INTRODUCTION TO HEAD-RELATED TRANSFER FUNCTIONS (HRTF'S): REPRESENTATIONS OF HRTF'S IN TIME, FREQUENCY, AND SPACE
(invited tutorial)

COREY I. CHENG[1] AND GREGORY H. WAKEFIELD[2]

[1]University of Michigan, Department of Electrical Engineering and Computer Science
Ann Arbor, Michigan, U.S.A.
coreyc@eecs.umich.edu ; http://www.eecs.umich.edu/~coreyc

[2] University of Michigan, Department of Electrical Engineering and Computer Science
Ann Arbor, Michigan, U.S.A.
ghw@eecs.umich.edu

This tutorial provides an introduction to Head-Related Transfer Functions (HRTF's) and their role in the synthesis of spatial sound over headphones. We define HRTF's and discuss how they provide spectral cues which reduce the ambiguity with which the classical duplex theory decodes a free-field sound's spatial location. We describe how HRTF's are measured and how they are typically used to synthesize spatialized sound. By comparing and contrasting representations of HRTF's in the time, frequency, and spatial domains, we highlight different analytic and signal processing techniques that have been used to investigate the structure of HRTF's.

1. INTRODUCTION AND BACKGROUND

The ability of humans to use sonic cues to estimate the spatial location of a target is of great practical and research importance. Recently, advances in computational power and acoustic measurement techniques have made it possible to empirically measure, analyze, and synthesize the spectral cues which influence spatial hearing. These spectral cues are called Head-Related Transfer Functions (HRTF's), and are the focus of much engineering and psychoacoustic research. This tutorial provides an introduction to the role of HRTF's in spatial hearing and provides an introduction to the structure of HRTF data.

This paper is organized as follows. Section 1.1 introduces duplex theory, a simple model introduced by Lord Rayleigh to explain directional hearing in the azimuthal (left-right) direction [64]. Despite its weaknesses, duplex theory provides a good foundational model to which we can compare and contrast HRTF/spectrally based models for directional hearing. Sections 1.2, 1.2.1, 1.2.2, and 1.2.3 define HRTF's, explain typical measurement procedures for HRTF's, discuss synthesis of spatial audio with HRTF's, and review some problems with current headphone-based spatial synthesis systems, respectively. Sections 2.1-2.3 provide an introduction to the structure of HRTF's by using the time, frequency, and spatial domains to compare HRTF's empirically measured from human subjects and HRTF's computed from an analytical, rigid sphere model of the head. Cross-comparison of HRTF data in these domains highlights two well-known structures in HRTF data: diffraction effects due to the head, and elevation effects.

1.1 Duplex Theory

Duplex theory is a model for estimating a free-field target's spatial location by two binaural cues: inter-aural time differences and interaural intensity differences [7][64]. An inter-aural time difference (ITD) is defined as the difference in arrival times of a sound's wavefront at the left and right ears. Similarly, an inter-aural difference (IID) is defined as the amplitude difference generated between the right and left ears by a sound in the free field.

It has been shown that both ITD's and IID's are important parameters for the perception of a sound's location in the azimuthal plane, e.g., perception of sound in the "left – right" direction [7]. In general, a sound is perceived to be closer to the ear at which the first wavefront arrives, where a larger ITD translates to a larger lateral displacement. In other words, for pure sinusoids, perceived lateral displacement is proportional to the phase difference of the received sound at the two ears. However, at approximately 1500 Hz, the wavelength of a sinusoid becomes comparable to the diameter of the head, and ITD cues for azimuth become ambiguous. At these frequencies and above, ITD's may correspond to distances that are longer than one wavelength. Thus, an aliasing problem occurs above 1500 Hz, and the difference in phase no longer corresponds to a unique spatial location, as can be seen in Figure 1 and Figure 3.

At frequencies above 1500 Hz, the head starts to shadow the ear farther away from the sound, so that less energy arrives at the shadowed ear than at the non-shadowed ear. The difference in amplitudes at the ears is the IID, and has been shown to be perceptually important to azimuth decoding at frequencies above 1500 Hz. The relationship of perceived location does not vary linearly with IID alone, as there is a strong dependence on frequency in this case. However, for a given frequency, perceived azimuth does vary approximately linearly with the logarithm of the IID [7]. See Figure 2 for details.

Although the simplicity and success of duplex theory are attractive, the theory only explains the perception of azimuth, or "left-right" displacement. If one attempts to apply Rayleigh's duplex theory to the estimation of a sound's location in free space where the sound is allowed to vary in elevation and distance, ITD and IID cues do not specify a unique spatial position, as there are an infinite number of locations along curves of equal distance from the observer's head which have the same associated ITD and/or IID. This ambiguity was noted by Hornbostel and Wertheimer in 1920, who loosely described the locus of all points sharing the same ITD as resembling a cone in the far field. This set of points is often called the "cone of confusion," since the location of all sounds originating from points on this cone are indistinguishable according to duplex theory. Figure 4 shows the cone of confusion.

The problem is acute in the median plane, which separates the two ears and runs vertically through the head. For a sound originating from any point in this plane, the IID and ITD are zero for an ideal model of the head, so that interaural difference information is at a minimum. However, because listeners can differentiate sounds originating from points in this plane, several authors have suggested that the ability to localize in the median plane is evidence for a monaural hearing mechanism which relies on the spectral coloration of a sound produced by the torso, head, and external ear, or pinna. For example, Blauert conducted a series of experiments which concluded that for a subject to correctly locate a sound's elevation in the median plane, the sound must be broadband, and contain frequencies at 7 kHz and above [7].

It is widely thought that the auditory system uses ITD, IID, and spectral cues to determine spatial location at all spatial positions, not just positions in the median plane. However, while psychoacoustic experiments have verified the relatively simple linear relationship between ITD and IID and perceived lateral displacement, the relationship between spectral content and perceived spatial location is not as simple. What spectral contours, frequencies, or combinations of frequencies correspond with which locations in space?

1.2 Head-Related Transfer Functions (HRTF's)

Whereas the linear relationship between ITD, IID, and perceived spatial location is reasonable to predict, there is less intuition as to how spectral structure and perceived spatial location relate mathematically. Consequently, as a first step toward understanding spectral cues in directional hearing, many researchers have tried to physically model [65], empirically measure [77], or more recently computationally simulate [42] the direction-dependent frequency response of the ear directly. These measurements are called Head-Related Transfer Functions (HRTF's), and summarize the direction-dependent acoustic filtering a free field sound undergoes due to the head, torso, and pinna. In this manner, researchers expect first to record the frequency response of the ear, and then analyze and uncover the data's perceptual structure later.

Formally, a single Head-Related Transfer Function (HRTF) is defined to be a specific individual's left or right ear far-field frequency response, as measured from a specific point in the free field to a specific point in the ear canal. Typically, HRTF's are measured from humans or mannequins for both the left and right ears at a fixed radius from

the listener's head. HRTF's are measured at several different azimuths (left-right direction) and elevations (up-down direction), which are both measured in degrees or radians. Figure 5 contains some relevant terminology, and depicts the spatial coordinate system used in much of the HRTF literature.

HRTF's are commonly specified as minimum phase FIR filters. Note that an HRTF subsumes both ITD and IID information: time delays are encoded into the filter's phase spectrum, and IID information is related to the overall power of the filter. However, HRTF's have been found empirically to be minimum-phase systems [45], which allows us to simplify the FIR specification of HRTF's in two important ways: 1) The minimum phase assumption allows us to uniquely specify an HRTF's phase by its magnitude response alone. This is because the log magnitude frequency response and the phase response of a minimum phase causal system form a Hilbert transform pair [56]. 2) The minimum phase assumption allows us to separate ITD information from the FIR specification of HRTF's. Since minimum phase filters have the minimum group delay property and minimum energy delay property [56], most of an HRTF's energy occurs at the beginning of its impulse response, so that the left and right ear minimum phase HRTF's both have zero delay. Thus, complete characterization of the auditory cues associated with a single spatial location involves the measurement of three quantities: left and right ear magnitude responses and the ITD.

1.2.1 Measurement of HRTF's

A common technique used to empirically measure right and left ear HRTF's is to insert probe tube microphones partially into a subject's ears, and then to perform a simple form of system identification by playing a known-spectrum stimulus through a loudspeaker placed at a specified azimuth θ, elevation ϕ, and distance from the subject's head [77]. In practice, the stimulus may be a simple click, pseudo-random binary sequences, or complementary Golay codes. Portions of the measured transfer functions due to the measurement apparatus, such as the microphone and speaker transfer functions, along with portions of the measured transfer functions which are the same for all locations, are called the common transfer function (CTF), and are removed from the raw measurements. The final result is the directional transfer function (DTF) at azimuth θ and elevation ϕ. The DTF is the quantity which contains spectral cues responsible for spatial hearing, and is often informally called the HRTF in much literature.

Mathematically, we can relate the DTF, CTF, stimulus signal, and measured signals as follows. Let *s(t)* be the known stimulus signal presented at azimuth θ and elevation ϕ. Let *c(t)* be the known CTF, $d_{l,\theta,\phi}(t)$ and $d_{r,\theta,\phi}(t)$ be the unknown left and right ear DTF's respectively; let $m_{l,\theta,\phi}(t)$ and $m_{r,\theta,\phi}(t)$ be the signals recorded from the left and right ear probe tube microphones, respectively. Then

$$m_{l,\theta,\phi}(t) = s(t) * c(t) * d_{l,\theta,\phi}(t) \qquad m_{r,\theta,\phi}(t) = s(t) * c(t) * d_{r,\theta,\phi}(t) \tag{1}$$

$$M_{l,\theta,\phi}(\omega) = S(\omega)C(\omega)D_{l,\theta,\phi}(\omega) \qquad M_{r,\theta,\phi}(\omega) = S(\omega)C(\omega)D_{r,\theta,\phi}(\omega) \tag{2}$$

Here, we assume that *c(t)* is spatially invariant, and can be computed from known measurements of the recording apparatus and spectrally averaged values of $m_{l,\theta,\phi}(t)$ and $m_{r,\theta,\phi}(t)$ for several locations. Hence, the left and right DTF's can be computed as follows.

$$\left|D_{l,\theta,\phi}(\omega)\right| = \frac{\left|M_{l,\theta,\phi}(\omega)\right|}{\left|S(\omega)\right|\left|C(\omega)\right|} \qquad \left|D_{r,\theta,\phi}(\omega)\right| = \frac{\left|M_{r,\theta,\phi}(\omega)\right|}{\left|S(\omega)\right|\left|C(\omega)\right|} \tag{3}$$

$$\angle D_{l,\theta,\phi}(\omega) = \angle M_{l,\theta,\phi}(\omega) - \angle S(\omega) - \angle C(\omega) \qquad \angle D_{r,\theta,\phi}(\omega) = \angle M_{r,\theta,\phi}(\omega) - \angle S(\omega) - \angle C(\omega) \tag{4}$$

$$D_{l,\theta,\phi}(\omega) = \left|D_{l,\theta,\phi}(\omega)\right| \exp(j\angle D_{l,\theta,\phi}(\omega)) \qquad D_{r,\theta,\phi}(\omega) = \left|D_{r,\theta,\phi}(\omega)\right| \exp(j\angle D_{r,\theta,\phi}(\omega)) \tag{5}$$

$$d_{l,\theta,\phi}(t) = F^{-1}(D_{l,\theta,\phi}(\omega)) \qquad d_{r,\theta,\phi}(t) = F^{-1}(D_{r,\theta,\phi}(\omega)) \tag{6}$$

Phase information from the computed time domain DTF's is used to compute ITD information associated with azimuth θ and elevation ϕ. Specifically, let $d_{l,\theta,\phi}(n)$ and $d_{r,\theta,\phi}(n)$ be the discrete, time-domain DTF's for the left and right ears corresponding to azimuth θ and elevation ϕ. Then the ITD $n_{ITD,\theta,\phi}$ is computed as the lag for which the cross-correlation function between $d_{l,\theta,\phi}(n)$ and $d_{r,\theta,\phi}(n)$ is maximized.

$$n_{ITD,\theta,\phi} = \arg\max_{\tau} \sum_{n} d_{l,\theta,\phi}(n) d_{r,\theta,\phi}(n+\tau) \tag{7}$$

After computing the ITD, minimum phase versions of the DTF's are often computed by windowing the real cepstrum of $d_{l,\theta,\phi}(n)$ and $d_{r,\theta,\phi}(n)$ [56][37]. Define the window

$$w(n) = \begin{cases} 1, & n = 1 \text{ or } n = N/2+1 \\ 2, & n = 2,\ldots,n = N/2 \\ 0, & n = N/2+2,\ldots,N \end{cases} \tag{8}$$

where N is an even length HRTF filter length. Compute the windowed cepstrum of $d_{l,\theta,\phi}(n)$ and $d_{r,\theta,\phi}(n)$

$$\begin{aligned} c_{l,\theta,\phi}(n) &= F^{-1}\left\{\log\left|F\left\{d_{l,\theta,\phi}(n)\right\}\right|\right\} = F^{-1}\left\{\log\left|D_{l,\theta,\phi}(k)\right|\right\} \\ c_{r,\theta,\phi}(n) &= F^{-1}\left\{\log\left|F\left\{d_{r,\theta,\phi}(n)\right\}\right|\right\} = F^{-1}\left\{\log\left|D_{r,\theta,\phi}(k)\right|\right\} \end{aligned} \tag{9}$$

$$\begin{aligned} \hat{c}_{l,\theta,\phi}(n) &= c_{l,\theta,\phi}(n) w(n) \\ \hat{c}_{r,\theta,\phi}(n) &= c_{r,\theta,\phi}(n) w(n) \end{aligned} \tag{10}$$

Finally, the minimum phase versions of $d_{l,\theta,\phi}(n)$ and $d_{r,\theta,\phi}(n)$ are given by

$$\begin{aligned} \mathrm{dmin}_{l,\theta,\phi}(n) &= F^{-1}\left\{\exp\left(F\left\{\hat{c}_{l,\theta,\phi}(n)\right\}\right)\right\} \\ \mathrm{dmin}_{r,\theta,\phi}(n) &= F^{-1}\left\{\exp\left(F\left\{\hat{c}_{r,\theta,\phi}(n)\right\}\right)\right\} \end{aligned} \tag{11}$$

1.2.2 Synthesis of Spatial Audio using HRTF's

Although all of the perceptually salient structures of empirically measured HRTF's are not yet fully understood, raw HRTF's have already been used extensively to synthesize spatialized sounds over headphones. Presumably, the left HRTF, right HRTF, and ITD associated with a specific location completely characterize the acoustic filtering of a sound originating from that location. Thus, assuming that the auditory system associates these quantities with a particular spatial location, HRTF's and ITD's can be used to filter a monaural sound into a binaural sound which will sound as though it originates from that location.

