OPTIMIZING THE COST AND PERFORMANCE OF NON-INDIVIDUALIZED BINAURAL CUES FOR SURROUND SOUND SIMULATIONS

PACS: 43.66.PN

Leung, N.M.[1]; Braasch, J.[2] and So, R.H.Y[1]
[1]Department of Industrial Engineering and Engineering Management
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon,
Hong Kong SAR, PRC.
Tel: ++852-23587105
Fax: ++852-23580062
Email: rhyso@ust.hk

[2]Institut für Kommunikationsakustik
Ruhr-Universität Bochum
Universitätsstr. 150
44787 Bochum, Germany
Tel.: ++49-234-3225387
Fax.: ++49-234-3214165
Email: braasch@ika.ruhr-uni-bochum.de

ABSTRACT
Two studies were conducted to optimize the binaural cues generated using non-individualized head related transfer functions (HRTFs). In the first study, the minimal number of coefficients of the impulse response (a determining factor of cost) was determined to be 32 or 64 depending on the direction. The second study applied 0, 12, and 18dB enhancements, in six frequency bands, to HRTFs and examined the effects on front-back confusion. Results from tests on 32 participants indicated that, for 0°, 45°, 135°, 180°, 225°, and 315° azimuth, the 12dB enhancements significantly reduced the occurrence of front-back confusion ($p<0.05$). A binaural surround sound simulator was developed on the Motorala_DSP56362 platform.

INTRODUCTION

The use of individualized and non-individualized head related transfer functions (HRTFs) to simulate binaural cues have been documented in many publications and will not be repeated here (e.g., Begault, 1994; Blauert, 1983, 1997). Data have indicated that while individualized HRTFs can simulate binaural cues in accurate directions, binaural cues generated using non-individualized HRTFs are associated with large localization errors (e.g., Flangan *et al.,* 1998; Wightman and Kistler, 1989; and Wenzel *et al.,*1993). Since 2000, researchers in the Department of Industrial Engineering and Engineering Management at the Hong Kong University of Science and Technology and the Institut für Kommunikationsakustik at Ruhr-Universität Bochum have been studying the possibility of mass-customizing non-individualized HRTFs. According to Tseng and Su (1998), mass-customization allows the production of customized products or services with near mass-production efficiency and cost. In the keynote speech of the first world congress on mass-customization, Prof. Nam Suh at MIT reported that more than 56% of the products made in US will be mass-customized in the next few years (Suh, 2001). Under the concept of mass-customization, a virtual headphone-based surround sound (VHSS) system based on non-individualized HRTFs can be customized for an individual listener at a cost and effort less than the cost and effort of measuring a set of individualized HRTFs. This paper reports on initial progress towards mass-customizing non-individualized HRTFs.

The aims of this study are two-fold: (i) to reduce the complexity of non-individualized HRTFs while maintaining the localization performance associated with their respective binaural cues;

and (ii) to reduce the occurrence of front-back confusion errors associated with simulated binaural cues by manipulating the spectra of the corresponding HRTFs.

OPTIMIZING THE COST OF NON-INDIVIDUALIZED BINAURAL CUES

In a virtual headphone-based virtual surround sound (VHSS) system, the main component is a set of non-individualized HRTF filters responsible for simulating accurate binaural cues. These filters are usually implemented by algorithms that run on a digital signal processing (DSP) chip and the cost of this chip is proportional to the complexity of the HRTF filters. In order to optimize the cost of simulating binaural cues, the complexity of the HRTF filters must be optimized.

Method

An experiment was conducted to examine the effects of reducing the complexity of HRTF filters on the localization errors associated with the resulting binaural cues. In this experiment, the complexity of HRTF filters were represented by the numbers of coefficients of the impulse responses representing the HRTF filters. The higher the number, the higher the level of complexity. The design of the experiment is summarized in Table 1 and further details can be found in Leung *et al.* (2002a). It was hypothesized that the relationship between the complexity of HRTF filters and the localization errors associated with the resulting binaural cues are not linear. In particular, when the complexity of the HRTFs used to produce the binaural cues reduced, listeners would still be able to perceive the correct directions of sound until the complexity was less than certain thresholds. The non-individualized HRTFs used in this experiment were taken from the open-copyrighted MIT KEMAR data set (Gardner and Martin, 1995).

Table 1: Summary of the design of experiment conducted to study the effects of reducing the complexity of HRTFs (adapted from Leung et al., 2002a).

