Optimizing front/back confusion rates in sound localization performance: cluster analyses and experimental studies

by

NGAN Kwok Hung

A Thesis Submitted to
The Hong Kong University of Science and Technology
In Partial Fulfillment of the Requirements for
The Degree of Master of Philosophy
In Industrial Engineering and Logistics Management

June 2005, Hong Kong

Authorization

I hereby declare that I am the sole author of the thesis.

I authorize the Hong Kong University of Science and Technology to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the Hong Kong University of Science and Technology to reproduce the thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

NGAN Kwok Hung

Optimizing front/back confusion rates in sound localization performance: cluster analyses and experimental studies

by

NGAN Kwok Hung

This is to certify that I have examined the above MPhil Thesis
and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by
the thesis examination committee have been made.

DR. Richard Hau Yue SO Supervisor

PROF. Chung Yee LEE Head

DEPARTMENT OF INDUSTRIAL ENGINEERING AND LOGISTICS MANAGEMENT

27 AUGUST 2005

Acknowledgements

I would like to express my sincere gratitude to my thesis supervisor Dr. Richard H. Y. So. He provided guidance, advice and encouragement with much patience throughout the years of my postgraduate study. He enabled me to conduct the research continuously while I faced critical problems. My vision has been considerably broadened through in depth discussion with him. I am indebted to him for not only his continual support but for this entire experience.

Special thanks must also be given to Dr. Xiangtong Qi and Dr. Lydia Ayers as they served as my thesis committee members and I very much appreciated their advice for my thesis and defense examination.

Sincere thanks are given to Prof. Jens Blauert, Dr. Jonas Braasch, Dr. Ercan Altinsoy and Antonio G. Avalos for their valuable suggestions and advice on my research while I was in Germany.

Deep appreciation is expressed towards Dr. Felix Lor and John Au. They spent many hours on discussing the research with me and assisted with solving many technical problems.

I would like to take this opportunity to thank Dr. Richard H. Y. So, Dr. Ravidra S Goonetilleke and all of the members in the advanced seminar group who gave valuable advice during my presentations.

I would like to express my appreciation to Denil Chan, K.C. Tin, Stanley Sze and K. K. Yung for their technical support in the experimental setup and for assistance with using many different types of equipment in the laboratories. I would like to take this opportunity to thank Dr. Aaron Leung and Diana at the Hong Kong Polytechnics University for spending many hours on the construction of ear models.

Moreover, I would like to thank Odan Wong, Calvin Chu, Hung, Daniel Lau, Rosary Chen, Jennifer Ji, Ada and Leo for their encouragement and support during the discussion of different research topics. In addition, I would like to thank Cherrie Yung, Wylie Tsang and Willie Shen for their support throughout the years.

I would like to take this opportunity to also thank Mr. Cao, a column writer for the Hong Kong Economic Journal. His writings inspired me to change my attitude towards life.

Last but not least my thanks must go to Janet and my family for their love and support.

TABLE OF CONTENTS

List of Figures

List of Tables

Optimizing front/back confusion rates in sound localization performance: cluster analyses and experimental studies

by

NGAN Kwok Hung

Department of Industrial Engineering and Logistics Management
The Hong Kong University of Science and Technology

Abstract

Non-individualized head-related transfer function (HRTF) devices cannot generate accurate directional sound cues. In particular, listeners often incorrectly perceive a sound cue from the front as coming from the back (referred to as front/back confusion). This study presents a methodology to reduce the occurrence of front/back confusion when listening to sound cues generated using non-individualized HRTF devices. The study investigates and identifies that front/back confusion occurs because of mis-matches between non-individualized HRTFs and the acoustics characteristics of the ears of individual listeners. Using literature concerning the acoustics characteristics of human ears, spectrum of 196 open-copyrighted

non-individualized HRTFs have been analyzed and clustered according to their abilities to generated directional sound cues that are coming from the front and from the back. These clusters are further processed to determine six standard directional sound cues for the center-front and center-back directions, respectively. These twelve standard sound cues enable listeners to choose and minimize the possible mis-match that leads to front/back confusion. Three experiments have been conducted and the results indicate that providing the choice of these six standard frontal cues and the six standard backward cues can significantly reduce the sound localization errors and the rates of front/back and back/front confusion. In addition, providing the choice of these twelve cues have been shown to achieve better sound localization performance than using the sound cues generated with the MIT KEMAR non-individualized HRTF set. This finding is important because the MIT KEMAR HRTF set is the most widely-used open-copyrighted HRTF data set. Results reported in this thesis enable engineers to design and develop customizable non-individualized HRTF devices in a cost-effective way. Currently, most digital audio products are manufactured in China but the profit margins for such products are decreasing annually. If industrialists in Hong Kong can acquire the ability to produce customizable virtual surround sound systems using non-individualized HRTF devices at low cost, new innovative products with much greater profit margins can be designed and produced.

Chapter 1. Introduction

1.1. The role of HRTFs in sound localization

The spatial filtering of sound waves, by the pinna, torso and head, before reaching the ear drum is termed the head-related transfer function (HRTF). It is composed of two parts: a frequency domain amplitude change and a time-delay. The complex construction of the pinna is the main cause of the spectral modification of the sound source. The content of the HRTF varies for each sound source position. The uniqueness of the HRTF at different sound source positions acts like fingerprints. It provides information for the brain to localize the sound source.

If the time-delay part of HRTFs is excluded, only the spectral part remains. These frequency-dependent amplitude changes are called spectral cues. Because of the uniqueness of the spectral part of HRTFs at different sound source positions, it can provide important information to distinguish front-back and up-down sound sources. This study focuses on analyzing spectral cues of human sound localization and identifying standard HRTF sets for listeners.

1.2. Simulation of binaural sound cues with headphones

HRTF technology is used for producing virtual sound with spatial information which was delivered by headphones. The process of binaural sound cues simulation was summarized in Figure 1.1. A mono sound source is filtered by HRTF for left ear and right ear. The filtered signals with spatial information were then delivered to left channel and right channel respectively. The filtration process is the convolution by FIR.

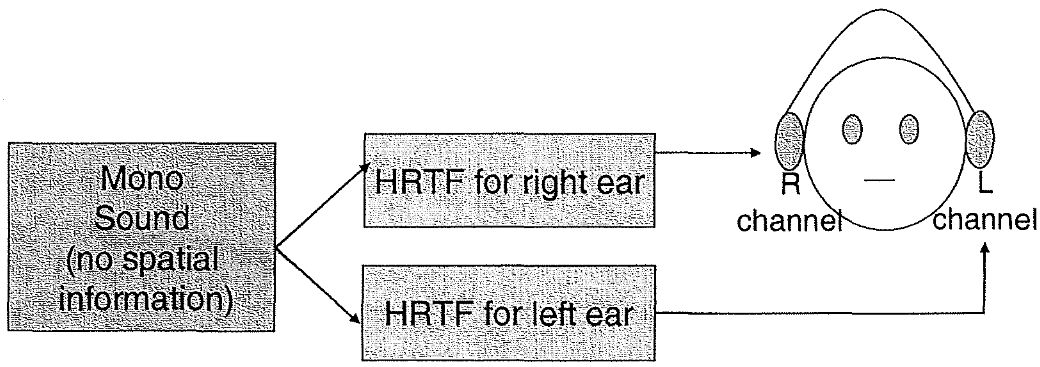

Figure 1.1 The process of binaural sound cues simulation

Each HRTF represents the spectral and temporal differences between the reference sound source and sound received at the opening of ear canal. Interaural level difference (ILD) and interaural time difference (ITD) are the differences between HRTF for left ear and HRTF for right ear. The calculation details can be found in Blauert (1997) [37].

1.3. Applications of HRTF technology

HRTF technology can be applied in a virtual surround sound system for home entertainment. With this technology, a pair of headphones can simulate the same surround sound effect as 6 or 8 speakers in Dolby™ 5.1 or 7.1 system. Also, listeners can experience the same sound effect in any locations while they have to stay in sweet spot if speaker system is used. HRTF technology can be used for military purposes. The virtual surround sound system can notify fighter pilots the directions of enemies and missiles. Then pilots can save time on monitoring the radar system.

1.4. Inter-subject variability in spectral cues and its possible consequences

Since each ear has different features, their acoustics effect and HRTF should be different. Inter-subject variability is a term to describe the differences among the population. To achieve the best sound localization performance, using individualized HRTFs provides a perfect solution. However it is expensive both in terms of time and money to collect individualized HRTFs. Several studies of using non-individualized HRTFs to simulate binaural sound cues were conducted. Since there are inter-subject variability in HRTFs, using non-individualized HRTFs has lead to an increase of front-back confusion and up-down (elevation) confusion. At the extreme, occluding pinnae is the most direct way to alter HRTFs. Musicant and Butler (1984) investigated the effect of occluding pinnae on sound localization. Eight subjects participated to localize sound with and without occluded pinnae. It was found that occlusion of both pinnae resulted in a dramatic increase in front-back confusion ($p<0.01$).

Wenzel *et. al.* (1988) [24] compared the sound localization performance of real sources for 3 subjects with stimuli synthesized from individualized and non-individualized HRTFs. A "good localizer" gave an accurate performance on locating the direction of sound and there was less front-back confusion with both free-field sources and individualized HRTFs. However, greater localization error and front-back confusion occurred when non-individualized HRTFs were used.

Further studies have demonstrated that non-individualized HRTFs result in greater front-back and elevation confusions. Asano *et. al.* (1990) [4] measured the HRTFs of 2 subjects and then presented virtual sound cues generated by these HRTFs to them through headphones. The rate of front-back confusion was higher when non-individualized HRTFs were used to synthesize virtual sound source. Wenzel *et. al.* (1993) [25] investigated the accuracy of locating the direction of sound on virtual stimuli synthesized with non-individualized HRTFs. The results from 16 participants showed that the virtual stimuli induced high rates of front-back and elevation perception error.

Middlebrooks *et. al.* (1989) [19] measured the HRTFs of six subjects and compared them at different frequency ranges. They found that the amplitude gains of a particular subject's ear at certain frequency ranges had a similar pattern in another subject's ear at a different frequency range. Since spectral cue relies on the shape of the HRTF, with the findings from Middlebrooks *et. al.* (1989) [19], the HRTF of different individuals have a high degree of similarity with minor inter-subject differences among the population.

In summary, there are two strands of argument for the position of inter-subject variability of HRTFs. On the one hand, the sound localization with virtual sources synthesized by non-individualized HRTFs has the problem of front-back and elevation confusion. It can be concluded that HRTFs are unique to each individual and hence spectral cues will be different for each listener. On the other hand, HRTFs have a general shape and high degree of similarity among the individuals. These points of view would appear to contradictory. However, it can be concluded that different individuals have neither totally different spectral cues nor the same spectral cues.

In other words, there is a degree of inter-subject variability on spectral cues. The observation and conclusion lead to a possibility to cluster individuals into different listener groups according to their spectral cues. It is expected that listeners who fall into the same cluster have similar spectral cues profiles.

1.5. Benefits of grouping individuals according to their spectral cues

Individualized head-related transfer function (HRTF) devices can accurately simulate directional binaural sound cues and generate personalized virtual surround sound systems (Begault, 1994 [6]; Hofman et al., 1998 [36]). Such a personalized virtual surround sound system enables listeners to enjoy the surround sound effects of Dolby™ 5.1 or 7.1 recordings using a pair of headphones instead of five to seven surround sound speakers correctly and spatially oriented. However, individualized HRTFs are costly to obtain in both terms of time and money (Begault, 1994 [6]; Blauert, 1997 [37]). On the other hand, non-individualized HRTFs are freely available (e.g., Algazi et al., 2001 [2]; Gardner and Martin, 1995 [11]) but each set of non-individualized HRTFs can only generate good quality virtual surround sound for a small group of listeners. For those who are not within the appropriate group, surround sound cues that are coming from the front may appear to have come from the back and vice versa. This is referred to as 'front-back confusion' (e.g., Middlebrooks, 1992 [18] and Wenzel *et. al.*, 1993 [25]). In addition, it has been established that the elevation angle of an incident sound cue cannot be accurately simulated (Hebrank and Wright, 1974 [12]).

A review of the literature indicates that, for a listener, the main factor accounting for the inaccuracies associated with sound cues generated using non-individualized HRTFs lies in the spectral similarity between the non-individualized HRTFs and the individualized HRTFs of that particular listener. The greater the similarity, the lessor the inaccuracy (Blauert, 1997 [37]). Consequently, if it is possible to determine clusters of non-individualized HRTFs and ways to assign listeners to appropriate clusters, then it would be possible to select a non-individualized HRTF device for a particular listener. Hence the spectral responses of the selected non-individualized HRTFs would be close to those of the individualized HRTFs of that listener. In other words, if one size does not fit all, then our objective is to design a way to produce as many sizes that will fit most of the population.

1.6. Purpose of this research

This study aimed to investigate the possibility of making use of cluster analysis on HRTFs sets to increase listeners' sound localization performance by looking for standard HRTF sets and grouping listeners. One of the applications of this technology is the headphone surround sound system. This study covered:

(i) setting up the definition of characteristics for directional bands in HRTFs;

(ii) investigating the performance of several cluster algorithms and their suitability on clustering the data of band characteristics;

(iii) defining a methodology to determine standard HRTF sets;

(iv) studying listeners' sound localization performances using different standard HRTF sets.

1.7. Organization of the thesis

The first chapter introduces the background and objectives of this research. The role of HRTFs, inter-subject variability in spectral cues and benefits of grouping individuals according to their spectral cues are also discussed in this chapter. The second chapter contains the literature review on the duplex theory, the oldest theory explaining how a person localizes sound, and its limitations. The literature review on spectral cues and frequency bands related to sound localization are also discussed.

In chapter three, the intra-subject and inter-subject variability of listeners' sound localization performance are studied. This chapter aimed to investigate the learning effect of listening tasks and the possibility of grouping listeners according to their sound localization performance, i.e. representing the population by more than one set of HRTFs.

The definitions of directional band characteristics are discussed in chapter four. These band characteristics are defined according to the results in the existing body of literature and have been divided into front, back and elevation cues.

Chapter five documents the overview of finding standard HRTF sets. Chapter six discusses the cluster analysis in details. Different cluster algorithms were compared and selected for analysis.

Chapter seven contains the technical details, including methodology, software and hardware, used in the experiments. Chapter eight to nine documents experiment details for verifying the standard HRTF sets. Chapter ten documents the experiment detail for the exhaustive study of localization performance of sound cues generated by HRTF sets. Finally the conclusions, discussions and recommendations are examined in chapter eleven.

Chapter 2. Literature Review

2.1. Duplex theory

The duplex theory is the earliest theory of human sound localization formulated by Lord Raleigh (1907) [17]. It assumes that the head is spherical with no pinnae. Since the two ears are separated on the left and right side of the head, unless the sound source is on the median plane, the distance traveled from the sound source to the left and right ear is different. This phenomenon causes the sound waves to reach the ipsilateral ear before reaching the contralateral ear. This leads to the phase difference at both ears which is known as the interaural time difference (ITD) and differences in sound level which is known as the interaural level difference (ILD).

However, the incompleteness of the duplex theory is illustrated by the cone of confusion illustrated in Figure 2.1 (Carlile 1996 [29]). The interaural time and level difference are ambiguous by choosing any two points on the cone of confusion. However, front-back and up-down direction cannot be distinguished with only the information of ITD and ILD. Other cues in addition to ITD and ILD are needed to explain the sound localization abilities of human beings.

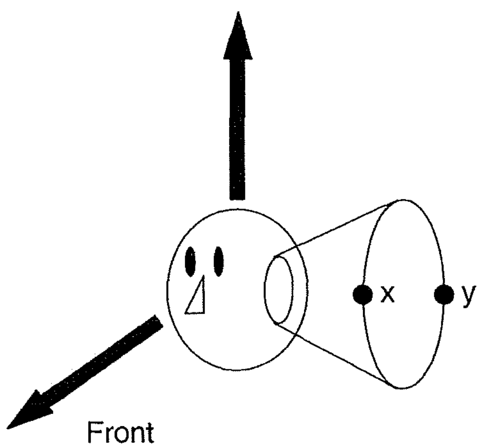

Figure 2.1 An illustration of a cone of confusion. Any two points on the plane, such as x and y, have no difference in time difference. Modified from Carlile, 1996 [29].

The limitation of the duplex theory leads to the study of the spectral response of pinna. The acoustical response of pinnae is the supplement to duplex theory. The folds of the pinnae, including the fossa triangularis, helix and the hollow area in concha, alter the spectral contents. These spectral differences are dependent on the sound direction. This leads to a series of studies on the role of pinnae and spectral cues in sound localization, i.e., Blauert 1969/70 [7]; Gardner and Gardner 1973 [9]; Hebrank and Wright 1974 [12]; Musicant and Butler 1984 [21]; Asano *et. al.* 1990 [4]; Middlebrooks 1992 [18]; Langendijk and Bronkhorst 2002 [16].

This thesis focused on studying of the sound localization on the median plane, where the effects of interaural level difference (ILD) and interaural time difference (ITD) are minimized. ILD and ITD can be modeled using a physical model (Kuhn 1987 [15]) that can be easily tuned for each individual. On the other hand, the inter-subject variability of the monaural spectral cues has not been studied systematically. Hence, the analysis of the monaural spectral cues is the focus of this study. Studying sound direction on the median plane only controls the main source of the inter-subject variability that solely comes from monaural spectral cues.

2.2. Characterizing the spectra of HRTFs into directional band

Blauert (1969/70) [7] was among the early studies characterizing the spectra of HRTFs into directional bands. Twenty participants listened to noise signals of one third octave bandwidth and were asked to judge whether they came from the front or the back. The signals were presented once from the front and once from the back through loudspeakers. Data showed that noise signals with energies in the frequency range of 280-560 Hz or 2900-5800 Hz only caused a bias towards frontal-judgment while the opposite is true for noise signals within the frequency range of 720-1800 Hz or 10300 – 14900 Hz.

Blauert further compared the spectra of sound pressure arriving at the entrance of ear canals when broadband noise was propagated from the front and from the back by speakers. Results from the participants showed that even though the same sound source was played, the sound pressures arriving the entrance of the ear canals from the front and back are significantly different in certain frequency bands ($p<0.05$, non-parametric statistical tests). In particular, Blauert found that if energies at certain frequency bands were amplified (i.e., boosted), the perceived direction of the sound could be manipulated to be more in the front or at the back depending on the bands. Blauert named those frequency bands as boosted bands and related them to frontal and back perception of sound. Table 1 summarizes the boosted bands reported in Blauert (1969/70) [7]. Blauert compared the positions of directional bands and boosted bands (Figure 2.2), and found that they adopted similar positions. Although Blauert (1969/70) [7] did not specifically refer to the boosted band as part of HRTFs since the term was not introduced at the time, his data provided strong basis on how to characterize the spectra of HRTFs into patterns that are responsible for giving listeners a frontal or back cue.

Table 2-1 Boosted bands in Blauert(1969/70) [7]

Direction	Frequency range
Frontal	150 – 540 Hz
	1900 – 2900 Hz
	3600 – 5800 Hz
Backward	720 – 1700 Hz
	7400 – 11100 Hz

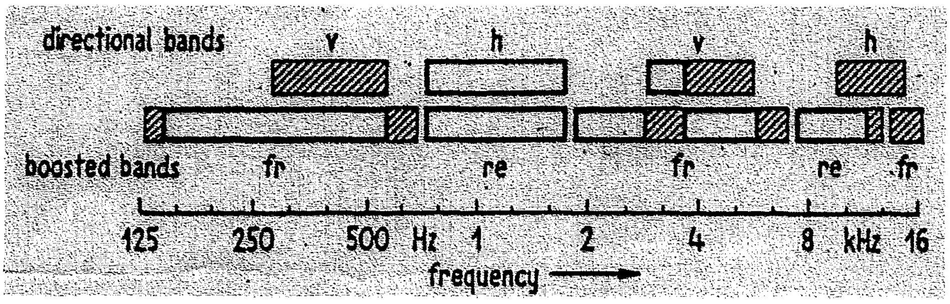

Figure 2.2 Positions of directional bands and boosted bands (extracted from Blauert (1969/70) [7]) v, fr: front, h,re: rear

Hebrank and Wright (1974) [12] conducted an experiment to investigate how the frequency content of a sound cue can affect the frontal, backward, above and below perceptions on the median plane. Twenty-eight participants were asked to localize white noises that were filtered by any one of six high-pass cutoffs (3.8, 5.8, 7.5, 10.0, 13.2 and 15.3 kHz), one of seven low-pass cutoffs (3.9, 6.0, 8.0, 10.3, 12.0, 14.5 and 16.0 kHz), one of twelve 1/12-octave bandpass filters (4.0, 4.5, 5.1, 5.7, 6.4, 7.2, 8.1, 9.1, 10.2, 11.4, 12.8 and 14.5 kHz) or one of twelve bandstop filters (6.2, 6.6, 7.0, 7.4, 7.9, 8.8, 9.8, 10.8, 12.0, 13.3, 15.0, and 17.8 kHz). The results are summarized in Table 2-2.

Table 2-2 Summary of the findings in Hebrank and Wright (1974) [12]. Different types of filters were applied to the white noise and then delivered to listeners. Listeners were biased to different direction in different type of filters.

Perceived direction	Type of filters applied to the white noise
Frontal elevation (Above)	3.9 – 8.0 kHz low-pass cutoffs 4.0 – 7.2 kHz bandpass peaks
Front	13.2 and 15.3 kHz high-pass cut-offs 14.5 kHz bandpass peak 7.4 – 10.8 bandstop notches
Behind (Back)	10.0 kHz high-pass cutoffs
Above	10.3 kHz low-pass cutoffs 8.1 – 9.1 kHz bandpass peak 12.0 – 17.8 kHz bandstop notches

However, the bandstop filters used in Hebrank and Wright (1974) [12] were generated by summing white noise with a time delay of itself. Distortions were made in the whole of the audible frequency spectrum by those filters in addition to the designated frequency. In Figure 2.3, the designated notch frequency is 17.8 kHz. However, the amplitude of frequency range from 0 to 9 kHz is magnified. From 9 kHz onwards, the amplitude is diminished. Those distortions affected the validity of the experiment. There was no data provided to describe how those distortions affected the sound localization performance.

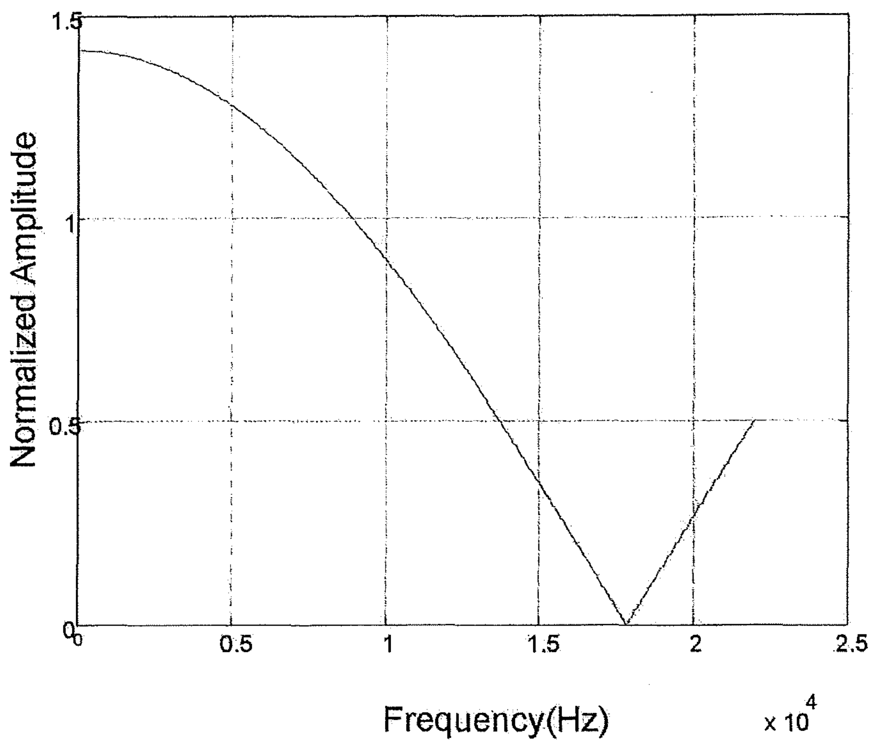

Figure 2.3 Frequency response of the bandstop filter at 17.8 kHz described in Hebrank and Wright (1974) [12]

More recently, Langendijk and Bronkhorst (2002) [16] studied the contribution of spectral cues to human sound localization by replacing spectral information in different frequency bands, which are above 4 kHz, with the average level of the replaced band. Eight participants' individualized HRTFs were measured and their own sets of HRTFs were used to synthesize the binaural cues. Nine experimental conditions were presented to subjects and outcome summarized in Table 2-3. The result showed that removing spectral cues in 4-16 kHz, 5.7-11.3 kHz or 8-16 kHz significantly increased elevation localization errors and front-back confusion rates ($p<0.05$).

The mean front-back confusion rate was significantly greater ($p<0.05$) in the band 8-16 kHz (condition no. 5) than the band 4-8 kHz and band 5.7-11.3 kHz (condition no. 3 and 4).The elevation error is significantly greater ($p<0.05$) in the band 5.7-11.3 kHz (condition no. 4) than in the band 4-8 kHz and band 8-16 kHz (condition no. 3 and 5). Hence, it was concluded that the most important elevation cues are at 5.7-11.3 kHz and that front-back cues are at 8-16 kHz.

Table 2-3 Experiment conditions in Langendijk and Bronkhorst (2002) [16]

Condition number	Frequency bands with spectral cues removed
1	No spectral cues removed
2	2-octave, 4-16 kHz
3	1-octave (low) 4-8 kHz
4	1-octave (middle) 5.7-11.3 kHz
5	1-octave (high) 8-16 kHz
6	1/2-octave (low) 4-5.7 kHz
7	1/2-octave (middle-low) 5.7-8 kHz
8	1/2-octave (middle-high) 8-11.3 kHz
9	1/2-octave (high) 11.3-16 kHz

Middlebrooks (1992) [18] analyzed 5 participants' sound localization performance of 1/6-octave bandpassed noise bursts centred at 6, 8, 10 and 12 kHz. Results indicated that most participants perceived 6 kHz stimuli as coming from the front and above. Many participants perceived 8 kHz and 10 kHz stimuli as coming from behind. The perceived directions of 8 kHz and 12 kHz stimuli were either on or below the horizontal plane. This suggests that the frequency band at 5500-6500 Hz, 7350-8650 Hz and 9150-10850 Hz, 7350-8650 Hz and 11000-13000 Hz are related to front and above perceptions, back perceptions, and down perceptions respectively.

Myers (1989) [27] characterized the audible frequency range in to 4 bands, i.e. A (225 – 682 Hz), B (682 – 2069 Hz), C (2069 – 6279 Hz) and D (6279 – 20000 Hz). Myers suggested that boosting and attenuating different frequency bands changes the spatial perception of sound. For frontal perception, bands A and C were boosted while bands B and D were attenuated; for backward perception, bands A and C were attenuated while band B and D were boosted. For elevation perception, a V-shape notch at a frequency centre from 6 kHz to 12 kHz for elevation angle of -45° to +45° attenuated by 15-20 dB was added. However, Myers did not provide any experimental results to prove his suggestion.

Tan and Gan (1998) [26] modified and tested the directional band characteristics suggested by Myers (1989) [27]. Results of ten participants indicated that the number of participants suffering from front-back confusion could be reduced from 8 to 4 if band boosting and attenuating suggested. The result showed that band characterization helps to reduce the rate of front-back confusion.

Chapter 3. Preliminary Experiment: Intra-subject and intra-subject variability on sound localization performance

3.1. Purpose of this experiment

The listener sound localization performance declines while using non-individualized HRTFs. However, the variation of the perceived error is not known. Investigating the intra-subject and inter-subject variability is essential for verifying and comparing the performance of different non-individualized HRTFs to subjects. This experiment aimed not only to verify that listeners have no learning effect in sound localization but also to investigate the possibility of dividing listeners into groups according to their sound localization performance.

3.2. Objective and hypothesis

The objective of this experiment is to investigate the inter-subject and intra-subject variability of listeners' sound localization performance with a generic non-individualized HRTFs set at different azimuth and elevation angles.

It was hypothesized that (H1) listeners have no learning effect in terms of eight types of azimuth and elevation localization errors in repetitions and (H2) the inter-subject variability is larger than the intra-subject variability.

It was natural to hypothesize that listeners have no learning effect because human beings have been using a pair of ears to localize sound since they were born. Since there are a great range of differences in ear shape among the population and those differences affect the HRTF content, it is believed that listeners have some variation in sound localization performance when a generic non-individualized HRTF set was used for generating virtual sound cues.

3.3. Participants

Ten male participants (PE1-PE10) were recruited in Hong Kong University of Science and Technology. Before the sound localization experiment, an audiometric test was carried out on each participant to determine auditory normality. The test procedures used are described in Abel and Paik (2004) [38]. Participants were eligible for the experiment only if their hearing thresholds were below 20dB HL in each ear at 125 Hz, 1 kHz and 8 kHz, and they had an interaural difference in threshold of less than 10dB. Participants were not informed about the purpose of the experiment. A sample consent form for experiments is shown in Appendix D.

3.4. Dependent and independent variables

Three independent variables were used in this experiment: (i) 8 azimuth angles (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°); (ii) 3 elevation angles (-40°, 0°, 40°); and (iii) 6 repetitions. The coordinate system is described in chapter 7.2.

The dependent variables were: (i) Absolute perceived localization errors along an azimuth direction with front/back confusion correction; (ii) Absolute perceived localization errors along an azimuth direction without front/back confusion correction; (iii) Perceived localization errors along an azimuth direction with front/back confusion correction; (iv) Perceived localization errors along an azimuth direction without front/back confusion correction; (v) Absolute perceived localization errors along an elevation direction with up/down confusion correction; (vi) Absolute perceived localization errors along an elevation direction without up/down confusion correction; (vii) Perceived localization errors along an elevation direction with up/down confusion correction; and (viii) Perceived localization errors along an elevation direction without up/down correction

Up/down confusion corrections were done by producing a mirror image of user response with the centre line at 0° elevation along elevation direction if up/down confusion occurs. Front/back confusion corrections were done by producing a mirror image of user response with the centre line along 270° and 90° azimuth if front/back confusion occurs. Errors were calculated by deducting the response angle with the incident angle.

3.5. Stimuli preparation

Each sound clip consists of a beep sound followed by the words "Thanks for participating this experiment" and its Cantonese translation. Sound clips were filtered by an MIT KEMAR HRTF set under Matlab environment.

3.6. Procedure

Participants were asked to sit comfortably for a period of 5 minutes. A sphere and a pen with an attached Polhemus Fastrak sensor attached was given to each participant. Instructions on using the pen as a measuring device were then given by the experimenter. Each participant then used the pen to indicate the elevation and azimuth angles.

