Close-range variation in binaural responses to orally-radiated sources

William L. Martens, Densil Cabrera, and Ken Stewart

Faculty of Architecture, Design and Planning, University of Sydney, Australia

ABSTRACT

Typical attempts to collect sets of head-related transfer functions (HRTFs) attempt to remove from the derived transfer functions the acoustical properties of the sound source used to make those measurements. This is done so that the directionally-dependent variations in the binaural response at the receiver's location can be independently characterised. For sound sources that are far from the receiver's location, this is appropriate and relatively straightforward to achieve. However, at close range (e.g., at distances between source and receiver position of less than 1 meter), characterising the variation in derived transfer functions that is dependent upon the acoustical characteristics of an orally-radiated source becomes potentially useful. The measurements reported here attempted to capture both source and receiver characteristics for a particular case, that in which the sound source is radiated from the mouth of an anthropomorphic manikin, and is received at the ear of a nearby manikin. Substantial range-dependent variation in the measured transfer functions was observed, clearly due to the presence of reflections between the surfaces of the source manikin and the receiver manikin's head. These results have implications for spoken telecommunication applications employing headphone-based virtual acoustic simulations.

INTRODUCTION

It is well established that measured head-related transfer functions (HRTFs) vary with source azimuth and elevation (Blauert, 1997), but less well studied is how HRTFs vary with range as well (see, e.g., Brungart & Rabinowitz, 1999). However, when measurements are made at close range (i.e., less than 1 meter from source to receiver position), substantial variations in the measured transfer functions can be observed under different measurement conditions, due to the difficulty in delivering sound from a loudspeaker without the loudspeaker's presence affecting the measurement in a range-dependent fashion. Theoretically, measurements would show range dependence even if it were possible to employ an ideal point source that would present no physical structure within the time window of analysis. Such range dependence has been studied analytically for the ideal spherical head, and the associated predicted responsess have agreed well with measurements under nearly ideal conditions (Duda & Martens, 1998). In additional to these range-dependent variations which occur even in the ideal case, there is an additional range dependent factor that is found in most practical HRTF measurements, which is a factor that is due to the physical structure of the test loudspeaker. In almost every measurement system, the loudspeaker presents a reflecting surface the acoustical effects of which are hard to remove from the desired HRTF to be derived form the raw test signal.

Rather than attempt to correct for the acoustical influence of the spatially extended transducer, the acoustical measurements reported here intentionally attempt to capture both source and receiver characteristics for a particular case of interest. The case under test is representative of an acoustical situation that is found in everyday life when a human listener is positioned within arm's reach of a human talker. Instead of using human subjects, however, the test case reported here used a pair of anthropomorphic manikins, so that many measurements could be taken at high spatial resolution over a wide range of receiver azimuth angles without fear of finding error variance that typically is introduced when using human subjects. So the orally-radiated source signal in this study was produced at the mouth of an anthropomorphic manikin, and received at the ear canal entrance of a second anthropomorphic manikin. The primary goal of this study is to observe variations that are clearly dependent upon the presence of reflections between the surfaces of the source manikin and the receiver manikin's head and torso.

In a previous study (Duda & Martens, 1998), an attempt was made to develop a better understanding of the close-range variation in the HRTF through a theoretical and experimental investigation of the response on the surface of an ideal rigid sphere. An algorithm was developed for computing the variation in sound pressure at the surface of the sphere as a function of direction and range to the sound source. Impulse responses were measured using a hard-surfaced sphere (in fact, a bowling ball) at a number of source ranges and many azimuth angles. The results may be summarized as follows: First, the experimental measurements were in close agreement with the theoretical solution. Second, the variation of low-frequency interaural level difference with range was found to be quite substantial for source ranges smaller than about five times the sphere radius. Third, the impulse response revealed the source of the ripples observed in the magnitude response, and provided direct evidence that the interaural time difference (ITD) is not a strong function of range. Finally, the transfer function for the ideal sphere appears to be minimum-phase, permitting exact recovery of the impulse response from the magnitude response in the frequency domain.

