Modeling distance-dependent individual head-related transfer functions in the horizontal plane using frontal projection headphones

Sunder, Kaushik; Gan, Woon-Seng; Tan, Ee-Leng

2015

Sunder, K., Gan, W.-S., & Tan, E.-L. (2015). Modeling distance-dependent individual head-related transfer functions in the horizontal plane using frontal projection headphones. *The Journal of the Acoustical Society of America*, 138(1), 150-171.

<https://hdl.handle.net/10356/81897>

<https://doi.org/10.1121/1.4919347>

© 2015 Acoustical Society of America. This paper was published in *The Journal of the Acoustical Society of America* and is made available as an electronic reprint (preprint) with permission of Acoustical Society of America. The published version is available at: [<http://dx.doi.org/10.1121/1.4919347>]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.

Downloaded on 26 Apr 2023 21:47:06 SGT

Modeling distance-dependent individual head-related transfer functions in the horizontal plane using frontal projection headphones

Kaushik Sunder,^{a)} Woon-Seng Gan, and Ee-Leng Tan

Digital Signal Processing Lab, School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore 639798, Singapore

(Received 24 August 2014; revised 10 April 2015; accepted 16 April 2015; published online 8 July 2015)

The veracity of virtual audio is degraded by the use of non-individualized head-related transfer functions (HRTFs) due to the introduction of front-back, elevation confusions, and timbral coloration. Hence, an accurate reproduction of spatial sound demands the use of individualized HRTFs. Measuring distance-dependent individualized HRTFs can be extremely tedious, since it requires precise measurements at several distances in the proximal region (<1 m) for each individual. This paper proposes a technique to model distance-dependent individualized HRTFs in the horizontal plane using “frontal projection headphones playback” that does not require individualized measurements. The frontal projection headphones [Sunder, Tan, and Gan (2013). *J. Audio Eng. Soc.* **61**, 989–1000] project the sound directly onto the pinnae from the front, and thus inherently create listener’s idiosyncratic pinna cues at the eardrum. Perceptual experiments were conducted to investigate cues (auditory parallax and interaural level differences) that aid distance perception in anechoic conditions. Interaural level differences were identified as the prominent cue for distance perception and a spherical head model was used to model these distance-dependent features. Detailed psychophysical experiments revealed that the modeled distance-dependent individualized HRTFs exhibited localization performance close to the measured distance-dependent individualized HRTFs for all subjects. © 2015 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4919347>]

[ELP]

Pages: 150–171

I. INTRODUCTION

The challenge of delivering realistic spatial audio through headphones has remained at the epicenter of 3D audio research for many decades now. Binaural audio is synthesized by using special spatial filters known as the head-related transfer functions (HRTFs) (Blauert, 1997). The spectral features in HRTFs are highly dependent on the listeners’ anthropometry. Rendering non-individualized binaural audio degrades the quality of the perceived spatial sound by introducing localization errors in the form of front-back confusions, up-down reversals, in-head localization, and timbral coloration. Thus, use of individualized HRTFs is recommended to alleviate these issues (Wenzel *et al.*, 1993; Wightman and Kistler, 1993; Begault, 2000; Kim and Choi, 2005; Morimoto and Ando, 1980). It is an immense task to measure these individualized HRTFs, as they demand very precise and tedious experiments to be carried out for a number of loudspeaker positions around the head. Moreover, unlike the far-field HRTFs (Mehrgardt and Mellert, 1977; Middlebrooks *et al.*, 1989; Wightman and Kistler, 1989; Carlile and Pralong, 1994; Møller *et al.*, 1995; Algazi *et al.*, 2001a), the near-field (<1 m) HRTFs vary significantly with distance (Brungart and Rabinowitz, 1999). The aim of this paper is to model the anechoic distance-dependent individualized HRTFs in the horizontal plane using frontal projection

headphones, without the need of individualized acoustical measurements (Fig. 1).

Unlike the measurement of HRTFs in the far-field, the near-field HRTF measurements require omni-directional acoustic point sources in order to avoid measurement errors due to scattering effects back and forth between the head and source. Researchers have developed various types of acoustic point sources to measure the near-field HRTFs. Some examples are, a combination of electro-dynamic driver and Tygon tubing (Brungart and Rabinowitz, 1999), micro-dodecahedral loudspeaker (Hosoe *et al.*, 2006), micro-dynamic-type loudspeaker (Hayakawa *et al.*, 2007), and a spark gap (Qu *et al.*, 2009). Other researchers, however, have used a Bose Acoustimass loudspeaker (Duda and

FIG. 1. (Color online) (a) Conventional side-projection (1), Frontal-projection (2) headphones. (b) Frontal projection headphones used in this study.

^{a)}Electronic mail: kaushik1@e.ntu.edu.sg

Martens, 1998) or a Tannoy loudspeaker (Calamia and Hixson, 1997), instead of a point source to measure the near-field HRTFs. Due to the requirement of an acoustic point source and the difficulty in obtaining individualized HRTFs; very few near-field HRTF databases are available. These issues, therefore, call for simpler techniques to obtain distance-dependent individualized HRTFs. In order to model the distance-dependent individualized HRTFs in this study, the perceptual cues that affect distance perception are first identified and simulated. Subsequently, a suitable HRTF individualization technique is developed.

A. Models for auditory distance perception

Binaural cues play a critical role for the perception of distance in anechoic conditions (Brungart and Rabinowitz, 1999). It is noted that with a decrease in source to head distance, the interaural level difference (ILD) rapidly increases, while the interaural time difference (ITD) remains more or less constant (Brungart and Rabinowitz, 1999). Several researchers have modeled the ILD and ITD cues that originated in the near-field (Stewart, 1911; Hartley and Fry, 1921; Hirsch, 1968; Molino, 1973; Tahara and Sakurai, 1974; Brungart, 2001; Hershkowitz and Durlach, 1969). Researchers have also identified the presence of auditory parallax effects in the near-field, which is a high-frequency phenomenon (Brungart, 1999b; Kim *et al.*, 2001). Kim *et al.* (2001) synthesized virtual sources up to a distance of 2 m using an “auditory parallax model” that is based on a simple geometrical model. In the study conducted by Kim *et al.* (2001), virtual sounds were synthesized using individualized HRTFs and played back over loudspeakers. The perceived distance of the virtual sound simulated by the auditory parallax model in this study (Kim *et al.*, 2001) was found to be similar to the free-field listening condition. However, several other studies have shown the lack of dependence of distance perception on high-frequency spectral cues (Zahorik, 2000; Brungart, 1999a; Sunder *et al.*, 2014), thereby leaving the exact role of auditory parallax cues for distance perception unclear, especially in the presence of other cues. This is further investigated in Sec. III.

Shaw and Teranishi (1968) first studied the pressure generated in an external ear replica and human ears by a nearby point source. The external replica of the ear used was made out of rubber holding the pinna, concha, and auditory meatus with dimensions comparable with those of real human ears. In order to model the near-field HRTFs, the distance-dependent spectral features in the proximal region are often analyzed by computing the sound pressure at the surface of a sphere (Duda and Martens, 1998; Rabinowitz *et al.*, 1993). Ze-Wei *et al.* (2012) used a spherical head with neck and torso to study the distance-dependent spectral features. Other researchers derived the spectral features in near-field HRTFs from the far-field HRTFs (Duraismami *et al.*, 2004; Menzies, 2009; Romblo and Cook, 2008; Kan *et al.*, 2009). Cues such as intensity (Zahorik, 2002a,b; von Békésy, 1949; Kim *et al.*, 2001; Simpson and Stanton, 1973), loudness (Stevens and Guirao, 1962; Zahorik and Wightman, 2001), and direct-to-reverberant energy ratio (DRR) (Shinn-Cunningham *et al.*,

2000a) have also been identified as important for distance perception (Begault, 2000). In this study, we consider modeling of HRTFs under anechoic conditions, since it is difficult to precisely simulate the distance-dependent DRR cues due to their heavy dependence on the listener’s acoustic space.

B. Individualization techniques

The idiosyncratic characteristics of HRTFs can be modeled from acoustical measurements, anthropometric features of the listener (Iida *et al.*, 2014), or by customizing generic HRTFs with some perceptual feedback. The most commonly used individualization technique involves measurement of listeners’ individualized HRTFs at different source positions (Møller *et al.*, 1995; Xu *et al.*, 2007), which is an extremely tedious process. Other techniques make use of optical descriptors, like magnetic resonance imaging, 3D mesh imagery, or a set of 2D images of the listener to derive the individualized anthropometrical features (Nicol, 2010; Rui *et al.*, 2013; Takemoto *et al.*, 2012). HRTFs are then obtained using numerical techniques, like the finite element method, the boundary element method (Katz, 2001; Kreuzer *et al.*, 2009), or by solving the problem of acoustic wave diffraction with the listener’s morphology (Lopez-Poveda and Meddis, 1996; Nicol, 2010; Kahana and Nelson, 2006). Middlebrooks (1999) developed the frequency scaling technique, in which subjects scaled the non-individualized HRTFs based on some anthropometric properties to obtain the individualized HRTFs. Tan and Gan (1998) customized the non-individualized HRTFs by asking listeners to tune the magnitude spectrum in various frequency bands until they achieved accurate localization. Another simple customization technique is to ask the listener to choose an HRTF from the database that suits his/her perception the best (Seeber and Fastl, 2003; Iwaya, 2006; Nicol, 2010). Individualized HRTFs can also be constructed using principal component analysis (Kistler and Wightman, 1992; Middlebrooks and Green, 1992) or independent component analysis techniques (Nicol, 2010). The weights used in these techniques are either derived from the source directions along the front-rear and up-down directions, or predicted from the descriptors of the anthropometric properties of the listener’s pinna (or head). All the above attempts to model distance-dependent individualized HRTFs require individualized far-field HRTFs, anthropometric features of the subject, or extensive near-field acoustical measurements. In this study, the individualization of HRTF is carried out using playback from the frontal projection headphones (Sunder *et al.*, 2013). An important advantage of the proposed individualization technique is that this technique does not require any individualized measurements.

