10-100 kHz. The authors have constructed a custom soundfield microphone for ultrasonic sound field recordings using four capacitive microphones (Knowles FG-23629-P16 series). The capsules were first calibrated for matching amplitude and phase responses in the anechoic chamber (discussed below). Once calibrated, the capsules were bonded into a 3D printed holder that placed the capsules on the faces of a tetrahedron with the capsule openings approximately 4 mm apart (Fig. 2.8). This gave the first iteration of the microphone an upper frequency range of roughly 40 kHz.

2.4.2 Microphone calibration

Direction estimation using a soundfield microphone depends on the individual capsules having the same far-field magnitude and phase response. Since the Knowles microphone capsules and associated preamplifiers/amplifiers were found to be quite different in their responses, the microphones were calibrated using a transfer function approach. The transfer functions between one capsule and the other three were estimated by driving the microphones with both white noise and a long (5 second duration) FM sweep from 20 kHz to 90 kHz.

Before the calibration test, the signal transmission performance of a capacitive ultrasonic speaker (SensComp 600 Series open face ultrasonic sensor) which was used as the emitter of audio signals in the calibration test (Fig. 2.5(a)) was verified. A measurement microphone (G.R.A.S 46DP-1, Fig. 2.5(a), positioned approximately 15 cm away from the speaker) recorded an FM signal (10-100 kHz, 10 second) from the SensComp transmitter, and an analog-to-digital

conversion board (NI 9222, National Instruments Corp) acquired the data. Sound pressure level was computed across the frequency range and plotted in the dB scale relative to a reference pressure of 20 μPa. RMS pressure values were calculated in 100 data point frequency-bins with 90% overlap in the frequency domain. The frequency response of the speaker is by no means spectrally flat in 50-100 kHz range, but it was able to emit energy across the band of interest of 20-100 kHz (Fig. 2.5(b)). The trend of this result also agreed with the frequency response curve provided by the manufacturer.

A calibration holder was 3D printed to hold the four capsules in a plane ensuring an identical distance (approximately 15 cm) between the capsules and the transmitter (Figs. 2.6(a)-(b)), and to record a signal through the four capsules simultaneously. A NI PCIe6351 Daq (National Instruments Corp.) was used to simultaneously transmit and record the signal. The sampling frequency for both digital to analog and analog to digital conversion was 300 kHz. The SensComp transmitter emitted an FM signal (20-90 kHz, 5 second). The measurement result shows that the capsules have significantly different responses across the test frequency range from each other (Figs. 2.6(c)-(d)). We set the first microphone as the reference and calculated transfer functions between it and the remaining three capsules (Figs. 2.7(a)-(b)). The magnitude and phase of the transfer functions illustrate how different the capsules are from one another in amplitude (Fig. 2.7(a)) and phase responses (Fig. 2.7(b)). The resulting transfer functions were used to correct the magnitude and phases of capsules 2-4 relative to capsule 1 (Fig. 2.7(c)). After calibration, the capsules were bonded and assembled into a tetrahedral soundfield microphone (Fig. 2.8(b)). A custom power supply was built to operate the microphone (Fig. 2.8(a)).

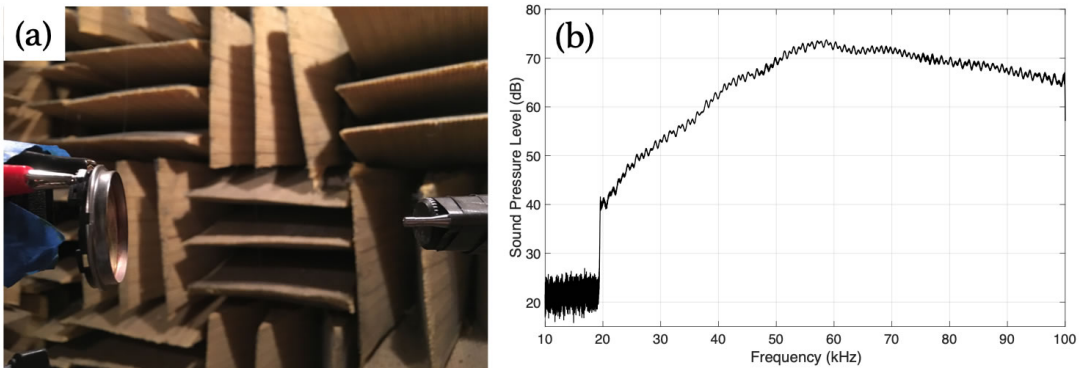


Figure 2.5. (a) A SensComp ultrasonic transmitter and a G.R.A.S 46DP-1 standard microphone in the anechoic chamber. (b) The frequency response of the transmitter emitting 10-100 kHz FM sweep (10 second) as recorded by the standard mic. The transmitter response was clear in 20-100 kHz range.

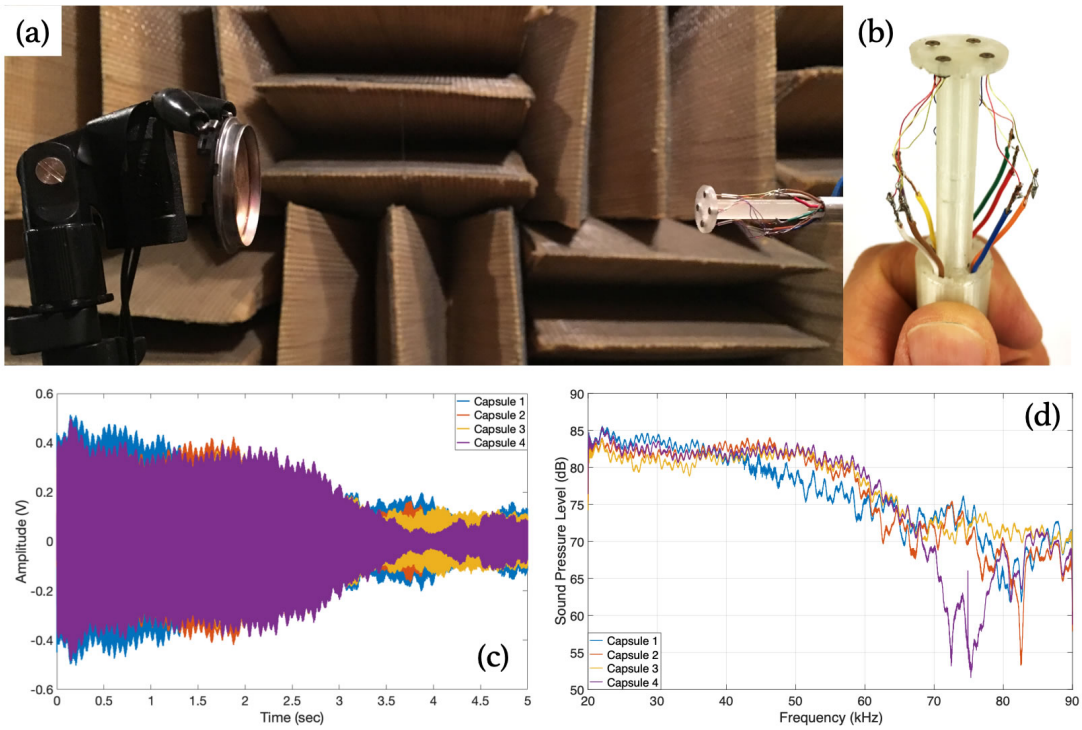


Figure 2.6. (a) The ultrasonic transmitter emitted a 20-90 kHz FM signal in the anechoic chamber. (b) The four capsules attached to a planar calibration holder were put approximately 15 cm from the transmitter and recorded the FM signal. (c) Raw measurement data and (d) power spectrum data of the four capsules showed significantly different responses between them.

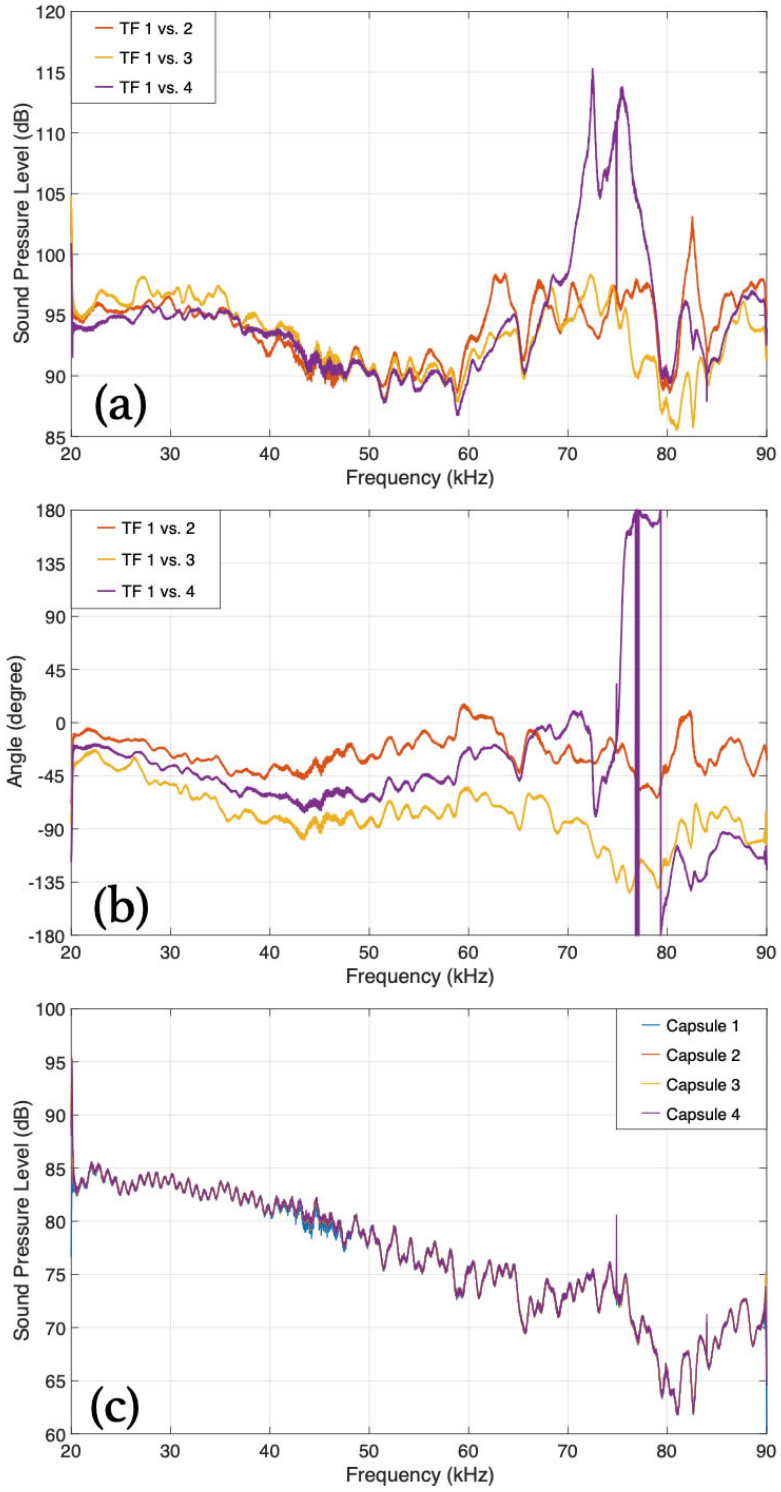


Figure 2.7. (a) Amplitude and (b) phase of the three transfer functions (1st capsule was set as the reference). (c) After correcting the amplitude and phase difference of each capsule (#2 – #4) using the transfer function approach, every capsule was virtually identical to the first (reference) capsule in response.

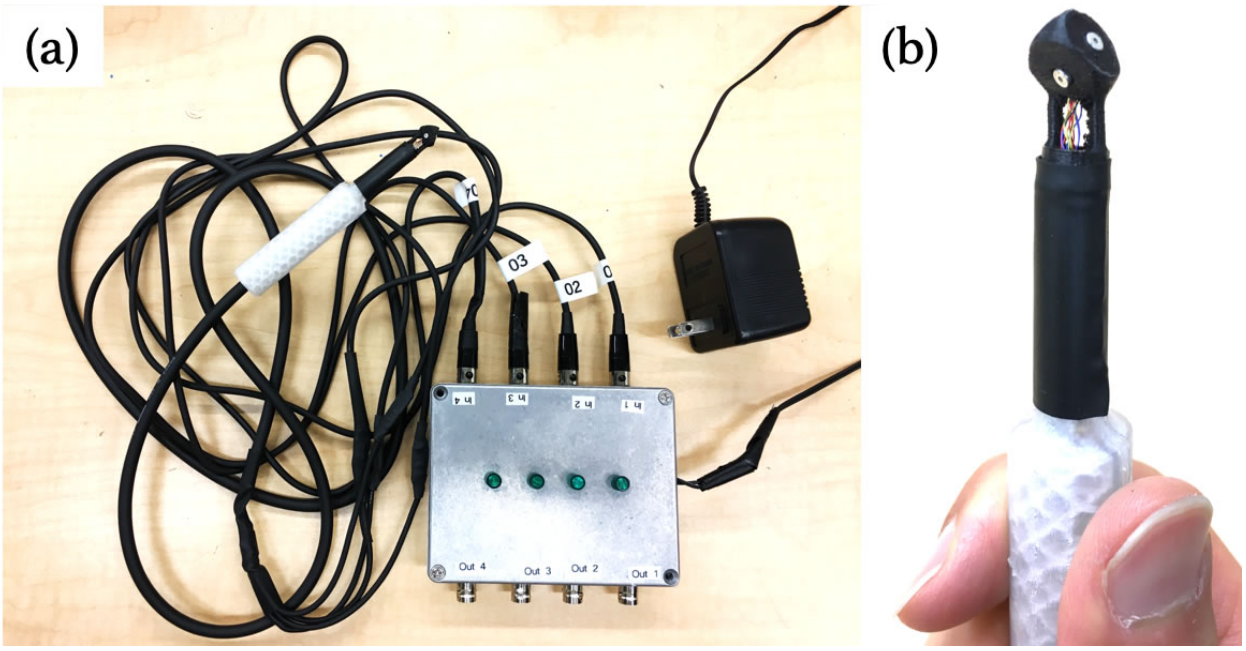


Figure 2.8. (a) The soundfield microphone with a custom-built power supply. (b) The custom-built ultrasonic tetrahedral soundfield microphone.

2.4.3 Validation

Once the calibration was done, the capsules were removed from the calibration holder and mounted in the 3D printed tetrahedron (Fig. 2.8). Directionality tests were done in a fairly anechoic room (a bat flight room at Brown University) to test the accuracy of the soundfield microphone. A two-millisecond downsweep chirp with two harmonics (the first harmonic of 50 to 25 kHz and second harmonic of 100 to 50 kHz in frequency) was transmitted using a loudspeaker (Kenwood KFC-XT15ie) at three different positions around the soundfield microphone (Fig. 2.9). The loudspeaker was connected with an amplifier (Harman Kardon PM645). The distance between the loudspeaker and microphone was 60 cm for all positions. At each position, 50 emissions were recorded with a sampling frequency of 250 kHz, and then averaged to increase the signal to noise ratio (SNR). An acquisition board (Measurement Computing USB-1208HS-2AO, resolution: 12 bit) and DASYLab 2016 software (Measurement Computing) were used for the acoustic measurements. Each of the four channels on the soundfield microphone was connected to a

high-pass filter with cutoff frequency of 22 kHz (Thorlabs EF123), and then to a microphone analog preamplifier (RME QuadMic II). Each audio channel was calibrated using its transfer function explained in the previous section during conversion from time domain to frequency domain. All recordings were processed and analyzed by the direction estimation system using the HARPEx algorithm. In each case, detected direction of the sound agreed with the direction of arrival (DOA) of the recorded sound as shown in Fig. 2.9, demonstrating the soundfield microphone functions correctly for ultrasonic waves.

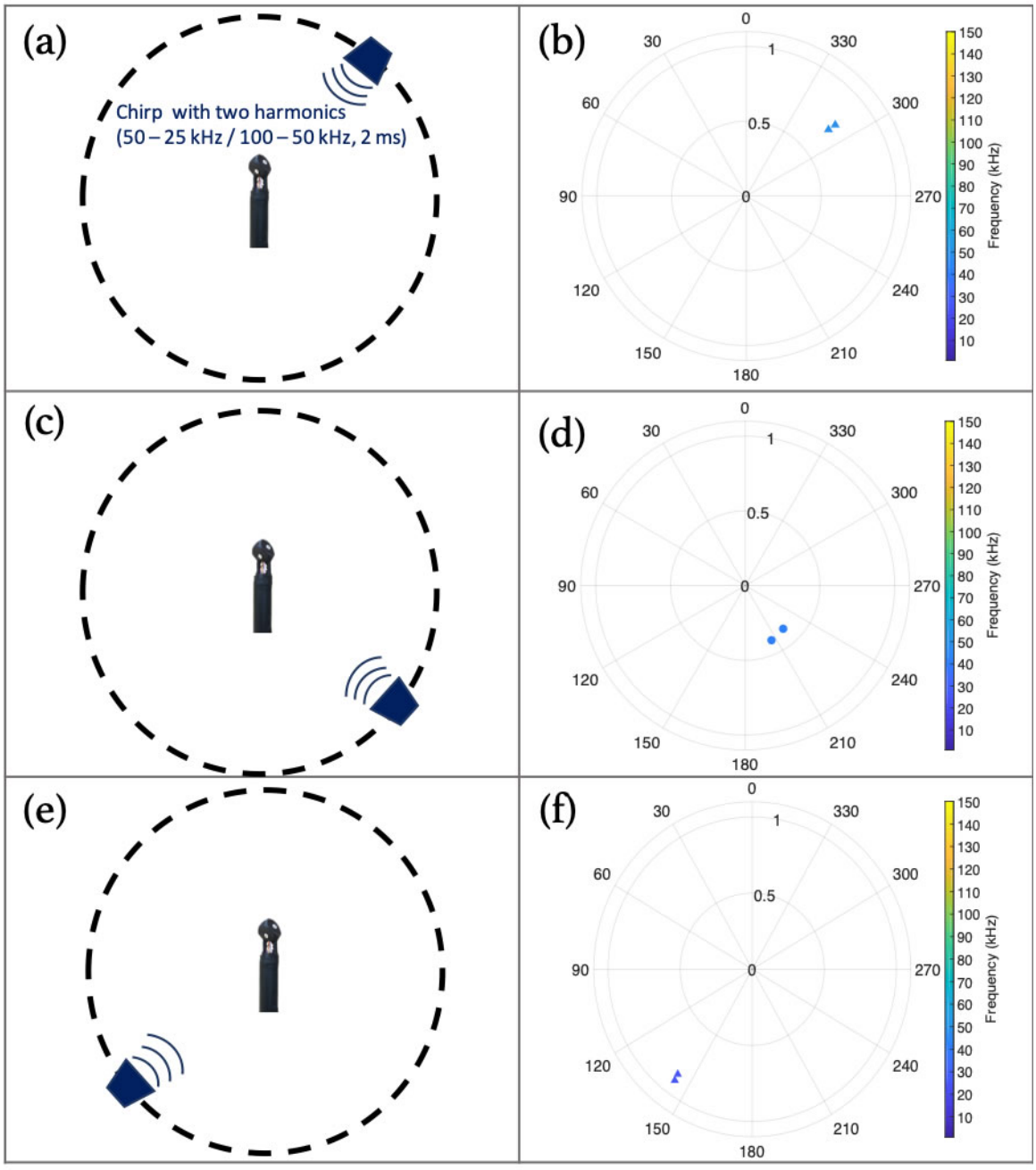


Figure 2.9. The custom-built ultrasonic microphone recorded linear frequency-modulated (LFM) sound waves containing two harmonics (100-50 and 50-25 kHz) transmitted by a loudspeaker. The loudspeaker emitted the sound at three positions (a, c, e). The direction estimation system using HARPEx identified the direction of the ultrasonic sound correctly in each case (b, d, f).