These prior results set the stage for the current investigation of a range dependent factor that was not included in the Duda & Martens (1998) study. Because the acoustics of real life situations include reflections between source and receiver that the prior study explicitly excluded, the current study addresses this issue specifically. In particular, the influence of these reflections on binaural responses measured for an orally-radiated source will be examined as a function of source range and receiver azimuth angle. Responses will be observed both in the time and frequency domains, so that the patterns of ripples in HRTF magnitude can be related to the reflection patterns.

METHOD

Measurements were made in an anechoic room. The receiving Head And Torso Simulator (HATS) was mounted on a turntable, with a pole supporting the HATS manikin at an ear height of 1.5 m above the acoustically transparent floor. This HATS model was Brüel & Kjær type 4100, which has microphones where the entrance of the ear canals would be in a human ear. For successive measurements, this receiving HATS was rotated in 2º increments, starting with the nose directly facing the sound source (0º), the configuration shown in Figure 1, and finishing with the nose facing away from the sound source (180º) – yielding 91 orientations. Note that the manikins were carefully positioned using a laser that is visible just over the left manikin's nose in Figure 1.

The receiving manikin was rotated clockwise (when viewed from above), meaning that the left ear was ipsilateral, and the right ear contralateral. It should be pointed out here that the employed two-manikin test situation requires a distinction to be made between azimuthal variation due to rotation of the receiver, and azimuthal variation due to rotation of the source manikin, the later potentially exhibiting directional dependence as a radiator relative to a fixed receiver location. Such directional dependence in oral radiation with source-manikin rotation was not investigated in the current study.

The sound source was a HATS (Brüel & Kjær type 4128C), which features a mouth simulator. A Brüel & Kjær type 4134 microphone was mounted at the source (The HATS has two possible microphone positions, and we used the position with the microphone right at the mouth). Calibration tones were recorded on all microphones so that the gains associated with transfer functions could be derived.

The sound source was positioned at four distances from the centre of the interaural axis of the receiving HATS when it was facing the source: 2 m, 0.71 m, 0.35 m and 0.25 m. Closer distances were not possible because the torso of the rotating HATS would have collided with the HATS sound source. Hence, in total, 91x4 binaural measurements were made. The distances mentioned here should be regarded as nominal distances, because in fact the distance increased as the HATS rotation angle increased (because the mounting position of the pole supporting the HATS was slightly behind the interaural axis).

The measurement was made using a logarithmic sinusoidal sweep (cf. Farina, 2000) with a frequency range extending between 50 Hz and 20 kHz, at a duration of 45 s and sampling rate of 44.1 kHz. Impulse responses were derived from the sweep recordings from each microphone. For measurements with the head-centre microphone, we synchronously averaged six impulse responses to further increase signal to noise ratio. For the measurements with HATS as receiver, we made a single sweep recording per angle per distance.

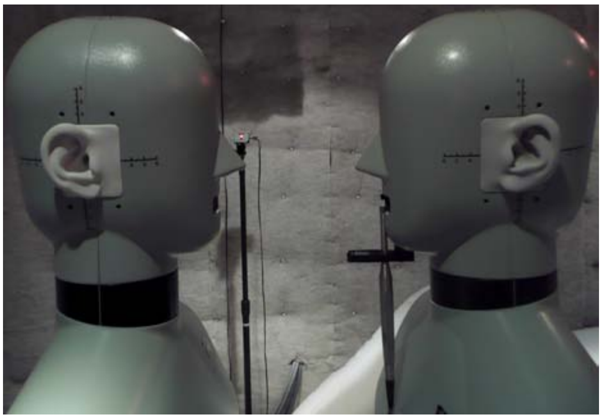

Figure 1. Photograph of the pair of anthropomorphic manikins that were employed to make measurements between the orally-radiated source (right manikin) and the ear canal entrance of the receiver (left manikin) as that receiver is rotated through a 180 deg. range of azimuth angles. Note the presence of a microphone at the mouth opening of the source manikin (on the right) that could be used to examine directly the oral-to-ear transfer function between manikins.

