Optimizing the performance of Virtual Headphone-based Surround Sound (VHSS) Systems by manipulating the spectra of non-individualized HRTFs

by

LEUNG NGAN MING

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 Engineering Management

December 2001, 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.

LEUNG NGAN MING

Optimizing the performance of Virtual Headphone-based Surround Sound (VHSS) Systems by manipulating the spectra of non-individualized HRTFs

by

LEUNG NGAN MING

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

PROF. CHUNG YEE LEE

INDUSTRIAL ENGINEERING AND ENGINEERING MANAGEMENT

15 JANUARY 2002

Acknowledgements

I would like to express my grateful gratitude to my supervisor Dr. Richard H. Y. So for his long lasting support, guidance, advice and continuous encouragement during the whole research. I am also grateful to Professor Jens Blauert and his student Dr. Jonas Braasch for their valuable suggestions and guidance to my research.

Deep appreciation is expressed towards Dr. Ravindra S. Goonetilleke, Dr. Heloisa H. O. M. Shih and Dr. Y. W. Wan for giving me advice during the defense examination and human factor seminars.

Thanks are extended to Denil Chan, Stanley Sze, Fred Chan and Darwin So for helping me to solve the technical problems in the prototype implementation and experimental setup.

I would like to thank Emily Au for helping me in conducting the double-blinded experiment. Sincere gratefulness to Edith Ng, Ling and YSL for their discussion and help on the experimental design and statistical analysis. Additional thanks are also extended to my UST friends and staff in the department.

Finally, I would like to take this opportunity to thank my parents, sister, girlfriend Moon and my friends for their encouragement and support throughout my study.

TABLE OF CONTENTS

List of Figures

List of Tables

Optimizing the performance of Virtual Headphone-based Surround Sound (VHSS) Systems by manipulating the spectra of non-individualized HRTFs

by

LEUNG NGAN MING

INDUSTRIAL ENGINEERING AND ENGINEERING MANAGEMENT

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Abstract

Traditional Dolby^{TM} 5.1 surround sound systems need six speakers (center, left, right, left rear, right rear and subwoofer) to deliver the surround sound effects. A Virtual Headphone-based Surround Sound (VHSS) System processes the raw signals of Dolby^{TM} 5.1 standard to deliver simulated surround sound effects to listeners through a pair of headphones. The key components of a VHSS system are non-individualized Head-related Transfer Function (HRTF) filters that contain important information necessary to simulate directional sound cues. The purpose of this research is to optimize the complexity and performance of a VHSS system through studying the effects of simplifying and manipulating the spectra of HRTFs.

Three experiments were conducted to achieve the goal. Four levels of complexities of the HRTF filters (128, 64, 32 and 18 coefficients) were studied for their effects on sound localization errors. Results indicated that reducing the number of coefficients in HRTFs significantly affected the localization errors for all directions *($p<0.05$)* except directions 0°, 90° and 270° *($p>0.1$)*. An optimized set of filters with different complexities was obtained and implemented as a prototype VHSS system. A double-blinded usability experiment verified that the prototype VHSS system produced significantly better surround sound effects than the corresponding Dolby™ stereo channels *($p=0.028$)*. Nevertheless, front-back confusions were still the major cause for the degraded performance. In the last experiment, spectra of HRTFs, divided into six frequency bands, were enhanced at three different levels to obtain higher localization accuracy. Results showed that 12dB and 18dB enhancements of five of the six frequency bands have significantly reduced localization errors for directions 0°, 45°, 135°, 180°, 225° than conditions without any enhancement *($p<0.05$)*. Actual or potential applications include the design of a cheaper and better VHSS system.

CHAPTER 1 INTRODUCTION

1.1 General introduction to simulation of binaural directional sound cues under headphone condition

In real life, humans localize directions of sounds depending on three important cues: inter-aural time different (ITD), inter-aural level difference (ILD) and spectral cues. Inter-aural time difference (ITD) is the cue due to the different paths of the sound source going to the left ear and right ear. Inter-aural level difference (ILD) is formed because of the shadowing effect of head on the sound source to the far ear. Spectral cues result from the filtering effects of specific shape of both ears. With sufficient information of these three cues, humans are able to localize the directions of sounds in daily lives. To simulate binaural directional sound sources in headphone condition, Head-related Transfer Functions (HRTFs) play an important part in this area. They are the frequency responses between sounds that originate from different positions and their corresponding destinations of sounds measured between the ear canals and the pinnae. In other words, they are digital filters consisting of those important localization cues as mentioned before. The traditional method to obtain the HRTFs of an individual or a mannequin head is to play a sound signal at a particular position at a certain distance of more than one meter and measure the impulse response with probe microphones mounted near the opening end of the car canals as shown in Figure 1.1. In order to achieve accurate HRTFs, an anechoic chamber is the most suitable environment for the measurement (Figure 1.2).

FIGURE 1.1 A probe microphone positioned near the opening end of the ear canal (adopted from Begault, 1994).

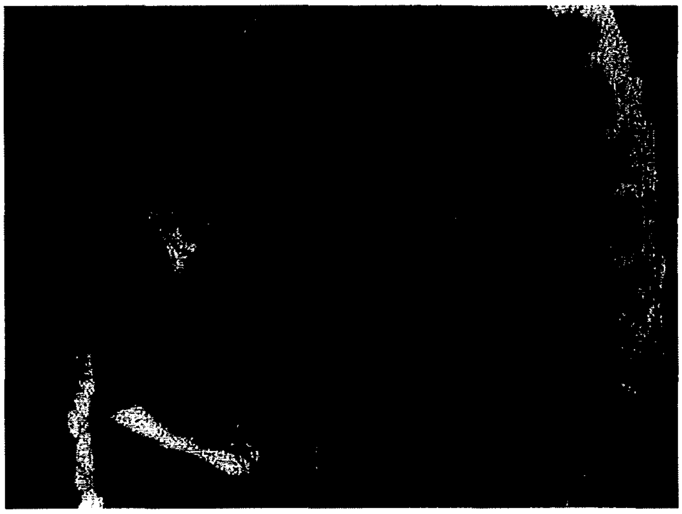


FIGURE 1.2 A photograph to illustrate the measurement of individualized HRTFs in an anechoic chamber (adopted from Begault, 1994).

Figure 1.3 shows a measured HRTF filter of 45° right direction to a listener's face (i.e. 0 degree, directly ahead of the listener) for right ear. To synthesize a binaural sound cue, convolution process as shown in Figure 1.4 will take place between an input mono sound and a pair of HRTF filters.

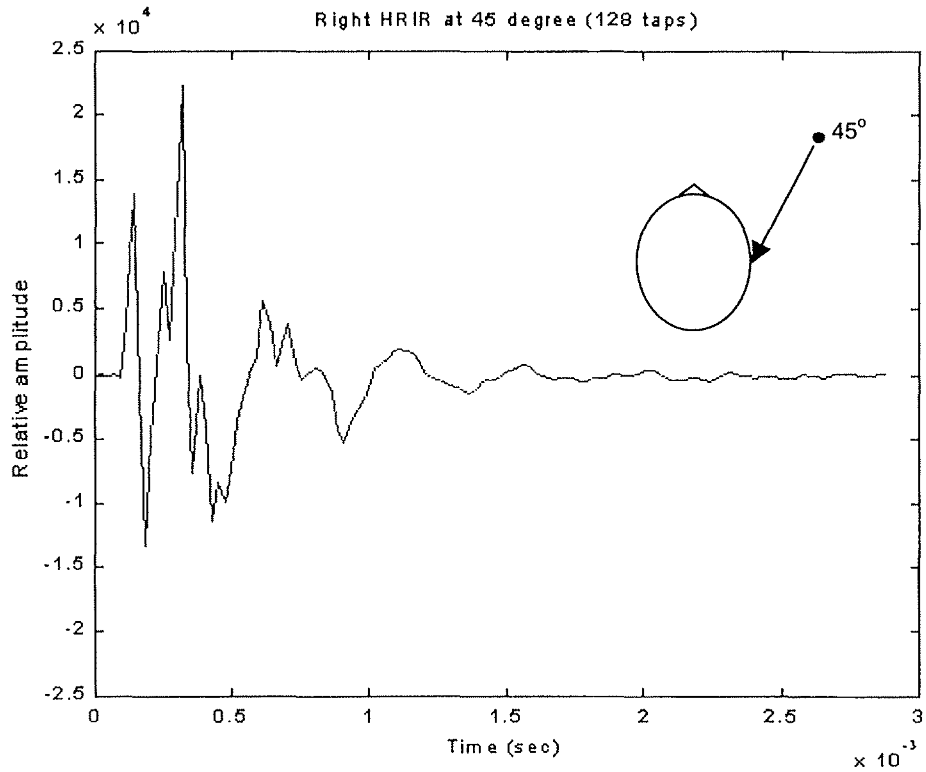


FIGURE 1.3 A time domain HRTF filter (45° direction, right ear, 128 coefficients) from Gardner and Martin (1995) measured in the ear canal in an anechoic chamber.

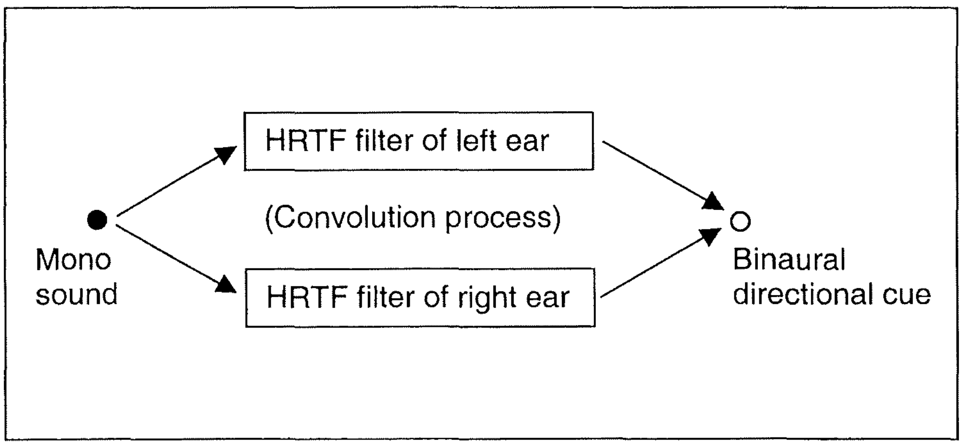


FIGURE 1.4 Convolution process between a mono sound and a pair of HRTF filters.

It was reported that individualized HRTFs could provide accurate localization cues to the humans when listening through the filtered sound signals under headphone condition (Wightman and Kistler, 1989b; Kulkarni and Colburn, 1998). However, the measurements involve tedious and time-consuming procedures requiring the person to remain motionless. It is also not feasible to collect the HRTFs from the public population when this technology is used in commercial applications. To cope with this difficulty, using non-individualized HRTFs of a mannequin head or an averaged group of individualized HRTFs are proposed.

1.2 Applications of simulated binaural directional sound cues

There are many uses for this technique in different application areas from a range of military actions to entertainment purposes. The main objective of using this technique to simulate binaural sound cues is to provide the listeners a correct perception for the locations of those directional sounds. Depending on the purposes used in different areas, either individualized or non-individualized HRTFs are used.

In military areas, simulating binaural directional cues that delivered to the pilots can allow them to localize the directions of sounds correctly in real time. The Traffic Alert and Collision Avoidance System (TCAS II) implemented this idea to simulate the sounds of other missiles or planes flying towards or behind the plane of a pilot, enhancing awareness of the external threat by synthesizing sound cues and allowing the pilots to escape from the danger (Begault, 1993b). As a very high degree of accuracy of sound directions should be required in simulating the sound cues, the HRTFs used should be measured from every pilot (i.e. individualized HRTFs) in this case. By using a similar idea, submarine operators can also locate the positions of any obstacles or other vessels by such kind of sound cues generated which can allow them to response external incidence more quickly.

There have been some applications of simulated sound for allowing vision-impaired people to handle in an environment of Graphical User Interface (GUI) (Burgess, 1992; Crispien and Petrie, 1993). Auditory icons were used to communicate with the Microsoft® Windows™ operating system, responding to different GUI interactions as described in Table 1.1. According to the position shown on the screen, the sound would be synthesized and presented to the blind people. Either individualized or non-individualized HRTFs were used depending on the degree of accuracy for the simulated sound cues.

TABLE 1.1 Some spatialized auditory icons utilized in the Graphical User Interface for the Blind (GUIB) system of Crispien and Petrie (1993) (adopted from Begault, 1994).

Interactions	**Auditory icons**
mouse tracking	steps walking
window pop-up	door opening
window moving	scratching
window sizing	elastic band
buttons	switches
menu pop-up	window shade

This simulation technique can also be used in the Virtual Headphone-based Surround Sound (VHSS) System. The main idea is to simulate the six decoded channels from a AC3 data stream according to the Dolby™ 5.1 standard (center, left, right, left rear, right rear and subwoofer) in a Digital Signal Processor (DSP) to generate a virtual surround sound effect, resulting in a two-channel left and right output. It is guaranteed that an excellent spatial sound effect can be obtained if individualized HRTFs are used. However, it is not possible to achieve this as it is difficult to collect HRTFs from each individual. From the standpoint of commercial budget, usually non-individualized HRTFs are used.

1.3 Problems associated with non-individualized simulated binaural directional sound cues and potential solutions

Although non-individualized HRTFs are widely used in commercial applications, the degraded localization cues lead to significant occurrence of front-back confusions. This has been proved by Laws (1974), Wightman and Kistler (1989b), Asano *et al.* (1990), Wenzel *et al.* (1993) and Begault and Wenzel (1993). This means that people localize the sound presented at 45° direction right to them as if coming from its mirror angle -135° direction behind them as shown in Figure 1.5. Also, there was a trend for more front-back confusions than back-front confusions. (Wightman and Kistler, 1989b; Wenzel *et al.*, 1993 and Begault and Wenzel, 1993).

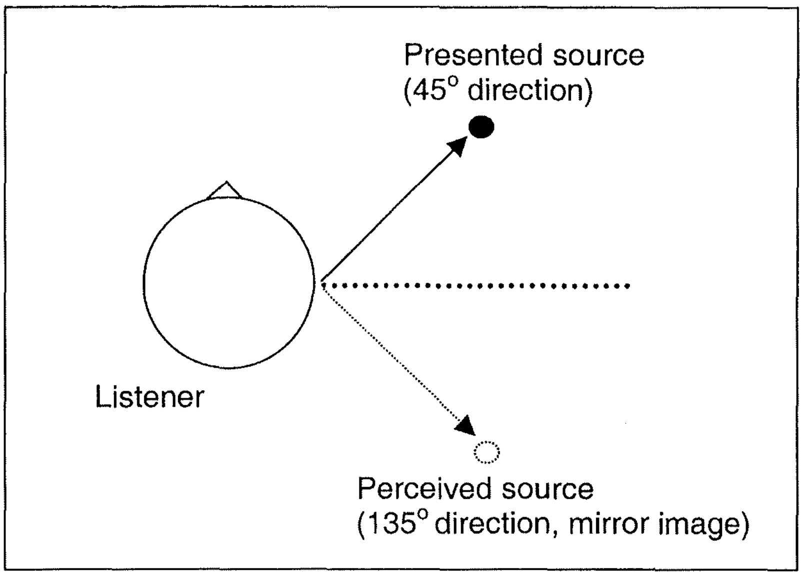


FIGURE 1.5 An illustration of the occurrence of front-back confusions.

Another problem for using non-individualized HRTFs is due to the large inter-subject differences. Figure 1.6 plots out the HRTF measurements of direction 0° from different authors in past literature. It was easy to observe that the relative difference at about 8.0 kHz was nearly up to 30 dB (indicated by the arrow). This explained why the localization performance of those listeners was largely degraded when non-individualized HRTFs were used. This was mainly due to large individual differences among people.

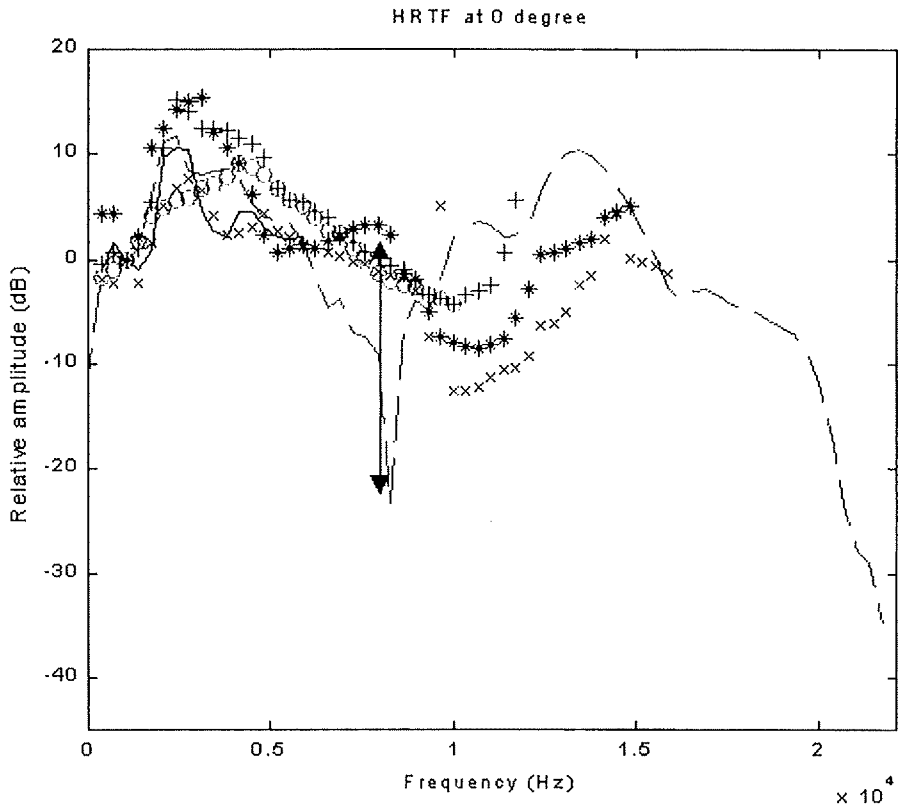


FIGURE 1.6 Modules of the HRTF form sound cues at 0° direction (i.e. directly ahead of the listener) from different authors ('-' Wiener, 1947; 'o' Robinson and Whittle, 1960; '.' Blauert, 1969a; '+' Shaw, 1974; '*' Mehergardt and Mellert, 1977; and '--' Gardner and Martin, 1995).

There have been many solutions proposed to solve the previously mentioned problems. One solution is to develop a low-cost and fast technique to measure individualized HRTFs. The Crystal River Engineering (CRE) has invented a portable HRTF measuring system called SNAPSHOT which can measure the HRTFs near the opening of ear canal, different from the traditional approaches measuring HRTFs deep in the ear canal. Wightman and Kistler (1995) compared the HRTFs obtained from SNAPSHOT and traditional methods. They stated that although there were some differences between the HRTFs measured by these two methods, SNAPSHOT preserved satisfactory characteristics of the measured HRTFs.

Another approach to solve the degraded performance of non-individualized HRTFs is to enhance those important cues found in the HRTFs. It was found that there are frontal and backward cues found in the HRTFs which lead listeners to disambiguate sounds coming from the front or at the back respectively. Correct promotion of those cues can ensure that the localization performance for non-individualized HRTFs can be improved. This research focuses on this direction to solve the unpleasant problems found when using non-individualized HRTFs.

1.4 Purpose of this research

This research aimed to study individual performance in localization performance as functions of spectral cues. Knowledge obtained will be used towards the development of a tuning function in the VHSS system. The specific objectives are:

i) to investigate how the spectral cues will affect the localization performance of binaural sound cues simulated by simplifying the spectra of non-individualized HRTF filters;

ii) to generate an optimized set of HRTF filters for the Virtual Headphone-based Surround Sound (VHSS) System;

iii) to study the localization performance of binaural sound cues by manipulating the frontal and backward cues in different frequency bands of non-individualized HRTFs; and

iv) to investigate if there is any difference in localization performance using participants with or without experienced musical training background.

1.5 Organization of the thesis

This thesis starts by introducing the principle of sound localization by human and simulation of binaural directional cues under headphone condition. The problems associated with non-individualized HRTFs are also mentioned briefly here. The second chapter contains the literature review on how humans localize sound and the use of individualized and non-individualized HRTFs to synthesize binaural directional sound cues. The review includes details on the three main localization cues (inter-aural time difference (ITD), inter-aural level difference (ILD) and spectral cues) and their relevant studies. In addition, reviews on simplifying the spectra of non-individualized HRTFs and manipulating the identified frontal and backward cues at different frequencies are presented. Lessons learnt from the literature review is documented in chapter three. Chapter four documents technical details related to the experiments and prototype implementations used in this study. In chapter five, Experiment one investigating the effects of simplifying the spectra of non-individualized HRTFs on their ability to produce accurate binaural directional cues is presented. Chapter six describes details of the implemented Virtual Headphone-based Surround Sound (VHSS) System and the usability experiment (i.e. Experiment two) on its performance using optimized set of HRTF filters based on results of the first experiment. Chapter seven reports experimental analysis of Experiment three that was conducted to study the use of manipulating the spectra of non-individualized HRTFs at different frequencies to further improve the localization performance associated with binaural directional cues generated with non-individualized HRTFs. Finally, general conclusions, discussion and recommendations will be mentioned in the last chapter.

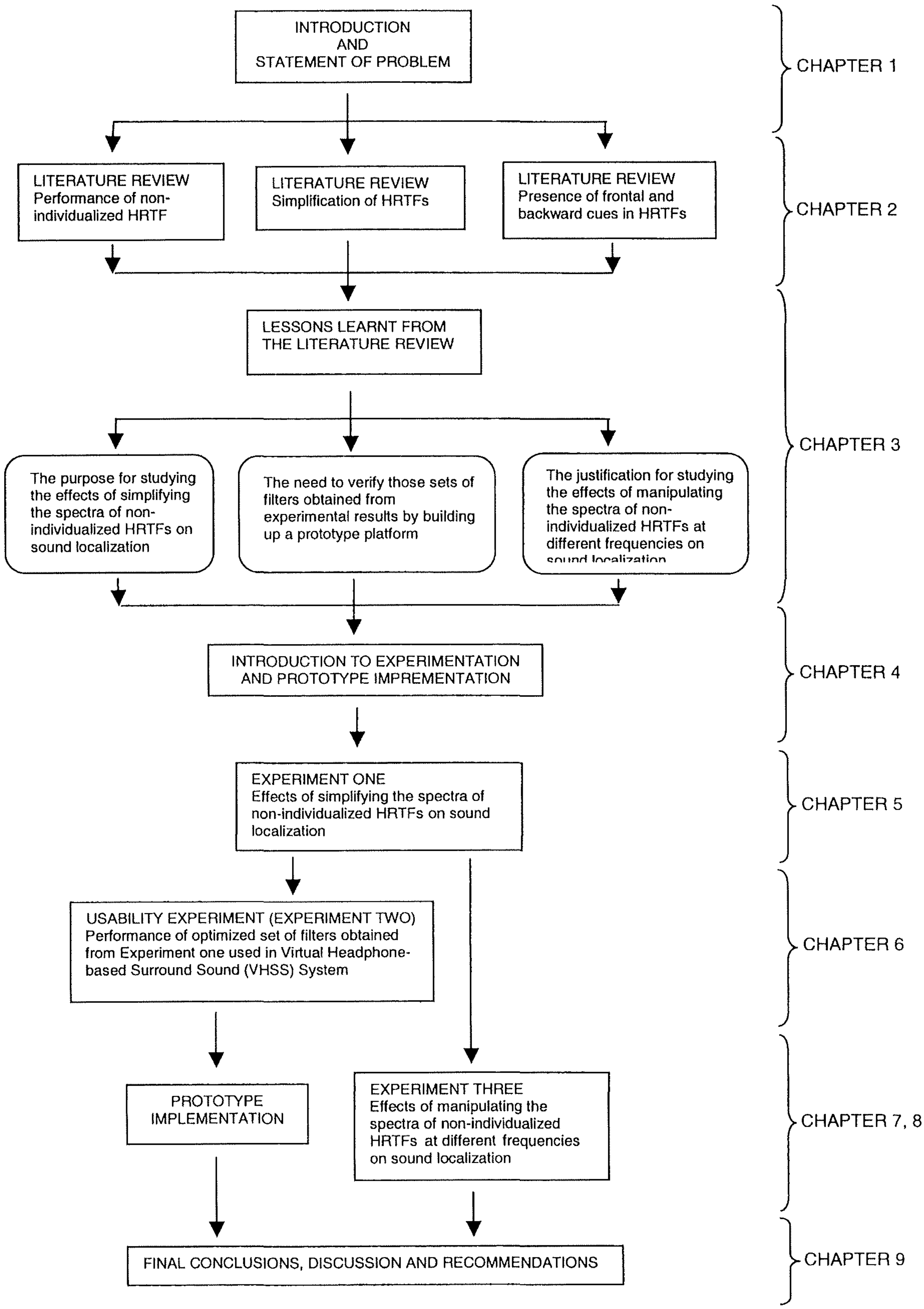


FIGURE 1.7 Outline of the thesis.

CHAPTER 2 LITERATURE REVIEW

2.1 How humans localize sound

Sound localization is of vital importance to humans. It allows them to judge the directions and distance of a sound source. Some important localization cues are needed for humans to localize sounds and can be divided into two types: (i) Binaural cues; and (ii) Monaural cues.

2.1.1 Binaural cues

About a century ago, the whole basis of sound localization was based on the Duplex theory purposed by Rayleigh (1907). He stated that sound localization depended on the inter-aural difference cues. As two ears are positioned on both sides of a relatively big head, when a sound source is presented from one side, there will be some differences in the path length from the sound source to both ears (Figure 2.1). This leads to a difference in arrival time to both ears. This difference is known as inter-aural time difference (ITD). The ITD initiated by the path length differences for the incident sound wave is calculated by this formula:

$$D = r * (\varnothing + \sin \varnothing),$$

where D = distance (m), r = radius of head size (m) and Ø = angle of sound wave from median plane (rad). The time difference (t) will be

$$t = D/c,$$

where c = speed of sound (ms^{-1}).

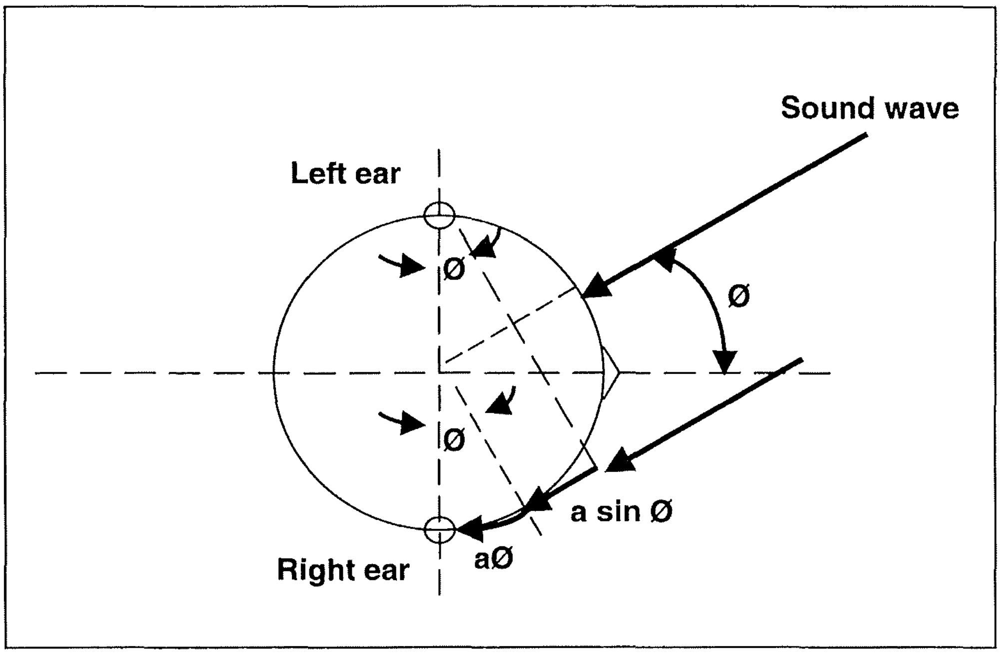


FIGURE 2.1 Calculation of inter-aural time difference (ITD) (adopted from Carlile, 1996).

The head of a listener tends to interrupt the path from the sound source to the far ear. The far ear will be effectively shadowed and this difference will be known as inter-aural level difference (ILD). The degree of shadowing is determined by the wavelength of the sound source compared with the physical dimensions of a head (Middlebrooks and Green, 1991; Carlile, 1996). Higher frequencies are attenuated more at the far ear because of this shadowing effect. It was revealed that ITD has significant effect at low frequencies (i.e. less than around 1.5 kHz), but not at high frequencies (i.e. larger than around 1.5 kHz) while ILD is more dominant at high frequencies than at low frequencies. This has been proved by Wightman and Kistler (1992) that they have produced some stimuli in which ITD cues signaled one particular direction while ITD belonged to another direction. Nine participants have been involved in this experiment. The result reported that the perceived directions of these cues which were conflicted with one another always followed the ITD if the stimuli consisted of low frequencies. The effect of ITD would be eliminated if components of low frequencies were removed from the stimuli; instead, ILD would become dominant in this case.

The advantage of using ITD and ILD as the only localization cues for a sound source is the ease of calculation for these values. However, Mills (1972) stated that the Duplex theory could not fully explain the cues for humans to localize a sound source because of the roughly spherical shape of a head leading to the phenomenon of "cone of confusion" as shown in Figure 2.2.

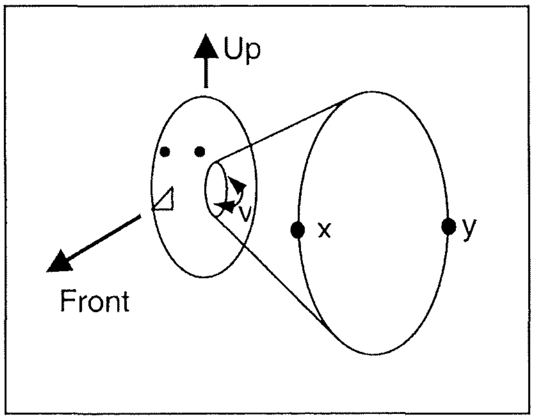


FIGURE 2.2 An illustration of a cone of confusion. Two opposite positions on any points of the conical surface have identical values of ITD and ILD. Points x and y show a front-back confusion. The maximum vertex angle v of the cone can extend to 180°, i.e. a pair of sound source position at 0° (directly ahead of the listener) and 180° (directly at the back) (modified from Carlile, 1996).

It was known that the sound source located on the conical surface of this cone of confusion has the same values of ITD and ILD. This makes people become unable to distinguish sound sources coming from the front and back or from the overhead and below where ITD and ILD cannot provide important information to localize the sound.

2.1.2 Monaural cues

Spectral cues help people to distinguish the sounds located on the cone of confusion by the filtering effects of the outer ears. The asymmetrical pinna shape causes unique time delays, resonances and diffractions, resulting in the formation of unique set of HRTFs for each sound source position. As stated by Blauert (1983), high frequencies (e.g. 10.0 kHz) would like to be more attenuated for sources positioned at the back than those situated in the front. Peaks and notches found in the spectrum of HRTFs could be considered as important cues to discriminate the sound directions coming from the front or at the back. (Bloom, 1977; Watkins, 1978; Blauert, 1983). However, when comparing to the calculation of ITD and ILD, it is not easy to obtain the spectral cues as it involves tedious and time consuming processes.

It was found that degraded spectral cues would result in decreased localization performance. Freedman and Fisher (1968) investigated the role of pinnae in sound localization. They have conducted the experiments to ask 35 participants to determine the direction of white noise under three different pinnae conditions: (i) own pinnae; (ii) no pinnae; and (iii) artificial pinnae (casts of human pinnae were placed on the ends of the tubes). With the head kept remained motionless, there was no significant difference between conditions (i) and (iii), which in turns, had significantly better localization performance than that in condition (ii). Nevertheless, there was no significant difference among all three conditions if head movements were allowed. All participants showed perfect performance on judging the directions. These findings also suggested that humans might be able to adapt other people's pinnae even they differed among individuals.

2.2 The simulation of binaural directional sound cues

2.2.1 Introduction

As mentioned briefly previously, Head-related Transfer Functions (HRTFs) are digital filters with coefficients which consist of those three important cues (ITD, ILD and spectral cues) that are used to simulate binaural directional sound cues. These transfer functions are measured from free-field to the eardrum at different source positions. It can preserve the inter-aural differences over the whole frequency ranges of the stimuli and capture the effects of filtering by the pinnae. Two types of HRTFs are categorized: individualized and non-individualized HRTFs.

2.2.2 Individualized HRTFs

2.2.2.1 Measurement of individualized HRTFs

Individualized HRTFs mean that the set of HRTF filters are recorded from a particular listener specifically preserving all the information about the inter-aural differences and filtering effects of his outer ears. The general approach to produce a set of HRTFs is based on the technique described by Wightman and Kistler (1989a). Figure 2.3 shows a probe microphone system consisting of a miniature microphone linked with a silicone rubber probe tube having less than 1 mm outer diameter.

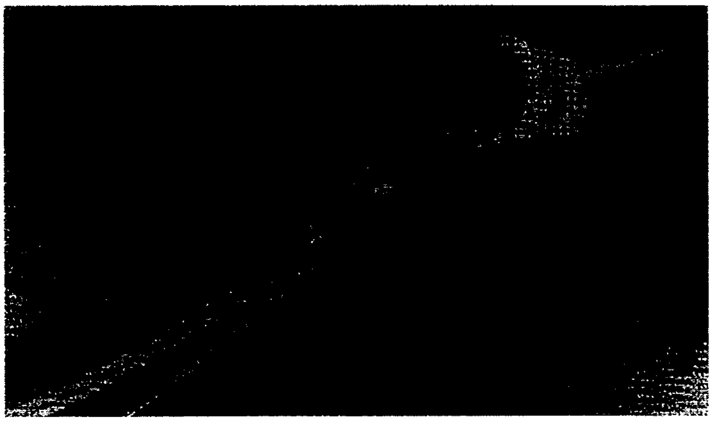


FIGURE 2.3 A probe microphone used for HRTF measurements in the study of Wightman and Kistler (1989a) (adopted from Begault, 1994).

As stated in the study of Wightman and Kistler (1989a), the probe microphone system had a high sensitivity and relative flat frequency response from 0.2 to 14.0 kHz. Two probe microphones were used in order to measure the acoustical responses for both ears simultaneously. One end of probe tubes was placed near the middle of the listener's ear canal at about 1 – 2 mm from the eardrum. This chosen position ensured that all direction-dependent effects could be captured. The outputs from microphones would be amplified and digitalized using 16 bit A/D converters manipulated by a personal computer. To avoid making any disturbance in the ear canals, the probe tubes were fixed with customized earmold shells. They were inserted into a thin, semi-rigid guide tube that mounted to the wall of the earmold shell. With this thorough preparation, the HRTFs could be measured without any unpleasant disturbance.