2.5 Application of the system to bat biosonar

2.5.1 Measurement using the soundfield microphone

Performance of the proposed measurement and analysis methodology to estimate direction of sound was validated by an experiment using a big brown bat (*Eptesicus fuscus*). In the experiment, the microphone (audio receiver) captured the ultrasonic sounds emitted by a flying bat and the estimation system detected its direction in angle. The soundfield microphone was placed in the center of a flight room, 1.5 m above the floor, and the bat was released from the corner and allowed to fly freely. The calls were recorded with the sampling frequency 250 kHz. The dimensions of the flight room are 4.4 m (L) by 4.2 m (W) by 2.5 m (H) (Barchi *et al.*, 2013). The data acquisition system was programmed to record the calls for 10 seconds. The bat flew around the soundfield microphone in the same direction for longer than 10 seconds. The walls of the flight room were covered with anechoic foam, and the floor was covered with artificial grass turf to reduce the reflection and reverberation of sound. The identical DAQ board, software, high-pass filter, and microphone amplifier used in section 2.4.3 were used for this measurement.

2.5.2 Analysis using the direction estimation system based on HARPEx

Recorded biosonar calls were processed using the estimation system using the HARPEx algorithm and analyzed to track the flight-path of the flying bat. The audio signals were filtered using a digital bandpass filter (20-100 kHz, FIR, length: 634, designed and implemented in MATLAB). The recorded audio signal was calibrated using its transfer function in frequency domain, and processed by the direction estimation system. Signals with clear first and second harmonics were identified as biosonar pulses from the bat. Length of each pulse to be processed was decided based on the duration of the first harmonic of the call. Depending on the length of

call, between 1,100 and 1,800 data points were used for analysis for each call. Result showed that estimated directions of the bat calls formed a smooth trajectory of the flying bat. If the direction at each call was incorrectly estimated, the result would show discontinuity on flight-path. The result confirmed that our ultrasonic soundfield microphone and HARPEx-based target detection system can be used for bat biosonar study to estimate angles of targets, which is the ultimate goal of this work. Five consecutive biosonar calls among many from the recording were selected to illustrate an example of trajectory of the bat, as shown in Fig. 2.10.

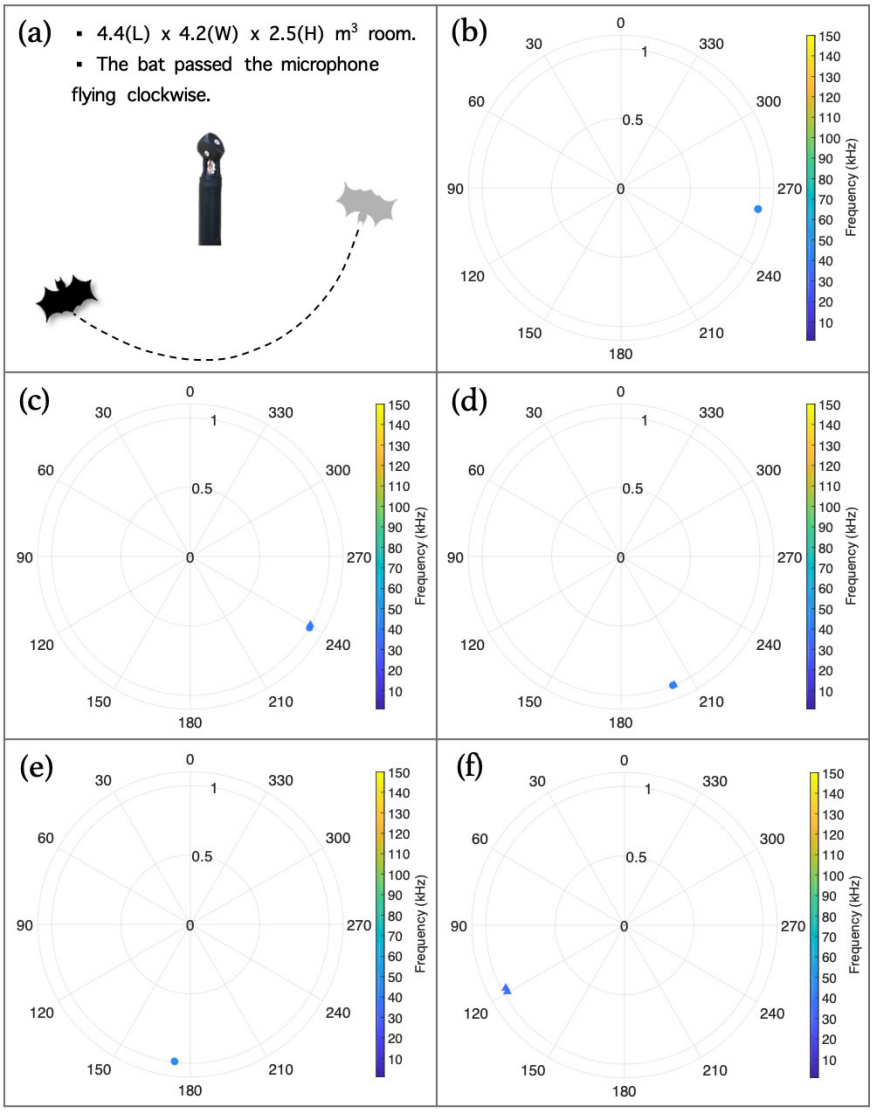


Figure 2.10. (a) A bat flying in a room was recorded by the custom-built ultrasonic soundfield microphone. The bat flew through the room passing by the microphone clockwise without elevation. (b-f) Five consecutive calls from the middle of the recording were analyzed by the direction estimation system. The estimation results illustrated a smooth trajectory of the bat's flight, demonstrating that the estimation system correctly detected the direction of bat position at each moment during the flight. This experiment validated that the developed hardware (the ultrasonic soundfield microphone) and software (the direction estimation system using HARPEx) based on spatial audio techniques can be used to estimate direction of target(s) by measuring and analyzing bat ultrasonic waves and echoes.

2.6 Conclusions and future work

The current work has developed and tested an ultrasonic tetrahedral soundfield microphone. The results show that such a device can be an effective for recording and analyzing ultrasonic bat transmissions and echoes. The approach gives accurate estimates of direction of arrival (azimuth and elevation) for up to two simultaneous (in time and frequency) signals. In parallel to the development of the soundfield microphone, we have implemented an audio processing and analysis method using the HARPEx algorithm in MATLAB. Using static tests in an anechoic chamber, and actual data captured from bats in flight, the system demonstrated the ability to accurately estimate bearing angles to the bat (target) and loudspeaker positions. Therefore, we believe that the system can be a useful additional tool to study bat echolocation.

Our ultimate goal in developing this system is to provide an image of the soundfield in close proximity to the bat's head using recorders such as telemetry microphones (Hiryu *et al.*, 2005; Hiryu *et al.*, 2007). Bats are capable of very high levels of adaptation of ear geometry which changes their head related transfer functions (Aytekin *et al.*, 2004). Our goal is to use soundfield analysis of the echoes arriving at the bat's location together with high-speed video to match physical adaptations with the changes in the soundfield. This would be especially useful in understanding how bats are able to find prey and avoid obstacles in high-density clutter.

Acknowledgments

This research was supported by the Multidisciplinary University Research Initiative (MURI) program from the Office of Naval Research (ONR) Grant No. N00014-17-1-2736.

References

Arlettaz, R., Jones, G., and Racey, P. A. (**2001**). "Effect of acoustic clutter on prey detection by bats," Nature **414**, 742.

Arteaga, D. (**2015**). "Introduction to Ambisonics."

Aytekin, M., Grassi, E., Sahota, M., and Moss, C. F. (**2004**). "The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation," The Journal of the Acoustical Society of America **116**, 3594-3605.

Barchi, J. R., Knowles, J. M., and Simmons, J. A. (**2013**). "Spatial memory and stereotypy of flight paths by big brown bats in cluttered surroundings," Journal of Experimental Biology **216**, 1053-1063.

Barclay, R. M. (**1985**). "Long-versus short-range foraging strategies of hoary (Lasiurus cinereus) and silver-haired (Lasionycteris noctivagans) bats and the consequences for prey selection," Canadian Journal of Zoology **63**, 2507-2515.

Barrett, N., and Berge, S. (**2010**). "A new method for B-format to binaural transcoding," in *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space* (Audio Engineering Society).

Berge, S., and Barrett, N. (**2010**). "High angular resolution planewave expansion," in *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics May*, pp. 6-7.

Fawcett, K., Jacobs, D. S., Surlykke, A., and Ratcliffe, J. M. (**2015**). "Echolocation in the bat, Rhinolophus capensis: the influence of clutter, conspecifics and prey on call design and intensity," Biology open **4**, 693-701.

Gerzon, M. A. (**1973**). "Periphony: With-height sound reproduction," Journal of the Audio Engineering Society **21**, 2-10.

Gerzon, M. A. (**1980**). "Practical periphony: The reproduction of full-sphere sound," in *Audio Engineering Society Convention 65* (Audio Engineering Society).

Hiryu, S., Hagino, T., Riquimaroux, H., and Watanabe, Y. (**2007**). "Echo-intensity compensation in echolocating bats (Pipistrellus abramus) during flight measured by a telemetry microphone," The Journal of the Acoustical Society of America **121**, 1749-1757.

Hiryu, S., Katsura, K., Lin, L.-K., Riquimaroux, H., and Watanabe, Y. (**2005**). "Doppler-shift compensation in the Taiwanese leaf-nosed bat (Hipposideros terasensis) recorded with a telemetry microphone system during flight," The Journal of the Acoustical Society of America **118**, 3927-3933.

Jakobsen, L., Ratcliffe, J. M., and Surlykke, A. (**2013**). "Convergent acoustic field of view in echolocating bats," Nature **493**, 93.

Jensen, M. E., and Miller, L. A. (**1999**). "Echolocation signals of the bat Eptesicus serotinus recorded using a vertical microphone array: effect of flight altitude on searching signals," Behavioral Ecology and Sociobiology **47**, 60-69.

Jones, G. (**1994**). "Scaling of wingbeat and echolocation pulse emission rates in bats: why are aerial insectivorous bats so small?," Functional Ecology, 450-457.

Knowles, J. M., Barchi, J. R., Gaudette, J. E., and Simmons, J. A. (**2015**). "Effective biosonar echo-to-clutter rejection ratio in a complex dynamic scene," The Journal of the Acoustical Society of America **138**, 1090-1101.

Ortolani, F. (**2015**). "Introduction to Ambisonics," Ironbridge Electronics.

Surlykke, A., Boel Pedersen, S., and Jakobsen, L. (**2008**). "Echolocating bats emit a highly directional sonar sound beam in the field," Proceedings of the Royal Society B: Biological Sciences **276**, 853-860.

Warnecke, M., Falk, B., and Moss, C. F. (**2018**). "Echolocation and flight behavior of the bat Hipposideros armiger terasensis in a structured corridor," The Journal of the Acoustical Society of America **144**, 806-813.

Waters, D. A., Rydell, J., and Jones, G. (**1995**). "Echolocation call design and limits on prey size: a case study using the aerial-hawking bat Nyctalus leisleri," Behavioral Ecology and Sociobiology **37**, 321-328.

Wheeler, A., Kloepper, L., Matsuo, I., and Simmons, J. A. (**2014**). "Temporal dynamics of echolocation during clutter navigation in Eptesicus fuscus," The Journal of the Acoustical Society of America **135**, 2208-2208.

Chapter 3

Performance of a DOA Estimator Using a Soundfield Microphone for Bat Biosonar

Hyeon Lee[1*], Michael J. Roan[1], Chen Ming[2], James A. Simmons[2]

[1]Department of Mechanical Engineering, Virginia Tech, 460 Old Turner Street, Blacksburg, VA 24061, USA

[2]Department of Neuroscience, Brown University, 185 Meeting Street, Providence, RI 02912, USA

Abstract

The authors previously introduced a direction of arrival (DOA) estimator for bat biosonar calls and echoes, based on a high-frequency soundfield microphone, Ambisonics, and high angular resolution planewave expansion (HARPEx). The estimator provided accurate bearing angles of sound sources (loudspeakers or a bat). In this work, the authors assess the performance of the estimator, using simulations and statistical analyses. Virtual microphone capsule responses to a source emitting a chirp (3 ms, two harmonics: 40-22 kHz and 80-44 kHz) at a 30° angle are used to estimate the source angle in the simulations. Medians and RMSEs (root-mean-square error) of 10,000 simulations for each case (out of 49 cases with variables *d*, distance between each capsule and the microphone center, and SNR) indicate how accurate and precise the estimations are, respectively. The Cramer-Rao lower bounds (CRLB) are also computed with various *d* and SNR conditions to be compared with the statistical study result. The CRLB does not reflect the effect of *d* properly since it is based on the time difference of arrival (TDOA) of signals, leading to incorrect performance assessments. Results suggest to ensure shorter *d* or/and higher SNR values to achieve higher performance of the estimator.

Keywords: soundfield microphone; HARPEx; angle estimation; CRLB; signal modeling; biosonar.

3.1 Introduction

We previously introduced a new method of measuring and processing bat biosonar calls and echoes to acquire directional information of the bat(s) or target(s) (Lee *et al.*, 2019). The method utilizes a high-frequency soundfield microphone, Ambisonics, and high angular resolution planewave expansion (HARPEx). The developed hardware and software system based on this method in the work accurately estimated the directions of high-frequency sound sources (loudspeakers or a bat). This proves the proposed method can be used for detecting the sound source direction of arrival (DOA). The next question is, how accurate estimations can we expect from the developed system?

The Cramer-Rao lower bound (CRLB) can be an option to answer the question. The most popular measurement and analysis method for bat biosonar study is using microphone arrays to measure bat sounds and the time differences of arrival (TDOA) of signals between the array sensors (microphones) for locating the bat(s) (Arlettaz *et al.*, 2001; Surlykke *et al.*, 2009; Jakobsen *et al.*, 2013; Warnecke *et al.*, 2018). For the signal measurement and analysis methods based on the TDOA, the CRLB is probably the most widely used tool to assess the performance of the system (estimator) by providing the minimum variance of estimations the estimator can achieve (Kay, 1993; Nielsen, 1999; Erling *et al.*, 2007).

The CRLB, however, is not a proper tool to assess our estimation system (entire hardware and software for measurements, signal processing, and analysis for estimations) because the system is developed based not on the TDOA, but on the intensity difference between the microphone capsule responses (Gerzon, 1973; 1980). Since the soundfield microphone of the estimation system consists of four cardioid capsules facing different directions from one another, the response (amplitude) of each capsule varies depending on the DOA of signals. The variation

of the response provides the directional information which is provided by the time delays of signals between sensors in the TDOA based systems. This also brings a substantial difference of the soundfield microphone from microphone arrays for the TDOA method; theoretically its four cardioid capsules are supposed to be placed at the same physical point (Arteaga, 2015; Ortolani, 2015). The demand of no physical gap between four capsules to achieve the optimal form of signal measurement tool (soundfield microphone) is not imposed on the TDOA based systems using the microphone arrays. Thus, we found an alternative way to assess the estimation system.

In this work, we conducted a statistical study of simulations with various scenarios (cases) to evaluate the performance of the estimator with the ultrasonic soundfield microphone, Ambisonics, and HARPEx. We created a virtual sound source with an angle, and generated signals imitating the responses of soundfield microphone capsules to the sound source. The generated signals with various SNR (signal-to-noise ratio) values were put into the system as measured data, and the estimation algorithm using HARPEx provided source angle estimations. A total of 10,000 simulations were performed for each scenario. We conducted a statistical analysis, and acquired the median and RMSE (root-mean-square error) of the estimation data for each scenario. The median and RMSE were used as indicators of the accuracy and precision of the estimation, respectively. We also computed CRLBs of our estimator in various situations to compare the CRLB method with the statistical study method.

This paper begins with how we build the signal model to generate the virtual capsule responses for the simulations. Next, we provide the simulation setup and the statistical study result of the estimations in the simulations. The CRLB calculation process and result for our estimator in various situations follow. The last part of this paper presents the comparison between the statistical study method and CRLB method, and conclusions from the results.

3.2 Signal model

3.2.1 Delays between the capsules

Knowing the speed of sound propagation (c = 343 m/s), differences in sound travel distance between capsules are calculated to acquire the time delays of signal arrival between the capsules, with taking capsule 1 as the reference. Four cardioid capsules form the tetrahedral soundfield microphone. Each capsule is located on the faces of an imaginary tetrahedron, with sharing an identical distance (d) between each capsule and the center of the soundfield microphone, as shown in Fig. 3.1. Taking the microphone center as the origin, the coordinates of capsules 1-4 are $\frac{d}{\sqrt{3}}(1,1,1)$, $\frac{d}{\sqrt{3}}(-1,1,-1)$, $\frac{d}{\sqrt{3}}(-1,-1,1)$, and $\frac{d}{\sqrt{3}}(1,-1,-1)$, respectively. Due to the difference in location, there are always delays in signal arrival time between capsules regardless of the direction from which sound comes.