Stimuli were presented through Sennheiser HD580 headphones. After each stimulus, each participant had to decide the direction that was most likely to be the sound source. A GUI program developed a under a Matlab environment was developed to input the responses. Trial measurements were conducted until participants fully understood the whole procedure. The first 5 readings were discarded as equipment testing.

3.7. Results and Discussion

3.7.1. Intra-subject variability

In this part of the study, the effect of repetition was studied for checking the existence of learning effect. Table 3-1 illustrates the ANOVA table to test the main effects on participants of azimuth cue angles, elevation cue angles and repetition as well as their two-way interactions on the absolute perceived localization error along the azimuth direction without front/back correction.

Table 3-2 demonstrates the summary of *p*-values obtained through ANOVAs to test the main effects on participants of azimuth cue angles, elevation cue angles and repetition as well as their two-way interactions on all eight types of perceived localization error. All significant main effects were confirmed to be consistent with the results of non-parametric Friedman two-way ANOVAs. As consistent with the literature, changes in cue directions along the azimuth and elevation directions had significant main effects on perceived localization errors (Leung, 2001 [34] and Wightman and Kistler, 1989 [35]). As these main effects are not the main objective of this thesis, they are not discussed in any detail. With the scope of this thesis, the aim of this preliminary experiment was to study the main effects of repetition, participant, and their two-way interactions.

With data from all 10 participants, for each of the 24 cue directions, repetition had no significant effect on all 8 types of perceived localization errors except in one case (ANOVAs: $p>0.05$, see Table 3-3 to Table 3-5). As the data distributions do not conform to the normal distribution ($p<0.0001$, Shapiro-Wilk test), these tests were repeated with Friedman two-way ANOVAs and the cases with significant effects of repetition increased to 6 and the significant case in ANOVA was included in these 6 cases out of 192 cases (Table 3-6 to Table 3-8). Further examination of the raw data of these 7 cases suggested no consistent learning trend (Table 3-9).

Table 3-1 ANOVA table to test the main effects of azimuth cue angles, elevation cue angles, repetition, and listeners as well as their two-way interactions on the absolute perceived localization error along azimuth direction without front/back correction

Source	**df**	**Sum of squares**	**Mean Square**	**F Value**	***p*-value**
REPEAT	5	2647.911	529.582	0.40	0.8511
LISTENER	9	25688.568	2854.285	2.14	0.0239
AZI	7	1477797.029	211113.861	158.26	<.0001
ELE	2	26225.132	13112.566	9.83	<.0001
AZI*ELE	14	29280.600	2091.471	1.57	0.0815
REPEAT*AZI	35	44095.879	1259.882	0.94	0.5623
REPEAT*ELE	10	8579.609	857.961	0.64	0.7774
REPEAT*LISTENER	45	64851.314	1441.140	1.08	0.3335
LISTENER*AZI	63	354775.775	5631.362	4.22	<.0001
LISTENER*ELE	18	62720.119	3484.451	2.61	0.0003
Error	1222	1630122.477	1333.979		
Corrected Total	1430	3726784.412			

Table 3-2 A summary table of p-values obtained through ANOVAs to test the main effects of azimuth cue angles, elevation cue angles, repetition, and listeners as well as their two-way interactions on all eight types of perceived localization error (FBC= error with front/back confusion correction; No FBC= error without front/back confusion correction; UDC= error with up/down confusion correction; No UDC= error without up/down confusion correction)

	Type of error							
	Azimuth				Elevation			
	Absolute		Relative		Absolute		Relative	
Main effect/ interactions	FBC	No FBC	FBC	No FBC	UDC	No UDC	UDC	No UDC
REPEAT	0.5647	0.8511	0.6910	0.5171	0.8668	0.5681	0.4715	0.1340
LISTENER	<.0001	0.0239	0.7505	0.1915	<.0001	<.0001	<.0001	<.0001
AZI	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0026	<.0001
ELE	0.1000	<.0001	0.6795	0.7977	<.0001	<.0001	<.0001	<.0001
AZI*ELE	0.3544	0.0815	0.0182	0.0897	0.0003	0.0373	<.0001	0.2248
REPEAT*AZI	0.1473	0.5623	0.0412	0.0249	0.4427	0.0127	0.9817	0.1212
REPEAT*ELE	0.3749	0.7774	0.8192	0.8355	0.6633	0.2835	0.3716	0.6531
REPEAT*LISTENER	0.0022	0.3335	0.0014	0.0892	0.9677	0.7240	0.8951	0.0006
LISTENER*AZI	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0006	<.0001
LISTENER*ELE	0.4463	0.0003	0.0457	0.1786	<.0001	<.0001	<.0001	<.0001

Table 3-3 *p*-values of ANOVAs conducted to test the main effects of repetition on all types of perceived localization errors along both azimuth and elevation directions for each of the 24 angles (Elevation angle = -40°)

Type of error			Ele	-40°							
			Azi	0°	45°	90°	135°	180°	225°	270°	315°
Azimuth	Absolute error	With front/back confusion correction		0.6235	0.1408	0.4680	0.9744	0.7368	0.4832	0.3757	0.3534
		Without front/back confusion correction		0.6066	0.6675	0.4680	0.8467	0.1244	0.5815	0.3757	0.5809
	Relative	With front/back confusion correction		0.8905	0.9914	0.7943	0.2534	0.0938	0.7818	0.3185	0.7049
		Without front/back confusion correction		0.3442	0.9638	0.3948	0.1460	0.2294	0.8027	0.3185	0.4145
Elevation	Absolute error	With up/down confusion correction		0.8775	0.5027	0.4770	0.8465	0.9025	0.5976	0.8931	0.9584
		Without up/down confusion correction		0.1299	0.9148	0.7921	0.7580	0.3944	0.3740	0.3854	0.8749
	Relative	With up/down confusion correction		0.2623	0.8010	0.2987	0.6776	0.3821	0.3213	0.3802	0.9001
		Without up/down confusion correction		0.1299	0.9140	0.8221	0.8786	0.3973	0.4565	0.3789	0.9463

Table 3-4 *p*-values of ANOVAs conducted to test the main effects of repetition on all types of perceived localization errors along both azimuth and elevation directions for each of the 24 angles (Elevation angle = 0°)

Type of error			Ele	0°							
			Azi	0°	45°	90°	135°	180°	225°	270°	315°
Azimuth	Absolute error	With front/back confusion correction		0.3005	0.5720	0.1831	0.7714	0.8449	0.8259	0.1419	0.5266
		Without front/back confusion correction		0.7201	0.5838	0.1831	0.7325	0.8800	0.8713	0.1419	0.9198
	Relative	With front/back confusion correction		0.4093	0.9067	0.9874	0.4785	0.1412	0.2505	0.6314	0.5930
		Without front/back confusion correction		0.8459	0.5505	0.9874	0.7647	0.6943	0.4489	0.6314	0.9052
Elevation	Absolute error	With up/down confusion correction		0.5990	0.4781	0.9493	0.4718	0.4350	0.8806	0.8332	0.7960
		Without up/down confusion correction		0.5990	0.4781	0.9493	0.4718	0.4350	0.8806	0.8332	0.7960
	Relative	With up/down confusion correction		0.5488	0.4354	0.7165	0.5947	0.3094	0.3304	0.9018	0.9922
		Without up/down confusion correction		0.5488	0.4354	0.7165	0.5947	0.3094	0.3304	0.9018	0.9922

Table 3-5 *p*-values of ANOVAs conducted to test the main effects of repetition on all types of perceived localization errors along both azimuth and elevation directions for each of the 24 angles (Elevation angle = 40°)

Type of error			Ele	40°							
			Azi	0°	45°	90°	135°	180°	225°	270°	315°
Azimuth	Absolute error	With front/back confusion correction		0.5355	0.9944	0.9506	0.6533	0.8474	0.0108	0.3993	0.9772
		Without front/back confusion correction		0.4211	0.2099	0.9506	0.2861	0.6899	0.2249	0.3993	0.3134
	Relative	With front/back confusion correction		0.5274	0.3880	0.5967	0.2287	0.1729	0.8515	0.8049	0.9532
		Without front/back confusion correction		0.9267	0.1995	0.5967	0.0735	0.0547	0.6697	0.8049	0.3997
Elevation	Absolute error	With up/down confusion correction		0.3793	0.7937	0.6559	0.9450	0.5466	0.9918	0.4233	0.9341
		Without up/down confusion correction		0.9441	0.9064	0.9555	0.9330	0.5546	0.8686	0.4569	0.9174
	Relative	With up/down confusion correction		0.9356	0.9665	0.6657	0.8154	0.9271	0.9102	0.4831	0.6898
		Without up/down confusion correction		0.9989	0.9155	0.9621	0.8637	0.9886	0.7003	0.5872	0.6247

Table 3-6 *p*-values of Friedman two-way ANOVAs conducted to test the main effects of repetition on all types of perceived localization errors along both azimuth and elevation directions for each of the 24 angles (Elevation angle= -40°)

Type of error			Ele	-40°							
			Azi	0°	45°	90°	135°	180°	225°	270°	315°
Azimuth	Absolute error	With front/back confusion correction		0.266	0.140	0.135	0.612	0.462	0.449	0.260	0.076
		Without front/back confusion correction		0.370	0.382	0.135	0.557	0.159	0.590	0.260	0.209
	Relative	With front/back confusion correction		0.606	0.773	0.647	0.096	0.143	0.696	0.053	0.323
		Without front/back confusion correction		0.370	0.382	0.427	0.086	0.437	0.981	0.053	0.062
Elevation	Absolute error	With up/down confusion correction		0.962	0.467	0.367	0.973	0.742	0.639	0.988	0.212
		Without up/down confusion correction		0.006	0.660	0.332	0.646	0.265	0.070	0.531	0.182
	Relative	With up/down confusion correction		0.250	0.776	0.199	0.798	0.732	0.482	0.990	0.510
		Without up/down confusion correction		0.006	0.606	0.314	0.739	0.265	0.107	0.427	0.300

Table 3-7 *p*-values of Friedman two-way ANOVAs conducted to test the main effects of repetition on all types of perceived localization errors along both azimuth and elevation directions for each of the 24 angles (Elevation angle= 0°)

Type of error			Ele	0°							
			Azi	0°	45°	90°	135°	180°	225°	270°	315°
Azimuth	Absolute error	With front/back confusion correction		0.061	0.324	0.088	0.743	0.462	0.908	0.068	0.491
		Without front/back confusion correction		0.447	0.553	0.088	0.833	0.437	0.750	0.068	0.833
	Relative	With front/back confusion correction		0.250	0.710	0.981	0.427	0.307	0.491	0.538	0.854
		Without front/back confusion correction		0.447	0.553	0.981	0.732	0.329	0.718	0.538	0.698
Elevation	Absolute error	With up/down confusion correction		0.314	0.620	0.436	0.532	0.015	0.948	0.541	0.187
		Without up/down confusion correction		0.314	0.620	0.436	0.532	0.015	0.948	0.541	0.187
	Relative	With up/down confusion correction		0.461	0.595	0.973	0.166	0.230	0.084	0.920	0.695
		Without up/down confusion correction		0.461	0.595	0.973	0.166	0.230	0.084	0.920	0.695

Table 3-8 p-values of Friedman two-way ANOVAs conducted to test the main effects of repetition on all types of perceived localization errors along both azimuth and elevation directions for each of the 24 angles (Elevation angle= 40°)

Type of error			Ele	40°							
			Azi	0°	45°	90°	135°	180°	225°	270°	315°
Azimuth	Absolute error	With front/back confusion correction		0.663	0.893	0.903	0.677	0.248	0.016	0.280	0.792
		Without front/back confusion correction		0.832	0.025	0.903	0.551	0.411	0.552	0.280	0.118
	Relative	With front/back confusion correction		0.420	0.571	0.564	0.529	0.256	0.079	0.097	0.783
		Without front/back confusion correction		0.832	0.017	0.564	0.176	0.062	0.767	0.097	0.245
Elevation	Absolute error	With up/down confusion correction		0.238	0.345	0.357	0.998	0.140	0.875	0.503	0.929
		Without up/down confusion correction		0.324	0.562	0.776	0.809	0.608	0.992	0.416	0.997
	Relative	With up/down confusion correction		0.721	0.847	0.640	0.912	0.736	0.615	0.263	0.259
		Without up/down confusion correction		0.840	0.982	0.615	0.636	0.201	0.763	0.706	0.519

Table 3-9 Friedman test result on type of error and angle which repetition had significant effect

Angle	Type of error	Chi-square value	*p*-value	Repetition number in ascending order of error
(0°,-40°)	Elevation localization error	16.230	0.006	5,2,1,4,3,6
	Absolute elevation localization error	16.230	0.006	5,2,1,4,3,6
(45°,40°)	Azimuth localization error	13.768	0.017	1,2,5,3,4,6
	Absolute azimuth localization error	12.794	0.025	1,2,5,3,4,6
(180°,0°)	Absolute elevation localization error	14.080	0.015	1,2,4,5,3,6
	Absolute elevation localization error with up-down confusion corrected	14.080	0.015	1,2,4,5,3,6
(225°,40°)	Absolute azimuth localization error with front-back confusion corrected	13.951	0.016	2,5,3,6,1,4

As shown in Table 3-1 and Table 3-2, there were significant two-way interactions between the effects of listeners and the effects of repetitions in some cases ($p<0.02$, ANOVA test, Table 3-2). Because the effects of repetitions have no significant two-way interactions ($p>0.13$, ANOVAs, Table 3-2), further ANOVAs had been conducted to test the effects of repetitions on data from each listener using cue directions as repeated measures.

The results indicated that significant main effects of repetition were found with data from four participants (PE3, 4, 6, 8: see Table 3-10). Further examinations of the trend of changes in perceived errors suggest that only participant PE3 showed a potential trend of reducing perceived error with increasing repetitions. In summary, as only one listener out of the 10 participants demonstrated a potential trend of reducing perceived localization error with repetitions, it can be concluded that the effects of repetition does not, in general, have any significant main effect on perceived localization error. This supports the hypothesis H1. However, to maintain complete accuracy, experiments one and two were repeated a total of 6 times in order for the effects of repetition to be measured.

Table 3-10 ANOVA test result on type of error and listeners which repetition had significant effect

Type of error	**Listener**	*p*-value	**Repetition number in ascending order of error**
Azimuth localization error with front-back confusion corrected	PE3	0.0084	2, 6, 5, 1, 3, 4
	PE4	0.0267	5, 1, 2, 6, 3, 4
	PE6	0.0079	6, 2, 1, 5, 3, 4
Absolute azimuth localization error with front-back confusion corrected	PE3	0.0153	6, 3, 4, 5, 2, 1
	PE8	0.0064	6, 1, 3, 5, 2, 4
Absolute azimuth localization error	PE8	0.0002	1, 6, 3, 5, 2, 4

3.7.2. Inter-subject variability

The objective of investigating the inter-subject variability was to discuss the possibility of dividing participants into groups according to their sound localization performance. After separating the participants into groups, suitable sets of HRTF sets were given to each group to improve their sound localization performance. As there will be a large variation within each group of sound localization, a standard HRTF set would not have been suitable. To divide listeners into groups, listeners' sound localization performance in the same set of HRTF should have large varieties. Otherwise, a single set of HRTF is already enough to fit the population.

The inter-subject and intra-subject standard deviation of localization errors were then calculated for each angle. The results showed that the inter-subject standard deviation of localization errors were statistically larger than the one of intra-subject. The significant levels of each type of errors were summarized in Table 3-11. The results indicated that the inter-subject variability was larger than the intra-subject variability in the area of sound localization performance. With a large inter-subject variability, it implied that using one set of non-individualized HRTF set would not be suitable for our group. This finding suggested that grouping participants according to their sound localization performance was possible.

Table 3-11 Result of Mann-Whitney U test comparing inter-subject standard deviation and intra-subject standard deviation of 8 types of error

Type of error			*p*-value
Azimuth	Absolute error	Front-back confusion correction	0.002
		No front-back confusion correction	<0.001
	Relative	Front-back confusion correction	<0.001
		No front-back confusion correction	<0.001
Elevation	Absolute error	Front-back confusion correction	<0.001
		No front-back confusion correction	<0.001
	Relative	Front-back confusion correction	<0.001
		No front-back confusion correction	<0.001

Chapter 4. Defining the characteristics of directional bands

HRTF sets were analyzed based upon directional band before clustering. Frequency band characteristics were defined referenced to Blauert (1969/70) [7], Hebrank and Wright (1974) [12], Middlebrooks (1992) [18] and Langendijk and Bronkhorst (2002) [16]. The band characteristics were summarized in Figure 4.1, 4.2 and Table 4-1.

Nine band characteristics, bands FA – FF and BA – BC, were defined for analyzing front and back cues. The boosted band defined by Blauert (1969/70) [7] were adopted in bands FA (150-540 Hz), FB (1900-2900 Hz), FC (3600-5800 Hz), BA (720-1700 Hz) and BB (7400 – 11100 Hz).

In Blauert's study, the average level difference between the HRTF at (0°,0°) and (180°,0°) were calculated. This calculation method was adopted to retrieve the band characteristics of bands FA-FC, BA and BB. Band FD (3800-8000 Hz) and FE (13200-16000Hz) are defined according to the findings in Hebrank and Wright (1974) [12]. Their findings concluded that 3.9 to 8 kHz low-pass cut-offs and 13.2 and 15.3 kHz high-pass cut-offs are related to frontal perception. Band FD starts from 3800 Hz because the authors found that only frequency band 3800-16000 Hz is related to sound localization. For the same reason, band FE stops at 16000Hz.

Langendijk and Bronkhorst (2002) [16] concluded that the most important

Langendijk and Bronkhorst (2002) [16] concluded that the most important front back cues are at 8-16 kHz. Band FF (8000-16000 Hz) is defined according to the result from Langendijk and Bronkhorst (2002) [16]. Langendijk and Bronkhorst (2002) [16] found that the variations in the spectral content in 8 – 16 kHz are essential for frontal perception. Pattern matching method was used in band FF to calculate the similarity between different set of HRTFs. For the methodology of pattern matching, please refer to Appendix A.

Band BC (10000-13200 Hz) is defined according to the result from Blauert (1983) and Hebrank and Wright (1974) [12]. Blauert (1997) [37] found that the probability of subjects perceiving sounds at the back is high when broadband noise with peaks at 10-14 kHz was presented. A similar conclusion was found in Hebrank and Wright (1974) [12]. They reported that 10.0 kHz high-pass cutoffs are related to back perception. However, another part of the result from their resharch showed that frequency above 13200 Hz is related to frontal perception. It is a contradiction with band FE if band BC covers the range from 13200 to 14000 Hz. Hence, band BC is truncated and ends at 13200 Hz.

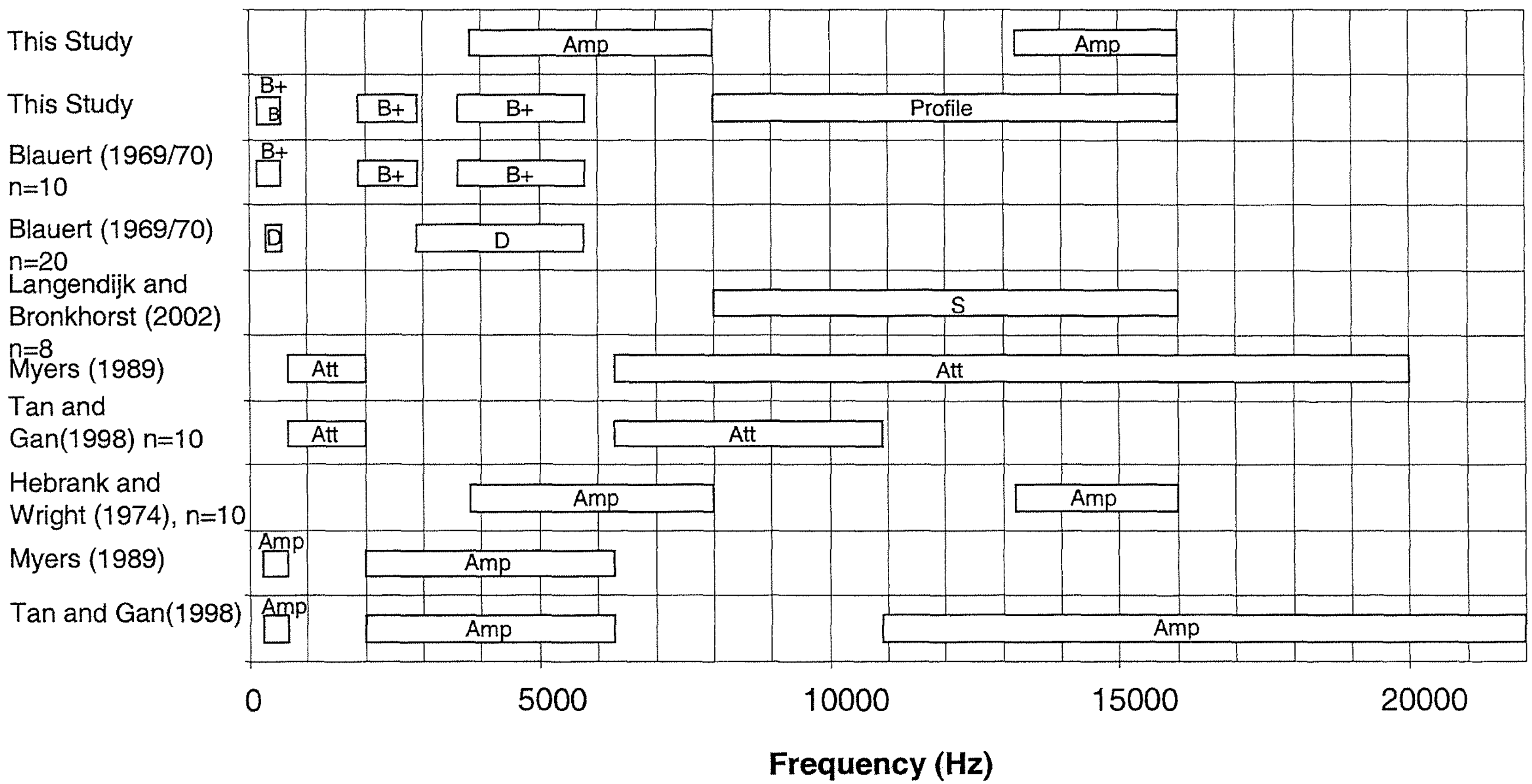

Figure 4.1 Summary of band characteristics for frontal cues (the key is located after Figure 4.2)

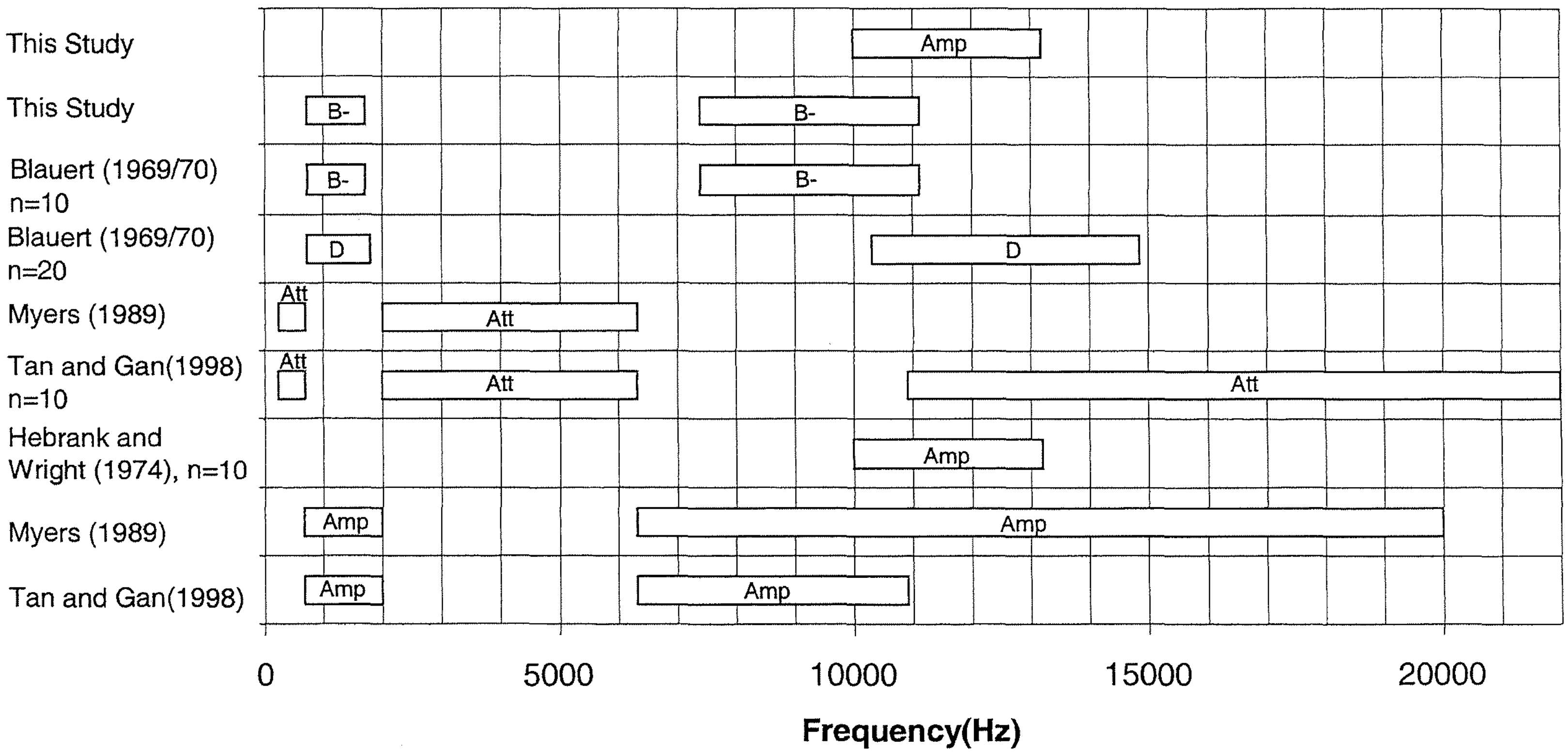

Figure 4.2 Summary of band characteristics for back cues

B+ B-	Boosted band: It indicates the frequency bands within which the modulus of a HRTF representing the front direction is significantly different from that of a HRTF representing the back direction. A 'positive' (B+) boosted band indicate that the front HRTF has a higher magnitude and a 'negative' (B-) boosted band indicate that the back HRTF has a higher magnitude. These definitions are adapted from Blauert (1969/70).
D	Directional band: It indicates the frequency ranges of audio sound that are associated with some dominating perceived directions. This definition is adapted from Blauert (1969/70) and the level of domination was set at a 90% confidence level.
Profile	Spectral profile: It indicates a frequency range of HRTF modulus where the spectral profile is expected to contribute towards the direction of the binaural cues that the HRTF will produce. See section 3.1 for further explanations.
S	Spectral content variations: Langendijk and Bronkhorst (2002) compared the sound localization performance of cues generated by HRTF with and without spectral content variations in certain frequency ranges.
Amp	Amplification: It means that HRTF modulus have higher average sound magnitudes within certain frequency ranges than the average level of the whole HRTF. The effects of increasing amplification magnitude were not studied.
Att	Attenuation: It means that HRTF modulus have lower average sound magnitudes within certain frequency ranges than the average level of the whole HRTF. The effects of decreasing amplification magnitude were not studied.
Peak	Peak: The peaks apply only to the upward cues reported by Hebrank and Wright (1974). Data indicated that as long as HRTFs have a peak of 1/12-octave bandwidth with centre frequencies within 4950 to 9350 Hz, listeners will have higher chances to report that the cue generated from the HRTFs are coming from above.

4.1. Rationale to use pattern matching methodology to quantify the spectral profile bands

This thesis proposed using the pattern matching method to capture the spectral profiles within pattern in designated frequency bands (Band FF). Langendijk and Bronkhorst (2002) [16] reported that the variations of spectral profiles of HRTFs within certain frequency bands can affect the perceived front-back and up-down direction of the binaural cues generated from those HRTFs. The patterns and variations in spectral profiles are caused by direction-dependent resonances of external ear, e.g., Shaw and Teranishi (1968) [22], Hebrank and Wright (1974) [12]; and Batteau (1967) [5]. Traditional principal component analysis (PCA) does not work in this situation. Comparison between PCA and pattern matching is described in Appendix B.

Inter-subject variations of the pinna dimensions are related to the inter-subject variations of the spectral profile of HRTF (Algazi *et. al.*, 2001 [2]). Prof. Algazi and his colleagues measured the HRTFs and the ear dimensions of 45 subjects. They found that significant correlations between the spectral profile in HRTF and ear dimensions. Since people have different ear dimensions, the spectral profile in HRTF could be different. This leads to our interest in investigating the pattern within various frequency bands of HRTF. We hope this will aid us to group listeners according to their sound localization

Table 4-1 Frequency band characteristics for front-back cues analysis

Band	Frequency	Nature of the band characteristics	Frontal/ Backward cue	Values to be retrieved
FA	150 – 540 Hz	Boosted	Frontal	1. Average level difference between HRTF at a0e0 and a180e0 (a0e0-a180e0)
FB	1900 – 2900 Hz	Boosted	Frontal	1. Average level difference between HRTF at a0e0 and a180e0 (a0e0-a180e0)
FC	3600 – 5800 Hz	Boosted	Frontal	1. Average level difference between HRTF at a0e0 and a180e0 (a0e0-a180e0)
FD	3800– 8000 Hz	Amplification	Frontal	1. Average level
FE	13200 – 16000 Hz	Amplification	Frontal	1. Average level
FF	8000 – 16000 Hz	Spectral content related	Frontal	1. Pattern matching method 2. Standard deviation of the modulus in the band
BA	720 – 1700 Hz	Boosted	Backward	1. Average level difference between HRTF at a180e0 and a0e0 (a180e0-a0e0)

BB	7400 – 11100 Hz	Boosted	Backward	1. Average level difference between HRTF at a180e0 and a0e0 (a180e0-a0e0)
BC	10000 – 13200 Hz	Amplification	Backward	1. Average level

Notes:
Average level: The average level of the frequency band
Centre frequency of the peak: The frequency of the peak which has the maximum level
Height of the peak: The level at the centre frequency of the peak minus the average level of the frequency band

The frequency band characteristics are quite different from the approach suggested by Myers (1989) [27]. In Myers (1989) [27], since it is a patent, the reference or the way to determine the band definitions is not well explained. The bands are not overlapped and the joint of all the bands covers the whole audible frequency range. In this study, the band characteristics are referenced to literatures which include experimental results to investigate the presence of spectral cues. Since a certain frequency range is important for both front-back and up-down cues, the bands were overlapped. This study focuses more on quantifying the band characteristics. Different values were retrieved from each frequency band in order to facilitate the clustering process.