Design of experiment:	Full-factorial within-subject design
Independent variables:	4 levels of HRTF complexity as represented by the numbers of coefficients in the respective impulse responses: 128, 64, 32, and 18 coefficients
	9 simulated cue directions: diotical-mono, and azimuth angles of 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315° at ear-level
	2 repetitions
Total number of runs per participant:	4 levels of complexity x 9 directions x 2 repeats = 72 runs
Number of participants:	24: 12 male and 12 female
Total number of data runs:	72 runs x 24 participants = a total of 1728 data runs

Results and discussion

As shown in Table 1, the experiment involved 24 participants. Only the initial results of the first 12 participants will be discussed here. The full results are reported in Leung *et al.,* 2002a. Figure 1 illustrates the median of the localization errors (in degrees) of the first 12 participants as functions of four cue directions and 4 levels of complexity. The figure indicates that although the binaural cues generated with HRTFs with 18 coefficients were associated with the highest localization errors in all four directions, the binaural cues generated using the HRTFs with 128 coefficients (i.e., the highest complexity) were not always associated with the lowest localization errors. In fact, for the directions of azimuth angles of 45° and 315°, the binaural cues generated using HRTFs of 64 coefficients were associated with the lowest errors. Results of Mann-Whitney statistical tests indicate that the reduction of HRTF coefficients significantly affected the measured localization errors of the corresponding binaural cues ($p<0.05$). These results partially support our hypothesis because the binaural cues generated from HRTFs with 18 coefficients were associated with significantly lower errors than those generated from HRTFs with 128 coefficients ($p<0.05$, Mann-Whitney tests). However, reducing the HRTF coefficients from 128 to 64 produced significantly more accurate binaural directional cues when the cue directions were at 45° and 315° azimuth angles.

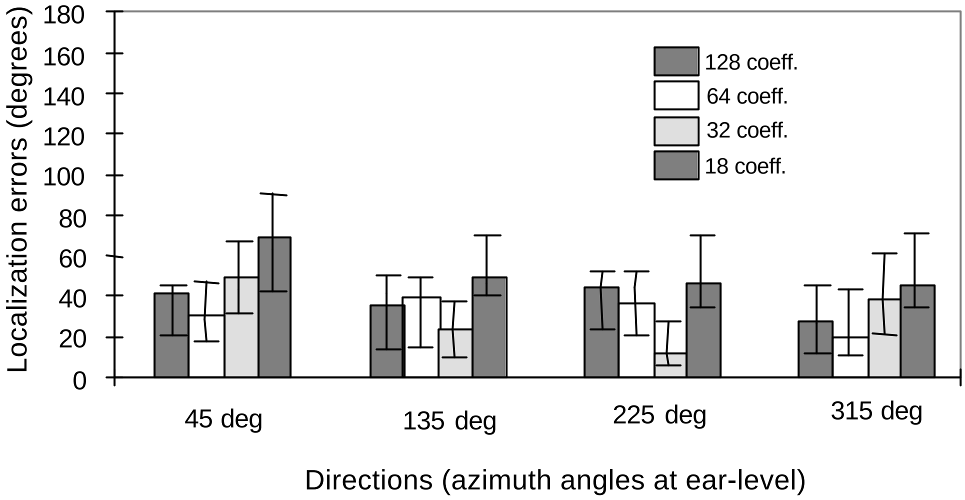

Figure 1: Median localization errors for the directions of 45°, 135°, 225°, and 315° with 4 levels of HRTF complexity: 128, 64, 32, and 18 coefficients in their respective impulse responses. Medians of the first 12 participants are shown. Full data can be found in Leung et al., (2002a).

OPTIMIZING THE PERFORMANCE OF NON-INDIVIDUALIZED BINAURAL CUES

This study attempted to reduce the occurrence of front-back confusion errors associated with binaural cues by amplifying and attenuating the appropriate spectral peaks and troughs of the corresponding HRTFs. This concept is not new. In 1989, Myers filed a patent that claims that if the signals of a binaural cue are amplified in the frequency ranges of 0.2 to 0.7kHz and 2 to 6.3kHz as well as attenuated in the frequency ranges of 0.7 to 2kHz and 6.3 to 22kHz, then the perceived position of the sound source will appear more in the front of a listener. Tan and Gan (1998) followed Meyer's work and reported that if signals of an non-individualized binaural cue from a frontal direction are amplified in the frequency ranges of 0.2 to 0.7kHz, 2 to 6.3kHz, and 11 to 22kHz as well as attenuated in the frequency ranges of 0.7 to 2kHz and 6.3 to 11kHz, the number of listeners perceiving this frontal cue as coming from the back would be reduced from 8 out of 10 to 4 out of 10. Although this concept has been studied, none of the past studies verified the findings using statistics. Also, the interactions between the effects of cue directions and the effects of spectral manipulations have not been studied.