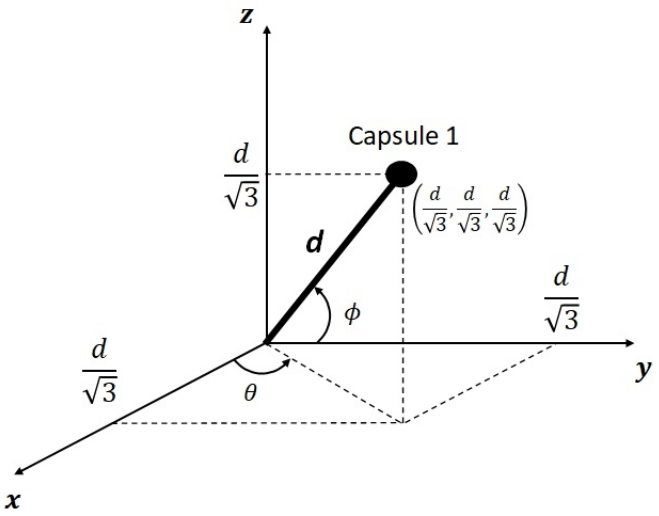


Figure 3.1. The coordinate of capsule 1 (the reference capsule) of the soundfield microphone. The center of the soundfield microphone is considered as the origin. d is the distance between a capsule and the origin. θ and ϕ denote azimuth and elevation, respectively, in this work.

Assume a situation with a monopole sound source as shown in Fig. 3.2. We consider a far-field condition where sound propagates as a planewave. The normal vector of the planewave that is the DOA is $(cos\theta cos\phi, sin\theta cos\phi, sin\phi)$ where θ and ϕ are azimuth and elevation, respectively. Thus, the equation of the plane when the acoustic planewave arrives at capsule 1 is defined as Eq. (3.1), taking the capsule as a point on the plane.

$$cos\theta cos\phi \left(x - \frac{d}{\sqrt{3}}\right) + sin\theta cos\phi \left(y - \frac{d}{\sqrt{3}}\right) + sin\phi \left(z - \frac{d}{\sqrt{3}}\right) = 0. \qquad (3.1)$$

The travel distance of the signal from capsule 1 to another capsule is computed as Eq. (3.2) using the equation of distance between a plane and a point.

$$\delta_{1j} = \frac{\left|cos\theta cos\phi * x_j + sin\theta cos\phi * y_j + sin\phi * z_j - \frac{d}{\sqrt{3}}(cos\theta cos\phi + sin\theta cos\phi + sin\phi)\right|}{\sqrt{cos^2\theta cos^2\phi + sin^2\theta cos^2\phi + sin^2\phi}}, \qquad (3.2)$$

where (x_j, y_j, z_j) is the coordinate of the j^{th} capsule location. Hence, the delay between the reference capsule and the j^{th} capsule is $\frac{\delta_{1j}}{c}$. Assuming the source has zero elevation ($\phi = 0°$, Fig. 3.2), the distances between the reference capsule and the others, and corresponding delays (τ_{1i}) are:

$$\delta_{12} = \frac{2d}{\sqrt{3}} cos\theta \implies \tau_{12} = \frac{2d}{\sqrt{3}c} cos\theta, \qquad (3.3a)$$

$$\delta_{13} = \frac{2d}{\sqrt{3}} (cos\theta + sin\theta) \implies \tau_{13} = \frac{2d}{\sqrt{3}c} (cos\theta + sin\theta), \qquad (3.3b)$$

$$\delta_{14} = \frac{2d}{\sqrt{3}} sin\theta \implies \tau_{14} = \frac{2d}{\sqrt{3}c} sin\theta. \qquad (3.3c)$$

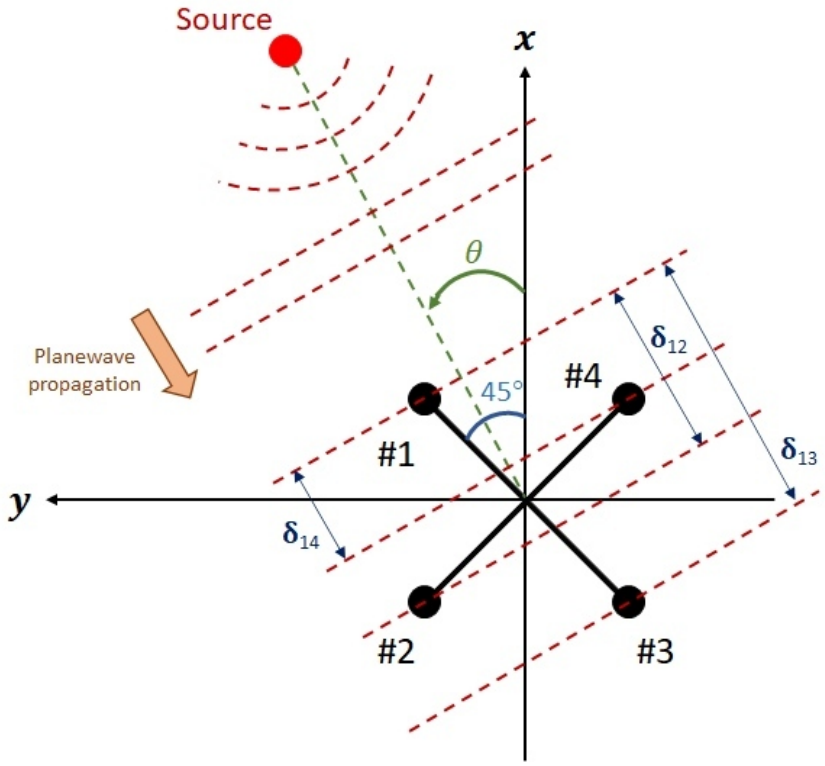


Figure 3.2. A situation of the soundfield microphone measuring a monopole sound source was considered in this work. The microphone capsules measure the planewave from the source at a certain angle. Since there are differences between the capsules in signal traveling distance (δ_{1i}), there exist time delays between the capsules (τ_{1i}).

3.2.2 Capsule sensitivity based on the cardioid pattern

Another factor to consider in the signal model is the sensitivity variation of the capsules depending on the sound direction. The variation is described as 'cardioid polar pattern,' as shown in Fig. 3.3. A cardioid microphone capsule responds most sensitively to sounds from the direction it faces, whereas it shows no response to sounds directly from behind. According to the direction each capsule faces based on the tetrahedron soundfield microphone structure, capsules 1-4 show the highest responses to sounds from 45°, 135°, 225°, and 315° angle, respectively.

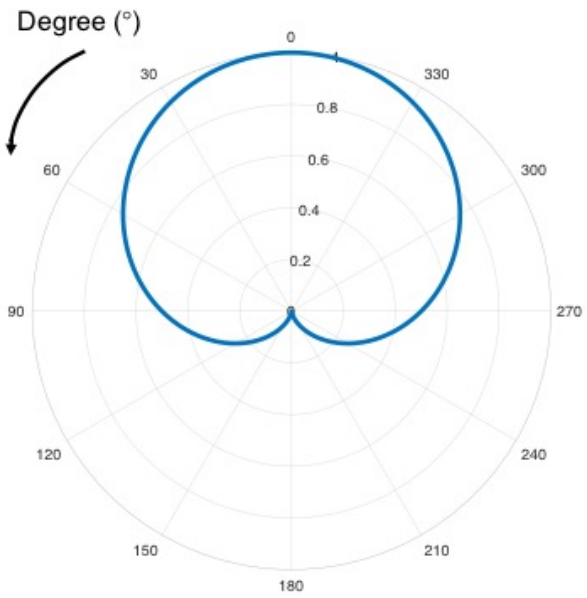


Figure 3.3. The response pattern of cardioid microphone capsules. It varies with the direction of sound. A cardioid microphone shows the highest response to sounds coming from the direction the microphone faces.

3.2.3 Quadratic frequency modulated (QFM) signal

A downward quadratic frequency modulated (QFM) signal was considered as the source signal with benchmarking FM-bat sonar calls. The 3-ms-long signal consists of two harmonics: 40-22 kHz and 80-44 kHz imitating the echolocation signals of the big brown bat (*Eptesicus fuscus*) in the terminal phase (Simmons *et al.*, 1975; Simmons *et al.*, 1995). The source signal, *s*, is defined as:

$$s = A_1 \exp\left[i2\pi f_{i1}\left(\frac{f_{f1}}{f_{i1}}\right)^{t/T} t\right] + A_2 \exp\left[i2\pi f_{i1}\left(\frac{f_{f2}}{f_{i2}}\right)^{t/T} t\right], \qquad (3.4a)$$

$$A_1 = \sigma\sqrt{SNR_1} \;\;\&\;\; A_2 = \sigma\sqrt{SNR_2}, \qquad (3.4b)$$

$$t = \frac{n}{f_s} \qquad (3.4c)$$

where A_1: amplitude of the first harmonic, A_2: amplitude of the second harmonic, f_{i1}: 40 kHz, f_{f1}: 22 kHz, f_{i2}: 80 kHz, f_{f2}: 44 kHz, t: time at the snapshot (s), T: 3 ms, f_s: sampling frequency, n: the number of nth data point, and σ: standard deviation of noise.

3.2.4 Completed signal model

Taking the delays between capsule 1 and the others, the direction sensitivity of the capsules, and the source signal (s) into account, expected signals at all channels (capsules) are described as:

$$c_1 = \left\{0.5 + 0.5cos\left(\theta_s - \frac{1}{4}\pi\right)\right\}s + e_1\,, \tag{3.5a}$$

$$c_2 = \left\{0.5 + 0.5cos\left(\theta_s - \frac{3}{4}\pi\right)\right\}\left\{A_1\exp\left[i2\pi f_{i1}\left(\frac{f_{f1}}{f_{i1}}\right)^{t/T}\left(t - \frac{2d}{\sqrt{3}c}cos\theta_s\right)\right] + A_2\exp\left[i2\pi f_{i1}\left(\frac{f_{f2}}{f_{i2}}\right)^{t/T}\left(t - \frac{2d}{\sqrt{3}c}cos\theta_s\right)\right]\right\} + e_2\,, \tag{3.5b}$$

$$c_3 = \left\{0.5 + 0.5cos\left(\theta_s - \frac{5}{4}\pi\right)\right\}\left\{A_1\exp\left[i2\pi f_{i1}\left(\frac{f_{f1}}{f_{i1}}\right)^{t/T}\left(t - \frac{2d}{\sqrt{3}c}(cos\theta_s + sin\theta_s)\right)\right] + A_2\exp\left[i2\pi f_{i1}\left(\frac{f_{f2}}{f_{i2}}\right)^{t/T}\left(t - \frac{2d}{\sqrt{3}c}(cos\theta_s + sin\theta_s)\right)\right]\right\} + e_3, \tag{3.5c}$$

$$c_4 = \left\{0.5 + 0.5cos\left(\theta_s - \frac{7}{4}\pi\right)\right\}\left\{A_1\exp\left[i2\pi f_{i1}\left(\frac{f_{f1}}{f_{i1}}\right)^{t/T}\left(t - \frac{2d}{\sqrt{3}c}sin\theta_s\right)\right] + A_2\exp\left[i2\pi f_{i1}\left(\frac{f_{f2}}{f_{i2}}\right)^{t/T}\left(t - \frac{2d}{\sqrt{3}c}sin\theta_s\right)\right]\right\} + e_4\,, \tag{3.5d}$$

where c_j : signal of j^{th} channel (the response of j^{th} capsule), θ_s: the angle of the sound source (radian), and e_j : noise in each channel. Generated signals using the equations (3.5a)-(3.5d) for each case of simulations were processed and analyzed by the estimator as input signals from all channels.

3.3 Simulations

3.3.1 Simulation setup

We created a virtual situation of a sound source emitting the 3-ms chirp signal at a 30° angle from x-axis (θ_s = 30°) for simulations as shown in Fig. 3.4. The source and the soundfield microphone are on the same level (ϕ = 0°). Study cases were created with various d values (1, 1.5,

2, 2.5, 3, 3.5, and 4 mm) and SNR values of the first harmonic signal (SNR_1: 0, 5, 10, 15, 20, 25, and 30 dB). There is a 5 dB gap in sound pressure level between the first and second harmonics, reflecting the sound intensity difference between the two harmonics depending on source angle. The sound intensity gap is observed when the source angle is over approximately 10°. The intensity gap increases or decreases with angle in 10-90° range forming a pattern with multiple lobes according to literature (Hartley and Suthers, 1989; Ghose and Moss, 2003; Bates *et al.*, 2011). The sampling frequency is 500 kHz referencing the specification of our data acquisition module (NI 9222, National Instruments Corp) for sound measurements. Random white Gaussian noise (σ = 0.5 V) with zero mean is included in the channels for each signal.

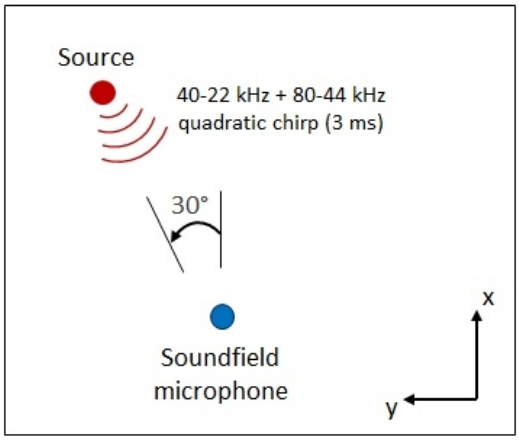


Figure 3.4. A situation of sound source emitting an QFM signal at an angle of 30° from x-axis was set for all simulations. The capsule responses of the soundfield microphone were generated using Eqs. (3.5a)-(3.5d) with imitating the soundfield microphone measures the source signal in the simulations.

Reflecting these setup conditions, signals of all channels (capsule responses) were generated using Eqs. (3.5a)-(3.5d) for each case. The generated signals were entered as measured data in the estimation system. A total of 49 cases with two variables of *d* and SNR (combinations of the seven values of *d* and the seven values of SNR) were studied. A total of 10,000 simulations

were performed (i.e. a total of 10,000 angle estimations were made) for each case. We conducted a statistical analysis for each case to assess the performance of the estimation system. The median and RMSE of 10,000 estimations were addressed to represent the accuracy and precision of the estimation, respectively, in each case. The RMSE is defined as:

$$RMSE = \sqrt{\frac{\sum_1^L (\theta_s - \theta_e)^2}{L}}, \qquad (3.6)$$

where L: the number of simulations for each case (10,000) and θ_e: angle estimation. Median was chosen instead of average because the probability density of estimations for each case do not fall into the normal distribution type. We studied the change of estimation performance with d and SNR.

3.3.2 Simulation results

Figure 3.5 shows the results of the statistical analyses with the various cases of simulations. The legend denotes the SNR values of the first harmonic (SNR_1). The corresponding SNR_2 values are 25, 20, 15, 10, 5, 0, and -5 dB in order (5 dB lower than SNR_1). Estimated angle (θ_e) of the sound source diverges from the actual angle value (30°) as d increases, as shown in Fig. 3.5(a). When d is 2.5 mm, an SNR_1 above 10 dB ensures the angle estimation accuracy of ± 10° range. Higher SNR values are required to retain the accuracy level as d increases further. If d increases to 3.5 mm, angle estimations even with 30 dB SNR_1 are clearly out of the ± 10° accuracy range (median: 15.4°). If SNR_1 is 5 dB, d is required to be 1.5 mm or less to achieve the tolerance of ± 10°. The RMSE of angle estimations increases as d increases, as shown in Fig. 3.5(b). Only SNR_1 of 30 dB delivers RMSEs lower than 10° when d is 2 mm or less. If d is 2.5 mm or greater, SNR_1 of 30 dB or lower leads to higher RMSEs than 30°. For our soundfield microphone, the capsules are roughly 2 mm (δ_c) apart from one another to achieve a frequency range up to approximately

80 kHz for the first microphone iteration. This gives roughly 1.2 mm as d (the distance between each capsule and the soundfield microphone center) using the geometry of tetrahedrons ($\delta_c \approx 1.63d$). Considering the soundfield microphone with the d of 1.2 mm, the result suggests to ensure SNR_1 higher than 25 dB, to achieve ± 10° tolerance in both accuracy and precision of angle estimations.

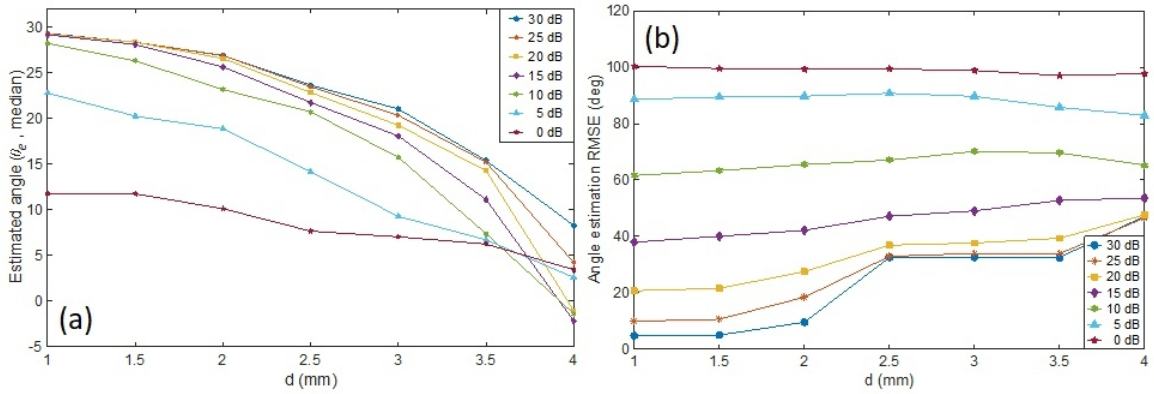


Figure 3.5. (a) The medians of angle estimations for all cases. Median diverges from the actual angle value as d increases or/and SNR decreases. (b) The RMSEs of angle estimations for all cases. RMSE increases as d increases or/and SNR decreases. These two plots suggest to ensure SNR_1 over 25 dB to achieve ± 10° tolerance in both accuracy and precision using our angle estimation system (d = 1.2 mm).