4.2. Cross examination of band characteristics in different sets of HRTFs

To cluster sets of HRTFs into groups with satisfactory results, the features extracted should have variations among the population. In order to verify the band characteristics selected are appropriate for clustering, a preliminary study was undertaken to investigate whether the band characteristics can be found in selected HRTFs from previous studies.

HRTFs from the frontal direction (0° azimuth, 0° elevation) were selected from the MIT KEMAR data set (Gardner and Martin 1995 [11]), CIPIC HRTF database (Algazi *et. al.* 2001 [2]), and our own HKUST dataset (the author's individualized HRTFs, referred as KHN) to test the frontal characteristics. Likewise, HRTFs for the back direction (180° azimuth, 0° elevation), and up and down directions (0° azimuth, 60° and -30° elevation) were used to test the backward cues and elevation cue characteristics, respectively. The results of this preliminary test are tabulated in Table 3-3. The results showed that the HRTF sets have variations in band characterization defined in previous sections.

Table 4-2 Preliminary study result of frequency band characteristics

Left/Right ear	MIT KEMAR		KHN		CIPIC021		CIPIC162	
	L	R	L	R	L	R	L	R
FA	×	✓	✓	×	×	×	×	×
FB	✓	✓	✓	×	✓	✓	×	✓
FC	✓	✓	✓	✓	✓	✓	×	✓
FD	1.6 dB	4.3 dB	-4.3 dB	-6.8 dB	1.8 dB	2.7 dB	-3.9 dB	-0.9 dB
FE	2.8 dB	0.5 dB	-6.5 dB	-7.5 dB	2.9 dB	3.0dB	-9.8 dB	1.3 dB
FF	N/A							
BA	×	×	✓	×	×	×	×	×
BB	×	×	✓	×	✓	✓	✓	✓
BC	0.5 dB	-13.3 dB	-9.0 dB	-4.7 dB	-0.9 dB	-2.9 dB	-2.8 dB	1.9 dB

Notes:

FA, FB, FC	✓	The average SPL of the band at 0° azimuth 0° elevation – The average SPL of the band at 180° azimuth 0° elevation > 0.7 dB
	×	The average SPL of the band at 0° azimuth 0° elevation – The average SPL of the band at 180° azimuth 0° elevation ≤ 0.7 dB
BA, BB	✓	The average SPL of the band at 180° azimuth 0° elevation – The average SPL of the band at 0° azimuth 0° elevation > 0.7 dB
	×	The average SPL of the band at 180° azimuth 0° elevation – The average SPL of the band at 0° azimuth 0° elevation ≤ 0.7 dB

Chapter 5. Overview of the processes of standard HRTF sets selection

Each set of standard HRTFs was expected to produce a 3-D sound with a satisfactory sound localization performance for a certain proportions of the population. To determine the suitable standard HRTF sets for listeners, the overview processes were shown in Figure 5.1.

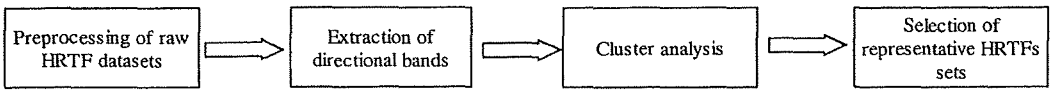

Figure 5.1 Overview processes to determine the suitable standard HRTF sets

5.1. Preprocessing of raw HRTF datasets

There are several sets of HRTF data used in this study. These include MIT KEMAR HRTF (Gardner and Martin 1995 [11]), CIPIC HRTF database (Algazi *et. al.* 2001 [2]), LISTEN HRTF database (Ircam and AKG acoustics, 2004 [13]) and the measured HRTF set of the author. The first three HRTF data sets are available online for research free of charge.

The HRTFs were equalized before extracting information from them. For the technical details of equalization of each HRTF data set, Kapralos *et. al.* (2003) [14] provides detailed information. For MIT KEMAR HRTFs set, since its HRIRs' average sound level is 84 dB while the level of other sets is 0dB, the levels of HRIRs were normalized.

Table 5-1 Number of coefficients in each HRTF data set

	Number of coefficients in HRIR
MIT KEMAR	512
CIPIC	200
LISTEN	512
HKUST	256

Since the number of HRIR coefficients varies in different sets of HRTF data sets (see Table 4-1), it is not fair to compare them without adjusting the number of coefficient of HRTFs sets at the same number. A simple rectangular windowing of the impulse response was performed. This method was described in Begault (1994) [6]. The first 200 coefficients were retained for analysis and experiment uses.

5.2. Extraction of directional bands

HRTFs of four angles from each ear, including frontal angle (0°,0°), backward angle (180°, 0°), upward angle (0°,60°) and downward angle (0°,-30°), were selected for analysis. The directional bands to be studied for each selected HRTF were summarized in Table 4-2. Although this study focuses on front-back and back-front confusion, upward and downward sound cues were added to increase varieties to listeners. It is worth noting that the upward and downward angles were selected as (0°, 60°) and (0°, -30°) respectively because the human perception deviation at those two angles are smaller among different upward and downward angles. (Algazi *et. al.* 2001 [2])

Table 5-2 Directional bands studied in different angles of HRTF

Angle of HRTFs	Directional bands studied
Frontal angle (0°,0°)	FD, FE, FF
Backward angle (180°, 0°)	BC
The band SPL of (0°,0°) – The band SPL of (180°, 0°)	FA, FB, FC
The band SPL of (180°,0°) – The band SPL of (0°, 0°)	BA, BB

Different features were extracted from each directional band. Table 4-1 summarized the features extracted. For each ear, these features were extracted and organized as a vector. The vectors were normalized before clustering. These vectors were used in cluster analysis and sets of HRTFs were grouped into clusters. Standard HRTF sets were selected from each cluster. Sets of standard HRTFs were used for producing spatial sound in the experiment described in the later part of this thesis.

5.3. Cluster analysis and selection of standard HRTF sets

The objective of cluster analysis on extracted data was to group the HRTF sets into clusters. After the analysis, standard HRTF sets were selected from each cluster. Details are described in Chapter 6.

The whole analysis process was organized using the ear as a base rather than the subject. Since left and right ears are not perfectly symmetrical, if a subject-based analysis was conducted, i.e. comparing each pair of ears, was conducted, the unique properties of each ear would be diluted. On the other hand, monaural properties in HRTFs were the focus of this study. Hence, ear-based analysis was selected.

Chapter 6. Cluster analysis on band characteristics

6.1. A review of cluster algorithm

There are several common hierarchical agglomerative cluster methods including single linkage, complete linkage, average linkage and Ward's minimum-variance method. Single linkage and complete linkage are the most common hierarchical agglomerative clustering methods. Single linkage algorithm joins an existing group to the other group with the highest similarity in any member in the group. In other words, the definition of the similarity between two existing groups is a pair of members, one member from each existing group, with the highest similarity among all possible combinations between two existing groups. On the other hand, a complete linkage algorithm is the opposite of single linkage algorithm. It joins two existing groups together with the highest similarity in all members between two groups. The similarity between two groups is defined as the lowest similarity among all possible combinations of the member pairs from two existing groups.

Ward's minimum-variance method (Ward 1963 [23]) is designed to minimize the intra-cluster variance. It has a strong tendency to split data into groups of roughly equal size with clusters shaped as hyperspheres. No cluster consists of only one or a few elements (Milligan 1980 [20]). It is an advantage in clustering raw HRTF data. If the number of elements in a cluster is too small, the standard HRTF set of this cluster can only represent a very small number of the population. The major drawback of Ward's method, however, is sensitive to outliers.

Complete linkage performs better than single linkage in terms of intra-cluster variance. Single linkage has a tendency to form long and elongated clusters. Also, it is likely to form a chain of large clusters. Complete linkage algorithm has a tendency to form relatively compact, hyperspherical clusters with lower intra-cluster variance. (Aldenderfer and Blashfield 1984 [1]) The properties of average linkage algorithm fall between complete linkage and single linkage algorithm.

In this chapter, these algorithms were compared with different metrics. The most suitable algorithm was selected to cluster the data.

6.2. Selection of cluster algorithm

Different cluster algorithms have different properties and hence each of them is suitable for a specific kind of data. An appropriate algorithm for this study should fulfill these criteria: i) The number of data contained in each cluster should not be too little. A small-sized cluster indicates that the standard HRTF sets satisfy only a small portion of listeners. It is not practical for industrial products; ii) Robust to outliers. The cluster result should be stable even if an outlier is added; iii) Satifactory cluster quality. The quality of the cluster result can be measured in terms of inter-cluster distance and intra-cluster distance. A larger inter-cluster distance indicates clusters were well separated. A smaller intra-cluster distance indicates data in the same cluster have high similarity.

The following subsections discuss the procedures of cluster algorithm selection based on those three criteria.

6.2.1. Robustness to outliers

A reliability test is essential for evaluating the robustness of the clustering result. A good clustering result should be robust to noise. In other words, taking out one of the data should not affect the form of clusters. This method is called "Leave-one-out". (Duda *et. al.*, 2001 [31]) The cluster result of leave-one-out is compared with the reference result which performed clustering with all of the data. The reliability score is defined as:

$$\text{Reliability Score} = \frac{\text{Number of data matches the reference result}}{\text{Total number of data - 1}}$$

The test was repeated for the number of times equal to the total number of data since one datum was taken out each time. Hence, the total reliability score is defined as:

$$\text{Total Reliability Score} = \frac{\sum_{i=1}^{N} \text{Number of data matches the reference result after taking data i out}}{N \times (N - 1)}$$

Where N is the total number of data

Table 6-1 Total reliability score of cluster algorithms in different number of clusters and directional bands (round up to 1 decimal place)

	Directional bands/ Number of cluster				
	Front			Back	
	5	8	9	5	9
Single	100.0%	100.0%	99.9%	100.0%	100.0%
Complete	80.8%	86.2%	87.9%	93.1%	93.0%
Average	90.8%	84.5%	88.6%	99.5%	97.9%
Ward's	83.1%	91.0%	93.2%	93.0%	90.1%

From Table 6-1, the rank of the cluster algorithms from the best to the worst is single linkage, average linkage, complete linkage and Ward's minimum variance method. Single linkage cluster algorithms produced excellent total reliability scores because it is likely that they produced singleton clusters. Those singletons are likely to be the outlier of the whole set of data and most of the data falls into one cluster. Hence, the clusters are stable even one piece of data was taken out.

6.2.2. Cluster quality

The quality of clustering can be measured in two dimensions, inter-cluster distance and intra-cluster distance. An appropriate clustering method should have a large inter-cluster distance with small intra-cluster distance. A large inter-cluster distance refers to clusters that are well separated and stable; a small intra-cluster distance refers to a high similarity of data in the same cluster.

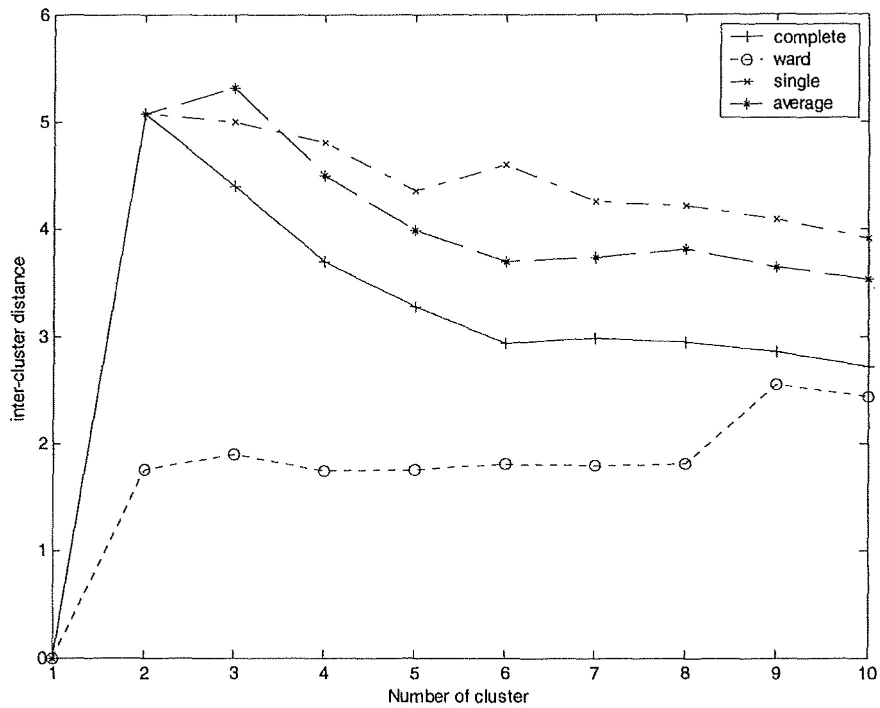

Figure 6.1 Inter-cluster distance of the HRTF data (Front bands) using different cluster algorithms

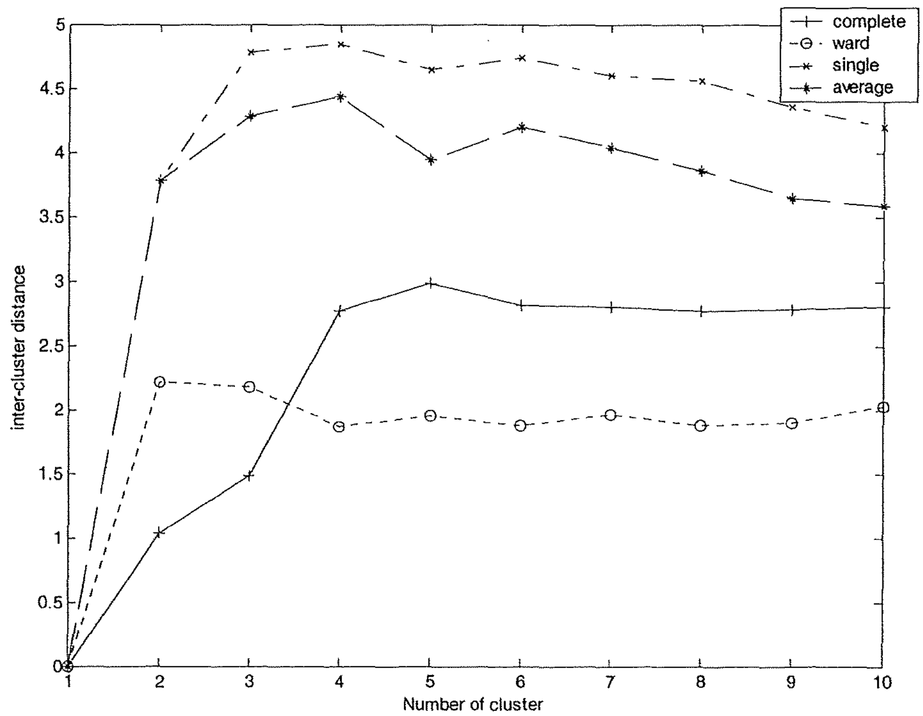

Figure 6.2 Inter-cluster distance of the HRTF data (Back bands) using different cluster algorithms

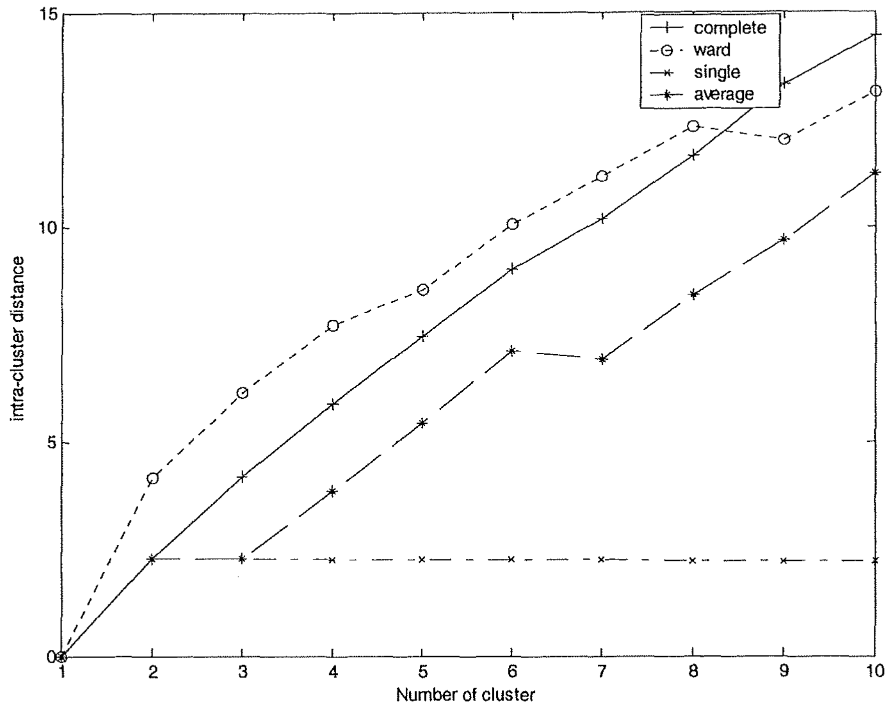

Figure 6.3 Intra-cluster distance of the HRTF data (Front bands) using different cluster algorithms

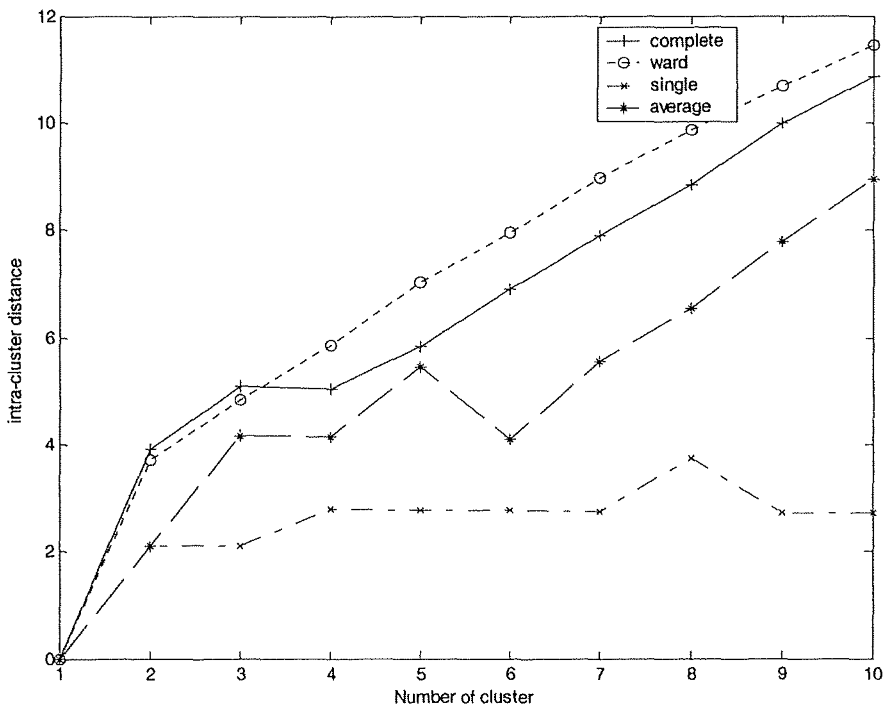

Figure 6.4 Intra-cluster distance of the HRTF data (Back bands) using different cluster algorithms

From Figure 6.1 to Figure 6.2, single linkage and average linkage algorithms outperform Ward's minimum variance method and complete linkage algorithms in terms of inter-cluster distance. Single linkage algorithms generally produced a larger inter-cluster distance rather than average linkage. This is because single linkage produced many singleton clusters, which are likely to be outliers, and hence the inter-cluster distances are larger than others.

It is also worth noting that single linkage algorithms produced especially small intra-cluster distances as shown in Figure 6.3 to Figure 6.4. This is because the single linkage algorithm has a tendency to produce a singleton cluster and hence the intra-cluster distance can be kept small. The average linkage produced smaller intra-cluster distances than the complete linkage and Ward's minimum variance method when the number of clusters is larger than three. From these observations, single linkage has the best overall performance and average linkage is the second best. Complete linkage algorithm and Ward's minimum variance method have similar inter-cluster and intra-cluster distances in their cluster result.

6.2.3. Number of data contained in each cluster

One of the objectives of this study is to provide sets of standard HRTFs which are suitable for most people. A standard HRTFs set which only satisfies a small portion of the population is not cost effective for manufacturers to produce as it requires virtual surround sound headphones. Hence, the cluster algorithms which only produce small-sized clusters are not appropriate to this study.

A single linkage algorithm does not satisfy this criterion. Because of its algorithm nature, the cluster result contains many singletons (Table 6-2 to Table 6-3). Hence, single linkage was not considered for analysis of the band characterization data.

Table 6-2 Number of data contained in each cluster at different number of clusters and cluster algorithm. (Front bands) Please refer to appendix C for the cluster results.

	Number of Clusters						
	2		5		9		
Single	195	1	192	1	188	1	1
			1	1	1	1	1
			1		1	1	1
Complete	195	1	40	43	24	5	11
			45	67	13	12	42
			1		43	45	1
Average	195	1	170	13	47	8	4
			11	1	10	1	111
			1		13	1	1
Ward's	78	118	20	22	15	1	14
			60	16	46	18	21
			78		39	20	22

Table 6-3 Number of data contained in each cluster at different number of clusters and cluster algorithm. (Back bands) Please refer to appendix C for the cluster results.

	Number of Clusters						
	2		5		9		
Single	195	1	191	1	1	1	187
			2	1	1	1	1
			1		2	1	1
Complete	121	75	51	5	35	13	45
			1	74	6	17	23
			65		51	5	1
Average	195	1	185	6	121	20	13
			1	3	31	2	1
			1		6	1	1
Ward's	138	58	26	32	22	11	13
			69	40	27	6	20
			29		36	32	29

6.2.4. Decision on cluster algorithm selection

The robustness of cluster algorithms was discussed in section 6.2.1. The performance ranks from the best to the worst were single linkage, average linkage, complete linkage and Ward's minimum variance method.

In section 6.2.2, it was found that the single linkage algorithm produced exceptionally large inter-cluster distances and small intra-cluster distances. The average linkage algorithm performs better than Ward's minimum variance method and complete linkage algorithm in terms of inter-cluster and intra-cluster distances.

However, from the discussion of section 6.2.3, the single linkage algorithm was excluded because all of its clusters produced are singletons. Therefore the average linkage algorithm was selected for this study.

6.3. Determination of the number of cluster

The average linkage algorithm produces dendrograms with a nested tree structure. The dendrograms for front band and back bands are shown in Figure 6.5 and Figure 6.6 respectively. It is important to determine the cutoff point of the tree for the optimal number of clusters. A commonly used heuristic method was employed in this study (Aldenderfer *et. al.*, 1984 [1]). Amalgamation coefficients were studied to find a “stable” number of clusters. The stable number of clusters, n, is the amalgamation coefficient at the cut off point for n clusters minus the amalgamation coefficient at the cut off point for n+1 clusters. However, the number of clusters should not be too little, otherwise, it is similar to the idea of using one non-individualized HRTF set to represent the whole population. The number of cluster of each direction was shown in Table 6-5.

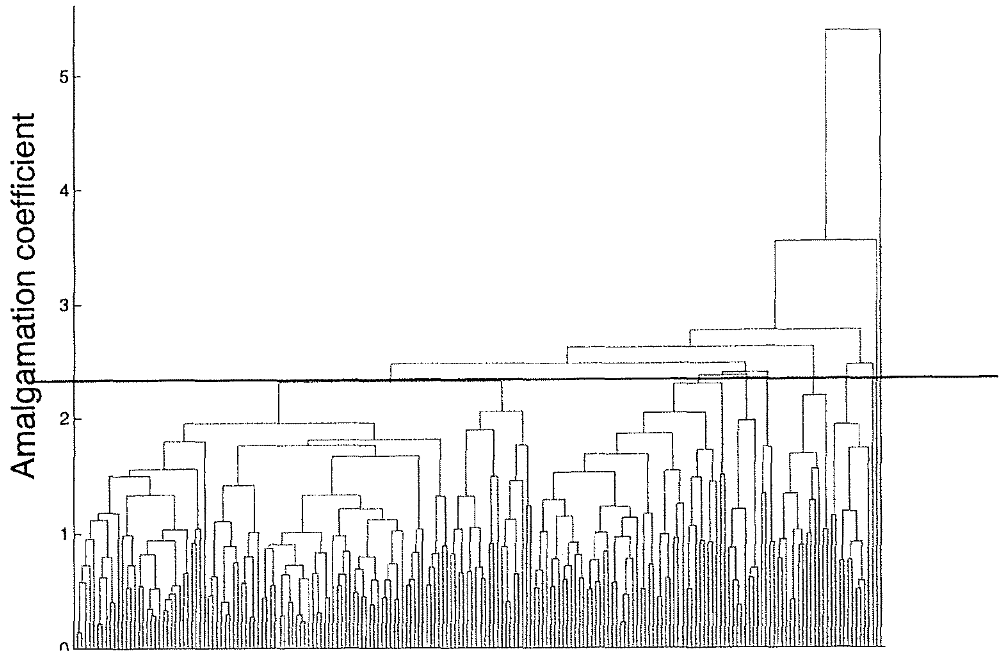

Figure 6.5 Dendrogram of clustering result of front bands with average linkage algorithm

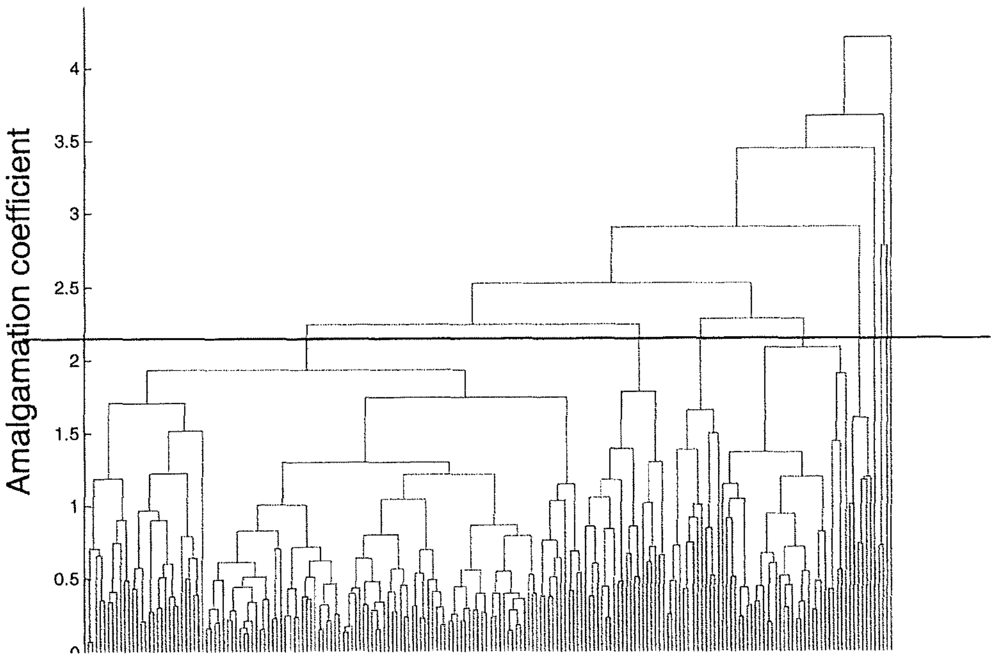

Figure 6.6 Dendrogram of clustering result of back bands with average linkage algorithm

Table 6-4 Number of cluster of each direction

Direction	**Number of cluster**	**Number of singletons**
Front	9	3
Back	9	3

6.3.1. Visualization of the cluster result with Principal Component Analysis

The band characteristics data have a high dimension and hence the cluster results are impossible to be visualized if it is not transformed to a lower dimension. Principal Component Analysis (PCA) was employed to reduce the data to 3-D and hence the data can be visualized and inspected in a convenient way. Figure 6.7 to Figure 6.8 shows the scatter plots of the cluster result; clusters were well separated and did not mix together.

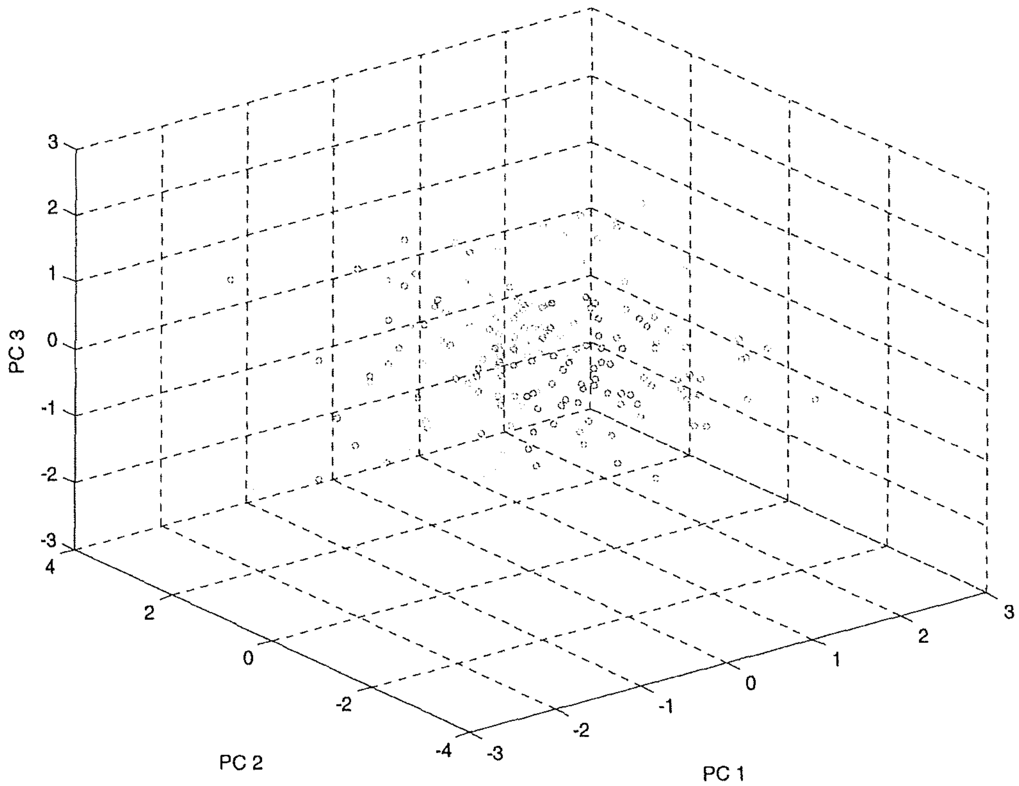

Figure 6.7 Scatter plot of first 3 principal components of cluster result using back bands (Average linkage algorithm, 9 clusters, 196 data), data having different colour represent belonging to different cluster

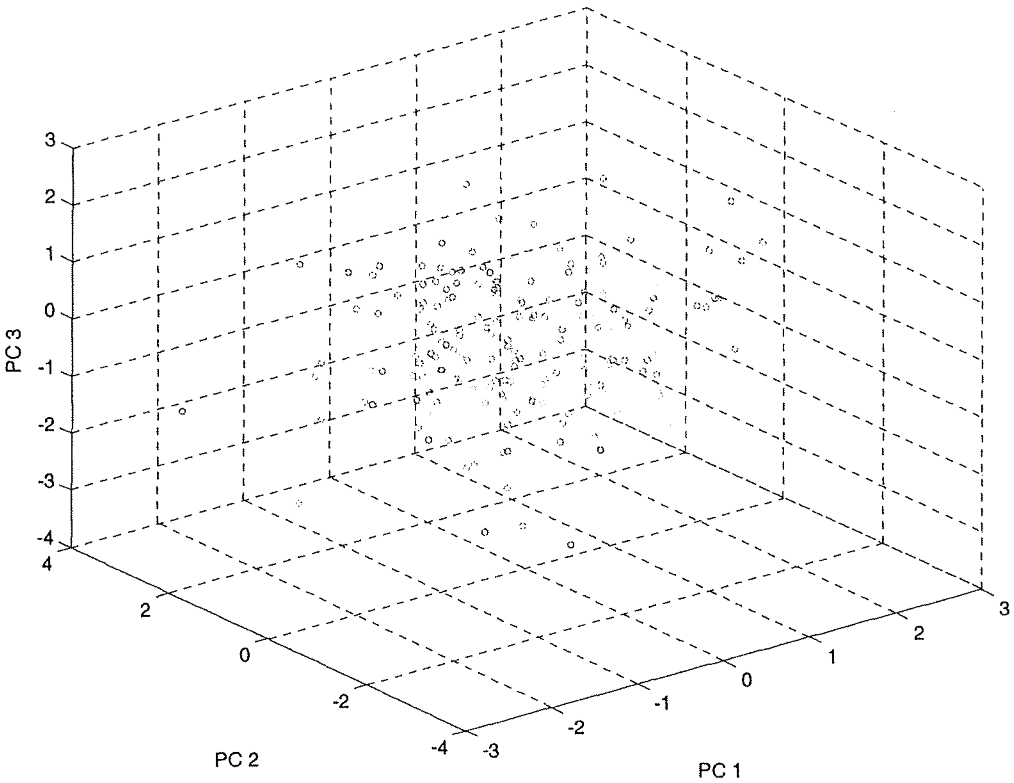

Figure 6.8 Scatter plot of first 3 principal components of cluster result using front bands (Average linkage algorithm, 9 clusters, 196 data), data having different colour represent belonging to different cluster

6.4. Standard set of HRTFs

Standard sets of HRTFs were selected from each cluster after running the cluster analysis. For each cluster, the HRTFs set with lowest sum of distances with other HRTFs sets in the same cluster were selected as standard set of HRTF. Table 6-5 illustrates the originates of the standard HRTF sets.