In this paper, we model the distance-dependent individualized HRTFs in two stages, as follows:

- (1) The most prominent cues for auditory distance perception in anechoic conditions are first identified, when played back over headphones. The role of auditory parallax and ILD cues are of particular interest. Using each of auditory parallax and ILD cues, distance-dependent HRTFs are synthesized from the measured individualized far-field

and measured non-individualized far-field HRTFs. Perceptual experiments are carried out with each of these synthesized HRTFs and their results are compared with those obtained using the measured distance-dependent individualized HRTFs (Sec. III).

- (2) Once the important cues that affect distance perception are identified (Sec. III), the distance-dependent individualized HRTFs are modeled using a cascade of signal processing blocks. The distance-dependent spectral cues due to head effects are simulated by the spherical head model (Duda and Martens, 1998). Individualized pinna spectral cues are then modeled by playback through frontal projection headphones (Sunder et al., 2013).

Unlike other individualization methods, the frontal projection headphones playback method does not require the measurement of individualized far-field measurements; nor does it require the anthropometric data of human subjects. Frontal projection headphones (Fig. 1) have a unique structure, as they project sound from the front directly onto the pinnae of the listener during playback, unlike the conventional side projection headphones [Fig. 1(a)]. In Sunder et al. (2013), the authors observed that the frontal projection headphones inherently create individualized pinna spectral features of HRTF during sound playback. The frontal projection headphones also help in reducing front-back confusions to a large extent, thereby enhancing the frontal perception of the virtual sound (Sunder et al., 2013). In addition, headphone equalization, low-frequency compensation, and rear-directional filters are required to model the distance-dependent individualized HRTFs.

This paper is organized as follows. The description of the acoustical measurements of distance-dependent individualized and non-individualized HRTFs is presented in Sec. II. In Sec. III, the role of auditory parallax and ILD cues are investigated through objective and subjective analyses. The model for distance-dependent individualized HRTFs is presented in detail in Sec. IV. The discussion and conclusions are finally presented in Sec. V.

II. MEASUREMENT OF DISTANCE-DEPENDENT HRTFS

The two main objectives of measuring distance-dependent HRTFs in this study follow:

- (1) To analyze the role of auditory parallax and ILD cues for distance perception using HRTFs synthesized from measured far-field HRTFs.
- (2) To validate the modeled distance-dependent individualized HRTFs using frontal projection headphones by comparing it with the measured distance-dependent individualized HRTFs.

A. Experimental setup

All of the HRTF measurements (non-individualized and individualized) were carried out in the anechoic chamber of Nanyang Technological University, Singapore. Ideally, a point sound source is required for the measurement of

near-field HRTFs in order to avoid multiple diffractions back and forth between the head and source. However, the effect of source scattering was found to be negligible for measurements of near-field HRTFs beyond 0.2 m. This is true, if the source radius does not exceed 0.03–0.05 m (Guang-Zheng et al., 2008). In our study, due to non-availability of a point source, near-field HRTF measurements were carried out using a Cambridge Soundworks mini cube speaker of dimension less than 0.03 m in radius. The HRTFs were precisely measured at distances of 0.35, 0.45, 0.50, 0.60, 0.75, 0.80, 1, and 1.4 m from the center of the head. Since all the HRTFs in this study were measured for distances beyond 0.2 m, and the source radius was less than 0.03 m, the effect of source scattering on the measurements was assumed negligible (Guang-Zheng et al., 2008).

The non-individualized HRTFs were measured at the eardrum of a Bruel & Kjaer (B&K) 4128-C head and torso simulator (HATS), equipped with B&K 4158-C (left) and 4159-C (right) ear simulators. Each ear simulator contained a $\frac{1}{2}$ in. microphone, which was connected to a microphone amplifier with an adapter. In addition, individualized HRTFs were measured at the blocked-entrance of the ear canal (Møller et al., 1995) using a B&K 4101 binaural microphone for three subjects. The measurements were controlled by a computer located in a control room next to the anechoic chamber. The computer was connected via a LAN cable to the B&K PULSE analyzer system, which generated the excitation signal and simultaneously recorded responses from the microphone. Exponential sine sweep was chosen as the excitation signal in these measurements, since they provided a better signal to noise ratio at low frequencies (Müller and Massarani, 2001). The signals were recorded with the microphones placed at the recording point (eardrum or blocked-entrance) at a sampling rate of 65 536 Hz. The recorded signal was first down-sampled to 44 100 Hz and then processed using a custom-made MATLAB function to extract the impulse response. The ITDs were obtained by first calculating the lag that maximized the cross-correlation function between the head-related impulse responses of the two ears within the range of ± 1 ms. The interaural cross-correlation coefficient is defined as

$$\zeta(\tau) = \frac{\int_{-\infty}^{\infty} HRTF_L(f) HRTF_R^*(f) e^{-i\omega\tau} df}{\left\{ \left[\int_{-\infty}^{\infty} |HRTF_L(f)|^2 df \right] \left[\int_{-\infty}^{\infty} |HRTF_R(f)|^2 df \right] \right\}^{1/2}}, \quad (1)$$

where $HRTF_L$ and $HRTF_R$ denote the HRTFs for the left and right ears, respectively. ω denotes the angular frequency in radians. The time and frequency samples are represented by τ and f , respectively. $\zeta(\tau)$ is the cross-correlation function in time domain, with $-1 \leq \zeta(\tau) \leq 1$. Prior to the computation of cross-correlation, the impulse responses for both the ears were low-pass filtered with a cut-off frequency of 2 kHz, since ITD is predominantly a low-frequency cue.

The angular convention used throughout this paper is defined as follows: The coordinate system has its origin at

the midpoint of the interaural axis. The azimuth angles were measured with respect to the mid-sagittal axis. 0° corresponds to the direction directly to the front of the head, 90° facing the right ear, 180° is behind the head, 270° facing the left ear, and 360° (0°) in front of the head. HRTFs were measured in the horizontal plane at a resolution of 5° . For every source location in the horizontal plane, 10 experimental trials were conducted to ensure consistency and repeatability of the measurements. Reference measurements were made in the free-field with a $\frac{1}{2}$ in. free-field microphone (B&K 4134 pressure microphone) located at the position of the center of the dummy head with the dummy head removed [Fig. 2(a)]. Reference measurements were obtained for each of the eight distances, for which the HRTFs were measured. The magnitude of the HRTF at each source location was obtained by frequency-wise division of the amplitude spectrum (measured on the dummy head/human subject) at that location and the free-field reference measurement corresponding to the source distance. The division by the free-field reference response is necessary to compensate for both the non-uniform loudspeaker response and the sound pressure level drop due to distance.

A series of calibration techniques was carried out prior to each set of experiments. The dummy/human head was first adjusted to point directly to the center of interaural axis. Correct placement is extremely important in the near-field, as small angular deviations between the center of the head and the axis of rotation can severely corrupt the measurements. With the help of acoustical measurements, it was ensured that the source was placed directly in front of the head (0° azimuth and 0° elevation). The dummy/human head was adjusted until the magnitude of the ITD was reduced below $2 \mu\text{s}$. Furthermore, the headphone transfer functions (HPTFs) for the frontal projection headphones and conventional open-back headphones (Sennheiser HD600) were also measured at the blocked-entrance of the ear canal [Fig. 2(b)] (Møller *et al.*, 1995). The headphone response varies with repositioning over the head due to change in headphone-ear coupling (Kulkarni and Colburn, 2000). Therefore, headphones

were repositioned over the ear after each measurement and ten such measurements were obtained. An average of the ten measurements was taken as the representative headphone response. To model the distance-dependent HRTFs, important cues that affect distance perception over headphones had to be identified. The role of auditory parallax and ILD cues on distance perception was investigated using the measured HRTFs and is described in Sec. III.

III. ROLE OF AUDITORY PARALLAX AND ILD CUES ON DISTANCE PERCEPTION

Auditory parallax and ILD cues are often considered as relevant cues contributing to the near-field distance perception in the absence of reverberant and intensity cues. Though the effect of ILD cues on distance perception is well studied, the perceptual effects of auditory parallax cues are still unclear. In this section, we investigate the role of each of these cues—individually and combined with the help of listening experiments. The distance-dependent HRTFs are first synthesized using ILD and auditory parallax cues from the measured far-field HRTFs.

A. HRTF synthesis using auditory parallax cues

In the near-field, the angle between the source and the ears ($\theta_{L,R}$) deviates significantly from that between the source and the center of the head (θ). This is known as auditory parallax effect. The auditory parallax effect can be approximately modeled by a simple geometrical formulation that relates both these angles for any source to head distance (Kim *et al.*, 2001). The left and right ear HRTFs in the near-field for a source at distance d_n from the head and at an azimuth angle of θ can be obtained by remapping it from the far-field HRTFs at angles θ_L and θ_R , respectively,

$$\theta_L = \tan^{-1} \left(\frac{d_n \sin \theta + a}{d_n \cos \theta} \right), \quad (2)$$

$$\theta_R = \tan^{-1} \left(\frac{d_n \sin \theta - a}{d_n \cos \theta} \right), \quad (3)$$

FIG. 2. (a) Free-field response of the loudspeaker for various distances. (b) Headphone responses of Sennheiser HD600 and the frontal projection headphone measured on the blocked ear-canal of a B&K HATS dummy head.

$$HRTF_{AP,L}(d_n, \theta) = HRTF_L(d_f, \theta_L), \quad (4)$$

$$HRTF_{AP,R}(d_n, \theta) = HRTF_R(d_f, \theta_R), \quad (5)$$

where $HRTF_{AP}$ is the near-field HRTF synthesized using auditory parallax (AP) cues, a is the radius of the head, and d_f is the source to head distance in the far-field. θ_L and θ_R correspond to the angles of the sound source as seen from left and right ears in the near-field, respectively. All the angles (θ , $\theta_{L,R}$) are measured with respect to the mid-sagittal axis passing through the center of the head. It is important to note that Eqs. (4) and (5) model only the high-frequency distance-dependent features of the outer ear.