Formally, suppose that one wants to process a monaural signal $s(n)$ such that it sounds as if it is located at azimuth θ and elevation ϕ. Specifically, let $\mathrm{dmin}_{l,\theta,\phi}(n)$ and $\mathrm{dmin}_{r,\theta,\phi}(n)$ be the minimum phase impulse responses measured at azimuth θ and elevation ϕ which have magnitude responses $|D_{l,\theta,\phi}(k)|$ and $|D_{r,\theta,\phi}(k)|$ respectively. Construct two sounds $s_l(n)$ and $s_r(n)$ as follows, and present $s_l(n)$ and $s_r(n)$ to the left and right ears simultaneously over headphones. Here, $n_{ITD,\theta,\phi}$ is defined to be negative for sounds arriving at the left ear first.

$$s_l(n) = s(n - n_{ITD,\theta,\phi}) * \mathrm{dmin}_{l,\theta,\phi}(n) \tag{12}$$

$$s_r(n) = s(n) * \mathrm{dmin}_{r,\theta,\phi}(n) \tag{13}$$

Several applications involve real time synthesis of spatial audio in which the sound source moves over time. Thus, in practice, high-speed DSP hardware is used to implement the convolutions in (12) and (13), while delay lines are used to implement the time delay in (12). In order to synthesize moving sounds, HRTF's and ITD's are dynamically updated in time to correspond to new spatial locations. Figure 6 shows a block diagram of a simple real-time headphone-based spatialization system.

1.2.3 Problems with HRTF-based Synthesis of Spatial Audio

Although the theory of using HRTF's to synthesize spatial audio is simple, there are still several problems that occur in practice. For example, simple HRTF-based spatialization algorithms such as the one in Figure 6 do not always produce sounds with the intended spatialization effects. Subjects often report that there is a lack of "presence" in spatially-synthesized sounds – sounds spatialized near the median plane (0° azimuth) sound as though they are "inside" the head instead of "outside" the head [32]. Sounds processed to sound as though they originate from the front of a listener actually sound like they originate from in back of the listener (the so –called "front-back" confusions) [78]. Synthesis of sounds with non-zero elevations is difficult. Also, since every individual has a unique set of HRTF's, a subject listening to a spatialized sound generated from a "generalized" HRTF set may not perceive the sound in the intended spatial location [76][62].

In addition to sound quality problems, HRTF-based sound synthesis faces several computational challenges as well. For example, dedicated real-time DSP hardware is often needed to implement even the simplest spatialization algorithms, so that high quality synthesis of virtual audio on low-cost generic computers is not often possible. Because HRTF's are typically measured at several hundred different spatial locations, there is a substantial amount of data that needs to be stored, accessed, and processed quickly. In moving sound synthesis as depicted in Figure 6, several "interpolated" HRTF's and ITD's may be required to produce a smoothly moving sound – how should we compute these interpolated quantities from a finite set of existing HRTF's and ITD's [19][8][48][34]?

Many researchers believe that the solution to the above problems involves a deeper understanding of the perceptual structure of HRTF data. By investigating, modeling, and parameterizing the structure of HRTF's, researchers expect to link salient features of HRTF's, such as peaks and dips in the magnitude responses and the impulse responses, to specific spatial parameters, such as azimuth, elevation, and distance. Future spatial audio synthesis algorithms could exploit this perceptual information to preprocess HRTF's in order to alleviate problems with existing systems. For example, a low order parameterization of HRTF's that maintains a sufficient number of perceptually relevant spectral cues could significantly lighten the computational and storage demands on hardware; understanding and pre-emphasizing the spectral cues corresponding to sounds in front of and in back of a subject could reduce front-back confusions; etc.

2. DIFFERENT REPRESENTATIONS OF HRTF DATA

There have been many attempts to understand the structure of HRTF's by displaying and analyzing HRTF data sets in the time, frequency, and spatial domains. In order to compare and contrast these different representations of HRTF data, we compare and contrast two different HRTF data sets – HRTF data empirically measured from a human subject, and HRTF data computed from a purely mathematical, spherical model of the head. In addition, in order to demonstrate some specific structures in HRTF data, we describe two well-known HRTF structures: head-related diffraction effects and elevation-related effects. We show how these structures can be found in each of the HRTF data sets to some degree, in all of the time, frequency, and spatial domains. Each type of data set and each of the structures to be investigated are summarized below:

1. Types of HRTF data

A. HRTF's derived from a spherical model of the head

The simplest analytical model of HRTF's is derived from a spherical model of the head. By solving the acoustic wave equation corresponding to a monotonic plane wave incident on a rigid sphere, we can compute the resultant pressure produced at the surface of the sphere [60]. In particular, we can compute the pressure at the two points on the sphere's surface which correspond to the left and right ears. By evaluating the pressure at these two points for different frequencies and different incident angles of the plane wave, we can systematically compute the left and right HRTF's at any given frequency and spatial location. Since the spherical model of the head is a mathematical construct, many authors have compared these theoretical HRTF's to measured HRTF's [21][65]. By looking for similarities between theoretical and measured data sets, researchers hope to learn how much of HRTF structure is due only to the head effects. A mathematical derivation of HRTF's derived from the spherical head model can be found in [7], and a recent article by Duda and Martens provides pseudo-code for the numerical computation of the "theoretical" HRTF's used in this paper [21]. In this study, the left and right ears are located at (azimuth, elevation) (-100°, 5°) and (+100°, +5°), respectively. The sound source is located 1.5 m from the head, which is assumed to have a radius of 10 cm, and the speed of sound is assumed to be 330 m/s.

B. HRTF's measured from a human subject

HRTF's were measured for a several subjects by Dr. John Middlebrooks at the Kresge Hearing Research Institute at the University of Michigan, using in-ear, probe-tube microphones. Measurements were taken in an anechoic chamber using the method described in 1.2.1 above. Complementary Golay codes were used as the stimulus signals and were presented from a loudspeaker approximately 1.5m from the subjects' heads. Left and right ear magnitude responses were measured at 400 different azimuth-elevation locations. Although irregularly spaced, these locations are roughly 10°-15° apart in the azimuth and elevation directions. The sampling rate was 50 kHz, the resolution of the data taken was 16 bits, and a 512-point FFT was used to compute the frequency response at each location.

2. Two well-known structures in HRTF data

A. Diffraction effects in the contralateral HRTF's due to the head

The spherical model of the head predicts HRTF's which exhibit diffraction effects. Specifically, for some frequencies and incident angles, the sphere has an amplifying effect on an incident plane wave at certain points near the sphere due to diffraction [60]. Surprisingly, there are some locations on the contralateral side of the head where this effect occurs, even though the head directly "blocks" or shadows the contralateral ear. Shaw refers to these contralateral locations as a "bright spot" in his analyses, since there is a local maximum in energy transmission that occurs in these areas on the contralateral side of the head [65]. An analytical derivation of the effect can be found in [7].

B. Elevation effects in the ipsilateral HRTF's due to the pinna

Spectral cues corresponding to elevation are thought to be related to the pinna, or external ear. Consequently, frequencies near 6-8 kHz are thought to be important for elevation decoding, since these frequencies have wavelengths which are similar to characteristic lengths of the pinna, and therefore interact strongly with the pinna [21]. There are noticeable patterns in HRTF data near these frequencies which have been shown psychophysically to be correlated with the perception of elevation [48].

2.1 Frequency domain representation of HRTF data

HRTF data is perhaps most easily understood in the frequency domain, where the magnitude response of various HRTF's are plotted as a function of frequency. Many studies have attempted to visualize HRTF data by examining how certain macroscopic properties of HRTF sets, such as peaks, notches, or other spectral shapes in particular locations of the magnitude frequency responses, associate and/or systematically vary with the perception of azimuth and/or elevation [7][43][44]. Consequently, many signal processing techniques have taken advantage of this domain in attempts to parameterize or compute interpolated HRTF's. For example, Principal Components Analysis of frequency-domain HRTF's has been performed [44], and pole-zero modeling of frequency-domain HRTF's has been attempted [9]. In addition, there have been several different schemes introduced to calculate interpolated HRTF's in the frequency domain, and a good summary is provided in [34].

The frequency domain representation of HRTF's clearly shows some major differences between the theoretical and measured HRTF data sets. Figures 7-10 show left and right ear frequency domain HRTF's computed from a spherical model of the head, as well as HRTF's measured from a human subject. Figures 7,9 and Figure 8,10 show HRTF's for locations in the horizontal and median planes (elevation = 0°, azimuth = 0°), respectively. Whereas the theoretical HRTF's are noiseless, the Signal to Noise Ratio (SNR) of the measured HRTF's seems to be a function of spatial location. Specifically, the contralateral HRTF's in Figure 9 and Figure 10 seem to be less smooth than the ipsilateral HRTF's, suggesting that SNR for ipsilateral HRTF's is generally higher than that for contralateral HRTF's. This is reasonable, since the contralateral ear typically receives less power than the ipsilateral ear during the measurement process. In general, the measured HRTF's are also more complex than the theoretical HRTF's in that there are several secondary peaks and notches in the magnitude spectra that are not found in the theoretical data set. These extra features are due to the filtering of the pinna and torso, which are not predicted in the spherical head model of HRTF's.

Elevation effects can also be seen in the frequency domain. For example, in the measured data in Figure 10, there is a notch at 7 kHz that migrates upwards in frequency as elevation increases. A shallow peak can be seen at 12 kHz for lower elevations in the median plane, and this peak "flattens out" for higher elevations. The theoretical data in Figure 8 also shows how the head alone can produce elevation-dependent features in HRTF's. Figure 8 shows that higher elevations have a slight high-pass characteristic, while lower elevations have a slight low-pass characteristic.

Diffraction effects can be seen most easily in Figure 7, where theoretical HRTF's are plotted in the frequency domain as a function of azimuth. Note the rippling shapes in the contralateral HRTF's corresponding to azimuths +90° and -90° for the left and right ears, respectively. The HRTF's corresponding to azimuths 127° to 60° for the left ear and –127° to –60° for the right ear contain a low frequency "main lobe" that attains its greatest width at azimuths 90° and -90° respectively. This main lobe is representative of an amplification effect the head has on lower frequencies due to diffraction on the contralateral side. These effects can also be seen in the measured HRTF's in Figure 9. In addition, amplification effects are expected and can be seen for high frequencies in ipsilateral HRTF's due to reflection and the ear's proximity to a rigid surface (e.g. the head) [60].

2.2 Time domain representation of HRTF data

The time domain version of the HRTF is the FIR filter which is computed by performing an inverse FFT on the frequency domain HRTF. Time domain HRTF's are sometimes called Head-Related Impulse Responses (HRIR's) in some literature [21][82]. HRIR's are in a form which can be inserted directly into sound spatialization systems such as the one shown in Figure 6. Because the complexity of a spatialization system depends largely on the length of the HRIR's, there has been some attempt to minimize the length of the HRIR while still preserving important spatially-related spectral cues. In addition, some researchers have smoothed (low-pass filtered) HRIR's in an effort to reduce the noise which the HRTF measurement process inevitably introduces [13].

Figures 11-14 show left and right ear HRIR's computed from a spherical model of the head, as well as HRIR's measured from a human subject. Figures 11, 12 and Figures 13,14 show HRIR's for locations in the horizontal and median planes (elevation = 0°, azimuth = 0°), respectively. Figure 11 and Figure 12 show that in general, a location which is farther away from the ipsilateral ear in azimuth and elevation has a corresponding HRIR which has a lower-

amplitude initial peak that occurs later in time. This is consistent with duplex theory, which predicts larger ITD's and IID's for sources with a larger absolute azimuth, or displacement from the median plane. In addition, The pronounced negative dip, or "overshoot" in some HRIR's in Figure 11 and Figure 12 indicates that high frequencies are boosted for these locations [21].

Diffraction effects and Shaw's "bright spot" can also be easily seen in Figure 11 and Figure 12. HRIR's corresponding to contralateral locations which lie in the direct shadow of the head have relatively higher-amplitude initial peaks. For example, Figure 11 shows how left-ear HRIR's associated with azimuths of +90 have relatively large amplitude initial peaks, even though these locations lie in the direct shadow of the head. One can also see elevation-related effects in the HRIR's of Figure 13 and Figure 14, as there is a slight difference in arrival times of for positive and negative elevation HRIR's.

Comparison of the theoretical and empirical data sets in each of Figures 11-14 reveals that although the general patterns are similar between the two data sets, measured data is much richer than the theoretical data. The measured HRIR's contain many secondary peaks in addition to the initial peak, which the theoretical HRIR's do not have. Although these effects could be related to the inherently noisy measurement process, these effects are also related to the complex acoustic structure of the outer ear, which a theoretical spherical model of the head only cannot predict.

2.3 Spatial domain representation of HRTF data [1]

HRTF's can be represented in the spatial domain in several different ways. Some authors plot ITD's and IID's as a surface, as a function of azimuth and elevation [82][14]. Others plot frequency domain HRTF's with a common azimuth or elevation as a surface, where the sequential layout of the data by elevation or azimuth, respectively, highlights patterns effectively [11].

In this paper, we focus on spatial representations which plot the magnitude response of all HRTF's in a data set for a fixed frequency as a function of azimuth and elevation. The current authors call such spatial representations of HRTF's Spatial Frequency Response Surfaces (SFRS's) [18][19]. Intuitively, these graphs indicate how much energy the right and left ears receive at a fixed frequency as a function of spatial location. Typically, SFRS's exhibit several local maxima, or "hotspots" which correspond to spatial areas from which the ears receive more energy than others. Some examples of SFRS's computed from both measured and theoretical HRTF's can be found in Figures 15-18. Other examples of this style of HRTF representation can be found in [14][34][18][19].

