8-1-2009

Parametric Model of Head Related Transfer Functions Based on Systematic Movements of Poles and Zeros with Sound Location for Pole/Zero Models

Bahaa W. Al-Sheikh Hussein
University of Denver

PARAMETRIC MODEL OF HEAD RELATED TRANSFER FUNCTIONS BASED ON SYSTEMATIC MOVEMENTS OF POLES AND ZEROS WITH SOUND LOCATION FOR POLE/ZERO MODELS

_____

A Dissertation

Presented to

The Faculty of Engineering and Computer Science

University of Denver

_____

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

_____

by

Bahaa W. Al-Sheikh Hussein

August 2009

Advisor: Mohammad A. Matin

Co-Advisor: Daniel J. Tollin

Author: Bahaa W. Al-Sheikh Hussein
Title: PARAMETRIC MODEL OF HEAD RELATED TRANSFER FUNCTIONS BASED ON SYSTEMATIC MOVEMENTS OF POLES AND ZEROS WITH SOUND LOCATION FOR POLE/ZERO MODELS.
Advisor: Mohammad A. Matin
Co-advisor: Daniel J. Tollin
Degree Date: August 2009

ABSTRACT

Directional transfer functions (DTFs), the directional components of the Head Related Transfer Functions (HRTFs), are generally measured at finite locations in azimuth and elevation. Thus models are needed to synthesize DTFs at finer spatial resolution than the measured data to create complete virtual auditory displays. In this research, minimum-phase all-pole and all-zero models were used for modeling both human and cat DTFs. Real cepstrum analysis has been used to represent minimum phase HRTFs in the time domain. For the human DTFs, model orders were chosen to achieve specific objective error criteria published in previous studies that were based on subjective listening tests. Because subjective listening tests are not always feasible in animals, objective methods must be used to assess the quality of the DTF reconstructions. The same error criteria reported in subjective tests of human DTF reconstructions was used to constrain models of cat DTFs on the assumption that if humans cannot discriminate reconstructed vs empirical DTFs for a given objective reconstruction error criteria, then cats won't be able to either. All-pole and all-zero models of orders as low as 25 were able to model DTFs with errors comparable to previous research findings and preserve the main spectral features in both human and cat DTFs.

A hypothesis that a systematic relation (i.e., parametric equations) can be found to describe the movements of the poles/zeros of the successful models with the change in sound source location was tested. Polynomials of different orders were extracted to describe the movements of the poles in all-pole models and zeros in all-zero models with the change in sound source location. The reconstructed DTFs were compared to the measured ones of same locations. The reconstructed DTFs preserved the main shape of the spectra, provided satisfactory RMS errors compared to the measured ones and accurately preserved the first notch spectral feature.

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my advisor, Professor Mohammad A. Matin, and my co-advisor, Professor Daniel J. Tollin, for their support and encouragement throughout my PhD program. Their valuable guidance helped me to overcome the difficulties in my academic as well as personal life. Without their continuous support, this dissertation would not have been possible.

I would like to thank Dr. Rahmat Shoureshi for his valuable supervision of the results in Chapter 4 of this dissertation and for his comments and feedback throughout my PhD program. I would like to thank Professors Sandra S. Eaton and Roger Salters for serving on my committee and reviewing this dissertation. I am grateful to Dr. Mohammad Mahoor for being in my PhD committee and for his advice and feedback during my comprehensive exam. Special thanks are due to the National Institutes of Deafness and Other Communicative Disorders for funding part of this project under Grant No. DC-006865 to D.J.T.

I am also grateful to my Masters Degree advisor, Dr. V. Chandrasekar, of the Electrical and Computer Engineering department at Colorado State University for giving me the invaluable experience in doing the scientific research as soon as I started my post-graduate life. I would like to acknowledge Mary Boevers and the staff of the Int'l House at DU for their kindness and help throughout my program.

My very special thanks to the two persons whom I owe everything I am today, my father and my mother. To my brother for adding the nice flavor to my life and for being not only an older brother, but also a great friend. My special thanks to my wife for the support and patience she showed throughout my academic journey and specifically through the completion of this dissertation. To my son, Omar, for keeping me smiling and laughing even when I was working on most difficult problems. To all of you and to the memory of my friend, Diaa' Al-titi, I dedicate this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Background

The main acoustic cues for sound localization by humans and animals are the interaural time difference (ITD), interaural level difference (ILD) and the monaural spectral cues (Searle et al., 1976). The first two are “binaural” cues and generated by spatial separation of the ears on both sides of the head (Rayleigh, 1907). The monaural spectral cues, which have been shown to be important in resolving the ambiguity of sound source localization (SSL) when described only by the duplex theory (Rayleigh, 1907) especially on the cone of confusion (see Figure 1), are generated by the spatial- and frequency-dependent filtering or interaction of the propagating sound waves by the external ears, head, shoulders and the torso (e.g., Musicant et al., 1990, Blauert, 1997; Rice et al., 1992). These acoustic transformations can be captured in measurements of the Head Related Transfer Function (HRTF), defined as the complex ratio of the spectrum at the ear drum to the spectrum of the sound source. The left and right ear pair of HRTFs comprises all acoustical cues to sound source location available from that location (e.g., Blauert, 1969; Middlebrooks, 1992). Figure 1.2 shows examples of measured HRTFs from human *SOW-L-ear data* of location given by azimuth and elevation angle (az=-40°,el=0°) and *cat 1107-R-ear data* of location (az=0°,el=-30°).

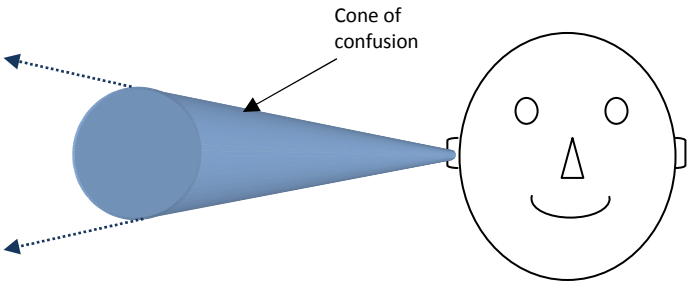


Figure 1.1. An illustration of cone of confusion. All points on the cone that have the same distance from the cone's vertex have the same ITD and ILD on both ears.

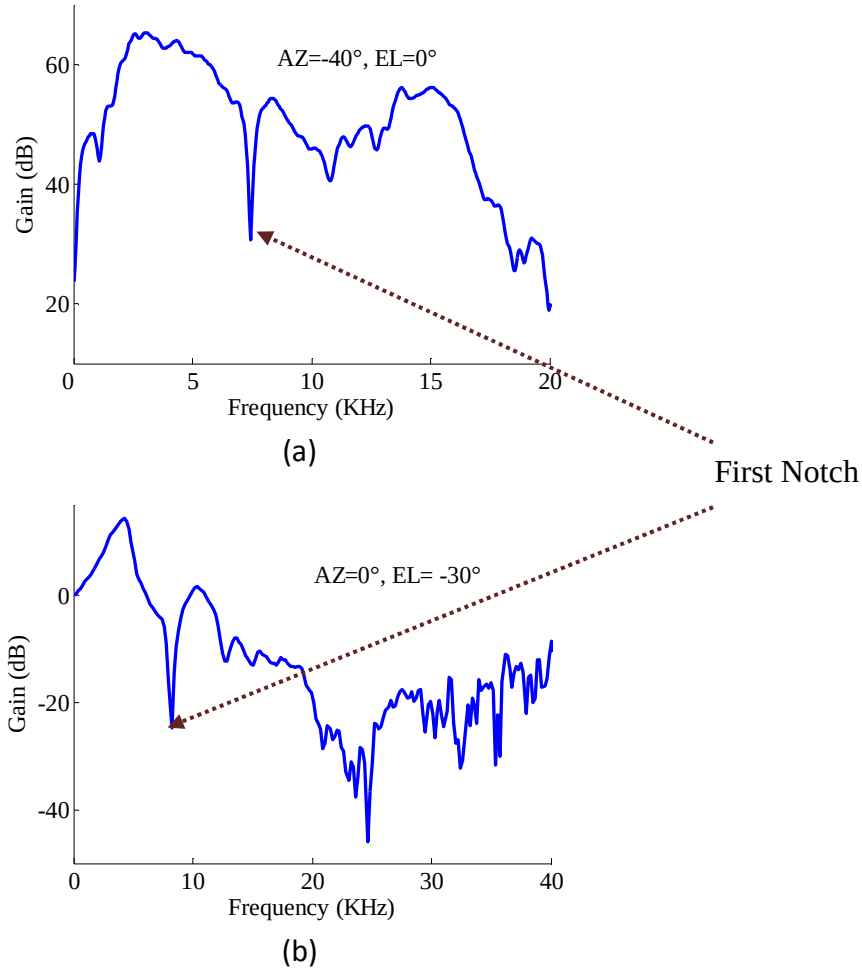


Figure 1.2. Measured HRTFs for (a) human *SOW-L-ear data* of location (az=-40°,el=0°) and (b) *cat 1107-R-ear data* of location (az=0°,el=-30°).

Because of their dependence on shape and geometry of the head, external ears, and body parts that interact with the received sound waves, HRTFs can be quite different for different individuals for a given location in space (Shaw, 1982; Wenzel et al., 1993; Langendijk and Bronkhorst, 2002; Middlebrooks, 1999; Jin et al., 2000).

HRTFs can be used to create virtual auditory displays (VADs) by presenting over headphones arbitrary sounds that have been filtered with a left and right ear pair of HRTFs to artificially recreate at the eardrums the appropriate spectral and temporal cues for a sound source in any direction. When this is done properly, the sound pressure waves presented at the two ears over headphones are virtually identical to those that would have been present if the same sound was presented from a loudspeaker at the desired location. Modeling HRTFs or directional transfer functions (DTFs), the directional components of the HRTFs, is important in many applications. For example, physiological studies using stimuli processed through HRTFs have lead to important insights into the functioning of the auditory nervous system (Tollin and Yin, 2002; Chase and Young, 2006). Behavioral studies in animals using HRTF-filtered stimuli have also been conducted (Poganiatz et al., 2001).

But arguably the most major limitation in current HRTF measurement and virtual acoustical display (VAD) implementation schemes concerns the resolution of spatial sampling during the empirical measurement of the HRTFs. Very high spatial resolution is required for simulating reverberation, sound source motion, and when the observer is able

to move interactively relative to the virtual sources. The ability to interact behaviorally in VADs continues to be a major area of auditory research. The clear advantage of simple pole/zero models (or any parametric model) of HRTFs is that the huge database of all possible left and right ear HRTFs does not need to be physically stored in memory and retrieved as needed (i.e., a look-up table) to implement the VAD; but rather the filters can be created in real time by simply inputting the desired source azimuth and elevation. (Otani and Ise, 2006) have shown that a hard drive space of approximately 450 GB is needed to store HRTFs measured with 43Hz increments and with calculation of up to 12KHz. So, if HRTFs can be rapidly recreated from a limited number of empirical measurements, not only can reconstruction of natural sound source and observer movements be realized (Nishino et al., 1999), but also data reduction can be achieved. VADs are currently used in psychoacoustic and physiological research, industry, some medical applications, military simulations, and for entertainment. Moreover, one of the features that has been studied to be restored in the hearing aids is the ability to improve directional hearing in individuals wearing hearing aids by finding the ratio of the unaided head related transfer function to the aided head related transfer function and then designing a hearing aid filter that is the inverse of that derived insertion effect (Soli et al., 1994). So, finding relations between the sound source direction movement and the change in the modeled HRTFs will simplify the needed design for this hearing aid filter to restore the directional hearing to a wearer of a hearing aid.

1.2 Statement of The Problem and Purpose of The Study

The empirical HRTFs necessary for VADs are generally measured at a limited number of finite source locations with a constant distance from the center of the head. Therefore, for full implementation of a complete VAD, efficient models, or interpolation, are required to synthesize HRTFs from locations not measured empirically and at finer resolution than the measured data.

The purpose of this study is divided into two parts. The first one is to find suitable low-order models for the HRTFs according to several error criteria. The models are of two types, Finite Impulse Response (FIR) filters and Infinite Impulse Response (IIR) filters. The modeling will seek the lowest possible order to keep the model simple and accurately preserve the main direction-dependent peaks and valleys in the measured spectra, but not to over fit the HRTFs in a way to follow noise in the measured data. The second part is to test the hypothesis that a systematic relation (i.e., parametric equations) can be found to describe the movements of the poles/zeros of the successful pole-zero models with the change in the sound source location (in elevation or azimuth angle) which can reduce the whole set of HRTFs to a small number of parametric equations describing the trajectory of the poles and zeros with changes in spatial location.

1.3 Literature Review

1.3.1 HRTF/DTF modeling

As with any transfer function, in order to have an accurate description of the HRTFs or the DTFs, we need to specify both its magnitude spectrum and its phase spectrum

independently (Oppenheim and Schafer, 1989). Studies have shown, however, that HRTFs or DTFs can be accurately modeled by just their minimum phase spectra because the auditory system is not sensitive to the absolute phase of a sound presented to a single ear (Kulkarni et al., 1999; Kistler and Wightman, 1992). That is why most empirical experiments on human HRTFs focused on the magnitude components in their research (Mehrgardt and Mellert, 1977; Midllebrooks et al., 1989; Shaw, 1974; Wightman and Kistler, 1989a). To preserve the ITD cues (given by the left- and right-ear phase *difference* spectra, which humans are very sensitive to) for any location, the phase difference can first be computed from the empirical impulse responses before applying the minimum phase transformation, and then added back to the magnitude of the modeled HRTFs as a real time delay (Kulkarni et al., 1999). The energy in minimum-phase systems is optimally concentrated in the beginning of the impulse response which allows for shortest filter lengths for a specific amplitude response (Huopaniemi et al., 1999) and that is another attraction of minimum phase systems, in general. For these reasons, here, only the magnitude spectra of the DTFs are modeled.

DTFs have been preferred to be modeled and then added to the common part of all HRTFs over modeling the complete HRTFs at once for many reasons. For example, (Kulkarni and Colburn, 2004) proposed that it is better to have a fixed response of a direction-*independent* transfer function in series with the direction-*dependent* filters that model the DTFs. It has been also shown that it is most efficient for both storage and computation to process stimuli through desired DTFs and then through a common filter

stage characterizing the mean direction-independent function (which captures the resonances of the ear canal) because of lower-order approximations needed to successfully model DTFs compared to the ones needed for HRTFs (Haneda et al., 1999; Kulkarni and Colburn, 2004). Moreover, DTFs remove the small variations in the probe tube microphone placement in the ear canal during HRTF data acquisition.

Several different techniques have been used by researchers to model HRTFs/DTFs. For example, (Kistler and Wightman, 1992; Martens, 1987; Jin et al., 2000; Langendijk and Bronkhorst, 1997) used models based on principal components analysis (PCA). (Kistler and Wightman, 1992) reconstructed DTFs from linear combination of principal components and then added the common part to form the reconstructed HRTF and showed that HRTFs (reconstructed DTFs with means added back) can be adequately approximated by a linear combination of as few as five basis functions (eg, 5 different spectral shapes). (Brown and Duda, 1998; Lopez-Poveda and Meddis, 1996) used structural models that analyze the sound waves propagation and diffraction from the sound source to the ear drum.

Methods based on interpolation in the time and/or frequency domains have also been used. Pole/zero approximations have been widely used also for HRTFs modeling (Blommer and Wakefield 1997; Haneda et al. 1999; Kulkarni and Colburn, 2004; Raykar et al., 2005; Asano et al., 1990; Durant and Wakefield, 2002). Finite impulse (FIR) and/or infinite impulse (IIR) response filters with different orders have been used. Despite the

fact that IIR filters are more complicated to implement on digital signal processors compared to FIR, or all-zero models, they have been shown to represent transfer functions with lower orders, or fewer parameters (Lyons, 1997; Oppenheim and Schafer, 1989). All-pole models seem to better approximate peaks than nulls in the DTFs (Kulkarni and Colburn, 2004; Raykar et al., 2005, Carlile, 1996). On the other hand, all-zero model is an efficient choice for representing HRTFs with deep valleys, but not those with sharp peaks (Carlile, 1996). (Blommer and Wakefield, 1997) have shown that combined pole-zero approximations can be used to synthesize HRTFs with even fewer total parameters than is typically achieved by all-zero approximations. On the other hand, all-zero models have been chosen for HRTF implementation due to straightforward interpolation, relatively good spectral performance, simplicity of implementation (Huopaniemi et al., 1999), and for being always stable. It should be mentioned though that both all-pole and all-zero models used in this work are stable since only minimum-phase filters are used as HRTF models (Oppenheim and Schafer, 1989).

IIR filters models have been applied to DTF and HRTF modeling using different techniques, such as Yule-Walker methods (Sandvad and Hammershoi, 1994; Jot et al., 1995), Prony's method (Sandvad and Hammershoi, 1994; Kulkarni and Colburn, 2004), weighted-least-squares Autoregressive Moving Average (ARMA) model (Kulkarni and Colburn, 2004; Blommer and Wakefield, 1994), Balanced Model Truncation (Mackenzie et al., 1997), and a spherical basis function neural network to approximate the model parameters (Jenison, 1995) . (Runkle et al., 1995) used Cepstrum techniques for all-zero

approximations to model HRTFs. (Haneda et al., 1999; Liu and Hsieh, 2001) represented HRTF by a direction-independent part (common acoustical poles) and a direction-dependent part (all-zero model). In that model, in addition to the fewer parameters that will be dependent on the sound source direction using common-acoustical-pole and zero modeling of HRTFs, it can also extract even the zeros that are missed in conventional pole/zero models due to pole zero cancellation which means that this model can trace well the zero variations due to changes in sound source direction (Haneda et al., 1999). (Grantham et al., 2005) used the same idea of common poles for the entire set of HRTFs and then used a technique based on the singular value decomposition of Head Related Impulse Responses (HRIRs) to create the final HRTF model. (Susnik et al., 2003) showed that HRTF could be realized with a simple set of IIR filters, time delay element and an amplifier.

1.3.2 Important spectral features of the DTFs for sound source localization - the first notch frequency

It has been shown in previous studies on humans that there are prominent spectral “notches” and “peaks” above 4-5 KHz that vary systematically with the changes in elevation and azimuth angles of sound source location which are essential for sound localization especially for source elevation and for determining whether a sound is in front of or behind the observer (Shaw, 1974 and 1982; Shaw and Teranishi, 1968; Hebrank and Wright, 1974; Kuhn, 1987; Langendijk and Bronkhorst, 2002; Raykar et al., 2005). While some notch frequencies vary smoothly with elevation, the spectral peaks do

not show this smooth trend (Raykar et al., 2005). It has been also shown that the spectral location of the first prominent spectral minima, which is known in the literature as the First Notch (FN), changes in frequency from low (~6 kHz) to high (~12 kHz) as the angle of sound source varies from -15° to 45° in elevation with fixed azimuth angle in human data (Shaw, 1982; Shaw & Teranishi, 1968; Langendijk and Bronkhorst, 2002). Similar behavior of spectral notch movement with the change in elevation angles have been shown in measurements in animals commonly used in anatomical and neuro-physiological studies, including cats (Musicant et al., 1990; Rice et al., 1992; Tollin and Koka, 2009; Young et al., 1992), gerbils (Maki and Furukawa, 2005), monkey (Spezio et al., 2000), and rats (Koka et al., 2008) with a difference in the change in the frequency range of the FN due to the difference in the size and the shape of different species. Cats are often used in studies of localization and are ideal models for this purpose because of their ability to localize sound source accurately and quickly with a performance close to humans (e.g., Tollin et al., 2005; May and Huang, 1996). Behavioral studies suggest that cats use spectral cues for sound localization (Casseday and Neff, 1973; Aitkin and Martin, 1987; Aitkin and Martin, 1990).

In cats, the FN changes from about 8 to 18 kHz with the change in elevation from low to high, with similar behavior as the sound source is moved laterally from the midline in the azimuth with fixed elevation (Musicant et al., 1990; Rice et al., 1992). The frequency shift of this notch is greater for angular displacement in elevation (EL) compared to that occurs in azimuth (AZ) and the notch becomes broader and shallower above 54º EL and

becomes less distinct as azimuth shifts more laterally (Musicant et al., 1990; Rice et al., 1992). (Huang and May, 1996) suggested that the mid-frequency region in cats' HRTF (5-18 KHz), which is the region that the FN occurs in, is the region that sound localization cues are derived from and that the spectral notches play an important role in it. In behavioral experiments in cats, Tollin and Yin (2003) showed that cats actually do rely on the FN frequency to localize sound sources that vary in elevation. (Musicant et al., 1990) indicated that pinna is essential for the presence of the notches at lower frequencies and it alters the notches at high frequencies. (Tetsufumi, et al., 2000) showed that this FN, which is regarded as an important localization cue (Rice et al., 1992), is diminished when subject's pinnae are covered with silicone putty. (Brown and Duda, 1998) mentioned that the pinna is very important for elevation estimation in addition to providing some azimuth information.

Rice and his colleagues (1992) mentioned that depending on the systematic movement of the FN with the change in sound source location, there will be an ambiguity if FN is used alone for localization because both increases in elevation and increases in azimuth cause the FN to move to a higher frequency. They also mentioned that this ambiguity can be resolved by information from the frequency regions outside the frequency range that the FN is observed and it can be also resolved by a knowledge of the FN frequency in the contra-lateral ear and referred to a localization scheme based on knowledge of FN frequencies in the two ears as the binaural FN direction code.

1.3.3 HRTF interpolation and reconstruction

Several techniques have been investigated or proposed by researchers for HRTF interpolation or reconstruction at unmeasured locations and at finer resolution compared to the measurement experiments. Wenzel and Foster (1993) compared three subjects' localization judgments for stimuli synthesized from non-interpolated HRTFs, simple linear interpolations of the empirical HRTFs, stimuli synthesized from non-interpolated minimum-phase approximations of the HRTFs and linear interpolations of the minimum-phase HRTFs. Two kinds of errors have been considered, small errors on the order of 5° to 20° and confusions errors (front positions heard in the rear; down locations heard as up. etc.). In general, the overall level of confusions was quite high. In that study, Wenzel and Foster mentioned that localization accuracy is largely unaffected by interpolation of either normal or minimum-phase HRTFs. The approximated HRTFs using linear interpolation from neighboring spatial locations are often acoustically discriminable from measured HRTFs at the same location (Kulkarni and Colburn, 1993). Interpolation in the time domain cannot, however, accurately reconstruct the spectral shape corresponding to the FN. For example, a sharp spectral notch cannont be recreated via interpolation of the sharp notches at two neighboring locations; interpolation simply results in a less sharp notch, which will reduce sound localization accuracy.

The relative performance of interpolation methods of all-zero and pole-zero modeling of DTFs has been studied in (Runkle et al., 1995). (Wang et al., 2008) proposed an interpolation for HRTF pole-zero models based on all-zero linear interpolation methods.

(Martin and McAnally, 2007) proposed an interpolation technique that involves the application of an inverse-distance-weighted averaging algorithm to HRTFs represented in the frequency domain. (Hacihabiboglu, 2002) presented an HRTF filter interpolation for FIR filters with a vectorial zero-displacement measure to synthesize HRTFs at unmeasured positions. Zeros interpolation was proposed for all-zero model of 64 zeros and has been shown how the general shape of the notches and the peaks can be captured (Hacihabiboglu, 2002). (Nishino et al., 1999) evaluated simple linear interpolation and *Spline* interpolation methods for HRTFs in the median plane. Nishino and his colleagues (1999) showed that the linear method is more effective when the total number of elevations is large and that *Spline* method is effective when the intervals among elevations are large and equivalent. (Cheng and Wakefield, 1999) used a combination of HRTFs to derive spatial frequency response surfaces to get interpolated HRTFs. Nishino and his colleagues (2007) estimated HRTFs on the horizontal plane by investigating the relationship between HRTFs and the physical features of the subject by multiple regression analysis. (Zhong and Xie, 2009) proposed a reconstruction relation for HRTFs at unmeasured azimuths for a given elevation as a linear combination of weighted azimuthal harmonics under the condition of having enough azimuthal measurements.

1.4 Dissertation Organization

Chapter 2 presents an overview of the used methods in this dissertation. It includes a description of the empirical datasets, data processing and pre-conditioning, DTFs modeling techniques and models' validation and evaluation techniques. A brief

description of how the effect of the movement of the poles and zeros in the z-plane on the DTFs will be studied for the purpose of building the parametric relations that describe those movements.

Following Chapter 2, the contents of the dissertation can be divided into two main parts. The first part which covers the results of the all-pole and all-zero modeling and the models evaluation which are presented in Chapters 3 and 4. The second part of the dissertation tests and proves the hypothesis that there is systematic movements of poles and zeros of pole-zero models that can be described by parametric relations for those poles and zeros in the z-plane as a function of the azimuth and elevation angles. This part is covered in Chapter 5.

Chapter 3 presents the results of the modeling techniques, the validation of these models and provides a detailed evaluation for the fitted DTFs using some error criteria and spectral features preservation testing. An evaluation for the modeled DTFs by comparing them to the measured DTFs using Artificial Neural Networks (ANN) is provided in Chapter 4. Chapter 4 also introduces using ANN in sound source localization from the monaural spectral features of the DTFs.

Chapter 5 shows how DTFs can be rebuilt from extracted parametric equations which describe the movement of the poles/zeros in the z-plane with the change in the sound source location. A comparison of the reconstructed and the measured DTFs is provided

also in Chapter 5. Chapter 6 presents the summary and the conclusions of the work done in this dissertation and outlines suggestions for future work.

Chapter 2

METHODOLOGY

2.1 Spatial Coordinate System for Sound Source Location

In this study, the coordinate system reference will be as follows. Source elevation corresponds to latitude and source azimuth corresponds to longitude where location (0°,0°) is directly in front of the subject. Positive elevations are above the horizontal plane and negative elevations are below. Directly above the subject's head is 90° EL. Positive azimuths are to the right side of the subject and negative azimuths are to the left.

2.2 Empirical HRTF/DTF Data Sets

One of the goals of the present study was to determine efficient low-order models for animal HRTFs so that these HRTFs could be subsequently used in physiological studies and in future psychophysical studies. We use here HRTFs from cats. Since we cannot do behavioral tests or behavioral research on cats, DTF models are applied on cats and human data in this study. Cats are often used in studies of localization and are ideal models for this purpose because of their ability to localize sound source accurately and quickly with a performance close to humans (e.g., Tollin et al., 2005; May and Huang, 1996). Some previous studies suggest that cats use spectral cues for sound localization (Tollin and Yin, 2003; Casseday and Neff, 1973; Aitkin and Martin, 1987; Aitkin and

Martin, 1990). HRTFs and DTFs datasets used in this study come from two species; humans and cats.

2.2.1 Human data

Human HRTF data were kindly provided by Drs Doris Kistler and Fred Wightman from the Department of Psychology at the University of Louisville. These data, which have been used in psychophysical sound localization studies (Wightman and Kistler, 1989a and 1989b), correspond to subject SDO. In that dataset, HRTFs were measured at 144 locations with elevation angle ranges from +54º to -36º with 18º step, and azimuth angles from +180º to -165º with 15º step and a 256-tap filter at a sampling rate of 50 kHz. The second dataset corresponds to subject SOW used in the DTF modeling and psychophysical studies of (Kulkarni and Colburn, 2004). The SOW dataset consists of measurements at 505 locations with elevation angle ranges from +80° to -50° and azimuths angles from +180° to -170° with 10° step in both directions, in addition to a measurement from directly above (+90° EL). These data are 256-tap filters at a sampling rate of 50 kHz.

2.2.2 Cat data

One set of cat HRTF data is from “cat 1107” from the study of (Rice et al., 1992), provided by Dr Eric Young at Johns Hopkins University. Source positions were 259 locations in the frontal field, between -75º and 75° AZ. Another set of cat data was provided from one adult cat from the study of Tollin and Koka (2009). This dataset was

sampled 3° in vertical steps and 7.5^{o} horizontal steps with a nominal sampling rate of 100 kHz. This dataset will be referred as "*UCHSC dataset*" since it has been measured by Prof. Daniel Tollin at University of Colorado Health Sciences Center (UCHSC).

2.3 Data Processing and Conditioning

Signal Processing Toolbox software of Matlab, The Mathworks Inc., has been used for signal processing, conditioning, modeling techniques and models evaluation algorithms in this project.

2.3.1 HRTFs to DTFs

Regardless of the source of the HRTFs used here (cat or human), the data have been converted to DTFs (Middlebrooks and Green, 1992) before applying the modeling techniques. This has been done by finding the average of the measured HRTFs in logarithmic scale in the frequency domain at all locations, then subtracting this average from each HRTF at each location in the logarithmic scale,

$$(DTF)_i = (HRTF)_i - \frac{1}{n}\sum_{j=1}^{n}(HRTF)_j \qquad (2.1)$$

which is equivalent to division in the linear scale. '*n*' in Equation (2.1) represents the number of the measured HRTFs for a certain dataset at one ear. DTFs isolate the direction-dependent compontents of the HRTFs – that is, those components that actually change as sound source location is changed. The average, or common component, isolated the direction-independent components of the HRTF, such as those due to the resonance of the ear canal (which does not depend on source direction) and the spectral

consequences of the exact placement of the microphone tip in the ear canal. Figure 2.1 shows the HRTFs at all 259 locations for one ear of *cat 1107* from (Rice et al., 1992) in addition to the direction-independent, or the common, component of the ensemble of HRTFs.

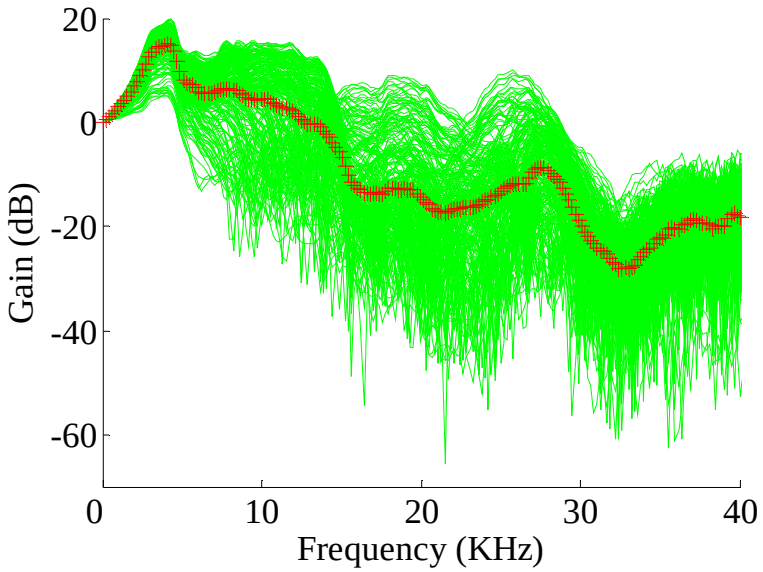


Figure 2.1. HRTFs at all locations from (Rice et al., 1992) dataset for *cat 1107* (right ear) in addition to the direction independent, or common, component for this dataset (red plusses).

2.3.2 Windowing and filtering

All DTFs have been windowed using a half Hanning window in order to remove any reflection in the impulse response. All DTFs have been also low-pass filtered to remove frequencies above 40 KHz for cat data and 20 KHz for human data which are the average upper hearing limits for these species. Figure 2.2 shows the Head Related Impulse Response (HRIR) from human *SDO-L-ear* data of location (az=-165°, el=-30°) (solid line) in addition to the used half-Hanning window (dotted line).

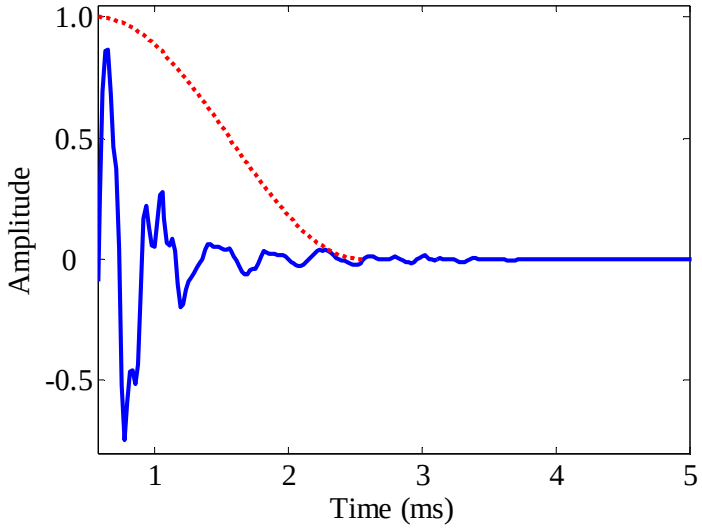


Figure 2.2. Windowing of HRIR from human *SDO-L-ear* data of location (az=-165°, el=-30°) (solid line) using half-Hanning window (dotted line).

2.3.3 Minimum-phase DTFs - r*eal Cepstrum*

Real *Cepstrum* analysis (Oppenheim and Schafer, 1989) has been used to represent the minimum phase in time domain for all used DTFs. It is defined as the inverse Fourier transform of the magnitude of the Fourier transform, as given in Equation 2.2. Runkle and his colleagues (1995) used the *Cepstrum* technique for all-zero approximation to model HRTFs. The resulting minimum-phase sequence is real, causal and stable where all poles and zeros are inside the unit circle.

$$c_x[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|X\left(e^{j\omega}\right)\right| e^{j\omega n} d\omega \qquad (2.2)$$

2.3.4 DTFs smoothing

To ensure low-order filters, DTFs were “smoothed” before applying our modeling techniques. It has been shown in previous studies (Kulkarni and Colburn, 1998;

Hacihabiboglu et al., 2002) that HRTFs can be smoothed significantly in the frequency domain without affecting the perception of sound stimuli filtered by HRTFs. In other words, the very fine spectral detail in HRTFs is not important for sound localization as long as the major features are preserved (Kulkarni and Colburn, 1998; Jin et al., 2000; Asano et al., 1990). Smoothing also simulates the spectral filtering process of the cochlea in the peripheral auditory system that eliminates sharp spectral peaks and notches. Some researchers used partial Fourier series at different levels to smooth the HRTFs (Kulkarni and Colburn, 1998). A bank of triangular band-pass filters which removes details that would be excluded by cochlear filtering and give a smoothed version to DTFs is another technique that has been used to smooth the HRTFs (Schnupp et al., 2003; Koka et al., 2008; Tollin and Koka, 2009).

HRTFs also have been successfully smoothed using different methods depending on wavelet transforms (Hacihabiboglu et al., 2002; Hacihabiboglu, 2002). Three different wavelet-transform methods were used and tested in (Hacihabiboglu et al., 2002): wavelet denoising with Stein's Unbiased Risk Estimation (SURE), wavelet approximation and redundant wavelet transform (RWT). Based on Hacihabiboglu and his colleagues' subjective listening test, they reported in their (2002) paper that wavelet-based spectral smoothing methods are beneficial in reducing the filter order by providing better localization cues than the direct design. They also mentioned that the RWT technique increased the sound localization accuracy without substantially decreasing the sound quality and it was also the preferred choice for all of the subjects in terms of the mean

localization cues compared to the other wavelet-based smoothing methods. Using that technique for smoothing (i.e. RWT), Hacihabiboglu and his colleagues mentioned that even for an IIR filter order as low as 10, the behavioral sound localization performance was satisfactorily high (Hacihabiboglu et al., 2002).

In this dissertation, "Symmlet 17" filter bank was applied to the magnitude responses of the DTFs using RWT technique, which is also known as Stationary Wavelet Transform (SWT) (Mallat, 1998), for spectral smoothing purpose. "Symmlet 17" has been chosen after comparing the results of different wavelets including Haar, discrete Meyer, Daubechies 2-9, Coiflet 1-5 and Symmlet 1-17 wavelets. Figure 2.3 illustrates a diagram of a three-level stationary wavelet transform decomposition, where *H* and *G* are wavelet basis lowpass and highpass filters.