3.4 Cramer-Rao Lower Bound (CRLB) analysis

We also computed CRLBs of our estimator for diverse cases with the variables of d and SNR_1. The CRLB study method is to be compared with the statistical study method from section 3.3 in performance assessment.

3.4.1 Measurement model

The method and formulae for acoustic vector sensor arrays from (Nehorai and Paldi, 1994; Lai and Bell, 2007) were used. All the details can be found from those references. Since the parameter of our interest is the source angle (θ), the B-format input signals can be in the format of:

$$S(t) = B(\theta)s(t) + e(t), \qquad (3.7)$$

where $B(\theta)$ is the transfer matrix, $e(t)$ is noise, and $s(t)$ is from Eq. (3.4a). Considering the monopole source, four-channel (microphone) inputs, and only the first harmonic (the main) signal, Eq.(3.7) is extended as:

$$\begin{matrix} S(t) \\ \begin{bmatrix} W(t) \\ X(t) \\ Y(t) \\ Z(t) \end{bmatrix} \end{matrix} = \begin{matrix} B(\theta) \\ \begin{bmatrix} r_1 + r_2 + r_3 + r_4 \\ r_1 - r_2 - r_3 + r_4 \\ r_1 + r_2 - r_3 - r_4 \\ r_1 - r_2 + r_3 - r_4 \end{bmatrix} \end{matrix} s(t) + \begin{matrix} e(t) \\ \begin{bmatrix} e_w \\ e_x \\ e_y \\ e_z \end{bmatrix} \end{matrix}, \qquad (3.8)$$

$$r_1 = \left\{0.5 + 0.5cos\left(\theta - \tfrac{1}{4}\pi\right)\right\}s + e_1, \qquad (3.9a)$$

$$r_2 = \left\{0.5 + 0.5cos\left(\theta - \tfrac{3}{4}\pi\right)\right\}\exp\left[-\frac{i4\pi d f_{i1}}{\sqrt{3}c}\left(\frac{f_{f1}}{f_{i1}}\right)^{t/T} cos\theta\right] + e_2, \qquad (3.9b)$$

$$r_3 = \left\{0.5 + 0.5cos\left(\theta - \tfrac{5}{4}\pi\right)\right\}\exp\left[-\frac{i4\pi d f_{i1}}{\sqrt{3}c}\left(\frac{f_{f1}}{f_{i1}}\right)^{t/T} (cos\theta + sin\theta)\right] + e_3, \qquad (3.9c)$$

$$r_4 = \left\{0.5 + 0.5cos\left(\theta - \tfrac{7}{4}\pi\right)\right\}\exp\left[-\frac{i4\pi d f_{i1}}{\sqrt{3}c}\left(\frac{f_{f1}}{f_{i1}}\right)^{t/T} sin\theta\right] + e_4, \qquad (3.9d)$$

where e_w, e_x, e_y, and e_z are noises in the corresponding signal. The elements of $B(\theta)$ include each microphone capsule's cardioid pattern, and the relationship between angle (θ) and the first order harmonics (Manikas *et al.*, 2001; Arteaga, 2015). We assume the source signal, $s(t)$, has zero mean, and its sequence is a Gaussian process. The covariance of the signal (P) is defined as:

$$P = E[s(t)s(t)^H]\,, \qquad (3.10)$$

where superscript H denotes Hermitian conjugate (or conjugate transpose). The CRLB for angle estimation (the only parameter: θ) of a monopole sound source using a soundfield microphone composed of four microphone capsules is defined as:

$$CRLB\ (\theta) = \frac{\sigma^2}{2N}[Re\{U \odot (D^H \Pi_c D)^T\}]^{-1} \qquad (3.11)$$

where superscript T: transpose, $\odot$: Hadamard product, and N: the total number of the signal data points,

$$U = P(B^H BP + \sigma^2 I)^{-1} B^H BP\,, \qquad (3.12a)$$

$$\Pi_c = I - \Pi\,, \qquad (3.12b)$$

$$\Pi = \mathrm{B}(B^H B)^{-1} B^H, \qquad (3.12c)$$

$$D = \begin{bmatrix} \partial B_1/\partial\theta \\ \partial B_2/\partial\theta \\ \partial B_3/\partial\theta \\ \partial B_4/\partial\theta \end{bmatrix}, \qquad (3.12d)$$

where I denotes the identity matrix, and B_1, B_2, B_3, and B_4 are the elements of matrix $B(\theta)$.

3.4.2 CRLB computation results

The monopole source scenario for simulations in section 3.3 was considered for the CRLB calculations. The source was emitting the 3-ms chirp signal at a 30° angle and zero elevation (ϕ = 0°). The soundfield microphone of our estimation system was considered as a vector sensor. CRLBs for the first snapshot (N=1) were computed using the same values from section 3.3 for all the terms. A range of 1-4 mm was considered for d, and five SNR values (-6, -3, 0, 3, and 6 dB) were used. The CRLB decreases as d increases, as shown in Fig. 3.6. The lowest SNR (-6 dB)

showed the highest CRLB of approximately 68° at d = 1 mm. The CRLB is approximately 3° or less with the SNR of 6 dB across the range of d.

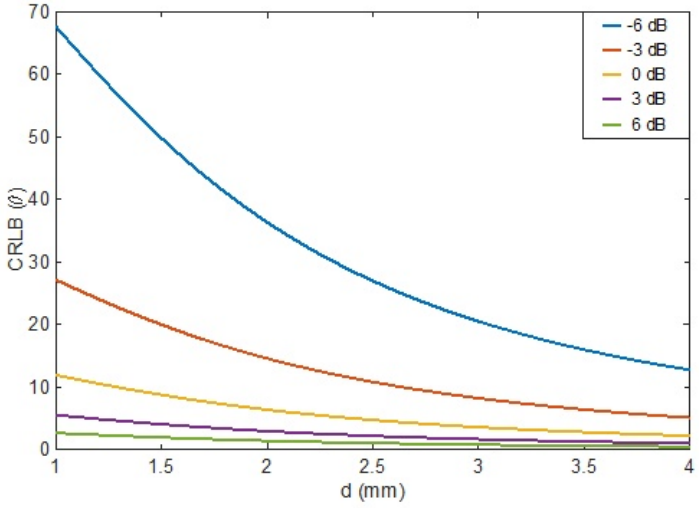


Figure 3.6. CRLB decreases as d increases with any SNR values. The highest SNR of 6 dB shows the lowest CRLBs across the range of d.

3.5 Discussion (comparison between two analysis methods)

The statistical study (section 3.3) and CRLB study (section 3.4) were conducted to assess the performance of our direction estimation system. Both studies agree that a higher SNR ensures a greater performance in angle estimation. The studies, however, disagree on the effect of d on the estimator performance. The statistical study demonstrates the accuracy and precision of angle estimation drops (higher gaps between the actual and estimated angles, and higher RMSEs) as d increases (Fig. 3.5). In contrast, the CRLB study shows the performance grows with d. The conflicting results are attributed to the improper use of the CRLB method to our estimator as previously mentioned in section 3.1. The CRLB is a method to assess the performance of an estimator by providing the minimum variance of estimations the estimator can achieve (Kay, 1993;

Nielsen, 1999; Erling *et al.*, 2007). The minimum variance indicates the minimum estimation error the estimator has. The important fact is the CRLB is a tool for estimators based on the TDOA method. Thus, the CRLB is not able to properly assess an estimator based on measurements and signal processing by a soundfield microphone, Ambisonics, and HARPEx. These are the crucial characteristics of our DOA estimator the CRLB method does not take into account, presented in the next paragraphs.

The fundamentals and basic theory of Ambisonics assume audio signals are measured by a soundfield microphone whose capsules are placed at a point together. While each capsule faces a different direction, they are assumed to measure signals from the same physical point regardless of the number of the capsules (regardless of the order of the microphone). The point all capsules are supposedly located at is the center of the soundfield microphone which is the signal measurement point. The DOA estimation mechanism using Ambisonics is established under this assumption. Cardioid microphone capsules have the response pattern with respect to angle as previously discussed in section 3.2.2 (Fig. 3.3). Since source angles determine capsule responses, the angle can be estimated from the capsule responses. Figure 3.7 illustrates a sound source at an angle (θ) with no elevation, and the cardioid capsule response patterns of a tetrahedron soundfield microphone. We are not able to detect the source angle by analyzing the response of only one microphone capsule because of the symmetry of the cardioid pattern, i.e. a unique response amplitude corresponds to two different incident angles. The comprehensive analysis of four cardioid capsule responses using Ambisonics, however, enables to narrow down to one incident angle. This is how the source angle estimation mechanism works.

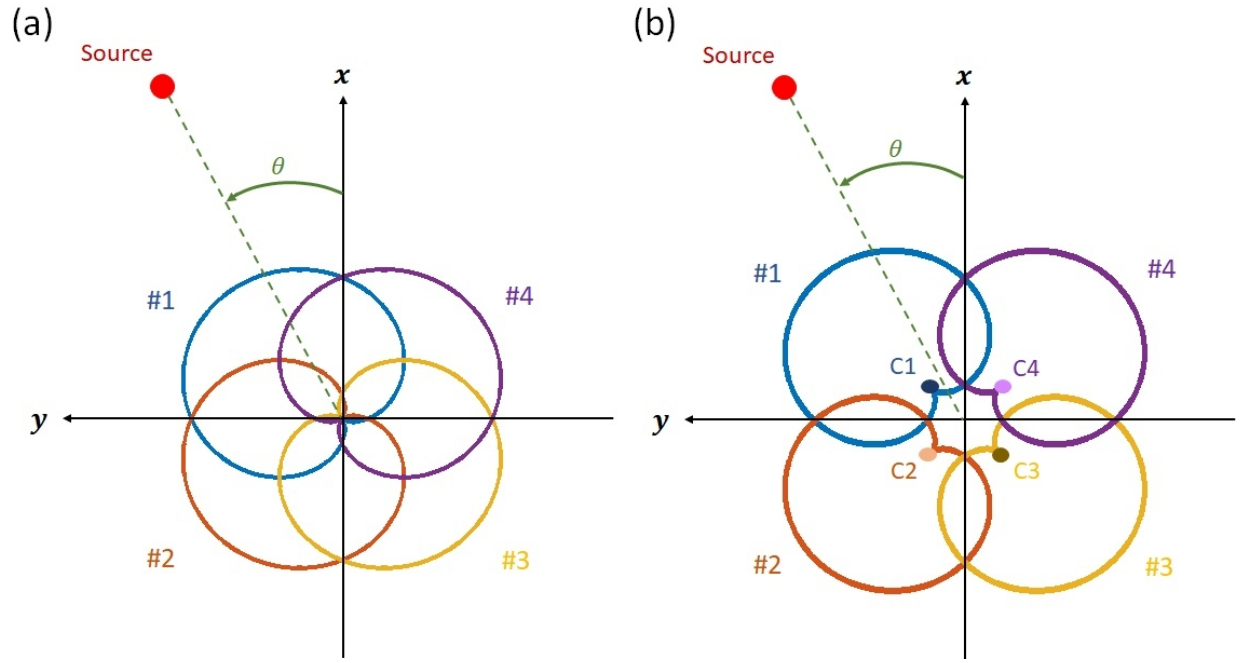


Figure 3.7. (a) Capsule response patterns when they are placed all together at the origin ($d = 0$). (b) When d is not zero, all capsules are located apart from one another.

However, it is nearly impossible for capsules of a soundfield microphone to be placed at a single point in practice due to physical constraints. This leads to a configuration of capsules of a tetrahedron soundfield microphone as shown in Fig. 3.7(b). Each capsule is apart from the center of the soundfield microphone (the origin) with a distance defined as d in this work. The capsules are also apart from one another. This means the measurement reference point of each capsule (the capsule center) does not match the others. C1, C2, C3, and C4 (Fig. 3.7(b)) that represent corresponding capsule centers, respectively, are four different points. See Fig. 3.8(a) to analyze what occurs to capsule 1 with a measurement reference point that is not the microphone center (the origin). Considering the new (translated) reference axes x' and y' for capsule 1, the capsule has a relatively narrower incident angle of sound arrival (θ') than the incident angle (θ) it would have if it was placed at the origin. The performance of estimating the actual source angle based on the

capsule response to the distorted incident angle (θ') surely decreases. Let's see the case of capsule 4 as shown in Fig. 3.8(b). The incident angle (θ') based on the new reference axes x' and y' for capsule 4 is greater than the actual source angle (θ). The incident angle for capsule 4 is different from not only the actual source angle, but also the incident angle for capsule 1. The disagreement between the capsules in incident angle observation hinders accurate DOA estimations. This would diminish the estimation performance more than the situation where the responses of the two capsules claim the identical incident angle could. Since there are four capsules in the soundfield microphone, incident angles the capsule responses claim are most likely all different from one another. Considering the different incident angles the capsules are exposed to and the symmetry of the capsule response pattern (cardioid), there could be maximum eight candidate angles for an estimation. This certainly lowers the chance for the estimator to detect the actual source angle. The accuracy and precision of source angle estimations drop more with the rise of *d*.

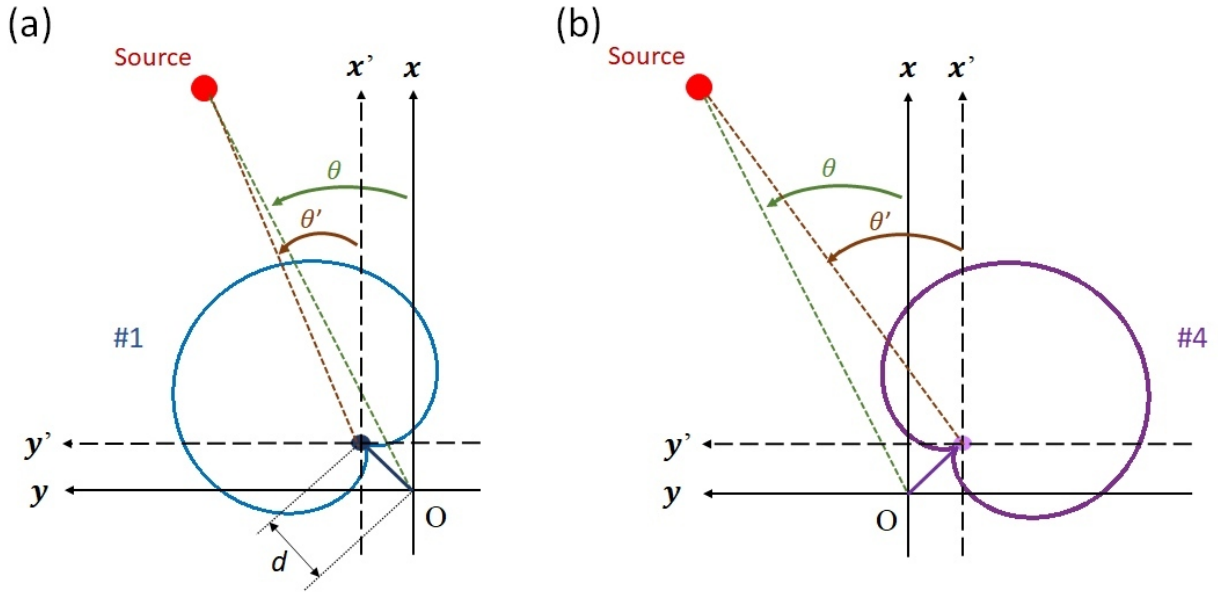


Figure 3.8. (a) The incident angle to capsule 1 is narrower than the actual source angle. (b) The incident angle to capsule 4 is broader than the actual source angle. The disagreement in incident angle such as this illustration occurs between all capsules. This leads to increases of estimation errors.

Signal delays in measurements between capsules due to the distance between them is another factor decreasing the estimation accuracy and precision. Theoretically there is supposed to be no signal delays between capsules since the capsules were assumed to be located at a point together. This could be significantly minimized by having less than the half-wavelength of the highest frequency in interest between the capsules. It is more critical and difficult to have less than the half-wavelength of an incident frequency between capsules for bat biosonar measurements, considering their high sonar call frequency ranges. We achieved the half-wavelength of the highest frequency in interest as the inter-capsule distance, approximately 2 mm ($d \approx 1.2$ mm) for 80 kHz, of our soundfield microphone. Nonetheless, there still exists a non-zero d resulting in possibly multiple source angles claimed by the four capsule responses even after achieving the half-wavelength spacing.

These factors (different incident angles to the capsules and signal delays between the capsules) are affected by a non-zero d. A higher d boosts the negative effects of these factors on the estimation process. This explains why the angle estimation performance decreases as d increases, observed in section 3.3.2. This also explains the disagreement in estimation performance assessment between the statistical study with simulations (section 3.3.2) and the CRLB study (section 3.4.2). The CRLB study does not provide a proper evaluation of our estimation system since the CRLB method does not take into account these factors affecting the system performance. This result also emphasizes the importance of the variable d. In the given situation where d is not able to be zero due to physical constraints, it is crucial to ensure as short d as possible. This ensures the shortest distance between the capsules ($\delta_c \approx 1.63d$), enhancing the estimator performance.

3.6 Conclusion

The current work shows the performance assessment study results of our DOA estimation system based on a soundfield microphone, Ambisonics, and HARPEx. The statistical study of the estimations from the simulations demonstrates a shorter *d* or/and higher SNR is required to achieve higher estimation accuracy and precision. For our system with the soundfield microphone having *d* of 1.2 mm, a SNR_1 higher than 25 dB helps achieving ± 10° tolerance in both accuracy and precision of angle estimations according to the result. The CRLB study fails to deliver an accurate assessment of the estimator performance. This is because the CRLB method does not cover the effects of the factors (associated with *d*) on the performance of our estimation system. Our system is a sound intensity-based estimator that does not rely on the TDOA method the CRLB can be used for. This is the reason we chose to conduct the statistical study with simulations as an alternative way to assess the performance of our estimator. This work demonstrates the importance of the variable *d* since the variable substantially affects the estimator performance. The shortest *d* in a given situation must be achieved in the soundfield microphone design and construction process to attain the highest estimation performance. The result from 3.3.2 also provides guidance on SNRs to be achieved for a desired performance level in a given situation. This work provides insights into designs and measurements considering *d* and SNR when building or/and using the DOA estimation system discussed here for bat biosonar study or any other applications.

