Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space*

COREY I. CHENG, AES Student Member, **AND GREGORY H. WAKEFIELD**

University of Michigan, Department of Electrical Engineering and Computer Science, Ann Arbor, MI 48109, USA

In this tutorial, head-related transfer functions (HRTFs) are introduced and treated with respect to their role in the synthesis of spatial sound over headphones HRTFs are formally defined, and are shown to be important in reducing the ambiguity with which the classical duplex theory decodes a free-field sound's spatial location. Typical HRTF measurement strategies are described, and simple applications of HRTFs to headphone-based spatialized sound synthesis are given. By comparing and contrasting representations of HRTFs in the time, frequency, and spatial domains, different analytic and signal processing techniques used to investigate the structure of HRTFs are highlighted

0 INTRODUCTION

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 that influence spatial hearing. These spectral cues are called head-related transfer functions (HRTFs) and are the focus of much engineering and psychoacoustic research. This tutorial provides an introduction to the role of HRTFs in spatial hearing and to the structure of HRTF data.

The paper is organized as follows. Section 1.1 presents the duplex theory, a simple model introduced by Lord Rayleigh to explain directional hearing in the azimuthal (left–right) direction [1]. 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 HRTFs, explain typical measurement procedures for HRTFs, discuss the synthesis of spatial audio with HRTFs, and review some problems with current headphone-based spatial synthesis systems, respectively. Sections 2.1–2.3 provide an introduction to the structure of HRTFs by using the time, frequency, and spatial domains to compare HRTFs mea-

sured empirically from human subjects and HRTFs 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 BACKGROUND

1.1 Duplex Theory

Duplex theory is a model for estimating a free-field target's spatial location by two binaural cues: interaural time differences and interaural intensity differences [1], [2]. An interaural time difference (ITD) is defined as the difference in arrival times of a sound's wavefront at the left and right ears. Similarly, an interaural intensity 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 ITDs and IIDs are important parameters for the perception of sounds originating from the horizontal plane, which includes both ears and separates the top and bottom halves of the head [2]. 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, the 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, ITDs 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 Figs. 1 and 2.

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 nonshadowed 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, the perceived azimuth does vary approximately linearly with the logarithm of the IID [2]. See Fig. 3 for details.

Although the simplicity and success of the duplex theory are attractive, the theory only explains the percep-

tion 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 locations of all sounds originating from points on this cone are indistinguishable according to the duplex theory. Fig. 4 shows the cone of confusion for a particular ITD.

The problem is acute in the median plane, which separates the two ears and runs vertically through the head.
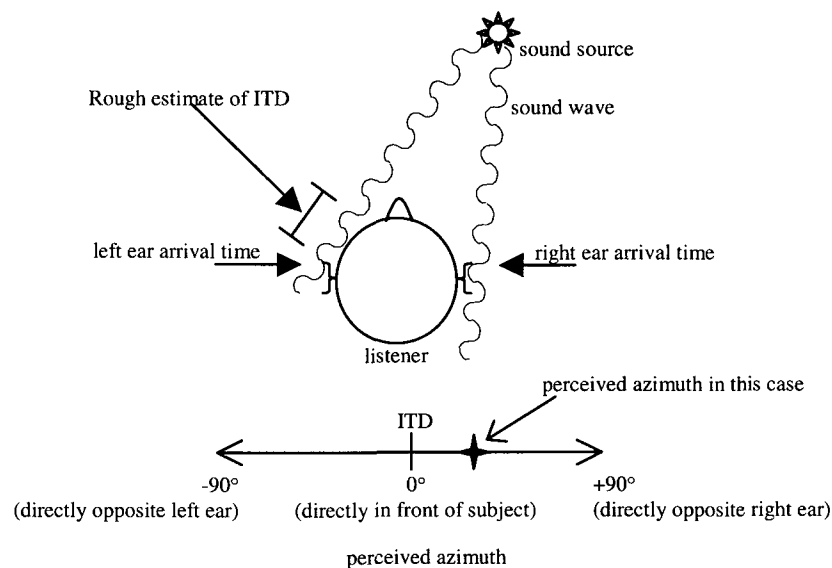

Fig 1. Using interaural time differences (ITDs) to estimate the azimuth of a sound source. In general, a source is perceived to be closer to the ear at which the first wavefront arrives The larger the magnitude of the ITD, the larger the lateral displacement.
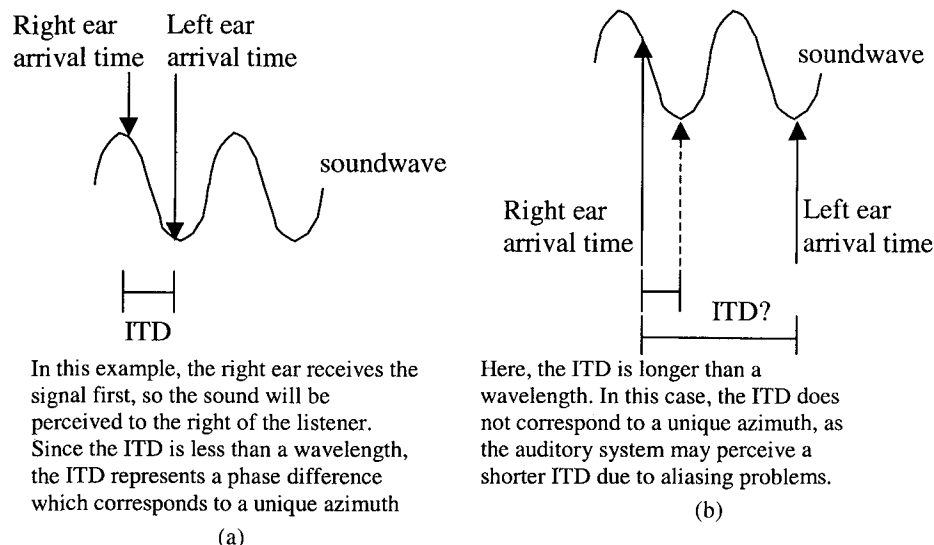

Fig 2 Ambiguity of ITDs in determining lateral position for higher frequencies (a) Below 1500 Hz (low frequency) (b) Above 1500 Hz (high frequency).

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 locate correctly a sound's elevation in the median plane, the sound must be broadband and contain frequencies of 7 kHz and above [2].

It is widely thought that the auditory system uses ITD, IID, and spectral cues to determine the 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 (HRTFs)

Whereas it is reasonable to predict the linear relationship between ITD, IID, and perceived spatial location, there is less intuition as to how spectral structure and perceived spatial location relate mathematically. As a first step toward understanding spectral cues in directional hearing, many researchers have tried to physically model [3], empirically measure [4], or more recently computationally simulate [5] the direction-dependent frequency response of the ear directly. These measurements are called head-related transfer functions (HRTFs) 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 to analyze

and uncover later the perceptual structure of the data.

Formally a single 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 HRTFs are measured from humans or mannequins for both the left and right ears at a fixed radius from the listener's head. HRTFs are measured at several different azimuths (left–right direction) and elevations (up–down direction), which are both measured in degrees or in radians. Fig. 5 contains some relevant terminology and depicts the spatial coordinate system used in much of the HRTF literature.

HRTFs 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, HRTFs have been found empirically to be minimum-phase systems and minimum-phase filters have been found to produce perceptually acceptable HRTFs [6]. This observation allows us to simplify the FIR specification of HRTFs 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 mag-
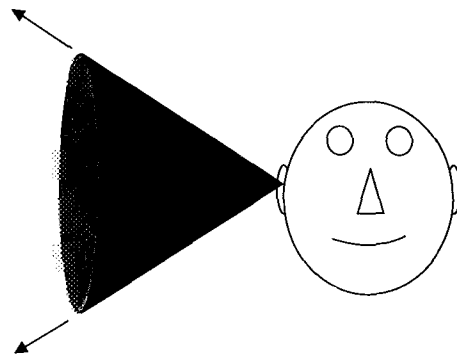

Fig. 4 Cone of confusion All points on cone at the same distance from the cone's apex share the same ITD and IID, and are therefore indistinguishable according to duplex theory.
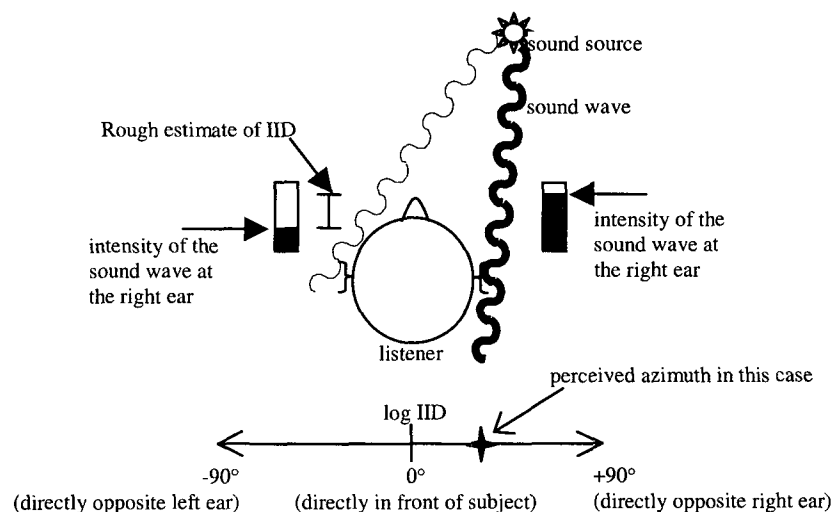

Fig 3 Using interaural intensity differences (IIDs) to estimate the azimuth of a sound source In general, a source is perceived to be closer to the ear at which more energy arrives. The larger the magnitude of the IID, the larger the lateral displacement

nitude frequency response and the phase response of a minimum-phase causal system form a Hilbert transform pair [7]. 2) The minimum-phase assumption allows us to separate ITD information from the FIR specification of HRTFs. Since minimum-phase filters have the minimum group-delay property and minimum energy-delay property [7], 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 HRTFs

A common technique used to measure right- and left-ear HRTFs empirically is to insert small microphones partially into a subject's ear canals, and then perform a simple form of system identification by playing a known-spectrum stimulus $s(n)$ through a loudspeaker placed at a specified azimuth θ, elevation ϕ, and distance from the subject's head [4]. A general description of this process is given in Fig. 6. There are many different methods that can be used to perform the system identification required to measure HRTFs. Since it is generally assumed that HRTFs are well modeled by linear, time-invariant systems, many researchers have employed standard LTI system identification methods such as direct excitation and cross-correlation analyses to measure HRTFs. A summary and brief comparison of four methods that have been used to measure HRTFs are given in Table 1.

