A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function

Alan Kan, a, Craig Jin, b and André van Schaik c
Computing and Audio Research Laboratory, School of Electrical and Information Engineering, University of Sydney, Sydney, New South Wales 2006, Australia

(Received 3 September 2008; revised 6 January 2009; accepted 20 January 2009)

A method for synthesizing near-field head-related transfer functions (HRTFs) from far-field HRTFs measured using an acoustic point-source of sound is presented. Near-field HRTFs are synthesized by applying an analytic function describing the change in the transfer function when the location of a sound source changes from the far-field to the near-field: the distance variation function (DVF). The DVF is calculated from a rigid sphere model and approximates the change in the frequency-dependent interaural level cues as a function of the change in sound source distance. Using a sound localization experiment, the fidelity of the near-field virtual auditory space (VAS) generated using this technique is compared to that obtained by simply adjusting the intensity of the VAS stimulus to simulate changes in distance. Results show improved distance perception for sounds at simulated distances of up to 60 cm using the DVF compared to simple intensity adjustment, while maintaining directional accuracy. The largest improvement for distance perception were for sound sources located to the side and within 40 cm. When intensity was removed as a cue for sound source distance from near-field sounds generated using the DVF, results showed some discrimination of sound source distances but, in general, distance perception accuracy was poor. © 2009 Acoustical Society of America. [DOI: 10.1121/1.3081395]

PACS number(s): 43.66.Qp, 43.66.Pn [JCM] Pages: 2233–2242

I. INTRODUCTION

A head-related transfer function (HRTF) or its time-domain representation, the head-related impulse response (HRIR) describes the change in the sound pressure from a point in the free-field to a listener’s ear drum. A separate HRTF exists for each sound source position. Commonly, the HRTF is defined for a planar sound source; however, in this paper we will consider the HRTF for an acoustic point-source of sound. For an acoustic point-source of sound, HRTFs do not vary substantially for distances beyond 1 m since the incoming sound waves are approximately planar, but Brungart and Rabinowitz (1999) showed using acoustic measurements on a manikin that HRTFs can vary substantially for distances within 1 m. This variation is due to the curvature of the wavefront from the point-source of sound and its interaction with the subject’s head, torso, and pinnae. Hence, it can be seen that the HRTFs within 1 m are dependent on the distance from the head due to the wavefront characteristics of the sound source.

For an acoustic point-source of sound located off the mid-sagittal plane, Brungart and Rabinowitz (1999) showed that the interaural level difference (ILD) increases as the distance to the head decreases, especially when the sound source is located within 0.5 m from the head. In particular, the low-frequency ILD (<500 Hz), which is usually small at sound source distances greater than 1 m, can be on the order of 20 dB for a sound source located at 0.12 m. Also, the HRTF for the contralateral ear shows increasing high-frequency attenuation with decreasing sound source distance. Results from human sound localization experiments conducted by Brungart et al. (1999) and Brungart (1999) suggest that the level differences in the HRTFs are a salient cue for distance localization for sound source distances within 1 m. From acoustic manikin measurements, Brungart and Rabinowitz (1999) also showed the existence of a shift in the spectral features of the HRTF, particularly at the high frequencies, as a function of distance, which is due to an “acoustic parallax” effect. This occurs when the angle between a sound source and the ear changes substantially when the sound source is moved closer to the head. The measurements also exhibit a small increase in the interaural time difference (ITD) at very close sound source distances (<0.25 m). However, Brungart and Rabinowitz (1999) argued that the small increase in ITD is likely to be insignificant and that the ILD in this region is the more dominant cue for sound source distance localization. To distinguish the region where the HRTFs exhibit significant variations with distance from the region where the HRTFs are relatively constant, the terms “near-field” and “far-field” will be used in this paper to refer to these two regions, respectively. A distance of 1 m is taken as the boundary between the two regions.

For the generation of virtual auditory space (VAS), it is common practice to acoustically measure HRIRs for a modest number of different sound source directions at a fixed radius of 1 to 2 m from the center of the listener’s head. To estimate an HRIR for a sound source direction that has not been measured, numerous interpolation techniques have been proposed [for example, see Wenzel and Foster (1993); Har-
of sound from any direction and distance in space to the surface of a rigid sphere was derived by Rabinowitz et al. (1993) and has been experimentally verified on a bowling ball by Duda and Martens (1998). The pressure at a point \( X \) on the surface of a rigid sphere of radius \( a \) due to a sinusoidal point-source of sound of angular frequency \( \omega \), at a distance \( r \) away from the center of the sphere, is given by

\[
p(a, \omega, r) = -kr \sum_{m=0}^{\infty} (2m+1) \frac{h_m(ka)}{h_m'(ka)} P_m(\cos \theta) e^{-ikr},
\]

