AALBORG UNIVERSITY
INSTITUTE OF ELECTRONIC SYSTEMS
DEPARTMENT OF ACOUSTICS

TITLE: 3D-sound by means of binaural signals presented through loudspeakers
PERIOD: September 4th - December 19th, 2002
GROUP: 02gr960

Abstract

This report describes a project dealing with the implementation of a system, that playback binaural recordings through loudspeakers in a multimedia setup. The usual means of reproducing binaural sound is headphones, since no crosstalk exist between left and right channel. However in light of the advancement in the computational power, it is now possible to make a system, that can reduce the crosstalk from a standard pair of loudspeakers, and thereby recreating the 3D sound image.
A crosstalk cancellation system using the shuffler filter topology implemented. The system was set to work in a frequency range of 500-6000 Hz and the lower and higher frequencies was added into the processed sound without correction.
Simulations showed a channel separation, between the two ears, of 104.5 dB more than the natural channel separation from the shadowing of the head. Physical measurements showed a channel separation of 19.5 dB more than the natural channel separation.
The system was compared with headphones in a listening test, and the hypothesis of both systems being equal could not be rejected, which indicates that it is not possible with a confidence level of 95 % to conclude that the headphone representation is better than the implemented system.

AUTHORS:
Christian Corfits Jensen
Pablo Faundez Hoffmann
Hans Bjergegaard
Morten Jørgensen

SUPERVISOR:
Flemming Christensen

NUMBER OF COPIES: 8
NUMBER OF PAGES: 51 pages
HANDED IN: December 19th, 2002

Aalborg University • Fredrik Bajers Vej 7 • 9220 Aalborg Ø

Preface

This report documents the project work by group 02gr960 in the period from September 4th to December 19th 2002 at Aalborg University's Master of Science Acoustics program.

The aim of the project is to study theoretical and practical methods for performing crosstalk cancellation in a multimedia setup, for playback of binaural recordings.

In order to keep a consistent layout the report is made using three levels: chapter, section and subsection. Though only the first two levels are shown in the table of contents.

Readers of the report are expected to be familiar with digital signal processing as well as fundamentals of acoustics and psychoacoustics.

References to literature are denoted in brackets [ ]. Bibliography is referred to [Author, Year] including the name of the author and the year of either edition in case of books, or publication in case of papers, technical reports, etc.

References to figures are denoted by Figure x.x and equations as (y.y), where the first number is the chapter number and the second number is the figur/equation number in that chapter.

The artificial head used for all measurements in this project is "Valdemar".

A CD-ROM containing data from measurements as well as MATLAB® code and test files is enclosed in the back of the report. Furthermore a PostScript version of this report is included on the CD-ROM.

Aalborg, December 19th, 2002

Christian Corfits Jensen Pablo Faundez Hoffmann

Hans Bjergegaard Morten Jørgensen

Contents

1 Introduction and Project Overview **1**

1.1 Background 1

1.2 Scope of the Project 2

1.3 Multimedia Environment 3

2 Human Sound Perception **5**

2.1 Introduction to HRTF 5

2.2 Localization of Sound Sources 6

2.3 Summary 9

3 Crosstalk Cancellation Algorithms **11**

3.1 Atal-Schroeder Crosstalk Canceller 13

3.2 Shuffler Crosstalk Canceller 15

3.3 Recursive Crostalk Canceller 16

4 Bandwidth for Crosstalk Cancellation **19**

4.1 Crosstalk Basics 19

4.2 Considerations about Bandwidth 21

4.3 Choice of Bandwidth 24

5 System Implementation **27**

5.1 Specifications and Considerations 27

5.2 Inverse Filtering 28

5.3 Filter Design 30

5.4 Simulations 32

5.5 Final Implementation 36

6 Listening Test **41**

6.1 Aim of Test 41

6.2 Method 41

6.3 Results 43

7 Conclusion **47**

7.1 Further Investigations 48

7.2 Improvements 48

Bibliography **51**

A Measurement Report **55**

A.1 Purpose of the Measurements 55

A.2 Measurement Setup 56

A.3 Measurement Results 59

A.4 Conclusion on Measurements 63

B Head Related Transfer Function Calculation **67**

B.1 HRTF Measurements 68

B.2 Gain Correction 68

B.3 Frequency Domain Processing 68

B.4 DC Correction 69

C Binaural Test Signals **71**

C.1 Source Signal 71

C.2 Recording Method 72

C.3 Recording Environment 72
C.4 Recording Equipment 72
C.5 Summary 73

CHAPTER 1

Introduction and Project Overview

This report is about binaural reproduction using a multimedia setup with two loudspeakers. This introduction serves as a short survey on binaural technique and the problems related to the use of loudspeakers in binaural reproduction.

1.1 Background

There are several techniques for playback of sound: monoaural, stereophonic, surround and binaural sound reproduction. This report revolves about binaural sound. The basic principle behind binaural sound is depicted on Figure 1.1.

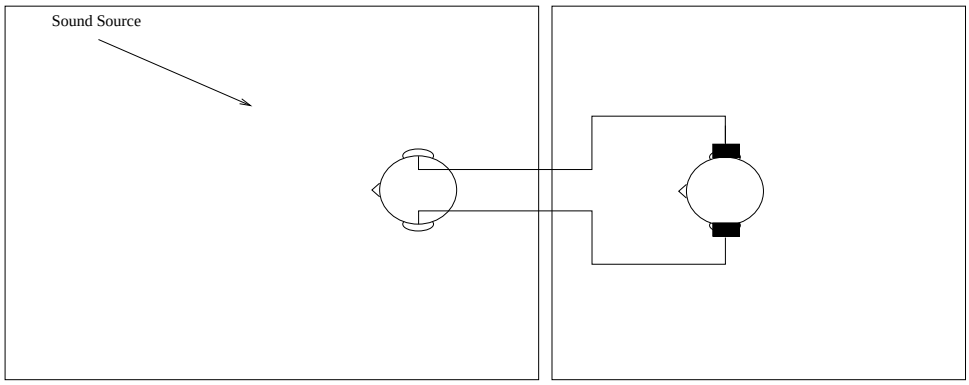

Figure 1.1: The principle behind binaural sound.

As seen the recording is made with a special recording technique. The microphones are located inside the ear of a person or an artificial head. This means that the microphones pick up the exact same signals as the person would hear without the microphones placed.

When playing this signal back at the same place it was recorded (at the ear canal) the soundfield of the original recording will be reproduced 1:1. This gives the most accurate high fidelity system possible. The problem with binaural sound reproduction is that in order to playback the left signal to left ear, and right signal to right ear, the method of playback must be by the use of headphones. Using loudspeakers will give interference (crosstalk) because left loudspeaker plays to both left and right ear and the same goes for the right loudspeaker.

In the recent years the development in computational power has reached a level where there is enough computational power to use digital signal processing algorithms for cancelling the crosstalk from the loudspeakers. These crosstalk cancellation methods reduce the crosstalk from the loudspeaker, thereby recreating the binaural sound at the listeners ear.

The aim of this project is to make a system that can reproduce binaural sound through a pair of loudspeakers in a multimedia setup,consisting of: a desk, a computer, a pair of PC loudspeaker and a computer monitor.

There are a number of reasons for choosing a multimedia setup. First: the implementation is somewhat easier than a standard stereo setup, because the listener usually is situated in a fixed position in front of the screen and relatively close to the loudspeakers. This makes the system more static. Second: a multimedia setup has the required computational power needed in the implementation of this binaural reproduction system. Third: binaural sound is entering the computer games. For example the new PC game "Hitman 2" features binaural sound.

1.2 Scope of the Project

The scope of this project is to investigate whether or not it is possible to recreate the binaural experience using a multimedia setup.

The hypothesis set in this project is:

Is there a significant difference between the perception of binaural sound when comparing a crosstalk cancellation system situated in the anechoic chamber with headphones?

The method of measuring the perception of sound, will be explained in the listening test section.

Before choosing and designing different solutions , it is necessary to investigate the human hearing, with focus on the binaural sound perception. This knowledge will also be used during the development of a listening test for evaluating the system.

A survey on the previous work on binaural sound reproduction using loudspeakers will be performed, leading to possible solutions for the system. This will lead to the implementation of a system, used for reproduction of 3D sound. However, first the multimedia environment is described.

1.3 Multimedia Environment

As mentioned above, it is decided to consider a standard multimedia setup as the playback system in the implementation of a system that can reproduce binaural sound. The chosen setup consist of a computer monitor and a keyboard placed on a computer desktop. The loudspeaker system used is a subwoofer/satellite loudspeaker system, which was chosen because it is very often used in multimedia setups. The loudspeakers used are Creative SoundWorks SW320, and measurements of the loudspeakers and setup can be found in appendix A on page 55.

The setup is depicted on Figure 1.2, and Figure 1.3:

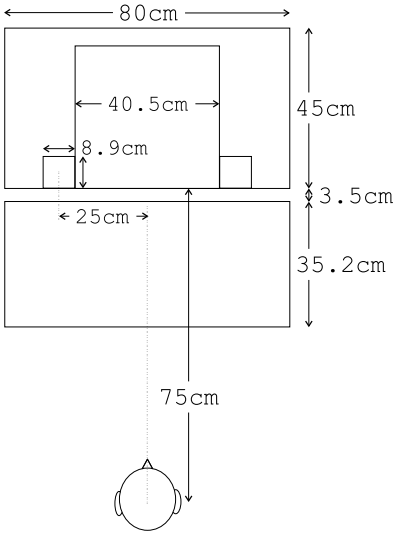

Figure 1.2: *Setup seen from the top.*

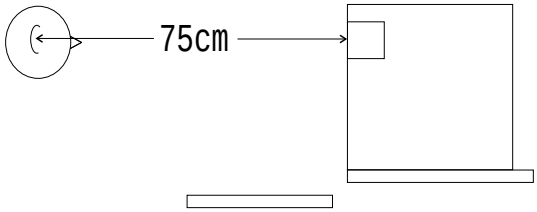

Figure 1.3: *Setup seen from the side.*

As it can be seen on the figures, the loudspeakers are mounted on the sides of the monitor, facing forward. This means that when a person is listening to the system, the loudspeakers will be at the same height as the ears of the listener and about 20° off axis. This was chosen as a compromise between difference in distance from each loudspeakers to each ear and the angle of the loudspeakers, because the optimal setup requires different distances between the loudspeakers according to which frequency is considered [Ward and Elko, 1999]. This is also explained in chapter 4. The angle of the loudspeakers should also be considered: the loudspeakers should ideally be on axis for high frequencies and off axis for low frequencies [Nielson and Koch-Henriksen, 2000].
Since the satellite loudspeakers only has one driver unit each, it is impossible to fullfill these demands and therefore, the solution where the satellite loudspeakers were mounted on the sides of the monitor was chosen. It was also considered to resemble a real life

situation quite good. Furthermore, due to the radiation pattern of the louspeaker, this was also considered to give a good natural crosstalk cancelation.

CHAPTER 2

Human Sound Perception

This section revolve about human sound perception. This topic is however very extensive, and only a brief overview is done, focusing on the sound perception features that are significant in creating a crosstalk cancellation system. This is primarily the way humans loalize sound from different directions. For information about the human hearing, the book [Hougaard et al., 1995] will give a good introduction.

2.1 Introduction to HRTF

In order to investigate the acoustics at the ears, the term:"Head Related Transfer Function" (HRTF) is often used. The HRTF is defined as:

$$HRTF(\phi, \theta) = \frac{P_{ear}}{P_{free}} \tag{2.1}$$

where ϕ is the azimuth angle (in the horizontal plane), and θ is the angle in the median plane, P_{ear} is the pressure measured at the ear of the listener, and P_{free} is the free field measurement of the sound pressure in the center of the head with no head present. In the time domain representaion of this transfer function, all the information humans use to localize sound sources is present.

Figure 2.1 shows the HRTF and Head Related Impulse Response (HRIR) obtained from a sound source located to the left of an artificial head ($\phi = 90°$ and $\theta = 0°$).

This Figure shows that the sound differs very little in the low frequency area, which is because the wavelength is long. At higher frequencies the difference in response appear due to the frequency dependent shadowing of the head. At higher frequencies a lot of fluctuations can be seen, which is caused by the structure of the outer ear. For example

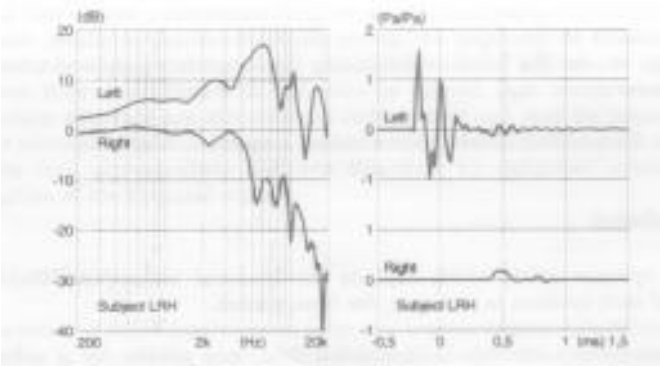

Figure 2.1: HRTF and HRIR measured with a sound source located beside the left ear. (This picture is from [Hammershøi, 1995] p.3)

the notches at 6 and 12 kHz at the left ear is due to the concha reflection. The broad peak at 4-6 kHz the resonance of the concha gives a peak [Gardner, 1998]. The frequency band which is used for the crosstalk cancellation system is discussed in depth in chapter 4 on page 19.

Looking at the time domain representation, it can be seen that the amplitude of the signal at the left ear is higher in amplitude and arrives sooner than to the right ear. This illustrates two of the most important cues for the localization of sound: the Interaural Level Difference (ILD) and the Interaural Time Difference (ITD).

2.1.1 Individual HRTFs

The artificial head used in this project is "Valdemar". This artificial head is designed to have the acoustic behaviour of a typical human being, which is a difficult task since the HRTF differs from person to person. It has been shown [Hammershøi, 1995] that the interindividual HRTFs change dramatically above 8kHz, which is depicted on Figure 2.2, where the deviation (see the bottom figures) increase with increasing frequency.

Also shown in Figure 2.2 is that the sound pressure measured with a blocked ear canal entrance has smaller interindividual deviation than when the sound pressure is measured with an open ear canal or at the eardrum. This is why "Valdemar" is constructed using the blocked ear canal principle.

2.2 Localization of Sound Sources

The information provided in the following section is taken from [Gardner, 1998].

The means available to humans in the localization of sound sources can be sorted into 2 groups: the interaural cues and the spectral cues.

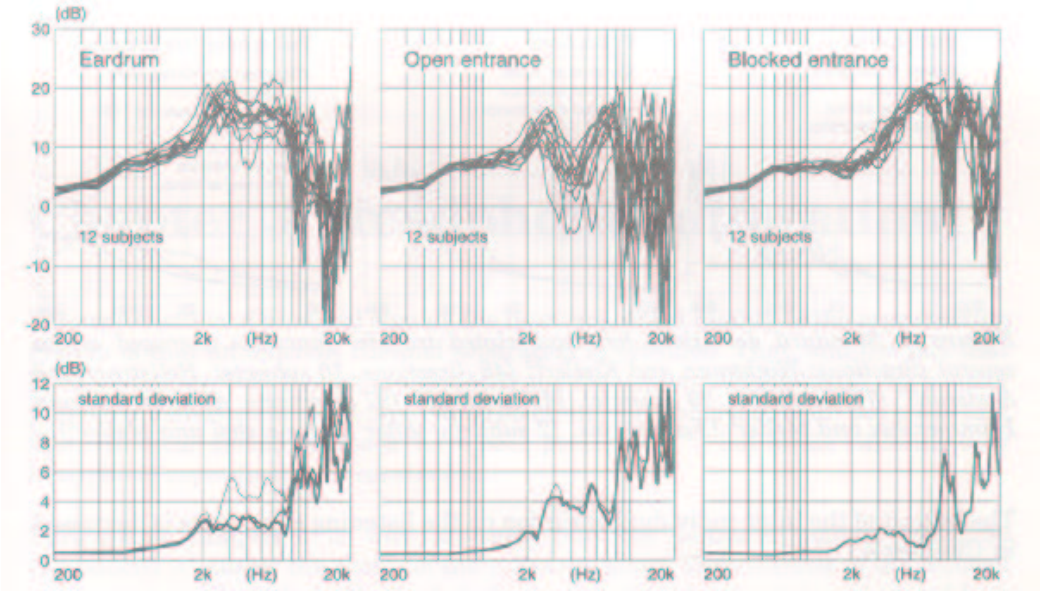

Figure 2.2: Head related transfer functions measured with both blocked and open earcanal. The bottom figures show the deviation [Hammershøi, 1995].

2.2.1 Interaural Cues

The Interaural Time Difference (ITD) and the Interaural Level Difference (ILD) are the main cues for sound localization, particularly when the sound source is located in the horizontal plane. The ITD is the difference in time from when a signal arrives at one ear, to when it arrives at the other ear. The ILD is the level intensity difference at the two ears.

The ITD cue is less dominant at higher frequencies due to the shadowing of the head but the ILD works at all frequencies. However, since there is little level difference at lower frequencies, the effect of ILD will go down as the frequency decreases. According to [Gardner, 1998] studies have shown that the ITD is the most dominant cue from 1500 Hz and downwards, and the ILD is the most dominant from 1500-6000 Hz.

2.2.2 Spectral Cues

The spectral cues are primarily generated by the physical environment of the outer ear and the pinna depending on the sound location. Spectral cues from the pinna contribute to localization in the median plane. A study by Hebrank and Wright [Gardner, 1998] have shown that the spectral cues are most dominant in the localization of sound sources in the median plane in the frequency range 4-16 kHz.

Even though it is possible to distinguish between localization cues, the overall conclusion is that all cues interact at all frequencies to give the best possible localization.

2.2.3 Minimum Audible Angle

When dealing with localization of sound sources, the term localization, and localization blur is often used.
Localization is a specific attribute in the sound that correlates with a response in an auditory event, and localization blur is the smallest change in a specific attribute of a sound event, that correlates with a response in an auditory event [Blauert, 1983].
Many attempts have been made to find the localization blur or the minimum audioable angle, and many different results have been presented. Figure 2.3 shows a study by Preibisch-Effenberger, Haustein and Schirmer, where the localization in the horizontal plane was investigated using white noise as a sound source.