Some early studies attempted to measure HRTFs directly by playing clicks or impulses over loudspeakers and recording the response at the ears. Other early stud-

ies attempted to increase the signal-to-noise ratio (SNR) during HRTF measurement by using a white-noise stimulus having a wideband, flat frequency response, with signal energy spread more evenly over a longer period of time. More recently, the maximal-length shift (MLS) register method of system identification has been employed for HRTF measurement. This technique maintains the advantage of using a stimulus with more evenly distributed energy for higher SNR while alleviating the problems caused by using a random stimulus [8]. Golay codes, or complementary codes, have also been recently used to measure HRTFs in an attempt to alleviate problems the MLS measurement technique has at low frequencies, while still providing good SNR [9], [10].

No matter which system identification procedure is used to measure HRTFs, portions of the measured transfer functions due to the measurement apparatus, such as the microphone and loudspeaker transfer functions, need to be removed from the measured response during HRTF postprocessing. These transfer functions can be measured with precision sound calibration equipment, and the inverses of these functions can be used to equalize the raw HRTF measurements. In addition, spectral features of the raw HRTFs which are the same for all locations are assumed not to contain important directional psychophysical cues, and can also be removed from the raw HRTFs. These features comprise the diffuse transfer function, which can be calculated by averaging the equalized, raw HRTFs from all spatial locations. We follow the conventions in [11] by defining the cascade of these two spatially invariant systems to be the common transfer function (CTF). During postprocessing, the CTF is removed from the raw HRTFs to yield the directional transfer function (DTF). The DTF is a function of azimuth θ and elevation ϕ, and is the quantity that contains
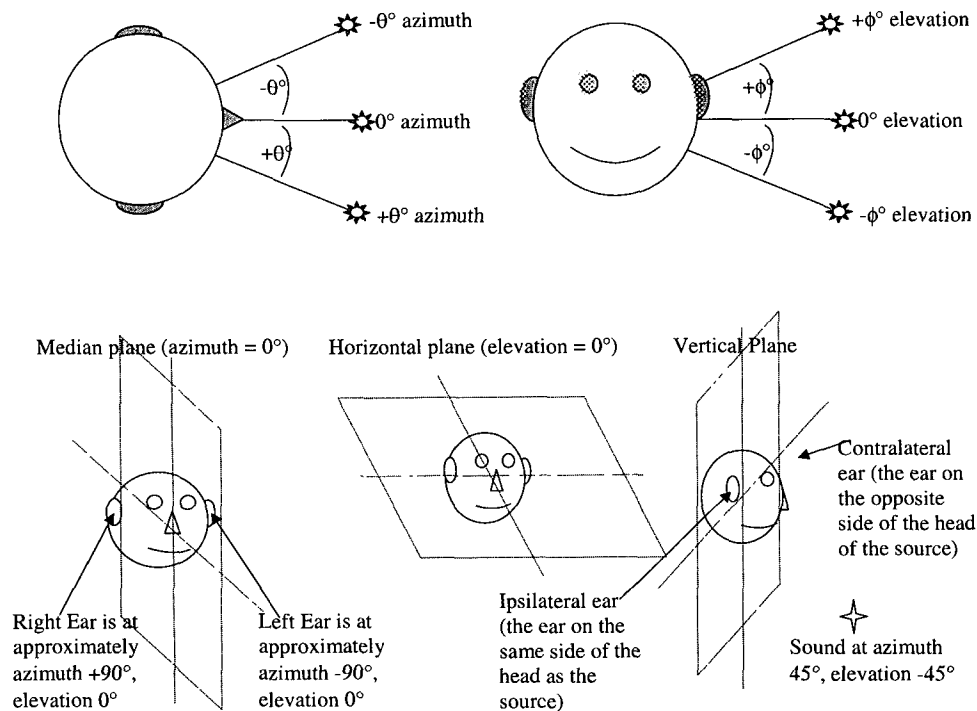

Fig. 5 Spatial coordinate system and terminology used in HRTF literature

spectral cues responsible for spatial hearing.

Mathematically, we can describe the postprocessing of the raw HRTF measurements as follows. Let $s(n)$ be the known stimulus signal presented at azimuth θ and elevation ϕ. Let $c(n)$ be the known CTF and $d_{l,\theta,\phi}(n)$ and $d_{r,\theta,\phi}(n)$ be the unknown left and right ear DTFs respectively; let $\hat{h}_{l,\theta,\phi}(n)$ and $\hat{h}_{r,\theta,\phi}(n)$ be the estimated left and right raw HRTFs, respectively, measured with one of the system identification methods discussed above. Then

$$\hat{h}_{l\,or\,r,\theta,\phi}(n) = s(n) * c(n) * d_{l\,or\,r,\theta,\phi}(n) \quad (1)$$

Or, equivalently expressed in the frequency domain,

$$\hat{H}_{l\,or\,r,\theta,\phi}(k) = S(k)C(k)D_{l\,or\,r,\theta,\phi}(k) \quad (2)$$

Here we assume that $c(n)$ is spatially invariant, and can be computed from known measurements of the recording apparatus and spectrally averaged values of $\hat{h}_{l,\theta,\phi}(n)$ and $\hat{h}_{r,\theta,\phi}(n)$ for several locations. Hence, compute the left and right DTFs as follows.

$$|D_{l\,or\,r,\theta,\phi}(k)| = \frac{|\hat{H}_{l\,or\,r,\theta,\phi}(k)|}{|S(k)||C(k)|} \quad (3)$$

$$\angle D_{l\,or\,r,\theta,\phi}(k) = \angle \hat{H}_{l\,or\,r,\theta,\phi}(k) - \angle S(k) - \angle C(k) \quad (4)$$

$$D_{l\,or\,r,\theta,\phi}(k) = |D_{l\,or\,r,\theta,\phi}(k)| \exp[j\angle D_{l\,or\,r,\theta,\phi}(k)] \quad (5)$$

$$d_{l\,or\,r,\theta,\phi}(n) = F^{-1}[D_{l\,or\,r,\theta,\phi}(k)] \quad (6)$$

Phase information from the computed time domain DTFs can be used to compute the ITD information associated with azimuth θ and elevation ϕ. Specifically, compute the ITD $n_{ITD,\theta,\phi}$ as the lag for which the cross-correlation function between $d_{l,\theta,\phi}(n)$ and $d_{r,\theta,\phi}(n)$ is maximized,

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

Minimum phase versions of the DTFs can be computed by windowing the real cepstrum of $d_{l,\theta,\phi}(n)$ and $d_{r,\theta,\phi}(n)$ [7], [12]. Define the window

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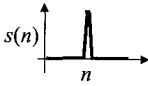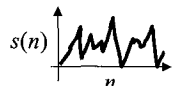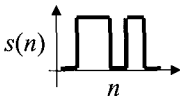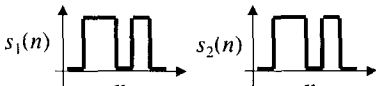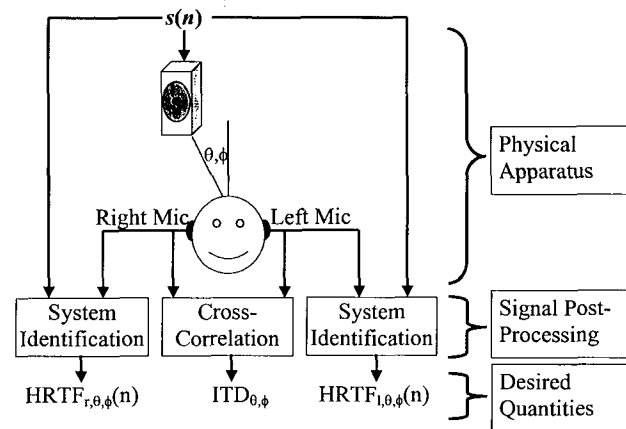

Fig 6 Measurement of HRTFs.

Table 1. Comparison of different system identification methods for measuring HRTFs

Stimulus signal $s(n)$	Shape and Type	Advantages	Disadvantages
Impulse (delta function)	Deterministic	Simple, intuitive meaning	$s(n)$ is a low-energy signal; may force nonlinear effects in loudspeakers
White gaussian noise	Random	More energy	$s(n)$ is random; need to average several measurements to identify $h(n)$
Maximal length sequences (MLS codes)	Deterministic, binary	More energy; not random	Problems identifying dc components of $h(n)$; $s(n)$ is an infinite-length signal
Golay (complementary) codes	Deterministic, binary	More energy; not random; finite length	Need to present two different stimuli to system in order to identify $h(n)$ (increased complexity)

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

$$c_{l\,\text{or}\,r,\theta,\phi}(n) = [F^{-1}\,(\log|F\{d_{l\,\text{or}\,r,\theta,\phi}(n)\}|)] \tag{9}$$

$$\hat{c}_{l\,\text{or}\,r,\theta,\phi}(n) = c_{l\,\text{or}\,r,\theta,\phi}(n)w(n) \tag{10}$$

Finally, compute the minimum-phase, time-domain versions of $d_{l,\theta,\phi}(n)$ and $d_{r,\theta,\phi}(n)$ as

$$d_{\min_{l\,\text{or}\,r\,\theta\,\phi}}(n) \triangleq F^{-1}\,[\exp(F\{\hat{c}_{l\,\text{or}\,r,\theta,\phi}(n)\})] \tag{11}$$

1.2.2 Synthesis of Spatial Audio Using HRTFs

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

Formally suppose that one wants to process a monaural signal $x(n)$ such that it sounds as if it were located at azimuth θ and elevation ϕ Specifically, let $d_{\min_{l\,\theta\,\phi}}(n)$ and $d_{\min_{r\,\theta\,\phi}}(n)$ 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 $x_l(n)$ and $x_r(n)$ as follows, and present $x_l(n)$ and $x_r(n)$ to the left and right ears simultaneously over headphones. Here $n_{\text{ITD},\theta,\phi}$ is defined to be negative for sounds arriving at the left ear first

$$x_l(n) = x(n - n_{\text{ITD},\theta,\phi}) * d_{\min_{l,\theta,\phi}}(n) \tag{12}$$

$$x_r(n) = x(n) * d_{\min_{r,\theta,\phi}}(n) . \tag{13}$$

Several applications involve the 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 Eqs. (12) and (13), whereas delay lines are used to implement the time delay in Eq. (12). In order to synthesize moving sounds, HRTFs and ITDs are dynamically updated in time to correspond to new spatial locations. Fig. 7 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 HRTFs 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 shown in Fig. 7 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" [13]. 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) [14]. The synthesis of sounds with nonzero elevations is difficult. Also, since every individual has a unique set of HRTFs, a subject listening to a spatialized sound generated from a "generalized" HRTF set may not perceive the sound in the intended spatial location [15], [16].

