Ergonomics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713701117

Toward orthogonal non-individualised head-related transfer functions for forward and backward directional sound: cluster analysis and an experimental study

R. H. Y. So[a]; B. Ngan[a]; A. Horner[b]; J. Braasch[c]; J. Blauert[c]; K. L. Leung[d]
[a] Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Hong Kong, PRC [b] Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, PRC [c] Institut für Kommunikationsakustik, Ruhr-Universität Bochum, Bochum, Germany [d] Rehabilitation Engineering Center, Hong Kong Polytechnic University, Hong Kong, PRC

Online publication date: 21 May 2010

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

To link to this Article: DOI: 10.1080/00140131003675117

URL: http://dx.doi.org/10.1080/00140131003675117

Ergonomics
Vol. 53, No. 6, June 2010, 767–781

Taylor & Francis
Taylor & Francis Group

Toward orthogonal non-individualised head-related transfer functions for forward and backward directional sound: cluster analysis and an experimental study

R.H.Y. So[a]*, B. Ngan[a], A. Horner[b], J. Braasch[c], J. Blauert[c] and K.L. Leung[d]

[a]*Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Hong Kong, PRC;* [b]*Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, PRC;* [c]*Institut für Kommunikationsakustik, Ruhr-Universität Bochum, Bochum, Germany;* [d]*Rehabilitation Engineering Center, Hong Kong Polytechnic University, Hong Kong, PRC*

(*Received 6 May 2008; final version received 21 December 2009*)

Individualised head-related transfer functions (HRTFs) have been shown to accurately simulate forward and backward directional sounds. This study explores directional simulation for non-individualised HRTFs by determining orthogonal HRTFs for listeners to choose between. Using spectral features previously shown to aid forward–backward differentiation, 196 non-individualised HRTFs were clustered into six orthogonal groups and the centre HRTF of each group was selected as representative. An experiment with 15 listeners was conducted to evaluate the benefits of choosing between six centre-front and six centre-back directional sounds rather than the single front/back sounds produced by MIT-KEMAR HRTFs. Sound localisation error was significantly reduced by 22% and 65% of listeners reduced their front–back confusion rates. The significant reduction was maintained when the number of HRTFs was reduced from six to five. This represents a preliminary success in bridging the gap between individual and non-individual HRTFs for applications such as spatial surround sound systems.

Statement of Relevance: Due to different pinna shapes, directional sound stimuli generated by non-individualised HRTFs suffer from serious front–back confusion. The reported work demonstrates a way to reduce front–back confusion for centre-back sounds generated from non-individualised HRTFs.

Keywords: HRTFs; binaural cues; cluster analyses; spectral cues; sound localisation

1. Introduction

1.1. *Simulating directional sounds using individualised and non-individualised head-related transfer functions*

The direction of a sound is defined by the angle in which the sound is perceived as coming from the left or right, known as the azimuth direction. It is also defined by the angle in which the sound is perceived as coming from above or below, known as the elevation direction. By convention, 0° azimuth and 0° elevation corresponds to centre-front and 180° azimuth and 0° elevation corresponds to centre-back (see Figure 1).

Past studies have identified three key individualised binaural cues responsible for allowing listeners to accurately perceive the direction of a binaural sound (e.g. Begault 1994, Blauert 1997). These three cues are: (i) the average inter-aural time difference (ITD); (ii) the average inter-aural level difference (ILD); (iii) the spectral content of the sound (Blauert 1997). The ITD is the average difference in time for a sound to reach one ear before the other. The ILD is the average difference in sound level as measured at the two ears. Both the ITD and the ILD are key cues for determining whether a binaural sound comes from the left or right. The third cue, the spectral cue, determines whether a sound comes from the front or back, as well as whether it comes from above or below. Unlike the ITD and ILD, the spectral cue is effective for both monaural and binaural directional sounds. In other words, even in the absence of any inter-aural difference, a change in the spectral content of a sound can significantly influence whether that sound is perceived as coming from the front or the back (e.g. Blauert 1969/70), as well as whether it comes from above or below (e.g. Hebrank and Wright 1974). This spectral cue has also been called the 'pinna effect' because the shape of the pinna affects the spectral content of the sound arriving in the ear canal (e.g. Lopez-Poveda and Meddis 1996, Blauert 1997).

For a particular listener, the three cues (the ITD, ILD and spectral content) can be measured and stored as a pair of individually measured head-related transfer functions (HRTFs) (e.g. Begault 1994, Gardner and Martin 1995). The measurement of individualised HRTFs requires an anechoic chamber with specialised equipment and can cost up to US$1500 per subject (J. Blauert, personal communication 1999). An

*Corresponding author. Email: rhyso@ust.hk

ISSN 0014-0139 print/ISSN 1366-5847 online

DOI: 10.1080/00140131003675117
http://www.informaworld.com

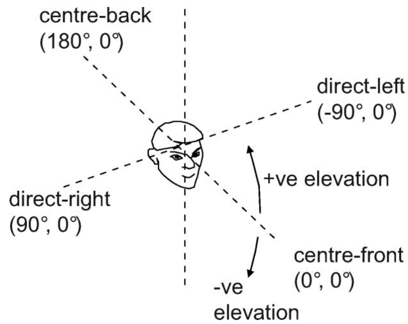

Figure 1. A diagram illustrating the centre-front and the centre-back incident angles for particular listeners.

alternative is to generate generic directional sounds using non-individualised HRTFs measured using a mannequin (e.g. the open-copyrighted and freely-downloadable HRTF data collected at the Massachusetts Institute of Technology (MIT) using the Knowles electronics manikin for acoustic research (KEMAR); Gardner and Martin 1995).

1.2. *Spectral cues and the aim of this study*

The folds of the pinnae alter the spectral content of the incoming sound according to its directional angle. Spectral cues of this sort are highly individualised, which explains why listeners report large front–back confusion rates when they localise directional sounds generated using non-individual HRTFs.

The aim of this study is to determine ways to reduce localisation error in non-individualised forward and backward directional sounds by providing some orthogonal choices for the listener to choose between. The individual spectral cues of 196 previously published non-individual HRTFs were extracted. The spectral features for forward directional sounds were then used to group the 196 HRTFs into six near-orthogonal groups. Another six groups were grouped according to the spectral features for backward directional sounds. An experiment was conducted to determine the minimum number of orthogonal choices required to reduce sound localisation error among a group of 15 listeners

1.3. *Front–back confusion using non-individualised head-related transfer functions and past efforts to tackle the problem*

Binaural directional sounds generated using non-individualised HRTFs have higher front–back confusion rates than those generated using individualised HRTFs (e.g. Wenzel *et al.* 1988, 1993). Front–back confusion occurs when a listener incorrectly perceives a sound coming from the front as coming from the back. The cause of front–back confusion has been attributed to mismatches between the spectral cues of non-individualised HRTFs and the spectral cues of individualised HRTFs for particular listeners (e.g. Blauert 1969/70, Gardner and Gardner 1973, Hebrank and Wright 1974, Asano *et al.* 1990, Langendijk and Bronkhorst 2002). Tan and Gan (1998) attempted to manipulate the spectral content of directional sounds to reduce front–back confusion. They used filtering instructions reported by Myers (1989) and successfully demonstrated a reduction in front–back confusion among 10 listeners. However, no statistical analysis was reported. In 2002, Zotkin and his colleagues attempted to match non-individualised HRTFs to particular individuals according to their pinnae dimensions. However, no significant improvement in sound localisation error was found when compared to the MIT-KEMAR HRTFs (Zotkin *et al.* 2002).

1.4. *A gap in the literature and specific objectives of this study*

The previous literature indicates that expensive individualised HRTFs can generate accurate directional sounds, while free non-individualised HRTFs can only produce inaccurate directional sounds with high rates of front–back confusion. By industrial engineering standards, individualised HRTFs represent a high-end, tailor-made solution with a high-quality performance, whereas non-individualised HRTFs represent a low-end, mass-produced solution with a low-quality performance. Previous efforts in bridging the performance gap between them have either failed (Zotkin *et al.* 2002) or not been verified by statistics (Tan and Gan 1998).

