EFFECTS OF SPECTRAL MANIPULATION ON NON-INDIVIDUALIZED HEAD-RELATED TRANSFER FUNCTIONS (HRTFs)

R.H.Y. So[1*], N.M. Leung[1], A. Horner[2], J. Braasch[3] and K.L. Leung[4]

[1]Department of Industrial Engineering and Logistics Management,
Hong Kong University of Science and Technology, Hong Kong, PRC

[2]Department of Computer Science and Engineering,
Hong Kong University of Science and Technology, Hong Kong, PRC

[3]Institut für Kommunikationsakustik, Ruhr-Universität Bochum
Universitätsstraβe 150 D-4630 Bochum, Germany

[4]Department of Health Technology and Informatics, Hong Kong Polytechnic University,
Hong Kong, PRC

***** Corresponding author: Dr. Richard So, email: rhyso@ust.hk

ABSTRACT

Background: Directional sounds simulated using non-individualized head-related transfer functions (HRTFs) often result in front-back confusion. **Objective:** This study was designed to examine how manipulating these non-individualized HRTF spectra can reduce front-back confusion in headphone-simulated directional sounds. **Methods:** HRTFs of six ear-level directions were studied (angles of 0°, 45°, 135°, 180°, 225°, and 315°). The HRTF gains in each of six frequency bands (200-690 Hz, 690-2,400 Hz, 2,400-6,500 Hz, 6,500-10,000 Hz, 10,000-14,000 Hz, and 14,000-22,000 Hz) were amplified or attenuated by 0, 12, or 18 dB. Each manipulated HRTF generated a directional sound stimulus. Thirty-two participants were invited to localize the randomly-ordered stimuli. **Results:** The results indicate that a 12 or 18 dB manipulation of five of the six frequency bands produced significantly better directional accuracy, with significantly less front-back confusion. A reduction of up to 70% in localization error was obtained, along with 66% less front-back confusion. Significant interactions were found between the manipulation level and frequency. **Conclusion:** A 12 dB spectral manipulation of selected HRTF frequency bands produces better directional accuracy. **Application:** The results of this research could be applied to the development of tunable non-individualized HRTFs for audio products.

Keywords: spectral cues, HRTFs, virtual surround sound, sound localization, binaural directional cues

* All correspondence to Dr. Richard H.Y. So (Email: rhyso@ust.hk)

1. INTRODUCTION

1.1 Background

Dolby Digital (DD) 5.1 and 7.1 have become the standard surround sound format for most audio CDs and movie DVDs. In the case of DD 5.1, carefully recorded sounds are played through three front and two rear speakers to create a surround sound effect. Such a setup requires physical space and is not portable. To save space, directional sounds coming from the five speakers may be simulated using a binaural signal that employs HRTFs and played through a pair of headphones. This is referred to as virtual surround sound technology (Begault, 1994; So *et al.*, 2006). Since individually measured HRTFs can cost in excess of US$2000 to obtain, free non-individual HRTFs have been commonly used in virtual surround sound systems (So *et al.*, 2006). However, studies have shown that non-individual HRTFs cannot accurately simulate directional sounds. In particular, the simulated directional sounds suffer from frequent front-back confusion (Begault and Wenzel, 1993; Wenzel *et al.,* 1993). Unlike listening to sounds in real environments where head movements can help reduce front-back confusion (Blauert, 1997), the binaural simulation of directional sounds is not affected by head movements. Consequently, listeners can only rely on acoustic spectral cues resulting from the shape of their pinnae in order to discern whether a sound is coming from the front or the back.

This study aimed to reduce the front-back confusion in simulated directional sounds by enhancing the spectral cues in non-individualized HRTFs. The problem of front-back confusion has hampered the wide-spread use of non-individualized HRTFs in applications such as virtual reality (Begault, 1994) and surround sound systems (So *et al.*, 2006). Front-back confusion in virtual reality (VR) systems has been so routinely accepted that experimenters conducting VR studies with 3D audio cues simply informed participants of whether sounds were coming from the front or the back (Nguyen *et al.*, 2009). It was hoped in this study that by reducing front-back confusion in binaural sounds generated using non-individualized HRTFs, accurate and low-cost directional sounds can be produced.

1.2 HRTFs and front-back confusion

An HRTF is a type of frequency response which describes the changes in magnitude and phase a sound undergoes when it travels from a source to a listener's ear in an anechoic environment (Blauert, 1997). The angle of the sound relative to the listener's ear is called its incident angle, where 0° corresponds to a sound coming from directly in front of the listener, and 180° corresponds to a sound coming from directly behind (see Figure 1). For a

particular angle, a pair of HRTFs, one for each ear, is measured. The HRTF pair can be used to filter a monophonic sound in order to simulate a binaural sound coming from a particular direction. Wightman and Kistler (1989) and Kulkarni and Colburn (1998) reported that individualized HRTFs can accurately simulate binaural directional sounds. However this is not always the case. For example, Asano *et al.* (1990) reported that a listener was excluded from their experiment because the subject could not accurately localize simulated directional sounds generated using the listener's own individualized HRTFs.

In general, listeners can accurately discern whether a simulated sound is coming from the left or the right, but perhaps not if it is coming from the front or the back. Such errors are referred to as front-back (or back-front) confusion, and are most serious when directional sounds are simulated using non-individualized HRTFs (Begault and Wenzel, 1993; Wenzel *et al.,* 1993). Studies have shown that the forward-backward perception of an HRTF-simulated directional sound depends on certain spectral cues (e.g., Begault, 1994; Blauert, 1969 and 1997; Bronkhorst, 1995; Hebrank and Wright, 1974; Myers, 1989; Musicant and Butler, 1983; So *et al.*, 2010; Tan and Gan, 1998). When HRTFs that lack these spectral cues are used to generate a binaural directional sound, front-back confusion may occur. Many studies have been made to identify HRTF spectral cues for forward-backward perception. Some of them are reviewed in the next section.

1.3 Spectral cues for forward-backward sound perception

Acoustically, the pinnae of our ears amplify and attenuate sound waves depending on their incoming angles (e.g., Blauert, 1997; Lopez-Poveda and Meddis, 1996; Shaw and Terabishi, 1968). Oldfield and Parker (1984) reported that when both pinnae were covered with moulds in order to neutralize their acoustic effect, front-back confusion greatly increased for listeners who were asked to localize sounds.

This section reviews the HRTF spectral peaks and notches researchers have found to help listeners localize forward and backward directional sounds. Sounds are either amplified or attenuated depending on an HRTF's spectral peaks and notches. Consequently, our literature review includes studies that have examined how sound spectra influence their perceived direction.