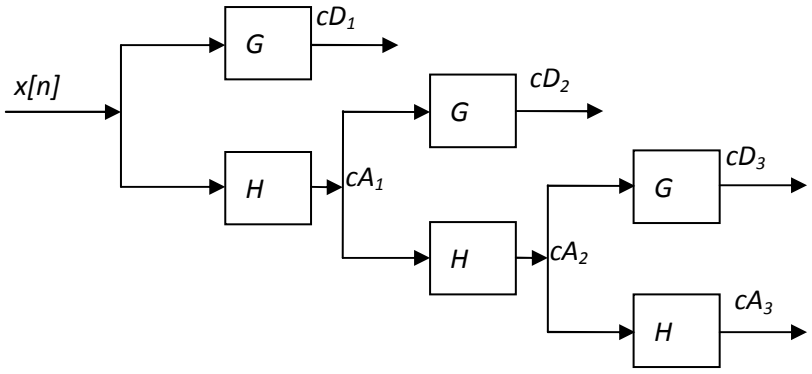


Figure 2.3. A three level stationary wavelet transform.

In SWT decomposition, the coefficients are not downsampled but the impulse responses of the highpass and lowpass filters are upsampled for each level, which yields the coarse

coefficient cA_j and detail coefficient cD_j level j. The coarse and detail coefficients can be computed recursively by

$$cA_{j+1}(n) = \sum_m h(m) cA_j(n - 2^j m) \quad (2.3)$$

$$cD_{j+1}(n) = \sum_m g(m) cD_j(n - 2^j m) \quad (2.4)$$

Figure 2.4 shows an example of the measured and the smoothed DTFs for *cat 1107* dataset at location (az=0°, el=37.5°) using RWT level 3 "Symmlet 17" filter bank.

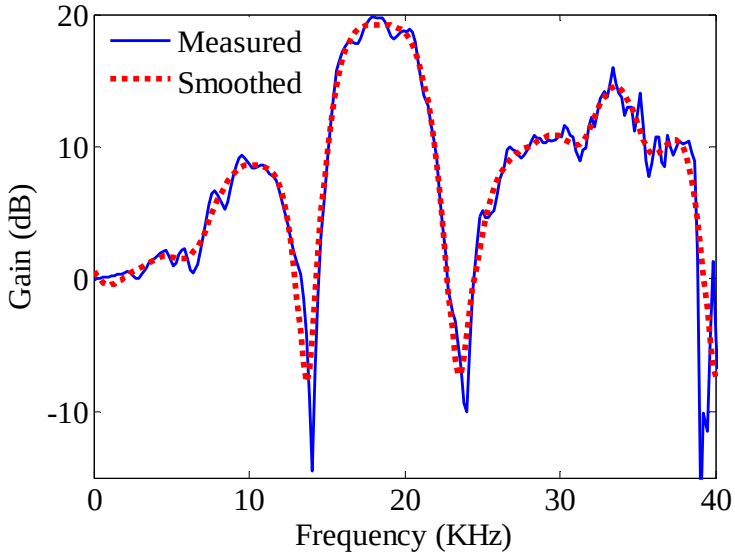


Figure 2.4. An example of a RWT smoothed DTF al location (az=0°, el=37.5°) from *cat 1107* dataset using "Symmlet 17" filter bank.

2.4 DTFs Modeling Techniques

In this study, we will use the pole/zero approximations for modeling because of the success of these approximations in modeling HRTFs/ DTFs as reported in previous research described earlier. We will model DTFs, not HRTFs, because of the advantages mentioned above of modeling DTFs compared to HRTFs. The models used in this work

are the ones whose Fourier transforms are minimum-phase where the log-magnitude spectrum and phase spectrum are Hilbert transform of each other. Accordingly, the minimum-phase system will have a unique phase function if the magnitude function is given (Oppenheim and Schafer, 1989; Kulkarni et al., 1999). Minimum-phase modeled DTFs have all their poles and zeros within the unit circle which adds the advantage of being always stable.

Because some important spectral features of the HRTFs (i.e., notches and peaks) change systematically with source location at least over a limited range, using the pole/zero approximation is a desirable technique for modeling since these systematic movements can be reflected into concomitant tractable movements of the poles and zeros positions in the pole/zero complex plane plot of these models. In this study, only all-pole and all-zero models were used to fit DTFs in order to avoid any pole zero cancellation that may happen in the models that have both poles and zeros combined in one model (Haneda et al., 1999). This gives better movement tracing for the direction dependent poles and zeros (Hacihabiboglu, 2002) that may accompany the systematic changes in the modeled DTFs (i.e. peaks or notches) with the change in the sound source elevation and/or azimuth.

The goal is to find the lowest possible order for all-pole and all-zero models to keep the model simple, efficient and accurately preserve the main spectral features of the DTFs, such as the direction-dependent peaks and valleys, but not to over fit the DTFs in a way to follow noise in the measured data. Finding lower-order models not only reduce the

computational complexity, but also simplify the tracking of the changes of the poles and zeros locations in pole/zero models with the change in the location of the sound source, which is the ultimate purpose of choosing suitable pole/zero models in this project.

Linear Predictor Coefficients (LPC) technique has been used for all-pole models by computing the least squares error solution with the autocorrelation method of autoregressive modeling to find the model coefficients (Jackson, 1989). The all-zero model uses the impulse response of the DTF as its coefficients with a rectangular window to truncate the original impulse response and constructed using an FIR filter (Oppenheim and Schafer, 1975). These models are used to fit the magnitude spectra.

2.5 Models Validation and Evaluation

The proposed models used in this project are validated and evaluated based on human and cat measured HRTFs/DTFs. Model orders are chosen according to error criteria published in previous studies that were supported by human subjective tests, and to the ability of these models to preserve the main spectral features that have been proven in previous researches as important spectral localization cues. Modeling animal DTFs has some limitations because it can be impractical to perceptually validate the modeled DTFs using subjective listening tests. Here, we propose to use the models and the associated objective error metrics based on human DTF models to constrain the models of cat DTFs. It is assumed that as long as the objective (i.e., quantitative) DTF reconstruction errors for the cat DTFs are constrained to be at or less than the DTF reconstruction errors

known in human listening tests to result in DTF reconstructions that are discriminable from empirical DTFs, the cats would also not be able to discriminate the modeled vs. the empirical DTFs. It has been shown in this study that this constraint can be met for both human and cat DTF models with model orders as low as 25 for all-pole and all-zero models. These models were evaluated by comparing the resulting errors to the errors in a previous study (Kulkarni and Colburn, 2004) and the effectiveness of these models to preserve the main spectral features in the empirical DTFs.

2.5.1 RMS error on a dB log scale

The modeled DTFs have been compared to the measured ones using the same error expression that have been used in (Kulkarni and Colburn, 2004) and computed in decibel (dB) scale,

$$Error = 20\left|\int_{-\pi}^{\pi}(\log_{10}|H(e^{j\omega})|-\log_{10}|\hat{H}(e^{j\omega})|)^2\frac{d\omega}{2\pi}\right|^{1/2} \qquad (2.5)$$

where, $H(e^{j\omega})$ and $\hat{H}(e^{j\omega})$ represent the measured and the modeled DTFs, respectively. One objective criteria used here was that the orders of all-pole or all-zero models were chosen to keep the mean error of Equation 2.5 computed across all DTF locations to be less than the mean error for human DTF reconstructions that (Kulkarni and Colburn, 2004) could be discriminated from empirical DTFs in their human psychophysical listening tests.

2.5.2 Spectral features preservation – the first notch frequency

It is crucial for any successful model to preserve the spectral features that are known to be important for sound localization, such as the first prominent spectral minima, which is known in the literature as the First Notch (FN; Rice et al., 1992). Because of the importance of the FN as a spectral cue for sound localization, we will test the preservation of this notch and its systematic movement with the change in sound source location using the proposed models. To our knowledge, this form of DTF reconstruction validation has not been used. We believe this analysis is important because if the important features for sound localization are not preserved in the reconstructions, then the whole point of the modeling has been defeated because localization errors will be made when these DTFs are used to implement a VAD. To test the validity of these models in preserving this feature at all locations, the FN frequency was extracted from the measured DTFs at all locations and compared to the extracted ones from the modeled DTFs and plotted against each other. Regression lines have been drawn for those plots using simple linear regression after forcing the regression line to pass through the origin:

$$Y_i = \beta X_i + \varepsilon_i \tag{2.6}$$

where β is the slope of the regression line, the term ε indicates the vertical deviation of a particular point from the regression line. The resulting regression lines, the values of the regression lines' slopes and the coefficient of determination, "r^2", values will be tested to check the preservation of FN systematic movement. The coefficient of determination is a measure of how well the regression line represents the data. The RMS error of Equation 2.5 does not by itself guarantee that features such as the spectral notches have been

preserved. Bootstrap resampling has been applied to the FN frequency of all locations in the frontal field to derive confidence intervals for the regression lines' slopes between the FN frequency of the measured DTFs and the ones of the all-pole and all-zero modeled DTFs.

2.5.3 Spectral shape comparison – coherence function

In addition to comparing our modeled DTFs to the measured ones using the error metric of equation 2.5 and the FN feature and its systematic movement preservation, the coherence between the modeled and the measured DTFs is calculated to compare the shape of the spectra between modeled and measured DTFs including all peaks and notches in those spectra. The coherence function measures the correlation between two signals as a function of the frequency components they contain (Shaw, 1981). For that purpose, the Magnitude Squared Coherence (MSC) which is defined in equation 2.7 (Van Drongelen, 2006) and uses *Welch's averaged, modified periodogram method* (Percival and Walden, 1993) will be used:

$$C(\omega) = \frac{\left|S_{xy}(\omega)\right|^2}{S_{xx}(\omega)S_{yy}(\omega)} \tag{2.7}$$

where $S_{xy}(\omega)$ is the cross spectral density, and $S_{xx}(\omega)$ and $S_{yy}(\omega)$ are the auto spectra of the modeled and measured DTFs. MSC estimate is a function of frequency with values ranging from 0 to 1 and provides an assessment of the linear relation at each frequency. MSC of 1.0 indicates perfect coherence and 0 indicates no coherence. The coherence is

estimated practically by averaging over several segments as indicated in Equation 2.8, where '*n*' represents the number of the segments in time domain.

$$C(\omega) = \frac{\left|\left\langle S_{xy}(\omega)\right\rangle_n\right|^2}{\left\langle S_{xx}(\omega)\right\rangle_n \left\langle S_{yy}(\omega)\right\rangle_n} \tag{2.8}$$

The data were divided into 64-point segments for analysis with a 60% overlap of adjacent segments using Hamming window.

2.6 Tracking the Systematic Movements of Poles/Zeros With Changes in The Sound Source Direction

To the extent to which the spectral peaks and notches of the DTFs move systematically with source location, which they do, the poles and zeros would be expected to also move systematically. The effect of the change in pole/zero pairs location on DTF spectra is studied to find the pairs that have significant effect on the spectral main features. Those pairs are considered the dominant pairs. To initially simplify the tracking of poles and zeros, their movements with several elevation angles at fixed source azimuths have been tracked and recorded. Then the same procedure has been repeated for several azimuths but with fixed source elevations. Poles and zeros movement tracking is performed in the 2-D polar coordinate system where the change in poles/zeros location will be described by an angle (β) and a distance (r) (i.e., a vector), as shown in Figure 2.5.

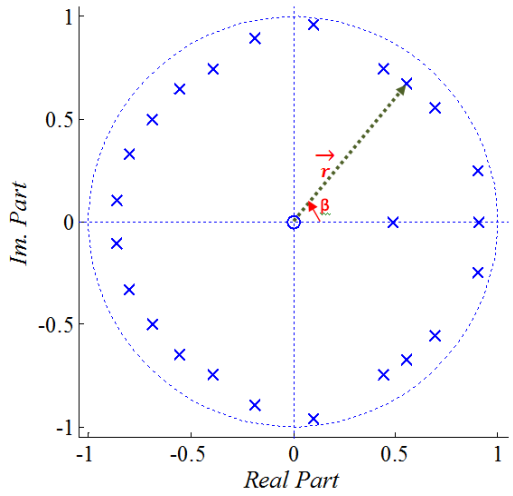


Figure 2.5. An illustration of the angle and distance of the used polar coordinate system for tracking the movements of poles/zeros in the z-plane.

The observations of the relations between the locations of the poles/zeros in the complex z-plane and the spectra of the modeled DTFs have been used to track the systematic movement of the poles/zeros with the change in the magnitude frequency response as a function of the sound source location (azimuth and/or elevation angle). The contribution of each pole/zero pair to a certain spectrum has been studied by examining the consequences of systematically moving that pole/zero pair in certain directions and the effect of their removal. A GUI developed by the members of the *Center for Signal and Image Processing (CSIP) at the Georgia Institute of Technology*, under the name *"PEZdemo"*, has been also used to study the effect of the poles/zeros removal or change in location on the resulting DTF.

Chapter 3

RESULTS AND DISCUSSION I: HRTF/DTF Modeling

3.1 Computing DTFs from HRTFs

Figure 1 shows an example of how DTFs are computed based on the HRTFs at all 259 locations for one ear of *cat 1107* from (Rice et al., 1992). The direction-independent, or the common, component of the ensemble of HRTFs was determined by taking the average across all locations and is shown in Figure 3.1 and Figure 3.2. DTFs are computed by subtracting the common component from the HRTF measured at each location.

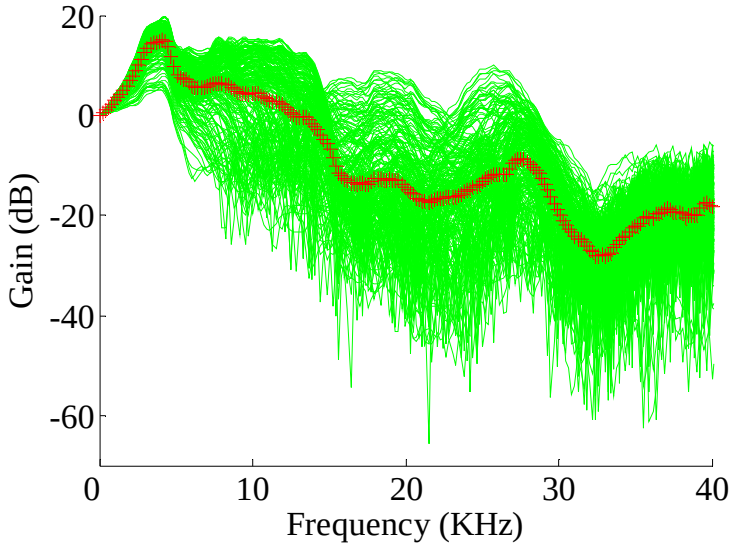


Figure 3.1. HRTFs at all locations from (Rice et al., 1992) dataset for *cat 1107* (right ear) in addition to the direction independent, or common, component for this dataset (red plusses).

Figure 3.2 shows how an HRTF measured at (0°,-30°) can be reconstructed from the sum of the modeled DTF (using techniques proposed in this study) and the common component. Note in Figure 3.2 that the first prominent peak in the HRTF, which is due to the non-directional resonance of the ear canal and concha (Tollin and Koka, 2009), is apparent in the common component but has been removed or minimized in the DTF.

3.2 Smoothing

Redundant Wavelet Transform (RWT) was used here to smooth DTFs in order to better preserve the main spectral features, including the frequency of the FN, when compared to the measured DTFs. Smoothing removes the "noise" in the recorded DTFs. It also simulates the bank-of-bandpass-filtering of sounds as produced in the cochlea (~1/3 – 1/6 octave band-pass resolution). Figure 3.3 shows an example of DTF smoothing on human and cat data using level 3 "Symmlet 17" wavelet filter bank. As noticed, smoothing retains the main features of the DTFs and eliminates or reduces the noise or the very fine details in the spectra (i.e., the very sharp and deep spectral notches) that the auditory system cannot resolve.

3.3 Modeling Human DTFs

Pole/zero models were applied to SDO and SOW datasets and the results were compared to the results in (Kulkarni and Colburn, 2004). Kulkarni and Colburn also modeled the same SDO and SOW datasets and, according to their human subject listening tests,

> "*…IIR models can provide efficient descriptions of HRTFs and whereas subjects could discriminate a 20-pole all-pole model reconstruction and a 4-pole 4-zero pole-zero model from the original HRTF, they found the task to be very difficult when a 25-pole all-pole model and a 6-pole, 6-zero pole-zero model reconstruction were used.*"

In other words, with HRTFs reconstructed using a 25-pole all-pole model, across the locations used in their psychophysical tests, the subjects could not consistently discriminate perceptually stimuli filtered through the reconstructed HRTFs from the empirical HRTFs. Thus, the reconstructed HRTFs were as good as the empirical. Thus, HRTFs reconstructed using that model order adequately preserved the important spectral components necessary for veridical (i.e., realistic and correct) perception. Across all HRTF locations, the worst reconstruction errors (using Equation 2.5) Kulkarni and Colburn (2004) reported from the SDO dataset using the successful models had errors of 3.5 and 5.3 dB for the left and the right ear, respectively.

3.3.1 dB error metric

Here, it has been assured that none of the modeled DTFs at any location exceed the minimum value of those worst cases (i.e. 3.5 dB) and that the mean error across all locations is comparable or less than the mean error from Kulkarni and Colburn (2004). Different model orders of all-pole and all-zero were systematically applied to the measured DTFs for the SDO and SOW datasets. Figure 3.4(a) shows plots for the mean dB error calculated using Equation (2.5) over the 144 locations versus model order for SDO human dataset (left ear) using all-pole and all-zero models.

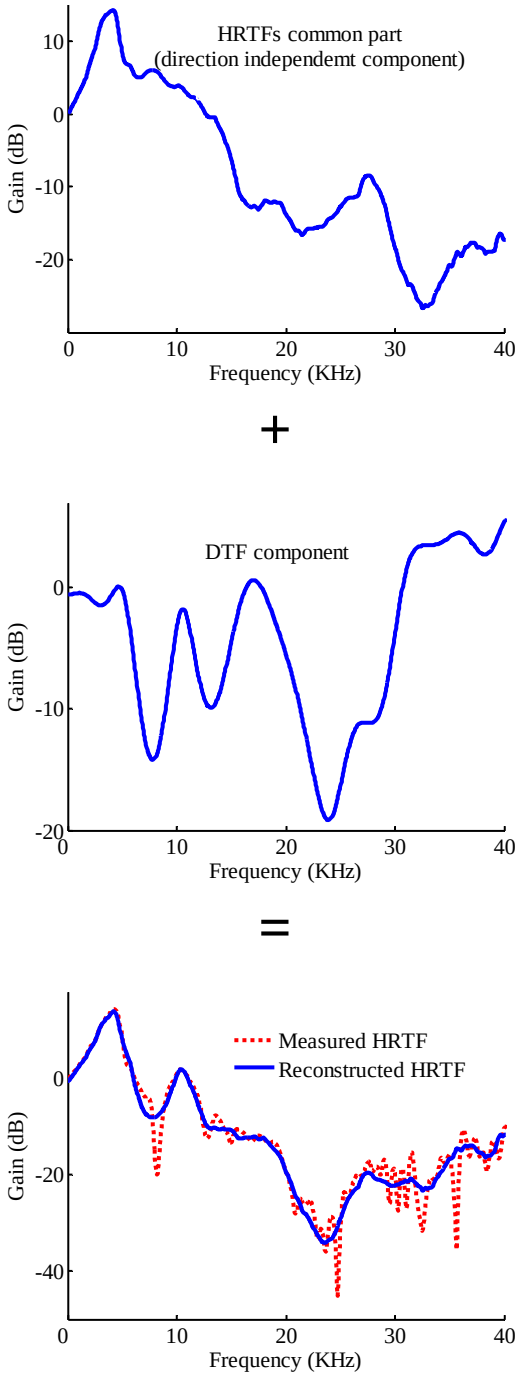


Figure 3.2. DTF at (0º,-30º) modeled using all-pole model of order 25 (middle) is added to the common component (top) of the HRTF measurements (*cat 1107*) resulting in an accurate reconstruction (bottom, solid line) of the measured HRTF (bottom, dotted line) at that location.

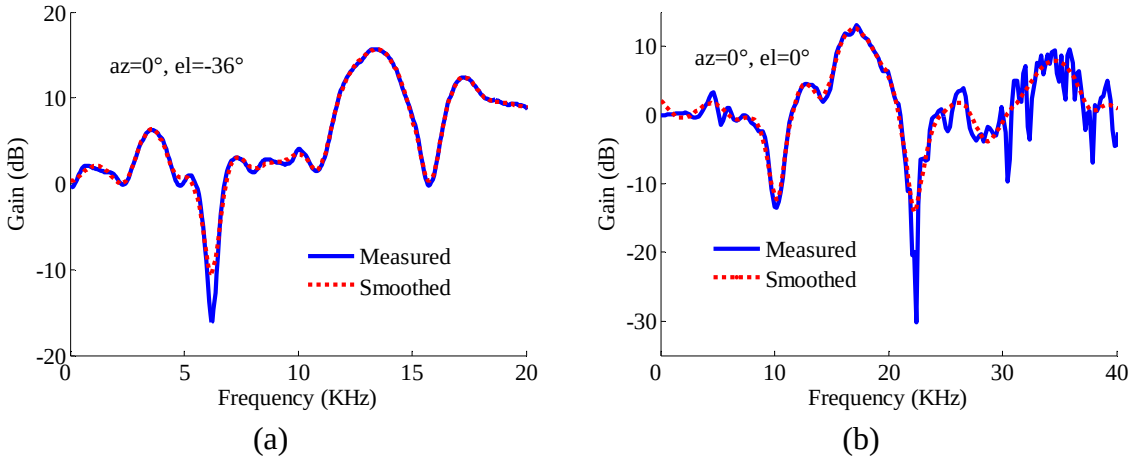
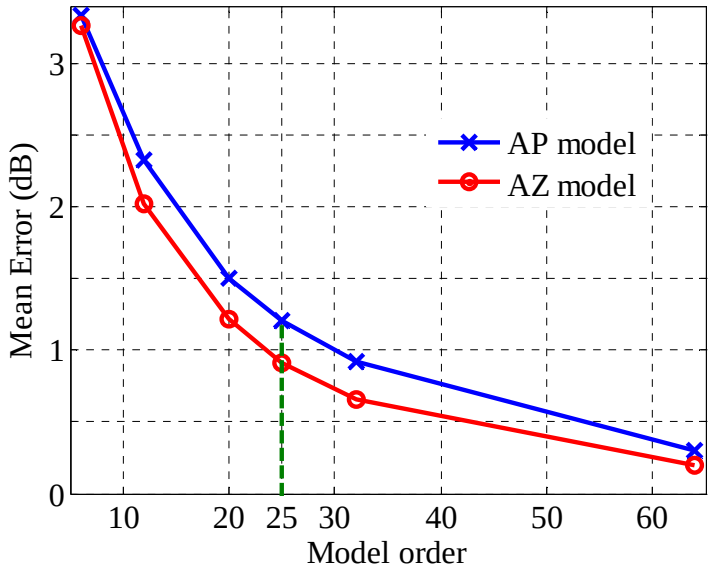


Figure 3.3. DTF smoothing using "*Symmlet 17*" filter bank. (a) Human (SDO) right-ear dataset. (b) *Cat 1107* right-ear dataset.

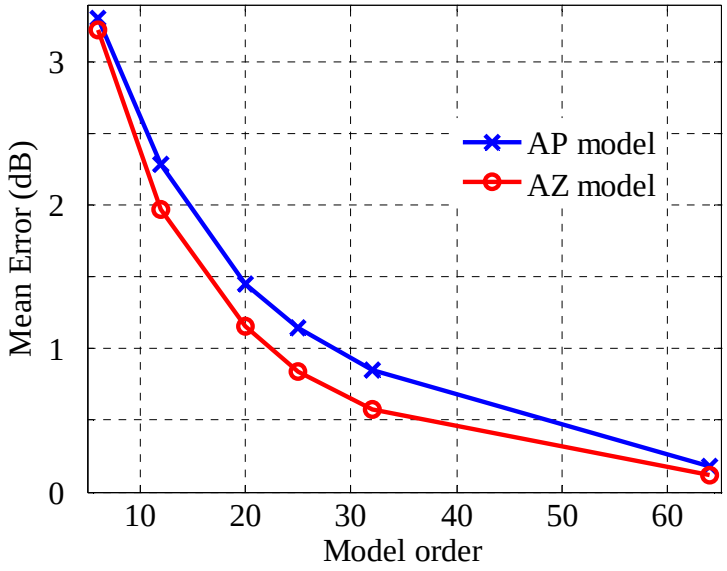
Error was computed with respect to the raw unsmoothed DTFs. Order 25 of all-pole model is marked in Figure 3.4(a) by a dashed line with an equivalent mean error of 1.17 dB. For a given order, all-zero models give smaller error than all-pole when applied to the same dataset. Figure 3.4(b) shows the same plots but with modeled DTFs compared to smoothed-empirical DTFs. In that case, the error curves for a given model order are shifted down and give less error relative to those in Figure 3.4(a) and give less error. This was expected because the sharp peaks and notches of the raw DTFs (e.g., Figure 3.2), that are likely not perceptually resolvable (Kulkarni and Colburn, 1999), are reduced after smoothing. A mean error of 1.10 dB at order 25 was obtained for the all-pole model when modeled DTFs are compared to smoothed-empirical DTFs.

The results of Figure 3.4 show that a model order of 25 for both all-pole and all-zero models produces mean DTF reconstruction error values that are at or below the error values mentioned above from the psychophysical studies of Kulkarni and Colburn

(2004). Based on this error and the subjective listening tests performed by Kulkarni and Colburn (2004) we believe that the modeled DTFs will be perceptually similar to the empirical DTFs. Our all-pole model of order 25 gives mean errors of 1.17 and 1.04 dB and worst case fit error of 3.23 and 3.50 dB for the left and the right ear, respectively. All-zero model of order 25 gives mean errors of 0.91 and 0.83 dB and worst case fit errors of 2.25 and 2.53 dB for the left and the right ear, respectively. Although all-zero models orders lower than 25 could yield comparable error values to order 25 with all-pole model, these lesser orders produced substantially larger peak errors in reconstructed DTFs at some locations with values much higher than 3.5 dB.



(a)



(b)

Figure 3.4. DTF-reconstruction model order vs. mean error in dB for both all-pole and all-zero models calculated using Equation (2.5) between modeled and empirical DTFs (a), and modeled and smoothed-empirical DTFs (b).

The spatial location dependent distributions of errors and the resulting descriptive statistics of these errors for all-pole model with 25 poles as a function of the source

position for SDO and SOW datasets are shown in Figures 3.5-3.8 and Table 3.1. To make the comparison consistent with the results in (Kulkarni and Colburn, 2004), only frequencies between 300 and 15,000 Hz have been considered in the error calculations. The error calculations for the models in this research are directly comparable to the ones reported in Kulkarni and Colburn (2004) using the same error criteria of Equation (2.5). In other words, when applied to the same datasets (SDO and SOW), our models and those of Kulkarni and Colburn (2004) produce the same results, thus validating our model.

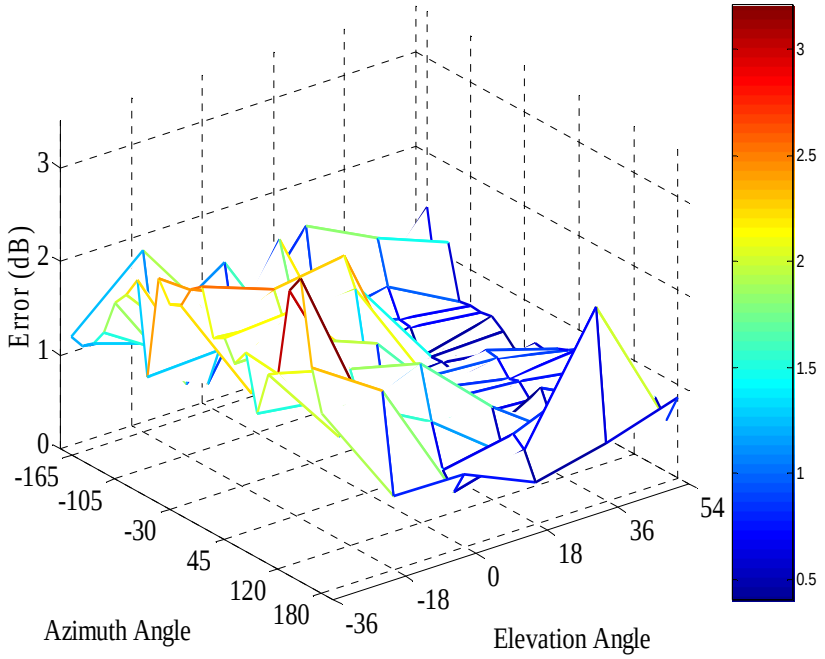


Figure 3.5. Error distributions for all-pole model with 25 poles as a function of the source position for SDO dataset-left ear.

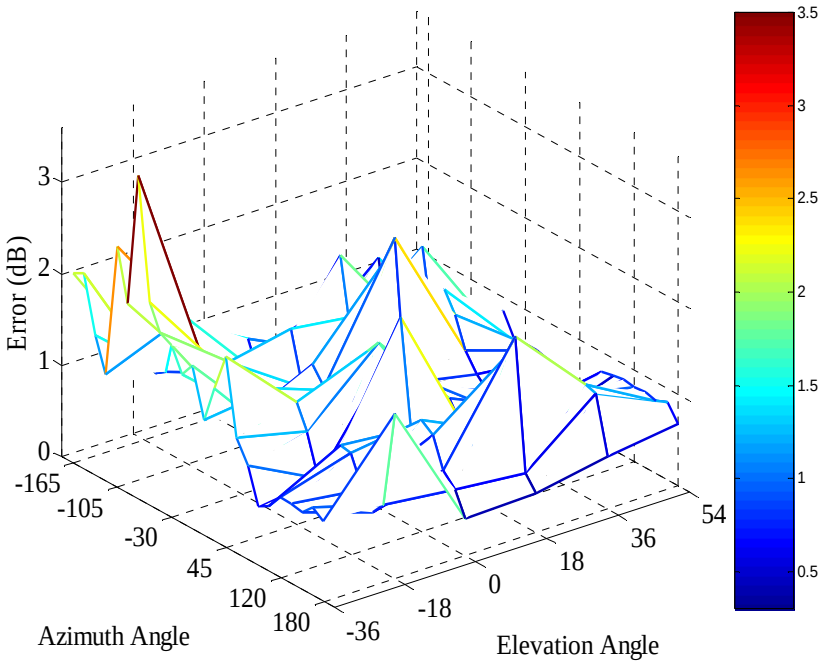


Figure 3.6. Error distributions for all-pole model with 25 poles as a function of the source position for SDO dataset- right ear.

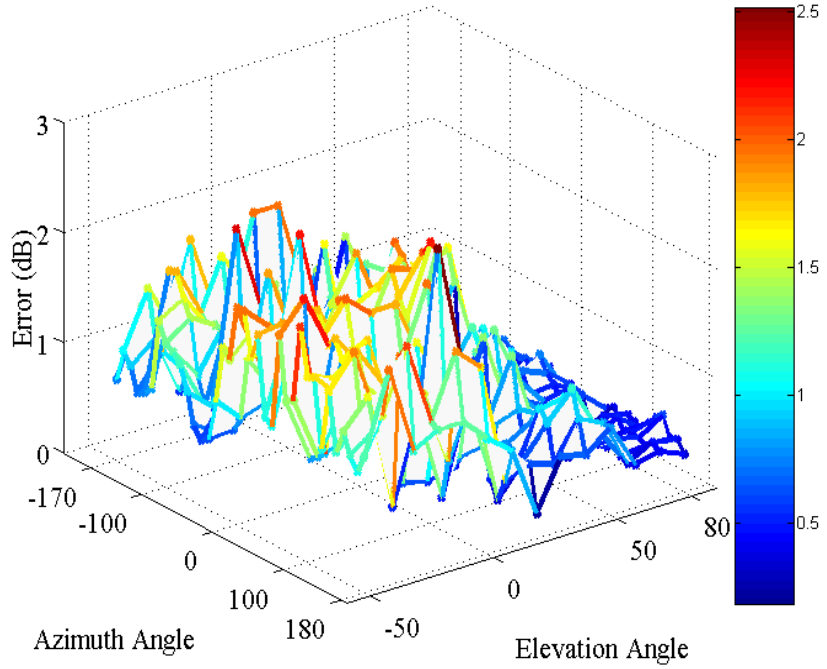


Figure 3.7. Error distributions for all-pole model with 25 poles as a function of the source position for SOW dataset- left ear.

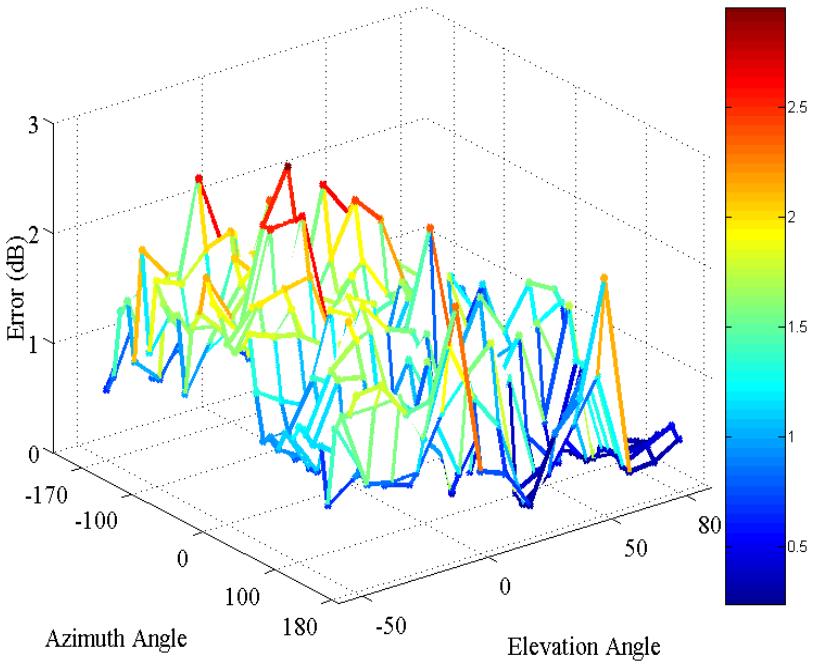


Figure 3.8. Error distributions for all-pole model with 25 poles as a function of the source position for SOW dataset- right ear.

Data set	Mean (dB)	STD (dB)	Max (dB) location (az,el)
SDO (L-ear)	1.17	0.65	3.23 (150°,-36º)
SDO (R-ear)	1.04	0.53	3.50 (-75°,-36º)
SOW (L-ear)	0.95	0.51	2.53 (110°,0º)
SOW (R-ear)	1.08	0.55	2.98 (-30°,-40º)

Table 3.1. Descriptive statistics for the error values calculated using Equation (2.5) between empirical and order 25 all-pole modeled DTFs for SDO and SOW data sets. SDO data set has 144 measurements while SOW data set has 505 measurements for each ear.

Figure 3.9 shows an example of a modeled DTF from the right ear SDO dataset at (0°,0º) using all-pole and all-zero models with order 25. The worst case fits of the SDO dataset for each model, all-pole or all-zero are shown in Figures 3.10(a) and 3.10(b), respectively. In all cases, the worst fit errors were located contralateral to the ear used for DTF measurement, locations that typically have low signal to noise ratio to begin with. The large errors were due entirely to mis-fitting the very narrow and deep spectral notches.