where \( h_m(ka) \) is the spherical Hankel function of the first kind of order \( m \) and \( h_m'(ka) \) is its first derivative at radius \( a \), \( k = \omega/c \) is the wavenumber, \( c \) is the speed of sound in air, \( P_m(\lambda) \) is the Legendre polynomial of degree \( m \), and \( \theta \) is the angle between a vector from the center of the sphere to the point \( X \) on the surface of the sphere and a vector from the center of the sphere to the sound source. By evaluating Eq. (1) for all frequencies of interest, the transfer function at point \( X \) can be obtained for a particular source location. In our experiments, we calculated the transfer function for 256 frequency bins up to the Nyquist frequency.

To synthesize a near-field HRTF, the DVF for a sound source at a near-field distance \( d_n \) is calculated as the ratio of the pressure arising from a source at distance \( d_n \) to the pressure arising from a source at a distance \( d_f \), where \( d_f \) is the distance at which the far-field HRTF was measured. In other words, we calculate

\[
\text{DVF} = \frac{p_n(a, \omega, \theta, d_n)}{p(a, \omega, \theta, d_f)}
\]

using Eq. (1), for all frequencies of interest, where \( a \) is the radius of the listener’s head and can be determined through measurement. In our experiment, the radius of the listener’s head was estimated using the ITD model proposed by Kuhn (1977):

\[
\text{ITD} = \frac{3a}{c} \sin \theta_{\text{inc}},
\]

where \( \theta_{\text{inc}} \) is the angle of incidence of the sound source relative to the mid-sagittal axis. ITD values were obtained by cross-correlation of the measured left and right HRIRs below 5 kHz. To determine the angle \( \theta \) to be used in Eq. (2), the positions of the ears were assumed to be at an angle of 100° on either side of the mid-sagittal plane on the audio-visual horizon for all subjects, in line with Duda and Martens (1998) and Blauert (1997). A separate DVF is calculated for each ear. To synthesize a near-field HRTF from a far-field HRTF, we calculate

\[
\text{HRTF}(d_n) = \text{DVF} \times \text{HRTF}(d_f).
\]

To synthesize sounds at multiple locations and distances, a DVF needs to be calculated for every sound source distance and direction and for each ear.

II. METHODS

A. Calculation of the DVF

The DVF is calculated using a rigid sphere model. The function describing the pressure of a sinusoidal point-source of sound from any direction and distance in space to the surface of a rigid sphere was derived by Rabinowitz et al. (1993) and has been experimentally verified on a bowling ball by Duda and Martens (1998). The pressure at a point \( X \) on the surface of a rigid sphere of radius \( a \) due to a sinusoidal point-source of sound of angular frequency \( \omega \), at a distance \( r \) away from the center of the sphere, is given by

\[
p(a, \omega, r) = -kr \sum_{m=0}^{\infty} (2m+1) \frac{h_m(ka)}{h_m'(ka)} P_m(\cos \theta) e^{-ikr},
\]

where \( h_m(ka) \) is the spherical Hankel function of the first kind of order \( m \) and \( h_m'(ka) \) is its first derivative at radius \( a \), \( k = \omega/c \) is the wavenumber, \( c \) is the speed of sound in air, \( P_m(\lambda) \) is the Legendre polynomial of degree \( m \), and \( \theta \) is the angle between a vector from the center of the sphere to the point \( X \) on the surface of the sphere and a vector from the center of the sphere to the sound source. By evaluating Eq. (1) for all frequencies of interest, the transfer function at point \( X \) can be obtained for a particular source location. In our experiments, we calculated the transfer function for 256 frequency bins up to the Nyquist frequency.

To synthesize a near-field HRTF, the DVF for a sound source at a near-field distance \( d_n \) is calculated as the ratio of the pressure arising from a source at distance \( d_n \) to the pressure arising from a source at a distance \( d_f \), where \( d_f \) is the distance at which the far-field HRTF was measured. In other words, we calculate

\[
\text{DVF} = \frac{p_n(a, \omega, \theta, d_n)}{p(a, \omega, \theta, d_f)}
\]

using Eq. (1), for all frequencies of interest, where \( a \) is the radius of the listener’s head and can be determined through measurement. In our experiment, the radius of the listener’s head was estimated using the ITD model proposed by Kuhn (1977):

\[
\text{ITD} = \frac{3a}{c} \sin \theta_{\text{inc}},
\]

where \( \theta_{\text{inc}} \) is the angle of incidence of the sound source relative to the mid-sagittal axis. ITD values were obtained by cross-correlation of the measured left and right HRIRs below 5 kHz. To determine the angle \( \theta \) to be used in Eq. (2), the positions of the ears were assumed to be at an angle of 100° on either side of the mid-sagittal plane on the audio-visual horizon for all subjects, in line with Duda and Martens (1998) and Blauert (1997). A separate DVF is calculated for each ear. To synthesize a near-field HRTF from a far-field HRTF, we calculate

\[
\text{HRTF}(d_n) = \text{DVF} \times \text{HRTF}(d_f).
\]