The data most simply derived from the measurements was impulse responses from the system output to the receiver HATS' ear microphones (by convolving the recorded signals with the inverse-filter of the test signal). Following this, head-related impulse responses (HRIRs) were derived from the transfer function between the head-centre microphone and the ear microphones. As reflective surfaces become close to high impedance (constant volume velocity) sound sources, the sound pressure may increase, and this would be seen in the HRIRs derived in this way. Since a real mouth is best modelled as constant volume velocity, this effect may be regarded as a beneficial contribution to the measurements. However, using the reference microphone on the sound source, it was also possible to derive HRIRs for a constant pressure (low impedance) sound source.

In deriving the impulse responses and transfer functions mentioned above, we truncated the signal in both frequency and time. In the frequency domain, very low and very high spectral components were suppressed (using a window function) to remove noise that was outside of the measurement range. In the time domain, similarly a window function was used to suppress sound outside the plausible time response of the system (e.g., from rogue reflections in the room). Examination of the results for such artefacts revealed some small unwanted reflections still remained within the analysis window; however, these were much lower in level than those due to the reflections of interest, those being the reflections between the head and torso of the two manikins. Time-domain display of these substantial reflections of interest, and their range-dependent and azimuth-dependent effects upon the obtained magnitude reponses, are presented in the following section.

