ELSEVIER

Applied Ergonomics 37 (2006) 695–707

APPLIED ERGONOMICS

www.elsevier.com/locate/apergo

A low cost, non-individualized surround sound system based upon head related transfer functions: An ergonomics study and prototype development

R.H.Y. So[a,*], N.M. Leung[a], J. Braasch[b], K.L. Leung[c]

[a]*Department of Industrial Engineering and Logistics Management, Hong Kong University of Science & Technology, Clear Water Bay, Kowloon, Hong Kong SAR*

[b]*Institut für Kommunikationsakustik, Ruhr-Universität Bochum, Universitätsstraße 150, D-4630 Bochum, Germany*

[c]*Rehabilitation Engineering Centre, the Hong Kong Polytechnic University, Kowloon, Hong Kong SAR*

Received 11 January 2005; accepted 29 December 2005

Abstract

This paper reports on the types and magnitudes of localization errors of simulated binaural direction cues generated using non-individualized, head-related transfer functions (HRTFs) with different levels of complexity. Four levels of complexity, as represented by the number of non-zero coefficients of the associated HRTF filters (128, 64, 32, 18 non-zero coefficients), were studied. Experiment 1 collected 1728 data runs that were exhaustive combinations of the four levels of complexity, nine simulated directions of sound (no direction (i.e., diotical-mono), 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315° azimuth angles at 0° elevation), two repetitions, and 24 participants). Binaural cues generated from HRTFs of reduced complexity (from 128 to 18 non-zero coefficients) produced significantly higher localization errors for the directions of 45°, 135°, 225°, and 315° azimuth angles ($p<0.01$). From the directions of 0°, 90°, and 270° azimuth angles, the cues produced by HRTFs with reduced complexity did not affect the localization error ($p>0.2$). Surprisingly, cues produced by HRTFs of 128 non-zero coefficients did not have the lowest number of errors. From 45°, 135°, 225°, and 315°, the lowest numbers of errors were obtained from cues produced by HRTFs of 64, 32, 32, and 64 non-zero coefficients, respectively. Based on these findings, a prototype virtual headphone-based surround-sound (VHSS) system was developed. A double-blind usability experiment with 32 participants indicated that the prototype VHSS system received significantly better surround-sound ratings than did a Dolby^{TM} stereo system ($p<0.02$). This paper reports results from an original ergonomics study and the application of these results to the design of a consumer product.

Keywords: Spectral complexity; HRTF; Virtual surround sound; Sound localization; Binaural direction cue

1. Introduction

1.1. Background on sound localization and head-related transfer functions (HRTFs)

Humans identify the direction of an incident sound cue by interpreting three aspects of an incoming sound: (i) differences in the times of the arrival of the sound in the two ear canals; (ii) differences in sound magnitudes in the two ear canals; and (iii) the frequency content (i.e., spectra) of the sounds in the two ear canals (Begault, 1994; Blauert, 1997). These three incoming sound aspects have been referred to as the (i) inter-aural time differences (ITDs), (ii) inter-aural level differences (ILDs), and (iii) pinna responses, respectively (Blauert, 1997). The values of ITDs and ILDs are mainly affected by the distances a sound wave travels from its source to the openings of the two ear canals: the shorter the travel distance, the earlier the arrival time and the higher the sound magnitude. However, sound sources whose locations form mirror images on the two sides of either the coronal plane (i.e., front vs back) or the

[*] Corresponding author.
E-mail address: rhyso@ust.hk (R.H.Y. So).

0003-6870/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apergo.2006.01.001

two sides of the transverse plane at ear-level (i.e., top vs bottom) would produce similar ITDs and ILDs. These sound locations have been called the cones-of-confusion (Blauert, 1997; Mills, 1972). In order to differentiate these sound sources, humans rely on information embedded within the pinna responses. Acoustically, our pinnas distort incident sound waves in the time and frequency domains through reflection, shadowing, dispersion, diffraction, interference, and resonance (Blauert, 1997; Lopez-Poveda and Meddis, 1996). Shaw and Teranishi (1968) identified a number of resonance frequencies of an artificial outer ear and these resonances can change the spectra (i.e., the frequency modulation) of the incident sound according to their incident angles. In particular, sound pressures arriving at the opening of the ear canal from sources located in the front, back, above, and below the listener are subjected to different patterns of frequency modulation (e.g., Begault, 1994; Blauert, 1997; Middlebrooks, 1992). Lopez-Poveda and Meddis (1996) reported a mathematical model to simulate the acoustics interactions between an incident sound wave and relevant parts of the pinna. To summarize, humans rely on ITDs and ILDs to determine the azimuth direction of an incident sound and rely on pinna responses to determine the sound's front–back and elevation directions.

With a miniaturized microphone inserted in each of the ear canals, ITDs, ILDs, and pinna responses to incident sounds of different angles can be measured in an anechoic environment. For a particular incident angle of sound, the ITDs, ILDs, and pinna response information are embedded in a pair of transfer functions called the HRTFs (e.g., Schröter et al., 1986; Wightman and Kistler, 1989; Middlebrooks and Green, 1990; Gardner and Martin, 1995). These HRTFs are usually stored in the form of impulse responses (Begault, 1994). The numbers of non-zero coefficients in an impulse response determines how many patterns of frequency modulations an HRTF can carry. As these patterns of frequency modulations contain information associated with the pinna response, an HRTF impulse response with more non-zero coefficients should be able to produce more accurate binaural sound cues. The non-zero coefficients in an HRTF impulse response are called the coefficients of an HRTF in this paper. Here, the number of HRTF coefficients is the key independent variable. The effects of these HRTF coefficients on the ability of an HRTF to produce an accurate binaural directional sound cue are studied. Further discussion is in Section 1.3.

1.2. Non-individualized HRTFs and their usages

Studies have shown that HRTFs measured for an individual listener (called individualized HRTFs) can be used as filters to simulate accurate binaural directional cues for that individual (e.g., Wightman and Kistler, 1989; Kulkarni and Colburn, 1998). However, the measurements of individualized HRTFs involve tedious, time-consuming, and expensive (in the range of US$1000, Blauert, 2000) procedures requiring a listener to remain motionless for about 1 h inside an anechoic chamber. This limits the use of individualized HRTFs in commercial products. As a cost-saving alternative, non-individualized HRTFs measured with a mannequin have been used (e.g., the open-copy-righted HRTF data collected using the Knowles electronics manikin for acoustic research (KEMAR) at the Massachusetts Institute of Technology (MIT): Gardner and Martin, 1995; the HRTF data set collected at the Center for Image Processing and Integrated Computing: Algazi et al., 2001). Examples of applications of non-individualized HRTFs include: (i) three-dimensional (3D) auditory displays for the blind (e.g., Lumbreras and Sanchez, 1999); (ii) virtual headphone-based surround-sound (VHSS) systems for entertainment (e.g., Begault, 1992; Lambrecht, 2001; Lowe et al., 1994; Scofield and Saounder, 1996); (iii) a spatial auditory system for virtual conferencing (e.g., Greenhalgh and Benford, 1995; Hollier et al., 1997), simulations (e.g., Flanagan et al., 1998; Hendrix and Barfield, 1996); and (iv) commercial aviation (e.g., Begault, 1993). Among the various applications, VHSS entertainment systems have been the first to reach the shelves of consumer electronics stores. For example, the "virtual Dolby^{TM} 5.1 digital" headphone system (model: MDR-DSS00000) and the "dts^{TM} virtual 5.1" headphone system (model: MTDR-DS5100) from Sony^{TM} Ltd. (Sony, 2004) have been released for sale to consumers. In Section 1.3, brief details of a VHSS system are reviewed and the motivation and the aims of this study are reported.

