Contents

Preface

This report is a documentaion for a 9^{th} semester project work at the Department of Acoustics at Aalborg University. It is addressed to anyone with an interest in transaural systems.

The report contains a main report, appendices and enclosures. The main report contains an analysis and description of the developed system, test and a conclusion. The appendices include a description of the measurements and some of the results. The enclosure contains the program setup for measurements and the important parts of the Matlab programs. In the back of the report a CD is included containing parts of the Matlab programs and the datas from the measurements.

References to literature will be shown in squared brackets in the report, e.g. [4]. In the brackets the reader will find a number refering to the list of bibliographies. Figures, tables and equations are numbered according to chapters. As an example figure 4 in chapter 3 is written as: figure 3.4.

Group ACO-963, Aalborg, December 20^{th} 2000

Allan Bøgh Christensen	Karen Reuter
Carsten Højgård Jensen	Stephanie Salaun

Chapter 1

Introduction

For the last decades, scientists and engineers have investigated different methods to reproduce sound from a real life situation for a listener. It is possible to create the impression of virtual sound sources at given positions in space where no real sources exist. This can be achieved by reproducing the sound pressures at the ears of a listener, as they would have been with a real source at the same position as the virtual source [13], [3].

Different methods to reproduce the sound pressures at the ears of a listener exist. First of all it is necessary to have a signal containing the same spatial information as the real life signal. For this a binaural recording is needed. A binaural recording is a signal, which is recorded by placing microphones in both ears of a listener or using an artificial head with microphones in the "ears". Binaural signals can also be synthesized on a computer [7].

The simplest method to reproduce the binaural sound is to use headphones. When using headphones the sound is played back with only a small transmission loss, so the signal is almost reproduced correctly.
It is also possible to play back binaural sound for one listener from at least two loudspeakers by use of crosstalk cancellation [12]. Crosstalk means that a loudspeaker placed to the right of the listener is heard not only with the right ear, but also the left ear and vice versa. This is an unwanted effect in the binaural sound reproduction, because only the binaural signals are wanted to be reproduced at the two eardrums. Therefore it is desired to cancel the crosstalk.
Representing binaural signals over loudspeakers with crosstalk cancellation is often referred to as *transauralstereo*. One of the problems with transaural stereo is the high sensitivity to head movements of the listener compared to the loudspeaker position. A conventional transaural system is a system consisting of two loudspeakers reproducing a binaural sound for (two points or) one head position, called the "sweet spot". Such a system is also denoted a 2 by 2 system. If the listener moves the head outside the sweet spot, the three dimensional experience deteriorates.

The purpose of this project is to develop and test a method for maintaining a good spatial impression for small head movements of the listener. The theory for the algorithm used is inspired

by previous research made by Cooper and Bauck [2]. The main idea of this research is to use N loudspeakers for crosstalk cancelling N points.
An extension to this theory is to use two loudspeakers for N points. An algorithm for this has been implemented off-line for systems containing the same loudspeaker placement, but for different numbers of N, and simulated in Matlab.

In order to validate the developed systems both objective and subjective tests have been performed. These tests are carried out to determine whether this algorithm is an improvement to a conventional crosstalk cancellation system.

Chapter 2

Sound localisation and crosstalk cancellation theory

When we talk about binaural sound reproduction, it is necessary to know about some psychoacoustic aspects of the human hearing, especially how sound is perceived and localised. Different parameters for sound localisation have to be taken into account. Some of these parameters, relevant to binaural sound, are known as sound perception, shape of the ear and where the sound comes from. This last one can be a problem because errors can occure in the perception of the sound. The method to reproduce the spatial information as close to real life as possible is to use binaural recordings. However, reproduction of binaural sound through loudspeakers creates an unwanted signal called crosstalk. This phenomenon occures when the sound signal from one loudspeaker reaches the opposite ear of the listener.
In this chapter the relevant parameters for sound localisation are described as well as the theory of crosstalk and crosstalk cancellation.

2.1 Sound perception and localisation

The human hearing uses different cues to localise sound. In order to describe these different cues some definitions have to be presented. In order to define sound source positions relative to the head, a coordinate system is illustrated in figure 2.1.

The azimuth is given by the angle φ, which is positive for rightwards directions and negative for leftwards directions. The elevation is described by the angle δ, positive for upward directions and negative for downwards directions. The distance to the centre of the head is r. Three planes are relevant for the description of sound localisation, the median, the horizontal and the frontal plane. The median is the plane in which the azimuth angle is 0°, in the horizontal plane the elevation angle is 0°, and the plane in which the azimuth angle is 90° is called the frontal plane. For sound sources located outside the median plane, the ear closest to the source is called the ipsilateral ear

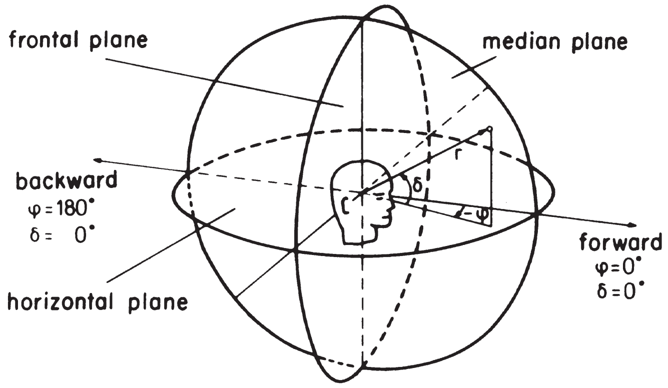

Figure 2.1: Illustration of the coordinate system used to define the position of sounds relative to the head.

while the opposite ear is called the contralateral ear.
If a sound source is positioned outside the median plane the distances from the source to the ipsilateral and contralateral ear are not the same. The difference in path length between the signals to the ipsi- and contralateral ear can be approximately calculated by the diameter of the head $2 \cdot l$ and the angle of sound incidence θ. The geometry can be seen in figure 2.2, taken from [14]. In the figure the head is described as a sphere, and the path difference can be calculated from any angle of sound incidence.

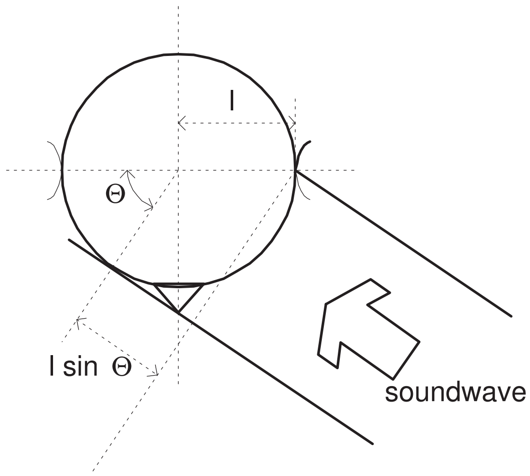

Figure 2.2: Illustration of the method for calculating the difference in arrival time at the two ears for a distant sound source at an angle θ to the observer.

The path difference d between the two ears is approximately given by:

$$d = l \cdot \theta + l \cdot sin(\theta),$$

where θ is the angle in radians.

Because of the shape of our pinna, head and torso the spectrum of incoming sound is influenced. The spectra differ depending on the angle of incidence of the sound relative to the head. Since the sound has to travel different distances to reach the ears when outside the median plane the two signals differ in arrival time, phase and level. These differences between the ears are called interaural differences. For low frequencies, when the wavelength is long compared to the dimensions of a human head, the sound wave diffracts or bends around the head which causes only a small interaural level difference (ILD). From [14] the ILDs are negligible for frequencies below 500Hz. In these cases the interaural time difference (ITD) or interaural phase difference (IPD) is used to localise sound. For high frequencies the sound does not bend around the head. Instead a sound shadow at the contralateral ear exists. This shadow causes a large ILD between the two signals and is used for sound localisation.

2.2 Sound localisation errors

One thing, which makes it difficult to localise a sound correctly, is that a certain ITD can be caused by several source positions. The sources, leading to the same ITDs, lie on the so called cones of confusion, as shown in figure 2.3 [14]. This means that the distances from two sound sources to the ipsilateral ear are equal as well as the distances from these sound sources to the contralateral ear are also equal.