There have been several attempts to process HRTF data in the spatial domain. For example, principal components analysis (PCA) of HRTF's has been performed in the spatial domain as an attempt to parameterize HRTF's [14]. Spherical basis functions have also been used to parameterize HRTF's in the spatial domain [41]. There have been attempts to parameterize HRTF's in the spatial domain using a beamformer, where a virtual sensor array models the spatial and temporal characteristics of HRTF's simultaneously [15]. Spatially-based HRTF interpolation methods have also been developed which produce perceptually reasonable HRTF's [48][34][19].

Elevation perception can be linked to some of the local maxima or hotspots in SFRS's at specified frequencies. An SFRS with one dominant hotspot might suggest that the auditory system favors the location of that hotspot perceptually when presented with narrowband noise centered at that frequency. This is consistent with the theory of directional bands, which states that certain narrowband signals are correlated with perceptually preferred spatial directions. For example, the 6.8 kHz SFRS's in Figure 17 corresponding to measured HRTF's contain dominant hotspots which are positive in elevation, near (azimuth, elevation) (-90°, +20°) and (+90°, +20°) for the left and right ears, respectively. Similarly, the 8.7 kHz SFRS's in Figure 18 corresponding to measured HRTF's contain dominant hotspots which are negative in elevation, near (azimuth, elevation) (-90°, -30°) and (+90°, -30°) for the left and right ears, respectively. Therefore, one might guess that subjects listening to narrowband sounds centered at 6.8 and 8.7 kHz would perceive the sounds as coming from above or below them, respectively. Indeed, one psychophysical study designed to test the theory of directional bands for several frequencies found that subjects tended to localize narrowband sounds centered at 6 and 8 kHz as originating from positive and negative elevations

[1] This section contains some material previously discussed in [18] and [19].

respectively, regardless of the actual free-field location of the source [48]. Thus, the theory of directional bands is consistent with the perception of elevation in this case, and is illustrated by the hotspots in the 6.8 and 8.7 kHz SFRS's.

Diffraction effects are easily seen in SFRS's computed from both measured and theoretical HRTF's. The 1.9 kHz SFRS and 2.4 kHz SFRS's in Figure 15 and Figure 16 both contain a "hotspot" on the contralateral side, near (azimuth, elevation) (90°, 10°) and (-90°, 10°) for the left and right ears, respectively. This is the well-known "bright spot" that Shaw refers to in his analyses [65].

Comparison between theoretical and measured SFRS's again shows that the measured data is much richer than the theoretical data. Figures 15-18 show that the measured data has several hotspots in SFRS's which are not found in the theoretical data. Furthermore, the measured data in Figure 15 shows a local minimum on the contralateral side at lower elevations, unlike the theoretical data in Figure 15. This minimum is caused by torso shadowing, and can be found at (azimuth, elevation) (+100°, -40°) and (-100°,-40°) for the left and right ears, respectively. These observations reinforce the fact that the spherical model of the head cannot predict the effects of the pinna and torso, which are responsible for the added detail in the SFRS's computed from measured data.

3. CONCLUSIONS

This paper introduced HRTF's and discussed their role in the synthesis of spatial audio over headphones. The need for spectral cues / HRTF's was motivated by an inability of duplex theory to resolve spatial locations uniquely from ITD's and IID's alone. We discussed the measurement of HRTF's from human subjects and HRTF-based synthesis of spatial audio over headphones. We reviewed several sound quality and computational limitations of current spatial audio synthesis systems including the high computational power and storage requirements for such systems, lack of externalization of spatialized sounds, and "front-back" confusions of spatialized sounds.

This paper also compared HRTF's measured from a human subject and HRTF's computed from an analytic spherical model of the head. We examined these data sets in the time, frequency, and spatial domains, and highlighted several signal processing techniques that take advantage of each domain to parameterize, interpolate, or otherwise model salient HRTF structures. By further examining these data sets in the time, frequency, and spatial domains, we were able to see two well-known structures in HRTF data: diffraction effects due to the head and elevation effects. Finally, measured HRTF's were more complex and contained many pinna and torso related effects which the theoretical HRTF's did not contain.

4. ACKNOWLEDGEMENTS

The authors thank Dr. John C. Middlebrooks at the Kresge Hearing Research Institute of the University of Michigan for providing the data used in this research. We also thank Dr. Michael A. Blommer for his early investigations of the interpolation problem. This research was supported by a grant from the Office of Naval Research in collaboration with the Naval Submarine Medical Research Laboratory, Groton, Connecticut.

5. REFERENCES

[1] Abel, Sharon and Hay, Valerie H. "Sound localization: The interaction of aging, hearing loss, and hearing protection." *Scandinavian Audiology*, **25**(1): 1996.

[2] Avendano et al. "Modeling the Contralateral HRTF." *Proceedings of the 16th Audio Engineering Society (AES) International Conference on Spatial Sound Reproduction*, Rovaniemi, Finland: 1999.

[3] Begault, Durand R. *3-D Sound for Virtual Reality and Multimedia*. Academic Press, Inc. Cambridge, Massachusetts: 1994.

[4] Begault, Durand R. and Wenzel, Elizabeth M. "Headphone localization of speech." *Human Factors*, **35**(2): 1993.

[5] Bernstein, Leslie R. and Constantine Trahiotis. “Binaural beats at high frequencies: Listeners’ use of envelope-based interaural temporal and intensitive disparities.” *Journal of the Acoustical Society of America*, **99**(3): March 1996.

[6] Bernstein, Leslie R. and Trahiotis, Constantine. “Binaural interference effects measured with masking-level difference and with ITD- and IID-discrimination paradigms.” *Journal of the Acoustical Society of America*, **98**(1): July 1995.

[7] Blauert, Jens. *Spatial Hearing*. The MIT Press, Cambridge: 1983.

[8] Blommer, A. and Wakefield, G. “A comparison of head related transfer function interpolation methods.” *1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics.* (IEEE catalog number: 95TH8144).

[9] Blommer, Michael Alan. *Pole-zero Modeling and Principal Component Analysis of Head-Related Transfer Functions.* Ph.D. Dissertation for the University of Michigan, Dept. of Electrical Engineering and Computer Science, Systems division, Ann Arbor, Michigan: 1996.

[10] Buell, Thomas N. et al. “Lateralization of bands of noise as a function of combinations of interaural intensitive differences, interaural temporal differences, and bandwidth.” *Journal of the Acoustical Society of America*, **95**(3): March 1994.

[11] Carlile, S. and Pralong, D. “The location-dependent nature of perceptually salient features of the human head-related transfer functions.” *Journal of the Acoustical Society of America*, **95**(6): June 1994.

[12] Chandler, David W. and Grantham, Wesley. “Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity.” *Journal of the Acoustical Society of America*, **91**(3): March 1992.

[13] Chapin, William L. Personal Communication: 1999.

[14] Chen, Jiashu et al. “A spatial feature extraction and regularization model of the head-related transfer function.” *Journal of the Acoustical Society of America*, **97**(1): January 1995.

[15] Chen, Jiashu et al. “External ear transfer function modeling: A beamforming approach.” *Journal of the Acoustical Society of America*, **92**(4) Pt. 1: October 1992.

[16] Chen, Jiashu et al. “Representation of external ear transfer function via a beamforming model.” *Proceedings of the 1991 International Conference on Acoustics, Speech, and Signal Processing (ICASSP).* New York (IEEE catalog number 91CH2977-7): 1991.

[17] Chen, Jiashu et al. “Synthesis of 3D virtual auditory space via a spatial feature extraction and regularization model.” *Proceedings of the IEEE Virtual Reality Annual International Symposium*. Seattle, Washington (IEEE catalog number 93CH3336-5): 1993.

[18] Cheng, Corey I. and Wakefield, Gregory H. “Spatial Frequency Response Surfaces: An Alternative Visualization Tool for Head-Related Transfer Functions (HRTF’s).” *Proceedings of the 1999 International Conference on Acoustics, Speech, and Signal Processing (ICASSP99)*, Phoenix, Arizona: 1999.

[19] Cheng, Corey I. and Wakefield, Gregory H. “Spatial Frequency Response Surfaces (SFRS’s): An Alternative Visualization and Interpolation Technique for Head-Related Transfer Functions (HRTF’s).” *Proceedings of the 16th Audio Engineering Society (AES) International Conference on Spatial Sound Reproduction*, Rovaniemi, Finland: 1999.

[20] Dasarathy, Belur V. *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*. IEEE Computer Society Press, Los Alamitos, California: 1991.

[21] Duda, Richard O. and Martens, William M. “Range dependence of the response of a spherical head model.” *Journal of the Acoustical Society of America*, 104(5): November 1998.

[22] Duda, Richard O. et al. “An Adaptable Ellipsoidal Head Model for the Interaural Time Difference.” *Proceedings of the 1999 International Conference on Acoustics, Speech, and Signal Processing (ICASSP99)*, Phoenix, Arizona: 1999.

[23] Endsley, Mica R. and Rosiles, Armida S. “Auditory localization for spatial orientation.” *Journal of Vestibular Research*, **5**(6): 1995.

[24] Fay, Richard R. and Popper, Arthur N. *Comparative Hearing: Mammals*. Springer-Verlag, New York: 1994.

[25] Fisher, N.I. et al. *Statistical Analysis of Spherical Data*. Cambridge University Press, New York: 1987.

[26] Fisher, N.I. *Statistical Analysis of Circular Data*. Cambridge University Press, New York: 1993.

[27] Giguère, Christian and Abel, Sharon M. “Sound localization: Effects of reverberation time, speaker array, stimulus frequency, and stimulus rise/decay.” *Journal of the Acoustical Society of America*, **94**(2) Pt. 1: August 1993.

[28] Gilkey, Robert H. “Some considerations for the design of auditory displays.” *1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics.* (IEEE catalog number: 95TH8144).

[29] Gilkey, Robert H. and Anderson, Timothy R. “The accuracy of absolute localization judgments for speech stimuli.” *Journal of Vestibular Research*, **5**(6): 1995.

[30] Golay, Marcel J.E. “Complementary Series.” *IRE Transactions on Information Theory* **7**: 1961.

[31] Green, David M. and Swets, John A. *Signal Detection Theory and Psychophysics*. Robert E. Krieger Publishing Company, New York: 1974.

[32] Griesinger, David. “Objective Measures of Spatiousness and Envelopment.” *Proceedings of the 16th Audio Engineering Society (AES) International Conference on Spatial Sound Reproduction*, Rovaniemi, Finland: 1999.

[33] Hartmann, William Morris and Rakerd, Brad. “On the minimum audible angle – A decision theory approach.” *Journal of the Acoustical Society of America*, **85**(5): May 1989.

[34] Hartung, Klaus et al. “Comparison of Different Methods for the Interpolation of Head-Related Transfer Functions.” *Proceedings of the 16th Audio Engineering Society (AES) International Conference on Spatial Sound Reproduction*, Rovaniemi, Finland: 1999.

[35] Hawkins, Harold L. et al, eds. *Auditory Computation*. Springer-Verlag, Inc., New York: 1996.

[36] Heller, Laurie M. and Trahiotis, Constantine. “Extents of laterality and binaural interference effects.” *Journal of the Acoustical Society of America*, **99**(6): June 1996.

[37] Huopaniemi, Jyri, and Smith, Julius O. “Spectral and time-domain preprocessing and the choice of modeling error criteria for binaural digital filters.” *Proceedings of the 16th Audio Engineering Society (AES) International Conference on Spatial Sound Reproduction*, Rovaniemi, Finland: 1999.

[38] Jenison, Rick L. “A spherical basis function neural network for approximating acoustic scatter.” *Journal of the Acoustical Society of America*, **99**(5): May 1996.

[39] Jenison, Rick L. “A spherical basis function neural network for pole-zero modeling of head-related transfer functions.” *1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics.* (IEEE catalog number: 95TH8144).

[40] Jenison, Rick L. and Fissell, Kate. “A Comparison of the von Mises and Gaussian basis functions for approximating spherical acoustic scatter.” *IEEE Transactions on Neural Networks* **6**(5): September 1995.

[41] Jenison, Rick L. and Fissell, Kate. “A spherical basis function neural network for modeling auditory space.” *Neural Computation*, **8**: 1996.

[42] Kahana, Yuvi et al. “Numerical Modelling of the Transfer Functions of a Dummy-Head and of the External Ear.” *Proceedings of the 16th Audio Engineering Society (AES) International Conference on Spatial Sound Reproduction*, Rovaniemi, Finland: 1999.

[43] Kendall, Gary S. “A 3-D sound primer: Directional hearing and stereo reproduction.” *Computer Music Journal*, **19**(4): Winter 1995.

[44] Kistler, Doris J. and Wightman, Frederic L. “A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction.” *Journal of the Acoustical Society of America*, **91**(3): March 1992.

[45] Kulkarni, A. et al. “On the minimum-phase approximation of head-related transfer functions.” *1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics.* (IEEE catalog number: 95TH8144).

[46] Loomis, Jack M. “Some research Issues in Spatial Hearing.” *1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics.* (IEEE catalog number: 95TH8144).

[47] Martin, Keith D. “Estimating azimuth and elevation from interaural difference.” *1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics.* (IEEE catalog number: 95TH8144).

[48] Middlebrooks, John C. “Narrow-band sound localization related to external ear acoustics.” *Journal of the Acoustical Society of America*, **92**(5): November 1992.

[49] Middlebrooks, John C. and Green, David M. “Directional dependence of interaural envelope delays.” *Journal of the Acoustical Society of America*, **87**(5): May 1990.

[50] Middlebrooks, John C. et al. “Directional sensitivity of sound-pressure levels in the human ear canal.” *Journal of the Acoustical Society of America*, **86**(1): July 1989.

[51] Morse, Philip M. *Vibration and Sound*. McGraw-Hill Book Company, Inc., New York: 1948.

[52] Musicant, Alan D. and Butler, Robert A. “Influence of monaural spectral cues on binaural localization.” *Journal of the Acoustical Society of America*, **77**(1): January 1985.