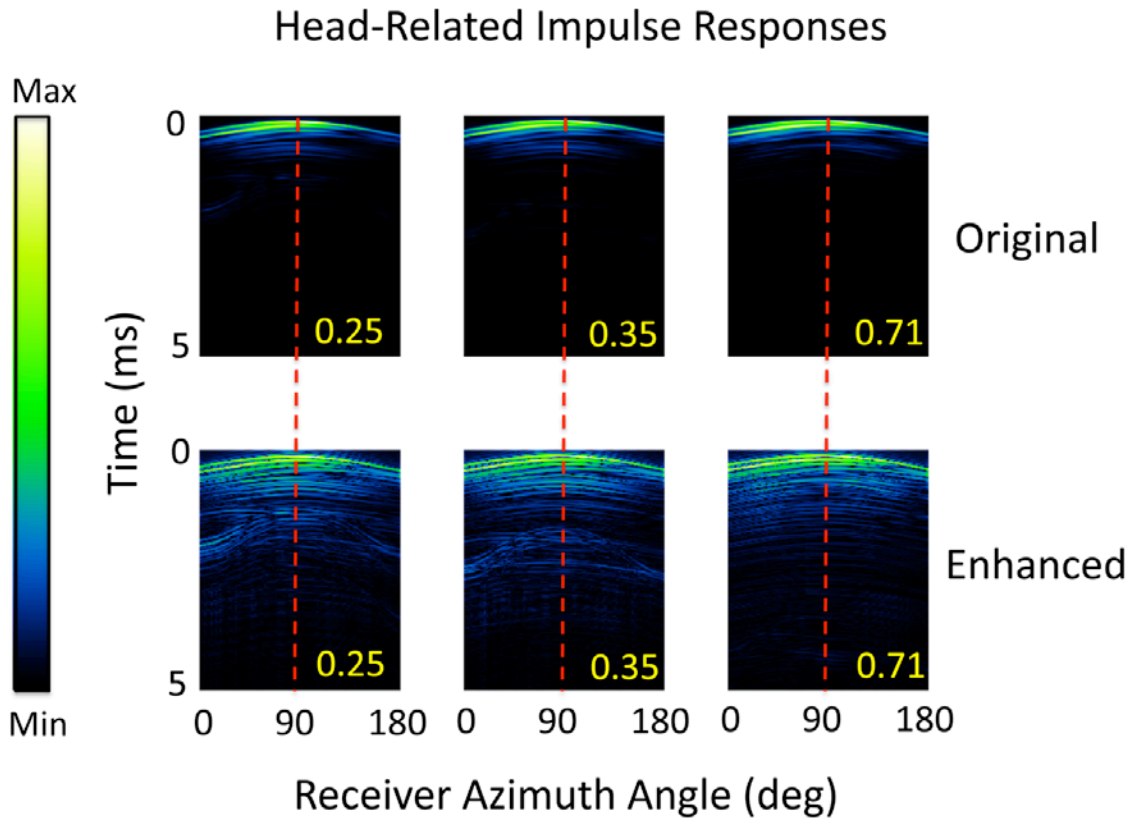

Figure 2. Two graphical perspectives on the HRIRs derived from the measurements made at three source ranges are displayed separately here as a function of azimuth angle (on the *x*-axis) and time (on the *y*-axis). The *y*-axis values range from 0 ms, corresponding to the arrival time of the source at 90 deg incidence, through the first 5 ms of the impulse response. The top row of three images shows from left to right the original amplitude of the envelope functions observed at source ranges of .25 m, .35 m, and .71 m, respectively (indicated by the yellow characters internal to each image). The row of three images on the bottom present an enhanced display of the low-level reflections that occur between 1 and 3 ms after arrival of the first wavefront (see text for a description of the enhancement method). The colorbar on the left shows the colormap that was used to display the amplitude of the envelope function for all six of the images displayed here, with colours ranging from the lowest brightness for minimum data values to the highest brightness for maximum amplitudes observed.

RESULTS

In order to enable visual inspection of the measured impulse responses, two graphical perspectives on the HRIRs are displayed in Figure 2, both using a 'geological' colour map indicating amplitude over a response surface with time and azimuth angle as independent parameters. The HRIRs are displayed over the first 5ms following the arrival of the first wavefront at each of the 91 azimuth angles at which measurements were taken. The top row of images shows the amplitude of the envelope functions for the HRIRs as observed at source ranges of .25 m, .35 m, and .71 m. These images are labelled as 'original' to distinguish this simple visualization from the 'enhanced' version of each displayed in the second row. The enhancement uses a technique commonly used in image processing to highlight edges in photographic images (using the 'unsharp mask' approach, which subtracts a Gaussian blur of the original image from itself, enabled by the MATLAB routine *fspecial*). It is the bottom row of images that shows best the pattern of the reflections of interest between the two manikins. In the leftmost plot of HRIRs in Figure 2 that were observed at a source range of .25 m the longest latency reflections that are clearly visible are those that occur between 1 and 3 ms after arrival of the first wavefront. Furthermore, the pattern of variation shows the expected modulation of a primary reflection attributed to the proximity of the two manikins, reaching its longest delay when the receiving manikin faces either directly towards or directly away from the source manikin. The pattern looks almost the same in the middle image visualizing the results observed at a source range of .35 m, except that the delays are longer and not quite as pronounced, as expected from the increased source range. At a source range of .71 m the reflection pattern is not so clearly visible, as is expected from the loss in level of the reflection relative to the direct sound. This detail should also make a clear distinction in the frequency domain as well: Whereas the higher level reflection found at the two smaller source ranges should result in a clear comb filtering effect in the magnitude response, at the larger source range, the ripples in the magnitude response due to the reflection should not show so clearly. This detail will be examined next in this section, but will not be examined for the whole 180-degree set of azimuth angles in the following. Note that the reflection pattern was more pronounced in the frontal region (between 0 and 90 degrees azimuth), and so the magnitude response data will be examined only for these data.