Blauert (1969) was among the first to establish the relationship between sound spectra and directional perception. He used loudspeakers placed at the front and rear to study how a change in sound spectra would affect a listener's forward-backward perception. Noise bands of 1/3 octave and with center frequencies of 125, 160, 200, 250, 315, 400, 500, 650, 800,

1000, 1250, 1600, 2000, 2500, 3150, 4000, 5000, 6300, 8000, 10k, 12.5k, and 16k Hz were presented from the front and rear at random. Listeners were asked to determine whether the sound was coming from the front or the rear. The results indicated that sounds with energy in the frequency bands of 280-560 Hz and 2,900-5,800 Hz tend to be perceived as coming more from the front, while sounds with energy in the frequency bands of 720-1,800 Hz and 10,300-14,900 Hz tend to be perceived as coming more from the back. A comparison between the 1/3 octave bands used by Blauert (1969) and critical bands can be found in Ballou (2002). The range of center frequencies used by Blauert (1969) covered the entire range of critical bands.

Myers (1989) filed a US patent on boosting and attenuating particular frequency bands to improve localization performance for three-dimensional auditory display. Four frequency bands were defined: A (200-682 Hz), B (682-2,069 Hz), C (2,069-6,279 Hz), and D (6,279-22,050 Hz). The patent claimed that amplifying the energy in frequency bands A and C while simultaneously attenuating the energy in bands B and D of an incident sound makes the sound more likely to be perceived as coming from the front. When the energy of an incident sound in frequency bands B and D are amplified while simultaneously attenuate its energy in bands A and C, the sound is more likely to be perceived as coming from the back. The patent cited Bauer's work (1969) and did not report any new empirical data.

Tan and Gan (1998) followed up on the work of Myers (1989) and implemented a set of filters to reduce front-back confusion in directional sounds generated using non-individualized HRTFs. They divided the frequency range into five regions: A (225-680 Hz), B (680-2,000 Hz), C (2,000-6,300 Hz), D (6,300-10,900 Hz), and E (10,900-22,000 Hz). Bands A, B, and C were basically the same as Myer's, while band D was subdivided into two bands. To enhance forward perception, the energy in bands A, C, and E was amplified by 12 dB while the energy in bands B and D was simultaneously attenuated by 12 dB. The reverse was done to enhance backward perception. The filters were evaluated by ten listeners, and the number of listeners who reported front-back confusion was reduced from eight to four. Unfortunately, experimental details were not fully reported, and statistical analysis of localization error was not carried out.

While the forward and backward directional bands of the above three studies are in close agreement, this is not true for all studies on the subject. Hebrank and Wright (1974) conducted a study similar to the one by Blauert (1969) and reported that removing frequencies below 3,800 Hz did not affect forward nor backward perception. This contradicts Blauert's findings that sounds between 280-560 Hz tend to be perceived as coming more

from the front, and sounds between 720-1,800 Hz tend to be perceived as coming from the back. Like Hebrank and Wright (1974), Musicant and Butler (1983) also reported that frequencies below 1,000 Hz did not carry important cues for sound localization. Hebrank and Wright (1974) also reported that a spectral notch situated between 4,000-8,000 Hz contributed to forward perception, while others associated this notch with the perceived elevation angle (Blauert, 1969).

Moreover, Bronkhorst (1995) reported that sound energy at frequencies above 9,000 Hz should not affect the perceived direction, while both Blauert (1969) and Hebrank and Wright (1974) reported that sounds with energy concentrated at about 12,000 Hz were more likely to be perceived as coming from the back. In Bronkhorst's study, headphone-based binaural sounds were used while both Blauert (1969) and Hebrank and Wright (1974) used loudspeakers. This could account for the difference in findings as headphones bypass the pinnae, while sound broadcasts from loudspeakers is transformed by the pinnae.

1.4 Motivations and aims of this study

This study was designed to examine how manipulation of the HRTF spectra affects front-back confusion in headphone-simulated directional sounds. The only study to successfully use both headphones and frequency bands for front-back differentiation was that by Tan and Gan (1998). Unfortunately they did not report any statistical results. Instead, they simply indicated the number of participants who experienced front-back confusion. Furthermore, individual spectral cues were not manipulated. In our study, spectral cues were individually manipulated one by one, and their effects were studied for six different angles.

2. THE EXPERIMENT

2.1 Directional bands selected for this study

Based on the findings of Blauert (1969), Myers (1989), and Tan and Gan (1998), we divided the frequency range into six bands. Band 1: 170-680 Hz; band 2: 680-2,400 Hz; band 3: 2,400-6,300 Hz; band 4: 6,300-10,300 Hz; band 5: 10,300-14,900 Hz; and band 6: 14,900-22,000 Hz.

Band 1 is similar to band A of Myers (1989) and Tan and Gan (1998). Amplifying the energy in band 1 should make a sound more likely to be perceived as coming from the front (Blauert, 1969; Tan and Gan, 1998). Band 2 was adapted from band B of Myers (1989) and Tan and Gan (1998), with the upper limit extended to 2,400 Hz. We extended it to 2,400 Hz because this is the midpoint between 1,800 and 2,900 Hz, the upper and lower edges of Blauert's

directional bands. Attenuating the energy in band 2 should make a sound more likely to be perceived as coming from the front (Tan and Gan, 1998). The upper limit of band 3 was adopted from band C of Myers (1989). Amplifying energy in band 3 should make a sound more likely to be perceived as coming from the front (Blauert, 1969). Band 4 was adopted from band D of Tan and Gan (1998).

Based on the studies by Hebrank and Wright (1974) and Tan and Gan (1998), attenuating the energy in band 4 should make a sound more likely to be perceived as coming from the front. Band E of Tan and Gan (1998) was subdivided into bands 5 and 6 since Blauert (1969) reported a backward directional band between 10,300-14,900 Hz, while Hebrank and Wright (1974) reported that amplification at 12,000 Hz could cause a sound to be perceived as coming more from the back. Therefore attenuating the energy in band 5 should make a sound more likely to be perceived as coming from the front. All in all, we hypothesized that if the energy in bands 1, 3, or 6 is amplified and that in bands 2, 4, or 5 is attenuated, the resulting sound will more likely be perceived as coming from the front. On the other hand, if the energy in bands 1, 3, or 6 is attenuated and that in bands 2, 4, or 5 is amplified, the resulting sound will more likely be perceived as coming from the back. Specific hypotheses are given in the next section.

2.2 Hypotheses

Our experiment manipulated (amplified or attenuated) the six spectral cues (bands 1 to 6) as defined in the previous section. We tested three levels of manipulation (0, 12, 18 dB). Manipulating the spectra in bands 1 to 5 of HRTF-simulated directional sounds was hypothesized to reduce both localization error and front-back confusion (H1 to H5). These hypotheses were based on the findings of Blauert (1969). For spectral cues in band 6, manipulation was hypothesized to have no significant effect (H6). This hypothesis is consistent with the findings of Blauert (1969), but inconsistent with the findings of Tan and Gan (1998). We also hypothesized that as the level of manipulation increased from 12 to 18 dB, the directional effect would increase (H7). Note that in this study manipulation can mean either amplification or attenuation as explained in the previous section.

2.3 Method

The experiment had four independent variables: manipulation level (0, 12, 18 dB); frequency band to which the manipulation was applied (bands 1 to 6; see Section 2.1); sound direction (angles of 0°, 45°, 135°, 180°, 225°, and 315° at ear level, as shown in Figure 1); and repetition (two repeats). Their exhaustive combinations gave 108 unique conditions. The

dependent variables were sound localization error and the percentage of front-back confusion.