This study attempts to determine a set of near-orthogonal, non-individualised HRTFs for listeners to choose between in order to achieve lower localisation error than simply using the MIT-KEMAR HRTFs. In particular, this study (i) quantifies the spectral features among the 196 non-individualised HRTFs, (ii) clusters these spectral features and determines orthogonal HRTFs among the 196 HRTFs and (iii) examines the benefits of allowing listeners to choose from these orthogonal sets. In short, the logic of this study is: 'if one size does not fit, then determine a range of sizes and choose between them'.

2. Head-related transfer function spectral features for generating forward and backward directional sounds: a literature review

2.1. *Identifying spectral features from the literature*

The spectrum of a HRTF is associated with its ability to produce a forward or backward directional sound (e.g. Shaw and Teranishi 1958, Blauert 1969/70, Hebrank and Wright 1974, Myers 1989, Tan and Gan 1998, Langendijk and Bronkhorst 2002). Previous literature has basically identified six HRTF spectral features as producing sounds with forward characteristics and three HRTF spectral features with backward characteristics (see Figures 2 and 3 and Table 1). The authors would like to stress that these spectral features were not selected based upon opinions or preferences, but according to the findings of these previously published studies. Studies relevant to the spectral features are summarised in Table 1 and are reviewed below.

2.1.1. *Hebrank and Wright and features F1 and F2*

Hebrank and Wright (1974) conducted an experiment to investigate how the frequency content of a sound can affect forward, backward, above and below perception in the median plane. A total of 28 participants were asked to localise white noise that had been filtered by high-pass, low-pass, band-pass and

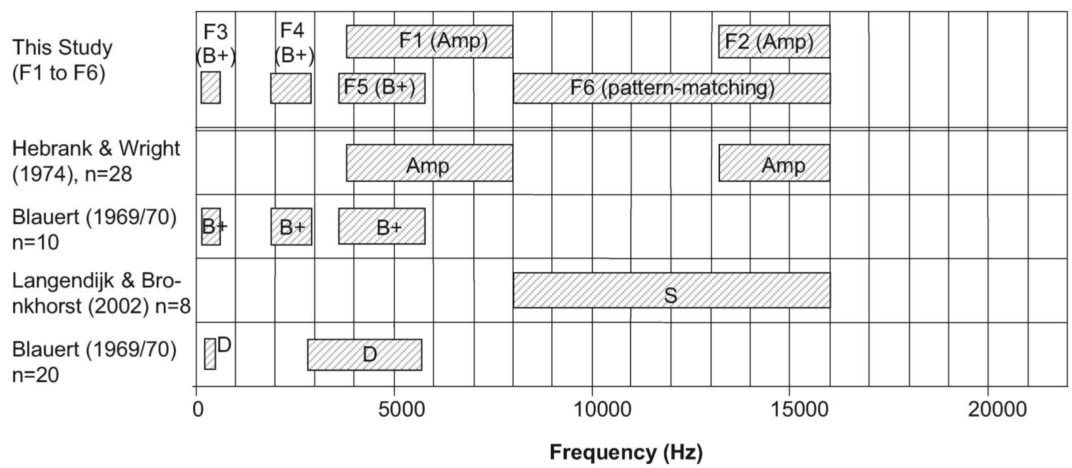

Figure 2. A summary of head-related transfer function (HRTF) spectral features that produce sounds likely to be perceived as coming from the front. Spectral features defined in this study (F1 to F6) are shown with relevant findings reported in the literature. If the literature includes empirical studies, the number of listeners used (*n*) is also shown. Amp = the frequency range in which the average HRTF magnitude is positively related to forward perception; B+ = the positive boosted band defined by Blauert (1969/70); Pattern-matching = the frequency range in which the HRTF modulus shape is related to forward perception; D = the directional band also defined by Blauert (1969/70); S = the frequency range in which spectral variations are related to forward perception as reported in Langendijk and Bronkhorst (2002); Att = the frequency range in which the average HRTF magnitude is negatively related to forward perception.

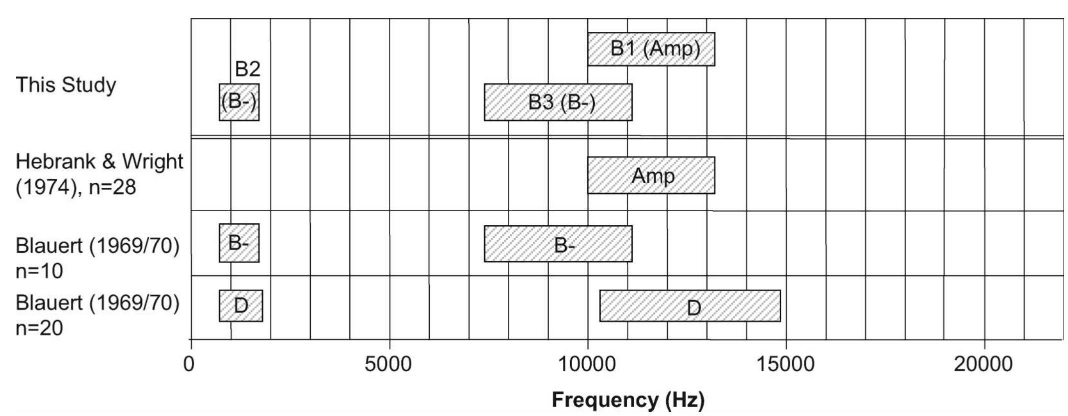

Figure 3. A summary of head-related transfer function (HRTF) spectral features that produce sounds likely to be perceived as coming from the back. Spectral features defined in this study (B1 to B3) are shown with relevant findings reported in the literature. If the literature includes empirical studies, the number of listeners used (*n*) is also shown. Amp = the frequency range in which the average HRTF magnitude is positively related to backward perception; B− = the negative boosted band defined by Blauert (1969/70); D = the directional band also defined by Blauert (1969/70).

Table 1. A summary of nine head-related transfer function (HRTF) spectral features for producing directional sounds with forward (F1 to F6) or backward (B1 to B3) directional characteristics.

Band	Frequency range	Nature of the band features	Direction of sound	Parameters attributed to features
F3	150–540 Hz	Boosted band (positive) (B+)	Forward	1. Difference between the average* HRTF level at azimuth 0° ear level and that of azimuth 180° ear level
F4	1900–2900 Hz	Boosted band (positive) (B+)	Forward	1. Difference between the average* HRTF level at azimuth 0° ear level and that of azimuth 180° ear level
F5	3600–5800 Hz	Boosted band (positive) (B+)	Forward	1. Difference between the average* HRTF level at azimuth 0° ear level and that of azimuth 180° ear level
F1	3800–8000 Hz	Amplification	Forward	1. Average level
F2	13200 –16,000 Hz	Amplification	Forward	1. Average level
F6	8000–16,000 Hz	Spectral variations	Forward	1. similarity matrix containing cosine similarity between each pair of HRTFs 2. Standard deviation of the modulus within the band
B2	720–1700 Hz	Boosted band (negative) (B–)	Backward	1. Difference between the average* HRTF level at azimuth 0° ear level and that of azimuth 180° ear level
B3	7400–11,100 Hz	Boosted band (negative) (B–)	Backward	1. Difference between the average* HRTF level at azimuth 0° ear level and that of azimuth 180° ear level
B1	10,000–13,000 Hz	Amplification	Backward	1. Average* level

*Average means the average HRTF magnitude within the band frequency range.
Note: The parameters extracted for each spectral feature are also listed.

band-stop filters with various cut-off frequencies. Their results showed that white noise containing more energy in the frequency range 3.9–9.0 KHz or 13.2–15.3 KHz is more likely to be perceived as coming from the front. Based on this finding, it is assumed that a HRTF with a higher relative gain at these frequencies will produce a directional sound with a higher relative chance of being perceived as coming from the front (F1 and F2 in Table 1).

2.1.2. *Blauert and features F3, F4, F5, B2, and B3*

Blauert (1969/70) was among the earlier studies characterising HRTF spectra into directional bands. A total of 20 participants listened to noise signals with one-third octave bandwidths and were asked to judge whether the sounds came from the front or back. The signals were presented once from the front and once from the back through loudspeakers. The data showed that noise signals with more energy in the frequency range 0.28–0.56 KHz or 2.9–5.8 KHz caused a bias towards forward-judgement while the opposite was true for noise signals in the frequency range 0.72–1.8 KHz or 10.3 KHz–14.9 KHz. Blauert referred to these as 'directional bands' and they are shown in Figures 2 and 3 (labelled 'D').