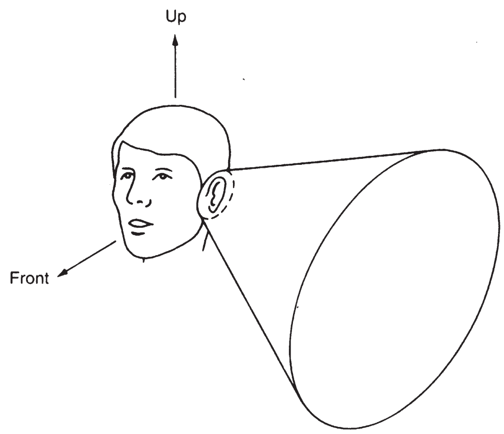

Figure 2.3: Cone of confusion for a spherical head and a particular ITD: All sound sources on the surface of the cone would produce that interaural time delay.

A special cone of confusion is the median plane in which all sound sources have the same path length to both ears. Errors in this plane are called median plane errors. The most typical median plane error is the back-to-front confusion where sound from the back is believed to come from the

front. Vice versa, when sound coming from the front is localised in the back, is called front-to-back confusion.
The correct position of the sound source can be defined by using small head movements. When the head is rotated the ITD changes, and the sound source can be localised. If no head movements are allowed (e.g. in a subjective listenening test) the listener often mix up sound sources lying on the cone of confusion.

If a sound source is localised in a position which does not lie on the same cone of confusion as the sound source it is called an out of cone error.

2.3 Binaural Recordings

As already mentioned the input to the hearing consists of two signals, namely the sound pressures at each of the eardrums. If these two signals are recorded at the eardrums of a human being and played back via headphones for the same person and position, the complete auditory experience including timbre and spatial impression should be exactly the same in case of ideal microphone and headphone frequency responses. This type of recordings, a set of left and right channel signals that are recorded at the eardrums, are called binaural recordings. For practical reasons, and because the human torso, head and pinna have different shapes and sizes from person to person, an artificial head is used for the binaural recordings. The artificial head is produced to represent an average human being.

2.4 Crosstalk and Crosstalk Cancellation

If two binaurally recorded signals are played back through loudspeakers instead of headphones the room influences the sound signal. In addition to the direct sound from the loudspeakers reflections from the walls and other surfaces are received. A more important aspect for the loudspeaker reproduction is that while the right signal is perceived at the right ear and the left signal at the left ear in headphone reproduction, there is also a transmission from the left loudspeaker to the right ear and vice versa in loudspeaker reproduction. This phenomenon is called crosstalk and can be seen in figure 2.4. Crosstalk is not an unwanted effect in stereo reproductions but unwanted in the reproduction of binaural signals.

In a normal stereo loudspeaker setup two loudspeakers are placed in front of the listener with a certain azimuth angle to the median plane. In this setup there are four transmission paths from each loudspeaker to each ear: two direct transmissions from each loudspeaker to the ipsilateral ear and two indirect transmissions from each loudspeaker to each contralateral ear.
From a certain azimuth loudspeaker angle the listener's head forms a "barrier" between the transmission paths from each loudspeaker to each contralateral ear. This effect increases with an in-

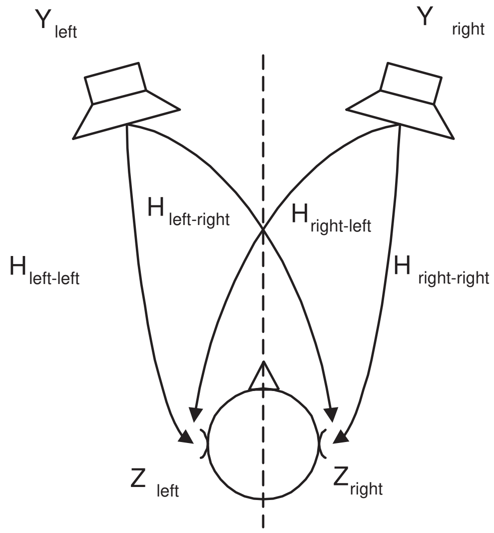

Figure 2.4: Crosstalk in a loudspeaker setup

creasing azimuth loudspeaker angle until a horizonzal angle of 90° is reached where the contralateral ear is on the opposite side of the head than the sound source. In these cases the ILD between ipsi- and contralateral ear not only results from the different sound path lengths but also from the shadowing effect of the head. The shadowing effect mainly appears at high frequencies when the dimension of the head is large compared to the wavelength of the sound. The low frequencies bend around the head and do not cause a large ILD (see 2.1). The large ILD at high frequencies which increases with increasing azimuth loudspeaker angles is also called natural crosstalk cancellation.

In order to implement total crosstalk cancellation, that means to reproduce the recorded signals separately at each ear, it is necessary to manipulate the original signals. It is possible to remove the crosstalk by adding an artificial crosstalk, which cancels the natural crosstalk. A system which can perform the task of cancelling crosstalk from a loudspeaker setup, is often referred to as a *transaural system*. The following parameters are important to describe how the crosstalk can be cancelled. All parameters are described in the frequency domain. X_{left} and X_{right} are the signals to be reproduced as sound pressures at the two eardrums. Y_{left} and Y_{right} are the loudspeaker inputs, P is the sound pressure at the centre of the head position with the head absent, Z_{left} and Z_{right} are the sound pressures at the two eardrums, and H is the head-related transfer function (HRTF), defined as the division of Z and P [7]. This division eliminates the loudspeaker responses included in both Z and P. Consequently, only the head information forms the HRTF, also called a true HRTF. $H_{right-right}$ is the HRTF from the right loudspeaker to the right ear, $H_{right-left}$ is the HRTF from the right loudspeaker to the left ear. As it can be seen from figure 2.4 the sound pressure at the ears, while playing back a signal through two loudspeakers, are:

$$Z_{left} = H_{left-left} \cdot Y_{left} + H_{right-left} \cdot Y_{right} \tag{2.1}$$
$$Z_{right} = H_{left-right} \cdot Y_{left} + H_{right-right} \cdot Y_{right} \tag{2.2}$$

The wanted signals at the ear drums are:

$$Z_{left} = X_{left} \tag{2.3}$$

$$Z_{right} = X_{right} \tag{2.4}$$

If Z_{left} and Z_{right} in equation 2.1 and 2.2 are replaced with equations 2.3 and 2.4 and solved with respect to Y, we get:

$$Y_{left} = \frac{X_{left} - H_{left-right} \cdot Y_{right}}{H_{left-left}} \tag{2.5}$$
$$Y_{left} = \frac{X_{right} - H_{left-left} \cdot Y_{right}}{H_{left-right}}. \tag{2.6}$$

Equations 2.5 and 2.6 are set equal and are solved with respect to Y_{left} and Y_{right}. We receive:

$$Y_{right} = \frac{X_{right} \cdot H_{left-left} - X_{left} \cdot H_{left-right}}{H^2_{left-left} - H^2_{left-right}}$$
$$Y_{left} = \frac{X_{left} \cdot H_{right-right} - X_{right} \cdot H_{left-right}}{H^2_{right-right} - H^2_{right-left}},$$

or in another form:

$$Y_{right} = \frac{H_{left-left}}{H^2_{left-left} - H^2_{left-right}} \cdot X_{left} - \frac{H_{left-right}}{H^2_{left-left} - H^2_{left-right}} \cdot X_{left} \tag{2.7}$$
$$Y_{left} = \frac{H_{right-right}}{H^2_{right-right} - H^2_{right-left}} \cdot X_{left} - \frac{H_{right-left}}{H^2_{right-right} - H^2_{right-left}} \cdot X_{left}. \tag{2.8}$$

This signal processing manipulates the two input signals in a way that the natural crosstalk is cancelled at the two ear drums. If the equations are represented in a flow graph it looks like figure 2.5.