Thirty-two university students (12 males, 20 females) volunteered for the experiment. They were aged between 19 and 27. All of them passed an audiometric test at 20 dB, and declared no known hearing problem. Participants were compensated at HK$50 (about US$7) per hour for their time and travel expenses. The purpose of the experiment was not disclosed to them. To increase their motivation, participants were informed that the top 10% who scored the highest accuracy would each get an extra HK$100 as a cash prize. At the end of the experiment, all of them received the extra HK$100 as the university human subjects committee requested us not to link participants' performance with monetary reward. The experiment had a full-factorial within-subject design, and each participant was subjected to all 108 conditions twice—once on the first day (the first repetition), and again on the second day (the second repetition). Participants were given a five-minute rest after every 36 stimuli, and the order the stimuli were presented was randomized.

After listening to each stimulus, participants indicated the perceived sound direction relative to a circle representing the top-view of their heads. The top-view shown to participants on a computer screen was similar to the one in Figure 1 but without the text and numbers. The participants were asked to move the mouse cursor to indicate the perceived incoming direction of the sound. They were instructed to place the cursor inside the circle if they perceived the sound source to be inside their heads. Otherwise they should position the cursor outside the circle such that the relative angle between the cursor position and the center of the circle best indicated the perceived angle of the source. The interface was adapted from Braasch (2001) and So *et al.* (2006) and was developed using the Matlab package. At the beginning of the experiment, participants were informed that all sounds would be coming from the ear level. They received no feedback whatsoever on their performance throughout the experiment. The experiment was conducted inside an acoustic chamber (a 1400-A-CT chamber custom-built by the Industrial Acoustics Company) with a background noise level of about 32 dBA. Sound stimuli were presented to each participant using a pair of Sennheiser HD545 headphones. The average sound pressure level measured at one side of the headphone for the stimulus projected at 0^o with 0 dB manipulation was 68 dBA (i.e., $L_{Aeq,13s}$ = 68 dBA, measured using a Quest sound level meter—Model 1800). After each spectral manipulation, the resulting stimulus was scaled so that the overall sound pressure level remained at the same level as its corresponding stimulus with 0 dB manipulation. This controlled for the overall loudness of each stimulus.

---

insert Figure 1 about here

2.4 Spectral cue manipulation and stimuli preparation

The experiment used the MIT KEMAR non-individualized HRTFs (Gardner and Martin, 1995) as a starting point. Six pairs of HRTFs corresponding to six directions were used (angles of 0^o, 45^o, 135^o, 180^o, 225^o, and 315^o at ear level). Each of these six HRFT pairs was manipulated in 18 different ways producing 108 HRTFs. The 108 HRTFs were the exhaustive combinations of six directions, six manipulation frequency bands (bands 1 to 6) and three manipulation levels (0, 12, and 18 dB). For each combination of direction, frequency band, and manipulation level, the following steps were performed to manipulate the HRTF pair and produce the manipulated sound stimulus:

Step 1. The HRTF pair corresponding to the selected direction was extracted from the MIT KEMAR database.

Step 2. The HRTF of the left ear was manipulated first.

Step 3. If the stimulus was incident from the front and the manipulated frequency band was either band 1, 3, or 6, the peak with the largest amplitude within the selected band was identified. Otherwise, the spectral notch with the minimum amplitude was identified. Matlab code was used to identify these peaks and notches. Any change in slope with amplitude of less than 0.3 dB was ignored (Zwicker and Fastl, 1990).

Step 4. Depending on the level of manipulation, a ramp function of 0, 12, or 18 dB was applied to the identified peak (or subtracted from the identified notch). This ramp function peaked or notched at the identified point and had a symmetric lower and upper bandwidth. Its bandwidth was defined as the smaller of the distances between the center frequency of the ramp and the band's lower and upper bounds. Figure 2 shows an example with the original HRTF (solid line) and its manipulation (dashed line).

Step 5. The summed gain and original phase of the HRTF (both a function of frequency) were used to calculate a complex number representing the newly manipulated HRTF using the following formulae:

c(f) = Gain(f) * cosine [phase(f)]

d(f) = Gain(f) * sine [phase(f)]

where

c + jd is the complex number representing the manipulated HRTF at f Hz.

Gain(f) is the gain of the manipulated HRTF at f Hz.

phase(f) is the phase of the original HRTF at f Hz before manipulation.

Step 6. Steps 3 to 5 were repeated for the HRTF of the right ear.

Step 7. The manipulated HRTF pair was transformed into a corresponding pair of impulse responses. A 13-second-long monophonic sound was then convolved with the HRTF impulse response to produce the binaural directional sound corresponding to that particular combination of sound direction, frequency band, and manipulation level. The monophonic sound consisted of consecutive clips of a female voice saying in Cantonese: “請留意我聲音和以下的音樂從邊一個方向來” (“Please pay attention to the incident direction of my voice and the following music”), followed by a short musical excerpt from an Irish folk dance, and a male voice saying in Cantonese: “而家憑你的感覺指示出來” (“Now please indicate the perceived incident direction of the sound”). The short musical excerpt consisted of hammered dulcimer music. The sound of the hammered dulcimer is similar to that of the harp, but is produced with small felt-tipped hammers rather than by plucking. The sound file is available at: http://www.ielm.ust.hk/dfaculty/so/HF-stimulus.wav. This music was selected because it contains acoustic energy at around 16 kHz (see Figure 3). Music was used instead of simple sound pulses since it was judged to be more exciting in our pilot tests. The authors acknowledge that the study did not control for which part of the stimulus particular listeners concentrated on. Listeners were instructed to pay attention to the whole stimulus (including both the speech and music). As the experiment used a within-subject design, effects of individual preferences on speech or music should be averaged out.

---

insert Figures 2 and 3 about here

Following the above steps, the spectral cues of six HRTF pairs from the KEMAR database were systematically manipulated to produce 108 HRTF pairs. Manipulation can also affect

the overall gain, altering the inter-aural level difference (ILD). In order to preserve the ILD, the ILD with 0dB manipulation (i.e., no manipulation) was used as reference for each combination of sound direction and frequency band. Sound stimuli manipulated at 12 and 18 dB levels were scaled to the same root mean square magnitude as the reference. The original ITDs embedded in the HRTFs taken from the MIT KEMAR database were preserved with the exception of the stimuli for three listeners whose head-width deviated by 1.5 cm or more from that of the MIT manikin (their head-widths were 16.9, 17.1, and 17.2 cm). The procedure for correcting the ITDs of the stimuli for these three listeners is explained in Appendix A.