Acknowledgments

This research was supported by the Multidisciplinary University Research Initiative (MURI) program from the Office of Naval Research (ONR) Grant No. N00014-17-1-2736.

References

Arlettaz, R., Jones, G., and Racey, P. A. (**2001**). "Effect of acoustic clutter on prey detection by bats," Nature **414**, 742-745.

Arteaga, D. (**2015**). "Introduction to Ambisonics," (Escola Superior Politècnica Universitat Pompeu Fabra, Barcelona, Spain).

Bates, M. E., Simmons, J. A., and Zorikov, T. V. (**2011**). "Bats use echo harmonic structure to distinguish their targets from background clutter," Science **333**, 627-630.

Erling, J. G., Roan, M. J., and Gramann, M. R. (**2007**). "Performance bounds for multisource parameter estimation using a multiarray network," IEEE Transactions on Signal Processing **55**, 4791-4799.

Gerzon, M. A. (**1973**). "Periphony: With-height sound reproduction," Journal of the audio engineering society **21**, 2-10.

Gerzon, M. A. (**1980**). "Practical periphony: The reproduction of full-sphere sound," in *Audio Engineering Society Convention 65* (Audio Engineering Society).

Ghose, K., and Moss, C. F. (**2003**). "The sonar beam pattern of a flying bat as it tracks tethered insects," The Journal of the Acoustical Society of America **114**, 1120-1131.

Hartley, D. J., and Suthers, R. A. (**1989**). "The sound emission pattern of the echolocating bat, E ptesicusfuscus," The Journal of the Acoustical Society of America **85**, 1348-1351.

Jakobsen, L., Ratcliffe, J. M., and Surlykke, A. (**2013**). "Convergent acoustic field of view in echolocating bats," Nature **493**, 93-96.

Kay, S. M. (**1993**). *Fundamentals of statistical signal processing* (Prentice Hall PTR).

Lai, H., and Bell, K. (**2007**). "Cramer-Rao lower bound for DOA estimation using vector and higher-order sensor arrays," in *2007 Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers* (IEEE), pp. 1262-1266.

Lee, H., Roan, M. J., Ming, C., Simmons, J. A., Wang, R., and Müller, R. (**2019**). "High-frequency soundfield microphone for the analysis of bat biosonar," The Journal of the Acoustical Society of America **146**, 4525-4533.

Manikas, A., Sleiman, A., and Dacos, I. (**2001**). "Manifold studies of nonlinear antenna array geometries," IEEE Transactions on Signal Processing **49**, 497-506.

Nehorai, A., and Paldi, E. (**1994**). "Acoustic vector-sensor array processing," IEEE Transactions on signal processing **42**, 2481-2491.

Nielsen, R. O. (**1999**). "Cramer-Rao lower bounds for sonar broad-band modulation parameters," IEEE journal of oceanic engineering **24**, 285-290.

Ortolani, F. (**2015**). "Introduction to Ambisonics," Ironbridge Electronics.

Simmons, J. A., Howell, D. J., and Suga, N. (**1975**). "Information content of bat sonar echoes: recent research on echolocation in bats identifies some of the kinds of information conveyed by echoes of their sonar sounds," American Scientist **63**, 204-215.

Simmons, J. A., Saillant, P. A., Wotton, J. M., Haresign, T., Ferragamo, M. J., and Moss, C. F. (**1995**). "Composition of biosonar images for target recognition by echolocating bats," Neural Networks **8**, 1239-1261.

Surlykke, A., Boel Pedersen, S., and Jakobsen, L. (**2009**). "Echolocating bats emit a highly directional sonar sound beam in the field," Proceedings of the Royal Society B: Biological Sciences **276**, 853-860.

Warnecke, M., Falk, B., and Moss, C. F. (**2018**). "Echolocation and flight behavior of the bat Hipposideros armiger terasensis in a structured corridor," The Journal of the Acoustical Society of America **144**, 806-813.

Chapter 4

High-Frequency Soundfield Microphone and Matched-filter to Reproduce Bat Biosonar

Hyeon Lee[1*], Michael J. Roan[1], Chen Ming[2], James A. Simmons[2]

[1]Department of Mechanical Engineering, Virginia Tech, 460 Old Turner Street, Blacksburg, VA 24061, USA

[2]Department of Neuroscience, Brown University, 185 Meeting Street, Providence, RI 02912, USA

Abstract

The authors introduce a sonar-like system that estimates the angle, range, and velocity of a target using a high-frequency soundfield microphone, spatial audio, and matched-filter techniques to reconstruct bat biosonar. A tetrahedral soundfield microphone was built to measure sounds in the bat sonar frequency range (20-80 kHz). Recorded signals were processed and reproduced using Ambisonics to reconstruct sound fields. An algorithm was built in the estimation system to matched-filter reproduced signals for detecting a target. The matched-filter result provides target information based on sonar fundamentals. Experiments recording transmitted signals from a loudspeaker (emitter) and echoes from an artificial or natural target (either stationary or moving) were conducted. The estimation system processed the signals from the experiments, and estimated the target's angle and range accurately. The authors also performed some simulations with a generated sonar call, benchmarking the big brown bat pulses and generated echoes based on a computed sound level using the sonar equation. The estimation system processed the generated signals as input data that imitate specific scenarios and accurately detected the target's angle, range, and velocity. The simulation results also appear to recommend an SNR level over -3 dB for accurate and precise estimations.

Keywords: biosonar; echolocation; matched-filter, soundfield microphone; Ambisonics; spatial audio.

4.1 Introduction

Bat echolocation behavior has been a topic of extensive research for several decades. One of the primary tools used in studying bat behavior in relation to navigation and hunting behaviors is an acoustic array. Bat transmissions, the corresponding location of the bat, and echo returns from the bats' prey must be monitored simultaneously. Microphone array outputs processed using time difference of arrival (TDOA) enables recording of bat pulses and simultaneous localization of the flying bat (Surlykke and Kalko, 2008; Barchi *et al.*, 2013). TDOA is likely the most widely used method of biosonar (Lee *et al.*, 2019). The method requires the arrival-time difference between signals at the array elements of the microphone array. The time differences among the elements make it possible to compute distances, and obtain hyperbolas representing possible ranges of the bat's location (Bard and Ham, 1999; Gustafsson and Gunnarsson, 2003; Kim *et al.*, 2012). The crossing points of several hyperbolas provides an estimation of the location from which the bat emitted the pulse at the snapshot. Connecting the estimated locations of each snapshot creates the reconstruction of bat's flight path over time. Although this method localizes a flying bat, it does not provide insights into the acoustics of the space as observed by the bat. Limitations on the accuracy of the reconstruction depend directly on the array size and number of array elements, which themselves may be limited by the flight-room size. In this work, we introduce a sound localization system using an alternative method. This is an energy-intensity based method to localize sound (e.g. sonar calls from bats and echoes from targets) and reproduce the sound fields around bats. This system employs spatial audio techniques – a soundfield microphone and Ambisonics signal processing – and wideband matched-filter techniques.

Spatial audio techniques provide estimated of sound source direction of arrival directly from measurements, and enables reconstruction of the sound field on a sphere around the subjects

of interest (bats or/and targets). The approach provides localization of transmitted pulses as well as echoes and provides a unique perspective on data interpretation. Using this approach, investigators can observe the acoustic response of targets, the space (i.e. room or clutter reflections) from the bat's point of view.

The soundfield microphone and Ambisonics technique introduced by Gerzon (Gerzon, 1973; 1980) is a spatial audio technique for recording three-dimensional (3D) audio and reproducing sound fields to provide more realistic aural experience. This technique records and reproduces 3D audio from the perspective of the listener in our case, the bat.

We previously demonstrated a hardware and software system for bat biosonar study using a custom-built ultrasonic soundfield microphone, a signal processing chain based on Ambisonics, and HARPEx (High Angular Resolution Planewave Expansion) (Lee *et al.*, 2019). The system estimated bearing angles of multiple simultaneous targets accurately. The system provides estimates of the direction to a source (azimuth and elevation), relative intensity, and frequency of the sound. The expanded system in this article provides further estimates of target range and velocity. The matched-filter technique widely used in sonar and radar plays a central role in the system to determine target range and velocity. The sound localization system with a soundfield microphone, Ambisonics signal processing, and matched-filter localized static and moving targets by providing the acoustical map of sound fields around a bat directly from the bat's perspective.

The paper is organized as follows: first we review the soundfield microphone and the related Ambisonics-based signal processing. Second, the fundamentals of the matched-filter technique and the matched-filter used in this work are introduced. Next, we provide the data processing procedure: soundfield decomposition using Ambisonics, matched-filtering, and data

interpretation. And lastly results from experiments in various cases that validate the performance of the target localization system are provided.

4.2 Soundfield microphone construction

An ultrasonic soundfield microphone was constructed to record the signals in the Big Brown Bat's (*Eptesicus fuscus*) frequency range of 20-80 kHz, based on the method introduced in our previous work (Lee *et al.*, 2019). The details of the construction process can be found in section IV of (Lee *et al.*, 2019). Four condenser microphones were used (Knowles FG-23629-P16 series) as the four microphone capsules of the tetrahedron soundfield microphone.

The microphone capsules were calibrated using a transfer function approach to assure the capsules have common amplitude and phase responses. The capsules recorded an LFM signal (10-90 kHz, 8 second) emitted from a loudspeaker (Kenwood KFC-XT15ie, Fig. 4.1(a)). Figure 4.1(b) shows the transmitted signal recorded by a commercial standard microphone (G.R.A.S 46DP-1). The transmission response of the loudspeaker was relatively flat up to 60 kHz. An analog-to-digital conversion module (NI 9222, National Instruments Corp) collected the data with a sampling frequency of 500 kHz, for both measurements using the standard microphone and using the capsules. The responses between capsules to the emitted signal were different as shown in Figs. 4.1(c)-(d). We computed transfer functions between the first capsule (the reference) and the other three then used them to correct the differences in far-field magnitude and phase shown in Figs. 4.2(a)-(b). After the relative correction of three capsules to the reference, the responses of all capsules were virtually identical to each other as shown in Fig. 4.2(c).

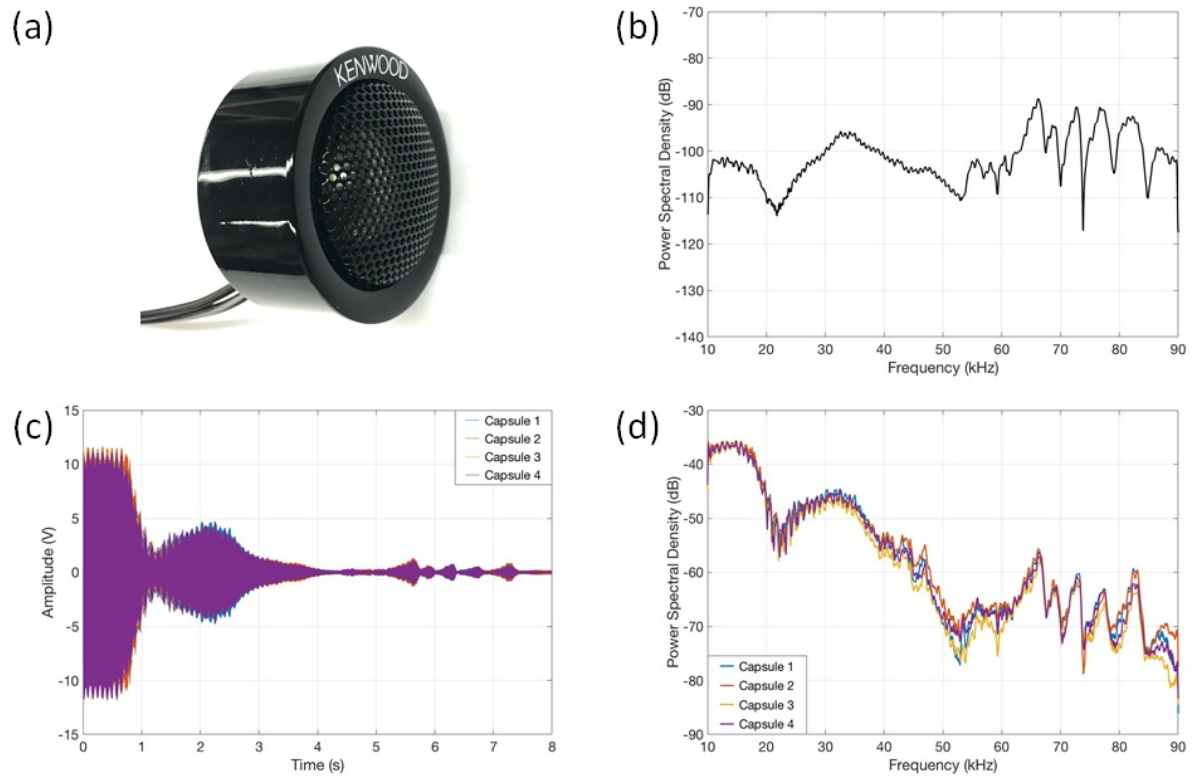


Figure 4.1. (a) A Kenwood tweeter was used for calibration measurement and other tests in this work. (b) A G.R.A.S standard microphone recorded an LFM signal (10-90 kHz, 8 second) emitted from the tweeter to evaluate the transmission performance of the tweeter across the frequency range (the measured signal was band-pass filtered, 10-90 kHz). (c) The four microphone capsules measured the same LFM signal (the measured signal was band-pass filtered, 10-90 kHz). (d) There were significant microphone output voltage differences between the capsules in response across the frequency range.

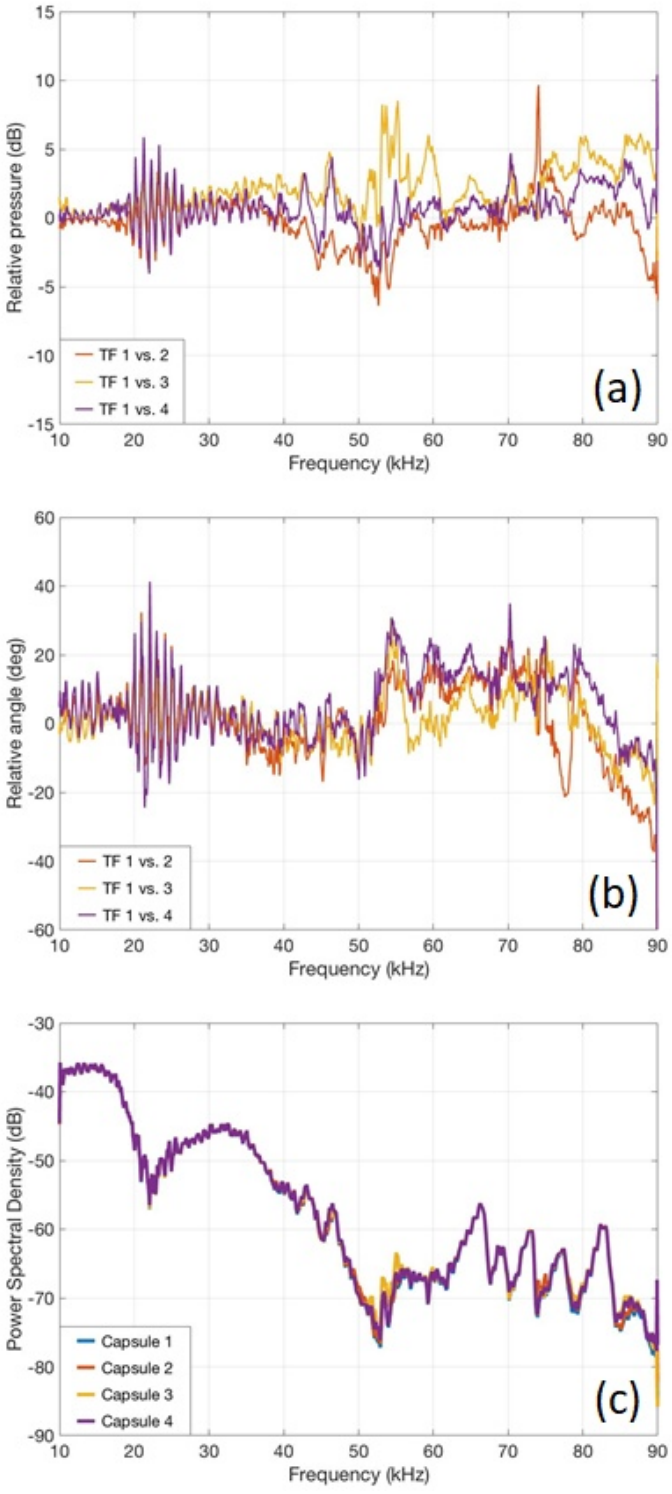


Figure 4.2. The transfer functions between the first capsule (the reference capsule) and the other three in (a) amplitude and (b) phase. (c) The responses of all capsules were virtually identical to each other after correcting the difference of each capsule relative to the reference using the transfer functions.

The capsules were combined as a soundfield microphone as shown in Fig. 4.3(a). The distance between adjacent capsule openings was roughly 2 mm, ensuring a high frequency range of approximately 80 kHz for the first microphone iteration (Fig. 4.3(a)). We constructed a custom four-channel pre-amplifier with a flat frequency response to 100 kHz to power the microphone capsules provide a gain of 20 to the capsule outputs (Fig. 4.3(b)).

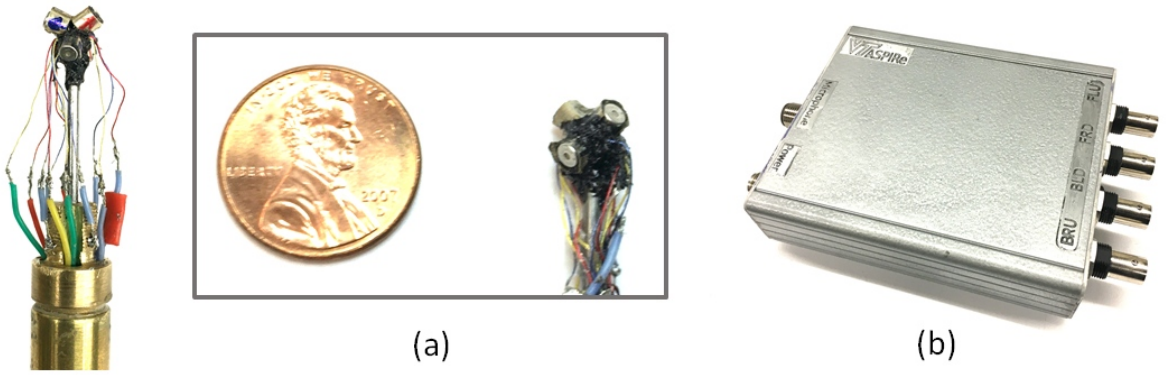


Figure 4.3. (a) The constructed 1st order (four channel) ultrasonic soundfield microphone. (b) The custom-built four-channel pre-amplifier including power supply for the microphone.