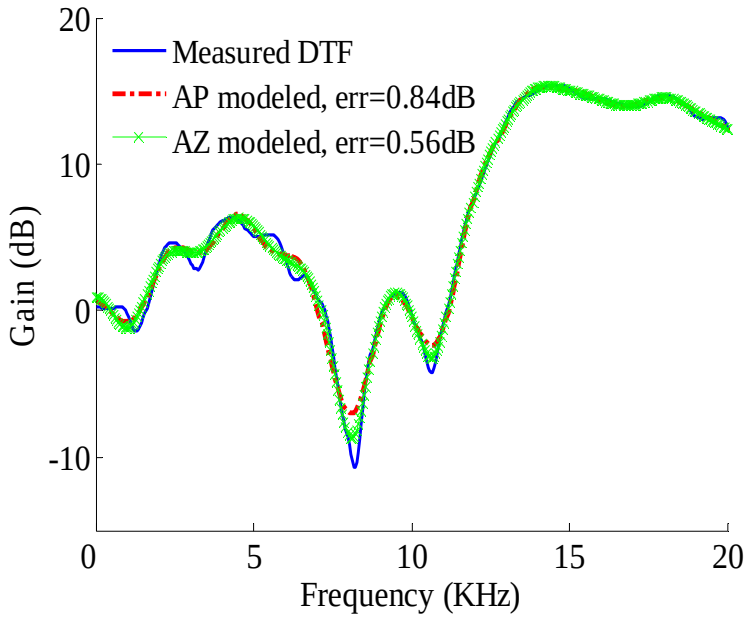
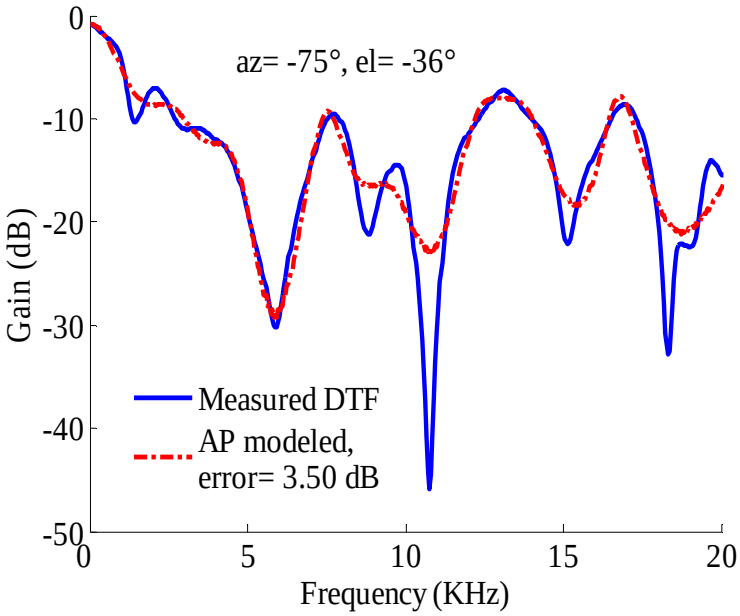
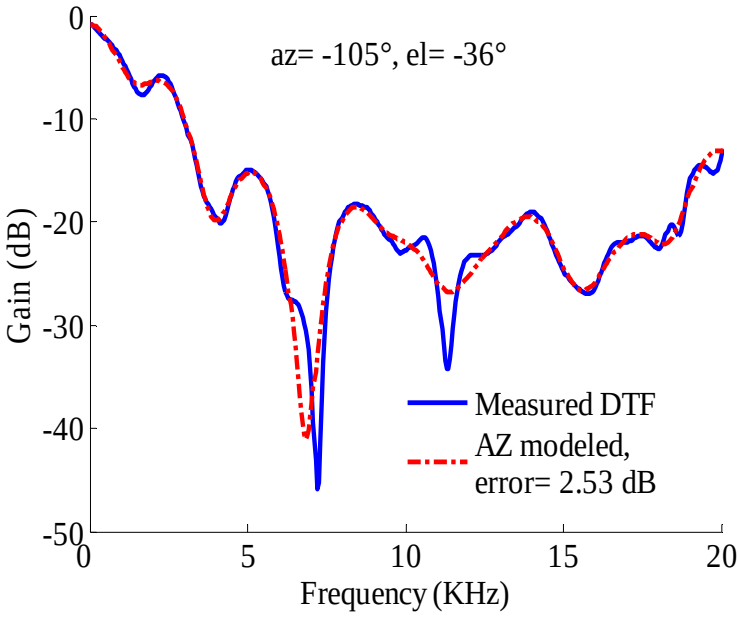


Figure 3.9. An example of all-pole and all-zero (order 25) fitting to DTF of location (0°,0º).



(a)



(b)

Figure 3.10. The worst case fits for (a) all-pole and (b) all-zero models.

The spatial location dependent distributions of errors and the resulting descriptive statistics of these errors for all-zero model of order 25 as a function of the source position for SDO dataset are shown in Figures 3.11, 3.12 and Table 3.2. In general, for a given model order, all-zero models gave lower errors than all-pole models when used for this dataset (see also Figure 3.4).

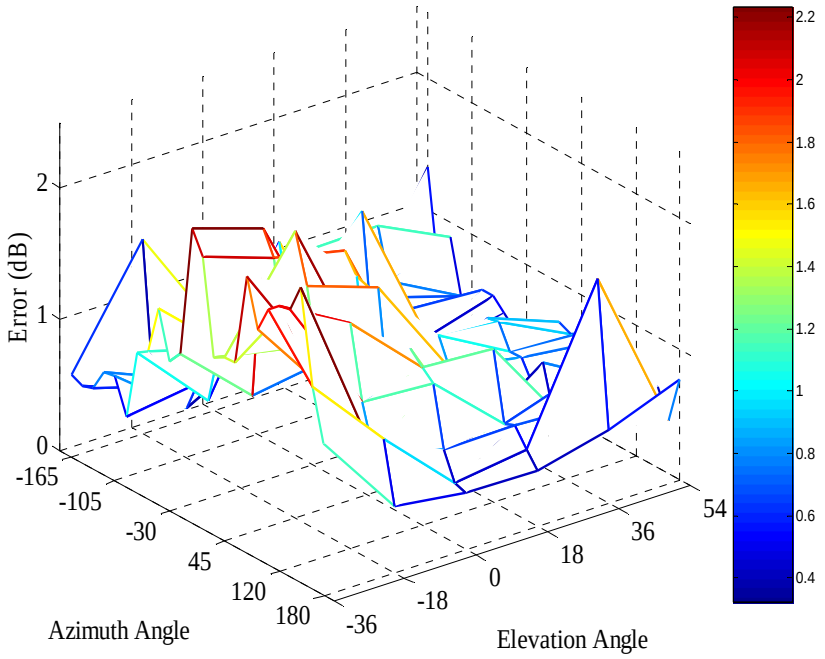


Figure 3.11. Error distributions for all-zero model with 25 zeros as a function of the source position for SDO dataset-left ear.

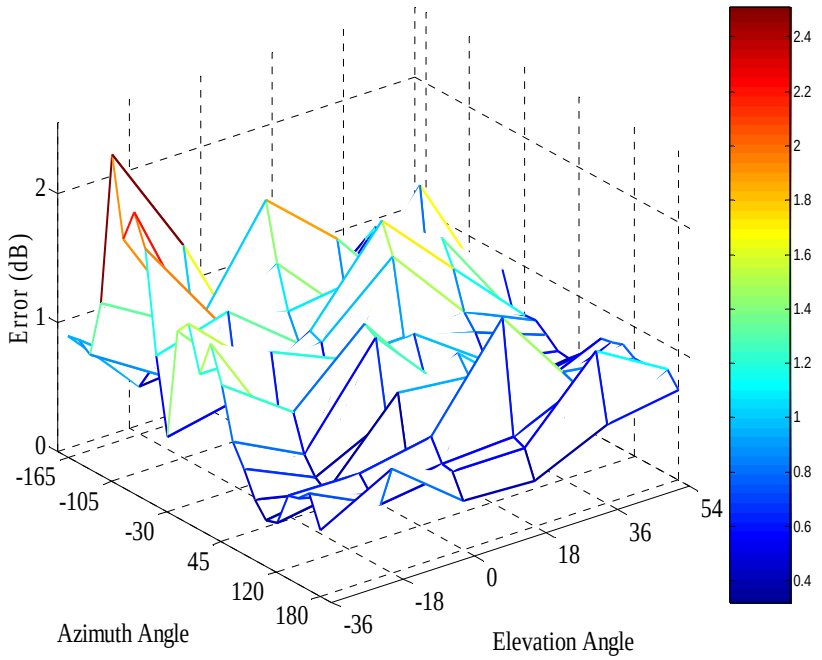


Figure 3.12. Error distributions for all-zero model with 25 zeros as a function of the source position for SDO dataset-right ear.

Data set	Mean (dB)	STD (dB)	Max (dB) location (az,el)
SDO (L-ear)	0.91	0.48	2.25 (150°,-36º)
SDO (R-ear)	0.83	0.40	2.53 (-105°,-36º)

Table 3.2. Descriptive statistics for the error values calculated using Equation (2.5) between empirical and order 25 all-zero modeled DTFs for SDO data set. SDO data set has 144 measurements for each ear.

In summary, using the exact same HRTF datasets and general modeling strategy as Kulkarni and Colburn (2004), we show here that HRTF-reconstruction error for all-pole and all-zero models decreases systematically with increasing model order. Moreover, for a given order, all-zero models produce lower error than all-pole. In their modeling and

psychophysical studies of HRTF reconstructions Kulkarni and Colburn (2004) showed that HRTFs reconstructed with a 25^{th} order all-pole model could not be reliably discriminated from empirical HRTFs. In other words, perceptually the empirical HRTFs and the modeled HRTFs are the same when modeled with order 25. One can think of this model order, and the associated across-location error metric, as a ceiling error below which reconstructed HRTFs are no longer likely to be indiscriminable from empirical HRTFs. The reconstruction error produced by a 25^{th} order model is used here as a ceiling error and ensure that all HRTFs reconstructed using models produce less error than this, for both human and cat HRTFs.

3.3.2 Magnitude squared coherence error

Figure 3.13 shows an example of the magnitude squared coherence (MSC) plot at (0°,0°) between the measured and the all-pole (25) modeled DTFs from SDO (L-ear) dataset, in addition to the plots of the measured and the modeled DTFs at that location. While the error calculated using Equation (2.5) describes the mean log difference between the dB values of the measured and the modeled DTFs, the MSC describes the difference between the spectral shapes themselves quantitatively. Thus, the MSC is a more global metric to assess the fidelity of the model reconstruction.

In general, a good model for DTF reconstruction should provide both low dB error values and high MSC values. But note that these two metrics need not be correlated. For example, a simple multiplication factor in a reconstructed DTF relative to the empirical DTF would yield large dB errors, but may have no impact at all on the MSC error

because the spectral shape had been maintained. And because perception of HRTFs and DTFs depends upon the shape of the spectrum and not the absolute values of the magnitude, the reconstruction with the high MSC error will be more perceptually similar to the empirical HRTF than the reconstruction with the high dB error. The mean MSC value (±1 SD) and the mean RMS error (±1 SD) for all locations in the SDO (L-ear) dataset for the all-pole model of order 25 compared to other values for orders 64, 46, 32, 20 and 12 are shown in Figure 3.14. Plots of the mean MSC between the measured and the all-pole and all-zero modeled DTFs over all locations versus the model order are shown in Figure 3.30 where they are also compared to the data from *cat 1107*.

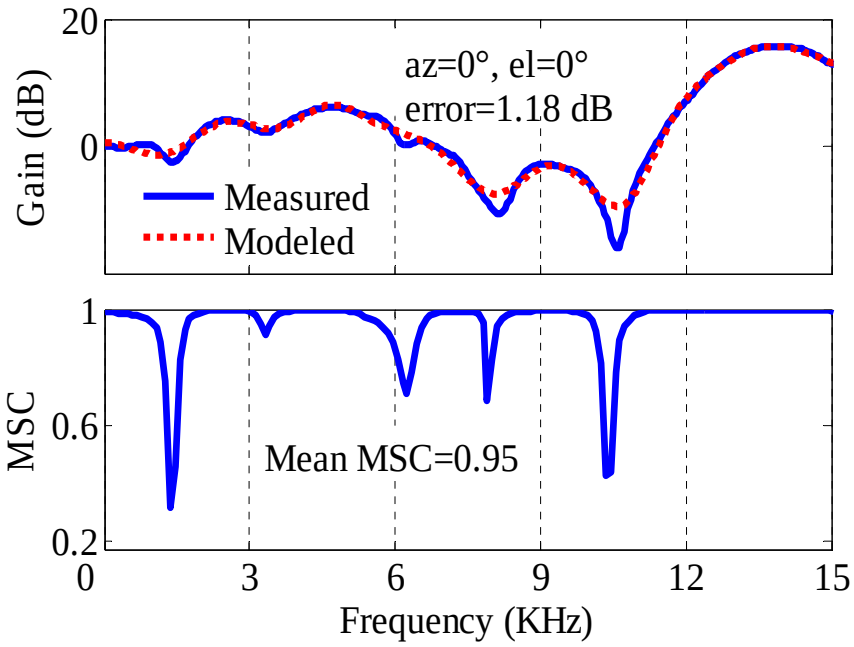


Figure 3.13. An example of the coherence spectrum: Measured, all-pole (25) modeled DTFs and the magnitude-squared coherence (MSC) at location (0°,0°) from SDO (L-ear) dataset. The mean MSC is calculated over the frequency range (2-15 kHz).

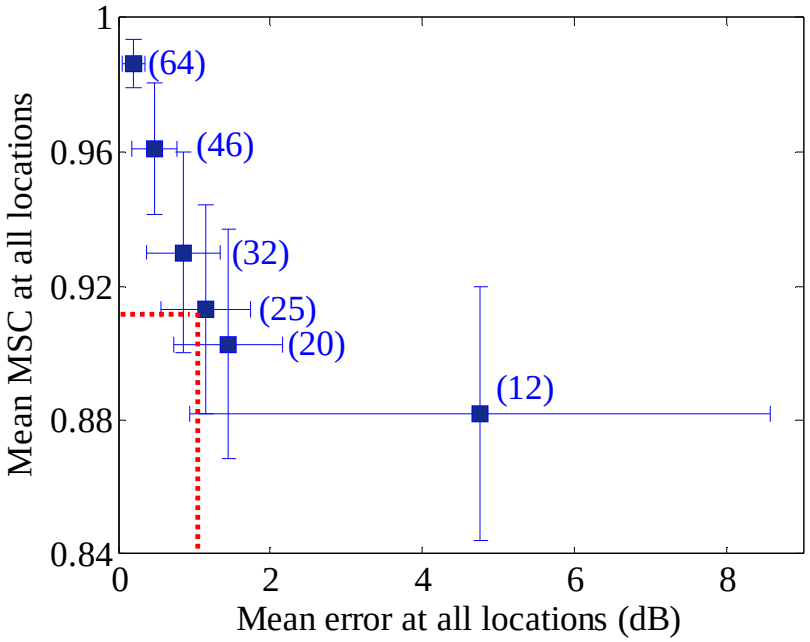


Figure 3.14. The mean magnitude-squared coherence (MSC) of all locations vs. the mean RMS error of all locations for all-pole model using Equation (2.5) at orders (64, 46, 32, 25, 20 and 12). These values at order 25 are marked by dotted lines. The error bars show the standard deviations of the mean across the 144 locations for each order.

3.4 Modeling Cat DTFs

Our modeling techniques were also applied to one of the cat datasets from (Rice et al., 1992), *cat 1107,* to test whether these DTF reconstruction models applied to animal data not only produce low error metrics based on Equation (2.5), but also preserve the main spectral features, such as the deep spectral notches, described in that paper. Measured, all-pole and all-zero modeled DTFs with order 25 for both models for *cat 1107-right ear* dataset are plotted in Figure 3.15 for sound source locations in the frontal field. Model order 25 was chosen initially because this order provided good fits to the human HRTFs above. These DTFs are plotted in the range of (2-33 KHz).

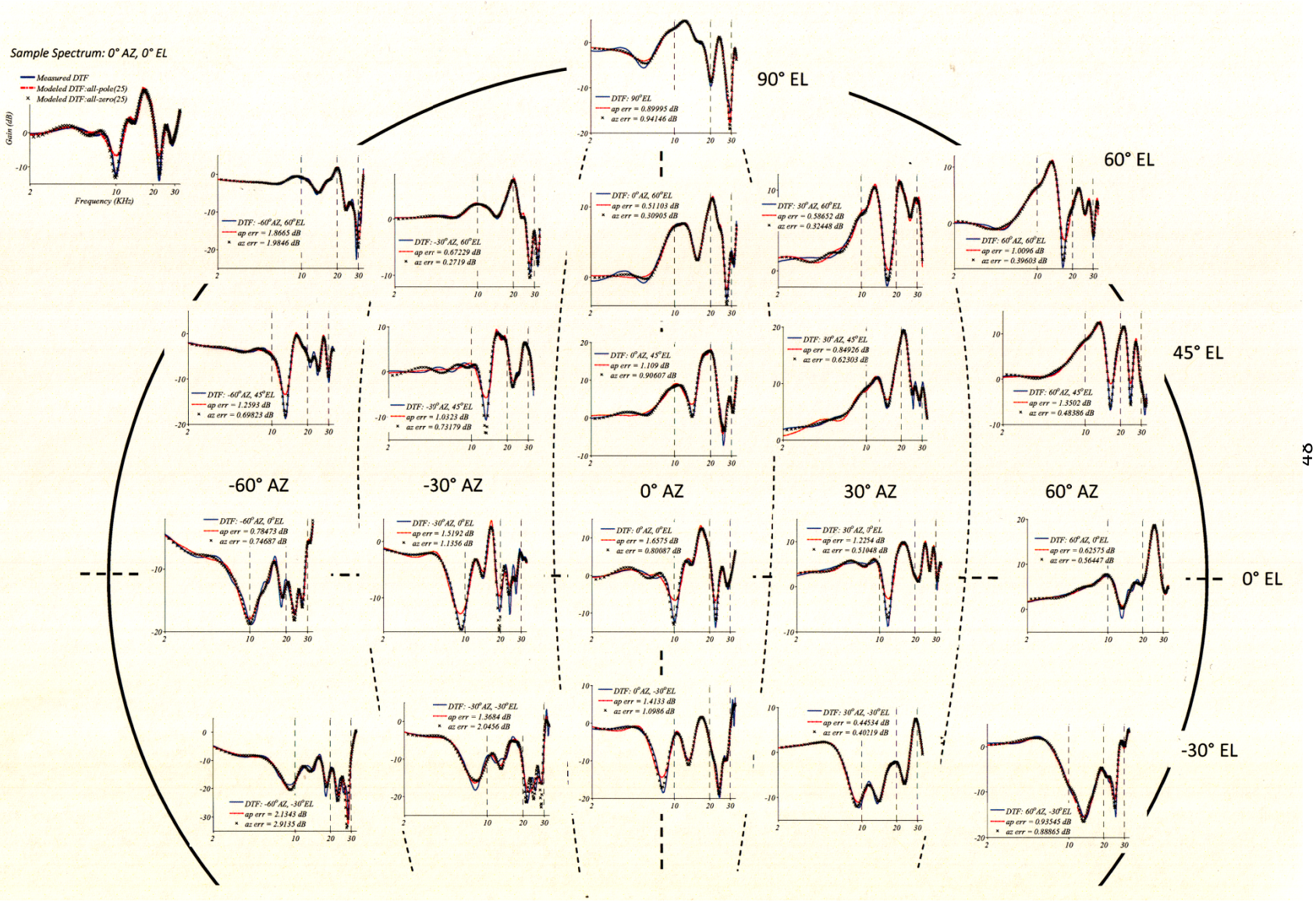
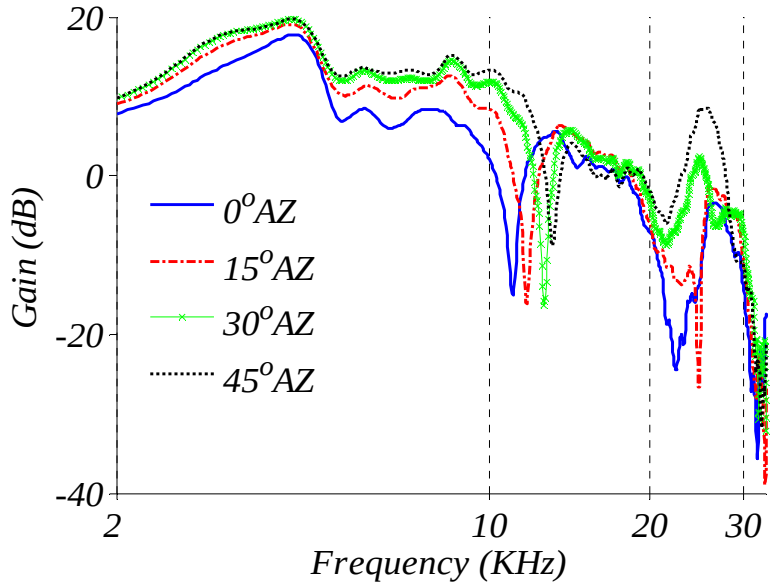

Sample Spectrum: 0° AZ, 0° EL
Measured DTF
Modeled DTF:all-pole(25)
Modeled DTF:all-zero(25)
Gain (dB)
Frequency (KHz)
90° EL
60° EL
45° EL
0° EL
-30° EL
-60° AZ
-30° AZ
0° AZ
30° AZ
60° AZ
DTF: 90°EL
ap err = 0.89995 dB
az err = 0.94146 dB
DTF: -60°AZ, 60°EL
ap err = 1.8665 dB
az err = 1.9846 dB
DTF: -30°AZ, 60°EL
ap err = 0.67229 dB
az err = 0.2719 dB
DTF: 0°AZ, 60°EL
ap err = 0.51103 dB
az err = 0.30905 dB
DTF: 30°AZ, 60°EL
ap err = 0.58652 dB
az err = 0.32448 dB
DTF: 60°AZ, 60°EL
ap err = 1.0096 dB
az err = 0.39603 dB
DTF: -60°AZ, 45°EL
ap err = 1.2593 dB
az err = 0.69823 dB
DTF: -30°AZ, 45°EL
ap err = 1.0323 dB
az err = 0.73179 dB
DTF: 0°AZ, 45°EL
ap err = 1.109 dB
az err = 0.90607 dB
DTF: 30°AZ, 45°EL
ap err = 0.84926 dB
az err = 0.62303 dB
DTF: 60°AZ, 45°EL
ap err = 1.3502 dB
az err = 0.48386 dB
DTF: -60°AZ, 0°EL
ap err = 0.78473 dB
az err = 0.74687 dB
DTF: -30°AZ, 0°EL
ap err = 1.5192 dB
az err = 1.1356 dB
DTF: 0°AZ, 0°EL
ap err = 1.6575 dB
az err = 0.80087 dB
DTF: 30°AZ, 0°EL
ap err = 1.2254 dB
az err = 0.51048 dB
DTF: 60°AZ, 0°EL
ap err = 0.62575 dB
az err = 0.56447 dB
DTF: -60°AZ, -30°EL
ap err = 2.1343 dB
az err = 2.9135 dB
DTF: -30°AZ, -30°EL
ap err = 1.3684 dB
az err = 2.0456 dB
DTF: 0°AZ, -30°EL
ap err = 1.4133 dB
az err = 1.0986 dB
DTF: 30°AZ, -30°EL
ap err = 0.44534 dB
az err = 0.40219 dB
DTF: 60°AZ, -30°EL
ap err = 0.93545 dB
az err = 0.88865 dB

Figure 3.15. Measured and modeled DTFs from the frontal field of cat *(1107-right ear)* dataset provided by Rice and his colleagues (Rice et al., 1992). The position indicated in the figure represents the location of the sound source which ranges from -60° to 60° in the azimuthal plane and from -30 ° to 90° in the vertical plane. At each position, there are three plots, smoothed-measured DTF (solid), modeled DTF with all-pole of order 25 (dash-dot line) and modeled DTF with all-zero of order 25 (cross markers) with the error resulted from each fitting.

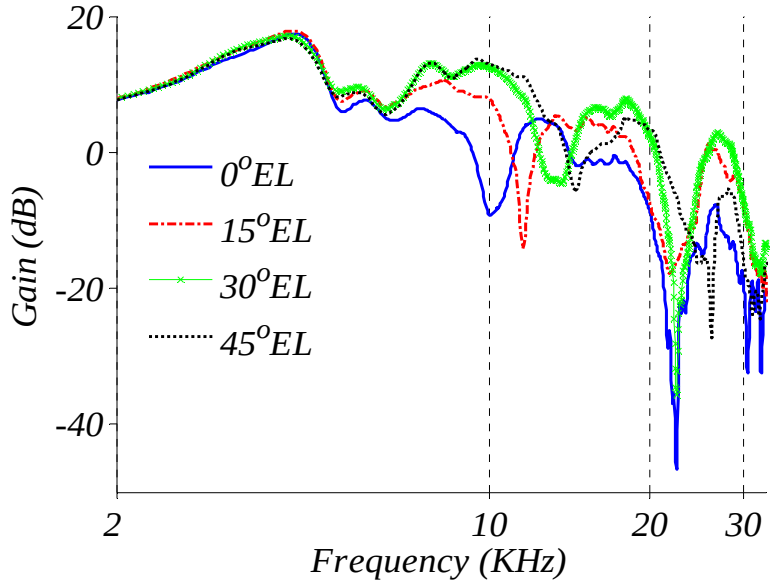
Initial observations from Figure 3.15, reveal that both all-pole and all-zero models with order 25 provide excellent fits to the measured DTFs in terms of the error metric (Equation (2.5)) and that both models also preserve the main features of the measured DTFs (peaks and notches). The mean error at all locations from *cat 1107-right ear* dataset using all-pole and all-zero models of order 25 are 1.71 and 1.47 dB and the worst case fits have errors of 3.27 and 3.20 dB, respectively.

In cats, as shown in Figure 3.15, the FN changes from about 8 to 18 kHz with the change in elevation from low to high, with similar behavior as the sound source is moved laterally from the midline in the azimuth with fixed elevation (e.g., Musicant et al., 1990; Rice et al., 1992; Tollin and Koka, 2009).

Figure 3.16 shows examples from *cat 1107* measured HRTFs of the increase in the FN frequency with the increase in the elevation angle and as the source moves horizontally in the frontal field toward the ear used for HRTF measurements.



(a)



(b)

Figure 3.16. (a) cat DTFs at azimuths (0°, 15°, 30° and 45°) with fixed elevation at 7.5°, (b) DTFs at elevations (0°, 15°, 30° and 45°) with fixed azimuth at 0º.

The systematic movement of the FN with the change in elevation and azimuth angle of the sound source is clarified in Figures 3.17-3.20. Smoothed-measured DTFs (Figure 3.17) and modeled DTFs using all-zero model of order 25 (Figure 3.18) from *cat 1107* dataset are plotted at a fixed 0° AZ and elevations from -30° to 60° with 7.5° step. A 20 dB shift has been added between each consecutive DTFs. Smoothed-measured (Figure 3.19) and modeled (Figure 3.20) DTFs using all-zero model of order 25 are plotted at a fixed 0° EL and azimuths from -30° to 60° with 15° step. The dotted arrow shows the systematic tendency of the FN movement with the increase in the elevation and the azimuth angle of the sound source.

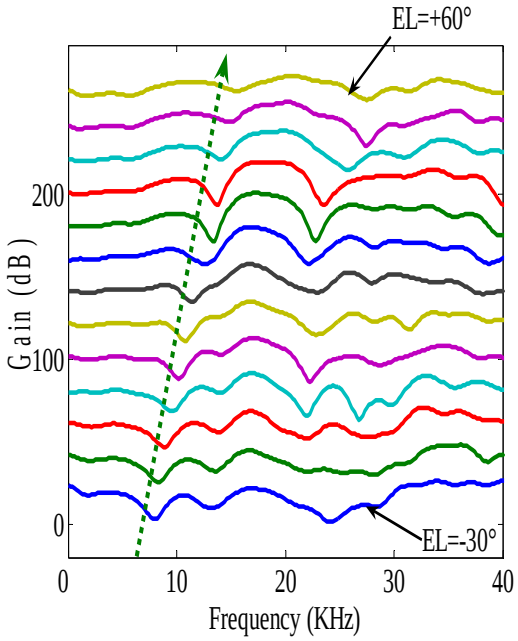


Figure 3.17. Smoothed measured cat DTFs at 0° AZ and elevations from -30° (bottom) to 60° (top) with a 7.5° step. The dotted arrow shows the systematic tendency of the FN movement with the change in elevation.

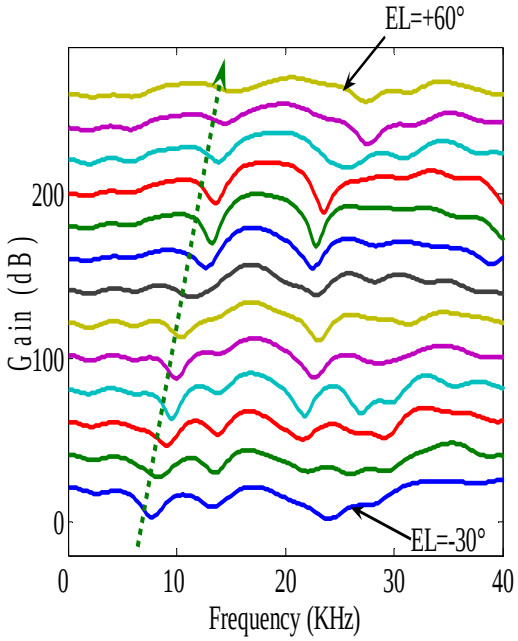


Figure 3.18. All-zero modeled DTFs at 0° AZ and elevations from -30° (bottom) to 60° (top) with a 7.5° step. The dotted arrow shows the systematic tendency of the FN movement with the change in elevation.

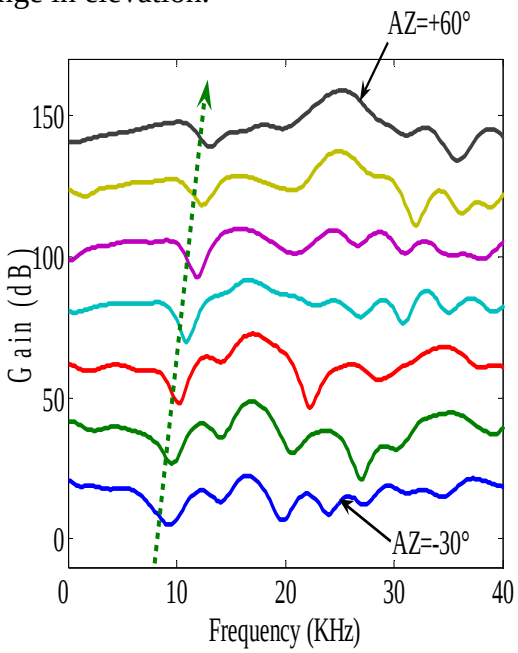


Figure 3.19. Smoothed measured cat DTFs at 0° EL and azimuths from -30° (bottom) to 60° (top) with a 15° step. The dotted arrow shows the systematic tendency of the FN movement with the change in azimuth.

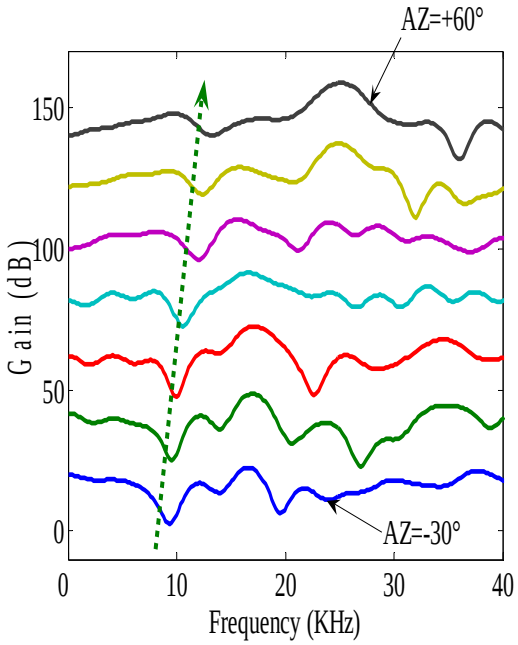


Figure 3.20. All-zero modeled DTFs at 0° EL and azimuths from -30° (bottom) to 60° (top) with a 15° step. The dotted arrow shows the systematic tendency of the FN movement with the change in azimuth.

To test more generally the hypothesis that the modeled DTFs preserved the spectral notch feature, the degree to which the systematic movement of the first notch frequency (FNF) with source location was preserved was examined for both all-pole and all-zero models with sources confined to the frontal field, which is where the FN feature in cats is very clear and prominent (Rice et al., 1992; Tollin and Koka, 2009). Both models preserved this feature quite well as shown in detail in Figures 3.21-3.28. The FN frequency changes in the measured, smoothed, all-pole and all-zero modeled DTFs with the change in the elevation angle at fixed azimuth angle 0° are shown in Figure 3.21. Figure 3.22 shows the same, but with the change in azimuth angle at fixed elevation angle of 15°. To test the validity of these models in preserving this feature at all locations in the frontal field, the FN frequencies extracted from smoothed DTFs (Figure 3.23), all-pole modeled DTFs

(Figure 3.24) and all-zero modeled DTFs (Figure 3.25) were plotted as a function of the FN frequencies from the empirical DTFs. Regression lines were computed for each relation using simple linear regression after forcing the regression line to pass through the origin (see Chapter 2 for details).

The values of the regression lines' slopes (β) and the resulting r^2 values plotted in Figures 3.23-3.25 show that the systematic movement of the FN was not affected by the smoothing technique and that it was well preserved by both models. The linear regression was performed with 1000 bootstrap replications for 209 locations. The resulting distributions of bootstrapped regression line slopes were compiled, from which the 95% confidence intervals was computed. Figures 3.26 and 3.27 show the distributions of regression line slopes from order 25 all-pole and all-zero models, respectively. The 95% confidence intervals for these distributions were [0.989-1.001] for the all-pole and [0.997-1.006] for the all-zero model. Because the regression line slope of value 1.0 is included within the 95% confidence intervals, both models accurately preserve this important spectral feature. Moreover, boxplots of the FN frequency absolute difference between measured and modeled DTFs in both models are shown in Figure 3.28. So, in addition to the low dB error values that are comparable to the results in (Kulkarni and Colburn, 2004), we show for the first time that these kinds of low-order DTF models also preserved the important FN frequency and its systematic movement with the change in the sound source location accurately.

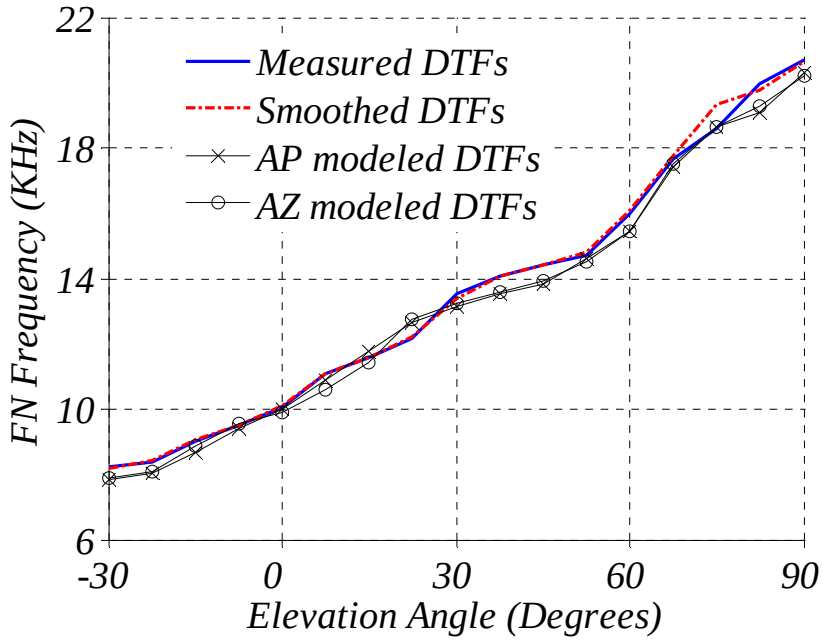


Figure 3.21. Change of FN frequency in measured, smoothed, all-pole (order 25) and all-zero (order 25) modeled DTFs with the change in elevation at fixed 0° azimuth.