3. RESULTS AND DISCUSSION

3.1 The main effects

Localization error is the absolute difference between the simulated stimulus angle and the perceived incident angle. Front-back confusion occurs when sounds simulated to be coming from the front (i.e., with incident angles of 0°, 45°, and 315°) were perceived as coming from the back and vice versa. Because the error data did not follow a normal distribution (p<0.001, one-sample Kolmogorov-Smirnov test), non-parametric statistical tests were used. When the incident angles were either 0° or 180°, some participants perceived the sound source to be located inside their heads and were thus not able to indicate its incident angle. Such cases were treated as missing data. Out of 6,912 data points, 186 were treated as missing data involving seven of the 32 participants. In this study, all participants were subjected to 108 unique conditions in a within-subject design. Friedman and Wilcoxon tests on related samples were used to analyse the data. Because both tests require the number of samples to be the same across different conditions, data associated with the missing data were removed. Since only the stimuli projected at 0° or 180° resulted in missing data, data removal only affected tests involving those stimuli.

A Wilcoxon signed-rank test was conducted to compare data taken from the two repetitions. Partial data from the seven participants were excluded due to missing data. Results indicated that there was no significant difference between data collected in the two repetitions *(p>0.15)*. Figures 4 and 5 illustrate the median localization errors and the percentages of front-back confusion. The 108 conditions represented the exhaustive combinations of six sound directions, six manipulated frequency bands, and three levels of manipulation. Inspection of Figures 4 and 5 shows that for most conditions, a manipulation of 0 dB (i.e., no manipulation) resulted in the highest error and percentage of front-back

confusion, while a manipulation of 12 dB resulted in the lowest error and front-back confusion In particular, Figure 5 indicates that although front-back confusion was reduced by manipulation, the reduction in localization error was more substantial (Figure 4). For example, a 12 dB manipulation of the 180° stimuli in band 1 reduced one listener's localization error from 88° to 6° (a 93% reduction). For that particular listener, front-back confusion had not occurred either before or after the 93% reduction in error. In other words, front-back confusion was a less sensitive measure than localization error.

Insert Figures 4 and 5 about here

Friedman two-way ANOVAs reported significant main effects of manipulation level, stimuli direction, and manipulated frequency on localization error (p<0.001). However, only manipulation level had a significant main effect on front-back confusion (p<0.05). There are two possible reasons for this: (i) the number of samples used in the statistical test was 1/64 of that used for localization error since front-back confusion was calculated from data collected in 64 trials (32 listeners x 2 repeats), and (ii) front-back confusion was a less sensitive measure than localization error as explained in the previous paragraph. Examination of the main effects of manipulation indicated that both manipulations of 12 and 18 dB significantly reduced localization error and the percentage of front-back confusion (p<0.01, Wilcoxon signed-rank tests). However, no significant difference was found between manipulations of 12 and 18 dB, rejecting hypothesis H7.

Insert Table 1 about here

3.2 Interaction effects among manipulation level, manipulated frequency, and stimulus direction

Wilcoxon signed-rank tests were conducted to examine the effects of spectral manipulation for each combination of stimulus direction and manipulation frequency (see Table 1). Tests were conducted for localization error only because there were not enough degrees of freedom in the percentage of front-back confusion. The results indicated that a 12 dB manipulation of frequency bands 1, 2, 3, 5, or 6 produced significantly better directional

accuracy at angles of 45°, 135°, 180°, and 225° (p<0.05). In 19 out of 20 combinations, a 12 dB manipulation produced the lowest error. A 51% average reduction in median localization error was achieved. In particular, for an incident angle of 180°, the median localization error was reduced by about 70% for a 12 dB manipulation of bands 1, 2, and 3. These results support hypotheses H1, H2, H3, and H5, but reject hypotheses H4 and H6.

Among the six frequency bands, band 5 had the largest reduction in error while band 4 had the smallest. This is interesting since bands 4 and 5 were grouped into a single band in Myers (1989). We had hypothesized that manipulation of both bands 4 and 5 would significantly reduce localization error (i.e., hypotheses H4 and H5). However, our results support H5, but reject H4. This suggests that bands 4 and 5 should not be combined. A possible reason for the lack of reduction in localization error for band 4 may be found in the study by Blauert (1969), who reported that frequencies between 6,000-10,000 Hz (i.e., our band 4) were associated with perceived elevation rather than forward-backward perception. This may explain why manipulation of band 4 did not affect localization error substantially. Another possible reason may be the lack of acoustic energy at frequencies around band 4 in the stimulus (Figure 3). Future work is needed to confirm this.

For stimuli projected at an angle of 315°, only manipulation of bands 2 and 5 produced significantly better results *($p<0.001$)*. One possible reason for this is that listeners were able to accurately localize stimuli from this direction even before manipulation.

Although front-back confusion was a less sensitive measure than localization error, the average percentage of front-back confusion was reduced by 45-69% for sounds with incident angles of 45°, 135°, 180°, and 225° and whose energy in frequency bands 2 and 5 was manipulated by 12 dB.

For sounds with simulated angles of 0° (i.e., direct-front), amplification of the energy in bands 1, 3, or 6 by 12 dB as well as attenuation of the energy in band 5 by 12 dB improved localization accuracy. On average, a 71% reduction in localization error was obtained (see Table 1). As for other angles, manipulation of band 4 did not result in any significant improvement. Contrary to the findings of Myers (1989) and Tan and Gan (1998), attenuation of the energy in band 2 did not affect localization accuracy. While the improved accuracy resulting from the amplification of the energy in band 6 was consistent with the findings of Tan and Gan (1998), the opposite was found for band 5, where amplification caused the sound to be perceived as coming more from the back. This confirms the findings of Blauert

(1969) and Hebrank and Wright (1974) who reported that amplification of sounds between 10,300 and 14,900 Hz causes the sounds to be perceived as coming more from the back.

In summary, our results support hypotheses H1, H2, H3, and H5, but reject H4, H6, and H7.

3.3 General discussion and limitations

In this study, the effects of changing the magnitudes, but not the frequencies, of spectral cues were studied. Middlebrooks (1999a, b) studied the effects of shifting the center frequencies of spectral cues while keeping the magnitudes constant. They reported a way to systematically scale the frequency of spectral cues to minimize inter-subject variability in individual HRTFs (Middlebrooks *et al.*, 2000). Although the two approaches are very different, they are related in their manipulation of spectral cues. Future studies to integrate the two approaches may provide even more directional accuracy.

A sequential presentation of human speech and music was used as the audio stimulus in this study. This stimulus is close to the directional audio signals that are typically used in virtual reality applications (e.g., spatial audio presentations in virtual conference applications: Shen *et al.*, 2010; and virtual surround sound simulation: So *et al.*, 2006). Results of this study suggest that the percentage front-back confusion of binaural directional sound can be reduced by manipulating the spectra of the directional sound. Since changing the spectra of sound can also affect other perceived quality (Horner *et al.,* 2004, 2006), future studies should include measurements of perceived sound quality.

4. CONCLUSIONS AND FUTURE WORK

Directional sounds generated using HRTFs can be made more directionally accurate by amplifying or attenuating selected frequency bands. In particular, for ear-level directional sounds with a simulated angle of 45°, amplifying the frequency bands of 170-680 Hz, 2,400-6,300 Hz, or 14,900-22,000 Hz by 12dB, or attenuating the frequency bands of 680-2,400 Hz or 10,300-14,900 Hz by 12dB produces significantly better directional accuracy. Localization error was reduced by as much as 54%, and the percentage of front-back confusion was reduced by up to 45%.