B. HRTF synthesis using ILD cues

The rapid rise of ILD cues with decreasing source to head distance in the near-field can be modeled using the spherical head model (Duda and Martens, 1998). These distance-dependent ILD features in the near-field are obtained by computing the ratio of the pressure at the surface of the sphere arising from a sound source at distance d_n to the pressure arising from a sound source at a distance d_f in the far-field. This function, also known as the distance variation function (DVF) (Kan et al., 2009), can be computed as follows:

$$DVF(\theta, d_n, d_f) = \frac{p_n(a, \omega, \theta, d_n)}{p_f(a, \omega, \theta, d_f)}, \quad (6)$$

$$HRTF_{ILD}(d_n, \theta) = DVF(\theta, d_n, d_f) \cdot HRTF(d_f, \theta), \quad (7)$$

where $HRTF_{ILD}$ is the near-field HRTF synthesized using ILD cues, a is the radius of the head, θ is the angle of incidence of the sound source with respect to the mid-sagittal axis, and ω is the angular frequency in radians. p_n and p_f are the pressures obtained from the spherical head model in near-field and far-field, respectively. d_n and d_f are the source to head distances in near-field and far-field, respectively.

C. HRTF synthesis using both ILD and auditory parallax cues

The synthesized HRTF containing both ILD and auditory parallax cues ($HRTF_{AP+ILD}$) can be obtained by cascading the auditory parallax and DVF models. The pinna spectral features are first remapped according to the parallax angles to obtain the HRTFs containing auditory parallax cues. The DVF model is then applied to $HRTF_{AP}$ to further incorporate the distance-dependent ILD cues, as given in the equation below:

$$HRTF_{AP+ILD}(d_n, \theta) = DVF(\theta, d_n, d_f) \cdot HRTF_{AP}(d_n, \theta). \quad (8)$$

The synthesized distance-dependent HRTFs were also compared with the measured HRTFs using objective and subjective measures. The low-frequency spectral cues (0.1–1 kHz) and the high-frequency pinna spectral cues (5–16 kHz) of the synthesized and measured HRTFs were compared by calculating the root mean square spectral difference (RSD) as an objective criterion,

$$RSD_{low,high} = \sqrt{\frac{1}{N} \sum_{f=f_i}^{f_h} \left(20 \log_{10} \left| \frac{HRTF_{measured}(f)}{HRTF_{modeled}(f)} \right| \right)^2}, \quad (9)$$

where f_i , f_h indicate the frequency range of comparison, N represents the number of frequency bins between f_i and f_h . In this paper, RSD_{low} is computed for $f_i = 0.1$ kHz and $f_h = 1$ kHz, while RSD_{high} is computed for $f_i = 5$ kHz and $f_h = 16$ kHz.

The DVF mainly models the low-frequency distance-dependent spectral effects of the head. However, the DVF model does not simulate any of the distance-dependent spectral effects of the pinna. On the other hand, the auditory parallax effect only models the distance-dependent effects of the pinna but does not simulate the spectral effects due to head.

From Fig. 3, it can be seen that the HRTFs synthesized using auditory parallax cues fail to simulate the low-frequency near-field spectral cues well ($RSD_{low} > 6$ dB). However, the HRTFs synthesized using auditory parallax cues model the high-frequency pinna cues reasonably well ($RSD_{high} = 2$ dB). The low-frequency spectral cues in the measured near-field HRTFs are modeled accurately ($RSD_{low} = 2$ dB) by the DVF model. However, there is a mismatch of small degree at high frequencies in the case of HRTF synthesis using DVF model ($RSD_{high} < 4$ dB). We note that the HRTFs synthesized using both DVF and auditory parallax models match with the measured HRTFs much better ($RSD_{low} = 1$ dB and $RSD_{high} = 2$ dB), as illustrated in Fig. 3. More importantly, it was observed that the auditory parallax contributes little to the synthesized distance-dependent HRTFs. This can be further analyzed with the help of detailed subjective experiments, as shown in Sec. IIID.

D. Subjective experiments

The roles of auditory parallax and of ILD cues for distance perception using both individualized and non-individualized HRTFs were investigated with the help of perceptual experiments. A total of 15 subjects participated in the experiments carried out using non-individualized HRTFs. HRTFs measured on the B&K HATS dummy head were used as non-individualized HRTFs in this experiment. Three subjects, whose personal HRTFs were measured, participated in the listening experiments using individualized HRTFs. All subjects had normal hearing. They were aged between 18 and 30 years, and had past experience in similar tests. A double blind procedure was adopted whereby neither the evaluator nor the subjects were aware of the order and nature of the test stimuli.

The stimuli were sequences of four different 300-ms bursts of white Gaussian noise (WGN) separated by 30-ms intervals of silence. Stimuli were gated with cosine-squared ramps of 30-ms onset and offset in order to avoid transient effects. The gated stimuli were then convoluted with the synthesized HRTFs obtained using both individualized and

FIG. 3. Comparison of measured and modeled near-field HRTFs for the HATS dummy head. Top: HRTF at a distance of 35 cm and 50° azimuth. Bottom: HRTF at a distance of 45 cm and 60° azimuth.

non-individualized far-field HRTFs. The four stimulus conditions used in this experiment are shown below:

- (1) Synthesized distance-dependent HRTFs containing auditory parallax cues ($HRTF_{AP}$).
- (2) Synthesized distance-dependent HRTFs containing bin-audal ILD cues ($HRTF_{ILD}$).
- (3) Synthesized distance-dependent HRTFs containing both auditory parallax and ILD cues ($HRTF_{AP+ILD}$).
- (4) Measured distance-dependent HRTFs ($HRTF_{measured}$).

For the first three stimulus conditions, virtual sounds were synthesized for nine distances at 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1, and 1.4 m. Stimuli synthesized from twenty directions (0° to 340° at a resolution of 20° including those at 90° and 270°) were presented in a random order for each of the nine distances. For the experimentally measured distance-dependent individualized and non-individualized HRTFs, the simulated distances were 0.35, 0.45, 0.50, 0.60, 0.75, 0.80, 1, and 1.4 m, as the HRTFs were measured only for these distances. The synthesized stimuli were then normalized to eliminate amplitude-based distance cues. In the near-field, the distance between source and the ears deviates significantly from the distance between source and the center of the head. The scaling factor used was based on the distance of the source from the left and right ears of the subject, similar to that adapted by Brungart (1999a). The correction was such that a distance of 1 m had a scaling factor of 1. The scaling factor (N) was

$$N = \frac{1}{\frac{50}{\text{Distance to left ear (cm)}} + \frac{50}{\text{Distance to right ear (cm)}}} \quad (10)$$

Amplitude scaling was carried out by multiplying the noise waveform with the scaling factor prior to playback. In addition to distance normalization, source amplitude was randomized in order to ensure that subjects' perception was based only on the spectral features of HRTF and not biased by source levels (Brungart, 1999a). To achieve this, the source amplitude was roved randomly by an additional 15 dB (from 0 to 15 dB in steps of 1 dB). Since the maximum amplitude of stimulus was approximately 60 dBA at 1 m, the effective stimulus amplitude ranged from 45 to 60 dBA after randomization and correction due to scaling factor.

To ensure consistency, three measurements were made for each of the 9 distances and 20 azimuth angles making it a total of 540 stimulus presentations (20 directions $\times$ 9 distances $\times$ 3 trials) for modeled HRTFs, and 480 stimulus presentations (20 directions $\times$ 8 distances $\times$ 3 trials) for measured HRTFs. All the listening experiments were conducted in a semi-anechoic chamber of dimension 5.5 m (length) $\times$ 3.1 m (width) $\times$ 2.4 m (height). The level of the background noise in the semi-anechoic chamber was 19 dBA.

Synthesized virtual sounds were played back using Sennheiser HD600 headphones. Headphone equalization was carried out prior to the start of experiment to compensate for the non-uniform headphone response. A graphical user interface (GUI) was used to present the stimuli in a random order to the subject. Before the experiment started, subjects were asked to stretch their hands and the distance between the head and the tip of their hands was taken as reference. For each condition, they had to identify the "apparent distance" of the sound from the head. Once the stimulus was presented, subjects were prompted to record their judgement of distance (in cm). They were allowed to remove and

reposition the headphones on their head in between trials. However, they were instructed to minimize any head movement during the experiment. Subjects were not given feedback on their responses.

E. Results

The main objective of this experiment was to examine the role of auditory parallax and ILD cues on distance perception over headphones in anechoic conditions. Distance perception was separately investigated for lateral (40° to 140° and 220° to 320° clockwise) and near-medial (340°, 0°, 20° in the front and 160°, 180°, 200° in the rear) directions.

Distance responses using non-individualized and individualized HRTFs for all the four stimulus conditions are shown in Figs. 4 and 5, respectively. The responses are plotted in Figs. 4 and 5 on a logarithmic scale because the magnitude of distance errors tended to increase with increasing source to head distance. Two different measures were used to evaluate the response accuracy:

- (1) Correlation coefficient between the logarithm of actual distances and the response distances.
- (2) Slope of line best fitting the log-stimulus data to the log-response data.