1.3. Benefits and challenges of VHSS systems and the motivations for this study

Reproducing realistic surround sound in a cost-effective way has been a challenge for the audio engineering and entertainment industry. The two most popular surround-sound systems for digital video disc (DVD) home entertainment equipment are the Dolby^{TM} 5.1 digital system and the dts^{TM} surround 5.1 systems. Both systems simulate surround-sound effects by playing five channels of appropriately recorded sound from five separate speakers located at the centre-front, left-front, right-front, left-surround, and right-surround positions relative to a listener (Dolby^{TM}, 2004). Although both systems have a sixth sound channel containing very low frequency (VLF) content (below 100 Hz), the incident angle of this VLF sound is not critical as humans do not use such low-frequency sound for sound localization (Blauert, 1997). Both the Dolby^{TM} 5.1 digital system and the dts^{TM} surround 5.1 system require listeners to place the five speakers physically apart at appropriate positions. In situations where positioning the five speakers is difficult (e.g., watching a movie on a passenger plane or from a portable personal DVD player), the five channels (plus the VLF channel) of the Dolby^{TM} system are down mixed into two channels called the Dolby^{TM} stereo system. For example, the left channel of a Dolby^{TM} stereo is the sum of (i) half of the centre-front channel, (ii) the left-front channel, (iii) the left-surround channel, and (iv) the VLF channel reduced by 10 dB (Flanagan, 2000). This down mixing procedure removes the front–back directional information of the five surround-sound channels and can degrade the surround-sound effects. With non-individualized HRTFs, an alternative to the Dolby^{TM} stereo system can be developed. This alternative is called the VHSS channel. To make the VHSS channel, each of the five surround-sound channels are processed by a pair of appropriate HRTFs so that they are transformed into five binaural sound tracks carrying the content of the sound with appropriate directional information. When the five binaural sound tracks and the VLF channel are played through the same pair of headphones, the directional information embedded in the binaural signals can simulate a surround-sound effect.

Reducing the cost of a VHSS system without reducing the quality of its performance is a challenge (Wong and Choi, 2003). Reducing the number of coefficients of HRTFs involved in a VHSS system is the most direct way to reduce the cost because this will reduce the rate of computational operations and a cheaper digital signal processing (DSP) integrated circuits (IC) can be used. This chain of relationships is still valid even when the absolute cost of a DSP IC is decreasing over time. Therefore, there is long-lasting value in optimizing the number of coefficients of HRTFs in a VHSS system against its surround-sound performance.

A review of the signal-processing literature suggests that reducing the number of coefficients used to represent a transfer function will reduce its spectral details (Oppenheim et al., 1997). In other words, a reduction in the coefficient in an HRTF will reduce the details of the pinna response information that it can carry. Past studies have shown that when the pinna response information was completely removed from a pair of HRTFs (i.e., HRTFs with a flat spectral response), binaural cues generated from this pair of HRTFs did not accurately indicate whether the sound source was coming from the front or the back (e.g., Blauert, 1997; Wightman and Kistler, 1997; Begault and Wenzel, 1993; Oldfield and Parker, 1984; Wenzel et al., 1993). This suggests that when the number of non-zero coefficients of a pair of HRTFs is reduced to one, their associated binaural cues will lose the front–back directional information. The detailed spectral patterns of HRTFs responsible for a frontal cue or a backward cue have been presented by Blauert (1997) and will not be repeated here. The aims of this study are two-fold: (i) to study the accuracies of binaural cues generated using non-individualized HRTFs of reduced coefficients; and (ii) to develop a VHSS system with optimal complexity that and can provide better rated surround-sound effects than the Dolby^{TM} stereo system produces. In Section 1.4, relevant past studies are reviewed.

1.4. Previous studies on the effects of reducing the number of coefficients of non-individualized HRTFs

The physical effects of reducing the number of coefficients in a non-individualized HRTF correspond to a reduction in the levels of spectral details of the corresponding HRTF. A review of the literature shows that there are only a few studies that have varied the levels of spectral details in HRTFs and have examined the effects on their abilities to produce accurate binaural cues. Asano et al. (1990) tested the localization performance of binaural cues generated from HRTFs of simplified spectra. They used four different orders of an auto-regressive moving-average model to simplify the spectra of HRTFs. Eighteen directions ranging from 0° to 180° in the upper hemisphere of the median plane were tested. They observed an increase in the occurrence of front–back confusion as the degree of simplification increased. However, only two participants were used and no statistical conclusion was drawn. Begault (1992) compared the localization performance of simulated binaural cues generated using a set of HRTFs with 512 coefficients and a set of HRTFs with only 65 coefficients. In contradiction to the study by Asano et al. (1990), Begault did not find any significant differences in the localization performance between the two sets of binaural cues. Thirty-three participants and 12 directions at the ear-level (30° interval from 0° to 360° azimuth angles) were used in Begault's study. In 1998, Kulkarni and Colburn reported that four participants were not able to distinguish a real binaural cue from a simulated binaural cue when the simulated binaural cue was generated using individualized HRTFs of 32 or more coefficients. However, this study involved only four listeners and results of statistical tests were not reported. Also, this study did not test non-individualized HRTFs. In summary, there are only a few studies investigating the effects of the spectral details of HRTF impulse responses and only one study used more than four participants. This current study examines the effects of reducing HRTF coefficients to smaller than 65 (e.g., 64, 32, and 18 coefficients) with a statistically significant sample of listeners.

2. Experiment 1: reducing the number of coefficients in non-individualized HRTFs

2.1. Objectives and hypotheses

Experiment 1 examined the accuracies of binaural cues generated from non-individualized HRTFs of reduced numbers of coefficients from 128 to 64, 32, and 18. The binaural cues of the same directions but generated from HRTFs with different coefficients were calibrated to contain the same levels of ITDs and ILDs. This eliminates the confounding effects of ITDs and ILDs from the effects of the HRTF coefficients. It was hypothesized that reducing the number of HRTF coefficients would not affect the accuracies of cues at the ear-level from azimuth angles of 90° and 270° (H1). This hypothesis was based on the logic that reducing the number of coefficients in HRTFs will flatten the HRTFs' spectra and, hence, reduce the pinna response information that is essential for constructing a binaural cue coming from the front or the back (see Section 1). Following this logic, the accuracies of cues with directions coming from the front (i.e., cues at ear-level with incident azimuth angles of 0°, 45°, and 315°) or the back (i.e., cues at ear-level with incident azimuth angles of 180°, 135°, and 225°) were hypothesized to decrease when the HRTF coefficients were reduced (H2). The percentages of front–back confusion and back–front confusion were also hypothesized to increase as the HRTF coefficients were reduced (H3).

2.2. Methods and design

2.2.1. Participants

Twenty-four participants (12 males, 12 females) served as volunteers for this experiment. All of them were university students between 19 and 27 years old and all had passed the audiometric test for normal hearing. The audiometer used was a Voyager 522 made by Madsen Electronic (Copenhagen, Denmark). The participants were not informed about the specific objective of the experiment but were asked to localize a series of binaural sound cues. The participants had also been screened for the absence of noticeable hearing loss, recent exposure to loud noise, and a medical history of hearing problems (Begault and Wenzel, 1993).