2.2.2.2 Individual differences

It was expected that individual anatomical features, namely head size, pinna breadth and depth, varied considerably among humans. Burkhard and Sachs (1975) recruited 12 males and 12 females in his measurement and concluded that the mean concha length and mean breadth was 2.63 cm and 1.8 cm respectively with standard deviations in both measurements of about 0.23 cm. As the resonances of the concha play an indispensable role in the filtering effects of the outer ears (Shaw and Teranishi, 1968), such kind of size variations would be considered as significantly high in this case. Middlebrooks and Green (1990) reported that head radius ranged from about 6.1 to 7.3 cm among six adults and the resultant maximum ITD lied between 645 to 750 μs. It was found that an ITD of 645 μs would present a sound cue at 90° direction lying in horizontal plane for those participants with smaller head whereas at about 75° direction for other participants. Past studies also showed that listening through other ears resulted in the changes in the perceived localization of sounds (Wenzel *et al.* 1988, 1993). Occurrence of front-back confusions was reported to be increased.

Using individualized HRTFs on a specific listener to synthesize binaural sound sources results in excellent localization performance as the degradation of spectral cues is largely minimized. Nevertheless, it may not be feasible to measure the HRTFs for each listener as it involves tedious procedures to achieve it, leading to considerable amount of time and money.

2.2.3 Non-individualized HRTFs

2.2.3.1 Measurement of non-individualized HRTFs

There are two methods for obtaining non-individualized HRTFs: one is measured from a mannequin head manufactured based on averaged HRTFs from a general population. Another method is to use the HRTFs of a particular person for simulating the binaural sound cues for other people, a commonly used practice in previous studies. For instance, authors in past studies used the non-individualized HRTFs from a representative person who showed good localization accuracy in azimuth directions (Wightman and Kistler, 1989b; Begault, 1992a; Begault and Wenzel, 1993; Begault, 1995; Wenzel *et al.*, 1993).

2.2.3.2 The problem of front-back confusions

To overcome the difficulty in collecting individualized HRTFs from general public, the use of non-individualized HRTFs to simulate binaural sound cues was proposed. It can avoid the waste of time and money spent for measuring individualized HRTFs. However, measured HRTFs vary between different participants mainly due to the differences of pinna shapes and head sizes. This resulted in increased occurrence of front-back confusions. There have been many studies on the comparison of localization performance of stimuli using individualized and non-individualized HRTFs. Wenzel *et al.* (1988b) investigated the localization performance of sources in free-field with that simulated from individualized and non-individualized HRTFs. For sources in free-field and stimuli synthesized from their own HRTFs, those participants acted as "good localizers" showing accurate localization performance and less front-back confusions. However, when using non-individualized HRTFs, there would be larger localization errors and more front-back confusions. Wightman and Kistler (1989b) recruited eight participants and compared the perceived positions of sounds presented in free-field and under headphone condition simulated by a given subject's HRTFs measured in the study (Wightman and Kistler, 1989a). Results indicated that the performance of azimuth localization was satisfactory for both free-field and headphone simulation. However, the occurrence of front-back confusions was still found to increase for headphone simulation.

In 1993, Wenzel *et al.* synthesized the stimuli using non-individualized HRTFs of the representative subject (a good localizer in both free-field and headphone conditions) based on the study by Wightman and Kistler (1989b) and compared the localization performance with their free-field condition. Results revealed that all 16 participants showed accurate performance in azimuth for both free-field and headphone conditions. However, a high rate of front-back confusions could be found significantly between the free-field and headphone conditions for nearly all 16 participants. This confirmed that using non-individualized HRTFs would lead to the problem of increased front-back confusions. Asano *et al.* (1990) has measured the individualized HRTFs for two participants. It was found that when simulating virtual sources with another set of HRTFs, the rate for front-back confusions would be higher.

2.2.3.3 The problem of elevation error

Apart from experiencing the front-back confusions, Wightman and Kistler (1989b) also observed larger elevation errors from the eight participants when using non-individualized HRTFs in simulating the stimuli in their study. In the study of Wenzel *et al.* (1993), they simulated the sound source using non-individualized HRTFs of the representative subject based on the study by Wightman and Kistler (1989b) as mentioned before and compared the localization performance with their free-field condition. Results indicated that 12 of 16 participants had good elevation judgment in both cases. Two of 16 participants showed poor elevation accuracy in both cases while another two participants had good elevation performance in free-field condition but not for the case of headphone condition. Middlebrooks (1999b) tried to compare the localization accuracy for stimuli synthesized by own-ears' HRTFs, other-ears' HRTFs and scaled-ears' HRTFs. Eighteen listeners were participated in this experiment. It was found that the elevation errors increased when participants listening to other-ears' HRTFs. Moller, Sorensen, Jensen and Hammershoi (1996) have compared the localization performance on sound sources in the following conditions: (i) free-field; (ii) individualized HRTFs; (iii) non-individualized HRTFs and (iv) mixed individualized and non-individualized HRTFs. Results of the eight participants revealed that stimulating with non-individualized HRTF performed poorer in localizing sounds than those with own-ears' HRTFs.

2.2.3.4 The problem of inside-the-head locatedness (IHL)

Begault and Wenzel (1993) used non-individualized HRTFs from the representative subject in the study of Wightman and Kistler (1989b) to synthesize the stimuli which were presented to 11 participants to judge the perceived positions. Results showed that the range of the responses heard inside the head (i.e. inside-the-head locatedness, IHL) was about 15% to 46%. One possible reason for the lack of externalization was the absence of environmental cues such as reverberation in the original HRTF measurements. Begault (1992a) used five participants to compare the percentage of having IHL between anechoic and reverberated 3-D processed speech stimuli using non-individualized HRTFs of the representative subject in the study of Wightman and Kistler (1989b); and found that the overall percentage of IHL occurrence was reduced from 25% to about 3% when reverberation was added to simulate the headphone stimuli.

Durlach and Colburn (1978) have mentioned that the externalization of a sound source was difficult to be predicted unless binaural HRTFs, head movement and reverberation were generated as close as natural condition under headphone stimulation. Also, familiar sound sources were found to produce more accurate distance estimates than unfamiliar sources (Coleman, 1962; McGregor, Horn and Todd, 1985). In addition, natural reverberation or artificial generated reverberation could result in enhanced externalization of virtual sound under headphone (Plenge, 1974; Toole, 1970; Sakamoto, Gotoh and Kimura; 1976; Begault, 1992a).

2.2.3.5 Past studies on improving localization performance of sound cues simulated using non-individualized HRTFs

There have been some studies which tried to improve the performance of non-individualized HRTFs by scaling the transfer functions in frequency (Middlebrooks, 1999a; 1999b). As there were spectral differences among the 45 participants, they tried to find out an optimal frequency scaling factor that aligned the spectral features between a pair of participants and minimized the spectral differences between them. Results indicated that the spectral differences after using this frequency scaling could be reduced by about

15.5% among all pairs of participants. This optimal scale factor between each participant pair had a high correlation with the ratios of participants' maximum inter-aural delays, their external ear sizes and head widths. In order to verify the effect of optimal scaling factor, Middlebrooks conducted an experiment to study the localization performance of virtual sources simulated by HRTFs under three conditions: (i) listeners' own ears; (ii) other participants' ears; and (iii) scaled ear (using scaling frequency factor to minimize the differences between own-ears' and other-ears' HRTFs). Eighteen participants were involved. He discovered that localization performance including the occurrence of front-back confusions could be improved using the scaled-ear condition when compared with that of using ears of other participants.

2.3 Simplification of HRTF filters

Asano *et al.* (1990) have simplified the spectral cues and tested the localization performance of two participants on the sound cues using these simplified filters. They have used four different orders of the auto-regressive moving-average (ARMA) model to simply the spectral cues. Eighteen directions ranging from 0° to 180° in top hemisphere of median plane were tested. A noticeable increase of front-back error was observed when the degree of simplification increased. Kulkarni and Colburn (1998) focused on the sensitivity of participants to the details of individualized HRTFs. They asked four participants to distinguish real sources from loudspeakers and virtual sources from tube-phone attached to the openings of ear canals of them. Four azimuth directions (0°, 45°, 135° and 180°) were used and ITD has been kept to be the same when reducing the number of coefficients (512, 256, 128, 64, 32, 16 and 8 coefficients). Results reported that they were unable to discriminate the real and virtual sources when coefficients of filter having more than 16 coefficients. Begault (1992b) evaluated the localization performance on their measured HRTFs of 512 coefficients with its simplified version of 65 coefficients among 13 participants. No statistically significant difference was found in localization accuracy between these two types of HRTFs for all directions (30° interval from 0° to 360°). This showed that using a lower coefficient number in the filters resulted in similar localization performance when compared with a longer one.

2.4 Presence of frontal and backward cues in HRTF filters

There have been some past studies reporting which frequency bands in the spectral cues can provide the frontal and backward cues for humans to localize sound sources coming from the front or at the back. It is believed that if a thorough understanding of the importance of these frequency bands in spectral cues can be achieved, listeners are able to distinguish the sounds coming from the front or at the back by boosting or attenuating those frequency bands. However, it seems that there are still some disagreements between past studies.

Bronkhorst (1995) tried to investigate the localization performance for the real and virtual sources. Totally eight participants were asked to point the directions of the stimuli by their head which could be detected by Polhemus tracker on top of their heads. It was found that front-back confusions became nearly the same when the low-pass cutoff frequency shifted from 7.0 kHz to 15.0 kHz. He has analyzed the spectral cues from all the participants and indicated that a front-back cue was situated in the frequency band around 4.0 kHz, however, few cues could be found in the frequency band above 9.0 kHz.

Musicant and Butler (1983) studied the role of pinna-based spectral cues on sound localization. They have tried to see whether there was a significant effect on the localization accuracy if one of participant's ears was occluded for stimuli consisted of different frequency ranges. Eight participants were recruited to conduct this experiment. Their results revealed that high frequencies contained in the stimuli (i.e. broadband and 4.0 kHz high-pass noise) were necessary for a better localization performance than those with only low frequencies (i.e. 1.0 kHz and 4.0 kHz low-pass noise). However, when the stimulus didn't contain frequencies above 4.0 kHz, it was found that the frequencies between 1.0 and 4.0 kHz also act as a cue to prevent front-back confusions. This showed that frequency bands above 1.0 kHz contained important cues for sound localization.

Hebrank and Wright (1974b) applied different kinds of filters listed in Table 2.1 to the white noise that presented to the 30 participants for localization through loudspeakers. It was reported that the frequencies above 16.0 kHz or below 3.8 kHz didn't affect the localization performance. Frontal cues were present as a notch between 4.0 and 8.0 kHz and as a peak above 13.0 kHz. On the other hand, a backward cue should be located as a small peak near 12.0 kHz.

TABLE 2.1 Types of filters and frequencies used in the study of Hebrank and Wright (1974b).

Type of filter used	Frequency used (kHz)
Low-pass cutoff filter	3.9, 6.0, 8.0, 10.3, 12.0, 14.5, 16.0
High-pass cutoff filter	3.8, 5.8, 7.5, 10.0, 13.2, 15.3
Band-pass filter	4.0, 4.5, 5.1, 5.7, 6.4, 7.2, 8.1, 9.1, 10.2, 11.4, 12.8, 14.5
Band-stop filter	6.2, 6.6, 7.0, 7.4, 7.9, 8.8, 9.8, 10.8, 12.0, 13.3, 15.0, 17.8

Blauert (1969) investigated the features of spectral cues that were responsible for the perception of sound sources. Loudspeakers were placed at the front, rear, left, right and overhead of the participants and pulses were presented to the 20 participants who localized the directions of sources. He defined "directional band" as "more than half of participants giving one answer in certain frequency bands more often than they gave two other possible answer combined" and "boosted band" as "average higher sound pressure level at eardrum in some frequency bands for sound coming from the front than that at the back". It was found that both distributions of directional band and boosted band along the whole frequency range agreed with each other. According to the results, frontal cues could be found in frequency bands approximately in between 0 – 0.75 kHz with a peak at around 0.4 kHz, 2.0 – 6.3 kHz with a peak at around 4.0 kHz and 13.0 – 16.0 kHz with a peak at around 16.0 kHz. At the same time, frequency bands between 0.75 – 2.0 kHz (i.e. a notch at around 1.1 kHz) and 10.0 – 12.0 kHz (i.e. a peak at around 12.0 kHz) consist of backward cues.

With the information on which frequency bands can provide the frontal and backward cues, Myers (1989) tried to boost and attenuate certain frequency bands in order to promote the localization performance on his three-dimensional auditory display. They have divided the whole frequency band into four regions: A (200-682 Hz), B (682-2069 Hz), C (2069-6279 Hz) and D (6279-22050 Hz). Boosting the frequency bands A and C while simultaneously attenuating bands B and D could give out a front cue to the signal. Conversely, a backward cue could be delivered to the signal by boosting the frequency bands B and D while simultaneously attenuating bands A and C. Following the work of Myers, Tan and Gan (1998) would like to customize the performance of the 3D sound system using non-individualized HRTFs. First, they let their participants choose a suitable set of HRTFs for themselves. The selected HRTFs would pass through another filter called "biaser". He defined the whole frequency range of biaser into five main regions: A (225-680 Hz), B (680 Hz-2.0 kHz), C (2.0-6.3 kHz), D (6.3-10.9 kHz) and E (10.9-22.0 kHz). Frequency bands A, C and E were boosted while simultaneously attenuating bands B and D to give frontal perception. Opposite practice took place for backward perception. The biaser would convolve with the selected HRTFs, forming the resultant HRTFs for the participants. The tested signals were synthesized by this HRTFs and then listened by inexperienced participants to see if the ability of distinguishing the front and back direction could be enhanced. Results revealed that lower number of participants experienced front-back confusions if the HRTFs they selected previously have been passed through the biaser.

2.5 Anatomical explanation for frontal and backward cues in HRTFs

The formation of those peaks and notches in HRTFs is due to the resonances in the cavum conchae (Yamagushi and Sushi, 1956; Blauert, 1967; Shaw and Teranishi, 1968; Shaw, 1974a; Shaw, 1975; and Shaw 1980). An external ear structure was shown in Figure 2.4. Shaw and Teranishi conducted an experiment to investigate the sound processing inside the pinnae using two different types of model: a human-ear replica and real human ears. Figure 2.5 shows the experimental apparatus used for the measurements of transfer functions for the human-ear replica. A pinna made of rubber was attached to the base plate with an external ear canal inside it. Sound pressure in the ear canal and pinna could be measured by the probe microphone.

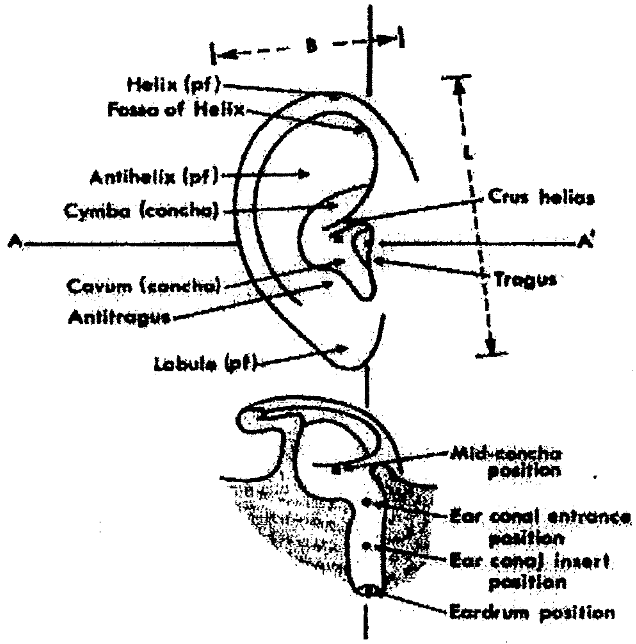


FIGURE 2.4 An illustration of different parts in an external ear (adopted from Shaw, 1975).

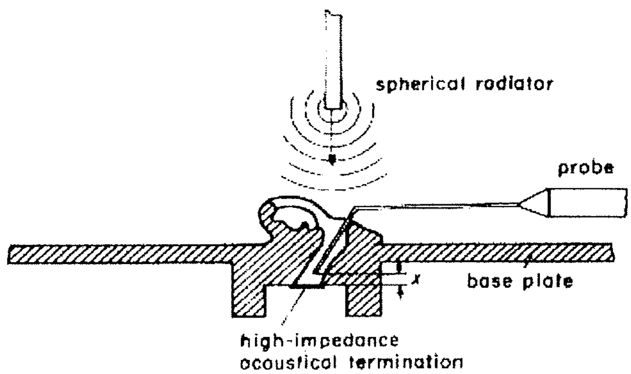


FIGURE 2.5 Experimental apparatus used in the study of Shaw and Teranishi (1968) (adopted from Blauert, 1983).

From this model, Shaw and Teranishi could find out a number of resonance frequencies ($f_{01} \sim 2.9$ kHz, $f_{02} \sim 5.5$ kHz, $f_{03} \sim 9.3$ kHz, $f_{04} \sim 11.2$ kHz, $f_{05} \sim 12.8$ kHz) as shown in Figure 2.6. It was observed that these resonance frequencies were closely matched with the frequencies where peaks and notches for frontal and backward cues located. The f_{01} resonance was due to the fundamental longitudinal resonance of ear canal while f_{02} resonance was caused by fundamental depth of the cavum conchae. For the resonances at higher frequencies, they were caused by the transverse modes of the cavum conchae leading to the high dependence on the angle of incidence. Six real natural ears have also been used in Shaw and Teranishi's experiment. The first two resonances f_{01} and f_{02} could be confirmed in real ears. For the resonances f_{03}, f_{04} and f_{05}, they were combined into a single peak with board increase in the transfer function plot of real ears.

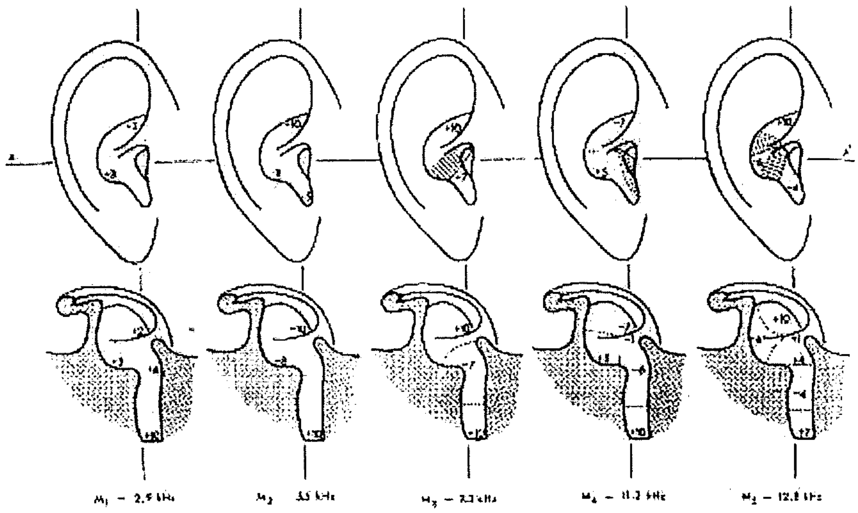


FIGURE 2.6 Several resonance frequencies found from Shaw and Teranishi's (1968) model (adopted from Shaw and Teranishi, 1968).

Yamagushi and Sushi (1956) and Blauert (1967) were found to confirm Shaw and Teranishi's result. Similar experimental apparatus in studies of Shaw and Teranishi have been used. Using the second resonance frequency f_{02} as an example, they claimed that the peak of the transfer function around 5 kHz was due to the cavum conchae, and a dip would be found in the same frequency if cavum conchae were filled up with putty.

CHAPTER 3 LESSONS LEARNT FROM THE LITERATURE REVIEW

3.1 Main findings of past studies

According to the literature review in the previous chapter, some studies have been done to simplify the complexities of HRTFs. Asano *et al.* (1990) reported that more front-back confusions were experienced by the participants when further reducing the number of coefficients used in the HRTF filters. Kulkarni and Colburn (1998) revealed that participants were unable to distinguish the sound source between real and virtual sources when the number of coefficients used in the individualized HRTF filters was reduced by up to 16 coefficients. Begault (1992b) found no significant difference in the localization accuracy for all the directions after reducing the number of coefficients to 65 coefficients. For manipulating the spectral cues in the HRTF filters, it was shown that there are frontal and backward cues positioned in different frequency ranges of the HRTF filters, leading to the indication for people to distinguish sound sources coming from the front or at the back. A peak should be found in frequency ranges 0 – 0.75 kHz with center frequency at around 0.4 kHz and 2.0 – 6.3 kHz with center frequency at around 4.0 kHz (Blauert, 1969; Myers, 1989; Bronkhorst, 1995; Tan and Gan, 1998) for frontal perception. Blauert's experiment also indicated that there should be a peak at frequency around 16.0 kHz. For the backward cues, a notch should be located in frequency range 0.75 – 2.0 kHz with center frequency at around 1.1 kHz (Blauert, 1969; Myers, 1989; Tan and Gan, 1998) while a peak was found in frequency range 10.0 – 12.0 kHz with center frequency at around 12.0 kHz (Blauert, 1969; Hebrank and Wright, 1974; Myers, 1989; Tan and Gan, 1998). However, Bronkhorst (1995) discovered that no important cues could be found for distinguishing sounds from the front or at the back at frequencies above 9.0 kHz. Musicant and Butler (1983) mentioned that frequencies less than 1.0 kHz did not contain any useful information for frontal and backward cues.

3.2 Problems with previous research

3.2.1 Lack of statistical support on achievement of using simplified HRTFs

According to the results of study in Asano *et al.* (1990), front-back confusions were found to increase greatly when reducing the complexities of spectral cues. For Kulkarni and Colburn's experiment, participants were unable to discriminate the real sound from loudspeakers and virtual sound from tube-phones until 16 coefficients were used. However, individualized HRTFs were used in this experiment. Also, no information on whether the localization accuracy and the occurrence of front-back confusions were affected by simplification of HRTF could be obtained from the experimental results. They only reported that whether participants were able to distinguish the sources of sounds between the loudspeakers and tube-phones. At the same time, there was no statistical analysis to support their results.

3.2.2 Lack of studies on further simplifying non-individualized HRTFs

In the study of Begault (1992b), they have reduced the number of coefficients used by up to 65 coefficients. They indicated that simplifying the spectra of HRTFs had no significant effect on localization performance for all directions between measured HRTFs of 512 coefficients and simplified version of 65 coefficients. A further investigation should be continued in order to study the localization performance of simplified HRTFs.

3.2.3 Lack of definite agreement on locations of frontal and backward cues in past literature

Although most of the studies gave us the impression that the locations of those frontal and backward cues were nearly the same, there was still some literature that disagreed with other results. Musicant and Butler (1983) revealed that those frequencies less than 1.0 kHz did not seem to carry any important cues for sound localization which contradicted with what other authors proposed. Also, Bronkhorst (1995) discovered that there should be no important cues for frequencies above 9.0 kHz. Hebrank and Wright (1974) mentioned that a notch situated between 4.0 and 8.0 kHz was attributed to the existence of frontal cues, while other studies claimed that it was the cues for elevation perception. On the other hand, results of Herbank and Wright (1974) seemed to indicate that spectral components below 4.0 kHz were less important. Myers (1989) defined those frequencies above 6.3 kHz consisting of backward cues only, however, other literature showed that there were some peaks or notches for frontal and backward cues above this frequency range.

3.2.4 Lack of evidence showing improved performance when enhancing frontal and backward cues

Tan and Gan (1998) have tried to test the sound localization performance of the participants when they listened to the sound cues synthesized from HRTFs with enhanced frontal and backward cues. They discovered that the number of participants who experienced front-back confusions was lowered; and claimed that the occurrence of front-back confusions was reduced. Previously, 8 out of 10 participants experienced front-back confusions before enhancing the frontal and backward cues. When the frontal and backward cues in the HRTF filters were enhanced, only 4 out of 10 participants experienced front-back confusions. However, such kind of comparison could not prove whether the localization performance of the participants was improved because of the lack of statistical test and proper experimental design.

CHAPTER 4 BACKGROUND TECHNICAL DETAILS OF EXPERIMENTS AND PROTOTPYE IMPLEMENTATION

4.1 Introduction

In this study, azimuth (i.e. sound sources located on the horizontal plane at ear level) localization error and the occurrence of front-back confusions are considered. From past experiences, front-back confusions are one of the major problems when synthesizing sound sources using non-individualized HRTFs; and azimuth localization error can be used to indicate the severity of front-back confusions. Elevation error will not be considered in this thesis.

The central hypotheses of this thesis are (i) simplifying the spectra of non-individualized HRTFs will not affect the sound localization performance; (ii) manipulating the spectra of non-individualized HRTFs at different frequencies will exert effect on the localization performance. Cost and localization performance can be optimized if reducing complexities of HRTFs does not significantly affect the localization performance. Problem of front-back confusions can be solved if localization improvement is found after enhancing frontal and backward cues in HRTF filters. At the same time, the individual effect of frontal and backward cues on sound localization performance in each of the frequency band could be examined.

The purpose of this chapter is to introduce the experimental details and prototype implementation. The experimental details and prototype implementation related to the structure of the thesis can be referred to in Figure 1.7.

4.2 Overview of experimentation and prototype implementation

Three laboratory experiments were conducted in this thesis. The rationale for conducting those experiments in this thesis will be described as follows:

4.2.1 Experiment 1: Effects on HRTF complexity

This experiment was conducted to study the effects of simplifying the spectra of non-individualized HRTFs (i.e. changing the HRTF complexity) on sound localization accuracy. The reason of conducting this experiment at the beginning of this thesis was to investigate if changing the whole spectra of the non-individualized HRTFs had the effects on the sound localization accuracy. Once it was proved that changing the spectra of non-individualized HRTFs exerted the effects on the sound localization accuracy, further experiment could be proceed to investigate the effects of changing individual sub-bands along the whole frequency range. Details of Experiment one are fully described in Chapter 5.

4.2.2 Experiment 2: Prototype implementation

A prototype Virtual Headphone-based Surround Sound (VHSS) System has been developed as shown in Figure 4.1. The main contribution of this VHSS system is to test the virtual surround sound effects simulated by the set of HRTF filters loaded in the system. Since an optimized set of filters has been obtained from the first experiment, Experiment two (a usability experiment) was conducted to compare the processed two-channel outputs of surround sound effects by the VHSS system and those produced by the Dolby™ stereo outputs. The ratings could be used to verify the results from the first experiment. Further experiment can be continued if the performance of the optimized set of filters obtained from the first experiment was verified in Experiment two. Details of Experiment two are described in Chapter 6.

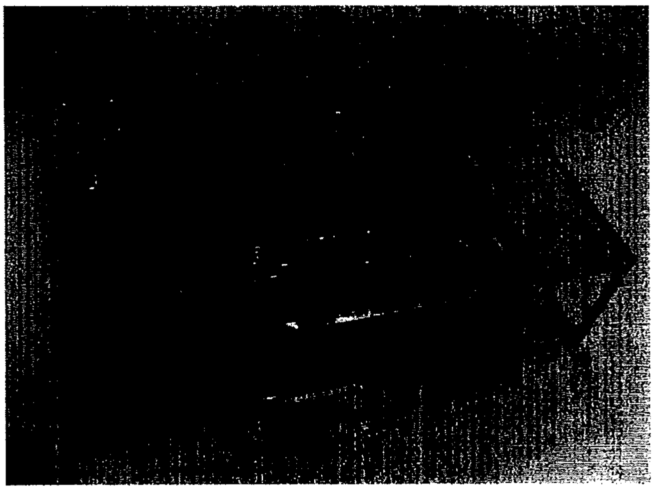


FIGURE 4.1 Overall structure of the prototype Virtual Headphone-based Surround Sound (VHSS) System.

4.2.3 Experiment 3: Effects of enhancing frontal and backward cues

The purpose of this experiment was to investigate the use of manipulating the spectra of non-individualized HRTFs at different frequencies to further improve the localization performance. The reason of conducting this experiment as the last experiment in this thesis because changing the whole spectra of the non-individualized HRTFs was needed to be proved to have the effects on the localization accuracy at the beginning of this thesis. This experiment was then conducted to examine the effects of changing the spectra at different frequencies on the localization performance. Detailed procedures and results of this experiment are reported in Chapter 7.

4.3 Equipment used in sound localization experiments

4.3.1 System requirements

In order to carry out the studies mentioned before, a sound localization system was set up. Figure 4.2 shows a block diagram on the sound localization system. The system had the following capabilities in those experiments:

(i) to present the sound cues randomly to participants via a pair of headphones;

(ii) to record the perceived position responded by participants; and

(iii) to change the sound cues presented conveniently in the system.

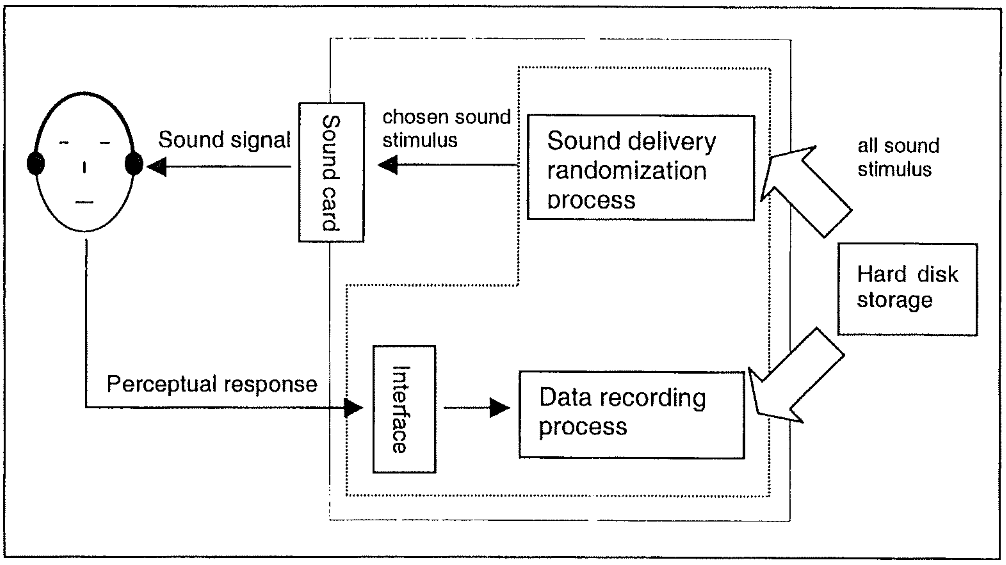


FIGURE 4.2 Block diagram on sound localization system.

4.3.2 Hardware

A Pentium III 600 Mhz personal computer was used. Different sound cues were stored in the hard disk for sound presentation. It has been equipped with Creative Live sound card. A pair of headphones (HD545, Sennheiser, Hannover, Germany) which connected to the audio output of the sound card was worn by the participants.

4.3.3 Software

A MATLab program has been modified from Braasch (2001) to control the experimentation. It is the interface (Figure 4.3) for the participants to indicate their perceived positions of each presented sound cue. The program has the following characteristics:

(i) sound cues were presented to participants through a pair of headphones in random order;

(ii) data on perceived directions are recorded; and

(iii) next sound cue is presented to participants automatically after participants respond to the previous sound cue.

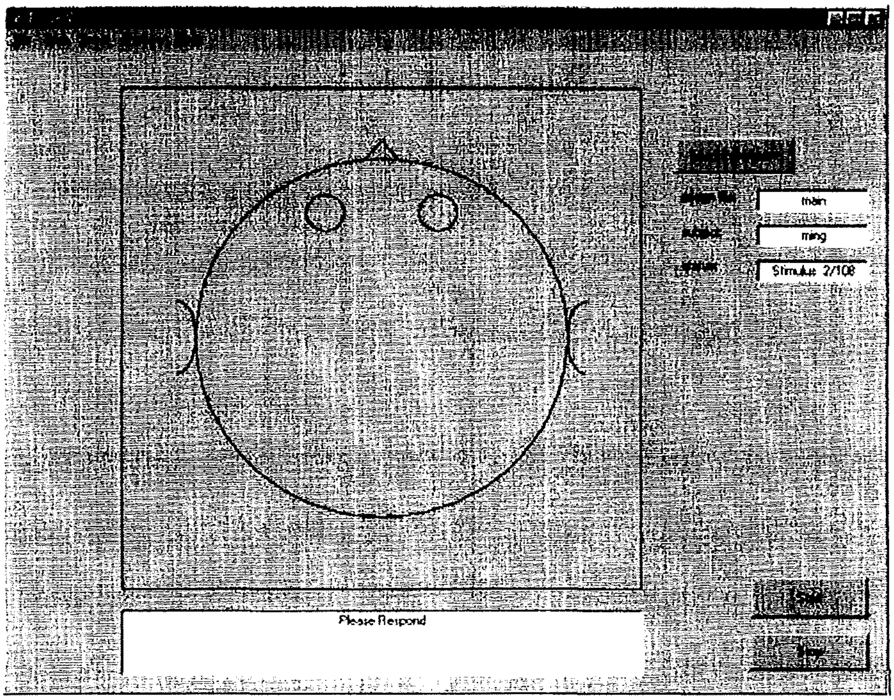


FIGURE 4.3 Interface for sound localization experiment.

4.3.4 Sound stimulus generation

The original source used in Experiment one was a clip of human speech in English ("Thank you for your participation in this experiment") and its translation in Cantonese. For the source used in Experiment three, a clip of music together with a male ("Please indicate your perceived direction on the interface") and a female voice ("Please pay attention to the perceived direction of this clip") spoken in Cantonese was used. For the HRTF filters, the popular MIT KEMAR non-individualized HRTF data (Gardner and Martin, 1995) were used to increase the general validity of the findings. Once different sets of filters were prepared, they were convolved with the original source to become binaural directional sound cues for the experiments. The volume level of the sound cues presented in the two experiments was about 60 dBA

4.3.5 Safety precautions

There is no anticipated danger in the experimental setup. In order to protect participants from hearing damage, sound cues listened presented to the participants were not greater than 80 dBA.

All experiments conducted were approved by the Human Subject Committee of the Hong Kong University of Science and Technology. This ensures that there is no foreseeable risk of impairment to the health of participants found from any experiments. Participants were instructed before experiments and they were informed that they may withdraw from the experiments at any time without the necessity to provide any reasons. A sample of the subject consent form used in these experiments can be referred to in Appendix A4.1.