Zahorik (2002a) and Zahorik *et al.* (2005) found that the following compressive function represents a good fit for listeners' distance responses under a variety of stimulus conditions:

$$r_{res} = k(r_{actual})^m, \quad (11)$$

where r_{res} is the estimated response distance, r_{actual} is the actual distance, k and m are the fit parameters of the power function. Distance responses of the subjects were fitted with Eq. (11). The value of exponent m was typically less than one and the value of constant k was close to 1, thereby indicating a strong perceptual bias in distance estimation.

It was observed that there was a general tendency to overestimate distances in near-field and underestimate distances in far-field. The distance perception using HRTFs synthesized with the auditory parallax model was poor for both lateral and near-medial directions. Using non-individualized HRTFs (Fig. 4), the mean correlation coefficient values between the estimated and actual distances were 0.31 for lateral and 0.30 for near-medial directions, respectively. Similarly, for individualized HRTFs (Fig. 5), the correlation coefficient values obtained for distance responses with auditory parallax modeled HRTFs were 0.34 for lateral and 0.31 for near-medial directions, respectively. Both these observations are consistent with the results of Brungart (1999a), who observed poor correlation coefficient values between the perceived and the simulated distance for monaural and high-pass filtered stimuli. This finding suggests that auditory parallax cues alone are not sufficient for good distance perception.

With the DVF-synthesized HRTFs obtained using non-individual HRTFs, distance perception improved with a correlation coefficient value of 0.59 for lateral directions.

However, the correlation coefficient value was poor (0.24) for the near-medial directions (Fig. 4). This is due to the absence of sufficient binaural cues for positions near the medial plane. The distance responses with the DVF synthesized HRTFs obtained using individualized HRTFs was found to be similar to that using non-individual HRTFs. The corresponding correlation coefficient values for the DVF synthesized HRTFs using individualized HRTFs were found to be 0.61 and 0.25 for lateral and near-medial directions, respectively.

When both auditory parallax and distance-dependent ILD cues were present in HRTF, subjects were able to localize distances better compared to the case when only auditory parallax cues were present in HRTF. The correlation coefficient values for distance responses with the modeled HRTFs containing auditory parallax and ILD cues were found to be 0.63 (lateral), 0.34 (near-medial) using non-individualized HRTFs, and 0.64 (lateral), 0.32 (near-medial) using individualized HRTFs. With the measured HRTFs, the correlation coefficient values were found to be 0.69 (lateral), 0.41 (near-medial) using non-individualized HRTFs, and 0.72 (lateral), 0.46 (near-medial) using individualized HRTFs. These values indicate that the distance localization with the modeled HRTFs containing distance-dependent binaural ILD cues was comparable to that with measured HRTFs. However, some in-head localization of sound was observed in distance responses for all the stimulus cases using both non-individualized and individualized HRTFs, as indicated in Figs. 4 and 5. Distance responses reported by the subjects that are less than 15 cm are not shown in Figs. 4 and 5 for the sake of clarity.

Figure 6 shows the correlation coefficient between the logarithm of actual and responded distances, and the slopes of the linear regression line fitting the log-stimulus data to the log-response data. The correlation coefficient values and slopes are plotted for all the azimuth directions between 0° and 180°. Correlation coefficient values were poor across all the azimuth angles with the auditory parallax modeled HRTFs using both non-individualized HRTFs and individualized HRTFs. For other stimulus cases, the correlation coefficient values increased for lateral angles due to the presence of ILD cues. The slope indicates the scaling bias in responses. A slope of less than 1 indicated that responses varied over a narrow range of distances than the actual locations, showing compression in the responses. Similar to the correlation coefficient, the slope was relatively high for lateral locations and low for near-medial locations. This was true for all the stimulus conditions except for the stimulus condition synthesized with auditory parallax modeled HRTFs. Due to the presence of sufficient ILD cues in the lateral directions, subjects tended to be more sensitive to the actual source location in these directions resulting in a high correlation coefficient value between the responded distance and the actual distance. The slope was less than 1 even for lateral directions indicating a tendency to compress distance responses for all azimuthal directions. However, slope of the linear regression line fitting the log-stimulus to the log-response data with synthesized HRTFs (non-individualized and individualized) containing only auditory parallax cues

FIG. 4. Stimulus-response pairs for distance (in cm) using non-individualized HRTFs. Note the log-log scale. The solid line in each panel represents a linear regression fit. The dashed line in each panel corresponds to the “correct” responses. Also shown is the correlation coefficient (ρ) between the log-stimulus and the log-response distances and the compressive function fit parameters k and m [Eq. (11)]. Left column corresponds to the lateral angles and the right column represents the data from the near-medial angles. Note that only the distance responses greater than 15 cm are displayed for visual clarity.

was low, indicating a poor fit for all azimuth angles. This key observation suggests that in the presence of binaural cues, auditory parallax cues are not critical for accurate distance perception.

An analysis of variance (ANOVA) test was carried out to further investigate the significance of auditory parallax and ILD cues for distance perception. The effect of auditory parallax cues on distance perception in the presence of other

FIG. 5. Stimulus-response pairs for distance (in cm) using individualized HRTFs. The layout is as in Fig. 4.

distance cues was first studied. The mean values of the responded distances with measured HRTFs ($HRTF_{measured}$), and with the modeled HRTFs containing only binaural ILD cues ($HRTF_{ILD}$), auditory parallax and ILD cues ($HRTF_{AP+ILD}$), were all similar for both non-individualized and individualized synthesis cases. This was evident from the ANOVA analysis, which returned no significant differences in distance responses with $HRTF_{measured}$, $HRTF_{ILD}$, and $HRTF_{AP+ILD}$ for non-individualized [$F(2,25)=0.55$, $p=0.58$] and individualized cases [$F(2,24)=0.5$, $p=0.55$], respectively. Note that F is the standard F -distribution. The p value refers to the probability of significance, which is compared to the threshold critical value to determine the significance of the factor. Throughout this paper, the significance of a factor is determined for critical significance levels (α) of 0.01 and 0.05 corresponding to 99% and 95%

confidence intervals, respectively. The auditory parallax cues did not have a significant effect on distance perception in the presence of ILD cues using both non-individualized [$F(1,15)=0.28$, $p=0.6$] and individualized HRTFs [$F(1,16)=0.34$, $p=0.56$]. However, the effect of ILD cues simulated using the DVF model had a significant effect in both the individualized, as well as the non-individualized [$p \ll 0.001$] synthesis cases.

All the above observations suggest that auditory parallax cues in the presence of ILD cues do not have any prominent effect in anechoic headphone playback conditions. The correlation coefficient values also indicate poor distance perception for all directions, when auditory parallax is the only available cue. Therefore, it can be concluded that auditory parallax cues can be neglected and it is sufficient to include ILD cues in the model for the distance-dependent

FIG. 6. (Color online) Top: Correlation coefficient between the logarithm of actual and responded distances. Bottom: Slope of the linear regression line best fitting the log-stimulus data to the log-response data for azimuth angles (between 0° and 180°) for both (a) non- individualized HRTFs and (b) individualized HRTFs.

individualized HRTFs. This also means that the angular position of frontal projection headphone transducer with respect to the ear need not be physically changed to incorporate the parallax effects in the near-field HRTFs.

In Sec. IV, the proposed technique to model the distance-dependent individualized HRTFs using frontal projection headphones is discussed. The model is subsequently validated with the help of objective and subjective analyses.

IV. MODELING DISTANCE-DEPENDENT INDIVIDUALIZED HRTFS

The frontal projection headphones project sound directly onto the pinnae from the front, emulating a loud-speaker setup in the far-field. Individualized pinna spectral cues are automatically modeled at the eardrum during frontal projection playback (Sunder *et al.*, 2013). The frontal projection headphones used in this study were open-type with free-air equivalent coupling (FEC) characteristics, as shown in Fig. 1 (Møller *et al.*, 1995; Sunder *et al.*, 2013). The role of frontal projection headphones in modeling the individualized pinna spectral cues has been explained in detail in Sunder *et al.* (2013).

From the analysis carried out in Sec. III, it is clear that ILD cues are critical for distance perception in anechoic conditions, and the effect of auditory parallax can be neglected. In the model described in Fig. 7(a), mono sound source is

first filtered with ipsilateral and contralateral ear transfer functions derived from the spherical head model. The spherical head model simulates the distance-dependent near-field effects and head shadow effects in the contralateral ear well, but it does not simulate the pinna spectral cues. The frontal projection headphones, however, inherently model the highly idiosyncratic frontal ipsilateral pinna cues at the eardrum during sound playback. Thus, in effect, ipsilateral spherical head transfer function is filtered with the frontal projection headphone transfer function in order to obtain the frontal ipsilateral cues in the HRTF. However, additional headphone equalization and low-frequency compensation filters are required to ensure accurate modeling of the HRTF. In this way, the frontal projection headphones can synthesize the frontal directional distance-dependent individualized HRTFs without the requirement of extensive measurements. To model the rear directional HRTFs, a rear directional filter is additionally required, as shown in Fig. 7(b). Each of the processing blocks [Figs. 7(a) and 7(b)] involved in the modeling process is explained below.

A. Distance-dependent head effects

In the near-field, the reflections, and scattering of the acoustic waves with the head have significant spectral effects. The spherical head model (Duda and Martens, 1998) is used to model these distance-dependent head effects. In

FIG. 7. Approaches to model distance-dependent individualized frontal (a) and rear (b) HRTFs.

frontal projection headphones playback, idiosyncratic frontal pinna cues are generated at both the contralateral ear and ipsilateral ear. However, this does not affect contralateral response much, because of the large attenuation at high frequencies due to the head shadow effect. [Humanski and Butler \(1988\)](#) studied in detail the contribution of near-ear and far-ear in sound localization in the sagittal plane. They found that listeners could localize well with just the ipsilateral ear open and contralateral ear occluded. As the sound source moves towards lateral positions, the energy at high frequencies (above 4 kHz) in the contralateral ear is attenuated due to head shadowing. As a result, the exact response at the shadowed ear becomes less critical ([Avendano et al., 1999](#)). [Avendano et al. \(1999\)](#), thus, showed that the contralateral ear HRTF could be modeled by a simple transformation of the ipsilateral ear HRTF using the spherical head model. The spherical head model also simulates the low-frequency ILD variation, which is a very prominent cue for distance perception.