Table 6-5 Originates of standard HRTF sets

Direction cue	**Standard HRTF set number**	**HRTF database**	**Subject number**	**Left/Right ear**
Front	1	CIPIC	15	R
	2	CIPIC	28	R
	3	CIPIC	148	L
	4	LISTEN	1054	L
	5	LISTEN	1046	R
	6	LISTEN	1059	L
	7	MIT KEMAR	N/A	L+R*
Back	1	CIPIC	12	L
	2	CIPIC	20	R
	3	CIPIC	51	L
	4	CIPIC	147	L
	5	CIPIC	155	L
	6	LISTEN	1003	R
	7	MIT KEMAR	N/A	L+R*

*MIT KEMAR HRTF was used as a standard HRTF set in front, back and elevation cue. The left channel and right channel of MIT KEMAR HRTF were used to generate left channel and right channel of virtual sound cues respectively.

Please refer to appendix C for the cluster results.

Chapter 7. Technical details of experiments

7.1. Introduction

The objective of this chapter is to introduce the experimental setup for collecting listeners' perceptions and the method of data analysis. Front-back confusion, back-front confusion, up-down confusion, down-up confusion and inside-the-head locatedness are the index of sound localization performance. Azimuth error was used to indicate front-back confusion and back-front confusion; elevation error was used to indicate up-down confusion and down-up confusion. To collect listeners' azimuth error and elevation error, a sound localization measurement system was set up.

7.2. Coordinate system

All of the sound source positions used in this study were specified with respect to the single pole polar system (Figure 7.1). The origin is the centre of the listener's head. The coordinates are specified with an azimuth (θ), an elevation (ø) and a range. Since this study focuses on the sound source direction, the ranges of sound sources are omitted in the rest of this thesis. Hence, the incident angles of sound sources are represented by (θ,ø).

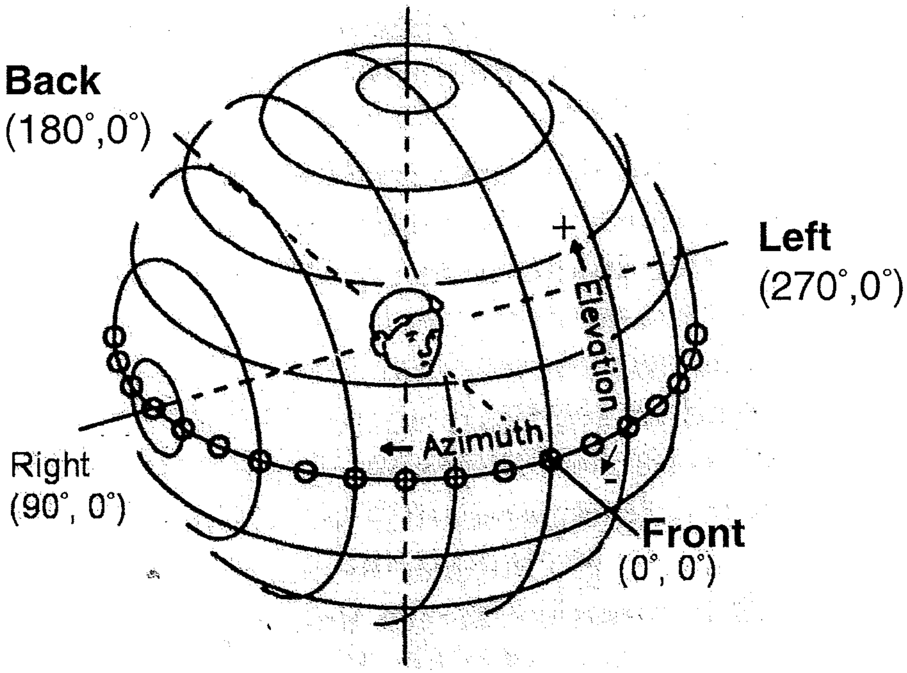

Figure 7.1 Single pole polar coordinate system. Reprinted from Carlile (1996) [8]. From topview, azimuth angle increases in clockwise direction. Elevation angle is positive if it is higher than the centre of the head and negative if it is lower than the centre of the head.

Azimuth and elevation angles represent the sound source direction. The range of azimuth is from 0° to 360°. An azimuth angle of 0° is directly in front and moving clockwise from 0° increases the azimuth angle. An azimuth angle of 90° is directly to the right and azimuth angle at 180° is directly to the back. The range of elevation is from -90° to 90°. The horizontal plane is at an elevation of 0°. The elevation angle increases when moving upwards and decreases when moving downwards. With a 90° elevation angle, the sound source direction is directly on top of the head. The sound source direction is directly below the head with -90° elevation angle although it is not likely happen as this is the direction pointing to the torso.

7.3. Equipment used in sound localization experiments

7.3.1. Functions of the equipment

A sound localization measurement system was designed to collect listeners' perception including:

(i) Perceived azimuth angle

(ii) Perceived elevation angle

The system setup was using the God's eye localization pointing technique (Gilkey *et. al.*, 1995 [30]). A block diagram (Figure 7.2) showed the overview of the system.

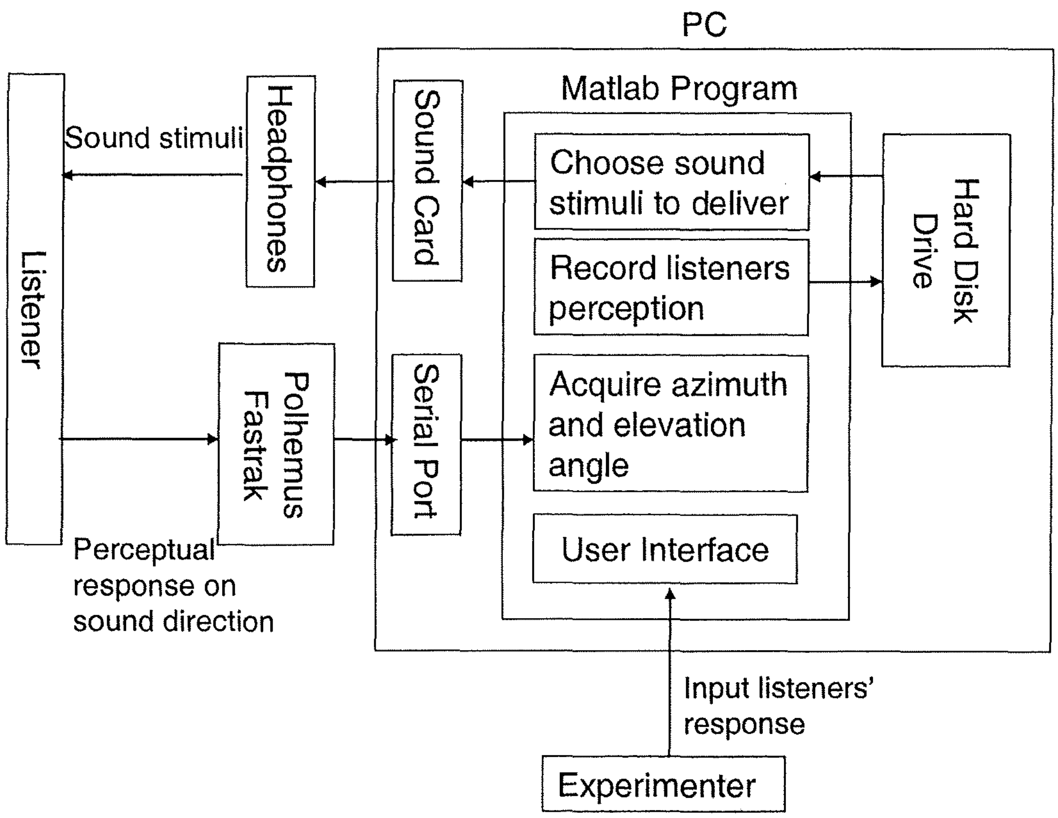

Figure 7.2 Block diagram of sound localization measurement system

7.3.2. Hardware

A Pentium III 600MHz personal computer was used to control the whole system. A Polhemus Fastrak was connected to the computer which was. It is used for collecting azimuth and elevation angles of listeners' perceptions. Figure 7.3 illustrates the hardware setup. Listeners indicated the perceived sound direction by pointing the sphere with a sensor connected to the Polhemus Fastrak.

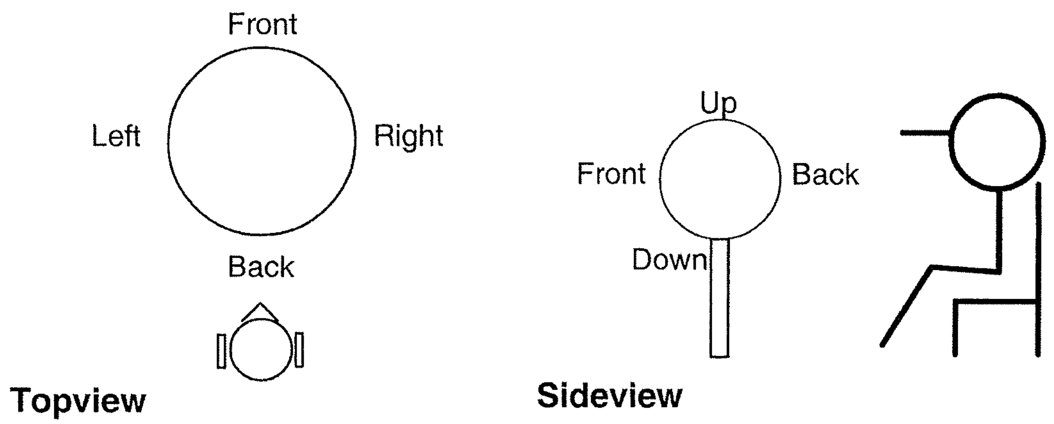

Figure 7.3 Hardware setup for collecting azimuth and elevation angles

The computer was equipped with a Creative Live sound card. A pair of Sennheiser HD 580 headphones (Hannover, Germany) were connected to the speaker output of the sound card, which was used to deliver sound stimuli to listeners.

7.3.3. Software

MATLAB programs were developed for the experimenters to control the system. The user interfaces for experiment 1 to 3 were shown in Figure 7.4 and Figure 7.5. The program was used for several functions i.e., to present sound clips in random order, to control the Polhemus Fastrak to collect azimuth and elevation angles of listeners' perception and to record listeners' perceptions. A set of Matlab code for convolution of sound is provided in Appendix H for demonstration.

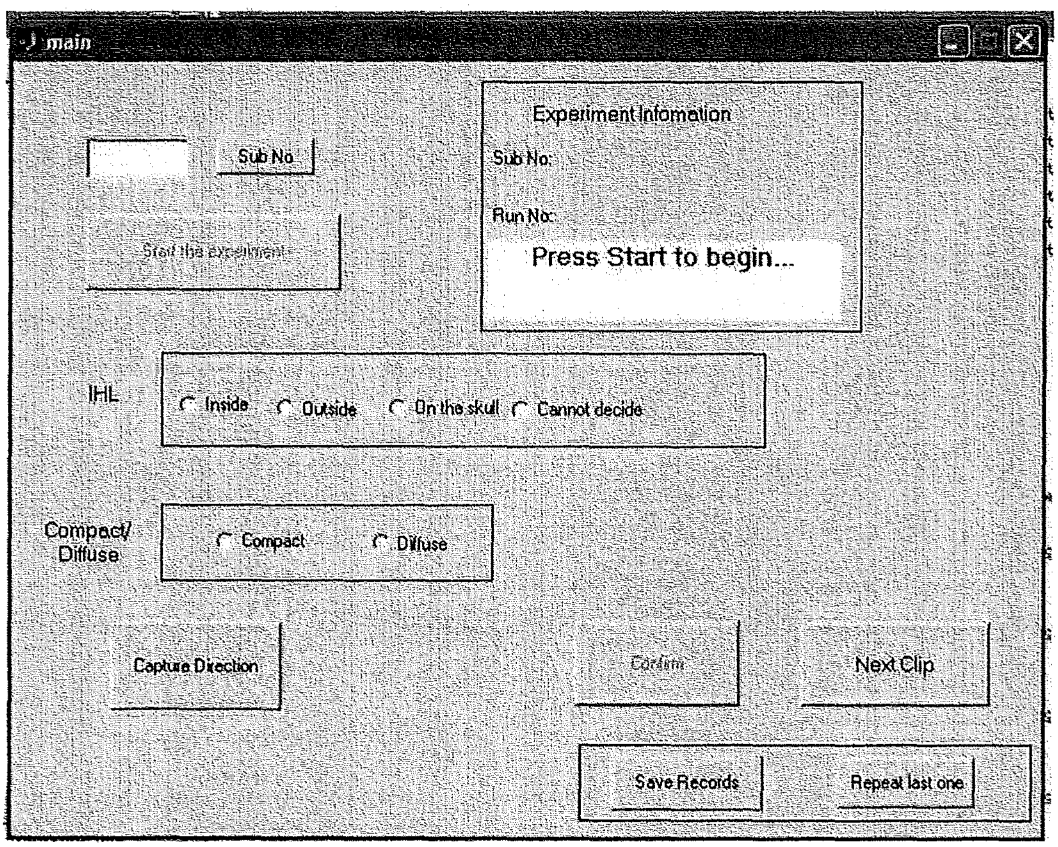

Figure 7.4 User interface for experiment 1 and 3

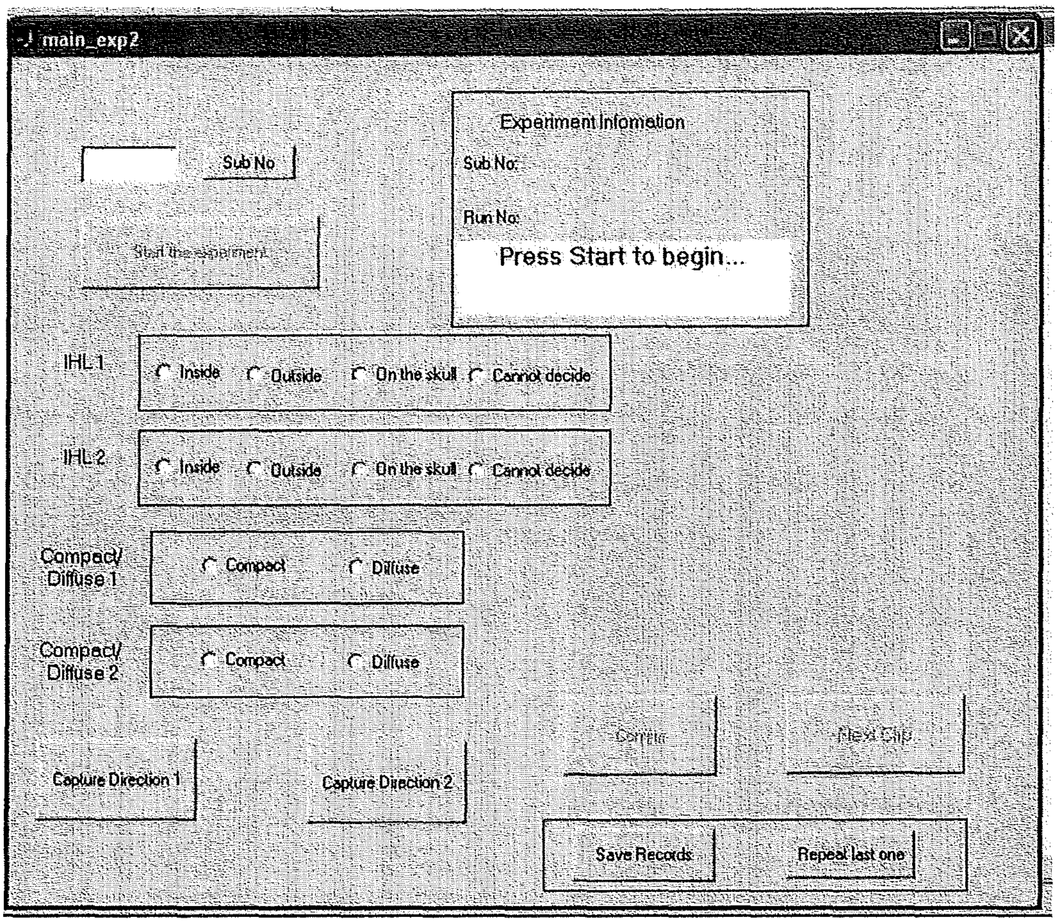

Figure 7.5 User interface for experiment 2

7.3.4. Safety precautions

There was no known hazard in the experimental setup. All of the sound clips were not greater than 80 dBA in order to protect participants from hearing damage.

Experiments were approved by the Human Subject Committee of Hong Kong University of Science and Technology. Participants were instructed before the experiments and had the option to withdraw from these experiments at any time without the necessity to provide any reasons. A sample of the subject consent form used in the experiments can be referred to in the Appendix D.

Chapter 8. Experiment 1: Verification of the standard HRTF sets: Single sound cue presentation

8.1. Purpose of this experiment

This experiment aimed to investigate the sound localization performance of listeners in different directional cues generated by standard HRTF sets.

8.2. Objective and hypothesis

The objective of this experiment was to compare the front-back confusion rate, back-front confusion rate and elevation error that listeners perceived under two conditions, (1) synthesizing binaural directional cues with standard sets of HRTFs, and (2) synthesizing binaural directional cues with a generic set of non-individualized HRTFs (KEMAR).

From the literature, it was hypothesized that (H3) listeners perceive lower front-back confusion rate and (H4) back-front confusion rate with binaural directional cues synthesized by suitable standard sets of HRTFs than a generic set of non-individualized HRTFs.

8.3. Participants

Fifteen (S4-S18) participants were recruited in Hong Kong University of Science and Technology. Six male and nine female volunteers aged 19-25 were recruited for the study. Before the sound localization experiment, they filled a questionnaire detailing their medical history concerning any auditory problems. Audiometric tests were carried out to test the participants' auditory normality. These test procedures are described in Abel and Paik (2004) [38]. Participants were eligible for the experiment only if their hearing thresholds are less than 20dB HL in each ear at 125 Hz, 1 kHz and 8 kHz, and an interaural difference in threshold is less than 10dB. Participants were not informed about the purpose of the experiment. A questionnaire about medical history was filled by participants. A sample is shown in Appendix E. The background of listeners participated can refer to Appendix G.

8.4. Dependent and independent variables

Three independent variables were used in this experiment: (i) the 4 directions of cues ((0°,0°), (0°,60°), (0°,-30°), (180°,0°)); (ii) HRTF sets (standard sets including 6 sets for front, 6 sets for back and MIT KEMAR set for front and back); and (iii) repetitions.

The dependent variables were: (i) Perceived incident angles along the azimuth direction; (ii) Front-back confusion status for each data run with cue direction in the front: 0 – no front back; 1- front-back (one data run refers to a single sound localization trial during which a listener has to report the perceived incident direction of a sound cue that he or she has heard); (iii) Perceived absolute localization errors along azimuth direction without front-back confusion corrections; (iv) Percentages of front-back confusion occurrence in the six repeated runs per each combination of conditions with cues in the front. In this experiment, there are 105 conditions which are the exhaustive combinations of 15 listeners, 7 standard HRTF sets, and 1 cue direction directly in the front with 0 degree elevation; (v) Back-front confusion status for each data run with cue direction at the back; (vi) Percentages of back-front confusion occurrence in the six repeated runs per each combinations of conditions with cues at the back. In this experiment, there are 105 conditions which are the exhaustive combinations of 15 listeners, 7 standard HRTF sets, and 1 cue direction directly at the back with 0 degree along elevation direction.

8.5. Stimuli preparation

Each stimulus included ten 0.3 s duration white noise pulses and the inter-pulse duration was 0.3s, filtered by predefined HRTFs sets or MIT

KEMAR HRTF set under Matlab environment.

8.6. Procedure

Participants were asked to sit comfortably and take rest for 5 minutes. A sphere and a pen with a Polhemus Fastrak sensor attached to it were given to each participant. Instructions on using the pen as measuring equipment were given by the experimenter. For the instruction sheets for participants, please refer to Appendix F. This measuring equipment was used for measuring the elevation and azimuth angle that listeners perceived.

Stimuli were presented through Sennheiser HD580 headphones. After each stimulus, each participant had to decide whether (i) the sound source was inside the head, outside the head or on the skull and (ii) the direction which was most likely to be the sound source. A GUI program developed under Matlab environment was developed to input the responses. Trial measurements were conducted until listeners can fully understood the whole procedure. Then, first 5 measurements were discarded. Participants were then asked to close the eyes while listening to the stimuli.

To avoid participants becoming tired, the experiment was divided into 2 sessions. There was a 5-minute break in the middle of each session. Between the 2 sessions, there was a rest of half an hour. Each participant completed 3 repetitions in the first session and 3 repetitions in the second session.

8.7. Results and Discussion

8.7.1. Front cue

Plots of perceived incident angles along the azimuth directions for each of the 15 participants, while listening to the sound cue directly in the front (0°,0°), are shown in Figure 8.1 to Figure 8.15 as functions of the 7 standard HRTF sets. Data from all six repetitions are shown on the same figure. The data points that had been circled indicated a front-back confusion percentage of less than 50%.

With inspections of Figure 8.1 to Figure 8.15, it indicated that in most conditions, the front-back confusion rates were greater than 50%. However, four out of fifteen listeners (S8, S11, S13 and S14) achieved less than 50% front-back confusion rate with one or more standard HRTF sets for frontal direction. Table 8-1 summarizes these rates for four participants and the corresponding standard HRTF sets that had produced frontal binaural cues that had been tested to be associated with less than 50% front-back confusion rates.

As shown in Table 8-1, there are five standard HRTF sets that had successfully produced a measured front-back confusion rate of less than 50% for participants S8, S11, S13, and S14. For each of these five standard HRTF sets, a Kruskal-Wallis one-way ANOVA was conducted to test for the main effects of listener on the front-back confusion status. Results indicated that listeners (participants) have significant main effects in all of the five tests ($p<0.03$) and the four listeners were ranked with the least total number of front-back confusion error occurrence. This suggested that for these four participants (S8, S11, S13, and S14), there exists at least one standard HRTF set that can produce binaural cues with frontal directions that are associated with less than 50% front-back confusion rates. Four out of fifteen participants represents 27% of the sampled population. Although these 27% of the participants supported our hypothesis H3, it is nonetheless disappointing because we were expecting 100% of the participants to support H3. Further examinations of the data indicate that six out of fifteen participants (S5, S6, S9, S12, S15 and S17) had 100% front-back confusion rates for all binaural cues produced by all seven standard HRTF sets in all repetitions.

Another four participants (S4, S10, S16 and S18) localized most of their sound cues at the back. This is an unexpected finding as the seven standard HRTF sets represent 196 non-individualized HRTF sets taken from the MIT KEMAR HRTF (Gardner and Martin 1995 [11]); the CIPIC HRTF database (Algazi *et. al.* 2001 [2]), the LISTEN HRTF database (Ircam and AKG acoustics, 2004 [13]), and the measured HRTF set of the author. One possible reason is the sound localization task may be too difficult because participants were asked to localize binaural cue containing pulsed broadband noise. Listening to sound cues in pairwise can provide a reference point for participants. It may increase the localization performance.

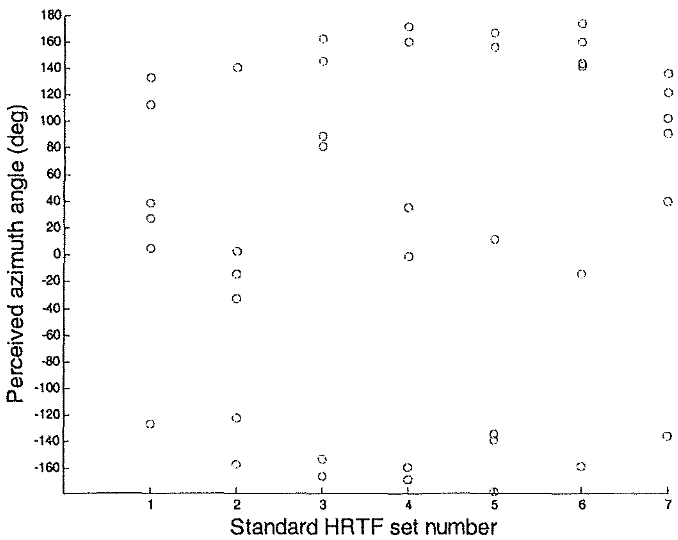

Figure 8.1 Plot of perceived incident angles along the azimuth directions for listener S4 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

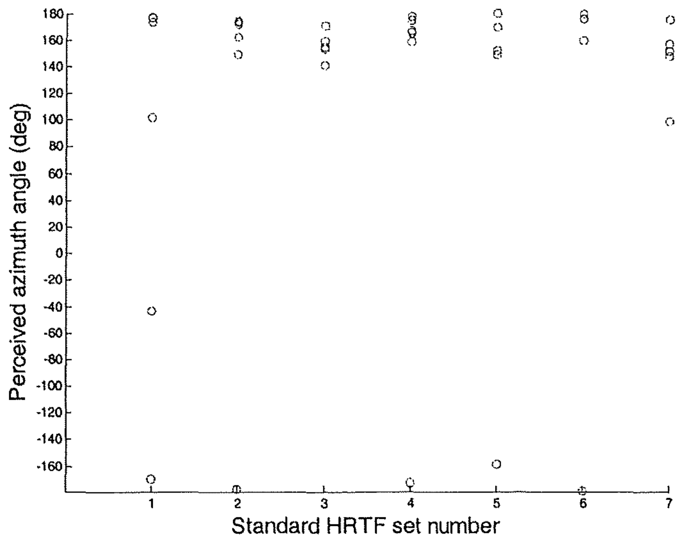

Figure 8.2 Plot of perceived incident angles along the azimuth directions for listener S5 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

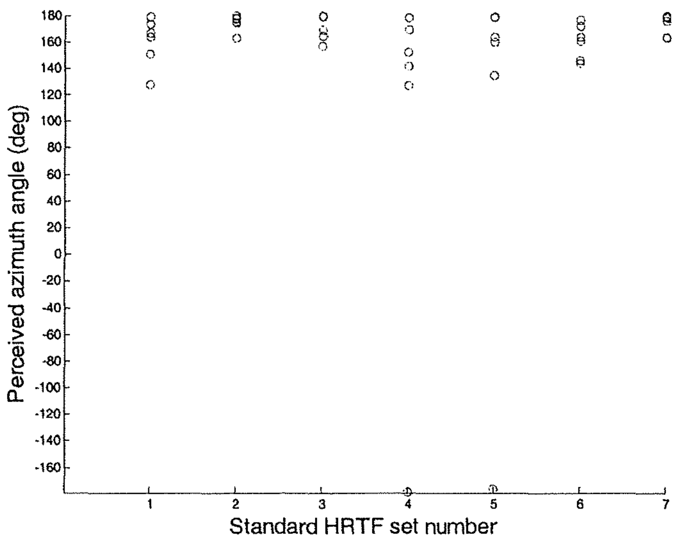

Figure 8.3 Plot of perceived incident angles along the azimuth directions for listener S6 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

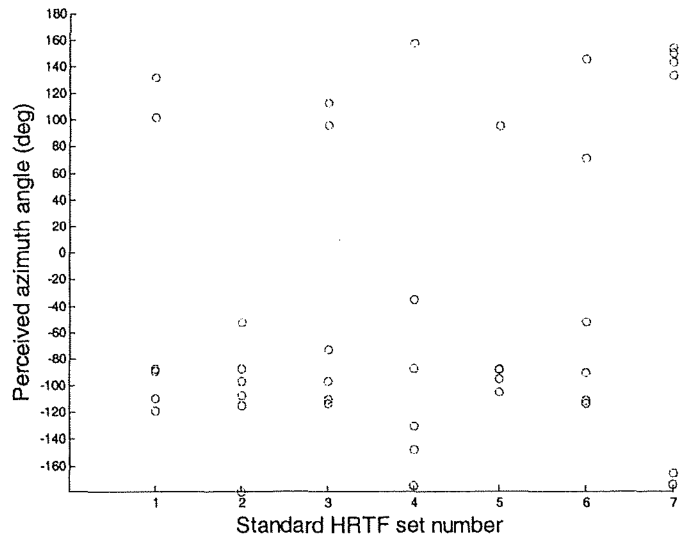

Figure 8.4 Plot of perceived incident angles along the azimuth directions for listener S7 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