For ear-level directional sounds with simulated angles of 135°, 180°, and 225°, amplifying the frequency bands of 680-2,400 Hz or 10,300-14,900 Hz by 12 dB, or attenuating the

frequency bands of 170-680 Hz, 2,400-6,300 Hz, or 14,900-22,000 Hz by 12 dB produces significantly better directional accuracy. For angles of 135°, 180°, and 225°, localization error was reduced by 51-75% while the percentage of front-back confusion was reduced by 50-69%.

A spectral manipulation of 12 dB resulted in better directional accuracy than a spectral manipulation of 18 dB, although there were no significant differences between the two levels of manipulation in most cases.

A primary contribution of this work is the systematic measurement of the interaction effects between the manipulation level of HRTFs, the manipulated frequency, and the stimuli direction. These interaction effects are useful in the development of spectrally tunable non-individual HRTFs. In particular, the specific spectral amplifications that have been proven to improve directional accuracy can be implemented as a set of add-on filters in applications such as virtual surround sound systems or as spatial audio sound servers in virtual reality applications. Data in this study suggests a reduction of front-back confusion by 50 to 69%.

5. ACKNOWLEDGEMENT

This work was supported by the Industrial Innovation Fund of the HKSAR Government through the Consumer Media Centre at the Hong Kong University of Science and Technology as well as the Research Grants Council of the HKSAR Government (HKUST6219/02E) and its counter-part in Germany (DAAD) through the Joint HK/Germany Research Grant.

Appendix A: Procedure for correcting the inter-aural time difference (ITD) for listeners with a head-width that deviated from that of the KEMAR manikin

For listeners with a head-width that deviated from that of the KEMAR manikin by 1.5 cm or more (the head-width of the KEMAR manikin was 15.2 cm according to Burkhard and Sachs, 1975), corrections to the ITD were made using simple geometry. For example, if a listener had a head-width of 17.2 cm, the additional ITD for a stimulus with an incident angle of 45° degrees was calculated as follows:

Assumption: Gardner and Martin (1995) reported that the distance between the virtual sound source and the head was 140 cm.

With a head-width of 15.2 cm, the distances between the source and the left and right ears were calculated using similar triangles (Figure A1) as 134.733 cm and 145.267 cm, respectively. With a head-width of 17.2 cm, the distances between the source and the left and right ears were 134.057 cm and 145.949 cm, respectively. Consequently:

$$\begin{aligned}\text{The additional ITD} &= [(145.949 - 134.057) - (145.267 - 134.733)]\ \text{cm} / 34{,}320\ \text{cm/s} \\ &= 1.358\ \text{cm} / 34320\ \text{cm/s} \\ &= 0.0396\ \text{ms}\end{aligned}$$

where 34,320 cm/s is the speed of sound in air.

The additional ITD was implemented in steps of 0.02 ms (0.02 ms was the sampling interval for the head-related impulse response since the sampling rate was 44.1 kHz). Adding a ‘zero’ to the front of a head-related impulse response (HRIR) increases its delay by 0.02 ms and removing a leading ‘zero’ advances the HRIR by 0.02 ms. In this case, one ‘zero’ was removed from the HRIR of the right ear, and one ‘zero’ was added to the HRIR of the left ear to facilitate the additional ITD of 0.04 ms.

Insert Figure A1 about here

---

6. REFERENCES

Begault, D. R. (1994) 3-D sound for virtual reality and multimedia. AP Professional, Cambridge MA.

Begault, D. R. and Wenzel, E. M. (1993) Headphone localization of speech. Human Factors, vol. 35, no. 2, pp.361-76.

Blauert, J. (1997) Spatial hearing: the psychophysics of human sound localization. MIT Press.

Blauert, J. (1969) Sound localization in the median plane. Acustica, vol. 22, no. 4, pp.205-13.

Braasch, J. (2001) Auditory localization and detection in multiple sound-source scenarios. PhD thesis, Ruhr University, Bochum.

Bronkhorst, A. W. (1995) Localization of real and virtual sound sources. Journal of the Acoustical Society of America, vol. 98, no. 5, pt. 1, pp.2542-53.

Gardner, W. G. and Martin, K. D. (1995) HRTF measurements of a KEMAR. Journal of the Acoustical Society of America, vol. 97, no. 6, pp.3907-8.

Hebrank, J. and Wright, D. (1974) Spectral cues used in the localization of sound sources on the median plane. Journal of the Acoustical Society of America, vol. 56, no. 6, pp.1829-34.

Horner, A., Beauchamp, J., and So, R. (2004) Detection of random alterations to time-varying musical instrument spectra. Journal of the Acoustical Society of America, 116(3), pp.1800-1810.

Horner, A., Beauchamp, J., and So, R.H.Y. (2006) A Search for Best Error Metrics to Predict Discrimination of Original and Spectrally Altered Musical Instruments Sounds. Journal of the Audio Engineering Society, 54(3), pp.140-156.

Kulkarni, A. and Colburn, H. S. (1998) Role of spectral detail in sound-source localization. Nature 396 (6713), 24-31, pp.747-749.

Middlebrooks, J.C. (1999) Individual differences in external-ear transfer functions reduced by scaling in frequency, J. Acoust. Soc. Am., vol. 106(3), pp. 1480-1492.

Middlebrooks, J.C. (1999) Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency, J. Acoust. Soc. Am., vol. 106(3), pp. 1493-1510.

Middlebrooks, J.C., Macpherson, E.A. and Onsan, Z.A. (2000) Psychophysical customization of directional transfer functions for virtual sound localization, J. Acoust. Soc. Am., vol. 108(6), pp. 3088-3091.

Musicant, A. D. and Butler, R. A. (1984) The influence of pinnae-based spectral cues on sound localization. Journal of the Acoustical Society of America, vol. 75, no. 4, pp.1195-200

Myers, P. H. (1989) Three-dimensional auditory display apparatus and method utilizing enhanced bionic emulation of human binaural sound localization. US Patent No. 4817149.

Oldfield, S.R. and Parker, S.P.A. (1984) Acuity of sound localization: a topography of auditory space. III. Pinnae cues absent. Perception, vol. 13, pp. 602-17.

Shen, L., Xu, Y.T. and Hang, B. (2010) Distributed spatial audio reconstruction for virtual conference. International Conference on Machine Vision and Human-machine Interface, 24-25 April, Kaifeng, PRC. pp.748-751.

So, R. H. Y., Leung, N.M., Braasch, J. and Leung, K.L. (2006) A low cost, Non-individualized surround sound system based upon head related transfer functions. An Ergonomics study and prototype development. Applied Ergonomics, 37, pp.695-707.

So, R.H.Y., Ngan, B., Horner, A., Leung, K.L., Braasch, J. and Blauert, J. (2010) Toward orthogonal non-individualized head-related transfer functions for forward and backward directional sound: cluster analysis and an experimental study. Ergonomics, 53(6), pp.767-781.