B. Rear directional HRTFs

The frontal projection headphones due to the frontal positioning of the emitters emulate only the ipsilateral frontal pinna spectral cues. Besides, the spherical head model is based on a perfect symmetrical sphere and thus, the spherical head model fails to simulate the rear directional spectral cues. Therefore, an additional rear directional filter that can modify the spectral properties of the frontal HRTF to model the rear directional HRTF is required (Fig. 8). The rear directional filter is based on [Blauert's directional bands \(Blauert, 1997\)](#), which suggest that certain frequency bands are boosted or attenuated characterizing frontal or rear directions. Rear directional filter is developed by computing the average spectral differences between the HRTFs in the frontal and its symmetrical direction in the rear, for all subjects selected from the CIPIC HRTF database ([Algazi et al., 2001a](#)). The rear directional filter is computed as follows:

$$RDF(\theta, \varphi) = \frac{\sum_{i=1}^{n\_subjects} [HRTF_{civic\_rear}(\theta, \varphi) - HRTF_{civic\_frontal}(180^\circ - \theta, \varphi)]}{N\_s}, \quad (12)$$

where N_s is the number of subjects in the CIPIC database (43 subjects), θ is the rear azimuth angle, and φ is the elevation angle. All angles are measured with respect to the mid-sagittal axis. $HRTF_{civic\_rear}(\theta, \varphi)$ is the rear HRTF and $HRTF_{civic\_frontal}(180^\circ - \theta, \varphi)$ corresponds to the symmetrical frontal HRTF, both of which are selected from CIPIC

database. The frequency responses of the rear directional filters are shown in Fig. 8,

$$HRTF_{rear}(\theta, \varphi) = HRTF_{frontal}(180^\circ - \theta, \varphi)RDF(\theta, \varphi). \quad (13)$$

FIG. 8. Rear directional filter for all the azimuth angles for the ipsilateral ear. Black: rear directional filters of all the subjects of the CIPIC database. Gray (bold): average rear directional filter of all the subjects.

The rear directional HRTFs can be obtained by filtering the frontal modeled HRTFs with the rear directional filter, as shown in Eq. (13).

C. Type-2 equalization

Typically, a HPTF comprises of the headphones transducer response and the acoustic coupling between headphones and listener's ears. Conventional headphone equalization methods often aim for a flat target response, so that the spectrum of the HRTF used remains undistorted at the eardrum. The equalization of the frontal projection headphones plays a critical role in the modeling process. In case of using frontal projection headphones, it is desirable to preserve the

highly individualized frontal pinna cues at the eardrum, which are created during playback. Thus, using conventional headphone equalization is not advisable; as such an equalization technique would equalize even the individualized pinna spectral features generated during playback. For the frontal projection headphones, only the distortion created by headphone transducer along with the resonant modes of the earcup has to be compensated. Thus, the individual pinna cues (*pinnae*) due to frontal projection are preserved after equalization. This type of equalization is known as the type-2 equalization. Type-2 equalization improves the performance of binaural audio playback using frontal projection headphones by reducing the front-back confusions and in-head localization (Sunder *et al.*, 2013). Type-2 equalization filter

[Fig. 9(a)] is obtained by computing the inverse of the free-field response (FFR) of the frontal projection headphones [Eq. (15)]. The frontal projection headphone with the transfer function $HPTF_{frontal}$, when equalized using a type-2 equalization filter, models the unique personal pinna cues (*pinnae*), as shown in Eqs. (14) and (15). FFR is obtained by measuring the free-field response of the headphones in the absence of dummy head. FFR captures only the spectral effects of transducer and resonant modes of the earcup (Sunder *et al.*, 2013). To measure the FFR of frontal projection headphones, a B&K 4961 Multi-Field $\frac{1}{4}$ in. microphone was placed at the center of the earcup just outside its peripheral surface. The distance of the microphone from the transducer was 5 cm, which is approximately equal to the distance at which the ears would have been present if the headphones were in place. Type-2 equalization can be described in mathematical terms as follows:

$$HPTF_{frontal}(\omega) = FFR(\omega) \times pinnae(\omega), \quad (14)$$

$$Type2EQ(\omega) = \frac{1}{FFR(\omega)}, \quad (15)$$

where $pinnae(\omega)$ corresponds to the individualized pinna cues due to frontal projection. $HPTF_{frontal}(\omega)$ and $FFR(\omega)$ are the frontal projection headphone response and its free field response, respectively. $Type2EQ(\omega)$ refers to the type-2 equalization filter.

D. Low-frequency compensation filter

The low-frequency spectral cues and ILD cues are key for distance perception and have to be modeled correctly. The frontal projection headphones have a poor low-frequency response due to their open structure. The inability of the type-2 equalization filter to completely correct the poor low-frequency response results in a noticeable deviation in the modeled HRTFs from the measured HRTFs. Deviation in the low-frequency response can have adverse effects on distance perception (Sunder *et al.*, 2014). Thus, the low-frequency deviation has to be corrected in order to

ensure that the auditory distance perception with modeled HRTFs is close to that with the true measured HRTFs (Sunder *et al.*, 2014). The low-frequency response of the HRTF is non-idiosyncratic and is mainly due to head, shoulder, and torso effects. These distance-dependent low-frequency effects can be approximated using the spherical head model, as shown in Figs. 7(a) and 7(b). To begin with, the low-frequency deviation from the true measured HRTF has to be estimated. Low-frequency compensation filter is obtained by taking spectral inverse of the low-frequency response (up to 1 kHz) of the type-2 equalized frontal projection headphone, which is measured on the B&K HATS dummy head. This filter model is valid as both dummy heads and human subjects display similar spectral properties in the HRTF in the low-frequency range. Note that the compensation filter is designed to have a flat response beyond 1 kHz, as shown in Fig. 9(b), to avoid any effect on the modeled HRTF spectrum at high frequencies.

E. Interaural time differences

HRTF is often approximated as a minimum-phase function with a position-dependent, frequency-independent time delay (Mehrgardt and Mellert, 1977). In this model, lateral azimuth perception is simulated by adding a frequency-independent ITD to the minimum-phase representation of the modeled HRTF. Wightman and Kistler (1992) carried out localization experiments using minimum-phase description of HRTFs and reported that the sound localization with these models using headphone simulation was close to that obtained in free-field conditions. Kulkarni *et al.* (1995) also validated the minimum-phase representation of HRTFs with the help of psychophysical experiments and modeled the ITD as frequency-independent time delay. The ITD used in this study is obtained by taking the average ITD of all subjects selected from the CIPIC database for a particular location. The effect of using a non-individualized ITD is not as critical as the use of non-individualized pinna cues. This is evident from the fact that the azimuth localization performance with non-individualized HRTFs is comparable to that

FIG. 9. (a) Type-2 equalization filter calculated by taking the inverse of an average FFR based on ten trials. (b) Low-frequency compensation filter to compensate for the deviation in the low-frequency response of the modeled HRTF from the measured individualized HRTF.

with the individualized HRTFs (Wenzel *et al.*, 1993). Moreover, ITD is primarily a low-frequency cue and is not affected much by the variations of individual's pinna, which contributes mainly to the high-frequency cues. Thus, the use of an average ITD is a valid approximation in this model. As the variation of ITD values is minimal with distance, we consider a constant ITD value across all distances in this model. The variation in the interaural cues with distance can also be measured in terms of their respective just noticeable differences (jnd). Hershkowitz and Durlach (1969) found that listeners could discriminate changes in ILD on the order of 0.8 dB over a wide range of ILDs. Thus, the distance-dependent changes in ILD from 0.12 m to 1 m span a range of up to 15 jnd's at 500 Hz and 30 or more jnd's at 3 kHz. However, the jnd for ITD was approximately 15 μ s at ITDs below 400 μ s and increased rapidly for ITDs greater than 400 μ s. The variation in ITD in the near-field spans only 2–3 jnd's. Researchers have speculated that listeners may make use of distance-invariant ITDs to first determine the azimuth location and then use the ILD cues to estimate the distance (Brungart and Rabinowitz, 1999; Shinn-Cunningham *et al.*, 2000b).

Figure 10 shows a series of spectra that illustrate a step-by-step process for modeling the ipsilateral frontal, contralateral frontal, and ipsilateral rear HRTFs. The model can be mathematically described as follows:

$$Y_{frontal}(\omega) = X(\omega)HPTF_{frontal}(\omega)Type2EQ(\omega) \times SHM(\omega)LfCF(\omega), \quad (16)$$

$$Y_{frontal}(\omega) = X(\omega)pinnacues(\omega)SHM(\omega)LfCF(\omega), \quad (17)$$

$$Y_{rear}(\omega) = X(\omega)pinnacues(\omega)SHM(\omega)RDF(\omega)LfCF(\omega), \quad (18)$$

where $X(\omega)$, $Y_{frontal}(\omega)$, and $Y_{rear}(\omega)$ are the input mono signal, and the binaural signals synthesized for frontal and rear directions, respectively. The type-2 equalization filter is denoted by $Type2EQ(\omega)$. The spherical head model that models the distance-dependent head effects is denoted by $SHM(\omega)$. The rear-directional filter and low-frequency compensation filter are represented by $RDF(\omega)$ and $LfCF(\omega)$, respectively.