[53] Nandy, Dibyendu and Ben-Arie, Jezekiel. “An auditory localization model based on high-frequency spectral cues.” *Annals of Biomedical Engineering*, **24**: 1996.

[54] Oldfield, Simon R. and Parker, Simon P.A. “Acuity of sound localisation: a topography of auditory space. I. Normal listening conditions.” *Perception*, **13**: 1984.

[55] Oldfield, Simon R. and Parker, Simon P.A. “Acuity of sound localisation: a topography of auditory space. II. Pinna cues absent.” *Perception*, **13**: 1984.

[56] Oppenheim, Alan V. and Schafer, Ronald W. *Discrete-Time Signal Processing*. Prentice Hall, Englewood Cliffs, New Jersey: 1989.

[57] Papoulis, Athanasios. *Probability, Random Processes, and Stochastic Processes*. McGraw-Hill, Inc., New York: 1991.

[58] Perrett, Stephen and Noble, William. “The contribution of head motion cues to localization of low-pass noise.” *Perception and Psychophysics*, **59**(7): 1997.

[59] Perrett, Stephen and Noble, William. “The effect of head rotations on vertical plane sound localization.” *Journal of the Acoustical Society of America*, **102**(4): October 1997.

[60] Pierce, Allan D. *Acoustics*. Acoustic Society of America, Woodbury, New York: 1991.

[61] Powell, M.J.D. “Radial basis functions for multivariable interpolation: a review.” *Algorithms for Approximation*. Oxford University Press, New York: 1987

[62] Pralong, Danièle, and Carlile, Simon. “The role of individualized headphone calibration for the generation of high fidelity virtual auditory space.” *Journal of the Acoustical Society of America*, **100**(6): December 1996.

[63] Rao, K. Raghunath and Ben-Arie, Jezekiel. “Optimal head related transfer functions for hearing and monaural localization in elevation: A signal processing design perspective.” *IEEE Transactions on Biomedical Engineering*, **43**(11): November 1996.

[64] Rayleigh, L. “On our perception of sound direction.” Philosophical Magazine **13**: 1907.

[65] Shaw, E.A.G. “The External Ear.” *Handbook of Sensory Physiology* ***V****/1: Auditory System, Anatomy Physiology(Ear).* Springer-Verlag, New York: 1974.

[66] Speyer, Gavriel, and Furst, Miriam. "A model-based approach for normalizing the head related transfer function." *Proceedings of the 1996 19th Convention of Electrical and Electronics Engineers in Israel* Jerusalem, Israel (IEEE catalog number 96TH8190): 1996.

[67] Stern, Richard et al. "Lateralization of complex binaural stimuli: A weighted-image model." *Journal of the Acoustical Society of America*, **84**(1): July 1988.

[68] Stern, Richard et al. "Lateralization of rectangularly modulated noise: Explanations for counterintuitive reversals." *Journal of the Acoustical Society of America*, **90**(4) Pt. 1: October 1991.

[69] Stern, Richard M. and Shear, Glenn D. "Lateralization and detection of low-frequency binarual stimuli: Effects of distribution of iternal delay." *Journal of the Acoustical Society of America*, **100**(4) Pt. 1: October 1996.

[70] Therrien, Charles W. *Decision, Estimation, and Classification.* John Wiley and Sons, New York: 1989.

[71] Trahiotis, Constantine and Bernstein, Leslie R. "Lateralization of bands of noise and sinusoidally amplitude-modulated tones: Effects of spectral locus and bandwidth." *Journal of the Acoustical Society of America*, **79**(6): June 1996.

[72] Trahiotis, Constantine and Stern, Richard M. "Across-frequency interaction in lateralization of complex binaural stimuli." *Journal of the Acoustical Society of America*, **96**(6): December, 1994.

[73] Van Trees, Harry L. *Detection, Estimation, and Modulation Theory, Part I*. John Wiley and Sons, New York: 1968.

[74] Van Veen, Barry D. and Jenison, Rick L. "Auditory space expansion via linear filtering." *Journal of the Acoustical Society of America*, **90**(1): July 1991.

[75] Wenzel, Elizabeth M. "The relative contribution of interaural time and magnitude cues to dynamic sound localization." *1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics.* (IEEE catalog number: 95TH8144).

[76] Wenzel, Elizabeth M. et al. "Localization using nonindividualized head-related transfer functions." *Journal of the Acoustical Society of America*, **94**(1): July 1993.

[77] Wightman, Frederic L. and Kistler, Doris J. "Headphone simulation of free-field listening. I: Stimulus synthesis." *Journal of the Acoustical Society of America*, **85**(2): February 1989.

[78] Wightman, Frederic L. and Kistler, Doris J. "Headphone simulation of free-field listening. II: Psychophysical validation." *Journal of the Acoustical Society of America*, **85**(2): February 1989.

[79] Wightman, Frederic L. and Kistler, Doris J. "The dominant role of low-frequency interaural time differences in sound localization." *Journal of the Acoustical Society of America*, **91**(3): March 1992.

[80] Wotton, Janine M. and Jenison, Rick L. "A backpropagation network model of the monaural localization information available in the bat echolocation system." *Journal of the Acoustical Society of America*, **101**(5) Pt. 1: May 1997.

[81] Wotton, Janine M. and Jenison, Rick L. "A backpropagation network model of the monaural localization information available in the bat echolocation system." *Journal of the Acoustical Society of America*, **101**(5): May 1997.

[82] Wu, Zhenyang et al. "A time domain binaural model based on spatial feature extraction for the head-related transfer function." *Journal of the Acoustical Society of America*, **102**(4): October 1997.

[83] Yost, William A. and Gourevitch, George eds. *Directional Hearing.* Springer-Verlag, New York: 1987.

[84] Zahorik et al. "On the discriminability of virtual and real sound sources." *1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics.* (IEEE catalog number: 95TH8144).

[85] Zhou et al. "Characterization of external ear impulse responses using Golay codes." *Journal of the Acoustical Society of America*, **92**(2) Pt 1: August 1992.

[86] Ziomek, Lawrence J. *Fundamentals of Acoustic Field Theory and Space-Time Signal Processing*. CRC Press, Ann Arbor: 1995.

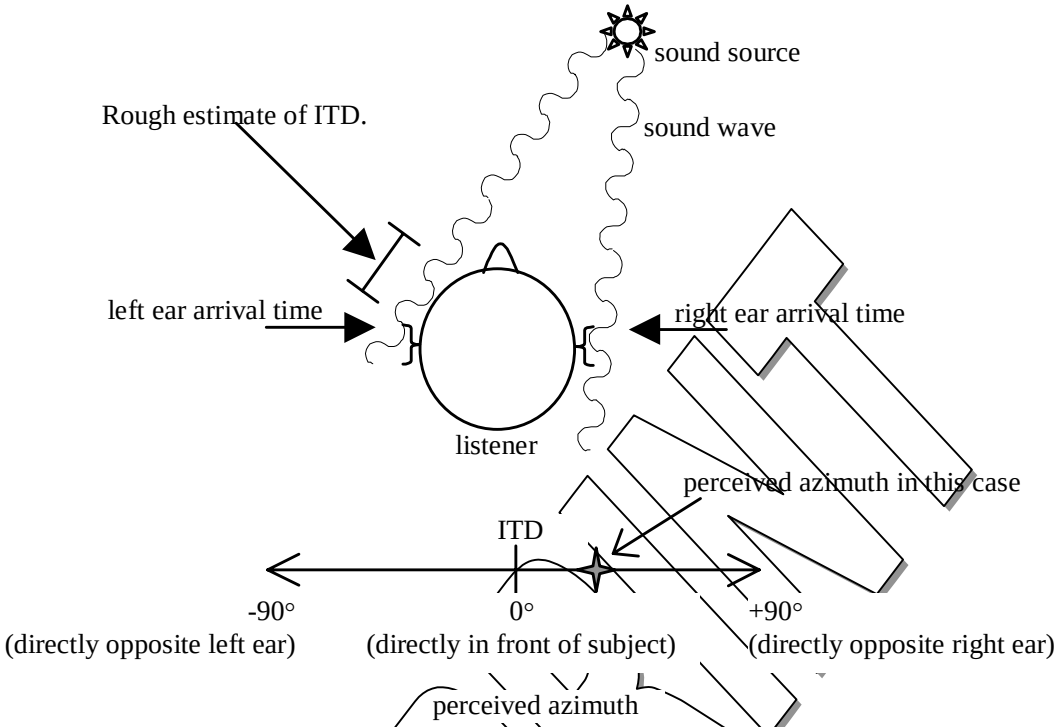


Figure 1. Using Interaural Time Differences (ITD's) to estimate the azimuth of a sound source. In general, a source is perceived to be closer to the ear at which the first wavefront arrives. The larger the magnitude of the ITD, the larger the lateral displacement.

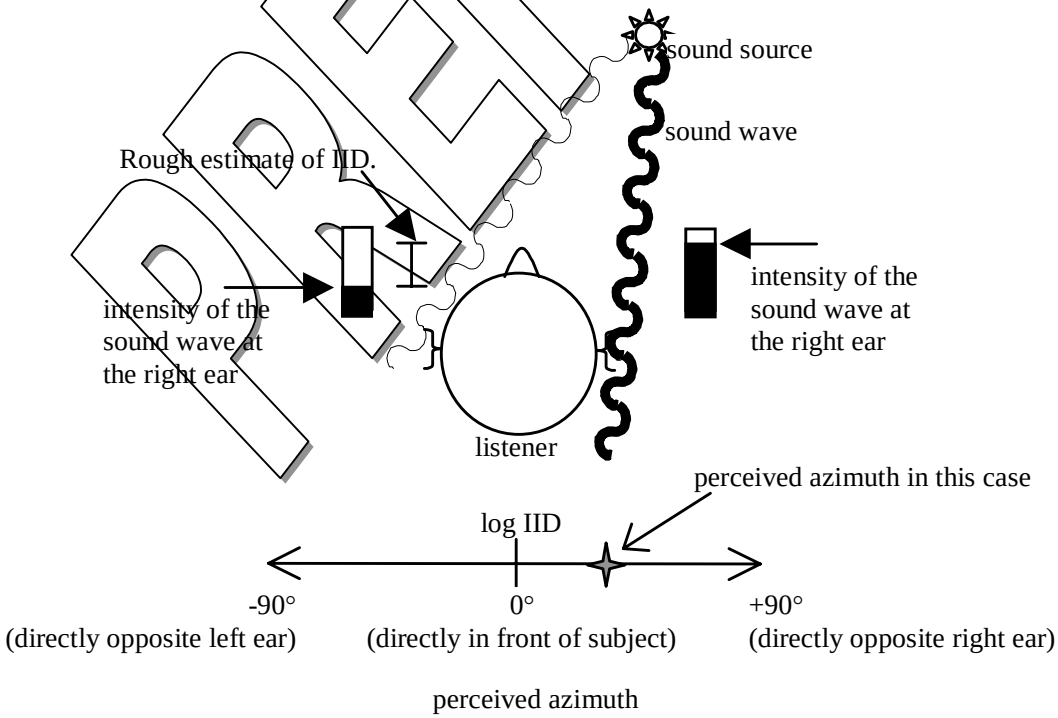
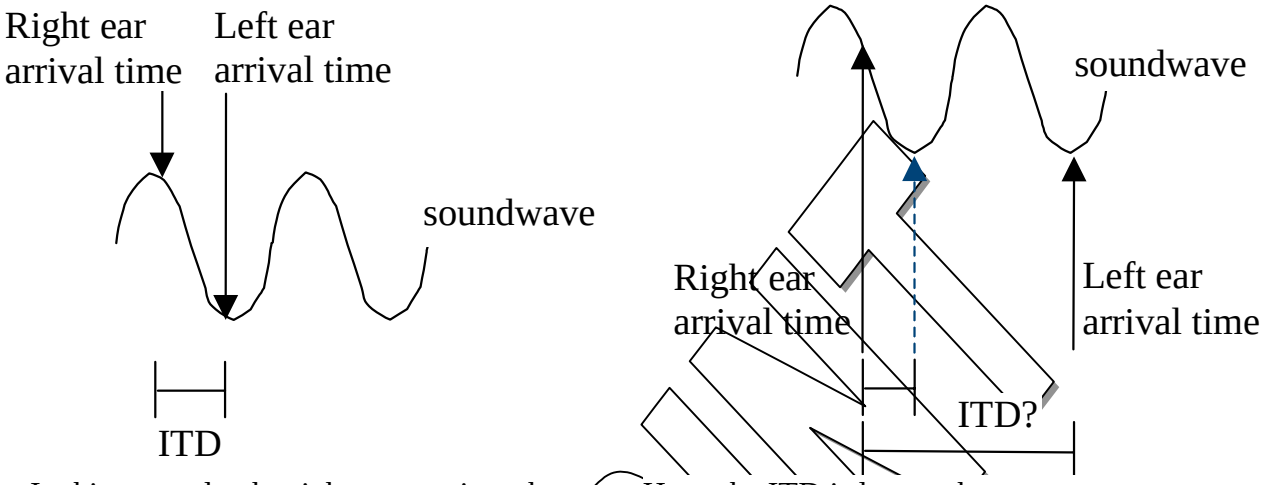


Figure 2. Using Interaural Intensity Differences (IID's) to estimate the azimuth of a sound source. In general, a source is perceived to be closer to the ear at which more energy arrives. The larger the magnitude of the IID, the larger the lateral displacement.

1. Below 1500 Hz (“low-frequency”) 2. Above 1500 Hz (“high-frequency”)



In this example, the right ear receives the signal first, so the sound will be perceived to the right of the listener. Since the ITD is less than a wavelength, the ITD represents a phase difference which corresponds to a unique azimuth.

Here, the ITD is longer than a wavelength. In this case, the ITD does not correspond to a unique azimuth, as the auditory system may perceive a shorter ITD due to aliasing problems.

Figure 3. The ambiguity of ITD’s in determining lateral position for higher frequencies.