4.3 Ambisonics signal processing

Previous work discussed the fundamentals of 1st order Ambisonics (Lee *et al.*, 2019), A-format and B-format, conversion between the two formats, and spherical harmonics. Ambisonics is a technique to codify directional properties of the soundfield into the spherical harmonics (W, X, Y, and Z) of the soundfield given by Eqs. (4.1a)-(4.1d), where θ and φ are azimuth and elevation, respectively (Arteaga, 2015; Ortolani, 2015). Physically, for a 1st order microphone, W is the omnidirectional signal and X, Y, and Z are figure-of-eight microphones aligned with each of the respective axes, giving the directional response along these axes.

$$W(t) = \frac{s(t)}{\sqrt{2}}, \tag{4.1a}$$

$$X(t) = s(t)\cos\theta cos\varphi, \tag{4.1b}$$

$$Y(t) = s(t)sin\theta cos\varphi, \tag{4.1c}$$

$$Z(t) = s(t)\sin\varphi. \tag{4.1d}$$

Equation (4.2) is an expanded version of Eq. (4.1) that provides a way to reproduce signals in arbitrary directions, where *i* is i[th] signal and *N* is the number of signals to be reproduced (Ortolani, 2015). When a signal is measured in B-format, the solution of *B=MS* for *S* provides a reproduced signal in each direction. Because *M* is not necessarily a square matrix, a pseudo-inverse matrix of *M* is required to solve the equation. In short, Ambisonics can decompose a soundfield into multiple directional components S. This can be done with theoretically limitless *N* using Eq. (4.2), to reconstruct the sound fields around the observer (listener), but is realistically limited in terms of resolution. Another way to look at this approach is that one is processing the B-format signal to form virtual directional microphones in the directions (θ_i, φ_i).

$$\overset{B}{\begin{bmatrix} W(t) \\ X(t) \\ Y(t) \\ Z(t) \end{bmatrix}} = \overset{M}{\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & \dots & 1/\sqrt{2} & \dots & 1/\sqrt{2} \\ cos\theta_1 cos\varphi_1 & cos\theta_2 cos\varphi_2 & \dots & cos\theta_i cos\varphi_i & \dots & cos\theta_N cos\varphi_N \\ sin\theta_1 cos\varphi_1 & sin\theta_2 cos\varphi_2 & \dots & sin\theta_i cos\varphi_i & \dots & sin\theta_N cos\varphi_N \\ sin\varphi_1 & sin\varphi_2 & \dots & sin\varphi_i & \dots & sin\varphi_N \end{bmatrix}} \overset{S}{\begin{bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_i(t) \\ \vdots \\ s_N(t) \end{bmatrix}} \tag{4.2}$$

Once the soundfield is decomposed into the directional components *S*, processing can be continued with a matched filter to determine range to the target and Doppler shift. This is the topic of the next section.

4.4 Matched-filter Approach

4.4.1 Wideband ambiguity function (WAF)

An ambiguity function (AF) is the cross-correlation between two waveforms using multiple Doppler (or in the wideband case, time-scale) replicas of the transmitted signal. The AF is used to study waveforms of signals for sonar or radar because it provides the insight of the Doppler (frequency shift) and range resolution of transmitted signals. Since most bats use echolocation signals with broad bandwidths considering their spectral centroids (Sibul and Titlebaum, 1981), the wideband ambiguity function (WAF) was used in this work. The WAF is a function of the Doppler scale factor (η) and time delay (τ) as defined (Kelly and Wishner, 1965; Lin, 1988):

$$\chi_s(\eta,\tau) = \eta^{1/2} \int s(t)s^*\big(\eta(t-\tau)\big)dt, \qquad \text{where } \eta = \frac{c+v}{c-v}, \qquad (4.3)$$

where t is time, v is radial velocity, and c is the propagation speed of sound. The two compared signals are generally a transmitted signal (s) and a received signal (s*), such as echoes. The WAF is calculated using the original signal (the transmitted signal) and modified replicas, with various Doppler shift hypotheses and time delays. The replicas represent potential waveforms of returning signals (echoes) that are distorted (delayed or/and frequency shifted) versions of the transmitted signal due to target motion. The WAF as the function of η and τ forms a surface plot such as Fig. 4.4(a) which is an example of WAF of a linear frequency modulated (LFM) signal (100-20 kHz, 20 ms duration). The peak of WAF occurs with no Doppler shift (η=1) and zero time delay (τ=0) because there is no difference in Doppler shift and time delay when the original signal is compared to itself (the auto-ambiguity function). The auto-ambiguity function provides the localization properties of the transmitted signal. For example, if a long tone were transmitted, the WAF would have multiple peaks in range spaced at the period of the transmitted waveform, but would be quite narrow in terms of Doppler. A broadband, white-noise transmitted signal would have a

"thumbtack" ambiguity function providing high resolution range and Doppler resolution. For a detailed discussion of wideband ambiguity functions see (Young, 2012). In this work, the WAF computation algorithm was implemented in MATLAB. After a desired waveform (transmitted signal) is chosen, the algorithm creates replicas with various frequency shifts in Doppler scale factor range where increment: 0.0001 and range: 0.9713~1.0296 (±5 m/s velocity range) considering the speed of bats and their prey. The algorithm then computes the WAF values between the original signal and the replicas based on Eq. (4.3). This produces the auto-ambiguity function. When the algorithm computes the WAF values between the transmitted signal and the received echo, the matched filter we produce the matched filter output.

4.4.2 Matched-filter

A matched-filter was implemented based on the cross ambiguity function of the transmitted signal and the received echo data (Trees, 2001). The measured return signals (echoes) are cross-correlated with the Doppler hypothesis shifted replicas of the original signal. The output of the matched filter provides a detection statistic that is a function of delay, or range to target, and Doppler, or target velocity (Trees, 2001). For instance, if the Doppler scale factor is 0.7455 and time delay is 5.07 ms for the replica showing the highest correlation value with a received signal (Fig. 4.4(c)), the Doppler scale factor and time delay of the received signal are 0.7455 and 5.07 ms, respectively.

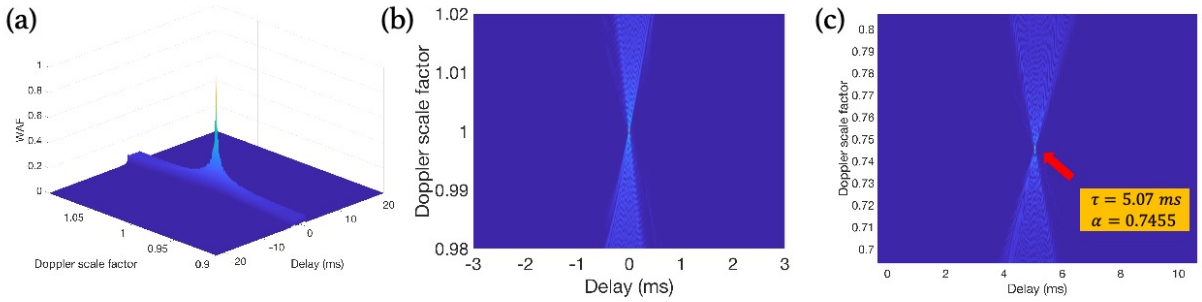


Figure 4.4. (a) An example of WAF plot of an LFM signal (100-20 kHz, 20 ms duration). (b) WAF between the original signal and the exact replica (with neither Doppler shift nor time delay) of the original signal brings a peak at η=1 and τ=0. (c) Matched-filter finds a received signal the most matched replica with specific η and τ values. In this example, the peak formed at η=0.7455 and τ=5.07 ms indicates η and τ values of the matched replica, meaning η and τ of the received signal are 0.7455 and 5.07 ms, respectively.

4.5 Signal processing procedure

The signal processing chain in this paper consisted of three main stages: soundfield decomposition, matched-filtering, and parameter estimation/display processing. Several experimental signals were processed and analyzed through this procedure. The processing chain was implemented in MATLAB.

4.5.1 Soundfield decomposition

The soundfield microphone recorded signals in A-format. These were filtered using digital bandpass filters (20-80 kHz) and converted into frequency domain using the DFT. They were calibrated using the transfer functions of the capsules, converted back into time domain, and put into B-format using Eqs. (4.1a)-(4.1d) from (Lee *et al.*, 2019). Equation (4.2) was used to compute time series for 21,600 (360 azimuths (0-360°) × 60 (±30°) elevations) signals in all directions (the

increments in both azimuth and elevation were 1°). Each time series computed in this stage of the processing chain was then passed to the matched filter.

4.5.2 Matched-filtering

Each direction-dependent time series computed in the previous step was match-filtered with the original transmitted signal and all of its Doppler hypothesis waveforms. For each reproduced signal, the algorithm calculated matched-filter responses of a sliding block at each snapshot throughout the signal in times-series. The length of the block (s* in Fig. 4.5) was the same as the entire length of emitted signal (s in Fig. 4.5), n data points ($n=T_t \times f_s$ where T_t: the duration of the emitted signal and f_s: acquisition sampling frequency). The sliding block covered the first n data points (1st to n^{th} data points) of the signal for the first snapshot. Then the block slid over 1 data point and covered the 2nd to $(n+1)^{th}$ data points for the second snapshot (overlap between a block for a snapshot and next one: n-1 data points). The algorithm calculated the matched-filter response at the j^{th} snapshot ($j \times f_s$ seconds) when the block covered the j^{th} to $(n+j)^{th}$ data points. The block kept moving for the next snapshot until the tail of the block hit the end of the signal. Estimated values of range, angle, elevation, and Doppler shift corresponding to each case (each snapshot block) were acquired using the matched-filter outputs.

4.5.3 Data visualization

The signal of the time-block with the highest matched-filter response was identified as the echo from the target because the highest response indicates the strongest correlation of the received signal with the emitted signal. The algorithm plotted the matched-filter responses on a Cartesian plane (X-Y) to visualize the location (angle and range), and read the estimated values of range,

azimuth, elevation, and Doppler shift as the final step of signal processing. Ranges (R) was computed using an equation, $R = \frac{t_e - t_o}{c}$, where t_e: time the echo front was detected and t_o: time the signal emission started. Azimuth and elevation were determined by reading the azimuth and elevation of the reproduced signal from which the detected echo was. Doppler shift information was delivered from the matched-filter result, and the velocity was computed using the relationship between Doppler shift and velocity (Eq.(4.3)).

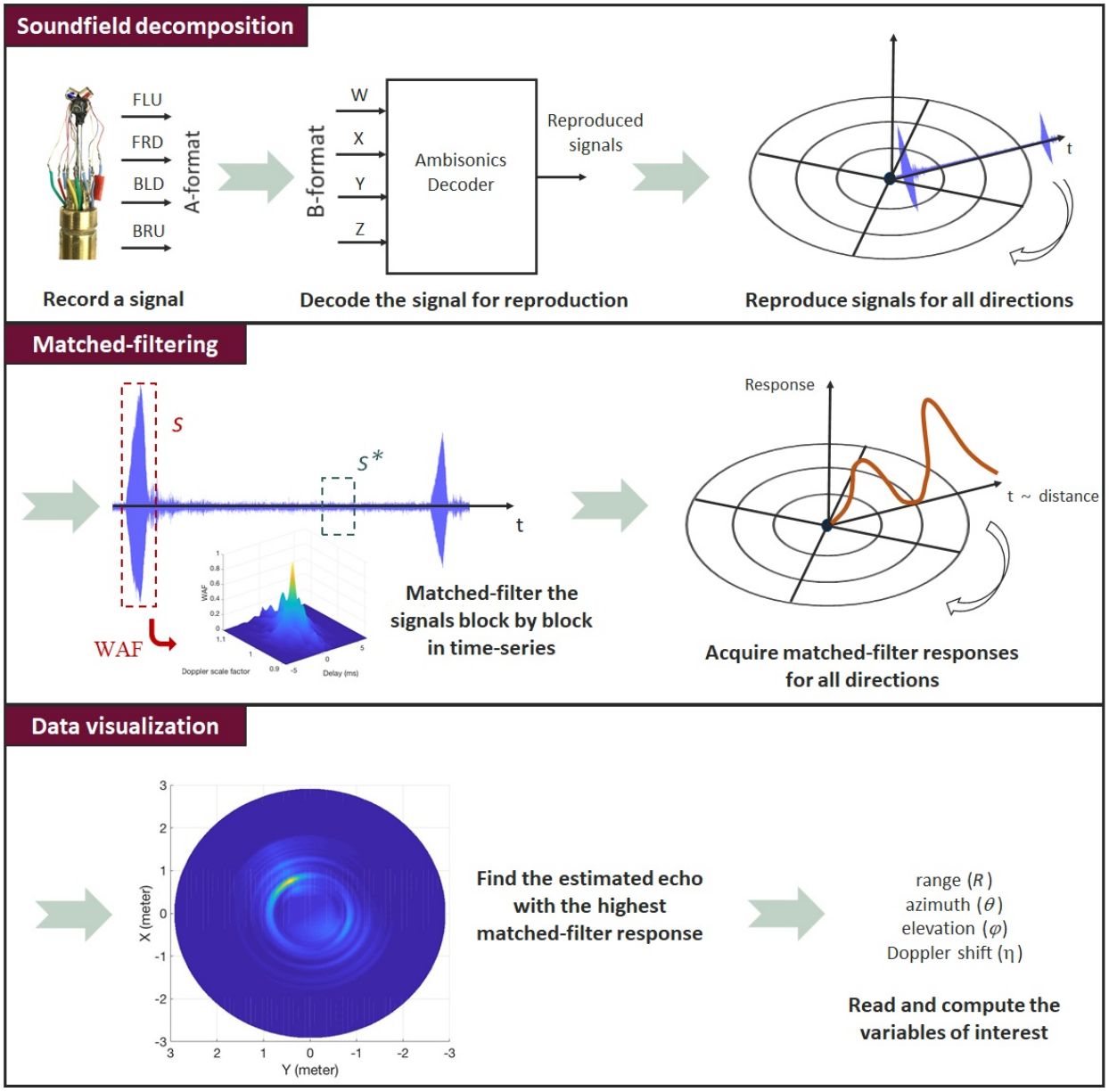


Figure 4.5. Diagram of the signal processing chain. Soundfield microphone recording in A-format is converted to B-format, and decomposed into multiple direction-dependent time series. These time series are passed through the matched filter block by block. Peak matched filter output is used to estimate the target location in azimuth, elevation, range and the normal velocity of the target.

4.6 Validation tests

We implemented several controlled tests with the ultrasonic transmitter (previously mentioned in section 4.2) and artificial target to validate the application capability of the developed system to bat biosonar study. The previously mentioned Kenwood loudspeaker (Fig. 4.1(a)) was mounted on a stand, and placed as the emitter in an anechoic chamber (in the ASPIRe lab at Virginia Tech). The soundfield microphone was also mounted on the stand, the same location as the loudspeaker, mimicking the effect of attaching the microphone near a bat's head. The microphone was approximately 5 cm above the loudspeaker. A 3D printed sound reflector which is a combination of three disks (1 inch diameter) on three principal planes was included in the setup as an artificial target (Fig. 4.6(a)). The target was hung from the ceiling of the chamber. The loudspeaker emitted signals from a DAC (Digital-to-Analog Converter) connected to a Raspberry Pi 2 as the transmission controller at the sampling frequency of 384 kHz, toward the target. We measured the transmitted signals from the emitter and echoes from the target using the soundfield microphone. The ADC module (NI 9222) previously used in section 4.4 recorded the microphone response at the sampling frequency of 500 kHz and 16 bits of resolution. We used this protocol described here for all the following experiments in section 4.6.

4.6.1 Visualization of frequency shift

The target, approximately 60 cm away from the loudspeaker, was swung back and forth in a pendulum motion normal to the transmitter/receiver. The loudspeaker continuously emitted a tone of 50 kHz for 10 seconds. The transmitted signal and echoes were recorded. We plotted a spectrogram of the measured signal (W in B-format) as shown in Fig. 4.6(b). The straight yellow line stretched from the start to the end of the period represents the frequency of the tone. The

sinusoidal line appeared around the 50 kHz tone represents the frequency variation of the echoes, reflecting the pendulum motion of the target. It clearly illustrates the frequency shift (Doppler shift) of the echoes from the swinging target. Since frequency shifted in a range of -200 ~ 200 Hz, the velocity of the target varied in a range of -1.4 ~ 1.4 m/s using the relationship between velocity and Doppler shift $\left(\frac{\Delta f}{f_o} = \frac{\Delta v}{c}\right)$, where f: frequency and f_o: original frequency.