4.4 Equipment used in prototype implementation

4.4.1 System requirements

A Virtual Headphone-based Surround Sound (VHSS) System was set up to carry out the usability experiment. Detailed descriptions can be referred to sections 6.1 and 6.2. The system had the following features:

(i) to decode the AC3 data stream format from DVD player into six digital sound signals according to the Dolby™ 5.1 standard (center, left, right, left rear, right rear and subwoofer);

(ii) to undergo convolution process on each channel in the Digital Signal Processor (DSP); and

(iii) to generate a resultant two-channel (i.e. left and right channels) output to a pair of headphones.

4.4.2 Hardware

A Motorola digital signal processing evaluation development board (DSP56362EVM, Motorola, Austin, USA) is the core part in this system as referred to Figure 4.1. It performs all operations including convolutions and signal decoding on those six input signals. The subwoofer channel (below 120 Hz) is not processed in the VHSS system as it does not provide much directional information for the human (Dolby 2001). A DVD player (DVD-A560EN, Panasonic, Osaka, Japan) was used to provide the AC3 data stream to the board. A Pentium III 600 Mhz personal computer and the Sennheiser headphone that used in Experiment one and three were prepared.

4.4.3 Software

A software called PPP Development Interface was used for the communication with the evaluation board. Commands were typed in the interface to load the filters to the digital signal processor (DSP) for convolution processes. Also, activating and deactivating the filtering processes were controlled by typing in the commands.

4.4.4 Virtual surround sound generation

As mentioned before, with the AC3 data stream provided by a DVD player, the AC3 decoder on the evaluation board could decode the data stream into six sound signals according to the Dolby^{TM} 5.1 standard. Those channels, except for the subwoofer, would undergo convolution processes with the filters loaded to the DSP, forming the resultant two-channel output. The volume level of the Dolby^{TM} stereo channels and the simulated two-channel outputs by the VHSS system was about 80 dBA.

4.4.5 Safety precautions

As the usability experiment asked for participants to give their ratings on the surround sound effects with the questionnaire provided, the safety precaution was similar to those mentioned in section 4.3.5.

4.5 Tasks

In Experiment one and three, participants were required to listen to the sound cues through a pair of headphones and respond their perceived positions for each of them. Same procedures continued until all sound cues were presented to the participants. In Experiment two, participants were required to watch the DVD tracks from a TV and listen to the virtual surround sound with a pair of headphones. Questionnaires were provided to fill in their ratings on the surround sound performance.

4.6 Participants

Male and female paid volunteers with normal hearing were involved in those experiments. Their hearing abilities were verified by audiometric screening of at most 20 dB. They were university students with ages between 19 and 27. For Experiments two and three, participants were categorized into groups:

with and without musical training background. The participants were not informed about the purpose of the experiment.

4.7 Data analysis

Both objective and subjective measures were used in these studies. Localization error and percentage of the occurrence of front-back confusions were measured to assess the effects of simplifying and manipulating the spectra of non-individualized HRTFs. Questionnaires were completed by the participants to judge their ratings on the virtual surround sound effects simulated from the VHSS system. In this case, objective measure was used in both Experiment one and Experiment three while subjective measure was used in the Experiment two.

4.7.1 Localization error

Localization error is defined as the angular difference between the perceived angle and the presented angle of a binaural directional sound cue.

4.7.2 Percentage of occurrence of front-back confusions

The occurrence of front-back confusions was counted when a sound cue presented in the frontal hemisphere of a listener was perceived at the back hemisphere and vice versa. The percentage of the occurrence of front-back confusions is the total number of front-back confusions occurred divided by the total number of binaural cues presented. This data reveals the severity of front-back confusions in the sound localization experiments. At the same time, data on whether front-back confusions occurred more frequently than back-front confusions or vice versa were obtained.

4.7.3 Questionnaire on surround sound ratings

Two questionnaires were prepared for participants to complete during the usability experiment (i.e. Experiment two). One questionnaire with seven questions focusing on the directional effects was designed for the part one of the experiment. Another questionnaire with 12 questions focusing on the spatial and surround sound effects was designed for part two. These questionnaires were adapted from the study of Berg and Rumsey (1999). All questions were rated by a 7-point scale asking for the ratings of the simulated surround sound effects. An example of a 7-point scale is shown as follows:

|------------------|------------------|------------------|------------------|------------------|------------------|

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

Questions were stated as a comparison format describing whether the characteristics of the first version was better than that of the second one. Participants would fill up their agreement on those statements during the pair-wise comparisons between the processed two-channel outputs of surround sound effects by the VHSS system and those produced by the Dolby^{TM} stereo outputs. Neutral (4) meant that there was no difference between the two versions. The questions focused on the directional effects, spatial effects and surround sound effects that they experienced.

4.7.4 Statistical test

A normality test was performed to examine if the data was normal or not. Statistical tests of parametric or non-parametric analysis were used to analyze the data, depending on whether the data was normal or not. Two-way analysis of variance (i.e. ANOVA) tests were used if the data followed normal distribution while Kruskal-Wallis one-way analyses of variance and Mann-Whitney tests were used if the data was not normal. Although the experiment used a within-subject design, Friedman two-way ANOVA tests were not used because in some conditions, the participants perceived the sound cues as located at the center of their heads and resulted in missing data in terms of localization errors and the occurrence of front-back confusions. With the statistical tests, it provided evidence on whether the factor had significant effect on localization performance.

CHAPTER 5 EXPERIMENT ONE: EFFECTS OF HRTF COMPLEXITY ON SOUND LOCALIZATION

5.1 Purpose of this experiment

Spectral cues have a close relationship with localization performance, especially the occurrence of front-back confusions. This experiment aimed to determine the effects of simplifying the spectra of non-individualized HRTFs on sound localization.

5.2 Objectives and hypotheses

An experiment was conducted to study the effects of simplifying the spectra of non-individualized HRTFs on sound localization performance. The objective was to determine whether there was a significant effect on localization accuracy if the number of coefficients used in the HRTF filters was reduced. According to what we have learnt from the literature, it was hypothesized that (i) directional sound cues generated by HRTFs with reduced number of coefficients would produce higher localization errors, and (ii) the increase in errors due to the reduction of HRTF coefficients would be associated with front-back confusions. These hypotheses were based on the studies by Wightman and Kistler (1989b), and Wenzel *et al.* (1993) and Bronkhorst (1995). They mentioned that if HRTFs could capture all the acoustical features as if the stimuli were presented in free-field, the localization performance would be accurate. Since the non-individualized HRTFs did not preserve all the important characteristics, they were unable to stimulate the virtual sound cues in a satisfactory way when compared with individualized HRTFs. In addition, Asano *et al.* (1990) discovered that the front-back errors increased when simplifying the complexities of HRTF filters, leading to higher localization errors. This confirmed that the performance of the HRTF filters would be degraded if the number of coefficients used in the filters was reduced.

5.3 Participants

Twenty-four participants (12 males, 12 females) served as volunteers for this experiment. All of them were university students with ages between 19 and 27. All of them passed the audiometric test for normal hearing of at most 20 dB. The participants were not informed about the purpose of the experiment.

5.4 Dependent and independent variables

In this experiment, two independent variables were used: (i) HRTF complexity (i.e. number of coefficients used in HRTF filters) of 128, 64, 32 and 18 coefficients; and (ii) the 8 azimuth directions of cues (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°) as shown in Figure 5.1.

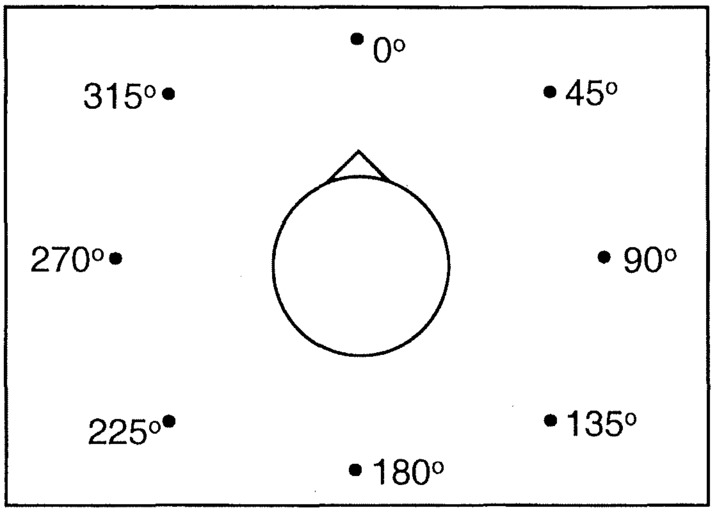


FIGURE 5.1 The 8 azimuth directions of sound cues that studied in Experiment one.

There have been many researchers who have measured the individualized HRTFs. Data on several sets of published HRTFs have been collected and extracted from the following authors: Wiener (1947); Robinson and Whittle (1960); Shaw (1974); Mehergardt and Mellert (1977); Blauert (1983); and Gardner and Martin (1995). Figure 5.2 shows the published data related to 0° direction. It was observed that the collected individualized HRTFs did not cover the whole frequency range (i.e. from 0 to 22.0 kHz) for some authors. The frequency range of HRTFs they collected varied among them.

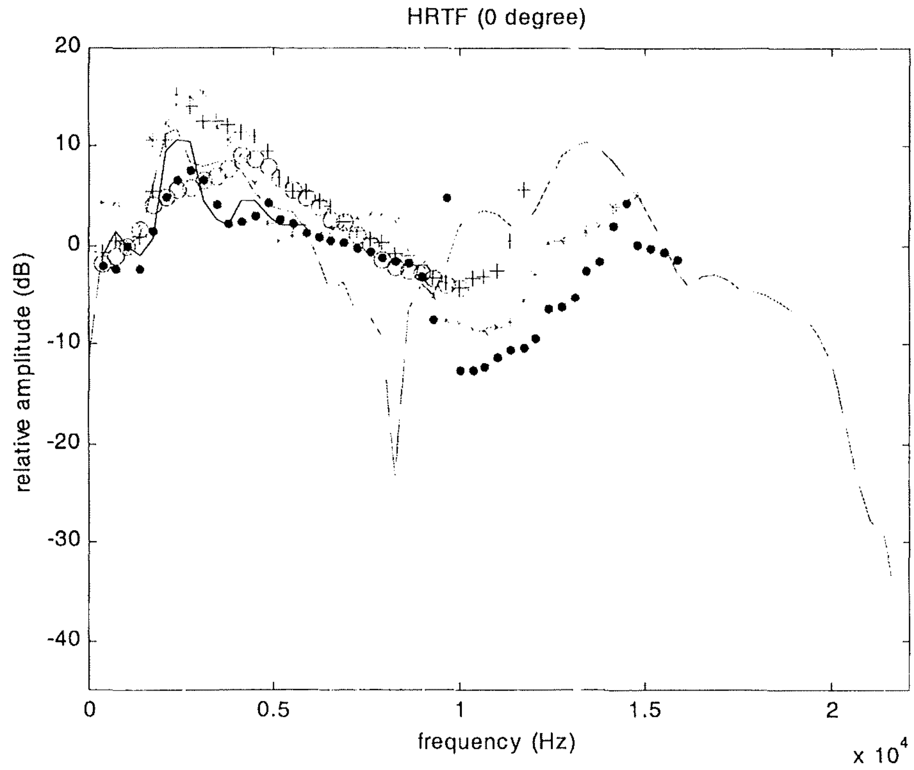


FIGURE 5.2 Modules of the HRTFs form sound cues at 0° direction (i.e. directly ahead of the listener) from different authors. For the HRTFs collected from some authors, they did not cover the whole frequency range ('-' Wiener, 1947; 'o' Robinson and Whittle, 1960; '.' Blauert, 1969a; '+' Shaw, 1974; '*' Mehergardt and Mellert, 1977; and '--' Gardner and Martin, 1995).

In order to increase the general validity of the findings, the popular MIT KEMAR non-individualized HRTF data (Gardner and Martin, 1995) were used in this experiment. Figure 5.3 illustrates the modules of HRTFs with different numbers of coefficients (128, 64, 32, 18 coefficients) for 0° direction (i.e. directly ahead of the listener). Filters of 18 coefficients were used, instead of 16 coefficients, because of the failure to obtain the normal sound quality. These HRTF filters were used to convolute with the original stimulus to generate sound cues of 8 azimuth directions as mentioned before (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°). Appendix A5.2 shows the modules of HRTFs with different numbers of coefficients for the other seven directions. The reason of choosing these directions is that directions of 0°, 45°, 90°, 270° and 315° are the possible locations for surround sound speakers according to Dolby™ 5.1 standard. For directions of 135°, 180° and 225°, they were used to examine the effect of front-back confusions.

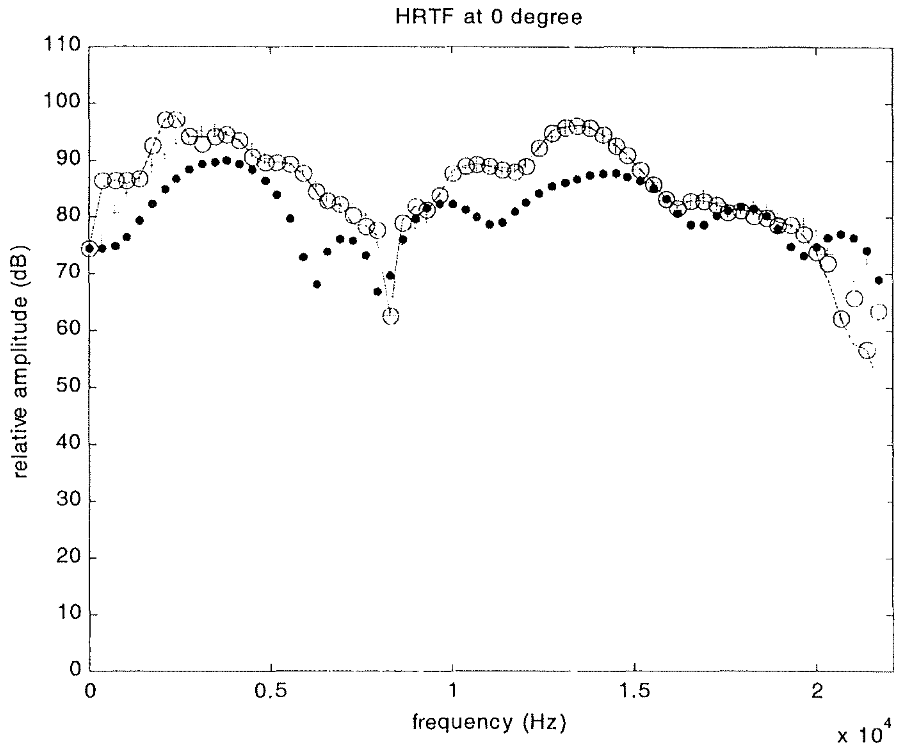


FIGURE 5.3 Modules of the HRTFs form sound cues at 0° direction ('-' 128 coefficients; 'o' 64 coefficients; '+' 32 coefficients; and '...' 18 coefficients).

The dependent variables were: (i) localization error; and (ii) percentage of the occurrence of front-back confusions. Localization error can be determined by the angular difference between the perceived angle and the presented angle of a binaural directional sound cue. The occurrence of front-back confusions was counted when a sound cue presented in the frontal hemisphere of a listener was perceived at the back hemisphere and vice versa.

5.5 Stimuli preparation

The prepared binaural sound cues lasted about 10 seconds each and contained a clip of human speech in English ("Thank you for your participation in this experiment") and its translation in Cantonese. There was a pause of about one second between the English and Chinese sentence. For each direction, while the details of the spectral cues embedded in the stimulated sound cues varied with the length (i.e. number of coefficients) of the HRTF filters, the inter-aural time difference (ITD) and inter-aural level difference (ILD) embedded in the sound cues were maintained to be the same regardless of the HRTF complexity. This was to ensure that, for the same direction, the only differences among the sound cues would be the level of complexity of the spectral cues. In addition to the 8 sound directions, a mono condition (i.e. no direction) was also added as a control condition. In total, there were 36 sound cues (9 sound directions * 4 levels of spectral complexities of HRTFs).

5.6 Procedure

The 36 prepared stimuli were presented via a MATLab program modified from Braasch (2001) for this study. Participants sat in front of a PC monitor and put on a pair of headphones (HD545, Sennheiser, Hannover, Germany). Stimuli were presented to them one by one in random order. After listening to each sound stimulus, the participants were then required to indicate their response by clicking on a diagram presented on the PC monitor (Figure 4.3). Participants were instructed to move the mouse to a position that best reflected (i) the azimuth direction of the sound stimulus in relation to the center of the head; and (ii) whether the stimulus appeared to come from inside or outside of the head. After listening to the 36 stimuli, the participants were given a five-minute rest. After the rest, the experiment was repeated and the order of presenting the next 36 stimuli was also randomized. In total, 72 stimuli were presented to each participant. The participants received no feedback on their performance during the entire experiment. The experiment was conduced inside an acoustic chamber with a background noise level of about 35 dBA.

5.7 Results and Discussion

The localization error data did not follow normal distribution and non-parametric statistical tests were used. Results of a Wilcoxon Signed Ranks Test showed that there was no significant difference between data collected in the two replications *(p=0.133)*. As a result, data from the two replications were combined in subsequent analyses. Figure 5.4 shows the median localization errors as functions of the 8 sound directions and the 4 levels of HRTF complexities as indicated by the number of filter coefficients. As mentioned in section 4.7.5, Kruskal-Wallis one-way analyses of variance were used for the within-subject design, the participants perceived the sound cues as located at the center of their heads in some conditions, resulting in missing data in terms of localization errors and occurrence of front-back confusions.

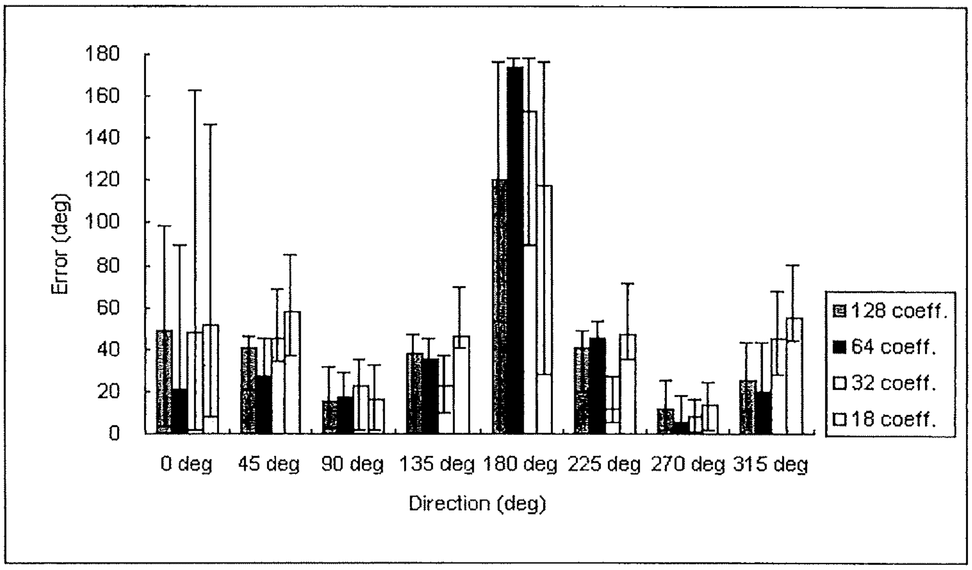


FIGURE 5.4 Median localization errors for 8 sound directions and 4 levels of HRTF complexities.

Kruskal-Wallis one-way analyses of variance were conducted to examine the effects of simplifying the spectra of HRTFs on localization errors. Simplifying the spectra of HRTFs had significant influence on the localization errors for the 45° *(p<0.001)*, 135° *(p<0.001)*, 180° *(p<0.05)*, 225° *(p<0.001)* and 315° *(p<0.001)* directions. For the 0°, 90° and 270° directions, simplifying the spectra of HRTFs had no significant effect on the localization errors *($p>0.1$)*. Mann-Whitney tests were conducted to compare the errors in pairs. The results of the tests are illustrated in Table 5.1. In Table 5.1, median localization errors associated with the same direction are listed in the same row. Along the same row, the condition with the highest error is listed first from left to right.

TABLE 5.1 Median localization errors for 4 levels of HRTF complexities (coefficients within lines did not have significant difference to each other).

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

Inspections of Table 5.1 indicate that HRTF filters of 18 coefficients produced significantly higher localization errors than the corresponding filters of 128, 64 and 32 coefficients for directions 135° *($p<0.005$)* and 315° *($p<0.05$)*. It was observed from the table that HRTF filters of 18 coefficients have higher localization errors than the corresponding filters of 128, 64 and 32 coefficients for directions 0°, 45°, 135°, 225°, 270° and 315°. These results agrees with Begault's work (1992b) which indicated that the reduction of HRTF complexity by up to 65 coefficients had no significant difference in localization errors for all directions ranged from 0° to 360° with 30° interval. Apart from the case for 180° direction, simplifying the spectra of HRTFs had no significant difference in the localization errors between 128 and 64 coefficients for all directions.

Since the problem of front-back confusion is known to be a serious problem, the localization errors were examined to exclude the errors introduced by front-back confusions (i.e., a correction in perceived directions of 180 degrees for 0° and 180° target directions, and a correction in perceived directions of 90 degrees for 45°, 135°, 225°, and 315° target directions if the corrections resulted in reductions in errors). Figure 5.5 illustrates the median localization errors with front-back confusions excluded. Figure 5.6 shows the percentages of front-back (back-front) confusions for the 6 sound directions and the 4 levels of HRTF complexities as indicated by the number of filter coefficients.

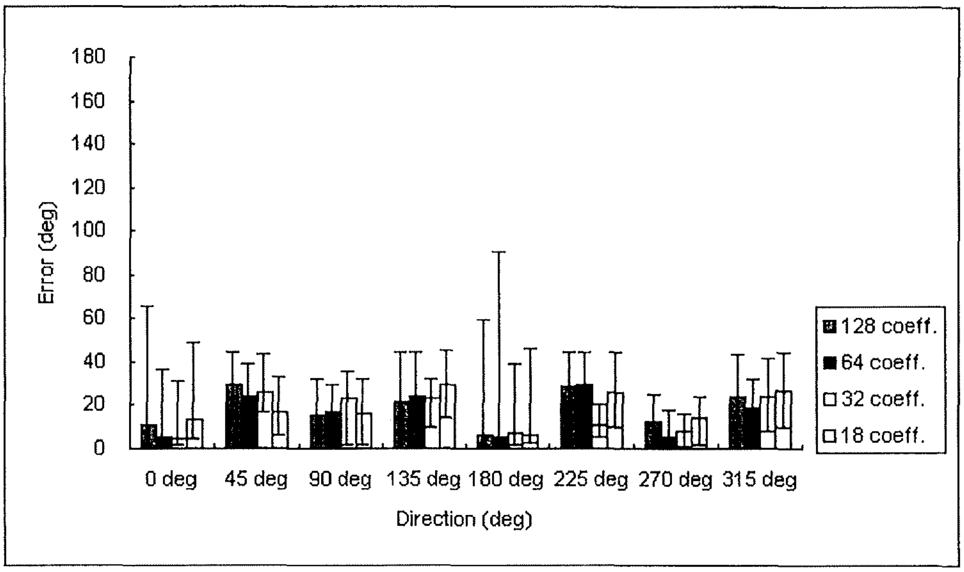


FIGURE 5.5 Median localization errors with front-back confusions excluded for 8 sound directions and 4 levels of HRTF complexities.

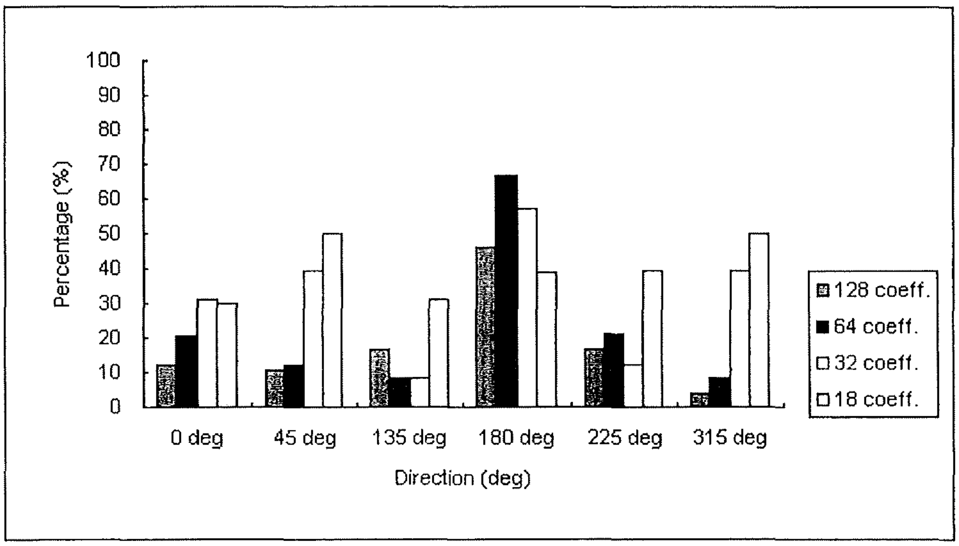


FIGURE 5.6 Percentages of the occurrence of front-back (back-front) confusions for 6 sound directions and 4 levels of HRTF complexities.

Inspections of Figures 5.5 and 5.6 indicate that most of the front-back confusions occurred with the 180° direction. The results of Kruskal-Wallis one-way analyses of variance showed that simplifying the spectra of HRTFs had no significant effect on localization errors with front-back confusions excluded *(p>0.1)* except for the direction 225°. This suggests that simplifying the spectra of HRTFs affected the occurrence of front-back confusions more than the localization errors with front-back confusions excluded. This agrees with the past literature because in this experiment, the changes in HRTF complexity were limited to the changes in the spectral complexity of HRTFs which were responsible for providing frontal and backward cues for a binaural sound cue. The same rationale could explain the lack of significant effect with 90° and 270° directions. It can be observed that for 180° direction, the percentages of front-back confusions were high even with the highest two levels of HRTF complexity. Among the four levels of HRTF complexity, 18 coefficients showed more front-back confusions for nearly all directions except the 180° direction. In fact, the percentage of front-back confusions at the 180° direction with HRTF filters of 18 coefficients was less than those with filters of 32, 64, 128 coefficients. This straight phenomenon is supported by the results of Mann-Whitney tests (Table 5.1). Inspection of Table 5.1 indicated that for the 180° direction, the localization error with HRTF filters of 18 coefficients was significantly lower than those filters with 64 coefficients and was ranked the least among the 4 levels of filter complexity. In order to confirm this straight effect, further small-scale experiment was conducted. Six listeners (different from the 24 listeners participated in Experiment one) were randomly selected. Similar to the experimental procedures in previous experiment, they were asked to show their response on the MATLab interface after listening to each stimulus. A total of 12 sound cues (1 sound direction (180°) * 4 levels of HRTF complexity * 3 repetitions) were presented to each participant.

Figure 5.7 shows the median localization errors as functions of the 4 levels of HRTF complexity for 180° direction. Results indicated that HRTF filters of 128 and 18 coefficients had similar localization errors to each other while filters of 64 coefficient was the highest errors. Although statistical analysis revealed that simplifying the spectra of HRTFs had no significant effect on localization error for 180° direction *(p=0.217)* due to small sample size. The similar pattern could still be found. It might be due to the fact that the notches located between 5.0 – 10.0 kHz could still exert its effect to act as a backward cue. This can be observed in Figure 5.8 showing different levels of HRTF complexities at 180° direction.

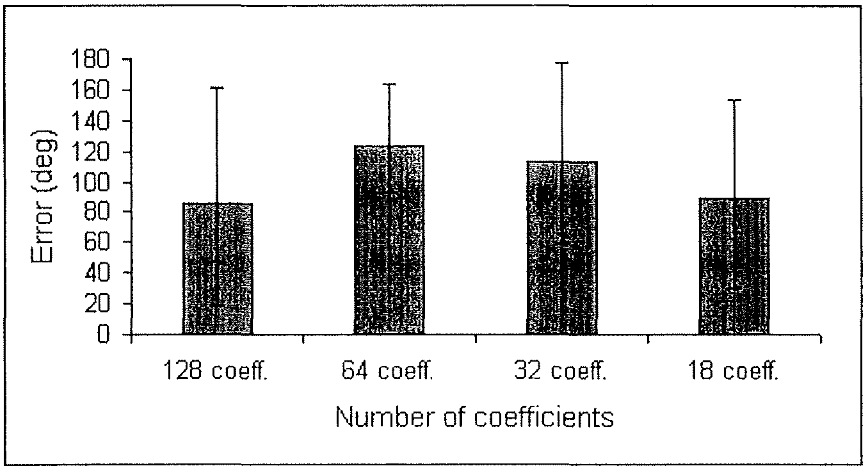


FIGURE 5.7 Median localization errors for 4 levels of HRTF complexities of 180° direction (data of 6 participants in further investigated small-scale experiment).

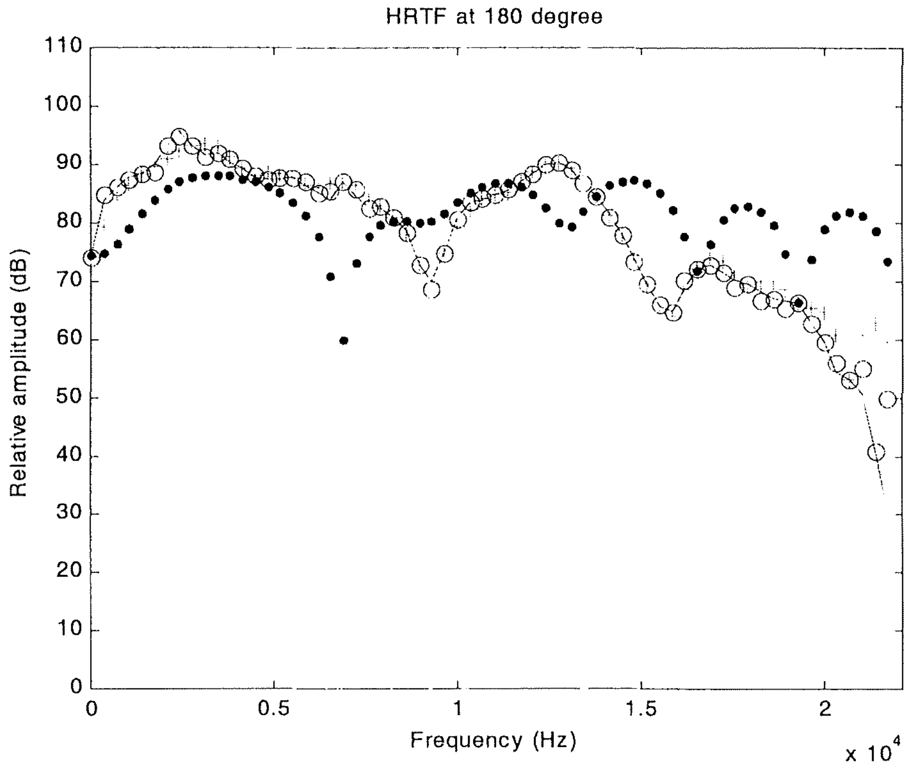


FIGURE 5.8 Modules of the HRTFs form sound cues at 180° direction ('-' 128 coefficients; 'o' 64 coefficients; '+' 32 coefficients; and '...' 18 coefficients).

For HRTF filters of 18 coefficient, the notch was situated at around 7.0 kHz which was different from the filters of other coefficients. Although the HRTF filters of 128 and 64 coefficients matched with a high degree, the difference in localization errors between the two coefficients might be due to the presence of a small peak located at around 22.0 kHz for filters of 64 coefficient. Further analysis is required in the future.

To conduct the statistical tests for the percentage of the occurrence of front-back confusions, data for all directions for each participant have been grouped. Results showed that simplifying the spectra of HRTFs had significant effect on the percentage of the occurrence of front-back confusions *($p<0.001$)*. HRTF filters of 128 and 64 coefficients produced significantly lower percentages of the occurrence of front-back confusions when compared with those filters of 32 coefficients *($p<0.01$)*. On the other hand, HRTF filters of 32 coefficients had significantly lower percentages than those with filters of 18 coefficients *($p<0.05$)*. These results are in agreement with the study reported by Asano *et al.* (1990). Their experiment showed that higher percentages of front-back confusions were experienced by participants when the number of coefficients used in the filters were reduced.

Results of Kulkarni and Colburn (1998) was different from this study. They claimed that participants were unable to differentiate the real and virtual sounds until the coefficients of filters used were reduced to 16 coefficients. As reported previously, HRTF filters of 32 coefficients had significant difference in the localization errors when compared to those filters of higher coefficients (i.e. 128 and 64 coefficients) for directions 45° and 315°, which was different from the study of Kulkarni and Colburn (1998). One of the possible reasons was due to their small sample size of only four participants for the experiment. Also, individualized HRTFs were used in the study of Kulkarni and Colburn.

From the previous results, the optimized set of filters for the prototype Virtual Headphone-based Surround Sound (VHSS) System can be chosen. To synthesize the left (315°) and right (45°) channels, it was proposed to use HRTF filters of 64 coefficients as the corresponding localization errors and were the significantly lowest compared with those filters of 32 and 18 coefficients *($p<0.01$)*. Also, when HRTF filters of 64 coefficients compared with 128 coefficients, no significant difference in errors was found *($p>0.5$)*. For the center (0°), left rear (270°) and right rear (90°) channels, HRTF filters of 32 coefficients was chosen. This was because filters of 32 coefficients were not significantly different from those errors associated with HRTF filters of 128 and 64 coefficients for directions 0° *($p=0.54$)*, 90° *($p=0.513$)* and 270° *($p=0.112$)*. Although HRTF filters with 18 coefficients showed no significant difference in localization errors with those filters of 32 coefficients for these three channels, the calculating power can be fully utilized when using 32 coefficients.

For the MATLab program that modified from Braasch (2001), people might claim that participants were unable to indicate the correct perception for the stimuli as there was no marking of the degree (e.g. 0^{o}, 45^{o}, 90^{o} etc) on the interface. This would affect the localization accuracy for the sound perception. Nevertheless, as this experiment is a within-subject experiment, there should be no effect on the localization accuracy for the participants to express their perceived directions of sound sources.

5.8 Summary

Front-back confusions were considered to be the major cause for the localization errors as simplifying the spectra of HRTFs had no significant effect on the localization errors with front-back confusions excluded except for direction 225°. Simplifying the spectra of HRTFs had significant effect on the localization errors except for directions 0°, 90° and 270°. HRTF filters of 32 coefficients were not significantly different from those localization errors with HRTF filters of 128 and 64 coefficients for directions 0°, 90°, 135°, 180° and 270°. As HRTF filters of 64 coefficients had significantly lower localization errors than the corresponding filters of 32 and 18 coefficients for directions 45° and 315°, an optimized set of HRTF filters for the VHSS system were obtained: HRTF filters of 64 coefficients were used for left (315°) and right (45°) channels while filters of 32 coefficients were chosen for center (0°), left rear (270°) and right rear (90°) channels.