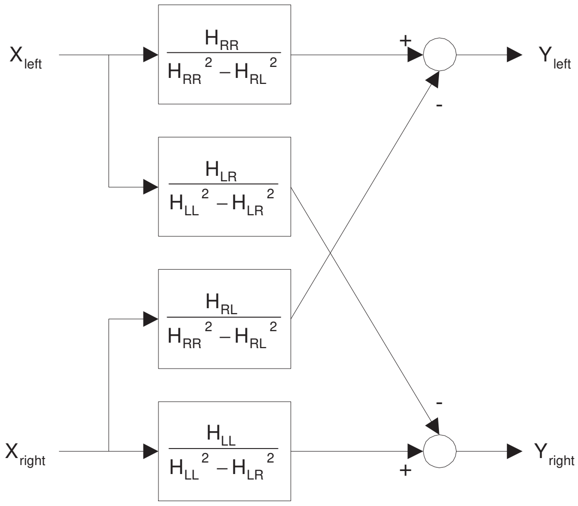

Figure 2.5: Signal processing of eqs. 2.7 and 2.8 in blockdiagram form

Ideally these equations provide perfect crosstalk cancellation and reproduction of binaural signals.

A description of the relevant sound localisation parameters for binaural sound has been made, together with an analysis of the problems regarding crosstalk when reproducing binaural sound through loudspeakers. As described, the transaural system solves this problem of playing binaural sound through loudspeakers. Such a system will be analysed in to more detail in the next chapter.

Chapter 3

Solution to head movement problem

As mentioned in chapter 2, crosstalk is calculated for two points in a conventional system, namely the two ear entrances of the listener. The position in which the listener has to be placed to get perfect crosstalk cancellation is called the sweet spot. As soon as the head moves outside the sweet spot the crosstalk cancellation is not perfect any more (e.g. [16]). Several crosstalk cancellation systems have been realised [15], [6], [16] which performed very well in the sweet spot. But it is not comfortable for a listener to stay in one position where no head movements are allowed during the play back. Furthermore when the listener position is not exactly in the sweet spot, the binaural sound reproduction deteriorates. Consequenlty the head position and head movements are a big consideration in the realisation of binaural sound reproduction.

3.1 Idea of solving the head movement problem

Cooper and Bauck [2] have developed a general theory used in crosstalk cancellation systems. The theory makes it possible to perform crosstalk cancellation for more than one listener. Using N loudspeakers, crosstalk can be cancelled for N/2 listeners (N ears).
The new idea in this project, is to move these N points from N/2 listeners to N points placed closely around the head of one listener. This should improve the sensitivity to head movements, because a movement of the head from the original position, will move the head into, or near, another point, where the crosstalk is cancelled.
A second idea, which is not considered in the article by Cooper and Bauck, is to implement a system, which uses fewer loudspeakers than points. To obtain perfect crosstalk cancellation, it is necessary to have as many loudspeakers as points. Since the points considered are now close to the head, and thereby close to each other, it is expected that their transfer functions will be so similar, that two loudspeakers can cancel the crosstalk in these points.
In this project, the application of the algorithm is to have one listener and N different head positions deviating from 5 to 10 cm around the initial position. Then the listener would have several points

around the sweet spot in which he should receive perfect crosstalk cancellation. In this way it should be possible to make head movements without loosing the spatial sound impression.
Therefore, an important parameter is to decide what area to choose around the centre position. However, it is not known how good the crosstalk canceller performs for the N points and what happens with the crosstalk cancellation between the chosen points. It could dependent on the distance between the points and perhaps also on the frequency. If the points are close to each other, the crosstalk cancellation is expected to work well, while the performance most likely will decrease with increasing distance between the points.
Since it is chosen only to use two loudspeakers, the setup is much simpler to implement regarding computational power. Also the equipment is now much cheaper, since not so many loudspeakers and amplifiers are needed. Finally the new system is not taking up a lot of space and will therefore be suited for an end user application.

3.2 Development of 2 by N system

As mentioned, a solution to the problem of many loudspeakers is to realise a system which is able to make crosstalk cancellation for N points with only two loudspeakers. In this project a system has been realised consisting of two loudspeakers, which perform crosstalk cancellation for N points: A 2 by N system. For this system the Cooper and Bauck algorithm has been used [2]. The algorithm is explained in the following section, first with a 2 by 2 system which represents a conventional crosstalk cancellation system (see figure 3.1) and then extended to a larger system.

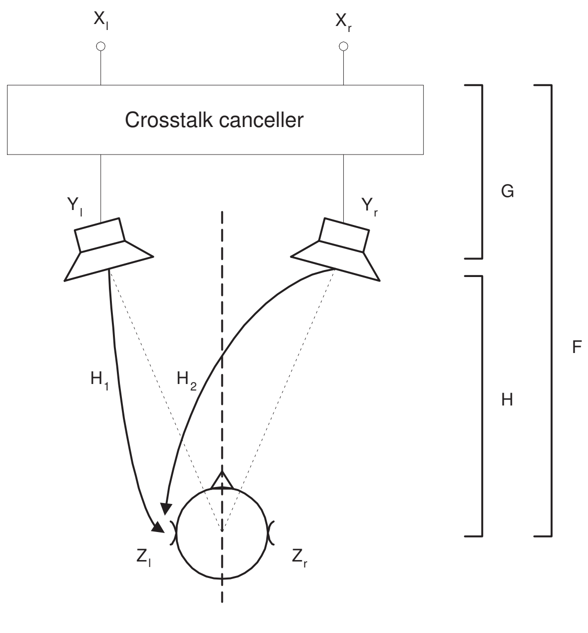

Figure 3.1: Symmetrical 2 by 2 crosstalk cancellation system

3.2.1 The 2 by 2 crosstalk canceller

X_l and X_r are the system input signals, Y_l and Y_r are the signals coming out of the crosstalk cancellation system and going into the loudspeakers. H_1 and H_2 are the transfer functions from the loudspeaker input signals to the sound pressures at the eardrums, that is the output from the microphones, (Z_l and Z_r). H corresponds to the HRTFs as they are described in [7]. In this case symmetry of the HRTFs is assumed, which means:

$$H_1 = H_{left-left} = H_{right-right}$$
$$H_2 = H_{left-right} = H_{right-left}$$

Index 1, in H_1, describes the direct transmissions from right to right and left to left while index 2 describes the indirect transmissions from right to left and left to right. In the 2 by 2 system the sound of both loudspeakers contribute to the sound pressure level at each eardrum. It can be written as:

$$Z_l = Z_{left} = H_1 \cdot Y_{left} + H_2 \cdot Y_{right}$$
$$Z_r = Z_{right} = H_2 \cdot Y_{left} + H_1 \cdot Y_{right}$$

As it can be seen from figure 3.1 the transmission from the signal input X to the signals at the eardrums Z is described by the transmission F. This transmission is divided into two parts: G, which is the crosstalk cancellation system from the system input X to the loudspeaker input Y and H, the transmission from the loudspeaker output Y to the sound pressures at the eardrums Z. For simplification we can put up 2×2 matrices to describe the transmissions.

$$\begin{bmatrix} Z_l \\ Z_r \end{bmatrix} = \begin{bmatrix} H_1 & H_2 \\ H_2 & H_1 \end{bmatrix} \cdot \begin{bmatrix} Y_l \\ Y_r \end{bmatrix} \tag{3.1}$$

The crosstalk canceller (G) described in section 2.4, which is the inverse of eq. 3.1 can be described as:

$$\begin{bmatrix} Y_l \\ Y_r \end{bmatrix} = \begin{bmatrix} G_1 & G_2 \\ G_2 & G_1 \end{bmatrix} \cdot \begin{bmatrix} X_l \\ X_r \end{bmatrix} \tag{3.2}$$

$$\tag{3.3}$$

F is then defined as a multiplication in the frequency domain of G and H:

$$\begin{bmatrix} Z_l \\ Z_r \end{bmatrix} = \begin{bmatrix} F_1 & F_2 \\ F_2 & F_1 \end{bmatrix} \cdot \begin{bmatrix} X_l \\ X_r \end{bmatrix} = \begin{bmatrix} H_1 & H_2 \\ H_2 & H_1 \end{bmatrix} \cdot \begin{bmatrix} G_1 & G_2 \\ G_2 & G_1 \end{bmatrix} \cdot \begin{bmatrix} X_l \\ X_r \end{bmatrix} \tag{3.4}$$