Methods

The design of the second experiment is summarized in Table 2. Further details of the experiment can be found in Leung *et al.* (2002b). In this study, the spectra of the HRTFs were dividied into 6 frequency bands: band one - 0.2 to 0.7kHz; band two - 0.7 to 2kHz; band three - 2 to 6.5 kHz; band four - 6.5 to 10kHz; band five - 10 to 14kHz; and band six - 14 to 22kHz. This division was finer than those of Myers (1989) and Tan and Gan (1998). The spectra of HRTFs in each frequency band were manipulated in turn and the exact method of manipulation is documented in Leung *et al.* (2002b). It was hypothesized that the higher the level of manipulation (from 0dB to 18dB), the lesser the occurrence of front-back confusion errors.

Table 2: Summary of the design of experiment conducted to study the effects of manipulating the spectra of HRTFs (adapted from Leung et al., 2002b).

Design of experiment:	Full-factorial within-subject design
Independent variables:	3 levels of spectral manipulation: 0dB, ±12Db, ±18dB (+dB for manipulating spectral peaks and –dB for troughs)
	6 frequency bands for manipulation: only one frequency band is manipulated at one time. The bands range from (0.2-0.7kHz) to (14-22kHz).
	6 simulated cue directions: azimuth angles of 0°, 45°, 135°, 180°. 225°, and 315° at ear-level.
	2 repetitions
Total number of runs per participant:	3 levels of manipulation x 6 bands x 6 directions x 2 repeats = 213 runs
Number of participants:	32: 12 male and 20 female (16 with previous musical training)
Total number of data runs:	216 runs x 32 participants = a total of 6912 data runs

Results and discussion

Figure 2 illustrates the percentages of occurrences of front-back confusion errors when participants listened to the binaural cues at 45° and 315° azimuth angles. The percentages were calculated from the data of the first 16 participants. The figure indicates that, in all frequency bands except band four, the occurrences of front-back confusion reduced after the spectra of HRTFs were manipulated. Results of Mann-Whitney non-parametric statistical tests show that for binaural cues of 0°, 45°, 135°, 180°, and 225° azimuth angles, the localization errors reduced significantly after the 12dB manipulations in all frequency bands except band four ($p<0.05$). This result is consistent with the finding by Blauert (1983, 1997) that spectral peaks and troughs at band four were responsible for detecting the elevation angles and not the azimuth angles of a sound source. For the binaural cues from the 315° azimuth angle, only the spectral manipulations in band two and five produced significant reductions in localizaton errors ($p<0.05$). A full discussion of the results can be found in Leung *et al.* (2002b).

DEVELOPMENT OF A PROTOTYPE VIRTUAL SURROUND SOUND SYSTEM

Results of the first experiment suggest that the optimal HRTF coefficient for simulating accurate binaural directional cues of 45° and 315° azimuth angles at the ear-level should be 64. These two directions were of special interest because they represented the relative directions between the left-front and the right-front speakers to a listener in a typical Dolby™ 5.1 digital surround sound arrangement (Dolby™ Laboratories, 2002). In addition, the results of the first experiment also indicate that for binaural cues from 0°, 90°, and 270° azimuth angles, reducing the HRTF coefficients from 128 to 32 did not significantly affect the localization performance ($p<0.05$). Consequently, the optimal HRTF coefficient for simulating accurate binaural cues for the centre (i.e., 0°), right-surround (i.e., 90°), and left-surround (i.e., 270°) speaker directions was 32. Using these results, a prototype virtual headphone-based surround sound (VHSS) system was developed on the Motorala_DSP56362 platform. Results of a double-blind usability test of 40 listeners indicated that this prototype VHSS system produced significantly better surround sound effects than the corresponding Dolby™ stereo system ($p<0.05$). Currently, the results from the second experiment are being urilized to reduce the occurrence of front-back confusion errors in the prototype VHSS system.

ACKNOWLEDGEMENT

This work was partially supported by the Innovation and Technology Fund of the HKSAR Government through the Consumer Media Center at the Hong Kong University of Science and Technology (AF/264/97) as well as the Research Grants Council of the HKSAR Government

(HKUST6219/02E) and its counter-part in Germany (DAAD) for the Joint HK / Germany Research Grant (G-HK99/00.EG03; G-HK002/00E).