For ease of explanation, we start with the frontal projection headphone response [$HPTF_{frontal}(\omega)$] that emulates the idiosyncratic frontal ipsilateral cues due to frontal projection. The response $HPTF_{frontal}(\omega)$ apart from the pinna cues also contains the inherent spectral distortion of transducer and the resonant modes of the ear cup that are compensated by the type-2 equalization filter. A spherical head model filter [$SHM(\omega)$] is then applied to model the distance-dependent spectral effects of head in the ipsilateral ear [Fig. 10(a)]. The spherical head model also simulates the head shadow effects in the contralateral ear [Fig. 10(b)]. To compensate for the deviation at low frequencies, a low-frequency compensation filter [$LfCF(\omega)$] is applied to obtain the correct low-frequency response of the HRTF. However, to model the rear directional HRTFs, an additional rear directional filter [$RDF(\omega)$] is required that modifies the frontal HRTF response to a HRTF at a symmetrical rear position [Fig. 10(c)]. The accuracy of the modeled and measured HRTFs are compared and analyzed in Sec. IV F.

FIG. 10. Different steps involved in the modeling process. (a) Subject K, source is placed at 40° azimuth angle and 70 cm from the center of the head. (b) Subject A, source is placed at 60° azimuth, 50 cm from the center of the head. (c) Subject R, source is placed at 100° azimuth, 60 cm from the center of the head. The azimuth angle is measured with respect to the sagittal plane.

F. Objective analysis

To validate the proposed model, the modeled distance-dependent individualized HRTFs were compared with the measured distance-dependent individualized HRTFs for all the three subjects, whose HRTFs were measured. The modeled individualized frontal ipsilateral, frontal contralateral, and rear ipsilateral HRTFs are shown in Figs. 11–13, respectively, along with the measured individualized HRTFs. It should be noted that the modeled HRTFs using the frontal projection headphones exhibit a good match with measured HRTFs. The high-frequency individualized pinna spectral cues modeled by frontal projection headphones match accurately with that of the actual frontal pinna spectral cues of the listener (Fig. 11). The spherical head model simulates the distance-dependent low-frequency features and the head shadow effects accurately. The head shadow effects modeled by the spherical head model is clearly visible in the modeled contralateral ear HRTFs at high frequencies (Fig. 12). For the rear ipsilateral HRTFs, it is observed that the spectral pattern of the modeled HRTFs is similar to that of the measured individualized HRTFs (Fig. 13).

RSD between modeled and measured HRTFs was calculated as an objective measure for comparison [Eq. (9)]. Average RSD scores in dB for the modeled ipsilateral and contralateral ear HRTFs at all source distances and across all the subjects are plotted as a function of azimuth in Fig. 14.

For the ipsilateral ear, RSD values were less than 4 dB in the frontal directions and increased for directions behind the head. The average RSD score for the ipsilateral HRTFs across all azimuth and all distances was found to be less than 4.5 dB (Fig. 14). The RSD errors for the contralateral ear (5 dB) were always greater than that for the ipsilateral ear HRTF for both frontal and rear directions (Fig. 14). The high-frequency mismatch in the modeled contralateral ear HRTFs with the measured individualized HRTFs is due to the model generating ipsilateral pinna cues even in the contralateral ear HRTFs. The RSD scores for the modeled rear directional ipsilateral HRTFs (5 dB) were higher than that for the modeled frontal ipsilateral HRTFs (3 dB), as shown in Fig. 14. For the contralateral HRTFs in the rear directions, the RSD scores were again higher than that for the frontal directions due to the generation of ipsilateral cues during playback. The higher RSD values in the rear directions is due to the fact that the rear directional filter is only an approximate method to model the rear directional HRTFs. The modeled HRTFs were also validated using perceptual experiments. These experiments are discussed in Sec. IV G.

G. Subjective experiments

Detailed subjective experiments were carried out to validate the proposed model. Subjects judged the perceived

FIG. 11. Comparison of modeled individualized frontal ipsilateral HRTFs and measured individualized frontal HRTFs of all the three subjects for different azimuthal angles and distances, as indicated in each panel.

FIG. 12. Comparison of modeled individualized frontal contralateral HRTFs and measured individualized HRTFs of all the three subjects for different azimuthal angles and distances, as indicated in each panel.

direction and distance of the virtual source synthesized using the modeled distance-dependent individualized HRTFs that was played back from the frontal projection headphones. The measured distance-dependent individualized HRTFs were reproduced over an equalized (conventional individualized equalization) Sennheiser HD600 headphone. Four tests were carried out to validate the proposed model

- (1) Localization experiment using modeled and measured distance-dependent individualized HRTFs.
- (2) Distance perception experiment using modeled distance-dependent individualized HRTFs.
- (3) Distance perception experiment using modeled distance-dependent individualized HRTFs with intensity scaling.
- (4) Distance perception experiment using real sound sources.

A total of 25 subjects participated in experiments 1 to 3, and 5 subjects participated in experiment 4. All subjects had normal hearing. The basic stimulus used was a sequence of four bursts of cosine square-gated (30-ms) WGN of 300-ms each, separated by an interval of 30-ms of silence, as in the

experiments in Sec. III. The stimulus for experiments 1, 2, and 4 were normalized using a scaling factor N , where N is obtained from Eq. (10). The scaling factor mainly compensates for the intensity cues that arise due to variation in distance between the source and head. In the intensity scaling condition, additional intensity cues were added to the un-normalized modeled HRTF to aid distance perception. A scaling factor of $(1/N)$ was used to account for the fact that the difference in distance between the sound source and the two ears can be substantially different, when the source is in the near-field and close to the interaural axis (Kan *et al.*, 2009). For the distance perception experiment with real sources, a Cambridge Soundworks mini cube speaker was used to playback the stimulus.

Distance perception was tested for all the eight speaker distances for which the HRTFs were measured in Sec. III. Apart from the loudspeaker from which the stimulus was played, additional dummy loudspeakers were placed at random positions in order to avoid visual bias.

All the subjects were blindfolded for the experiment, which was conducted with real sound sources. All the

FIG. 13. Comparison of modeled individualized rear ipsilateral HRTFs and measured individualized HRTFs of all the three subjects for different azimuthal angles and distances, as indicated in each panel.

experiments were carried out in the semi-anechoic chamber that was used for the experiments described in Sec. III. The highest stimulus level was approximately 60 dBA at 1 m distance. Subjects gave their responses for the perceived direction and distances on a GUI, similar to the one used in experiments in Sec. III.

The analysis of results for directional localization was carried out by dividing the distances into three different

categories as SET1: near (20, 30, 40 cm), SET2: medium (50, 60, 70 cm), and SET3: far (80, 90, 140 cm). The front-back confusions in the localization responses with the modeled individualized HRTFs, measured individualized HRTFs, and non-individualized HRTFs were analyzed for all the distance sets (Fig. 15).

An ANOVA analysis for front-back confusions revealed that the mean differences in front-back confusions between modeled and measured HRTFs were statistically significant for near (SET 1) [$F(1,32) = 26, p \ll 0.01$] and medium distances (SET 2) [$F(1,32) = 7, p \ll 0.01$]. In the far-field (SET 3), front-back confusions in the localization responses with modeled and measured individualized HRTFs were similar with no significant differences [$F(1,32) = 7, p = 0.03$] for a critical significance level of 0.01. However, their differences were statistically significant, if a critical significance level of 0.05 was considered. Overall, front-back confusions for localization with modeled individualized HRTFs were reduced by more than 50% as compared to those with non-individualized HRTFs. In fact, the front-back confusions with modeled individualized HRTFs tended to be closer to the measured individualized HRTFs (Fig. 15).

Figure 15 shows the localization responses for modeled individualized HRTFs and measured (non-individualized and individualized) HRTFs. In this study, localization error is defined as the angular difference between the perceived angle reported by the subject and the actual azimuth angle of

FIG. 14. RSD between modeled and measured individualized HRTFs for the ipsilateral and contralateral ear.

FIG. 15. Localization responses for modeled individualized, measured individualized, and measured non-individualized HRTFs for all the three sets of distances. The distances are divided into three categories: near (20 cm, 30 cm, 40 cm), medium (50 cm, 60 cm, 70 cm), and far (80 cm, 90 cm, 140 cm). Modeled distance dependent individualized HRTFs were played back using frontal projection headphones, while the measured individualized and non-individualized HRTFs were played back using Sennheiser HD600 headphones.

the source. Front-back confusions were corrected prior to the calculation of the localization error. An ANOVA test was carried out to investigate the directional localization performance using modeled and measured HRTFs. Comparing the modeled and measured individualized HRTFs, the localization errors (with front-back confusions removed) were not significantly different across all the three sets for a critical significance level of 0.01; SET 1 [$F(1,26) = 0.39$, $p = 0.53$], SET 2 [$F(1,26) = 0$, $p = 0.9644$], SET 3 [$F(1,26) = 5.91$, $p = 0.02$]. The localization error differences between the modeled and measured individualized HRTFs in the far field (SET 3) were found to be significant for a higher critical significance level of 0.05.

The distance responses for modeled individualized HRTFs (with and without intensity scaling), and for the real sound sources are plotted in Fig. 16. The distance responses for the measured individualized HRTFs are shown in Fig. 5. The correlation coefficient values between the log-stimulus and log-response distance and the slopes of the linear regression line fitting the log-stimulus to the log-response data for the modeled individualized HRTFs with and without intensity cues are plotted across the azimuth in Fig. 17. The correlation coefficient values for the distance responses using modeled HRTFs without intensity cues (0.68 lateral, 0.35 near-medial), were of an order close to that of the measured

individualized HRTFs, for both lateral (0.72) and near-medial (0.46) directions (Fig. 5). When the modeled individualized HRTFs were intensity scaled and played back, the mean correlation coefficient values increased to 0.83 and 0.54 in the lateral and near-medial directions, respectively. This result suggests that with intensity cues added, the distance perception could be improved for the modeled individualized HRTFs.