To synthesize sounds at multiple locations and distances, a DVF needs to be calculated for every sound source distance and direction and for each ear.
equipped with a loudspeaker (VIFA-D26TG-35) mounted on a robotic arm. The robotic arm can accurately position the loudspeaker to within a fraction of a degree, at any point on the surface of an imaginary sphere of 1 m radius except for positions greater than 45° below the audio-visual horizon. The subject was seated with his head in the center of the measurement sphere and aligned to the axes of the measurement system with the aid of a laser-alignment system. A chin-rest was provided and an electromagnetic sensor mounted on a headband worn by the subject monitored his head orientation during the measurement process. A light emitting diode (LED) display provided feedback to the subject with regard to the alignment of his head.

HRIRs were recorded using a “blocked-ear canal” technique, as described by Møller (1992), by placing miniature microphones (AuSim) at the entrance of the ear canal in both ears. Complementary Golay codes were used as stimuli and the HRIRs were obtained using the steps described by Zhou et al. (1992). In order to improve the signal-to-noise ratio, 16 repetitions of the Golay codes, each with a 1024 sample length, were recorded at each position. Tucker Davis Technology (TDT) system II hardware, interfaced with customized MATLAB software, was used to play and record the codes at an 80 kHz sampling rate. The stimulus duration at each recording position was just under 1 s. HRIRs were recorded for 393 different sound source directions, upwards of 45° below the audio-visual horizon and equally distributed around the subject’s head.

After recording, a number of post-processing steps were applied to the HRIRs. First, the HRIRs were low-pass filtered at 16 kHz and resampled at 48 kHz. Second, since the HRIRs recorded with our system are at the limits of the noise floor for frequencies below 500 Hz, the recorded HRIRs were compensated below 500 Hz for each direction according to the frequency response derived from the rigid sphere model in Eq. (1) by replacing the HRTF magnitudes for frequencies below 500 Hz with those calculated by Eq. (1) and scaled relative to the level of the measured HRTF at 500 Hz. It should be noted that the HRIRs used in this experiment contain the transfer function of the measurement loudspeaker. Third, the magnitude response of the ear canal resonance up to 4 kHz, measured on a Briel and Kjær Head and Torso Simulator (HATS) manikin (Type 4128), was also added to the HRIRs to compensate for the missing ear canal resonance caused by the blocked-ear recording and the use of in-ear tube phones for VAS presentation. The magnitude of the ear canal resonance was obtained from a blocked-ear recording made with the Ausim recording microphones on the HATS manikin and an open-ear recording made with the ear simulator microphones inside the HATS manikin. Assuming both the Ausim microphones and ear simulator microphones have a flat frequency response, the ear canal resonance can be obtained by subtracting the blocked-ear canal HRTF from the open-ear canal HRTF in the log-magnitude domain. The ear canal resonance obtained has a similar response as that presented in Hammershøi and Møller (1996) for the mean ratio of the pressure at the blocked-ear to the pressure at the ear-drum for 12 subjects, and was applied to the recorded far-field HRIRs as a minimum-phase filter.

C. Stimulus generation

A psychoacoustic experiment was conducted to test the perceptual fidelity of the VAS generated by the synthesized near-field sounds. White noise bursts, 400 ms in duration with 5 ms raised-cosine onset and offset ramps, were used as stimuli for the experiment. A different white noise burst was generated for each sound source location. In this experiment three conditions were tested, in which the near-field stimuli were generated by (1) filtering white noise bursts with near-field HRIRs synthesized by the DVF, henceforth referred to as the DVF-generated stimulus condition; (2) filtering white noise bursts with far-field HRIRs and scaling the intensity of the stimulus inversely proportional to the simulated sound source distance, henceforth referred to as the intensity-scaled stimulus condition; and (3) filtering white noise bursts with near-field HRIRs synthesized by the DVF but with constant intensity, henceforth referred to as the constant-intensity DVF-generated stimulus condition.

To generate stimuli for the DVF-generated stimulus condition, white noise bursts were convolved with near-field HRIRs synthesized by the DVF according to Eq. (2) and (4). A constant post-scaling factor was then applied to the stimuli to ensure that the stimuli at positions closest to the ears would be at a comfortable listening level.