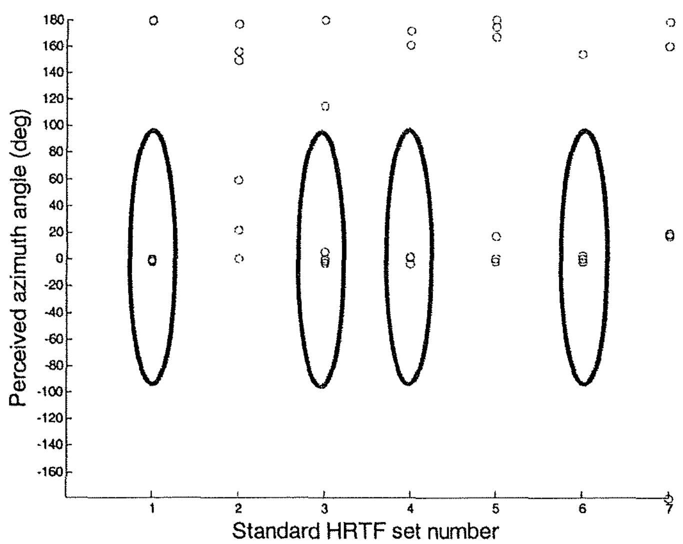

Figure 8.5 Plot of perceived incident angles along the azimuth directions for listener S8 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

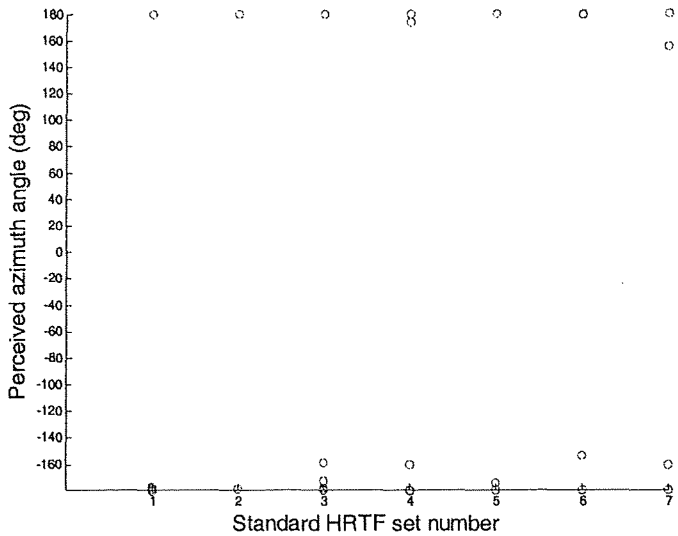

Figure 8.6 Plot of perceived incident angles along the azimuth directions for listener S9 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

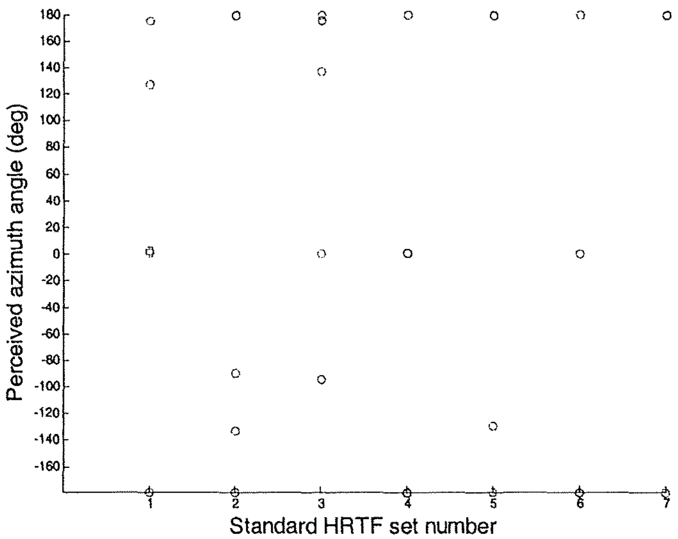

Figure 8.7 Plot of perceived incident angles along the azimuth directions for listener S10 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

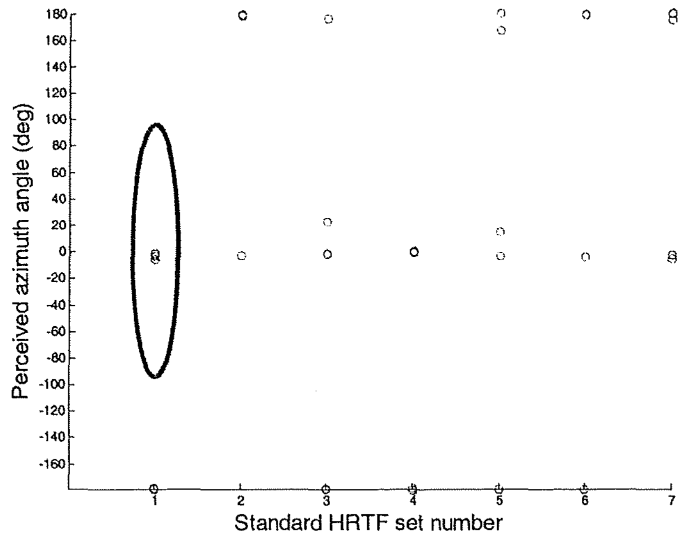

Figure 8.8 Plot of perceived incident angles along the azimuth directions for listener S11 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

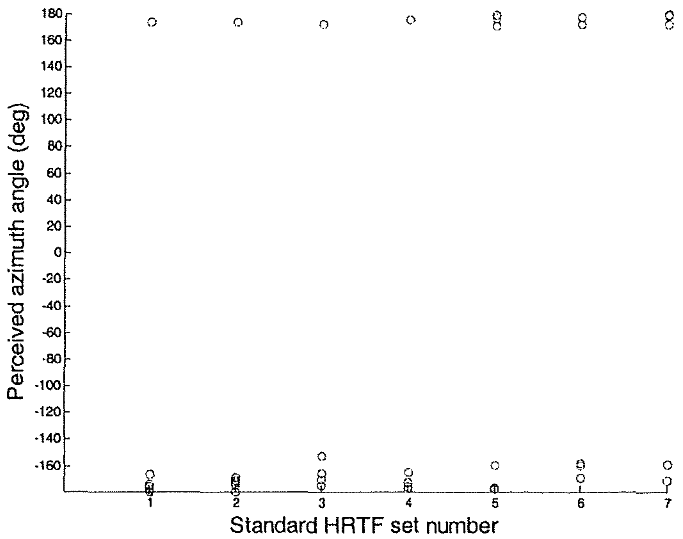

Figure 8.9 Plot of perceived incident angles along the azimuth directions for listener S12 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

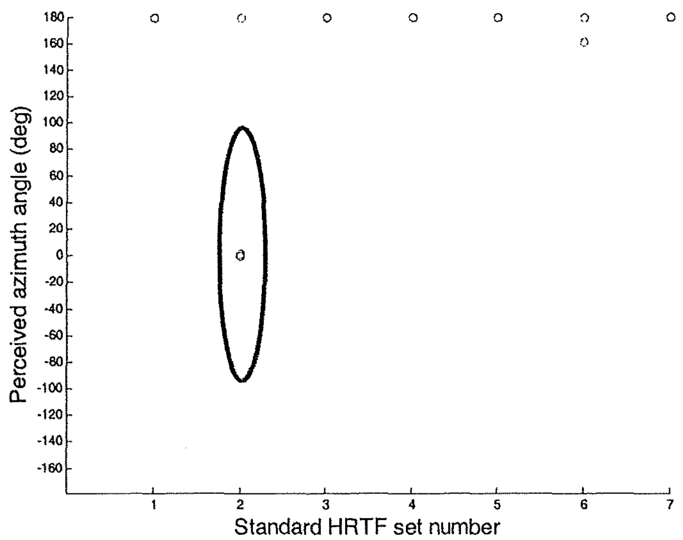

Figure 8.10 Plot of perceived incident angles along the azimuth directions for listener S13 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

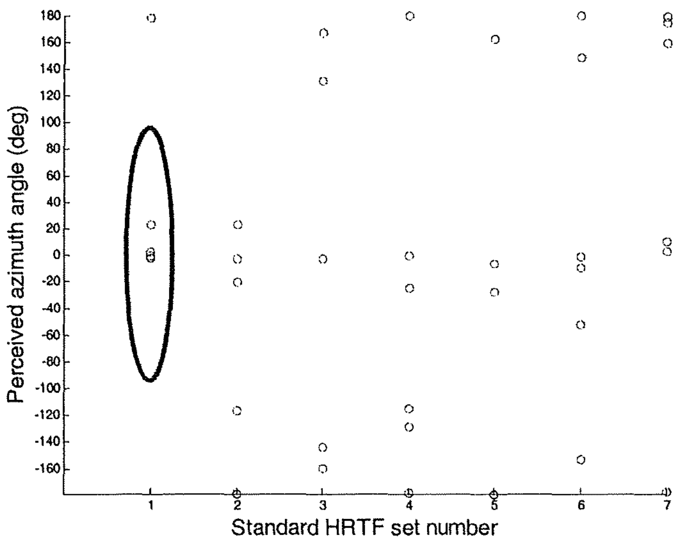

Figure 8.11 Plot of perceived incident angles along the azimuth directions for listener S14 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

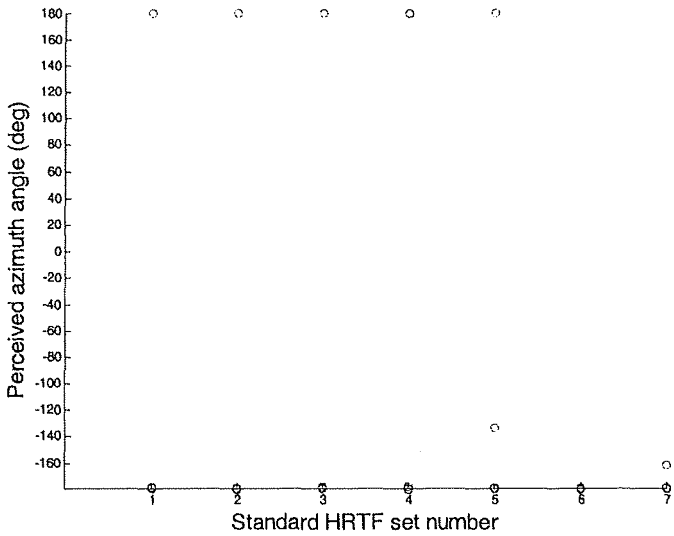

Figure 8.12 Plot of perceived incident angles along the azimuth directions for listener S15 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

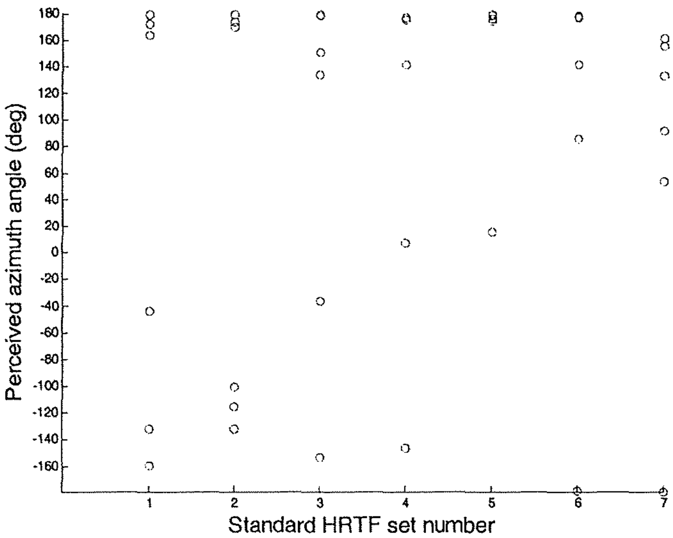

Figure 8.13 Plot of perceived incident angles along the azimuth directions for listener S16 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

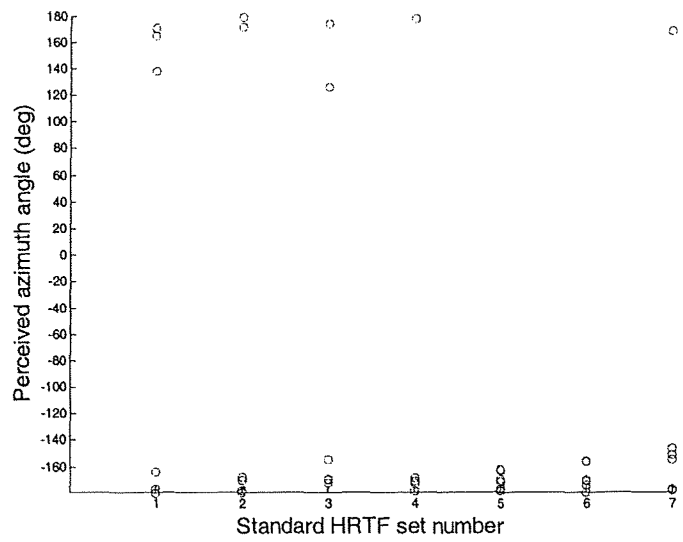

Figure 8.14 Plot of perceived incident angles along the azimuth directions for listener S17 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

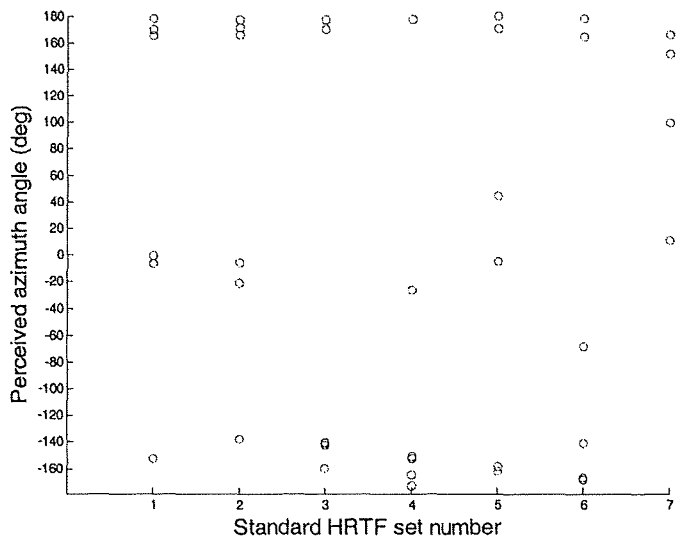

Figure 8.15 Plot of perceived incident angles along the azimuth directions for listener S18 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

Table 8-1 The standard HRTF set for front number and number of repetition perceived at the front while listeners perceived more front direction than back direction

Listener	Standard HRTF set for front number	Number of repetition perceived at the front out of six repetitions
S8	1	4
	3	4
	4	4
	6	5
S11	1	4
S13	2	5
S14	1	4

The absolute azimuth error of the best standard HRTF set of each listener was statistically smaller (p<0.003, Wilcoxon signed ranks test) than corresponding errors of other standard HRTF sets. However, the absolute azimuth error with front-back confusion correction of the best standard HRTF set of each listener was not statistically smaller than corresponding errors of other standard HRTF sets. Hence, given the choice of 7 standard sets of sound cues, assuming listeners have means to choose the most appropriate sets for themselves, the sound localization performance of 15 randomly selected listeners were significantly improved.

8.7.2. Back cue

Plots of perceived incident angles along the azimuth directions for each of the 15 participants, while listening to the sound cue directly in the back (180°,0°), are shown in Figure 8.16 to Figure 8.30 as functions of the 7 standard HRTF sets. Data from all six repetitions are shown on the same figure.

In general, all 15 participants performed quite well with less than 50% back-front confusion rate in at least one set of back cues generated by standard HRTF for back. There were 7 participants (S5, S6, S7, S9, S12, S15 and S17) who perceived all back perception in all sound cues provided. Hence, they have 0% back-front confusion in all standard HRTF sets for back. For 4 participants (S8, S11, S13 and S14) who had less than 50% front-back confusion in at least one set of standard HRTF for front, they also had less than 50% back-front confusion in at least one set of standard HRTF for back. Their result was summarized in Table 8-2. For the rest of participants (S4, S10, S16 and S18), they had at least one set of standard HRTF for back which the back-front confusion was 0%. From this result, participants had no difficulty in localizing back cues in at least one set of standard HRTF for back. Their result was summarized in Table 8-3.

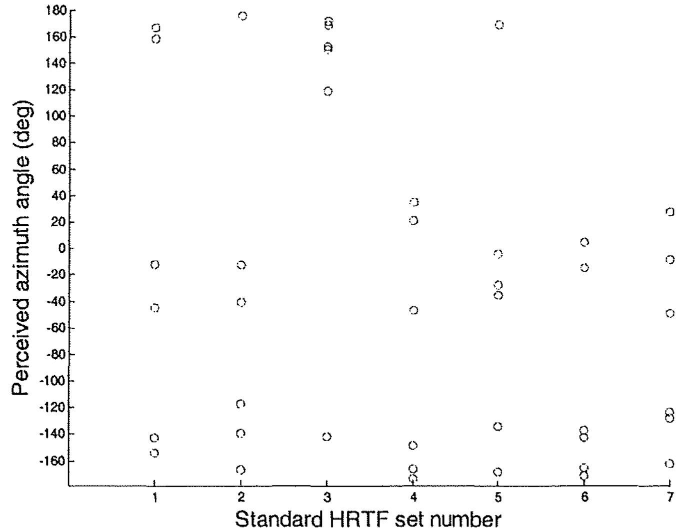

Figure 8.16 Plot of perceived incident angles along the azimuth directions for listener S4 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

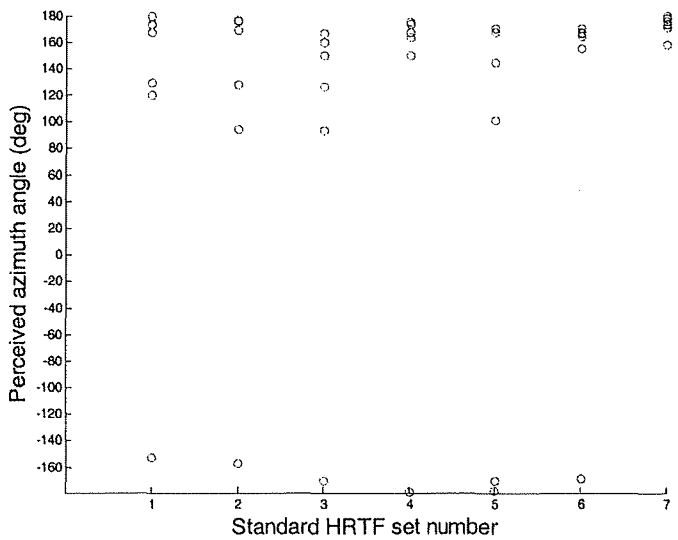

Figure 8.17 Plot of perceived incident angles along the azimuth directions for listener S5 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

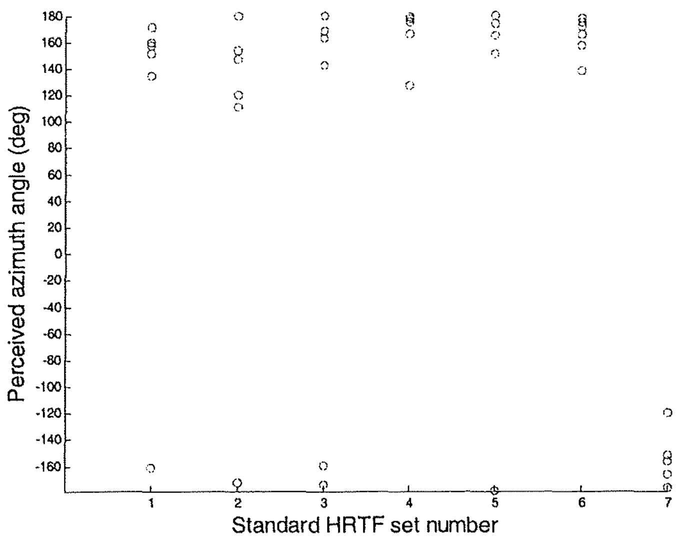

Figure 8.18 Plot of perceived incident angles along the azimuth directions for listener S6 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

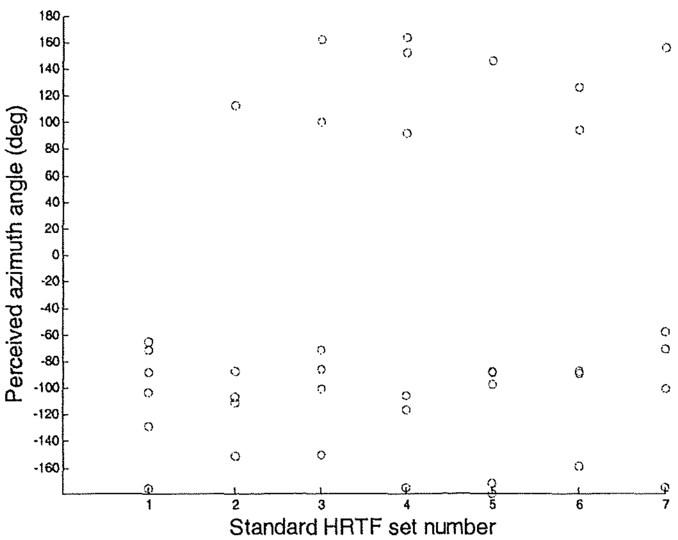

Figure 8.19 Plot of perceived incident angles along the azimuth directions for listener S7 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

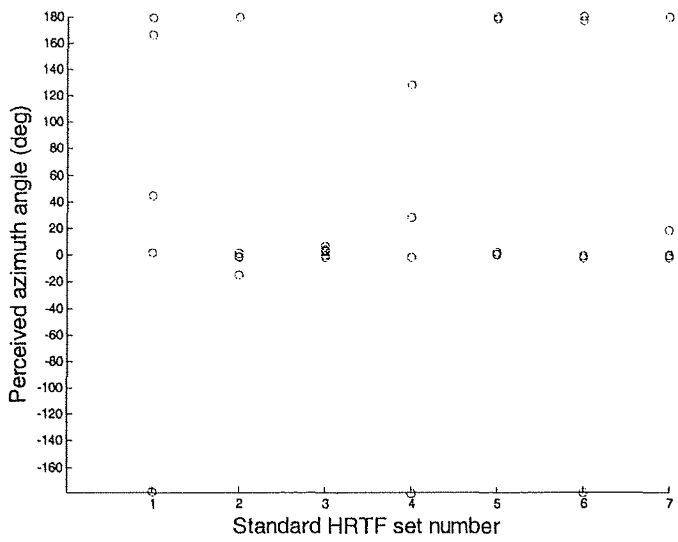

Figure 8.20 Plot of perceived incident angles along the azimuth directions for listener S8 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

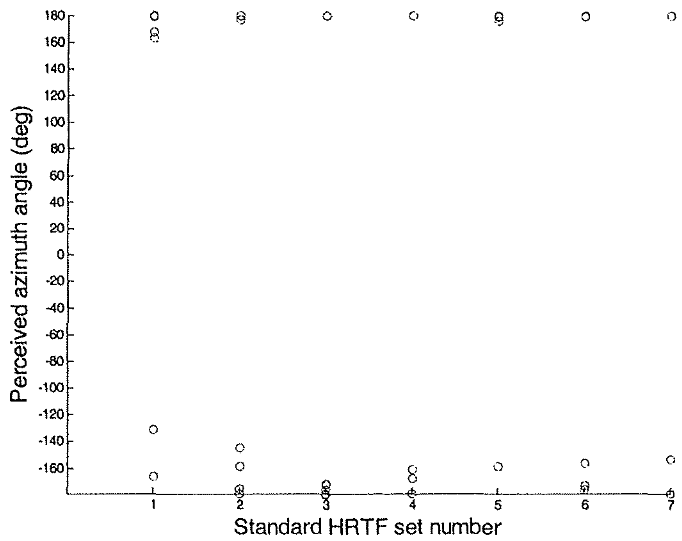

Figure 8.21 Plot of perceived incident angles along the azimuth directions for listener S9 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

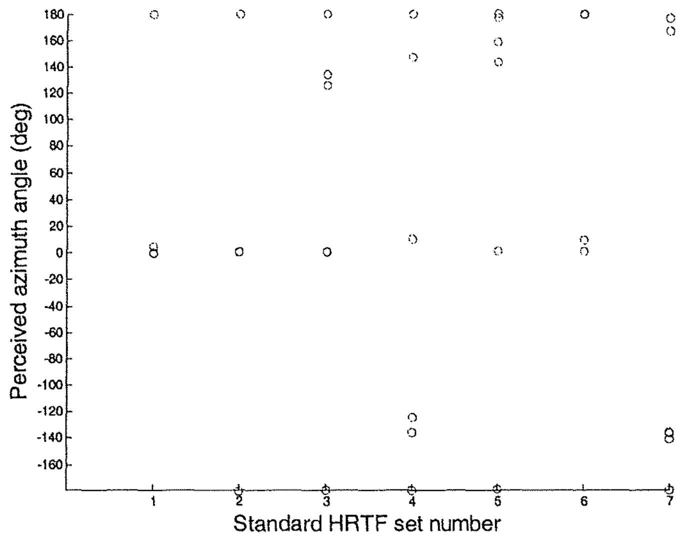

Figure 8.22 Plot of perceived incident angles along the azimuth directions for listener S10 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

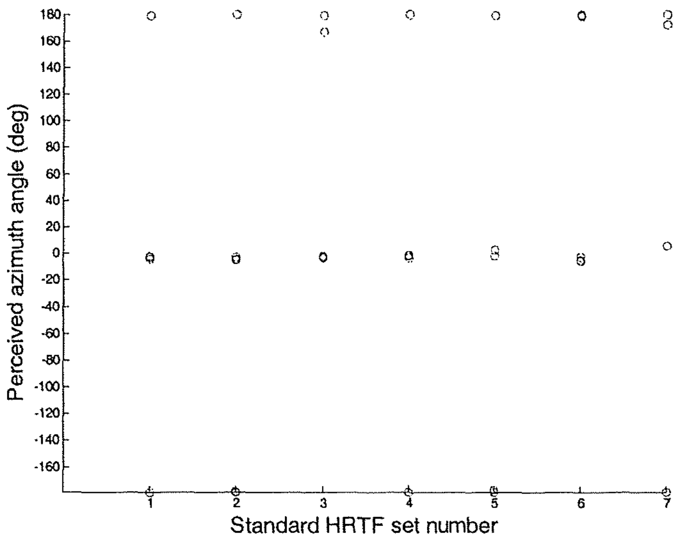

Figure 8.23 Plot of perceived incident angles along the azimuth directions for listener S11 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

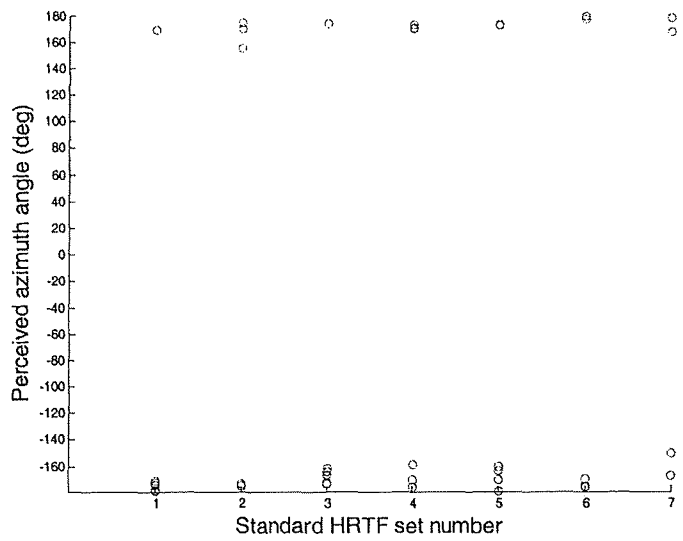

Figure 8.24 Plot of perceived incident angles along the azimuth directions for listener S12 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

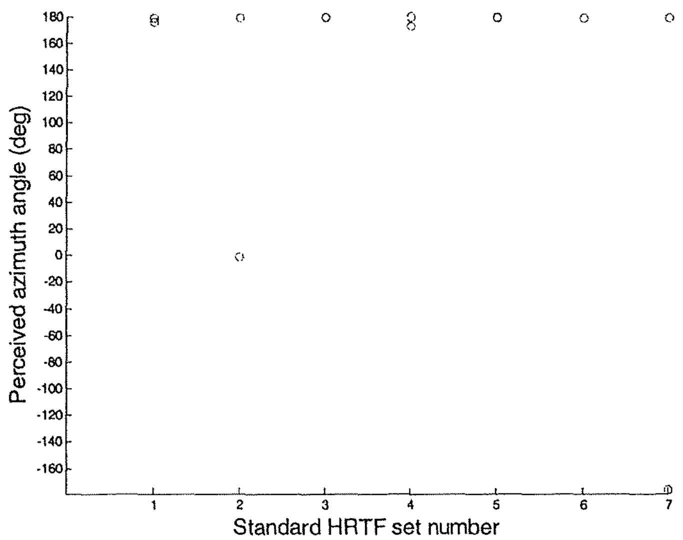

Figure 8.25 Plot of perceived incident angles along the azimuth directions for listener S13 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

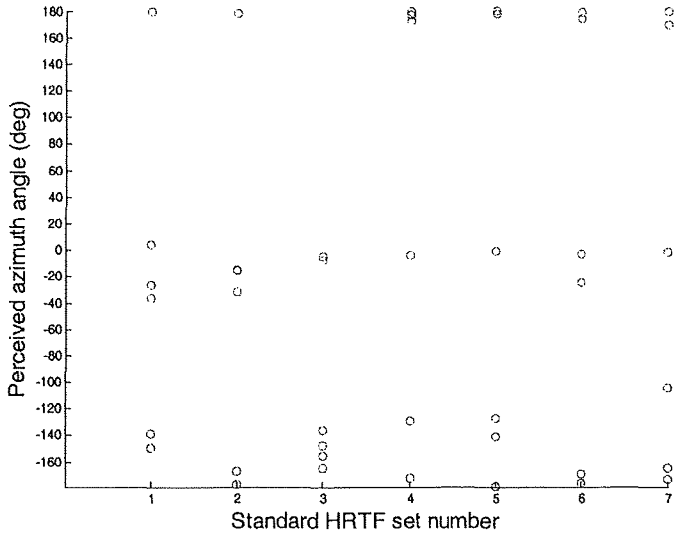

Figure 8.26 Plot of perceived incident angles along the azimuth directions for listener S14 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