2.2.2. Stimuli: binaural sound cues

The open-copyrighted MIT KEMAR non-individualized HRTF data (Gardner and Martin, 1995) were used for this experiment because these HRTFs were well documented and have been made freely available to researchers. The original HRTF data had 128 coefficients and they were re-sampled in the frequency domain to produce HRTFs with 64, 32, and 18 coefficients. These HRTFs can be downloaded from the Internet (http://sound.media.mit.edu/KEMAR.html). HRTFs of 18 coefficients were used instead of 16 coefficients because data from pilot experiments indicated that HRTFs with 16 coefficients produced binaural cues with noticeable distortion. In total, 32 pairs of HRTFs were generated, representing combinations of the four numbers of coefficients (128, 64, 32, 18) and the eight pre-defined cue directions (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315° azimuth angles at ear-level). These 32 HRTFs were then convolved with a mono sound clip of human speech lasting 10 s to generate the corresponding binaural cues. The mono clip contained two sentences. The first sentence was "thank you for your participation in this experiment" spoken in English; the second sentence was this sentence translated into Cantonese, the native language of the participants. In addition to the eight cue directions, a mono-sonnolrik (i.e., no HRTF filtering) direction was added as the ninth cue direction. In total, there were 36 binaural cue sound stimuli (i.e., a full-factorial combination of nine cue directions by four levels of HRTF coefficients).

2.2.3. Procedure and apparatus

All the signal-processing procedures (e.g., re-sampling of HRTFs and convolution between HRTFs and the original mono sound clips) were done in the Matlab^{TM} programming environment (Version 5.3 by MathWorks, Natick, MA, USA). In one experimental session, the 36 binaural cues were presented randomly to seated participants through a pair of headphones (HD545, Sennheiser Ltd., Hannover, Germany). Each participant took part in two experimental sessions with a 5-min break between them. The randomized order of the cues was not repeated across repetitions or participants. The HD545 Sennheiser headphone was used because the MIT KEMAR non-individualized HRTF data had been equalized to the headphone's responses (Gardner and Martin, 1995). After listening to each sound cue, the participants were then required to indicate the incident azimuth angle of the sound cue and its perceived relative distance to the centre-of-the-head on a graphical interface. The graphical interface was adopted from Braasch (2001). In this study, the main dependent variables of interest were the perceived azimuth directions of the binaural sound cues. The relative positions of the perceived sound sources to the centre-of-the-head were measured as controls. When participants reported that the sources were located at the centre-of-the-head, the reported perceived azimuth incident directions became meaningless. More details about this control measure are provided in Section 2.3.2. The participants received no feedback on their performance during the entire experiment. Before the start of the experimental sessions, each participant was given a chance to practice with the apparatus and the experimental procedure but no feedback on the accuracy of their sound localization performance was given. The experiment was conducted inside an acoustic chamber (a custom-made 1400-A-CT series chamber from Industrial Acoustic Company Inc., New York, USA) with a background noise level of about 30 dBA. In summary, the three independent variables were as follows: (i) cue directions with nine levels, (ii) spectral complexity of HRTFs with four levels, and (iii) repetition with two levels.

2.3. Results, data analyses, and discussions

2.3.1. Effects of the repeated run

A sound localization error for a particular binaural cue is defined as the angular differences (in absolute value) between the perceived azimuth angles and the cue direction. In this study, non-parametric statistical tests were used because the data did not follow a normal distribution. Results of Wilcoxon signed rank tests showed that there was no significant difference between data collected in the two periods ($p < 0.2$, Wilcoxon test). As a result, data from the two replications were combined in subsequent analyses.

2.3.2. Exclusions of data

In this study, the centre-of-the-head was defined as an inner sphere with a radius corresponding to 20% of the radius of the head. The rationale was that if a sound source was perceived as located within an inner sphere of 20% of the radius of a head, the perceived direction of that sound cue was not accurate. When the data collected under the mono-sonnolrik condition were excluded, the occurrence of incorrect perceptions of the sound source location in the centre-of-the-head was less than 5% of the total data. These 5% incorrect perceptions were mainly associated with cues from 0° and 180° azimuth directions. As explained in Section 2.2.3, these 5% incorrect perceptions as well as the data obtained from the mono-sonnolrik condition were excluded from subsequent analyses.

2.3.3. Effects of reducing HRTF complexity on sound localization errors and their interactions with cue directions

Fig. 1 illustrates the median sound localization errors as functions of cue directions and levels of HRTF complexity represented by the numbers of HRTF coefficients. The authors observed that there were large individual variations in the data and therefore used statistical tests to verify any repeated effect. Only those effects that are statistically significant are reported. Fig. 1 suggests that binaural cues with 180° and 0° azimuth angle directions were associated with larger localization errors than were other cues. Results of a Kruskal–Wallis one-way analysis of variance (ANOVA) confirm that the choice of cue direction had significant effects on the sound localization errors ($p<0.001$, Kruskal–Wallis one-way ANOVA). The main objective of this study was to examine the effects of reducing the coefficient of the HRTF. Results of the Kruskal–Wallis one-way ANOVA indicate that the reduction of HRTF complexity significantly affected the localization errors ($p<0.001$, Kruskal–Wallis one-way ANOVA). Fig. 1 suggests that there could be interactions between the effects of HRTF complexity and cue directions. Kruskal–Wallis one-way ANOVAs were used to test for the effects of HRTF complexity under conditions with the same cue directions. Results indicated that for the 0°, 90°, and 270° azimuth angles, reducing HRTF complexity had no significant effect on localization errors ($p<0.2$). For the 45°, 135°, 180°, 225°, and 315° azimuth angles, reducing HRTF complexity resulted in significant changes in localization errors ($p<0.01$, Kruskal–Wallis one-way ANOVA). For data obtained under the same cue direction, Mann–Whitney U-tests were utilized to compare the localization errors collected from each level of HRTF complexity. The results were used to group the conditions such that errors collected from experimental conditions within the same grouping were not significantly different from each other at the $p<0.05$ level. The grouping results are presented in Table 1 where the median errors collected from conditions with the same cue direction are listed in the same row. In each row, the condition with the highest error is listed on the left. Table 1 indicates that reducing the complexity of the non-individualized HRTFs from 128 to 18 coefficients did not significantly affect the localization errors for the 90° and 270° cue directions. This finding supports the first hypothesis (H1). For cues in the front of the participants (i.e., cues with azimuth angles of 0°, 45°, and 315°), reducing HRTF complexity from 128 to 32 coefficients or to 18 coefficients significantly increased sound localization errors with the exception of the 0° condition ($p<0.05$, Mann–Whitney U-tests, Table 1). This result partially supports the second hypothesis (H2). Fig. 1 shows that the errors collected for the 0° condition cover a large inter-quartile range, which explains the lack of significant effects of HRTF complexity. For cues coming from the back of the participants (i.e., cues with azimuth angles of 135°, 180°, and 225°), reduction of HRTF complexity to 18 coefficients from 128 coefficients resulted in significant increases in localization errors except when the cue direction was 180° ($p<0.05$, Mann–Whitney U-tests, Table 1). For the 180° condition, binaural cues generated from the HRTFs with 18 coefficients were associated with the lowest localization errors. More discussion on the 180° condition is presented in Section 2.4.