As for the HRTFs, symmetry between direct and indirect transmissions are assumed.
Because it is desired to receive the system input signal X_l at the left ear and X_r at the right ear as described in eqs. **??** and **??** ($Z_l = X_l$ and $Z_r = X_r$) the desired F will be:

$$\begin{bmatrix} Z_l \\ Z_r \end{bmatrix} = \begin{bmatrix} F_1 & F_2 \\ F_2 & F_1 \end{bmatrix} \cdot \begin{bmatrix} X_l \\ X_r \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_l \\ X_r \end{bmatrix} \tag{3.5}$$

The matrix $\underline{\underline{F}}$ is often referred to as the target matrix, because this is how the transfer function from the input of the crosstalk canceller to the sound pressure at the eardrums ideally should be. The matrix $\underline{\underline{H}}$ is known. It consists of the HRTFs which can be calculated from measurements. Then $\underline{\underline{G}}$ can be calculated by taking the inverse of $\underline{\underline{H}}$:

$$\underline{\underline{G}} = \underline{\underline{H}}^{-1} = \frac{1}{{H_1}^2 - H_2^2} \cdot \begin{bmatrix} H_1 & -H_2 \\ -H_2 & H_1 \end{bmatrix} \tag{3.6}$$

This is the same result as obtained in section 2.4, except now symmetrical HRTFs are assumed. Cooper and Bauck ([2]) only shows examples on systems with more loudspeakers than points or the same number of loudspeakers and points, whereas the intention in this project is that the crosstalk cancellation system can be extended to more points than loudspeakers. A matrix has to be created consisting of all transmission paths from each loudspeaker to each point. As an example a 2 by 4 system, consisting of two loudspeakers and four points, is explained in the following.

3.2.2 2 by 4 Crosstalk cancellation

An example of the 2 by 4 system is seen in figure 3.2. The points are placed next to each other symmetrically to the loudspeaker mirror line.

In total there are eight transmission paths from the loudspeakers to all the eardrums. From these eight paths only four are unique because symmetry is assumed. H_1 and H_3 describe the transmission from the left loudspeaker to the ears of the left listener while H_2 and H_4 describe the transmission to the right listener. The transmissions from the right loudspeaker are, because of the symmetry, assumed to be the same as the corresponding from the left loudspeaker. They are shown with dashed lines in figure 3.2. The sound pressure at each eardrum is the sum of the signals from both loudspeakers. This gives the following equations:

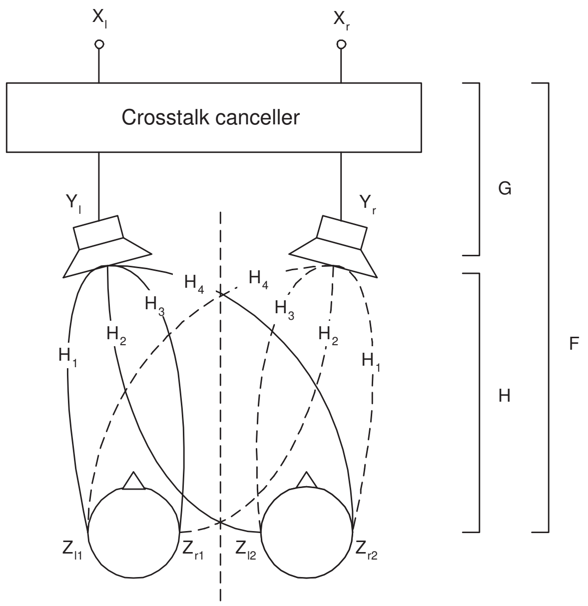

Figure 3.2: Symmetrical 2 by 4 crosstalk cancellation system

$$
\begin{aligned}
Z_{l1} &= H_1 \cdot Y_l + H_2 \cdot Y_r \\
Z_{r1} &= H_3 \cdot Y_l + H_4 \cdot Y_r \\
Z_{l2} &= H_4 \cdot Y_l + H_3 \cdot Y_r \\
Z_{r2} &= H_2 \cdot Y_l + H_1 \cdot Y_r
\end{aligned}
$$

or in matrix form:

$$
\begin{bmatrix} Z_{l1} \\ Z_{r1} \\ Z_{l2} \\ Z_{r2} \end{bmatrix} = \begin{bmatrix} H_1 & H_2 \\ H_3 & H_4 \\ H_4 & H_3 \\ H_2 & H_1 \end{bmatrix} \cdot \begin{bmatrix} Y_l \\ Y_r \end{bmatrix} \tag{3.7}
$$

To receive a transmission without crosstalk only four transfer functions are required, the transmission from the left loudspeaker to both left ears and from the right loudspeaker to both right ears of the two listeners. The wanted signals are the left input signal X_{l1} at Z_{l1} and Z_{l2} and the right input signal X_r at Z_{r1} and Z_{r2}.

$$Z_{l1} = Z_{l2} = X_l \tag{3.8}$$
$$Z_{r1} = Z_{r2} = X_r \tag{3.9}$$

The matrix $\underline{\underline{F}}$ is created with respect to equations 3.8 and 3.9.

$$\underline{\underline{Z}} = \underline{\underline{F}} \cdot \underline{\underline{X}} \Rightarrow \begin{bmatrix} Z_{l1} \\ Z_{r1} \\ Z_{l2} \\ Z_{r2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_l \\ X_r \end{bmatrix} \tag{3.10}$$

The matrix needed for the crosstalk cancellation system $\underline{\underline{G}}$ can now be calculated from the inverse $\underline{\underline{H}}$ matrix and the $\underline{\underline{F}}$ matrix.

$$\underline{\underline{G}} = \underline{\underline{H}}^{-1} \cdot \underline{\underline{F}}$$

With this solution the transmission from the system input signal $\underline{\underline{X}}$ to the loudspeaker input signal $\underline{\underline{Y}}$ can be implemented, and crosstalk cancellation is reached.

The Cooper and Bauck algorithm can easily be extended to any number of listeners and loudspeakers by creating the $\underline{\underline{F}}$ matrix for the desired transmission paths between each loudspeaker and each calculated point.
In this project it was chosen to realise a practical crosstalk cancellation system which is not too expensive and could be used as an end user application. It consists of two loudspeakers and has the possibility of being extended to different numbers and positions of crosstalk cancelled points, a 2 by N system.

Chapter 4

Transaural prospects

When reproducing binaural sound through loudspeakers it is necessary to investigate what kind of setup should be chosen for the system depending on the application. As mentioned in the previous chapter it has been chosen to implement a system that could be used as an end user application. This could be a multimedia setup or a normal living room. Many different parameters have to be taken into account to get a system which can reproduce binaural signals through loudspeakers. In this chapter these parameters are described, discussed and a setup is chosen. When the setup has been considered the head movements should be analysed. In order to implement a crosstalk cancellation system the different positions should be taken into account.

4.1 Aspects of loudspeaker choice

The most important parameter, when building a transaural system, is the loudspeakers. Some earlier work has been about how to exploit the advantages of performing crosstalk cancellation for multichannel systems, [15] and [8]. The idea was to split the frequency range up into sections, where the crosstalk is cancelled by single loudspeaker units for each frequency band. The advantage of these systems is that the units can be placed at different positions around the head. Where to place the units and which units to use is individual from application to application. To obtain e.g. a large amount of natural crosstalk a good idea would be to place e.g. the tweeters with a large azimuth angle, so that the head will block the path to the contralateral ear. However, there is also problems in placing the loudspeakers with a large azimuth angle. If crosstalk cancellation is performed on these loudspeakers, then this will be extremely sensitive to sideways head movements because even small head movements will change the time properties of the loudspeaker-ear path significantly. An alternative, which is robust to head movements would be to place the tweeters very close to each other, with a small azimuth angle. These tweeters would then obtain stereo dipole properties [17], which is robust to head movements. This is because head movements only give a small change in the distance to loudspeakers with a small spacing compared to head move-

ments for a large spacing of the loudspeakers. This is however only a good idea with the tweeters. If it is assumed that the same is done for low frequencies, where the wavelength is long compared to the distance between the units, the dipole effect would cause the loudspeakers to almost cancel each other and therefore makes this a very bad solution. Instead it could be chosen to use a subwoofer for the lowest frequencies (e.g. up to 150Hz), and then not perform crosstalk cancellation in this range. Even though no crosstalk is cancelled in this frequency range it is still supposed to work well in the reproduction of binaural signals because not much spatial information is contained in the low frequencies [4].
As it might appear from this discussion, it is often a compromise to choose the setup of the loudspeakers and also very dependent on the application for the system. In this project it is chosen to implement a system consisting of two loudspeakers. It is also chosen not to use a subwoofer for the low frequencies to simplify the setup, the crosstalk filters and also because it is assumed that this extension could easily be made, if the system performs well.