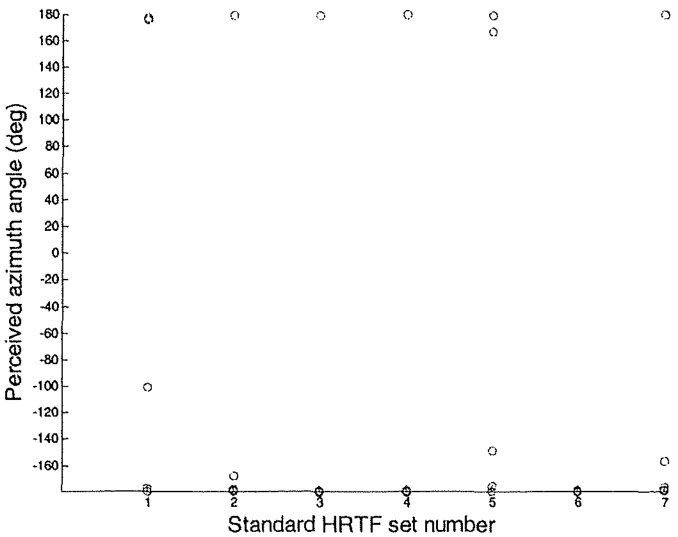

Figure 8.27 Plot of perceived incident angles along the azimuth directions for listener S15 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

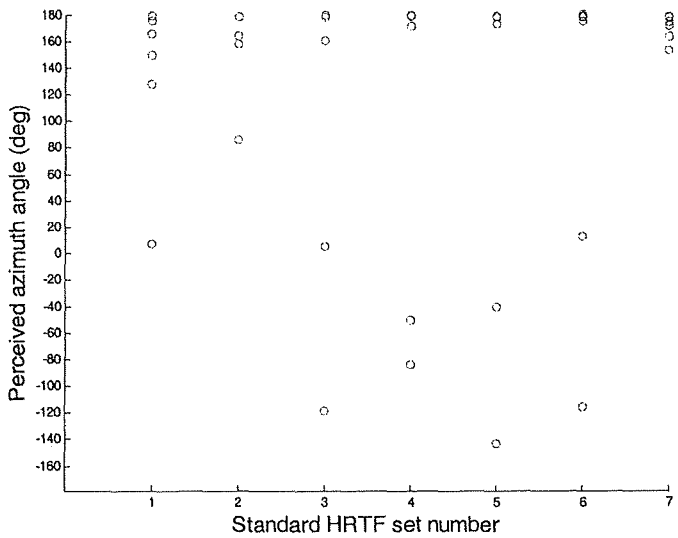

Figure 8.28 Plot of perceived incident angles along the azimuth directions for listener S16 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

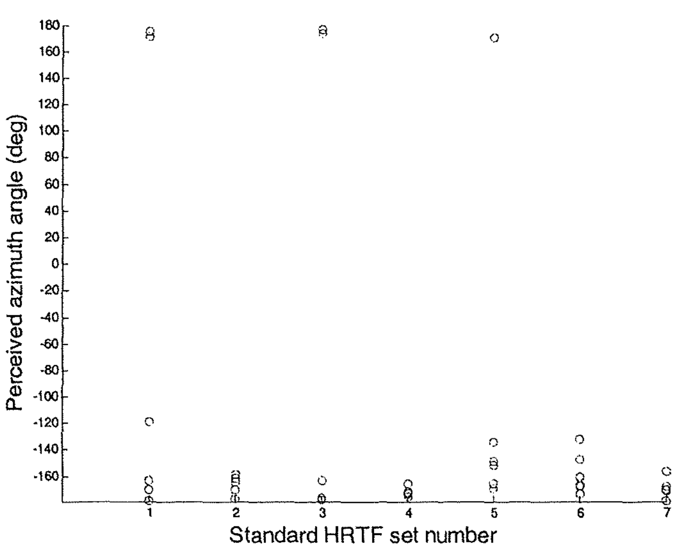

Figure 8.29 Plot of perceived incident angles along the azimuth directions for listener S17 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

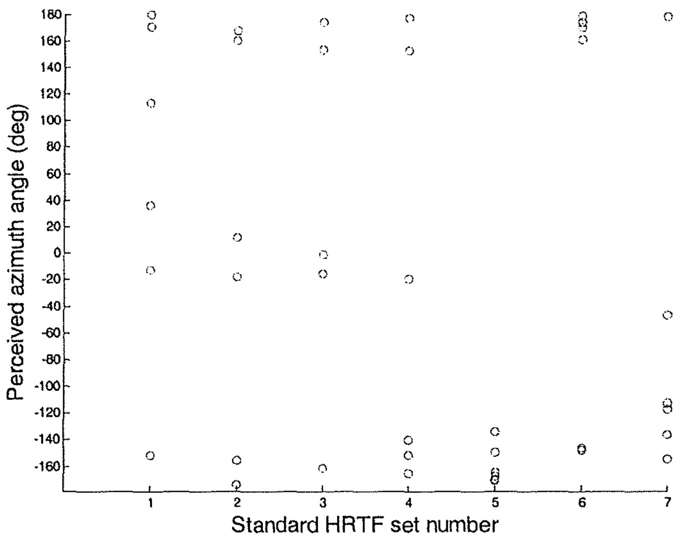

Figure 8.30 Plot of perceived incident angles along the azimuth directions for listener S18 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

Table 8-2 The standard HRTF set number for back which the listener S8, S11, S13 and S14 perceived less than 50% back-front confusion

Listener	Standard HRTF set number for back which the listener perceived less than 50% back-front confusion
S8	1
S11	5, 7
S13	1, 2, 3, 4, 5, 6, 7
S14	3, 4, 5, 6, 7

Table 8-3 The standard HRTF set number for back which the listener S4, S10, S16 and S18 perceived 0% back-front confusion

Listener	Standard HRTF set number for back which the listener perceived 0% back-front confusion
S4	3
S10	7
S16	7
S18	5,6

The absolute azimuth error absolute azimuth error with back-front confusion correction of the best standard HRTF set of each listener was statistically smaller ($p<0.007$ and $p<0.008$, Wilcoxon signed rank test) than corresponding errors of other standard HRTF sets. Hence, given the choice of 7 standard sets of sound cues, assuming listeners have means to choose the most appropriate sets for themselves, the sound localization performance of 15 randomly selected listeners were significantly improved.

8.7.3. Learning effect

The perceived azimuth angle and perceived elevation angle did not follow normal distribution ($p<0.0001$, Shapiro-Wilk test), hence a Friedman two-way analysis of variance by ranks was employed to test the effect of repetition on perceived incident angles obtained from each participant for each cue direction. Data from listening to the cues generated from the seven standard HRTF sets was used as repeated measures in the tests for effects of repetition. Results indicated that 20 out of the 120 cases (15 listeners x 4 cue directions x 2 types of perceived angle) reported significant effects of repetitions ($p<0.048$).

The listeners and angles of which repetition had significant effect were shown in Table 8-4 and Table 8-5. Similar to the analysis of intra-subject variability in preliminary experiment, the repetition number in ascending order of perceived angle had shown no evidence to support that there was a learning effect.

Table 8-4 Friedman test result on listeners and angles which repetition had significant effect on perceived angle along azimuth direction

Listener	**Angle**	***p*-value**	**Repetition number in ascending order of perceived azimuth angle**
S4	(0,-30)	0.007	2, 3, 6, 5, 4, 1
S5	(0,-30)	0.004	1, 4, 2, 3, 6, 5
	(180,0)	0.048	1, 2, 5, 4, 3, 6
S7	(0,-30)	0.028	1, 6, 4, 2, 5, 3
	(0,60)	0.003	1, 5, 6, 2, 4, 3
	(180,0)	0.023	4, 6, 1, 2, 5, 3
S8	(0,-30)	0.022	2, 1, 6, 3, 5, 4
S9	(0,60)	0.013	3, 4, 2, 1, 6, 5
	(180,0)	0.018	4, 1, 2, 6, 3, 5
S12	(0,0)	0.018	4, 6, 5, 2, 3, 1
	(180,0)	0.045	3, 5, 4, 1, 2, 6
S13	(0,-30)	0.019	2, 5, 1, 4, 3, 6
S15	(0,60)	0.006	2, 3, 5, 6, 4, 1
S17	(0,-30)	0.048	6, 1, 3, 4, 5, 2

Table 8-5 Friedman test result on listeners and angles which repetition had significant effect on perceived angle along elevation direction

Listener	Angle	p-value	Repetition number in ascending order of perceived elevation angle
S4	(0,-30)	0.033	4, 3, 6, 5, 1, 2
S5	(0,-30)	0.044	6, 3, 5, 2, 4, 1
S6	(0, 0)	0.014	1, 3, 5, 4, 6, 2
S7	(180,0)	0.031	1, 2, 3, 6, 5, 4
S9	(0,-30)	0.025	1, 3, 6, 5, 4, 2
	(0,0)	0.039	5, 4, 3, 6, 2, 1
S10	(180,0)	0.018	2, 6, 5, 1, 3, 4
S11	(0,0)	0.042	2, 4, 3, 5, 1, 6
S13	(0,-30)	0.019	2, 5, 1, 4, 3, 6
	(0,0)	0.019	1, 4, 3, 6, 2, 5
	(0,60)	0.019	1, 4, 3, 6, 2, 5
S16	(0,60)	0.031	6, 4, 1, 5, 2, 3

8.7.4. Summary of experiment 1

The objective of this experiment was to verify the representative standard of HRTF set using front, back and elevation cues for a selected group of participants.

Participants performed well in back cues generated by the standard HRTF sets for back direction. All 15 participants' back-front confusion rate were less than 50% in at least one set of standard HRTF set for back. Eleven participants scored 0% back-front confusion in at least one set of standard HRTF set for back.

The standard HRTF sets for front cues did not perform as expected. A satisfactory result was obtained for front cues, as four participants out of the total of fifteen had less than 50% front-back confusion in at least one set of standard HRTF for front. Other participants localized most of the sound cues at the back or had unstable perception of front and back on the same set of HRTF. It is interesting to note that participants tended to localize virtual sound sources at the back even individualized HRTF was used (Asano *et. al.* 1990 [4]). The listening task may be too difficult for listeners to distinguish sound cues were from the front or back. To ease the difficulties of sound localization tasks, additional information, such as the delivery of two sound sources in sequence and let participants to compare them, can be added to help participants to distinguish sound sources at the front or the back. This idea leads to the experiment described in next chapter.

Finally, there was no significant learning effect found in this experiment. This result was consistent with the result in the preliminary experiment.

Chapter 9. Experiment 2: Verification of the standard HRTF sets: pairwise sound cues presentation

9.1. Purpose of this experiment

This experiment aimed to investigate the sound localization performance of human beings towards different standard HRTF sets, in particular, in terms of front-back and back-front confusions.

9.2. Objective and hypothesis

According to the result of the last experiment, most of the listeners recognized sound sources at the back, independent of the spectral cues provided. In the daily life, sound sources come from both front and back. Additional information was given to human to localize the sounds by comparing sound from front and back. It is hypothesized that listeners performed better in sound localization in terms of (H6) front-back and (H7) back-front confusions if both front and back sound sources were given and compared.

The objective of this experiment was, similar to the last experiment, investigating the sound localization performance of listeners towards different standard HRTF sets in terms of (i) front-back and (ii) back-front confusions.

9.3. Participants

Fifteen listeners (S4-S18) who took part in the last experiment agreed to take part in this one. All of the listeners for this experiment passed the audiometric test.

9.4. Dependent and independent variables

Three independent variables were used in this experiment: (i) HRTF sets (standard sets including 6 sets for front, 6 sets for back and MIT KEMAR set) and (ii) repetitions. The dependent variables were: (i) Perceived incident angles along the azimuth direction; (ii) Front-back confusion status for each data run with cue direction in the front: 0 – no front back; 1- front-back (one data run refers to a single sound localization trial during which a listener have to report the perceived incident direction of a sound cue that they hear); (iii) Perceived absolute localization errors along azimuth direction without front-back confusion corrections; (iv) Percentages of front-back confusion occurrence in the six repeated runs per each combination of conditions with cues in the front. In this experiment, there are 105 conditions which are the exhaustive combinations of 15 listeners, 7 standard HRTF sets, and 1 cue direction directly in the front with 0 degree elevation; (v) Back-front confusion status for each data run with cue direction at the back and; (vi) Percentages of back-front confusion occurrence in the six repeated runs per each combinations of conditions with cues at the back.

In this experiment, there are 105 conditions which are the exhaustive combinations of 15 listeners, 7 standard HRTF sets, and 1 cue direction directly at the back with 0 degree along elevation direction.

9.5. Stimuli preparation

Each stimulus was combined using three parts: (i) eight 0.3 s duration white noise pulses with 0.3 s inter-pulse duration filtered by particular set of HRTFs at (0,0) or (180,0); (ii) 0.6 s pause; and (iii) eight 0.3 s duration white noise pulses with 0.3 s inter-pulse duration filtered by particular set of HRTFs at (0,0) if part (i) was (180,0) or (180,0) if part (i) was (0,0). All stimuli were prepared under Matlab environment.

9.6. Procedure

Listeners were asked to sit comfortably and relax for a total of 5 minutes. A sphere and a pen with an attached Polhemus Fastrak sensor were given to each of the listener. Instructions on using the pen as a measuring device were explained by the experimenter to each participant. This measuring device was used for measuring the elevation and azimuth angle that listeners perceived. Stimuli were presented through Sennheiser HD580 headphones. After each stimulus, each listener had to decide whether (i) the first and the second sound source was inside the head, outside the head or on the skull and (ii) the direction of the first and the second sound source which is most likely to be the sound source. A GUI program developed under Matlab environment was developed to input the responses. Trial measurements were conducted until listeners can fully understand the whole procedure. After that, the first 5 measurements were discarded. Listeners were asked to close their eyes while listening to the stimuli.

To avoid listeners becoming tired, the experiment was divided into 2 sessions. Between the 2 sessions, there was a 5 minute break. Each listener completed 3 repetitions in the first session and 3 repetitions in the second session.

9.7. Results and Discussion

9.7.1. Front cue

Plots of perceived incident angles along the azimuth directions for each of the 15 listeners, while listening to the sound cue directly in the front, are shown in Figure 9.1 to Figure 9.15 as functions of the 7 standard HRTF sets. Data from all six repetitions are shown on the same figure. The data points that have been circled indicated that a front-back confusion percentage of less than 50%. A thorough inspections of the data from Figure 9.1 to Figure 9.15 indicated that in addition to the four listeners identified in Experiment 1 who reported of less than 50% front-back confusion rates with cues from one or more standard HRTF sets (S8, S11, S13, S14), in addition, two more listeners experienced less than 50% front-back confusion rate (S16, S18).

Table 9-1 contains a summary of these six listeners and the corresponding standard HRTF sets that produced the frontal binaural cues that were tested to be associated with less than 50% front-back confusion rates. As shown in Table 9-1, there were six standard HRTF sets that had successfully produced a measured front-back confusion rate of less than 50% from listeners S8, S11, S13, S14, S16 and S18. For each of these six standard HRTF sets, a Kruskal-Wallis one-way ANOVA was conducted to test for the main effects of listener on the front-back confusion status.

Results indicated that listeners had significant main effects in all of the six tests ($p<0.007$) and the six listeners were ranked with the least total number of front-back confusion error occurrence. This suggests that for these six listeners (S8, S11, S13, S14, S16 and S18), there exists at least one standard HRTF set that can produce binaural cues with frontal directions that are associated with less than 50% front-back confusion rates.

Those six listeners (S5, S6, S9, S12, S15 and S17) who localized all of the sound cues at the back in experiment 1, all of them still localized the front cues with 100% or near 100% front-back confusion rates. These 6 listeners had a total of 194 attempts with front-back confusion occurring a total of 193 times. This result indicated that broadcasting sound cues in pairs did not help improving their sound localization performance.

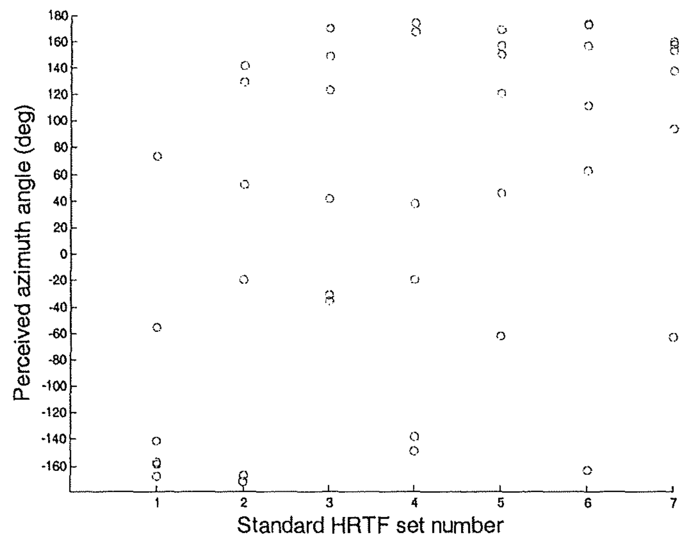

Figure 9.1 Plot of perceived incident angles along the azimuth directions for listener S4 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

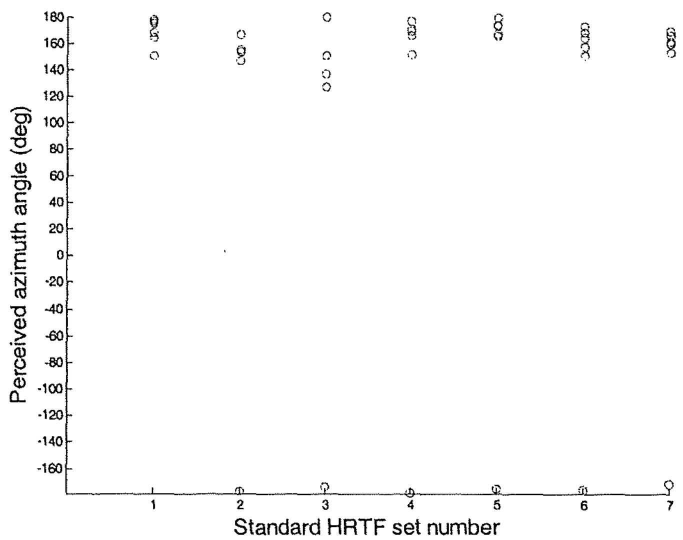

Figure 9.2 Plot of perceived incident angles along the azimuth directions for listener S5 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

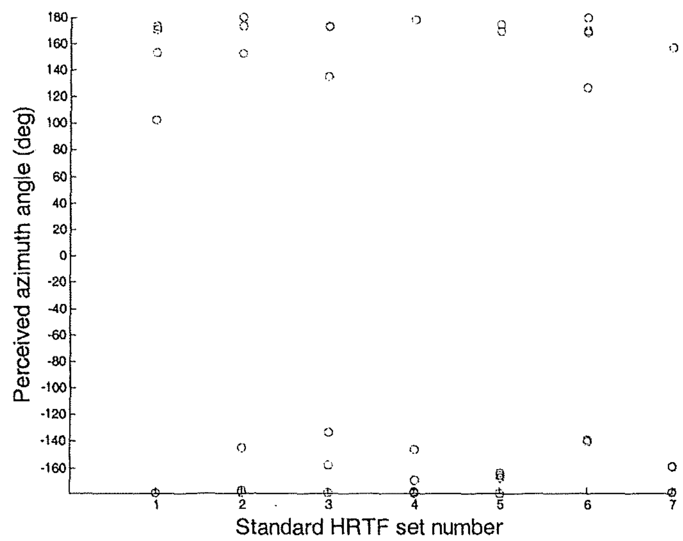

Figure 9.3 Plot of perceived incident angles along the azimuth directions for listener S6 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

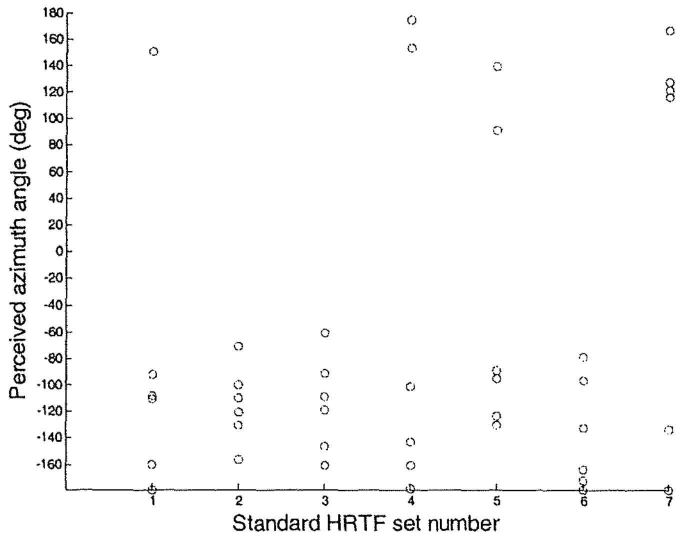

Figure 9.4 Plot of perceived incident angles along the azimuth directions for listener S7 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

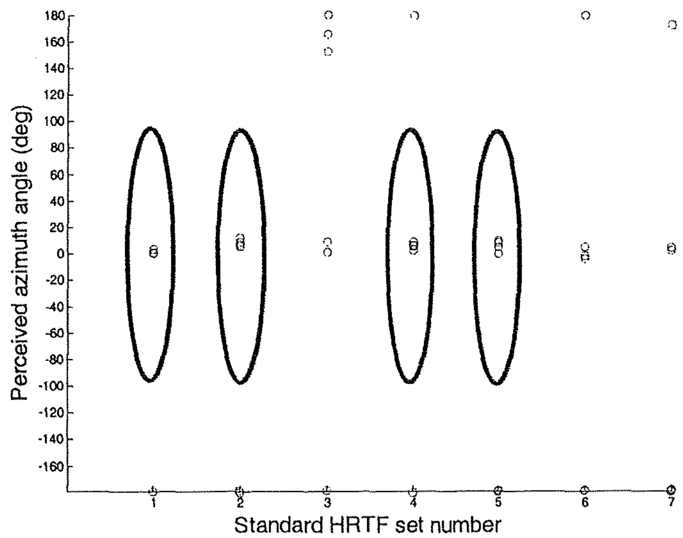

Figure 9.5 Plot of perceived incident angles along the azimuth directions for listener S8 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

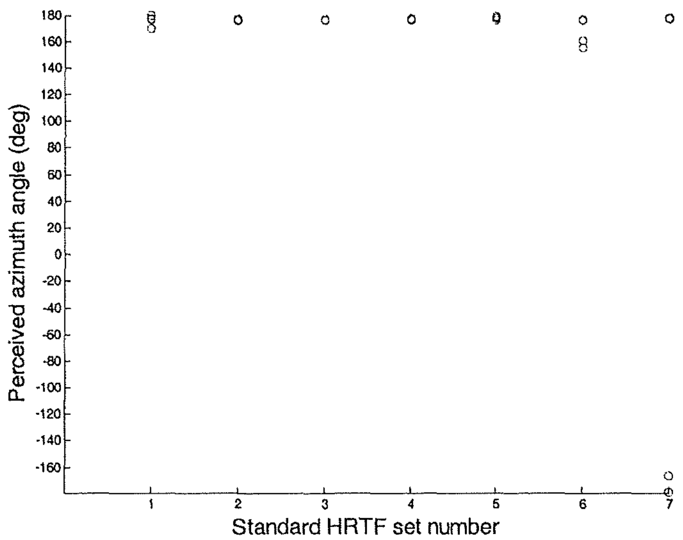

Figure 9.6 Plot of perceived incident angles along the azimuth directions for listener S9 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

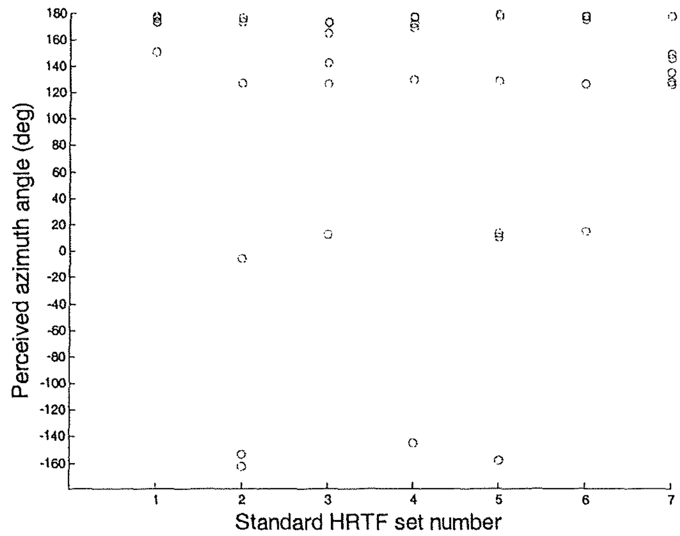

Figure 9.7 Plot of perceived incident angles along the azimuth directions for listener S10 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

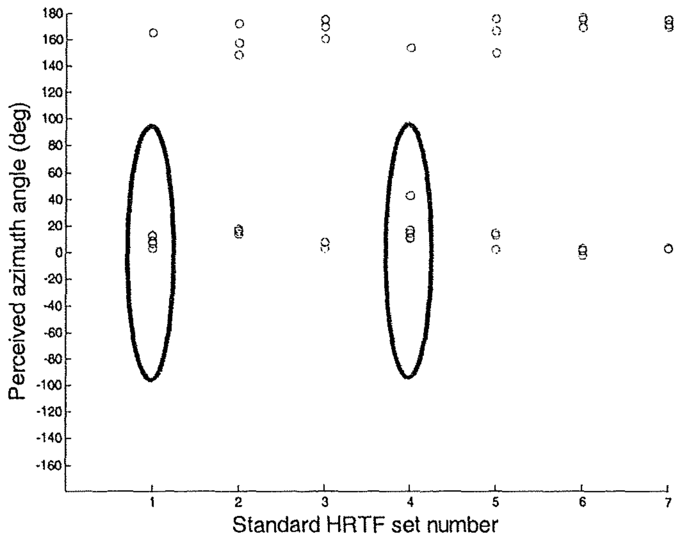

Figure 9.8 Plot of perceived incident angles along the azimuth directions for listener S11 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

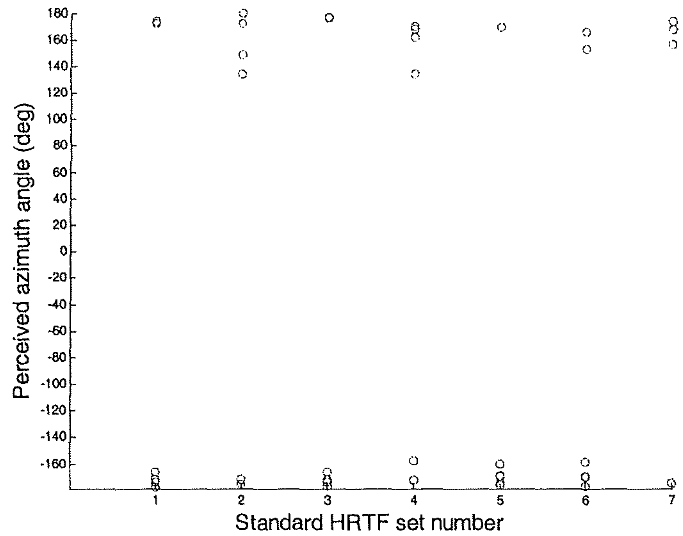

Figure 9.9 Plot of perceived incident angles along the azimuth directions for listener S12 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

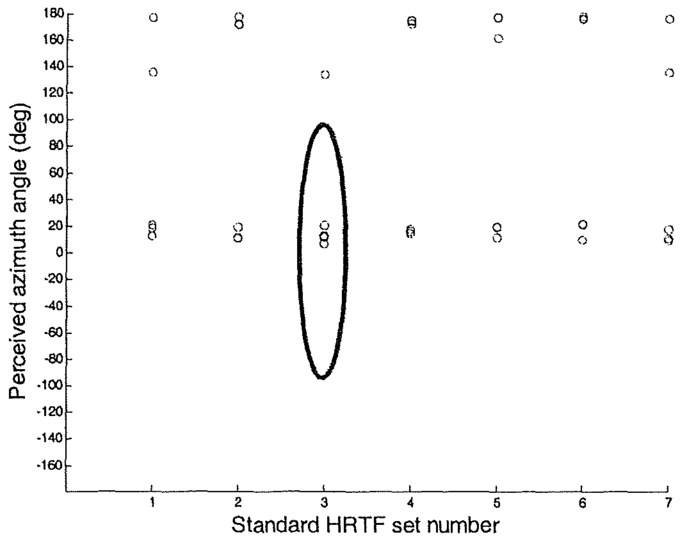

Figure 9.10 Plot of perceived incident angles along the azimuth directions for listener S13 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

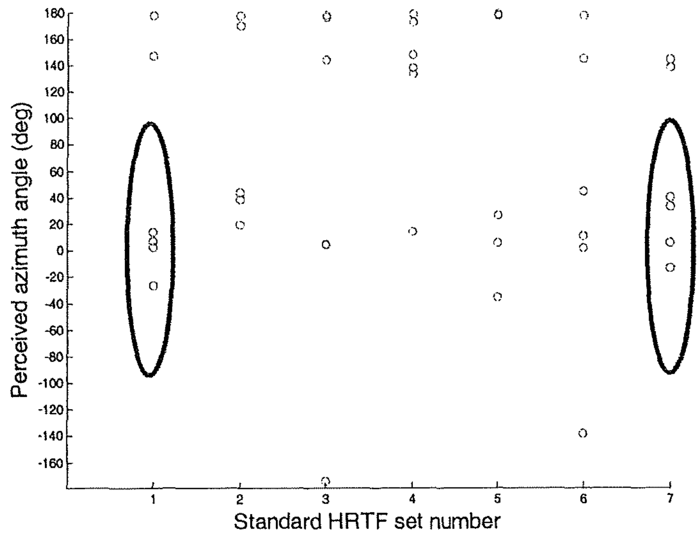

Figure 9.11 Plot of perceived incident angles along the azimuth directions for listener S14 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

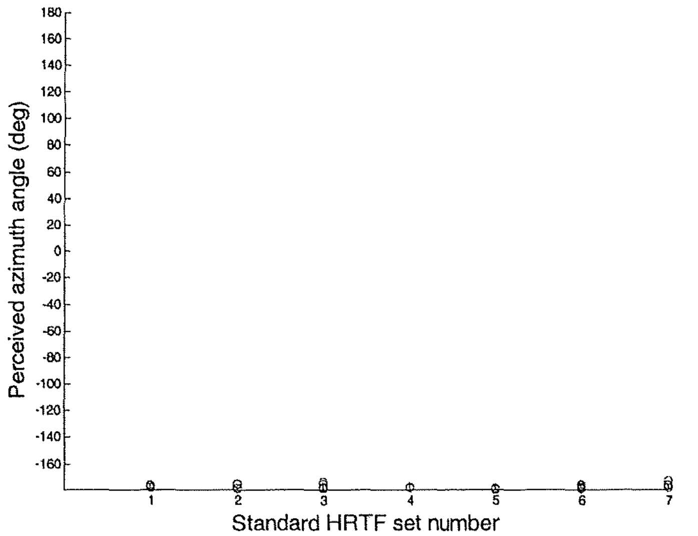

Figure 9.12 Plot of perceived incident angles along the azimuth directions for listener S15 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

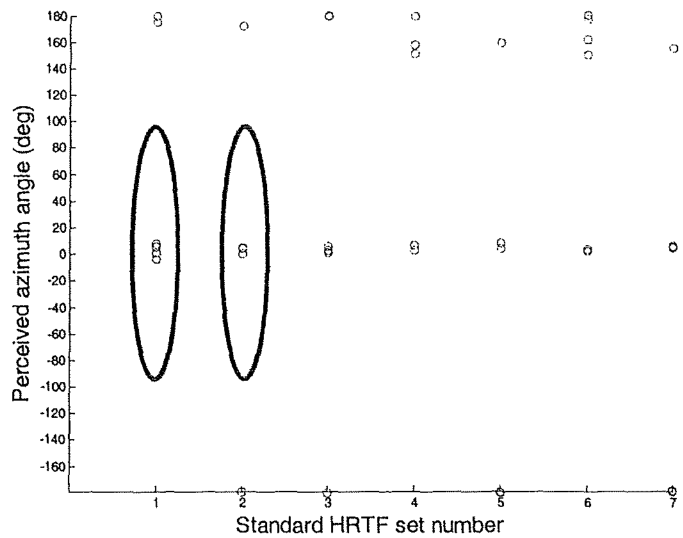

Figure 9.13 Plot of perceived incident angles along the azimuth directions for listener S16 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

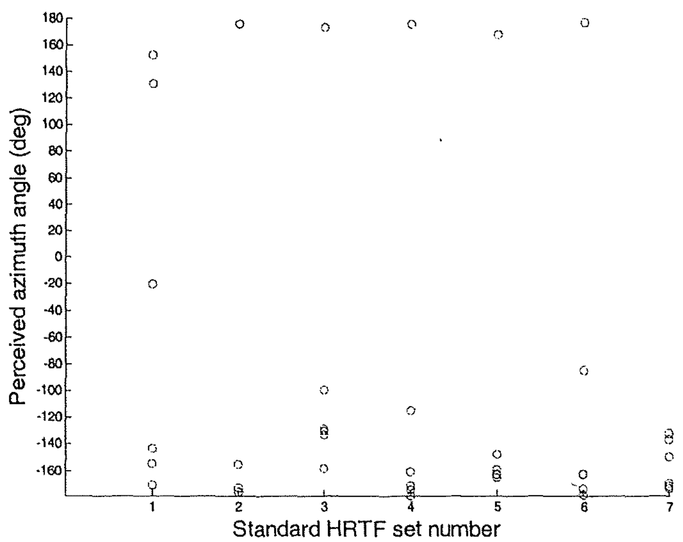

Figure 9.14 Plot of perceived incident angles along the azimuth directions for listener S17 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

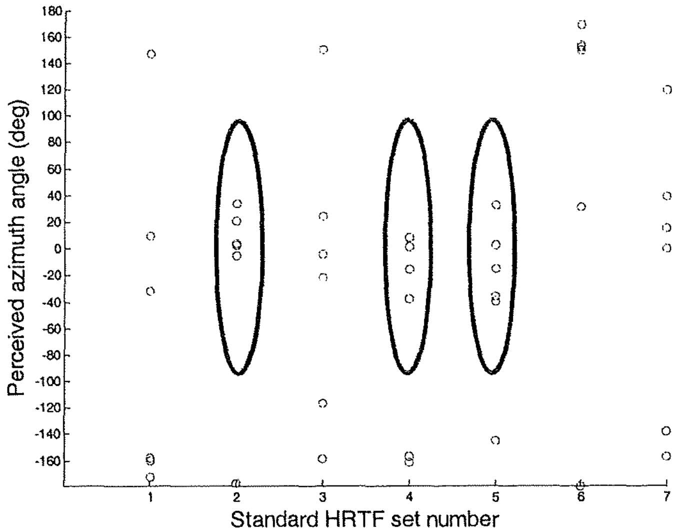

Figure 9.15 Plot of perceived incident angles along the azimuth directions for listener S18 while listening to the sound cue directly in the front (0°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

Table 9-1 The standard HRTF set for front number and number of repetition perceived at the front while listeners perceived more front direction than back direction

Listener	Standard HRTF set for front number	Number of repetition perceived at the front out of six repetitions
S8	1	4
	2	4
	4	4
	5	4
S11	1	5
	4	5
S13	3	5
S14	1	4
	7	4
S16	1	4
	2	4
S18	2	5
	4	4
	5	5

9.7.2. Back cue

Plots of perceived incident angles along the azimuth directions for each of the 15 listeners, while listening to the sound cue directly in the back (180°,0°), are shown in Figure 9.16 to Figure 9.30 as functions of the 7 standard HRTF sets. Data from all six repetitions are shown on the same figure.