Fig. 2 illustrates the effects of HRTF complexity on localization errors with cues coming from the front, back, and sides of the listeners. It can be observed that the localization errors obtained with the cues coming from the side directions (i.e., 90°, 270° azimuth angles) were not affected by the changes in HRTF complexity. This is

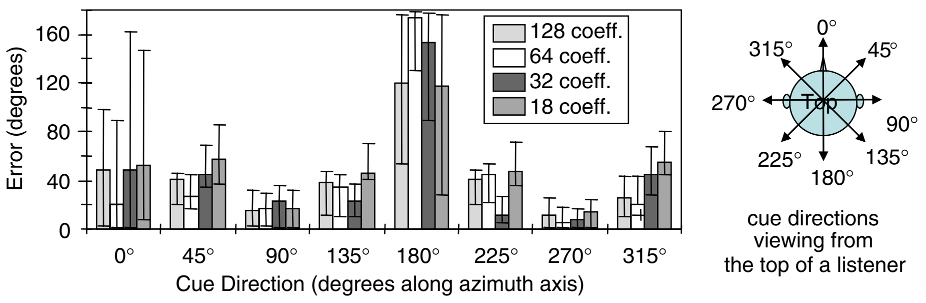

Fig. 1. Median localization errors for eight sound directions and four levels of HRTF complexity (data from 24 participants). Error bars representing the inter-quartile ranges are also shown. The directions of the sound are illustrated using a top-view of a listener's head.

Table 1
Median localization errors for four levels of HRTF complexity (numbers underlined or covered by the same line are not significantly different from each other, Mann–Whitney tests)

Direction	Median localization errors with HRTFs of different coefficients (ranked)						
0°	52 (18 coeff.)	>	49 (128 coeff.)	>	48 (32 coeff.)	>	21 (64 coeff.)
45°	58 (18 coeff.)	>	45 (32 coeff.)	>	41 (128 coeff.)	>	27 (64 coeff.)
90°	23 (32 coeff.)	>	17 (64 coeff.)	>	16 (18 coeff.)	>	15 (128 coeff.)
135°	46 (18 coeff.)	>	38 (128 coeff.)	>	35 (64 coeff.)	>	23 (32 coeff.)
180°	174 (64 coeff.)	>	153 (32 coeff.)	>	120 (128 coeff.)	>	118 (18 coeff.)
225°	47 (18 coeff.)	>	45 (64 coeff.)	>	41 (128 coeff.)	>	12 (32 coeff.)
270°	14 (18 coeff.)	>	12 (128 coeff.)	>	8 (32 coeff.)	>	5 (64 coeff.)
315°	55 (18 coeff.)	>	45 (32 coeff.)	>	25 (128 coeff.)	>	20 (64 coeff.)

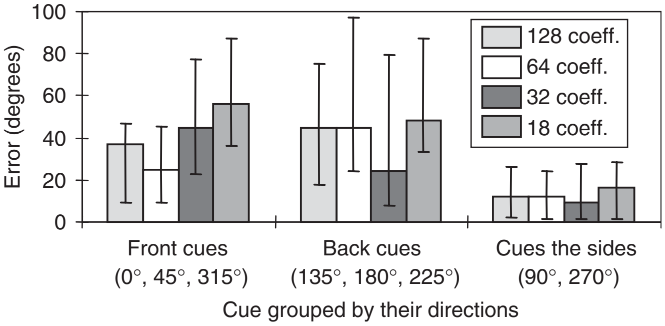

Fig. 2. Median localization errors for regional directions (front, back, and side) and four levels of HRTF complexity (data from 24 participants).

confirmed by statistical results ($p = 0.785$, Kruskal–Wallis tests). However, significant effects of HRTF complexity on errors obtained with cues from the front and back directions were found ($p<0.001$, Kruskal–Wallis tests). One possible explanation for these results is that the changes in HRTF complexity mainly affected the front–back confusion errors and these errors do not occur when the cues are coming from the sides of a participant. This explanation supports H3. In order to test the effects of HRTF complexity on front–back confusion errors, percentages of front–back (or back–front) confusion were calculated. In this experiment, for the same level of HRTF complexity, each participant listened twice to three different cues coming from the front (i.e., 0°, 45°, and 315° azimuth angles) and the percentage of front–back confusion for each listener was the percentage of occurrence of front–back confusion among these six listening trials. Similarly, the percentage of back–front confusion was the percentage of occurrence of back–front confusion when listening to the six cues coming from the back (i.e., 135°, 180°, and 225° with two repetitions). The median and inter-quartile ranges of these percentages are plotted as functions of HRTF complexity in Fig. 3. Results of Friedman two-way analyses of variances show that as the number of the HRTF coefficients was reduced, the binaural cues produced were associated with significantly higher percentages of front–back confusion ($p<0.001$) while the percentages of back–front confusion remain unchanged. This suggests that H3 is only true for front–back confusion and not true for back–front confusion. As discussed above, localization errors collected for cues simulating the 180° azimuth direction did not follow the

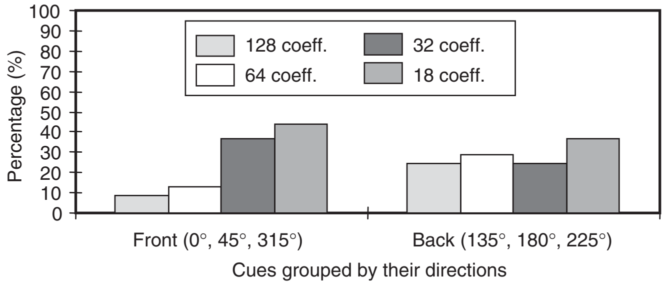

Fig. 3. Percentages of front–back (back–front) confusion for regional directions (front and back) and four levels of HRTF complexity (data from 24 participants).

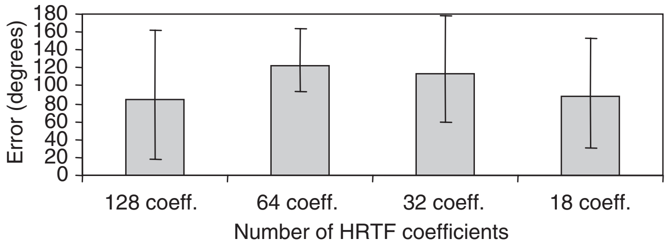

Fig. 4. Median localization errors for four levels of HRTF complexity from the 180° azimuth direction collected from six participants in the small-scale supplementary experiment.

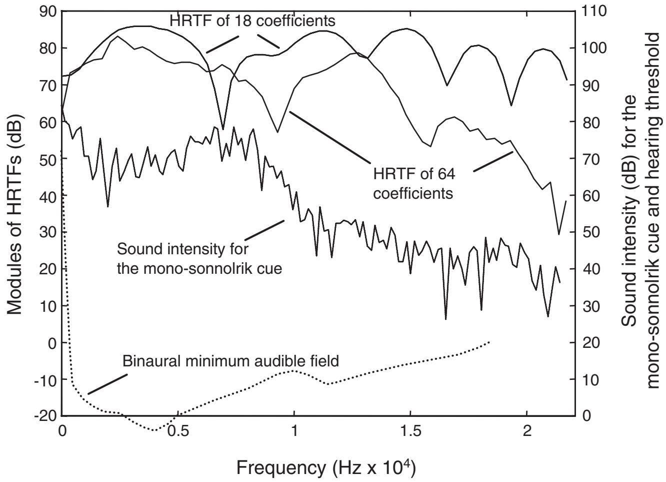

Fig. 5. Modules of HRTFs for the incident angle of 180° azimuth with 64 coefficients (—) and 18 coefficients (- - - -). The sound intensity of the mono-sonnolrik cue (–––) and the binaural minimum audible field adapted from Robinson and Dadson (1957) (----). The modules of the two HRTFs were obtained using the transfer function methods by passing white noise through the respective HRTF filters.

prediction of H2. Reducing the complexity of the HRTFs from 64 to 18 coefficients surprisingly reduced both the localization errors ($p<0.05$, Mann–Whitney U-tests, Table 1) and the percentages of back–front confusion (from 66.7% to 39%) with a cue direction of 180°. This finding does not support H2 and H3. A small-scale experiment conducted to explore this effect further is reported in Section 2.4.