In addition to its 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 a high-quality synthesis of virtual audio on low-cost generic computers is often not possible. Because HRTFs are typically measured at several hundred different spatial locations, there is a substantial amount of data that need to be stored, accessed, and processed quickly. In moving sound synthesis as depicted in Fig. 7, several "interpolated" HRTFs and ITDs may be required to produce a smoothly moving sound. How should we compute these interpolated quantities from a finite set of existing HRTFs and ITDs [17]–[20]?
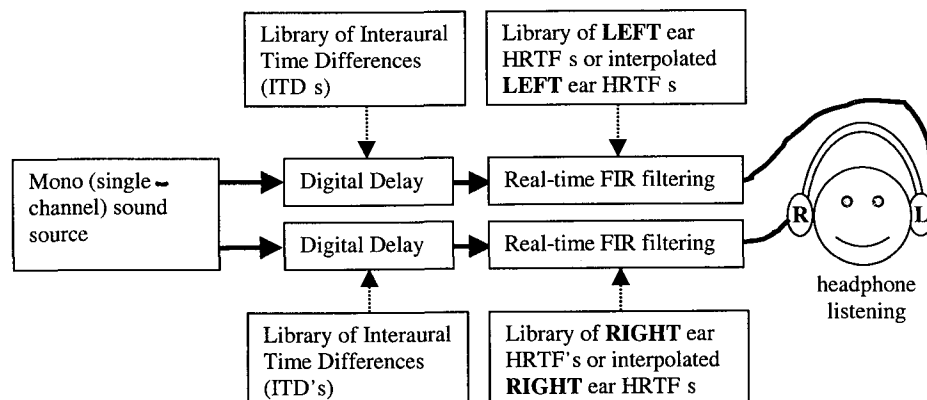
Many researchers believe that the solution to these

Fig. 7 Block diagram of simple HRTF-based spatial sound synthesis system Delays and FIR filters can be updated in real time to synthesize moving sound sources.

problems involves a deeper understanding of the perceptual structure of HRTF data. By investigating, modeling, and parameterizing the structure of HRTFs, researchers expect to link salient features of HRTFs, 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 HRTFs in order to alleviate problems with existing systems. For example, a low-order parameterization of HRTFs that maintains a sufficient number of perceptually relevant spectral cues could significantly lighten the computational and storage demands on hardware; understanding and preemphasizing the spectral cues corresponding to sounds in front and in back of a subject could reduce front–back confusions, and so on.

2 DIFFERENT REPRESENTATIONS OF HRTF DATA

There have been many attempts to understand the structure of HRTFs 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 measured empirically on 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 all the structures to be investigated are summarized next.

- *Types of HRTF Data*
  HRTFs derived from a spherical model of the head. The simplest analytical model of HRTFs 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 [21]. 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 HRTFs at any given frequency and spatial location. Since the spherical model of the head is a mathematical construct, many authors have compared these theoretical HRTFs to measured HRTFs [22], [3]. By looking for similarities between theoretical and measured data sets, researchers hope to learn how much of the HRTF structure is due only to the head effects. A mathematical derivation of HRTFs from the spherical head model can be found in [2], and in a recent paper Duda and Martens provide pseudo-code for the numerical com-

putation of the "theoretical" HRTFs used in this paper [18]. 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.

HRTFs measured from a human subject. HRTFs were measured for several subjects by 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 Golay code method [9]. Complementary Golay codes were used as the stimulus signals and were presented from a loudspeaker approximately 1.5 m from the subjects' heads. Left- and right-ear magnitude responses were measured at 400 different azimuth–elevation locations. Although spaced irregularly, these locations were roughly $10°–15°$ apart in the azimuth and elevation directions. The sampling rate was 50 kHz, the resolution of the data taken was 16 bit, and a 512-point FFT was used to compute the frequency response at each location.

- *Two Well-Known Structures in HRTF Data*
  Diffraction effects in the contralateral HRTFs due to the head. The spherical model of the head predicts HRTFs that 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 [21]. 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 "bright spots" in his analyses, since there is a local maximum in energy transmission that occurs in these areas on the contralateral side of the head [3]. An analytical derivation of the effect can be found in [2].

Elevation effects in the ipsilateral HRTFs 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 that are similar to the characteristic lengths of the pinna, and therefore interact strongly with the pinna [22]. There are noticeable patterns in HRTF data near these frequencies, which have been shown psychophysically to be correlated with the perception of elevation [19].

2.1 Frequency-Domain Representation of HRTF Data

HRTF data are perhaps most easily understood in the frequency domain, where the magnitude responses of various HRTFs 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 or elevation [2], [23], [24]. Consequently many signal processing techniques have taken advantage of this domain in attempts to parameterize or compute interpolated HRTFs. For example, principal components analysis of frequency-domain HRTFs has been performed [24], and pole–zero modeling of frequency-domain HRTFs has also been attempted [25]. In addition, several different schemes have been introduced to calculate interpolated HRTFs in the frequency domain, and a good summary is provided in [20].

The frequency-domain representation of HRTFs clearly shows some major differences between theoretical and measured HRTF data sets. Figs. 8–11 show left- and right-ear frequency-domain HRTFs computed from a spherical model of the head, as well as HRTFs measured from a human subject. Figs. 8, 10 and Figs. 9, 11 show HRTFs for locations in the horizontal and median planes (elevation = 0°, azimuth = 0°), respectively. Whereas the theoretical HRTFs are noiseless, the signal-to-noise ratio (SNR) of the measured HRTFs seems to be a function of spatial location. Specifically, the contralateral HRTFs in Figs. 10 and 11 seem to be less smooth than the ipsilateral HRTFs, suggesting that the SNR for ipsilateral HRTFs is generally higher than that for contralateral HRTFs. This is reasonable, since the contralateral ear typically receives less power than the ipsilateral ear during the measurement process. In general,

the measured HRTFs are also more complex than the theoretical HRTFs 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 HRTFs.

Elevation effects can also be seen in the frequency domain. For example, in the measured data in Fig. 11 there is a notch at 7 kHz that migrates upward in frequency as the 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 Fig. 9 also show how the head and ear locations can produce elevation-dependent features in HRTFs. Fig. 9 shows that higher elevations have a slight high-pass characteristic, whereas lower elevations have a slight low-pass characteristic.

Diffraction effects can be seen most easily in Fig. 8, where theoretical HRTFs are plotted in the frequency domain as a function of azimuth. Note the rippling shapes in the contralateral HRTFs corresponding to the azimuths +90° and −90° for the left and right ears, respectively. The HRTFs 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
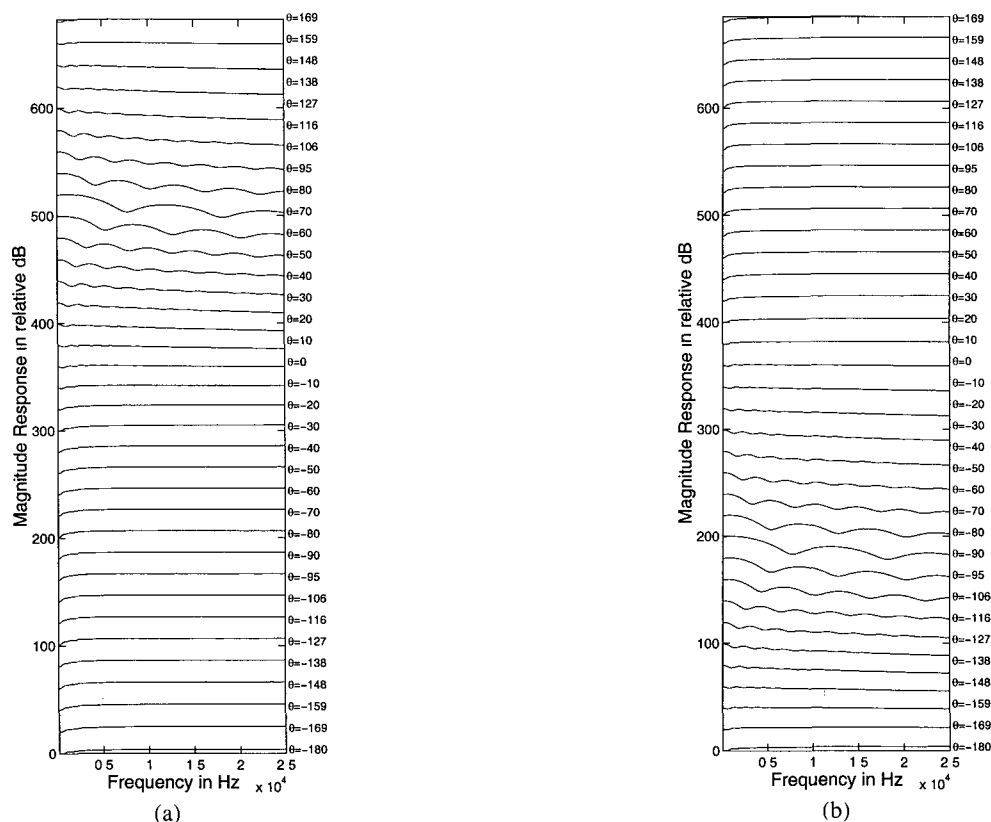

(a)

(b)