CHAPTER 6 EXPERIMENT TWO: USABILITY EXPERIMENT ON SURROUND SOUND PERFORMANCE OF VIRTUAL HEADPHONE-BASED SURROUND SOUND (VHSS) SYSTEMS

6.1 Overview of the Virtual Headphone-based Surround Sound (VHSS) system

The Virtual Headphone-based Surround Sound (VHSS) System consists of the following parts: a DVD player (DVD-A560EN, Panasonic, Osaka, Japan), a Motorola Digital Signal Processing evaluation development board (DSP56362EVM, Motorola, Austin, USA) and a pair of headphones (HD545, Sennheiser, Hannover, Germany). Figure 6.1 shows a photograph of the Motorola evaluation board used in the VHSS system.

FIGURE 6.1 The Virtual Headphone-based Surround Sound (VHSS) DSP56362EVM evaluation board.

When a DVD disc containing Dolby™ 5.1 surround sound tracks was played in the DVD player, the player would transmit the digitally coded surround sound information through the AC3 data stream format. The AC3 digital signal was fed into the Motorola DSP56362EVM evaluation board. Inside the DSP56362EVM board, the AC3 encoded data stream would be decoded into six digital sound signals according to the Dolby™ 5.1 standard (center, left, right, left rear, right rear and subwoofer). Those decoded signals would then be convolved with the nine appropriate HRTF filters. The outputs of the nine convolution processes would then be combined into a left channel and a right channel. The procedures of signal processing within the DSP56362EVM board are illustrated in Figure 6.2. Inspections of Figure 6.2 show that the center, left, right, left rear and right rear channels were filtered by appropriate HRTF filters so that they would appear to come from the directions of 0°, 315°, 45°, 270°, 90° respectively. As the subwoofer channel (below 120 Hz) does not provide much directional information for the human (Dolby 2001), this channel was not filtered. Inspections of Figure 6.2 indicate the use of nine HRTF filtering algorithms (i.e. four filters with 32 coefficients and five filters with 64 coefficients). The abilities of these nine HRTF filters to simulate accurate binaural directional cues have been studied in Experiment one.

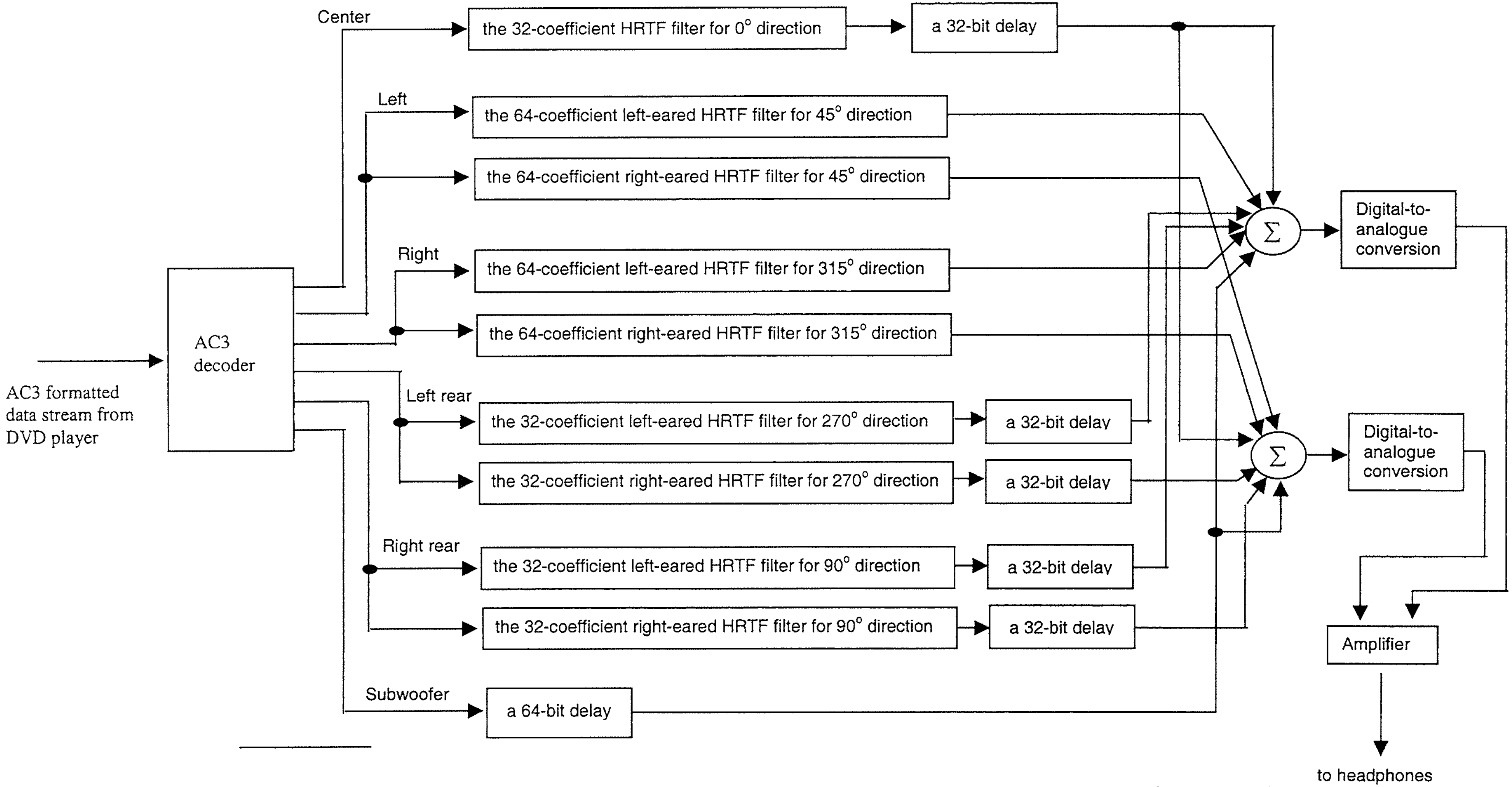


FIGURE 6.2 An illustration of the signal processing paths inside the Virtual Headphone-based Surround Sound (VHSS) DSP56362EVM evaluation board.

6.2 Purpose of this experiment

As mentioned before, the prototype Virtual Headphone-based Surround Sound (VHSS) system can decode the Dolby™ AC3 data stream into six digital sound signals and simulate virtual surround sound into a two-channel output. Having higher complexities of HRTF filters can ensure an excellent surround sound effect simulated from the VHSS system. However, higher calculating power (million instructions per second, MIPS) is needed for the DSP used. The calculating power and cost of VHSS system are optimized if the complexity of HRTF filters can be reduced while keeping satisfactory sound localization performance. As the dominant factor affecting the complexity of the filters is the details of the spectral cues, reducing the number of coefficients used in the filters can simplify the system. From the results of the first experiment, the left (315^o) and right (45^o) channels were simulated with filters of 64 coefficients while 32 coefficients of HRTF filters were used for the center (0^o), left rear (270^o) and right rear (90^o) channels.

6.3 Objectives and hypothesis

This experiment was a double-blinded usability experiment conducted to compare the surround sound ratings between the Dolby™ stereo channels and the simulated two-channel outputs by the VHSS system. It was hypothesized that the VHSS system would produce significantly higher surround sound ratings than the Dolby™ stereo channels. The double-blinded refer to the lack of knowledge on the order of presentation of conditions in both the participants and the experimenter. Besides the comparison on surround sound ratings between the Dolby™ stereo channels and the VHSS system, a second objective was to examine whether a history of musical training background (e.g. playing musical instruments or singing in a choir) would affect a listener's ability to assess the surround sound effects.

6.4 Participants

Forty participants served as paid volunteers for this experiment. As the authors had no reason to expect gender difference in the ability to appreciate surround sound, there was no attempt to balance the number of male and female participants. There were 16 male and 24 female listeners. Twenty of the 40 participants were experienced in playing musical instruments or singing in a choir while the other 20 had no musical training background. The experienced ability of playing musical instruments or singing in a choir was measured by a pre-exposure questionnaire. All participants were university students with aged between 19 and 27. All of them passed the audiometric test for normal hearing of at most 20 dB. The participants were not informed about the purpose of the experiment.

6.5 System and stimuli preparation

The VHSS system was used to deliver the simulated surround sound effect condition while the Dolby™ stereo channels was taken directly from the audio outputs of the DVD player. Detailed descriptions on how the VHSS system works can be found in sections 6.1 and 6.2.

Two sound tracks were used: (i) the Dolby™ Surround 5.1 channels demonstration from Dolby™ demonstration DVD for part one of the experiment (Hong Kong & Taiwan Chinese Productions, 1999), and (ii) the movie Air Force One™ DVD for part two of the experiment (Bernstein *et al.*, 1999). The former contained music played sequentially from each of the six surround channels (center, left, right, left rear, right rear and subwoofer) and the latter one contained surround sound effects from air-to-air combat scenes. Both sound tracks had their corresponding videos which were shown on the same TV during all conditions. Depending on the surround sound quality, listeners should be able to match the spatial orientations of the perceived audio cues with those perceived from the TV. For the first demonstration with the Dolby™ 5.1 DVD, this meant a match between the audio and visual cues of music played sequentially from each of the six virtual speakers located at center, left, right, left rear, right rear and subwoofer locations. As stated in section 6.1, the subwoofer audio channel was not meant to carry any spatial information. Figure 6.3 shows the sequence for the six surround channels delivering corresponding sound effects to the participants in part one of the experiment.

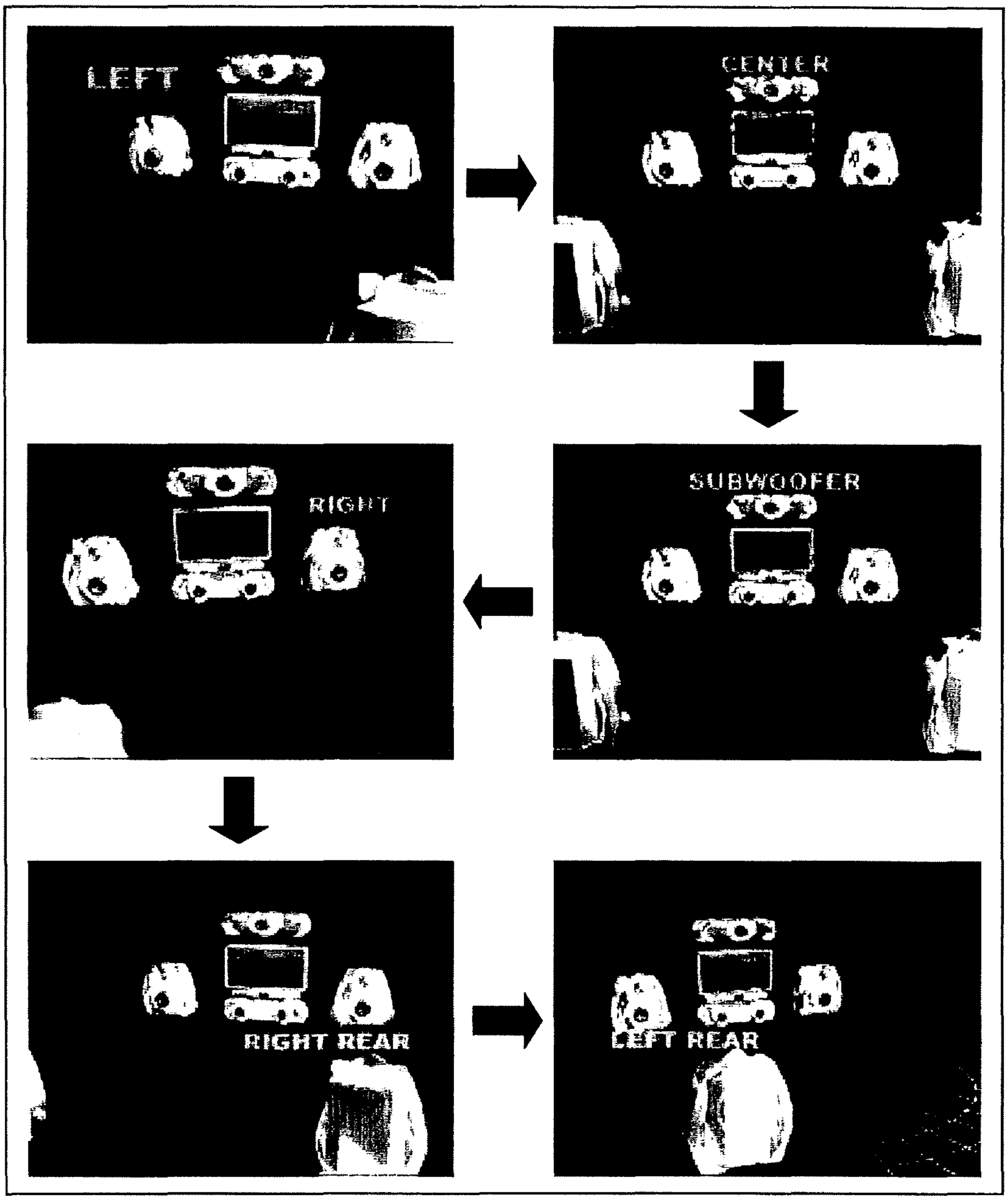


FIGURE 6.3 Sequence of the six surround channels (left → center → subwoofer → right → right rear → left rear) delivering corresponding sound effects in part one of Experiment two.

For the second demonstration with Air Force One™, this meant a match between the audio and visual cues of each missile, shrapnel and fighter planes. Figure 6.4 shows the sequence of important sound cues appeared in the selected clip of movie. The following described the important cues for part two of the experiment.

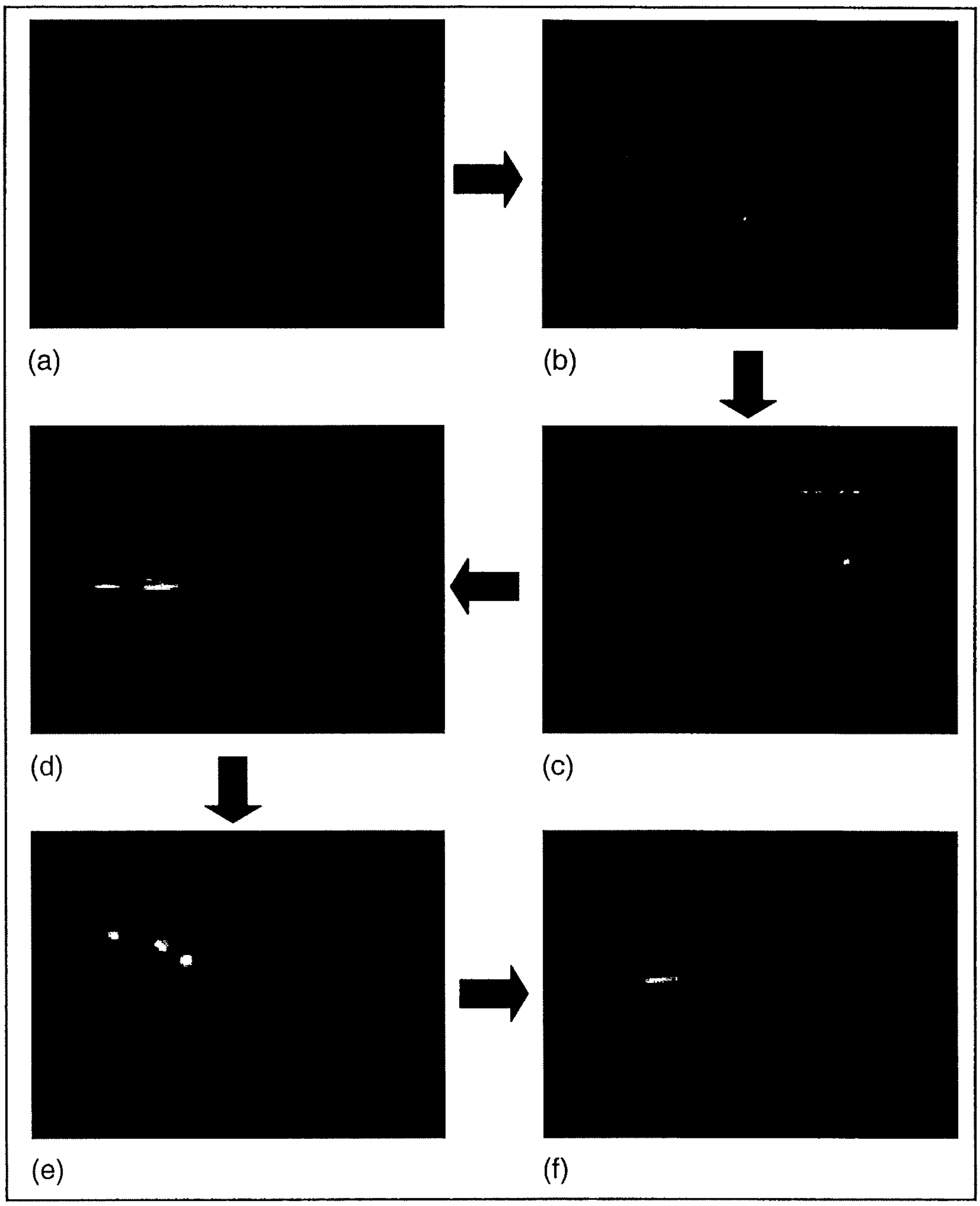


FIGURE 6.4 Sequence of important sound cues appeared in part two of Experiment two.

Figure 6.4a indicates a missile being shot from the front side towards the direction of audience's left hand side, whereas a missile flied from the direction of the audience towards the front side as shown in Figure 6.4b. Figure 6.4c shows a missile passing through the right hand side of the audience and flying towards the front side. Figure 6.4d captures the scene of several fighter planes shooting the missiles from the left hand side to the right hand side of the screen. Figure 6.4e shows a plane flying towards the audience and figure 6.4f indicates a fighter plane shooting a missile from the right hand side to the left hand side of the audience.

6.6 Procedure

For the entire experiment, doubled-blinded design was used. This eliminated any bias due to the brand-name preference by the participants and the experimenter. As explained in section 6.5, this experiment had two tasks: (i) to listen, watch and compare two similar DVD tracks from the Dolby™ surround 5.1 channels demonstration DVD, (ii) to listen, watch and compare two similar DVD tracks from Air Force One™. During the experiment, participants were instructed to wear a pair of headphones and sit in front of a TV to watch the movie simultaneously. Although the volume levels of the sounds presented between the VHSS system and the Dolby™ stereo channels were tuned to give the same rms magnitude levels, listeners were given the opportunities to tell the experimenter should they perceived that the sound levels from the two systems were different. Only eight out of 40 participants raised such concerns on the volume issue. In order to ensure the volume level remained control for the experiment, data of those eight participants have been excluded. Comparison of 32 participants with those of 40 participants was found in detail in results and discussions part.

For each pair-wise comparison, participants listened to the Dolby™ stereo channels and the simulated two-channel outputs by the VHSS system one by one. The sequence for presenting the two versions was randomized. An experimenter without knowing the sequence of the two versions presented would interview the participants to complete the questionnaire (Appendices A6.1 and A6.2) as described in section 4.7.3 for part one and part two of the experiment. Participants had to compare the directional effects between Dolby™ stereo channels and the simulated two-channel outputs by the VHSS system for questionnaire A in part one, whereas the surround sound effects and spatial effects were compared for questionnaire B in part two of the experiment. Ratings were marked down on a continuous scale for each question. Effects of subwoofer were told to be ignored as sounds from subwoofer have not been processed for directional effects. Comparisons were repeated for part one and part two of the experiment. If the sound effects from the VHSS system were presented first in the 1st trial, then it would be presented as the second one in the 2nd trials and vice versa. Participants were given a five-minute rest before the part two started.

6.7 Results and Discussion

As mentioned in previous section, 8 out of 40 participants requested for adjusting the volume levels in the experiment. To prevent from suffering the effects on the surround sound ratings by different volume levels, data of all 40 participants were analyzed to compare with those of 32 participants (i.e. data of those eight participants who adjusted the volume levels were excluded).

The Cronbach alpha values were 0.7384 and 0.9462 for questionnaires A and B respectively. When data of those participants who adjusted the volume levels were excluded (i.e. 32 participants), the Cronbach alpha values were 0.7501 and 0.9467 for questionnaires A and B respectively. This indicated that the data were reliable enough. One interesting findings was that when question Q2 in questionnaire A and question Q9 in questionnaire B were deleted, the alpha values were found to increase in both questionnaires. It might be due to the observation that participants were unable to get the idea of "inside-the-head locatedness (IHL)" since some participants spent more time to respond to these questions.

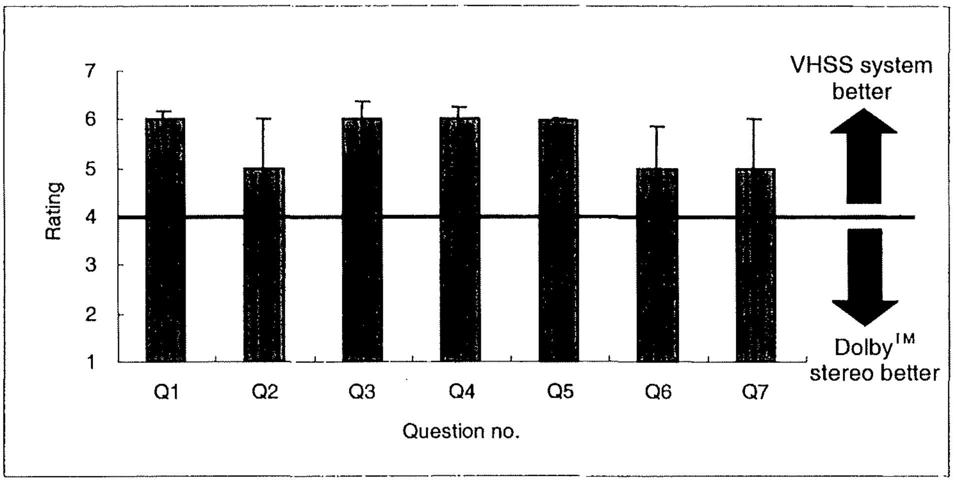


FIGURE 6.5 Median ratings for directional effects of the VHSS system when playing the Dolby™ 5.1 Demonstration DVD (40 participants, included data of those eight participants who adjusted volume levels).

The data did not follow normal distribution and non-parametric statistical tests were used. Results of a Wilcoxon Signed Ranks Test showed that there was no significant difference between data collected in the two replications *($p>0.05$)*. As a result, data from the two replications were combined in subsequent analyses. Figures 6.5 and 6.7 show the median ratings for the surround sound performance of the VHSS system for part one playing the Dolby™ 5.1 Demonstration DVD and part two playing the Air Force One™ DVD respectively. Figure 6.6 and 6.8 shows same kind of information but data of those eight participants who adjusted volume are excluded.

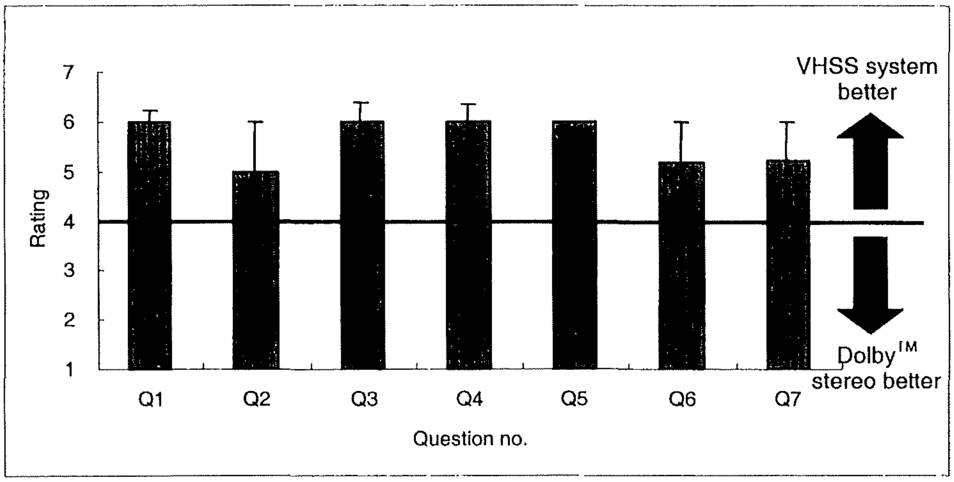


FIGURE 6.6 Median ratings for directional effects of the VHSS system when playing the Dolby[TM] 5.1 Demonstration DVD (32 participants, excluded data of those eight participants who adjusted volume levels).

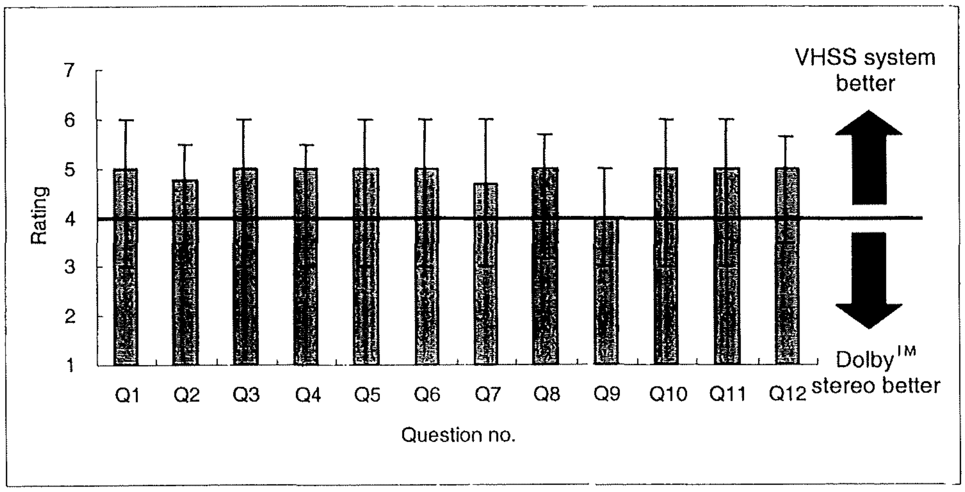


FIGURE 6.7 Median ratings for surround sound and spatial effects of the VHSS system when playing the Air Force One[TM] DVD (40 participants, included data of those eight participants who adjusted volume levels).

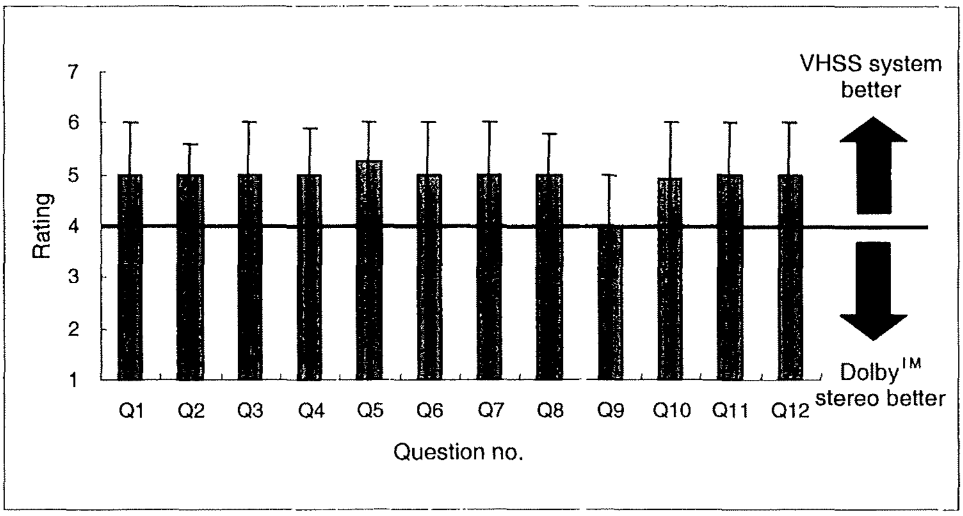


FIGURE 6.8 Median ratings for surround sound and spatial effects of the VHSS system when playing the Air Force One™ DVD (32 participants, excluded data of those eight participants who adjusted volume levels).

Inspections of Figures 6.9 and figure 6.10 indicated that there was no significant difference in the surround sound ratings between data of 32 participants (i.e. excluded those eight data who adjusted volume levels) and 40 participants (i.e. included those eight data who adjusted volume levels) *($p>0.5$)*. Nevertheless, the following data analysis would base on those data of 32 participants who remained the volume levels unchanged throughout the whole experiment. This ensured that the volume level was under controlled.

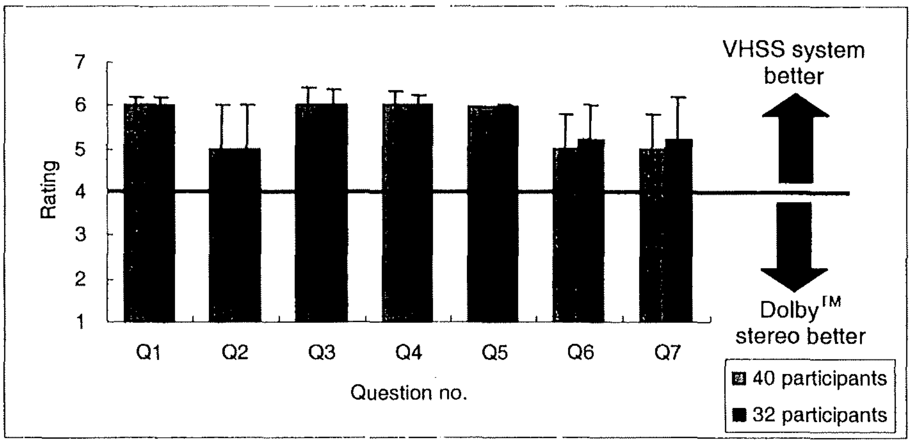


FIGURE 6.9 Median ratings for directional effects of the VHSS system between data of 32 participants (i.e. excluded data of those eight participants who adjusted volume levels) and 40 participants (i.e. included data of those eight participants who adjusted volume levels) when playing the Dolby™ 5.1 Demonstration DVD.

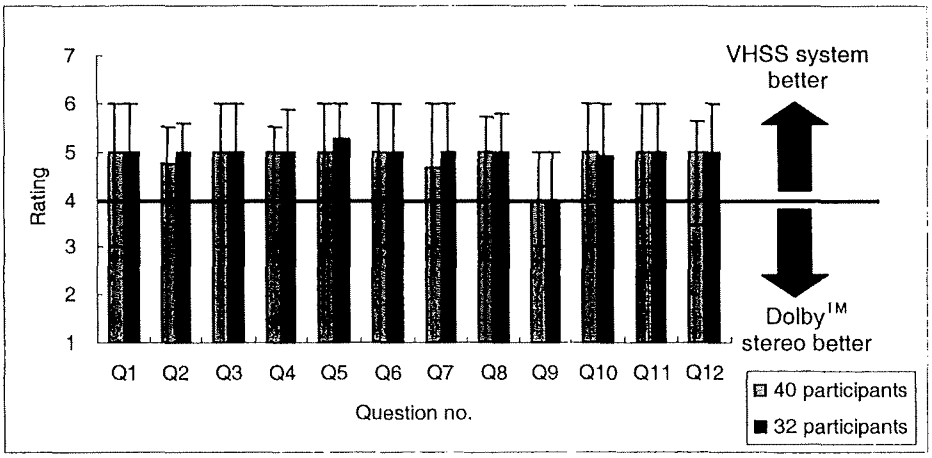


FIGURE 6.10 Median ratings for surround sound and spatial effects of the VHSS system between data of 32 participants (i.e. excluded data of those eight participants who adjusted volume levels) and 40 participants (i.e. included data of those eight participants who adjusted volume levels) when playing the Air Force One™ DVD.

As rating score "4" meant that there was no difference between the surround sound ratings between the Dolby™ stereo channels and the simulated two-channel outputs by the VHSS system, this implied that the VHSS system showed higher surround sound ratings if the median values were found to be greater than 4 for each question. It was observed that the median values for 18 out of 19 questions were larger than 4 except for question Q9 in questionnaire B. Results of Wilcoxon Signed Rank tests indicated that the simulated surround sound effects from the VHSS system showed significantly higher surround sound ratings than the Dolby™ stereo channel for all questions except for question Q9 ($p<0.05$). The following showed the question Q9 that extracted from the questionnaire B (i.e. part two of the experiment):

9. Sounds in the 1st version appear to originate **MORE** from **"Inside of your head"** than the sounds in the 2nd version.

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

It was found that this question was related to the issue on the externalization of sound effects. As the VHSS system did not focus on improving the externalization performance of the synthesized virtual sound, this led to why the VHSS system did not have a higher rating than the Dolby™ stereo channel for this question. In conclusion, these results provided the evidence that the optimized set of filters obtained from previous experiment showed satisfactory performance in the VHSS system.

Figures 6.11 and 6.12 illustrate the comparison of the surround sound ratings between participants with and without musical training background for part one and part two of the experiment respectively. Results of the Wilcoxon Signed Rank tests showed that there was no significant difference in the ratings between the two groups of participants when playing the Dolby™ 5.1 Demonstration DVD *($p>0.1$)*. For part two playing the Air Force One™ DVD, it was observed that participants without musical training background had higher surround sound ratings than those participants with musical training background. However, statistical test showed that significant difference was only found for questions Q3, Q5, Q7, Q10 and Q12 *($p<0.05$)*.

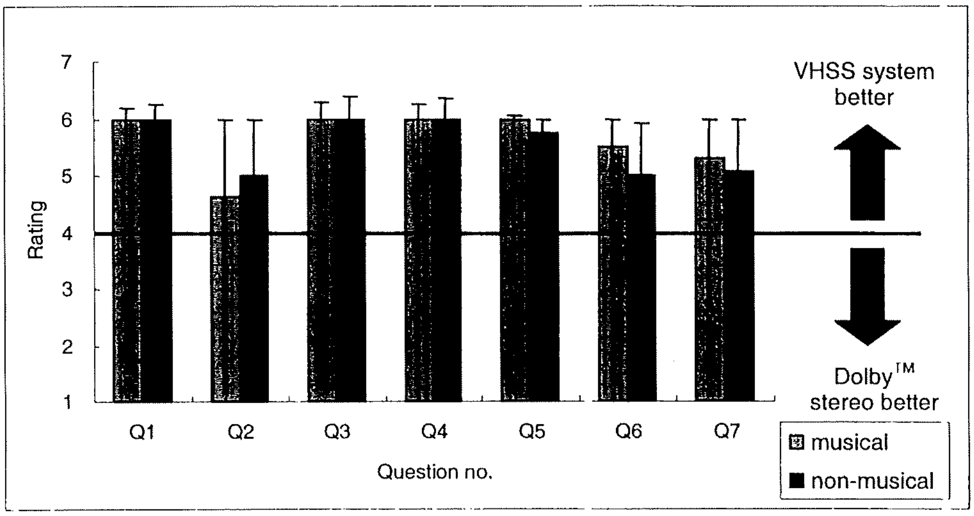


FIGURE 6.11 Median ratings for directional effects of the VHSS system between participants with and without musical training background when playing the Dolby™ 5.1 Demonstration DVD.