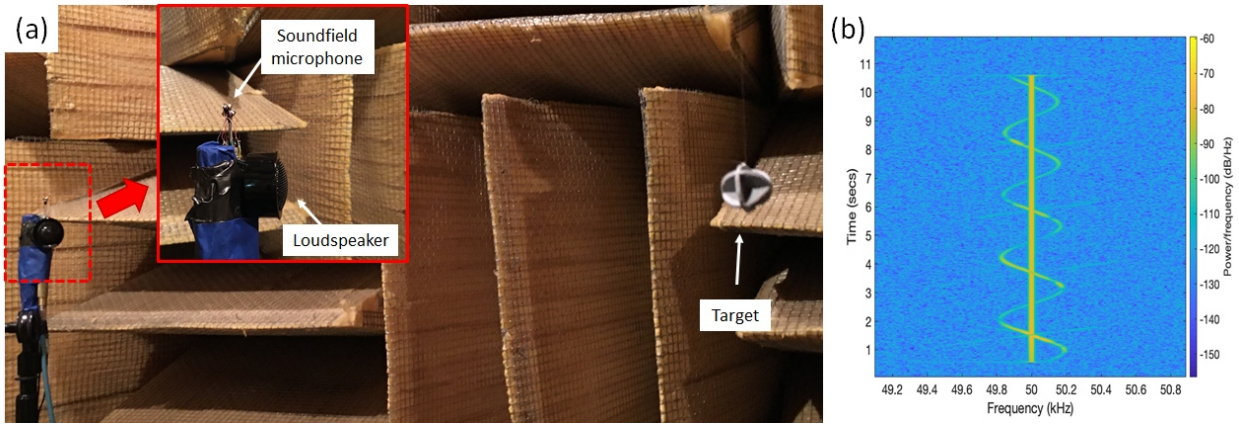


Figure 4.6. (a) The Kenwood tweeter and soundfield microphone were placed at the same spot on a stand, while a 3D printed artificial target was hung on from the chamber ceiling approximately 60 cm away from them. The soundfield microphone measured a 50 kHz tone emitted from loudspeaker, and echoes reflected from the swinging (normal to the transmitter/receiver) target for 10 seconds. (b) A spectrogram of the measured signal clearly illustrates the frequency shift (Doppler shift) of the echoes from the pendulum motion of the target.

4.6.2 Static target

The target, hung approximately 1 m away from the loudspeaker and at a 20° angle, remained still (Fig. 4.7(a)). The loudspeaker, which was on the same level as the target, emitted a 3 millisecond LFM (22-40 kHz) signal with a harmonic of 44-80 kHz. We reached the length and frequency range of the signal considering the big brown bat's terminal (or buzz) calls for targets

in short ranges (Simmons *et al.*, 1995; Parsons *et al.*, 1997). The matched-filter used the recorded 3 ms chirp as the replica. The 15 ms window immediately following the chirp was taken from each reproduced signal corresponding to each direction, as the received signal for matched-filtering. The matched-filter responses of the reproduced signals were plotted on the X-Y Cartesian plane (Fig. 4.7(b)). Figure 4.7(b) shows the brightest color indicating the location from which the signal (echo) with the highest matched-filter response comes. The analysis through the data processing algorithm estimated that the target was in 1.098 m range and at 34° angle with zero elevation and 0.7 m/s velocity. The estimation result was fairly accurate for every variable.

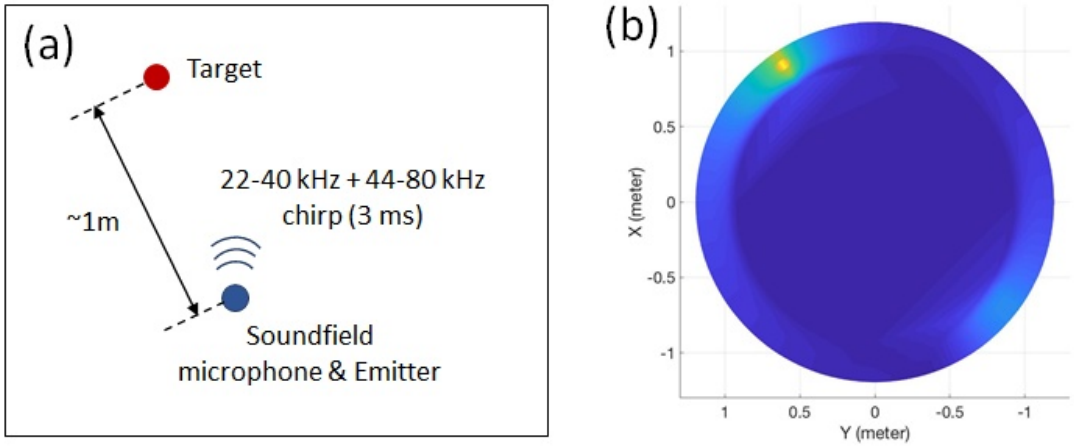


Figure 4.7. (a) The soundfield microphone collocated with the emitter recorded a 3 ms chirp from the emitter and echo from the still target. (b) The matched-filter responses of the reproduced signals clearly point a direction, estimating the angle (34°) and range (1.098 m) of the target.

4.6.3 Moving target

The exact experimental setup and signal processing method for the static target test was used for a moving target test. The target was swinging back and forth in the distance direction during the measurement (Fig. 4.8(a)). The loudspeaker emitted a 3-ms chirp every 18 ms for a second. An 18-ms window, from the start point of a 3-ms chirp to the start point of the next chirp,

in the raw data in time-series was taken as a frame. We analyzed the first six frames and plotted the estimated angle and range of the target in the frames as black dots in Fig. 4.8(b). The numbers next to the dots indicate the corresponding frame numbers. The estimation result demonstrates a smooth trajectory of the swinging target. The inset of Fig. 4.8(b) describes the oscillation, clearly showing the attenuation trend of traveling distance through the frames. The magnitude of traveling distances between the frames is also reasonable considering the estimated velocity range (±1.4 m/s) of the swinging target from Fig. 4.6(b).

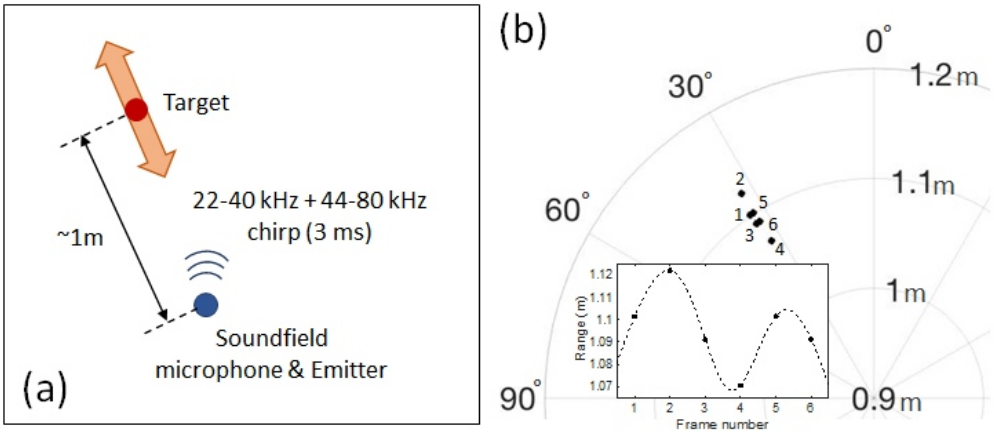


Figure 4.8. (a) The soundfield microphone measured 3 ms chirps (in 18 ms cycle) and echoes from the swinging target for 1 second. (b) The angle and range estimation result of the first six frames (each number next to a dot denotes the corresponding frame) shows a smooth trajectory of the swinging target. The inset illustrates the oscillation of the target with an attenuating sinusoidal curve.

4.6.4 Natural target

We conducted the same experiments as those from 4.6.2 and 4.6.3, but using a natural target instead of an artificial target. We selected a Madagascar hissing cockroach as the target (Fig. 4.9(a)). The cockroach suspended from the ceiling remained still for one experiment, and swung back and forth in the radial direction for the other experiment, respectively (Fig. 4.9(b)). Figure

4.9(c) shows the matched-filter response of reproduced signals in various directions from the static target experiment. The result shows as clear estimation as in Fig. 4.7(b) when the artificial target was used, indicating the range and angle of the target are 1.0948 m and 28°, respectively. The angle and range estimation of the moving target also illustrates the pendulum motion of the target as shown in Fig. 4.9(d). The inset of Fig. 4.9(d) shows a trajectory of sinusoidal oscillation although some data points are slightly off the trend.

The reasonable estimation results from the experiments in this section validate that our estimation system using the matched-filter approach with an artificial emitter can detect the angle and range of either artificial or natural targets regardless of their movement.

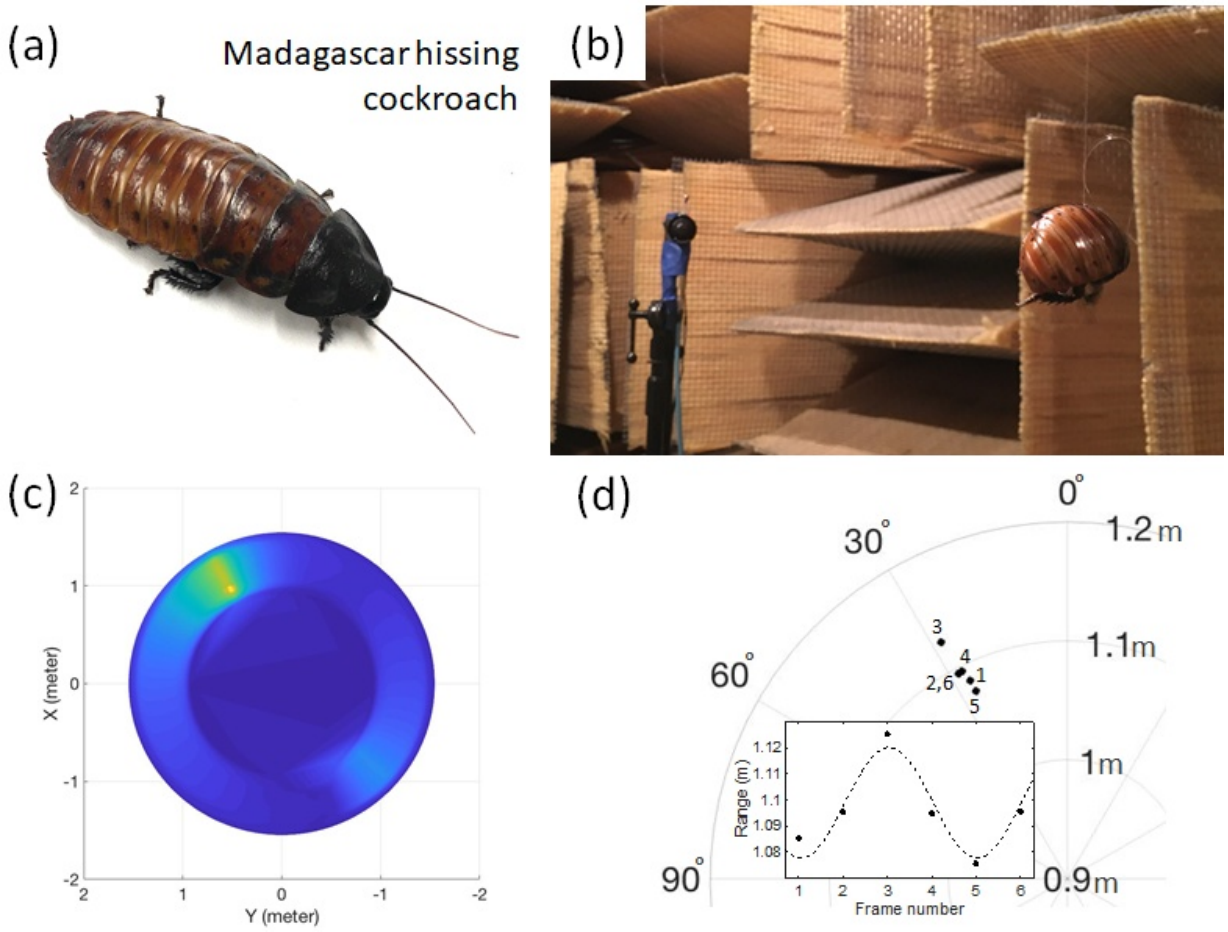


Figure 4.9. (a) A Madagascar hissing cockroach having sufficient sound reflection area (approximately 2-inch long) and high reflectivity with a hard exoskeleton (b) was used as the target for static target and moving target experiments. (c) The matched-filter response of reproduced signals estimated 1.0948 m and 28° as the range and angle of the target, respectively. (d) The trajectory of the estimated ranges of the moving target illustrates a sinusoidal oscillation reflecting the pendulum motion of the target.

4.7 Validation simulations

In order to test the signal processing chain in a controlled manner, simulations of bat transmission/reception were developed. In these simulations, a target was placed (either static or

moving) at a 30° angle from the axis of a bat's central broadcast beam (x-axis). The simulated soundfield microphone was situated 1 cm directly ahead of the bat (i.e. at 0 degrees) (Fig. 4.10(a)).

Bat pulses and echoes were generated with realistic sound pressure levels using the sonar equation (Eq. (4.4)),

$$EL = SL - 2TL + TS - NL + AG - DL \qquad (4.4)$$

where *EL*: echo level, *SL*: source level, *2TL*: two-way transmission loss, *TS*: target strength, *NL*: noise level, *AG*: array gain, and *DL*: directionality loss. To acquire the sonar call level of the big brown bat, as the source level (*SL*), we first averaged the sound levels of 27 independent sonar calls of a bat (Fig. 4.10(b)) that are recorded approximately 10 cm away from the bat in a soundproof room at Brown University (Barchi *et al.*, 2013). Then, the computed sound level for a location 1 cm away from the bat using the inverse distance law was taken as the source level (*SL* = 129 dB). The sensitivity of the microphone capsules used in the calculation is 22.4 mV/Pa (-53 dB @ 74 dB SPL). This bat sonar call measurement result shows a consistency with previously reported measurements (Surlykke *et al.*, 1999; Surlykke and Kalko, 2008; Surlykke *et al.*, 2009; Madsen and Surlykke, 2013). The two-way transmission loss is the sound level drop during a round-trip of a transmitted pulse through a the medium. The loss of 42 dB was used as *2TL* from Fig. 2 of (Madsen and Surlykke, 2013) for a target 1 m away from a bat assuming a 25°C room temperature and 60% humidity in the air. A target strength of -20 dB was taken for the *TS* in the simulations assuming the target is a small moth (Surlykke *et al.*, 1999; Surlykke and Kalko, 2008). The *NL* was acquired by analyzing a pure noise section of the data from Fig. 4.10(b), *NL* = 61 dB. *NL* contains all of the acoustical, mechanical, and electrical noises. As for the *AG*, we took 6 dB simply considering a double gain like the gain for bats with two ears. The *DI* takes into account the variation of sonar call intensity depending on the angle from the axis of the central broadcast

beam. Considering the 30° angle of the target in the simulations, 6 dB was taken as the *DI* value according to bat sonar beamwidth measurements from literature (Hartley and Suthers, 1989; Ghose and Moss, 2003; Bates *et al.*, 2011). The sonar equation with the all values for the terms yields 6 dB as the *EL*. The computed *EL, NL* and *SL* were used for all simulations conducted in this section.

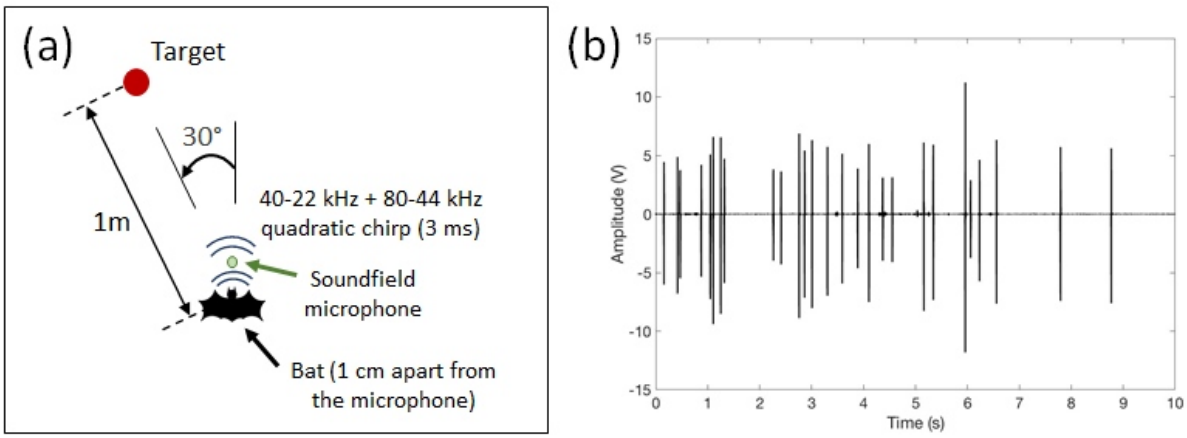


Figure 4.10. (a) Geometry of the emitter (bat), soundfield microphone, and target for all simulations in section 4.7. (b) A total of 27 pulses from a recording of a big brown bat approximately 10 cm away from the soundfield microphone. *SL* and *NL* were calculated using this data.

The simulated chirp was a 3-ms quadratic FM, 40-22 kHz, signal with a harmonic of 80-44 kHz and the sound level of the *SL* (Eqs. (4.5a)-(4.5c)).

$$s = A_{c1}\cos\left[2\pi f_{i1}\left(\frac{f_{f1}}{f_{i1}}\right)^{t/T} t\right] + A_{c2}\cos\left[2\pi f_{i1}\left(\frac{f_{f2}}{f_{i2}}\right)^{t/T} t\right], \quad (4.5a)$$

$$A_{c1} = \sqrt{2}\left(20\mu Pa * 10^{\frac{SL}{20}}\right) * microphone\ sensitivity * preamp\ gain, \quad (4.5b)$$

$$A_{c2} = \sqrt{2}\left(20\mu Pa * 10^{\frac{SL-5}{20}}\right) * microphone\ sensitivity * preamp\ gain, \quad (4.5c)$$

where f_{i1}: 40 kHz, f_{f1}: 22 kHz, f_{i2}: 80 kHz, f_{f2}: 44 kHz, t: time, T: 3 ms, *microphone sensitivity*: 22 mV/Pa, and *preamp gain*: 17. There was a 5 dB difference in sound level between the first and second harmonics considering the target's angle (30°) for both the chirp signal and echoes. This is

because the variation pattern of bat sonar call intensity in angle changes with frequency; the intensity of the second harmonic is weaker than the first harmonic, with the angle from the axis of the central broadcast beam (Hartley and Suthers, 1989; Ghose and Moss, 2003; Bates *et al.*, 2011). The generated chirp signal and echoes were windowed using a Hanning window. While echoes were also generated using Eq. (4.5a), the amplitudes of the first (A_{e1}) and second (A_{e2}) harmonics were determined using the SNR (signal-to-noise ratio) equation (Eq.(4.6a)) and Eqs. (4.6b)-(4.6d).

$$SNR = 10log\left[\left(\frac{A_{RMS,signal}}{A_{RMS,noise}}\right)^2\right], \qquad (4.6a)$$

$$A_{e1} = \sqrt{2}\left(A_{RMS,noise} * \sqrt{10^{\frac{EL}{10}}}\right), \qquad (4.6b)$$

$$A_{e2} = \sqrt{2}\left(A_{RMS,noise} * \sqrt{10^{\frac{EL-5}{10}}}\right), \qquad (4.6c)$$

$$A_{RMS,noise} = \left(20\mu Pa * 10^{\frac{NL}{20}}\right) * microphone\ sensitivity * preamp\ gain. \qquad (4.6d)$$

For the echoes of moving targets, the signal length of echoes was modified to include the Doppler (or time-scale) effect in the same manner used in 4.4.1. White Gaussian noise of 61 dB (*NL*) was added into signals. A signal consisted of the chirp, a pure noise section, and an echo. The length of the pure noise section was determined depending on the distance from bat to target. Simulated transmissions and related echoes were generated in B-format using Eqs. (4.1a)-(4.1d). The algorithm processed the signals, and estimated the angle, range, and velocity of the target. A total of 100 simulations were performed for each case for statistical analysis. RMSE (root mean square error) was computed to judge the precision of estimations.