4.1.1 Choice of loudspeakers

The choice of loudspeakers for the system is based on three parameters:

1. When implementing a crosstalk system it is necessary to use loudspeakers with almost identical frequency responses [17].

2. It is preferred to have loudspeakers with a smooth and flat frequency response, because this will cause least problems in the inversion process.

3. The loudspeakers should be able to deliver/handle enough power, without distortion, when the HRTFs are inverted in the crosstalk cancellation.

The loudspeakers could be any type, from cheap PC-loudspeakers to advanced and expensive 2- or 3-way loudspeakers. To find a usable pair of loudspeakers measurements were made on four different types. See appendix **??** for more details. The measured loudspeakers were two normal PC loudspeakers, four BOSS MA-12V and two Jolly FPT loudspeakers. These three types are all active loudspeakers with build in amplifiers. Furthermore the Jolly and the PC loudspeakers are 2-way systems, which contain cross-over networks. Also four ball-shaped loudspeakers with Vifa MD10 units were included in the test. These are passive loudspeakers, and therefore need external amplifiers. The measured frequency responses are shown in figure 4.1

From the measurements it has been found that the ball loudspeakers have the most similar and smooth characteristics, so they were chosen for the system.

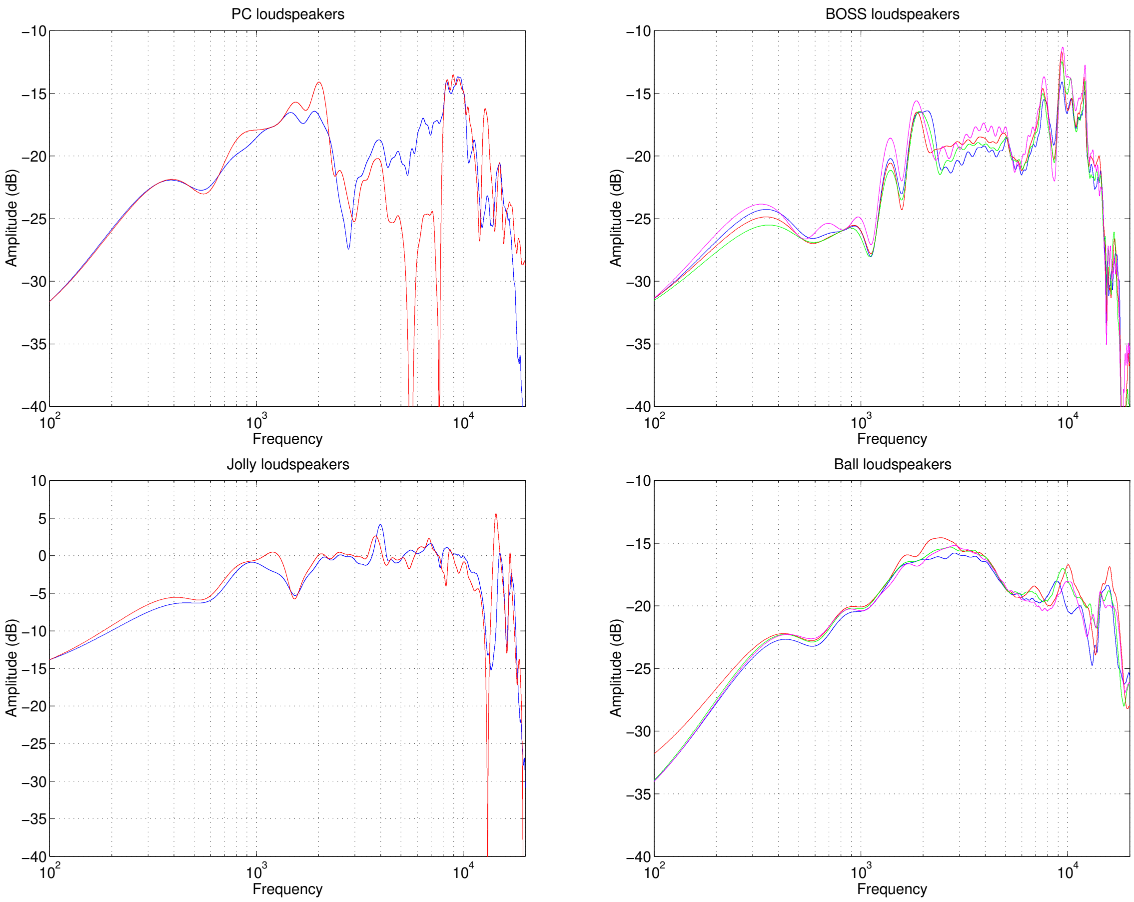

Figure 4.1: Frequency responses for four types of loudspeakers

4.2 Placement of loudspeakers

The placement of the loudspeakers is also essential for the spatial impression of a crosstalk cancellation system. This applies for both the elevation and the azimuth angle to the listener and also the distance from the listener to the loudspeakers.

4.2.1 Elevation angle

To decide how the loudspeakers should be positioned in the vertical plane, transfer functions were measured with a loudspeaker at different elevation angles. For the measurements the loudspeakers were placed at elevation angles of 0° in front of an artificial head, Valdemar, and up to 90° above Valdemar with intervals of 15°. The aim of these measurements was to find a loudspeaker position where the frequency response of the transfer functions was as smooth and flat as possible. The more smooth and flat the transfer function, the simpler inverse filters for implementation of the crosstalk canceller and the less influence small head movements should have for the system.
All of the measured transfer functions were compared and from these, two were chosen to be implemented in the crosstalk system. An elevation angle of 90° was chosen because its transfer function was the smoothest and flattest. Since it was not known how well the system would perform for the 90° angle an elevation angle of 0° was chosen. The tranfer functions for these two angles are shown in figure 4.2. The rest of the measured tranfer functions can be seen in appendix **??**.

From figure 4.2 it is seen that none of the transfer functions are particularly smooth and flat, but the transfer function measured at 90° has less and smaller dips in the frequency response. However, the sound pressure level difference from the ipsilateral side to the contralateral side, in some regions, is almost 10dB.