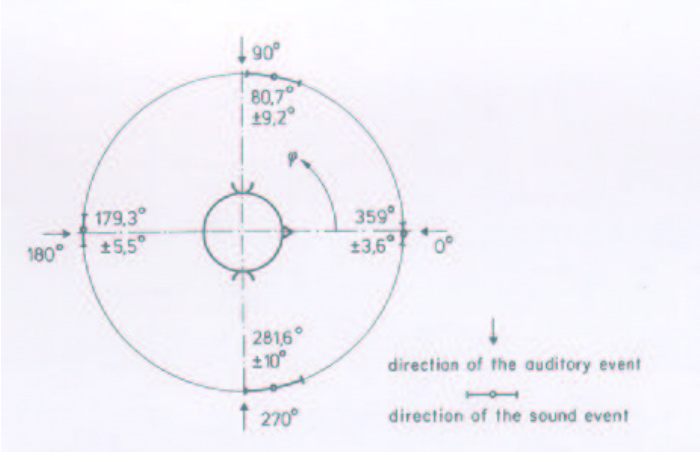

Figure 2.3: Localization blur in the horizontal plane. Studies carried out by Preibisch-Effenberger, Haustein and Schirmer in 1968-1970.

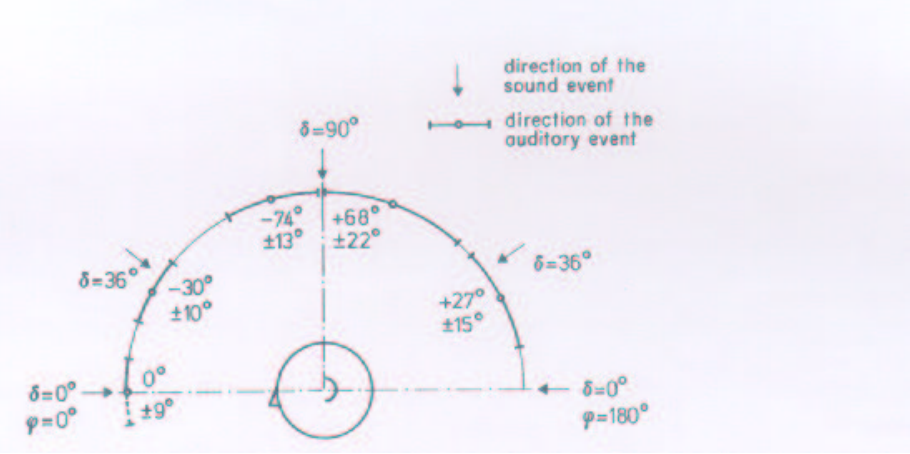

Figure 2.4: Minimum audible angle in the vertical plane, according to Damaske and Wagener in 1969.

Another study from Damaske and Wagener investigating localization in the median plane have lead to the results depicted on Figure 2.4. Here, speech signals were used as a stimuli, which means that Figure 2.3 and 2.4 are not directly comparable. As it can be seen the minimum audible angle is dependent on location. The localization blur is most dominant

when having sound sources in the median plane with no horizontal displacement. This is due to the fact that the spectral cues are the most dominant. The interaural cues (ITD and ILD) are almost zero.

2.2.4 Typical Localization Errors

The errors found when testing sound localization ability for humans can be split up into the following categories:

- Within-cone errors
- Out of cone errors
- Median-plane errors

The cone mentioned is the socalled "cone of confusion" [Møller et al., 1995]. This error is can appear from the fact that the interaural time difference in a cone around the ear is the same between the two ears. This gives a confusion area, since the there is only dominant spectral cues. Within-cone errors are therefore errors those that can appear on the cone of confusion. Out of cone errors are may occur where the cone of confusion is not active. A special case of the cone of confusion is the median plane, where the interaural cues are very small. This is called the median-plane errors. Median-plane errors and within-cone errors are usually most dominant for bandlimited signals. As shown by [Christensen, 2001] the effect of reducing the upper frequency limit to less than 8 kHz will increases the amount of median-plane errors and the within cone errors.

2.3 Summary

In the localization of sound sources both spectral cues and interaural cues must be considered. Ideally, the binaural recording/reproduction should be done with the same head in order to obtain good results. However, since this is usually not practically possible, an artificial head is often used. This gives a problem at higher frequencies, where the individual HRTF's differs the most. The localization blur is different depending on which angle the sound source is comming from.

CHAPTER 3

Crosstalk Cancellation Algorithms

The playback of a binaural recording is typically done using equalized headphones, in order not to add influence of the surrounding acoustic environment. Alternatively, playback can utilize a conventional stereo setup, which is often present in a standard multimedia system, which will obviously involve a more complicated playback scenario. In the following, a brief discussion about aspects that need to be considered and compensated for in this type of playback will be given.

Deficiencies of playback equipment The loudspeakers do not perform a linear electrical to acoustical transduction of the signal, meaning that some frequencies are represented with different amplitude than others. This can be compensated for by the use of inverse filtering applied to the signal corresponding to the transfer function of the loudspeakers.

Influence of the listener The presence of the listener will colorize the sound image by adding an additional HRTF to the binaural signal, which of course needs to be neutralized. This means that the inverse HRTF of the listener must be used to prefilter the signal.

Crosstalk This phenomenon occurs when the ear on one side of the head receive sound signal from a speaker on the other side. Physically, crosstalk can be removed by using sound absorbing panels in order to separate the two different sound fields related to each loudspeaker respectively. Other approaches are simply by using headphones or by applying signal processing implemented in crosstalk cancellation algorithms. The latter method will be used in this project.

To facilitate the comprehension of the crosstalk cancellation principle further explanation

is appropriate. Assuming a standard stereo setup in anechoic conditions, the acoustical information received by the ears of the listener will be a signal characterized by the frequency response of the measurement system, loudspeakers, the head and torso of the listener mainly. Suppose H_{LL} as the transfer function from the left loudspeaker to the left ear, H_{LR} as the transfer function from the left loudspeaker to the right ear and vice versa. If y_L and y_R correspond to the signals delivered to the left and right loudspeakers respectively, then the signals e_L received by the left ear and e_R received by the right ear, can be expressed in terms of the loudspeaker frequency responses as

$$\begin{bmatrix} e_L \\ e_R \end{bmatrix} = \begin{bmatrix} H_{LL} & H_{RL} \\ H_{LR} & H_{RR} \end{bmatrix} \begin{bmatrix} y_L \\ y_R \end{bmatrix} \tag{3.1}$$

where

$$\mathbf{H} = \begin{bmatrix} H_{LL} & H_{RL} \\ H_{LR} & H_{RR} \end{bmatrix} \tag{3.2}$$

is the acoustic transfer function matrix. This scenario is depicted in Figure 3.1.

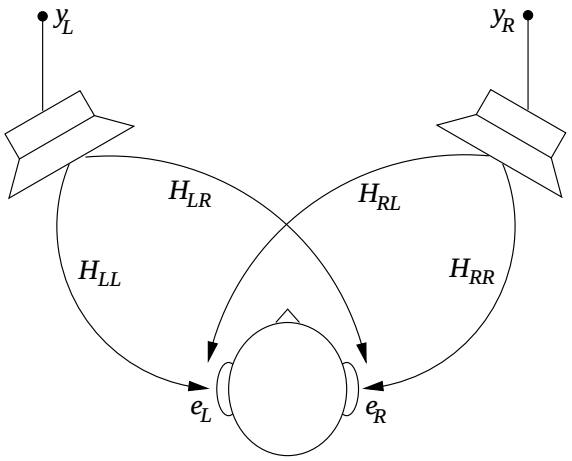

Figure 3.1: Model of a stereo playback with labels on individual transfer functions and signals.

Now that we have established our scenario, denote x_L and x_R as the corresponding left and right binaural signals that we want to reproduce at the corresponding left and right ear of the listener. Therefore, in order to deliver accurate binaural signals to the ears, the playback system will have to implement a crosstalk canceller $\mathbf{C}$ in the transmission path between x_L and x_R and the output from the loudspeakers y_L and y_R

$$\mathbf{x} = \begin{bmatrix} x_L \\ x_R \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} y_L \\ y_R \end{bmatrix}$$

$$\mathbf{y} = \mathbf{C} \cdot \mathbf{x} \tag{3.3}$$

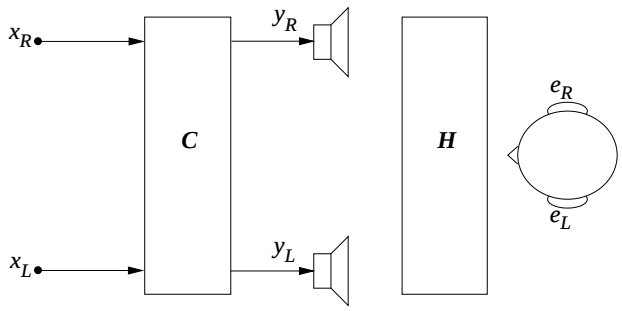

Figure 3.2: Binaural playback system including crosstalk canceller (C) and acoustic transfer function matrices (H).

as shown in Figure 3.2. The main function of this crosstalk canceller is to perform an inversion of the acoustical transfer function matrix **H**. Thus by using (3.1) and (3.2), (3.3) can be rewritten as:

$$\mathbf{e} = \begin{bmatrix} e_L \\ e_R \end{bmatrix}$$

$$\mathbf{x} = \mathbf{H^{-1}} \cdot \mathbf{H} \cdot \mathbf{e} \tag{3.4}$$

this way, the signal received by the ears correspond to the binaural signals. Now, the success in the implementation of a crosstalk cancellation system will strongly depend on the invertibility of the acoustical transfer function matrix. As explained before, this matrix contains the information of the loudspeaker response and the influence of the listener, therefore, a usual method to carry out a crosstalk cancellation involves the equalization of the loudspeakers, and the inversion of HRTF.
Crosstalk cancellation can be classified, depending on the implementation, under either asymmetrical or symmetrical conditions, regarding the listener position compared to the loudspeakers. This chapter is focused on the symmetric crosstalk cancellation situation which assume the same distance from the listener to each loudspeaker. Several techniques which exploit the symmetry of the listening situation have been proposed, discussed and implemented. In the following sections three different filter topologies for symmetric crosstalk cancellation will be reviewed.

3.1 Atal-Schroeder Crosstalk Canceller

In order to suppress the influencing phenomena described before, an effective inverse filtering scenario must be found. Assuming symmetry, the playback scenario at hand is sketched in Figure 3.3. The ipsilateral transfer functions are thought to be equal, then

$$H_{LL} = H_{RR} = S \tag{3.5}$$

and similarly for contralateral transfer fuctions

$$H_{LR} = H_{RL} = A \tag{3.6}$$

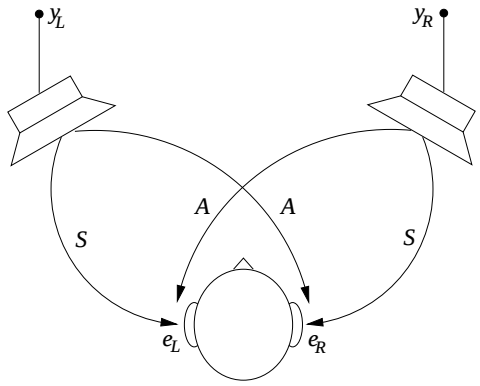

Figure 3.3: Stereo setup for playback. S is the ipsilateral transfer function and A is the contralateral transfer funtion.

The first crosstalk cancellation algorithm focusing on audio was presented by B.S. Atal and M.R. Schroeder in 1966 [Schroeder, 1975]. Their method was based on a symmetrical allignment of the loudspeakers and a good playback result was obtained in an anechoic environment. Figure 3.4 is a block diagram of how this algorithm is constructed.

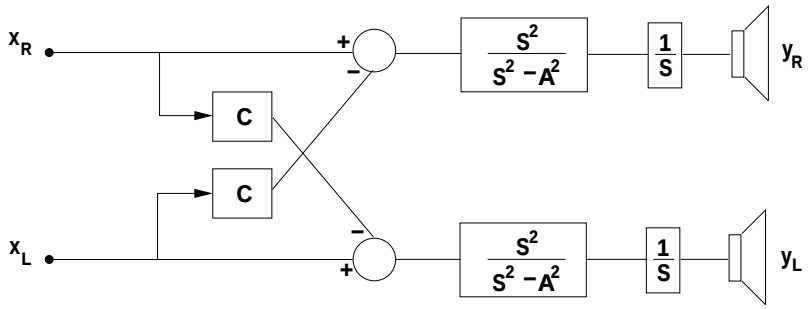

Figure 3.4: The Atal-Schroeder crosstalk cancellation filter array.

In this block diagram, C is the ratio between the contralateral transfer function and the ipsilateral transfer function

$$C = \frac{A}{S} \tag{3.7}$$

which by definition correspond to the interaural transfer function [Møller, 1992]. The signal presented at the ear should be x_R and x_L in order to remove the contralateral transfer funtion, (crosstalk). The relation between the left input source to the left loudspeaker terminal in the Atal Schroeder crosstalk cancellation algorithm is given in the expression

below.

$$y_L = \frac{S \cdot x_L - A \cdot x_R}{S^2 - A^2} \tag{3.8}$$

This is equivalent for the right channel given by

$$y_R = \frac{S \cdot x_R - A \cdot x_L}{S^2 - A^2} \tag{3.9}$$

The relation between the loudspeaker outputs to the ear of the listener are given by

$$e_L = S \cdot y_L + A \cdot y_R \tag{3.10}$$

$$e_R = A \cdot y_L + S \cdot y_R \tag{3.11}$$

This means that the relation between the left input signal to the signal present at the left ear is given by

$$e_L = \frac{S \cdot (S \cdot x_L - A \cdot x_R) + A \cdot (-A \cdot x_L + S \cdot x_R)}{S^2 - A^2} \tag{3.12}$$

which can be expanded to

$$e_L = \frac{S^2 \cdot x_L - A \cdot S \cdot x_R - A^2 \cdot x_L + A \cdot S \cdot x_R}{S^2 - A^2} \tag{3.13}$$

and that is equal to

$$e_L = \frac{(S^2 - A^2) \cdot x_L}{S^2 - A^2} = x_L \tag{3.14}$$

and it is thereby proven that e_L is equal to x_L, meaning that no coloration of the signal occurs. A similar relationship exist between x_R and e_R.

3.2 Shuffler Crosstalk Canceller

This system was implemented by [Cooper and Bauck, 1989] as a simplified structure of the Atal-Schroeder crosstalk canceller. The filter topology of a shuffler crosstalk canceller is based on two filters which operate on the sum and difference of the input binaural signals respectively. As depicted in Figure 3.5, these filters take the form of

$$F_L = \frac{1}{S - A} \tag{3.15}$$

$$F_R = \frac{1}{S + A} \tag{3.16}$$

so the output from the loudspeakers are

$$y_L = \frac{x_R + x_L}{S + A} - \frac{x_R - x_L}{S - A} \tag{3.17}$$

$$y_R = \frac{x_R + x_L}{S + A} + \frac{x_R - x_L}{S - A} \tag{3.18}$$

Thus the prefiltered signals received by the ears after the crosstalk occurs will be

$$e_L = S \cdot y_L + A \cdot y_R = 2 \cdot x_L \tag{3.19}$$

$$e_R = A \cdot y_L + S \cdot y_R = 2 \cdot x_R \tag{3.20}$$

where the factor of two represents a uniform gain and can be ignored [Kotorynski, 1990].

In order to succesfully implement these filters some requirements must be fullfilled. As F_L and F_R correspond to an inversion of the sum and difference of S and A, the transfer functions $S+A$ and $S-A$ must be minimum-phase to avoid instability in the inversion of this filters, and must share a common excess phase since a difference between them will reduce the crosstalk cancelling ability of the shuffler configuration [Kotorynski, 1990].

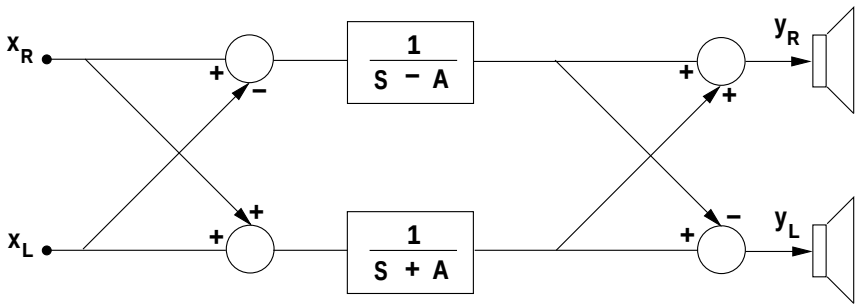

Figure 3.5: Shuffler crosstalk canceller topology.

3.3 Recursive Crostalk Canceller

The topology of a recursive crosstalk canceller is based on two filters. One filter corresponds to the inverse of the ipsilateral transfer function S. The other filter is represented by the interaural transfer function. The schematic of a recursive topology is shown in Figure 3.6. The major difference between this topology and the two discussed anteriorly, is that the design of the filters does not depend on the differences $S^2 - A^2$ and $S - A$ avoiding problems at frequencies where S and A are approximately equal.
The system equations corresponding to this topology are

$$y_L = \frac{1}{S} x_L - \frac{A}{S} y_R \tag{3.21}$$

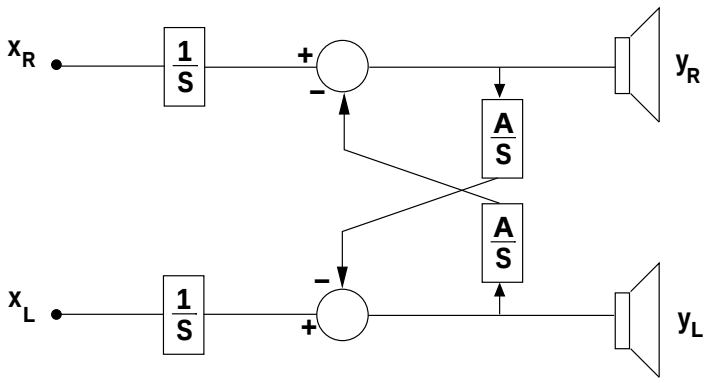

Figure 3.6: Recursive implementation of a symmetric crosstalk cancellation filter.

$$y_R = \frac{1}{S}x_R - \frac{A}{S}y_L \tag{3.22}$$

Now, by replacing (3.21) and (3.22) into (3.10) and (3.11) respectively, it can be proven that the desired signals x_L and x_R will be delivered at the ears of the listener ideally. Furthermore, as the cancellation is carried out by the interaural transfer function, it can be seen that this kind of topology feeds the cancellation signal back to the opposite channel's input allowing an automatic cancellation of higher order crosstalks [Gardner, 1998].