4.7.1 Static target

A static target was placed 1 m away from a bat and on the same elevation of the bat in this simulation. The estimation algorithm found the accurate angle of target (30°) in every individual simulation (Fig. (4.11a)). Statistical analysis result showed estimations were accurate in both range (median: 0.98 m and RMSE: 0.03 m) and velocity (median: 0.03 m/s m and RMSE: 0.02 m/s).

4.7.2 Moving target

A target was moving directly toward a bat at a speed of 1 m/s (velocity, v = -1 m/s) and on the same elevation as the bat. Due to the pulse repetition period, there were three opportunities for detection as the target closed on the bat. The target was sequentially detected at the positions of 100 cm, 97.7 cm, and 95.7 cm away from the bat. In this situation, the bat detects a moving target with its first chirp when the target was 1 m away, emits the second chirp in 20 ms, and emits the third chirp in another 20 ms. The estimated target angle was 30° for each detection. The estimated velocities at all three spots were -1 m/s with various error rates (RMSE: 0.03 m/s, 0.04 m/s, and 0.05 m/s in order). The estimated ranges of the consecutive incidents were 98.4 cm (RMSE: 3.8 cm), 96.6 cm (RMSE: 3.6 cm), and 93.7 cm (RMSE: 3.9 cm). The echolocated target positions form a clear moving path of the target in the simulations as shown in Fig. 4.11(b).

4.7.3 Echo level (*EL*) threshold for the estimation system

We studied the effect of the *EL* (or SNR) on the performance of our estimation system using a simulation scenario where a target is 1 m away from a bat and moving directly toward the bat at a speed of 1 m/s (v = -1 m/s). Various noise levels (0, -3, -6, -9, -12, -15, and -18 dB) were used as the *EL*. While the system continued accurate estimation of the angle of the target (30°),

estimates of range and velocity had increasing error with decreasing *EL* as shown in Fig. 4.11(c). As the *EL* drops to -18 dB the RMSE increases to 8.5%. The velocity estimation is accurate down to -12 dB. However, the RMSE (the precision error) of the velocity estimation exceeds 30% at -6 dB. Based on the result, the echo level (*EL*) of over -3 dB is required to achieve accurate and precise estimations (below 15% RMSE for both range and velocity).

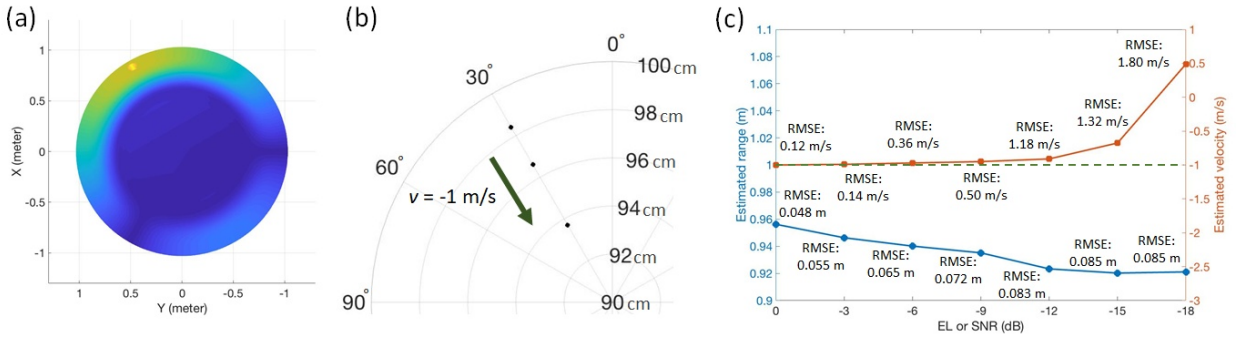


Figure 4.11. (a) The matched-filter response of the static target simulation indicates approximately 0.98 m and 30° as estimations in range and angle, respectively. (b) The estimated range and angle of a moving target with a velocity of -1 m/s constructs a clear trajectory of the target in the simulations. (c) Range and velocity estimations gradually diverge from the actual values as the *EL* decreases; two plots run off the guideline (the dotted line) as the *EL* drops. The range estimation remains both high precision (RMSE < 9%) and accuracy (difference between the estimated and actual values < 9%), whereas the velocity estimation remains only high accuracy down to -12 dB (estimated velocity: -0.91 m/s at -12 dB). The *EL* of over -3 dB must be achieved for accurate and precise estimations.

4.8 Conclusions and future work

In this work, we developed a system (hardware/software) to reproduce bat biosonar, which is an upgraded version of the angle estimation system from our previous work (Lee *et al.*, 2019). The built system includes an upgraded ultrasonic soundfield microphone, and a matched-filter and

Ambisonics signal processing algorithm. The algorithm reproduced high-frequency sound fields the soundfield microphone recorded, through the developed signal processing procedure. Analyzing the reproduced sound fields, the system estimated angles, ranges, and velocities in a fairly accurate level. In the experiments with the ultrasonic emitter, the system showed fairly accurate angle and range estimations of the artificial and natural targets, although velocity estimations were slightly off. Relatively unstable transmissions of short LFM signals could be the cause. The simulation results demonstrated that accurate and precise estimations in every parameter using the matched-filter approach are feasible as long as the *EL* is over -3 dB.

Testing of the estimation system using real bats (real bat sonar calls and their echoes) should be a follow-on future work. Real bat experiment results would provide insightful information for the comparison between signal transmissions/receptions by the man-made ultrasonic emitter and a bat. A compact soundfield microphone package (including power supply, preamplifier, and data transmitter or recorder) onboard a flying bat should be further developed for signal measurements in close proximity to a bat's head. This enables the collection of signal data and the ability to reproduce sound fields directly from the bat's view, helping to understand the acoustics around the bat and its adaptation behavior associated with the acoustical environment changes. Experiments with multiple targets could be also another useful future work to find the maximum number of sources the system is capable of discriminating as a gauge of its estimation capacity.

To summarize, this work introduced a matched-filter approach alongside spatial audio techniques for bat biosonar. The developed estimation system based on the techniques provides information about the angle, range, and velocity of a target(s) similar to man-made sonar. This

brings a possible application of the developed system using a vector acoustic sensor (the soundfield microphone), spatial audio signal processing technique, and matched-filter to man-made sonars.

Acknowledgments

We are grateful for Ian Sandum, M.S. (Department of Entomology, Virginia Tech), for his technical assistance on the preparation and implementation of the natural target experiments in VI D. This research was supported by the Multidisciplinary University Research Initiative (MURI) program from the Office of Naval Research (ONR) Grant No. N00014-17-1-2736.

References

Arteaga, D. (**2015**). "Introduction to Ambisonics," (Escola Superior Politècnica Universitat Pompeu Fabra, Barcelona, Spain).

Barchi, J. R., Knowles, J. M., and Simmons, J. A. (**2013**). "Spatial memory and stereotypy of flight paths by big brown bats in cluttered surroundings," Journal of Experimental Biology **216**, 1053-1063.

Bard, J. D., and Ham, F. M. (**1999**). "Time difference of arrival dilution of precision and applications," IEEE transactions on Signal Processing **47**, 521-523.

Bates, M. E., Simmons, J. A., and Zorikov, T. V. (**2011**). "Bats use echo harmonic structure to distinguish their targets from background clutter," Science **333**, 627-630.

Gerzon, M. A. (**1973**). "Periphony: With-height sound reproduction," Journal of the audio engineering society **21**, 2-10.

Gerzon, M. A. (**1980**). "Practical periphony: The reproduction of full-sphere sound," in *Audio Engineering Society Convention 65* (Audio Engineering Society).

Ghose, K., and Moss, C. F. (**2003**). "The sonar beam pattern of a flying bat as it tracks tethered insects," The Journal of the Acoustical Society of America **114**, 1120-1131.

Gustafsson, F., and Gunnarsson, F. (**2003**). "Positioning using time-difference of arrival measurements," in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).* (IEEE), pp. VI-553.

Hartley, D. J., and Suthers, R. A. (**1989**). "The sound emission pattern of the echolocating bat, E ptesicusfuscus," The Journal of the Acoustical Society of America **85**, 1348-1351.

Kelly, E., and Wishner, R. (**1965**). "Matched-filter theory for high-velocity, accelerating targets," IEEE Transactions on Military Electronics **9**, 56-69.

Kim, E., Kang, C., Jung, H.-S., and Yang, Y.-M. (**2012**). "TDOA-based distance estimation for mobile beacon-assisted localization in large-scale sensor networks," in *SENSORS, 2012 IEEE* (IEEE), pp. 1-4.

Lee, H., Roan, M. J., Ming, C., Simmons, J. A., Wang, R., and Müller, R. (**2019**). "High-frequency soundfield microphone for the analysis of bat biosonar," The Journal of the Acoustical Society of America **146**, 4525-4533.

Lin, Z. b. (**1988**). "Wideband ambiguity function of broadband signals," The Journal of the Acoustical Society of America **83**, 2108-2116.

Madsen, P., and Surlykke, A. (**2013**). "Functional convergence in bat and toothed whale biosonars," Physiology **28**, 276-283.

Ortolani, F. (**2015**). "Introduction to Ambisonics," Ironbridge Electronics.

Parsons, S., Thorpe, C. W., and Dawson, S. M. (**1997**). "Echolocation calls of the long-tailed bat: a quantitative analysis of types of calls," Journal of mammalogy **78**, 964-976.

Sibul, L., and Titlebaum, E. (**1981**). "Volume properties for the wideband ambiguity function," IEEE Transactions on Aerospace and Electronic Systems, 83-87.

Simmons, J. A., Saillant, P. A., Wotton, J. M., Haresign, T., Ferragamo, M. J., and Moss, C. F. (**1995**). "Composition of biosonar images for target recognition by echolocating bats," Neural Networks **8**, 1239-1261.

Surlykke, A., Filskov, M., Fullard, J. H., and Forrest, E. (**1999**). "Auditory relationships to size in noctuid moths: bigger is better," Naturwissenschaften **86**, 238-241.

Surlykke, A., Ghose, K., and Moss, C. F. (**2009**). "Acoustic scanning of natural scenes by echolocation in the big brown bat, Eptesicus fuscus," Journal of Experimental Biology **212**, 1011-1020.

Surlykke, A., and Kalko, E. K. (**2008**). "Echolocating bats cry out loud to detect their prey," PLoS one **3**.

Trees, H. L. (**2001**). *Detection, Estimation, and Modulation Theory: Part III. Radar-sonar Signal Processing and Gaussian Signal in Noise* (John Wiley & Sons, Incorporated).

Young, R. K. (**2012**). *Wavelet theory and its applications* (Springer Science & Business Media).

Chapter 5

Conclusion

5.1 Overall conclusions

An ultrasonic tetrahedral soundfield microphone was designed and constructed to capture sounds in the high-frequency ranges of bat echolocation signals. Ambisonics and HARPEx algorithm were programed in MATLAB for signal processing and DOA estimations. The hardware (soundfield microphone) and software (Ambisonics and HARPEx) system measured static sources, moving sources, and a flying bat, and provided accurate estimates of bearing angles to the sources.

The performance of the estimation system using the soundfield microphone, Ambisonics, and HARPEx was assessed by a statistical study of simulations. A signal model was built to capsule responses to a virtual monopole emitting a 3-ms chirp with two harmonics (40-22 kHz and 80-44 kHz) at an angle of 30°. The accuracy and precision represented by medians and RMSEs, respectively, drop as d increases. To achieve ± 10° tolerance in both accuracy and precision using the estimation system, a SNR_1 higher than 25 dB is required. The CRLB study result showed a disagreement with the statistical study result in the effect of d. This is because the CRLB method

which is for TDOA-based systems does not cover the effects of the two factors, different incident angles to the capsules and signal delays between capsules, caused by a non-zero *d* on the angle estimation.

Another version of an estimation system using a matched-filter approach was introduced. It is a sonar-like system that reproduces sound fields and detects a target using matched-filter responses of the reproduced transmission and echo signals. The soundfield microphone design was upgraded by having 2 mm for the inter-capsule spacing, giving the first microphone iteration a frequency range up to approximately 80 kHz. The system measured LFM signals transmitted from an artificial emitter (loudspeaker) and echoes from an artificial (the 3D printed reflector) or natural target (the Madagascar hissing cockroach). It then processed the measured data and provided accurate estimates of the target's range and direction for each case. In the simulations with the capsule responses to the virtual bat pulse (3 ms, and two harmonics: 40-22 kHz and 80-44 kHz) and echo (from a target at a 30° angle), the system accurately estimated the range, direction, and velocity of the target. Using the estimation system, a SNR_1 higher than -3 dB for echoes is required to achieve accurate and precise estimations (below 15% RMSE for all parameters).

5.2 Original contributions

1. An ultrasonic first-order soundfield microphone was designed and constructed to measure bat echolocation signals and their echoes. Approximately 2 mm for the distance between the microphone capsules was achieved, providing the first microphone iteration a frequency range up to approximately 80 kHz. This enables the microphone to capture sounds in the frequency range

of 20-80 kHz which is the echolocation signal frequency range of our study subject, the big brown bat (*Eptesicus fuscus*). A custom-built four-channel pre-amplifier magnifies the capsule response signals (approximately 17 times), enhancing SNR in measurements.

2. Ambisonics was introduced to process and analyze the bat echolocation signals measured by the soundfield microphone. It handles signals in time- or/and frequency-domains. It reproduces the sound fields around the signal measurement point containing the directional properties (azimuth and elevation) of the measured signals.

3. HARPEx, a spatial audio decoding technique, was introduced to find dominant signals and their directions in bat biosonar studies. This approach is used as a direction estimation method, and provides accurate estimates of DOA for up to two simultaneous sources.

4. Another DOA estimation system using the matched-filter approach alongside the spatial audio measurement and processing techniques was developed for bat biosonar studies. The system estimates the direction, range, and velocity of a target(s) using the sonar fundamentals and the matched-filter responses of reproduced signals in various directions.

5. Performance assessments of the estimation systems were delivered using simulations. This provides insights into designs and measurements considering d and SNR to be achieved when building or/and using the estimation systems for bat biosonar study or any other applications. The assessment result also emphasizes the importance of ensuring a shorter distance between

microphone capsules (associated with *d*) to enhance the performance of the estimators using ultrasonic soundfield microphones.

In summary, the primary contribution of this work is the introduction of spatial audio techniques, which are widely used for 3D audio applications, for bat biosonar. This work proposes two feasible measurement and analysis methods/systems to provide the visualization of sound fields from a bat's perspective, which is the main goal of this research. Various experiments and simulations demonstrated the developed methods/systems to detect a bat or target conveying diverse information (direction, range, and velocity). This proves the proposed methods/systems can be independently or complementarily (to the currently existing methods) used for bat echolocation studies potentially in various ways.

5.3 Limitations and future work

Measuring sonar signals in close proximity to a bat's head is the ultimate goal in measurements because it enables us to collect signal data and reproduce sound fields directly from the bat's view. The soundfield analysis from the bat's view would be useful in understanding the acoustics around the bat and its adaptation behavior associated with the acoustics. To achieve this goal, the development of a compact soundfield microphone package (including power supply, preamplifier, and data transmitter or recorder), such as telemetry microphones (Hiryu *et al.*, 2005; Hiryu *et al.*, 2007) is essential. An onboard high-speed video camera alongside the microphone will be also required to match physical adaptations with the changes in the soundfield. The methods developed in the current work will be the foundations of this suggested future work.

For the second estimation system using the matched-filter approach, experiments with multiple sound sources could be useful. The maximum number of sources the system is capable of discriminating can be found to gauge its capacity in estimation. Experiments using bats are also recommended to validate the performance of the system further. The motive for this work was to provide complementary or alternative methods to TDOA method. In parallel to the comparison of the developed matched-filter approach/system with TDOA method, its comparison to FDOA (frequency difference of arrival) method could be also a useful study since the system analyzes and uses Doppler shift for estimations.

Improving transmission quality is an ongoing goal. To build a sonar system closely imitating the bat sonar, a transmitter that can clearly emit short and wideband signals is essential. The transmitter is further required to emit quadratic and downward sweep chirps just as bat FM echolocation signals are. Finding higher quality components for the transmission system or/and improving transmission process would enhance the transmission quality.

Showing some exemplary methods that are combinations of the methods introduced in this work and the current TDOA based methods could be another useful work. It would be a guidance on feasible ways one can adapt for his/her own study. It could also provide insights into developing new ways of combinations or finding another applications.

Outdoor measurements would be beneficial to study more realistic situations for bats. Since any indoor space is limited, the signals we can observe from bats inside are limited to short range sonar calls. The soundfield microphone onboard a bat is also advantageous to open-field measurements since there is no spatial limit for the microphone unlike microphone arrays, either inside and outside.

References

Hiryu, S., Hagino, T., Riquimaroux, H., and Watanabe, Y. (**2007**). "Echo-intensity compensation in echolocating bats (Pipistrellus abramus) during flight measured by a telemetry microphone," The Journal of the Acoustical Society of America **121**, 1749-1757.

Hiryu, S., Katsura, K., Lin, L.-K., Riquimaroux, H., and Watanabe, Y. (**2005**). "Doppler-shift compensation in the Taiwanese leaf-nosed bat (Hipposideros terasensis) recorded with a telemetry microphone system during flight," The Journal of the Acoustical Society of America **118**, 3927-3933.