Tan, C. J. and Gan, W. S. (1998) User-defined spectral manipulation of HRTF for improved localization in 3D sound systems. Electronic Letters, vol. 34, no. 25, pp. 2387-89.

Wightman, F. L. and Kistler, D. J. (1989) Headphone simulation of free-field listening. II. Psychophysical validation. Journal of the Acoustical Society of America, vol. 85, no. 2, pp.868-78.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F.L. (1993) Localization using non-individualized head-related transfer functions. J. Acoust. Soc. Am., 94, pp.111-123.

Zwicker, E. and Fastl, H. (1990) Psychoacoustics, Facts and Models. Springer-Verlag Berlin Heidelberg, Munich, Germany.

TABLE 1 Results of Wilcoxon signed-rank tests conducted for sound localization error. The table compares differences between the manipulation level for each combination of stimulus direction and frequency band. Median errors (in brackets) of the three manipulation levels (0, 12, and 18 dB) are listed in descending order. Errors that are significantly different from each other are indicated by their levels of significance (**: p<0.001; *: p<0.01, +: p<0.05).

Stimulus Direction	Manipulated Frequency Band					
	Band 1 170-680Hz	Band 2 680-2,00Hz	Band 3 2,400-6,300Hz	Band 4 6,300-10,300Hz	Band 5 10,300-14,900Hz	Band 6 14,900-22,000Hz
0° (front)	0dB (96°) 18dB (36°) 12dB (29°)	0dB (86°) 18dB (39°) 12dB (33°)	0dB (90°) 18dB (46°) 12dB (27°)	0dB (92°) 18dB (71°) 12dB (48°)	0dB (83°) 12dB (19°) 18dB (5°)	0dB (85°) 18dB (57°) 12dB (27°)
45° (front)	0dB (43°) 18dB (32°) 12dB (21°)	0dB (47°) 18dB (29°) 12dB (26°)	0dB (44°) 18dB (33°) 12dB (28°)	18dB (43°) 0dB (43°) 12dB (41°)	0dB (46°) 18dB (22°) 12dB (21°)	0dB (40°) 18dB (24°) 12dB (24°)
135° (back)	0dB (43°) 18dB (24°) 12dB (23°)	0dB (52°) 18dB (17°) 12dB (13°)	0dB (44°) 18dB (24°) 12dB (19°)	0dB (51°) 18dB (45°) 12dB (45°)	0dB (53°) 18dB (30°) 12dB (30°)	0dB (49°) 18dB (44°) 12dB (42°)
180° (back)	0dB (103°) 18dB (33°) 12dB (30°)	0dB (109°) 12dB (36°) 18dB (33°)	0dB (112°) 18dB (46°) 12dB (32°)	18dB (122°) 0dB (122°) 12dB (111°)	0dB (124°) 18dB (65°) 12dB (58°)	0dB (109°) 18dB (91°) 12dB (31°)
225° (back)	0dB (47°) 18dB (26°) 12dB (23°)	0dB (43°) 18dB (22°) 12dB (22°)	0dB (51°) 18dB (30°) 12dB (28°)	18dB (47°) 0dB (47°) 12dB (44°)	0dB (45°) 18dB (37°) 12dB (24°)	0dB (49°) 18dB (37°) 12dB (31°)
315° (front)	0dB (35°) 18dB (35°) 12dB (30°)	0dB (42°) 12dB (25°) 18dB (18°)	0dB (36°) 18dB (33°) 12dB (27°)	0dB (34°) 18dB (30°) 12dB (30°)	0dB (37°) 18dB (23°) 12dB (18°)	0dB (34°) 18dB (32°) 12dB (29°)

FIGURE CAPTIONS

Figure 1 A top-view of the six stimuli directions (at ear-level with incident angles of 0°, 45°, 135°, 180°, 225°, and 315°). The actual figure that was shown on a computer screen to participants did not include the text and numbers as seen here. Participants were instructed to move the mouse cursor to indicate the perceived incoming direction of the sound.

Figure 2 Manipulation of peaks and notches for the direct-front (0°) HRTF (original HRTF denoted by a solid line; enhanced peaks and notches denoted by a dashed line).

Figure 3 The power spectral density of the stimulus before it is filtered with HRTFs (FFT length = 256).

Figure 4 Median localization error (with inter-quartile ranges) for the six sound directions and three manipulation levels in the six frequency bands (band 1: 170-680 Hz; band 2: 680-2,400 Hz; band 3: 2,400-6,300 Hz; band 4: 6,300-10,300 Hz; band 5: 10,300-14,900 Hz; and band 6: 14,900-22,000 Hz).

Figure 5 The percentage of front-back confusion for the six sound directions and three manipulation levels in the six frequency bands (band 1: 170-680 Hz; band 2: 680-2,400 Hz; band 3: 2,400-6,300 Hz; band 4: 6,300-10,300 Hz; band 5: 10,300-14,900 Hz; and band 6: 14,900-22,000 Hz).

Figure A1 Triangulation of the distances between the source and the left and right ears of the KEMAR manikin.