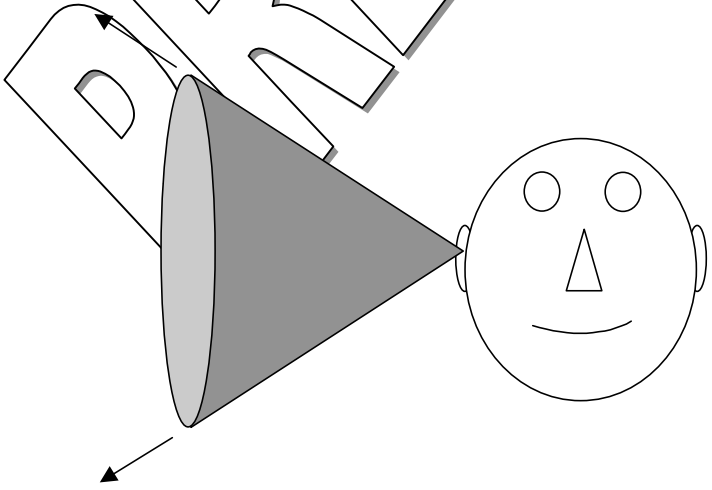


Figure 4. The Cone of Confusion. All points on the cone the same distance from the cone’s apex share the same ITD and IID, and are therefore indistinguishable according to duplex theory.

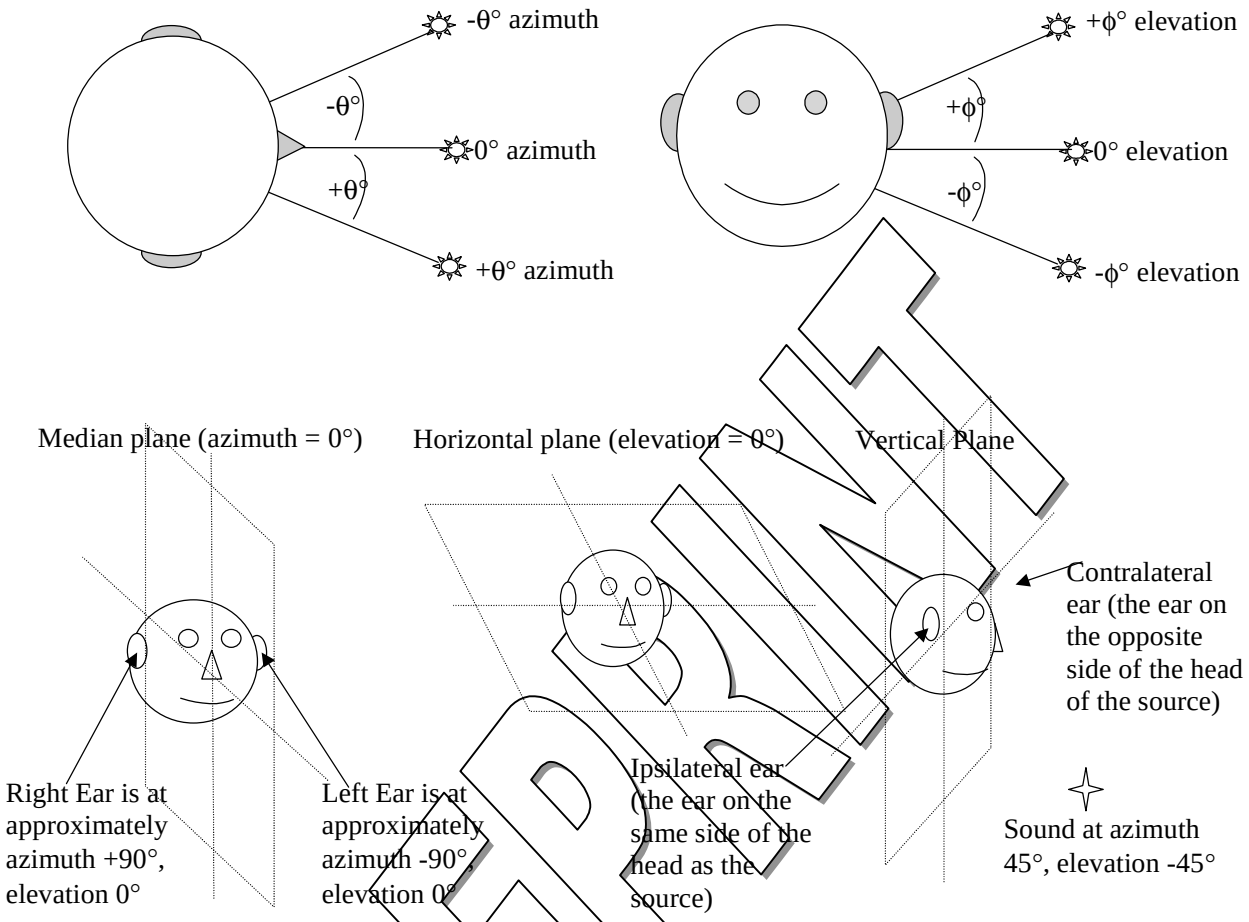


Figure 5. Spatial coordinate system and terminology used in much HRTF literature.

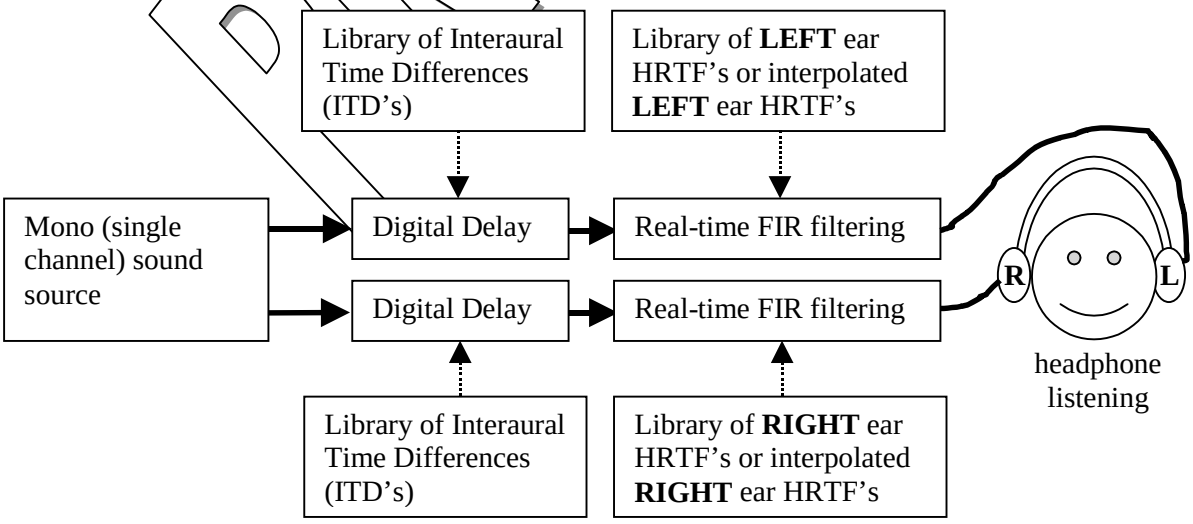


Figure 6. Block diagram of a simple HRTF-based spatial sound synthesis system. The delays and FIR filters can be updated in real time to synthesize moving sound sources.

Frequency Domain Representations of HRTF's

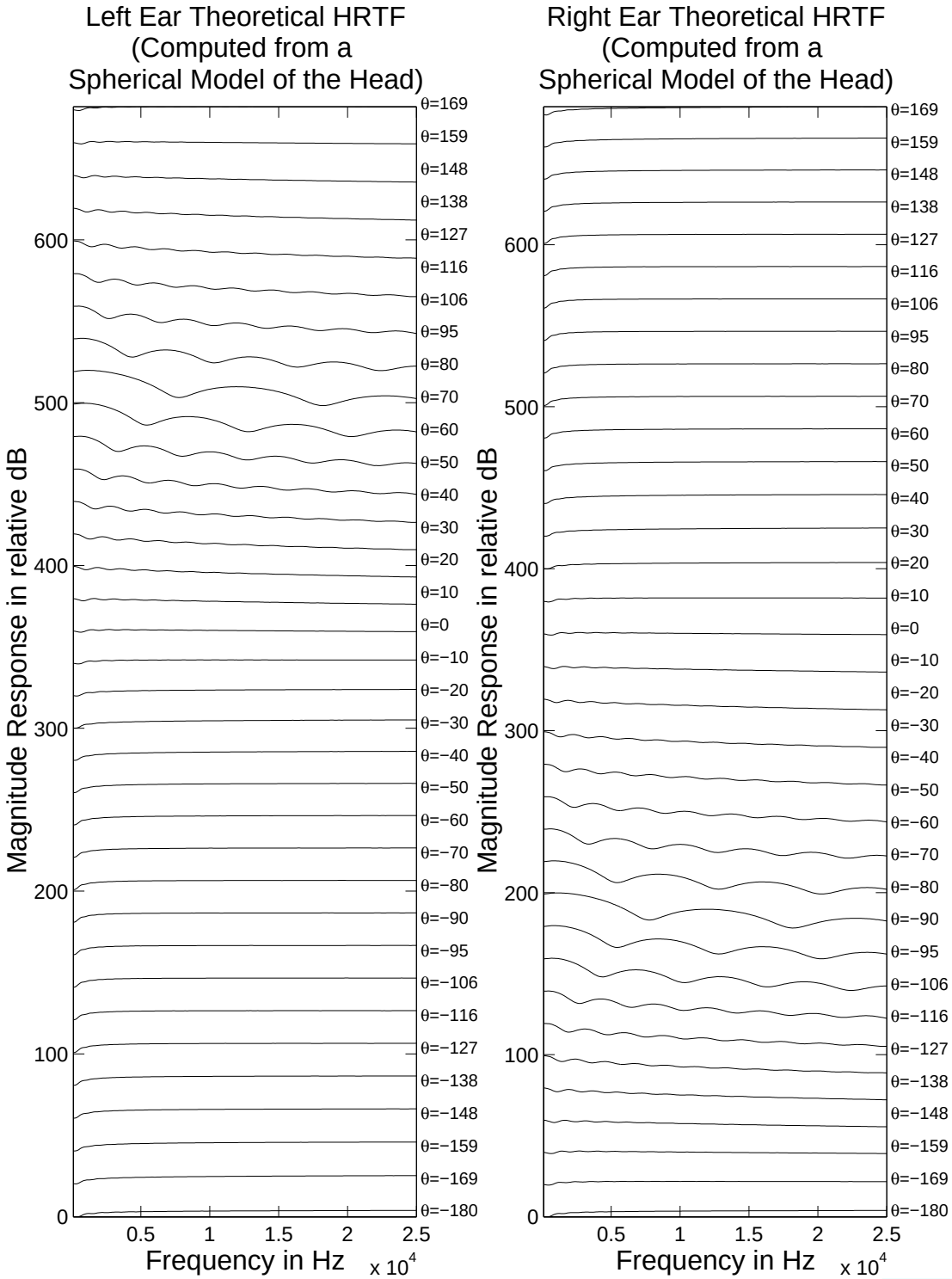


Figure 7. Frequency domain comparison of theoretical HRTF's as a function of azimuth in the horizontal plane (elevation = 0°). Diffraction effects for low frequencies can be seen on the contralateral side at azimuths +90° and –90° for the left and right ears, respectively. Amplification effects due to the ear's proximity to the head can be seen on the ipsilateral side. The contralateral HRTF's are more complex than the ipsilateral HRTF's.

Frequency Domain Representations of HRTF's

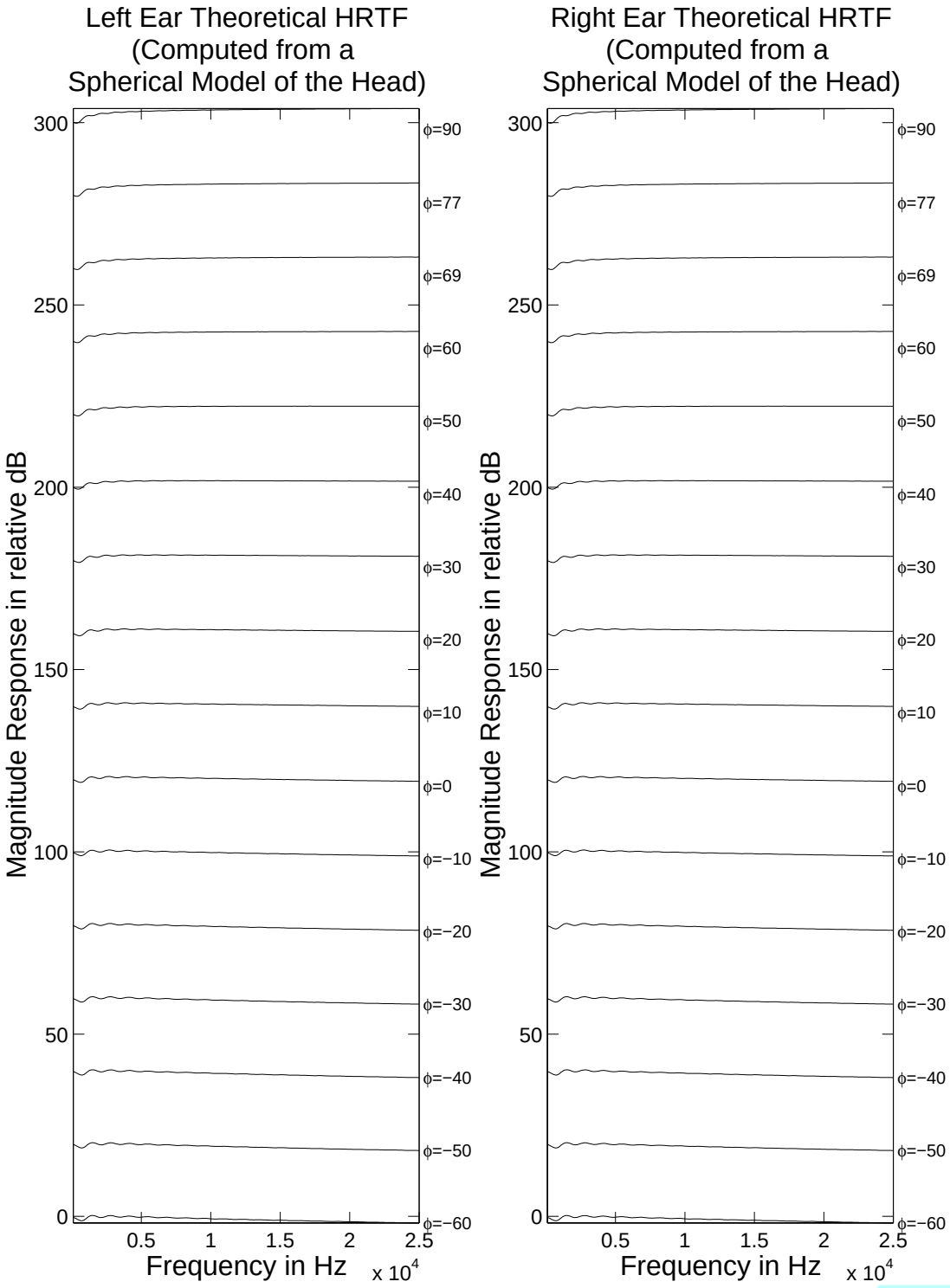


Figure 8. Frequency domain comparison of theoretical HRTF's as a function of elevation in the median plane (azimuth = 0°). There is a slight high-pass characteristic to HRTF's with higher elevations, and a slight low-pass characteristic to HRTF's with lower elevations.