Blauert also compared the spectra arriving at the entrance to the ear canal when broadband noise was propagated from the front and back. Results showed that, even though the same sound source was played, the spectra were significantly different in certain frequency bands. In particular, Blauert found that if energy in certain frequency bands was amplified (i.e. boosted), the perceived direction of the sound could be manipulated to be more in the front or more at the back. Blauert named these frequency bands 'boosted bands' and related them to forward and backward perception of sound.

To identify boosted bands associated with forward perception, Blauert subtracted the centre-back HRTF spectrum from the centre-front HRTF spectrum. He found that if the average differences were positive in the frequency ranges 0.15–0.54 KHz, 1.9–2.9 KHz and 3.6–5.8 KHz, listeners would tend to perceive forward directional sound. These three frequency bands were referred to as the 'positive boosted bands' (labelled as 'B+' in Figure 2). These have been adopted as forward spectral features in this study (and are labelled F3, F4 and F5 in Table 1).

Blauert (1969/70) also reported that if the average differences were negative in the frequency ranges 0.7–1.7 KHz (B2) and 7.4–11.1 KHz (B3), listeners would tend to perceive backward directional sounds. These frequency bands were referred to as the 'negative boosted bands' (labelled as 'B–' in Figure 2). These have been adopted as backward spectral features in this study (and are labelled B2 and B3 in Table 1).

2.1.3. *Langendijk and Bronkhorst and feature F6*

While both Blauert (1969/70) and Hebrank and Wright (1974) investigated the effect of overall gain for particular frequency bands on forward directional perception, Langendijk and Bronkhorst (2002) studied the effect of HRTF spectral variations on forward directional perception. Eight individualised centre-front HRTFs were measured and the spectral variations of eight selected frequency bands were averaged one by one (the bands were 4.0–5.7 KHz, 4.0–8.0 kHz, 4.0–16.0 KHz, 5.7–8.0 KHz, 5.7–11.3 KHz, 8.0–11.3 kHz, 8.0–16.0 KHz and 11.3–16.0 KHz). A control condition with no averaging was added for a total of nine conditions, where listeners were presented with sound stimuli filtered using their own individualised HRTFs. The results showed that removing spectral variations from 4.0 to 16.0 kHz, 5.7 to 11.3 kHz and 8.0 to 16.0 kHz significantly increased front–back confusion. They concluded that spectral variations from 8.0 to 16.0 kHz were essential for producing accurate forward directional sounds. This feature has been adopted in the current study as F6 (it is shown in Figure 1, labelled 'S').

2.1.4. *Spectral feature B1*

Finally, the backward spectral feature was defined according to the results in Blauert (1969/70) and Hebrank and Wright (1974). Blauert (1969/70) reported that if a broadband noise had more energy between 10 and 14 KHz, more listeners would perceive that the noise was coming from the back. However, Hebrank and Wright (1974) reported that more energy above 13.2 kHz tends to cause a sound to be perceived as coming from the front. As a compromise, the frequency range for backward feature B1 was defined as 10–13 KHz in this study (see Figure 3 and Table 1).

2.2. Parameters used to quantify the spectral features

2.2.1. *Quantifying spectral features F1 and F2*

Hebrank and Wright (1974) suggested that if a HRTF had a higher relative gain at frequencies between 3.8 to 8 kHz (feature F1) and between 1.3 to 16 kHz (feature F2), the directional sound would tend to be perceived as coming from the front (see Table 1). To quantify the F1 and F2 features for each of the 196 HRTFs, the HRTF measured for the centre-front incident angle (0° azimuth and 0° elevation) was extracted. Then the average HRTF magnitudes of the F1 and F2 bands were measured. This produced 392 parameters (two for each of the 196 HRTFs) to be used for further analysis in this method.

2.2.2. *Quantifying spectral features F3, F4, F5, B2, and B3 based on Blauert*

As explained above, the three positive boosted bands reported by Blauert (1969/70) have been adopted as features F3 (0.15–0.54 KHz); F4 (1.9–2.9 KHz); and F5 (3.6–5.8 KHz) for this study. The two negative boosted bands have been adopted as features B2 (0.7–1.7 KHz) and B3 (7.4–11.1 KHz). For each of the 196 non-individualised HRTFs, the procedure reported in Blauert (1969/70) was followed by subtracting the centre-back HRTF spectrum from the centre-front HRTF spectrum. The average difference was calculated for each of the F3, F4, F5, B2 and B3 bands over all 196 HRTFs. This produced 980 parameters (five bands for each of the 196 HRTFs) to be used for further analysis.

2.2.3. *Quantifying spectral feature F6 based on Langendijk and Bronkhorst*

Langendijk and Bronkhorst (2002) reported that HRTF spectral variations between 8 to 16 kHz can affect its ability to produce accurate forward directional sounds. This has been adopted as spectral feature F6 in this study (see Table 1). In order to quantify the shape of the HRTF spectral variation between 8 and 16 KHz, amplitude variation as a function of frequency (between 8 and 16 KHz) was extracted as a vector. A cosine similarity was then calculated between each pair of vectors (Apostolico and Galil 1997), where a value of '-1' indicates that the two vectors vary in the opposite way and a value of '1' indicates that the two vectors vary in the same way. This produced a 196×196 spectral variation similarity matrix. In addition to this similarity matrix, the standard deviation of the spectral variations within the F6 frequency range was extracted from the 196 HRTFs at the centre-front incident angle. This produced 196 parameters for further analysis. The authors acknowledge that, besides cosine similarity, there are other ways to quantify similarity between spectral variation patterns of two HRTFs. Cosine similarity was chosen for two reasons: (1) its outputs are compatible with most clustering algorithms; (2) it has been widely used in data mining to quantify similarity between two data vectors (e.g. Zamir *et al.* 1997, Caverlee *et al.* 2004).

2.2.4. *Quantifying spectral feature B1*

To quantify the backward spectral feature B1 for each of the 196 HRTFs, the centre-back HRTF was extracted. Then the average magnitude within the frequency range of 10–13 KHz was measured for each HRTF. This produced 196 parameters for further analysis.

3. Cluster analysis

Once the HRTF spectral features have been quantified as parameters, the HRTFs can be clustered into near-orthogonal groups.

3.1. Non-individualised head-related transfer function sources

In the cluster analysis of this study, an HRTF is defined to include the HRTFs measured at the same ear for incident sounds of different angles. This ear-based definition is different from the subject-based definition that is commonly used. The ear-based definition is used since the objective of cluster analysis is to group HRTFs according to their ability to produce forward and backward directional perception (as discussed in section 2.1).

HRTFs measured at 196 ears were used as the raw data for cluster analysis. In total, 90 HRTFs were taken from the CIPIC HRTF database (Algazi *et al.* 2001, half were measured at the left ear of 45 subjects and the other half measured at the right ear of the same subjects). A total of 102 HRTFs were taken from the HRTF database collected by the European project entitled 'Augmenting everyday environments through interactive soundscapes (LISTEN)' (Ircam and AKG Acoustics, 2004, again half measured at the left ear of 51 subjects, and the other half measured at the right ear of the same subjects). Two HRTFs were taken from the MIT-KEMAR data (Gardner and Martin 1995; one measured at the left ear of a KEMAR and one measured at the right ear of the same mannequin). The last two HRTFs were measured at the second author's left and right ear, respectively (Ngan 2005). These 196 HRTFs contained different numbers of coefficients in their impulse response functions (from 200 to 512 coefficients). In this study, only the first 200 coefficients of the impulse responses were used and a simple rectangular window truncated the data (Begault 1994). No further scaling was applied to the 196 HRTFs except that a -84 dB scaling factor was applied to the MIT-KEMAR HRTFs to remove the 84 dB constant offset in the original data downloaded from the MIT website (Gardner and Martin 1995).