2.4. Effects of HRTF complexity on the localization of cues with an incident angle of 180° azimuth: a follow-up experiment

2.4.1. Aim

Table 1 shows that reducing the complexity of HRTFs used to produce the cues from 64 to 18 coefficients reduced the localization errors significantly for cues simulating the 180° azimuth direction ($p<0.05$, Mann–Whitney U-tests). This result is opposite to the prediction of the second hypothesis (H2) and an additional small-scale experiment with six participants was conducted to verify this finding.

2.4.2. Method

Six university students who did not take part in Experiment 1 were randomly invited to participate in this follow-up experiment. Similar to the experimental procedures in Experiment 1, the participants were asked to listen to binaural cues and then to indicate their perceived sound source positions. A total of 12 binaural cues (one 180° direction × four levels of HRTF complexity × three repetitions) were presented to each participant. It was hypothesized that (i) the finding of Experiment 1 would be repeatable. In other words, for the 180° azimuth direction, the binaural cue produced by the 18-coefficient HRTFs would be more accurate than that produced by the 64-coefficient HRTFs (hypothesis H4).

2.4.3. Results

Fig. 4 shows the median localization errors as functions of the four levels of HRTF complexity for the 180° azimuth direction condition. Figs. 2 and 4 indicate that the effects of HRTF complexity on localization errors were similar in both experiments: HRTFs of 128 and 18 coefficients were associated with the lowest errors and the HRTF of 64 coefficients was associated with the highest errors. This supports H4 and verifies the repeatability of Experiment 1. In order to investigate the significant reduction of localization errors when HRTFs with 18 instead of 64 coefficients were used to produce binaural cues from the 180° azimuth direction, the modules of HRTFs with different coefficients were examined (Fig. 5). The sound intensity of the un-filtered stimuli and the minimum binaural audible level are also shown in the figure. The frequency responses of HRTFs were calculated using the transfer function method in Matlab$^{\mathrm{TM}}$ (Jackson, 1996). Fig. 5 indicates that the modules of HRTFs with 64 and 18 coefficients have peaks and notches at different frequencies. A review of the literature indicates that binaural cues with energies between 0.7 and 1.75 kHz and between 9.5 and 14.4 kHz are more likely to be perceived as coming from the back (Blauert, 1969/70). Fig. 5 shows that the 18-coefficient HRTF had a larger magnitude than the 64-coefficient HRTF in both frequency ranges. This could be the reason why the binaural cue from the 180° azimuth direction produced by the 18-coefficient HRTF had significantly fewer localization errors than that produced by the 64-coefficient HRTF. In addition, Hebrank and Wright (1974) suggested that a band-stop response between 7 and 11 kHz is a characteristic of HRTFs from frontal directions. It can be observed from Fig. 5 that the 64-coefficient HRTF has a notch between 7 and 10 kHz while the 18-coefficient HRTF does not. This may explain why more participants perceived the 180° azimuth binaural cues generated by the 64-coefficient HRTF as from the front (the percentages of back–front confusion for binaural cues of the 180° azimuth angle produced by the 18- and 64-coefficient HRTFs are 40% and 68%, respectively). For the 18-coefficient HRTF, the notch occurs at 6.7 kHz and falls outside the band-stop frequency range (7–11 kHz) suggested by Hebrank and Wright (1974) for frontal cues. This finding is very interesting because it suggests that an accurate binaural cue from 180° can be generated using HRTFs with just 18 coefficients as long as the spectra of the HRTFs meet certain criteria.

2.4.4. *Conclusions of the follow-up experiment*

The results of the follow-up experiment are consistent with those of Experiment 1. In other words, for the 180° azimuth direction, the binaural cue produced by the 18-coefficient HRTFs was more accurate than that produced by the 64-coefficient HRTFs. The frequency spectra of the corresponding HRTFs suggest that cues generated using the 18-coefficient HRTFs were more accurate because the 18-coefficent HRTFs have spectral peaks and notches at the 'appropriate' frequencies as suggested by Blauert (1969/70) and Hebrank and Wright (1974) for HRTFs carrying information to generate backward binaural cues.

3. Experiment 2: usability testing of a prototype VHSS system

3.1. *The development of a prototype VHSS system using the results obtained from Experiment 1*

A prototype VHSS system was developed based on the results from Experiment 1. This system comprises the following parts: a Motorola DSP evaluation development board (DSP56362EVM, Motorola Ltd., Austin, USA) and a pair of headphones (HD545, Sennheiser Ltd., Hannover, Germany). The system takes in the digitally coded surround-sound data in AC3 format from a DVD player and outputs a stereo channel of virtual surround sound. When an AC3 digital signal is fed into the Motorola DSP56362EVM evaluation development board, it is decoded into six digital surround-sound signals according to the Dolby™ 5.1 standard (centre-front, left-front, right-front, left-surround, right-surround, and subwoofer, Dolby™, 2004). Each of the decoded surround-sound signals except for the subwoofer signal is then convolved with one appropriate pair of HRTFs tested in Experiment 1. The HRTF pairs add the appropriate directional information to the five decoded surround-sound signals. The directions of the HRTF pairs used in the prototype are given in Table 2. The outputs of the 10 convolution processes and the un-filtered subwoofer (after synchronization) are then combined into a left channel and a right channel. The subwoofer signal (below 100 Hz) was not convolved with any HRTFs because humans do not use such low-frequency sound to localize binaural cues (Blauert, 1997). The combined left and right channels are then presented to listeners via a pair of headphones. The objective of the VHSS system is to simulate the experience of listening to a Dolby™ 5.1 surround-sound track playing from the six surround speakers (i.e., centre-front, left-front, right-front, left-surround, right-surround, and subwoofer speakers) appropriately placed relative to the participants (Sony™, 2004). The heart of this prototype VHSS system is the five pairs of HRTFs whose directions are listed in Table 2.

Table 2
Directions of HRTF pairs used to simulate the five surround sound channels in the prototype VHSS system

Surround sound levels	Direction of HRTF pairs (degree azimuth at ear-level)	Complexity of HRTFs as indicated by the number of non-zero coefficients
Centre-front	0	32
Left-front	315	64
Right-front	45	64
Left-surround	270	32
Right-surround	90	32

From the results in Experiment 1, HRTFs with 64 coefficients were used to synthesize the right-front (i.e., 45° azimuth) and left-front (i.e., 315° azimuth) channels. The binaural cues generated with these coefficients resulted in significantly fewer localization errors than did the corresponding cues generated by HRTFs with 32 and 18 coefficients ($p < 0.001$, Kruskal–Wallis tests). HRTFs with 128 coefficients were not used because they demanded higher computational resources than HRTFs with 64 coefficients as well as offering no significant improvement in performance. Results on the centre-front (0°), right-surround (90°), and left-surround (270°) channels suggested that HRTFs with reduced complexity did not have any significant effect on their accuracies ($p > 0.2$, Kruskal–Wallis tests). To minimize the computation complexity, a logical choice would be to use HRTFs with 18 coefficients in these positions. However, after consulting field experts in the audio industry, we decided to use HRTFs with 32 coefficients for fear that HRTFs with 18 coefficients might introduce undesirable frequency distortion although there was no evidence of this in Experiment 1 (Wong and Choi, 2003). Because this prototype was developed for the industry, advice from field experts was given much weight. In summary, the trade-off in choosing the numbers of HRTF coefficients used in the prototype was between the levels of complexity of HRTFs (directly related to the cost of manufacturing) and the performance of the resulting sound cues.