With real sources, the correlation coefficient values between log-stimulus distance and log-response distance for lateral and near-medial directions were found to be 0.76 and 0.50, respectively. It was noted that the distance localization accuracy with the modeled HRTFs (without intensity cues) was close to that with real sound sources for lateral directions. However, for near-medial directions, localization with real sources was much better. One of the probable reasons for better performance with real sources could be the presence of some reverberant cues, as these experiments were carried out in the semi-anechoic chamber (Bronkhorst *et al.*, 1999; Mershon *et al.*, 1989; Mershon and Bowers, 1979; Mershon and King, 1975; Shinn-Cunningham *et al.*, 2000a). Other researchers have also reported that the distance perception with real sources, is usually better than that in virtual auditory display (Brungart and Brian, 2001).

FIG. 16. Stimulus-response pairs for distance (in cm) using modeled individual HRTFs (with and without intensity cues) and real sound sources (without intensity cues). The layout is as in Fig. 4.

FIG. 17. Correlation coefficient between the log-stimulus distance and log-response distance and slope of the linear regression line best fitting the log-stimulus data to the log-response data at each azimuth, for the modeled individualized HRTFs with and without intensity cues.

The slope was less than one for both lateral and near-medial directions with and without the presence of intensity cues, indicating a general tendency to compress the distance responses for all azimuthal directions. The correlation coefficient values between actual and responded distance was poor for positions near the medial plane, and it improved for more lateral angles in both the cases, as shown in Fig. 17. An ANOVA test was carried out to compare the distance perception between the modeled individualized, measured individualized, and real sources. Critical significance levels of 0.01 and 0.05 were considered. For a critical significance level of 0.01, the distance perception with the modeled and measured individualized HRTFs were found to be similar with no significant differences [$F(1,8) = 7.51$, $p = 0.02$]. However, these differences were significant for a higher critical significance level of 0.05. The distance responses with real sources for lateral directions were compared to that with modeled HRTFs. ANOVA results returned their differences [$F(1,8) = 9$, $p = 0.015$] to be statistically not significant for a critical significance level of 0.01 and significant for a significance level of 0.05. In the median plane, significant differences were found between the distance responses with modeled and measured individualized HRTFs [$F(1,8) = 16$, $p < 0.01$]. The effect of intensity scaling was found to be significant [$F(1,8) = 11.36$, $p = 0.009$] and the distance perception improved for both lateral and near-medial locations in the presence of intensity cues.

V. DISCUSSIONS AND CONCLUSIONS

In this paper, a technique has been introduced to model distance-dependent individualized HRTFs in the horizontal plane. The main results are as follows:

- Auditory parallax cues at high frequencies displayed negligible perceptual effects for distance perception in the presence of ILD cues. Hence, the auditory parallax cues were not incorporated in the proposed model. The low-frequency ILD cues were found to be a prominent distance cue in anechoic headphone listening conditions.
- Distance-dependent individualized HRTFs in the horizontal plane were modeled using frontal projection headphones along with a cascade of signal processing blocks. Frontal projection of sound inherently modeled the frontal ipsilateral pinna cues at the eardrum spectra during sound playback. The head shadow effects in the contralateral ear were modeled by using the spherical head model. A rear-directional filter was developed to model rear-directional HRTFs from frontal HRTFs. In addition, an equalization filter (type-2 equalization) and a low-frequency compensation filter were required to accurately model the distance-dependent individualized HRTFs.
- The proposed individualization technique modeled the frontal ipsilateral HRTFs well. Modeling errors were larger for the contralateral ear, as the ipsilateral pinna cues are generated in the modeled contralateral HRTF due to frontal projection. For the rear directions, the modeling errors were larger than in the frontal directions. This is because the rear-directional filter used to model the rear directional HRTFs is a generic filter, which is based on

the average front-back spectral differences in the CIPIC database.

- Directional localization performances with the modeled and measured individualized HRTFs were similar with no significant differences for all the three sets of distances (near, medium, and far) for a significance level of 0.01. However, their differences were statistically significant for a significance level of 0.05. Larger errors in the modeled contralateral responses due to frontal projection did not have any perceptual effect in terms of localization and distance perception. Moreover, subjects could perceive the rear directions well, albeit with slight increase in error.
- Distance judgements for the modeled and measured individualized HRTFs were similar with no statistical differences for both lateral and near-medial directions for a significance level of 0.01. However, these differences were statistically significant for a significance level of 0.05. Presence of intensity cues had a significant effect and improved the distance perception for both lateral and near-medial directions.
- In comparison with the real sources, distance judgements in lateral directions for both modeled and real sources were similar with no significant differences for a significance level of 0.01. However, their differences were statistically significant for a significance level of 0.05. For near-medial directions, distance localization with real sound sources were more accurate than for virtual sounds synthesized with modeled distance-dependent HRTFs. The improved performance with real sound sources was probably due to reverberant cues present in the semi-anechoic chamber and absent in our model.

The frontal projection headphones used in this study are of open-type with FEC characteristics. FEC characteristics are vital particularly when the HPTF measurements are carried out at the blocked ear canal. Use of FEC headphones is critical in making auralization realistic and to ensure that sound pressure levels with and without headphones are similar (Møller *et al.*, 1995). Since the position of the transducer in frontal projection headphone does not vary with changes in virtual source azimuth, the pinna spectral cues introduced due to frontal projection also remain the same with changes in azimuth angle. This assumption is valid since the pinna-based spectral notch positions do not vary much with changes in azimuthal position in the horizontal plane (Lopez-Poveda and Meddis, 1996). The proposed method made use of a distance-invariant, frequency-independent, azimuth-dependent ITD value taken as the average of ITDs of all subjects from the CIPIC database to simulate lateral azimuthal perception in the horizontal plane. Other techniques can further be used to individualize the ITD to achieve a more accurate azimuthal perception (Algazi *et al.*, 2001b).

To conclude, a technique has been introduced in this work, to model distance-dependent individualized HRTFs in the horizontal plane. A direct extension of this work is to model the HRTFs at any elevation by manipulating the frontal notches created by frontal projection, which is an essential cue for elevation. Additional techniques can be used to increase the apparent source width of the auditory image as

the source moves towards the head, which will also be considered in future works.

ACKNOWLEDGMENTS

This work was supported by the Singapore Ministry of Education Academic Research Fund Tier-2, under research Grant No. MOE2010-T2-2-040.