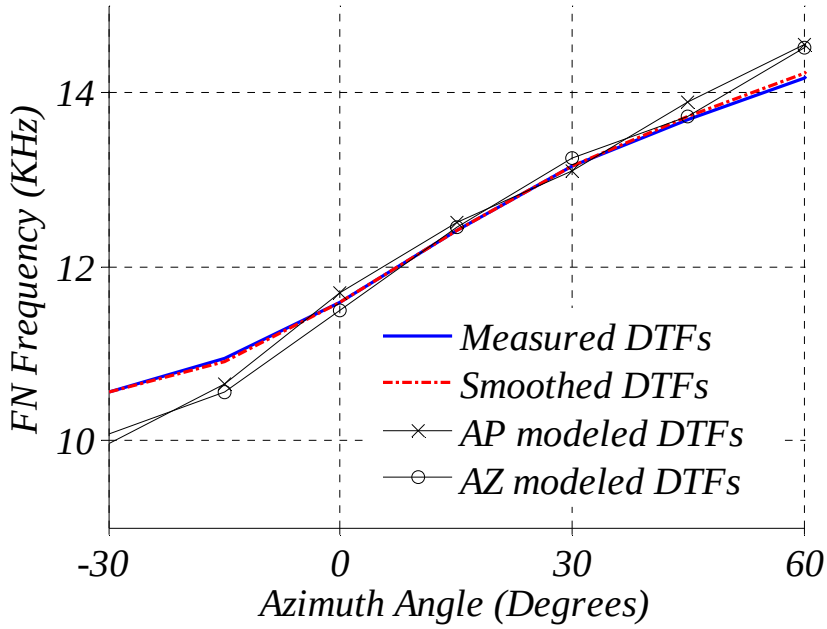


Figure 3.22. Change of FN frequency in measured, smoothed, all-pole (order 25) and all-zero (order 25) modeled DTFs with the change in azimuth at fixed 15° elevation.

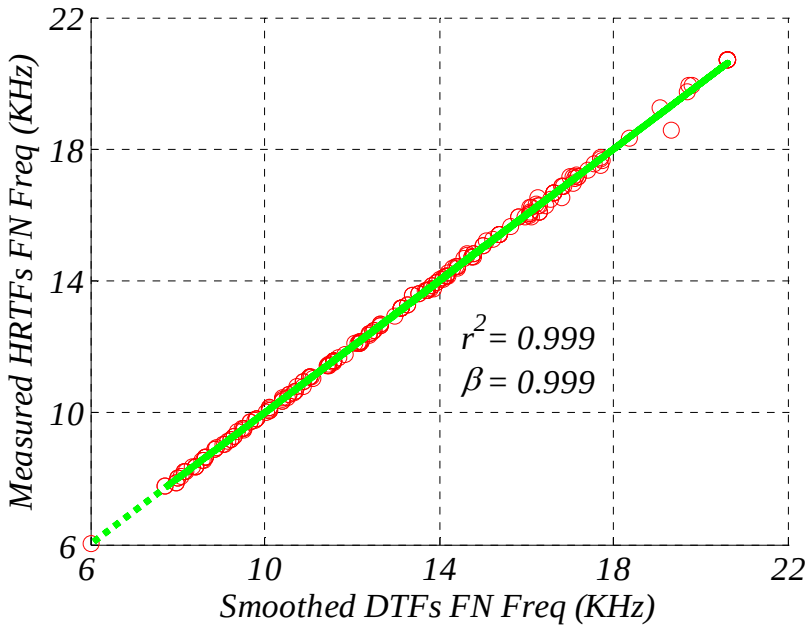


Figure 3.23. FN frequency in the measured HRTFs plotted with the ones extracted from smoothed DTFs. The green line represents the linear regression.

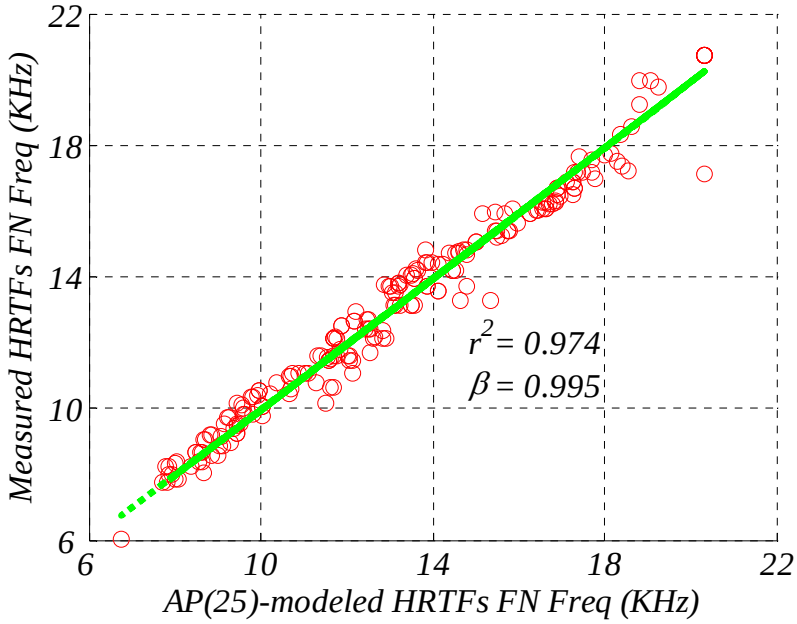


Figure 3.24. FN frequency in the measured HRTFs plotted with the ones extracted from all-pole modeled DTFs. The green line represents the linear regression.

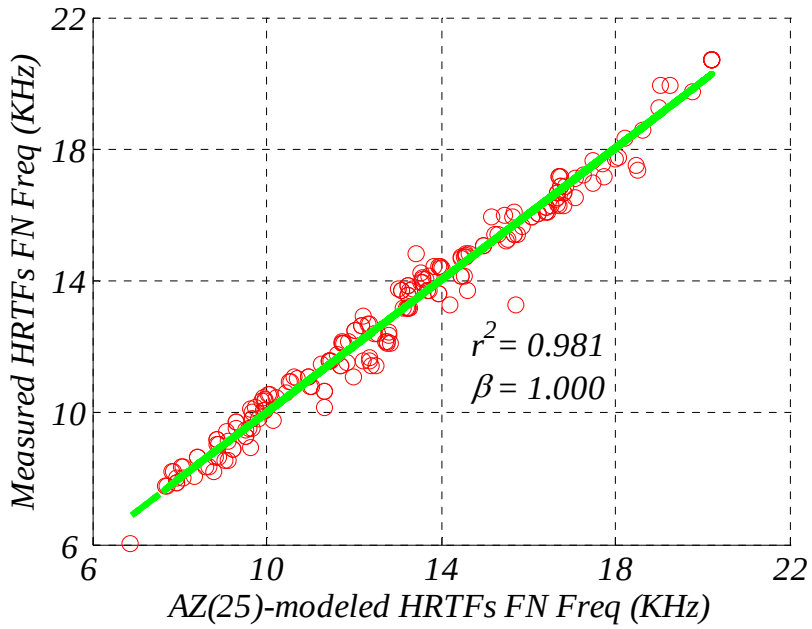


Figure 3.25. FN frequency in the measured HRTFs plotted with the ones extracted from all-zero modeled DTFs.

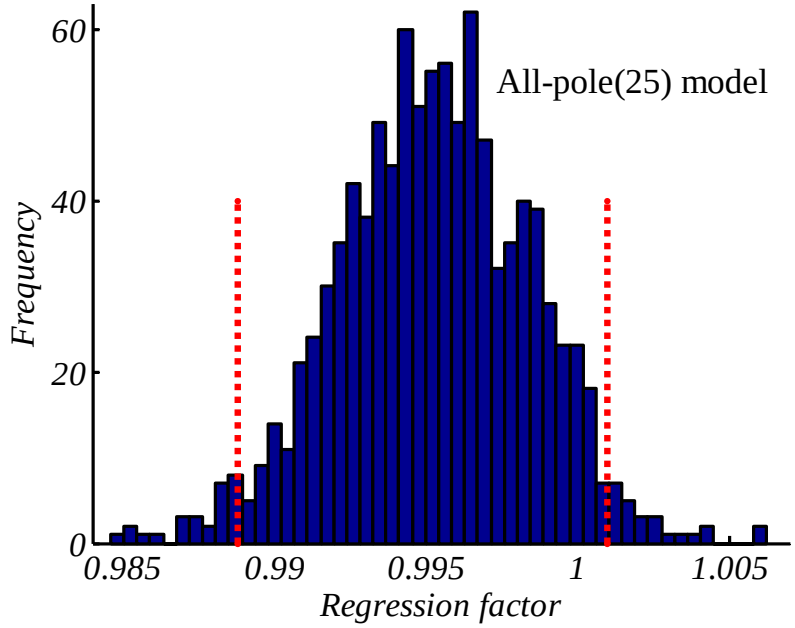


Figure 3.26. Results of bootstrap resampling of 1000 replications for the regression line slope between FN frequency in the measured DTFs and all-pole modeled DTFs.

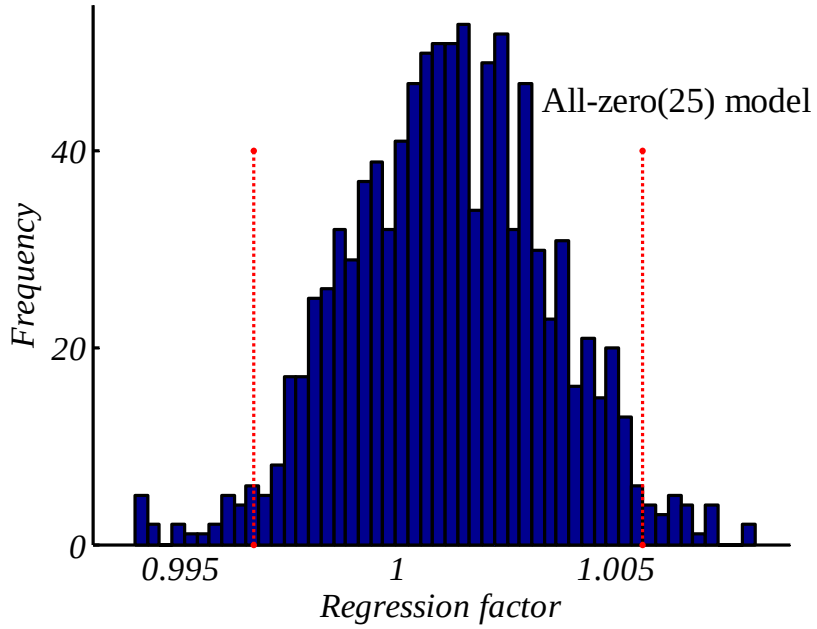


Figure 3.27. Results of bootstrap resampling of 1000 replications for the regression line slope between FN frequency in the measured DTFs and all-zero modeled DTFs.

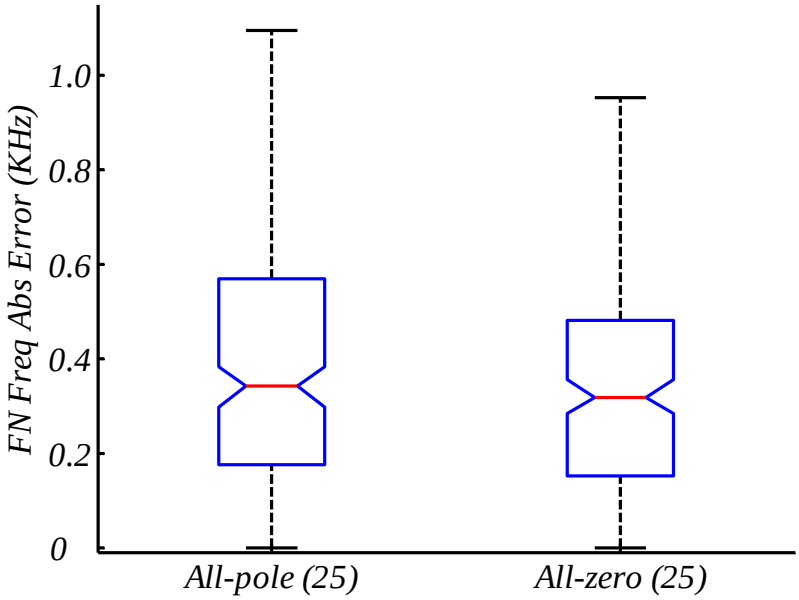


Figure 3.28. Boxplots of the FN frequency absolute difference between measured and all-pole and all-zero modeled DTFs. The line in the middle of each box represents the median and the lines at the upper and the lower edge of each box represent the upper and the lower quartiles, respectively.

In cats, the change in the FN frequency for changes in source elevation is larger than that with the changes in azimuth. For example, while the FN frequency in the measured DTFs increases from 8.16 to 16.08 kHz for elevation angles from -30° to 60° at 0° azimuth (resulting in ~88 Hz/degree for empirical DTFs and ~84 for both 25^{th} order all-pole and all-zero models), it increases from 8.85 - 13.15 kHz for the same angular displacement in azimuth at 0° elevation (~48 Hz/degree for empirical and ~50 for all-pole and ~45 for all-zero).

The reconstruction of the broadband spectral patterns of the DTFs using these models has been also tested using the MSC metric. Figure 3.29 shows an example of the MSC plot at location (0°,45°) from *cat 1107* (R-ear) dataset between the measured and the 25^{th} order all-pole modeled DTFs, in addition to the plots of the measured and the modeled DTFs at that location. Similar results were observed in the models of UCHSC dataset from Tollin and Koka (2009).

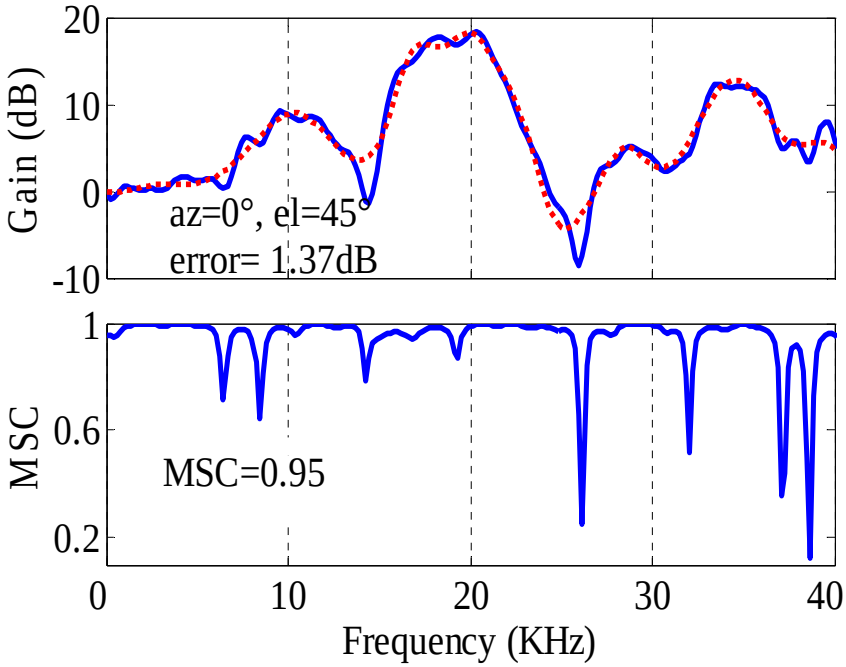


Figure 3.29. An example of the coherence spectrum: Measured, all-pole modeled DTFs (order 25) and the magnitude-squared coherence (MSC) at location (0°,45°) from *cat 1107* (R-ear). The mean MSC is calculated over the frequency range (2-33 kHz).

Plots of the mean MSC between measured and all-pole and all-zero modeled DTFs over all locations versus the model order using cat dataset (*cat 1107*) and human dataset (SDO) are shown in Figure 3.30. At order 25, the all-pole model gave mean MSC values ranged from 0.79 to 0.98 (mean = 0.92 ±0.04) for SDO (R-ear) dataset over the frequency range (0.3-15 kHz), and ranged from 0.68 to 0.96 (0.86 ±0.06) for *cat 1107* (R-ear) dataset over the frequency range (2-33 KHz). All-zero model had mean MSC values ranged from 0.28 to 0.98 (0.85±0.14) for SDO dataset, and ranged from 0.66 to 0.97 (0.86±0.07) for *cat 1107* dataset. Order 25 for both all-pole and all-zero models show satisfactory MSC values for the used human and cat datasets.

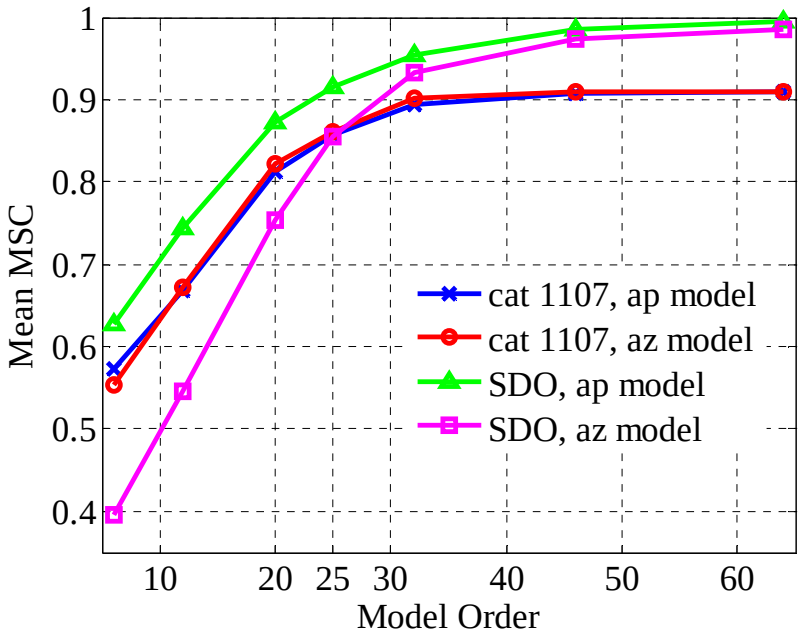


Figure 3.30. Mean magnitude-squared coherence (MSC) vs. all-pole and all-zero model order for human (SDO) and *cat 1107* datasets.

On average, low RMS error is associated with high MSC, as might be expected (although this relationship need not be always true). Some DTFs at some locations show low RMS error and also low MSC value because the DTF reconstruction misses a notch or a peak despite the fact that it fits the rest of the spectrum well. On the other hand, at some other locations, high RMS error is associated with high MSC value when the model follows the spectral shape of the empirical DTF and model the deep notches for example by shallow notches. In summary, 25^{th} order all-pole and all-zero models gave low RMS error (Figures 3.4-3.12), high MSC values (Figures 3.13, 3.14, 3.29 and 3.30) and preserved one of the important sound localization cues, the FN (Figures 3.15, 3.17-3.28), successfully for the tested human and cat datasets.

Chapter 4

Sound Source Localization Prediction from Head Related Transfer Functions Using Artificial Neural Networks

"The work done in this chapter was supervised by Prof. Rahmat Shoureshi, School of Engineering and Computer Science, University of Denver."

4.1 Overview

In this chapter, the performance of different architectures of feed-forward back propagation artificial neural networks (ANNs) trained by either *batch gradient descent (BGD)* or *Levenbergh-Marquardt (LM)* algorithm are evaluated for estimating the sound source direction (i.e. azimuth and elevation angles) from Head Related Transfer Function (HRTF) corresponds to that direction. Cat data are used in this study. The trained NNs efficiency will be measured by root-mean-squared (RMS) error and regression analysis between NNs outputs and the corresponding targets. The localization of the sound source direction from HRTFs facilitates the task of building virtual auditory displays which are used in psychoacoustic and physiological research, industry, some medical applications, military simulations, and entertainment.

4.2 Literature Review

Many techniques have been used to find the relations between the HRTFs and the sound source direction, how to build HRTFs at locations that are not measured or at finer resolution in the space, and how to predict the accurate location of the sound source given the corresponding HRTF of that location. Some of these techniques are based on artificial neural networks that have been used successfully in many fields such as modeling, classification and data processing.

Neti and his colleagues (1992) trained a three-layer NN to map spectra of stimuli (mainly, broad band noise) to sound source locations and they have used that mapping to study the importance of particular cues by systematically modifying or eliminating cues in different frequency regions and examined the response properties of the neurons after those modifications. Cat HRTFs in the frontal field have been used in that study. It has been shown that the frequency region (5 to 18 KHz) is the most important in supporting the calculations of HRTF models (Neti et al., 1992). The main feature in this region is what is known as the first notch (FN) in the literature (Rice et al., 1992), which is defined as the first prominent spectral minima. In a similar study but using human data, a two-layer feed-forward back propagation ANN has been trained to transform localization cues (spectral cues in addition to the inter-aural time difference cues) to a two-dimensional map that gives the direction of the sound source (Chung et al., 2000). Alim and Farag (2000) trained a NN using the binaural cues, interaural time and level differences, to

localize sound in the horizontal plane only. Human data were used for that purpose and the results have been evaluated using a human subjective test.

Hofman and his colleagues (1998) used a two layer feed-forward network to map pinna filter functions of a new modified ear (input layer), by adding a well-fitting custom-made molds within the concha of the regular ear, onto the elevation domain (one output unit that resembles the elevation). Other researchers (Furukawa et al., 2000) used ANN to identify the sound source location using spike patterns of cortical neural ensembles as inputs to that network. The input layer takes spike patterns from the brain and uses a feed-forward NN to find the sound location in the azimuth plane only.

In this chapter, different NN architectures will be trained to transfer the spectral presentation of the head related transfer functions into two outputs that resemble the azimuth and the elevation angles of the sound source direction using only the spectral cues for sound localization.

4.3 Methods

4.3.1 Used data

Some of the data used in this study are taken from a dataset for one of the cats measured at the University of Colorado Health Science Center (UCHSC), provided by Prof Daniel Tollin and published in (Tollin and Koka, 2009). That dataset was sampled at 3° in vertical steps and 7.5º horizontal steps with a nominal sampling rate of 100 kHz. The other data set used here is from "*cat 1107*" from the study of (Rice et al., 1992). Source

positions for that dataset were 259 locations in the frontal field, between -75° and 75° AZ. All measurements are taken at a fixed distance of 1 meter from the subject's head. Here, only samples of frequency range of (2-33 KHz) will be used as inputs since it has been shown that there is little direction dependent information and minor effect in sound localization outside this frequency range (Kistler and Wightman, 1992; Rice et al., 1992; Langendijk and Bronkhorst, 2002).

4.3.2 Data preprocessing

For the purpose of this study, DTFs, the directional components of the HRTFs, will be used instead of the entire HRTF because they only have the directional components which carry the features that will change with the sound direction and exclude the common component among all HRTFs for a certain ear of a certain subject. That will also reduce the complications needed to be taken into consideration for each HRTF. This has been done by finding the average of the measured HRTFs in log scale frequency domain at all locations, then subtracting this average from each HRTF at each location in the logarithmic scale,

$$(DTF)_i = (HRTF)_i - \frac{1}{n}\sum_{j=1}^{n}(HRTF)_j \tag{4.1}$$

which is equivalent to division in the linear scale. In Equation (4.1), '*n*' is the number of the measured HRTFs for a certain dataset at one ear. All inputs will be normalized to fall in the range of -1.0 and 1.0 before applying them to the network. All DTFs have also

been smoothed using redundant wavelet transform to get rid of the unwanted details in the DTFs as described in the Methods section (2.3.4).

It has been shown in previous studies (Kulkarni and Colburn, 1998; Hacihabiboglu et al., 2002) that HRTFs can be smoothed significantly in frequency domain without affecting the perception of sound stimuli filtered by HRTFs. In other words, the very fine spectral detail in HRTFs is not important for sound localization as long as the major features are preserved (Kulkarni and Colburn, 1998; Jin et al., 2000; Asano et al., 1990). A "*symmlet* 17" filter bank is applied to the magnitude responses of the DTFs using Redundant Wavelet Transform (RWT) technique for smoothing, which helps to remove the spectral details that are not important in sound direction perception. The performance of the NN will be evaluated by the values of the root-mean-squared (RMS) errors between the network outputs and the targets and given in degrees. In addition to the RMS error, some regression analysis between the network response and the corresponding targets are performed. When the network gives outputs that are the same as the targets, then the slope of the regression line would be 1 and the y-intercept would be 0.

4.3.3 Neural network architecture

Two-layer feed-forward back propagation neural networks with tan-sigmoid function used in the hidden layer and linear function used in the output layer will be used. In this study, either *Batch Gradient Descent* or *Levenbergh-Marquardt* algorithm will be used as training techniques. Matlab 7.0 has been used for data conditioning and preprocessing, NNs training, and NNs evaluation.

In back propagation, a gradient vector that points to the steepest descent from the current point was calculated which means that as the training moves along this vector, it will decrease the error. These sequential movements will eventually find a minimum error. The inputs to the NN are samples of the normalized DTFs (in frequency domain). There will be two outputs, the elevation angle and the azimuth angle. Each output, azimuth and elevation, has its own trained network. The resulting angles will be given in radian which is simply converted to the angle in degrees for testing and evaluation of the NN for each output separately. Several numbers of neurons have been used in the hidden layer and the results have been compared.

4.4 Results and Discussion

Four different tests were done in this study. A description of each of them with the corresponding results is presented.

4.4.1 Test1: Spectra from median and horizontal plane-high spatial resolution:

The dataset used in this test consists of DTFs from the median plane (0° azimuth at various elevations) and the horizontal plane (0° elevation at various azimuths) from the dataset of UCHSC (Tollin and Koka, 2009). An advantage of using this dataset is that it has high resolution (3° and 7.5° steps in elevation and azimuth planes, respectively). In the first part of this test, the training data are DTFs at locations of 0° azimuth and elevations from -45° to 45° with 6° step. The test data are DTFs at locations of 0° azimuth and elevations from -45° to 45° with 3° step. Initially, two-layer feed-forward back propagation ANN with 8 neurons in the hidden layer and one neuron in the output

layer has been trained to transform the input DTFs into an angle that resembles the elevation angle of sound source location as shown in Figure 4.1. The empirical DTFs of this dataset have been down-sampled by the factor of 4 in the frequency domain (i.e. from 512 to 128 points) in order to reduce the inputs samples.

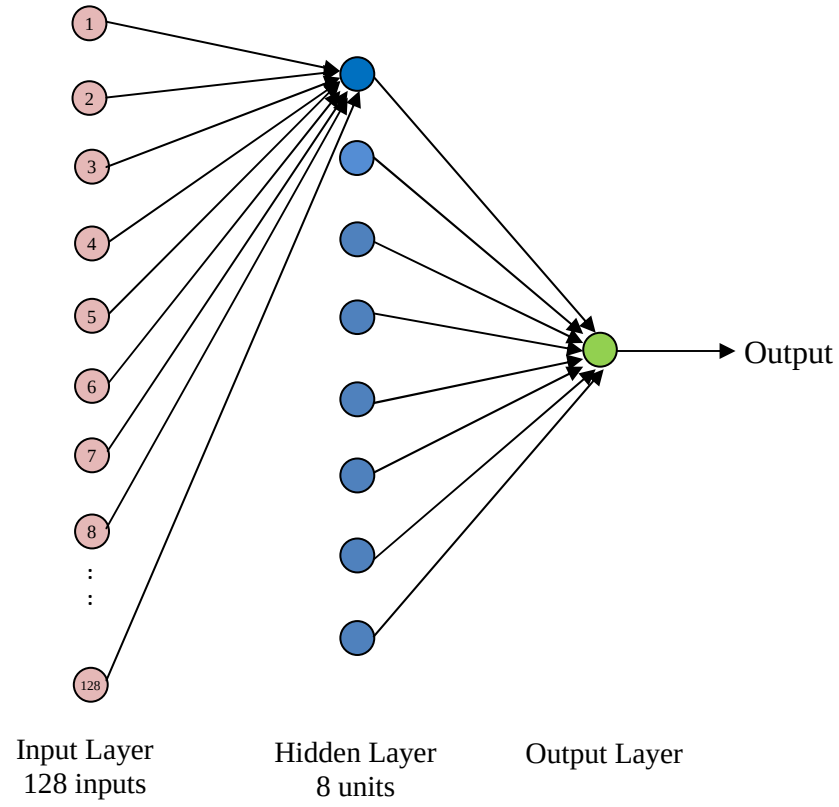


Figure 4.1. A schematic diagram of the neural network used in *Test1* with a 128 input frequency samples, 8 units in the hidden layer and one unit in the output layer. The output predicts the elevation or the azimuth angle of the sound source.

Different learning rates have been tested for the best performance of the network by comparing the resulting error between the targets and the simulated outputs from the testing DTF dataset with same number of epochs for network training. A learning rate of 0.005 has been used according to this comparison. Table (4.1) shows the performance of the networks with monaural inputs. Because it is possible to find a number of solutions

that provide a small final error E for the training dataset, we will focus more on the testing datasets errors that include the data that are not used for the training and it will be used to examine the performance of the networks with more generalization. The last column in Table (4.1) shows the mean and the standard deviation of the RMS error values between the predicted outputs from the NN and the targets in degrees. Both training techniques, *batch gradient descent* and *Levenbergh-Marquardt*, gave close error results to each other when used with 8 neurons in the hidden layer. When larger number of neurons in the hidden layer were tested on the same data sets (hidden layer with 24, 64 and 128 neurons), they gave lower error values on the training data set but didn't show any improvement on the testing data sets compared to the NN with 8 neurons in the hidden layer. Networks trained by *Levenbergh-Marquardt* show an increase in the RMS error as the number of the neurons in the hidden layer goes below 4. On the other hand, using smaller number of neurons for the networks trained by *batch gradient descent*, improved the performance of the network with the same NN parameters by giving less testing error. In addition to the error criterion used to evaluate the networks response, a regression analysis on the trained networks between the networks output and the corresponding targets has been performed. Figure 4.2(a) and 4.2(b) show the regression analysis results for the lowest RMS error case for both training techniques using 8 neurons in the hidden layer. The regression analysis of the lowest RMS error with the *batch gradient descent* training algorithm using one neuron in the hidden layer is shown in Figure 4.2(c).

The same idea has been applied to DTFs in the horizontal plane at fixed elevation 0°. Training data are the monaural inputs trained with DTFs at locations of 0° elevation and azimuths from -90° to 90° with 15° step and tested on locations of 0° elevation and azimuths from -90° to 90° with 7.5° step. The results are shown in Table (4.2).

Testing dataset	# neurons in hidden layer	Training Technique	# solutions	$E_{testing}$ (degrees)
DTFs in median plane	8	*BGD*	10	5.82±2.64
DTFs in median plane	8	*LM*	10	5.85±2.03
DTFs in median plane	4	*BGD*	10	5.82±2.62
DTFs in median plane	4	*LM*	10	4.10±2.41
DTFs in median plane	2	*BGD*	10	3.58±1.51
DTFs in median plane	2	*LM*	10	7.69±3.22
DTFs in median plane	1	*BGD*	10	3.03±0.85
DTFs in median plane	1	*LM*	10	7.23±3.47

Table 4.1. Testing dataset RMS error with different number of neurons in the hidden layer and different training techniques for a NN used to estimate the elevation angle in the median plane.

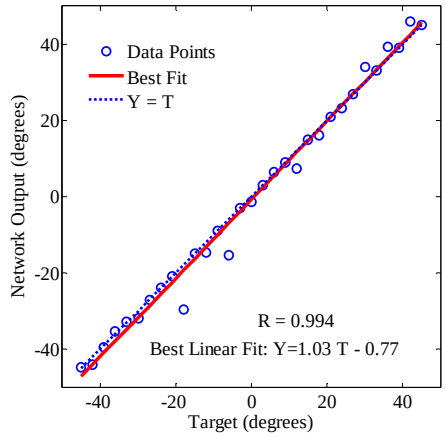
The higher resulting errors compared to the ones tested on the median plane can be justified by the fact that spectral features are clearer and show systematic movement with the change in elevation in a smoother way compared to those with the change in the horizontal plane. That may also explain why spectral features are very important for localization with the change in elevation (e.g., Musicant et al., 1990; Blauert, 1997; Rice et al., 1992), while the binaural cues are very important for sound source localization in the azimuthal plane. In addition, in this dataset, DTFs in the horizontal plane have lower resolution compared to those in the vertical plane, which may cause higher error values.

Testing dataset	# neurons in hidden layer	Training Technique	# solutions	$E_{testing}$ (degrees)
DTFs in horizontal plane	8	*BGD*	10	11.73±2.67
DTFs in horizontal plane	8	*LM*	10	10.42±3.55
DTFs in horizontal plane	4	*BGD*	10	9.25± 2.34
DTFs in horizontal plane	4	*LM*	10	12.79±3.88
DTFs in horizontal plane	2	*BGD*	10	9.96±2.37
DTFs in horizontal plane	2	*LM*	10	21.31±9.76
DTFs in horizontal plane	1	*BGD*	10	14.68±2.86
DTFs in horizontal plane	1	*LM*	10	29.09±8.43

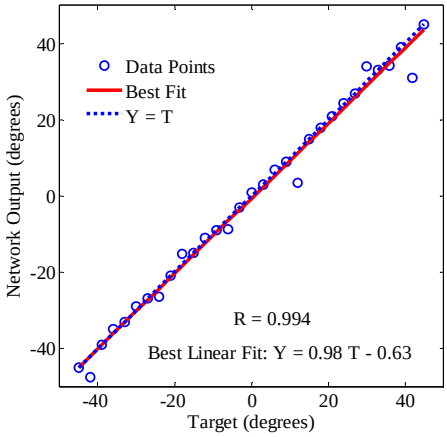
Table 4.2. Testing dataset RMS error with different number of neurons in the hidden layer and different training techniques for a NN used to estimate the azimuth angle in the horizontal plane.

4.4.2 Test2: Spectra from median and horizontal plane- lower resolution:

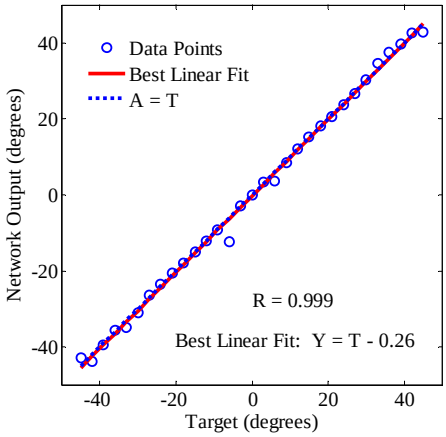
In this test, DTFs of "*cat 1107*" dataset from study (Rice et al., 1992) were used. Table 3 shows the performance of the networks with monaural inputs trained with DTFs at locations of 0° azimuth and elevations from -30° to 90° with 15° step and tested on locations of 0° azimuth and elevations from -30° to 90° with 7.5° step from the dataset measured by Rice et al. (1992). Only the error of the testing datasets will be mentioned to examine the performance of the networks with more generalization. The resulting error is higher when Rice dataset is used compared to the ones from UCHSC dataset and that can be justified by the lower spatial resolution for *cat 1107* data. When a larger number of neurons were used in the hidden layer, it improved the training data error but gave higher error values on the testing dataset.



(a)



(b)



(c)

Figure 4.2. Regression analysis for the network output versus the targets of elevation angles in the median plane with (a) 8 neurons in the hidden layer and using BGD training, (b) 8 neurons in the hidden layer and using LM training, and (c) 1 neuron in the hidden layer and using BGD training.

That is caused by the over-fitting for the training data. In order to improve the network generalization in this case, the number of neurons is kept low enough to provide an adequate fit without over-fitting. Table (4.3) shows a summary of the performance of the networks with monaural inputs in the median plane with 8 neurons in the hidden layer.

Testing dataset	# neurons in hidden layer	Training Technique	# solutions	$E_{testing}$ (degrees)
DTFs in median plane	8	*BGD*	10	9.31±2.72
DTFs in median plane	8	*LM*	10	12.35±4.37

Table 4.3. Testing dataset RMS error with 8 neurons in the hidden layer and different training techniques for a NN used to estimate the elevation angle in the median plane.