3.2. Combining the parameters for cluster analysis

As described in the previous section, nine parameters for each of the 196 HRTFs and a 196×196 spectral variation similarity matrix (for feature F6) were extracted to represent similarity in how each HRTF characterises forward directional sounds. To combine these parameters for clustering, 196×196 similarity matrices were calculated for each of the nine parameters (F1, F2, F3, F4, F5, F6's standard deviation, B1, B2 and B3). The coefficients of the nine matrices were calculated as follows (shown here for spectral feature F1):

$$\text{coefficient}_{i,j} = e^{-\text{abs}\ [(p[i]-p[j])/(\text{mean of 196 parameters extracted for F1})]}$$

where $\text{coefficient}^{i,j}$ is the coefficient at i^{th} row and j^{th} column (i and j range from 1 to 196), abs = absolute value, p[i] = parameter quantifying feature F1 for the i^{th} HRTF, and p[j] = parameter quantifying feature F1 for the j^{th} HRTF.

These coefficients range in value from 0 to 1 with:

- 0 very different;
- +1 the same (or nearly the same if coefficient near +1).

In other words, coefficient(37, 196) = 1 indicates that the parameters extracted from HRTF(37) are the same as that from HRTF(196) for a particular feature. In general, a larger coefficient(i,j) indicates that HRTF(i) is closer to HRTF(j), while a smaller coefficient(i,j) indicates that HRTF(i) is different to HRTF(j).

The F6 spectral variation similarity matrix contains coefficients ranging in value from -1 to 1 with:

- -1 related in opposition;
- 0 not related, very different;
- +1 perfectly related, exactly the same (or nearly the same if coefficient near +1).

In other words, coefficient(i,j) = 0 indicates that HRTF(i) is the same as HRTF(j) for feature F6.

The nine 196×196 similarity matrices were then combined together with the F6 spectral variation similarity matrix to form a single 196×196 master similarity matrix. Matrix coefficients from each of the 10 similarity matrices were treated with equal weight and simply added together to form the master similarity matrix. Theoretically, the value of the coefficient in the master similarity matrix can be negative since the F6 spectral variation matrix can be negative. In practice, this did not happen after summing the nine other positive-valued matrices. The master similarity matrix was then imported to MatLab (The Mathworks Inc., Natick, MA, USA) for cluster analysis.

The authors acknowledge that normalising the coefficients of the F6 spectral variation similarity matrix to values between 0 and 1 would make all 10 matrices uniform. That was not done because, within the F6 frequency band, the HRTF gain could be amplified to increase frontal perception (e.g. feature F2; Hebrank and Wright 1974) or reduced to increase backward perception (e.g. feature B3; Blauert 1969/70). In other words, spectral changes with the F6 band could have the opposite effect. The cosine algorithm uses negative values to represent opposite relationships between two spectra. In addition, absolute changes in spectral variation within the F6 frequency band are already represented by standard deviations in the F6 similarity matrix.

3.3. *Selection of clustering algorithms and verification of clustering results*

Cluster analysis was conducted using the four most common hierarchical agglomerative clustering algorithms: (i) single linkage; (ii) complete linkage; (iii) average linkage; (iv) Ward's minimum-variance method (Aldenderfer and Blashfield 1984). In order to select the most suitable clustering algorithm, the results were evaluated based on the following four criteria: (1) minimisation of singleton clusters; (2) maximum robustness in cluster formations; (3) maximum inter-cluster distance; (4) minimisation of intra-cluster distance. The level of robustness was calculated using the 'leave-one-out' method, which represents the level of robustness in cluster formations when a single data point is deleted (Duda *et al*. 2001). Results indicate that the single linkage algorithm had the best scores for robustness, inter-cluster distance and intra-cluster distance. The average linkage algorithm was second in these categories. However, in the end, the average linkage algorithm was selected as the best overall clustering algorithm because the single linkage algorithm produced many more singleton clusters (Aldenderfer and Blashfield 1984).

3.4. *Determining the number of clusters and results*

To recap, the similarity of the 196 HRTFs was represented by a 196 × 196 master similarity matrix. The average linkage algorithm used this matrix to group the 196 HRTFs into clusters. The results of cluster analysis are shown in the dendrogram in Figure 4. HRTFs with similar spectral features for producing forward directional sounds were grouped into the same cluster. As shown in Figure 4, as the amalgamation coefficient decreases, the number of clusters increases. This coefficient can be interpreted as the lower threshold for the inter-cluster distance. In other words, when the distance between two data points is greater than this threshold, they are grouped in different clusters.

The number of clusters selected in this study was the result of a compromise between the stability of the cluster formation and the number of clusters. The stability index for n clusters is defined as the difference between the maximum amalgamation coefficient and the maximum amalgamation for $n + 1$ clusters (Aldenderfer and Blashfield 1984). A larger index

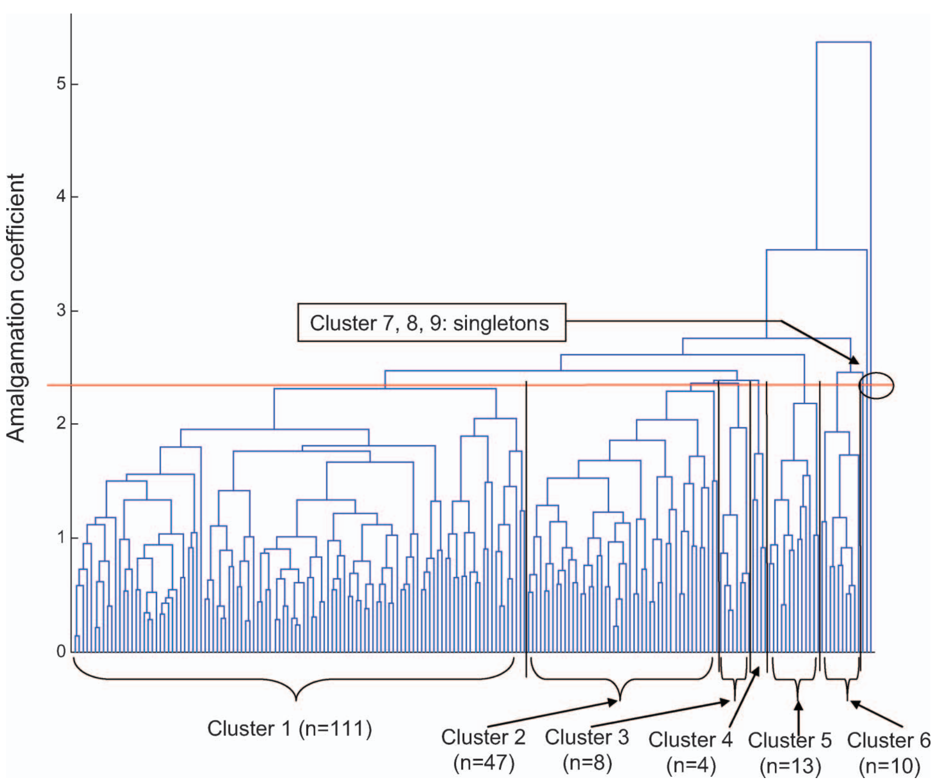

Figure 4. A dendrogram of the head-related transfer function clustering results for forward directional spectral features. The average linkage algorithm was used for clustering.

indicates a more stable cluster formation. Figure 4 indicates that, when there are two clusters, the stability index is at its maximum. However, this causes the 196 HRTFs to be separated into one singleton cluster together with a large cluster containing the other 195 HRTFs. An examination of the stability index indicates that when the number of clusters is 9, 11, 12 and 14, the stability indices are at local maximums.

Nine clusters (including three singleton clusters) were chosen arbitrarily so that six near-orthogonal HRTFs could be determined. The rationale for this is that if a significant reduction in sound localisation error can be achieved by allowing the listener to choose between six near-orthogonal HRTF-generated stimuli, similar or even larger benefits can be expected using more clusters and more choices. Therefore, the number of clusters was set to nine. Similarly, nine clusters were also chosen for backward directional sound. The authors acknowledge that future studies might use even more clusters.

The number of HRTFs grouped within each cluster is shown in Table 2. For both forward and backward directions, nine clusters were chosen, with three of them singleton clusters. The singleton clusters were ignored because they were outliers. Each non-singleton cluster was analysed to determine the HRTF closest to its centre. This was identified and labelled as an orthogonal HRTF. Consequently, six orthogonal HRTFs were determined according to their ability to produce forward directional sound and six more for backward directional sound. Table 3 contains the source of these 12 near-orthogonal HRTFs so that readers can download them for research purposes.