4.2.2 Azimuth angle

From previous research [17] it was found that the smaller the spacing of the loudspeakers - the smaller the azimuth angle of the loudspeaker to the listener - the more robust the system is to small head movements. The reason for this is that the distance from one loudspeaker to the contralateral ear compared to the distance to the ipsilateral ear is changed more when the loudspeakers are more off centre.
The disadvantage of small azimuth angles is that less natural crosstalk cancellation is obtained because of less shadowing effect of the head for high frequencies. Therefore, more work has to be done by the crosstalk cancellation system. Furthermore, a small spacing can cause problems with the power handling because of two things: The system, which has to be inverted to obtain crosstalk cancellation often demands large amplitudes in the low frequency region, furthermore the loudspeakers will act as a stereo dipole. A stereo dipole effect occures for frequencies where

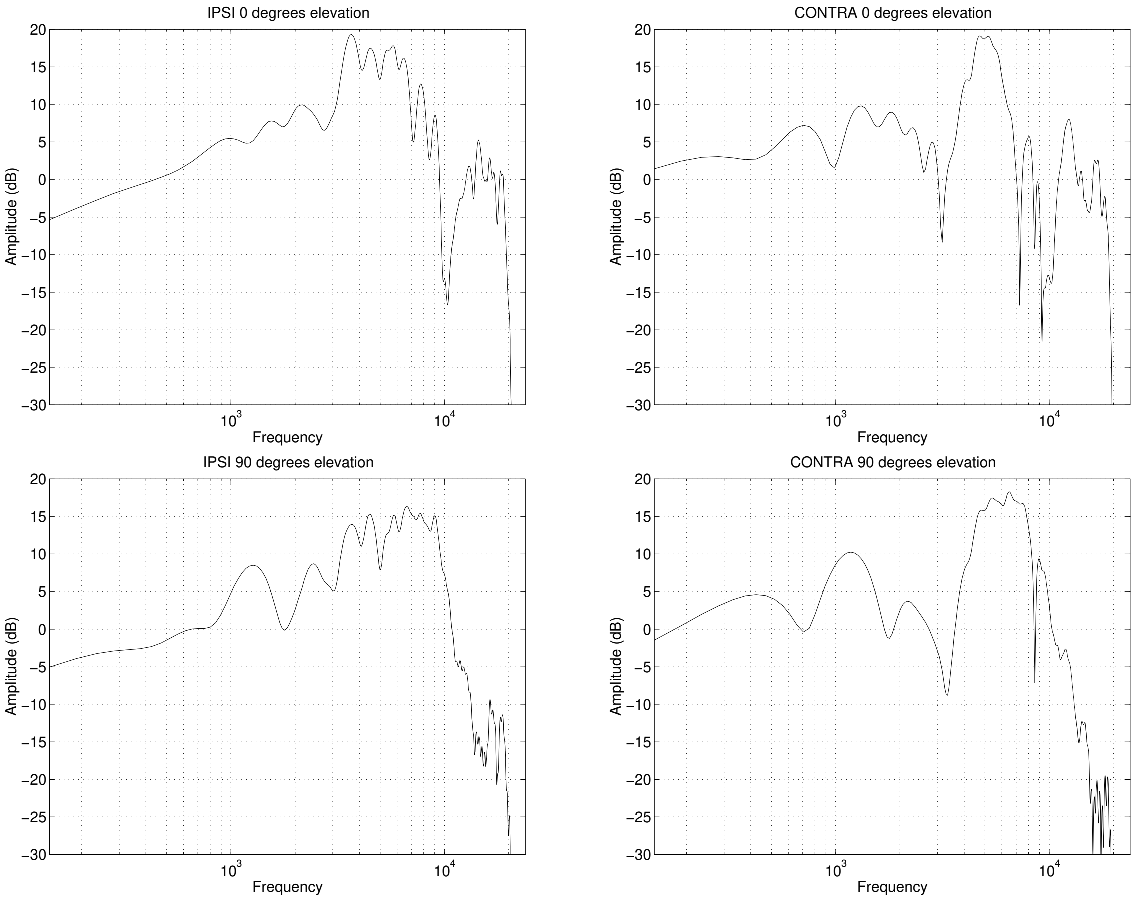

Figure 4.2: HRTFs for 0°and 90°elevation, measured on Valdemar with 15°azimuth between loudspeaker and artificial head

half the wavelength is equal to or more than the spacing between the loudspeakers. Therefore it is neccesary to increase the power considerably at these frequencies to produce the same sound pressure level as in the rest of the frequency range.
Therefore, a compromize between natural crosstalk cancellation and power handling for wide loudspeaker angles, and robustness to head movements for narrow angles need to be needed. It was decided to place the ball loudspeakers with an azimuth angle of 15° from the center to the listener. The natural crosstalk cancellation for wide azimuth angles did not have a high priority in the decision because the crosstalk should be cancelled for an area aroundthe head in this project.

4.2.3 Distance and influence of the room

The choice of distance from the listener to the loudspeaker depends on the room in which the application will be used. It is desired only to have the direct sound from the loudspeakers at the listeners ears. But in a normal room there will always be reflections from the walls reaching the listeners ears and influencing the sound. This is why the setup was placed in the VR-room at the Department of Acoustics. This room is damped and thereby has a short reverberation time. It has been measured in [1] and has an average of around 0.2sec. However, the reverberation time for high and low frequencies is very different.
The loudspeakers were placed at a distance of 1.5m to the center of listener's head. To verify this distance as a good choice, two demands have to be met:

1. *The distance has to be larger than five times the diameter of the head.* According to work done in [5] this is the minimum distance to ensure that the low-frequency ILD variation outside this distance is not significant in the HRTFs.
2. *The distance has to be smaller than the reverberation distance of the room.* The reverberation distance is the radius where the level of the direct sound and the diffuse sound are the same. Therefore it is desirable to be inside this area.

The first demand is fulfilled since five times the radius of a head is a around 1m. The second demand we have to check with the formula for the reverberation distance [11]:

$$r_d = \frac{1}{4}\sqrt{\frac{A}{\pi}},$$

where A is the arbsorption area of the room, calculated from Sabine's formula [9] to $134.4m^2$. This gives a reverberation distance of $1.64m$, and thereby the second demand is also kept.

4.2.4 Summary of loudspeaker setup

In the previous sections it has been discussed which type of loudspeakers to use and where to put them in an implementation of a crosstalk cancellation system. The choices made are:

- It is chosen to use the passive ball loudspeakers with an external amplifier for the system because they are the most similar to each other of the considered types.
- Two different elevation angles are chosen to be implemented: 0° and 90°. 90° because this elevation has the most smooth and flat HRTfs and 0° because this solution is well documented and is known to work.
- The azimuth is chosen to be 15° from the centre because this angle is expected to be rather robust to movements of the head.
- The distance from listener to loudspeakers is chosen to be 1.5m.

From this decision the system is set up according to figure 4.3.

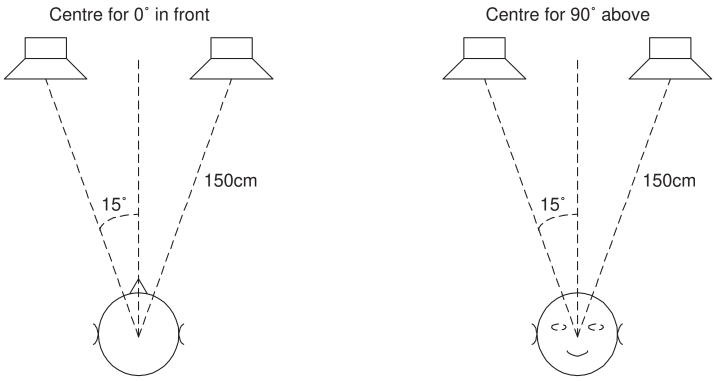

Figure 4.3: The placement of loudspeakers for the systems chosen to be implemented.

4.3 Designing the crosstalk cancellation system by taking head movements into account

As described in section 3.1, it is chosen to realize a crosstalk cancellation system which is robust to head movements by extending it to more head positions around the sweet spot. To do this it is needed to measure N HRTFs for these N/2 head positions for use in the 2 by N crosstalk canceller. Since a good choice of N must be based on simulations or perhaps even tests on a real system, it is chosen to measure the HRTFs in a grid with enough points, making it more than sufficient for the implementation. A grid has been chosen to be able to record the HRTFs for different positions. The dimensions of the grid should reflect small head movements of the listener. From the illustrated grid in figure 4.4 an area of $20 \cdot 20cm$ including 41 points can be seen.