In the intensity-scaled stimulus condition, we desired a scaling factor that would change the intensity of the stimuli with respect to the simulated distance and account for the fact that the distance between a sound source and the two ears can be substantially different when the source is located in the near-field and close to the interaural axis. For a real sound source, the intensity of the sound is inversely proportional to the distance, that is \( I \propto \frac{1}{d} \). To account for the different distances to the two ears, the intensity can be assumed to be approximately proportional to the mean of the two distances, that is,

\[
I \approx \frac{1}{2} \left[ \frac{1}{d_l} + \frac{1}{d_r} \right],
\]

where \( I \) is the intensity and \( d_l \) and \( d_r \) are the distances between the sound source and the left and right ears, respectively. The scaling factor, \( S \), can be defined as the ratio of the intensity level for a sound source at a near-field distance over the intensity level for a sound source at 1 m. Assuming the distance to the two ears are identical for a sound source at 1 m (e.g., along the mid-line), then

\[
S = \frac{I_{\text{near field}}}{I_{1 \text{ m}}} = \frac{50}{d_l} + \frac{50}{d_r},
\]

where \( d_l \) and \( d_r \) are in centimeters. The stimuli for the intensity-scaled stimulus condition were thus obtained by convolving white noise bursts with far-field HRIRs and scaling by the numerical factor \( S \). An additional post-scaling factor was applied to the stimuli to ensure that a sound directly in front of the subject at a distance of 1 m would have the same root-mean-squared energy as a sound for the DVF-generated stimulus condition for the same position.

In the constant-intensity DVF-generated stimulus condition, near-field sounds were generated as described for the
DVF-generated stimulus condition with an additional scaling factor of $1/S$ [refer to Eq. (6)] applied to the sounds to achieve approximately constant intensity for all simulated sound source distances. This condition was included to investigate the salience of the remaining cues introduced by the DVF for sound source distance perception when the intensity cue is absent.

Figure 1 shows the range of stimulus levels at the left and right ears for a sound source located directly in front of a subject at different simulated distances. Similar stimulus levels were presented to the left ear for the same sound source positions.

DVF

FIG. 1. (Color online) The stimulus level presented to the right ear for a sound source located directly in front of a subject at different simulated distances. Similar stimulus levels were presented to the left ear for the same sound source positions.

D. Experimental setup

A sound localization task was used for the experiment whereby the subjects indicated the perceived location of the near-field sounds by placing an electromagnetically-tracked sensor, mounted on the tip of a wand, at the perceived location. Brungart et al. (2000) found this method to be the least biased and more accurate than other types of response methods for indicating the perceived location of a sound. An electromagnetic tracking system (Polhemus Fastrak), with two sensors, was used to measure the perceived sound source location relative to the center of the listener's head. One of the sensors was mounted on the tip of the wand, and the other on a rigid headband worn by the listener. The transmitter of the electromagnetic tracking system was located about 50 cm behind the subject’s head on a wooden pole, as shown in Fig. 2, and connected to a computer controlling the experiment from an adjacent room. A handheld pushbutton was given to the subject to trigger the computer and an LED display was used to provide head-orientation feedback to the listener prior to each stimulus presentation. Sounds were presented through in-ear tube phones (Etymotic Research ER-2) and played through an RME HDSP soundcard connected to the computer. The experiments were conducted in an ordinary room with no acoustical treatment. The room was lit during the experiment so that the subject could see the exact placement of the sensor at the end of the wand.

At the start of each trial the position of the center of the subject’s head relative to the sensor on the headband was estimated by sequentially placing the sensor on the wand on the left and right ears and recording the sensor positions. The center of the head was then calculated as the midpoint between the recorded locations of the left and right ears. Prior to each stimulus presentation, the subject aligned his head to a prescribed starting position with the aid of an LED display [Fig. 2(a)]. The subject then pressed the handheld pushbutton causing the position of the sensor on the headband worn on the subject’s head to be recorded by the computer to determine the current location of the center of the head and the stimulus to be played. The subject then responded by placing...
the sensor at the end of the wand at the perceived location, as shown in Fig. 2(b), and pressed the handheld pushbutton again. A reading was then taken from the sensor at the end of the wand to capture the perceived location. The perceived direction and distance of the sound stimulus were calculated relative to the location of the center of the subject’s head prior to stimulus presentation. The subject then returned to the prescribed starting position for the next stimulus presentation.

In each test session, subjects would localize near-field sounds at each of the four distance ranges for 76 different directions. The stimuli for all directions and distances were presented in random order in each session. Subjects completed five sessions each for each stimulus condition described in Sec. II C. Since the different stimulus conditions were tested in separate sessions, subjects were aware of the method that had been applied to produce the stimuli during the test session. A total of five subjects participated in these experiments. The subjects were all male, between the ages of 26 and 37, and had normal hearing in both ears. All subjects had prior training and experience in human auditory localization experiments involving a head-pointing paradigm [see Carlile et al. (1997)] but no new training was given to the subjects for this new localization paradigm.