Fig. 8 Frequency-domain comparison of theoretical HRTFs (computed from spherical model of head) as a function of azimuth in horizontal plane (elevation = 0°). Diffraction effects for low frequencies can be seen on contralateral side at azimuths +90° and −90° for left and right ears, respectively. Amplification effects due to ear's proximity to head can be seen on ipsilateral side. Contralateral HRTFs are more complex than ipsilateral HRTFs. (a) Left-ear theoretical HRTF (b) Right-ear theoretical HRTF
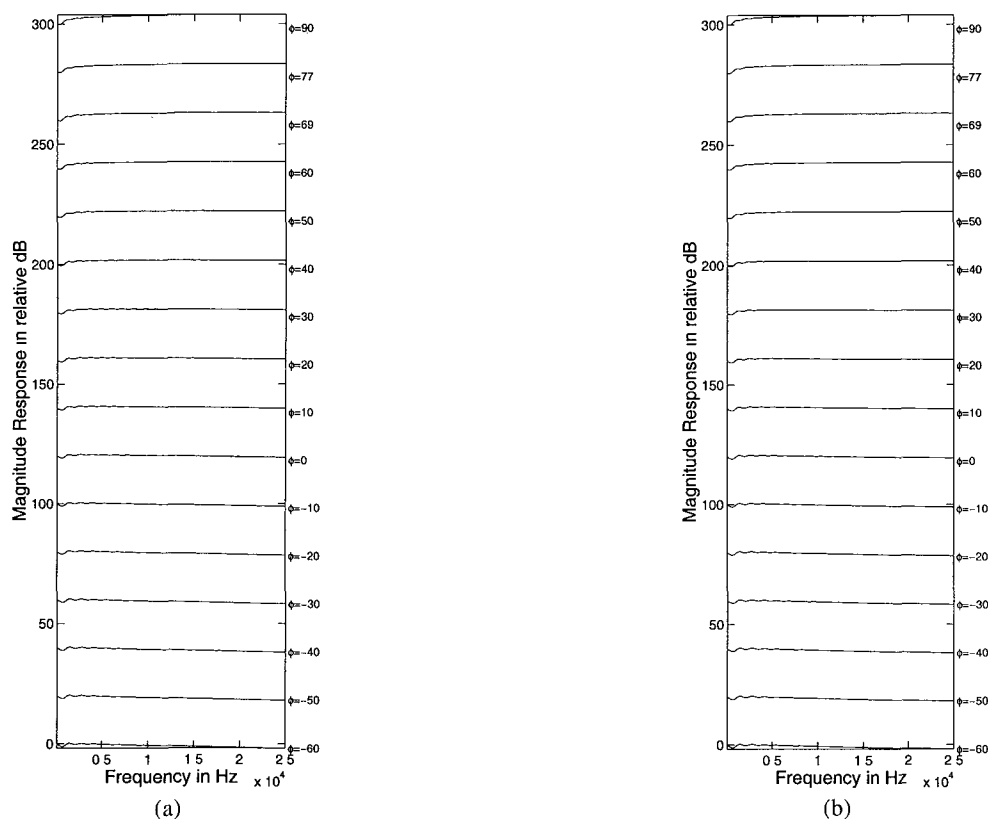
(a)

(b)

Fig. 9. Frequency-domain comparison of theoretical HRTFs (computed from spherical model of head) as a function of elevation in median plane (azimuth = 0°). Due to nonsymmetric ear locations on the head, there is a slight high-pass characteristic to HRTFs with higher elevations, and a slight low-pass characteristic to HRTFs with lower elevations (a) Left-ear theoretical HRTF. (b) Right-ear theoretical HRTF.
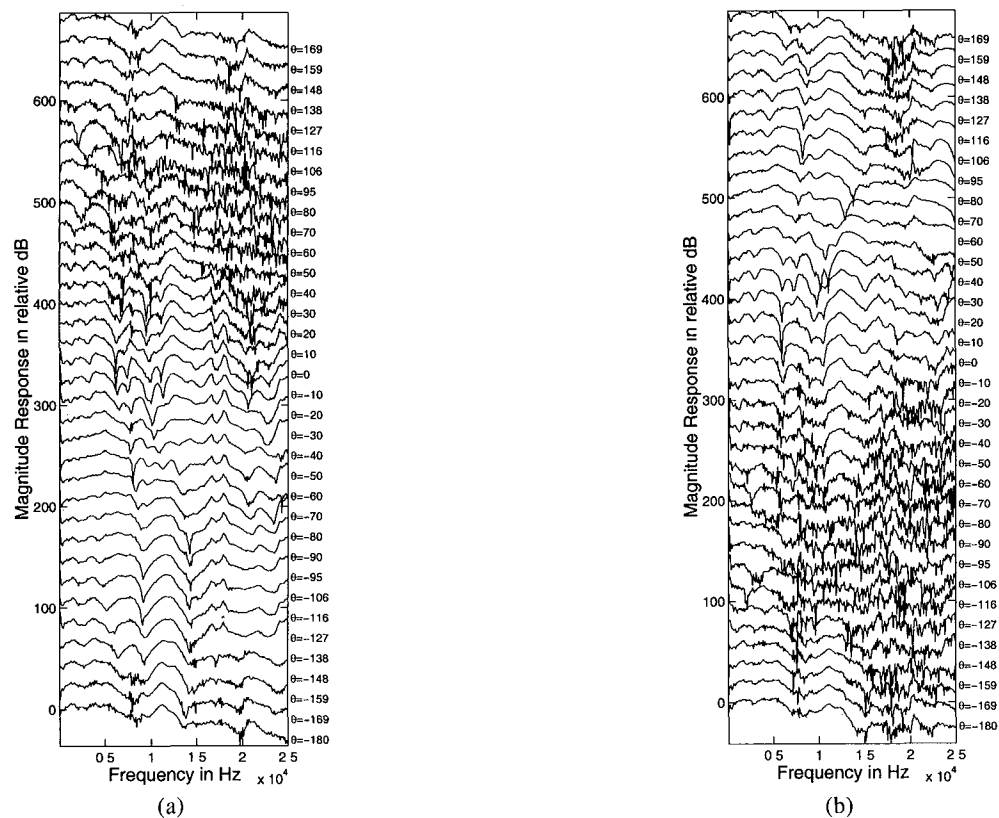

(a)

(b)

Fig. 10 Frequency-domain comparison of measured HRTFs as a function of azimuth in horizontal plane (elevation = 0°) These HRTFs are more complex than their theoretical counterparts in Fig. 7. Diffraction effects for low frequencies can be seen on contralateral side at azimuths +90° and −90° for left and right ears, respectively In general, contralateral HRTFs have a lower signal-to-noise ratio (SNR) than ipsilateral HRTFs (a) Left-ear measured HRTF (b) Right-ear measured HRTF.

also be seen in the measured HRTFs in Fig. 10. In addition, amplification effects are expected and can be seen for high frequencies in ipsilateral HRTFs due to reflection and the ear's proximity to the head, a rigid surface [21].

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 HRTFs are sometimes called head-related impulse responses (HRIRs) in some literature [22], [26]. Because the complexity of a spatialization system depends largely on the length of the HRIRs, 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) HRIRs in an effort to reduce the noise that the HRTF measurement process inevitably introduces [27].

Figs. 12–15 show left- and right-ear HRIRs computed from a spherical model of the head, as well as HRIRs measured from a human subject. Figs. 12, 13, and Figs. 14, 15 show HRIRs for locations in the horizontal and median planes (elevation = 0°, azimuth = 0°), respectively. Figs. 12 and 13 show that in general, a location that 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 ITDs and IIDs for sources with a larger absolute azimuth, or displacement from the median plane. In addition, the pronounced negative dip, or "overshoot," in some HRIRs in Figs. 12 and 13 indicates that high frequencies are boosted for these locations [22].

Diffraction effects and Shaw's "bright spot" can also be seen easily in Figs. 12 and 13. HRIRs corresponding to contralateral locations that lie in the direct shadow of the head have relatively higher amplitude initial peaks. For example, Fig. 12 shows how left-ear HRIRs 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 HRIRs of Figs. 14 and 15, as there is a slight difference in the arrival times for positive- and negative-elevation HRIRs.

A comparison of the theoretical and empirical data sets in each of Figs. 12–15 reveals that although the general patterns are similar between the two data sets, the measured data are much richer than the theoretical data The measured HRIRs contain many secondary peaks in addition to the initial peak, which the theoretical HRIRs 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 alone cannot predict.
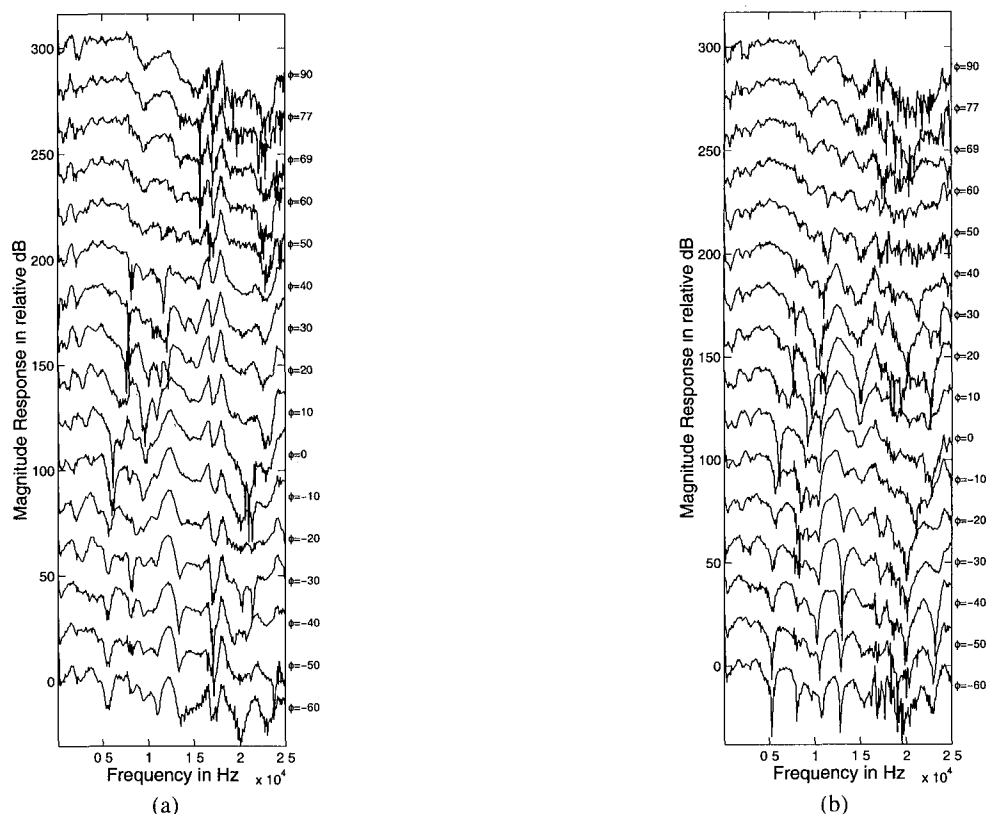

(a)   (b)