When an ANN of 8 neurons in the hidden layer used to estimate the azimuth angle in the horizontal plane, the testing error values increased as expected. The used training data were at 0° elevation and azimuth angles of -75°, -45°, -15°, 7.5°, 22.5°, 37.5°, 52.5°, 67.5° and 75°, and tested on locations of 0° elevation and azimuth angles -75°, -60°, -45°, -30°, -15°, 0°, 7.5°, 15°, 22.5°, 30°, 37.5°, 45°, 52.5°, 60°, 67.5° and 75°. The results are summarized in Table (4.4).

Testing dataset	# neurons in hidden layer	Training Technique	# solutions	$E_{testing}$ (degrees)
DTFs in horizontal plane	8	*BGD*	10	20.95±7.59
DTFs in horizontal plane	8	*LM*	10	22.22 ±5.21

Table 4.4. Testing dataset RMS error with 8 neurons in the hidden layer and different training techniques for a NN used to estimate the azimuth angle in the horizontal plane.

4.4.3 Test 3. Spectra in the frontal field

In this test, the used DTFs are from (Rice et al., 1992) study. The used data are DTFs from the frontal field which range over azimuths between -75° and 75° and elevations from -30° to 90°, spaced every 15° in azimuth and 7.5° in elevation. The locations of the DTFs used for NN training are shown in Appendix A. The rest of the locations in the frontal field except the DTF at location of elevation 90° are used to test the NN and to evaluate the response of the network in estimating the elevation and the azimuth angles of the sound source. So, in this part, none of the testing data are used in the training of the network. The networks are evaluated according to their ability to predict the location that corresponds to the given DTF as an azimuth and an elevation angle of the sound source. The total number of the training data is 211 DTFs and the networks are tested on the DTFs of 43 locations distributed in the frontal space. The results are summarized in Table (4.5). The lowest testing error for elevation angle estimation is (mean±std) 5.15°±0.91° for the network architecture of 32 neurons in the hidden layer and using the *Levenbergh-Marquardt* training technique. A network with 16 neurons in the hidden layer which was trained by *batch gradient descent* showed the lowest error with a value of (15.36°±0.91°) in estimating the azimuth angle for the testing data. On average, the model gets the result correct. The largest errors appear to be in the contralateral side where the SNR is poor. This might be expected. On the ipsilateral side, the NN does quite a good job.

Testing dataset	# neurons in hidden layer	Training Technique	# sol's	Elevation angle prediction RMS error (degrees)		Azimuth angle prediction RMS error (degrees)	
				$E_{training}$	$E_{testing}$	$E_{training}$	$E_{testing}$
DTFs in frontal field	4	*BGD*	10	7.20±1.05	8.19±1.34	17.71±1.12	15.39±0.96
DTFs in frontal field	4	*LM*	10	0.72±1.14	27.66±14.45	10.89±0.00	78.41±69.15
DTFs in frontal field	8	*BGD*	10	5.86±0.44	7.29±0.66	16.14±0.73	15.56±1.01
DTFs in frontal field	8	*LM*	10	0.34±1.08	13.78 ±4.73	14.04±6.64	30.06±5.45
DTFs in frontal field	16	*BGD*	10	4.87±0.30	6.49±0.73	14.93±0.33	15.36±0.91
DTFs in frontal field	16	*LM*	10	0.00± 0.00	5.93±0.82	10.89±0.00	26.66±2.89
DTFs in frontal field	32	*BGD*	10	3.85±0.50	5.76±0.53	13.80±0.28	16.14±0.91
DTFs in frontal field	32	*LM*	10	0.00±0.00	5.15±0.59	10.89±0.00	24.81±1.76
DTFs in frontal field	64	*BGD*	10	3.08±0.24	6.25±0.56	13.03±0.20	17.62±0.88
DTFs in frontal field	64	*LM*	10	0.57±1.39	5.24±1.16	13.52±6.23	23.95±2.12
DTFs in frontal field	128	*BGD*	10	2.57±0.14	7.98±0.65	12.44±0.18	19.21±0.43
DTFs in frontal field	128	*LM*	10	NA*	NA*	NA*	NA*

Table 4.5. Training and testing datasets' errors with different number of neurons in the hidden layer and different training techniques for a NN used to estimate the direction of the sound source (i.e. the azimuth and the elevation angle of the sound source). *result not available because computer did not have enough memory.

4.4.4 Test 4. Measured and modeled DTFs comparison

Now that we have established the general applicability of the NN model of DTF recognition, in this part, NNs will be used to evaluate modeling techniques for measured DTFs. NNs are trained with the entire set of measured DTFs in the frontal field from Rice et al. (1992) and then the modeled DTFs were simulated on these trained networks to evaluate the used modeling techniques. DTFs modeling using all-pole and all-zero techniques will be evaluated here. This test can facilitate the evaluation of the models used for DTFs fitting by checking how the neural network can differentiate the measured from the modeled DTFs and give the average error of sound source localization for the modeled DTFs. This test was exclusive to elevation angles because we are emphasizing on spectral cues in this analysis and elevation is represented exclusively by spectral cues. To the best of my knowledge, this kind of evaluation for the modeled DTFs has not been

performed before. A network with 32 neurons in the hidden layer with *Levenbergh-Marquardt* training algorithm was used in this test because it has shown very good results in the previous tests. The network has been trained over all measured DTFs in the frontal field to give an error of $(5x10^{-5})°$. The regression analysis for the trained data is shown in Figure 4.3(a). The resulting RMS errors of the predicted elevation angles using modeled DTFs simulated on the trained network are summarized in Table (4.6). Figure 4.3(b) shows plots of the RMS error between the predicted and the actual elevation angles vs. the order of all-pole and all zero models used to model the DTFs. Figures 4.3(c)-4.3(f) show the regression analysis of the predicted elevation angles using the NN from all-pole and all-zero modeled DTFs and the exact targets of the corresponding elevation angles for the model orders 64 and 32.

Testing dataset	Elevation angle prediction RMS error, $E_{testing}$ (degrees)	
	All-pole model	All-zero model
Modeled DTFs (order=64)	2.55	0.86
Modeled DTFs (order=46)	4.92	2.44
Modeled DTFs (order=32)	5.15	5.29
Modeled DTFs (order=25)	8.94	7.95
Modeled DTFs (order=20)	14.47	20.16
Modeled DTFs (order=16)	21.04	25.83
Modeled DTFs (order=10)	31.67	33.53
Modeled DTFs (order=6)	48.73	75.81

Table 4.6. RMS error values for elevation angle prediction from modeled DTFs simulated on a NN trained by measured DTFs in the frontal field.

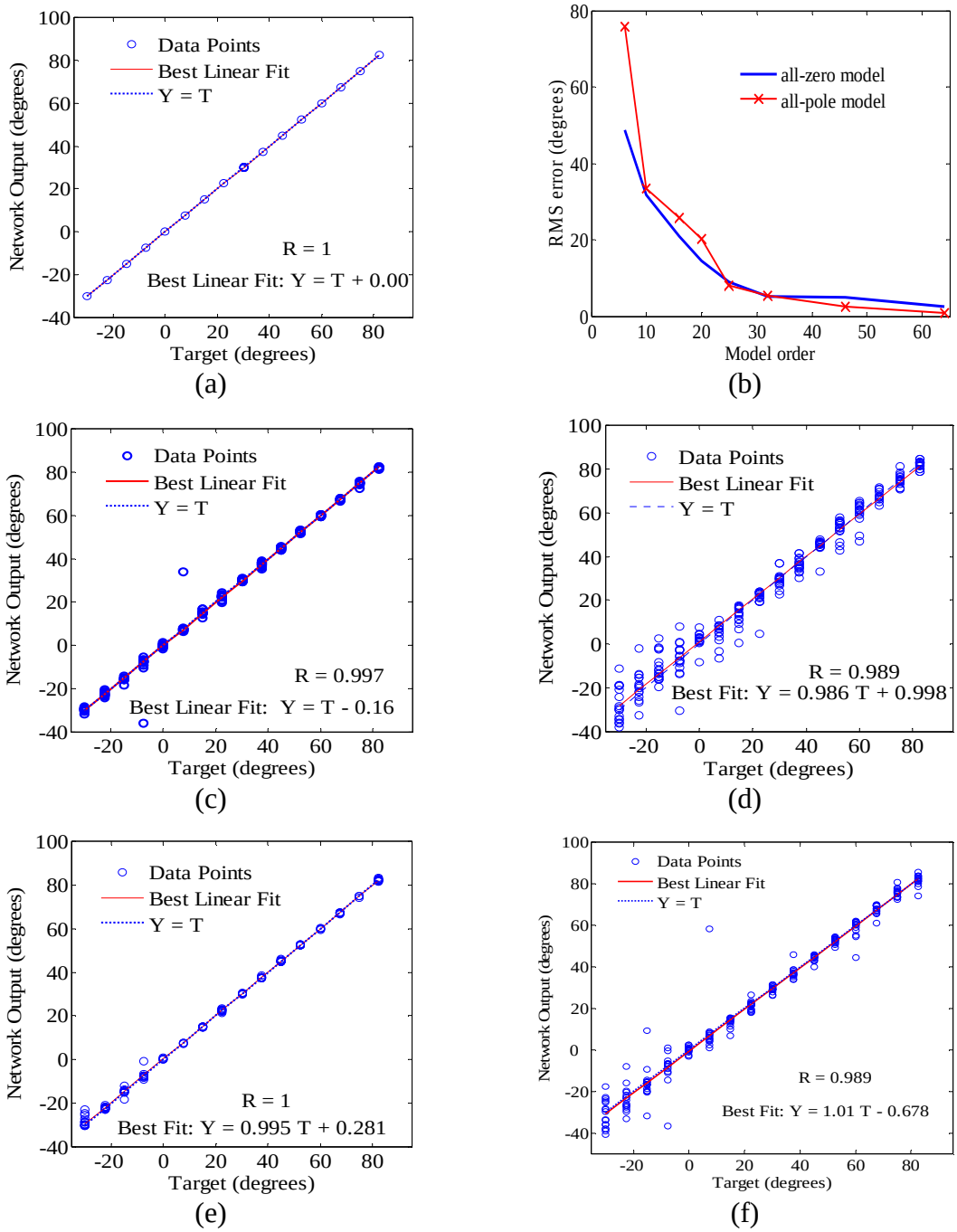


Figure 4.3. (a) Regression analysis for the simulated training data by NN versus the actual targets of elevation angles in the frontal field. (b) plots of the RMS error between the predicted and the actual elevation angles vs. the order of all-pole and all zero models used to model the DTFs. Regression analysis for the exact targets of the elevation angles in the frontal field with predicted elevation angles using the NN from DTFs modeled by (c) all-pole (order 64), (d) all-pole (order 32), (e) all-zero (order 64), (f) all-zero (order 32).

4.5 Conclusions

In general, NN showed much better performance in predicting elevation angles than azimuth angles using the spectral features of HRTFs, even when same spatial resolution is used in both planes. A network with more than one neuron in the hidden layer to estimate the elevation angle in the median plane only showed slightly poorer performance than that with one neuron when BGD training technique is used. A feed-forward back-propagation NN with 32 neurons showed a very good performance in predicting the elevation angle of the sound source from given HRTFs in the frontal plane with an average RMS error of 5.15° according to a randomly chosen 43 locations in the frontal field used in test 3. The trained network showed a high correlation between actual and predicted elevations. On the other hand, the performance was not as good in azimuth angle prediction where the average RMS error with the same network architecture was 16.14°. Furthermore, the neural networks can be a useful tool to evaluate modeling techniques for measured DTFs. According to this evaluation, high orders of all-zero models outperforms the ones of all-pole models with the same orders in modeling the measured DTFs by giving lower RMS error values in elevation angle prediction and better correlation with real elevation angles.

Chapter 5

RESULTS AND DISCUSSION II:

Systematic Movements of Poles/Zeros with Sound Source Direction

5.1 Poles/Zeros Movement Effect on DTF Spectrum

To simplify the idea of the relation that will be used here between the locations of the poles/zeros on the pole/zero plot and the main features (peaks and notches) in the transfer function that those poles/zeros resemble, an example of a pair of zeros and the effect of their movement on the transfer function is examined. Figure 5.1(b) shows a spectral notch created by the location of the pair of zeros in Figure 5.1(a) assuming a sampling frequency of 100 KHz. A plot of the notch frequency (frequency at the minimum point of the notch) versus the polar angle (β) (Figure 5.1(a)) for the zero in the upper half of the unit circle is shown in Figure 5.1(c). A similar behavior is found when two pairs of poles are moved systematically. Moving the two poles in the same way as the zeros above creates a systematic moving notch between the two peaks that these two pairs of poles creates. Thus, where the data are modeled using zeros or poles, the systematic movement of the first notch frequency component of the DTF spectra can be modeled in the way shown in Figure 5.1.

Systematic movements of some poles and zeros, but not others, have been noticed as elevation or azimuth angle changes in the modeled DTFs. The systematic movements in

poles and zeros in pole/zero models are not always simple to describe even with this systematic change in DTF spectral features with the change in elevation or azimuth. Also, at certain locations or ranges these poles and zeros show systematic movements with the increase in the elevation for example by moving counter clockwise but high order relations are needed to describe these systematic movements. In this study, focus is given to the systematic movements of poles and zeros of the modeled DTFs in the frontal field.

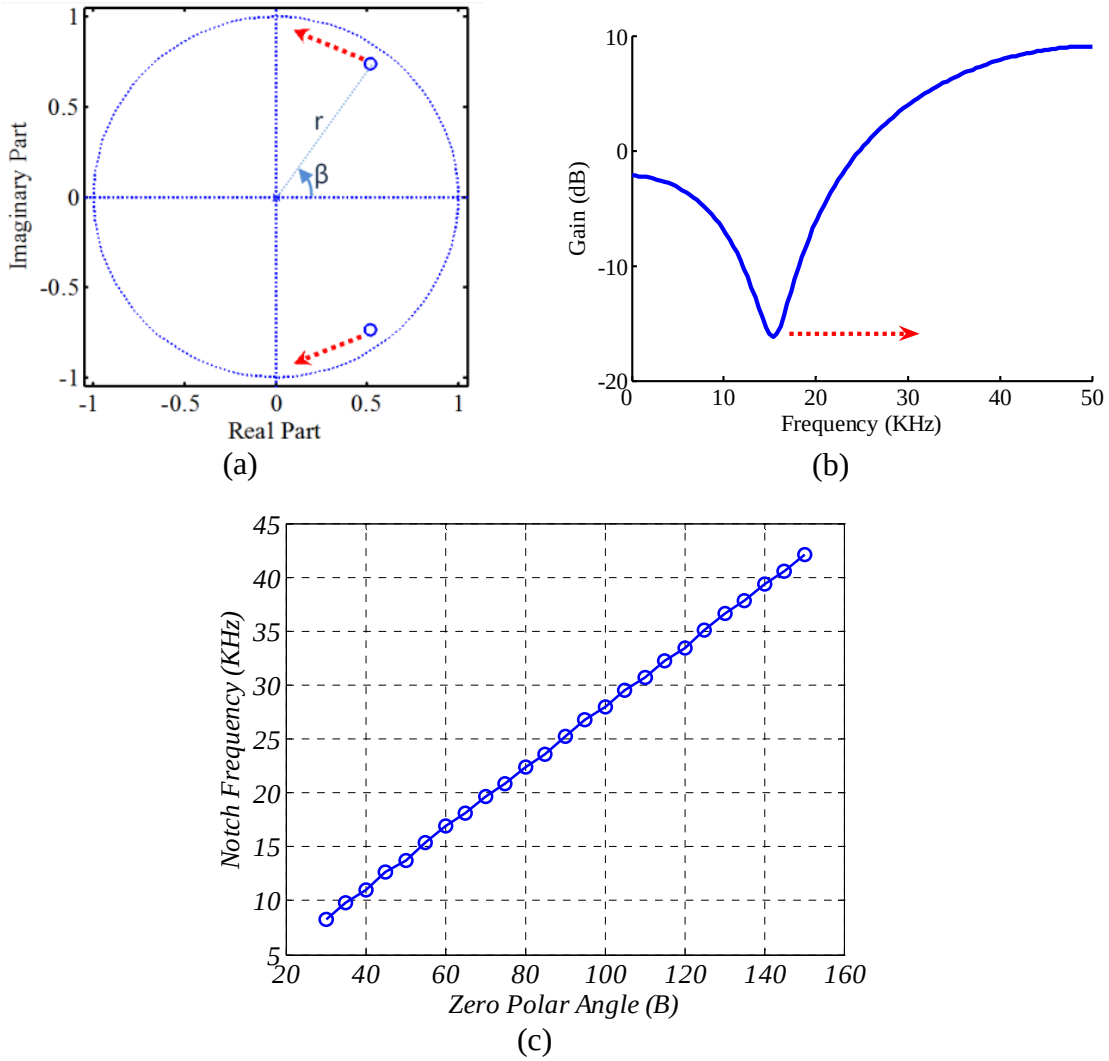


Figure 5.1. Demonstration of the movement of a spectral notch (b) with the change in the location of a pair of zeros given by the polar angle β (b). (c) A plot of the "notch" frequency as a function of the polar angle (β) of the moving zero in (a).

5.2 Poles/Zeros Systematic Movement in Human DTFs

Because of the low spatial resolution of SDO dataset (18° step in the vertical plane and 15° in the horizontal plane), SOW human dataset, which has 10° step in both direction, will be used in this part of the study. Since the systematic movement of poles and zeros accompany the systematic changes in DTFs, it is useful to pay attention to the range that one of the important features in DTFs that show systematic movement, the first notch, is noticed clearly. In SOW dataset, FN exhibits systematic behavior in the frontal field in the range between azimuth angle of 130° and -60° and elevation angles of -50° and 50° for the right ear. In locations outside this range, the systematic movement of the FN was not clear and sometimes several notches are noticed in the region of the FN that merge to one notch at some location. The two notches in the DTFs of elevations between -50° and 20° with fixed azimuth angle of 160° is an example of those cases. Where these two notches frequencies increase systematically with the increase in the elevation till they merge in one notch at location (az=160°, el=30°). Groups of poles in all-pole model and zeros in all-zero model have shown systematic movements with the change in azimuth and elevation angles within the same range that the systematic FN movement has been noticed for these datasets. Some other systematic movements of notches as well as poles and zeros in all-pole and all-zero models outside those ranges that diminish at certain locations are also noticed. Because of that, those locations are not included in the ranges specified above as clear systematic movement of poles and zeros with the change in sound source direction.

Figure 5.2 shows the smoothed measured DTFs of the SOW data for the right ear at locations of 0° azimuth and elevations from -50° (bottom in the plot) to +40° (top in the plot) with a 10° step. The green dotted arrow shows the systematic tendency of the FN movement with the increase in the elevation of the sound source. Figure 5.3 shows the all-pole modeled DTFs using 25 poles for the same locations of Figure 5.2. As shown in Figure (5.3), spectral notches and peaks, including the main first notch, are accurately modeled. Since it will not be practical to track the movements of all poles/zeros of all locations in one plot at the same time, the poles/zeros movement tracking will be performed within limited ranges. Figure 5.4 shows the poles of the all-pole model of order 25 in the upper half of the unit circle for modeled DTFs at locations of fixed 0° azimuth and elevation angles of 10°, 20°, 30°, 40° and 50° which are given by colors, magenta, cyan, red, green and blue, respectively. The counter clockwise systematic movement of the poles located under the four groups (marked as G1, G2, G3 and G4) shown in Figure 5.4 is remarkable. Since the responses of all used models in this project are real-valued, the complex poles and zeros occur in conjugate symmetric pairs, i.e., if there is a complex pole (zero) at $p=p_o$, there is also a complex pole (zero) at $p=p_o^*$ (Hayes, 1998). Because of this, only the poles in all-pole models and zeros in all-zero models in the upper half of the z-plane unit-circle in addition to the real axis will be shown for the pole/zero plots.

Other systematic movements of other poles with the change in the elevation angle can be noticed but it is not as clear as it is in these four groups. Figures 5.5 and 5.6 show the

systematic movement of the first notch with the change in the azimuth angle at the locations of fixed 0° elevation and azimuth angles between -20° (bottom in the plots) and 50° (top in the plots) for measured and all-pole modeled of order 25 DTFs, respectively. The green dotted arrows in Figures 5.5 and 5.6 show the systematic tendency of the FN movement with the change in azimuth angle of the sound source. Figure 5.7 shows the poles of the all-pole model of order 25 in the upper half of the unit circle for modeled DTFs at locations of fixed 0° elevation and azimuth angles of -10°, 0°, 10°, 20° and 30° which are given by colors, magenta, cyan, red, green and blue, respectively. The counter clockwise systematic movement of some poles located under the four groups shown in Figure 5.7 is noticed.

By comparing the results with the change in location of both azimuths and elevation in Figures 5.2-5.7, it is clear that the systematic movement of the poles within each group is larger with the change in elevation compared to that with the change in azimuth. The change in FN as an important spectral feature that accompanies the change in elevation or azimuth can justify that. FN frequency increases from 6.80 KHz to 8.07 KHz with the change in the elevation angle from 10° to 50° with fixed azimuth at 0°, while the FN frequency changes from 6.41 KHz to 6.85 KHz when there is the same angular displacement in azimuth, from -10° to 30° with fixed elevation at 0°.

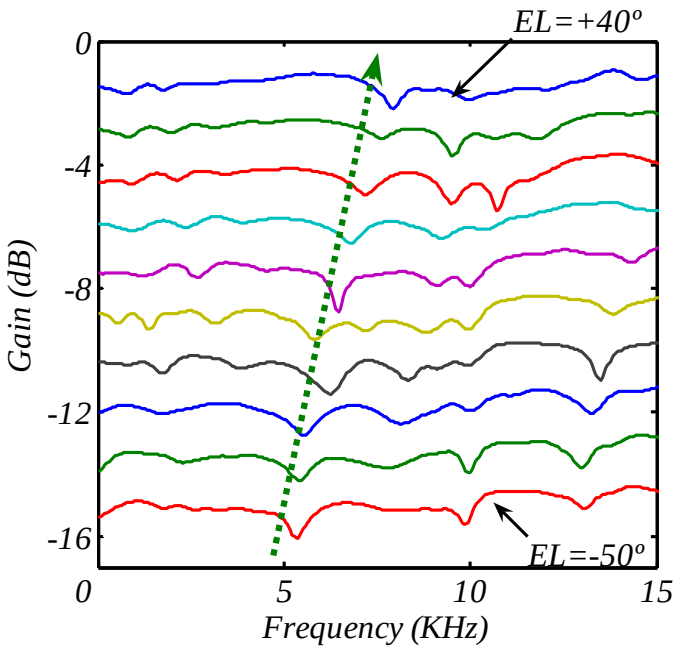


Figure 5.2. Smoothed measured DTFs of SOW data for the right ear at locations of 0° azimuth and elevations from -50° (bottom) to +40° (top) with a 10° step. The green dotted arrow shows the tendency of the FN movement with the change in the elevation angle.

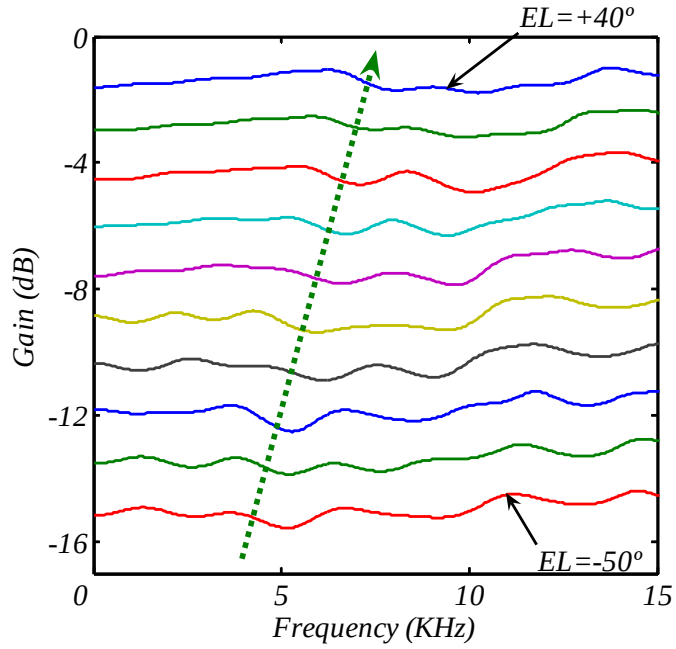


Figure 5.3. All-pole modeled DTFs (order 25) of SOW data for the right ear at locations of 0° azimuth and elevations from -50° (bottom) to +40° (top) with a 10° step. The green dotted arrow shows the tendency of the FN movement with the change in the elevation angle.

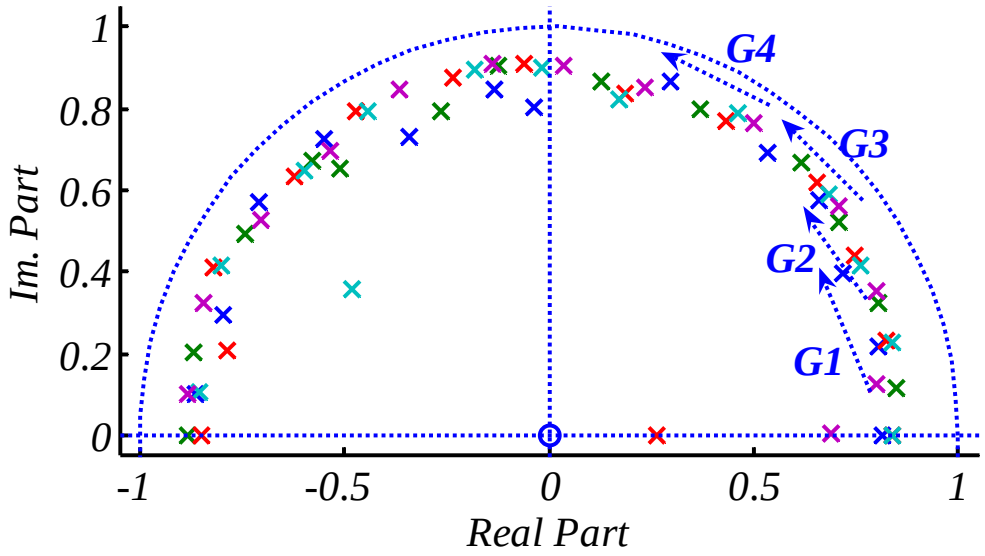


Figure 5.4. Pole/zero plots for DTFs of locations of 0° azimuth and elevations 50° (blue), 40° (green), 30° (red), 20° (cyan), 10° (magenta). Dotted arrows show the tendency of the systematic movement of the poles in groups 1-4. G1 to G4 represent the groups of poles that show systematic movements with the change in sound source elevation.

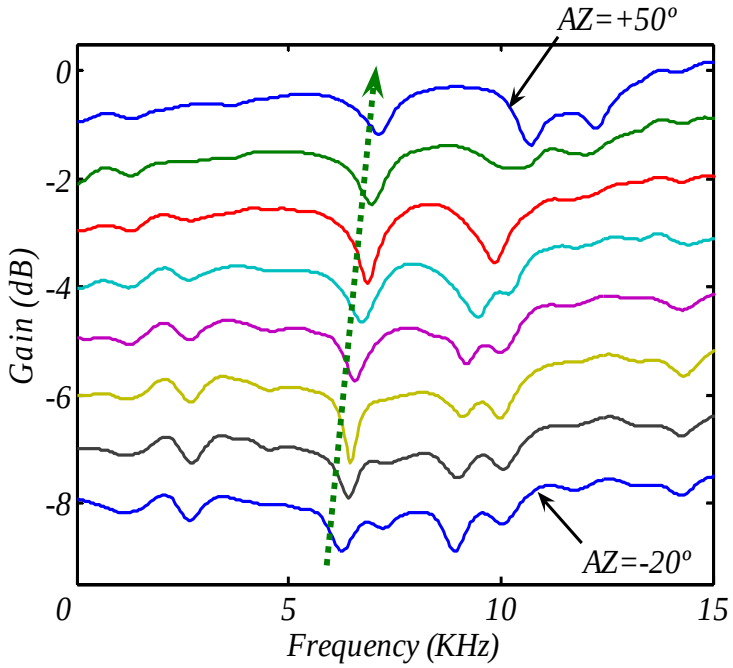


Figure 5.5. Smoothed measured DTFs of SOW data for the right ear at locations of 0° elevation and azimuths from 50° (top) to -20° (bottom) with a 10° step. The green dotted arrow shows the tendency of the FN movement with the change in the azimuth angle.

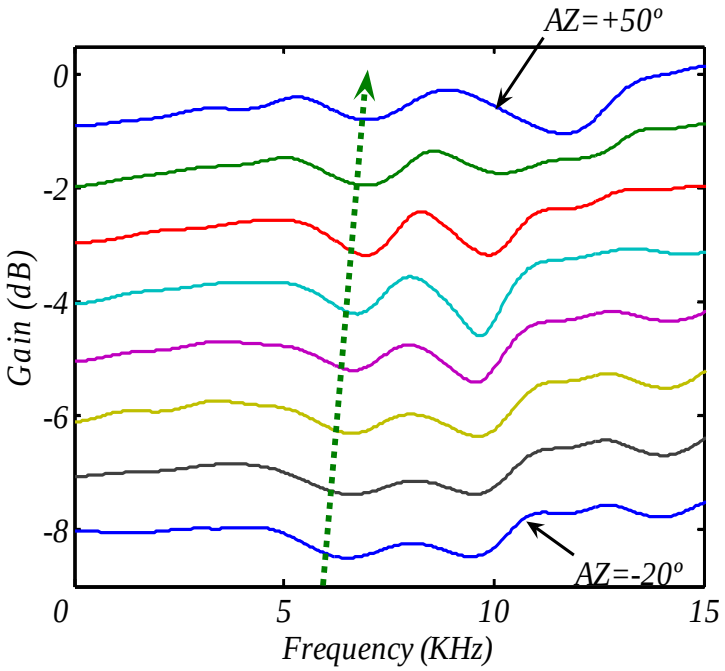


Figure 5.6. Modeled DTFs (using all-pole of order 25) of SOW data for the right ear at locations of 0° elevation and azimuths from 50° (top) to -20° (bottom) with a 10° step. The green dotted arrow shows the tendency of the FN movement with the change in the azimuth angle.

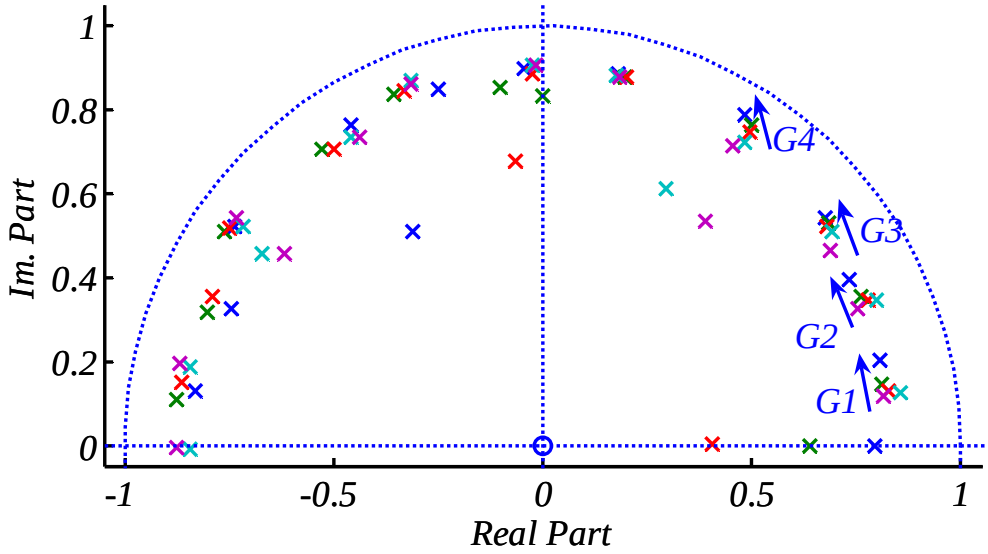


Figure 5.7. Pole/zero plots for DTFs of locations of 0° elevation and azimuths 30° (blue), 20° (green), 10° (red), 0° (cyan), -10° (magenta). Dotted arrows show the tendency of the systematic movement of the poles in groups 1-4.

To track the systematic movement of poles or zeros in all-pole or all-zero models, polynomials of different orders have been used to describe the change in poles/zeros locations with the change in elevation or azimuth angle within each group that shows these systematic movements. To evaluate the reconstructed DTFs from the estimated locations of poles/zeros, they have been compared to modeled and measured DTFs at the same locations. The two-dimensional polar coordinate system is used in building the relations of the changes in poles/zeros locations. To test the efficiency of reconstructing DTFs using parametric relations to describe the systematic movement of the poles in all-pole model or zeros in all-zero model, few tests will be performed.

5.2.1 Test#1: Systematic movements of poles in all-pole models with changes in the elevation of sound sources in the median plane

Objective: *"Extract relations between the locations of the poles in the groups that show systematic movements with the change in elevation and the change in the elevation angle (β) in the median plane, reconstruct DTFs using the extracted (fitted) relations and compare the reconstructed DTFs to the measured ones."*

Using right ear SOW dataset, the systematic movement of the poles with the change in the polar angle (β) in the polar coordinate system (see Figure 5.8) is described. A third-order polynomial has been used to describe the change in the polar angle (β) with the change in the elevation angles described in Figures 5.2-5.4 for the four groups of poles. The length of the vectors, **r1** and **r2** in the used polar coordinate system (Figure 5.8) are fixed and chosen to be equal to the average of the distances of the poles in each group

from the origin. (alternatively, the vector magnitudes could be an additional free parameter, but we did not test for this here.) This test is applied on locations (azimuth= 0° and elevations 10°, 20°, 30°, 40° and 50°). The pole/zero plots that show the locations of the poles in the four groups are shown in Figure 5.9. The estimated locations of the poles in the rebuilt models using third order polynomials are shown in Figure 5.10. The fitted third order polynomials used to find the location of the poles in the rebuilt models for poles groups, one, two, three and four are implemented in Equations 5.1, 5.2, 5.3, and 5.4, where the polar angle (β) and the elevation angle (θ) are given in radian. Those polynomials have been fitted to give the lowest mean squared error between the predicted and the observed polar angles. The first notch frequency versus the elevation angle is shown in Figure 5.11 for both, all-pole modeled DTFs and rebuilt DTFs.

$$\beta = 1.672\theta^3 - 2.291\theta^2 + 1.315\theta - 0.005 \quad (5.1)$$

$$\beta = 0.627\theta^3 - 0.836\theta^2 + 0.720\theta + 0.312 \quad (5.2)$$

$$\beta = 0.305\theta^3 - 0.209\theta^2 + 0.283\theta + 0.623 \quad (5.3)$$

$$\beta = 0.884\theta^3 - 1.009\theta^2 + 0.577\theta + 0.918 \quad (5.4)$$

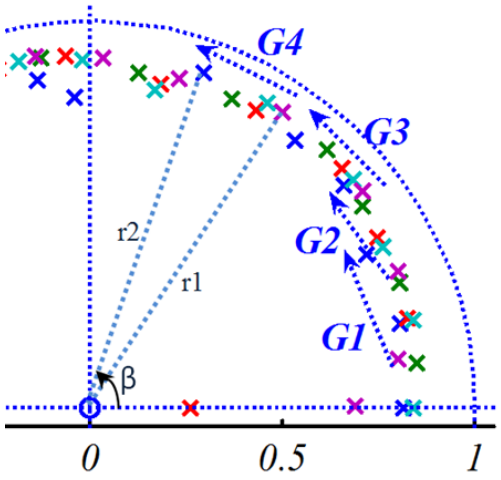


Figure 5.8. Demonstration of the vectors and the polar angle (β) that is used to extract relations between poles/zeros movement with the change in the sound source elevation/azimuth.