Frequency Domain Representations of HRTF's

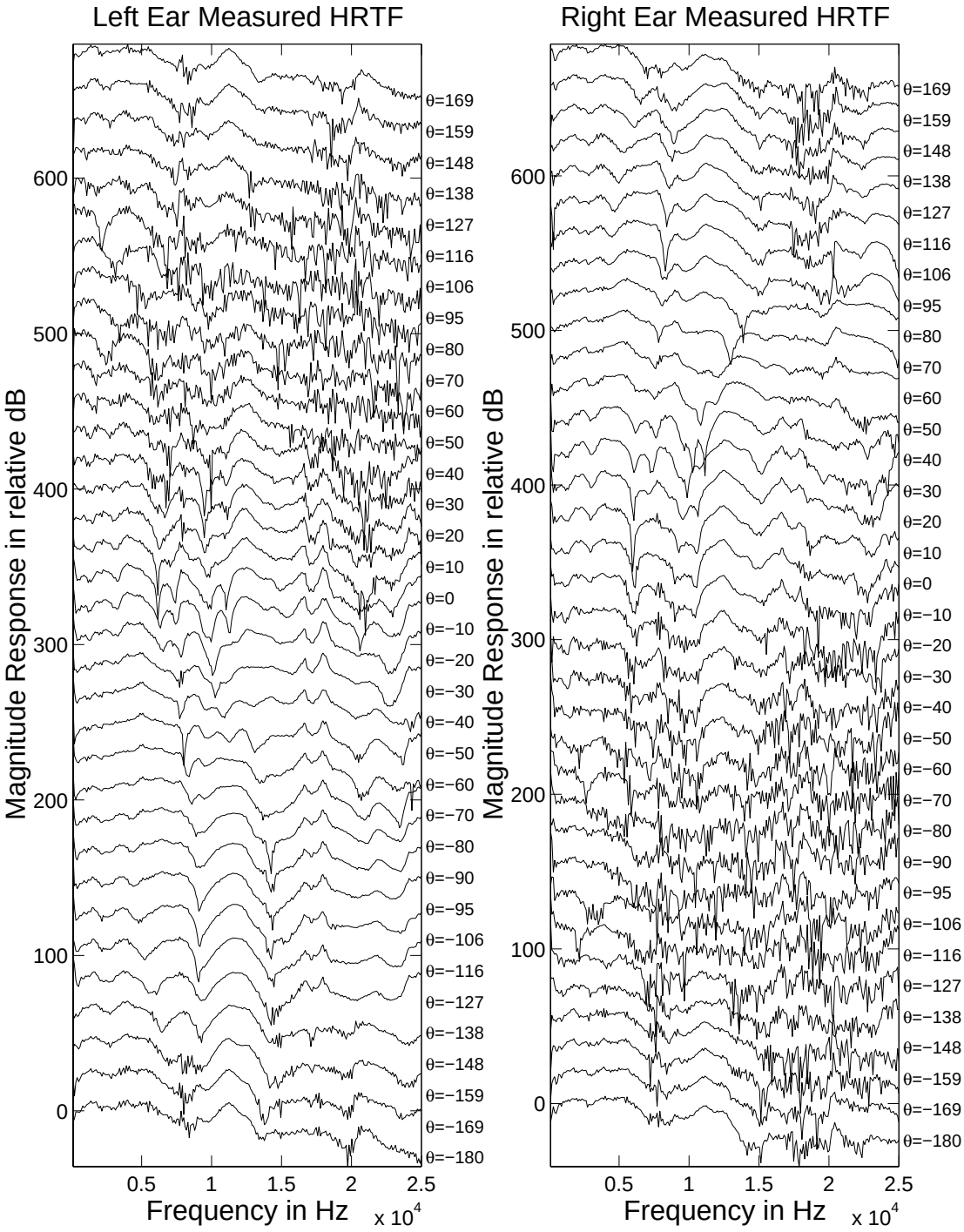


Figure 9. Frequency domain comparison of measured HRTF's as a function of azimuth in the horizontal plane (elevation = 0°). These HRTF's are more complex than their theoretical counterparts in Figure 7. Diffraction effects for low frequencies can be seen on the contralateral side at azimuths +90° and –90° for the left and right ears, respectively. In general, the contralateral HRTF's have a lower Signal to Noise Ratio (SNR) than the ipsilateral HRTF's.

Frequency Domain Representations of HRTF's

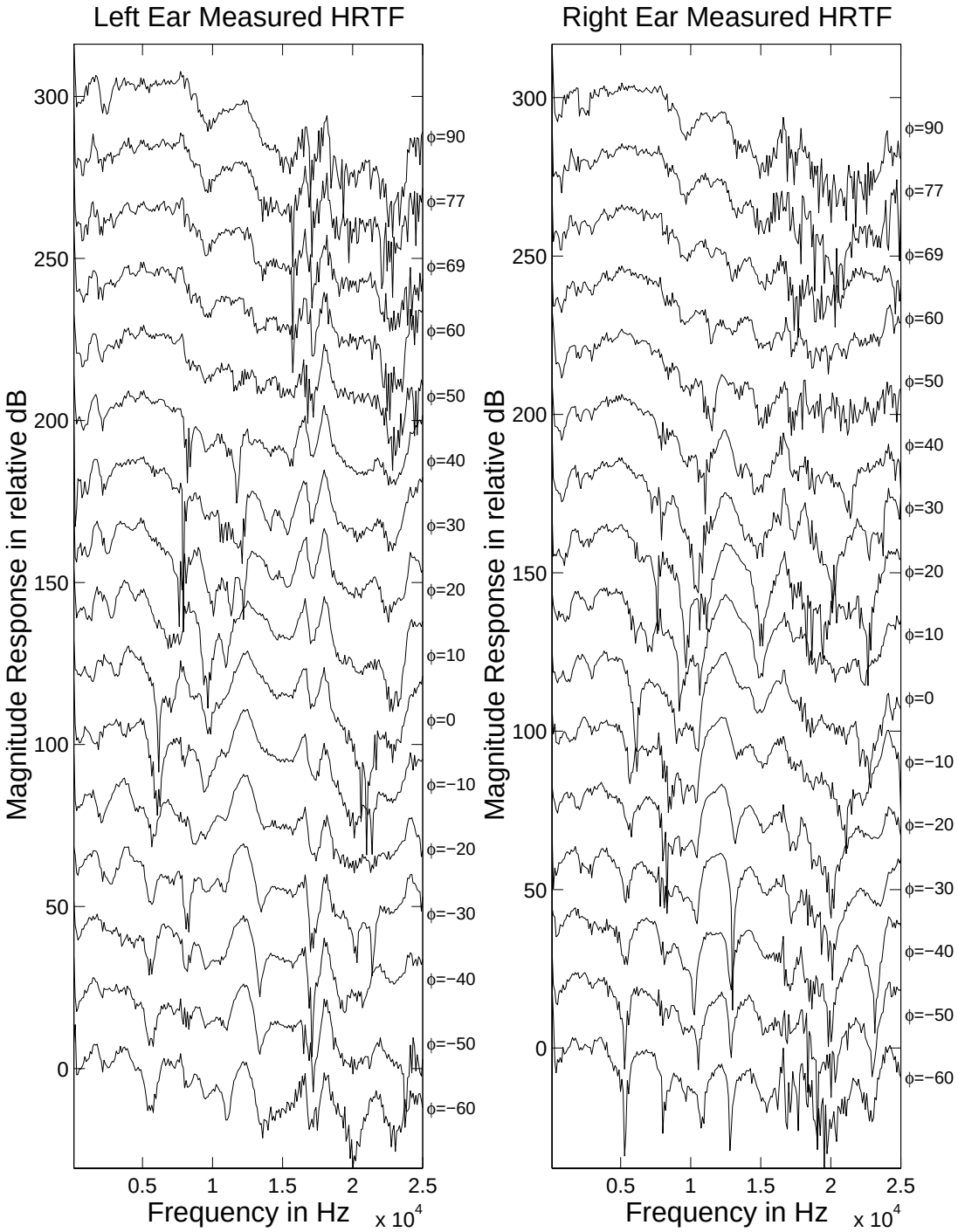


Figure 10. Frequency domain comparison of measured HRTF's as a function of elevation in the median plane (azimuth = 0°). These HRTF's are more complex than their theoretical counterparts in Figure 8. There is a notch at 7 kHz that migrates upward in frequency as elevation increases. There is also a shallow peak at 12 kHz which "flattens out" at higher elevations. The more complex structure of measured HRTF's is due to pinna and torso interactions, which are not predicted in the spherical head model of HRTF's.

Time Domain Representations of HRTF's

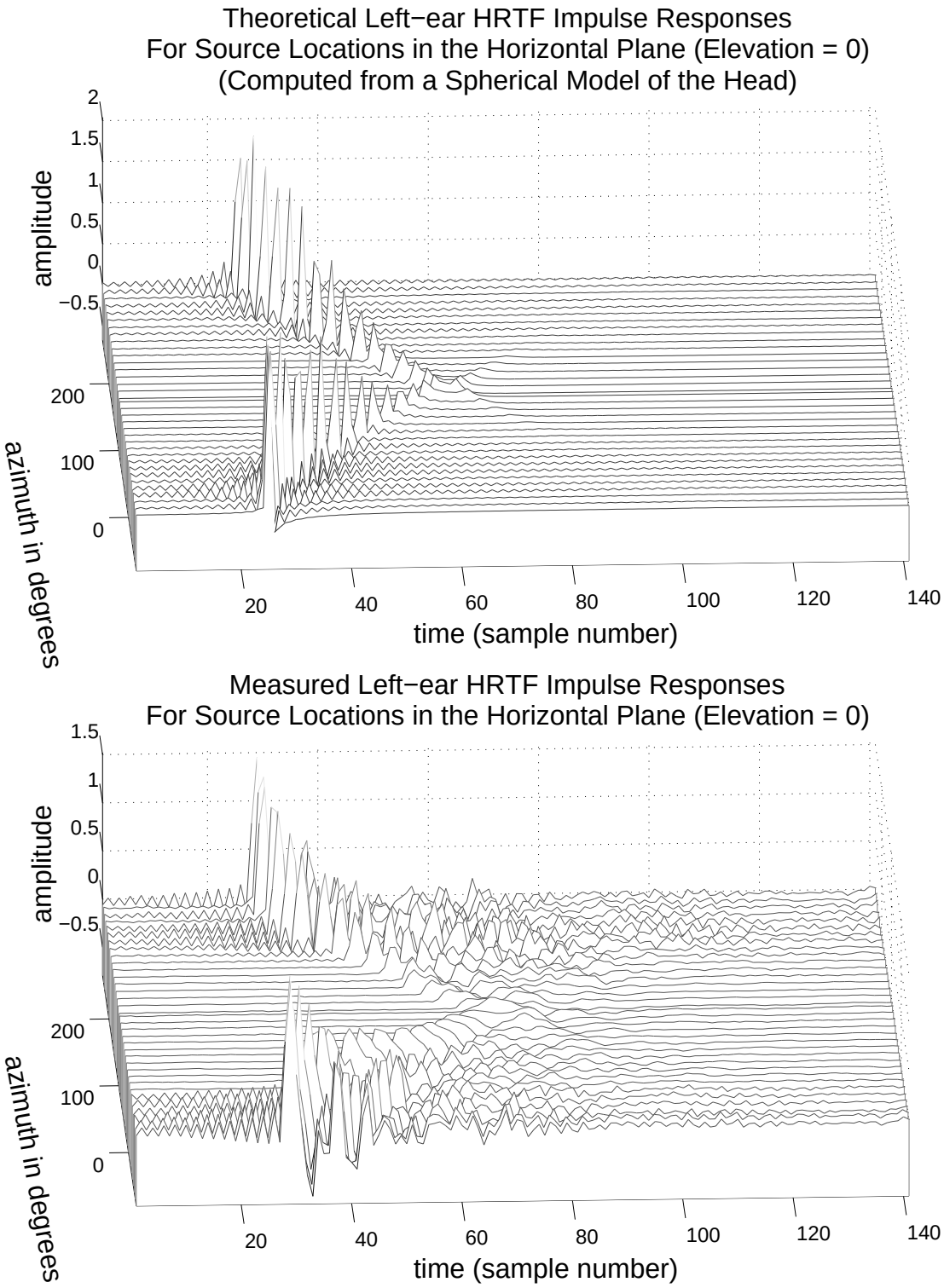


Figure 11. Time domain comparison of measured and theoretical left ear HRTF's as a function of azimuth in the horizontal plane (elevation = 0°). Significant energy arrives at the left ear from some contralateral locations due to diffraction effects: note the relatively large amplitude of the initial peak in the impulse responses corresponding to azimuth +90°.