Table 2. Number of head-related transfer functions (HRTFs) grouped within each of the six forward and backward clusters.

Direction of sound	Cluster number	Number of HRTFs grouped within each cluster
Forward	1	111
	2	47
	3	8
	4	4
	5	13
	6	10
Backward	1	121
	2	20
	3	13
	4	31
	5	6
	6	2

4. The main experiment (experiment 1)

4.1. *Objective and hypotheses*

An experiment was conducted to verify improvement in localisation performance when listeners were given the choice of six centre-front sound stimuli (or six centre-back stimuli) as compared to just one centre-front (or centre-back) stimuli generated by MIT-KEMAR HRTFs. It was hypothesised that, for both directions, the choice of the six sound stimuli would improve sound localisation performance compared to the single stimuli generated by MIT-KEMAR HRTFs. More formally, assuming that there exists a means for

Table 3. The source of the 12 near-orthogonal head-related transfer functions (HRTFs) determined by cluster analysis for forward and backward directional sound generation.

Direction of sound	HRTF no.*	Number of HRTFs within cluster	HRTF source: HRTF Database	HRTF source: Subject no.	HRTF source: Left/right ear
Forward	1	111	LISTEN‡	1054	Left
	2	47	CIPIC†	15	Right
	3	8	CIPIC†	148	Left
	4	4	LISTEN‡	1059	Left
	5	13	LISTEN‡	1046	Right
	6	10	CIPIC†	28	Right
Backward	1	121	CIPIC†	51	Left
	2	20	LISTEN‡	1003	Right
	3	13	CIPIC†	147	Left
	4	31	CIPIC†	155	Left
	5	6	CIPIC†	20	Right
	6	2	CIPIC†	12	Left

*The HRTF number is equivalent to the cluster number (which follows the order in the dendrogram of Figure 4).
†CIPIC head-related impulse responses (HRIRs) are free to download at the CIPIC Interface Laboratory home page (http://interface.cipic.ucdavis.edu/data/CIPIC_hrtf_database.zip).
‡LISTEN HRIRs are free to download at the LISTEN HRTF Database home page (http://recherche.ircam.fr/equipes/salles/listen/download.html).

listeners to choose their best-matched directional sound, the localisation error of participants in perceiving their 'best-matched' sound stimuli were hypothesised to be significantly lower than the localisation error for stimuli generated using the MIT-KEMAR HRTFs (H1: for stimuli at the centre-front direction; H2: for stimuli at the centre-back direction). It was also hypothesised that the improvement in localisation performance would be negated when the error was corrected for front–back confusion. In other words, H1 and H2 would not be supported after the error for front–back confusion had been corrected (H3 and H4 for, respectively, H1 and H2).

4.2. Method

4.2.1. Design of experiment, variables and participants

The experiment used a within-subject full-factorial design with three independent variables. The three independent variables were: (i) the four sound directions (centre-front: 0° azimuth and 0° elevation; centre-back: 180° azimuth and 0° elevation; decoy direction A: 0° azimuth and +60° elevation; decoy direction B: 0° azimuth and −30° elevation); (ii) the stimuli generated from seven HRTFs (the six orthogonal HRTFs and the MIT-KEMAR HRTF); (iii) six repetitions. There were 168 conditions that comprised the exhaustive combinations of four sound directions, seven HRTFs and six repetitions. Listeners were told that sound could come from all possible incident angles and they were instructed to report the perceived incident angles for each presented sound stimulus. Only the data collected for the centre-front and centre-back sound stimuli are reported in this paper.

The main measurements in this experiment were the perceived incident angles of the directional sounds along the azimuth and elevation directions (see Figure 1). These data made it possible to calculate the following dependent variables: (i) the localisation error along the azimuth direction (i.e. the left–right error), which is the angular separation between the perceived incident angle and its known direction; (ii) the left–right error after front–back confusion correction; (iii) the percentage of front–back confusion in the six repeated trials for each condition. The front–back confusion correction procedure identified occurrences of front–back confusion and adjusted for their effect on the measurements. For example, a centre-front sound stimulus incorrectly perceived to come from 160° azimuth and +30° elevation (i.e. from behind, slightly above and to the right) would be corrected to have come from 20° azimuth and +30° elevation (from the front, slightly above and to the right).

Six male and nine female listeners participated in this experiment. They were university students aged between 19 and 25 years. These listeners had been screened via questionnaires to be free from hearing problems (Begault 1994). In addition, audiometric tests were conducted (Abel and Paik 2004) and all 15 listeners had hearing thresholds of 20 dB or less at 125 Hz, 1 kHz and 8 kHz. The inter-aural difference was 5 dB or less for all 15 listeners. The experiment was approved in advance by the Human Subjects Experimentation Committee at the Hong Kong University of Science and Technology (HKUST). Listeners received US$7 per hour to compensate for their time and travel expenses.

4.2.2. Stimuli, presentation, and the listening environment

The original mono sound source for the binaural sound stimuli was a sequence of 10 white noise pulses of 0.3 s duration separated by 0.3 s of silence. Noise pulses have been used as directional sound stimuli in many previous studies on sound localisation (e.g. Middlebrooks and Green 1991, Horman and van Opstal 1998, Jin *et al.* 2004). To produce a directional sound for a particular combination of the four sound directions and seven HRTFs (the six orthogonal HRTFs plus the MIT-KEMAR HRTF), the appropriate HRTF was selected and transformed into a head-related impulse response (HRIR). The mono source was convoluted with the HRIR to form a mono sound, which was then presented to both the left and right ears. This procedure was documented by Begault (1994) and was implemented using MatLab. The sound stimuli were purposely anechoic since echo and reverberation can influence sound localisation performance (Begault 1994, Blauert 1997).

Listeners were seated and the sound stimuli were presented via a pair of Sennheiser HD580 headphones (Sennheiser Electronic, Wennebostal, Germany). This headphone was selected because of its flat spectral response and it was also recommended for use with the MIT-KEMAR HRTFs (Gardner and Martin 1995). The experiment was conducted inside the inner room of an air-conditioned, double-chamber acoustic room (a custom-made 1400-A-CT chamber built by the Industrial Acoustics Company, Winchester, Hants, UK). The chamber sits on an anti-vibration floor and has a double-walled panel for noise isolation. The background noise was measured to be 38 dBA. This chamber has an inner wall and ceiling filled with sound absorption material 6 inches thick. It is acknowledged that there may still be some slight mismatch between the semi-anechoic chamber and the anechoic sound stimulus. However, every measure was taken to reduce any mismatch. For example, listeners were instructed to close their eyes when they listened to the sound stimuli. Moreover, the procedure for presenting the sound stimuli and collecting their incident angles has been used previously (Gilkey *et al.* 1995).

4.2.3. Procedure

The 168 conditions were separated into six repeated sessions of 28 conditions that represent the exhaustive combinations of four incident directions (centre-front, centre-back and two decoy directions) and seven HRTFs. Each session lasted for about 10 min with 5-min break between each session. Within each session, the presentation order for the 28 conditions was randomised.

The subjects were asked to close their eyes while listening to each sound stimulus. A 2-D level meter was attached to the headphones to minimise tilting of the head. After listening to each sound, subjects were asked to indicate the perceived incident angle of the sound using a hand-held pointer. In front of the listener was a basketball-like sphere. The subject was instructed to treat the centre of the sphere as the centre of his/her head and to touch the pointer on the surface of the sphere to indicate the perceived incident angle of the sound. A Polhemus 3-SPACE position sensor (Colchester, VT, USA) was used to measure the position of the pointer tip. MatLab code was developed to calculate the incident angle from the tip of the pointer to the centre of the sphere. This method of measuring perceived incident sound angles was adopted from Gilkey *et al.* (1995). A photograph of a listener using the pointer to indicate the perceived incident direction of a sound is shown in Figure 5. Written instructions were given and listeners were allowed to become familiar with the procedure and apparatus before the experiment.