E. Analysis of results

The results of the psychoacoustic experiments will be presented in terms of direction and distance separately. The directional localization data will be presented using a lateral-polar angle coordinate system [see Fig. 3]. The lateral angle for a particular point \( T \) is defined as the horizontal angle away from the \( XZ \)-plane given by \( \angle AOX \) in Fig. 3. Negative lateral angles (down to \(-90^\circ\)) and positive lateral angles (up to \(90^\circ\)) define the left and right hemispheres, respectively. The polar angle corresponding to the point \( T \) is given by \( \angle TBA \) in Fig. 3. Polar angles are defined with \(0^\circ\) representing the front, \(90^\circ\) representing directly above, \(180^\circ\) representing the back, and \(270^\circ\) representing directly below. The distance localization data will be presented with reference to the center of the head. Scatter plots will be used to display the data in terms of lateral angle, polar angle, and distance localization, where the target angle or distance is plotted against the response angle or distance. Perfect localization, that is, when the response location matches the target location, is indicated by circles along the diagonal from the bottom left to top right corner of the scatter plot. Mean differences between target and response locations across the localization data for all subjects will also be presented along with results from statistical significance tests. Kruskal–Wallis non-parametric analysis of variance (ANOVs) and post hoc analyses (Tukey HSD) were conducted on the data within each condition, as well as across the conditions. An alpha value of 0.05 was used for significance testing. The corresponding non-parametric effect size was estimated by calculating Cliff’s \( d \), as described by Cliff (1993), where \( d \) estimates the fraction of the total number of samples where the samples in condition \( A \) are greater than the samples in condition \( B \), minus the fraction of the total number of samples in condition \( B \) that are greater than the samples in condition \( A \). That is, we calculate

\[
d = \frac{1}{mn} \sum_i \sum_j d_{ij},
\]

where \( m \) and \( n \) are the number of samples in conditions \( A \) and \( B \), respectively, and

\[
d_{ij} = \text{sign}(x_i - x_j),
\]

where \( i = 1, 2, \ldots, m \) and \( j = 1, 2, \ldots, n \). Equation (8) compares a sample \( x_i \) from condition \( A \) with a sample \( x_j \) from condition \( B \) and assigns \( d_{ij} = 1 \) if \( x_i > x_j \), \( d_{ij} = -1 \) if \( x_i < x_j \), and \( d_{ij} = 0 \) if \( x_i = x_j \).

III. RESULTS

A. Directional localization

The directional localization results in terms of lateral angle are presented in Fig. 4(a) for the three stimulus conditions. The lateral angle data have been rounded to the nearest \(5^\circ\) and the data for all the subjects have been grouped together since all subjects showed similar responses in each condition. Each column shows the localization data grouped into the different distance ranges and each row contains the results for the different conditions. The size of the circle indicates the number of responses in a particular angular region, as described by the legend at the bottom of the figure. In all stimulus conditions, the localization data show that the lateral angles of the responses were close to the target lateral angles at the different distance ranges implying that there were no major lateral angle localization errors in all stimulus conditions. Figure 4(b) shows the mean absolute difference between the target and response lateral angle for the different distance ranges for the three stimulus conditions. In all conditions there is a statistically significant increase in absolute lateral angle difference at the 10–20 cm distance range but the effect is small (Table I). Comparing across the conditions, there is a statistically significant increase in the absolute lateral angle difference in the two DVF stimulus conditions compared to the intensity-scaled stimulus condition \( [\chi^2(2, 7600) = 17.64, p < 0.005] \) but again the effect is small \( (d = 0.08) \).
The directional localization results in terms of polar angle are presented in Fig. 5(a), where the data are shown in a similar format as Fig. 4(a). The polar angles have been rounded to the nearest 10°. The data show that, in general, the polar angles of the responses were close to the target polar angles implying that there were no major polar angle localization errors in all stimulus conditions. Some of the localization differences between the target and response polar angles can be attributed to front-back confusions, where the lateral angle of the response is correct but the perceived direction is incorrectly placed in the wrong front-back hemisphere (Carlile et al., 1997). In our analysis, a front-back confusion is identified when the response lateral angle is within 10° of the target lateral angle but the response is located in the wrong front-back hemisphere. The percentage of front-back errors for the different stimulus conditions are shown in Fig. 6. In all conditions, the percentage of front-back errors are of a similar order as free-field sound localization experiments using a head-pointing response paradigm reported elsewhere (Carlile et al., 1997). The mean absolute difference between the target and response polar angle in each distance range for the three stimulus conditions is shown in Fig. 5(b), where front-back confusions have been removed. There are statistically significant increases in the