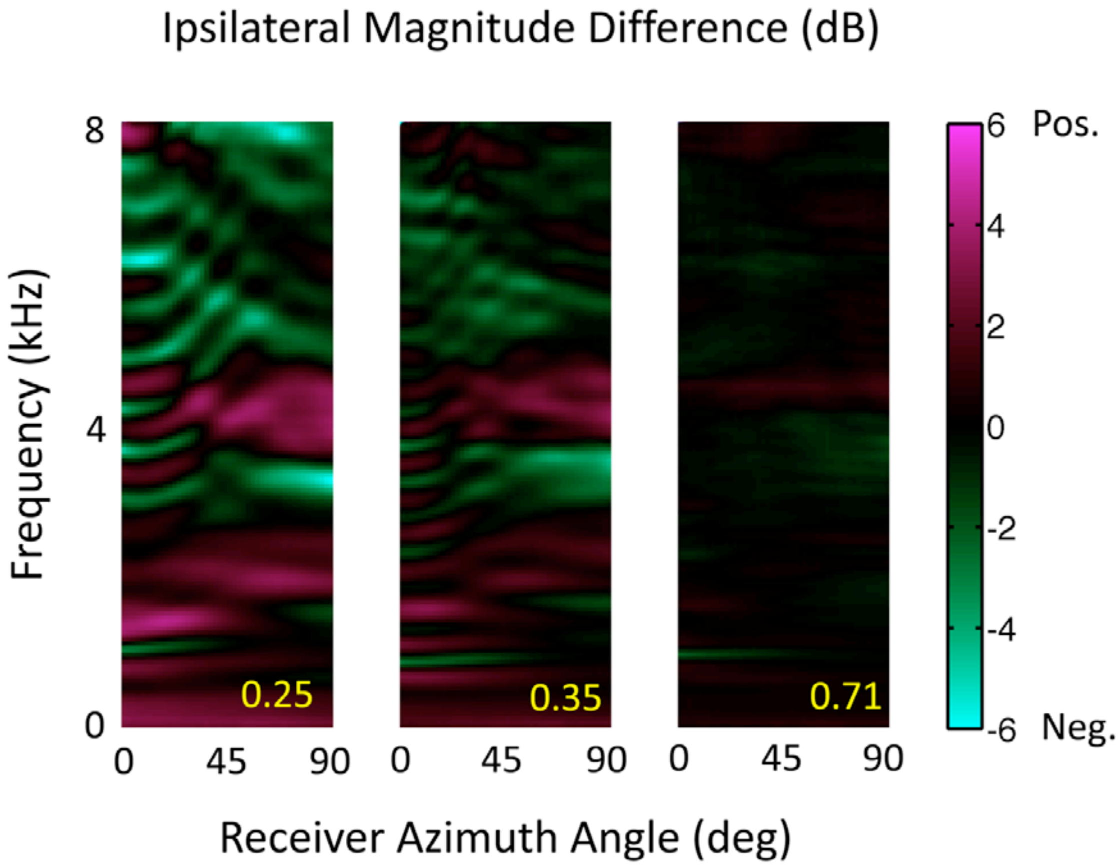

Figure 3. Images showing the difference in magnitude between ipsilateral HRTFs measured at close range relative to a reference set of HRTFs measured at a source range of 2 m. As in Figure 2, the results from the measurements at three ranges are displayed as a function of azimuth angle (on the *x*-axis), but only for incidence in the frontal hemifield. In contrast to Figure 2, the *y*-axis parameter of each image is frequency, with values ranging from 0 to 8 kHz. Again, the three images show results observed at source ranges of .25 m, .35 m, and .71 m, respectively (labelled in yellow characters). Since the displayed data is a magnitude differences relative to a reference measure, a different colormap is used in this figure, so that positive and negative deviations from the reference magnitude can be easily distinguished. Positive deviations (indicating responses greater than the reference) are always colour coded in shades of magenta, and negative deviations (responses below that of the reference) are always colour coded in shades of cyan. The colorbar on the right shows that data values near zero dB are colour coded as black.

In order to enable visual inspection of the range dependence of HRTF magnitude response, the magnitude response surface (in decibels) with azimuth and frequency as parameters associated with the reference measurements made at 2 m was subtracted from the HRTF magnitude responses data measured at the three smaller source ranges. Just as in Figure 2, but in the frequency domain instead of the time domain, the three images show results observed at source ranges of .25 m, .35 m, and .71 m, respectively. At the largest of the three range values (.71 m), the differences from the reference measurements are so small that very few details are visible. In effect, the response surface visualized in this rightmost image appears nearly black because the deviations are all near 0 dB (as indicated in the colorbar on the right of the figure). At the two smaller source range values, however, the deviations approach extremes of 6 dB, as the responses were modulated above and below the magnitude of the reference measurement in a manner of the ripple pattern associated with a comb filtering effect. Indeed, the regular pattern of peaks and troughs that is clearly visible for azimuth angles between 0 and 35 degrees is just what would be expected when a reflection with an amplitude just a bit less than the direct sound were to summed with it a relatively constant delay. However, as the receiver HATS was rotated so that the receiving ear faced the source more directly, up to around 75 degrees azimuth, the pattern of the ripple shifts toward the higher frequencies, as the reflection latency is reduced. When the azimuth angle approaches 90 degrees, however, the pattern is not so clear. The same type of behaviour is seen at the .35-m source range, although the peaks and troughs are more closely space. This is what is expected for the slightly longer reflection latencies in this case. The conclusion regarding the contribution of reflections of interest between the two manikins is that substantial modulation of the response is to be expected only when the source manikin is within arm's reach of the receiving manikin, since the modulation all but disappears as the source range increases through the .71 m case that was observed here. But the pattern of variation in responses measured at the ipsilateral tells only half the story; therefore, modulation in contralateral ear responses was also observed.