3.2. Objectives and hypotheses of Experiment 2

This experiment was a double-blind usability test conducted to compare the surround-sound effects of the Dolby^{TM} stereo system and the prototype VHSS system. It was hypothesized that the outputs of the VHSS system would provide better surround-sound effects than would the Dolby^{TM} stereo outputs (H5). In the double-blind design, both the participants and the experimenter did not know the order of the presentation of conditions. This eliminated any bias due to brand-name preference. In this experiment, the number of participants with and without musical training (e.g., playing musical instruments or singing in a choir) was controlled to be the same and it was hypothesized that a musical background would not affect the surround-sound ratings (H6).

3.3. Methods and design

3.3.1. Participants

Thirty-two participants served as volunteers for this experiment. As the authors had no reason to expect that gender would make a difference in the ability to appreciate surround sound, there was no attempt to balance the number of male and female participants. There were 16 male and 24 female participants. Sixteen of the 32 participants were experienced in playing musical instruments or singing in a choir for more than 3 years while the other 16 had no musical experience. The experience of playing musical instruments or singing in a choir was measured by a pre-exposure questionnaire. The participants were university students between 19 and 27 years old. All of them passed the audiometric test for normal hearing. The audiometer used was the same as that used in Experiment 1. The participants were not informed about the real purpose of the experiment but were instructed to give ratings on the surround-sound effects of a series of movie clips.

3.3.2. Apparatus and stimuli

The VHSS prototype system described in Section 3.1 was used to deliver the virtual surround-sound condition while the Dolby^{TM} stereo system with a Panasonic (DVD-A560EN) DVD player was used to deliver the Dolby^{TM} stereo condition. Two movie samples with Dolby^{TM} 5.1 digital surround-sound effects were used as the raw stimuli: (i) the Dolby^{TM} Surround 5.1 channel demonstration from the Dolby^{TM} demonstration DVD for part 1 of the experiments (Hong Kong and Taiwan Chinese Productions, 1999) and (ii) selected clips from the movie Air-Force One^{TM} DVD (Bernstein et al., 1999) for part 2 of the experiment. The former movie clip contained a 2-min test music sequence illustrating music played from each of the six surround-sound channels in sequential order (i.e., left-front, centre-front, right-front, left-surround, right-surround, and subwoofer). The music was accompanied by movie shots illustrating the correct spatial direction of the appropriate speaker. The latter movie clip contained a 4-min selected air-to-air combat scene taken from the movie Air-Force One^{TM}. In both movie clips, videos containing visual cues that were appropriate for the audio directional cues were presented on a TV.

3.3.3. Procedure and surround-sound questionnaire

During the experiment, participants were instructed to wear a set of headphones (HD545, Sennheiser Ltd., Hannover, Germany) and sit in front of a TV. The volume levels of the sound clips presented from the VHSS prototype system and the Dolby^{TM} stereo channels were calibrated to have the same r.m.s. magnitude levels.

The experiment was divided into four sessions. In the first session, participants watched and listened to the DVD demonstration movie clips twice: one with the Dolby^{TM} stereo sound track and one with the VHSS two-channel virtual surround-sound track. The order of presenting the two tracks was randomized. The visual component of the movie was the same for both audio presentation conditions. After watching and listening to the presentations, participants were asked to complete a questionnaire asking them to compare the surround-sound effects between the two sound tracks (see Appendix A1). In the second session, the procedure was the same except that selected clips from the movie Air-Force One^{TM} were used instead of the DVD demonstration clip and a second questionnaire was used (see Appendix A2). Both questionnaires were developed based on the semantics used to describe surround-sound effects extracted using the repertory grid technique (Berg and Rumsey, 1999). Each of the two questionnaires was customized to suit the context of the two movie clips, namely the DVD demonstration and the Air-Force One^{TM} film. The first one focused more on the perceived direction of the sound and the second questionnaire focused more on the overall surround-sound effects. The first clip contained incident sound of speakers placed at known locations relative to the participant while the second clip contained movie scenes of missiles, planes, as well as shrapnel flying from left-to-right, right-to-left, back-to-front, and front-to-back. The third and fourth sessions were repeats of the first two sessions.

3.4. Results and discussion

The Cronbach α values of the internal consistencies were calculated for the data obtained using the two questionnaires; both values were greater than 0.75, indicating that the questionnaires were reliable (Miller, 1995). One interesting finding was that when the answers to Questions QA1.2 and QA2.9 were deleted, the α values increased for both questionnaires. This indicates that the perception of the sound source being 'inside-my-head' was not correlated with the perception of 'surround-sound effects', '3D sound

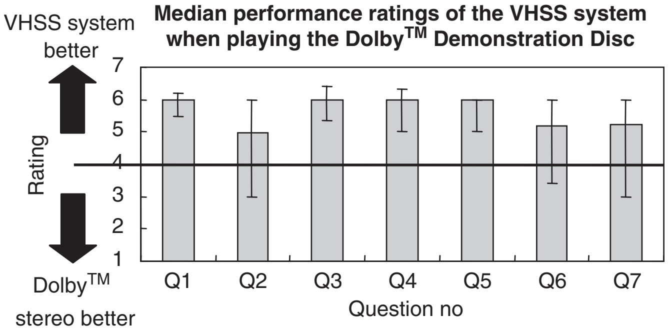

Fig. 6. Median performance ratings of the VHSS system when playing the Dolby™ Demonstration Disc (data from 32 participants).

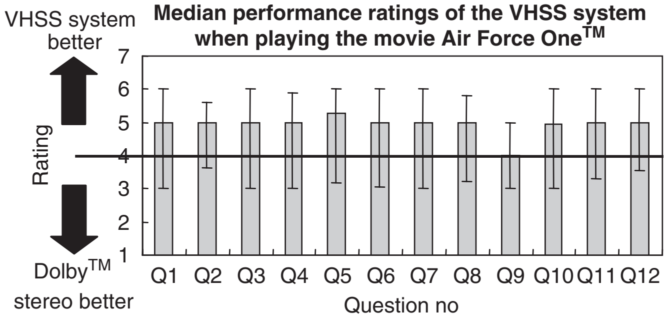

Fig. 7. Median performance ratings of the VHSS system when playing the movie Air Force I (data from 32 participants).

effects', and 'feeling of presence'. As the data did not follow a normal distribution, non-parametric statistical tests were used. Results of a Wilcoxon signed ranks test showed that there was no significant difference between data collected in the two replications ($p>0.1$, *Wilcoxon test*). Since the order of presenting the Dolby™ stereo version and the VHSS version were randomized, data were reorganized so that a rating of greater than four meant that the VHSS version was considered better and a rating of less than four meant that the Dolby™ stereo version was deemed better. The median data are shown in Figs. 6 and 7 and it can be observed that participants reported better surround-sound ratings for the VHSS system than for the Dolby™ stereo system on all questions except Question QA2.9. Results of statistical tests indicate that the ratings were significantly greater than '4' for all 17 questions except for Questions QA1.2 and QA2.9 ($p<0.03$, Wilcoxon signed ranks test). This suggests that, in this study, the VHSS system achieved significantly higher surround-sound ratings than did the Dolby™ stereo system. The bandwidth of the tested clips may affect the directional accuracy of the surround-sound cues (e.g., Blauert, 1997; Begault, 1994). In other words, if a test clip containing only a very narrow bandwidth was used, the significant difference in the surround-sound ratings between the two systems may be reduced. Because the two selected movie clips, namely the DVD demonstration and the Air-Force One™ film, are typical sound clips that listeners will hear while they are watching DVDs with surround-sound tracks, the findings of this study can be generalized to test clips with embedded surround-sound tracks.