TABLE I. The results of a Kruskal–Wallis non-parametric ANOVA conducted on the absolute lateral angle differences across the distance ranges are shown for each stimulus condition. The table shows the \( \chi^2 \), degrees of freedom (dof) and \( p \)-values. A post hoc analysis (Tukey HSD) revealed a statistically significant difference in localization performance between the 10–20 cm distance range and the other distance ranges in all stimulus conditions. The index measuring the size of this effect, \( d \), is also shown in the table, and is small for all conditions.

<table>
<thead>
<tr>
<th>Stimulus condition</th>
<th>( \chi^2 )</th>
<th>dof</th>
<th>( p )</th>
<th>( d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity-scaled</td>
<td>8.53</td>
<td>3</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td>DVF</td>
<td>28.93</td>
<td>3</td>
<td>&lt;0.005</td>
<td>0.08</td>
</tr>
<tr>
<td>Constant-intensity DVF</td>
<td>20.76</td>
<td>3</td>
<td>&lt;0.005</td>
<td>0.06</td>
</tr>
</tbody>
</table>

FIG. 4. (a) shows the directional localization results in terms of lateral angle. The data for the three stimulus conditions are shown in each row and are grouped into different distance ranges. (b) shows the mean absolute difference between the target and response lateral angle. The bars are grouped into different distance ranges and within each group, and the mean absolute angular difference is shown for the three stimulus conditions. The error bars indicate the 95% confidence interval of the mean.

FIG. 5. (a) shows the directional localization results in terms of polar angle and (b) shows the mean absolute difference between the target and response polar angle. The data are shown in the same format as in Fig. 4.
absolute polar angle difference between the target and response angle in the 10–20 cm distance range compared to the other distance ranges for the two DVF-generated stimulus conditions and in the 60–100 cm distance range for the constant-intensity stimulus condition though the effect is small (Table II). Comparing across the stimulus conditions, there are statistically significant differences in absolute polar angle localization performance at the different distance ranges but again the effect is small (Table III).

**B. Distance perception**

The perceived distances for the three stimulus conditions are shown in Fig. 7. Distances have been rounded to the nearest 5 cm. The data for both the left and right-hand sides have been grouped and divided into one of three azimuth regions shown in each column: front (0°–60°), side (60°–120°), and back (120°–180°), where 0° is directly in front. Each row shows the data for each stimulus condition. In all conditions, the localization data show a high degree of overestimation of the intended distance of the sound source indicated by the responses above the dashed line which indicates the theoretical responses for accurate distance perception. The mean perceived distance for all stimulus conditions separated into different distance ranges and azimuth regions is shown in Fig. 8(a). In the intensity-scaled and DVF-generated stimulus conditions, the mean perceived distance follows the expected trend; that is, at the further distance ranges the mean perceived distance was larger than the closer distance ranges. However, the mean perceived distance is generally outside the intended distance range. In the constant-intensity DVF-generated stimulus condition, the mean perceived distance only fits the expected trend when the stimulus was located toward the side region. In the front and back regions, subjects generally perceived the stimulus as being at 50 and 65 cm, respectively, for all target distance ranges. From Fig. 7, it can also be seen that subjects exhibited a high amount of variability in perceived distance accuracy indicated by the spread of responses. The mean difference between the target and response distance for the different distance ranges and azimuth groups is shown in Fig. 8(b), where a positive difference indicates overestimation of the simulated distance. It can be seen that in all conditions subjects tended to overestimate the intended sound source distance when the stimulus was within 60 cm with slightly

<table>
<thead>
<tr>
<th>Distance range</th>
<th>$\chi^2$</th>
<th>dof</th>
<th>$p$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10–20 cm</td>
<td>7.84</td>
<td>2</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td>25–35 cm</td>
<td>0.03</td>
<td>2</td>
<td>0.999</td>
<td>na</td>
</tr>
<tr>
<td>40–60 cm</td>
<td>11.74</td>
<td>2</td>
<td>&lt;0.005</td>
<td>0.04</td>
</tr>
<tr>
<td>60–100 cm</td>
<td>49.53</td>
<td>2</td>
<td>&lt;0.005</td>
<td>0.07</td>
</tr>
</tbody>
</table>

![FIG. 6. The percentage front-back errors for the three stimulus conditions. The data are shown in the same format as in Fig. 4(b).](image-url)