The outermost points in the grid are chosen to see how the crosstalk cancellation performs for large head movements. HRTFs are measured for all points in the grid and for both 0° and 90°

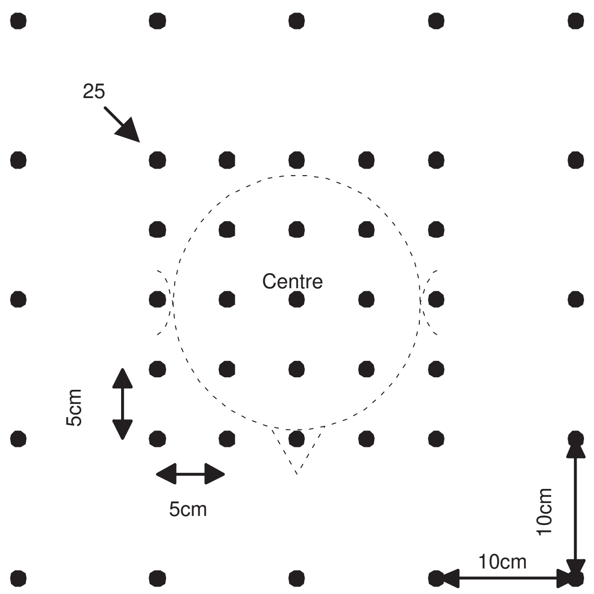

Figure 4.4: The grid used for measuring HRTFs 0° and 90° elevation. In total the grid contains 41 points.

elevation to be able to realize both systems. The method of measuring the HRTFs is found in **??**.

In figure 4.5 two of the ipsilateral head related impulse responses (HRIRs) and their respective HRTFs measured in the grid at 0° elevation is shown. The measurements are made with in the center position and in position 25, which is marked with an arrow in figure 4.4. Figure 4.6 shows the contralateral HRIRs and HRTFs for the same measuring positions.

It appears from figure 4.5 and 4.6 that the delay of the HRIR measured in position 25 is larger than the centre position. This is true because the head is further away from the loudspeaker in position 25, which means that the delay must be longer.
The frequency response does not change much though from the centre position to position 25, but there is a significant difference from the ipsi- to the contralateral side of the head. The shape has not changed much but the amplitude is smaller on the contralateral side which is caused by the shadow of the head.
From these HRTF measurements many different 2 by N systems can be realized. A 2 by 18 system could e.g. be implemented with the HRTFs measured for nine head positions closest to the center. It could also be done with the center position and the corresponding 8 points from the outer square of points. But in general the grid makes it possible to use any combination of points.
Simulation and test have been made to determine how many and which N points to include in the final system.

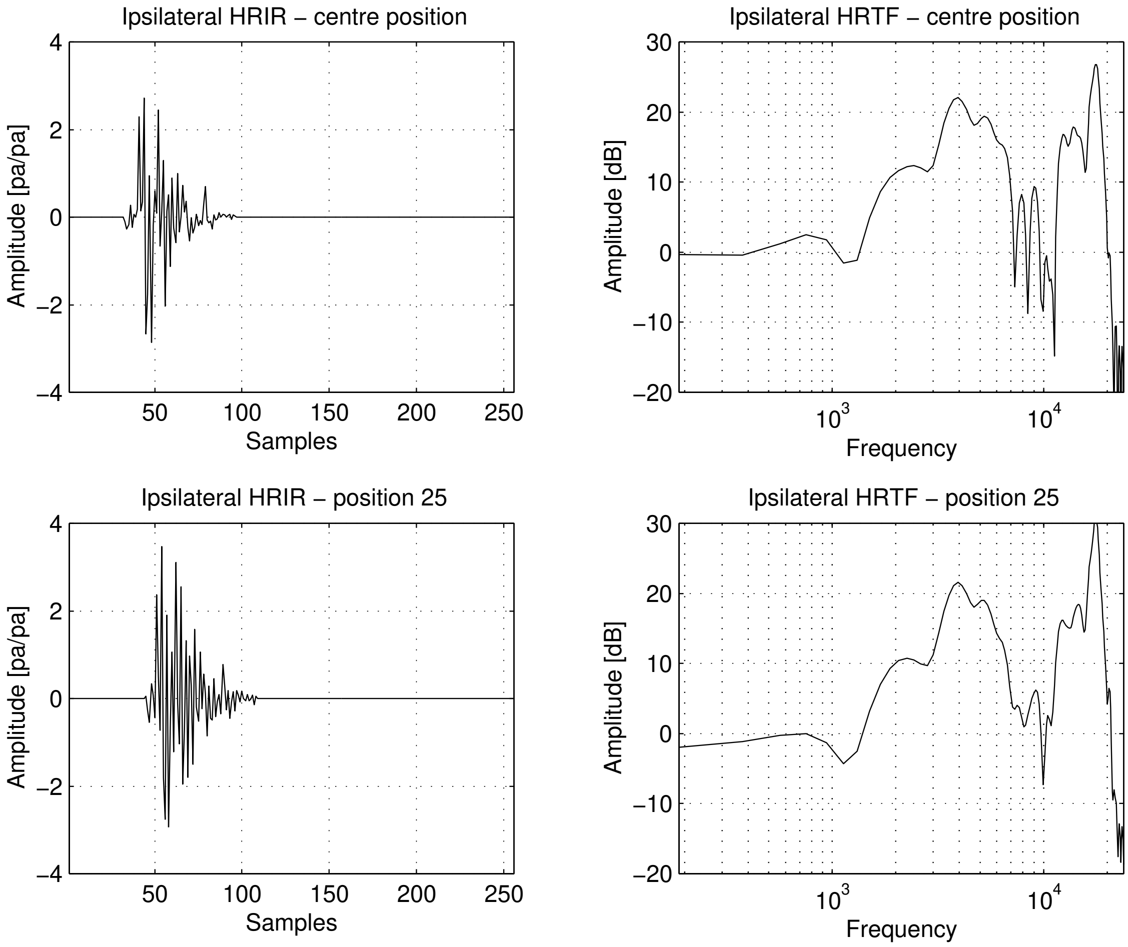

Figure 4.5: Ipsilateral HRTFs measured in grid position 1 (centre) and 25.

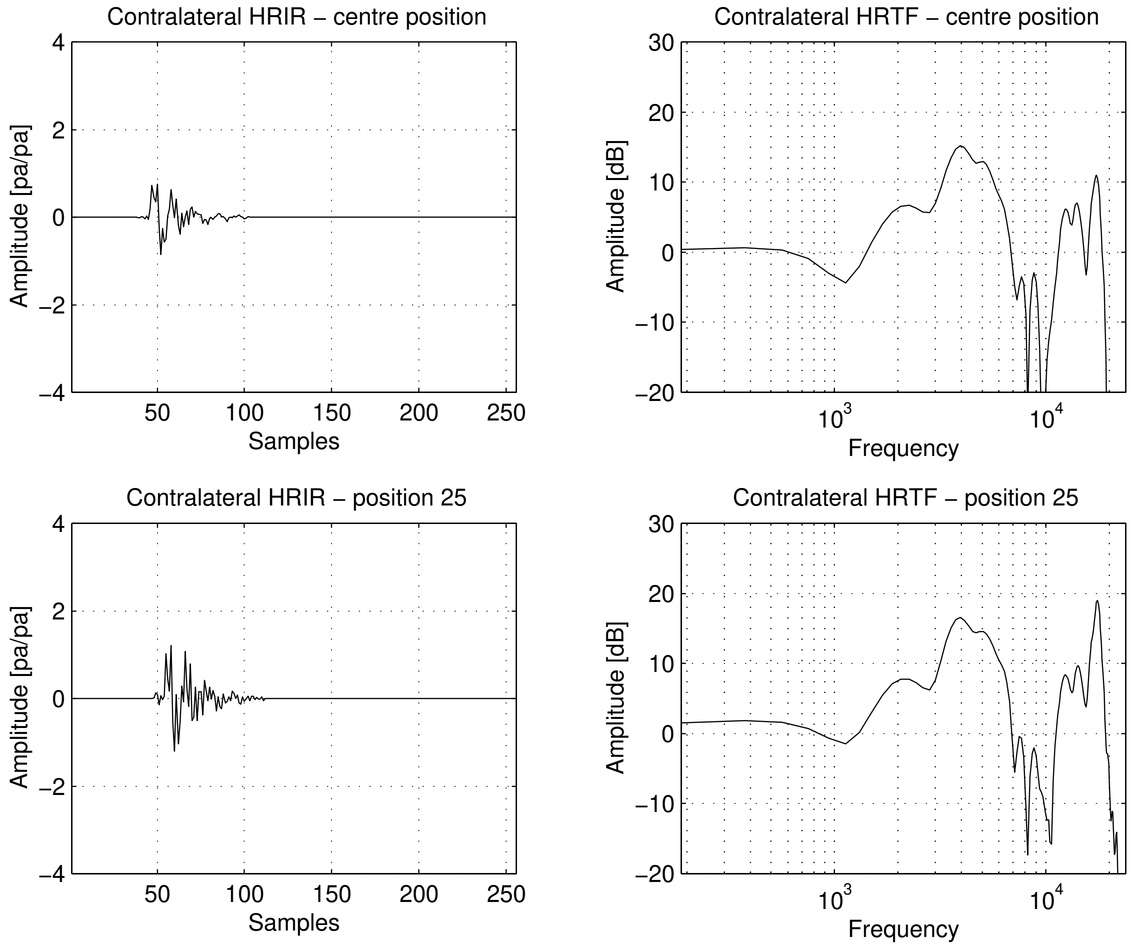

Figure 4.6: contralateral HRTFs measured in grid position 1 (centre) and 25.