In general, all 15 listeners had similar results to those in experiment 1. All of them performed satisfactorily with less than 50% back-front confusion rate in at least one set of back cues generated by standard HRTF for back.

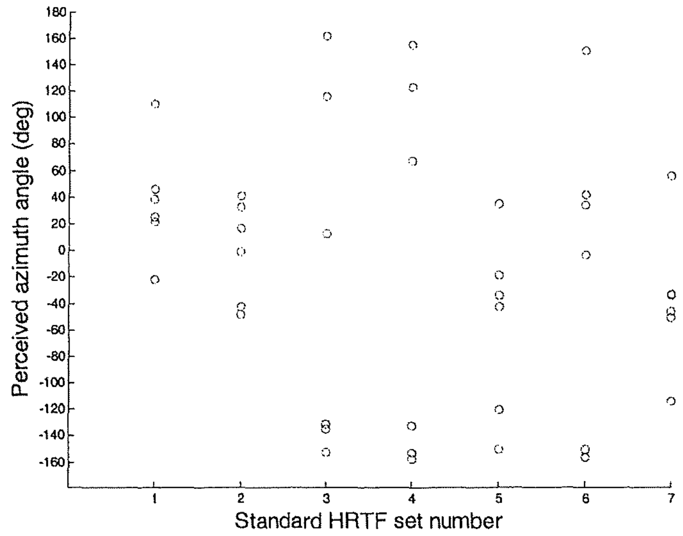

Figure 9.16 Plot of perceived incident angles along the azimuth directions for listener S4 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

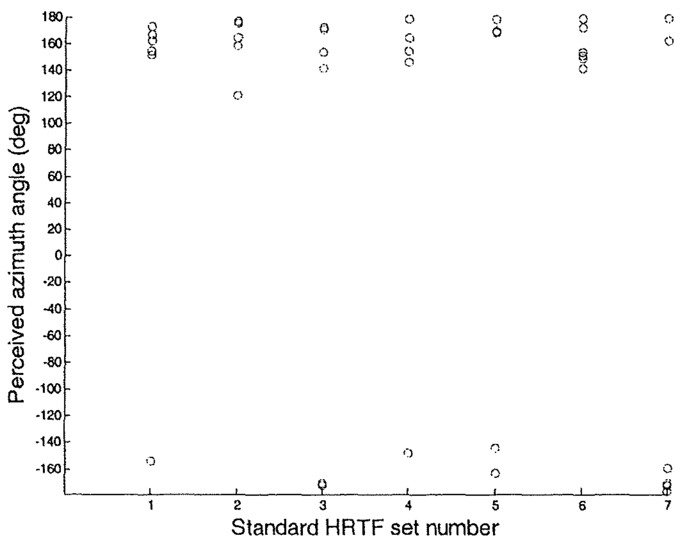

Figure 9.17 Plot of perceived incident angles along the azimuth directions for listener S5 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

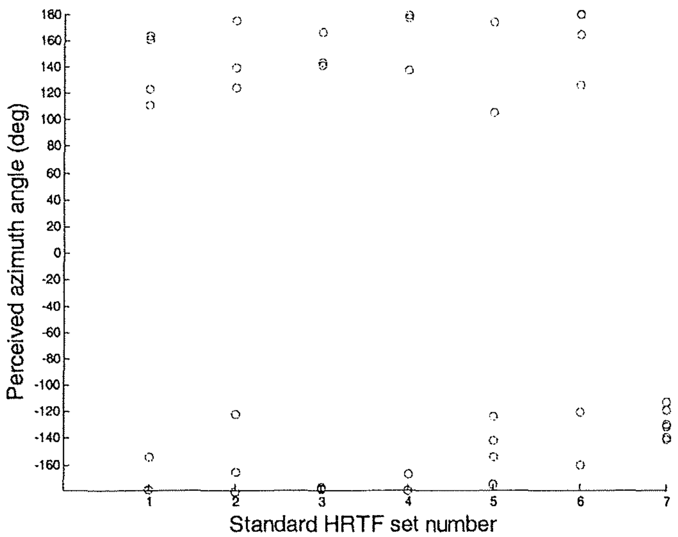

Figure 9.18 Plot of perceived incident angles along the azimuth directions for listener S6 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

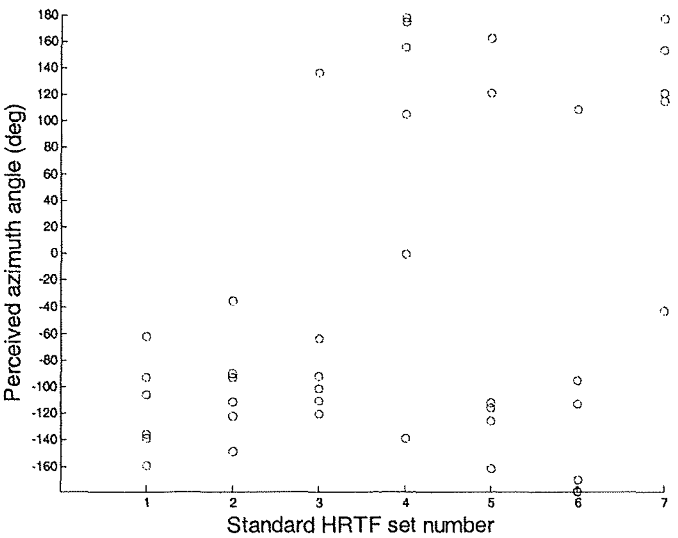

Figure 9.19 Plot of perceived incident angles along the azimuth directions for listener S7 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

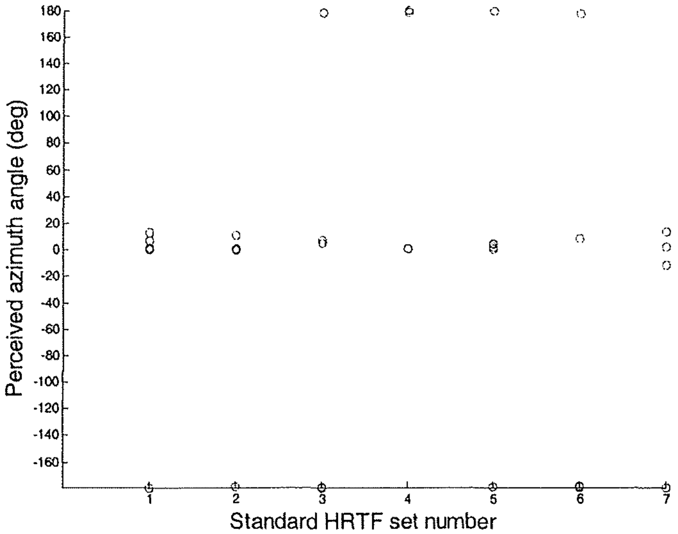

Figure 9.20 Plot of perceived incident angles along the azimuth directions for listener S8 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

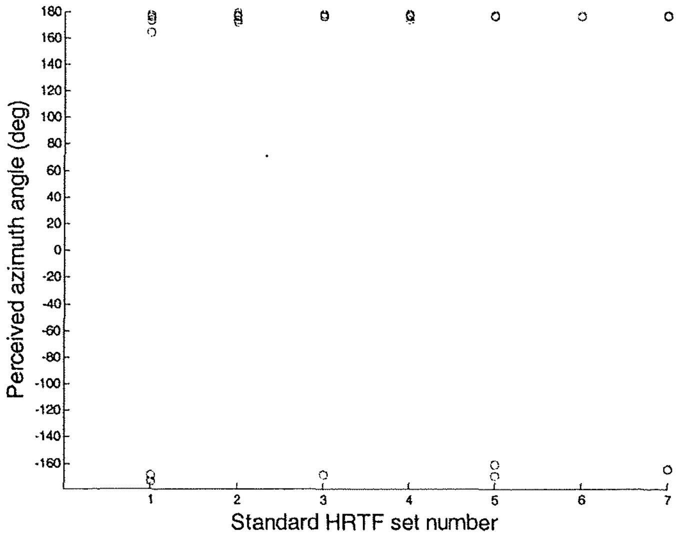

Figure 9.21 Plot of perceived incident angles along the azimuth directions for listener S9 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

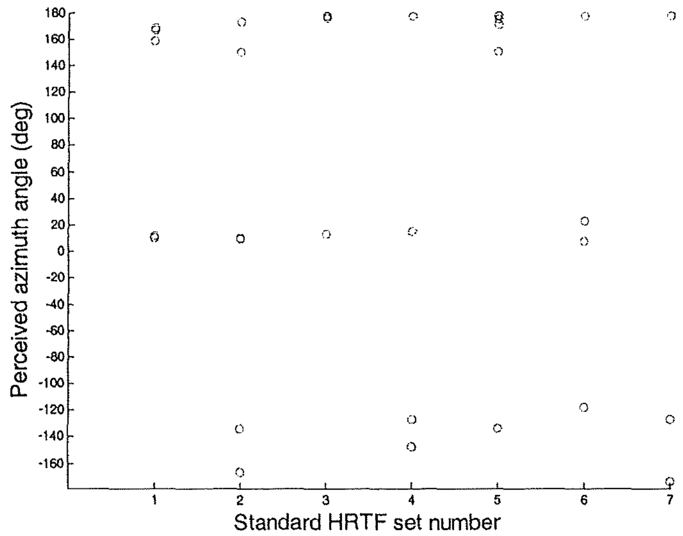

Figure 9.22 Plot of perceived incident angles along the azimuth directions for listener S10 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

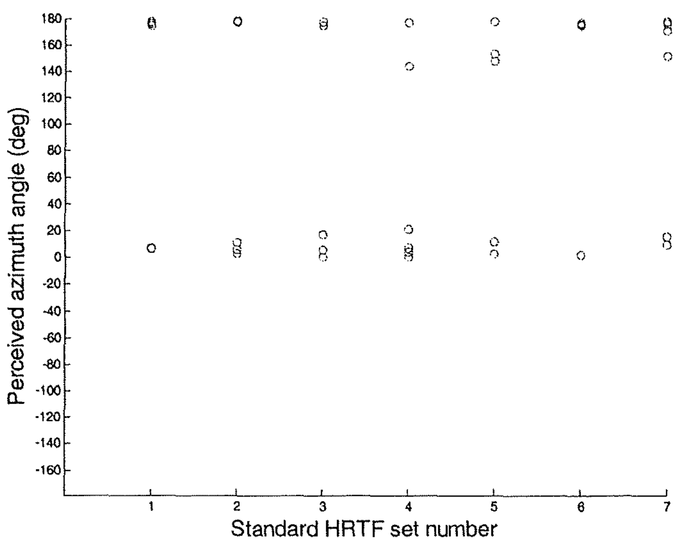

Figure 9.23 Plot of perceived incident angles along the azimuth directions for listener S11 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

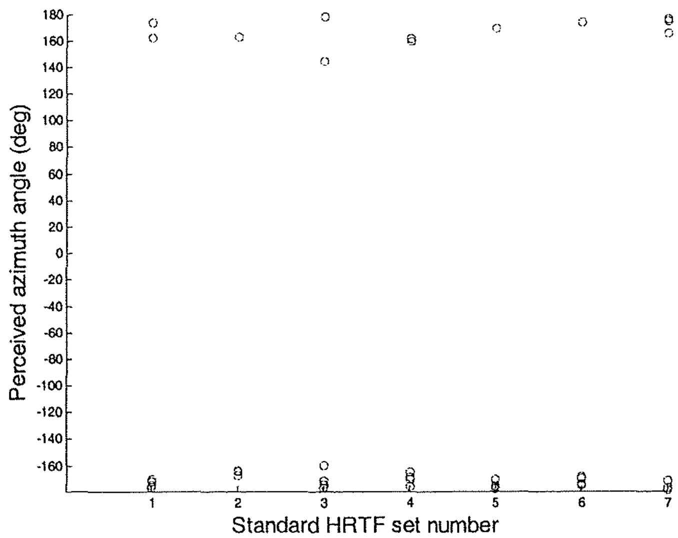

Figure 9.24 Plot of perceived incident angles along the azimuth directions for listener S12 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

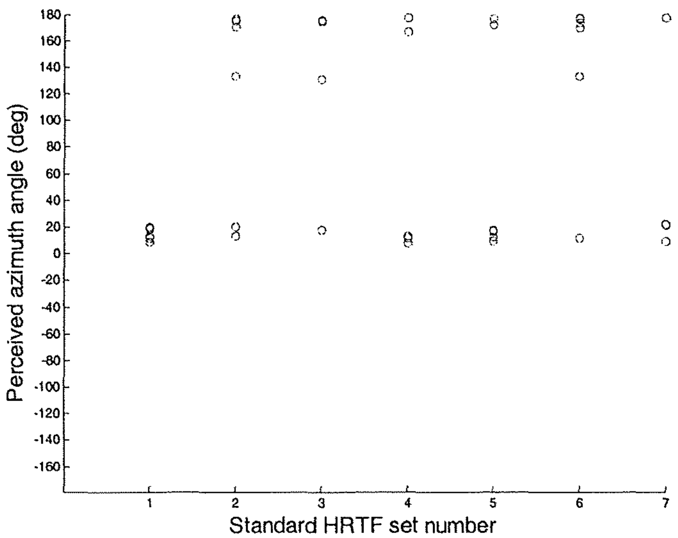

Figure 9.25 Plot of perceived incident angles along the azimuth directions for listener S13 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

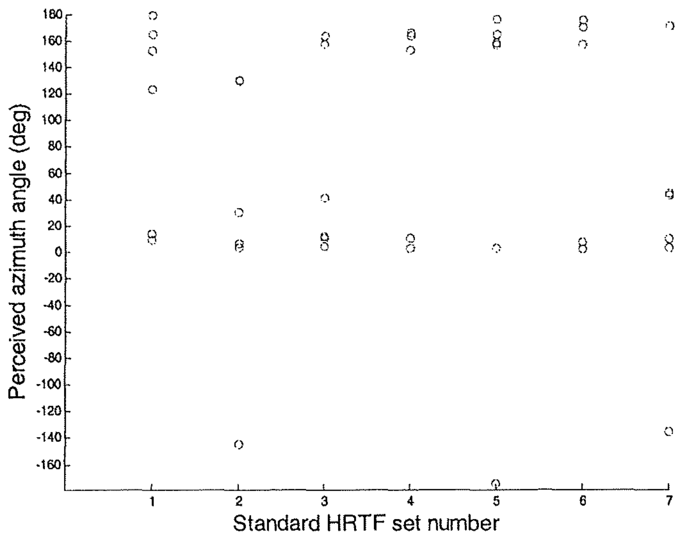

Figure 9.26 Plot of perceived incident angles along the azimuth directions for listener S14 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

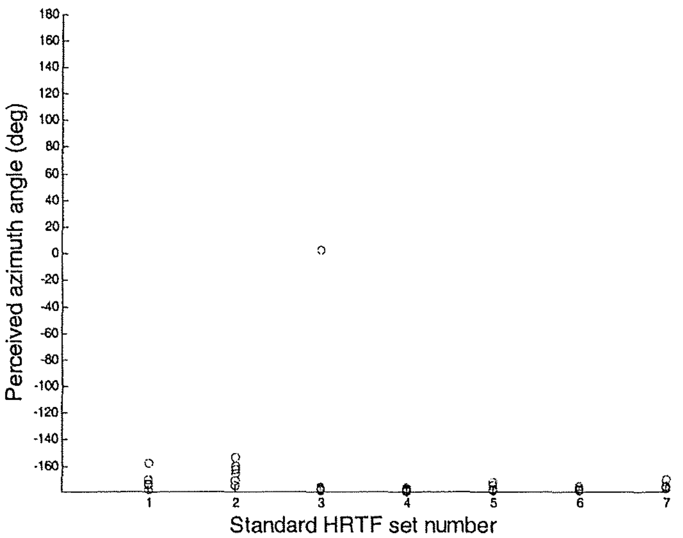

Figure 9.27 Plot of perceived incident angles along the azimuth directions for listener S15 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

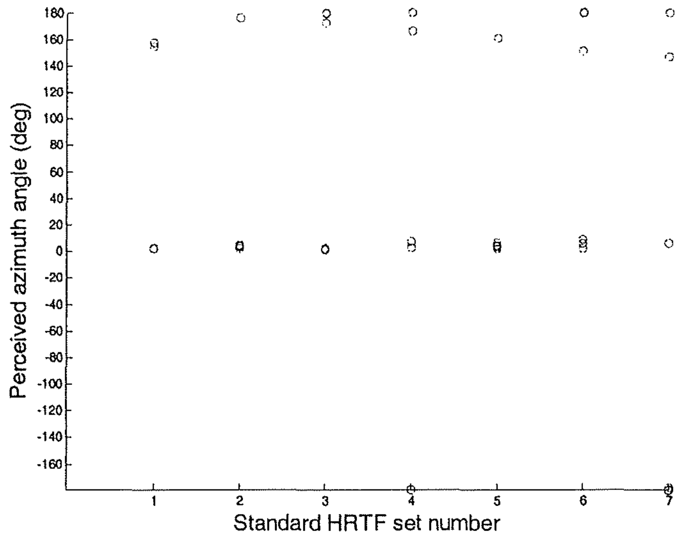

Figure 9.28 Plot of perceived incident angles along the azimuth directions for listener S16 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

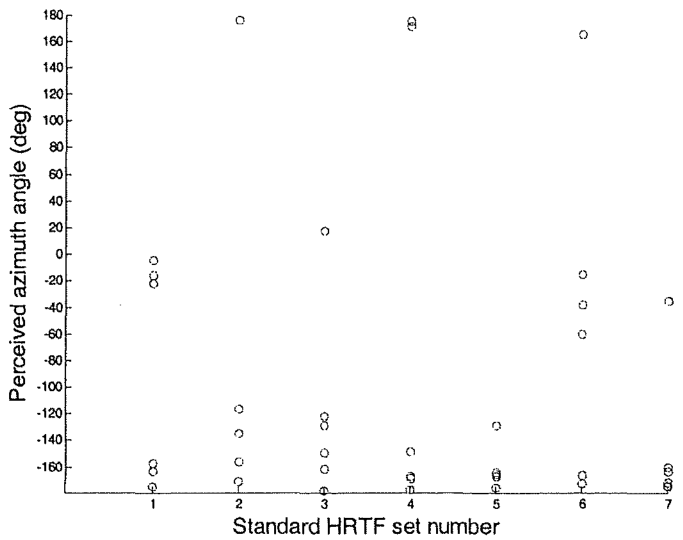

Figure 9.29 Plot of perceived incident angles along the azimuth directions for listener S17 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

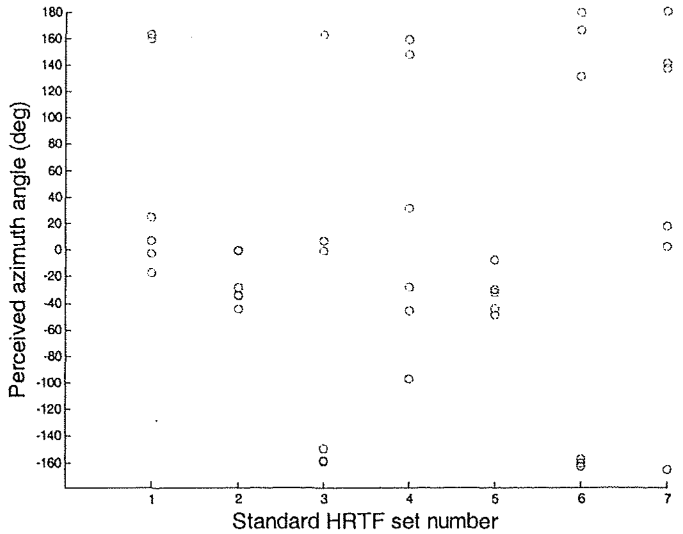

Figure 9.30 Plot of perceived incident angles along the azimuth directions for listener S18 while listening to the sound cue directly in the back (180°,0°) and as functions of the 7 standard HRTF sets. Data from all six repetitions are also shown and the meaning of the 7 standard HRTF sets can be found in Table 6-6. Frontal perceived azimuth angle: from -90° to 90°; Backward perceived azimuth angle: from -180° to -90° and from 90° to180°.

There were 6 listeners (S5, S6, S9, S12, S15 and S17) who perceived all back perception in all sound cues provided in experiment 1. Four of them (S5, S6, S9 and S12) had a consistent result, i.e. 0% back-front confusion rate. The back-front confusion rate of listeners S7, S15 and S17 increased in some of the back cues generated by standard HRTF sets for back. The result was tabulated in Table 9-2.

Table 9-2 The standard HRTF set for back and the corresponding back-front confusion rate in experiment 2 for listener S7, S15 and S17

Listener	Standard HRTF set number for back	Back-front confusion rate in experiment 2
S7	1	17%
	2	33%
	3	17%
	4	17%
	7	17%
S15	3	17%
S17	1	50%
	3	17%
	6	50%
	7	17%

For 4 listeners (S8, S11, S13 and S14) who had less than 50% back-front confusion rate in at least one set of standard HRTF for back in experiment 1, they still had less than 50% back-front confusion rate in at least one of the back cues generated by standard HRTF set for back in experiment 2.

Table 9-3 The standard HRTF set number for back which the listener S8, S11, S13 and S14 perceived less than 50% back-front confusion

Listener	Standard HRTF set number for back which the listener perceived less than 50% back-front confusion
S8	3, 6
S11	1, 5, 6, 7
S13	2, 3, 6
S14	1, 5

For those listeners (S4, S10, S16 and S18) who had 0% back-front confusion rate in at least one set of back cue generated by standard HRTF sets for back in experiment 1, two of them (S4 and S16) had degraded performances such that they had no standard HRTF sets for back which gave them 0% back-front confusion rate. Nonetheless, they still had less than 50% back-front confusion rate in at least one set of back cues generated by standard HRTF sets for back.

9.7.3. Learning effect

The perceived azimuth angle and perceived elevation angle did not follow normal distribution ($p<0.0001$, Shapiro-Wilk test), Friedman two-way analysis of variance by ranks was employed to test the effect of repetition in perceived azimuth angle and perceived elevation angle. Data from listening to the cues generated from the seven standard HRTF sets were used as repeated measures in the tests for effects of repetition. Results indicated that 16 out of the 120 cases (15 listeners x 4 cue directions x 2 types of perceived angle) reported significant effects of repetitions ($p<0.05$). The listeners and the angles of which repetition had most significant effect are shown in Table 9-4 and Table 9-5. Similar to the analysis of intra-subject variability in preliminary experiment, the repetition number in ascending order of perceived angle had shown no evidence to support there was learning effect.

Table 9-4 Friedman test result on listeners and angles which repetition had significant effect on perceived angle along azimuth direction

Angle	Listener	*p*-value	Repetition number in ascending order of perceived azimuth angle
(0,0)	S5	0.031	4, 1, 6, 3, 2, 5
	S6	0.050	4, 5, 2, 3, 1, 6
	S13	0.001	4, 2, 6, 3, 5, 1
	S14	0.021	6, 1, 2, 4, 5, 3
	S16	0.014	4, 6, 3, 2, 1, 5
(180,0)	S5	0.041	4, 6, 3, 5, 1, 2
	S6	0.005	2, 6, 3, 5, 1, 4
	S11	0.021	1, 5, 3, 2, 6, 4
	S12	0.037	4, 6, 2, 5, 1, 3

Table 9-5 Friedman test result on listeners and angles which repetition had significant effect on perceived angle along elevation direction

Angle	Listener	*p*-value	Repetition number in ascending order of perceived elevation angle
(0,0)	S5	0.004	1, 6, 3, 4, 5, 2
	S6	0.003	1, 3, 6, 5, 4, 2
	S9	0.028	1, 2, 3, 4, 6, 5
	S11	0.028	5, 1, 3, 6, 4, 2
	S17	0.003	1, 3, 6, 5, 2, 4
(180,0)	S7	0.029	5, 2, 4, 3, 6, 1
	S16	0.005	4, 6, 2, 3, 1, 5

9.7.4. Summary of experiment 2

Listening to the front cue and back cue in a pairwise way was supposed to create a reference point for each listener in the hope that their localization performance would increase. However, 5 listeners who perceived 100% front-back confusion in experiment 1 were still suffering from serious front-back confusion in this experiment. It was suspected that the HRTF sets used in the cluster analysis were all not at all suitable for most of the listeners. The performance in back cues were similar to the performance in experiment 1. It was also consistent with preliminary experiment and experiment 1 that no learning effect was found in experiment 2.

Chapter 10. Experiment 3: Exhaustive study of listeners' localization performance with simulated binaural cues

10.1. Purpose of the experiment

The result of the last two experiments indicated that the method of clustering HRTF sets and using standard HRTF sets as binaural cues was successful for a minority of listeners. There are several possible reasons for this i.e. too few HRTF sets for cluster, incomplete study of spectral cues in the literature, learning effects and spectral cues is not the only factor for front-back and elevation confusions. This experiment aimed to verify listeners' sound localization performance on different sets of HRTFs used in band characteristics extraction and cluster analysis.

10.2. Objective and hypothesis

This experiment was conducted to study listeners' sound localization performance, specifically on front cues, on different sets of HRTFs used in band characteristic extraction and cluster analysis. In previous experiments, listeners localized most of the sound cues at the back. Since most of the HRTF sets were collected in either Europe or America and all of the listeners who participated in this study were Chinese, one of the possible reasons of having high rate of front-back confusion in experiments is the incompatibility of European's HRTF and Asian's HRTF. Hence, it was hypothesized that (H8) all of the HRTFs sets used in this study are not suitable for generating front and back cues for listeners who had near 100% front-back confusion rate in experiment 1 and 2.

10.3. Participants

Fifteen listeners (S4-S18) who were involved in the last two experiments took part in this experiment. All of the listeners passed the audiometric test.

10.4. Dependent and independent variables

There is one dependent and one independent variable relating to this study. The occurrence of front-back confusion was the dependent variable and the independent variable was the 196 HRTFs sets.

The HRTFs sets include CIPIC HRTF database, IRCAM LISTEN HRTF database, MIT KEMAR HRTFs set, author's individualized HRTFs set and another set of individualized HRTFs of a Chinese male.

10.5. Stimuli preparation

The format of each stimulus is the same as stimuli used in experiment 2. Each stimulus included ten 0.3 s duration white noise pulses and the inter-pulse duration was 0.3s, filtered by predefined HRTF sets under Matlab environment.