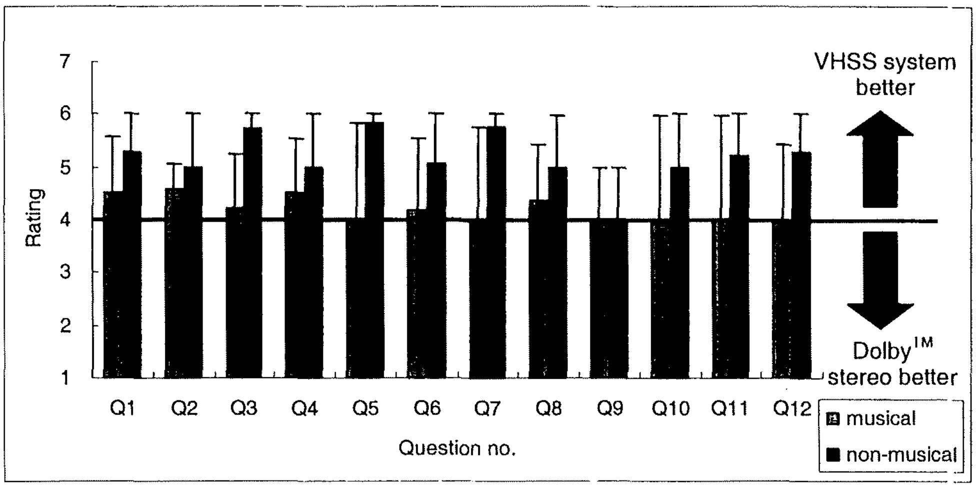


FIGURE 6.12 Median ratings for surround sound and spatial effects of the VHSS system between participants with and without musical training background when playing the Air Force One™ DVD.

Although the number of male and female were the same for the group without musical training background, more female were found in the group with musical training background. As female tend to have less excitement towards surround sound and spatial effects when compared with male, the unbalanced number of male and female for the group with musical training background may be attributed to the phenomenon that participants with musical training background showed lower surround sound ratings for the VHSS system when playing the Air Force One™ DVD.

The surround sound ratings for the VHSS system between male and female when playing the Air Force One™ DVD are shown in Figure 6.13. Inspections of Figure 6.13 indicate that the ratings between male and female participants for all the questions were similar to each other. Results of Wilcoxon Signed Rank tests showed that gender had no significant effect on the surround sound ratings *(p>0.2)*.

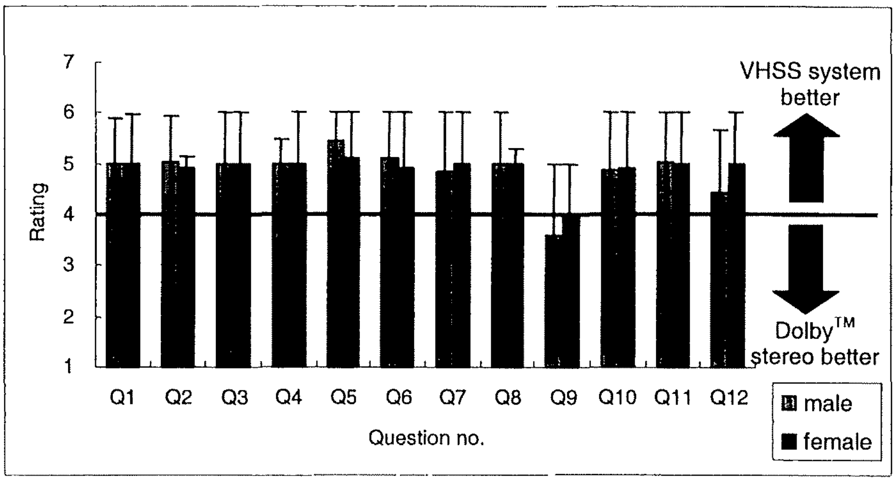


FIGURE 6.13 Median ratings for surround sound and spatial effects of the VHSS system between male and female participants when playing the Air Force One™ DVD.

Figure 6.14 indicates the median ratings for surround sound and spatial effects of the VHSS system between participants of gender difference with and without musical training background when playing the Air Force One™ DVD. By observation, male with musical training background showed comparatively lower surround sound ratings for all questions. This contradicted to the previous assumption that female tend to have less excitement towards surround sound and spatial effects.

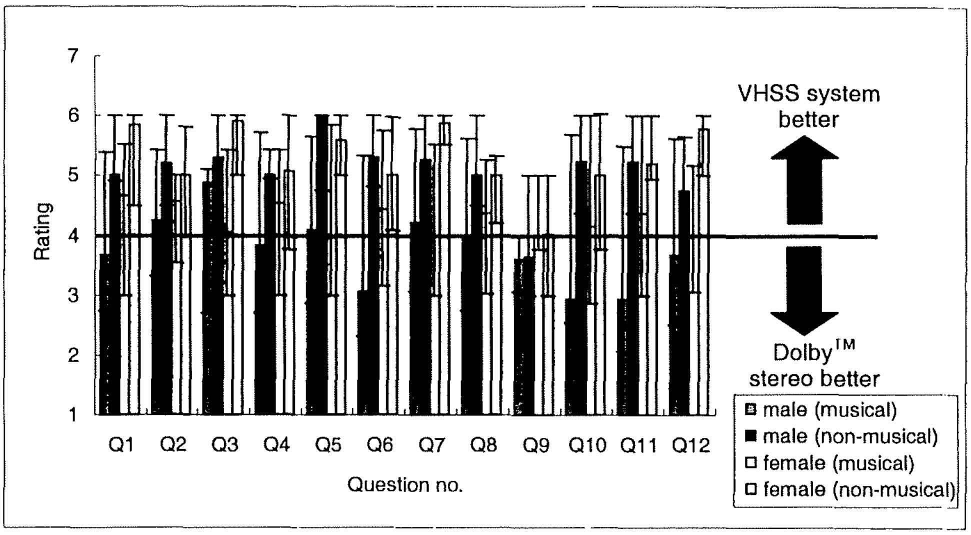


FIGURE 6.14 Median ratings for surround sound and spatial effects of the VHSS system between participants of gender difference with and without musical training background when playing the Air Force One™ DVD.

It would be necessary to collect the information on the preference for the participants towards enjoyment of surround sound effects. It is hypothesized that the ratings will be higher if participants interest on issues with surround sound effects. They will enjoy more and immerse into the movie to answer the questions related to surround sound and spatial effects.

6.8 Summary

The simulated surround sound effects from the optimized filters based on the first experiment in the VHSS system showed significantly higher performance ratings than the Dolby™ stereo channels for 18 out of 19 questions. Results showed that there was no significant difference in the surround sound ratings between participants with and without musical training background except for 5 out of 12 questions in part two when playing the Air Force One™ DVD.

CHAPTER 7 EXPERIMENT THREE: EFFECTS OF FRONTAL AND BACKWARD CUES ON SOUND LOCALIZATION

7.1 Purpose of this experiment

Frontal and backward cues can help listeners to distinguish sounds coming from the front or at the back directions, this experiment was conducted to investigate the effects of manipulating the spectra of non-individualized HRTFs at different frequencies to further improve the localization performance.

7.2 Objectives and hypothesis

The objective of this study was to investigate the effects of enhancing the magnitude of spectral cues at different frequency bands along non-individualized HRTFs on sound localization accuracy. It was hypothesize that the localization errors would be lower for those binaural direction cues prepared using HRTFs with enhanced spectral cues. This hypothesis was consistent with the results of Tan and Gan (1998) who reported that the number of participants experienced front-back confusions reduced if the spectral cues were enhanced.

7.3 Participants

Thirty-two participants (12 males, 20 females) served as paid volunteers for this experiment. They were university students with ages between 19 and 27. All of them passed the audiometric test of at most 20 dB and were therefore regarded as having normal hearing. Participants were not aware the purpose of the experiment.

7.4 Dependent and independent variables

The independent variables used were (i) levels of enhancements (OdB, 12dB, 18dB) for peaks and notches; (ii) ranges of frequency bands (band 1, 0.2 – 0.69 kHz; band 2, 0.69 – 2.4 kHz; band 3, 2.4 – 6.5 kHz; band 4, 6.5 – 10.0 kHz; band 5, 10.0 – 14.0 kHz; band 6, 14.0 – 22.0 kHz); and (iii) azimuth directions of sound cues (0^o, 45^o, 135^o, 180^o, 225^o, 315^o) as shown in Figure 7.1. Directions 90^o and 270^o were not considered as front-back confusions could not be counted. Dependent variables used were similar to Experiment one: (i) localization error; and (ii) percentage of the occurrence of front-back confusions.

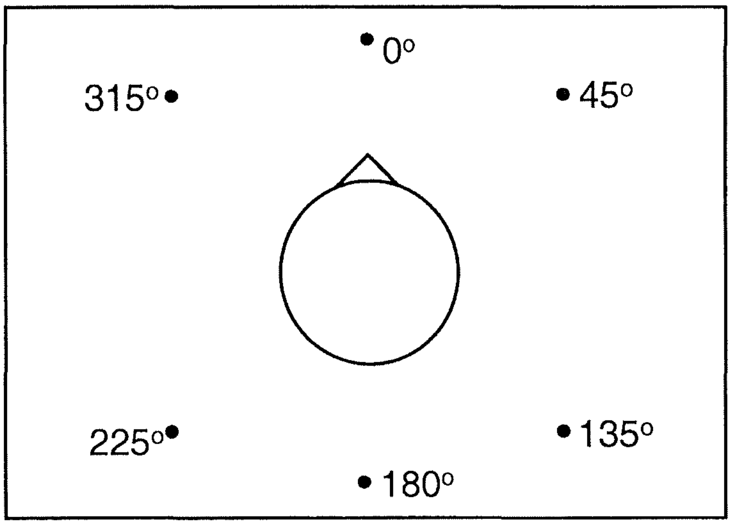


FIGURE 7.1 The 6 azimuth directions of sound cues that studied in Experiment three.

The reasons for choosing those three levels of enhancements were based on the results obtained from a small-scale preliminary test with six participants involved. The levels of enhancements used were 0dB, 6dB, 12dB and 18dB. The detailed descriptions of sound cue preparation and experimental procedures could be found in latter sections 7.5 and 7.6. Figure 7.2 shows the median localization errors as functions of 6 directions and 4 levels of enhancements at 6 frequency bands.

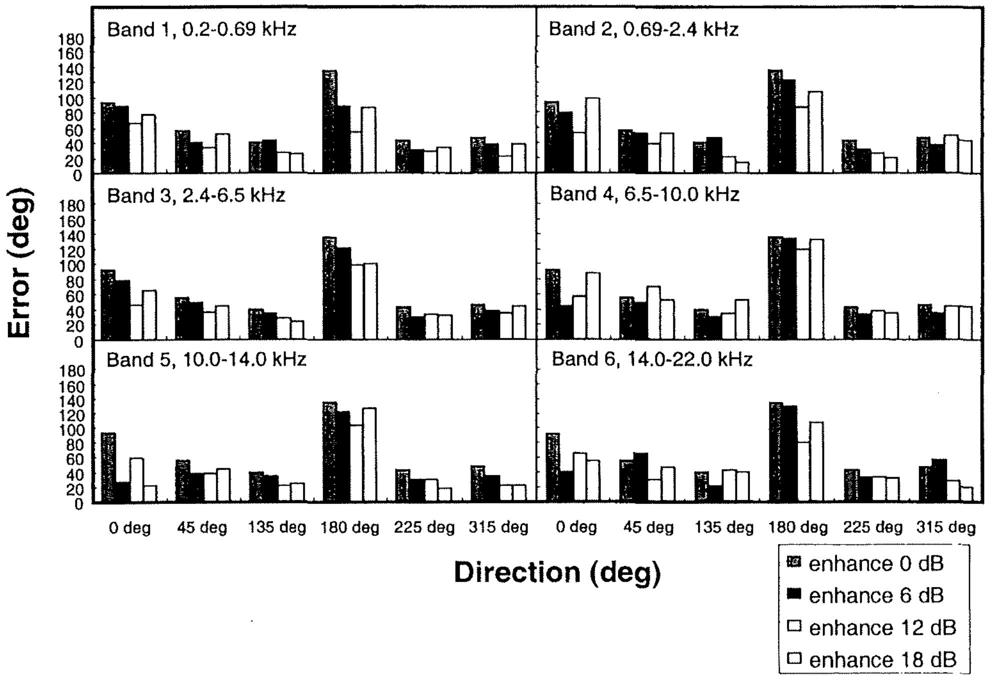


FIGURE 7.2 Median localization errors for 6 sound directions and 4 levels of enhancements at 6 frequency bands in the preliminary test.

It was easy to observe that the pattern for the localization errors among different levels of enhancements could be found from the graphs. For 6dB, 12dB and 18dB enhancement, the localization errors could be reduced. HRTF filters with 12dB enhancement were found to have the highest improvement in localization errors for most of the conditions. When frequency bands were enhanced further to 18dB, improvement in localization errors was reduced when compared with those of 12dB enhancement. Considering the limitation of resources and fatigue effect, filters with 6 dB enhancement were not considered in this study as 12dB enhancement showed a higher improvement in localization errors in most of the cases.

7.5 Stimuli preparation

The MIT KEMAR non-individualized HRTF data published by Gardner and Martin (1995) were used in this experiment to increase the general validity for the results. The original stimulus used in this experiment consisted of a clip of music together with a male ("Please indicate your perceived direction on the interface") and a female voice ("Please pay attention to the perceived direction of this clip") spoken in Cantonese. The clip lasted for 12 seconds and all the binaural directional sound cues were obtained by passing the clip through the corresponding HRTF filters.

Those enhancements of peaks and notches would be applied to the frequency domain of the MIT HRTF data. The whole frequency range (i.e. 22.0 kHz) was divided into 6 frequency bands (band 1, 0.2 – 0.69 kHz; band 2, 0.69 – 2.4 kHz; band 3, 2.4 – 6.5 kHz; band 4, 6.5 – 10.0 kHz; band 5, 10.0 – 14.0 kHz; band 6, 14.0 – 22.0 kHz) since past literature showed that those peaks and notches contributed the frontal and backward cues in each frequency band (Blauert, 1969, 1983; Hebrank and Wright, 1974; Musicant and Butler, 1983; Bronkhorst, 1995; Myers, 1989; Tan and Gan, 1998). For frontal directions 0^{o}, 45^{o}, 315^{o}, peaks in bands 1, 3, 6, together with notches in bands 2, 4, 5 indicated the frontal perception. For backward directions 135^{o}, 180^{o}, 225^{o}, peaks in bands 2, 4, 5 together with notches in bands 1, 3, 6 indicated the backward perception, opposite to those of frontal directions. According to those literature, the peaks and notches in each frequency band of the MIT data were defined according to the following rules (Figure 7.3): (i)

The amplitude of peaks or notches must be at least 0.5 dB within a full range of 4.0 kHz or the full range of the frequency band whichever is smaller for a peak or notch as the just-noticeable level difference was around 0.3 dB and 0.7 dB (Zwicker and Fastl, 1990); (ii) peaks or notches defined in past literature were used instead if no peaks or notches found within the frequency band; and (iii) the boundary of frequency band limited the width of a peak or notch (indicated in Figure 7.4 by a dotted vertical line).

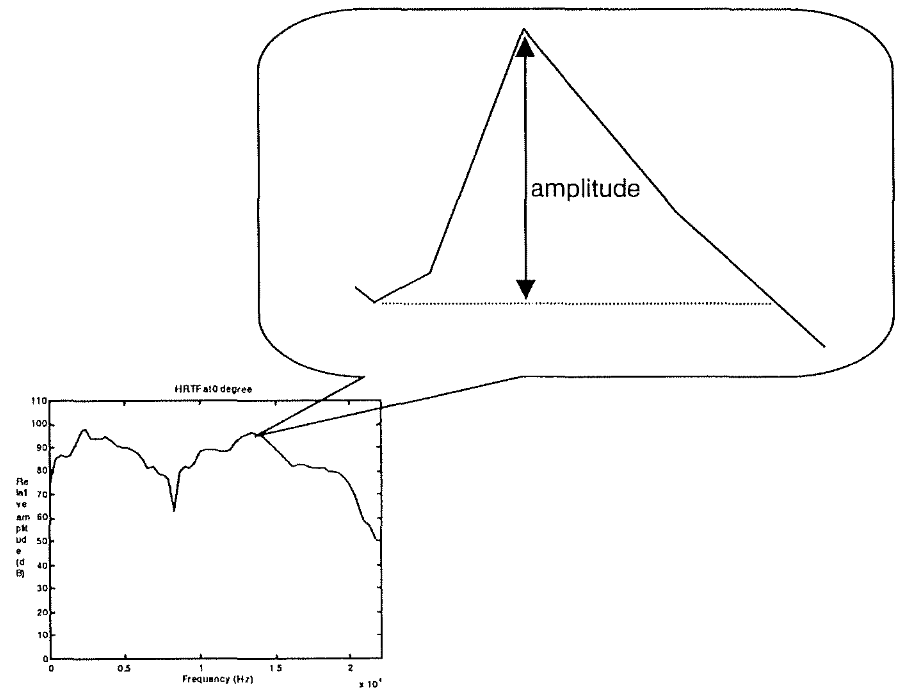


FIGURE 7.3 An illustration of a peak found in HRTF filters.

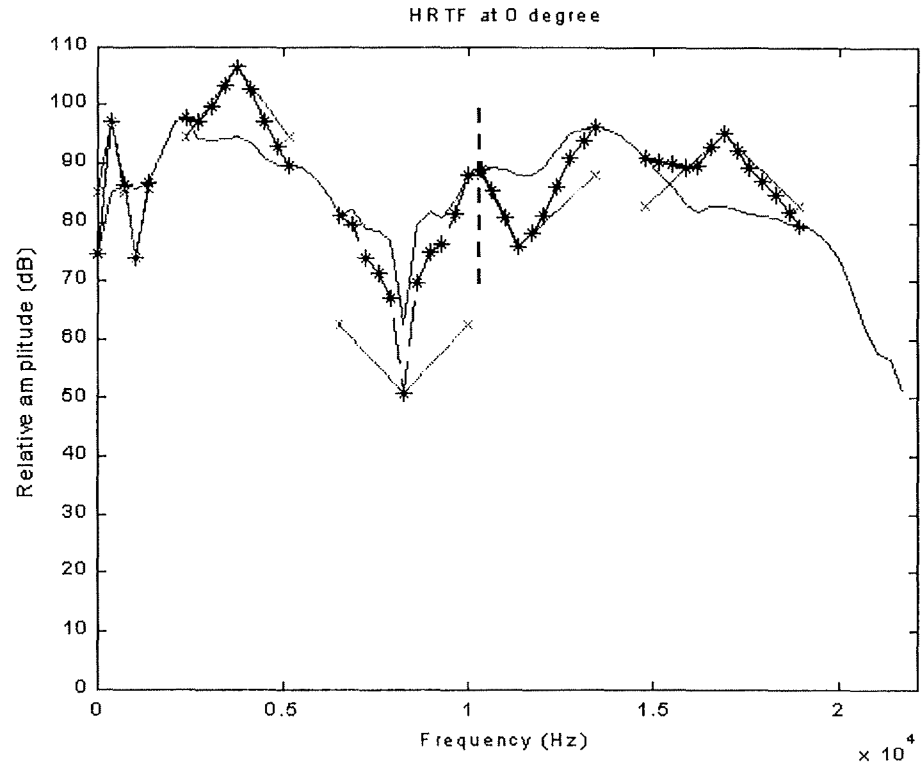


FIGURE 7.4 Enhancements of peaks and notches in the HRTFs at 0° direction ('-' original HRTF; '-x-' triangular function; '-*-' enhanced peaks or notches; and '---' bounded by boundary of frequency band).

These enhanced HRTF filters were convoluted with the original stimulus to generate different sound cues. There are three localization cues found in HRTFs that determine the sound localization: inter-aural time difference (ITD), inter-aural level difference (ILD) and spectral cues. While manipulating the spectral cues for the HRTF filters, ITD and ILD were kept unchanged.

The experiment used 108 pairs of HRTF filters. Each pair of HRTF filters represented the HRTF filters for the left and right ears of a binaural directional sound cue. The 108 pairs exhausted the combinations of 3 levels of enhancements (0dB, 12dB, 18dB), 6 frequency bands (0.2 – 0.69 kHz, 0.69 – 2.4 kHz, 2.4 – 6.5 kHz, 6.5 – 10.0 kHz, 10.0 – 14.0 kHz, 14.0 – 22.0 kHz), and 6 azimuth directions (0^o, 45^o, 135^o, 180^o, 225^o, 315^o). Totally 216 (3 levels of enhancements * 6 frequency bands * 6 sound directions * 2 repetitions) stimuli were listened to by each participant.

7.6 Procedure

Each of the 32 participants needed to listen to all the 108 binaural cues twice. In order to prevent participants from the effect of fatigue when listening to the 216 stimuli during the experiment, the entire experiment was divided into two days. Total 108 stimuli (first repetition) were tested in the first day and the remaining 108 stimuli (second repetition) were tested for the second day. Within the 108 stimuli, the participants were given a five-minute rest after listening to 36 stimuli. After the rest, the experiment was repeated again. The order of presenting the stimuli was randomized. Each stimulus was presented through a pair of headphones (HD 545, Sennheiser, Hannover, Germany) to the participants. After listening to each stimulus, the participants were then required to move the mouse and click on a diagram presented on the PC monitor (see Figure 4) using the diagram. Participants were instructed to move the mouse to a position that best reflected (i) the direction of the sound in relation to the center of the head; and (ii) whether the sound appeared to come from inside or outside of the head. The participants received no feedback on their performance during the entire experiment. The experiment was conduced inside an acoustic chamber with a background noise level of about 35dBA.

7.7 Results and Discussion

The localization error data did not follow normal distribution and non-parametric statistical tests were used *(p<0.001)*. Results of the Wilcoxon Signed Ranks Test showed that there was no significant difference between data in the two replications *(p=0.16)*. As a consequence, data from the two replications were combined in subsequent analyses. Figures 7.5 and 7.6 show the median localization errors and the percentages of the occurrence of front-back confusions as functions of the 6 sound directions and 3 levels of enhancements for frequency bands 1 to 6.

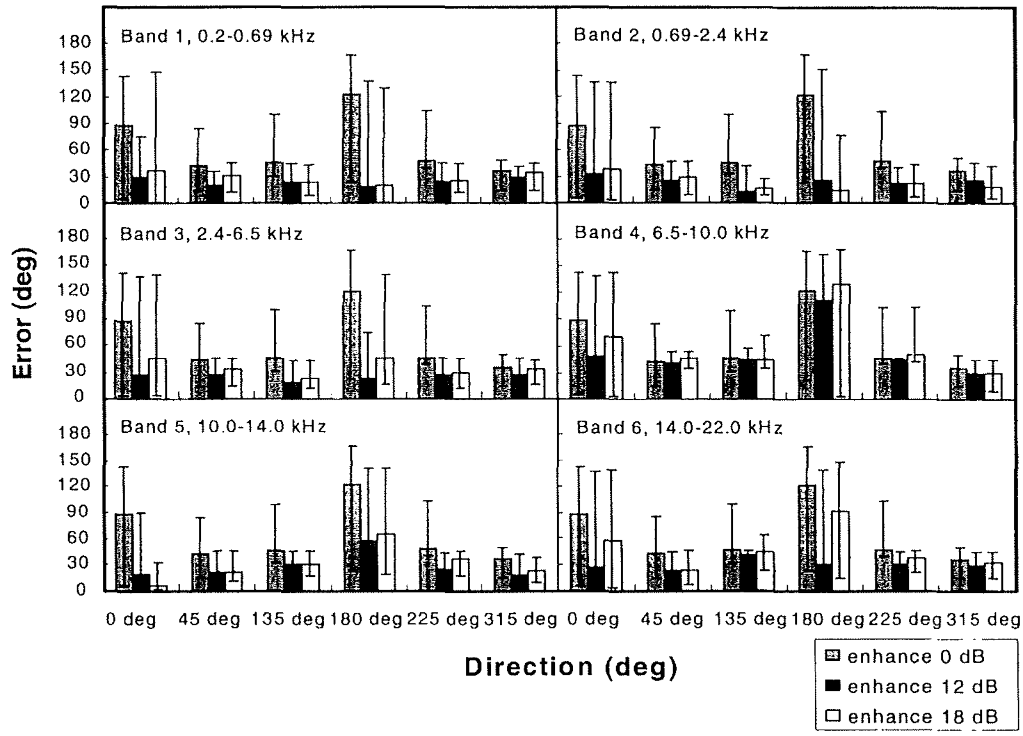


FIGURE 7.5 Median localization errors for 6 sound directions and 3 levels of enhancements at 6 frequency bands.

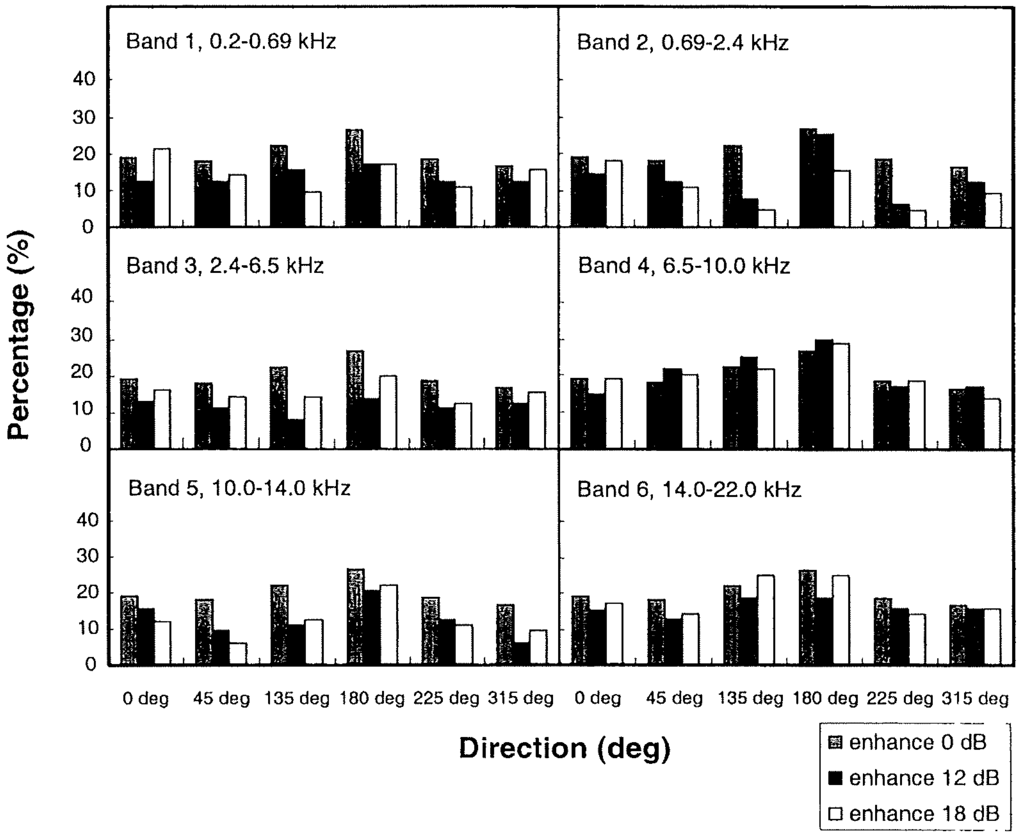


FIGURE 7.6 Percentages of the occurrence of front-back (back-front) confusions for 6 sound directions and 3 levels of enhancements at 6 frequency bands.

Inspections of Figures 7.5 and 7.6 indicate that, in general, the condition with 0dB enhancement (i.e. without enhancement) had the highest localization errors and percentages of the occurrence of front-back confusions while the condition with 12dB enhancement had the lowest localization errors and percentages of the occurrence of front-back confusions. Table 7.1 indicates that the significant levels of the main effects of enhancements (0dB, 12dB, 18dB), directions of cues (0°, 45°, 135°, 180°, 225°, 315°), and frequency bands at which the enhancement was done (0.2 – 0.69 kHz, 0.69 – 2.4 kHz, 2.4 – 6.5 kHz, 6.5 – 10.0 kHz, 10.0 – 14.0 kHz, 14.0 – 22.0 kHz).

TABLE 7.1 Results of Kruskal-Willis tests indicating the significant levels of the main effects of the 3 independent variables on sound localization errors and percentages of the occurrence of front-back confusions.

Source	df	Localization errors		Percentage of occurrence of front-back confusions	
		χ^2 value	p value	χ^2 value	p value
Enhancement	2	295.047	0.001	7.708	0.021
Cue direction	5	180.983	0.001	2.798	0.731
Frequency band	5	81.455	0.001	9.481	0.091

Results of Kruskal-Wallis tests indicate that all three main effects on the localization errors were significant. In some conditions, the participants perceived the sound cues as located at the center of their heads and resulted in missing data in terms of localization errors and occurrence of front-back confusions. This can explain why not all three main effects on the percentage of the occurrence of front-back confusions were significant. Missing data tended to affect the percentages of the occurrence of front-back confusions more than the localization errors as the occurrence of front-back confusions was sensitive to the number of trials. Having a significant main effect of cue direction is not new, but the significant main effects of enhancement and frequency bands are interesting and new. Although the effects of enhancing the spectra of non-individualized HRTFs have been studied before (e.g. Tan and Gan, 1998), statistical test was not used. Mann-Whitney tests were conducted to further compare the effects of individual levels of enhancements on localization errors for each combination of cue directions and frequency bands (Table 7.2).

TABLE 7.2 Results of Mann-Whitney tests comparing the differences between individual levels of enhancements on sound localization errors for each combination of cue directions (6 levels) and frequency bands (6 levels). The format of illustration follows that of a SNK test: for each combination of conditions, median errors of the 3 levels of enhancement (0dB, 12dB, 18dB) are listed in descending order and the significant levels of their differences are indicated by groupings (1-3), levels belonging different groupings are significant different at $p<0.05$ level.

Directions of cues	**Frequency bands at which the enhancements take place**					
	0.2 - 0.69 kHz	**0.69 - 2.4 kHz**	**2.4 - 6.5 kHz**	**6.5 - 10.0 kHz**	**10.0 - 14.0 kHz**	**14.0 - 22.0 kHz**
0	0dB (88°) 1 18dB (36°) 1 2 12dB (29°) 2	0dB (88°) 1 18dB (39°) 1 12dB (33°) 1	0dB (88°) 1 18dB (46°) 1 2 12dB (27°) 2	0dB (88°) 1 18dB (71°) 1 12dB (48°) 1	0dB (88°) 1 12dB (19°) 2 18dB (5°) 2	0dB (88°) 1 18dB (57°) 1 2 12dB (27°) 2
45	0dB (43°) 1 18dB (32°) 2 12dB (21°) 2	0dB (43°) 1 18dB (29°) 2 12dB (26°) 2	0dB (43°) 1 18dB (33°) 2 12dB (28°) 2	18dB (46°) 1 0dB (43°) 1 12dB (41°) 1	0dB (43°) 1 18dB (22°) 2 12dB (21°) 2	0dB (43°) 1 18dB (24°) 2 12dB (24°) 2
135	0dB (46°) 1 18dB (24°) 2 12dB (23°) 2	0dB (46°) 1 18dB (17°) 2 12dB (13°) 2	0dB (46°) 1 18dB (24°) 2 12dB (19°) 2	0dB (46°) 1 18dB (45°) 1 12dB (45°) 1	0dB (46°) 1 18dB (30°) 2 12dB (30°) 2	0dB (46°) 1 18dB (44°) 2 12dB (42°) 2
180	0dB (122°) 1 18dB (20°) 2 12dB (19°) 2	0dB (122°) 1 12dB (25°) 2 18dB (14°) 2	0dB (122°) 1 18dB (46°) 1 2 12dB (24°) 2	18dB (130°) 1 0dB (122°) 1 12dB (111°) 1	0dB (122°) 1 18dB (65°) 1 2 12dB (58°) 2	0dB (122°) 1 18dB (91°) 1 2 12dB (31°) 2
225	0dB (47°) 1 18dB (26°) 2 12dB (23°) 2	0dB (47°) 1 18dB (22°) 2 12dB (22°) 2	0dB (47°) 1 18dB (30°) 2 12dB (28°) 2	18dB (50°) 1 0dB (47°) 1 12dB (44°) 2	0dB (47°) 1 18dB (37°) 2 12dB (24°) 2	0dB (47°) 1 18dB (37°) 2 12dB (31°) 2
315	0dB (36°) 1 18dB (35°) 1 12dB (30°) 1	0dB (36°) 1 12dB (25°) 1 2 18dB (18°) 2	0dB (36°) 1 18dB (33°) 1 12dB (27°) 1	0dB (36°) 1 18dB (30°) 1 12dB (30°) 1	0dB (36°) 1 18dB (23°) 2 12dB (18°) 2	0dB (36°) 1 18dB (32°) 1 12dB (29°) 1

From Table 7.2, it can be observed that HRTF filters with 12dB or 18dB enhancement at one of the following five frequency bands (0.2 – 0.69 kHz, 0.69 – 2.4 kHz, 2.4 – 6.5 kHz, 10.0 – 14.0 kHz, 14.0 – 22.0 kHz) produced significantly lower localization errors than those filters with no enhancement for directions 0°, 45°, 135°, 180° and 225° *($p<0.05$)*. Enhancements at frequency band between 6.5 kHz and 10.0 kHz (i.e. frequency band 4) did not affect the localization errors *($p>0.1$)* except for 225° direction. In general, HRTF filters with 12 dB enhancement had significantly lower errors than those filters with 18dB enhancement in most of the combinations but the differences were not significant *($p>0.05$)*.

According to Blauert's (1969) study, he identified the locations of frontal and backward cues, together with elevation cues, along the whole frequency band. An elevation cue was identified to situate at center frequency of 8.0 kHz with center bandwidth of 4.0 kHz, which was nearly the same as the frequency band 4 (6.5 – 10.0 kHz) in this experiment. This explained why there was no improvement for the localization errors when enhancing the spectra of frequency band 4. Inspections of Table 7.2 indicate that HRTFs with 12dB or 18dB enhancement significantly reduced the localization errors than those with 0dB enhancement at only frequency band 2 (0.69 – 2.4 kHz) and 5 (10.0 – 14.0 kHz) for 315° direction *($p<0.005$)*. This was different from the study of Tan and Gan (1998) which suggested that enhancing all the frontal and backward cues at frequency bands 1 to 6 should have localization improvement.