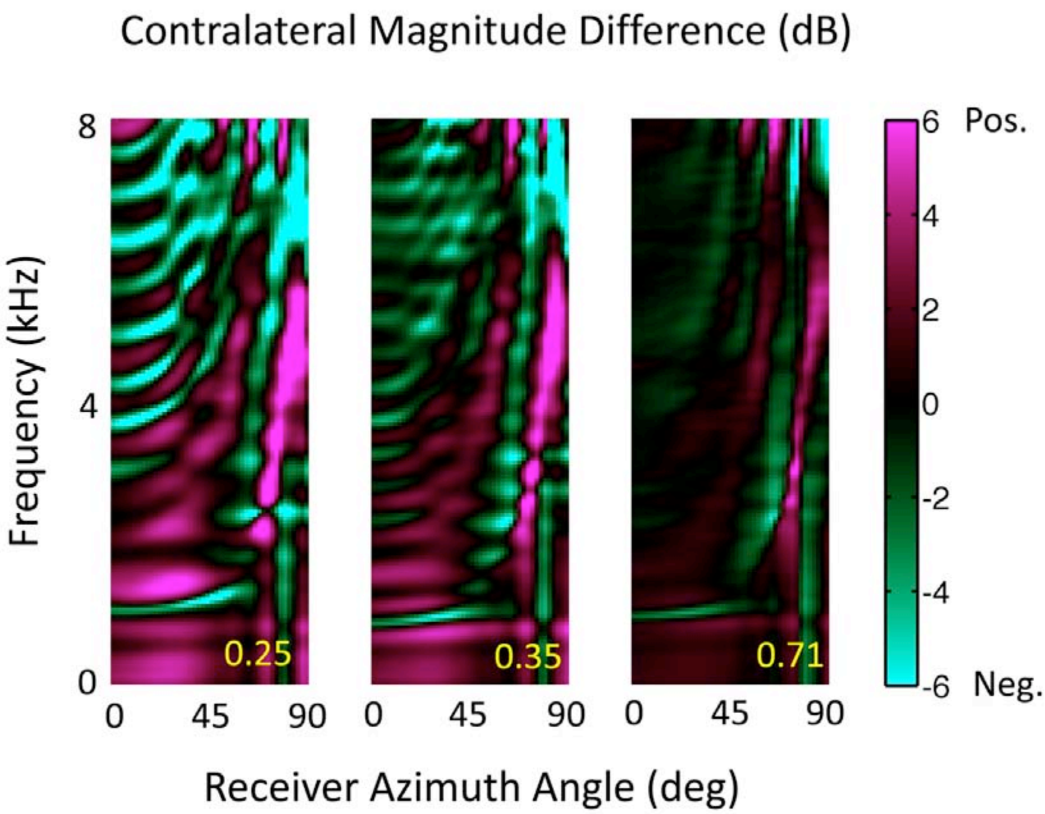

Figure 4. Images showing the difference in magnitude between contralateral HRTFs measured at close range relative to a reference set of HRTFs is displayed here just as was done for the ipsilateral HRTFs in Figure 3. Note that the images displayed here were generated using the same colormap as that used in Figure 3, so that the extreme positive and negative deviations from the reference magnitude reached at the contralateral ear could be directly compared with that of the ipsilateral ear.