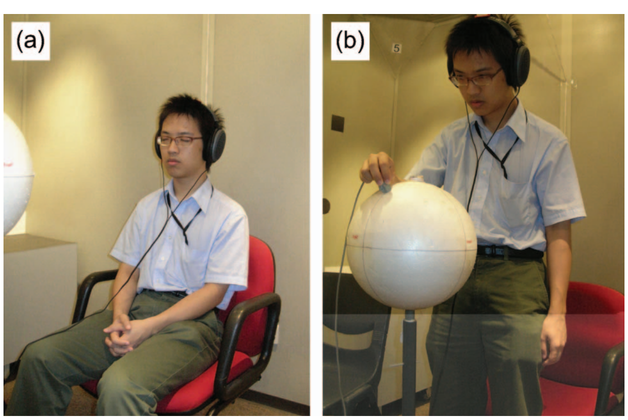

Figure 5. Two photographs showing a listener (a) listening to a sound and (b) using a pointer to indicate the perceived incident direction of the sound that he has just heard.

4.3. Results and discussion

4.3.1. Effects of repetition, listener and gender

The data did not follow a normal distribution and non-parametric statistical tests were used to analyse the data. Friedman two-way ANOVA indicated that repetition did not have a significant effect on the perceived localisation error along the azimuth and elevation directions for both sound directions ($p > 0.08$: azimuth errors; $p > 0.25$: elevation errors). No consistent learning trend among listeners was observed from the data. The absence of a significant effect for repetition was consistent with a previous study on intra-subject variability in sound localisation performance (Ngan 2005).

Results of Friedman two-way ANOVA indicated that female listeners had significantly more localisation error along the azimuth direction than male listeners ($p < 0.001$). The significant effect of gender is not due to a gender bias in the orthogonal HRTFs since the 12 selected sets were based on non-individualised HRTFs measured with both male and female listeners. It is possible that the significant main effect of gender is due to differences in ear dimensions, as Zotkin *et al.* (2002) reported a possible link between ear dimensions and localisation performance for binaural sound generated using non-individualised HRTFs. Results of a Mann–Whitney U test indicate that the pinna height of male listeners was significantly larger than that of female listeners ($p < 0.001$). However, no significant correlation was found between pinna height and localisation error along the azimuth direction ($p > 0.1$). Further study to investigate the effects of gender is needed. In the subsequent analysis, data from both genders are used, since this study aims to examine the benefits of providing orthogonal choices to the general population, not just males.

4.3.2. Benefits of providing the choice of six orthogonal stimuli vs. the MIT-KEMAR stimulus alone and the testing of hypotheses 1–4

For each of the four sound directions, the orthogonal stimulus associated with the smallest median localisation error in the six repeated trials for each listener was labelled as the 'best-matched' stimulus. Wilcoxon signed ranked tests were used to compare the localisation error along the azimuth direction for the best-matched stimuli with those from the MIT-KEMAR stimulus. The results indicate that, for both centre-front and centre-back directions, errors for the best-matched stimuli were significantly lower than for the MIT-KEMAR stimulus ($p < 0.005$). This result supports H1 and H2 and demonstrates the significant benefit of providing the choice of six orthogonal sound stimuli for listeners. The reduction in the average localisation error along the azimuth axis was about 22% for both directions (centre-front: from 147 to 115°; centre-back: from 144 to 112°).

To test H3 and H4, front–back correction was applied. In other words, if a centre-front sound was incorrectly perceived to be at the back (e.g. 130° azimuth and +20° elevation), the perceived angle would be mirrored back to the front using the sagittal plane as a mirror (i.e. mirrored to 50° azimuth and +20° elevation). Results of Wilcoxon tests indicated that, for both sound directions, the localisation errors for the best-matched stimuli were not significantly smaller than those from the MIT-KEMAT stimuli. This supports H3 and H4 and suggests that the benefit of providing a choice of orthogonal stimuli is that it reduces front–back confusion.

4.3.3. Effects of reducing the number of orthogonal stimuli choices from six to one

After demonstrating the significant benefits of providing a choice of six orthogonal stimuli, it would be interesting to find out whether the same benefits can be maintained with fewer choices. To answer this question, the orthogonal stimuli were ranked by the number of HRTFs that were clustered close to them (see Table 2). Then the stimulus with the smallest number of HRTFs was removed to give five orthogonal stimuli. This procedure was repeated to find the choices for four, three, two and one stimuli. Inevitably, reducing the number of stimuli reduced the scope of representation. It was hypothesised that as the number of choices decreased, fewer listeners would be able to find the most suitable stimuli and therefore the average localisation error for the best-matched stimuli among the 15 listeners would increase.

Figures 6 and 7 show the mean localisation error along the azimuth axis for the 15 listeners as the number of choices decreases from six to one for centre-front stimuli (Figure 6) and centre-back stimuli (Figure 7). The mean errors for the MIT-KEMAR stimuli are also shown for reference. Both figures indicate that, as the number of choices decreases, the localisation error for the best-matched stimuli also decreases significantly ($p < 0.001$, Friedman ANOVA). For each point in the graphs, Wilcoxon signed ranked tests were conducted to compare the localisation error for the best-matched stimuli with the MIT-KEMAR stimuli error. The results indicate that for both centre-front and centre-back directions, as the

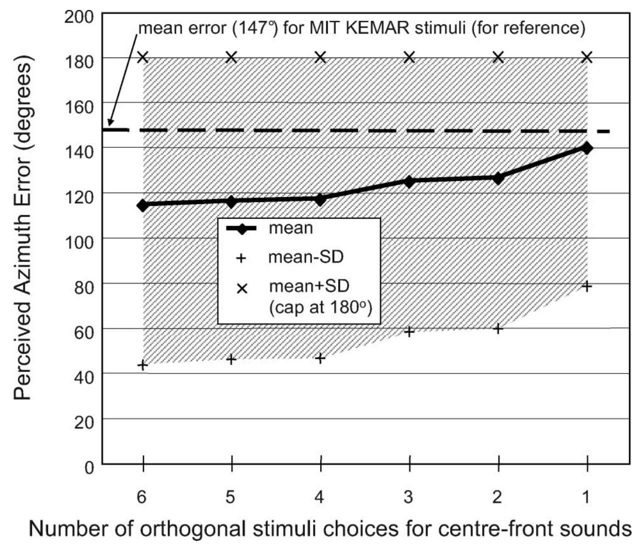

Figure 6. Plots of mean localisation error and standard deviation along the azimuth direction for the best-matched centre-front stimuli as a function of the number of orthogonal stimuli choices. Each mean and standard deviation is calculated from 90 data points taken from six repeated trials of 15 listeners. The mean error obtained for the Massachusetts Institute of Technology Knowles electronics manikin for acoustic research (MIT-KEMAR) stimuli are also shown.

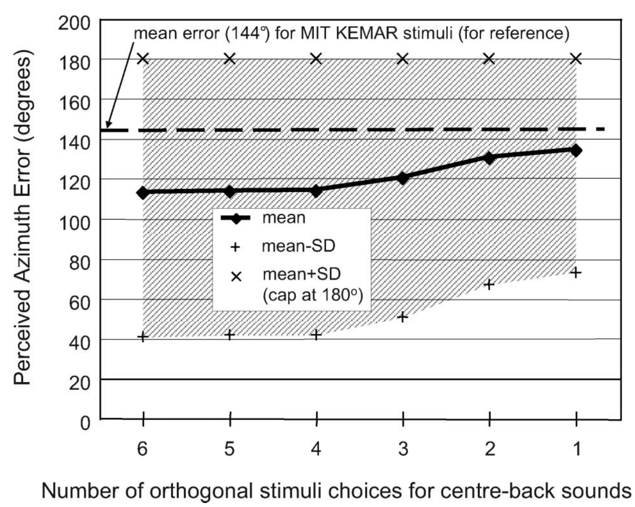

Figure 7. Plots of mean localisation error and standard deviation along the azimuth direction for the best-matched centre-back stimuli as a function of the number of orthogonal stimuli choices. Each mean and standard deviation is calculated from 90 data points taken from 6 repeated trials of 15 listeners. The mean error obtained for the Massachusetts Institute of Technology Knowles electronics manikin for acoustic research (MIT-KEMAR) stimuli are also shown.

number of choices is reduced to five, the significant reduction in localisation error for the best-matched stimuli over the MIT-KEMAR stimuli holds

($p < 0.05$, Wilcoxon). However, when the number of choices is reduced to four or less, the differences are no longer significant.