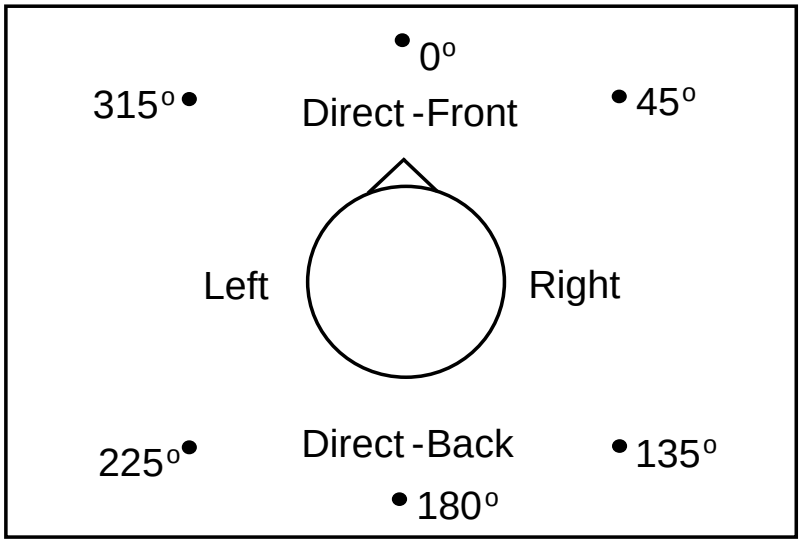

Figure 1

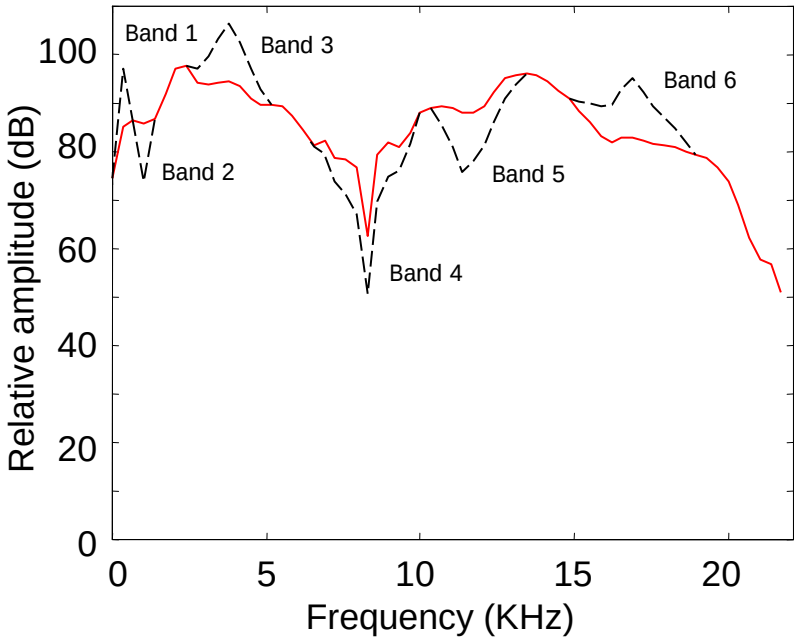
Band 1
Band 3
Band 6
Band 2
Band 5
Band 4
Relative amplitude (dB)
100
80
60
40
20
0
0
5
10
15
20
Frequency (KHz)

Figure 2

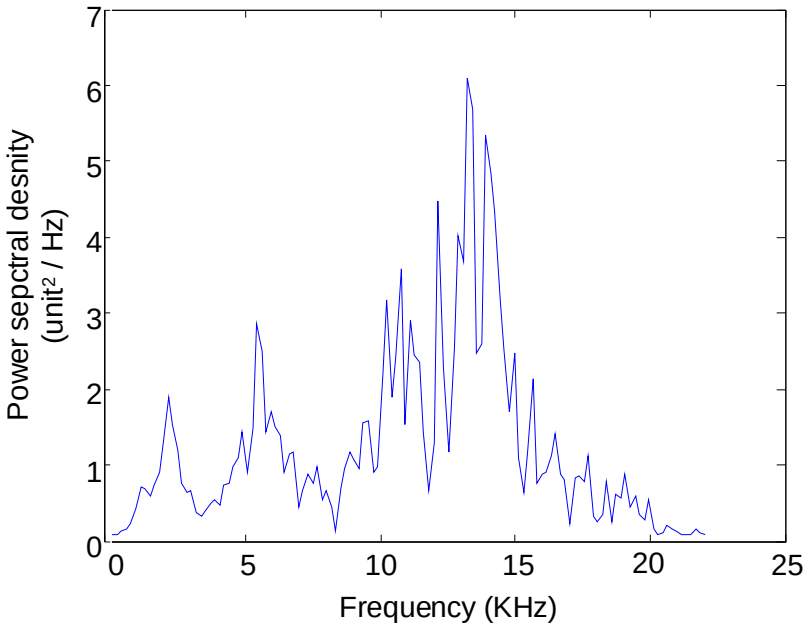
Power sepctral desnity
(unit² / Hz)
Frequency (KHz)
0
1
2
3
4
5
6
7
0
5
10
15
20
25

Figure 3

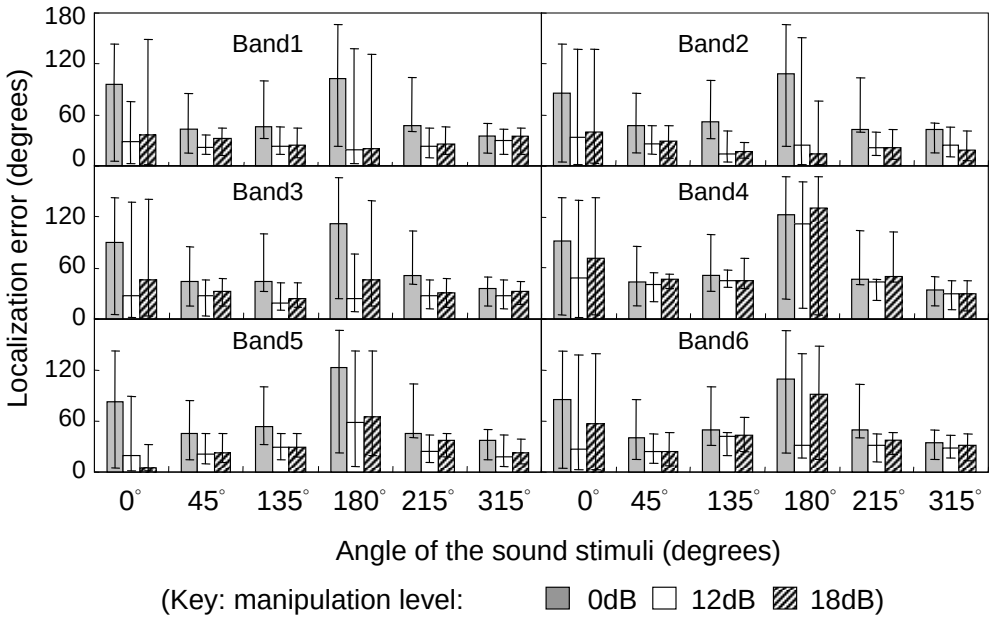
Band1
Band2
Band3
Band4
Band5
Band6
Localization error (degrees)
180
120
60
0
0°
45°
135°
180°
215°
315°
Angle of the sound stimuli (degrees)
(Key: manipulation level: 0dB 12dB 18dB)

Figure 4

]

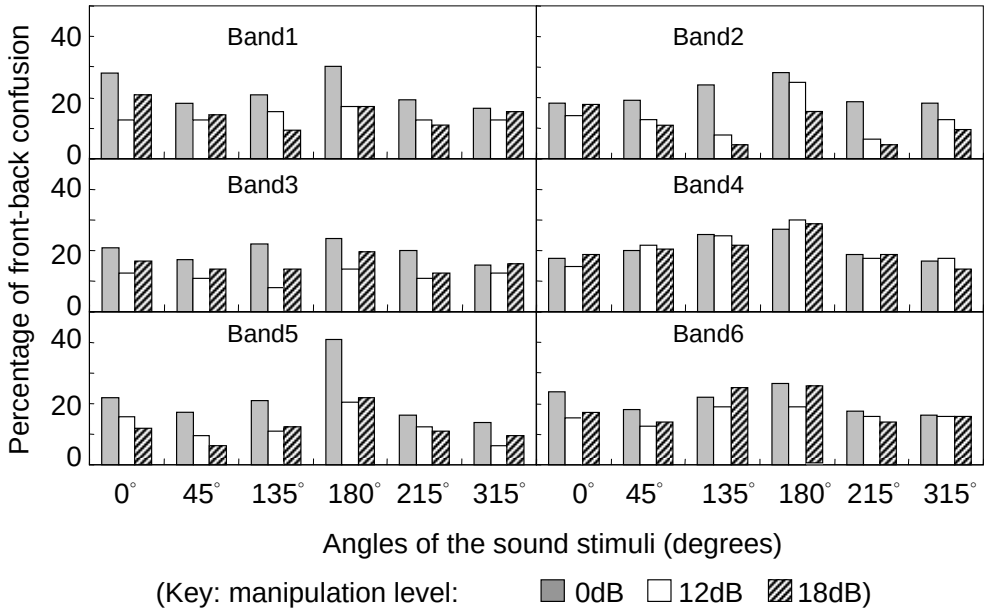
Band1
Band2
Band3
Band4
Band5
Band6
Percentage of front-back confusion
40
20
0
0° 45° 135° 180° 215° 315°
Angles of the sound stimuli (degrees)
(Key: manipulation level: 0dB 12dB 18dB)