CHAPTER 4

Bandwidth for Crosstalk Cancellation

This chapter describes the considerations which were done in order to choose the optimum bandwitdh for the use in the crosstalk cancellation system.

4.1 Crosstalk Basics

In this section, the fundamental aspects of a crosstalk cancellation scheme are described. The scenario under consideration is a symmetrical setup with two loudspeakers and two microphones in a free field setup. It is illustrated in Figure 4.1.

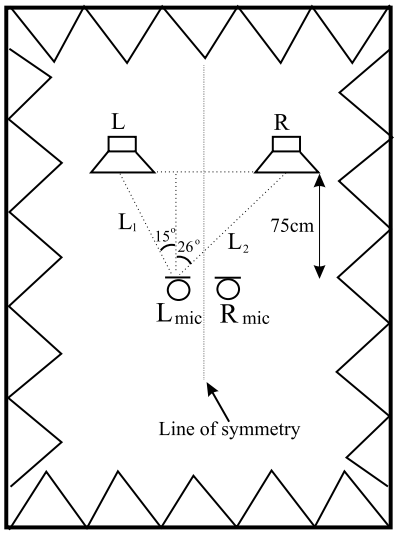

Figure 4.1: Setup used in the explanation of basic crosstalk cancellation.

On this figure the distance, L_1 can be calculated by:

$$L_1 = \frac{75cm}{cos(15°)} = 77.64 \quad cm \tag{4.1}$$

Correspondingly, L_2 is found to be 83.44 cm. This means that there is a difference of 5.8 cm in travelling distance, corresponding to a time delay, ITD of approximately 169 μsec. The signals in the following considerations are sinusoids, which are chosen in order to keep things simple.

Consider the case where it is desired to present a sound to R_{mic} and complete silence at L_{mic}. In this case, cancellation of sound at L_{mic} is desired, which means that L must radiate the same sound as R, in opposite phase and with a time delay of ITD. This is illustrated in Figure 4.2 with the loudspeaker diaphragm displacements as a function of time.

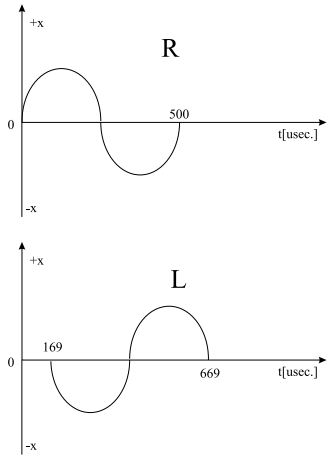

Figure 4.2: Diaphragm displacement versus time for generating a 2kHz sound at point R_{mic}.

From this it can be seen that in theory, it is possible to obtain crosstalk cancellation at the left ear for any frequency. This is however not desireable due to reasons that will be explained later in this chapter. In the considered case, the sound pressure at L_{mic} is the combination of sound from L and R arriving simultaneously and the sound pressure at R_{mic} is the combined response of L delayed by two ITD and R without delay. Expressed in formulas:

$$R_{mic} = R(0 \text{ sec}) + L(2 \cdot ITD \text{ sec}) \tag{4.2}$$

$$L_{mic} = R(ITD \text{ sec}) - L(ITD \text{ sec}) \tag{4.3}$$

From this it can be seen that the sound is cancelled at L_{mic} and at R_{mic} it is the sum of the R without delay and L with a time delay of 2 ITD. It can be found that in order to have perfect cancellation at L_{mic} and no cancellation at R_{mic}, 2 times the ITD should correspond to a phase shift of 180°. This would mean that the sound pressure at R_{mic} consist of two sounds of equal magnitude that are in phase, and at L_{mic} the resulting sound pressure (=0) would consist of two sounds with equal magnitude that are in antiphase. The frequency where optimal crosstalk cancellation occurs is found to be approximately 1.5 kHz. The effect of using tones with a lower frequency than this, is that the overall

phase difference between S and A at R_{mic}, named Δ-phase, which can be found using the following formula, is lowered.

$$\Delta\text{-phase} = 180° + \frac{2 \cdot ITD}{T} \cdot 360°. \tag{4.4}$$

, where T is the period of the sinusoidal sound.

4.2 Considerations about Bandwidth

In this section the band for which to do crosstalk cancellation is examined based on the physical outline of the setup.

4.2.1 Setup Without Listener

This subsection holds the investigation of the setup illustrated in 4.1. A graph showing the phase difference at R_{mic}, Δ-phase, as a function of frequency is illustrated in Figure 4.3.

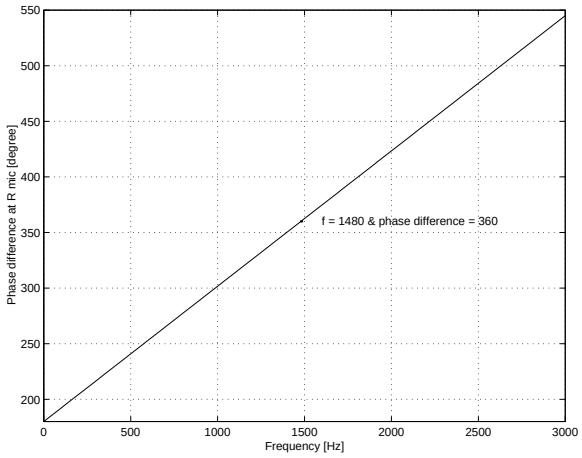

Figure 4.3: Phase difference between signals from R and L at R_{mic}.

At 1.5 kHz the signals are in exact phase (360°) resulting in a perfect cancellation at one ear and a doubling of the sound pressure at the other ear. At half the frequency(750 Hz), there will be a rise in amplitude of 3 dB at R_{mic} because Δ-phase is 270°. At 100 Hz for example, Δ-phase is approximately 192°, meaning that signals at both R_{mic} and L_{mic} will be almost equally cancelled. From 4.3 it can be seen that the ITD directly relates to the frequency for which optimal crosstalk cancellation occurs. The longer the ITD, the lower in frequency it is possible to do perfect crosstalk cancellation. The frequency for which optimal crosstalk cancellation is possible for the considered setup is 1.5 kHz, corresponding to an ITD of 169 μsec. The distance between the eardrums of a human is approximately 16cm (measured on the artificial head), which means that the maximum ITD for this dummy is 466 μsec. This means that the lowest frequency for which it

is possible to perform optimal crosstalk cancellation is approximately 536 Hz. In order to increase the ITD, the loudspeakers must be moved further apart, thereby increasing the difference in travelling time to one ear compared to the other. Different attempts focusing on loudspeaker spacing have been done, with the aim of getting better low frequency performance (for example: [Nielson and Koch-Henriksen, 2000]). Decreasing the ITD makes it physically impossible to equalize anything but high frequencies, but the advantage is that the sweetspot concerning the distance from the screen becomes larger. However, head movements sideways, destroying the symmetry of the setup, will be more critical for the sweetspot when the ITD is smaller, because the variation caused in the ITD is consirerably larger. In summary: The placement of the loudspeakers is fundamental for the performance of a crosstalk cancellation system:

Closely Spaced Loudspeakers Here, the ITD will be small and the lower limit for the crosstalk cancellation algorithm will be high. However, the system is stable for head movements towards or away from the screen. It is not stable regarding headmovements sideways.

Distantly Spaced Loudspeakers In this case, the ITD is large, which means that the lower limit for the crosstalk cancellation algorithm will be low. The system is then more sensitive to movements on the symmetrical plane (front/back) but less sensitive to head movements sideways. This means that loudspeaker spacing ideally should vary with frequency [Ward and Elko, 1999].

Taking advantage of the radiation pattern of the loudspeaker is also a possible method which can be used to passively increase the channel seperation. The simulated radiation pattern of the used satellite loudspeaker is illustrated in Figure 4.4. This simulation is based on the model of a point source radiation into a $2 \cdot \pi$ space. On the figure, the horizontal axis going from 0 to 180 °correspond to the baffle of the loudspeaker.

The magnitude of the mainlobe has a maximum at 94 dB SPL rel. 20 μPa. The first order sidelobes are in opposite phase to the direct sound and the second order sidelobes are in phase with the direct sound. Using this knowledge it is possible to construct a setup which is optimal for a specific set of loudspeakers in one frequency band. This method is very complicated, since the radiation pattern changes for different frequencies.

4.2.2 Setup With Crosstalk Cancellation Applied

In order to estimate an appropriate lower cutoff frequency, Δ-phase combined with the characteristics of the ITF should be considered. Therefore, the two microphones on the drawing (Figure 4.1 on page 19) are replaced by two eardrum entrances of a human listener, located in the horizontal center of the draving. Next, consider the case where a white noise signal convolved with the ipsilateral transfer function should be presented to the right ear of a listener. No signal should be presented to the left ear of the listener, which means that the following signals should be radiated by the loudspeakers.

- R = white noise.

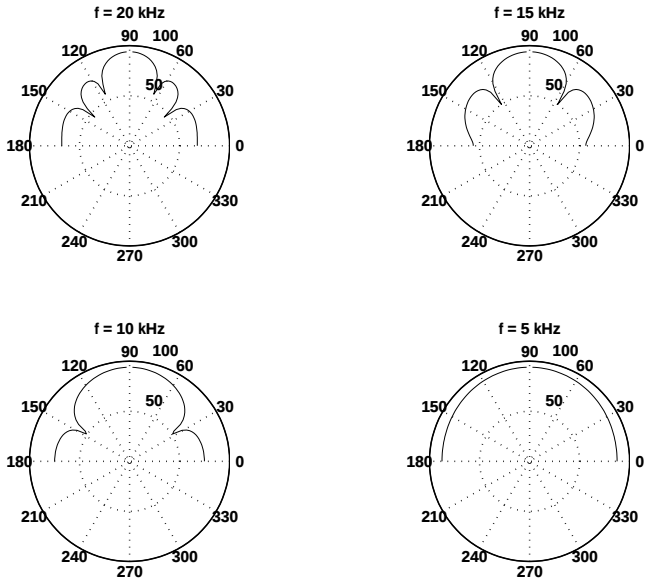

Figure 4.4: Simulated radiation pattern for the satellite loudspeaker at different frequencies.

- L = white noise $\cdot ITF$.

This means that the transfer function from L to left ear will be given by the ipsilateral transfer function convolved with the ITF resulting in a contralateral transfer function, appearing at the left ear. This signal is identical to the contralateral transfer function which appear due to R. The signal produced by the left loudspeaker should be delayed with the amount of ITD and have the opposite phase as that coming from R. This is illustrated in the time domain on the pictures below (SP is Sound Pressure).

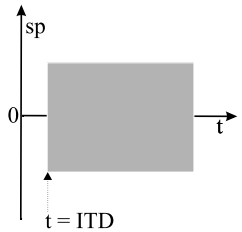

Figure 4.5: *Sound radiated from L.*

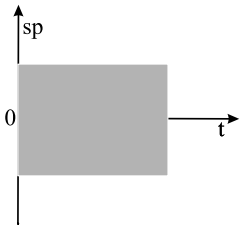

Figure 4.6: *Sound radiated from R.*

These sounds reach the left ear at time ITD + TD, where TD is the travelling distance time. The arrival time at the right ear is TD + $2 \cdot ITD$ for the sound coming from L and TD for the sound from R. The arrival of sound from L and R at the right ear is illustrated in the pictures below (Figure 4.7 and 4.8).

These signals are illustrated in the frequency domain in Figures 4.9 and 4.10 below, where TF is short for Transfer Function.

The signals at the left ear are:

- From R comes the contralateral signal = white noise $\cdot TF_{contralateral}$.

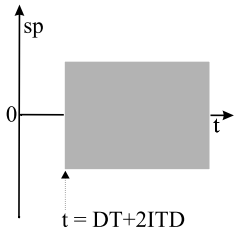

Figure 4.7: *Sound from L at right ear*

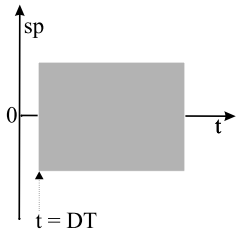

Figure 4.8: *Sound from R at right ear*

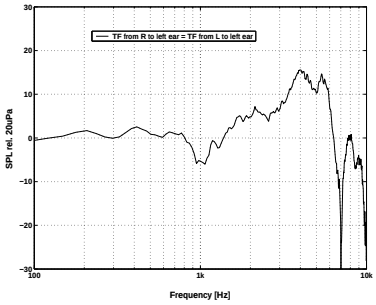

Figure 4.9: *TF's to the left ear*

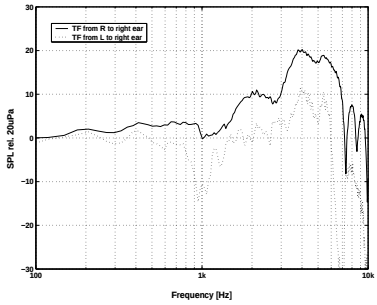

Figure 4.10: *TF's to the right ear*

- From L comes white noise $\cdot TF_{ipsilateral} \cdot ITF$.

These signals arrive simultaneously and have opposite phase resulting in a cancellation of the sound.

At the right ear the signals are:

- From R comes the ipsilateral signal = white noise $\cdot TF_{ipsilateral}$
- From L comes white noise $\cdot TF_{contralateral} \cdot ITF$

Here, the signal from L is delayed by $2 \cdot ITD$ and have opposite phase to the signal from R.

The resulting sound pressure at the right ear compared to the desired is illustrated in Figure 4.11.

On this plot it can be seen that the sound pressure at the right ear decreases due to the "required" crosstalk cancellation at the left ear.

4.3 Choice of Bandwidth

In this particular crosstalk cancellation system it is desired to obtain a optimally perceived channel separation. Since interindividual ITF's has most similar characteristics in the frequency range below 8 kHz, the best performance for arbitrary chosen listeners would probably be obtained if the upper cuttof frequency for the crosstalk cancellation algorithm

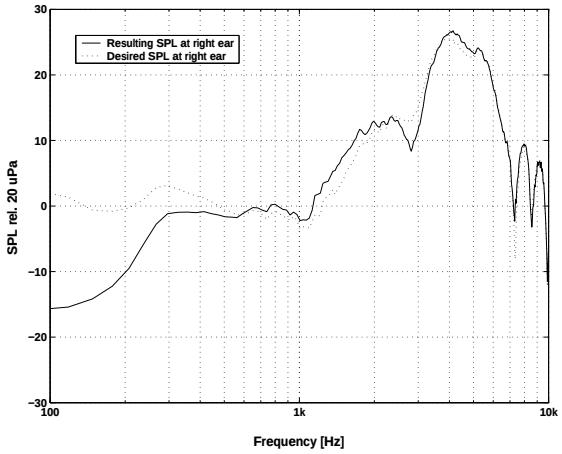

Figure 4.11: The resulting SPL compared with the desired at the right ear.

is chosen below this. The upper frequency for the crosstalk cancellation algorithm is chosen to be 6 kHz, and it is thereby possible to use downsampling by a factor of three which increases the power of the filter efficiency. The advantage of using 8 kHz as the upper limit would probably be minimal, since 7.5 kHz is the frequency where the two signals are in phase again, thereby providing the best crosstalk cancellation.

The lower cuttof frequency is chosen to be 500 Hz based on the figure 4.11 which indicate that the SPL at the right ear is lowered about 3 dB at this frequency, which is considered to be an acceptable attenuation considering the limitations of the human hearing, which, according to the equal loudness contour, is not that sensitive to sound pressures at 500 Hz. If the low frequency limit is decreased, the sound at both ears will be lowered for low frequencies, thus destroying the energy content of the signal.

The choice of band in which to perform crosstalk cancellation also satisfies the loudspeaker system which was chosen for the system. It is a loudspeaker system of modest quality with only one loudspeaker driver unit per satellite thereby, and therefore they are not good at representing the entire audioband (low and high frequencies, specifically).

CHAPTER 5

System Implementation

This chapter presents and describes the implementation of a crosstalk cancellation system. The implementation involves the choice of filter topology, filter design methodology for inverse function calculation, simulation and physical validation of the crosstalk cancellation performance.

5.1 Specifications and Considerations

As shown in Figure 5.1 the crosstalk cancellation system is intended to be implemented for a multimedia setup consisting of an ordinary desk and a desktop computer with two satellite loudspeakers located at the same high of the users' ears (see the CDROM for more information about the loudspeakers). Based on this configuration a symmetric crosstalk cancellation system was assumed, therefore, the shuffler structure proposed by [Cooper and Bauck, 1989] is chosen to be appropriated for the implementation of the crosstalk cancellation system.

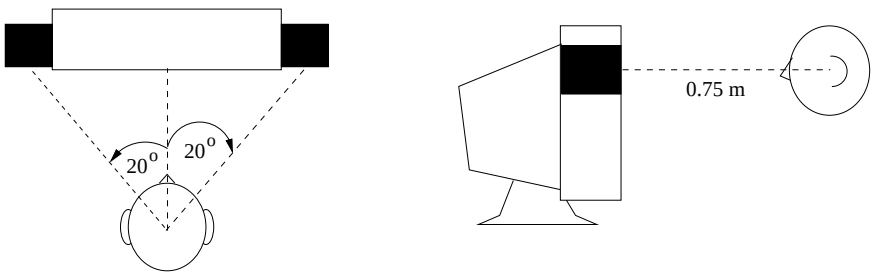

Figure 5.1: Schematic of the multimedia setup configuration.

The system configuration is planned to be optimized for just one position of the listener, the "sweet spot". Therefore, no dynamic systems that include head tracking are considered.

The crosstalk cancellation system will be implemented for a frequency range from 500 Hz to 6 kHz. This decission is based on a review of how the crosstalk cancelling frequency range is affected by the loudspeaker response, loudspeaker position and inter-individual HRTFs.