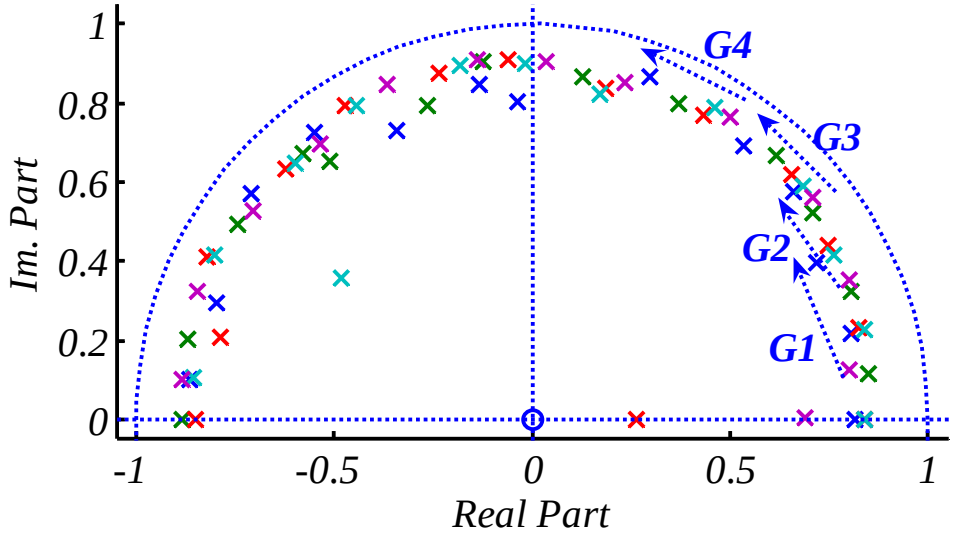


Figure 5.9. Pole/zero plots for DTFs of locations of 0° azimuth and elevations 50° (blue), 40° (green), 30° (red), 20° (cyan), 10° (magenta). Dotted arrows show the tendency of the systematic movement of the poles in groups 1-4.

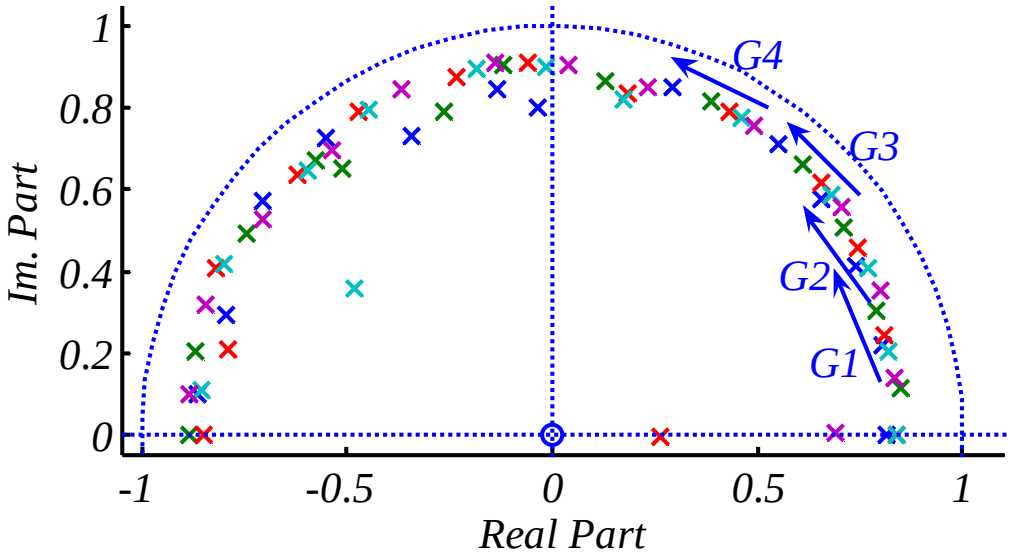


Figure 5.10. Pole/zero plots for DTFs of locations of 0° azimuth and elevations 50° (blue), 40° (green), 30° (red), 20° (cyan), 10° (magenta) where the locations of the poles in the four groups are rebuilt using 3rd order polynomials.

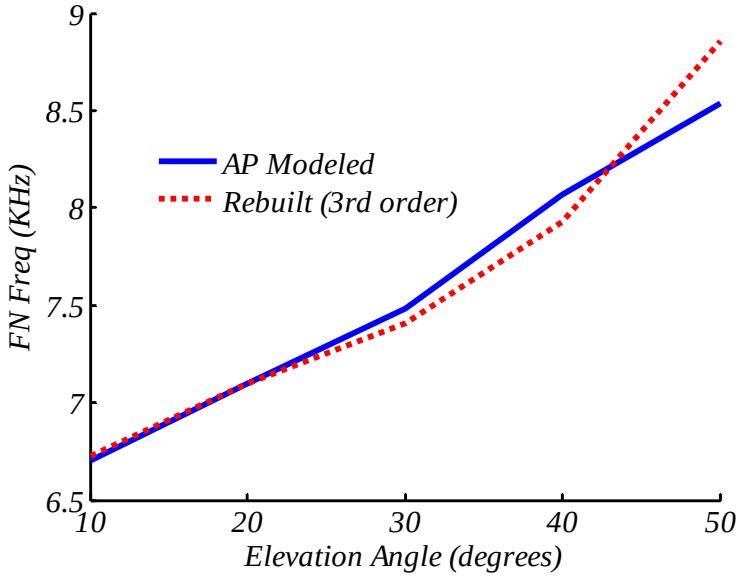


Figure 5.11. The first notch frequency versus the elevation angles for all-pole modeled DTFs and reconstructed DTFs from relations in Equations (5.1-5.4).

5.3 Cat Data

Rice and his colleagues (1992) reported that the region over which the FN exhibits systematic behavior spans a region in front of the cat located roughly between -45 ° AZ and 60° AZ and between -30° EL and 45° EL. *Cat 1107* dataset has been used to track the systematic movement of poles/zeros in that range of the frontal field. Figures 5.12 and 5.13 show the measured and the modeled DTFs, respectively, for the right ear of dataset provided kindly from (Rice et al., 1992, *cat 1107*) at fixed 0° azimuth and elevations from -30º (bottom in the plot) to +90º (top in the plot) with 7.5º step using all-pole model with order equals 25. To show these spectra in a suitable way, a 20 dB shift between the consecutive elevation angles' spectra was inserted. The systematic changes of some spectral notches are traced by dotted arrows in Figures 5.12 and 5.13. Plots of the FN frequency extracted from measured HRTFs, measured, smoothed and modeled DTFs versus the elevation angle are shown in Figure 5.14.

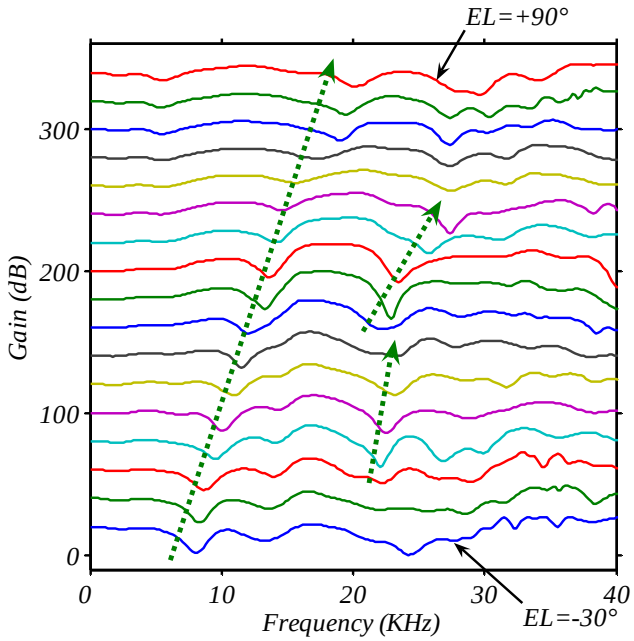


Figure 5.12. Smoothed measured DTFs of *cat 1107* dataset for the right ear at locations of 0° azimuth and elevations from 90° (top) to -30° (bottom) with a 7.5° step. The dotted arrows show the tendency of some prominent notches movements with the change in the elevation angle.

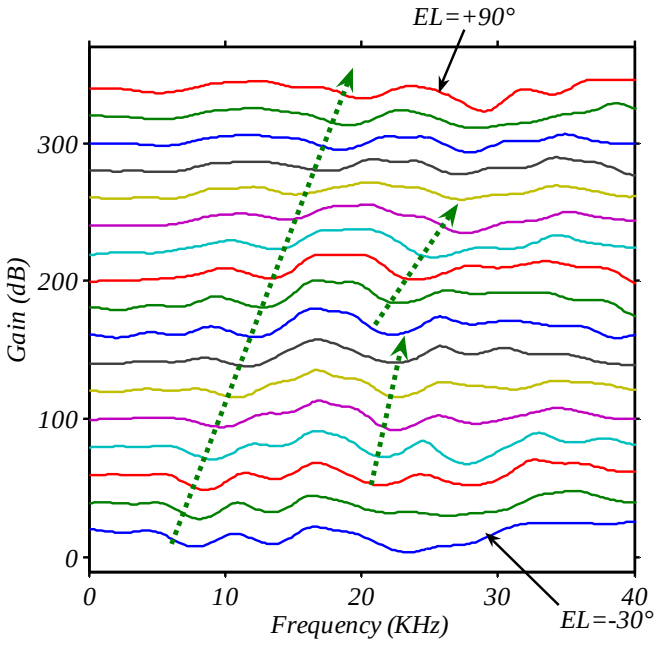


Figure 5.13. All-pole modeled DTFs of *cat 1107* dataset for the right ear at locations of 0° azimuth and elevations from 90° (top) to -30° (bottom) with a 7.5° step. The dotted arrows show the tendency of some prominent notches movements with the change in the elevation angle.

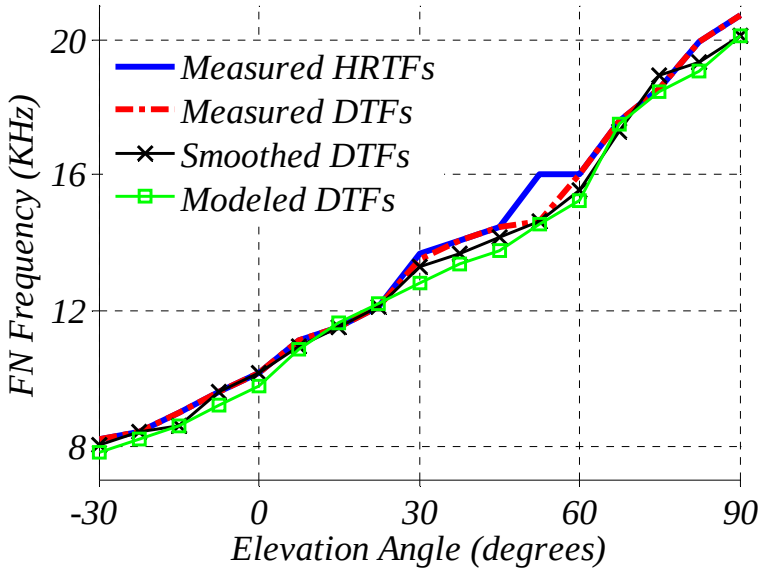


Figure 5.14. Plots of the FN frequency extracted from measured HRTFs, measured, smoothed and modeled DTFs versus the elevation angle for *cat 1107* dataset.

Since the dataset of UCHSC provided by Prof Daniel Tollin (Tollin and Koka 2009), one of the advisors of this study, has higher resolution compared to *cat 1107* dataset it will be used for detailed tests on poles/zeros systematic movement with the change in elevation or azimuth angles. Figures 5.15 and 5.16 show the measured and the modeled DTFs of the right ear at 0° Az and elevations from -45º (bottom in the plot) to +45º (top in the plot) with 3º step using all-pole model. The root-mean-squared (RMS) error between measured and all-pole modeled DTFs for these locations has a mean value of 0.682 dB with standard deviation of 0.251 dB and a maximum value of 1.444dB. The FN frequency increases from 7.74 KHz to 15.24 KHz as elevation angle increases from -45° to +45° and this systematic movement is accurately preserved in both the all-pole and the all-zero models as shown in Figure 5.17. The mean absolute error for the difference in FN frequency between measured and all-pole modeled DTFs has a mean value of 0.101 KHz with a standard deviation of 0.225 KHz and a maximum value of 0.475 KHz.

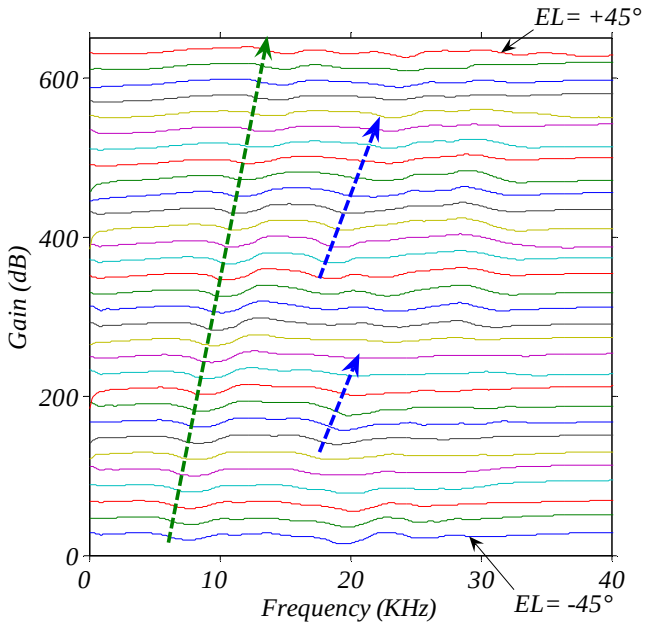


Figure 5.15. Smoothed measured DTFs of UCHSC dataset for the right ear at locations of 0° azimuth and elevations from 45° (top) to -45° (bottom) with a 3° step. The dotted arrows show the tendency of some prominent notches movements with the change in the elevation angle.

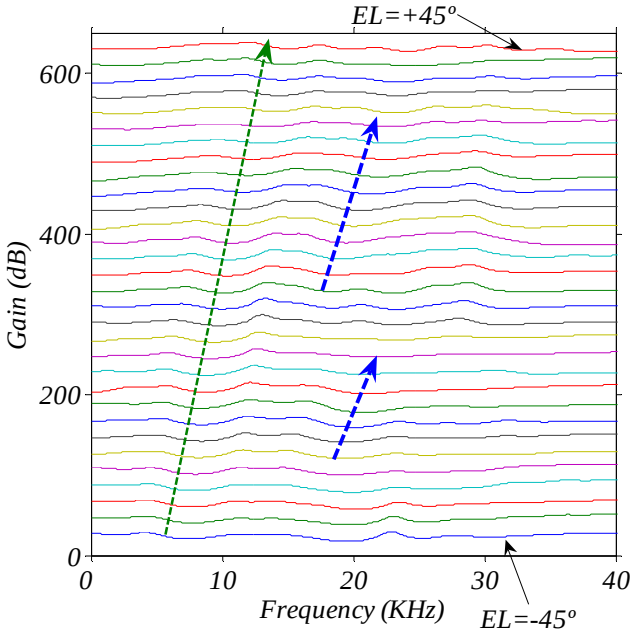


Figure 5.16. All-pole modeled DTFs of UCHSC dataset for the right ear at locations of 0° azimuth and elevations from 45° (top) to -45° (bottom) with a 3° step. The dotted arrows show the tendency of some prominent notches movements with the change in the elevation angle.

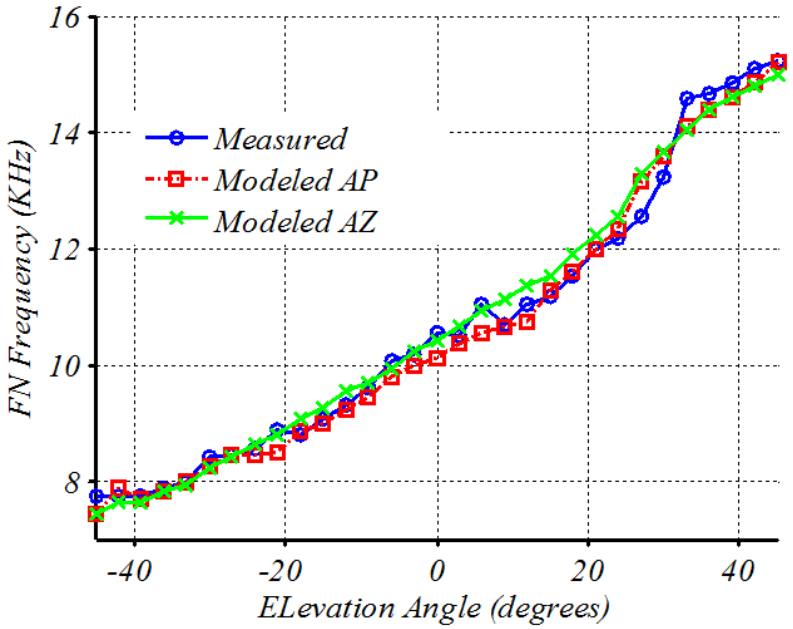


Figure 5.17. Plots of the FN frequency extracted from measured, all-pole modeled and all-zero modeled DTFs versus the elevation angle for UCHSC dataset.

5.3.1 Test#2: Systematic movements of poles in all-pole model with change in elevation of sound source in the median plane using linear relations of pole movements

Objective: *"Extract linear relations between the locations of the poles in the groups that show systematic movements with the change in elevation and the change in the elevation angle (β) in the median plane, reconstruct DTFs using the extracted (fitted) relations and compare the reconstructed DTFs to the measured ones."*

Figure 5.18 shows a pole/zero plot for all-Pole modeled DTFs at fixed azimuth angle 0º and elevation angles from +3º to +18º with 3º steps. The poles in the four marked groups show systematic movements with the increase in the elevation angle. Linear relations (Equations 5.5-5.8) have been used to describe the change in the poles' polar angle (β)

locations with the increase in the elevation angle from +3º to +18° for groups 1, 2, 3 and 4, respectively. The locations of the poles in the rebuilt models using Equations (5.5-5.8) are shown in Figure 5.19. The modeled and the reconstructed DTFs of locations at fixed azimuth angle 0º and elevation angles from +3º to +18º with 3º steps are shown in Figures 5.20 and 5.21. Figure 5.22 shows the FN frequency for modeled (blue), when the four groups of poles only moves and the rest are fixed at the poles of elevation of 3°(dotted red), and from DTFs rebuilt by the linear relations (black) at elevations 3º, 6º, 9º, 12º, 15º, and 18º.

$$\beta = 0.283\theta + 0.472 \tag{5.5}$$

$$\beta = 0.381\theta + 0.846 \tag{5.6}$$

$$\beta = 0.522\theta + 0.983 \tag{5.7}$$

$$\beta = 0.875\theta + 1.320 \tag{5.8}$$

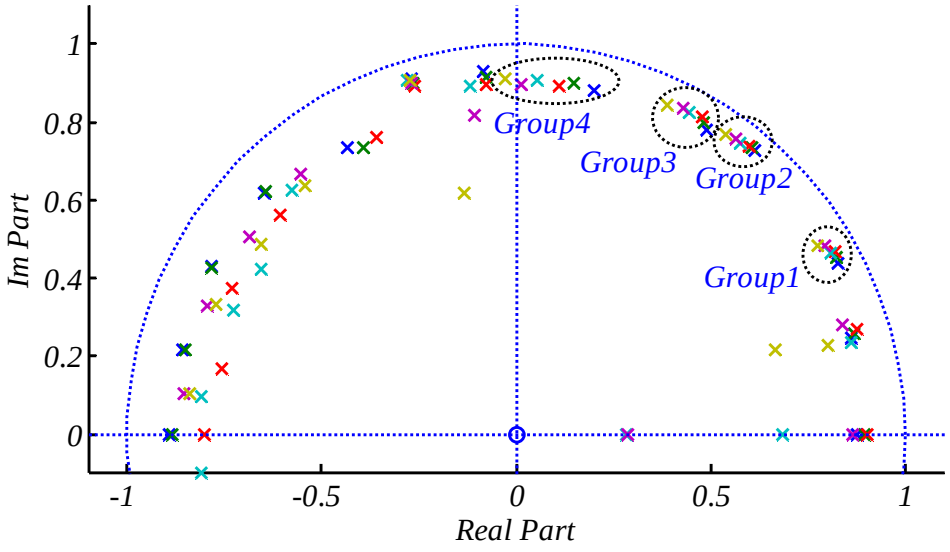


Figure 5.18. Pole/zero plots for DTFs of locations of 0° azimuth and elevations 3° (blue), 6° (green), 9° (red), 12° (cyan), 15° (magenta), 18° (yellow). Poles that show systematic movements are marked by Groups 1-4.

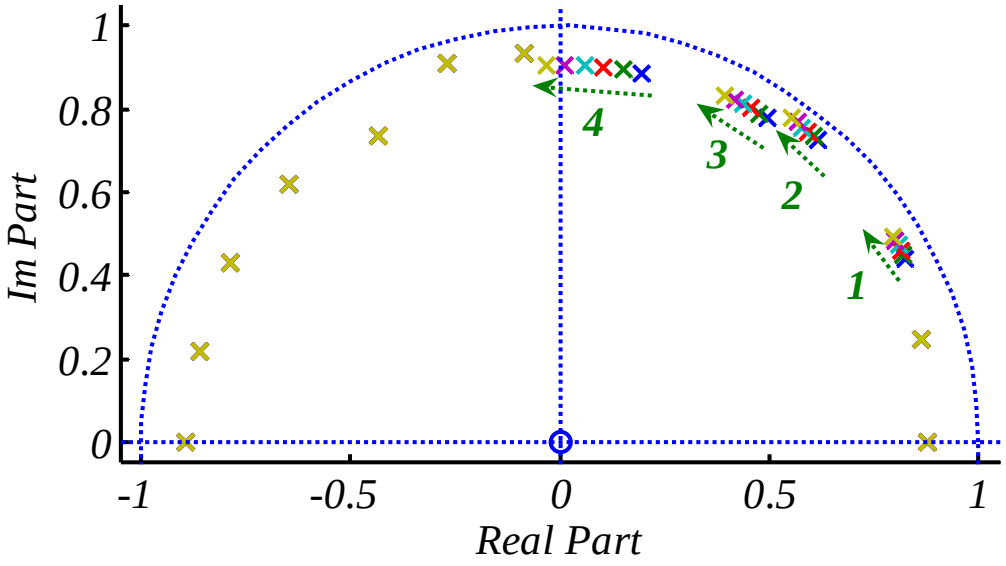


Figure 5.19. Pole/zero plots for DTFs of locations of 0° azimuth and elevations 3° (blue), 6° (green), 9° (red), 12° (cyan), 15° (magenta), 18° (yellow). Locations of the poles within the four groups are estimated using the fitted first order relations.

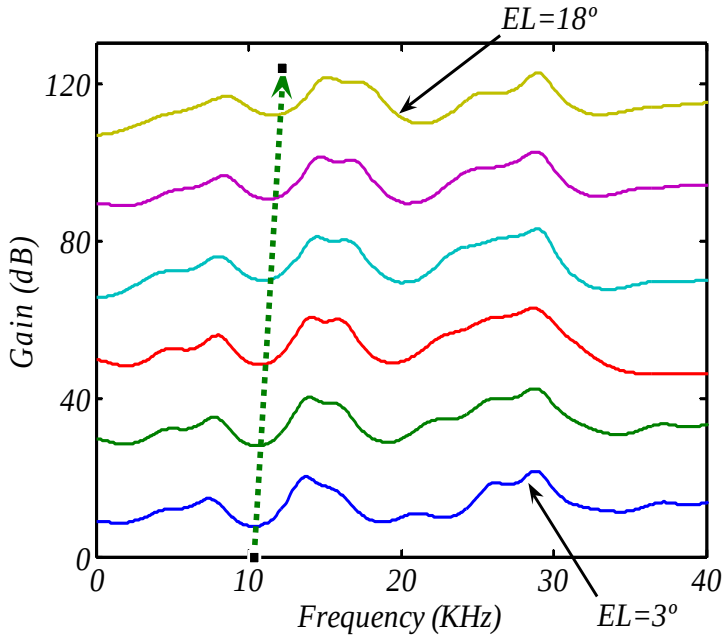


Figure 5.20. All-pole modeled DTFs of UCHSC dataset for the right ear at locations of 0° azimuth and elevations from 18° (top) to 3° (bottom) with a 3° step. The dotted arrow shows the tendency of the FN movement with the change in the elevation angle.

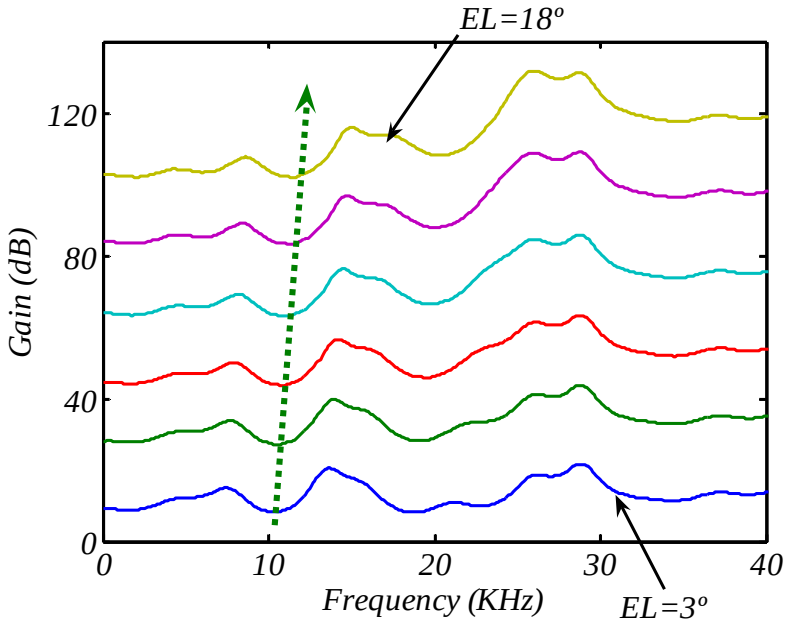


Figure 5.21. Reconstructed DTFs using linear relations of UCHSC dataset for the right ear at locations of 0° azimuth and elevations from 18° (top) to 3° (bottom) with a 3° step. The dotted arrow shows the tendency of the FN movement with the change in the elevation angle.

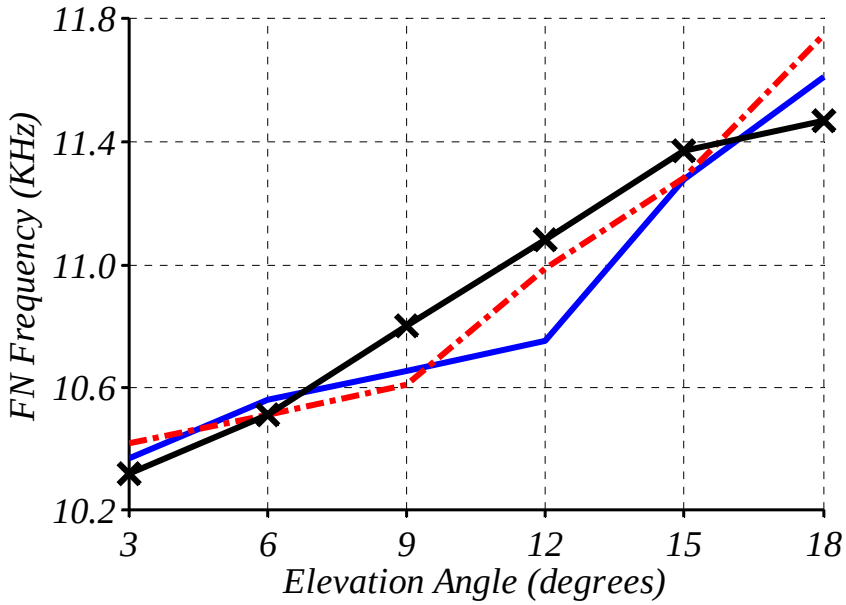


Figure 5.22. Plots of the FN frequency extracted from all-pole modeled DTFs (blue), DTFs reconstructed by moving the poles in the four groups (red) and DTFs reconstructed using linear relations for the poles in the four groups (black) versus the elevation angle using UCHSC dataset.

5.3.2 Test #3: Systematic movements of poles in all-pole model with change in elevation of sound source using 1st and 3rd order polynomials

Objective: *"Compare the efficiency in the preservation of the FN frequency in the reconstructed DTFs using first order and third order polynomials"*

In this test, the poles in the four groups for locations at 0° azimuth and elevations at 0°, 6°, 12°, 18°, and 24° were used to build mathematical relations with the change in the elevation angle and then use that relation to build DTFs at 0° azimuth and elevations 3°, 9°, 15°, 21° as follows. First order (linear) and third order polynomials are tested in building the relations between the elevation angles (θ) and the poles polar angle (β) in the upper half of the unit circle of the z-plane. The length of the vector in the polar coordinate system (the distance of the estimated pole location from the origin) is taken as the average of the lengths of the two neighboring poles vectors.

FN frequency extracted from the modeled DTFs and rebuilt DTFs are shown in Figure 5.23 with (mean±std) for the absolute error in FN frequency (153.37°±81.40°) Hz and (96.375°±55.40°) Hz, for first and third order polynomials, respectively. Both polynomial orders give satisfactory prediction for the locations of the poles in the reconstructed DTFs in regards of the FN preservation, but as expected, the 3rd order polynomial gives better reconstruction.

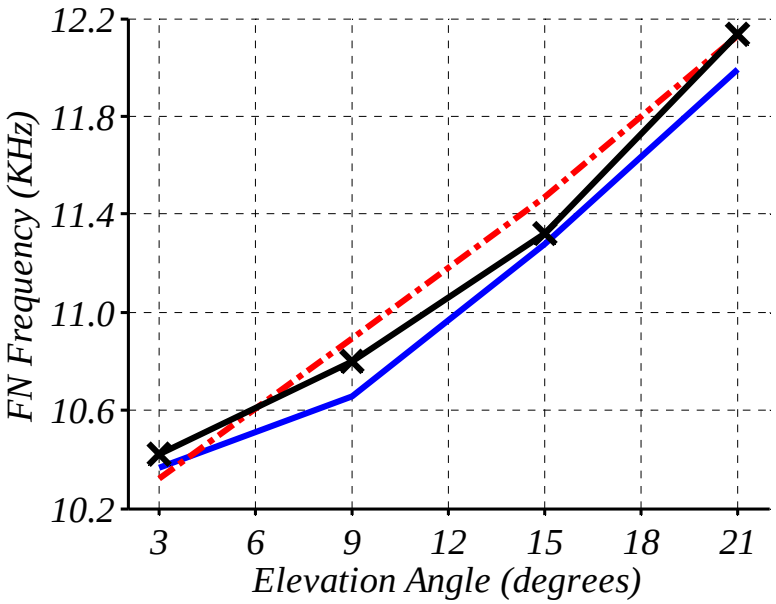


Figure 5.23. Plots of the FN frequency extracted from all-pole modeled DTFs (blue), reconstructed DTFs using linear relations for the poles in the four groups (red) and reconstructed DTFs using 3rd order relations for the poles in the four groups (black) versus the elevation angle.

5.3.3 Test #4: Systematic movements of zeros in all-zero model with change in elevation of sound source in the median plane

Objective: **"***Extract relations between the locations of the zeros in the groups that show systematic movements with the change in elevation and the change in the elevation angle in the median plane for DTFs at certain locations. Using those extracted relations, build DTFs at other locations in the median plane and compare the reconstructed DTFs to the measured ones."*

DTFs at fixed 0° azimuth and elevations of -45°, -39°, -33°, -27° and -21° are used to reconstruct DTFs at 0° azimuth and elevations at -42°, -36°, -30° and -24° (i.e. at locations of finer resolution of the measured DTFs). The reconstructed DTFs were evaluated by comparing them to the measured DTFs at those locations. Figure 5.24

shows the all-zero modeled DTFs of the right ear at 0° Az and elevations from -45º (bottom in the plot) to +45º (top in the plot) with 3º step using the all-zero model. The pole/zero plots for the modeled DTFs at locations of fixed 0° azimuth and elevations of -45°, -39°, -33°, -27° and -21° are shown in Figure 5.25.

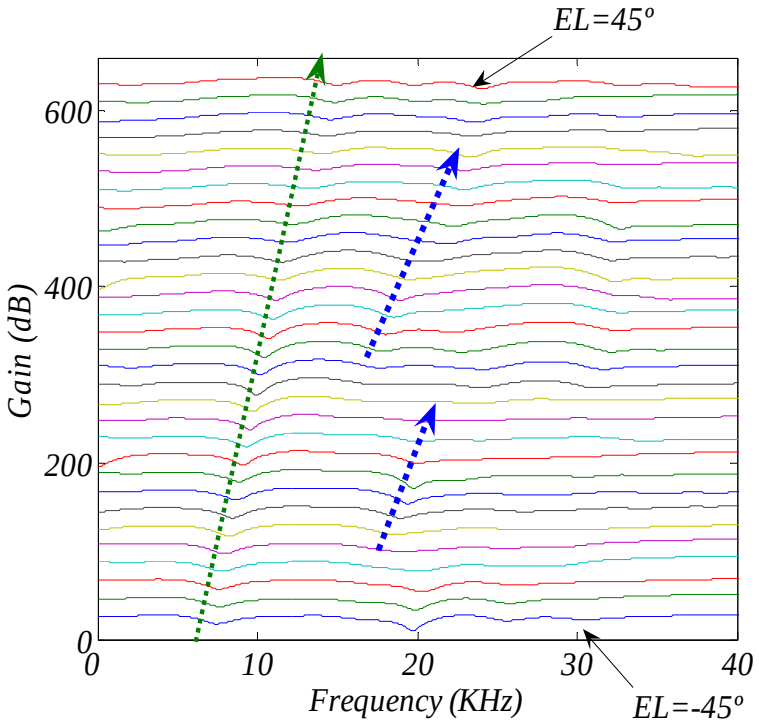


Figure 5.24. All-zero modeled DTFs of UCHSC dataset for the right ear at locations of 0° azimuth and elevations from 45° (top) to -45° (bottom) with a 3° step. The dotted arrows show the tendency of some prominent notches movements with the change in the elevation angle.

DTFs at locations of 0° azimuth and elevations of -42°, -36°, -30° and -24° have been reconstructed as follows:

1. The polar angle (β) for the locations of the zero groups that show systematic movements with change in elevation angle (five groups of zeros pairs in this test as shown in Figure 5.25) for each DTF is calculated using the extracted relations depending

on the group they are located within. Those equations are extracted using the zeros movements of the all-zero modeled DTFs of locations at 0° azimuth and elevations of -45°, -39°, -33°, -27° and -21°.

2. The distance of each zero in step number 1 from the origin is calculated as the average of the distances of the zeros locations in the same group for the locations directly above and directly below the reconstructed DTF location.

3. The rest of the zeros locations are chosen to be at the same locations of the zeros of the DTF directly in the lower elevation.

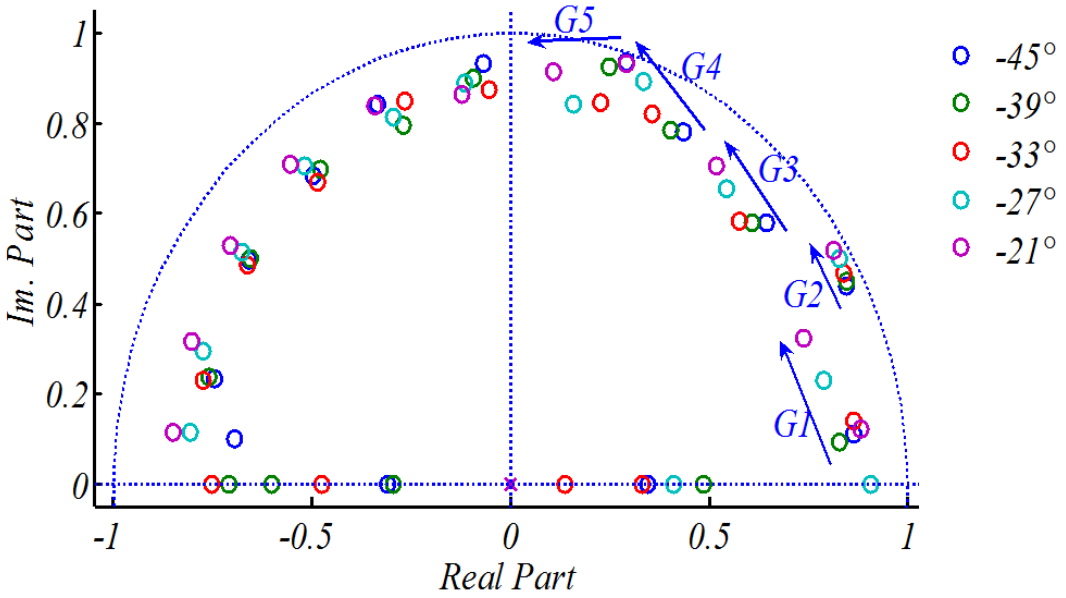


Figure 5.25. Pole/zero plots for DTFs of locations of 0° azimuth and elevations -45° (blue), -39° (green), -33° (red), -27° (cyan), -21° (magenta). The arrows in the figure show the tendency of the systematic movement of the zeros in groups 1-5.