![FIG. 7. The perceived distances are shown for the three stimulus conditions. Each column shows the data for one of three azimuth regions defined as front (0°–60°), side (60°–120°), and back (120°–180°), where 0° is directly in front. Each row shows the data for each stimulus condition. The dashed line indicates the theoretical responses for accurate distance perception.](image-url)
IV. DISCUSSION

We conducted a psychoacoustic experiment to investigate the perceptual fidelity of near-field VAS generated using near-field HRTFs synthesized using the DVF method compared to that generated by simply scaling the intensity of the sound stimulus. A comparison of the distance perception from the DVF-generated stimulus condition and the intensity-scaled condition shows a statistically significant difference in localization performance in the side region between 60° and 120° and within 40 cm of the subject. In this region, subjects exhibited substantial overestimation of the sound source distance when intensity-scaled stimuli were used (on average, 35 cm from the intended sound source distance) compared to less than an average of 25 cm for the DVF-generated stimulus condition. This indicates that simply adjusting the intensity of the stimulus may not be enough to indicate to subjects that the sound stimulus is intended to be within 40 cm. Also, calculation of the effect size index showed that the improvement in distance perception using the DVF-generated stimulus over the intensity-scaled stimulus is large. The improvement in distance perception can be understood by studying the ILD introduced by the DVF. The ILD is calculated by taking the difference in sound source distance when intensity-scaled stimuli were used on average, 35 cm from the intended sound source distance compared to less than an average of 25 cm for the DVF-generated stimulus condition.

The ILD was calculated for a rigid sphere of 7 cm radius located within 60 cm and calculation of the effect size index showed this improvement to be moderately large ($d=0.41$). The largest improvement in distance perception is in the side region within 40 cm of the subject ($d=0.61$).

![FIG. 8. (a) shows the mean perceived distance of each distance range for the three stimulus conditions. (b) shows the mean difference between the target and response distance, where the bars are grouped into the different azimuth regions and the results for the three stimulus conditions are shown within each group. The error bars indicate the 95% confidence interval of the mean.

![FIG. 8 continued.](image)

The ILD is calculated by taking the difference in sound source distance when intensity-scaled stimuli were used on average, 35 cm from the intended sound source distance compared to less than an average of 25 cm for the DVF-generated stimulus condition. This indicates that simply adjusting the intensity of the stimulus may not be enough to indicate to subjects that the sound stimulus is intended to be within 40 cm. Also, calculation of the effect size index showed that the improvement in distance perception using the DVF-generated stimulus over the intensity-scaled stimulus is large. The improvement in distance perception can be understood by studying the ILD introduced by the DVF. The ILD is calculated by taking the difference (in decibels) between the left and right DVFs for a particular sound source position. The magnitude of the ILD introduced by the DVF is shown as contour plots for different frequencies in Fig. 9. The DVF was calculated for a rigid sphere of 7 cm radius with “ears” located at ±100° azimuth from the mid-sagittal plane on the audio-visual horizon. Within each plot, the horizontal axis shows the distance between the sound source and the center of the head (on a logarithmic scale) and the vertical axis shows the lateral angle of the sound source. Areas of approximately equal ILD magnitude are indicated by color.

A comparison of the ILD in the area of improved distance perception using the DVF-generated stimulus with that outside this region shows that the ILDs are greatest in the area between 60° and 120° and within a distance of about 40 cm. Outside of this region, the ILDs are smaller. Hence, it would appear that the improvement in distance localization performance is related to the ILD introduced by the DVF. This ILD should approximate the true ILD of the near-field HRTF since the size of the rigid sphere used to calculate the DVF is similar in size to the subject’s head.

In the case where intensity was removed as a cue for distance in the constant-intensity DVF-generated stimulus condition, it was observed that there was some distance discrimination in the side region and within the 10–20 cm distance range, though performance was poor. It would appear...
that the ILD introduced by the DVF provides some cue for absolute distance discrimination but is not strongly salient. Hence, in using the DVF method to synthesize near-field HRIRs for generating VAS, the combination of intensity cues and ILD cues seem to be important for better distance perception.

Although there is an improvement in distance perception in the DVF-generated stimulus condition compared to the intensity-scaled stimulus condition, there is still a high degree of overestimation of the distance and variability for all subjects. However, a recent summary of human distance perception research by Zahorik et al. (2005) shows that it is quite common for listener’s to underestimate distance of far-field sources and overestimate the distance of near-field sounds. They also show that there is often high variability in judgments of sound source distance. Hence, it is difficult to determine whether the overestimation and variability in distance perception are due to the generated VAS or due to the generally poor distance discrimination by humans.