Time Domain Representations of HRTF's

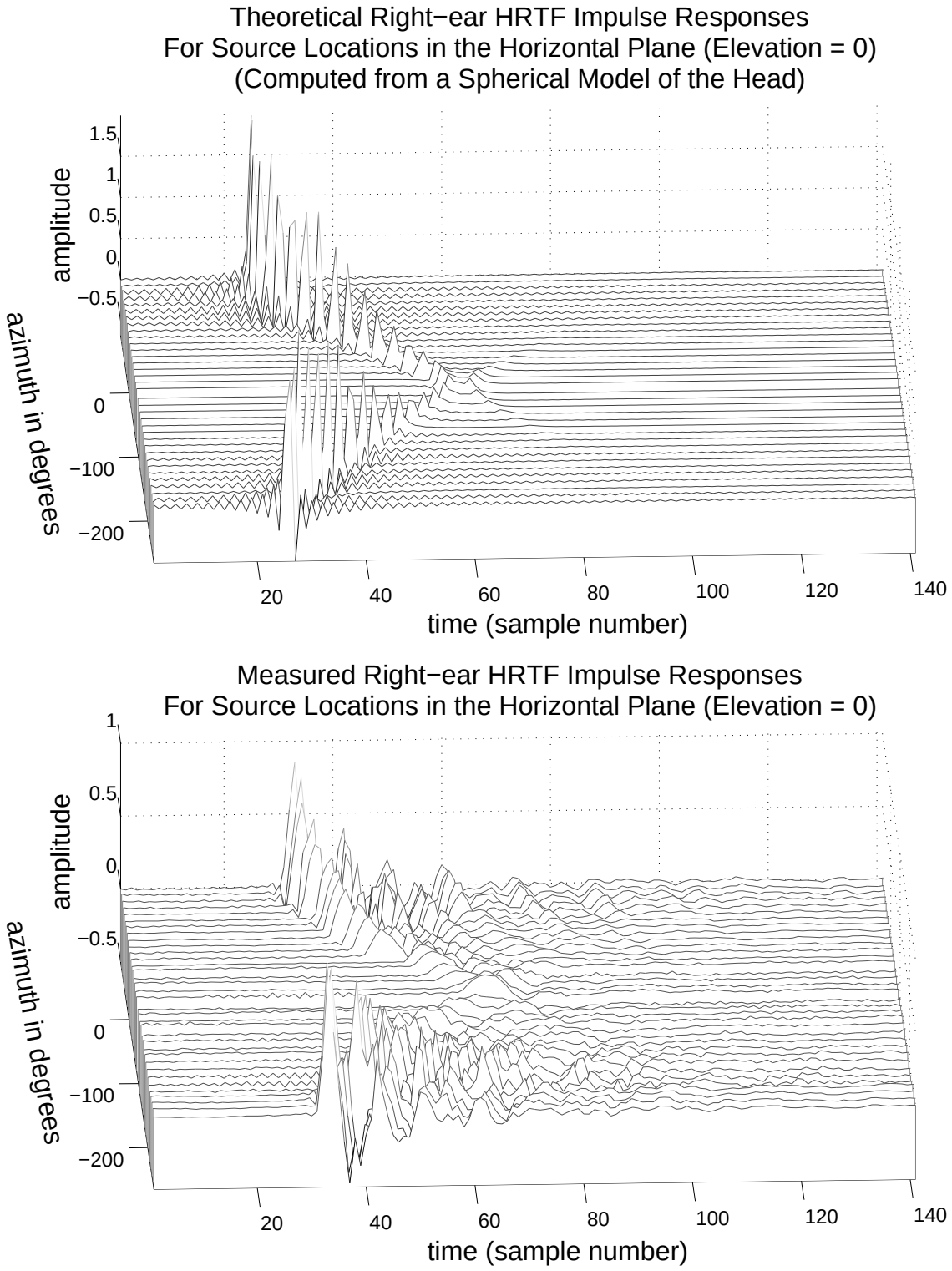


Figure 12. Time domain comparison of measured and theoretical right ear HRTF's as a function of azimuth in the horizontal plane (elevation = 0°). Significant energy arrives at the right ear from some contralateral locations due to diffraction effects: note the relatively large amplitude of the initial peak in the impulse responses corresponding to azimuth -90°.

Time Domain Representations of HRTF's

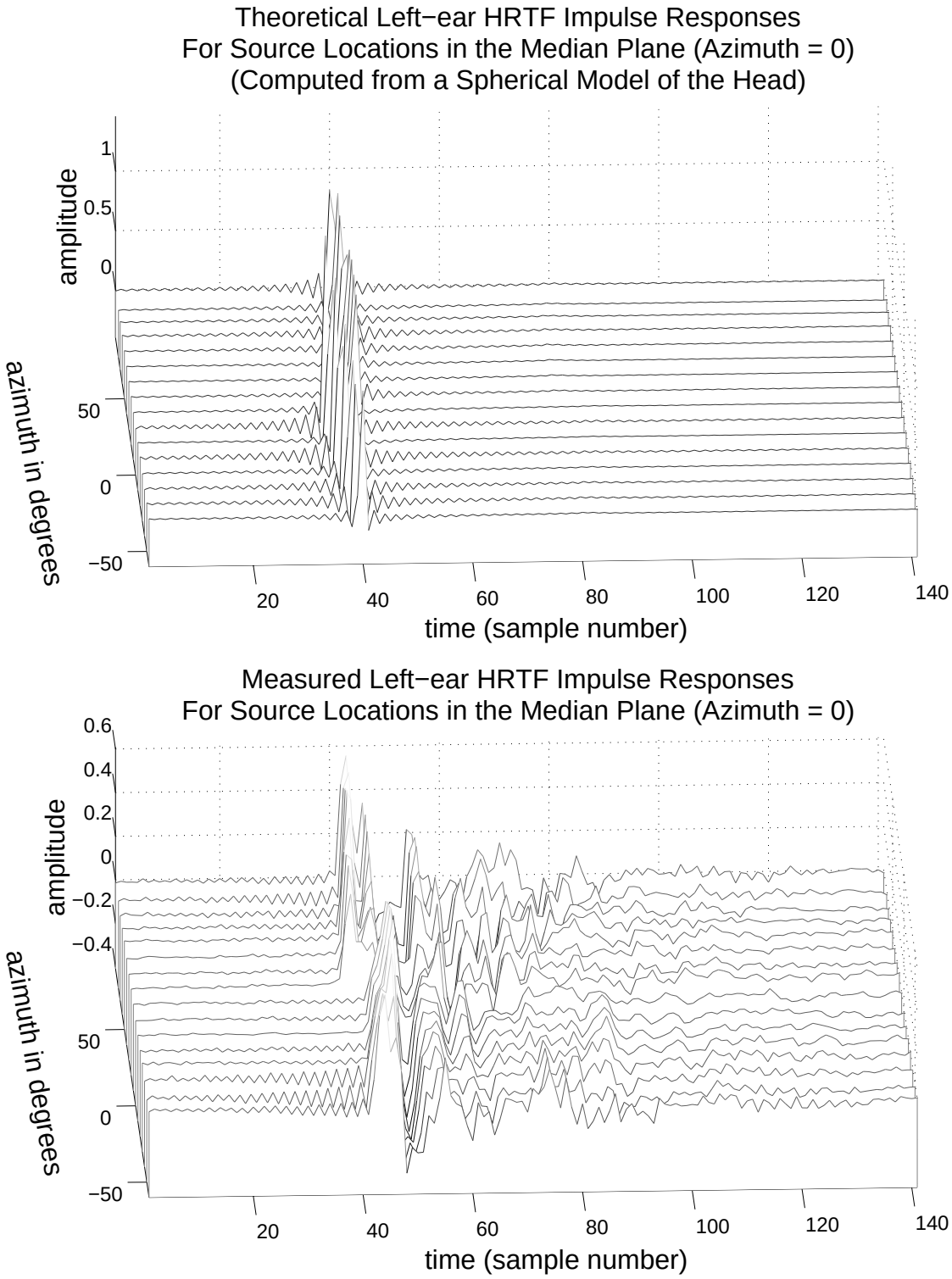


Figure 13. Time domain comparison of measured and theoretical left ear HRTF's as a function of elevation in the median plane (azimuth = 0°). Note the slight difference in arrival times associated with positive and negative elevations.

Time Domain Representations of HRTF's

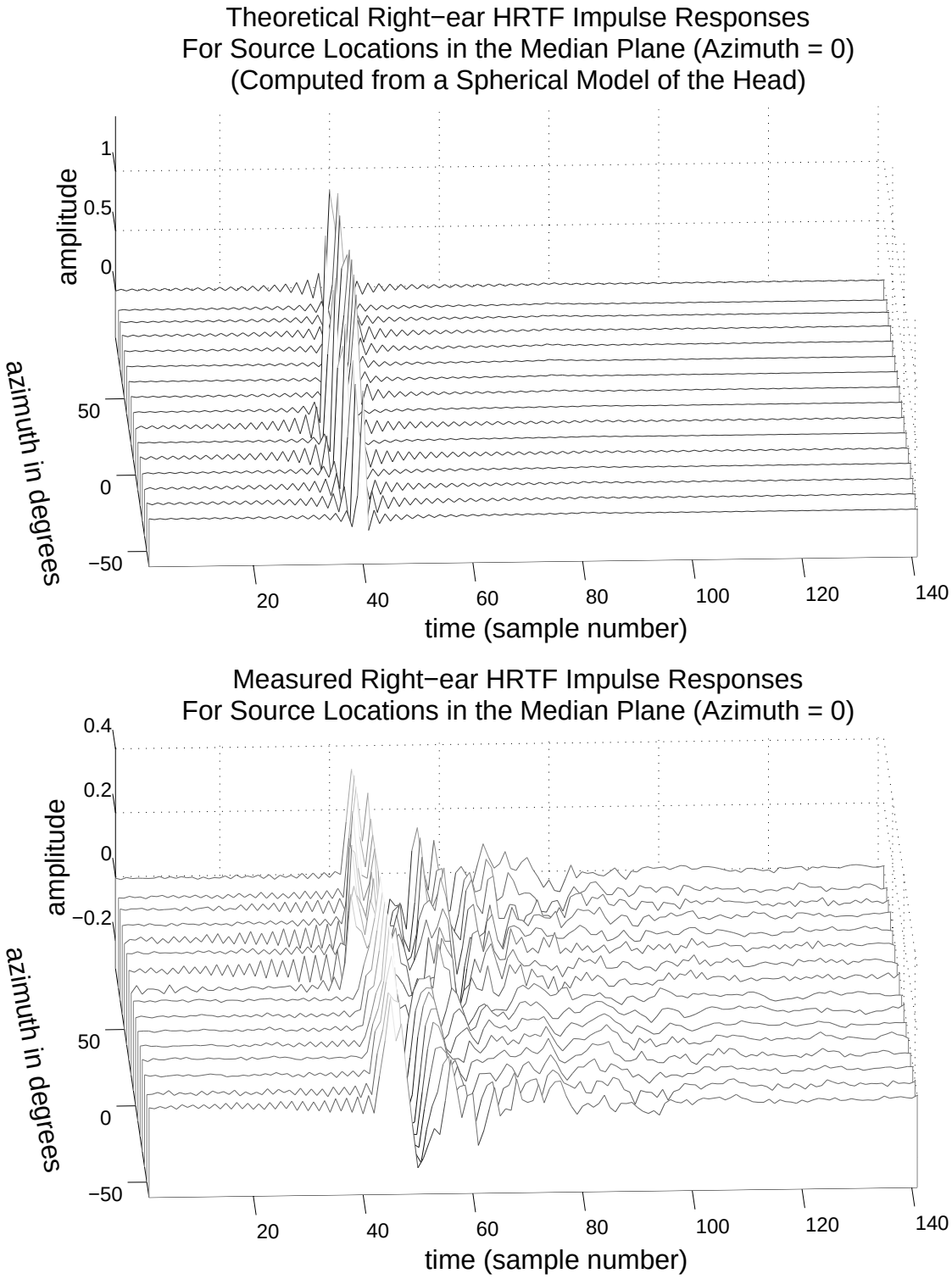


Figure 14. Time domain comparison of measured and theoretical right ear HRTF's as a function of elevation in the median plane (azimuth = 0°). Note the slight difference in arrival times associated with positive and negative elevations.

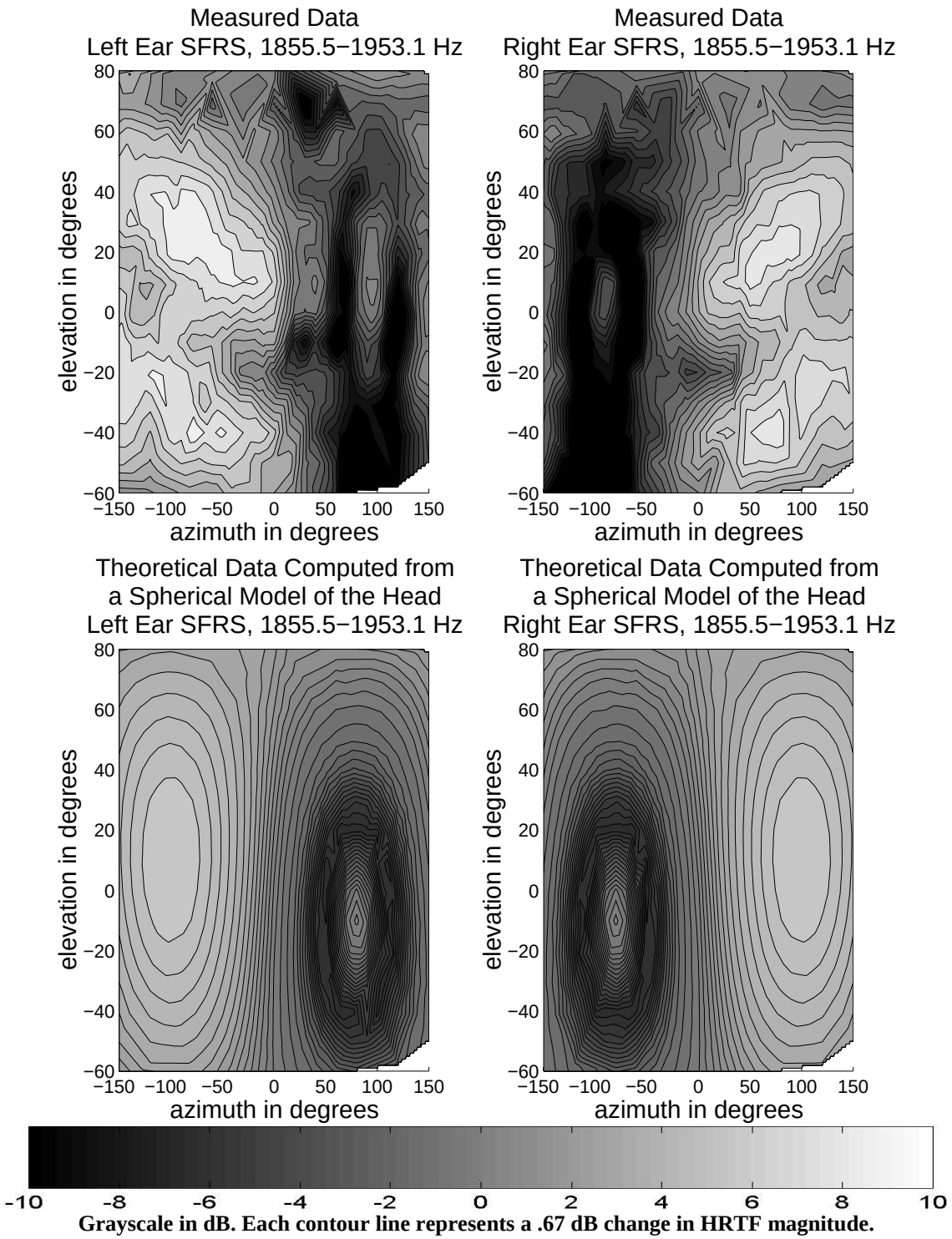


Figure 15. Diffraction effects in theoretical and measured HRTF's at 1.9 kHz as expressed in the spatial domain. Note the local maxima, or "hotspots" on the contralateral side in both the measured and theoretical HRTF's near azimuths +100°, -100° for the left and right ears, respectively. These hotspots are the "bright spots" discussed by Shaw [65].