The reconstructed DTFs are evaluated using RMS dB error comparison with the measured DTFs at those locations and also with the preservation of the FN frequency as an important spectral feature for sound localization. The zeros locations of the

reconstructed DTFs at -42°, -36°, -30° and -24° using 3rd order polynomials are shown in Figure 5.26.

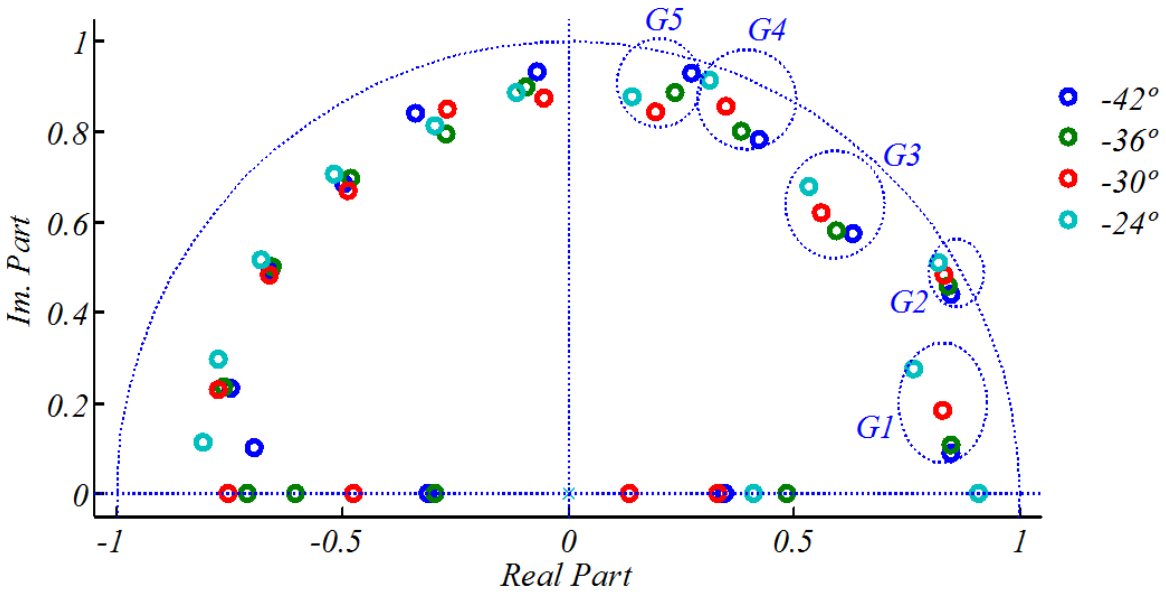


Figure 5.26. Pole/zero plots for reconstructed DTFs of locations of 0° azimuth and elevations -42° (blue), -36° (green), -30° (red), -24° (cyan). The locations of zeros in groups 1-5 were estimated using 3rd order polynomials.

The fitted third order polynomials used to estimate the zeros locations polar angle are demonstrated for the groups from 1 to 5 by Equations 5.9 to 5.13, respectively.

$$\beta = -4.323\theta^3 - 5.073\theta^2 - 0.673\theta + 0.635 \quad (5.9)$$

$$\beta = -1.204\theta^3 - 1.807\theta^2 - 0.613\theta + 0.527 \quad (5.10)$$

$$\beta = -2.181\theta^3 - 2.961\theta^2 - 0.659\theta + 0.988 \quad (5.11)$$

$$\beta = -1.946\theta^3 - 3.167\theta^2 - 1.134\theta + 1.179 \quad (5.12)$$

$$\beta = 2.650\theta^3 + 5.451\theta^2 + 3.974\theta + 2.312 \quad (5.13)$$

The measured, modeled and reconstructed DTFs for elevations of -42°, -36°, -30° and -24° are shown in Figure 5.27. The RMS error values between measured and reconstructed DTFs are shown in Table 5.1.

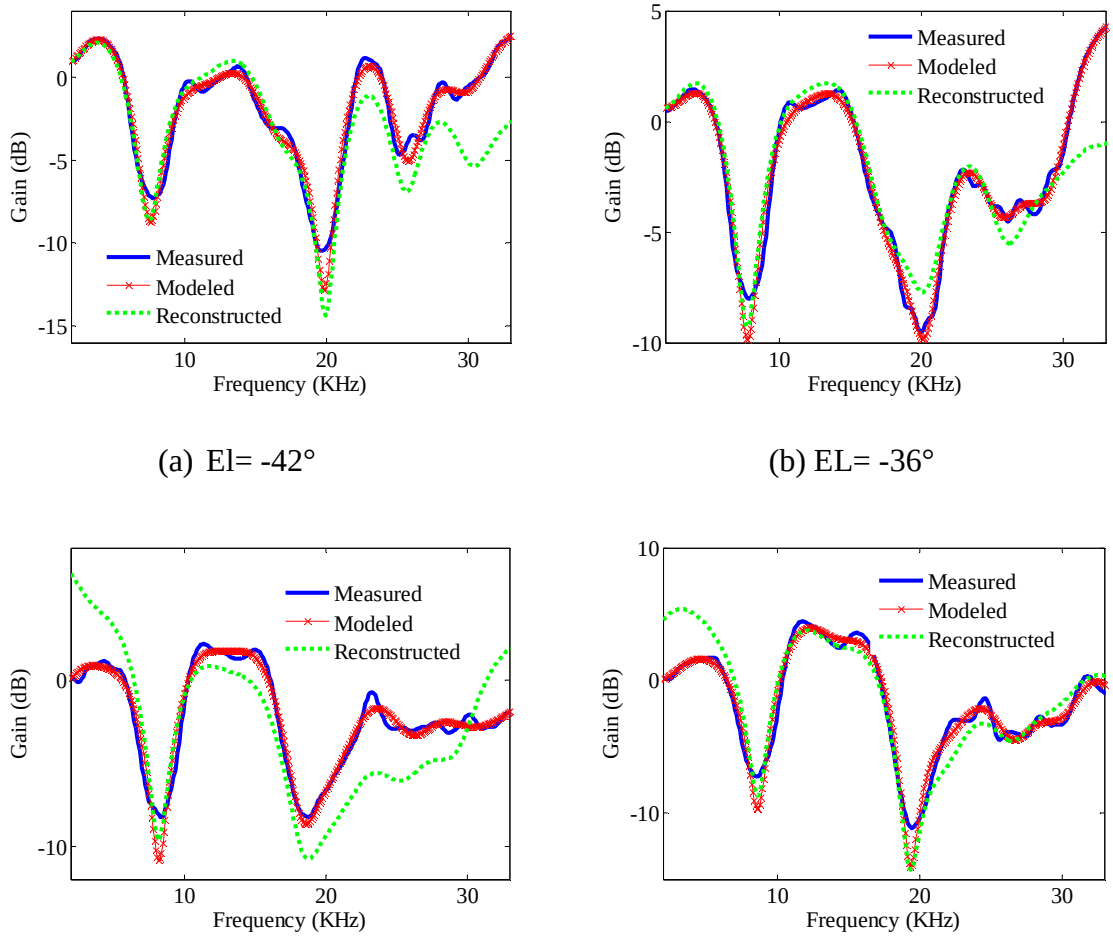


Figure 5.27. Measured, all-zero modeled and reconstructed DTFs using 3rd order polynomials for locations of 0° azimuth and elevations of -42°, -36°, -30° and -24°.

Elevation angle (degrees)	RMS error (2-33 KHz) (dB)	RMS error (2-22 KHz) (dB)
-42°	2.13	1.03
-36°	1.46	0.92
-30°	2.64	2.46
-24°	1.78	2.04

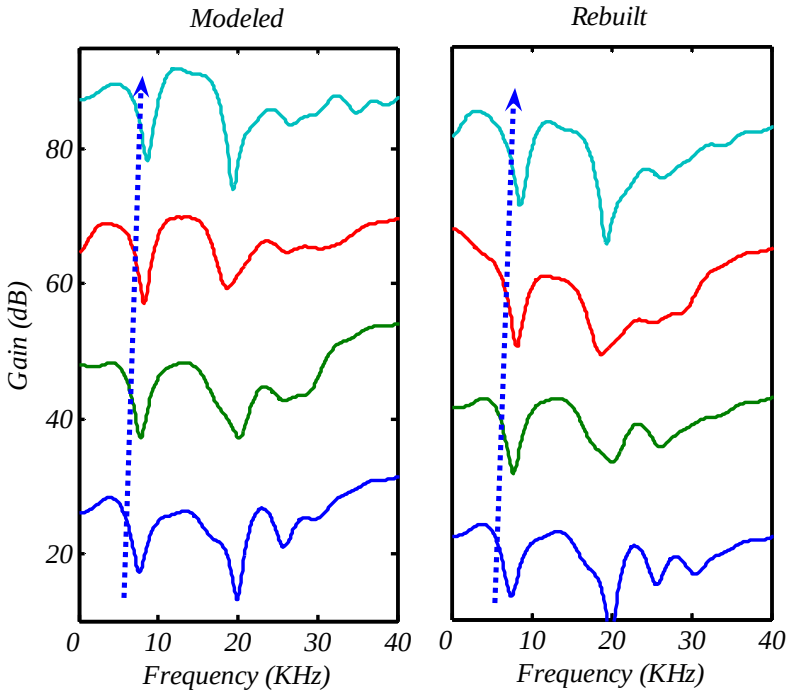
Table 5.1. RMS error values between measured and reconstructed DTFs using 3rd order polynomials for locations of 0° azimuth and elevations of -42°, -36°, -30° and -24°.

Figure 5.28(a) shows the plots for the modeled and the reconstructed DTFs at the four locations with dotted arrows that resemble the tendency of the FN movement with the change in elevation angle. Plots of the changes in the FN frequency with the elevation angle for measured, modeled and reconstructed DTFs are shown in Figure 5.28(b).

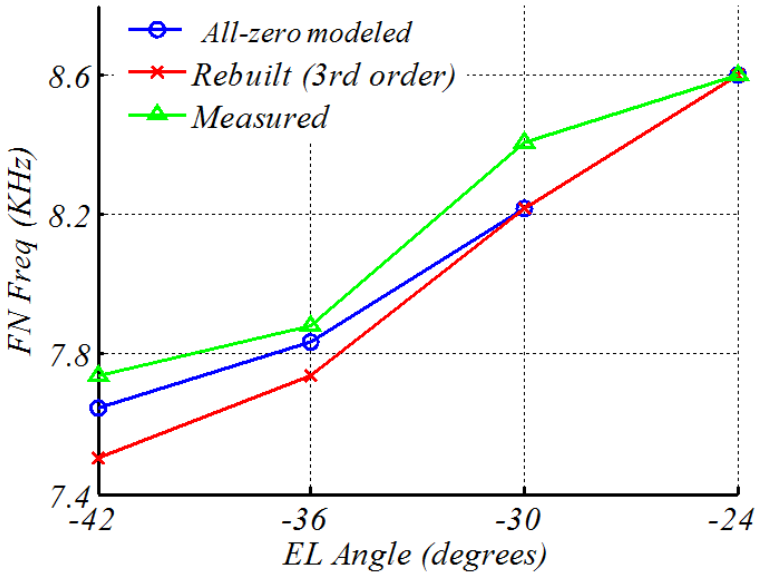
5.3.4 Test #5: Systematic movements of poles in all-pole model with changes in the elevation of sound source in the median plane

Objective: **"***Extract relations between the locations of the poles in the groups that show systematic movements with the change in elevation and the change in the elevation angle in the median plane for DTFs at certain locations. Using those extracted relations, build DTFs at other locations in the median plane and compare the reconstructed DTFs to the measured ones. In addition, a comparison is performed between the reconstructed DTFs using the all-pole and the all-zero modeled DTFs"*

In this part, *test #4* is repeated but with all-pole model used instead of all-zero. The pole/zero plot for locations of the all-pole modeled DTFs at 0° azimuth and elevations -45°, -39°, -33°, -27° and -21° are shown in Figure 5.29. The arrows show the tendency of poles movement with the increase in elevation angle.



(a)



(b)

Figure 5.28 (a) Plots of modeled and the reconstructed DTFs at locations of 0° azimuth and elevations of -42° (bottom), -36°, -30° and -24° (top). The dotted arrow shows the tendency of the FN movement with the change in the elevation angle. (b) Plots of the FN frequency extracted from measured DTFs (green), all-zero modeled DTFs (blue) and DTFs reconstructed by moving the zeros in the five groups (red) versus the elevation angle.

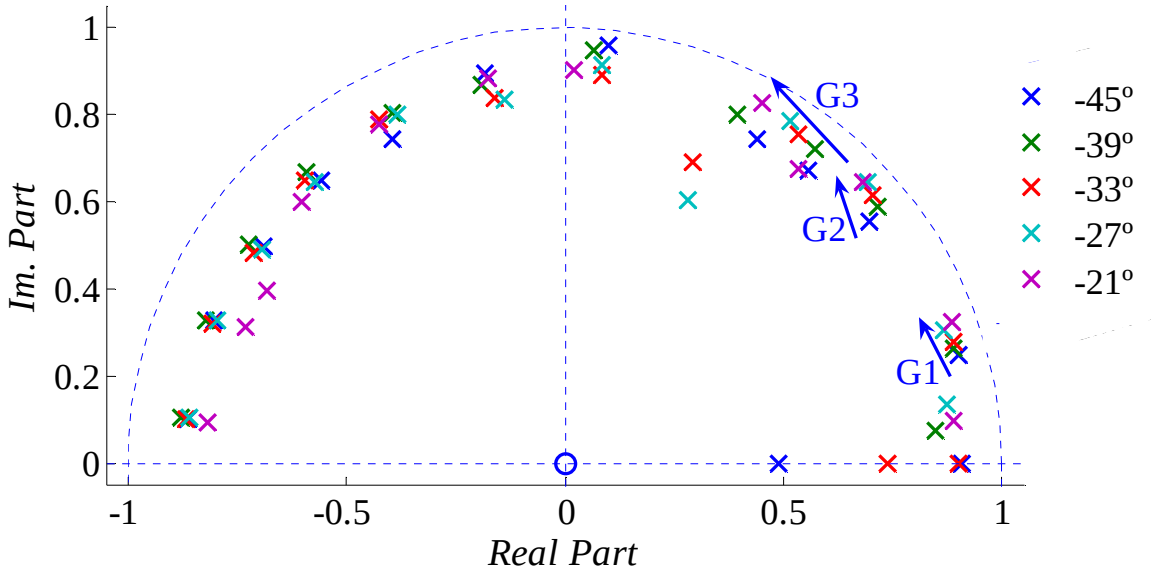


Figure 5.29. Pole/zero plots for all-pole modeled DTFs of locations of 0° azimuth and elevations -45° (blue), -39° (green), -33° (red), -27° (cyan), -21° (magenta). The arrows in the figure show the tendency of the systematic movement of the poles in groups 1-3.

The locations of the poles for the reconstructed DTFs at elevations of -42°, -36°, -30° and -24° using 3rd order fitted polynomials are shown in Figure 5.30. The third-order equations used to estimate the polar angle (β) as a function of the elevation angle (θ) for the locations of the poles in the rebuilt models are given in Equations (5.14-5.16) for poles' groups (1-3), respectively.

$$\beta = -1.247\theta^3 - 2.095\theta^2 - 0.924\theta + 0.234 \quad (5.14)$$

$$\beta = -2.768\theta^3 - 4.874\theta^2 - 2.535\theta + 0.348 \quad (5.15)$$

$$\beta = 0.957\theta^3 + 2.307\theta^2 + 2.121\theta + 1.584 \quad (5.16)$$

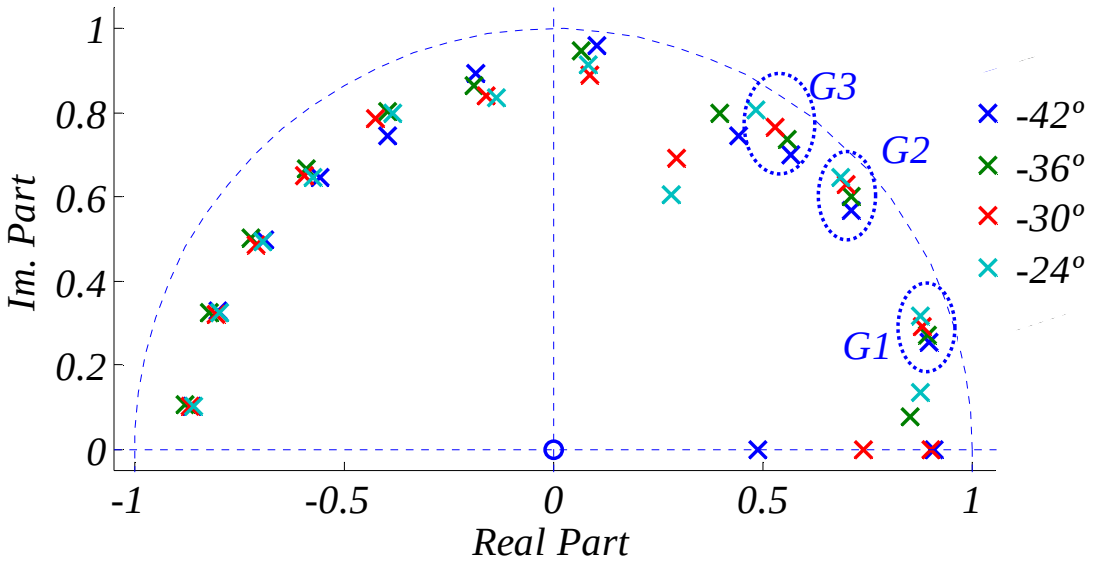
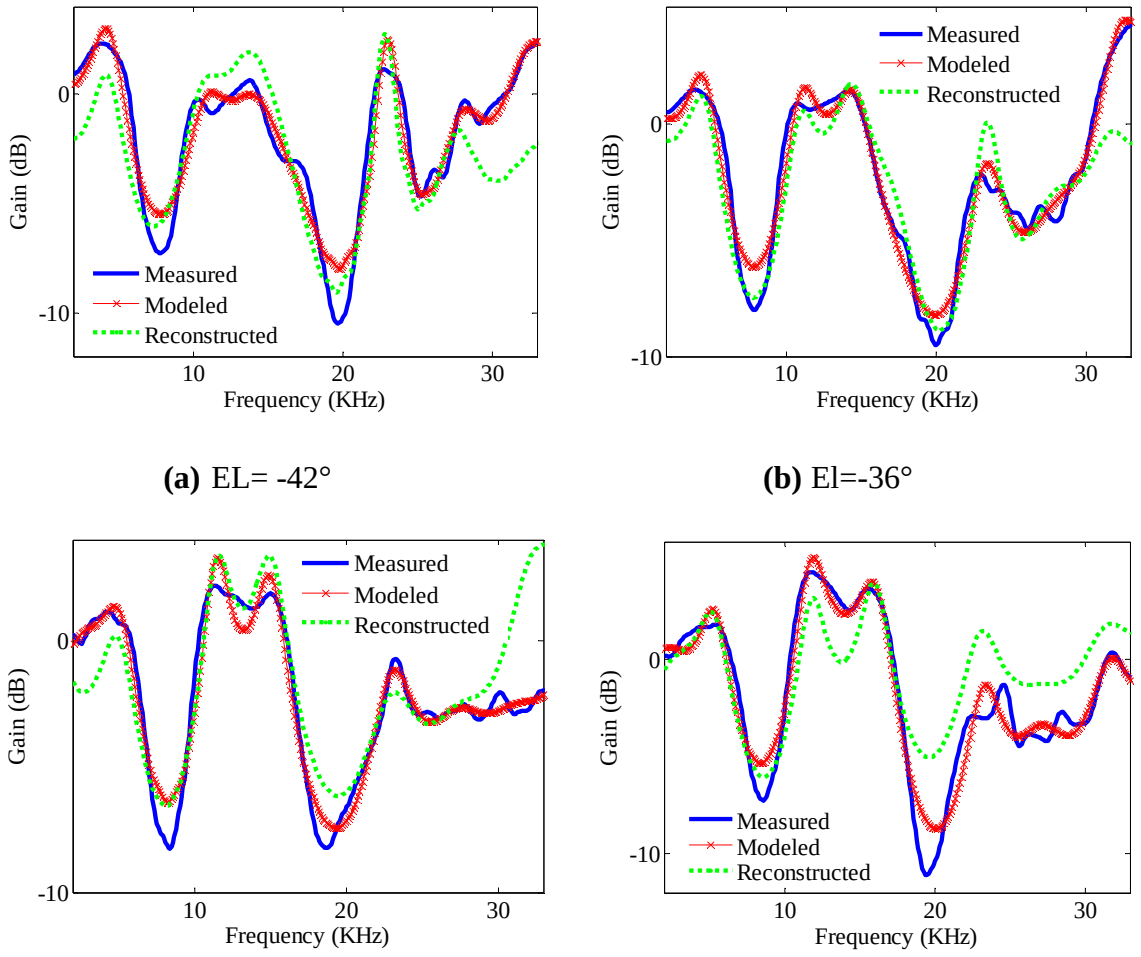


Figure 5.30. Pole/zero plots for reconstructed DTFs of locations of 0° azimuth and elevations -42° (blue), -36° (green), -30° (red), -24° (cyan). The locations of the poles in groups 1-3 were estimated using 3rd order fitted polynomials.

The measured and the reconstructed DTFs for elevations of -42°, -36°, -30° and -24° are shown in Figure 5.31. The RMS error values between measured and reconstructed DTFs are shown in Table 5.2.

Figure 5.32 shows the plots for the modeled and the reconstructed DTFs at elevations -42°, -36°, -30° and -24° with dotted arrows represent the tendency of the FN movement with the change in elevation angle. Plots of the changes in the FN frequency with the elevation angle for measured, reconstructed DTFs using all-pole modeled DTFs and reconstructed DTFs using all-zero modeled ones are shown in Figure 5.33.



(a) EL= -42° **(b)** El=-36°

(c) El= -30° **(d)** El= -24°

Figure 5.31. Measured, all-pole modeled and reconstructed DTFs using 3rd order polynomials for locations of 0° azimuth and elevations of -42°, -36°, -30° and -24°.

Elevation angle (degrees)	RMS error (2-33 KHz) (dB)	RMS error (2-22 KHz) (dB)
-42°	2.04	1.68
-36°	1.49	1.13
-30°	1.84	1.36
-24°	2.50	2.36

Table 5.2. RMS error values between measured and reconstructed DTFs using 3rd order fitted polynomials for locations of 0° azimuth and elevations of -42°, -36°, -30° and -24° using all-pole modeled DTFs.

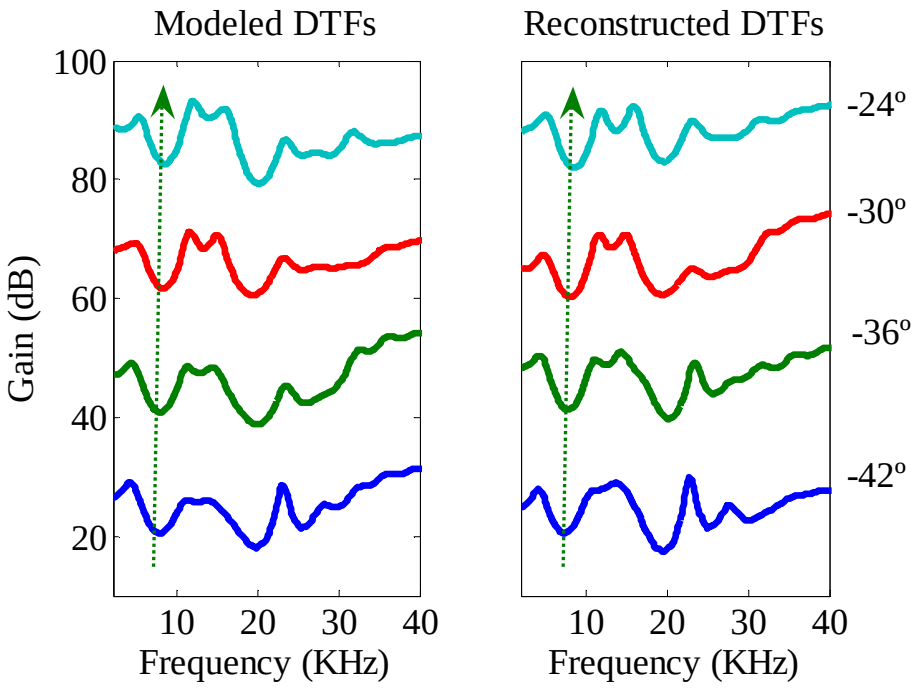


Figure 5.32. Plots for all-pole modeled and reconstructed DTFs at locations of 0° azimuth and elevations of -42° (bottom), -36°, -30° and -24° (top). The dotted arrow shows the tendency of the FN movement with the change in the elevation angle.

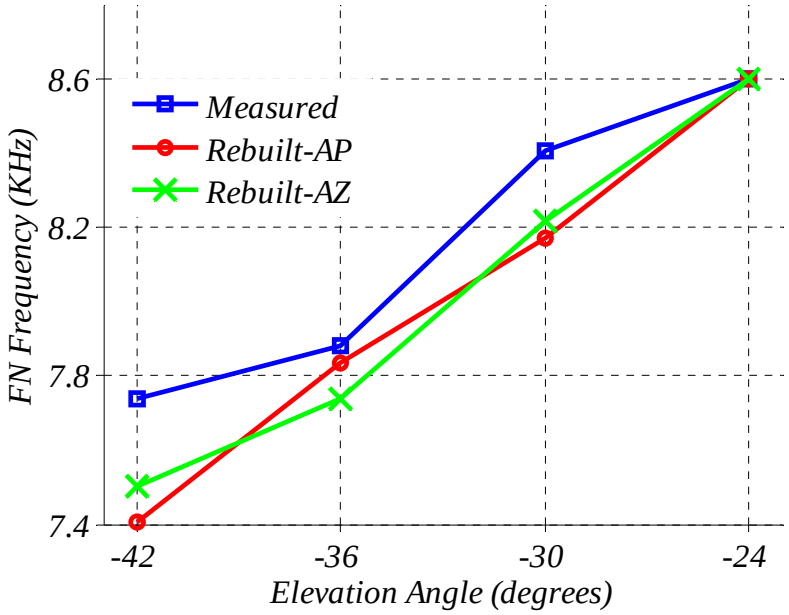


Figure 5.33. Plots of the FN frequency extracted from measured DTFs (green), DTFs reconstructed by moving the zeros of all-zero model in the five groups shown in Figure 5.26 (green), and DTFs reconstructed by moving the poles of all-pole model in the three groups shown in Figure 5.30 (red) versus the elevation angle in the median plane.

5.3.5 Test #6: Systematic movement of zeros in all-zero model with changes in azimuth angle of sound sources in the horizontal plane

Objective: **"***Extract relations between the locations of the zeros in the groups that show systematic movements with the change in azimuth and the change in the azimuth angle in the horizontal plane for DTFs at certain locations. Using those extracted relations, build DTFs at other locations in the horizontal plane and compare the reconstructed DTFs to the measured ones."*

In this test, zeros in two groups of all-zero modeled DTFs that show systematic movements for locations at 0° elevation and azimuths of -30°, -15°, 0°, 15°, and 30° are used to reconstruct DTFs at locations of 0° elevation and azimuths of -22.5°, -7.5°, 7.5° and 22.5°. The zeros plots for DTFs of 0° elevation and azimuths of -30°, -15°, 0°, 15°, and 30° with arrows that show the tendency of zeros movements with the change in azimuth angle are shown in Figure 5.34. Lower resolution in the horizontal plane (7.5° step) compared to the one of the vertical plane (3° step) results in not having the strong systematic movement that has been noticed in the vertical plane. Also, the smaller changes in the spectral features with the change in the azimuth angle compared to the ones with the change in the elevation angle over the same angular displacement makes the expected systematic movements of the poles/zeros weaker with the change in azimuth compared to the ones with the change in elevation. In addition, as previously mentioned, spectral cues are more fundamental in the vertical plane localization compared to the horizontal plane which may add the possibility of having clearer spectral cues changes with elevations compared to those with azimuths.

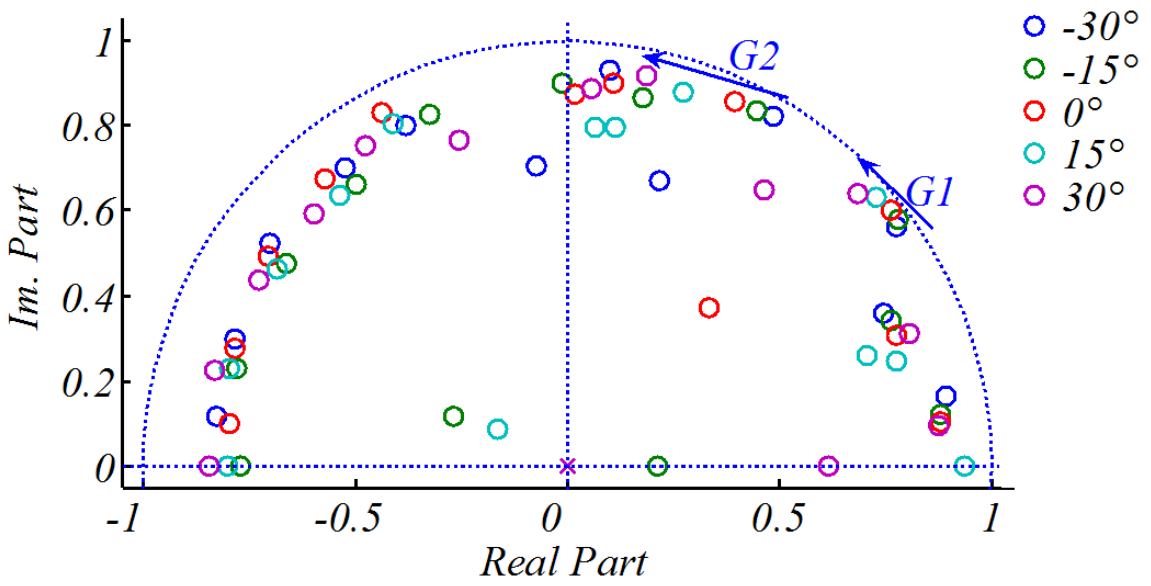


Figure 5.34. Pole/zero plots for DTFs of locations of 0° elevation and azimuths -30° (blue), -15° (green), 0° (red), 15° (cyan), 30° (magenta). The arrows in the figure show the tendency of the systematic movement of zeros in groups 1 and 2.

DTFs at locations of 0° elevation and azimuths of -22.5°, -7.5°, 7.5° and 22.5° have been reconstructed as follows:

1. The polar angle (β) for the locations of zeros' groups that show systematic movements with change in azimuth angle (two groups of symmetric conjugate zeros pairs in this test as shown in Figure 5.34) for each DTF is calculated using the extracted relations depending on the group they are located within.

2. The distance of the zeros in step number 1 from the origin is calculated as the average of the distances of the zeros locations in the same group for the DTFs locations on both sides of the reconstructed DTF location.

3. The rest of the zeros locations are chosen to be at the same locations of the zeros of the modeled DTF that is adjacent the one needs to be reconstructed from the negative azimuth axis.

The reconstructed DTFs were evaluated using the RMS error comparison with the measured DTFs at those locations and also with the preservation of the FN frequency as an important spectral cue for sound localization. The zeros locations of the reconstructed DTFs at 0° elevation and azimuths of -22.5°, -7.5°, 7.5° and 22.5° using 3rd order fitted polynomials are shown in Figure 5.35. The third order polynomials used for the two groups of zeros locations polar angle (β) are given for groups 1 and 2 in Equations 5.17 and 5.18, respectively.

$$\beta = -0.1267\theta^3 + 0.0668\theta^2 + 0.1526\theta + 0.671 \quad (5.17)$$

$$\beta = -0.2278\theta^3 + 0.1939\theta^2 + 0.3777\theta + 1.1533 \quad (5.18)$$

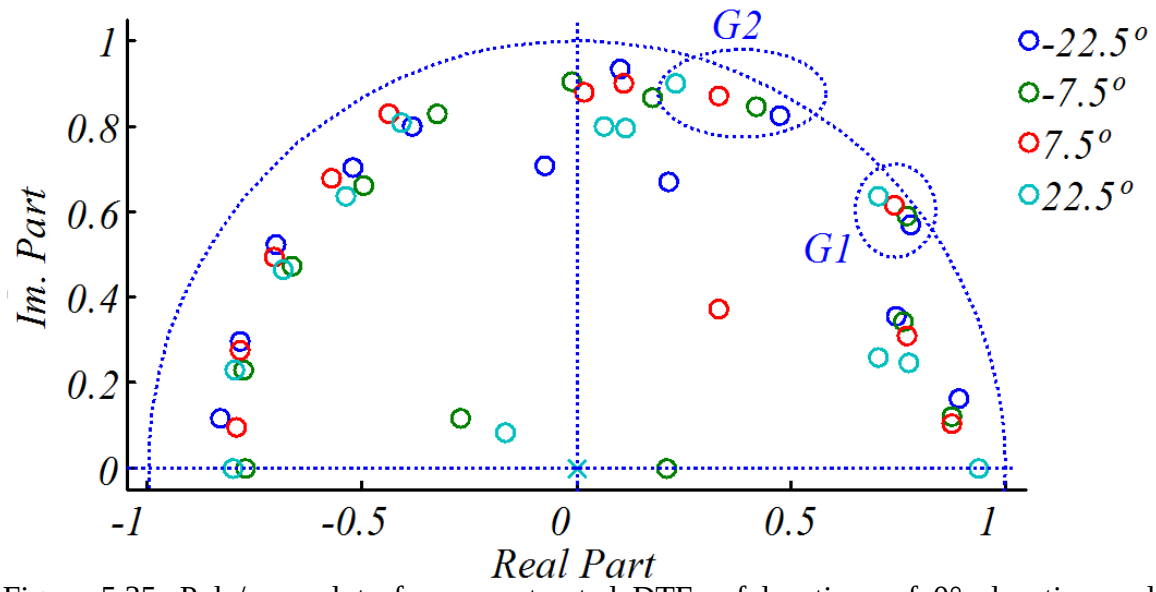


Figure 5.35. Pole/zero plots for reconstructed DTFs of locations of 0° elevation and azimuths -22.5° (blue), -7.5° (green), 7.5° (red), 22.5° (cyan). The locations of zeros in groups 1 and 2 were estimated using 3rd order polynomials.

The measured and the reconstructed DTFs for locations of 0° elevation and azimuths of -22.5°, -7.5°, 7.5° and 22.5° are shown in Figure 5.36. The RMS error values between measured and reconstructed DTFs are shown in Table 5.3.