For comparison of the range dependence of HRTF magnitude response deviations between those observed at the ipsilateral ear and those observed at the contralateral ear, the contralateral magnitude differences are displayed in Figure 4. Just as in Figure 3, the three images show results observed at source ranges of .25 m, .35 m, and .71 m, respectively. Again, at the largest of the three range values (.71 m), the differences from the reference measurements are very small at most azimuth angles, with the exception that increased modulation is observed as the receiver azimuth angle approaches 90 degrees. Although the response surface visualized in this rightmost image appears nearly black at smaller azimuth angles, more substantial modulation is observed between 70 and 90 degrees, swinging nearly 6 dB above and below the reference. The source of this modulation is the subject of further investigation, and no speculation as to the cause of this phenomenon will be presented here. At the two smaller source range values, however, the deviations in magnitude are quite similar to those in the ipsilateral ear, although the modulation appears to be a bit greater at the contralateral ear. This occurs most likely because of the more extreme contralateral attenuation of the direct sound that is observed at such close range. Indeed, quite substantial increases in the head shadowing effect are well known, even in the analytical solution in the case of the ideal sphere response, as explained by Duda and Martens (1998). In the case of the comb-filtering effects observed in the contralateral ear in the present study, it may be that the greater attenuation of direct sound, relative to that of the reflected sound, that brings these two acoustical components of the response closer in level, producing greater modulation than in the ipsilateral case, where the direct sound is relatively stronger than the reflected sound.

DISCUSSION

The directionally-dependent variation that was observed in the measured HRTFs for an orally-radiated source showed a clear range dependence just as expected from the results of related previous studies (Duda & Martens, 1998; Brungart & Rabinowitz, 1999; Qu et al., 2009). However, the range-dependent modulation of HRTF magnitude associated with the reflections between source and receiver has apparently received no attention as a matter of interest in its own regard, as was the perspective for the current study. Indeed, a relatively recent review of auditory distance perception in humans (Zahorik, et al., 2005), the summary of past and present research that was provided did not include any mention of the close-range reflections investigated in the current study. Whether the investigated phenomena are important in auditory distance perception or not should be established through subjective testing using human listeners. Suffice it to say that the substantial modulation in the HRTF magnitude that is observed at close range, when referenced to HRTFs measured at greater distance between source and receiver, suggests that the effects will be quite audible. More important than audibility, however, is the consideration of how effective the inclusion of such effects in a virtual acoustic simulation will be, as compared with the more straightforward extension of more conventional dry HRTF-based systems, such as that reviewed by Martens (2003). That review examined the development and evaluation of a binaural synthesis system providing close-range HRTF-based cues to source range, which system has been described also as a 'near-field virtual audio display' by Brungart (2002).

The reason this application of binaural technology is considered to be important is that control of perceived source range in auditory displays is a complicated matter that requires some more sophisticated treatment than the attention to level cues alone. The level-based cue is the only range-related parameter that typically is manipulated (i.e., that due to direct sound propagation attenuation). Of course there are many cues that can contribute to the modulation of the perceived range of a sound source. For example, Little, Mershon, & Cox (1992) have examined the role of spectral content as a cue to perceived auditory distance of the direct sound. Perhaps even more important, however, is the role of indirect sound resulting from room reflections in providing more graphic modulation of source range when more realistic spatial impressions are desired. In this case, it is crucial to consider the role of indirect sound that arrives soon after the direct sound when these room reflections are included in what the human listener hears (Bronkhorst & Houtgast, 1999). Indeed, Martens (2004) has shown that when simulated reflections based upon a small virtual acoustic environment are included in a headphone-based auditory display, much improved externalization of the virtual sources may be expected. Furthermore, as the indirect sound simulation is held constant for sources in the space quite nearby the listener, source range was shown to be manipulated in a manner that decouples direct sound level (i.e., loudness) from the range control made possible by including range-related HRTF variation in the sound processing. For example, the results reported by Martens (2004) indicated that the decrease in source range associated with a 9dB increase in interaural level difference (capturing the range-dependent variation of close-range HRTFs), could be counteracted by a 3dB decrease in direct sound level. One interesting implication of this finding is that source range and source loudness could be somewhat decoupled, at least for source quite nearby a listener's ear. What was not studied in that previous investigation of virtual source range control was the special situation in which a sound source is radiated from a talker's mouth, and received at the ear of a nearby listener.