Fig 11 Frequency-domain comparison of measured HRTFs as a function of elevation in median plane (azimuth = 0°). These HRTFs are more complex than their theoretical counterparts in Fig. 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 HRTFs is due to pinna and torso interactions, which are not predicted in the spherical head model of HRTFs (a) Left-ear measured HRTF (b) Right-ear measured HRTF
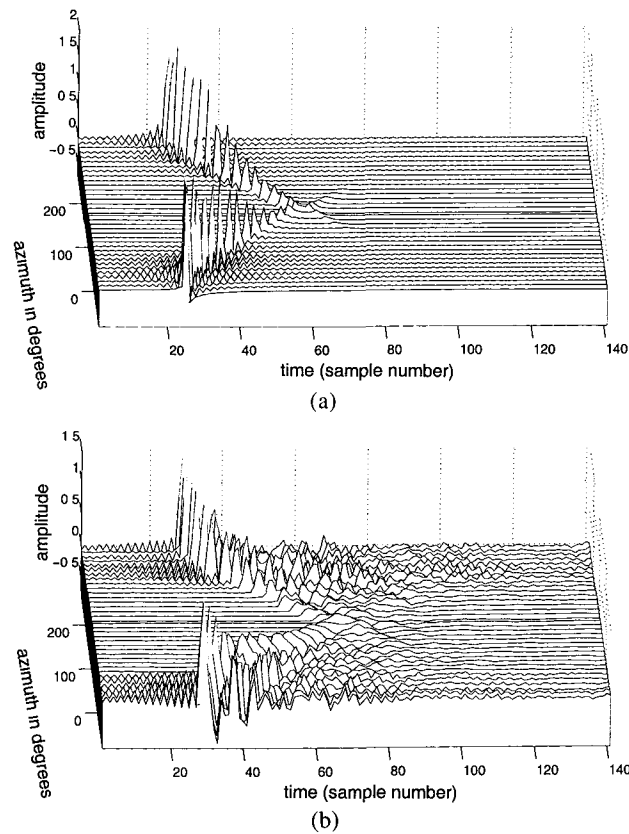
(a)

(b)

Fig. 12. Time-domain comparison of measured and theoretical left-ear HRIRs as a function of azimuth in horizontal plane (elevation = 0°) Significant energy arrives at left ear from some contralateral locations due to diffraction effects Note relatively large amplitude of initial peak in impulse responses corresponding to azimuth +90° (a) Theoretical left-ear HRIR impulse responses (b) Measured left-ear HRIR impulse responses
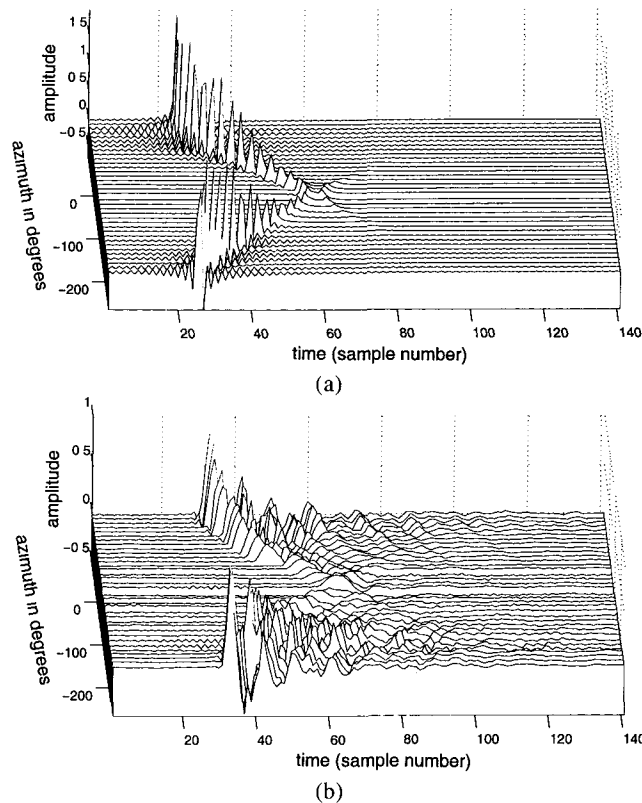

(a)

(b)

Fig. 13 Time-domain comparison of measured and theoretical right-ear HRIRs as a function of azimuth in horizontal plane (elevation = 0°). Significant energy arrives at right ear from some contralateral locations due to diffraction effects Note relatively large amplitude of initial peak in impulse responses corresponding to azimuth −90° (a) Theoretical right-ear HRIR impulse responses (b) Measured right-ear HRIR impulse responses.
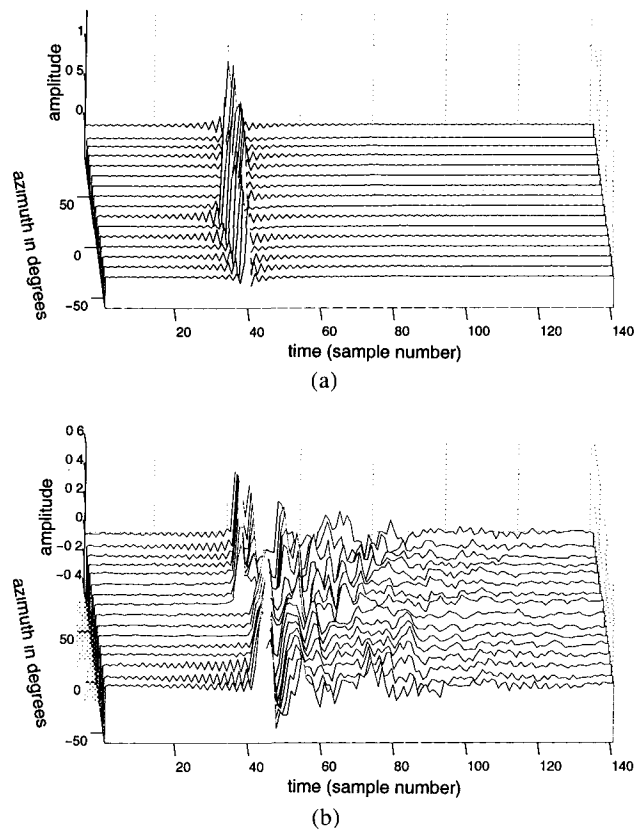
Fig 14. Time-domain comparison of measured and theoretical left-ear HRIRs as a function of elevation in median plane (azimuth = 0°). Note slight difference in arrival times associated with positive and negative elevations. (a) Theoretical left-ear HRIR impulse responses (b) Measured left-ear HRIR impulse responses
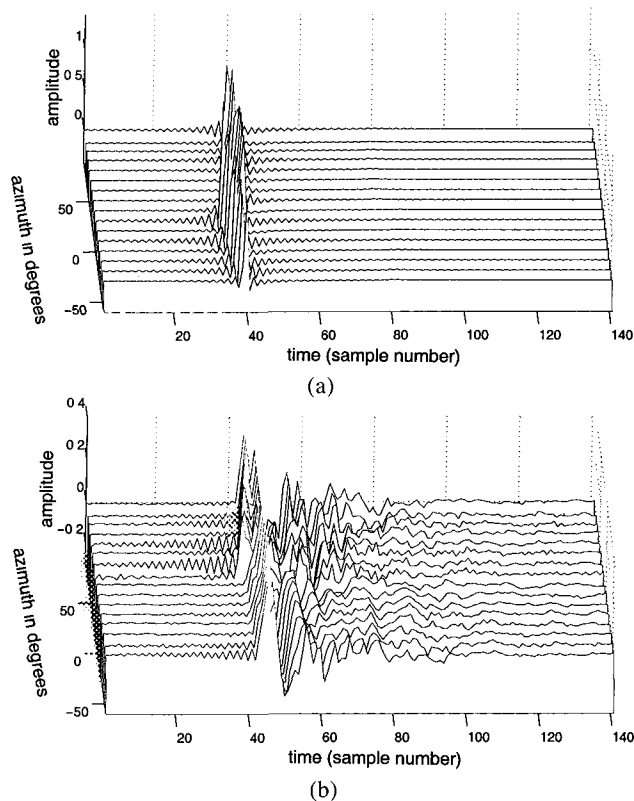

Fig 15 Time-domain comparison of measured and theoretical right-ear HRIRs as a function of elevation in median plane (azimuth = 0°) Note slight difference in arrival times associated with positive and negative elevations (a) Theoretical right-ear HRIR impulse responses (b) Measured right-ear HRIR impulse responses.

theory of directional bands, which states that certain narrow-band signals are correlated with perceptually preferred spatial directions. For example, the 6.8-kHz SFRSs in Fig. 18 corresponding to measured HRTFs 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 SFRSs in Fig. 19 corresponding to measured HRTFs 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 narrow-band 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 narrow-band sounds centered at 6 and 8 kHz as originating from positive and negative elevations, respectively, regardless of the actual free-field location of the source [19]. 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 SFRSs.

Diffraction effects are easily seen in SFRSs computed from both measured and theoretical HRTFs. The 1.9-kHz and 2.4-kHz SFRSs in Figs. 16 and 17 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 [3].