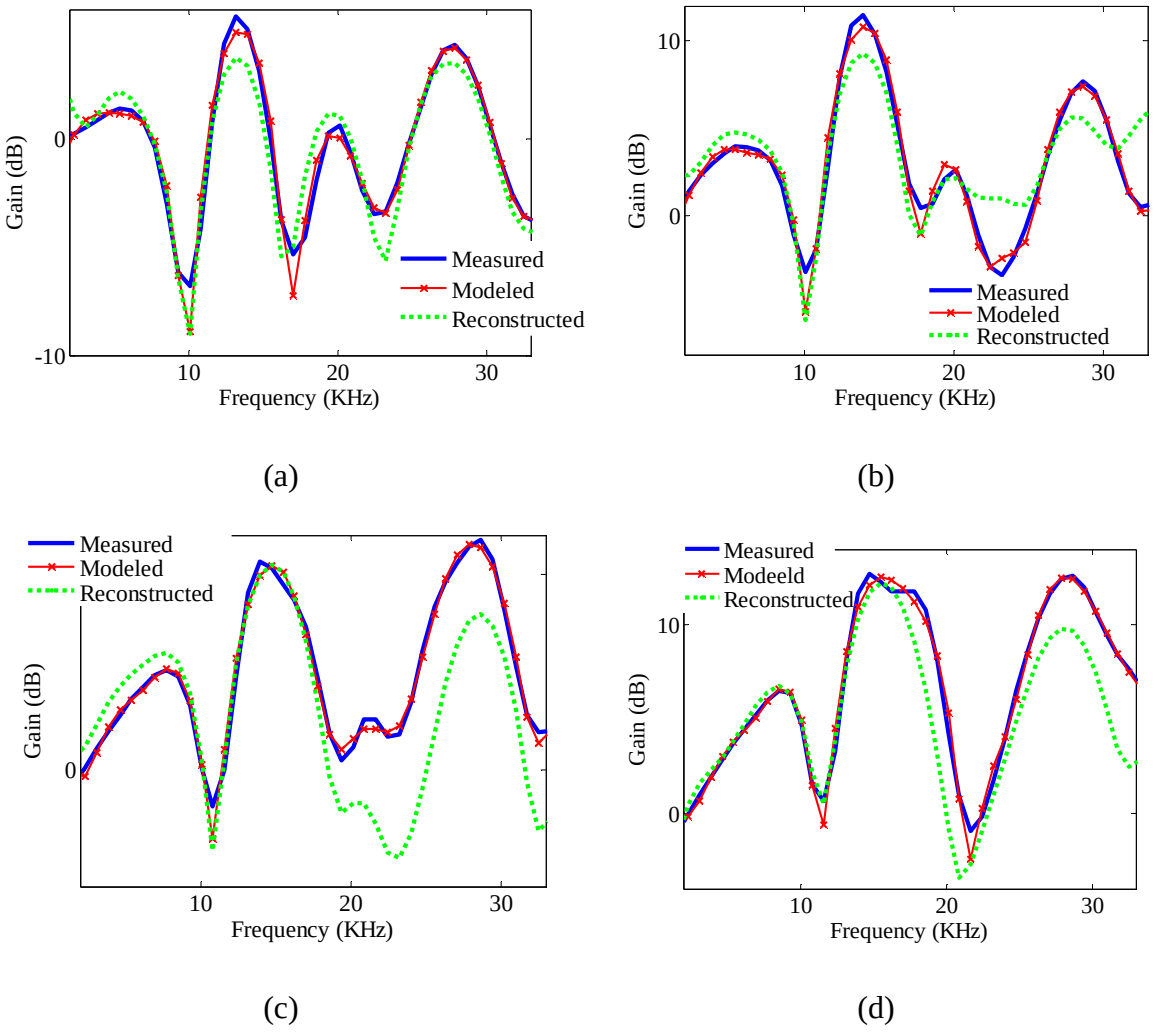


Figure 5.36. Measured, all-zero modeled and reconstructed DTFs using 3rd order polynomials for locations of 0° elevation and azimuths of (a) -22.5°, (b) -7.5°, (c) 7.5° and (d) 22.5°.

Azimuth angle (degrees)	RMS error (2-33 KHz) (dB)	RMS error (2-22 KHz) (dB)
-22.5°	1.17	1.29
-7.5°	1.86	1.30
7.5°	3.26	1.85
22.5°	2.34	1.96

Table 5.3. RMS error values between measured and reconstructed DTFs using 3rd order polynomials for locations of 0° elevation and azimuths of -22.5°, -7.5°, 7.5° and 22.5° for all-zero modeled DTFs.

Figure 5.37(a) shows the plots for the all-zero modeled and the reconstructed DTFs at the four locations with dotted arrows that resemble the tendency of the FN movement with the change in azimuth angle. Plots of the changes in the FN frequency with the azimuth angle for measured, modeled and reconstructed DTFs are shown in Figure 5.37(b).

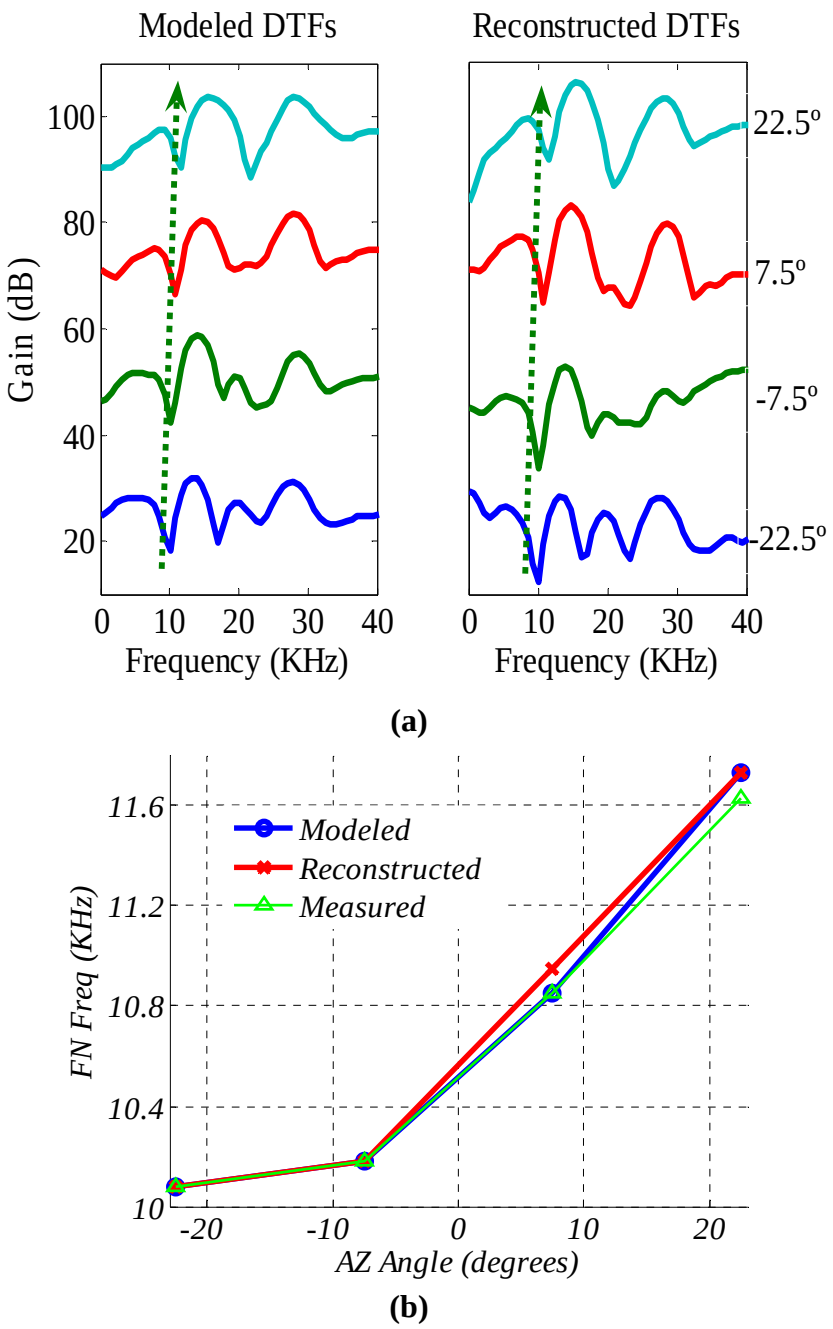


Figure 5.37. (a) Plots for all-zero modeled and reconstructed DTFs at locations of 0° elevation and azimuths of -22.5° (bottom), -7.5°, 7.5° and 22.5° (top). The dotted arrows show the tendency of the FN movement with the change in the azimuth angle. **(b)** Plots of the FN frequency extracted from measured DTFs (green), all-zero modeled DTFs (blue) and DTFs reconstructed by moving the zeros in the two groups (red) versus the azimuth angle in the horizontal plane.

Chapter 6

SUMMARY AND FUTURE WORK

6.1 Summary

All-pole and all-zero models were presented to model the acoustic directional transfer functions (DTFs). While high order models (as high as 64) of all-pole and all-zero gave accurate modeling for the DTFs, all-pole and all-zero models of orders as low as 25 were still able to model human and cat DTFs with errors comparable to previous research findings and with an accurate preservation of the broadband spectral shape and the first spectral notch (FN) feature, the latter which is considered an important direction-dependent spectral feature.

Modeling the direction dependent component of the head related transfer function (HRTFs), the DTFs, and using a suitable smoothing technique, RWT, ensured lower order modeling capability. "*Symmlet 17*" wavelet filter bank has been chosen after comparing the results of different wavelets according to the preservation of the spectral shape, especially the FN frequency in smoothed DTFs compared to measured ones.

An objective dB RMS error criterion was used for comparison with previous research regarding DTF modeling using pole/zero models that were also supported by human

subjective listening tests in order to perceptually validate the all-pole and all-zero modeled DTF. The modeled DTFs in this project yielded comparable RMS error values. In general, all-zero models produced lower error than all-pole models of the same order. In addition to validating the proposed all-pole and all-zero models of order 25 according to the mean logarithmic (dB) difference between the measured and the modeled DTFs, regression analysis and bootstrap resampling techniques showed that the proposed models provide an accurate preservation of the FN feature of the broadband spectral pattern of the DTFs and the systematic movement of the FN frequency with the change in sound source location. To our knowledge, this form of DTF reconstruction validation based on one particular cue to sound location (the FN frequency) has not been used. All-pole and all-zero models presented here have been evaluated using human and cat DTFs.

The proposed all-pole and all-zero models also showed accurate results when the modeled broadband spectral shapes were compared to the measured ones using MSC function. Here we evaluated the modeled DTFs with respect to the measured ones and showed high MSC values (higher than 0.85) for orders 25 of both all-pole and all-zero models when used for both human and cat DTFs. Thus, with order 25 both models accurately preserve the broadband spectral shapes of the DTFs that are necessary for accurate sound source localization.

Finally, a two-layer feed-forward back-propagation ANN with different number of neurons was presented to predict the location of the sound source (elevation and azimuth

angles) using the monaural spectral features of HRTFs of cat data. A NN with 32 neurons in the hidden layer and one neuron in the output layer showed a very good performance in predicting the elevation angle of the sound source from given HRTFs with an average RMS error of 5.15° according to a randomly chosen 43 locations in the frontal field. On the other hand, the performance was not as good in azimuth angle prediction where the average RMS error with the same network architecture was 16.14°.

A feed-forward back-propagation network with 32 neurons in the hidden layer and one neuron in the output layer was presented also as a tool to evaluate the proposed all-pole and all-zero modeling techniques by training the network with measured HRTFs as inputs and their corresponding locations as the targets. The all-pole and all-zero modeled DTFs were simulated by that trained network. According to that evaluation, all-zero models of orders 25 and higher (up to 64) outperforms the ones of all-pole models with the same orders in modeling the measured DTFs according to the regression test and the resulting RMS error values (given in degrees) between the predicted and the actual elevation angles.

In the second main part of this study we have tested and proved that there are clear systematic movements for the poles in the all-pole model and zeros in the all-zero model that accompany the changes in the sound source elevation or azimuth angle. Having low-order models (in the range of 25) simplified the description of these systematic

movements and the tracking for the moving poles and zeros. Both human and cat data have been used to support that hypothesis with the proposed all-pole and all-zero models. By comparing the results with the change in location of both azimuths and elevations, it has been shown that the systematic movement (polar angular displacement) of the poles within each group, that shows accompanying systematic movements, is larger with the change in elevation compared to that with the change in azimuth. The change in FN as an important spectral feature that accompanies the change in elevation or azimuth can justify that. FN frequency showed smaller change with the change in azimuth compared to the one with elevation over the same angular displacement which is clarified in detail in the text.

We have extracted polynomials of first and third order to describe the movements of the poles in all-pole models and zeros in all-zero models with the change in sound source location and used those equations to reconstruct DTFs of sound sources for other locations of finer resolution than the ones used to extract those relations. The reconstructed DTFs were compared to the measured ones of same locations. We have shown that the reconstructed DTFs preserved the main shape of the spectra, have a satisfactory RMS error compared to the measured ones and accurately preserved the FN feature in the reconstructed DTFs. The accurate preservation of the FN feature in the reconstructed DTFs can be justified by the fact that the systematic movement of the poles and zeros of the upper half of the unit circle in the z-plane are mainly noticed between the positive real axis and the positive imaginary axis and may extend a little bit to the left of

the positive imaginary axis which is the region that affects the FN region the most in both human and cat DTF.

6.2 Future Work

Regarding the techniques and the modeling discussed in this dissertation, further investigation is deserved for the following areas:

1. To assess the fidelity of the reconstructed DTFs fully, human subjective listening test should be done. The modeling framework and objective error analysis techniques have been fully developed in this dissertation. Modeled DTFs of different orders can be generated and human subjective listening tests can be systematically performed to validate the present results.

2. As the resolution of the DTFs used to extract the equations needed for the poles/zeros movement description increases, it is expected that the accuracy of the reconstructed DTFs increases and vice versa. Further investigation of what will be the minimum spatial resolution needed for the measured DTFs to allow successful reconstruction of DTFs from the measured ones can be done.

3. The joint dependence of the pole/zero trajectories with source location and head and pinnae dimensions is an area to investigate. Knowing how poles and zeros may change with the change in the size of the pinna and the head will add a useful tool to HRTF/DTF modeling among different individuals. Is there a simple scale factor that can be applied to

the pole/zero models with the scale factor based on measured head and pinnae dimensions?

4. Kulkanri and Colburn (1998) showed that the DTFs can be smoothed considerably without affecting the perception. Potentially, models of order much lower than 25 may be achievable after substantial smoothing on the order of what Kulkarni and Colburn describe. Even lower order models than 25 would be desirable for investigating and modeling the systematic movement of poles/zeros with source location since fewer poles or zeros must be tracked.

BIBLIOGRAPHY

Abdel Alim, O. and Farag, H. (2000). "Modeling non-individualized binaural sound localization in the horizontal plane using artificial neural networks," Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on, vol. 3, 642-647 vol.3.

Aitkin, L. M., and Martin, R. L. (1987). "The representation of stimulus azimuth by high best-frequency azimuth-selective neurons in the central nucleus of the inferior colliculus of the cat," J Neurophysiol 57, no. 4: 1185-1200.

Aitkin, L.M. and Martin, R.L. (1990). "Neurons in the inferior colliculus of cats sensitive to sound-source elevation," Hear. Res. 50, 97-106.

Al-Sheikh, B., Matin, M., and Tollin, D. J. (2007). "Are there systematic movements of poles/zeros with sound source location for pole/zero models of acoustic directional transfer functions?," The Journal of the Acoustical Society of America 122, no. 5: 3061.

Al-Sheikh, B. W., Matin, A. M. and Tollin, D. J. (2009). "All-pole and all-zero models of human and cat head related transfer functions," Accepted to SPIE Advanced Signal Processing Algorithms, Architectures, and Implementations XIX.

Asano, F., Suzuki, Y., and Sone, T. (1990). "Role of spectral cues in median plane localization," J. Acoust. Soc. Am. 88, no. 1: 159-168.

Blauert, J. (1969). ''Sound localization in the median plane,'' Acoustica, vol. 22, pp.205–213.

Blauert, J. (1997). "Spatial Hearing (Revised Edition): The psychophysics of human sound localization", (MIT Press, Cambridge, MA).

Blommer, M. A., and Wakefield, G. H. (1992). "Investigation of phase distortion in the synthesis of head-related transfer functions," J. Acoust. Soc. Am. 92, no. 4: 2297.

Blommer, M. A., and Wakefield, G. H. (1994). “On the design of pole-zero approximations using a logarithmic error measure,” Signal Processing, IEEE Transactions on 42, no. 11: 3245-3248.

Blommer, M. A., and Wakefield, G. H. (1997). “Pole-zero approximations for head-related transfer functions using a logarithmic error criterion,” Speech and Audio Processing, IEEE Transactions on 5, no. 3: 278-287.

Brown, C. P., and Duda, R. O. (1998). “A structural model for binaural sound synthesis,” Speech and Audio Processing, IEEE Transactions on 6, no. 5: 476-488.

Butler, R. A., and Belendiuk, K. (1977). “Spectral cues utilized in the localization of sound in the median sagittal plane,” J. Acoust. Soc. Am. 61, no. 5: 1264-1269.

Carlile, S. (1996). “Virtual Auditory Space: Generation and Applications”, R. G. Landes Company, Chapman and Hall, NY, USA.

Casseday, J. H., and Neff, W. D. (1973). “Localization of pure tones,” J. Acoust. Soc. Am. 54, no. 2: 365-372.

Chase, S. M., and Young, E. D. (2006). “Spike-timing codes enhance the representation of multiple simultaneous sound-localization cues in the inferior colliculus,” J. Neurosci. 26, no. 15: 3889-3898.

Cheng, C. I., and Wakefield, G. H. (1999). “Spatial frequency response surfaces: an alternative visualization tool for head-related transfer functions (HRTFs),” In: Proc. IEEE Internat. Conf. Acoustics, Speech Signal Process, 2: 961-964.

Chung, W., Carlile, S. and Leong, P. (2000). “A performance adequate computational model for auditory localization,” The Journal of the Acoustical Society of America 107, no. 1: 432-445.

Demuth, H., Beale, M. and Hagan, M. (2007). “Neural network toolbox 6: User's guide”, The MathWorks, Accelerating the pace of engineering and science.

Donoho, D. L. (1995). “Denoising by Soft-thresholding,” IEEE Trans. Inf. Theory, vol. 41, pp. 613-627.

Donoho, D. L., and Johnstone, I. M. (1995). “Adapting to Unknown Smoothness via Wavelet Shrinkage,” J. Am. Statist. Ass., vol. 90, 1200-1224.

Durant, E. A., and Wakefield, G. H. (2002). “Efficient model fitting using a genetic algorithm: pole-zero approximations of HRTFs,” Speech and Audio Processing, IEEE Transactions on 10, no. 1: 18-27.

Furukawa, S., Xu, L. and Middlebrooks, J. C. (2000). “Coding of Sound-Source Location by Ensembles of Cortical Neurons,” J. Neurosci. 20, no. 3: 1216-1228.

Gathercole, C., and Mantooth, H. A. (2001). “Pole-zero localization: a behavioral modeling approach,” In: Proc of the Fifth IEEE Internat. Workshop on Behavioral Modeling and Simulation, 59-65.

Grantham, D. W., Willhite, J. A., Frampton, K. D., and Ashmead, D. H. (2005). “Reduced order modeling of head related impulse responses for virtual acoustic displays,” J. Acoust. Soc. Am. 117, no. 5: 3116-3125.

Hacihabiboglu, H. (2002). “Interpolation of low-order HRTF filters using a zero displacement measure,” EAA Convention, Proc. Forum Acusticum Sevilla (CD-ROM), Sevilla, Spain.

Hacihabiboglu, H., Gunel, B., and Murtagh, F. (2002). “Wavelet-based spectral smoothing for head-related transfer function filter design,” Proc. of the AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio, Espoo, Finland.

Haneda, Y., Makino, S., Kaneda, Y., and Kitawaki, N. (1999). “Common-acoustical-pole and zero modeling of head-related transfer functions,” Speech and Audio Processing, IEEE Transactions on 7, no. 2: 188-196.

Hayes, M. H., Schaums’s Outline of Theory and Problems of DigitalSignal Processing. New York: Mc-Graw Hill, 1999.

Hebrank, J., and Wright, D. (1974). “Spectral cues used in the localization of sound sources on the median plane,” J. Acoust. Soc. Am. 56, no. 6: 1829-1834.

Hofman, P. M., Van Riswick, J. A. and Van Opstal, A. J. (1998). “Relearning sound localization with new ears,” Nature neuroscience, vol 1, no 5, 417-421.

Huang, A. Y., and May, B. J. (1996). “Sound orientation behavior in cats. II. Mid-frequency spectral cues for sound localization,” J. Acoust. Soc. Am. 100, no. 2: 1070-1080.

Huopaniemi, J., Zacharov, N., and Karjalainen, M. (1999). ‘‘Objective and subjective evaluation of head-related transfer function filter design,’’ J. Audio Eng. Soc. 47, 218–239.

Jackson, L. B. (1989). “Digital Filters and Signal Processing”, 2nd ed., (Kluwer Academic Publishers, Boston, MA).

Jenison, R. L. (1995). “A spherical basis function neural network for pole-zero modeling of head-related transfer functions,” In: IEEE Workshop on Applications of Signal Process to Audio and Acoustics, 92-95.

Jin, C., Leong, P., Leung, J., Corderoy, A., and Carlile, S. (2000). “Enabling individualized virtual auditory space using morphological measurements,” In Proceedings of the IEEE 2000 International Symposium on Multimedia Information Processing, 235-238.

Jot, J., Larcher, V., and Warusfel, O. (1995). “Digital signal processing issues in the context of binaural and transaural stereophony,” In Proc. 98th Audio Engr. Soc. Conv., preprint 3980, Paris, France.

Kaalund, C. J., and Peng, Gang-Ding. (2004). “Pole-zero diagram approach to the design of ring resonator-based filters for photonic applications,”, J. Lightwave Technology 22, no. 6: 1548-1559.

Kistler, D. J., and Wightman, F. L. (1992). “A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction,” J. Acoust. Soc. Am. 91, no. 3: 1637-1647.

Koka, K., Read, H. L., and Tollin, D. J. (2008). “The acoustical cues to sound location in the rat: measurements of directional transfer functions,” J. Acoust. Soc. Am. 123, no. 6: 4297-4309.

Kuhn, G. F. (1987). "Acoustics and measurements pertaining to directional hearing," in W. A. Yost and G. Gourevitch, Eds., Directional Hearing (Springer Verlag, New York, 1987), pp. 3-25.

Kulkarni, A., and Colburn, H. S. (1993) “Evaluation of a linear interpolation scheme for approximating HRTFs.” The Journal of the Acoustical Society of America 93, no.4.

Kulkarni, A., and Colburn, H. S. (1998). “Role of spectral detail in sound source localization,” Nature, London, 396, 747–749.

Kulkarni, A. and Colburn, H. S. (2004). “Infinite-impulse-response models of the head-related transfer function,” J. Acoust. Soc. Am. 115, no. 4: 1714-1728.

Kulkarni, A., Isabelle, S. K., and Colburn, H. S. (1999). “Sensitivity of human subjects to head-related transfer-function phase spectra,” J. Acoust. Soc. Am. 105, no. 5: 2821-2840.

Kulkarni, A. and Colburn, H. S. (2004). “Infinite-impulse-response models of the head-related transfer function,” J. Acoust. Soc. Am. 115, no. 4: 1714-1728.

Kulkarni, A., Isabelle, S. K., and Colburn, H. S. (1999). “Sensitivity of human subjects to head-related transfer-function phase spectra,” J. Acoust. Soc. Am. 105, no. 5: 2821-2840.

Langendijk, E. H., and Bronkhorst, A. W. (1997). “Collecting localization response with a virtual acoustic pointer,” J. Acoust. Soc. Am. 101, no. 5: 3106.

Langendijk, E. H., and Bronkhorst, A. W. (2002). “Contribution of spectral cues to human sound localization,” J. Acoust. Soc. Am. 112, no. 4: 1583-1596.

Liu, C. J., and Hsieh, S. F. (2001). “Common-acoustic-poles/zeros approximation of head-related transfer functions,” Acoustics, Speech, and Signal Processing, IEEE International Conference on, 5:3341-3344 vol.5.

Lopez-Poveda, E. A., and Meddis, R. (1996). “A physical model of sound diffraction and reflections in the human concha,” J. Acoust. Soc. Am. 100, no. 5: 3248-3259.

Lyons, R. G. (1997). “Understanding Digital Signal Processing”, Addison Wesley Longman, pp. 219-220.

Mackenzie, J., Huopaniemi, J., Valimaki, V., and Kale, I. (1997). “Low-order modeling of head-related transfer functions using balanced model truncation,” Signal Processing Letters, IEEE 4, no. 2: 39-41.

Makhoul, J. : Linear prediction: a tutorial review. Proc. IEEE 63, 561-580 (1975).

Maki, K., and Furukawa, S. (2005). “Acoustical cues for sound localization by the Mongolian gerbil, Meriones unguiculatus,” J. Acoust. Soc. Am. 118, no. 2: 872-886.

Mallat, S. (1998). “A wavelet tour of signal processing”, (Academic Press, London, 1998).

Martens, W. L. (1987). “Principal components analysis and resynthesis of spectral cues to perceived direction,” In Proc. Int. Comput. Music Conf., 1987, pp. 274-281.

Martin, R., and McAnally, K. (2007). “Interpolation of Head-Related Transfer Functions” Air Operations Division, Defence Science and Technology Organisation, Victoria 3207 Australia.

May, B. J., and Huang, A. Y. (1996). “Sound orientation behavior in cats: I. Localization of broadband noise,” J. Acoust. Soc. Am. 100, no. 2: 1059-1069.

Mehrgardt, S., and Mellert, V. (1977). “Transformation characteristics of the external human ear,” J. Acoust. Soc. Am. 61, no. 6: 1567-1576.

Middlebrooks, J. C. (1992). “Narrow-band sound localization related to external ear acoustics,” J. Acoust. Soc. Am. 92, no. 5: 2607-2624.

Middlebrooks, J. C. (1999). “Individual differences in external-ear transfer functions reduced by scaling in frequency,” J. Acoust. Soc. Am. 106, no. 3: 1480-1492.

Middlebrooks, J. C. and Green, D. M. (1992). “Observations on a principal components analysis of head-related transfer functions,” J. Acoust. Soc. Am. 92, no. 1: 597-599.

Middlebrooks, J. C., Makous, J. C. and Green, D. M. (1989). “Directional sensitivity of sound-pressure levels in the human ear canal,” J. Acoust. Soc. Am. 86, no. 1: 89-108.

Musicant, A. D., Chan, J. C. K., and Hind, J. E. (1990). “Direction-dependent spectral properties of cat external ear: new data and cross-species comparisons,” J. Acoust. Soc. Am. 87, no. 2: 757-781.

Neti, C., Young, E. D. and Schneider, M. H. (1992). “Neural network models of sound localization based on directional filtering by the pinna,” The Journal of the Acoustical Society of America, no. 6 (1992): 3140-3156.

Neural Networks, Statsoft, Inc., 1984-2008.

Nishino, T., Kajita, S., Takeda, K., and Itakura, F. (1999). “Interpolating head related transfer functions in the median plane,” In IEEE Workshop on Applications of Signal Process to Audio and Acoustics, 167-170.

Nishino, T., Inoue, N., Takeda, K. and Itakura, F. (2007). “Estimation of HRTFs on the horizontal plane using physical features.” Applied Acoustics 68, no. 8: 897-908.

Oppenheim, A. V., and Schafer, R. W. (1989). “Discrete-time signal processing”, (Prentice-Hall, N.J.).

Otani, M., and Ise, S. (2006). “Fast calculation system specialized for head-related transfer function based on boundary element method,” J. Acoust. Soc. Am. 119, no. 5: 2589-2598.

Percival, D. B., and Walden, A. T. (1993). “Spectral analysis for physical applications: multitaper and conventional univariate techniques”, Cambridge Univ. Press, New York.

Poganiatz, I., Nalken, I., and Wagner, H. (2001). ‘‘Sound-localization experiments with barn owls in virtual space: influence of interaural time difference on head-turning behavior,’’ J. Assoc. Res. Otolaryngol. 2, 1–21.

Raykar, V. C., Duraiswami, R., and Yegnanarayana, B. (2005). “Extracting the frequencies of the pinna spectral notches in measured head related impulse responses,” J. Acoust. Soc. Am. 118, no. 1: 364-374.

Rayleigh, L. (1907). “On our perception of sound direction,” Phil Mag, vol 13, 214-32.

Rice, J. J., Bradford J. M., Spirou, G. A., and Young, E. D. (1992). “Pinna-based spectral cues for sound localization in cat,” Hearing Research 58, no. 2 (March): 132-152. doi:doi: DOI: 10.1016/0378-5955(92)90123-5.

Runkle, P. R., Blommer, M. A., and Wakefield, G. H. (1995). “A comparison of head related transfer function interpolation methods.” In IEEE Workshop on Applications of Signal Process to Audio and Acoustics, 88-91.

Sandvad, J., and Hammershøi, D. (1994). “Binaural auralization: comparison of FIR and IIR filter representation of HIRs,” 96th Audio Engr Soc Conv, Amsterdam, The Netherlands, preprint 3862. Abstr in: J Audio Engr Soc 42:395.

Schnupp, J. W. H., Booth, J., and King, A. J. (2003). “Modeling individual differences in ferret external ear transfer functions,” J. Acoust. Soc. Am. 113 (4), 2021-2030.

Searle, C. L., Braida, L. D., Davis, M. F., and Colburn, H. S. (1976). “Model for auditory localization,” J. Acoust. Soc. Am. 60, no. 5: 1164-1175.

Shaw, E. A. (1974). “Transformation of sound pressure level from the free field to the eardrum in the horizontal plane” J. Acoust. Soc. Am. 56, no. 6: 1848-1861.

Shaw, J. C. (1981). “An introduction to the coherence function and its use in EEG signal analysis,” Journal of Medical Engineering & Technology. 5, 279-288.

Shaw, E. A. (1982). “External ear response and sound localization in localization of sound: Theory and applications”, edited by R. Gatehouse (Amphora, Groton, CT), pp. 30-41.

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

Soli, S. D., Jayaraman, S., Gao, S., and Jean, S. (1994). “Method of signal processing for maintaining directional hearing with hearing aids,” U.S. Patent 5,325,436 (Jun 28, 2004).

Spezio, M. L., Keller, C. H., Marrocco, R. T., and Takahashi, T. T. (2000). “Head-related transfer functions of the Rhesus monkey,” Hearing Research 144, no. 1-2 (June): 73-88. doi:doi: DOI: 10.1016/S0378-5955(00)00050-2.

Signal processing toolbox 6 software and user’s guide, MATLAB (The MathWorks Inc., Natick, MA, USA.

Susnik, R., Sodnik, J., Umek, A., and Tomazic, S. (2003). “Spatial sound generation using HRTF created by the use of recursive filters,” In EUROCON 2003. Computer as a Tool. The IEEE Region 8, 1:449-453 vol.1.

Tetsufumi, I., Shiro, I., Hidemi, I. (2000). “ Pinnae's contribution in head related transfer function (HRTF),” IEIC Technical Report (Institute of Electronics, Information and Communication Engineers), Vol.100; no.725, 153-159.

Tollin, D. J. (2003). “The lateral superior olive: A functional role in sound source localization,” Neuroscientist 9, no. 2: 127-143.

Tollin, D. J., and Koka, K. (2009). “Postnatal development of sound pressure transformations by the head and pinnae of the cat: monaural characteristics,” J. Acoust. Soc. Am. 125, no. 2, Feb. 2009.

Tollin, D. J., Populin, L. C., Moore, J. M., Ruhland, J. L., and Yin, T. C. T. (2005). “Sound-localization performance in the cat: The effect of restraining the head,” J Neurophysiol 93, no. 3: 1223-1234.

Tollin, D. J., and Yin, T. C. T. (2002). “The coding of spatial location by single units in the lateral superior olive of the cat: I. Spatial receptive fields in azimuth,” J. Neurosci. 22, no. 4: 1454-1467.

Tollin D. J. and Yin T. C. T. (2003). “Spectral cues explain illusory elevation effects with stereo sounds in cats.” J Neurophysiol 90: 525–530.

Van Drongelen, W. (2006). “Signal processing for neuroscientists: An introduction to the analysis of physiological signals.” Academic Press, pp: 142-147.

Wang, L., Yin, F. and Chen, Z. (2008). Acoustical Science and Technology 29, no. 5: 329-331.

Wenzel, E. M., and Foster, S. H. (1993). “Perceptual consequences of interpolating head-related transfer functions during spatial synthesis,” In IEEE Workshop on Applications of Signal Process to Audio and Acoustics, Final Program and Paper Summaries, 102-105.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (1993). “Localization using nonindividualized head-related transfer functions,” J. Acoust. Soc. Am. 94, no. 1: 111-123.

Wesley, G. D., Willhite, J. A., Frampton, K. D., and Ashmead, D. H. (2005). “Reduced order modeling of head related impulse responses for virtual acoustic displays,” J. Acoust. Soc. Am. 117, no. 5: 3116-3125.

Wightman, F. L., and Kistler, D. J. (1989a). “Headphone simulation of free-field listening. I: Stimulus synthesis,” J. Acoust. Soc. Am. 85, no. 2: 858-867.

Wightman, F. L., and Kistler, D. J. (1989b). “Headphone simulation of free-field listening. II: Psychophysical validation,” J. Acoust. Soc. Am. 85, no. 2: 868-878.

Wightman, F. L., and Kistler, D. J. (1999). “Explaining individual differences in head-related transfer functions” J. Acoust. Soc. Am. 105, no. 2: 1036.

Young, E. D., Spirou, G. A., Rice, J. J., and Voigt, H. F. (1992). “Neural organization and responses to complex stimuli in the dorsal cochlear nucleus,” Philos. Trans. R. Soc. Lond B Biol Sci, B 336(1278):407–413.

Zhong, X., and Xie. B. (2009) “Maximal azimuthal resolution needed in measurements of head-related transfer functions.” The Journal of the Acoustical Society of America 125, no. 4: 2209-2220.

Appendix A

Locations of DTFs used for NN training in *Test 3*, section 4.4.3.

(Angles are given in degrees):

AZ / EL	-75	-60	-45	-30	-15	0	7.5	15	22.5	30	37.5	45	52.5	60	67.5	75
-30	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
-22.5	**Y**		**Y**		**Y**	**Y**		**Y**		**Y**		**Y**		**Y**		**Y**
-15	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
-7.5	**Y**		**Y**		**Y**		**Y**		**Y**		**Y**		**Y**		**Y**	**Y**
0	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
7.5	**Y**	**Y**		**Y**		**Y**		**Y**		**Y**	**Y**		**Y**		**Y**	**Y**
15	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
22.5	**Y**		**Y**		**Y**		**Y**		**Y**	**Y**	**Y**		**Y**	**Y**		**Y**
30	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**		**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
37.5	**Y**		**Y**		**Y**		**Y**		**Y**	**Y**	**Y**		**Y**	**Y**		**Y**
45	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
52.5	**Y**		**Y**		**Y**	**Y**		**Y**	**Y**		**Y**		**Y**		**Y**	**Y**
60	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
67.5	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
75	**Y**		**Y**		**Y**	**Y**		**Y**	**Y**		**Y**	**Y**		**Y**		**Y**
82.5	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**

Appendix B

List of Abbreviations

ANN	Artificial Neural Network
AP	All-pole
ARMA	Auto-regressive Moving Average
AZ	All-zero
BGD	Batch Gradient Descent
dB	Decibel
DTF	Directional Transfer Function
EL	Elevation
FIR	Finite Impulse Response
FN	First Notch
GB	Giga Byte
HRIR	Head Related Impulse Response
HRTF	Head Related Transfer Function
Hz	Hertz
IIR	Infinite Impulse Response
ILD	Inter-aural Level Difference
ITD	Inter-aural Time Difference
LM	Levenbergh-Marquardt
LPC	Linear Predictor Coefficients
MSC	Magnitude Squared Coherence
RMS	Root Mean Squared
RWT	Redundant Wavelet Transform
SSL	Sound Source Location
SURE	Stein's Unbiased Risk Estimation
UCHSC	University of Colorado Health Science Center
VAD	Virtual Auditory Display