Since this project deals with the implementation of a prototype system for binaural playback, realtime constrains are not considered. This situation generates the possibility to separate the filtering and playback process. In addition, the filtering can be done in ordinary computers by utilizing MATLAB® as the implementation language. Furthermore, as there are no requirements for the length of the filters, longer filter could be implemented and applied. Since the playback is thought for a standard multimedia setup the obtained measurement were resampled from a sampling frequency of 60.61 kHz (used by the MLSSA setup during measurements) to 44.1 kHz (a ratio of 8/11 was used).

As already stated, the main purpose of the system is to deliver binaural signals to the listeners ears. This means that the crosstalk cancellation system should take into account the HRTF, the response of the loudspeakers and the reflection from the multimedia setup mainly. Due to these reflections longer filters might be applied in order to obtain a better crosstalk cancellation in terms of the magnitude differences in the frequency domain.

5.2 Inverse Filtering

When facing inverse filtering of binaural signals, it has been shown that good results can be accomplished by using least square methods [Sæbo, 2001]. One of these methods is the Wiener filter which tries to approximate an inverse by minimizing the square of the error between a desired signal and an estimation of it.

Consider the situation shown in Figure 5.2. It is desired to design a FIR filter with impulse

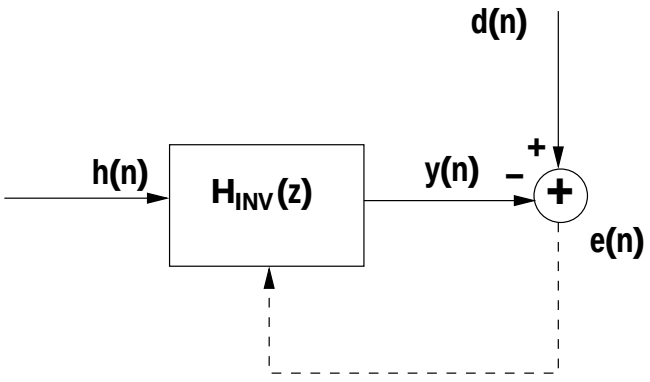

Figure 5.2: Configuration for the calculation of the Wiener filter.

response $h_{INV}(n)$ and transfer function $H_{INV}(z)$ such that

$$h(n) * h_{INV}(n) = \delta(n) \tag{5.1}$$

$$H(z)H_{INV}(z) = 1 \tag{5.2}$$

are satisfied. If b_k, $k = 0, 1..., M$, corresponds to the FIR filter coefficients then the approximate inverse impulse response $h_{INV}(n)$ is given by

$$h_{INV}(n) = \sum_{k=0}^{M} b_k z^{-k} \tag{5.3}$$

Now, as the estimation signal $y(n)$ is obtained from the convolution between the coefficients b_k and the input signal $h(n)$ then the error signal due to this approximation is given by

$$e(n) = d(n) - \sum_{k=0}^{M} b_k h(n-k) \tag{5.4}$$

thus the expression for the total squared error is

$$E = \sum_{n=0}^{\infty} [d(n) - \sum_{k=0}^{M} b_k h(n-k)]^2 \tag{5.5}$$

The total squared error E assumes a minimum if its partial derivatives with respect to the filter coefficients b_k are zeros

$$\frac{\partial E}{\partial b_k} = 0 \tag{5.6}$$

therefore, upon differentiating E with respect of b_k

$$\frac{\partial E}{\partial b_k} = -2 \sum_{n=0}^{\infty} d(n)h(n-l) + 2 \sum_{n=0}^{\infty} \left(\sum_{k=0}^{M} h(n-k) \right) h(n-l) \qquad l = 0, 1, ..., M \tag{5.7}$$

and by combining (5.7) with (5.6) and rearranging we obtain

$$\sum_{k=0}^{M} b_k \sum_{n=0}^{\infty} h(n)h(n-l) = \sum_{n=0}^{\infty} d(n)h(n-l) \tag{5.8}$$

where

$$\mathbf{p}_{dh}(n) = \sum_{n=0}^{\infty} d(n)h(n-l) \tag{5.9}$$

corresponds to the crosscorrelation vector, and

$$\mathbf{R}_{hh}(n) = \sum_{n=0}^{\infty} h(n)h(n-l) \tag{5.10}$$

corresponds to the autocorrelation matrix. Thus the optimal Wiener filter is given by

$$\vec{b} = \mathbf{R}_{hh}^{-1}(n)\mathbf{p}_{dh}(n) \tag{5.11}$$

where $\vec{b}$ is the column vector of the FIR filter coefficients.

5.3 Filter Design

The design of crosstalk cancelling filters corresponding to the shuffler topology, consists of computing the inverse of the sum and difference between the ipsilateral and contralateral signals. These signals are obtained from an elevation angle of 0° and an azimuth angle of 20°. The equations for these filters expressed in the frequency domain are given by

$$F_{INVSUM} = \frac{1}{S + A} \tag{5.12}$$

$$F_{INVDIF} = \frac{1}{S - A} \tag{5.13}$$

where $S = H_{LL} = H_{RR}$ and $A = H_{LR} = H_{RL}$ (see figure 3.1 on page 12) since symmetry is assumed and desired. To calculate these filters the algorithm of equation (5.11) is implemented and a bandpassed unit impulse sequence between 500 Hz and 6000 Hz with a certain delay is used as the desired signal. This delay is applied since the ipsilateral and contralateral responses are not exactly minimum phase which means that by allowing a delay, letting the desired response $d(n) = \delta(n - delay)$, a better approximation of the inverse can be achieved [Sæbo, 2001].

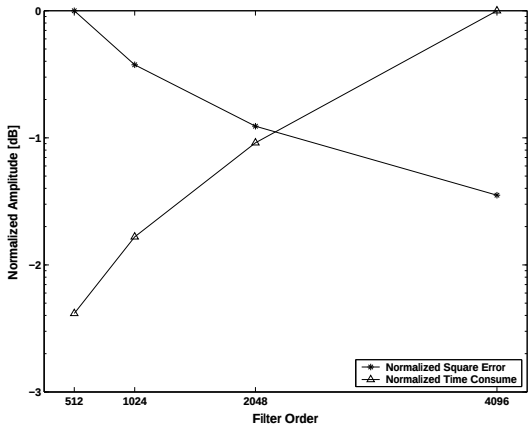

Figure 5.3: Comparison between the summed squared error and the computational complexity as a function of the filter order for the Wiener solution. Both parameter were obtained as an average from the computation of the shuffler filters based on pure HRTF for a direction of 20°. These calculations were made for 512, 1024, 2048 and 4096 filter order.

In the ideal case S and A correspond to HRTF only and a flat response from the loudspeakers with no reflections from the setup is assumed. In order to have a more realistic approach, simulation of the crosstalk cancellation system is divided in three cases. The first case calculates the shuffler filters from pure HRTF and one filter is implemented to equalize the loudspeaker response and setup reflections at one point. The second case calculates the shuffler filters from a combination of HRTFs and loudspeaker response, having an extra filter to compensate for the setup reflections. The third case calculates the shuffler filters assuming that S and A have all the acoustical information conformed by the HRTFs, loudspeaker response and setup reflections. For the three cases the shuffler filters are optimized for the frequency range from 500 Hz to 6000 Hz and the equalization filters

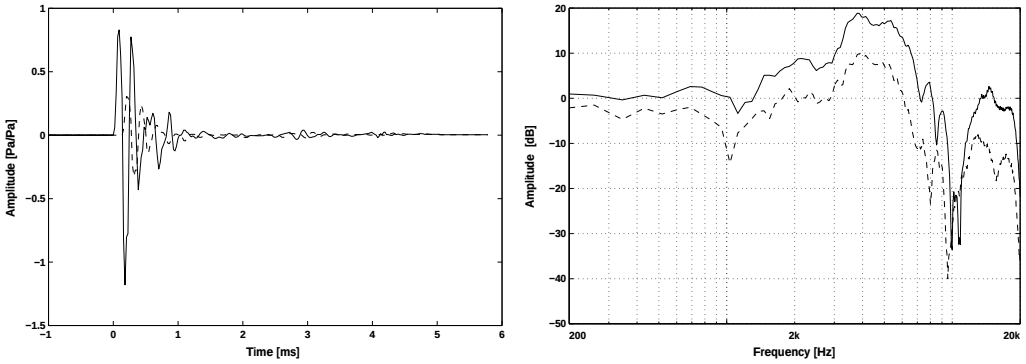

Figure 5.4: Head related impulse response and the corresponding HRTF for an elevation angle of 0°and azimuth angle of 20°. These correspond to the ipsilateral (solid line) and contralateral (dashed line) signals used in the shuffler filters.

are calculated trying to compensate for the audible frequency range. The length of the inverse FIR Wiener filters is set as a trade off between the summed squared error and the computational complexity. Figure 5.3 shows a comparison between this two parameters where the computational complexity is measured as the time needed by a SUN Unix machine to compute the shuffler filters as a function of the filter order. From this comparison the filter order is set to 2048, which gives a frequency resolution of $\Delta f = \frac{44100}{2048} \approx 21$Hz.

5.3.1 HRTFs Calculations

The HRTF are obtained by measuring the responses of the artificial head from a loudspeaker in anechoic conditions. The loudspeaker consisted of a single driver mounted in a hard plastic ball and this loudspeaker was located in the same horizontal plane as the ears of the artificial head. To ensure optimal accuracy in the measurements the artificial head was mounted in a mechanical turning device where the azimuth angle could be adjusted. The procedure to obtain the HRTF from the measurements is explained in appendix B on page 67. Prior the calculation the shuffler filters, the linear phase component of the HRTF is removed and the ITD is added to the contralateral impulse response. The ITD value is calculated by determining at which time the ipsilateral and contralateral impulse responses become non zero. Since the values measured at the beginning of the signals are not identically zeros, it is assumed that the first value which exceeds 1 % of the maximum absolute value of the signal will be considered as the first relevant sample. The sample number found for the ipsilateral impulse response is then substracted from the sample number of the contralateral impulse response and the ITD is estimated [Sandvad and Hammershøi, 1994]. The time domain representation of the HRTF is windowed by a rectangular window to obtain a pair of 256 coefficients impulse responses which are shown in Figure 5.4.

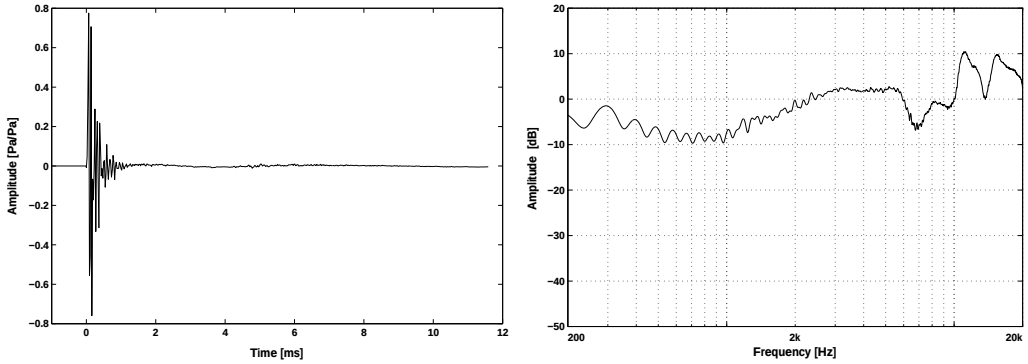

Figure 5.5: Impulse and frequency response of the satellite loudspeaker used in the multimedia setup. This measurement was done on axis and the ripples seen at low frequencies are due to the truncation of the signal.

5.3.2 Loudspeaker Responses and Setup Reflections

From the first set of measurements (see appendix A on page 55 for more details), one measurement in the center of the artificial head with the head absent was done. This measurement defined as $P1$ contains the information from the loudspeaker, and the reflections from the desk, monitor and keyboard mainly. Furthermore, measurements of the left and right loudspeaker responses were done on axis and at the same location of the closer ear (referenced to 15°) and the farther ear (referenced to 26°) with the head absent. In this measurement, shown in Figure A.7 on page 60 in appendix A, it is useful to observe that the variations between the on axis and off axis loudspeaker responses have noticeable differences outside the frequency range of interest (500 - 6000 Hz). Along with the loudspeaker response, an impulse response containing information of the setup reflections is also desired. To obtain this response the inverse of the on axis loudspeaker impulse response is calculated by using the algorithm of equation 5.11 and then convolved with $P1$, thereby an approximation of the setup impulse response is obtained. Prior to calculating the inverse filter, the linear phase component of the on axis loudspeaker response is removed and the resulting signal is windowed with a square window reducing the length of the signal to 512 taps. As symmetry is assumed, the equalization task is performed only for one point corresponding to the center of the head. The impulse and frequency response of the loudspeaker and setup reflection is shown in Figures 5.5 and 5.6 respectively.

5.4 Simulations

In this section, the performance of the crosstalk cancellation is evaluated by simulating the acoustics of the listening situation. As explained in section 5.3 three different cases are presented. To evaluate the effectiveness of the crosstalk cancellation, a frequency dependent description named channel separation is used. Channel separation is calculated by measuring the ipsilateral and contralateral impulse responses and then computing the ratio between the contralateral (numerator) and ipisilateral (denominator) frequency

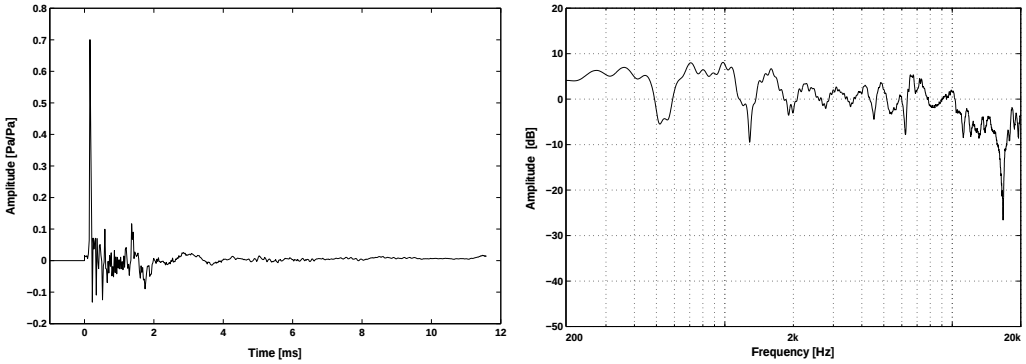

Figure 5.6: Calculated impulse and frequency response corresponding to the setup reflections. From the frequency response is possible to observe slight comb filter effects which might correspond to reflections from the desk mainly.

responses. The acoustic path between the loudspeakers and the listener is given in the frequency domain by

$$S = TF_{IPSI} \cdot H_{LSPK} \cdot H_{SETUP} \tag{5.14}$$

$$A = TF_{CONT} \cdot H_{LSPK} \cdot H_{SETUP} \tag{5.15}$$

where TF_{IPSI}, TF_{CONT}, H_{LSPK} and H_{SETUP} correspond to the ipsilateral and contralateral transferfunctions, the loudspeaker response and the response representing the reflections from the setup (Figures 5.4, 5.5 and 5.6).

To verify the performance of the crosstalk cancelling filters, an artificial signal corresponding to a positive pulse will be sent through the left channel and zero signal through the right channel. Only the left-right channel separation is considered since symmetry is assumed. As the frequency range of interest is 500 - 6000 Hz, the input signals to the shuffler filters are previously filtered by a bandpass FIR filter from 500 to 6000 Hz.

It can be seen that when no crosstalk cancellation is applied and a positive pulse is the input signal to one channel, the resulting channel separation correspond to the ITF which represents the natural crosstalk cancellation due to the head shadowing. Therefore a comparison between the ITF and the channel separation obtained when crosstalk cancellation is implemented will be the approach of the simulation. Furthermore, and in order to represent the crosstalk cancellation by one number, an average of the differences in dB between the ITD and crosstalk cancellation in the frequency range of interest is computed. The equation that describes this operation is given by

$$CC_{AVG} = E[20 * log(|ITF_{500-6kHz}|) - 20 * log(|CC_{500-6kHz}|)] \tag{5.16}$$

where CC is the channel separation due to the artificial crosstalk filtering and $E[\cdot]$ correspond to the mean value. CC is called artificial crosstalk cancellation to establish the difference with the natural crosstalk cancellation due to the head shadowing which is described by the ITF.

From these results it is expected to select one case in order to be implemented for a physical and psychophysical validation.

5.4.1 Case I

In this case the calculation of the shuffler filters (5.12) and (5.13) are based on pure HRTF. The equalization for the loudspeaker responses and setup reflections is carried out separately by convolving the loudspeaker response with the reflection response and posteriorly calculate the inverse of it, instead of approximate inverse functions for each response. Therefore this configuration is based on three filters. The topology for this case is depicted in Figure 5.7.

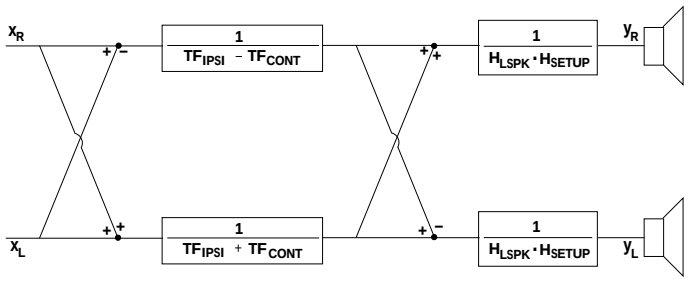

Figure 5.7: Shuffler configuration for case I.

5.4.2 Case II

In this case the calculation of the shuffler filters (5.12) and (5.13) are based on a combination between the HRTF and the loudspeaker response. The head related impulse responses are convolved with the same loudspeaker response and the resulting signals are used for computing the shuffler filters. An extra filter is calculated in order to equalize the response corresponding to the setup reflections. Therefore this configuration is also based on three filters. The topology corresponding to this case is depicted in Figure 5.8.

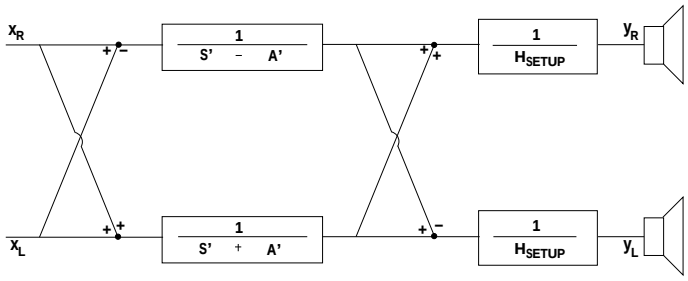

Figure 5.8: Shuffler configuration for case II. In this case the shuffler filters are based on $S' = TF_{IPSI} \cdot H_{LSPK}$ and $A' = TF_{CONT} \cdot H_{LSPK}$.