Figure 5

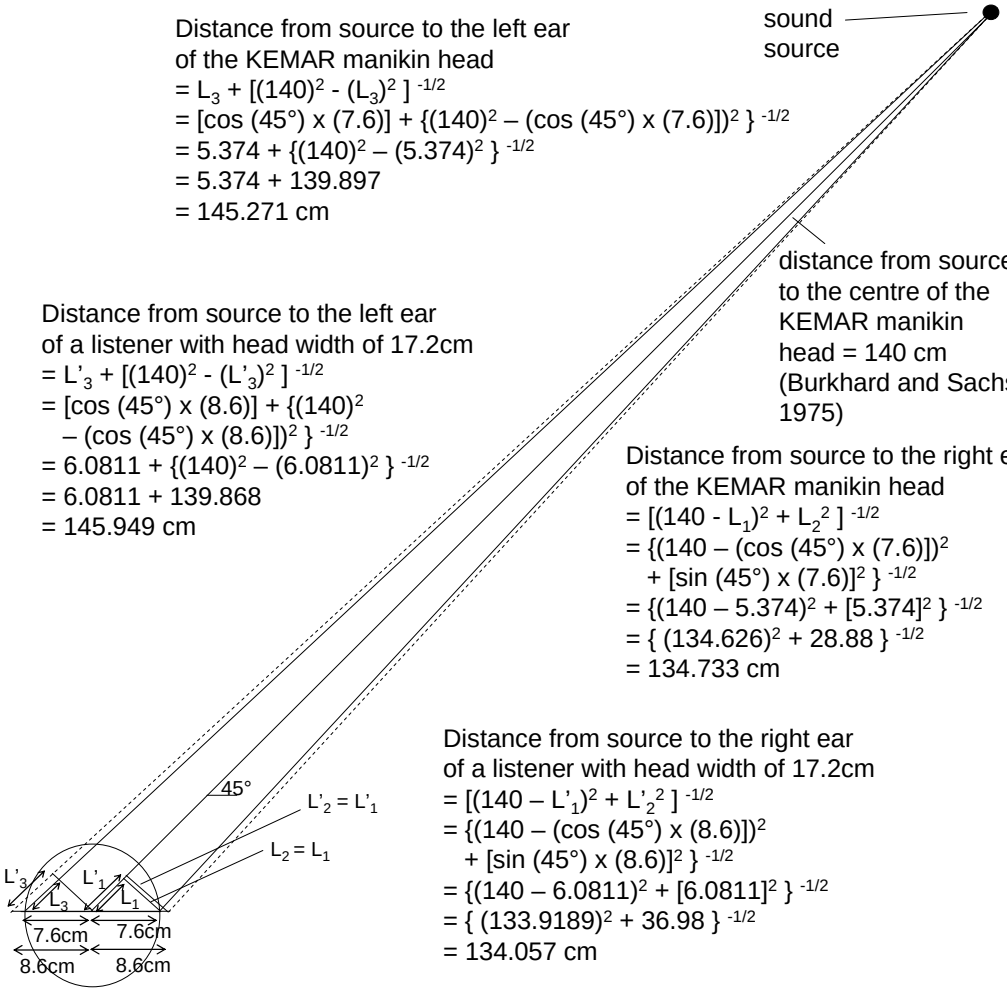
sound source
Distance from source to the left ear of the KEMAR manikin head
= $L_3 + [(140)^2 - (L_3)^2]^{-1/2}$
= $[\cos(45°) \times (7.6)] + \{(140)^2 - (\cos(45°) \times (7.6)])^2\}^{-1/2}$
= $5.374 + \{(140)^2 - (5.374)^2\}^{-1/2}$
= 5.374 + 139.897
= 145.271 cm
distance from source to the centre of the KEMAR manikin head = 140 cm (Burkhard and Sachs, 1975)
Distance from source to the left ear of a listener with head width of 17.2cm
= $L'_3 + [(140)^2 - (L'_3)^2]^{-1/2}$
= $[\cos(45°) \times (8.6)] + \{(140)^2 - (\cos(45°) \times (8.6)])^2\}^{-1/2}$
= $6.0811 + \{(140)^2 - (6.0811)^2\}^{-1/2}$
= 6.0811 + 139.868
= 145.949 cm
Distance from source to the right ear of the KEMAR manikin head
= $[(140 - L_1)^2 + L_2^2]^{-1/2}$
= $\{(140 - (\cos(45°) \times (7.6)])^2 + [\sin(45°) \times (7.6)]^2\}^{-1/2}$
= $\{(140 - 5.374)^2 + [5.374]^2\}^{-1/2}$
= $\{(134.626)^2 + 28.88\}^{-1/2}$
= 134.733 cm
Distance from source to the right ear of a listener with head width of 17.2cm
= $[(140 - L'_1)^2 + L'^2_2]^{-1/2}$
= $\{(140 - (\cos(45°) \times (8.6)])^2 + [\sin(45°) \times (8.6)]^2\}^{-1/2}$
= $\{(140 - 6.0811)^2 + [6.0811]^2\}^{-1/2}$
= $\{(133.9189)^2 + 36.98\}^{-1/2}$
= 134.057 cm
45°
$L'_2 = L'_1$
$L_2 = L_1$
L'_3
L'_1
L_3
L_1
7.6cm
7.6cm
8.6cm
8.6cm

Figure A1

Prof. Richard H.Y. So is an Associate Professor in Industrial Engineering. He obtained his PhD from the Institute of Sound and Vibration Research, University of Southampton. His research interests are in spatial vision and spatial audio.

Mr. N.M. Leung received his MPhil degree from the Hong Kong University of Science and Technology. His thesis was in spatial audio and his thesis supervisor was Prof. Richard So.

Prof. Andrew Horner is a Professor in Computer Science. He received his PhD in Computer Science from the University of Illinois at Champaign-Urbana. His research interests are in music synthesis and timbre.

Dr. J. Braasch received his PhD degree from the Institut für Kommunikationsakustik, Ruhr-Universität. His thesis was in binaural hearing and spatial audio.

Prof. K.L. Leung is an Associate Professor in Health Technology and Informatics. He received his PhD in Orthopedics and Traumatology from the Chinese University of Hong Kong. His research interests are in orthopedics and the physical modeling of ear shapes.