The performance ratings obtained from participants with and without musical training were compared and no significant difference was found for ratings given to the Dolby™ demonstration DVD ($p>0.19$, *Wilcoxon signed ranks tests*). This result supports H6. For the ratings given to the Air-Force One™ movie, the effect was marginally significant ($p=0.1$, Wilcoxon signed ranks tests). Further analyses showed that participants without musical experience gave significantly higher ratings than those with musical training on Questions QA2.3, QA2.5, QA2.7, QA2.10, and QA2.12 ($p<0.05$, *Wilcoxon signed ranks tests*). Further work is needed to determine the reason(s) for these differences. No significant effect was found for all ratings ($p>0.3$, *Wilcoxon signed ranks tests*).

4. Conclusions

Non-individualized HRTFs can produce binaural direction cues. Reducing the number of non-zero coefficients in HRTFs from 128 to 64, 32, or 18 did not significantly affect the perceived directions of the corresponding binaural cues when the simulated directions were at ear-level with azimuth angles of 0°, 90°, and 270° ($p>0.2$). This suggests the possibility of producing binaural cues at centre-front, left, and right directions using HRTFs with coefficients less than 128 without affecting their perceived accuracies.

For the directions of 45°, 135°, 225°, and 315° azimuth angles at ear-level, reducing the number of HRTF coefficients from 128 to 64 also did not have a significant effect. However, when the number of coefficients was reduced from 128 to 32 and then to 18, the corresponding binaural cues were associated with significantly higher localization errors ($p<0.05$). Results suggest that it is possible to produce binaural cues at front-left (315° azimuth), front-right (45° azimuth), back-left (225° azimuth), and back-right (135° azimuth) directions using HRTFs with 64 coefficients rather than 128 coefficients without affecting their perceived accuracies.

For binaural cues simulating the direction of a 180° azimuth angle at ear-level, reducing the number of coefficients from 128 to 18 produced a non-linear change in the cue accuracies. As the number of coefficients was decreased from 128 to 64 and then to 32, the corresponding binaural cues became less accurate and led to significantly higher localization errors. However, when the number of coefficients was reduced to 18, the corresponding cues became more accurate and led to significantly lower localization errors than did the cues generated with HRTFs with 64 coefficients. This surprising finding was repeated in an additional experiment. Spectral analyses of the 18-coefficient HRTFs indicated that an HRTF with a small number of coefficients can still produce an accurate binaural cue as long as it contains the appropriate spectral variations in some key frequency ranges. This explanation is consistent with the current understanding of spatial hearing (Begault, 1994; Blauert, 1997).

When the number of HRTF coefficients was reduced, the binaural cues produced were associated with significantly higher percentages of front–back confusion ($p<0.001$) while the percentages of back–front confusion remained unchanged. The significant increases in the front–back confusion rates when the HRTF coefficients were reduced are consistent with the theoretical prediction that reducing the number of HRTF coefficients reduces the pinna response information that is carried by the resulting binaural sound cues. The absence of significant changes in back–front confusion rates when the number of HRTF coefficients was reduced was due to the special case with cues coming from the back (i.e., 180° azimuth) as reported above.

Based on the results from Experiment 1, a prototype virtual headphone-based surround-sound (VHSS) system was developed. This system used non-individualized HRTFs of 32, 64, 64, 32, and 32 coefficients to simulate the appropriate directional information for the centre-front, left-front, right-front, left-surround, and right-surround-sound channels, respectively. A double-blind experiment indicated that the VHSS prototype system produced better surround-sound effects than did the Dolby^{TM} stereo channels ($p<0.05$).

This study demonstrates the benefits of applying ergonomics techniques in the design and development of audio systems. Novel data are reported. These data have potential applications in virtual surround-sound products.

5. Limitations and future work

The sound localization data collected in this study showed many individual variations. Although all participants were randomly selected and all reported significant effects have been verified by statistical tests, further work to include more participants is desirable. In particular, studies focusing on the inter- and intra-subject variability in sound localization performance are needed.

The unexpected results associated with cues coming from the back (i.e., 180° azimuth at ear-level) suggest that it is possible to represent an HRTF with only 18 non-zero coefficients and generate accurate directional sound cues as long as the spectral peaks and notches of the HRTFs are consistent with those reported by Blauert (1969/70) and Hebrank and Wright (1974). Further exploration of the effects of manipulating the spectrum of HRTFs on the accuracy of the resulting binaural cues is needed.

Acknowledgements

The work was jointly supported by the Innovation and Technology 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 (HKUST 6219/02E) and its counter-part in Germany, DAAD (G-HK99/00.EG03, G_HK002/00E). The authors would like to thank the DSP engineers Fred Chan Yiu Tong, Darwin So Wan Yun, and Douglas R. George of the Consumer Media Centre at HKUST. Parts of the preliminary results of the first experiment were presented at the 8th International Congress on Sound and Vibration. The 128-coefficient non-individualized HRTFs used in this study were adopted from the MIT KEMAR www site and the authors are grateful for the availability of these open-copyrighted data (Gardner and Martin, 1995). Last but not least, the authors would like to thank Kenneth Wong and Raymond Choi from Valencetech Ltd. (HK) for their advice.

Appendix A

A1. Questionnaire used to compare the surround-sound performance of two audio-visual presentation of a DVD demonstration movie clips

QA1.1 I can detect the correct direction* of sound in the 1st version **MUCH EASIER** than in the 2nd version.

Strongly Agree	Agree	Slightly Agree	Neutral	Slightly Disagree	Disagree	Strongly Disagree
(1)	(2)	(3)	(4)	(5)	(6)	(7)

QA1.2 The sound in the 1st version appears to come from the "inside" of my head **MORE** than the sound in the 2nd version.

QA1.3 I can tell that the sounds from each of the five surround loudspeakers (as shown on the TV) were coming from different directions **MUCH EASIER** in the 1st version than in the 2nd version.

QA1.4 Sounds in the 1st version give a **MUCH BETTER** sense of direction (方向感) than the 2nd version.

QA1.5 The level of matching between the directions of the loudspeakers (as shown on TV) and the perceived directions of the corresponding sounds is **MUCH WORSE** in the 1st version than in the 2nd version.

QA1.6 On the whole, the 1st version gives me **BETTER** surround-sound effects* (環迴效果) than the 2nd version.
QA1.7 All in all, the 1st version gives me **BETTER** three-dimensional (3D) sound effects* (三維空間效果) than the 2nd version.