5.4.3 Case III

In this case the calculation of the shuffler filters (5.12) and (5.13) are based on combination of HRTF, loudspeaker response and setup reflections response. Therefore all the acoustical

information is contained in the ipsilateral and contralateral impulse responses and no extra filters are required for equalization. Thus the filter structure for this case can be implemented by two filters. The topology for this case is depicted in Figure 5.9.

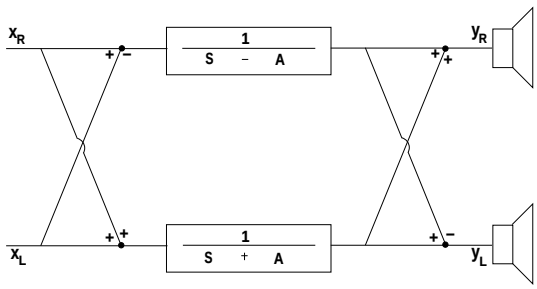

Figure 5.9: Shuffler configuration for case III.

5.4.4 Evaluation

Figures 5.10, 5.11 and 5.12 show the channel separation achieved by the simulation of case I, case II and case III respectively. The crosstalk cancellation averages obtained by using 5.16 are 38.7 dB, 31.9 dB and 39.4 dB. Figure 5.13 compares the channel separation for the three cases. It can be seen that the first case accomplish a good crosstalk cancellation but add unwanted peaks in the high frequency region between 7 kHz and 8 kHz. The second case shows a poor crosstalk cancellation performance between the frequency range of 1 kHz to 4 kHz but has a better performance for low frequencies compared with the other two cases. The third case accomplish the best crosstalk cancellation perfomance over the frequency range between 900 Hz and 3 kHz and the difference with the second

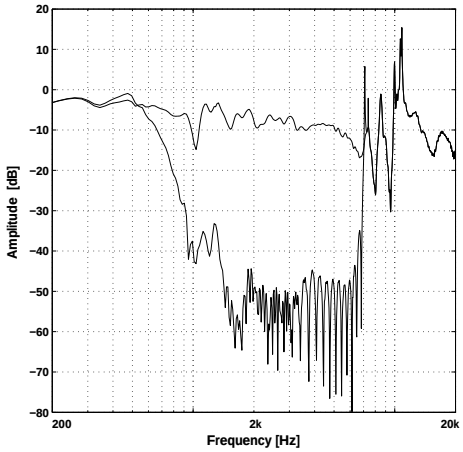

Figure 5.10: Left-right channel separation for shuffler filters calculated from pure HRTF and a separate equalization to compensate for the loudspeaker and reflections from the setup.

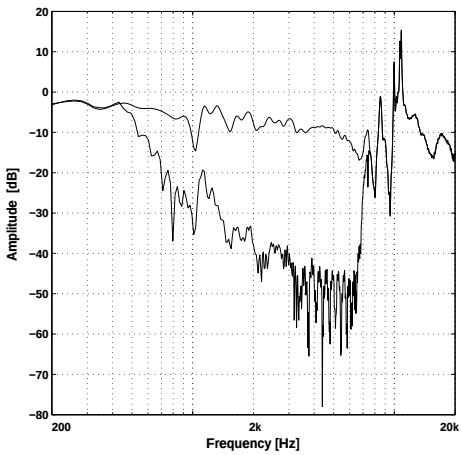

Figure 5.11: Left-right channel separation for shuffler filters calculated from a combination of HRTF and loudspeaker response and a separate equalization to compensate for the setup reflections.

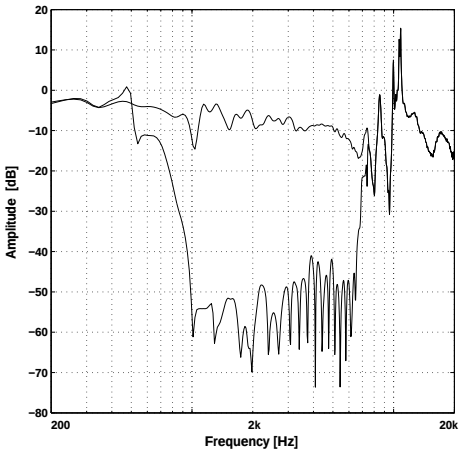

Figure 5.12: Left-right channel separation for shuffler filters calculated from ipsilateral and contralateral impulse response containing the HRTF, loudspeaker response and reflections from the setup. Large variations between -60 and -80 dB are due to the bandpass filtering.

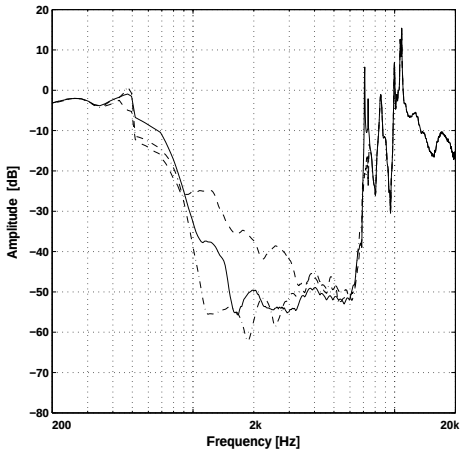

Figure 5.13: Comparison of channel separation for symmetric crosstalk canceller simulations in case I (solid line), case II (dashed-line) and case III (dash-dot line). For visual representation these channel separations were smoothed using a 16 points average of the frequency magnitude response.

case at low frequencies is relatively small. For the frequency range between 4 kHz and 6 kHz the channel separation of the three cases is similar and not noticeable differences are seen. Considering the results of this simulation and the fact that the third case uses just two filters for the implementation of the crosstalk canceller, it has been decided that the latter case should be used.

5.5 Final Implementation

After simulating various methods for the shuffler crosstalk cancellation implementation, it was decided to do the implementation using the contralateral and ipsilateral transferfunctions including all setup parameters (case III). This section deals with the implementation of this in MATLAB®, simulation of results and measured results.

5.5.1 Implementation in MATLAB®

As mentioned in chapter 4 on page 19 the band in which to do crosstalk cancellation is 500-6000 Hz. This means that this frequency area is only subject to the shufflers crosstalk cancellation system. However the lower and upper part of the frequency range should be added into the signal when the algorithm is finished. A diagram showing the program structure is depicted on Figure 5.14.

This implementation was initially performed, however since the upper frequency limit for the crosstalk cancellation system is 6 kHz, and the sampling frequency is 44.1 kHz,

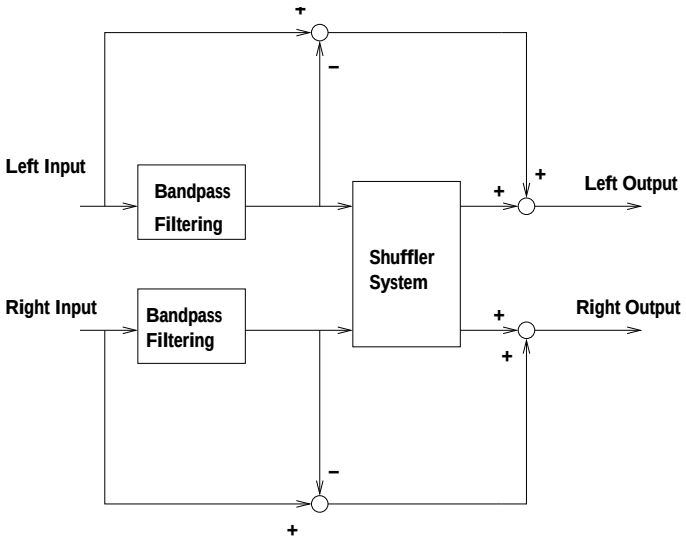

Figure 5.14: Implementation of the crosstalk cancellation system when adding high and low frequency band.

then using downsampling by a factor of 3 will give more power to the filters used in the crosstalk cancellation system. This was implemented and the diagram of this is depicted on Figure 5.15.

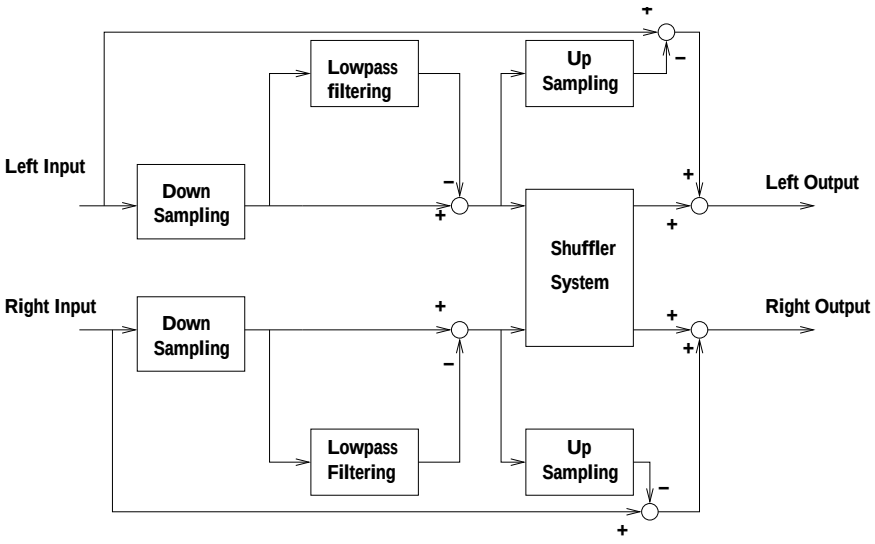

Figure 5.15: Implementation of the crosstalk cancellation system with the use of up/down-sampling.

The effect of the downsampling is that the frequency resolution becomes a factor of 3 better, which gives more power to the lower frequencies as well as saving computational power. The MATLAB® code for the crosstalk cancellation system was divided into two parts. One part that finds the filters (CCS1 and CCS2) and one that uses the filters and a wave file to create another wave file for use in the multimedia setup. The MATLAB® code can be found in the enclosure.

5.5.2 Simulation of Implementation

The simulation of the crosstalk cancellation system is done using an input file, where 10 seconds of random noise exists in the left channel and nothing in the right channel.

Before presenting the simulation results it should be noted, that when visually verifying the filters and their effect on the system, a dominant reflection around 1200 Hz, was discovered. This reflection could not be equalized, and a lot of power in the filter was used to try to cancel this out (without much effect). Therefore it was decided to remove that specific reflection (damping material), and the filter was made without this reflection.

The channel separation is found in the area from 500-6000 Hz to be 104.5 dB more than the natural ITF. A frequency plot of the channel seperation can be seen on figure 5.16. It should be noted that the figure is smoothed using windowing.

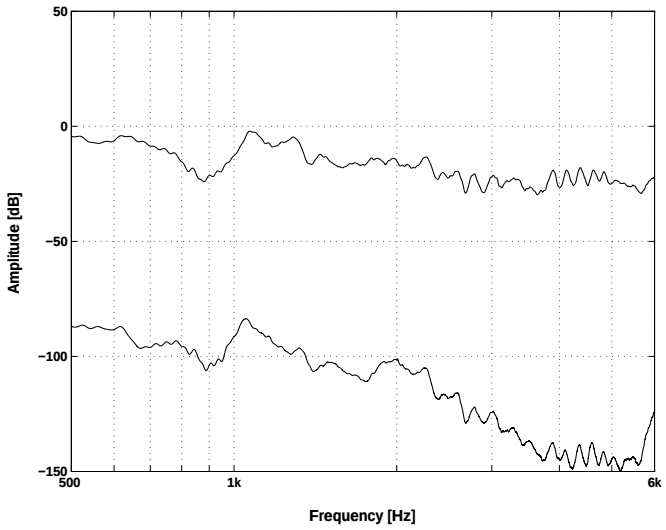

Figure 5.16: Channel separation (upper) vs. artificial channel separation (lower).

5.5.3 Physical Validation

The multimedia setup with the artificial head was used for the physical validation. The pysical validation is done using the same input file as used in the previous simulation. The input file is then used for playback in the multimedia setup. At the same time as playing the wave file, the sound comming from the two ears of the artificial head is recorded. Ideally there should be random noise in left ear and nothing in the right ear. The recorded signal from this experiment is the channel separation using the crosstalk cancellation system. Another measurement of the natural crosstalk cancellation is done. Here the unproccessed wave file used to generate the crosstalk cancellation file, is used. This will give the natural crosstalk cancellation measured physically. The two results are shown on figure 5.17, where the natural crosstalk cancellation is plotted with the crosstalk cancellation using the shufflers algorithm. This figure is also smoothed using windowing.

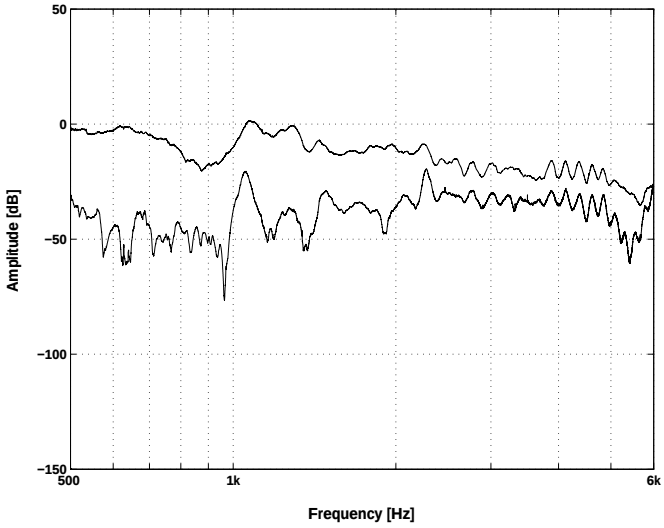

Figure 5.17: Natural crosstalk cancellation (upper) vs. shufflers crosstalk cancellation (lower) in the physical validation.

The channel seperation is found in the area from 500-6000 Hz to be 19.5 dB over the natural ITF.

From this it is possible to conclude that the system works both in simulation and in real life however, the crosstalk cancellation is less when measuring it physically.

CHAPTER 6

Listening Test

This chapter describes the design, conduction and results from a listening test, which was used to evaluate the subjective quality of the implemented system.

6.1 Aim of Test

The purpose of this test was to obtain knowledge about the subjective parameters of the system, and thereby evaluate if playback of binaural recordings, is possible via the chosen multimedia system.
Specifically, the results obtained with headphone playback are the measure of success, which is also stated in the project hypothesis. This hypothesis can be stated as a null hypothesis (H_0) and an alternate hypothesis (H_A):

H_0 : The crosstalk cancellation system has the same error rate in the localization of sound sources as headphone reproduction has.

H_A : The crosstalk cancellation system implemented has a different error rate than headphone reproduction.

6.2 Method

The method chosen for testing the quality of the constructed system is to test the localization of sound sources.

6.2.1 Stimuli

As mentioned above the method for testing the quality of the implemented system is sound comming from different directions. The directions selected are depicted on Figure 6.1

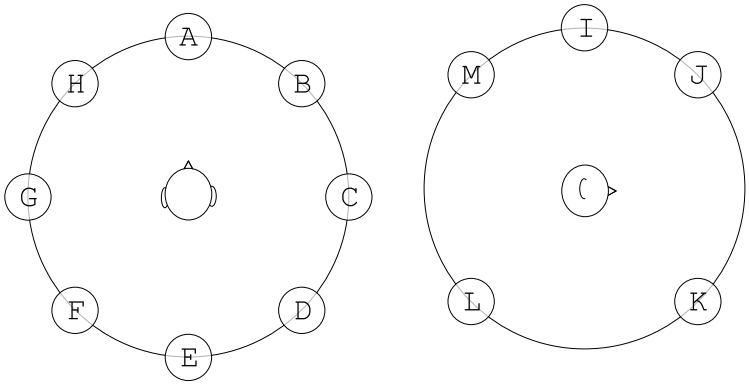

Figure 6.1: sound source locations for the listening test.

The horizontal plane (with no elevation) has sound sources 360° around the listener with 45° spacing intervals. In the median plane (with no azimuth angle) the sound sources are also located with 45° spacing, however the sound source located under the listener was omitted. This gives a total of 13 directions. This setup gives a good indication of the performance of the system both in the horizontal and the median plane. Similar setup spacing have been used in [Christensen, 2001] and [Møller et al., 1995], where more locations were used. All the sound sources are located at a distance of 1 meter, whereby the distance parameter of the perception is omitted.

The stimuli for the sound sources is selected to be pulsed white noise with a duration of 5 seconds. The signal was highpass filtered at 600 Hz, and the recording is done in a measuring room at the Department of Acoustics at AAU. Reason for stimuli and recording environment are explained in appendix C on page 71, Binaural Test Signals. The recorded stimuli was played at a level of 77 dBA (recorded with B&K 2238 mediator, lnbr.33948), and the headphone playback level was subjectively set to the same level using a headphone amplifier (Behringer HA903, lnbr.33239).

6.2.2 Subjects

8 subjects (7 male and 1 female) in the age of 24-28 years participated in the listening experiment. The test subjects had no training in listening experiments and the selection of these subject was not done to fit a certain profile. Furthermore, it was required that the test participants had normal hearing capabilities, but due to time limitations no audiometry was performed, which meant that the answers obtained during the training session were used as a quality measure of the listener. If a test subject had more that 70% errors in the training session, the test subject was disqualified for further testing. This happened with one of the subjects, and another subject was used instead.

6.2.3 Design of Test

The procedure of performing the test is described by the written instructions and the sheet for answers, which was handed out to the test subjects. These instructions and answering sheets can be found in the enclosure. The outline and content of these sheets are based on the suggestions, which were pointed out by Morten Lydolf during the course "Human Sound Perception and Psychometry". As seen on the sheet with instructions, 39 test tones (3 times run-through of each source) were done per test subject during a test. The 39 answers were treated as independent answers and all of the 39 sounds were randomized. This means that the answers from each test subject was treated as 3 independent evaluations. A familiarization run-through was done in the beginning of the test, to familiarize the test subjects with the method of answering. Two pictures showing the listening test scenario is depicted on Figure 6.2 and Figure 6.3.