Considering the percentages of the occurrence of front-back confusions, HRTFs with 12dB or 18dB enhancement at the five frequency bands were associated with significantly lower percentage of the occurrence of front-back confusions *($p<0.05$)* than those cues with no enhancement. However, effect on cue direction could not be examined as data of all directions were grouped to analyze the percentage of the occurrence of front-back confusions. Improvements between participants with and without musical training background were also examined. Results of Wilcoxon Signed Rank show that there was no significant in the localization errors and the percentages of the occurrence of front-back confusions between participants with and without musical training background *($p>0.1$)*. Participants with musical training background may not have much advantage for the improvement in localization errors when the spectra of HRTFs were enhanced.

One of the reasons of choosing the levels of enhancements for the peaks and notches to be 0dB, 12dB and 18dB was that the sound quality for the stimuli prepared using HRTFs with enhanced frontal and backward cues of more than 18dB became awful. It was believed that the sound quality of the synthesized stimuli would be affected if enhancements of spectral cues exceeded a certain limit.

Spectral cues were formed due to the filtering effects of outer ear. Dimensions of outer ears were expected to have close relationship on sound localization errors with binaural cues generated by non-individualized HRTFs

(Middlebrooks, 1999a). It would provide important information on improvements of errors if localization errors or improved errors (i.e. difference in localization errors between 0dB and 12dB enhancement) were found to have correlations with dimensions of outer ear. Improved errors for 18dB enhancement were not considered as previous results showed that reduction of localization errors with HRTF filters with 12dB enhancement was generally higher than those filters with 18dB enhancement. It would be beneficial if the physical sizes (e.g. height) of participants correlated with the localization errors or improved errors. Five dimensions of outer ear were measured for each participant according to the study of Middlebrooks (1999a). Figure 7.7 shows the physical structures of an outer ear that involved in the five measured dimensions.

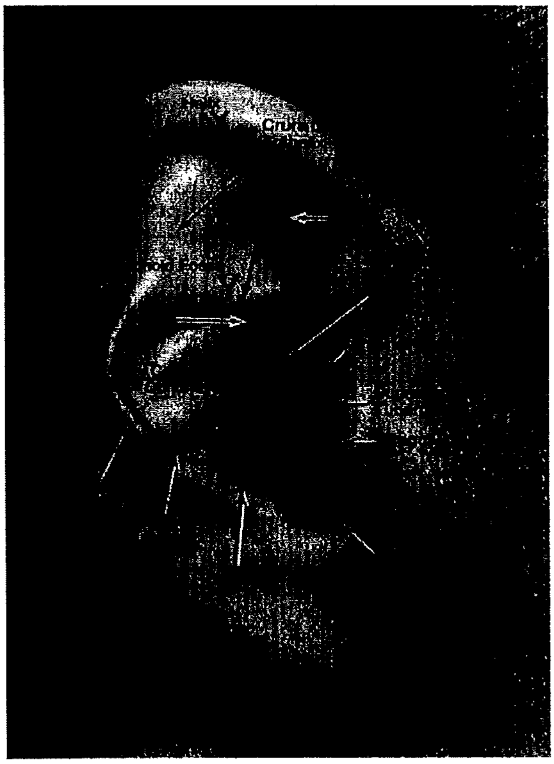


FIGURE 7.7 Illustration the physical structures that involved in the five measured dimensions of an outer ear (adopted by Schuknecht and Gulya, 1986).

The measurements consisted of (i) intertragal incisure to the rim of the helix – pinna-cavity height (dimension 1, I to X); (ii) intertragal incisure to the crura of antihelix (dimensions 2, I to Y); (iii) intertragal incisure to the crus of helix (dimension 3, I to Z); tragus to antitragus (dimension 4, T to U); and (v) tragus to inner surface of the back of concha – pinna-cavity breath (dimension 5, T to V). Digital photographs of both ears have been taken for each participant. The ear dimensions and photographs of each participant could be found in Appendix A7.4. Table 7.1 summarizes correlations of outer ear and height dimensions with localization errors and improved errors. Dimensions correlated with both left ears and right ears simultaneously were listed in Table 7.3.

TABLE 7.3 Summary on correlations of outer ear and height dimensions with localization errors of 0dB enhancement (e.g. 'no_enhance_0', no enhancement for 0^o direction) and improved errors (e.g. 'band1_0_improve', reduction of localization errors for 0^o direction with frequency band 1 enhanced).

Dimension	In correlation with
1, left ear	height (0.601**), band6_315_improve (-0.455**)
1, right ear	height (0.485**), band6_315_improve (0.432*)
2, left ear	-
2, right ear	-
3, left ear	band3_315_improve (-0.384*), band5_315_improve (-0.366*)
3, right ear	band3_315_improve (-0.386*), band5_315_improve (-0.399*)
4, left ear	band3_315_improve (-0.385*), band5_315_improve (-0.427*)
4, right ear	band3_315_improve (-0.381*), band5_315_improve (-0.397*)
5, left ear	band2_0_improve (-0.372*), band2_45_improve (0.401*)
5, right ear	band2_0_improve (0.372*), band2_45_improve (0.401*)
6, height	dimension1_left (0.601**), dimension1_right (0.485**)

Inspections of Table 7.3 show that pinna-cavity height (dimension 1) correlated with the height *($p<0.005$)*. The height dimension did not have any correlations with localization errors and improved errors. Improved errors of some frequency bands in certain directions had correlations with some of ear dimensions. Further works with a larger sample size are desirable.

7.8 Summary

Binaural sound cues simulated by HRTF filters with 12dB or 18dB enhancement at one of the following five frequency bands (0.2 – 0.69 kHz, 0.69 – 2.4 kHz, 2.4 – 6.5 kHz, 10.0 – 14.0 kHz, 14.0 – 22.0 kHz) were associated with significantly lower localization errors than those filters with no enhancement for directions 0°, 45°, 135°, 180° and 225°. Enhancements at frequency band between 6.5 kHz and 10.0 kHz (i.e. frequency band 4) were unable to reduce the localization errors except for 225° direction. For 315° direction, only enhancement at frequency bands 2 and 5 significantly reduced the localization errors. Generally, HRTF filters with 12dB enhancement had significantly lower errors than those filters with 18dB enhancement in most of the combinations but the differences were not significant. Also, improvements in localization errors when enhancing the spectra at the six frequency bands had no significant difference between participants with and without musical training background.

CHAPTER 8 FINAL CONCLUSIONS, DISCUSSION AND RECOMMENDATIONS

8.1 Summary of experimental findings

In this thesis, results of three experiments are presented. The first experiment was conducted to investigate the effects of simplifying the spectra of non-individualized HRTFs on their ability to produce accurate binaural directional sound cues. A usability experiment (i.e. Experiment two) was then performed to test the performance of a Virtual Headphone-based Surround Sound (VHSS) System using the optimized set of HRTF filter coefficients obtained from the first experiment. Experiment three was performed to investigate the use of manipulating the spectra of non-individualized HRTFs at different frequencies to further improve the localization performance associated with binaural directional cues generated with non-individualized HRTFs. The specific findings of the three experiments are summarized as follows:

(i) Simplifying the spectra of HRTFs significantly affected the localization errors of the corresponding binaural sound cues set up to simulate the azimuth directions of 45° *($p<0.001$)*, 135° *($p<0.001$)*, 180° *($p<0.05$)*, 225° *($p<0.001$)* and 315° *($p<0.001$)*.

(ii) HRTF filters of 18 coefficients produced significantly higher localization errors than the corresponding filters of 128, 64 and 32 coefficients for directions 135° *($p<0.005$)* and 315° *($p<0.05$)*. By observation, HRTF filters of 18 coefficients generally have higher localization errors than the corresponding filters of 128, 64 and 32 coefficients for directions 0°, 45°, 225° and 270°.

(iii) Localization errors associated with HRTF filters of 32 coefficients were not significantly different from those errors associated with HRTF filters of 128 and 64 coefficients for directions 0° *($p=0.54$)*, 90° *($p=0.513$)* and 270° *($p=0.112$)*.

(iv) For directions 45° and 315°, HRTF filters of 64 coefficients had significantly lower localization errors than the corresponding filters of 32 and 18 coefficients *($p<0.01$)*. When compared with 128 coefficients, no significant difference in errors was found *($p>0.5$)*.

(v) When the localization errors were examined to exclude errors due to front-back confusions, simplifying the spectra of HRTFs was found to have no significant influence on the localization errors with front-back confusions excluded *($p>0.1$)* expect for 225° direction. This suggests that simplifying the spectra of HRTFs will affect the occurrence of front-back confusions.

(vi) Simplifying the spectra of HRTFs had significant effects on the percentages of the occurrence of front-back confusions *($p<0.001$)*. HRTF filters of 128 and 64 coefficients produced significantly lower percentages of the occurrence of front-back confusions when compared with to those filters of 32 coefficients *($p<0.01$)* which, in turn, produced significantly lower errors than those filters of 18 coefficients *($p<0.05$)*.

(vii) The VHSS system using the optimized set of HRTF filters as recommended in the results of Experiment one produced significantly higher surround sound ratings than the Dolby™ stereo channels in 18 out of 19 questions *($p<0.05$)*.

(viii) There was no significant difference in the surround sound ratings between participants with and without musical training background *($p=0.127$)*.

(ix) For directions 0°, 45°, 135°, 180° and 225°, binaural directional cues generated by HRTFs whose spectra were enhanced (i.e. amplify the peaks and notches by 12dB or 18dB) at one of the following five frequency bands (0.2 – 0.69 kHz, 0.69 – 2.4 kHz, 2.4 – 6.5 kHz, 10.0 – 14.0 kHz, 14.0 – 22.0 kHz) were associated with significantly lower localization errors *($p<0.05$)* than those cues with no enhancement. Enhancement at frequency band at between 6.5 kHz and 10.0 kHz (i.e. band 4) did not affect the localization errors *($p>0.1$)* except for 225° direction. Considering the percentage of the occurrence of front-back confusions, HRTFs whose spectra were enhanced by 12dB or 18dB at the five frequency bands were associated with significantly lower localization errors *($p<0.05$)* than those cues with no enhancement.

(x) Generally, enhancement of 12dB produced lower errors than 18dB but the differences were not significant *($p>0.05$)*.

(xi) There was no significant difference in the localization errors and the percentages of the occurrence of front-back confusions between participants with and without musical training background *($p>0.1$)*.

8.2 Implications of experimental results

8.2.1 Limitations of the laboratory findings

For Experiments one and three where the participants had to listen to 72 and 216 binaural cue conditions respectively, it was unavoidable that some participants would feel fatigue and bored. Rest periods were inserted once the participants listened to 36 sound cues and in Experiment three, the experiment was divided into two one-hour sessions on consecutive days. In Experiment three, the interactions among spectral enhancements at two or more frequency bands were not studied in order to limit the already large number of conditions.

For the measurements of ear dimensions, the measurements were not precise enough since all the measurements were based on the photographs of participants' ears that were taken during the experiment. In order to obtain the most accurate ear dimensions, a portable 3D scanner is desirable.

For the three experiments reported in the thesis, all participants were students or staff at the university. Although the author does not expect the ability to localize sound will be different among different population groups (e.g. house wives, teenagers), it would be desirable to repeat Experiment two with listeners who have brought surround sound systems or intended to buy surround sound systems as well as retailers of surround sound systems.

8.2.2 An optimized set of non-individualized HRTF filters for the VHSS system

Experiment one in this study aimed at simplifying the spectra of the non-individualized HRTFs in order to simplify the complexity of the HRTF filters (i.e. the number of filter coefficients). This is important to the industry as the amount of devoted calculating power (measured in Millions Instructions Per Second, MIPS) of the DSP used in the VHSS system can be saved and utilized for other purposes. Also, using a DSP with less calculating power can reduce the cost of producing the VHSS system (e.g. DSPs with 40 MIPS, 80 MIPS, 160 MIPS typically cost around US$20, US$60, US$100 respectively – from website of Analog Devices Inc., http://www.analog.com). The results of Experiment one suggest that in implementing a VHSS system, the optimized number of coefficients to be used in HRTF filters for simulating the center, left, right, left rear and right rear channels are 32, 64, 64, 32 and 32 coefficients respectively. In the first experiment, it was found that front-back confusions were a major source of localization errors. Experiment three was conducted to investigate the use of enhancing certain peaks and notches in the spectra of HRTFs to reduce the occurrence of front-back confusions. These peaks and notches occur at six different frequencies and were previously reported to contain cues for determining whether a sound is from the front or at the back.

Table 8.1 lists out the results of Mann-Whitney tests on examining the significant difference in localization errors between HRTF filters with 0dB and

12dB enhancement (or 18dB enhancement whichever is better) for each frequency band at each cue direction.

TABLE 8.1 The levels of enhancements that can significantly reduce sound localization errors at each of the 36 combination of cue directions and frequencies of enhancements. Significant levels have been taken from Mann-Whitney tests as shown in Table 7.2).

Direction	Frequency band					
	band 1 (0.2 – 0.69 kHz)	band 2 (0.69 – 2.4 kHz)	band 3 (2.4 – 6.5 kHz)	band 4 (6.5 – 10.0 kHz)	band 5 (10.0 – 14.0 kHz)	band 6 (14.0 – 22.0 kHz)
0°	12dB**	-	12dB*	-	12dB**, 18dB**	12dB*
45°	12dB**, 18dB**	12dB**, 18dB**	12dB*, 18dB*	-	12dB**, 18dB**	12dB**, 18dB**
135°	12dB**, 18dB**	12dB**, 18dB**	12dB**, 18dB**	-	12dB**, 18dB**	12dB**, 18dB*
180°	12dB**, 18dB**	12dB*, 18dB**	12dB**	-	12dB*	12dB*
225°	12dB**, 18dB**	12dB**, 18dB**	12dB**, 18dB**	12dB**	12dB**, 18dB**	12dB**, 18dB**
315°	-	18dB**	-	-	12dB**, 18dB**	-

Inspections of Table 8.1 show that at most enhancements at most frequencies can significantly reduce errors *($p<0.05$)* except enhancement at frequency band 4 (i.e. 6.5 – 10.0 kHz). Enhancement at this band only showed improvements for 225° direction. Also, Table 8.1 showed that only frequency band 2 (0.69 – 2.4 kHz) and band 5 (10.0 – 14.0 kHz) showed improvements for 315° direction. This was different from the study of Tan and Gan (1998) as they suggested that enhancing all the frontal and backward cues at frequency band 1 to band 6 should have effects to reduce the occurrence of front-back confusions. Using the data in Table 8.1, the non-individualized HRTFs from MIT KEMAR data can be enhanced. For example, HRTFs for center (i.e. 0° direction) should be enhanced at frequency bands 1, 3, 5 and 6. After the enhancement, the HRTFs can be re-sampled using methods used in Experiment one to have 32 coefficients (again, according to the results of Experiment one). The result is an optimized HRTFs for center channel in surround sound system. Repeating this for left (315^{o}) and right (45^{o}) but using 64 coefficients will produce an optimized HRTFs for the left and right channels. For the left rear (270^{o}) and right rear (90^{o}) channels, the optimized HRTFs with 32 coefficients as suggested by Experiment one will be used. For these two channels, no spectral enhancement is needed because they should not be associated with front-back confusions.

8.2.3 Towards the possibility of mass-customizing the non-individualized HRTFs

Tables A8.1a to A8.1f in Appendix A8.1 illustrate the occurrence of improvement in localization errors for each of 32 participants in each of the 36 combinations of six frequencies of enhancements and six directions. For each combination by a participant, an improvement occurred when the mean (of repetitions) localization errors reduced after the 12dB enhancement was applied. A '+' sign in Tables A8.1a to A8.1f indicates that there was an improvement of errors with more than 2^{o}. Gardner (1968) and Blauert (1970b) reported that the minimum detectable change in angular position of a sound source broadcasting speech was between 0.9^{o} and 1.5^{o}. This suggests that an improvement of more than 2^{o} in this study would be detectable. A '-' sign in Tables A8.1a to A8.1f represents an increase in localization errors of more than 2^{o}. A symbol of '0' indicates no improvement (i.e. a difference less than or equal to ±2^{o}). When counting the total occurrence of improvement for each frequency band of a particular direction, the choice of which frequency bands should be enhanced to reduce the localization errors was in alignment with the statistical results shown in Table 8.1. Because of the existence of individual difference, there exist a possibility to categorize listeners into different groups so that a set of optimized HRTFs can be customized for each group of listeners. In practice, this means that the VHSS system will store a few set of HRTF filter coefficients and listeners can choose the best available sets for themselves.

In order to categorize the group more easily, the occurrence of improvement in localization errors for each of 32 participants for frontal directions (0°, 45° and 315°) were summed together. The same summing procedure was applied to the backward directions (135°, 180° and 225°). The rationale behind the summing procedure was that a listener generally showed an improvement in 0° direction as well as 45° and 315° directions as they all belonged to the frontal directions. Tables A8.2a and A8.2b in Appendix A8.2 show the occurrence of improvement in localization errors for each of the 32 participants for the frontal and backward directions. By observation, those top ten participants showing reduction in localization errors and another top ten participants showing increase in localization errors were identified and listed in the Tables 8.2a and 8.2b.

TABLE 8.2a Participants showing net reduction and increase in localization errors after enhancing the six frequency bands (0.2 – 0.69 kHz, 0.69 – 2.4 kHz, 2.4 – 6.5 kHz, 6.5 – 10.0 kHz, 10.0 – 14.0 kHz, 14.0 – 22.0 kHz) for frontal directions (i.e. 0^{o}, 45^{o}, 315^{o}).

Effect on localization error after enhancements at 6 frequency bands	Participants
A net reduction in errors	2, 4*, 5*, 7, 14*, 15*, 28*, 29, 30*, 32
A net increase in errors	8, 11, 12, 18*, 19*, 22*, 23, 24, 25*, 27

* These participants also appear in the appropriate section in Table 8.2b – a sister table for backward directions.

TABLE 8.2b Participants showing net reduction and increase in localization errors after enhancing the six frequency bands (0.2 – 0.69 kHz, 0.69 – 2.4 kHz, 2.4 – 6.5 kHz, 6.5 – 10.0 kHz, 10.0 – 14.0 kHz, 14.0 – 22.0 kHz) for backward directions (i.e. 135^{o}, 180^{o}, 225^{o}).

Effect on localization error after enhancements at 6 frequency bands	Participants
A net reduction in errors	4*, 5*, 8, 14*, 15*, 16, 24, 26, 28*, 30*
A net increase in errors	1, 2, 7, 18*, 19*, 22*, 25*, 29, 32

* These participants also appear in the appropriate section in Table 8.2a – a sister table for frontal directions.

By so doing, it was able to divide all the participants into several categories such that one group of the participants would use the enhanced HRTF filters as improvements in localization errors were observed while other groups might use the original HRTF filters because of the increase in localization errors after using enhanced HRTF filters. From Tables 8.2a and 8.2b, it can be observed that some participants had consistently showed improvements after spectral enhancements for cues at for both frontal and backward directions. Similarly, some participants exhibited increases in errors after spectral enhancements. On the other hands, some participants showed reductions in localization errors for the backward directions but increases in localization errors for the frontal directions or vice versa e.g. participants numbered 7, 8, 24, 29, 32). Future works on the grouping of listeners are desirable.

As stated previously in Chapter 7, dimensions of outer ears were considered to have influences on sound localization errors with binaural cues generated by non-individualized HRTFs (Middlebrooks, 1999a). It would be interesting to see if there was any significant difference on the ear dimensions between those participants who showed reduction and increase in localization errors. Based on the Tables 8.2a and 8.2b, those participants who showed reductions or increases in localization errors for both cues at frontal and backward directions were chosen and listed in Table 8.3.

TABLE 8.3 Participants showing net reduction and increase in localization errors after enhancing the six frequency bands (0.2 – 0.69 kHz, 0.69 – 2.4 kHz, 2.4 – 6.5 kHz, 6.5 – 10.0 kHz, 10.0 – 14.0 kHz, 14.0 – 22.0 kHz) for both frontal and backward directions.

Effect on localization error after enhancements of frequency bands	**Participants**
A net reduction in errors	4, 5, 14, 15, 28, 30
A net increase in errors	18, 19, 22, 25

Figure 8.1 compares the dimensions of outer ears between two groups of participants listed in Table 8.3. The detailed descriptions of the dimensions can be referred to in section 7.7. One group consistently demonstrated reductions in errors after enhancements and the other group consistently demonstrated increases in errors after enhancements. Because of the lack of

data points, no firm conclusion can be drawn in this study. Future works with a larger sample size are desirable.

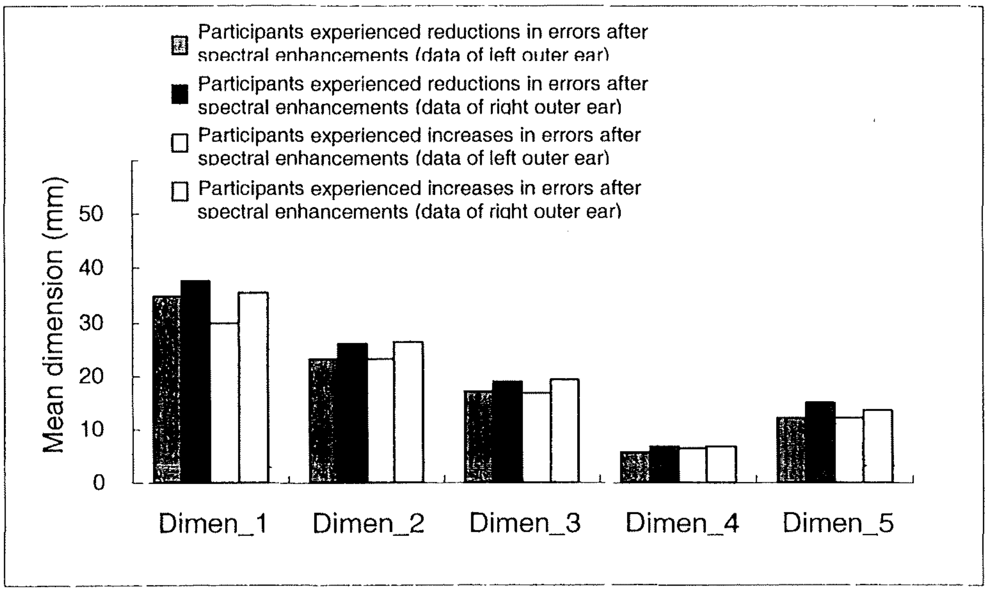


FIGURE 8.1 A comparison of outer ear dimensions between groups of participants exhibited reductions in sound localization with binaural cues of enhanced spectra (mean of 6 participants) and groups of participants who exhibited increases in sound localization errors with binaural cues of enhanced spectra (mean of 4 participants).

8.3 Conclusions

In this study, effects of changing the spectra of HRTF filters on the localization performance were examined. The first experiment studied the effects of simplifying the spectra of non-individualized HRTFs by reducing the number of HRTF filter coefficients from 128 coefficients to 64, 32, 18 coefficients. Results showed that HRTF filters of 32 coefficients did not have significant difference in the localization errors when compared with those filters of 128 and 64 coefficients for all directions except for directions 45° and 315°. For directions 45° and 315°, HRTF filters of 64 coefficients had significantly lower localization errors than the corresponding filters of 32 and 18 coefficients. At the same time, HRTF filters of 18 coefficients, showed comparatively higher localization errors than those filters of 128, 64 and 32 coefficients for most of the directions. Based on these results, HRTF filters of 32 coefficients were used for simulating the center (0^o), left rear (270^o) and right rear (90^o) channels of a Virtual Headphone-based Surround Sound (VHSS) System. Also based on the results of Experiment one, HRTF filters of 64 coefficients were selected for left (315^o) and right (45^o) channels. The usability experiment (i.e. Experiment two) indicated that the virtual surround sound simulated from the VHSS system had significantly higher surround sound ratings than the Dolby™ stereo channels.

To cope with problems of front-back confusions, effects of enhancements of spectra at six frequencies were studied. The spectra of HRTFs at these frequencies carry important information for listeners to determine the front or back directions of binaural cues. It was discovered that, for nearly all directions of cues and nearly all frequencies of enhancements, HRTFs with 12dB and 18dB enhancements produced binaural cues with significantly lower localization errors and percentages of the occurrence of front-back confusions than those cues generated with HRTFs of no enhancements. The exception was with enhancements between 6.5 to 10.0 kHz. There was no significant difference in localization errors and percentages of the occurrence of front-back confusions between the 12dB and 18dB enhancement conditions although, in general, 12dB enhancements were associated with lower errors. Using the results of the experiments, an optimized set of HRTFs for VHSS system can be produced. Lastly, this study combined both fundamental studies on human perception on binaural directional cues as well as the implementations and testings of a practical VHSS system. The author hopes to create the know-how for local consumer audio industries to enhance their competitive strength.

8.4 Future work recommendations

Further studies on investigating the interactive effects of enhancing each sub-band on the localization errors and the percentages of the occurrence of front-back confusions are desirable. This can enrich the understanding on the optimized performance on combined enhancements of those frequency bands.

Further work on categorizing listeners into different groups is recommended. A larger sample size will be collected to increase the general validity of the groupings. More data on the dimensions of outer ears are collected to examine if there is any difference in the outer ear's dimensions between different categorized groups.

REFERENCES

Asano F., Suzuki Y., Sone T. (1990). Role of spectral cues in median plane localization. Journal of the Acoustical Society of America, vol. 88, no. 1, pp.159-68.

Begault D. R. (1992a). Perceptual effects of synthetic reverberation on three-dimensional audio systems. Journal of the Audio Engineering Society, vol. 40, pp. 895-904.

Begault D. R. (1992b). Perceptual similarity of measured and synthetic HRTF filtered speech stimuli. Journal of the Acoustical Society of America, vol. 92, pp.2334.

Begault D.R. (1993b). Head-up Auditory Displays for Traffic Collision Avoidance System Advisories: A Preliminary Investigation. Human Factors, vol. 35, pp.707-17.

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

Begault D. R. (1995). Head-up auditory display research at NASA Ames Research Center. In Proceedings of the 39th Annual Meeting of the Human Factors and Ergonomics Society 1995, pp.114-118.

Berg J., Rumsey F. (1999). Spatial Attribute Identification and scaling by repertory grid technique and other methods. In Proceedings of the AES 16th International Conference on Spatial Sound Reproduction, May, 1999, Munich.

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

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

Blauert J. (1970b). An experiment in directional hearing with simultaneous optical stimulation. Acustica, vol. 23, pp.118-9.

Blauert J. (1983). Spatial Hearing: The Psychophysics of Human Sound Localization. MIT Press, London, England.

Bloom P. J. (1977). Creating source elevation illusions by spectral manipulation. Journal of the Audio Engineering Society, vol. 25, pp.560-65.

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

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

Burgess D.A. (1992). Real-time audio spatilization with inexpensive hardware. (Report No. GIT-GVU-92-20). Graphics Visualization and Usability Center, Georgia Institute of Technology.

Carlile S. (1996). The physical and psychophysical basis of sound localization. In Virtual Auditory Space: Generation and Applications.

Coleman P. D. (1962). Failure to localize the source distance of an unfamiliar sound. Journal of the Acoustical Society of America, vol. 34, pp. 345-46.

Crispien K. and Petrie H. (1993). Providing access to GUIs for blind people using a multimedia system based on spatial audio presentation. In 95th Audio Engineering Society Convention (Preprint No. 3738). New York: Audio Engineering Society.

Dolby Laboratories. (2001). High level description of the Low-Frequency Effects channel. Available URL: http://www.dolby.com/tech/

Fisher H. G., Freedman S. J. (1968). The role of the pinna in auditory localization. Journal of Auditory Research, vol. 8, no. 1, pp.15-26.

Gardner M. B. (1968a). Lateral localization of 0^{o} or near-0^{o}-oriented speech signals in anechoic space. Journal of the Acoustical Society of America, vol. 44, pp.797-803.

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

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

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

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

Kulkarni A., Colburn H. S. (1998). Role of spectral details in sound-source localization. Nature, vol. 396, pp.747-49.

Lord Rayleigh., On our perception of sound direction. Philos. Mag. 13 1994, pp.214-32.

McGregor P., Horn A. G., Todd M. A. (1985). Are familiar sounds ranged more accurately? Perceptual and Motor Skills, vol. 61, pp. 1082.

Mehrgardt S., Mellert V. (1977). Transformation characteristics of the external human ear. Journal of the Acoustical Society of America, vol. 61, no. 6, pp.1567-76.

Middlebrooks J. C., Green. D. M. (1990). Directional dependence of interaural envelope delays. Journal of the Acoustical Society of America, vol. 87, pp.2149-62.

Middlebrooks J. C., Green. D. M. (1991). Sound localization by human listeners. Annual Review of Psychology, vol. 42, pp.135-59.

Middlebrooks J. C. (1999a). Individual differences in external-ear transfer functions reduced by scaling in frequency. Journal of the Acoustical Society of America, vol. 106, no. 3, pt. 1-2, pp.1480-92.

Middlebrooks J. C. (1999b). Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. Journal of the Acoustical Society of America, vol. 106, no. 3, pt. 1-2, pp.1493-510.

Mills A. W. (1972). Auditory localization. In: J. Tobias (Ed.), Foundations of modern auditory theory, vol. 2, Academic Press, New York, pp.301-45.

Moller H., Friis S. M., Boje J. C., Hammershoi D. (1996). Binaural technique: Do we need individual recordings? Journal of the Audio Engineering Society, vol. 44, no. 6, pp.451-69.

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

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

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

Robinson D. W., Whittle L. S. (1960). The Loudness of Directional Sound Fields. Acustica, vol. 10, pp.74-80.

Sakamoto N., Gotoh T., Kimura Y. (1976). On "out-of-head localization" in headphone listening. Journal of the Audio Engineering Society, vol. 24, pp. 710-16.

Schuknecht H. F., Gulya A. J. (1986). Anatomy of the Temporal Bone with Surgical Implications. Lea & Febiger, Philadelphia, USA.

Shaw E. A. G., Teranishi R. (1968). Sound pressure generated in an external-ear replica and real human ears by a nearby point source. Journal of the Acoustical Society of America, vol. 44, no. 1, pp.240-9.

Shaw E. A. G. (1974a). The external ear. In: W. D. Keidel and W. D. Neff (Eds.), Handbook of sensory psychology, vol. 5, Springer-Verlag, Berlin-New York, pp.455-90.

Shaw E. A. G. (1974b). Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. Journal of the Acoustical Society of America, vol. 56, no. 6, pp.1848-61.

Shaw E. A. G. (1975). The external ear: New knowledge. In: S. Daalsgaard (Ed.), Earmolds and associated problems, Scand. Audiol. Suppl. 5, pp.24-48.

Shaw E. A. G. (1980). The acoustics of external ear. In: G. Studebaker and I. Hochberg (Eds.), Acostical factors affecting hearing aid performance, University Parkk Press, Baltimore, MD.

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

Toole F. E. (1970). In-head localization of acoustic images. Journal of the Acoustical Society of America, vol. 48, pp. 943-49.

Watkins, A.J. (1978) Psychoacoustical aspects of synthesized vertical locale cues. Journal of the Acoustical Society of America, vol. 63, pp.1152-1165.

Wenzel E. M., Arruda M., Kistler D. J., Wightman F. L. (1993). Localization using nonindividualized head-related transfer functions. Journal of the Acoustical Society of America, vol. 94, no. 1, pp.111-23.

Wiener F. M. (1947). On the diffraction of a progressive sound wave by the human head. Journal of the Acoustical Society of America, vol. 19, pp.143-146.

Wightman F. L, Kistler D. J. (1989a). Headphone simulation of free-field listening. I. Stimulus synthesis. Journal of the Acoustical Society of America, vol. 85, no. 2, pp.858-67.

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

Wightman F. L., Kistler D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. Journal of the Acoustical Society of America, vol. 91, no. 3, pp.1648-61.

Wightman F.L., Kistler, D.J., Foster S.H. et al. (1995). A comparison of head-related transfer functions measured deep in the ear canal and at the ear canal entrance. Proceedings of 18th ARO Midwinter Meeting. St. Peterburg Florida.

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

GLOSSARY

Front-back confusions – sounds originated in front of a person is localized at the back or vice versa

Individualized HRTFs – set of HRTF filters recorded from a particular person specifically preserving all the information about the inter-aural differences and filtering effects of his outer ears

Inside-the-head locatedness (IHL) – distance of perceived position smaller than the radius of the head (Blauert, 1983)

Localization blur (Minimum audible angle) – the smallest possible change of position of a sound source that produces a just-noticeable change of position perceived by a person (Blauert, 1983)

Non-individualized HRTFs – set of HRTF filters measured from a mannequin head or a particular person and simulated the virtual sources for another person

APPENDIX TO CHAPTER 4

A4.1 Sample of consent form

Consent form

1. Name ______________________________

2. Are you feeling ill in any way? Yes/No

3. Do you suffer from diabetics (糖尿病) or epilepsy (癲癇症)? Yes/No

4. Are you under medical treatment or suffering disability which affects your daily life? Yes/No

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

DECLARATION

I consent to take part in the 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 withdraw declared above.

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 experiment 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__________________________

(When completed this form should be filed in the Exposure Archive).

APPENDIX TO CHAPTER 5

A5.1 Sample of instruction sheets in Experiment one

Instructions for Sound Localization Experiment

Objective:

This experiment is to investigate the performance of subjects on localizing the angular directions of sound stimuli.

Procedure:

Audiometric test

- Please put on the specific headphones and sit straight.
- Hold the button in one of your hands and press down the button when you listen the "beep" sound from the headphones.
- Repeat the trials until end of the testing.

Experiment

- Sit straight in front of the monitor and put on the pair of headphones.
- Press the "start" button on the screen by mouse once you are ready to start the experiment.
- Sound stimuli will be delivered through the headphones (Close your eyes when listening to the sound stimuli).
- Once the sound stimulus stops (Open your eyes), please use the mouse to click ONCE on the "head" (top view) shown on the screen to indicate where you perceive the sound clip comes from (All sound stimuli are considered to come from your ear level, i.e. same horizontal plane as your ear level).