4.3.4. Front–back confusion rates, sensitivity and response bias

An examination of occurrences of front–back confusion indicated that for centre-back stimuli, eight out of the 15 listeners exhibited a reduction in front–back confusion rates with at least one of the six orthogonal stimuli. The other seven listeners exhibited 0% front–back confusion rates in all their sound localisation trials (for the six orthogonal stimuli and the MIT-KEMAR stimulus). This suggests that front–back confusion was not a problem for 46% of the participants; for the remaining 54%, providing five or more choices significantly reduced front–back confusion. The average front–back confusion rate for centre-back stimuli was 10% for six orthogonal choices, a 65% reduction compared to the 29% front–back confusion rate for the MIT-KEMAR stimulus alone.

When a similar examination was conducted on centre-front stimuli, an interesting result was obtained. Although 10 out of 15 listeners exhibited reductions in front–back confusion with at least one of the six orthogonal stimuli, only four listeners had a front–back confusion rate of 33% or less for their best-matched stimuli (i.e. two front–back confusions or less on the six trials).

Furthermore, five of the 15 listeners had 100% front–back confusion rates. This translates to a 56% average front–back confusion rate among the 90 trials (six repetitions among 15 listeners). This was obviously not a satisfactory finding and the MIT-KEMAR centre-front stimulus also resulted in a high average front–back confusion rate of 63%. Possible reasons for the failure of these five listeners to correctly perceive centre-front stimuli as coming from the front include: (i) the task was too difficult for these five listeners so they were insensitive to the incident angle of the sound; (ii) the six orthogonal stimuli were not representative enough because some spectral features were not selected in the feature extraction procedure; (iii) the six orthogonal stimuli were not representative enough because the 196 raw HRTFs used in the cluster analysis were biased samples; (iv) a certain percentage of listeners are simply incapable of accurately localising directional sound generated using a HRTF.

The first possible reason is related to the listeners' sensitivity towards the sound (i.e. its correct incident direction). During the main experiment, listeners were told that the sound could come from all possible directions and, indeed, there were sound stimuli coming from above and below (the decoy stimuli). While this should have reduced any response bias from the listeners, it might have reduced the listeners' sensitivity to the incident angle of the sound. To explore this possibility, a two-alternative forced choice (2AFC) version of the main experiment was conducted and this is referred to as experiment 2. To explore the third and fourth possibilities, a second follow-up study was conducted, in which the listeners were asked to localise the centre-front sound stimuli generated by all 196 non-individualised HRTFs. This study was referred to as experiment 3 (see section 4.3.6).

4.3.5. A two-alternative forced choice follow-up study (experiment 2)

The main experiment was repeated with an easier 2AFC sound localisation discrimination task. The same 15 listeners were invited back for the follow-up experiment. For this experiment, listeners were exposed to two 4.8-s sound stimuli (consisting of eight 0.3-s noise pulses and 0.3-s gaps). These were presented consecutively with time gaps of 1 s. Listeners were asked to discriminate which sound was from the front and which from the back. Listeners were told that one of the stimuli came from a forward incident angle and the other from a backward incident angle. Each pair of stimuli was repeated six times. The order of presentation of which stimulus was first within the pair was balanced. Similar to the main experiment, the centre-front and centre-back sound stimuli were filtered using the HRIRs extracted from the six orthogonal HRTFs and the MIT-KEMAR HRTF. Because the six orthogonal centre-front HRTFs were different from those for the centre-back (see Table 3), the two conditions with centre-front and centre-back sound stimuli could not be combined. In this experiment, listeners were exposed to 84 pairs of stimuli that were combinations of two sound directions (centre-front and centre-back), seven stimuli (generated from the six orthogonal HRTFs and the MIT-KEMAR HRTF) and six repetitions. As before, this experiment was approved by the Human Subjects Experimentation Committee of HKUST and listeners received US$7 per hour compensation.

The results indicate that, for centre-front stimuli, the five listeners who had 100% front–back confusion (i.e. perceived all centre-front stimuli to be coming from the back in all six repeated trials for all conditions) still had 100% front–back confusion. This confirms that replacing the sound localisation task with a simpler 2AFC discrimination task does not help such listeners improve their sound localisation accuracy. Furthermore, the 100% front–back confusion rate suggests that these listeners were consistently perceiving the wrong direction, not just randomly guessing (which would give a 50% error rate).

For centre-back stimuli, all 15 listeners had less than 50% front–back confusion on their best-matched stimuli (six listeners had 0%, six listeners 17% and three listeners 33%). This translates to an average front–back confusion rate of 13%. The authors were surprised to find that the front–back confusion rate in experiment 2 was higher than that in experiment 1, although they were not significantly different ($p > 0.4$, Wilcoxon). A possible reason might be that listeners did not find the sound localisation task in the main experiment difficult and, therefore, changing it to a 2AFC task did not change the performance. On the other hand, the results of the main experiment and the follow-up experiment are consistent.

4.3.6. *A second follow-up study (experiment 3)*

To explore whether the 100% front–back confusion rates for the same five listeners in experiments 1 and 2 were due to possible sample bias caused by the 196 HRTFs, a second follow-up study (experiment 3) was conducted. The aim was to test whether these five listeners would find a best-matched stimulus out of the 196 centre-front stimuli generated by the 196 HRTFs.

In this second follow-up study, all 15 listeners were invited back to listen to the centre-front stimuli generated using all 196 HRTFs, presented in a random order. The listeners were told that sound would come from all possible incident angles and were instructed to indicate the perceived incident angle in the same way as in the main experiment. They were also told that each angle might be repeated. The stimuli preparation and presentation were similar to those in the main experiment. The presentation of the 196 stimuli was separated into four sessions with 5-min breaks in between.

The results indicate that three of the five listeners still perceived all 196 centre-front stimuli as coming from the back, while the remaining two perceived 194 and 130 stimuli as coming from the back, respectively. These results partially support the third and the forth possible reasons listed in section 4.3.4 that either the 196 HRTFs were not representative enough or these listeners were simply not good at localising simulated binaural sound. Asano *et al*. (1990) reported that one listener had to be excluded from their experiment because that subject could not accurately localise the binaural cues generated using the listener's own individualised HRTFs despite repeated attempts. Asano's finding suggests that some listeners might fail to localise a binaural sound accurately even though it is generated using the listener's own individualised HRTFs. This does not mean that such listeners are unable to localise sound stimuli in real life, because other aids such as echoes and head movements were not included in the HRTF binaural sound stimuli. Further work is needed to consider the benefits of adding echoes and dynamic responses, such as head movement. The potential effects of bias in experiment 3 are addressed in section 5.

The four listeners in the main experiment who reported 33% or less front–back confusion rates for their best-matched centre-front stimuli also had consistent results in experiment 3.

5. Discussion and limitations

The main experiment (experiment 1) demonstrated that a choice of six orthogonal stimuli could significantly reduce azimuth sound localisation error by about 22% compared to simply using the stimuli generated by the MIT-KEMAR HRTFs. This reduction was negated when the perceived incident sound was corrected for front–back confusion. This indicates that the benefit of providing a choice of orthogonal stimuli is in reducing front–back confusion. In particular, the choice of orthogonal stimuli successfully reduces front–back confusion for centre-back stimuli, but not for centre-front stimuli. In fact, one could conclude that the problem of front–back confusion for centre-front stimuli is more serious than for centre-back stimuli – a finding that has not been reported before.

Two follow-up studies were conducted to explore possible reasons for 100% front–back confusion rates among five of the listeners. The first follow-up study (experiment 2) replaced the sound localisation task with a 2AFC task of front–back discrimination. It was hypothesised that the five listeners who reported 100% front–back confusion would manage to correctly discriminate the centre-front stimuli as coming from the front. Surprisingly, these five listeners still had 100% front–back confusion for all 42 centre-front stimuli (six repetitions of seven types of centre-front stimuli generated from the six orthogonal HRTFs plus the MIT-KEMAR HRTF).