Figure 6.2: *Picture from the listening test setup, left side.*

Figure 6.3: *Picture from the listening test setup, right side.*

The test was performed as a blind test with forced answers. One of the group member were in charge of instructions and questions from the test subjects. Another group member was in charge of playing the correct files to the subject. The test was fully balanced, which means that the test subjects were exposed to different randomized sound locations and a different system order (headphone system or multimedia system first). The test duration was approximately 40 minutes, with a 5 minute break between the two systems.

6.3 Results

The results are presented in a 13x13 matrix. The sound source location is marked on the abscissa and the response to the source location is marked on the ordinate. Figure 6.4 shows the results from the headphone reproduction and Figure 6.5 shows the results from the crosstalk cancellation system.

As it can be seen from Figure 6.4 and Figure 6.5 that the response from the crosstalk cancellation system is quite similar to the response from headphones. The errors in localization are mainly located in the median plane. The horizontal plane seems to work as expected, which indicate that the horizontal cues (ITD and ILD) are correct.

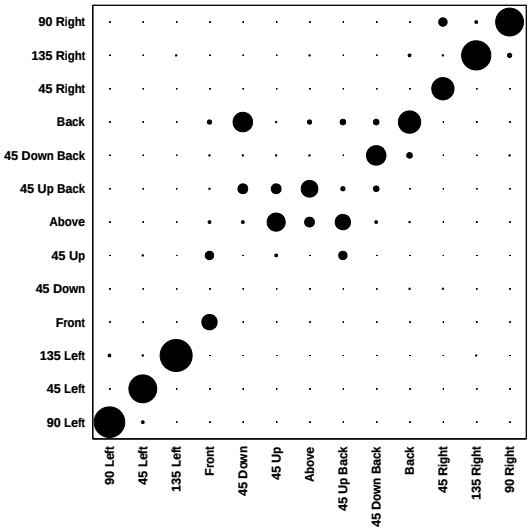

Figure 6.4: Result of the listening test, where the stimulus was presented in headphones. The stimulus directions are marked on the abscissa and the response is marked on the ordinate.

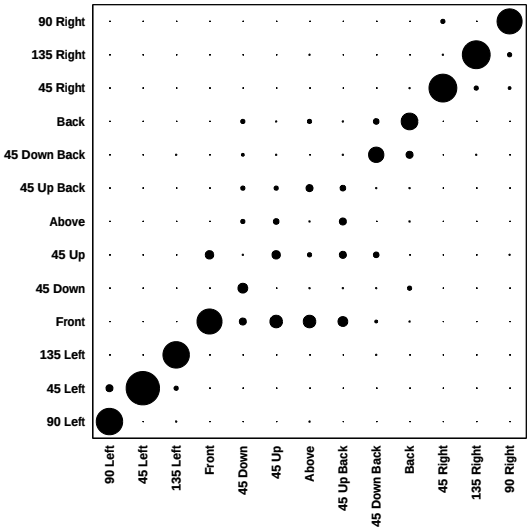

Figure 6.5: Result of the listening test, where the stimulus was presented on the multimedia setup. The stimulus directions are marked on the abscissa and the response is marked on the ordinate.

6.3.1 Statistical analysis

As mentioned in section 2.2.4 on page 9, Typical Localization Errors, the errors from a listening test can be sorted into groups. In order to see which errors are the most prominent, it is decided to divide the errors into 3 categories:

- Median-plane errors
- Cone of confusion errors
- Other errors

The median plane errors can appear, when a sound source is located in the median plane, and the answer is also located in the median plane. The cone of confusion can appear as sound comming from the left or right front is confused with sound comming from the left or right back. This is depicted on Figure 6.6, where the different errors are marked.

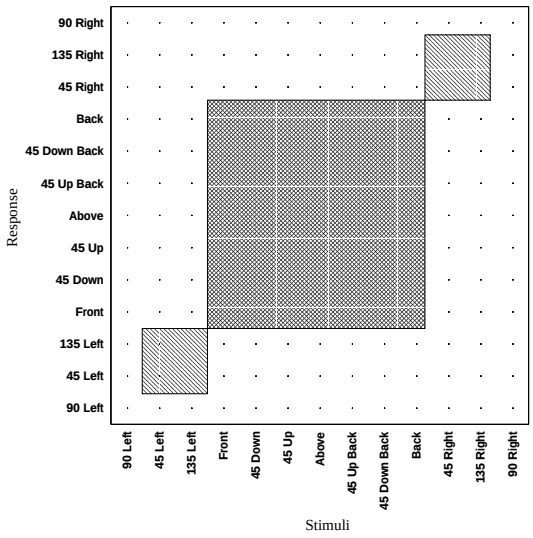

Figure 6.6: The errors are divided into three categories: median plane errors (darkest), cone of confusion errors (grey) and other errors (white). The diagonal is not included in the errors since it is the correct responses.

The error percentage and exact number of errors in the listening test are shown in table 6.1. Figure 6.7 shows the error percentage of the two systems tested.

Stimuli	Median plane errors	Cone of confusion errors	Other errors
Headphones	**66.7 %**	**2.1 %**	**7.4 %**
	112	2	23
Multimedia system	**61.9 %**	**7.3 %**	**7.0 %**
	104	7	22
Potential errors	168	96	312

Table 6.1: This table shows the errors in percentage (bold) and numbers. In the bottom of the table is the amount of possible errors in each category.

The error percentage the ratio between the number of errors within one category and the possible number of errors within that same category.

In order to evaluate the errors statistically a single sided Fisher-Irwin test was done. This test was selected since it uses Bernoulli variables. Bernoulli random variable are suitable to use when an experiment outcome can be classified in two different ways: as a "success" or a "failure". More literature regarding Fisher-Irwin can be found in [Ross, 2000]. The test was based on a significance level of 5 %, and the null hypothesis is used for each of the three categories: median plane errors, cone of confusion errors and other errors. If the Fisher-Irwin test shows that one of these categories can be rejected, the overall hypothesis is also rejected. Table 6.2 shows the results from the Fisher-Irwin test.

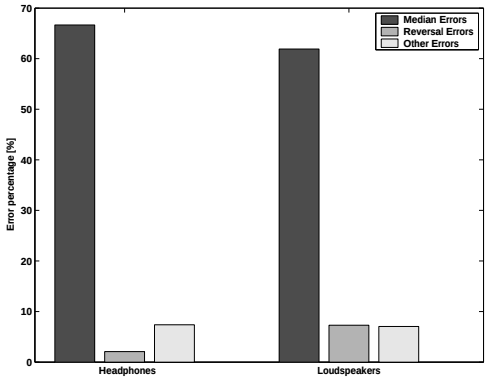

Figure 6.7: Error percentage of the data from the listening test. These are sorted into three categories: median-plane errors, cone of confusion errors and other errors.

Comparison	Median plane errors	Cone of confusion errors	Other errors
Headphones vs. Loudspeakers	Accepted	Accepted	Accepted

Table 6.2: Result of the Fisher-Irwin test.

As it can be seen from table 6.2, the statistic test shows, that the null hypothesis could not be rejected in any of the three categories. This means that there is no significant difference between the perception of binaural sound using a multimedia setup over headphone reproduction under the circumstances of which the test is performed.

6.3.2 Summary

In this chapter a listening test, evaluating the subjective performance of the crosstalk cancellation system, was performed. The test has been designed, carried out, and statistical analysis have shown that the null hypothesis could not be rejected.

The results also showed, that localization in the median plane is more difficult than localization in the horizontal plane.

6.3.3 Discussion

The errors are probably due to the recording of the binaural sound (appendix C on page 71). As mentioned the loudspeaker used for the recordings only have linear response up to 9 kHz and studies by [Christensen, 2001] have shown that the localization of sound sources deteriorate, when decreasing the upper frequency limit of a signal. It has also been shown, that the vertical localization depends primarily on the higher frequency content from 4-16 kHz, see section 2.2.2 on page 7, Spectral Cues.

CHAPTER 7

Conclusion

The preceding chapters in this report have described the design, implementation and test of a crosstalk cancellation system for using in a multimedia setup. The aim of this project was to investigate whether or not it was possible to make a multimedia system, that could reproduce 3D sound. The obtained results were compared to the results obtained using headphones.

Based on investigations regarding human sound perception and different crosstalk cancellation methods, the implemented system, was done using a shuffler crosstalk cancellation algorithm. The system was bandlimited from 500-6000 Hz and the implemented filter order was 2048. Different methods of implementing shufflers algorithm were simulated, and a final version using downsampling was implemented. The implemented system was simulated and a channel separation of 104.5 dB more than the natural channel separation by the shadowing of the head was obtained. However, physical measurements showed that the system in practice had 19.5 dB more than the natural channel separation.

A listening test was performed to investigate whether the hypothesis set in this project could be rejected or not. The hypothesis was stated as a null hypothesis and an alternate hypothesis:

H_0 : The crosstalk cancellation system implemented has the same error rate in the localization of sound sources as headphone reproduction has.

H_A : The crosstalk cancellation system implemented has a different error rate than headphone reproduction.

The listening test was performed as a forced choice blind test, where the participants, were asked to select a direction of a sound source. The sound source was played in headphones and on the implemented system. The errors from results of the listening test were

classified in three different categories, and on each of these categories a statistical analysis was made to observe if the null hypothesis could be rejected. If one of the three hypothesis was rejected the overall null hypothesis was rejected. A single sided Fisher-Irwin test was done with a level of significance of 5 %. The result of the Fisher-Irwin test showed that the null hypothesis could not be rejected in any of the three cases, meaning that the overall hypothesis could not be rejected.

From this it can be concluded that it is not possible, within a confidence level of 95 %, to determine that the localization using headphones is better than the localization using the implemented system.

7.1 Further Investigations

In the design of the implemented system, boundaries had to be set in order to avoid extensive investigations. In this section ideas for further investigations is presented.

A main part of the crosstalk cancellation system correspond to the loudspeakers that are used. Because of the fact that the loudspeakers only have one driver they have a limited frequency range. Further investigations showing the effect of the used loudspeakers would be interesting. This also leads to further investigations regarding the upper and lower frequency limit. In this project theoretical investigations have been made to determine the upper and lower frequency limit of a crosstalk cancellation system. However, it would be more feasible to document this using physical measurements. This would show where the channel separation of crosstalk cancellation system is deteriorated. At lower frequencies the cancellation at one ear would also have substantial effect on the other ear, which is not feasible. At higher frequencies the individual HRTFs will become dominant, however this project has not investigated whether or not that situation has substantial influence on the system performance. Also the angle of the loudspeaker could be investigated. For instance investigating the effect of placing the loudspeakers further back on the screen. This would give a channel separation caused by the shadowing of the monitor in the multimedia system.

7.2 Improvements

The improvements of this system has mainly two aspects: the stimuli used in the listening test and the listening test.

Test Signal The stimuli recorded for the listening test were done using a single loudspeaker unit (Vifa MD10 placed in a hard plastic ball). This loudspeaker is only linear from 600-9000 Hz. In order to achieve better results in the vertical plane in headphones, the upper frequency limit should be higher. The project group feels that this would lead to an improvement in the localization in the median plane in headphones. This improvement would probably not be influential on the performance of the multimedia system,

because of the poor higher frequency response of the satellite loudspeakers.

Listening Test The listening test performed uses the answering sheet located in the enclosure. A further elaboration in the listening test could be to use the multimedia system tested. This means that by implementing a graphical user interface would be possible to achieve a more interactive test, where the test subjects could select the duration of the sound stimuli and the duration of the pause. This would give the subject the necessary time to listen the stimuli, and the time to note the position. Regarding the position of the sound source, a 3D plot of the listening position could be used, where the possible locations are plotted. This idea for a new test is made because the results from the listening test performed in this project, showed some errors that can only be led back to the layout of the answering sheet. For instance when answering a sound source as comming from the back low (position L) when the sound source is located in the left back (position F). For source positions see enclosure.

Bibliography

[Blauert, 1983] Blauert, J. (1983). *Spatial Hearing : The Psychophysics of Human Sound Localization.* MIT Press. 0-262-02413-6.

[Christensen, 2001] Christensen, F. (2001). *Binaural technique with special emphasis on recording and playback.* PhD thesis, Department of Acoustics Institute of Electronics Aalborg University.

[Cooper and Bauck, 1989] Cooper, D. H. and Bauck, J. L. (1989). Prospects for transaural recordings. *Audio Engineering Society*, 37(1/2):3–19.

[Gardner, 1998] Gardner, W. G. (1998). *3D Audio Using Loudspeakers.* Kluwer Academic Publishers. ISBN: 0-7923-8156-4.

[Hammershøi, 1995] Hammershøi, D. (1995). *Binaural Technique - a method of true 3D sound reproduction.* Institute of electronic systems, Aalborg University, Acoustics Laboratory. ISBN: 87-7307-516-7.

[Hougaard et al., 1995] Hougaard, S., Jensen, O., Kristensen, M., Ludvigsen, C., Petersen, J., and Weis, P. (1995). *Sound & Hearing.* Widex. ISBN: 87-985233-1-7.

[Kotorynski, 1990] Kotorynski, K. (1990). Digital binaural/stereo conversion and crosstalk cancelling. *Proc. Audio Eng. Soc. Conv*, Preprint 2949.

[Møller, 1992] Møller, H. (1992). Fundamentals of binaural technology. *Applied Acoustics*, 36:171–218.

[Møller et al., 1995] Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (1995). Binaural technique: Do we need individual recordings? *Audio Engineering Society*, 44(6):451–469.

[Nielson and Koch-Henriksen, 2000] Nielson, J. and Koch-Henriksen, U. (2000). Role of loudspeaker angle in a crosstalk cancellation system. Master's thesis, Department of Acoustics, Aalborg University.

[Ross, 2000] Ross, S. M. (2000). *Introduction to Probability and Statistics for Engineers and Scientists*. Academic Press. ISBN: 0-12-598472-3.

[Sandvad and Hammershøi, 1994] Sandvad, J. and Hammershøi, D. (1994). Binaural auralization. comparison of FIR and IIR filter representation of HIRs. *Proc. Audio Eng. Soc. Conv. (Amsterdam)*, Preprint 3862.

[Sæbo, 2001] Sæbo, A. (2001). *Influence of Reflections on Crosstalk Cancelled Playback of Binaural Sound*. PhD thesis, Department of Telecommunications Norweing University of Science Technology. ISBN:82-471-5380-7.

[Schroeder, 1975] Schroeder, M. R. (1975). Models of hearing. *Proc. IEEE*, 63:1332–1350.

[Ward and Elko, 1999] Ward, D. B. and Elko, G. W. (1999). Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation. *IEEE Signal Processing Letters*, 6(5):106–108.

APPENDICES

APPENDIX A

Measurement Report

This appendix contains descriptions of the measurements done in the laboratory. These measurements were made in order to obtain necessary knowledge about the system setup which is used in this project. The appendix is separated into 4 sections: "Purpose of Measurements", "Measurement Setup", "Results" and "Summary". The first section describes the purpose of each of the two overall measurements, then it is described how the measurement is done and in "Results" the measurement results of each measurement is presented. Finally, a summary about the obtained results is given and the results are evaluated and discussed.

A.1 Purpose of the Measurements

The purpose of these measurements is to obtain knowledge about the parameters influencing the total system transfer function. Knowing these parameters, it is possible to separate the system transfer function into several parts, representing individual elements of the system model. The transfer functions which are desired correspond to the transfer function of the setup(desk, monitor and keyboard) called H_{SETUP}, the transfer function of the artificial head separated into an ipsilateral transfer function TF_{IPSI} and a contralateral transfer function TF_{CONT} and the transfer function of the loudspeakers, H_{LSPK} (nonlinearity in loudspeakers).

A.1.1 Loudspeaker Measurement

The aim of this measurement was to obtain the loudspeaker transfer function, H_{LSPK}. But beforehand it should be determined whether the loudspeaker system at hand could

be used in the crosstalk cancellation system. Evaluation of the loudspeaker system was based on the following requirements:

- Frequency response within ± 5 dB SPL rel. 20 uPa in the region for crosstalk cancellation.
- Crossover frequency between the subwoofer and the satellites must be much lower than the lower crosstalk cancellation frequency.
- Similar response from left and right loudspeaker.

The "flat frequency response" parameter is important, mainly because it reveals characteristics about the impulse response of the loudspeaker. A flat frequency response is not a necessity, because it will be equalized anyway, but large dips and peaks are undesireable, because they are difficult to equalize. The parameter of a "sufficiently low crossover frequency" is important, because the playback through the subwoofer is mono, and hence no binaural playback is possible in the frequency range of the subwoofer. The criterion about the similarity in response from left and right loudspeaker must be met if symmetry about the setup is going to be assumed.

A.1.2 Anechoic Measurements

These measurements were made with and without the artificial head in order to obtain H_{SETUP}, the ipsilateral and contralateral transfer functions of the artificial head (S and A). These transfer functions could be extracted, using various inverse filtering schemes.

A.2 Measurement Setup

The impulse responses (and frequency responses), were found by the use of a MLSSA system. MLSSA is an acronym for Maximum Length Sequence System Analyser. A MLS (Maximum Length Sequence) signal is used for comparison of a known signal with a processed signal (for example the coloration caused by a loudspeaker). By calculating the correlation between the two signals, the impulse response of the loudspeaker can be obtained. MLSSA parameters that were used during all measurements are described in Table A.1. The measurement equipment, which was used during all the measurements in the laboratory is listed in Table A.2.

A.2.1 Loudspeaker Measurements

Satellites Measuring the reponse of the satellites was done by connecting the MLSSA system to the SW320 power amplifier, which fed the satellites and subwoofer with a filtered and level adjusted signal. This signal was picked up by the microphone, B&K 4134 which was driven by the measuring amplifier, B&K 2636, and from there fed to the MLSSA system. The SW320 power amplifier is intended to be fed with a "line in" signal. Level

ACQUISITION	UNITS
mode: cross- correlation	acquisition: 1V/V
length: 16383	stimulus: 1V/V
sample rate: 60,61 kHz	
auto-range: Enabled	
ANTI-ALIASING FILTER	DIGITAL FILTER
width: 1 octave	center: 1000 Hz
type: Chebyshev	width: 1 octave
bandwith: 20 kHz	
gain: 10	
TRIGGER	STIMULUS
type: Stimulus trigger	burst: MLS
delay: 0 samples	amplitude: 1.5 V
	period: 16383

Table A.1: *The MLSSA parameters.*

Measurement system:	MLSSA
Microphone:	B&K 4134 and B&K 4133
Measuring amplifier:	B&K 2636
Phantom power source:	Neumann 48V
Artificial torso:	Valdemar w. filling
Artificial Head Mic.	GRAS 40 AD (left and right)
Sound level calibrator:	B&K 4230
Loudspeakers:	Creative SW320 satellite loudspeaker and subwoofer

Table A.2: *The equipment used during the measurements.*

adjustment of the signal output from the SW320 is controlled by a potentiometer which is connected in series with the "line in" signal. The level of the subwoofer is individually adjustable. These (level) parameters are required to be stationary throughout the signal processing done, which was ensured by covering the potentiometers with tape. The setup which was used for this measurement is illustrated in Figure A.1.