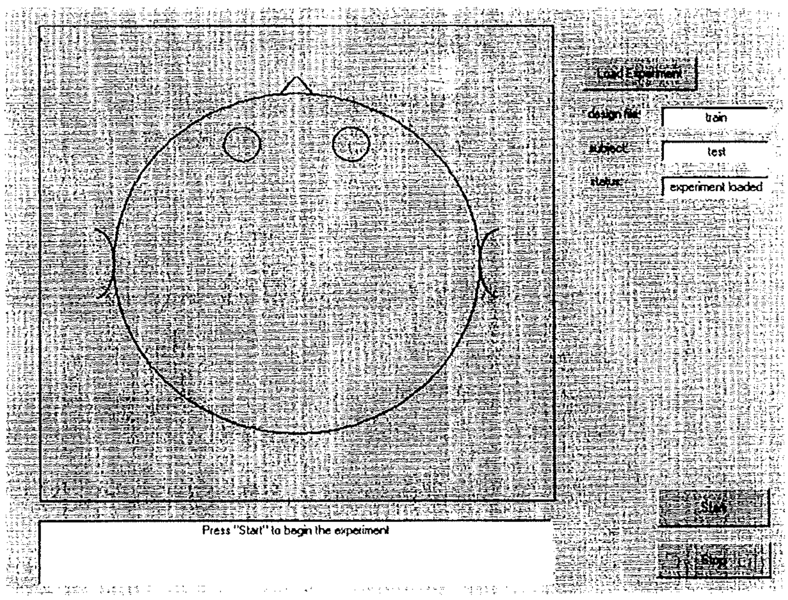


Diagram of Interface during experiment

Note:

1. If you perceive the sound coming from a wider region, please click the centroid of that region.
2. If you REALLY perceive that the sound has its origin inside your head, indicate it on the screen inside the "head". At the same time, if you find that the sound is perceived as coming at the center of your head, please indicate it at the center of "head". On the other hand, please indicate your answers outside the "head" if you perceive it coming from outside your head. Actual distance of perception will not be considered here.
3. You should click your perception only when the sound stops. Please don't click your answer during the delivery of sound.
4. There will be no response from the system if you click your answer on the diagram "line". In this case, please click again on the location nearby.

- Another sound stimulus will be delivered immediately (Close your eyes again) after you clicked the mouse.
- (Open your eyes again) Indicate your perception on the direction of sound it comes from.
- Trials will be repeated following this format.
- Several trials for training purpose will be given for familiarizing this interface (Results are not counted in this case).
- Repeat the same procedures when actual experiment starts until you finish half of the total presented stimuli.
- Have a rest of 5 minutes.
- Repeat the remaining trials until you see the word "End of the experiment" appeared on the "status" item.
- This will be the end for the experiment.

Thank you very much for participating in this experiment!!!
If you have any comments or questions, please contact Ming (ieming@ust.hk).

A5.2 Modules of HRTFs with different coefficients for directions 45°, 90°, 135°, 180°, 225°, 270°, 315°

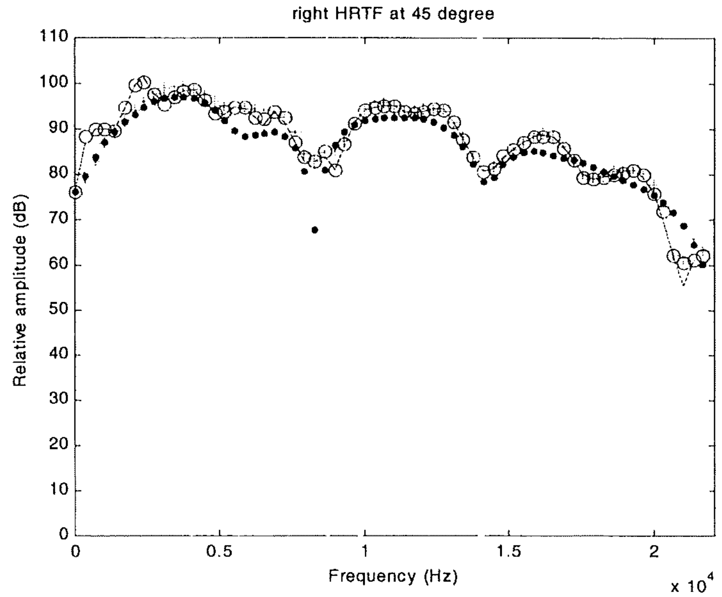


FIGURE A5.1 Modules of the HRTFs form sound cue at 45° direction ('-' 128 coefficients; 'o' 64 coefficients; '+' 32 coefficients; and '...' 18 coefficients).

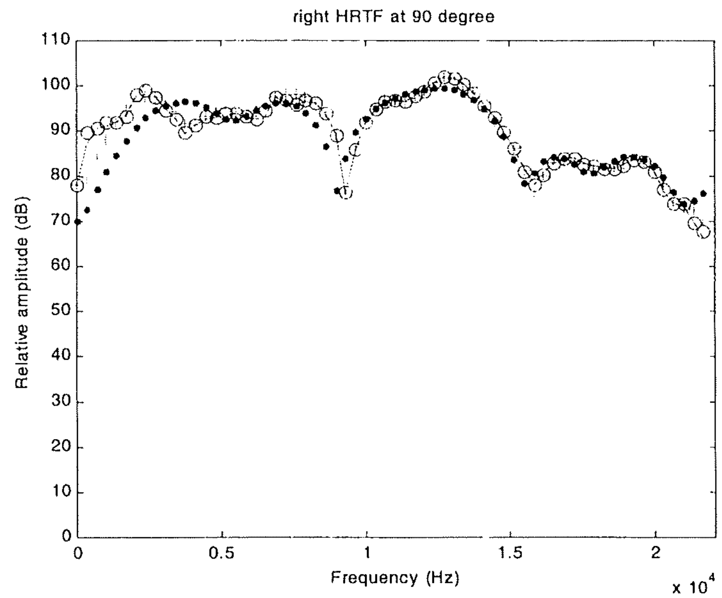


FIGURE A5.2 Modules of the HRTFs form sound cue at 90° direction ('-' 128 coefficients; 'o' 64 coefficients; '+' 32 coefficients; and '...' 18 coefficients).

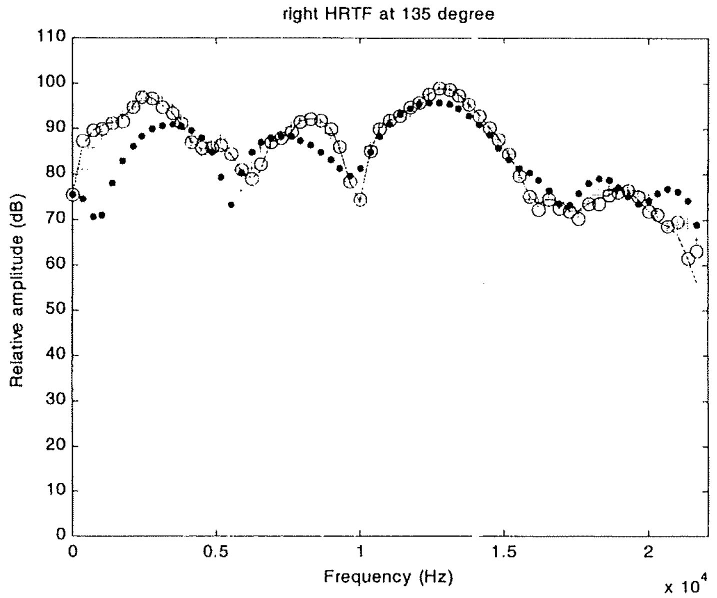


FIGURE A5.5 Modules of the HRTFs form sound cue at 135° direction ('-' 128 coefficients; 'o' 64 coefficients; '+' 32 coefficients; and '...' 18 coefficients).

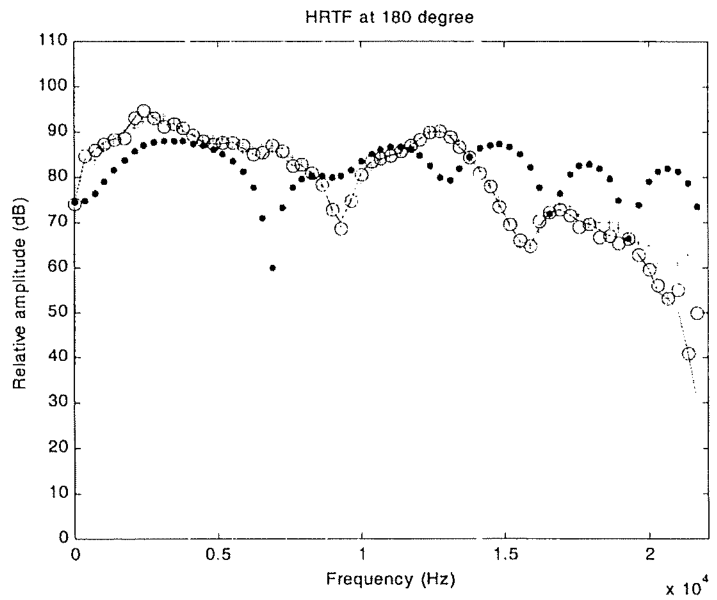


FIGURE A5.4 Modules of the HRTFs form sound cue at 180° direction ('-' 128 coefficients; 'o' 64 coefficients; '+' 32 coefficients; and '...' 18 coefficients).

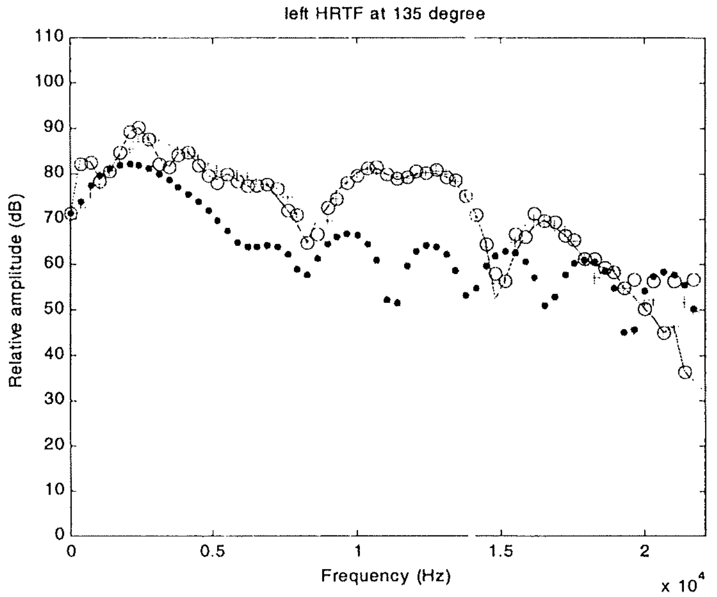


FIGURE A5.5 Modules of the HRTFs form sound cue at 225° direction ('-' 128 coefficients; 'o' 64 coefficients; '+' 32 coefficients; and '...' 18 coefficients).

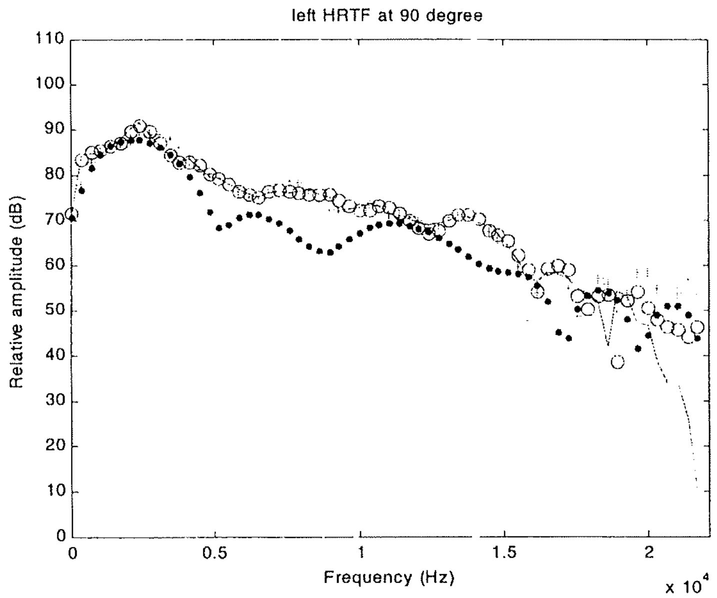


FIGURE A5.6 Modules of the HRTFs form sound cue at 270° direction ('-' 128 coefficients; 'o' 64 coefficients; '+' 32 coefficients; and '...' 18 coefficients).

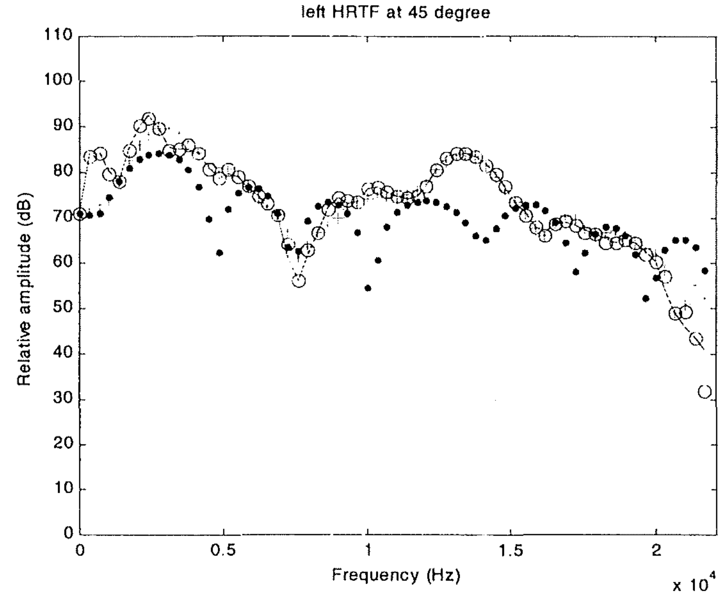


FIGURE A5.7 Modules of the HRTFs form sound cue at 315° direction ('-' 128 coefficients; 'o' 64 coefficients; '+' 32 coefficients; and '...' 18 coefficients).

A5.3 Detailed descriptions of program code to generate HRTFs with different coefficients in Experiment one (using direction 45° as an example)

```
fo=fopen('c:\sound\compact\elev0\H0e045a.dat','r','ieee-be');
```

% 'fo' was the variable of file that used to open the file containing the compact set of data for azimuth direction 45° at 0-degree elevation. 'r' meant opening the file with permission of 'read' function. 'ieee-be' meant that it was a IEEE floating point with big-endian byte ordering.

```
data=fread(fo,256,'short');
```

% This statement was used to read the HRTF data with the size of 256 data (128 data for left ear and 128 data for right ear) from the file variable 'fo' and then write it into the matrix 'data'. 'Short' meant that the data were represented as integers with 16 bits.

```
fclose(fo);
```

% The command 'fclose' was used to close the file after reading the HRTF data.

```
leftpulse=data(1:2:256);
```

% The variable 'leftpulse' was used to store the 128 data for the left ear. As the data for left and right ear were stored in an alternated order in the file, therefore, data of the left ear were read starting from the first data and then skipped one data to store the next data. The storing process stopped until the index reached the value 256.

```
rightpulse=data(2:2:256);
```

% This statement performed like the above-mentioned but for the right ear. The only difference between them was that the data started storing from the second data instead of the first one. All the data for the right ear were stored in the variable 'rightpulse'.

```
leftfft=fft(leftpulse);
rightfft=fft(rightpulse);
```

% These statements were used to perform the Fast Fourier Transform on the impulse response data.

```
leftsel=leftfft(1:4:128);
rightsel=rightfft(1:4:128);
```

% These statements were used to capture one impulse response data for every four of them for the left and right ears until the index reaches the value 128.

```
leftinv=real(ifft(leftsel));
rightinv=real(ifft(rightsel));
```

% Inverse Fast Fourier Transform was performed by using the command 'ifft'. The real part of the 're-sampled' complex number data were extracted for the left and right ears respectively.

```
leftinv(33:53)=leftinv(1:21);
leftinv(1:21)=[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0];
leftinv(54:128)=zeros(75,1);
```

% The above statements were used to rearrange the position of the 're-sampled' impulse response data as the '0' delay changed the normal positioning of the 're-sampled' impulse data after performing the inverse Fast Fourier Transform. Also, the time delay of 21 '0' was needed to include in the data for the left ear and 75 '0' were filled up the end part of the original 128 data space.

```
rightinv(33:37)=rightinv(1:5);
rightinv(1:5)=[0 0 0 0 0];
rightinv(38:128)=zeros(91,1);
```

% The above statements were similar to the previous one except for the right ear. In this case, the time delay of 5 '0' was needed to include in the data for the right ear and 91 '0' were filled up the end part of the original 128 data space.

A5.4 Normal probability plots of localization errors in Experiment one

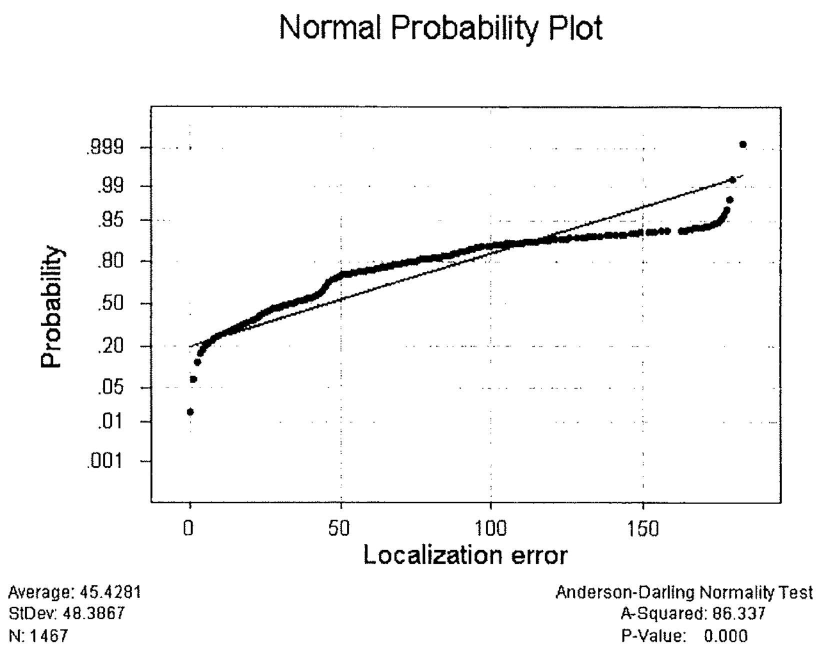


FIGURE A5.8 Normal probability plot of the localization error in Experiment one.

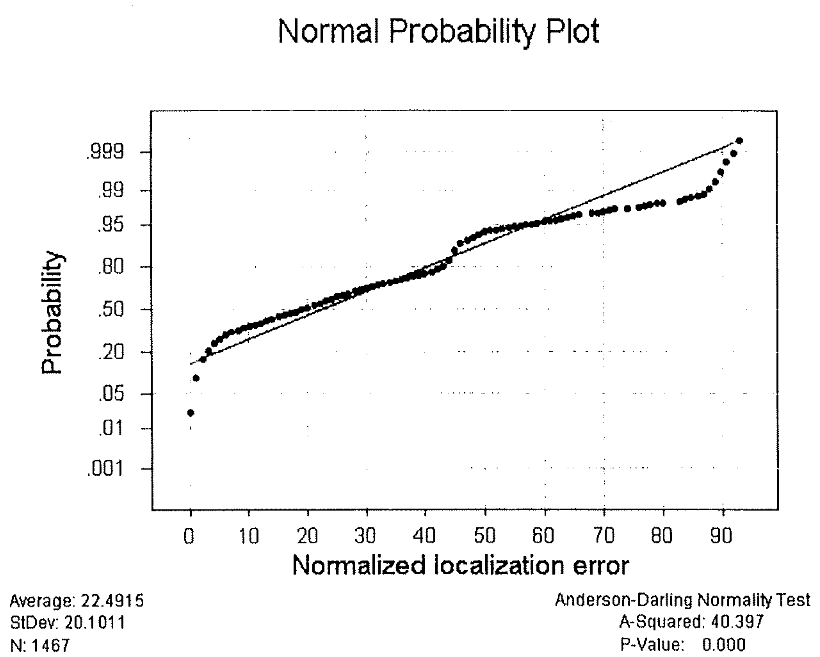


FIGURE A5.9 Normal probability plot of the localization error with front-back confusions excluded in Experiment one.

APPENDIX TO CHAPTER 6

A6.1 Samples of instruction sheets and questionnaires (part one: directional effect) used in Experiment two

Questionnaire A on virtual surround sound performance (directional effect) of Virtual Headphone-based Surround Sound (VHSS) System

Age: ____________
Sex: ____________
School: Engineering Business Science Humanities

This experiment is being conducted to examine the virtual surround sound performance **(directional effect, 方向性效果)** of the Virtual Headphone-based Surround Sound (VHSS) System. In this experiment, you are required to watch a demo clip (Dolby™ Surround Sound Demonstration Disc) about the delivery of surround sound from six speakers **(Center, Left, Right, Left Rear, Right Rear and Subwoofer)** consecutively. It lasts for about two minutes long. You will listen to the virtual surround sounds through a pair of headphones. There will be four versions of surround sound effects that you need to listen to. Once you finished enjoying and listening to the version 1, you will require to listen to the same part of movie clip for the version 2. Pair-wise comparisons of the surround sound performance on the directional effects between these two versions are needed. You will need to complete a set of questionnaire for the comparison. **(Please ignore the effects of Subwoofer (重低音喇叭), the speaker that located below the Center speaker (中置喇叭), when filling up the questionnaire as sounds from Subwoofer have not been processed for directions. You may not be able to hear the sound from that speaker in some versions.)** After finishing answering the questionnaire, you are required to undergo another pair-wise comparison between version 3 and version 4. Again, a set of questionnaire will be given for you to fill it up.

Have you ever learned any musical instrument(s) or joined any choir(s)?

If yes, please write down some details:

How long have you learned for that/those musical instrument(s) or joined the choir(s)?

Are you still practising that/those musical intrument(s) or choir(s) who have mentioned above in the past 6 months?

Do you think you are experienced in playing musical instrument(s) or singing?

Now, any questions you want to ask? If no, then please read through the first set of questionnaire on the following pages.

Please indicate to the experimenter that you are ready to start if you don't have any questions about the questionnaire.

Musical (Professional) participants: Yes / No
(Fill in by experimenter)

<u>Questionnaire A on virtual surround sound performance (directional effect) of Virtual Headphone-based Surround Sound (VHSS) System – 1st trial</u>

This questionnaire is being conducted to examine the virtual surround sound performance of the Virtual Headphone-based Surround Sound (VHSS) System. Please mark down your feeling that you think it is the most appropriate answer for that question. Continuous scale is used for marking down your answers. You are allowed to mark your feeling down anywhere in between two options in proportional scale. <u>(Please ignore the effects of Subwoofer when filling up this questionnaire.)</u>

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

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

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

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

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

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

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

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

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

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

6. On the whole, the 1st version gives me **BETTER** surround sound effects (環迴效果) than the 2nd version.

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

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

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

A6.2 Samples of instruction sheets and questionnaires (part two: surround effect and spatial effect) used in Experiment two

Questionnaire B on virtual surround sound performance (surround effect and spatial effect) of Virtual Headphone-based Surround Sound (VHSS) System

This experiment is being conducted to examine the virtual surround sound performance **(surround effect, 環迴效果 and spatial effect, 空間效果)** of the Virtual Headphone-based Surround Sound (VHSS) System. In this experiment, you are required to watch three short movie clips (Air-force One) that only last for about 30 seconds totally. At the same time, you will listen to the virtual surround sounds through a pair of headphones. There will be **four** versions of surround sound effects that you need to listen to. For the three short movie clips namely A, B and C respectively, you will listen to the version 1 of clip A (i.e. A1), and then followed by version 2 of clip A (i.e. A2). Afterwards, version 1 of clip B (i.e. B1) will be delivered and then followed by version 2 of clip B (i.e. B2). Same practice will be applied for clip C. The sequence for surround sound delivery will then be A1, A2, B1, B2, C1, C2. Pair-wise comparisons of the surround sound performance on the **surround effect and spatial effect** between these two versions are needed. You will need to complete a set of questionnaire for the comparison. After finishing answering the questionnaires, you are required to undergo another pair-wise comparison between version 3 and version 4 following the above-mentioned procedure. Again, a set of questionnaire will be given for you to fill it up.

Now, any questions you want to ask? If no, then please read through the first set of questionnaire on the following pages.

Please indicate to the experimenter that you are ready to start if you don't have any questions about the questionnaire.

<u>Questionnaire B on virtual surround sound performance (surround effect and spatial effect) of Virtual Headphone-based Surround Sound (VHSS) System – 1st trial</u>

This questionnaire is being conducted to examine the virtual surround sound performance of the Virtual Headphone-based Surround Sound (VHSS) System. Please mark down your feeling that you think it is the most appropriate answer for that question. Continuous scale is used for marking down your answers. You are allowed to mark your feeling down anywhere in between two options in proportional scale.

1. The 1st version sounds **MORE "Live" (現場感)** than the 2nd version.

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

2. I can distinguish sounds of different directions **MUCH EASIER** in the 1st version than in the 2nd version.

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

3. The 1st version of sound gives **MORE "Wide spatial effects" (空間效果闊)** than the 2nd version.

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

4. Sounds in the 1st version appear to have a **CLEARER** direction than sounds in the 2nd version.

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

5. The 1st version gives **BETTER "Surround sound effects" (環迴效果)** than the 2nd version.

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

6. The 1st version of sound gives **BETTER "Feeling of Presence" (置身其中的感覺)** than 2nd version.

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

7. Sounds in the 1st version appear to give a **MORE "Narrow spatial effects" (空間效果窄)** than sounds in the 2nd version.

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

8. Sounds in the 1st version appear to surround me **MORE** than sounds in the 2nd version.

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

9. Sounds in the 1st version appear to originate **MORE** from **"Inside of your head"** than the sounds in the 2nd version.

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

10. The 1st version of sound gives **BETTER** feeling of being inside the movie environment (置身其中) at the position of the cameraman than the 2nd version.

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

11. The 1st version is **MORE** enjoyable than the 2nd version.

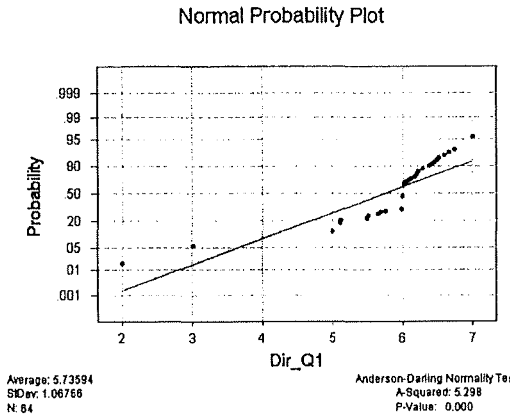
Strongly Agree	Agree	Slightly Agree	Neutral	Slightly Disagree	Disagree	Strongly Disagree
(1)	(2)	(3)	(4)	(5)	(6)	(7)

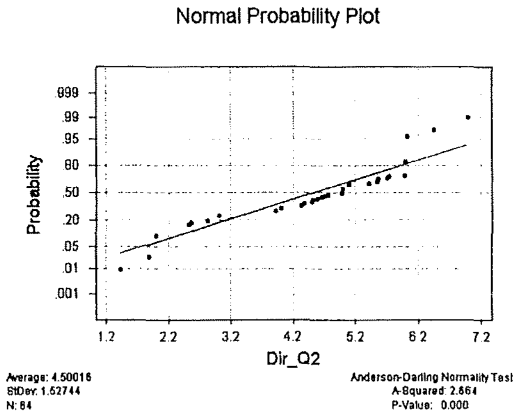
12. I can hear sounds coming from my back **MORE** in the 1st version than in the 2nd version.

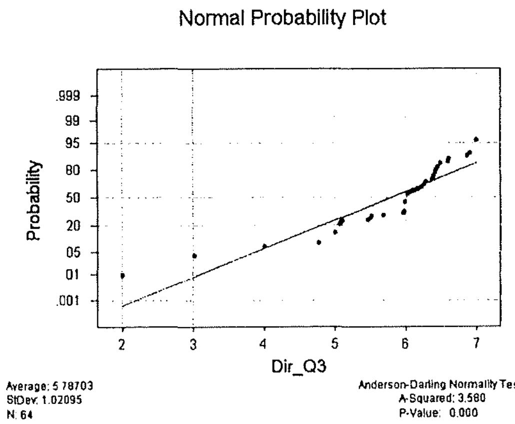
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|

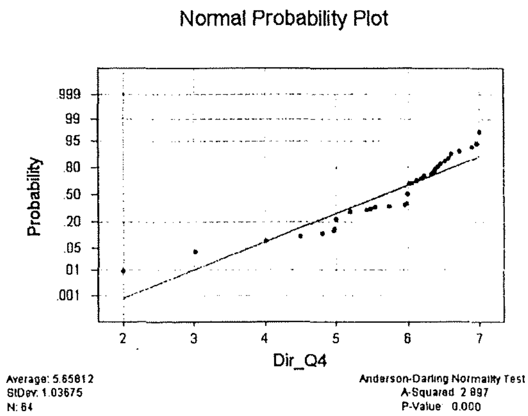
Strongly Agree	Agree	Slightly Agree	Neutral	Slightly Disagree	Disagree	Strongly Disagree
(1)	(2)	(3)	(4)	(5)	(6)	(7)

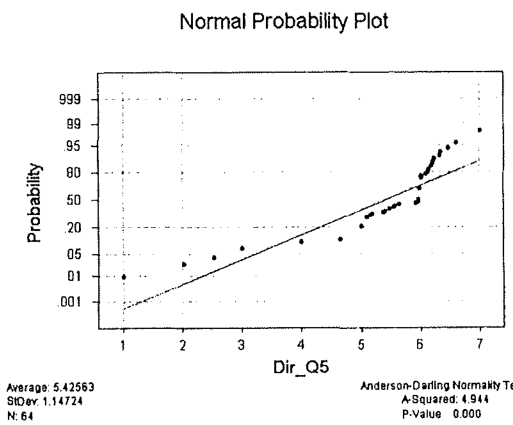
A6.3 Normal probability plots of surround sound ratings for questions 1 to 7 in part one of Experiment two

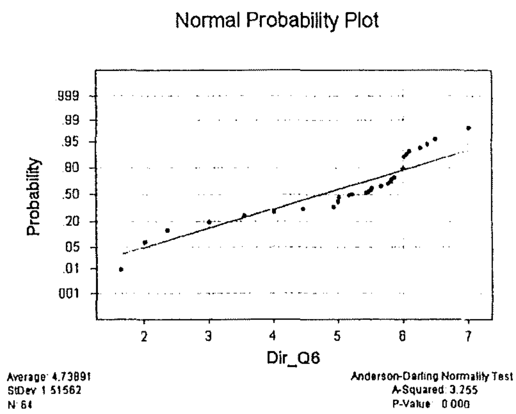


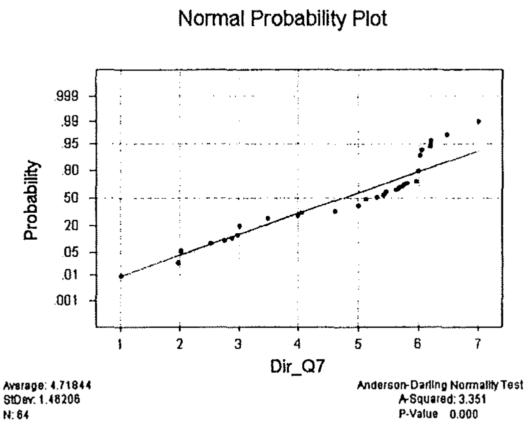






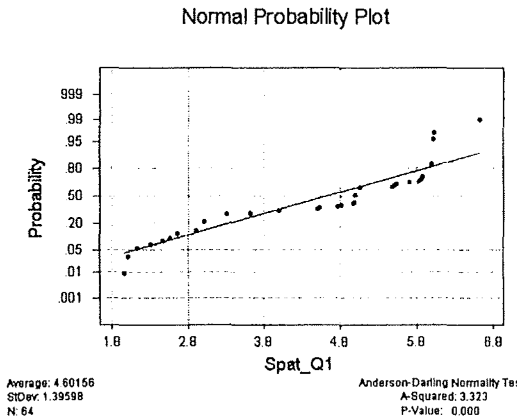


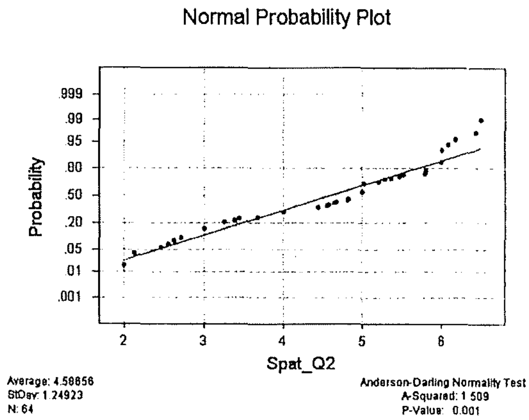


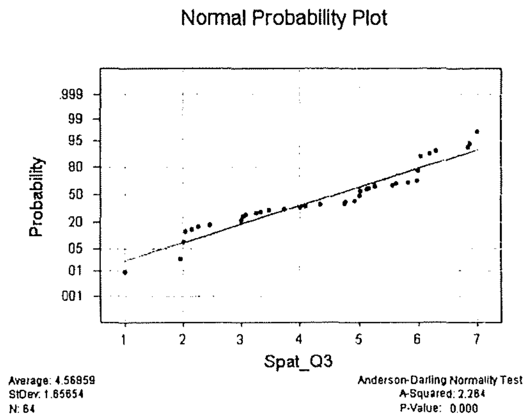


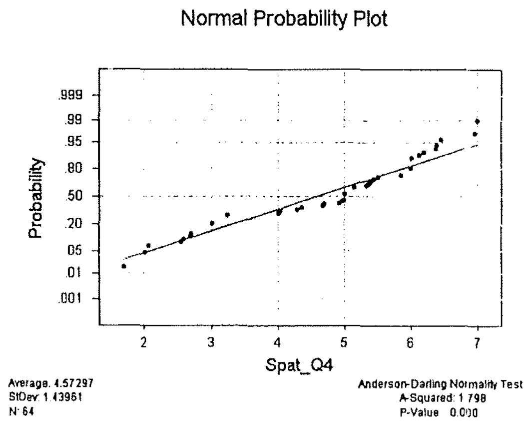
FIGURES A6.1 – A6.7 Normal probability plots of surround sound ratings for questions 1 to 7 in part one of Experiment two.