Of particular interest here is the potential application of the more comprehensive virtual acoustic simulation suggested here for applications of binaural technology in the development of spoken telecommunication systems, such as that described by Kan, Pope, Jin, & van Schaik (2004). But what may have been less well appreciated in recent literature on this topic are the benefits that might be derived from the inclusion of sophisticated application of binaural technologies such as those proposed in the current paper. For example, when the voice of a talker at one end of a transmission is treated as a virtual acoustic source that is allowed to approach quite close to the receiver's position, the possibility for delivering messages in confidence that sound like a whisper in the listener's ear may be enabled (Martens & Yoshida, 2000). It might be said that the natural sound of a source would afford immediate awareness by the listener that a message was being delivered in confidence, due to the apparent source proximity rather than any explicit statement confirming the talker's intention to deliver such a confidential message. It will be interesting to examine in more depth whether the reflections between talker and receiver will provide useful information to the listener when virtual acoustic simulations include acoustical cues resembling those that were measured between two manikins in the current study.

CONCLUSION

The directionally-dependent variations in the binaural response were measured for an orally-radiated source at a number of nearby source ranges, relative to a more distant source. At the two closest ranges tested (.25 m and .35 m), substantial modulation was observed that resembled the comb-filtering effect associated with the summing of a direct sound with a single reflection arriving at short latency (between 1 m and 3 ms), and at a level near that of the direct sound. This modulation in HRTF magnitude response was even more pronounced at the contralateral ear. A slightly more distant source (at .71 m) showed much less modulation in HRTF magnitude, and the reflection pattern was so low in level at this longer latency that an enhanced visualization did not render it visible. These results have implications for applications of binaural technology, especially for spoken telecommunication systems employing headphone-based virtual acoustic simulations. Although no subjective tests using human listeners have been run to establish the relative importance of these modulations, it is proposed that such patterns may make a clearly audible difference when compared to virtual acoustic rendering solutions that include only dry HRTFs in their close-range simulation of orally-radiated sources.

REFERENCES

Blauert, J 1997, 'Spatial Hearing,' MIT Press, 1997.

Bronkhorst A & Houtgast T 1999, 'Auditory distance perception in rooms,' Nature, 397:517-520.

Brungart D 2002, 'Near-field virtual audio displays,' Presence: Teleoperators & Virtual Environments, 11(1):93-106.

Brungart D & Rabinowitz W 1999, 'Auditory localization of nearby sources: Head-related transfer functions,' The Journal of the Acoustical Society of America, 106(3):1465-1479.

Duda RO & Martens WL 1998, 'Range dependence of the response of a spherical head model,' The Journal of the Acoustical Society of America, 104(5):3048-3058.

Farina A 2000, 'Simultaneous measurement of impulse response and distortion with a swept-sine technique,' Proceedings of the 108th AES Convention, Paris, France.

Kan A, Pope G, Jin C, & van Schaik A 2004, 'Mobile spatial audio communication system,' In Proceedings of ICAD 04-Tenth Meeting of the International conference on Auditory Display, Sydney, Australia, 2004.

Little AD, Mershon DH, & Cox PH 1992, 'Spectral content as a cue to perceived auditory distance,' Perception, 21(3):405-416.

Martens, WL 2003, 'Perceptual evaluation of filters controlling source direction: Customized and generalized HRTFs for binaural synthesis,' Acoustical Science and Technology, 24 (5), 220-232.

Martens WL 2004, 'Decoupled loudness and range control for a source located within a small virtual acoustic environment,' Proceedings of ICAD 04-Tenth Meeting of the International conference on Auditory Display, Sydney, Australia, 2004.

Martens WL & Yoshida A 2000, 'Psychoacoustically-based control of auditory range: Display of virtual sound sources in the listener's personal space,' Proceedings of the International Conference on Information Society in the 21st Century: Emerging Technologies and New Challenges (IS2000), Aizu-Wakamatsu, Japan.

Qu TS, Xiao Z, Gong M, Huang Y, Li XD, & Wu XH 2009, 'Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap,' IEEE Transactions On Audio Speech And Language Processing, 17(6):1124-1132.

Thurlow, W, Mangels, JW, & Runge PS 1967, 'Head movements during sound localization,' J. Acoust. Soc. Am., 42, 489-493.

Zahorik P, Brungart, D & Bronkhorst A 2005, 'Auditory distance perception in humans: a summary of past and present research,' Acta Acustica United with Acustica, 91:409-420.