Subwoofer This measurement was done without using the satellites in order to obtain the crossover frequency between the subwoofer and the satellites. The measurement setup was similar to A.1 except that the satellites were disconnected.

A.2.2 Anechoic Measurement

The complete setup illustrated in chapter 1 on page 1 section 1.3 was used during this measurement. In order to obtain the overall combined transfer function of H_{SETUP}, H_{LSPK} and the either TF_{IPSI} or TF_{CONT}, the artificial head was used. This setup is illustrated in Figure A.2.

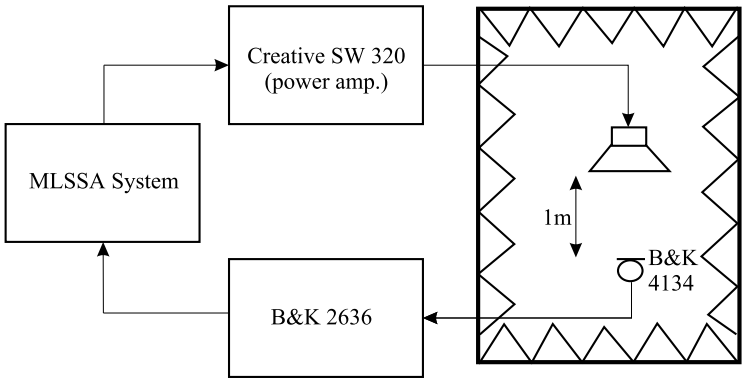

Figure A.1: Setup for measuring the loudspeaker system response.

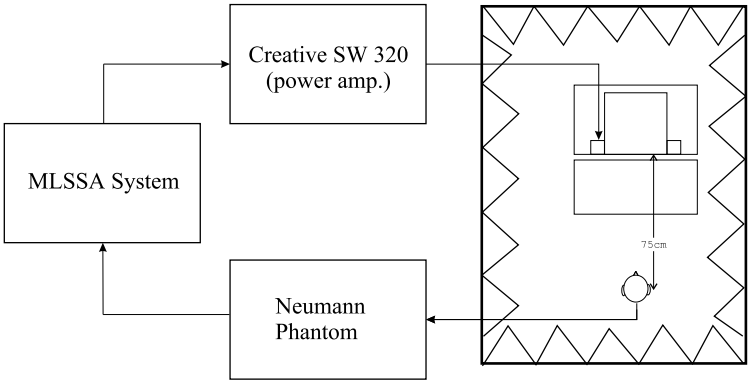

Figure A.2: Setup for measuring the combined response of H_{LSPK}, H_{SETUP} and TF_{IPSI} or TF_{CONT}.

In this setup, the contralateral and ipsilateral transfer functions for both ears were measured. The subwoofer, combined with one loudspeaker at a time was used during the measurement. Another measurement was done using the artificial head and a loudspeaker located in similar positions as during the setup. This was done in order to separate the transfer funtions of the artificial head from H_{SETUP} and H_{LSPK}. This measurement setup is illustrated in Figure A.3

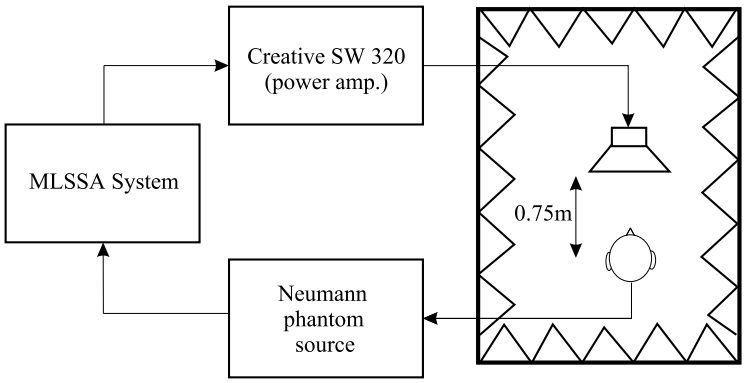

Figure A.3: Setup for measuring the combined response of H_{LSPK} and TF_{IPSI} or TF_{CONT}.

A.3 Measurement Results

The measurements were saved as ASCII files and imported to MATLAB®, thereby providing the data of the measured impulse responses. These impulse response data were then used to obtain the frequency responses (the transfer functions).

A.3.1 Loudspeaker Measurement

In the figures below, the frequency response of left and right loudspeaker on axis and 26 degrees off axis is plotted. For each of these measurements, the subwoofer response is included at a fixed level.

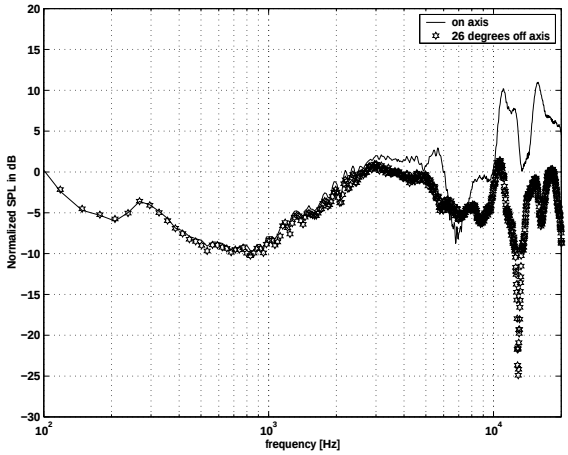

Figure A.4: Frequency response of the right satellite loudspeaker including subwoofer.

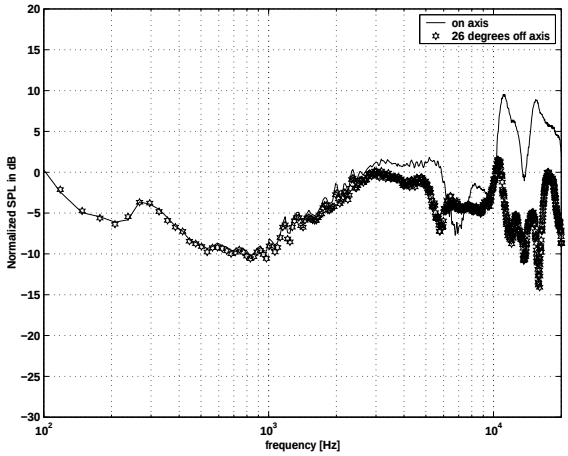

Figure A.5: Frequency response of the left satellite loudspeaker including subwoofer.

The measured impulse response of the right loudspeaker including the subwoofer is shown in Figure A.6 on the following page. In this figure, notice the time delay of approximately 3 ms, which correspond to the distance, that includes sound to travel 1m in air at 25°

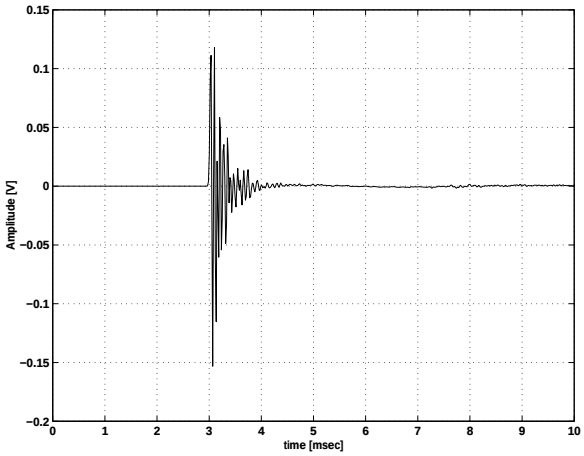

Figure A.6: On axis impulse response of the right satellite loudspeaker including subwoofer.

Celsius. Also, it can be seen that the impulse response is rather long compared to the ideal Dirac's delta impulse, which means that the system is poorly damped. In order to compare right and left channel with each other, a relationship between the left and right loudspeaker is found:

$$H_{LSPK_R} = H_{LSPK_L} \cdot \frac{H_{LSPK_R}}{H_{LSPK_L}} \tag{A.1}$$

The transfer function $\frac{H_{LSPK_R}}{H_{LSPK_L}}$ is plotted in Figure A.7

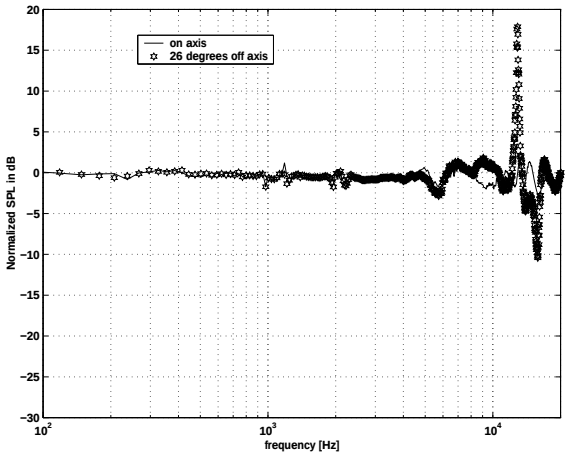

Figure A.7: $\frac{H_{LSPK_R}}{H_{LSPK_L}}$.

Figure A.7 indicates that the right and left loudspeaker are similar in the low and middle frequencies and differ a little in the response at the high frequency region. A final measurement was done in order to determine the crossover frequency between the satellites and the subwoofer. By observing the frequency response of the subwoofer alone and comparing this to the combined frequency response, the crossover frequency was estimated to be 150 Hz. For this measurement, the sampling frequency was lowered to 15.15 kHz in order to get a better low frequency resolution.

A.3.2 Anechoic Measurement

In the anechoic chamber, measurements on the complete setup with and without the artificial head were done. In Figure A.8 and A.9 the ipsilateral and contralateral transfer functions combined with H_{LSPK} and H_{SETUP} are depicted.

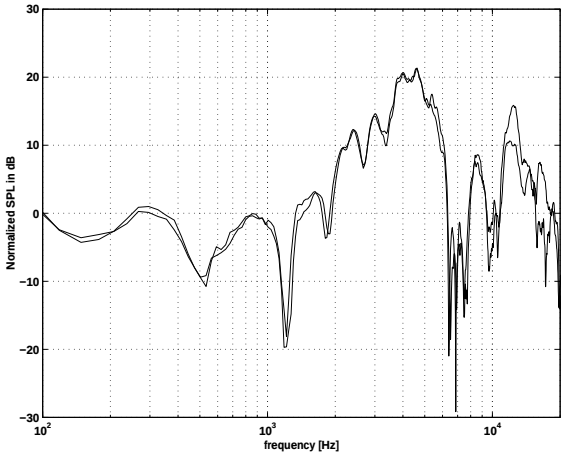

Figure A.8: Ipsilateral transfer function, TF_{IPSI} including H_{LSPK} and H_{SETUP}. These transfer functions correspond to the left and right ear measurements.

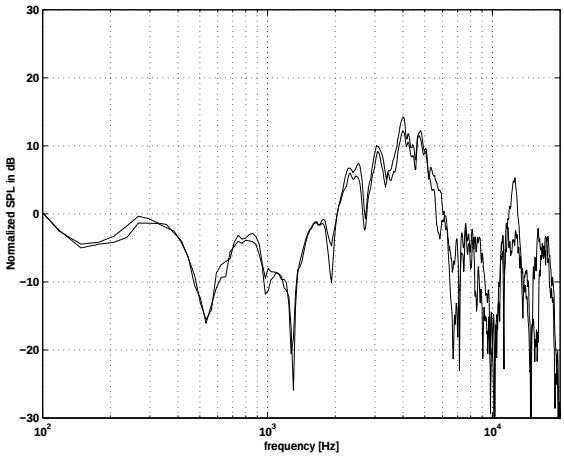

Figure A.9: Contralateral transfer function, TF_{CONT} including H_{LSPK} and H_{SETUP}. These transfer functions correspond to the left and right ear measurements.

Another measurement was done in order to find the combined transfer function of S or A and H_{LSPK}. This measurement was performed using the artificial head and a loudspeaker initially positioned 150° off axis and then 260° off axis.

The data obtained from this measurement is plotted in Figure A.10 and A.11.

With these measurements it is possible to separate the transfer function of the setup, H_{SETUP} by using the following equation:

$$H_{SETUP} = \frac{TF_{IPSI} \cdot H_{LSPK} \cdot H_{SETUP}}{H_S \cdot H_{LSPK}} \tag{A.2}$$

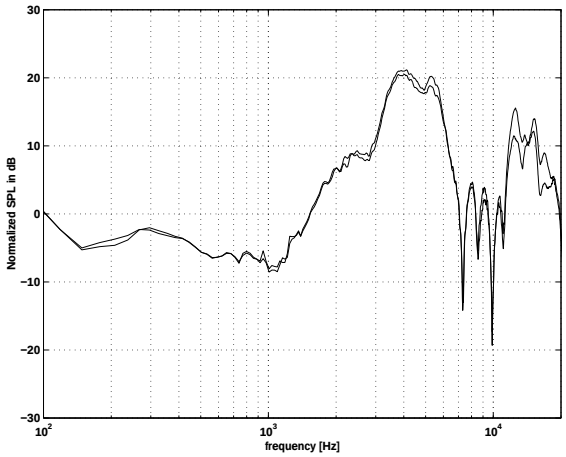

Figure A.10: Ipsilateral transfer function, TF_{IPSI} including H_{LSPK}.

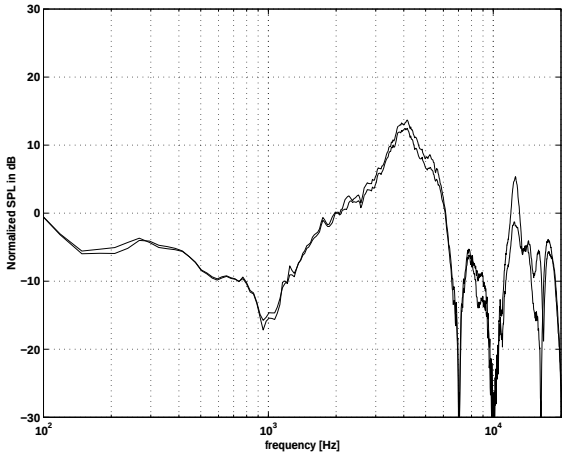

Figure A.11: Contralateral transfer function, TF_{CONT} including H_{LSPK}.

or, equivalently:

$$H_{SETUP} = \frac{TF_{CONT} \cdot H_{LSPK} \cdot H_{SETUP}}{H_A \cdot H_{LSPK}} \tag{A.3}$$

This means that H_{SETUP} can be obtained by a deconvolution in the time domain or a division in the frequency domain. The result is illustrated in Figure A.12 and Figure A.13. These data are obtained using the measurements from the right loudspeaker.

The HRTF are obtained by dividing the ipsilateral and contralateral transfer functions with the transfer function of the loudspeaker 20° off axis, corresponding to a microphone position in the middle to the artificial head. This HRTF is illustrated in Figure A.14 and A.15 for the ipsilateral and contralateral representation respectively.

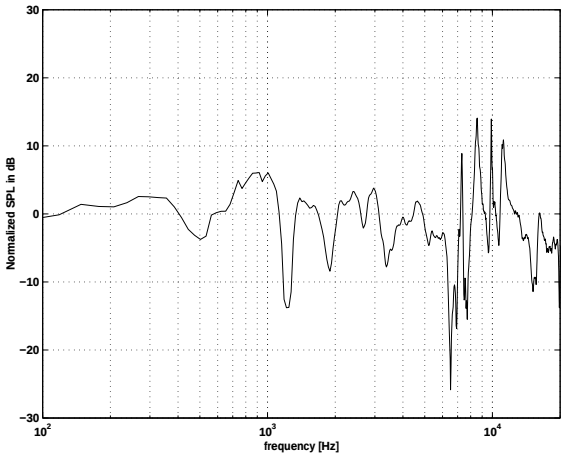

Figure A.12: Calculated H_{SETUP} based on the ipsilateral transfer function measurements.

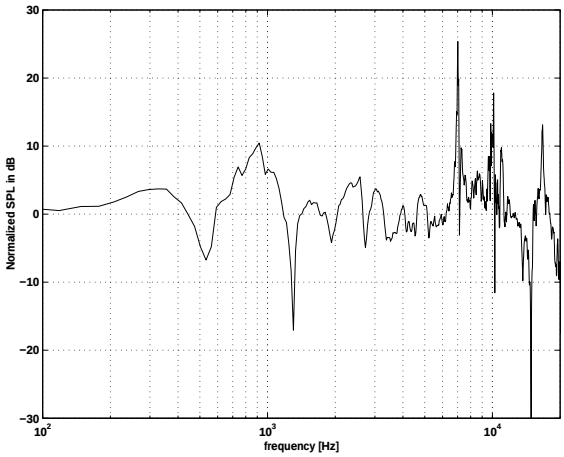

Figure A.13: Calculated H_{SETUP} based on the contralateral transfer function measurements.

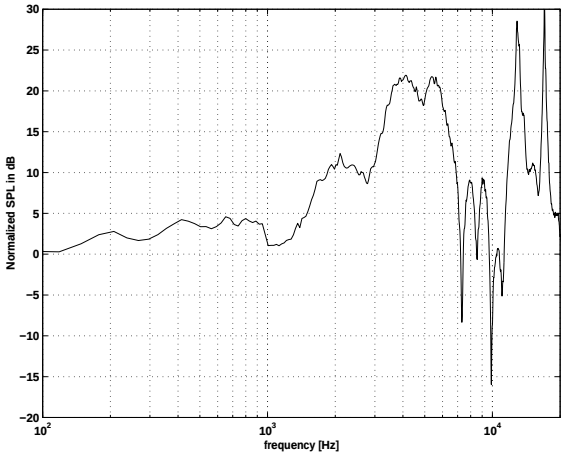

Figure A.14: Calculated HRTF for the ipsilateral transfer function.