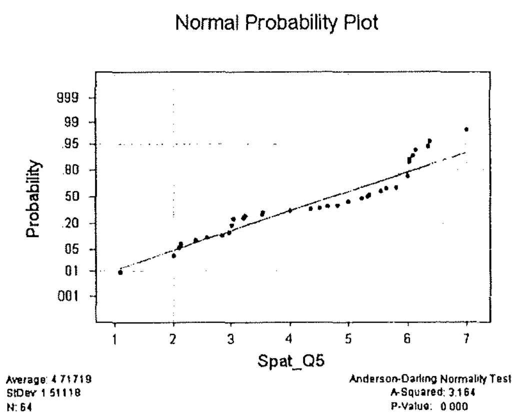
A6.4 Normal probability plots of surround sound ratings for questions 1 to 12 in part two of Experiment two

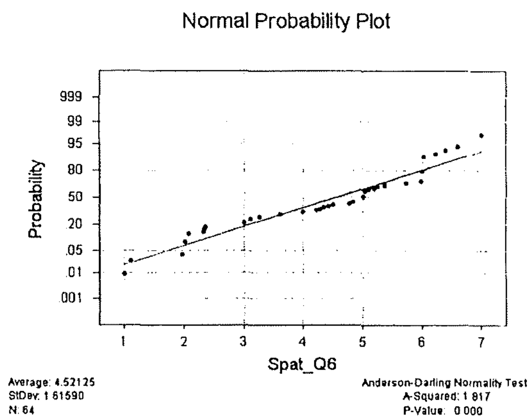


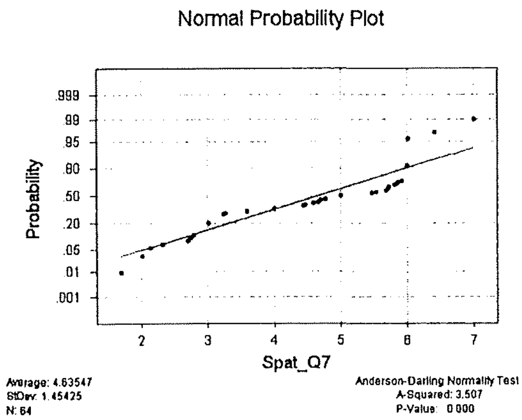


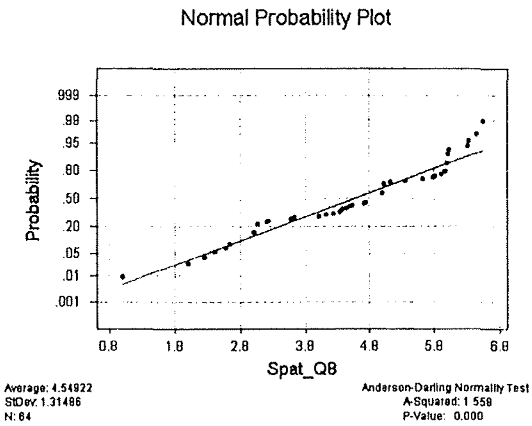


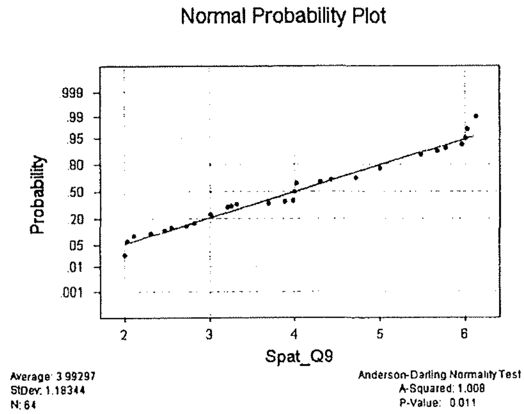


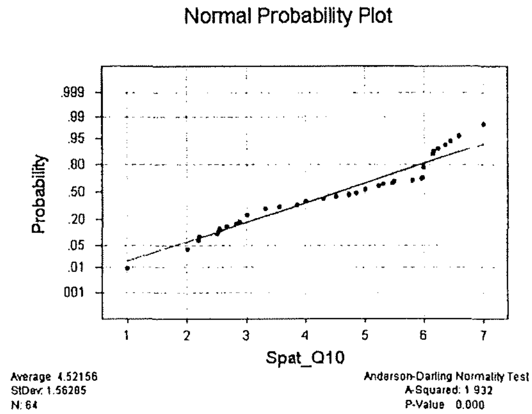


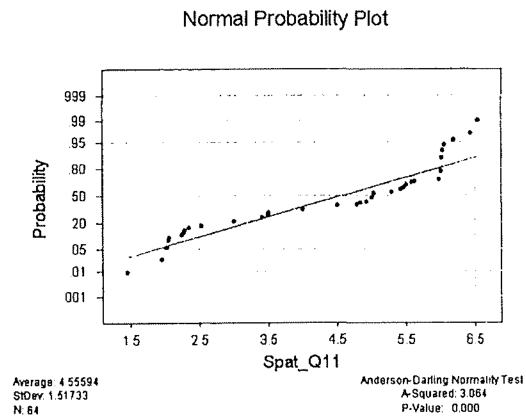


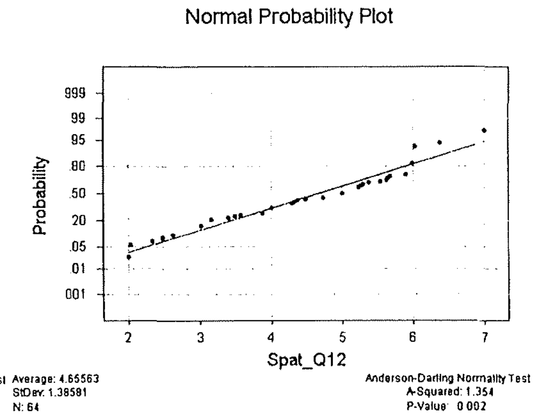












FIGURES A6.8 – A6.19 Normal probability plots of surround sound ratings for questions 1 to 12 in part two of Experiment two.

A6.5 Cronbach alpha analysis on surround sound ratings for all questions in part one and part two of Experiment two

Part one

****** Method 1 (space saver) will be used for this analysis ******

R E L I A B I L I T Y   A N A L Y S I S - S C A L E (A L P H A)

Item-total Statistics

	Scale Mean If Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
DIR_Q1	30.8283	23.8965	.6444	.6877
DIR_Q2	32.0641	29.7171	-.0172	.8363
DIR_Q3	30.7772	23.7244	.7035	.6791
DIR_Q4	30.9061	23.6771	.6949	.6797
DIR_Q5	31.1386	22.6014	.7192	.6677
DIR_Q6	31.8253	23.0334	.4422	.7288
DIR_Q7	31.8458	23.5074	.4315	.7302

Reliability Coefficients

N of Cases = 64.0 N of Items = 7

Alpha = .7501

Part two

****** Method 1 (space saver) will be used for this analysis ******

R E L I A B I L I T Y   A N A L Y S I S - S C A L E (A L P H A)

Item-total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
SPAT_Q1	49.8773	160.3601	.7844	.9410
SPAT_Q2	49.8923	162.4900	.8160	.9404
SPAT_Q3	49.9103	153.3358	.8278	.9394
SPAT_Q4	49.9059	158.2563	.8203	.9397
SPAT_Q5	49.7617	154.5380	.8841	.9374
SPAT_Q6	49.9577	153.4684	.8482	.9386
SPAT_Q7	49.8434	157.9016	.8215	.9397
SPAT_Q8	49.9297	161.1369	.8141	.9403
SPAT_Q9	50.4859	193.2438	-.1399	.9649
SPAT_Q10	49.9573	154.0411	.8650	.9380
SPAT_Q11	49.9230	153.8346	.9007	.9368
SPAT_Q12	49.8233	162.9245	.7122	.9433

Reliability Coefficients

N of Cases = 64.0 N of Items = 12

Alpha = .9467

APPENDIX TO CHAPTER 7

A7.1 Detailed descriptions of program code to generate the enhanced HRTFs in Experiment three (using direction 45° with enhancement of 12dB at frequency band 3 as an example)

```
fid=fopen('c:\sound\thesis\enhance_45l_3_12db.dat','r');
```

% 'fo' was the variable of file that used to open the file containing the compact set of data for azimuth direction 45° at 0-degree elevation. 'r' meant opening the file with permission of 'read' function.

```
[gain45l,count]=fscanf(fid,'%f');
```

% The resultant enhanced HRTFs of the left ear previously stored in the .dat data file were now read into the variable 'gain45l'.

```
fclose(fid);
```

% The command 'fclose' was used to close the file after reading the enhanced HRTF data.

```
fid=fopen('c:\sound\thesis\enhance_45r_3_12db.dat','r');
[gain45r,count]=fscanf(fid,'%f');
fclose(fid);
```

% Same as above-mentioned but for the right ear.

```
read045;
chgfft;
uwrap;
```

% These commands were used to capture the phase angle values from the original HRTF data of direction 45°.

```
x0l=10.^(gain45l/20);
x0r=10.^(gain45r/20);

for j=1:128;
   left_a(j)=x0l(j)*cos(leftphase(j));
   left_b(j)=x0l(j)*sin(leftphase(j));
   x0l_comx(j)=left_a(j)+i*left_b(j);
   right_a(j)=x0r(j)*cos(rightphase(j));
   right_b(j)=x0r(j)*sin(rightphase(j));
   x0r_comx(j)=right_a(j)+i*right_b(j);
end;
```

% The above commands were used to obtain back the complex number based on the formula: Gain = 20 * log(A) and complex_no = a + bj where a = A * cos(phase angle), b = A * sin(phase angle).

```
x0l_imp=real(ifft(x0l_comx));
x0r_imp=real(ifft(x0r_comx));
```

% Inverse Fast Fourier Transform was performed by using the command 'ifft'. The real part of the complex number data were extracted for the left and right ears respectively.

A7.2 Normal probability plots of the localization errors in Experiment three

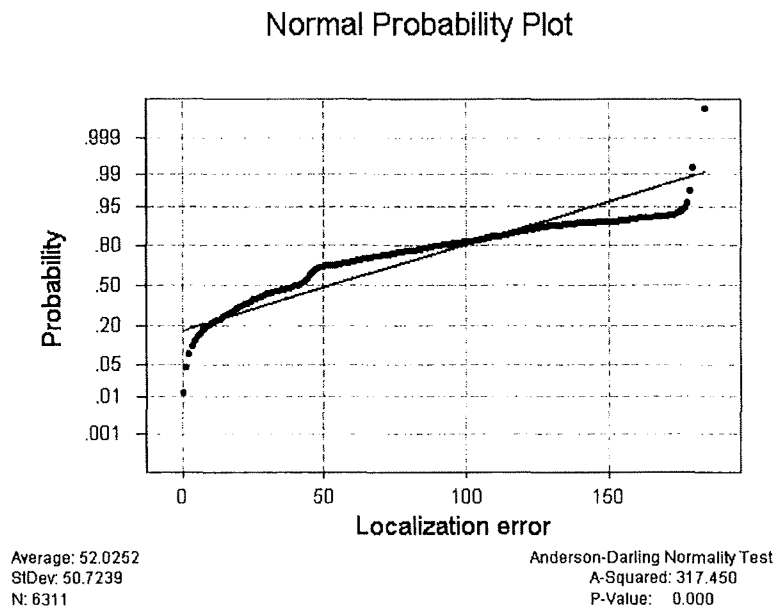


FIGURE A7.1 Normal probability plot of the localization error in Experiment three.

A7.3 Modules of HRTFs of enhanced peaks and notches in the spectra for 3 levels of enhancements, 6 frequency bands and 6 sound directions

The HRTF filters shown in this part are used for commercial applications and kept confidential.

A7.4 Ear dimensions of 32 participants in Experiment three

This appendix illustrates the dimensions of left and right ears of 32 participants who took part in Experiment three. Ear photographs of all participants (front view, side view and back view) can be found in the compact disc attached.

TABLE A7.1 Ear dimensions (in mm) for all 32 participants in Experiment three.

Participant	Ear	Pinna-cavity height (I to A)	Intertragal incisure to anti-helix (I to B)	Intertragal incisure to crus (I to C)	Tragus to antitragus (G to H)	Pinna-cavity breath (G to E)
1	Right	41.43	26.9	20	8.57	15.48
	Left	38.72	24.36	19.74	6.67	13.59
	Diff	2.71	2.54	0.26	1.9	1.89
2	Right	35	23.04	15.65	5.43	12.83
	Left	34.13	21.96	15.65	5.65	13.04
	Diff	0.87	1.08	0	-0.22	-0.21
3	Right	28.48	20.87	14.78	7.61	13.91
	Left	26.17	17.87	12.77	8.51	14.47
	Diff	2.31	3	2.01	-0.9	-0.56
4	Right	32.39	22.83	16.74	5	13.04
	Left	30.64	20.64	15.53	5.11	9.36
	Diff	1.75	2.19	1.21	-0.11	3.68
5	Right	44.88	30	21.95	7.56	14.15
	Left	40.27	26.76	18.65	4.05	10.81
	Diff	4.61	3.24	3.3	3.51	3.34
6	Right	37.32	25.12	17.07	8.54	15.85
	Left	37.75	26.25	18.5	5.25	13.75
	Diff	-0.43	-1.13	-1.43	3.29	2.1
7	Right	26.43	20.71	14.29	4.29	11.61
	Left	24.55	20.18	15.09	4.73	10.55
	Diff	1.88	0.53	-0.8	-0.44	1.06

Participant	Ear	Pinna-cavity height (I to A)	Intertragal incisure to anti-helix (I to B)	Intertragal incisure to crus (I to C)	Tragus to antitragus (G to H)	Pinna-cavity breath (G to E)
8	Right	37.83	26.52	21.3	5	13.04
	Left	31.25	22.29	17.08	7.08	13.13
	Diff	6.58	4.23	4.22	-2.08	-0.09
9	Right	33.13	21.04	16.88	5.21	10.21
	Left	29	20	14	5	8.2
	Diff	4.13	1.04	2.88	0.21	2.01
10	Right	47.91	30.47	20	9.53	15.35
	Left	42.44	26.83	17.32	8.54	14.39
	Diff	5.47	3.64	2.68	0.99	0.96
11	Right	33.1	20.24	16.67	7.14	14.52
	Left	31.89	18.11	12.83	5.28	10.75
	Diff	1.21	2.13	3.84	1.86	3.77
12	Right	42.14	27.62	21.67	10	15.95
	Left	38.81	25	19.05	8.81	13.57
	Diff	3.33	2.62	2.62	1.19	2.38
13	Right	42.7	28.65	20	8.65	16.22
	Left	33.81	22.38	16.19	8.33	15.48
	Diff	8.89	6.27	3.81	0.32	0.74
14	Right	33.49	26.74	19.07	6.28	13.49
	Left	32	23.2	16.8	5.8	12.4
	Diff	1.49	3.54	2.27	0.48	1.09
15	Right	39.53	25.35	17.44	6.28	16.74
	Left	39.53	24.42	17.67	6.51	14.19
	Diff	0	0.93	-0.23	-0.23	2.55

Participant	Ear	Pinna-cavity height (I to A)	Intertragal incisure to anti-helix (I to B)	Intertragal incisure to crus (I to C)	Tragus to antitragus (G to H)	Pinna-cavity breath (G to E)
16	Right	40	24.19	18.37	7.91	15.81
	Left	34.22	20.89	15.56	8.22	14
	Diff	5.78	3.3	2.81	-0.31	1.81
17	Right	28.6	22.79	17.21	6.28	12.79
	Left	27.5	22.27	16.82	5.45	12.73
	Diff	1.1	0.52	0.39	0.83	0.06
18	Right	39.5	24.75	19.25	7.75	13.25
	Left	37.18	23.33	17.44	7.18	12.82
	Diff	2.32	1.42	1.81	0.57	0.43
19	Right	33.8	29.2	22.2	6.4	13.2
	Left	27.78	23.56	17.33	6.44	10.44
	Diff	6.02	5.64	4.87	-0.04	2.76
20	Right	35.38	24.1	16.67	6.41	14.87
	Left	32.33	23.72	18.84	6.98	13.49
	Diff	3.05	0.38	-2.17	-0.57	1.38
21	Right	39.25	26.5	19.25	8.75	14.25
	Left	33.9	23.9	18.54	8.78	16.59
	Diff	5.35	2.6	0.71	-0.03	-2.34
22	Right	29.79	23.62	17.66	6.38	13.83
	Left	27.4	22.6	16.2	5.8	12.6
	Diff	2.39	1.02	1.46	0.58	1.23
23	Right	36.5	25.75	17.75	8.25	14.75
	Left	30.65	21.74	15.65	7.83	13.26
	Diff	5.85	4.01	2.1	0.42	1.49

Participant	Ear	Pinna-cavity height (I to A)	Intertragal incisure to anti-helix (I to B)	Intertragal incisure to crus (I to C)	Tragus to antitragus (G to H)	Pinna-cavity breath (G to E)
24	Right	33.86	25	18.19	7.95	13.41
	Left	30.44	22	16	7.11	11.78
	Diff	3.42	3	2.19	0.84	1.63
25	Right	38.33	27.62	18.1	6.19	13.81
	Left	38.26	26.96	18.91	6.3	11.96
	Diff	0.07	0.66	-0.81	-0.11	1.85
26	Right	41.4	28.14	21.4	7.21	13.72
	Left	34.42	21.46	16.04	6.67	11.46
	Diff	6.98	6.68	5.36	0.54	2.26
27	Right	39.09	24.09	18.86	4.55	11.59
	Left	38.1	23.1	16.43	6.19	10
	Diff	0.99	0.99	2.43	-1.64	1.59
28	Right	35.35	23.72	16.51	9.07	16.05
	Left	31.14	20.68	15.23	7.73	13.18
	Diff	4.21	3.04	1.28	1.34	2.87
29	Right	39.75	27	21	6.75	17
	Left	34.76	22.38	18.1	5.24	13.33
	Diff	4.99	4.62	2.9	1.51	3.67
30	Right	38.04	26.3	17.83	8.26	14.35
	Left	35	25	17.29	7.5	15.63
	Diff	3.04	1.3	0.54	0.76	-1.28
31	Right	40	25.64	18.97	8.21	14.1
	Left	38.46	23.85	17.69	5.64	12.82
	Diff	1.54	1.79	1.28	2.57	1.28

Participant	Ear	Pinna-cavity height (I to A)	Intertragal incisure to anti-helix (I to B)	Intertragal incisure to crus (I to C)	Tragus to antitragus (G to H)	Pinna-cavity breath (G to E)
32	Right	30.8	19.4	17	7.2	14.8
	Left	31.04	23.13	16.25	7.08	15.42
	Diff	-0.24	-3.73	0.75	0.12	-0.62

APPENDIX TO CHAPTER 8

A8.1 Modules of HRTFs illustrating an optimized set of non-individualized HRTF filters for the VHSS system

The HRTF filters shown in this part are used for commercial applications and kept confidential.

A8.2 Indications of improvement in sound localization errors from each participant after 12 dB enhancement of the spectra of non-individualized HRTFs at the 6 frequency bands in Experiment three

The tables show the Indications of improvement in sound localization errors from each participant after 12dB enhancement of the spectra of non-individualized HRTFs at the 6 frequency bands A '+' sign means that there was an reduction in localization errors of more than 2^o. A '-' sign represents an increase in the localization errors of more than 2^o. A symbol of '0' indicates that no improvement (2^o < difference < -2^o) was observed after enhancing the frontal and backward cues. 'Nil' means that no data was available because, after the enhancement, the participants perceived the sound source as at the center of the head.

TABLE A8.1a Indications of improvement in sound localization errors from each participant after 12dB enhancement of the spectra of non-individualized HRTFs at the 6 frequency bands for 0^{o} direction ('+' reductions in errors; '-' increases in errors; '0' no change in errors; 'nil' missing data).

Participant	band 1 (0.2 – 0.69 kHz)	band 2 (0.69 – 2.4 kHz)	band 3 (2.4 – 6.5 kHz)	band 4 (6.5 – 10.0 kHz)	band 5 (10.0 – 14.0 kHz)	band 6 (14.0 – 22.0 kHz)
1	+	+	+	nil	nil	nil
2	nil	+	+	nil	+	+
3	-	+	+	-	-	-
4	+	+	+	+	+	+
5	+	+	+	+	+	+
6	+	-	nil	+	+	-
7	+	0	+	+	+	+
8	nil	-	-	-	nil	-
9	+	-	+	+	-	+
10	+	+	+	-	+	nil
11	-	+	-	+	nil	-
12	-	-	-	-	+	+
13	+	nil	+	nil	+	-
14	+	+	-	+	+	+
15	+	+	+	+	-	+
16	-	-	-	nil	+	+
17	nil	+	+	+	+	+
18	+	+	-	-	0	0
19	-	+	+	-	+	-
20	+	+	+	nil	+	-
21	nil	+	-	+	+	+
22	nil	-	+	-	-	+
23	-	+	+	-	-	-
24	+	-	nil	+	+	nil
25	+	+	nil	+	-	-
26	+	+	+	0	-	+
27	+	-	-	-	-	+
28	+	+	+	nil	+	+
29	+	+	+	nil	+	+
30	+	+	+	-	+	+
31	+	+	-	-	-	+
32	nil	+	nil	+	+	+

TABLE A8.1b Indications of improvement in sound localization errors from each participant after 12dB enhancement of the spectra of non-individualized HRTFs at the 6 frequency bands for 45° direction ('+' reductions in errors; '-' increases in errors; '0' no change in errors; 'nil' missing data).

Participant	band 1 (0.2 – 0.69 kHz)	band 2 (0.69 – 2.4 kHz)	band 3 (2.4 – 6.5 kHz)	band 4 (6.5 – 10.0 kHz)	band 5 (10.0 – 14.0 kHz)	band 6 (14.0 – 22.0 kHz)
1	+	+	+	-	+	+
2	+	+	+	+	+	+
3	+	0	+	-	+	+
4	+	-	-	+	+	+
5	+	-	+	+	+	+
6	+	+	+	+	+	-
7	+	-	+	+	+	+
8	-	-	-	-	-	-
9	-	+	-	-	+	-
10	+	+	-	0	+	+
11	+	-	+	0	+	-
12	+	+	-	+	+	+
13	+	+	+	+	+	+
14	+	+	+	+	+	+
15	-	+	+	+	+	+
16	+	+	+	+	+	+
17	+	0	+	0	+	0
18	+	+	+	+	+	+
19	-	-	+	-	-	+
20	+	+	+	+	+	+
21	0	+	+	+	+	-
22	+	+	+	+	+	+
23	+	+	+	+	-	+
24	+	-	-	+	+	+
25	-	-	-	-	0	-
26	+	+	+	-	+	+
27	+	+	+	+	+	+
28	+	+	+	+	+	+
29	+	+	+	+	+	+
30	+	+	+	+	+	+
31	+	+	+	+	+	+
32	+	+	+	+	+	+

TABLE A8.1c Indications of improvement in sound localization errors from each participant after 12dB enhancement of the spectra of non-individualized HRTFs at the 6 frequency bands for 135° direction ('+' reductions in errors; '-' increases in errors; '0' no change in errors; 'nil' missing data).

Participant	band 1 (0.2 – 0.69 kHz)	band 2 (0.69 – 2.4 kHz)	band 3 (2.4 – 6.5 kHz)	band 4 (6.5 – 10.0 kHz)	band 5 (10.0 – 14.0 kHz)	band 6 (14.0 – 22.0 kHz)
1	+	-	+	+	+	-
2	-	+	+	0	+	+
3	+	+	+	+	+	+
4	+	+	+	+	+	+
5	+	+	+	+	+	+
6	+	+	+	+	+	+
7	0	+	+	+	+	-
8	+	+	+	-	+	+
9	+	+	+	+	+	+
10	+	+	+	+	-	+
11	+	+	+	+	+	+
12	+	+	+	+	+	+
13	+	+	+	-	+	+
14	+	+	+	-	+	+
15	+	+	+	+	+	+
16	+	+	+	+	+	+
17	+	+	+	-	+	+
18	+	+	+	-	+	+
19	-	-	-	-	0	+
20	0	+	0	-	-	+
21	+	+	+	-	+	+
22	-	-	+	0	0	+
23	+	+	0	-	+	+
24	+	+	+	+	+	+
25	+	+	+	-	+	+
26	+	+	+	0	+	+
27	+	+	+	-	+	+
28	-	+	+	-	+	+
29	+	+	+	+	+	-
30	+	+	+	+	+	+
31	+	+	+	+	+	+
32	+	+	-	-	+	-

TABLE A8.1d Indications of improvement in sound localization errors from each participant after 12dB enhancement of the spectra of non-individualized HRTFs at the 6 frequency bands for 180° direction ('+' reductions in errors; '-' increases in errors; '0' no change in errors; 'nil' missing data).

Participant	band 1 (0.2 – 0.69 kHz)	band 2 (0.69 – 2.4 kHz)	band 3 (2.4 – 6.5 kHz)	band 4 (6.5 – 10.0 kHz)	band 5 (10.0 – 14.0 kHz)	band 6 (14.0 – 22.0 kHz)
1	+	+	nil	+	nil	+
2	0	+	-	-	+	+
3	-	+	+	-	+	+
4	-	+	-	-	+	-
5	-	-	+	-	+	+
6	nil	nil	+	+	-	+
7	+	-	+	-	+	+
8	+	+	-	nil	+	+
9	+	+	+	+	+	-
10	+	nil	+	-	nil	nil
11	+	nil	0	nil	-	+
12	+	-	+	-	0	-
13	+	+	nil	nil	nil	+
14	+	+	+	+	+	+
15	-	+	+	+	-	+
16	+	+	+	nil	+	+
17	+	-	+	+	+	+
18	+	+	+	-	+	+
19	+	+	+	+	+	+
20	nil	+	+	+	+	+
21	+	+	+	+	+	+
22	+	-	-	+	-	+
23	+	+	-	+	+	+
24	+	+	+	-	-	+
25	+	-	+	-	+	-
26	+	+	+	-	+	+
27	+	-	+	-	+	-
28	+	+	+	-	+	+
29	+	-	+	+	+	+
30	+	+	+	+	+	-
31	-	+	+	-	-	nil
32	nil	nil	-	+	-	-

TABLE A8.1e Indications of improvement in sound localization errors from each participant after 12dB enhancement of the spectra of non-individualized HRTFs at the 6 frequency bands for 225° direction ('+' reductions in errors; '-' increases in errors; '0' no change in errors; 'nil' missing data).

Participant	band 1 (0.2 – 0.69 kHz)	band 2 (0.69 – 2.4 kHz)	band 3 (2.4 – 6.5 kHz)	band 4 (6.5 – 10.0 kHz)	band 5 (10.0 – 14.0 kHz)	band 6 (14.0 – 22.0 kHz)
1	0	+	+	-	+	+
2	-	+	+	0	+	+
3	+	+	+	+	+	+
4	+	+	+	+	+	+
5	+	+	+	+	+	+
6	+	+	+	+	+	+
7	+	+	-	-	+	-
8	+	+	+	+	+	+
9	+	+	-	0	+	+
10	+	+	+	+	+	+
11	+	+	+	+	+	+
12	+	+	+	+	+	+
13	+	+	+	+	+	+
14	+	+	+	+	+	+
15	+	+	+	+	+	+
16	+	+	+	-	+	+
17	+	+	+	+	+	+
18	-	+	-	0	+	0
19	+	+	+	-	+	+
20	+	+	+	+	+	+
21	+	+	+	-	+	+
22	-	+	-	-	0	-
23	+	+	+	+	+	+
24	+	+	+	+	+	+
25	+	0	+	-	+	+
26	+	+	+	+	+	+
27	+	+	+	+	+	+
28	+	+	+	+	+	+
29	+	+	-	+	0	-
30	+	+	+	+	+	+
31	+	+	+	+	+	+
32	+	+	+	-	+	+

TABLE A8.1f Indications of improvement in sound localization errors from each participant after 12dB enhancement of the spectra of non-individualized HRTFs at the 6 frequency bands for 315° direction ('+' reductions in errors; '-' increases in errors; '0' no change in errors; 'nil' missing data).

Participant	band 1 (0.2 – 0.69 kHz)	band 2 (0.69 – 2.4 kHz)	band 3 (2.4 – 6.5 kHz)	band 4 (6.5 – 10.0 kHz)	band 5 (10.0 – 14.0 kHz)	band 6 (14.0 – 22.0 kHz)
1	+	+	-	0	+	-
2	+	+	+	+	+	+
3	+	+	+	+	+	+
4	+	+	+	+	+	+
5	+	+	+	+	+	+
6	-	+	+	+	+	+
7	+	+	+	+	+	+
8	+	+	+	+	+	+
9	+	+	+	+	+	+
10	+	+	0	+	-	+
11	0	+	0	-	+	-
12	+	+	-	-	-	-
13	-	-	0	-	-	0
14	+	-	+	-	+	-
15	+	-	+	+	+	+
16	0	-	+	+	+	-
17	-	-	+	+	-	+
18	-	-	0	-	0	-
19	+	+	-	-	+	-
20	+	0	-	-	0	-
21	+	+	0	0	0	-
22	0	-	-	-	+	0
23	+	-	-	-	+	+
24	-	+	-	-	+	+
25	+	-	+	-	+	-
26	-	0	0	+	+	-
27	-	-	+	0	0	+
28	-	+	+	-	+	+
29	-	-	+	-	-	+
30	+	+	-	+	+	0
31	0	0	-	-	+	0
32	-	-	-	+	+	+

TABLE A8.2a Indications of improvement in sound localization errors from each participant after 12dB enhancement of the spectra of non-individualized HRTFs at the 6 frequency bands for frontal directions ('+' reductions in errors; '-' increases in errors; '0' no change in errors; 'nil' missing data). Data are the sums of improvement from the tables of 0°, 45° and 315°.

Participant	band 1 (0.2 – 0.69 kHz)	band 2 (0.69 – 2.4 kHz)	band 3 (2.4 – 6.5 kHz)	band 4 (6.5 – 10.0 kHz)	band 5 (10.0 – 14.0 kHz)	band 6 (14.0 – 22.0 kHz)
1	+++	+++	+	-	++	0
2	++	+++	+++	++	+++	+++
3	+	++	+++	-	+	+
4	+++	+	+	+++	+++	+++
5	+++	+	+++	+++	+++	+++
6	+	+	++	+++	+++	-
7	+++	0	+++	+++	+++	+++
8	0	-	-	-	0	-
9	+	+	+	+	+	+
10	+++	+++	0	0	+	++
11	0	+	0	0	++	---
12	+	+	---	-	+	+
13	+	0	++	0	+	0
14	+++	+	+	+	+++	+
15	+	+	+++	+++	+	+++
16	0	-	+	++	+++	+
17	0	0	+++	++	+	++
18	+	+	0	-	+	0
19	-	+	+	-	+	-
20	+++	++	+	0	++	-
21	+	+++	0	++	++	-
22	+	-	+	-	+	++
23	+	+	+	-	-	+
24	+	-	--	+	+++	++
25	+	-	0	-	0	---
26	+	++	++	0	+	+
27	+	-	+	0	0	+++
28	+	+++	+++	0	+++	+++
29	+	+	+++	0	+	+++
30	+++	+++	+	+	+++	++
31	++	++	-	-	+	++
32	0	+	0	+++	+++	+++

TABLE A8.2b Indications of improvement in sound localization errors from each participant after 12dB enhancement of the spectra of non-individualized HRTFs at the 6 frequency bands for backward directions ('+' reductions in errors; '-' increases in errors; '0' no change in errors; 'nil' missing data). Data are the sums of improvement from the tables of 135°, 180° and 225°.

Participant	band 1 (0.2 – 0.69 kHz)	band 2 (0.69 – 2.4 kHz)	band 3 (2.4 – 6.5 kHz)	band 4 (6.5 – 10.0 kHz)	band 5 (10.0 – 14.0 kHz)	band 6 (14.0 – 22.0 kHz)
1	++	+	++	+	++	+
2	--	+++	+	-	+++	+++
3	+	+++	+++	+	+++	+++
4	+	+++	+	+	+++	+
5	+	+	+++	+	+++	+++
6	++	++	+++	+++	+	+++
7	++	+	+	-	+++	-
8	+++	+++	+	0	+++	+++
9	+++	+++	+	++	+++	+
10	+++	++	+++	+	0	++
11	+++	++	++	++	+	+++
12	+++	+	+++	+	++	+
13	+++	+++	++	0	++	+++
14	+++	+++	+++	+	+++	+++
15	+	+++	+++	+++	+	+++
16	+++	+++	+++	0	+++	+++
17	+++	+	+++	+	+++	+++
18	+	+++	+	--	+++	++
19	+	+	+	-	++	+++
20	+	+++	++	+	+	+++
21	+++	+++	+++	-	+++	+++
22	-	-	-	0	-	+
23	+++	+++	0	+	+++	+++
24	+++	+++	+++	+	+	+++
25	+++	0	+++	---	+++	+
26	+++	+++	+++	0	+++	+++
27	+++	+	+++	-	+++	+
28	+	+++	+++	-	+++	+++
29	+++	+	+	+++	--	-
30	+++	+++	+++	+++	+++	+
31	+	+++	+++	+	+	++
32	++	++	-	-	+	-

TABLE A8.3 Indications of improvement in sound localization errors from each participant after 12dB enhancement of the spectra of non-individualized HRTFs at the 6 frequency bands for all directions ('+' reductions in errors; '-' increases in errors; '0' no change in errors; 'nil' missing data). Data are the sums of improvement from the tables of 0^{o}, 45^{o}, 135^{o}, 180^{o}, 225^{o} and 315^{o}.

Participant	band 1 (0.2 – 0.69 kHz)	band 2 (0.69 – 2.4 kHz)	band 3 (2.4 – 6.5 kHz)	band 4 (6.5 – 10.0 kHz)	band 5 (10.0 – 14.0 kHz)	band 6 (14.0 – 22.0 kHz)
1	+++++	++++	+++	0	++++	+
2	0	++++++	++++	+	++++++	++++++
3	++	+++++	++++++	0	++++	++++
4	++++	++++	++	++++	++++++	++++
5	++++	++	++++++	++++	++++++	++++++
6	+++	+++	+++++	++++++	++++	++
7	+++++	+	++++	++	++++++	++
8	+++	++	0	-	+++	++
9	++++	++++	++	+++	++++	++
10	++++++	+++++	+++	+	+	++++
11	+++	+++	++	++	+++	0
12	++++	++	0	0	+++	++
13	++++	+++	++++	0	+++	+++
14	++++++	++++	++++	++	++++++	++++
15	++	++++	++++++	++++++	++	++++++
16	+++	++	++++	++	++++++	++++
17	+++	+	++++++	+++	++++	+++++
18	++	++++	+	---	++++	++
19	0	++	++	----	+++	++
20	++++	+++++	+++	+	+++	++
21	++++	++++++	+++	+	+++++	++
22	0	--	0	-	0	+++
23	++++	++++	+	0	++	++++
24	++++	++	+	++	++++	+++++
25	++++	-	+++	----	+++	--
26	++++	+++++	+++++	0	++++	++++
27	++++	0	++++	-	+++	++++
28	++	++++++	++++++	-	++++++	++++++
29	++++	++	++++	+++	+++	++
30	++++++	++++++	++++	++++	++++++	+++
31	+++	+++++	++	0	++	++++
32	++	+++	-	++	++++	++