Comparison between theoretical and measured SFRSs again shows that the measured data are much richer than the theoretical data. Figs. 16–19 shows that the measured data have several hotspots in SFRSs which are not found in the theoretical data. Furthermore, the measured data in Fig. 16 show a local minimum on the contralateral side at lower elevations, unlike the theoretical data in Fig. 16. 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 SFRSs computed from measured data.
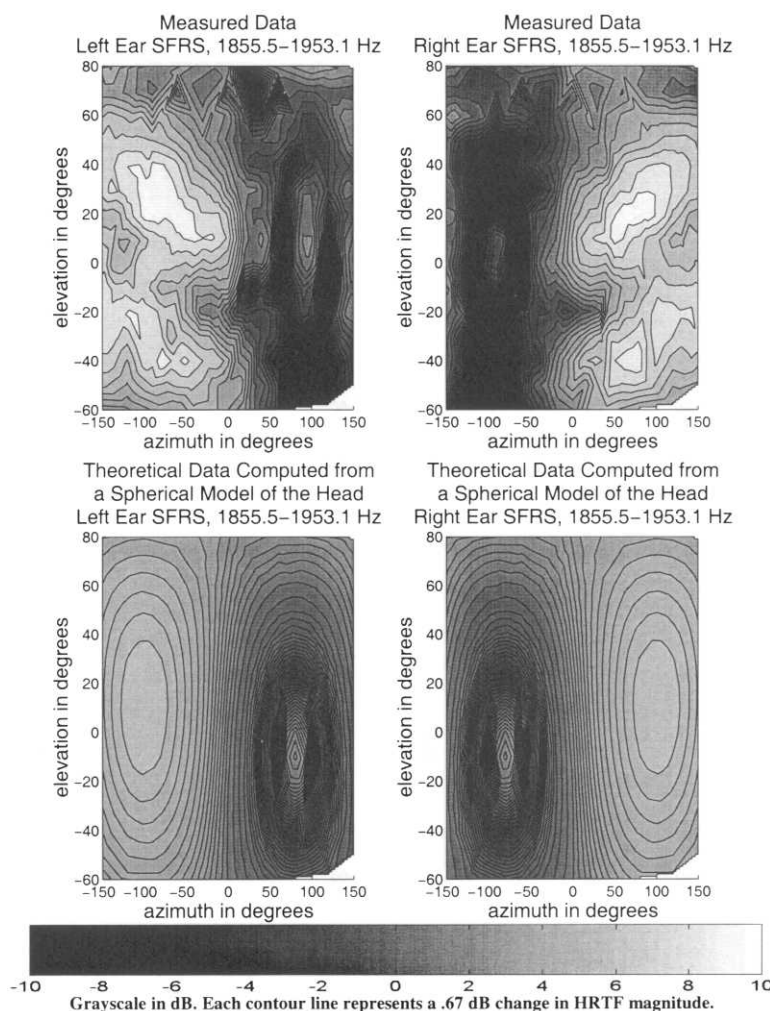

Fig. 17 Diffraction effects in theoretical and measured HRTFs at 2.4 kHz as expressed with SFRSs in spatial domain. Note local maxima, or "hotspots," on contralateral side in both measured and theoretical HRTFs near azimuths $+100°$, $-100°$ for left and right ears, respectively These hotspots are the "bright spots" discussed by Shaw [3]

2.3 Spatial-Domain Representation of HRTF Data[1]

HRTFs can be represented in the spatial domain in several different ways. Some authors plot ITDs and IIDs as a surface, as a function of azimuth and elevation [26], [28]. Others plot frequency-domain HRTFs with a common azimuth or elevation as a surface, where the sequential layout of the data by elevation or azimuth, respectively, highlights patterns effectively [29].

In this paper we focus on spatial representations, which plot the magnitude response of all HRTFs in a data set for a fixed frequency as a function of azimuth and elevation. The current authors call such spatial representations of HRTFs spatial frequency response surfaces (SFRSs) [17], [30]. Intuitively these graphs indicate how much energy the right and left ears receive at a fixed frequency as a function of spatial location. Typically, SFRSs exhibit several local maxima, or "hotspots," which correspond to spatial areas from which

the ears receive more energy than others. Some examples of SFRSs computed from both measured and theoretical HRTFs can be found in Figs. 16–19. Other examples of this style of HRTF representation can be found in [17], [20], [28], [30].

There have been several attempts to process HRTF data in the spatial domain. For example, principal components analysis (PCA) of HRTFs has been performed in the spatial domain as an attempt to parameterize HRTFs [28]. Spherical basis functions have also been used to parameterize HRTFs in the spatial domain [31]. There have been attempts to parameterize HRTFs in the spatial domain using a beamformer, where a virtual sensor array models the spatial and temporal characteristics of HRTFs simultaneously [32]. Spatially based HRTF interpolation methods have also been developed, which produce perceptually reasonable HRTFs [17], [19], [20].

Elevation perception can be linked to some of the local maxima or hotspots in SFRSs 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 narrow-band noise centered at that frequency. This is consistent with the

---

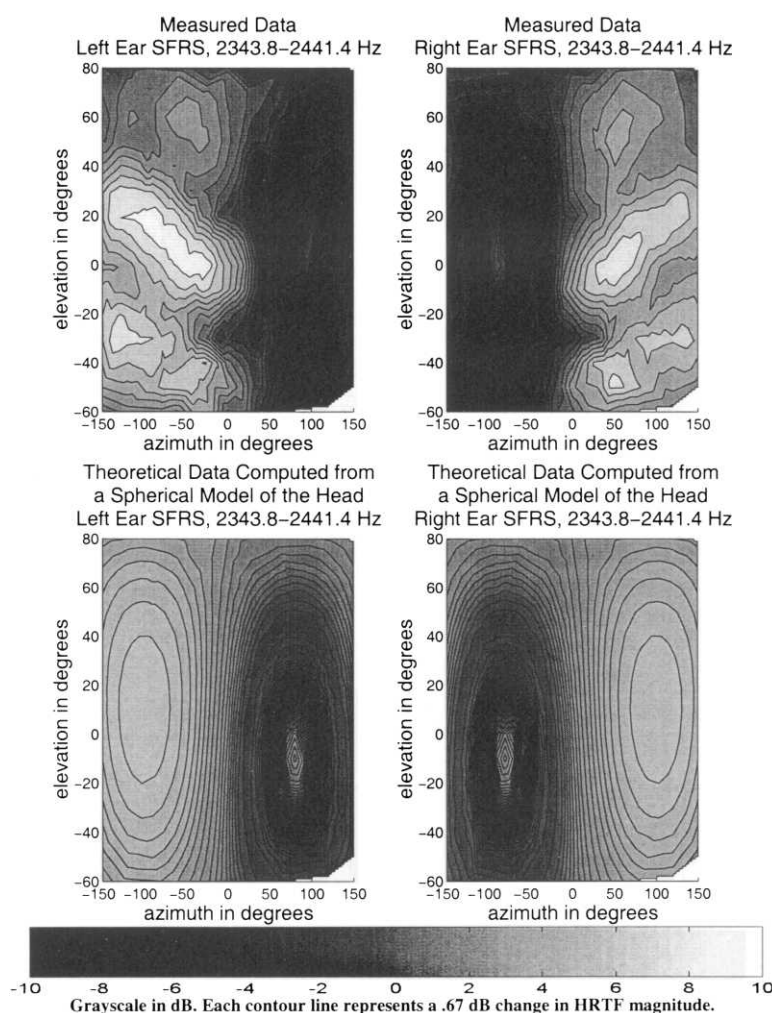[1] This section contains some material previously discussed in [17], [30]

Fig 16. Diffraction effects in theoretical and measured HRTFs at 1.9 kHz as expressed with SFRSs in spatial domain. Note local maxima, or "hotspots," on contralateral side in both measured and theoretical HRTFs near azimuths +100°, −100° for left and right ears, respectively These hotspots are the "bright spots" discussed by Shaw [3].

3 CONCLUSIONS

This paper introduced HRTFs and discussed their role in the synthesis of spatial audio over headphones. The need for spectral cues/HRTFs was motivated by an inability of the duplex theory to resolve spatial locations uniquely from ITDs and IIDs alone. We discussed the measurement of HRTFs 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, the lack of externalization of spatialized sound, and front–back confusions of spatialized sounds.

This paper also compared HRTFs measured from a human subject and HRTFs 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 HRTFs were more complex and contained many pinna and torso related effects which the theoretical HRTFs did not contain.

4 ACKNOWLEDGMENT

5 REFERENCES

[1] L. Rayleigh, "On Our Perception of Sound Direction," *Philosoph. Mag.*, vol. 13 (1907).

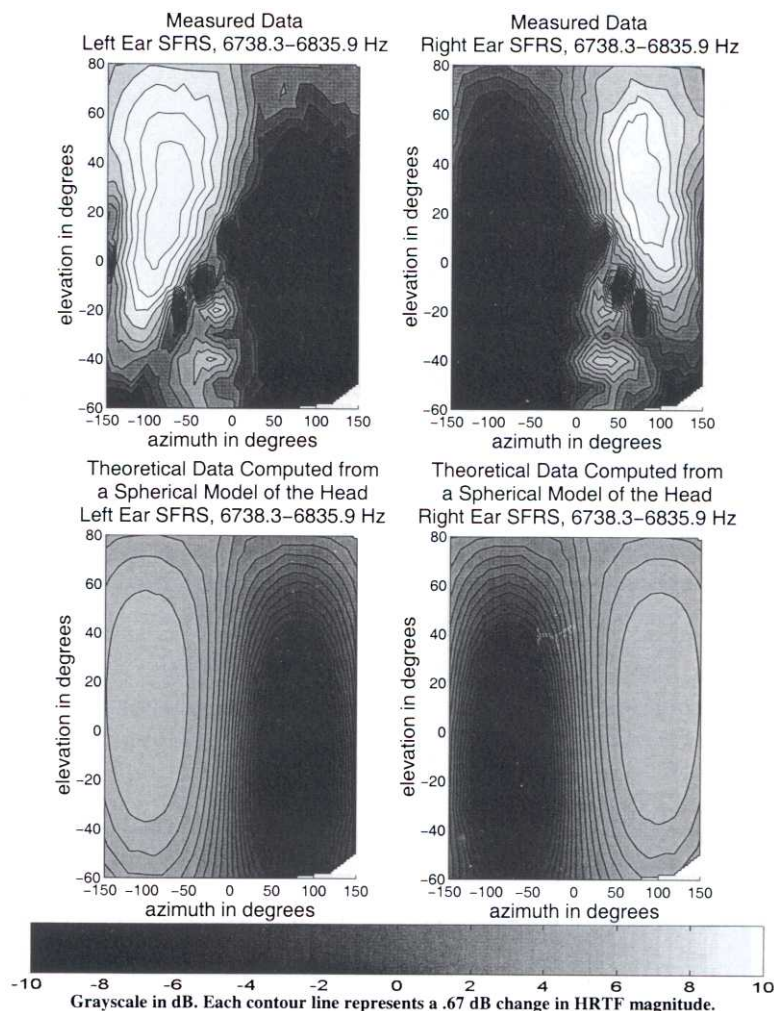[2] J. Blauert, *Spatial Hearing* (MIT Press, Cam-

Fig. 18. Elevation effects in theoretical and measured HRTFs at 6.8 kHz as expressed with SFRSs in spatial domain. Note prominent "hotspot," which occurs at a positive elevation on the ipsilateral side This hotspot corresponds to a preferred positive perceptual elevation for narrow-band sounds centered near this frequency.

bridge, MA, 1983).

[3] E. A. G. Shaw, "The External Ear," in *Handbook of Sensory Physiology*, vol. 1, *Auditory System, Anatomy Physiology Ear* (Springer, New York, 1974).