10.6. Procedure

The procedure is similar to the experiment 1 procedure. Listeners were asked to sit comfortably and take rest for 5 minutes. A sphere and a pen with a sensor of Polhemus Fastrak attached were given to the subject. Instructions on using the pen as measuring equipment were given by the experimenter. This measuring equipment was used for measuring the elevation and azimuth angle that listeners perceived.

Stimuli were presented through Sennheiser HD580 headphones. After each stimulus, subject had to decide whether (i) the sound source was inside the head, outside the head or on the skull and (ii) the direction of the sound source which is most likely to be the sound source. A GUI program developed under Matlab environment was developed to input the responses. Trial measurements were conducted until listeners can fully understand the whole procedure. Then, first 5 measurements were discarded. Listeners were asked to close the eyes while listening to the stimuli.

To avoid fatigue of listeners, the experiment was divided into 4 sessions. Between each session, there was a 5-minutes break.

10.7. Results and Discussion

Four listeners who reported a less than 50% front-back confusion rate in experiment one also reported 0% front-back confusion in experiment three with cues from their best matched standard HRTF sets identified in experiment one. It is interesting to find out that when these four listeners localized the front cues generated from 196 HRTF sets taken from the CIPIC, LISTEN, and MIT KEMAR HRTF database, they had reported front-back confusion with 123, 94, 169, and 137 out of the 196 cues. In other words, randomly speaking, the chances of finding a cue that is matches to these four listeners is 37%, 52%, 14%, and 30%. However, with the six standard HRTF sets (the MIT KEMAR HRTF set was excluded because it was not the direct result of the cluster analyses), two experiments have repeatedly verified that at least one HRTF set can produce matching frontal cues for these four listeners.

Among the six listeners who perceived all or nearly all back perceptions in front cues presented in experiment 1 and 2, three of them (S6, S15 and S17) localized all of the sound cues presented at the back and two of them (S5 and S9) localized nearly all (98%) of the sound cues presented at the back. Another two listeners (S7 and S12) had 51% and 34% respectively of HRTFs sets that they can localize the sound cues at the front. The result was summarized in Table 10-1.

Table 10-1 Percentage of HRTFs sets that listeners localized at the back in experiment 3

Listener	S5	S6	S9	S12	S15	S17
Percentage of HRTFs sets localized at the back (i.e. front-back confusion)	98%	100%	98%	66%	100%	100%

It was hypothesized that all of front cues generated by the HRTF sets used in the study gave back perception, i.e. front-back confusion, to listeners who only perceived back direction in experiment 1 and 2. Five out of 6 listeners perceived all or nearly all back perceptions in the virtual sound cues which were generated by the HRTF sets used in cluster analysis. This result indicated that the variety of HRTF sets used for analysis is not large enough. It was suggested more HRTF sets should be collected for further analysis. Listener S7 and S12 perceived front perception in 4 and 3 respectively out of 7 virtual sound cues generated by standard HRTF sets for front which were used in experiment 1 and 2. However, both perceived 100% back perception in experiment 1 and 2. The reason of having this phenomenon is unknown.

Chapter 11. Conclusion and future work

11.1. Discussion

The goal of this study was to optimize listeners' front/back confusion with standard HRTF sets. From the study of inter-subject and intra-subject variability, it was found that inter-subject standard deviation of elevation error and azimuth error were statistically larger (*p*=0.001 and 0.091 respectively) than the corresponding intra-subject standard deviations. This implied that using more than one set of HRTFs for listeners can achieve better sound localization performance.

It was hypothesized that at least one set of standard HRTF set for each of the listeners gave a better result than a generic non-individualized HRTF. This hypothesis held for 4 out of 15 listeners in experiment 1 and 6 listeners in experiment 2. The main effects of listener were significant ($p<0.03$, experiment 1; $p<0.007$, experiment 2, Kruskal-Wallis test) and there were 6 out of 15 listeners who perceived 100% back perception when single sound cues, which were filtered by standard HRTF sets for front, were presented. The remaining 5 listeners showed an improvement trend in sound localization performance with the choice of six standard cues. When the data from all 15 listeners are analyzed, providing the choice of six standard center-front cues significantly improved the sound localization performance. In other words, localization errors obtained when listeners were presented with the choice of six standard center-front cues were significantly lower than those errors obtained when listeners were presented with anyone of the standard center-front cues or the MIT KEMAR center-front cue.

To explore the possible cause for the 6 listeners who have reported 100% or near 100% front/back confusion rates, further studies were done on localizing the front cues exhaustively of all of the 196 HRTF sets used in the cluster analysis. For these 6 listeners, 3 of them perceived 100% and 2 of them perceived 98% backward perception in localizing each of the 196 frontal sound cue generated from the HRTF sets used in cluster analysis. This suggests that the 196 non-individualized HRTF sets, including HRTF sets downloaded from the CIPIC, LISTEN and KEMAR databases, did not match with the acoustics characteristics of the ears of these 5 listeners.

The proven benefits of providing a choice of six standard center-front and center-back cues make it possible to apply the mass-customization concept on designing non-individualized HRTF devices. The idea of mass customization is to provide a customized service or product at mass production cost and efficiency (Tseng and Piller, 2003 [33]). The application of this concept to non-individualized HRTF devices suggests a design methodology to customize inexpensive non-individualized HRTF devices to achieve better performance that is similar to those achieved by individualized HRTF devices. The latter is very expensive to make.

11.2. Conclusion

11.2.1. Cluster analysis on band characteristics

This thesis successfully reduces 196 published open- copyrighted HRTF sets (102 sets from LISTEN: IRCAM and AKG acoustics, 2004 [13]; 90 sets from CIPIC, Algazi *et. al.* 2001 [2]; 2 sets from author's individualized HRTF and 2 sets from KEMAR: Gardner and Martin 1995 [11]) into (i) 6 clusters according to their abilities to simulate binaural cues coming from the front and (ii) 6 clusters according to their abilities to simulate binaural cues coming from the back. The cluster analyses are consistent with the findings of spectral characteristics as reported in the core literature (e.g., Blauert, 1969/70 [7]; Hebrank and Wright, 1974 [12]; Middlebrooks, 1992 [18]; and Langendijk and Bronkhorst, 2002 [16]). These clusters are new, original and have not been reported before.

Traditionally, studies of perceived directions of binaural cues generated using non-individualized HRTFs have used a small number of selected listeners (e.g., Asano *et al.*, 1990 [4], n=2; and Middlebrooks, 1992 [18], n=5). This thesis studied inter-subject and intra-subject variability with 10 randomly selected listeners and sound localization performance with 15 randomly selected listeners. The larger number of listeners enable statistical analyses that make the findings from this thesis more robust than findings of previous studies.

11.2.2. Intra-subject and inter-subject variability

This study had studied and reported the inter-subject and intra-subject variability with 10 listeners. Significant two-way interactions between effects of listeners and effects repetition have been consistently found. Although evidence of learning on sound localization performance with binaural cues was not found, data suggest that future studies in sound localization experiments need to include repetitions so that the mentioned two-way interactions can be measured. This finding is important because the mentioned two-way interactions could affect any main effects that are of interests to future researchers (Chapter 3).

Inter-subject variability in sound localization performance with binaural cues generated using non-individualized HRTFs had been shown to be significantly greater than intra-subject variability. This indicated that listeners' sound localization performance are very different and using just one generic set of non-individualized device is not good enough to produce accurate sound cues. This finding is important because it supports the hypothetical benefits of optimizing sound localization performance of generically simulated binaural cues by providing a choice of cues generated using a range of standard HRTF sets (Chapter 3).

11.2.3. Front-back confusion

Twenty-seven percents of the sampled population benefited significantly from the choice of standard HRTF sets. In particular, if these listeners were given the choice to choose from the binaural cues generated using the standard HRTF sets, they were able to identify at least one HRTF set that can produce cues with less than 50% front/back confusion rate (Chapter 8). This percentage increased to forty percents when the experiment was repeated with a change in cue presentation method (Chapter 9). Besides the twenty-seven percent, thirty-three percentages of the listeners in Experiment one also showed some improvements in sound localization performance when a choice of six standard cues is provided. The combined effect is that when data from all 15 randomly selected listeners were analyzed, significant reductions in sound localization errors were found when listeners were presented with a choice of six standard cur rather than just one generic cue. This demonstrates the benefits of providing a choice of cues generated from the optimized sets of standard HRTFs (Chapters 8 and 9).

Forty percent of the sampled population reported near 100% front-back confusion rates with frontal cues generated from the six standard HRTFs representing the well-known CIPIC, LISTEN, and MIT KEMAR HRTF databases (Chapter 8). This finding was repeated in a second experiment (Chapter 9). This finding is important because currently commercial HRTF-based products are using the open-copyrighted HRTF sets from CIPIC, LISTEN, and MIT KEMAR HRTF databases and these databases represent the major sources for open-copyrighted HRTF sets. Among this forty percent of the sampled population, seventy-one percent of them reported greater than 98% of front-back confusion rate when they were asked to localize frontal cues generated from all the open-copyrighted HRTF sets from the CIPIC, LISTEN, and MIT KEMAR HRTF database. This suggests that either the scope of these open-copyrighted HRTF sets were too limited or some listeners were incapable to perceive frontal cues from binaural cues generated using non-individualized HRTFs. The latter reason is consistent with a documented report that a listener failed to correctly localize binaural cues generated from his own individualized HRTF set despite repeated attempts (Asano *et al.*, 1990 [4]). Because there has been no report to classify listeners according to their ability to localize HRTF-based binaural cues, the finding concerning the grouping of listeners according to their sound localization performance carries important contribution.

11.2.4. Back-front confusion

This thesis confirms that back-front confusion is not a problem associated with localize binaural cues generated using non-individualized HRTFs. Previous studies reporting similar results using only 2 or 3 listeners (Asano *et al.*, 1990 [4], n=2; and Middlebrooks [18], 1992, n=5) but this study used 15 listeners and the results were verified with robust statistical analyses.

11.3. Future work recommendations

Further studies on relationships between spectral cues and ear shape is suggested. Individualized HRTFs can be generated artificially by capturing the ear features if the relationships are known. Since measuring individualized

HRTFs using traditional method is expensive, artificial generated HRTFs assist with reducing the cost. On the other hand, grouping listeners can be done without acquiring HRTFs. Preliminary studies concerning clustering with ear dimensions have been carried out by Ngan *et. al.* (2005) [32].

REFERENCES

[1] M.S. Aldenderfer, R.K. Blashfield (1984). Cluster analysis. Sage publications, Inc., USA.

[2] V.R. Algazi, R.O. Duda, D.M. Thompson, C. Avendano (2001). The CIPIC HRTF database. In 2001 IEEE ASSP Workshop on Applications of Signal Processing to Acoustics, pp. 111-123, New Paltz, NY, USA, October 2001.

[3] A. Apostolico, Z. Galil (1997). Pattern Matching Algorithms. Oxford University Press, Oxford.

[4] F. Asano, Y. Suzuki, T. Sone (1990). Role of spectral cues in median plane, J. Acoust. Soc. Am., vol. 88(1), pp. 159-168.

[5] D.W. Batteau (1967). The role of pinna in human localization. Proc. R. Soc. London Ser. B 168, pp. 158-180.

[6] D.R. Begault (1994). 3-D sound for virtual reality and multimedia. Academic Press Professional, Inc., San Diego, CA.

[7] J. Blauert (1969/70). Sound localization in the median plane, Acustica, vol. 22, no. 4, pp. 205-13.

[8] S. Carlile (1996). Virtual auditory space: generation and applications, R.G. Landes Company and Chapman & Hall, Texas.

[9] M.B. Gardner, R.S. Gardner (1973). Problem of localization in the median plane: effect of pinnae cavity occlusion, J. Acoust. Soc. Am., vol. 53(2), pp. 400-408.

[10] W.G. Gardner (1998). 3-D audio using loudspeakers. Kluwer Academic Publishers, USA.

[11] W.G. Gardner, K.D. Martin (1995). HRTF measurements of a KEMAR. J. Acoust. Soc. Am., 97(6), pp. 3907-3908, 1995.

[12] J. Hebrank, D. Wright (1974). Spectral cues used in the localization of sound sources on the median plane, J. Acoust. Soc. Am., vol. 56(6), pp. 1829-34.

[13] Ircam and AKG Acoustics (2004). LISTEN HRTF Database. http://www.ircam.fr/equips/salles/listen/index.html.

[14] B. Kapralos, M. Jenkin, E. Milios (2003) Auditory Perception and Spatial (3D) Auditory Systems. Technical Report CS-2003-07, York University.

[15] G.F. Kuhn (1987), Physical acoustics and measurements pertaining to directional hearing, Directional hearing (ed. by Yost, W.A. and Gourevitch, G.), Springer-Verlag, New York, pp. 3-25.

[16] E.H.A. Langendijk, A.W. Bronkhorst (2002). Contribution of spectral cues to human sound localization, J. Acoust. Soc. Am., vol. 112(4), pp. 1583-96.

[17] Lord Raleigh, J.W. Strutt (1907). Our perception of sound direction. Philosophical Magazine, 12:214-232.

[18] J.C. Middlebrooks (1992). Narrow-band sound localization related to external ear acoustics, J. Acoust. Soc. Am., vol. 92(5), pp. 2607-24.

[19] J.C. Middlebrooks, J.C. Makosu, D.M. Green (1989) Directional sensitivity of sound-pressure levels in the human ear canal. Journal of the Acoustical Society of America, 86, pp.89-108.

[20] G.W. Milligan (1980), An Examination of the Effect of Six Types of Error Perturbation on Fifteen Clustering Algorithms, Psychometrika, 45, 325 -342.

[21] A.D. Musicant, R.A. Butler (1984). The influence of pinnae-based spectral cues on sound localization, J. Acoust. Soc. Am., vol. 75(4), pp. 1195-200.

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

[23] J.H. Ward (1963). Hierarchical Grouping to Optimize an Objective Function, Journal of the American Statistical Association, 58, 236 -244.

[24] E.M. Wenzel, F.L. Wightman, D.J. Kistler, S.H. Foster (1988). Acoustic origins of individual differences in sound localization behaviour. J. Acoust. Soc. Am., vol. 84, S79.

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

[26] C.J. Tan, W.S. Gan (1998). User-defined spectral manipulation of HRTF for improved localisation in 3D sound systems. Electronics letters, vol. 34, no. 25, 2387-2389.

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

[28] V.R. Algazi, C. Avendano, R.O. Duda (2001). Elevation localization and head-related transfer function analysis at low frequencies. J. Acoust. Soc. Am., vol. 109, 1110-1116.

[29] S. Carlile (1996). Virtual Auditory Space: Generation and Application. R.G. Landes Company, Austin, TX, USA.

[30] R.H. Gilkey, M.D. Good, M.A. Ericson, J. Brinkman, J.M. Stewart (1995). A pointing technique for rapidly collecting localization responses in auditory research. Behavior Research Methods. Instruments & Computers, vol. 27(1), 1-11.

[31] R.O. Duda, P.E. Hart, D.G. Stork (2001). Pattern Classification. John Wiley & Sons, Inc., New York, NY, USA.

[32] K.H. Ngan, R.H.Y. So, A.K.L. Leung, E. Altinsoy, A.G. Avalos (2005). Towards mass-customizing non-individualized head-related transfer functions: A preliminary Study. Annual Journal of IIE (HK), Vol. 25 (2004/2005), 14-25.

[33] M. M. Tseng, and F.T. Piller (2003). The customer centric enterprise: advances in mass customization and personalization. Springer-Verlag, Berlin.

[34] N.M. Leung (2001). Optimizing the performance of Virtual Headphone-based Surround Sound (VHSS) Systems by manipulating the spectra of non-individualized HRTFs. Thesis, Hong Kong University of Science and Technology, Hong Kong.

[35] F.L. Wightman, and D.J. Kistler (1989). Headphone simulation of free-field listening. II: Psychophysical validation. J. Acoust. Soc. Am. vol. 85, 868-878.

[36] P.M. Hofman, A.J.V. Opstal (1998). Spectro-temporal factors in two-dimensional human sound localization. J. Acoust. Soc. Am. vol. 103(5), 2634-2648.

[37] J. Blauert (1997). Spatial Hearing: The Psychophysics of Human Sound Localization. MIT Press, London, England.

[38] S.M. Abel, S.J.E. Paik (2004). The benefit of practice for sound localization without sight. Applied Acoustics. Vol. 65(3), 229-241.

APPENDICE

Appendix A Procedures of feature extraction of HRTFs

1. The frequency band characterization is defined according to the literatures.

For the band characterizations that the spectral profile content related to human sound localization (Band FF) pattern matching method (Apostolico, 1997) is used for feature extraction. The cosine similarities between the HRTFs sets are calculated. The formula is as follows:

$\vec{x} = [x_1, \ldots, x_n]^T$ is the frequency band characterization of a set of HRTFs, where n is sequence number of HRTF data.

$\vec{y} = [y_1, \ldots, y_n]^T$ is the frequency band characterization of another set of HRTFs, where n is sequence number of HRTF data.

$$PatternCosSim(\vec{x}, \vec{y}) = \sum_{i,j=1}^{i,j=n-4} CosSim[(x_i, x_{i+1}, x_{i+2}, x_{i+3}, x_{i+4}), (y_j, y_{j+1}, y_{j+2}, y_{j+3}, y_{j+4})]$$

$$CosSim(\vec{x}, \vec{y}) = \frac{\vec{x}.\vec{y}}{\|\vec{x}\| \cdot \|\vec{y}\|}$$

Then the similarity matrix is constructed.

2. For each of other band characterizations (Band FA, FB, FC, FD, FE, BA, BB, BC, UA), the parameters are retrieved according to Table 4-1. Those values are then normalized. If there is more than one parameter retrieved from a band, the average value of those parameters is used.
3. Then, the similarity matrix of each band is constructed. The formula for constructing the similarity matrix of each band is as follow. [formula f2]

$$Sim(\overrightarrow{x_1}, \overrightarrow{x_2}) = e^{\frac{-\|(\overline{x_1} - \overline{x_2})\|}{\bar{x}}}$$

4. The master similarity matrix for front cue is calculated by adding all of the similarity matrixes related to front cue (Band FA, FB, FC, FD, FE and FF) together. The master similarity matrix for back cue (Band BA, BB and BC) is calculated by the same method. This method assumes that all band characterizations are equally important. The master similarity matrix is then used for clustering.

Appendix B Failure of principal component analysis (PCA) to separate HRTF into different factors according to its spectral profile

Principal component analysis is a common tool for factor analysis. However it is not a suitable tool for analyzing different factors in spectral profiles. Here is the example to illustrate why PCA fails.

Four sets of data which have 5 features were generated. The basic sets are, A [10 20 30 40 50], B [14 24 34 44 54], C [20 30 40 50 60] and D [24 34 44 54 64]. Basic set C is the left-shifted version of set A. Sets D is the left-shifted version of set B. Each data was generated by adding a random error terms to one of the basic sets. There were 400 data generated. Then PCA was conducted and data were transformed. The transformed data were clustered into 2 groups. A good analysis should put sets A and C into one group and sets B and D into another.

PCA fails to detect the shifted data. The result showed that data from basic set A (100%) and B (92%) formed 1 group and data from set C (86%) and D (97%) formed another group.

From the idea of PCA, it helps noise reduction and normalization only. It doesn't detect the shift of the peak and trough, which are essential for this study. This study adopted the pattern matching method instead of PCA.

Appendix C Cluster analysis result of HRTF sets

Table C.1 Cluster analysis result of HRTF set.

HRTF database	Subject number	Left/Right ear	Cluster number (Front)	Cluster number (back)
CIPIC	3	L	1	1
		R	1	1
	8	L	1	9
		R	4	7
	9	L	1	3
		R	1	4
	10	L	6	1
		R	1	4
	11	L	1	3
		R	1	1
	12	L	5	5
		R	4	1
	15	L	6	1
		R	1	1
	17	L	1	4
		R	1	1
	18	L	6	2
		R	6	2
	19	L	6	1
		R	6	1
	20	L	6	1
		R	1	7
	21	L	1	3
		R	1	1
	27	L	6	1
		R	6	1
	28	L	4	1
		R	4	1
	33	L	1	4
		R	1	4
	40	L	6	1
		R	6	1

	44	L	6	7
		R	6	1
	48	L	1	4
		R	7	1
	50	L	1	1
		R	1	4
	51	L	1	1
		R	6	2
	58	L	6	1
		R	1	4
	59	L	1	1
		R	6	2
	60	L	4	1
		R	2	1
	61	L	7	7
		R	7	1
	65	L	1	4
		R	7	1
	119	L	7	1
		R	4	1
	124	L	1	4
		R	6	1
	126	L	6	7
		R	7	1
	127	L	6	1
		R	6	1
	131	L	6	6
		R	6	1
	133	L	1	4
		R	6	3
	134	L	1	1
		R	6	1
	135	L	1	1
		R	6	1
	137	L	2	1
		R	2	1
	147	L	1	3

		R	1	3
	148	L	2	4
		R	6	4
	152	L	8	3
		R	6	3
	153	L	6	1
		R	1	8
	154	L	6	1
		R	6	1
	155	L	1	4
		R	6	4
	156	L	6	1
		R	6	1
	158	L	6	1
		R	6	3
	162	L	1	1
		R	1	3
	163	L	6	3
		R	1	1
	165	L	1	1
		R	6	2
LISTEN	1002	L	6	1
		R	6	1
	1017	L	6	1
		R	6	2
	1033	L	1	1
		R	6	1
	1048	L	6	2
		R	6	1
	1003	L	6	2
		R	4	2
	1018	L	6	2
		R	1	1
	1034	L	7	1
		R	7	4
	1049	L	2	1
		R	6	1

	1004	L	6	2
		R	6	1
	1020	L	6	1
		R	6	4
	1037	L	6	1
		R	6	2
	1050	L	3	1
		R	6	2
	1005	L	6	1
		R	7	1
	1021	L	6	1
		R	6	1
	1038	L	6	1
		R	6	1
	1051	L	6	1
		R	6	4
	1006	L	1	4
		R	6	3
	1022	L	6	1
		R	6	1
	1039	L	6	2
		R	6	1
	1052	L	6	4
		R	6	1
	1007	L	6	1
		R	6	1
	1023	L	1	1
		R	1	4
	1040	L	6	4
		R	6	1
	1053	L	3	4
		R	6	1
	1008	L	2	1
		R	6	1
	1025	L	2	1
		R	2	4
	1041	L	6	1

		R	6	1
	1054	L	6	1
		R	6	1
	1009	L	6	1
		R	6	1
	1026	L	6	2
		R	6	1
	1042	L	6	1
		R	1	1
	1055	L	6	2
		R	6	2
	1012	L	6	1
		R	6	1
	1028	L	4	1
		R	6	4
	1043	L	6	1
		R	6	1
	1056	L	6	1
		R	6	1
	1013	L	1	1
		R	6	1
	1029	L	6	1
		R	1	7
	1044	L	6	1
		R	6	1
	1057	L	4	1
		R	6	1
	1014	L	6	1
		R	6	1
	1030	L	7	1
		R	6	1
	1045	L	1	1
		R	1	4
	1058	L	6	1
		R	6	1
	1015	L	6	1
		R	6	1

	1031	L	7	1
		R	6	1
	1046	L	7	1
		R	7	2
	1059	L	3	4
		R	6	2
	1016	L	6	1
		R	6	1
	1032	L	6	4
		R	6	4
	1047	L	6	1
		R	1	4
Author	N/A	L	9	5
		R	3	4
KEMAR	N/A	L	1	3
		R	4	2

Data with the same cluster number are in the cluster. There were 9 clusters in cluster for front and cluster for back.

Website for each HRTF database:

CIPIC http://interface.cipic.ucdavis.edu/CIL_html/CIL_whatis.htm

LISTEN http://recherche.ircam.fr/equipes/salles/listen/

KEMAR http://sound.media.mit.edu/KEMAR.html

Appendix D A sample of subject consent form for experiments

STATEMENT OF CONSENT TO TAKE PART AS A SUBJECT IN 3D SOUND LOCALIZATION EXPERIMENT

1. Name ________________________ (Mr/Mrs/Miss/)
2. Are you feeling ill in any way? Yes / No
3. Do you suffer from any form of hearing problems? Yes / No
4. Have you suffered any serious illness? Yes / No
5. Are you under medical treatment or suffering disability affecting your daily life? Yes / No

If you answer is 'Yes' to question (2), (3), (4) or (5), please give details to the Experimenter.

DECLARATION

I consent to take part in an experiment. My replies to the above questions are correct to the best of my belief, and I understand that they will be treated as confidential by the experimenter.

I understand that I may at any time withdraw from the experiment and that I am under no obligation to give reasons for withdrawal or to attend again for experimentation.

I undertake to obey the regulations of the laboratory and instructions of the experimenter regarding safety only to my right to withdraw declared above.

The purpose and methods of the research have been explained to me and I have had the opportunity to ask questions.

Signature of Subject ____________________ Date ____________________

This experiment conforms to the requirement of the University Research Ethic Committee.

Signature of Experimenter ________________ Date ____________________

Appendix E Questionnaire for experiment 1 to 3

Questionnaire on 3D sound localization experiment

Name:

Age:

Gender: M / F

School: Engineering Business Science Humanities
(please circle the appropriate choice.)

Do you suffer from noticeable hearing loss judged by medical staff?

Have you explored to loud noises (e.g. Amplified music, plane, factories) within the last 30 days? If you have, please give details.

Have you ever seek medical advice towards hearing problem? If you have, please give details.

Date:

Appendix F Instruction sheets in experiments

Experiment about 3D sound localization

Welcome for participating this experiment! Please read through this instruction first. If you have any question about the procedure, please ask the experimenter at any time.

The objective of this experiment is to investigate the 3D sound localization performance. The experiment is divided into 2 sessions. Two sessions are scheduled on two different days. Each session will take approximately 60 minutes of your time.

The procedures of the experiment are as follows.

1. You will be asked to sit comfortably and take rest for 5 minutes.
2. You will be asked to read this instruction and fill in the pre-experiment questionnaire and the consent form.
3. Audiometric test :
   Please put on a headphone. Hold the button in one of your hands and press the button down when you listen consecutive 'beep' sound from headphones. Please press the button until you cannot hear the sound. This process will repeat for several times. Further verbal instruction will be provided during the audiometric test.
4. 3D sound localization experiment:
   Detailed instructions will be given later. Briefly speaking, you will be listening to a series of sound clips. You have to point out whether this sound is coming from inside or outside of your head, the sound is compact or diffuse and also pointing out the incoming direction of the sound.
5. Ear measurement:
   We will measure your ears with a special designed equipment. You will be asked to sit comfortably.

If there is any problem during the experiment, please ask the experimenter. Thanks for your help!

IMPORTANT:
Please DO NOT mention any details of this experiment to the others

Details of 3D sound localization experiment

Part 1

You need to put on a headphone during the experiment.

The experimenter will broadcast a testing sound clip for testing purposes. Please make sure both the LEFT and RIGHT channels of the headphones have the same loudness by adjusting the headphone. If you have any difficulty, please let the experimenter know immediately.

After that, a series of sound clips will be played to test your ability to use the equipment.

Clips of sound will be played through the headphones. After listening to each sound clip, you have to decide i) it was coming from **INSIDE** or **OUTSIDE** or **on the skull** of your head ii) if the sound was coming from **INSIDE** of your head, the sound was **COMPACT** or **DIFFUSE** and iii) the **sound source direction**.

Please refer to the Figure 1 about the area of inside and outside of your head. Please say "**INSIDE**" or "**OUTSIDE**" or "**on the skull**" to indicate your choice. Next, if the sound was coming from **INSIDE** of your head, you have to decide the sound was **COMPACT** or **DIFFUSE**. If the sound is **perceived within a comparatively small region in space**, it will be precisely localized and described as '**COMPACT**. If the sound is **perceived as partially split in space**, e.g. echo, it is described as '**DIFFUSE**'.

To indicate the sound source direction, a sphere and a sensor will be given to you. Please refer to Figure 2 and 3 for instructions. This measuring equipment is for measuring the sound source that you perceived. Please point out the sound source direction that you think to be with the sensor.

If you have pointed out the direction, please say "**READY**" and hold the sensor until the experimenter says "**OKAY**". Then, next sound clip will be played and the procedures will be repeated. There are trials for you to familiarize with the equipments.
IMPORTANT:
Please **close your eyes** and **never move your head** while listening to the sound clip.

Details of 3D sound localization experiment

Part 2

You need to put on a headphone during the experiment.

The experimenter will broadcast a testing sound clip for testing purposes. Please make sure both the LEFT and RIGHT channels of the headphones have the same loudness by adjusting the headphone. If you have any difficulty, please let the experimenter know immediately.

After that, a series of sound clips will be played to test your ability to use the equipment.

Clips of sound will be played through the headphones. The clip is divided into **2 parts**. For **each part** of the clip, you have to decide i) it was coming from **INSIDE** or **OUTSIDE** or "**on the skull**" of your head and ii) the **sound source direction**.

Please refer to the Figure 1 about the area of inside and outside of your head. Please say "**INSIDE**" or "**OUTSIDE**" or "**on the skull**" to indicate your choice.

To indicate the sound source direction, a sphere and a sensor will be given to you. Please refer to Figure 2 and 3 for instructions. This measuring equipment is for measuring the sound source that you perceived. Please point out the sound source direction that you think to be with the sensor.

If you have pointed out the direction, please say "**READY**" and hold the sensor until the experimenter says "**OKAY**". Then, next sound clip will be played and the procedures will be repeated. There are trials for you to familiarize with the equipments.

IMPORTANT:

Please **close your eyes** and **never move your head** while listening to the sound clip.

Appendix G Background of listeners participated in experiment 1 to 3

Table G.1 Background of listeners participated in experiment 1 to 3

Listener	Gender	Age	School
S4	F	21	Business
S5	M	19	Business
S6	M	23	Engineering
S7	M	20	Science
S8	M	20	Engineering
S9	F	20	Business
S10	F	21	Business
S11	F	22	Business
S12	F	24	Engineering
S13	M	20	Science
S14	M	25	Engineering
S15	F	21	Engineering
S16	F	22	Engineering
S17	F	21	Business
S18	F	21	Business

Appendix H Demonstration matlab functions for filtering sound with HRTFs

The CD enclosed includes two matlab functions for filtering sound with HRTFs used in experiment 1 to 3. After filtering, the sound carries spatial information. Under the folder "demo program\demo1", there is a matlab function called conv_demo1. There are 3 parameters in this function. Parameter wavfile is the file name of a mono wave file which is going to be filtered. Parameter direction indicates the direction of the HRTF. "1" indicates front (0°,0°) and "2" indicates back (180°,0°). Parameter HRTF_num indicates the standard HRTF number. The range of it is from 1 to 7.

Under the folder "demo program\demo2", there is a matlab function called conv_demo2. There are 2 parameters in this function. Parameter wavfile is the file name of a mono wave file which is going to be filtered. Parameter HRTF_num indicates the standard HRTF number. The range of it is from 1 to 196.

The output stereo wave file is saved as output.wav.