A second follow-up study (experiment 3) was conducted to determine whether these five listeners were able to correctly perceive some of the centre-front stimuli generated by the 196 non-individualised HRTFs. The results indicate that four of the five listeners perceived all 196 centre-front stimuli as coming from the back. This suggests that the 196 HRTFs might not have covered all the different profiles for generating centre-front sound stimuli. The authors acknowledge that the second follow-up experiment might have biased listeners' responses, since one might expect a random distribution of angles and, therefore, this expectation would be inconsistent with the fact that all the stimuli were, in fact, coming from the front. However, this bias should have discouraged a listener from answering that all 196 centre-front stimuli were coming from the back. Obviously, such a bias did not prevent four of the listeners from doing exactly that in experiment 3, where they had 100% front–back confusion.

This study assumes that there exists a means for listeners to choose the stimuli that they consider most accurate and this has been modelled by the best-matched stimuli. In 2005, the authors reported an online sound localisation game for listeners to choose their best-matched forward and backward stimuli (Ngan *et al*. 2005). However, it is acknowledged that having to chose from six stimuli might be inconvenient in some applications.

6. Conclusions and future work

Based upon previous work, six spectral HRTF features for the generation of forward directional sounds were identified, defined and quantified for 196 non-individualised HRTFs. These include HRTFs measured for 90 ears in the CIPIC project (Algazi *et al*. 2001) and 102 ears in the LISTEN project (Ircam and AKG Acoustics 2004).

Similarly, three spectral HRTF features for the generation of backward directional sounds were identified, defined and quantified for the same 196 non-individualised HRTFs.

These 196 HRTFs were grouped into nine orthogonal groups (including three singleton groups) according to their ability to generate forward directional sounds. While ignoring the singleton groups, the HRTFs located at the centre of each of the six remaining groups were extracted and used to filter a white-noise pulse chain to make six orthogonal centre-front sound stimuli. Similarly, six orthogonal centre-back stimuli were also determined.

The results of the main experiment show that providing a choice of six orthogonal centre-back stimuli can significantly reduce sound localisation error when compared to using just the centre-back stimulus generated using the MIT-KEMAR HRTF ($p < 0.05$). The results also suggest that the average front–back confusion rate can be reduced by about 65% (from 29 to 10%) using a choice of six orthogonal centre-back stimuli. The result holds even with only five orthogonal centre-back stimuli.

When providing a choice of six orthogonal centre-front stimuli, the benefit is not as obvious. Although there is a significant reduction in sound localisation error, the average rate of front–back confusion does not come down much, remaining above 50%. In particular, 33% of listeners reported 100% front–back confusion in the main experiment as well as in two follow-up studies. Results of the second follow-up study showed that four of 15 listeners perceived the centre-front stimuli as coming from the back. Further work is needed to explore why these listeners have such a strong tendency to perceive forward directional sound as coming from the back.

In this study, only centre-front and centre-back stimuli have been considered, since the cluster analysis used HRTF data for these two directions only. Future work is needed to extend this work to other incident angles.

Acknowledgements

The authors would like to thank the Hong Kong Research Grants Council for partially supporting this work through an earmarked competitive grant (HKUST6219/02E). The authors would also like to thank Mr Denil Chan and Mr Stanley Sze for their technical support.

References

Abel, S.M. and Paik, S.J.E., 2004. The benefit of practice for sound localization without sight. *Applied Acoustics*, 65, 229–241.

Aldenderfer, M.S. and Blashfield, R.K., 1984. *Cluster analysis*. Thousand Oaks, CA: Sage Publications, Inc.

Algazi, V.R., *et al*., 2001. The CIPIC HRTF database. *In*: *Proceedings of IEEE ASSP workshop on applications of signal processing to acoustics*, October, New Paltz, NY, USA. New York: IEEE Press, 111–123.

Apostolico, A. and Galil, Z., 1997. *Pattern matching algorithms*. Oxford: Oxford University Press.

Asano, F., Suzuki, F., and Sone, T., 1990. Role of spectral cues in median plane. *Journal of the Acoustical Society of America*, 88 (1), 159–168.

Begault, D.R., 1994. *3D sound for virtual reality and multimedia*. Rochester, NY: AP Professional.

Blauert, J., 1969/70. Sound localization in the median plane. *Acustica*, 22, 205–213.

Blauert, J., 1997. *Spatial hearing: the psychophysics of human sound localization*. Cambridge, MA: MIT Press.

Caverlee, J., Liu, L., and Buttler, D., 2004. Probe, cluster, and discover: focused extraction of QA-Pagelets from the deep web. *Proceedings of 20th IEEE international conference on data engineering*, 2004, 30 March – 2 April, Boston, USA. New York: IEEE Press, 103–114.

Duda, R.O., Hart, P.E., and Stork, D.G., 2001. *Pattern classification*. New York: John Wiley & Sons, Inc.

Gardner, M.B. and Gardner, R.S., 1973. Problem of localization in the median plane: effect of pinnae cavity occlusion. *Journal of the Acoustical Society of America*, 53 (2), 400–408.

Gardner, W.G. and Martin, K.D., 1995. HRTF Measurements of a KEMAR. *Journal of the Acoustical Society of America of America*, 97, 3907–3908.

Gilkey, R.H., *et al*., 1995. A pointing technique for rapidly collecting localization responses in auditory research. *Behavior Research Methods*, 27 (1), 1–11.

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*, 56, 1829–1834.

Horman, P.M. and van Opstal, A.J., 1998. Spectro-temporal factors in two-dimensional human sound localization. *Journal of the Acoustical Society of America*, 103, 2634–2649.

Ircam and AKG Acoustics, 2004. *LISTEN HRTF database* [online]. Available from: http://recherche.ircam.fr/equipes/salles/listen/download.html [Accessed 25 June 2004].

Jin, C., *et al*., 2004. Contrasting monaural and interaural spectral cues for human sound localization. *Journal of the Acoustical Society of America*, 115, 3124–3142.

Langendijk, E.H.A. and Bronkhorst, A.W., 2002. Contribution of spectral cues to human sound localization. *Journal of the Acoustical Society of America of America*, 112 (4), 1583–1596.

Lopez-Poveda, E.A. and Meddis, R., 1996. A physical model of sound diffraction and reflections in the human concha. *Journal of the Acoustical Society of America of America*, 100, 3248–3260.

Middlebrooks, J.C. and Green, D.M., 1991. Sound localization by human listeners. *Annual Review of Psychology*, 42, 135–140.

Myers, P.H., 1989. *Three-dimension auditory display apparatus and method utilizing enhanced bionic emulation of human binaural sound localization*. US Patent No. 4817149. Assignee: American Natural Sound Company. Filed 22 January 1987, Published 28 March 1989.

Ngan, K.H., 2005. *Optimizing front/back confusion rates in sound localization performance: cluster analyses and experimental studies*. Unpublished thesis (MPhil). Hong Kong University of Science and Technology.

Ngan, K.H., *et al*., 2005. Mass-customisation of virtual headphone-based surround sound (VHSS) systems. *Proceedings of the third interdisciplinary world congress on mass-customization and personalization*, 18–21 September, Hong Kong, The Hong University of Science and Technology, CD-Rom.

Shaw, E.A.G. and Teranishi, R., 1958. Sound pressure generated in an external-ear replica and real human ears by a nearby sound source. *Journal of the Acoustical Society of America of America*, 44, 240–249.

Tan, C.J. and Gan, W.S., 1998. User-defined spectral manipulation of HRTF for improved localization in 3D sound systems. *Electronics Letters*, 34, 2387–2389.

Wenzel, E.M., *et al*., 1988. Acoustic origins of individual differences in sound localization behaviour. *Journal of the Acoustical Society of America of America*, 84, S79.

Wenzel, E.M., *et al*., 1993. Localization using non-individualised head-related transfer functions. *Journal of the Acoustical Society of America of America*, 94, 111–123.

Zamir, O., *et al*., 1997. Fast and intuitive clustering of web documents. *In*: *Proceedings of the 3rd international conference on knowledge and data mining*, Newport, CA, USA, 14–17 August. New York: ACM Press, 287–290.

Zotkin, D.N., *et al*., 2002. Virtual audio system customization using visual matching of ear parameters. ICPR 2002, Quebec City, Canada, August 2002. Washington, DC: IEEE Computer Society, 1003–1006.