[4] F. L. Wightman and D. J. Kistler, "Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis," *J. Acoust. Soc. Am.*, vol. 85 (1989 Feb.).

[5] Y. Kahana et al., "Numerical Modelling of the Transfer Functions of a Dummy-Head and of the External Ear," in *Proc. 16th Audio Eng. Soc. (AES) Int. Conf. on Spatial Sound Reproduction* (Rovaniemi, Finland, 1999).

[6] A. Kulkarni et al., "On the Minimum-Phase Approximation of Head-Related Transfer Functions," in *Proc. 1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics* (IEEE catalog no. 95TH8144).

[7] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1989).

[8] S. Foster, "Impulse Response Measurement Using Golay Codes," in *Proc. 1986 Int. Conf. on Acoustics, Speech, and Signal Processing* (ICASSP86) (New York, NY).

[9] Zhou et al., "Characterization of External Ear Impulse Responses Using Golay Codes," *J. Acoust. Soc. Am.*, vol. 92, pt. 1 (1992 Aug.).

[10] M. J. E. Golay, "Complementary Series," *IRE Trans. Inform. Theory*, vol. IT-7 (1961).

[11] J. C. Middlebrooks et al., "Directional Sensitivity of Sound-Pressure Levels in the Human Ear Canal," *J. Acoust. Soc. Am.*, vol. 86 (1989 July).

[12] J. Huopaniemi and J. O. Smith, "Spectral and Time-Domain Preprocessing and the Choice of Modeling Error Criteria for Binaural Digital Filters," in *Proc. 16th Audio Eng. Soc. (AES) Int. Conf. on Spatial Sound Reproduction* (Rovaniemi, Finland, 1999)

[13] D. Griesinger, "Objective Measures of Spatiousness and Envelopment," in *Proc. 16th Audio Eng. Soc. (AES) Int. Conf. on Spatial Sound Reproduction* (Rovaniemi, Finland, 1999).

[14] F. L. Wightman and D. J. Kistler, "Headphone Simulation of Free-Field Listening. II: Psychophysical Validation," *J. Acoust. Soc. Am.*, vol. 85 (1989 Feb.).

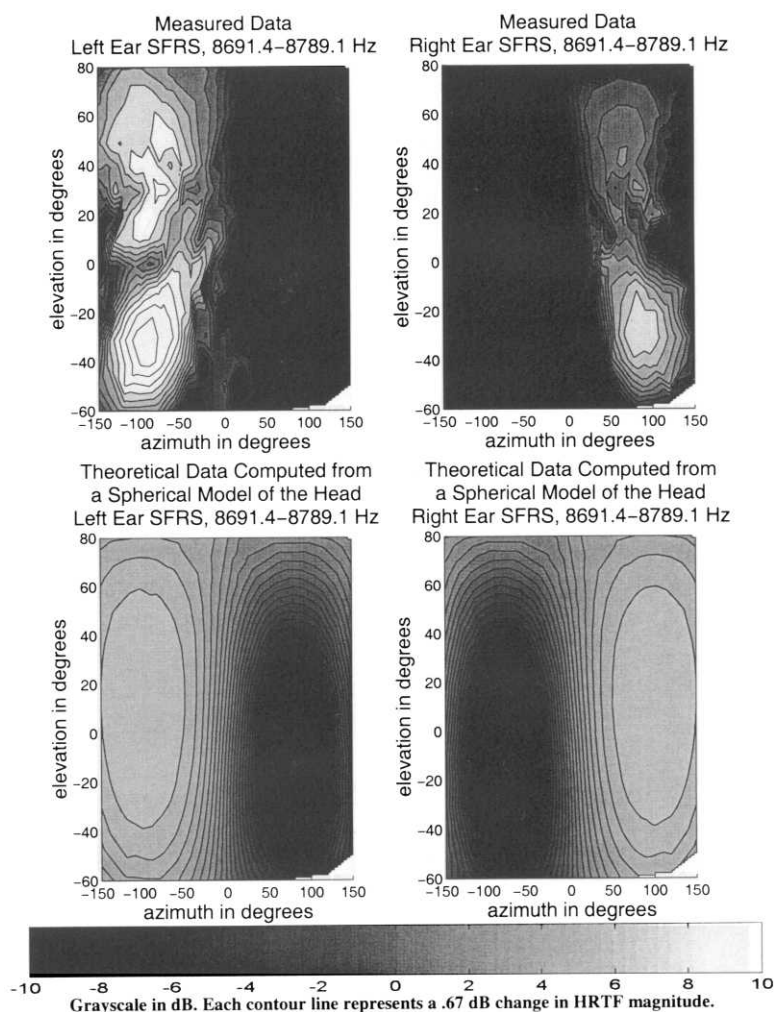[15] E. M. Wenzel et al., "Localization Using Nonindividualized Head-Related Transfer Functions," *J.*

Fig. 19 Elevation effects in theoretical and measured HRTFs at 8.7 kHz as expressed with SFRSs in spatial domain Note prominent "hotspot," which occurs at a positive elevation on the ipsilateral side This hotspot corresponds to a preferred positive perceptual elevation for narrow-band sounds centered near this frequency.

Acoust. Soc. Am., vol. 94 (1993 July).

[16] D. Pralong and S. Carlile, "The Role of Individualized Headphone Calibration for the Generation of High Fidelity Virtual Auditory Space," *J. Acoust. Soc. Am.*, vol. 100 (1996 Dec.).

[17] C. I. Cheng and G. H. Wakefield, "Spatial Frequency Response Surfaces (SFRS's): An Alternative Visualization and Interpolation Technique for Head-Related Transfer Functions (HRTF's)," in *Proc. 16th Audio Eng. Soc. (AES) Int. Conf. on Spatial Sound Reproduction* (Rovaniemi, Finland, 1999).

[18] A. Blommer and G. Wakefield, "A Comparison of Head Related Transfer Function Interpolation Methods," in *Proc. 1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics* (IEEE catalog no. 95TH8144).

[19] J. C. Middlebrooks, "Narrow-Band Sound Localization Related to External Ear Acoustics," *J. Acoust. Soc. Am.*, vol. 92 (1992 Nov.).

[20] K. Hartung et al., "Comparison of Different Methods for the Interpolation of Head-Related Transfer Functions," in *Proc. 16th Audio Eng. Soc. (AES) Int. Conf. on Spatial Sound Reproduction* (Rovaniemi, Finland, 1999).

[21] A. D. Pierce, *Acoustics* (Acoustic Society of America, Woodbury, NY, 1991).

[22] R. O. Duda and W. M. Martens, "Range Dependence of the Response of a Spherical Head Model," *J. Acoust. Soc. Am.*, vol. 104 (1998 Nov.).

[23] G. S. Kendall, "A 3-D Sound Primer: Directional Hearing and Stereo Reproduction," *Computer Music J.*, vol. 19 (Winter 1995).

[24] D. J. Kistler and F. L. Wightman, "A Model of Head-Related Transfer Functions Based on Principal Components Analysis and Minimum-Phase Reconstruction," *J. Acoust. Soc. Am.*, vol. 91 (1992 Mar.).

[25] M. A. Blommer, "Pole–Zero Modeling and Principal Component Analysis of Head-Related Transfer Functions," Ph.D. dissertation, University of Michigan, Dept. of Electrical Engineering and Computer Science, Systems, Div., Ann Arbor, MI (1996).

[26] Z. Wu et al., "A Time Domain Binaural Model Based on Spatial Feature Extraction for the Head-Related Transfer Function," *J. Acoust. Soc. Am.*, vol. 102 (1997 Oct.).

[27] W. L. Chapin, personal communication (1999).

[28] J. Chen et al., "A Spatial Feature Extraction and Regularization Model of the Head-Related Transfer Function," *J. Acoust. Soc. Am.*, vol. 97 (1995 Jan.).

[29] S. Carlile and D. Pralong, "The Location-Dependent Nature of Perceptually Salient Features of the Human Head-Related Transfer Functions," *J. Acoust. Soc. Am.*, vol. 95 (1994 June).

[30] C. I. Cheng and G. H. Wakefield, "Spatial Frequency Response Surfaces: An Alternative Visualization Tool for Head-Related Transfer Functions (HRTF's), in *Proc. 1999 Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP99)* (Phoenix, AZ, 1999).

[31] R. L. Jenison and K. Fissell, "A Spherical Basis Function Neural Network for Modeling Auditory Space," *Neural Comput.*, vol. 8 (1996).

[32] J. Chen et al., "External Ear Transfer Function Modeling: A Beamforming Approach," *J. Acoust. Soc. Am.*, vol. 92, pt. 1 (1992 Oct.).

6 BIBLIOGRAPHY

Abel, S., and V. H. Hay, "Sound Localization: The Interaction of Aging, Hearing Loss, and Hearing Protection," *Scand. Audiol.*, vol. 25, no. 1 (1996).

Avendano et al., "Modeling the Contralateral HRTF, in *Proc. 16th Audio Eng. Soc. (AES) Int. Conf. on Spatial Sound Reproduction* (Rovaniemi, Finland, 1999).

Begault, D. R., *3-D sound for Virtual Reality and Multimedia* (Academic Press, Cambridge, MA, 1994).

Begault, D. R., and E. M. Wenzel, "Headphone Localization of Speech," *Human Factors*, vol. 35, no. 2 (1993).

Bernstein, L. R., and C. Trahiotis, "Binaural Beats at High Frequencies: Listeners' Use of Envelope-Based Interaural Temporal and Intensitive Disparities," *J. Acoust. Soc. Am.*, vol. 99 (1996 Mar.).

Bernstein, L. R., and C. Trahiotis, "Binaural Interference Effects Measured with Masking-Level Difference and with ITD- and IID-Discrimination Paradigms," *J. Acoust. Soc. Am.*, vol. 98 (1995 July).

Buell, T. N., et al., "Lateralization of Bands of Noise as a Function of Combinations of Interaural Intensitive Differences, Interaural Temporal Differences, and Bandwidth," *J. Acoust. Soc. Am.*, vol. 95 (1994 Mar.).

Chandler, D. W., and W. Grantham, "Minimum Audible Movement Angle in the Horizontal Plane as a Function of Stimulus Frequency and Bandwidth, Source Azimuth, and Velocity," *J. Acoust. Soc. Am.*, vol. 91 (1992 Mar.).

Chen, J., et al., "Representation of External Ear Transfer Function via a Beamforming Model," in *Proc. 1991 Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)* (IEEE catalog no. 91CH2977-7).