- Algazi, V. R., Avendano, C., and Duda, R. O. (2001b). "Estimation of a spherical-head model from anthropometry," *J. Audio Eng. Soc.* **49**, 472–479.
- Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (2001a). "The CIPIC HRTF Database," in *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, 2001* (IEEE, New York), pp. 99–102.
- Avendano, C., Duda, R. O., and Algazi, V. R. (1999). "Modeling the contralateral HRTF," in *16th International Audio Engineering Society Conference on Spatial Sound Reproduction*, Rovaniemi, Finland.
- Begault, D. R. (2000). *3-D Sound for Virtual Reality and Multimedia* (Academic, Cambridge, MA), pp. 71–82, 99–115.
- Blauert, J. (1997). *Spatial Hearing* (Massachusetts Institute of Technology Press, Cambridge, MA), pp. 11, 12, 373.
- Bronkhorst, A. W., and Houtgast, T. (1999). "Auditory distance perception in rooms," *Nature* **397**(6719), 517–520.
- Brungart, D. S. (1999a). "Auditory localization of nearby sources. III. Stimulus effects," *J. Acoust. Soc. Am.* **106**, 3589–3602.
- Brungart, D. S. (1999b). "Auditory parallax effects in the HRTF for nearby sources," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 1999* (IEEE, New Paltz, NY), pp. 171–174.
- Brungart, D. S. (2001). "Preliminary model of auditory distance perception for nearby sources," in *Computational Models of Auditory Function* (IOS Press, Amsterdam), pp. 83–96.
- Brungart, D. S., and Brian, D. S. (2001). "Auditory localization of nearby sources in a virtual audio display," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2001* (IEEE, New Paltz, NY), pp. 107–110.
- Brungart, D. S., and Rabinowitz, W. M. (1999). "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Am.* **106**, 1465–1479.
- Calamia, P. T., and Hixson, E. L. (1997). "Measurement of the head-related transfer function at close range," *J. Acoust. Soc. Am.* **102**, 3117.
- Carlile, S., and Pralong, D. (1994). "The location-dependent nature of perceptually salient features of the human head-related transfer functions," *J. Acoust. Soc. Am.* **95**, 3445–3459.
- Duda, R. O., and Martens, W. L. (1998). "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.* **104**, 3048–3058.
- Duraiswami, R., Zotkin, D. N., and Gumerov, N. A. (2004). "Interpolation and range extrapolation of HRTFs [head related transfer functions]," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP*, Montreal, QC, Canada, Vol. 44, pp. iv-45–iv-48.
- Guang-Zheng, Y., Bo-Sun, X., and Dan, R. (2008). "Effect of sound source scattering on measurement of near-field head-related transfer functions," *Chin. Phys. Letters* **25**, 2926–2929.
- Hartley, R., and Fry, T. C. (1921). "The binaural location of pure tones," *Phys. Rev.* **18**, 431–444.
- Hayakawa, Y., Nishino, T., and Takeda, K. (2007). "Development of small sound equipment with micro-dynamic-type loudspeaker for HRTF measurement," in *Proceedings of the 19th International Congress on Acoustics*, Madrid, pp. 2–7.
- Hershkowitz, R., and Durlach, N. (1969). "Interaural time and amplitude JNDs for a 500-Hz tone," *J. Acoust. Soc. Am.* **46**, 1464–1467.
- Hirsch, H. R. (1968). "Perception of the range of a sound source of unknown strength," *J. Acoust. Soc. Am.* **43**, 373–374.
- Hosoe, S., Nishino, T., Itou, K., and Takeda, K. (2006). "Development of micro-dodecahedral loudspeaker for measuring head-related transfer functions in the proximal region," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP*, Toulouse, France, Vol. 5, pp. 329–332.
- Humanski, R. A., and Butler, R. A. (1988). "The contribution of the near and far ear toward localization of sound in the sagittal plane," *J. Acoust. Soc. Am.* **83**, 2300–2310.
- Iida, K., Ishii, Y., and Nishioka, S. (2014). "Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae," *J. Acoust. Soc. Am.* **136**, 317–333.
- Iwaya, Y. (2006). "Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears," *Acoust. Sci. Technol.* **27**, 340–343.
- Kahana, Y., and Nelson, P. A. (2006). "Numerical modeling of the spatial acoustic response of the human pinna," *J. Sound Vib.* **292**, 148–178.
- Kan, A., Jin, C., and van Schaik, A. (2009). "A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function," *J. Acoust. Soc. Am.* **125**, 2233–2242.
- Katz, B. F. (2001). "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," *J. Acoust. Soc. Am.* **110**, 2440–2448.
- Kim, H.-Y., Suzuki, Y., Takane, S., and Sone, T. (2001). "Control of auditory distance perception based on the auditory parallax model," *Appl. Acoust.* **62**, 245–270.
- Kim, S.-M., and Choi, W. (2005). "On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach," *J. Acoust. Soc. Am.* **117**, 3657–3665.
- Kistler, D. J., and Wightman, F. L. (1992). "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *J. Acoust. Soc. Am.* **91**, 1637–1647.
- Kreuzer, W., Majdak, P., and Chen, Z. (2009). "Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range," *J. Acoust. Soc. Am.* **126**, 1280–1290.
- Kulkarni, A., and Colburn, H. S. (2000). "Variability in the characterization of the headphone transfer-function," *J. Acoust. Soc. Am.* **107**, 1071–1074.
- Kulkarni, A., Isabelle, S., and Colburn, H. (1995). "On the minimum-phase approximation of head-related transfer functions," in *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, 1995* (IEEE, New Paltz, NY), pp. 84–87.
- Lopez-Poveda, E. A., and Meddis, R. (1996). "A physical model of sound diffraction and reflections in the human concha," *J. Acoust. Soc. Am.* **100**(5), 3248–3259.
- Mehrgardt, S., and Mellert, V. (1977). "Transformation characteristics of the external human ear," *J. Acoust. Soc. Am.* **61**, 1567–1576.
- Menzies, D. (2009). "Calculation of near-field head related transfer functions using point source representations," in *Ambisonics Symposium*, IEM Graz, pp. 23–28.
- Mershon, D. H., Ballenger, W. L., Little, A. D., McMurtry, P. L., and Buchanan, J. L. (1989). "Effects of room reflectance and background noise on perceived auditory distance," *Perception* **18**(3), 403–416.
- Mershon, D. H., and Bowers, J. N. (1979). "Absolute and relative cues for the auditory perception of egocentric distance," *Perception* **8**(3), 311–322.
- Mershon, D. H., and King, L. E. (1975). "Intensity and reverberation as factors in auditory perception of egocentric distance," *Percept. Psychophys.* **18**(6), 409–415.
- Middlebrooks, J. C. (1999). "Individual differences in external-ear transfer functions reduced by scaling in frequency," *J. Acoust. Soc. Am.* **106**, 1480–1492.
- Middlebrooks, J. C., and Green, D. M. (1992). "Observations on a principal components analysis of head-related transfer functions," *J. Acoust. Soc. Am.* **92**, 597–599.
- Middlebrooks, J. C., Makous, J. C., and Green, D. M. (1989). "Directional sensitivity of sound-pressure levels in the human ear canal," *J. Acoust. Soc. Am.* **86**, 89–108.
- Molino, J. (1973). "Perceiving the range of a sound source when the direction is known," *J. Acoust. Soc. Am.* **53**, 1301–1304.
- Møller, H., Sørensen, M. F., Hammershøj, D., and Jensen, C. B. (1995). "Head-related transfer functions of human subjects," *J. Audio Eng. Soc.* **43**, 300–321.
- Morimoto, M., and Ando, Y. (1980). "On the simulation of sound localization," *J. Acoust. Soc. Jpn. (E)* **1**, 167–174.
- Müller, S., and Massarani, P. (2001). "Transfer-function measurement with sweeps," *J. Audio Eng. Soc.* **49**, 443–471.
- Nicol, R. (2010). "Binaural technology," *Audio Engineering Society Monograph*, New York, NY, pp. 45–64.
- Qu, T., Xiao, Z., Gong, M., Huang, Y., Li, X., and Wu, X. (2009). "Distance dependent head-related transfer functions measured with high spatial

- resolution using a spark gap," *IEEE Trans. Audio Speech Lang. Process.* **17**, 1124–1132.
- Rabinowitz, W. M., Maxwell, J., Shao, Y., and Wei, M. (1993). "Sound localization cues for a magnified head—Implications from sound diffraction about a rigid sphere," *Presence (Camb.)* **2**, 125–129.
- Romblom, D., and Cook, B. (2008). "Near-field compensation for HRTF processing," in *125th Audio Engineering Society International Convention*, San Francisco, CA.
- Rui, Y., Yu, G., Xie, B., and Liu, Y. (2013). "Calculation of individualized near-field head-related transfer function database using boundary element method," in *134th Audio Engineering Society Convention*, Rome, Italy.
- Seeber, B. U., and Fastl, H. (2003). "Subjective selection of non-individual head-related transfer functions," in *Proceedings of the 2003 International Conference on Auditory Display*, Boston, MA, pp. 1–4.
- Shaw, E., and Teranishi, R. (1968). "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," *J. Acoust. Soc. Am.* **44**, 240–249.
- Shinn-Cunningham, B., Santarelli, S., and Kopco, N. (2000a). "Distance perception of nearby sources in reverberant and anechoic listening conditions: Binaural vs. monaural cues," in *23rd Mid-winter Meeting of the Association for Research in Otolaryngology*, St. Petersburg Beach, FL, p. 88.
- Shinn-Cunningham, B. G., Santarelli, S., and Kopco, N. (2000b). "Tori of confusion: Binaural localization cues for sources within reach of a listener," *J. Acoust. Soc. Am.* **107**, 1627–1636.
- Simpson, W., and Stanton, L. D. (1973). "Head movement does not facilitate perception of the distance of a source of sound," *Am. J. Psychol.* **86**, 151–159.
- Stevens, S., and Guirao, M. (1962). "Loudness, reciprocity, and partition scales," *J. Acoust. Soc. Am.* **34**, 1466–1471.
- Stewart, G. (1911). "The acoustic shadow of a rigid sphere, with certain applications in architectural acoustics and audition," *Phys. Rev. (Series I)* **33**, 467–479.
- Sunder, K., Tan, E.-L., and Gan, W.-S. (2013). "Individualization of binaural synthesis using frontal projection headphones," *J. Audio Eng. Soc.* **61**, 989–1000.
- Sunder, K., Tan, E.-L., and Gan, W.-S. (2014). "Effect of headphone equalization on auditory distance perception," in *137th Audio Engineering Society Convention*, Audio Engineering Society, 2014.
- Tahara, Y., and Sakurai, H. (1974). "A tentative model for the localization of sound based on the simultaneity of time difference and level difference of the sound between ears," in *Proceedings of the Congress of the Acoustical Society of Japan*, pp. 373–374.
- Takemoto, H., Mokhtari, P., Kato, H., Nishimura, R., and Iida, K. (2012). "Mechanism for generating peaks and notches of head-related transfer functions in the median plane," *J. Acoust. Soc. Am.* **132**, 3832–3841.
- Tan, C. J., and Gan, W. S. (1998). "User-defined spectral manipulation of HRTF for improved localisation in 3D sound systems," *Electron. Lett.* **34**, 2387–2389.
- von Békésy, G. (1949). "The Moon illusion and similar auditory phenomena," *Am. J. Psychol.* **62**, 540–552.
- Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (1993). "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.* **94**, 111–123.
- Wightman, F., and Kistler, D. (1993). "Multidimensional scaling analysis of head-related transfer functions," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 1993*, IEEE, pp. 98–101.
- Wightman, F. L., and Kistler, D. J. (1989). "Headphone simulation of free-field listening. II: Psychophysical validation," *J. Acoust. Soc. Am.* **85**, 868–878.
- Wightman, F. L., and Kistler, D. J. (1992). "The dominant role of low-frequency interaural time differences in sound localization," *J. Acoust. Soc. Am.* **91**, 1648–1661.
- Xu, S., Li, Z., and Salvendy, G. (2007). "Individualization of head-related transfer function for three-dimensional virtual auditory display: A review," in *Virtual Reality* (Springer, Berlin), pp. 397–407.
- Zahorik, P. (2000). "Distance localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.* **108**, 2597.
- Zahorik, P. (2002a). "Assessing auditory distance perception using virtual acoustics," *J. Acoust. Soc. Am.* **111**, 1832–1846.
- Zahorik, P. (2002b). "Auditory display of sound source distance," in *Proceedings of the 8th International Conference on Auditory Display*, Kyoto, Japan, pp. 1–7.
- Zahorik, P., Brungart, D. S., and Bronkhorst, A. W. (2005). "Auditory distance perception in humans: A summary of past and present research," *Acustica* **91**, 409–420.
- Zahorik, P., and Wightman, F. L. (2001). "Loudness constancy with varying sound source distance," *Nat. Neurosci.* **4**, 78–83.
- Ze-Wei, C., Guang-Zheng, Y., Bo-Sun, X., and Shan-Qun, G. (2012). "Calculation and analysis of near-field head-related transfer functions from a simplified head-neck-torso model," *Chin. Phys. Lett.* **29**, 034302.