It is also important to ensure that directional localization accuracy is not degraded by the synthesized near-field stimulus. In all conditions, there was a statistically significant difference in the mean difference between the target and response lateral and polar angles for different distance ranges. In particular, there was a statistically significant increase in the lateral and polar angle differences at the 10–20 cm distance range. However, calculation of the effect size index showed the effect of the differences in directional localization performance between the different distance ranges to be small for all conditions. It can be expected that directional localization errors would increase with decreasing sound source distance since a small error in the placement of the electromagnetic sensor to indicate the perceived location will translate into a larger angular difference at a close distance compared to a farther distance.

The advantage of the DVF technique is that it avoids the technical difficulties associated with measuring near-field HRIRs on human subjects which has hindered the development of high fidelity near-field VAS. There have been some other methods presented in the literature for generating near-field VAS without having to measure near-field HRIRs on human subjects. Martens and Yoshida (2000) presented a method for synthesizing near-field VAS using an approximation of the rigid sphere model. In their work a number of signal processing steps are applied to the stimulus to synthesize a sound source in the near-field. First, the ITD for the contralateral ear signal is adjusted for the intended near-field distance. Second, attenuation due to distance is applied to both the signals for both ears equally and an additional attenuation is applied to the contralateral ear signal to mimic the ILD in the near-field. Finally, a single-pole, single-zero filter that approximated the response of a rigid sphere is applied to model the frequency-dependent acoustical response of the head. Their method is based on the results of a simple psychoacoustic experiment that showed that the perceived distance of speech stimuli could be changed by manipulation of the ILD. However, the accuracy of the distance localization of the stimuli was not directly examined. It is also unclear whether directional localization accuracy would be affected by such adjustment of the ILD. The advantage of the DVF method over the method proposed by Martens and Yoshida (2000) is that the DVF is able to account for the ILD for sounds in the near-field, as well as low-frequency parallax effects of the HRTF. Since a human head and a rigid sphere of the same size have a similar low-frequency response, the DVF calculated from the rigid sphere model describes the change in the low frequencies of the HRTF as the sound source distance is changed from the far-field to the near-field. However, the DVF is unable to account for high-frequency parallax effects since pinnae are not modeled by the rigid sphere model. Although Möller et al. (1995) and Kim et al. (2001) suggested that it is possible to simulate a source in the near-field by choosing the appropriate far-field HRTF according to a simple parallax model, it is unclear that the high-frequency spectra of the HRTFs indeed do change in a manner that follows the simple parallax model they have presented and whether directional localization accuracy will be affected.

Brungart and Simpson (2001) conducted a psychoacoustic experiment to evaluate human sound localization performance of near-field VAS generated using near-field HRIRs measured on an acoustic manikin. They found that, in general, listener’s could discriminate sound source distance but directional localization performance was not as accurate as that of nearby sounds in the free-field. This is not surprising since Wenzel et al. (1993) had previously shown that virtual sound stimuli generated using HRIRs recorded on a subject’s own ears caused less directional localization errors than that generated using HRIRs measured on another person or a manikin. Brungart and Simpson (2001) investigation highlights the need for HRIRs measured on the subject for the generation of high fidelity near-field VAS. In our work, HR-
IRs measured on the individual subjects have been used to synthesize the near-field HRIRs and directional localization is preserved.

Duraiswami et al. (2004) presented a method for synthesizing near-field HRIRs from HRIRs measured in the far-field. By representing each measured HRTF as a series of multipole solutions to the Helmholtz wave equation and solving for the coefficients of the multipole solutions simultaneously, the HRTFs for any direction and distance from the head can theoretically be evaluated from the obtained coefficients. The major drawback of this technique is that close to 2000 far-field HRIRs need to be measured to accurately evaluate HRTFs up to 16 kHz. In our work, we have synthesized near-field HRIRs from far-field HRIRs measured on the subject using a rigid sphere model with the same radius as the head of the subject. Our method does not increase the number of far-field HRIRs that need to be measured to generate the near-field VAS.

V. CONCLUSION

We have presented a method for synthesizing near-field HRTFs using a rigid sphere model of the human head by calculating a DVF. The DVF approximates the ILD cues of the HRTF as a function of sound source distance. Using this method, the technical difficulties associated with measuring near-field HRTFs on human subjects are avoided. A human sound localization experiment was conducted to test the perceptual fidelity of the near-field VAS generated by this method as well as VAS generated by simply adjusting the intensity of the stimulus. Results show improved distance perception for sounds at simulated distances of up to 60 cm while maintaining directional accuracy, with the largest improvements for sound sources located to the side and within 40 cm. When intensity cues are removed from the DVF-generated stimulus, results still showed some discrimination of sound source distance but distance perception accuracy was poor.

ACKNOWLEDGMENTS

The DVF method is patent pending and thanks is given to Personal Audio Pty. Ltd. for permission to publish these results.