Chen, J., et al., "Synthesis of 3D Virtual Auditory Space via a Spatial Feature Extraction and Regularization Model," in *Proc. IEEE Virtual Reality Ann. Int. Symp.* (Seattle, WA, 1993) (IEEE catalog 93CH3336-5).

Dasarathy, B. V., *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques* (IEEE Computer Society Press, Los Alamitos, CA, 1991).

Duda, R. O., et al., "An Adaptable Ellipsoidal Head Model for the Interaural Time Difference," in *Proc. 1999 Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP99)* (Phoenix, AZ).

Endsley, M. R., and A. S. Rosiles, "Auditory Localization for Spatial Orientation," *J. Vestib. Research*, vol. 5, no. 6 (1995).

Fay, R. R., and A. N. Popper, *Comparative Hearing: Mammals* (Springer, New York, 1994).

Fisher, N. I., et al., *Statistical Analysis of Spherical Data* (Cambridge University Press, New York, 1987).

Fisher, N. I., *Statistical Analysis of Circular Data* (Cambridge University Press, New York, 1993).

Giguère, C., and S. M. Abel, "Sound Localization:

Effects of Reverberation Time, Speaker Array, Stimulus Frequency, and Stimulus Rise/Decay," *J. Acoust. Soc. Am.*, vol. 94, pt. 1 (1993 Aug.).

Gilkey, R. H., "Some Considerations for the Design of Auditory Displays," in *Proc. 1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics* (IEEE catalog no. 95TH8144).

Gilkey, R. H., and T. R. Anderson, "The Accuracy of Absolute Localization Judgments for Speech Stimuli," *J. Vestib. Research*, vol. 5, no. 6 (1995).

Green, D. M., and J. A. Swets, *Signal Detection Theory and Psychophysics* (Krieger, New York, 1974).

Hartmann, W. M., and B. Rakerd, "On the Minimum Audible Angle—A Decision Theory Approach," *J. Acoust. Soc. Am.*, vol. 85 (1989 May).

Hawkins, H. L., et al., Eds., *Auditory Computation* (Springer, New York, 1996).

Heller, L. M., and C. Trahiotis, "Extents of Laterality and Binaural Interference Effects," *J. Acoust. Soc. Am.*, vol. 99 (1996 June).

Jenison, R. L., "A Spherical Basis Function Neural Network for Approximating Acoustic Scatter," *J. Acoust. Soc. Am.*, vol. 99 (1996 May).

Jenison, R. L., "A Spherical Basis Function Neural Network for Pole–Zero Modeling of Head-Related Transfer Functions," in *Proc 1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics* (IEEE catalog no. 95TH8144).

Jenison, R. L., and K. Fissell, "A Comparison of the von Mises and Gaussian Basis Functions for Approximating Spherical Acoustic Scatter," *IEEE Trans. Neural Networks*, vol. 6 (1995 Sept.).

Loomis, J. M., "Some Research Issues in Spatial Hearing," in *Proc. 1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics* (IEEE catalog no. 95TH8144).

Martin, K. D., "Estimating Azimuth and Elevation from Interaural Difference," in *Proc. 1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics* (IEEE catalog no. 95TH8144).

Middlebrooks, J. C., and D. M. Green, "Directional Dependence of Interaural Envelope Delays," *J. Acoust. Soc. Am.*, vol. 87 (1990 May).

Morse, P. M., *Vibration and Sound* (McGraw-Hill, New York, 1948).

Musicant, A. D., and R. A. Butler, "Influence of Monaural Spectral Cues on Binaural Localization," *J. Acoust. Soc. Am.*, vol. 77 (1985 Jan.).

Nandy, D., and J. Ben-Arie, "An Auditory Localization Model Based on High-Frequency Spectral Cues," *Ann. Biomed. Eng.*, vol. 24 (1996).

Oldfield, S. R., and S. P. A. Parker, "Acuity of Sound Localisation: A Topography of Auditory Space. I. Normal Listening Conditions," *Perception*, vol. 13 (1984).

Oldfield, S. R., and S. P. A. Parker, "Acuity of Sound Localisation: A Topography of Auditory Space. II. Pinna Cues Absent," *Perception*, vol. 13 (1984).

Papoulis, A., *Probability, Random Processes, and Stochastic Processes* (McGraw-Hill, New York, 1991).

Perrett, S., and W. Noble, "The Contribution of Head Motion Cues to Localization of Low-Pass Noise," *Perception and Psychophys.*, vol. 59, no. 7 (1997).

Perrett, S., and W. Noble, "The Effect of Head Rotations on Vertical Plane Sound Localization," *J. Acoust. Soc. Am.*, vol. 102 (1997 Oct.).

Powell, M. J. D., "Radial Basis Functions for Multivariable Interpolation: A Review," in *Algorithms for Approximation* (Oxford University Press, New York, 1987).

Rao, K. R., and J. Ben-Arie, "Optimal Head Related Transfer Functions for Hearing and Monaural Localization in Elevation: A Signal Processing Design Perspective," *IEEE Trans. Biomed. Eng.*, vol. 43 (1996 Nov.).

Speyer, G., and M Furst, "A Model-Based Approach for Normalizing the Head Related Transfer Function," in *Proc. 1996 19th Conv. of Electrical and Electronics Engineers in Israel* (Jerusalem, Israel) (IEEE catalog no. 96TH8190).

Stern, R., et al., "Lateralization of Complex Binaural Stimuli: A Weighted-Image model," *J. Acoust. Soc. Am.*, vol. 84 (1988 July).

Stern, R., et al., "Lateralization of Rectangularly Modulated Noise: Explanations for Counterintuitive Reversals," *J. Acoust. Soc. Am.*, vol. 90, pt. 1 (1991 Oct.).

Stern, R. M., and G. D. Shear, "Lateralization and Detection of Low-Frequency Binaural Stimuli. Effects of Distribution of Internal Delay," *J. Acoust. Soc. Am.*, vol. 100, pt. 1 (1996 Oct.).

Therrien, C. W., *Decision, Estimation, and Classification* (Wiley, New York, 1989).

Trahiotis, C., and L. S. Bernstein, "Lateralization of Bands of Noise and Sinusoidally Amplitude-Modulated Tones: Effects of Spectral Locus and Bandwidth," *J. Acoust. Soc. Am.*, vol. 79 (1996 June).

Trahiotis, C., and R. M. Stern, "Across-Frequency Interaction in Lateralization of Complex Binaural Stimuli," *J. Acoust Soc. Am.*, vol. 96 (1994 Dec.).

Van Trees, H. L., *Detection, Estimation, and Modulation Theory*, pt. I. (Wiley, New York, 1968).

Van Veen, B. D., and R. L. Jenison, "Auditory Space Expansion Via Linear Filtering," *J. Acoust. Soc. Am.*, vol. 90 (1991 July).

Wenzel, E. M., "The Relative Contribution of Interaural Time and Magnitude Cues to Dynamic Sound Localization," in *Proc. 1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics* (IEEE catalog no. 95TH8144).

Wightman, F. L., and D. J. Kistler, "The Dominant Role of Low-Frequency Interaural Time Differences in Sound Localization," *J. Acoust. Soc. Am.*, vol. 91 (1992 Mar.).

Wotton, J. M., and R. L. Jenison, "A Backpropagation Network Model of the Monaural Localization Information Available in the Bat Echolocation System," *J. Acoust. Soc. Am.*, vol. 101, pt. 1 (1997 May).

Yost, W. A., and G. Gourevitch, Eds., *Directional Hearing* (Springer, New York, 1987).

Zahorik et al., "On the Discriminability of Virtual and Real Sound Sources," in *Proc. 1995 IEEE ASSP workshop on Applications of Signal Processing to Audio

and Acoustics (IEEE catalog no. 95TH8144).

Ziomek, L. J., *Fundamentals of Acoustic Field The-*

ory and Space–Time Signal Processing (CRC Press, Ann Arbor, MI, 1995).

THE AUTHORS

C I. Cheng

G H Wakefield
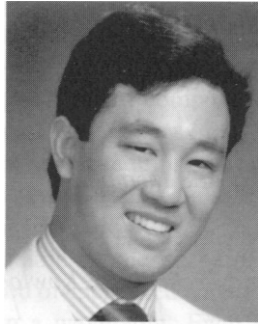
Corey I. Cheng was born in 1972. He studied physics at Harvard University from which he received a B.A in 1994, electroacoustic music at Dartmouth College from which he received an M A in 1996, and electrical engineering at the University of Michigan where he received an M.S.E. in 1998. He is currently a candidate for the Ph.D. in electrical engineering at the University of Michigan. Mr. Cheng has worked as a signal processing engineer for Fujitsu-Ten Corporation, the University of Michigan School of Business Administration, and the Naval Submarine Medical Research Laboratory, Groton, CT His technical research interests include spatial audio and head-related transfer functions (HRTFs), applications of wavelets to electroacoustic music, and econosonometrics, the study of sound level, trade volume, and price indices in "open cry-out" commodities trading pits. In addition, his electroacoustic music has appeared at various International Computer Music Conferences (ICMC) as well as at the Society of Electro-Acoustic Music in the United States (SEAMUS) Mr Cheng is a student member of the AES, IEEE, ICMA, SEAMUS, and is also a proud member of the University of Michigan ballroom dance team
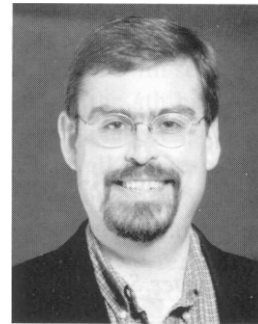
Gregory H Wakefield received a B A in mathematics and psychology in 1978, an M S E E in 1982, a Ph D in electrical engineering in 1985, and a Ph D in psychology in 1988, all from the University of Minnesota, Minneapolis, MN Dr Wakefield joined the faculty of the University of Michigan in 1986, where he is currently an Associate Professor of Electrical Engineering and Computer Science He also holds appointments in the Medical School (Otolaryngology) and in the School of Music (Vocal Arts Division and Performing Arts and Technology). He is the cofounder and codirector of the MusEn Project at the University of Michigan His main research interests are at the intersection of acoustics and signal processing, with applications drawn from auditory perception, vocal production, audio, music, and sound quality engineering Dr. Wakefield is a former NSF Presidential Young Investigator and a Millennium Medalist of the IEEE He consults regularly in sound quality engineering, with clients in the automotive, telecommunications, music, hearing aid, and sound appliance industries