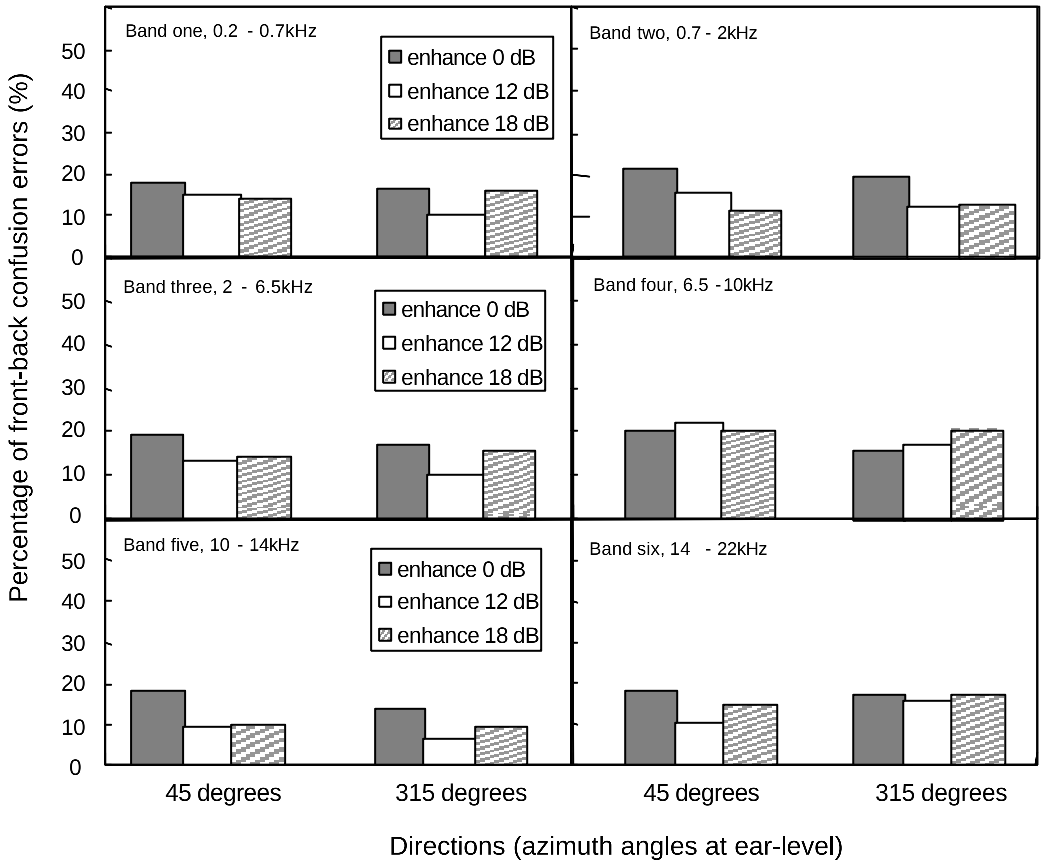

Figure 2: Percentages of occurrences of front-back confusion errors when participants listened to simulated binaural cues at 45° and 315° azmiuth angles with 3 levels of spectral manipulations at 6 frequency bands. The percentages were calculated using the data of the first 16 participants. Full data can be found in Leung et al., 2002b.

LITERATURE CITED

BEGAULT, D.R. **(1994)**: *3D sound for virtual reality and multimedia*. AP Professional.

BLAUERT, J. **(1983 or 1997)**: *Spatial hearing: the psychophysics of human sound localization*. MIT Press.

DOLBY™ LABORATORIES **(2002)**: *Speaker arrangement of Dolby™ 5.1 digital surround sound systems.* Available URL: http://www.dolby.com/tech/

FLANAGAN, P., MCANALLY, K.I., MARTIN, R.L., MEEHAN, J.W. and OLDFIELD, S.R. **(1998)**: *Aurally and visual guided visual search in a virtual environment*. Human Factors, 40, 461-8.

GARDNER, W. G., MARTIN, K.D., **(1995)**: *HRTF measurements of a KEMAR*. Journal of the Acoustical Society of America 97 (6), 3907-3908.

LEUNG, N. M., BRAASCH. J., SO, R. H. Y. **(2002a)**: *Optimising the complexity of non-indivdualized HRTFs for simulating surround sound effects: an ergonomics study and product implementation*. Awaiting publication. Submitted to Applied Ergonomics.

LEUNG, N. M., BRAASCH, J., SO, R. H. Y. **(2002b)**: *Effects of amplifying and attenuating spectral cues at different frequencies: a preliminary study towards mass-customizing non-individualized Head-Related Transfer Functions*. Awaiting publication. Submitted to Human Factors.

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

SUH, N.P. **(2001)**: *Design for mass-customization. Keynote*, 1st World Congr. Mass-customization & personalization, HK.

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

TSENG, M.M. and DU, X. **(1998)**: *Design by Customers for Mass Customization Products*. CIRP Annals, 47, 103-106.
WENZEL, E.M., ARRUDA, M., KISTLER, D.J., and WIGHTMAN, F.L. **(1993)**: *Localization using non-individualized head-related transfer functions*. J. Acoust. Soc. Am., 94, 111-123.
WIGHTMAN, F.L., KISTLER D.J. **(1989)**: *Headphone simulation of free-field listening II: Psychophysical validation*. J. Acoust. Soc. Am., 85, 868-78.