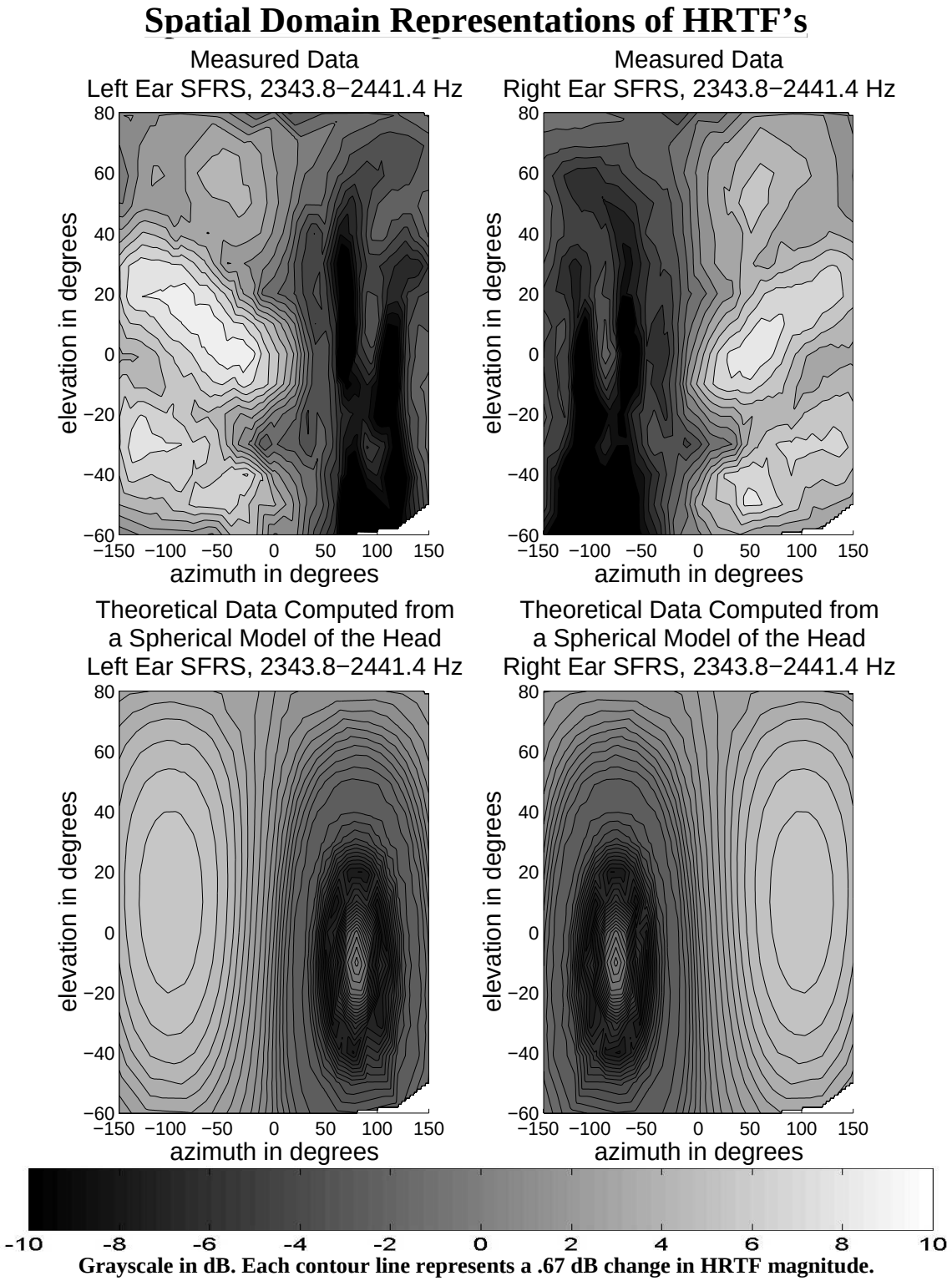


Figure 16. Diffraction effects in theoretical and measured HRTF's at 2.4 kHz as expressed in the spatial domain. Note the local maxima, or "hotspots" on the contralateral side in both the measured and theoretical HRTF's near azimuths +100°, -100° for the left and right ears, respectively. These hotspots are the "bright spots" discussed by Shaw [65].

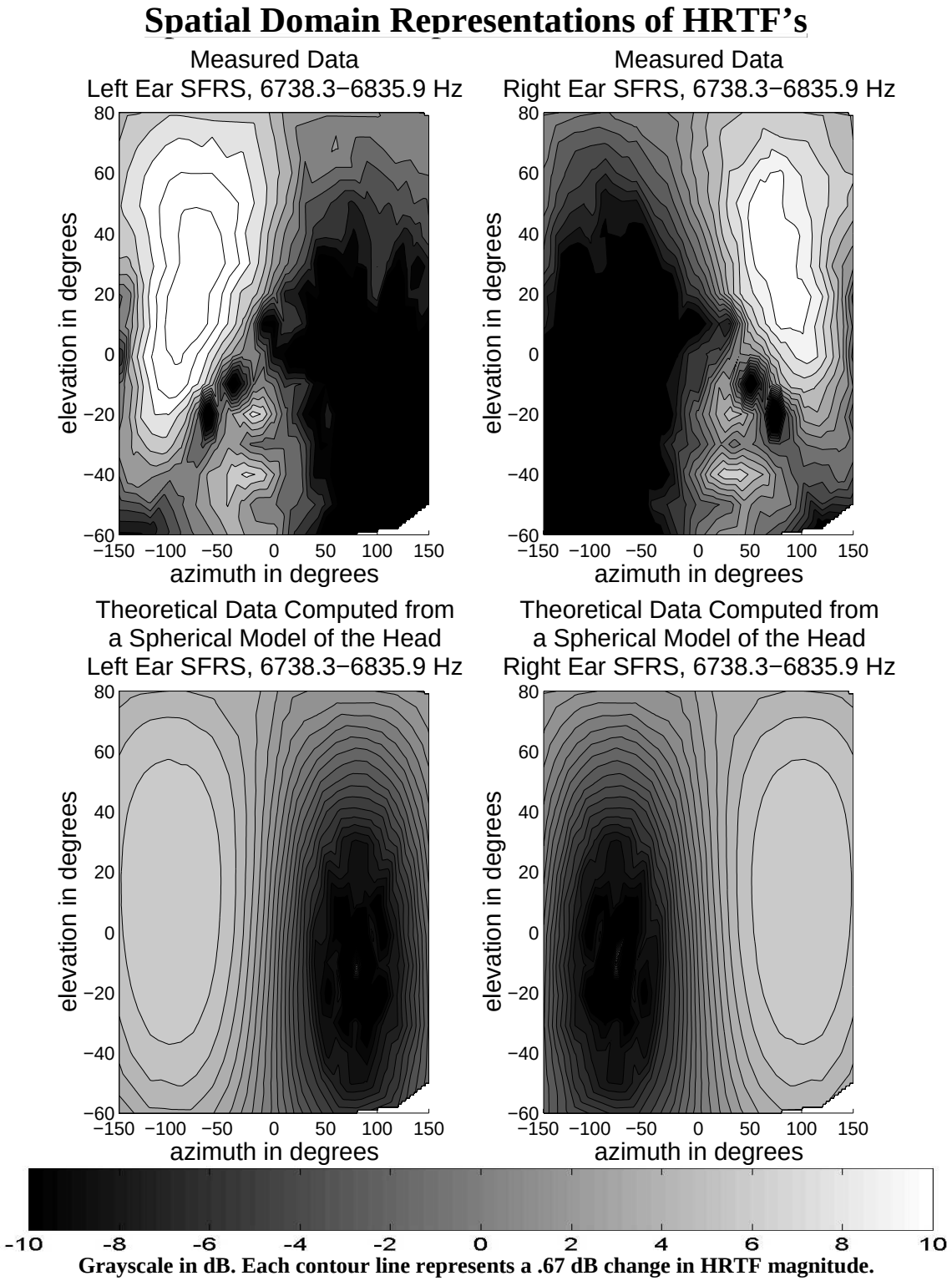


Figure 17. Elevation effects in theoretical and measured HRTF's at 6.8 kHz as expressed in the spatial domain. Note the prominent "hotspot" which occurs at a positive elevation on the ipsilateral side. This hotspot corresponds to a preferred positive perceptual elevation for narrowband sounds centered near this frequency.

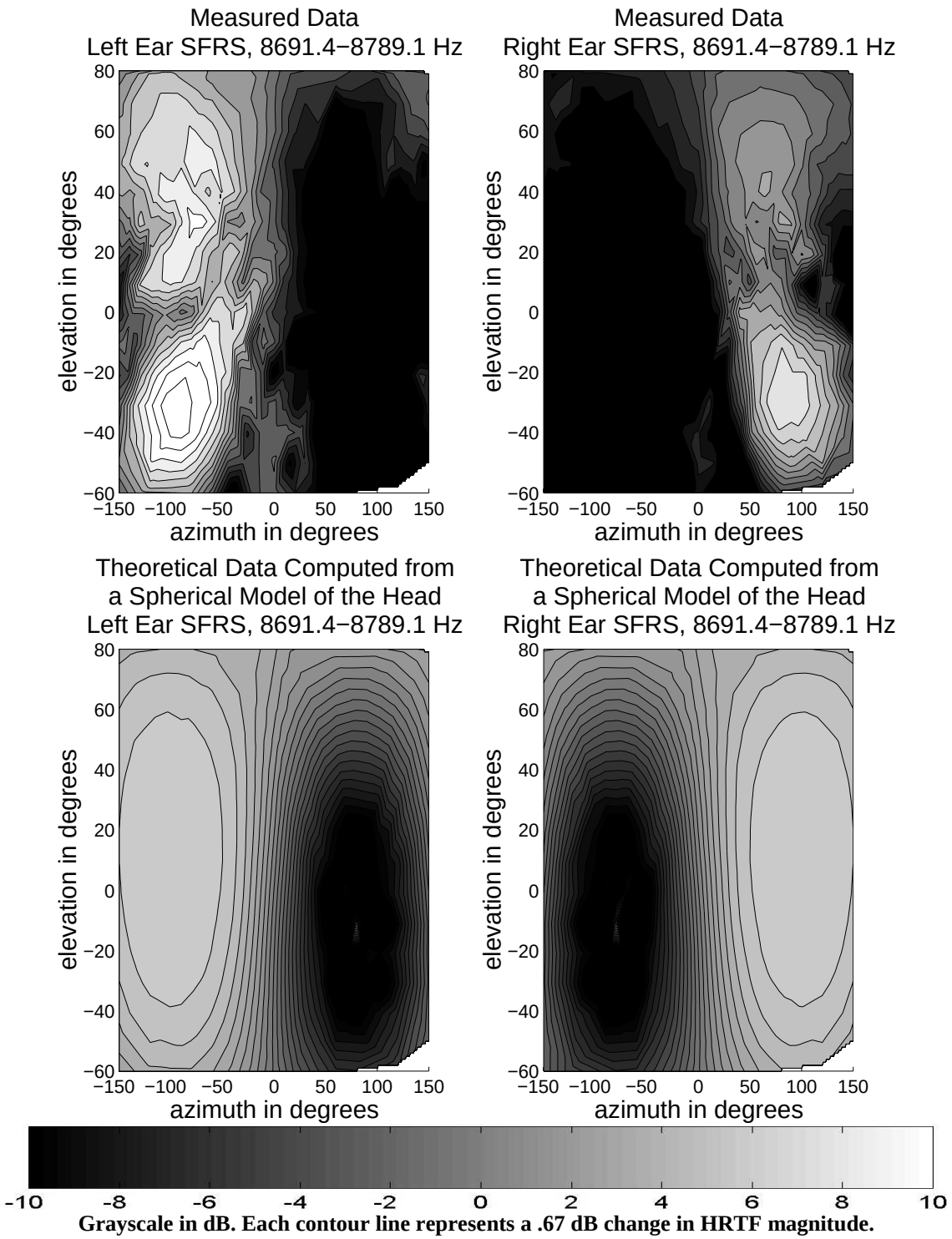


Figure 18. Elevation effects in theoretical and measured HRTF's at 8.7 kHz as expressed in the spatial domain. Note the prominent "hotspot" which occurs at a positive elevation on the ipsilateral side. This hotspot corresponds to a preferred positive perceptual elevation for narrowband sounds centered near this frequency.