A.4 Conclusion on Measurements

In this section, the results of these measurements are considered and discussed and a concluding remark is added.

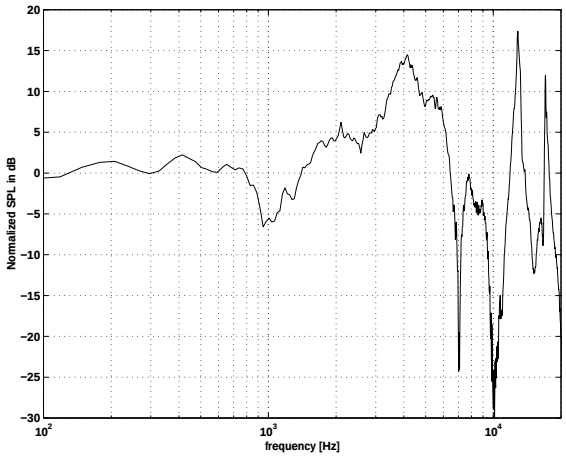

Figure A.15: Calculated HRTF for the contralateral transfer function.

A.4.1 Loudspeaker System

The satellite loudspeakers has a reasonably flat frequency response (to within ± 3 dB SPL rel. 20 μPa) in the frequency range from 500 Hz to 6 kHz. The point of crossover frequency between the satellite and the subwoofer was estimated to be around 150 Hz, which is sufficiently low. The right and left channel have similar responses but seems to differ in the high frequency area. However, this is not a significant disadvantage for being used in a crosstalk cancellation system, since they have similar responses in the frequency band of interest for crosstalk cancellation. On the basis of these measurement results, the performance of this specific loudspeaker system is evaluated as being acceptable, and it is therefore chosen for implementing in the crosstalk cancellation system.

A.4.2 Anechoic Measurements

The measurements on the setup in an anechoic environment proved to be succesfull because both the HRTF was found. Furthermore, the transfer function of the setup, H_{SETUP} was found. As it can be seen on the plot of H_{SETUP} this transfer function consist mainly of a combfilters, which are caused by the desktop, and the diffraction pattern caused by the edges of the monitor. By comparing the impulse response measured for the ipsilateral setup for both right and left ear a small deviation in symmetry was found. This error in the symmetry of the setup caused correspoond to an additional delay of the signal from the left loudspeaker of approximately 0.018 ms compared to the right loudspeaker, which corresponds to approximately 1 sample (when the sampling frequency is 60610 Hz). This means that we have made an error in our setup, causing the left loudspeaker to be positioned 0.62 cm further away from the left ear than the right loudspeakers distance to the right ear. The sampling rate allows a maximum accuracy given by:

$$\text{Sound travelling distance/sample} = \frac{1}{f_s} \cdot c$$

Where f_s is the sampling frequency and c is the speed of sound in the surrounding environment. In this case it means that the sound travels around 0.56 cm per sample. Even

though the setup was not perfectly symmetrically alligned during these measurements, the measurement error due to this deviation was estimated to be acceptable.

A.4.3 Final remark

In this measurement report, it has been documented how individual transfer functions of the system have been found. The following transfer functions have succesfully been found: TF_{IPSI}, TF_{CONT}, H_{SETUP},and H_{LSPK}. H_{LSPK} was found for the following conditions: on axis, 15°, 20° and 26° off axis. The method used for extracting the not measured transfer functions utilized in the system is explained in chapter 5.

APPENDIX B

Head Related Transfer Function Calculation

The aim of this appendix is to show a mathematical description of the procedure concerning the extraction of Head Related Transfer Functions (HRTFs) from raw measurements. Commonly, this procedure involves gain correction, division in the frequency domain, low pass filtering and DC correction. Following a standard terminology, time domain representations are described by lower letters and frequency domain representations are described by capital letters.

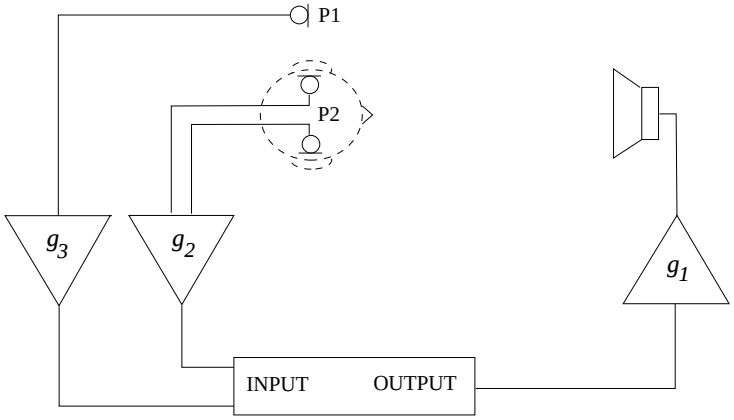

Figure B.1: Common setup used to measure Head Related Transfer Function. When measuring, P1 should be placed in the middle of the head, with the head absent.

B.1 HRTF Measurements

Head Related Transfer Functions describe the response for specific sound source localization relative to the head. The acoustical information obtained in a HRTF involves the influence of the head and torso mainly. However, when measuring the impulse response of a HRTF the obtained impulse response combine this influence with the response of the measuring system which includes amplifiers, loudspeaker, anti-alias filters, microphones and preamplifiers commonly. If we referred $P2$ as the measured response at the ears of the listener and $P1$ as the measured response at the centre of the head with the head absent, then by dividing $P2$ with $P1$ it is possible to extract the HRTF.

B.2 Gain Correction

Figure B.1 shows an electroacoustic setup for HRTF measurements. Since the gain of the different amplifiers could differ and also the sensitivity of the different microphones, a gain correction to compensate for gain differences is needed to be carried out. If $p2_{IPSI}$, $p2_{CONT}$ and $p1$ denote the ipsilateral, contralateral and free field time representations of $P2$ and $P1$, and g_1, g_2 and g_3 correspond to the loudspeaker amplifier gain and the different microphone amplifiers given in dB, then the gain correction is accomplished by the following equations:

$$p2_{IPSI_G} = p2_{IPSI} \cdot 10^{\frac{g_1+g_2}{20}} \cdot s_{IPSI} \tag{B.1}$$

$$p2_{CONT_G} = p2_{CONT} \cdot 10^{\frac{g_1+g_2}{20}} \cdot s_{CONT} \tag{B.2}$$

$$p1 = p1 \cdot 10^{\frac{g_1+g_3}{20}} \cdot s_{Ref} \tag{B.3}$$

where s_{IPSI}, s_{CONT} and s_{FF} are the sensitivity of the microphones used to measure $p2_{IPSI}$, $p2_{CONT}$ and $p1$ respectively, and $p2_{IPSI_G}$, $p2_{CONT_G}$ are the gain corrected.

B.3 Frequency Domain Processing

After the ipsilateral, contralateral and free field impulse responses have been gain corrected, they are transformed to the frequency domain by using the Discrete Fourier Transform (DFT). The length of the frequency representation is chosen in order to avoid aliasing when returning to the time domain. Moreover, since for some angles $P2_{IPSI}$ arrives to the microphone before the free field response $P1$ then it is necessary to delay $P2_{IPSI}$ ensuring a causal response when division is taking place in the frequency domain. By dividing $P2$ with $P1$ all the information common to both responses is removed. Therefore only the frequency response due to the presence of the listener is retained. After division the inverse Fourier Transform is applied to the ipsilateral and contralateral frequency responses. Then the resulting responses are low pass filtered since any anti-aliasing filter which could be used by the measuring setup has been removed. Finally, the appropriate new ipsilateral and contralateral time representations are obtained by applying a square window.

B.4 DC Correction

Due to a general poor performance of the measuring setup to accurately measure very low frequencies, a DC-value correction is necessary. The theoretical DC-value is 1 since at 0 Hz the sound pressure is static, which means that the sound pressure is equal at both ears. Considering the before explained, the DC-value is set to 0 dB by adding either a constant δ in the time domain representation of the HRTF or a sinc function in the frequency domain representation. The approach of this work has been a DC correction in the time domain where the constant δ is given by

$$\delta = \frac{-1}{N} \sum_{i=0}^{N-1} (h_i) \tag{B.4}$$

where h_i are the coefficients of the signal to be corrected. Thus the DC corrected HRTFs are obtained as

$$p2_{IPSI_{DC}} = p2_{IPSI} + \delta_{IPSI} \tag{B.5}$$

$$p2_{CONT_{DC}} = p2_{CONT} + \delta_{CONT} \tag{B.6}$$

APPENDIX C

Binaural Test Signals

This appendix deals with the recording of the signals used in the listening test. The signals needed are binaural wave files, that can be played back in headphones. These files are also input to the MATLAB® program developed in this project. From this MATLAB® program another wave file is created. This file is intended for playback in the multimedia setup. All recording is done at a sampling frequency of 44.1 kHz.

When discussing the sounds that needs to be evaluated by the listeners in the listening test, the following topics were evaluated:

- Source signal
- Recording method
- Recording environment
- Recording equipment

C.1 Source Signal

The source signal chosen for the listening test is pulsed random noise generated in MATLAB®. The random noise comes in 5 pulses of 0.8 seconds. In between each pulse is silence for 0.2 seconds. This give a total stimuli file of 5 seconds. This stimuli was high pass filtered at 600 Hz. The reason for this was the lower limit of the crosstalk cancellation system. A study by [Christensen, 2001] showed that highpass filtering a signal will not deteriorate the localization.

Different sound files where investigated, here in speech signals and the pulsed random

noise using fade in/fade out. Based on preliminary testing by the project group, the pulsed random noise without fade in/fade out was used.

C.2 Recording Method

The binaural recordings were done with the artificial head. The artificial head was mounted in a mechanical turning device, where the azimuth angle could be adjusted. As mentioned in chapter 6 on page 41, Listening Test, the sound sources located with 45° spacing in the horisontal plane and 45° spacing in the median plane. To record this using the artificial head, a "ball" loudspeaker (VIFA MD10 placed in a hard platic ball) was set in a distance of 1 meter from the artificial head and no azimuth or elevation angles. Then measurements were done in the horizontal plane, by turning the artificial head to the azimuth angle needed. Then the loudspeaker was placed in an elevation angle of 45° and 315° and measurements were done with azimuth angle 0° and 180°. Finally a measurement was done with the loudspeaker situated 1 meter above the artificial head.

It should be noted that no signal processing was used to correct the response from the ball loudspeaker. The response from the ball loudspeaker are resonably flat between 600-9000 Hz.

C.3 Recording Environment

The room used for the recordings was a measuring room located at the acoustics department of Aalborg university. The measuring room was picked because of the availability, at the time of recording. Two locations were recorded: the measuring room and the anechoic chamber. Again a preliminary test done by the project group, showed that the reflections from the room gave a better perception of where the sound was comming from. Therefore this was chosen as the recording environment.

C.4 Recording Equipment

The equipment used to make these recordings are:

- An artificial head
- A 200Mhz Pentium PC with sound card
- A mixer (Behringer MX802A, lnbr. 52641)
- Ball loudspeaker (VIFA MD10)
- Pioneer A-616 amplifier for the ball loudspeaker (lnbr. 08341)

C.5 Summary

The project group recorded three different types of stimuli in two different environments. A preliminary test by the project group was used to select the input stimuli.

It should be noted that these preliminary tests could be further investigated, since optimal recordings of binaural sound will lead to a decrease in the localization errors. This could include: recording environment, stimuli and equalization of the ball loudspeaker.

List of Figures

1.1 The principle behind binaural sound. 1

1.2 *Setup seen from the top.* 3

1.3 *Setup seen from the side.* 3

2.1 HRTF and HRIR measured with a sound source located beside the left ear. (This picture is from [Hammershøi, 1995] p.3) 6

2.2 Head related transfer functions measured with both blocked and open earcanal. The bottom figures show the deviation [Hammershøi, 1995]. . . . 7

2.3 Localization blur in the horizontal plane. Studies carried out by Preibisch-Effenberger, Haustein and Schirmer in 1968-1970. 8

2.4 Minimum audible angle in the vertical plane, according to Damaske and Wagener in 1969. 8

3.1 Model of a stereo playback with labels on individual transfer functions and signals. 12

3.2 Binaural playback system including crosstalk canceller (C) and acoustic transfer function matrices (H). 13

3.3 Stereo setup for playback. S is the ipsilateral transfer function and A is the contralateral transfer funtion. 14

3.4 The Atal-Schroeder crosstalk cancellation filter array. 14

3.5 Shuffler crosstalk canceller topology. 16

3.6 Recursive implementation of a symmetric crosstalk cancellation filter. . . . 17

4.1 Setup used in the explanation of basic crosstalk cancellation. 19

4.2 Diaphragm displacement versus time for generating a 2kHz sound at point R_{mic}. 20

4.3 Phase difference between signals from R and L at R_{mic}. 21

4.4 Simulated radiation pattern for the satellite loudspeaker at different frequencies. 23

4.5 *Sound radiated from L.* 23

4.6 *Sound radiated from R.* 23

4.7 *Sound from L at right ear* 24

4.8 *Sound from R at right ear* 24

4.9 *TF's to the left ear* 24

4.10 *TF's to the right ear* 24

4.11 The resulting SPL compared with the desired at the right ear. 25

5.1 Schematic of the multimedia setup configuration. 27

5.2 Configuration for the calculation of the Wiener filter. 28

5.3 Comparison between the summed squared error and the computational complexity as a function of the filter order for the Wiener solution. Both parameter were obtained as an average from the computation of the shuffler filters based on pure HRTF for a direction of 20°. These calculations were made for 512, 1024, 2048 and 4096 filter order. 30

5.4 Head related impulse response and the corresponding HRTF for an elevation angle of 0°and azimuth angle of 20°. These correspond to the ipsilateral (solid line) and contralateral (dashed line) signals used in the shuffler filters. 31

5.5 Impulse and frequency response of the satellite loudspeaker used in the multimedia setup. This measurement was done on axis and the ripples seen at low frequencies are due to the truncation of the signal. 32

5.6 Calculated impulse and frequency response corresponding to the setup reflections. From the frequency response is possible to observe slight comb filter effects which might correspond to reflections from the desk mainly. . 33

5.7 Shuffler configuration for case I. 34

5.8 Shuffler configuration for case II. In this case the shuffler filters are based on $S' = TF_{IPSI} \cdot H_{LSPK}$ and $A' = TF_{CONT} \cdot H_{LSPK}$. 34

5.9 Shuffler configuration for case III. 35

5.10 Left-right channel separation for shuffler filters calculated from pure HRTF and a separate equalization to compensate for the loudspeaker and reflections from the setup. 35

5.11 Left-right channel separation for shuffler filters calculated from a combination of HRTF and loudspeaker response and a separate equalization to compensate for the setup reflections. 35

5.12 Left-right channel separation for shuffler filters calculated from ipsilateral and contralateral impulse response containing the HRTF, loudspeaker response and reflections from the setup. Large variations between -60 and -80 dB are due to the bandpass filtering. 36

5.13 Comparison of channel separation for symmetric crosstalk canceller simulations in case I (solid line), case II (dashed-line) and case III (dash-dot line). For visual representation these channel separations were smoothed using a 16 points average of the frequency magnitude response. 36

5.14 Implementation of the crosstalk cancellation system when adding high and low frequency band. 37

5.15 Implementation of the crosstalk cancellation system with the use of up/down-sampling. 37

5.16 Channel separation (upper) vs. artificial channel separation (lower). 38

5.17 Natural crosstalk cancellation (upper) vs. shufflers crosstalk cancellation (lower) in the physical validation. 39

6.1 sound source locations for the listening test. 42

6.2 *Picture from the listening test setup, left side.* 43

6.3 *Picture from the listening test setup, right side.* 43

6.4 Result of the listening test, where the stimulus was presented in headphones. The stimulus directions are marked on the abscissa and the response is marked on the ordinate. 44

6.5 Result of the listening test, where the stimulus was presented on the multimedia setup. The stimulus directions are marked on the abscissa and the response is marked on the ordinate. 44

6.6 The errors are divided into three categories: median plane errors (darkest), cone of confusion errors (grey) and other errors (white). The diagonal is not included in the errors since it is the correct responses. 45

6.7 Error percentage of the data from the listening test. These are sorted into three categories: median-plane errors, cone of confusion errors and other errors. 46

A.1 Setup for measuring the loudspeaker system response. 58

A.2 Setup for measuring the combined response of H_{LSPK}, H_{SETUP} and TF_{IPSI} or TF_{CONT}. 58

A.3 Setup for measuring the combined response of H_{LSPK} and TF_{IPSI} or TF_{CONT}. 58

A.4 Frequency response of the right satellite loudspeaker including subwoofer. . 59

A.5 Frequency response of the left satellite loudspeaker including subwoofer. . . 59

A.6 On axis impulse response of the right satellite loudspeaker including subwoofer. 60

A.7 $\frac{H_{LSPK_R}}{H_{LSPK_L}}$. 60

A.8 Ipsilateral transfer function, TF_{IPSI} including H_{LSPK} and H_{SETUP}. These transfer functions correspond to the left and right ear measurements. . . . 61

A.9 Contralateral transfer function, TF_{CONT} including H_{LSPK} and H_{SETUP}. These transfer functions correspond to the left and right ear measurements. 61

A.10 Ipsilateral transfer function, TF_{IPSI} including H_{LSPK}. 62

A.11 Contralateral transfer function, TF_{CONT} including H_{LSPK}. 62

A.12 Calculated H_{SETUP} based on the ipsilateral transfer function measurements. 63

A.13 Calculated H_{SETUP} based on the contralateral transfer function measurements. 63

A.14 Calculated HRTF for the ipsilateral transfer function. 63

A.15 Calculated HRTF for the contralateral transfer function. 64

B.1 Common setup used to measure Head Related Transfer Function. When measuring, P1 should be placed in the middle of the head, with the head absent. 67

List of Tables

6.1 This table shows the errors in percentage (bold) and numbers. In the bottom of the table is the amount of possible errors in each category. . . . 45

6.2 Result of the Fisher-Irwin test. 46

A.1 *The MLSSA parameters.* 57

A.2 *The equipment used during the measurements.* 57

ENCLOSURES