A2. Questionnaire used to compare the surround-sound performance of two audio-visual presentations of selected movie clips from the Air-Force One^{TM} movie

QA2.1 The 1st version sounds **MORE "Live" (現場感)** than the 2nd version.
QA2.2 I can distinguish sounds of different directions **MUCH EASIER** in the 1st version than in the 2nd version.
QA2.3 The 1st version of sound gives **MORE "Wide spatial effects"* (空間效果闊)** than the 2nd version.
QA2.4 Sounds in the 1st version appear to have a **CLEARER** direction than sounds in the 2nd version.
QA2.5 The 1st version gives **BETTER "Surround sound effects" (環迴效果)** than the 2nd version.
QA2.6 The 1st version of sound gives **BETTER "Feeling of Presence" (置身其中的感覺)** than 2nd version.
QA2.7 Sounds in the 1st version appear to give a **MORE "Narrow spatial effects" (空間效果窄)** than sounds in the 2nd version.
QA2.8 Sounds in the 1st version appear to surround me **MORE** than sounds in the 2nd version.
QA2.9 Sounds in the 1st version appear to originate **MORE** from **"Inside of your head"** than the sounds in the 2nd version.
QA2.10 The 1st version of sound gives **BETTER** feeling of being inside the movie environment (置身其中) at the position of the cameraman than the 2nd version.
QA2.11 The 1st version is **MORE** enjoyable than the 2nd version.
QA2.12 I can hear sounds coming from my back **MORE** in the 1st version than in the 2nd version.

References

Algazi, V.R., Duda, R.O., Thompson, D.M., Avendano, C., 2001. The CIPIC HRTF database. In: Proceedings of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics. Mohonk Mountain House, New Paltz, NY, October 21–24, 2001, pp. 99–102.

Asano, F., Suzuki, Y., Sone, T., 1990. Role of pinna responses in median plane localization. J. Acoust. Soc. Am. 88, 159–168.

Begault, D.R., 1992. Perceptual similarity of measured and synthetic HRTF filtered speech stimuli. J. Acoust. Soc. Am. 2334.

Begault, D.R., 1993. Head-up auditory displays for traffic collision avoidance system advisories: a preliminary investigation. Hum. Factors 35, 707–717.

Begault, D.R., 1994. 3-D Sound for Virtual Reality and Multimedia. AP Professional, Cambridge, MA.

Begault, D.R., Wenzel, E.M., 1993. Headphone localization of speech. Hum. Factors 35, 361–376.

Berg, J., Rumsey, F., 1999. Spatial attribute identification and scaling by repertory grid technique and other methods. In: AES 16th International Conference on Spatial Sound Reproduction, Germany.

Bernstein, A., Petersen, W., Katz, G., Shestack, J., (Producer), and Petersen, W. (Director), 1999. Air Force One [DVD video]. Buena Vista International, USA. Code No. 4-717415-850019.

Blauert, J., 196970. Sound localization in the median plane. Acustica 22, 205–213.

Blauert, J., 1997. Spatial Hearing: the Psychophysics of Human Sound Localization. Massachusetts Institute of Technology, London, England.

Blauert, J., 2000. Personal Communication.

Braasch, J., 2001. Auditory localization and detection in multiple sound-source scenarios. Ph.D. Thesis, Ruhr University, Bochum.

Dolby^{TM}, 2004. Dolby surround in the age of Dolby digital. 〈http://www.dolby.com/ht/surr-age.pdf〉.

Gardner, W.G., Martin, K.D., 1995. HRTF measurements of a KEMAR. J. Acoust. Soc. Am. 97, 3907–3908.

Greenhalgh, C., Benford, S., 1995. MASSIVE: a collaborative virtual environment for teleconferencing. ACM Trans. CHI 2, 239–261.

Hendrix, C., Barfield, W., 1996. The sense of presence within auditory virtual environments. Presence 5, 290–301.

Hollier, M.P., Rimell, A.N., Burraston, D., 1997. Spatial audio technology for telepresence. BT Technol. J. 15, 33–41.

Flanagan, P., 2000. Personnel communication via sa-sig@leeds.ac.uk. Product Manager. Professional and Research, Lake Technology Ltd., Australia.

Flanagan, P., McAnally, K.I., Martin, R.L., Meehan, J.W., Oldfield, S.R., 1998. Aurally and visual guided visual search in a virtual environment. Hum. Factors 40, 461–468.

Hebrank, J., Wright, D., 1974. Spectral cues used in the localization of sound sources on the median plane. J. Acoust. Soc. Am. 56, 1829–1834.

Hong Kong and Taiwan Chinese Productions, 1999. Demonstration: Digital Video/Audio System Demo I [DVD video]. Cine-Asia Entertainment Co. Ltd., Taiwan. Code No. 4-717415-880030.

Jackson, L.B., 1996. Digital Filters and Signal Processing: with Matlab^{TM} Exercises. Kluwer Academic Publishers, Dordrecht.

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

Lambrecht, J.A., 2001. System and method for interactive approximation of head transfer function. US Patent 6181800 B1.

Lopez-Poveda, E.A., Meddis, R., 1996. A physical model of sound diffraction and reflections in the human concha. J. Acoust. Soc. Am. 100, 3248–3260.

Lowe, D.D., Cashion, T., Williams, S., 1994. Stereo headphone sound source localization system. US Patent 5371799.

Lumbreras, M., Sanchez, J., 1999. Interactive 3D sound hyperstories for blind children. In: Proceedings of the Conference on Human Factors in Computer Systems, Pittsburgh, May 15–20, pp. 318–325.

Middlebrooks, J.C., 1992. Narrow-band sound localization related to external ear acoustics. J. Acoust. Soc. Am. 92, 2607–2624.

Middlebrooks, J.C., Green, D.M., 1990. Directional dependence of interaural envelope delays. J. Acoust. Soc. Am. 87, 2149–2162.

Miller, M.B., 1995. Coefficient alpha: a basic introduction from the perspectives of classical test theory and structural equation modelling. Struct. Eq. Model. 2 (3), 255–273.

Mills, A.W., 1972. Auditory localization. In: Tobias, J.V. (Ed.), Foundations of Modern Auditory theory, vol. 2. Academic Press, New York.

Oldfield, S.R., Parker, S.P.A., 1984. Acuity of sound localization: a topography of auditory space. II. Pinnae cues absent. Perception 13, 601–617.

Oppenheim, A.V., Willsky, A.S., Nawab, H., 1997. Signals and Systems, second ed. Prentice-Hall, Upper Saddle River, NJ.

Robinson, D., Dadson, R., 1957. Threshold of hearing and equal-loudness relations for pure tones, and the loudness function. J. Acoust. Soc. Am. 29, 1284–1288.

Schröter, J., Pösselt, C., Opitz, H., Divenyi, P., Blauert, J., 1986. Generation of binaural signals for research and home entertainment. In: Proceedings of the 12th International Congress on Acoustics, vol. 1, Toronto, B1-6.

Scofield, W.C., Saounder, S.O., 1996. Head mounted surround sound system. US Patent 5661812.

Shaw, E.A.G., Teranishi, R., 1968. Sound pressure generated in an external ear replica and real human ears by a nearby sound source. J. Acoust. Soc. Am. 44, 240–249.

Sony, 2004. 〈http://www.sony.com.hk/〉.

Wenzel, E.M., Arruda, M., Kistler, D.J., Wightman, F.L., 1993. Localization using nonindividualized head-related transfer functions. J. Acoust. Soc. Am. 94, 111–123.

Wightman, F.L., Kistler, D.J., 1989. Headphone simulation of free-field listening. II. Psychophysical validation. J. Acoust. Soc. Am. 85, 868–878.

Wightman, F.L., Kistler, D.J., 1997. Monaural sound localization revisited. J. Acoust. Soc. Am. 101, 1050–1063.

Wong, K., Choi, Y.M.R., 2003. Personal communication. Kenneth Wong was the Vice-President of ValenceTech Ltd. (HK) and Raymond Choi was the President of Valence Semiconductor Design Ltd.