4.4 Choice of crosstalk cancellation systems

From the measurements of HRTFs in the grid of figure 4.4 different 2 by N systems can be implemented. To get an impression of how well the crosstalk cancellation algorithm works for different numbers of N and for different positions in the measured grid, it was decided to implement four different systems for both 0° and 90° elevation. The chosen systems are seen in figure 4.7.

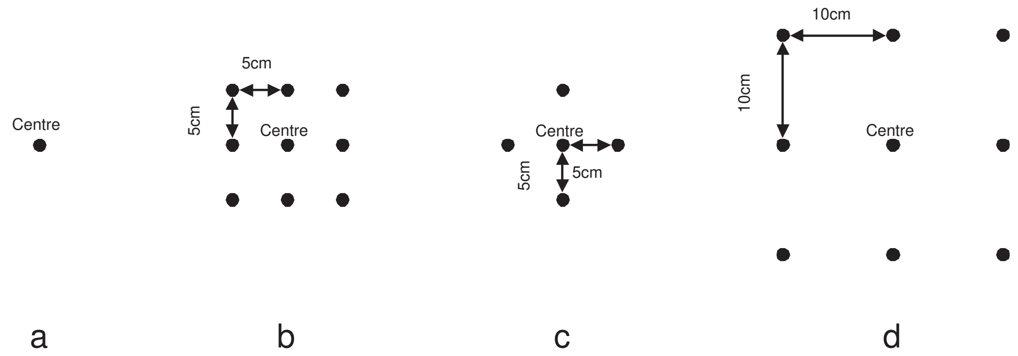

Figure 4.7: The systems which are chosen to be implemented in the crosstalk canceller.

System *a* in figure 4.7 is the normal 2 by 2 system, where the crosstalk is only cancelled for one position of the listener. System *b* includes nine positions on the grid, all with 5cm to the nearest point. This system is denoted as: *2 by 18-5cm* in the rest of the report. The *c* system includes five positions, also with a distance of 5cm to the nearest point. This system is denoted: *2 by 10-5cm* in the rest of the report. System *d* includes 9 positions, like system *b*, but here the distance to the nearest point is 10cm. This system is denoted: *2 by 18-10cm*.
The reason for choosing these systems, with a different number of N and a different spacing between the N/2 positions, is to be able to compare a conventional 2 by 2 system to different 2 by N systems. It is decided not to use the points in the outer square of the measured grid, because the differences in the HRTFs are so large that the algorithm is expected not to give a satisfactory crosstalk cancellation. Therefore a maximum distance between each of the N/2 positions is 10cm. No more than nine positions are chosen, because the effect of the algorithm is not expected to be significant. Furthermore the two loudspeakers should have limitations of functionement. The more positions that are crosstalk cancelled the more the crosstalk cancellation deteriorate.

To sum up, two setups of loudspeakers, 0° angle and 90° angle, have been carried out as well as four crosstalk cancellation systems, 2 by 2, 2 by 18-5cm, 2 by 10-5cm and 2 by 18-10cm. In total eight implementations have been considered.

Chapter 5

Implementation of the crosstalk canceller

The different systems, described 4.4 were implemented with the algorithm described in chapter 3. The systems have all been implemented with the same algorithm, making it possible to compare them to each other. In this chapter is is desribed how this algorhtm works and how the signal was processed with the crosstalk filters.

5.1 Implementation of 2 by N system

The four crosstalk cancellation systems have been implemented for both 0° and 90° elevation of loudspeakers in Matlab. Consequently eight systems are implemented. The general structure of the program is the same for each system. The purpose of this section is to describe step by step the implementation of one system as shown in figure 5.1. The explanation of the algorithm is illustrated by the calculation of the crosstalk filters for the 2 by 2 system, 0° elevation.

5.1.1 Loading the transfer functions

The binaural transfer functions measured in the grid are loaded at the beginning of the program. Both the transfer functions for 0° and 90° elevation angle are loaded. For one position of the loudspeaker and each position in the grid (41 positions), two transfer functions are loaded:

- The contralateral transfer function H_{LL},
- The ipsilateral transfer function H_{LR}.

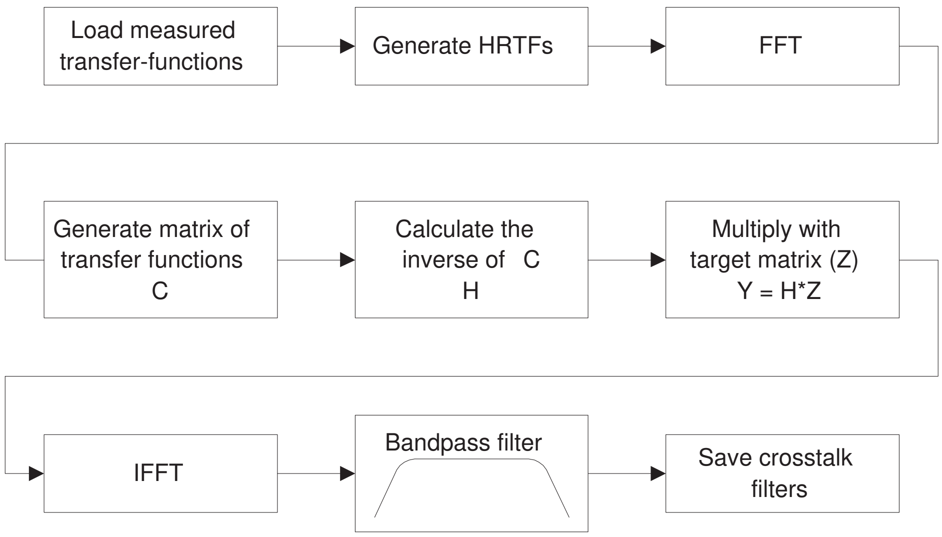

Figure 5.1: General program structure.

In order to convert these transfer functions to true HRTFs and to remove the response of the loudspeaker, the transfer function called the free field correction of the head, P (see section 2.4), is needed for the center position of the grid. As there are two setups of loudspeakers there are two Ps recorded and loaded. In total, N H_{LL}, N H_{LR} and one P are loaded for one setup.
for the example of the 2 by 2 system the loaded impulse responses for the center position is plotted in figure 5.2.

5.1.2 Generation of the HRTFs

As the true HRTF does not include the loudspeaker characteristics, the measured $H_L L$ and $H_L R$ are transformed in order to isolate the head caracteristics.
The calculation consists in deconvolving H_{LL} and H_{LR} with P [7]. The method is the same for each $H_L L$ and $H_L R$ and can be decomposed as follow:

1. 1.Normalise $H_L L$, $H_L R$ and P by multiplying the measured transfer functions by taking the sensitivity of the microphones and the measurement equipment into account.

2. 2.Remove the delay present at the beginning of the impulses responses corresponding to the distance loudspeaker/ear. This is done to ensure, that the effective length of the impulse responses is the same for all positions in the grid.
   It is however preferable, that the delay of $H_L L$ and $H_L R$ is bigger than the delay of P to get a causal result of the deconvolution.

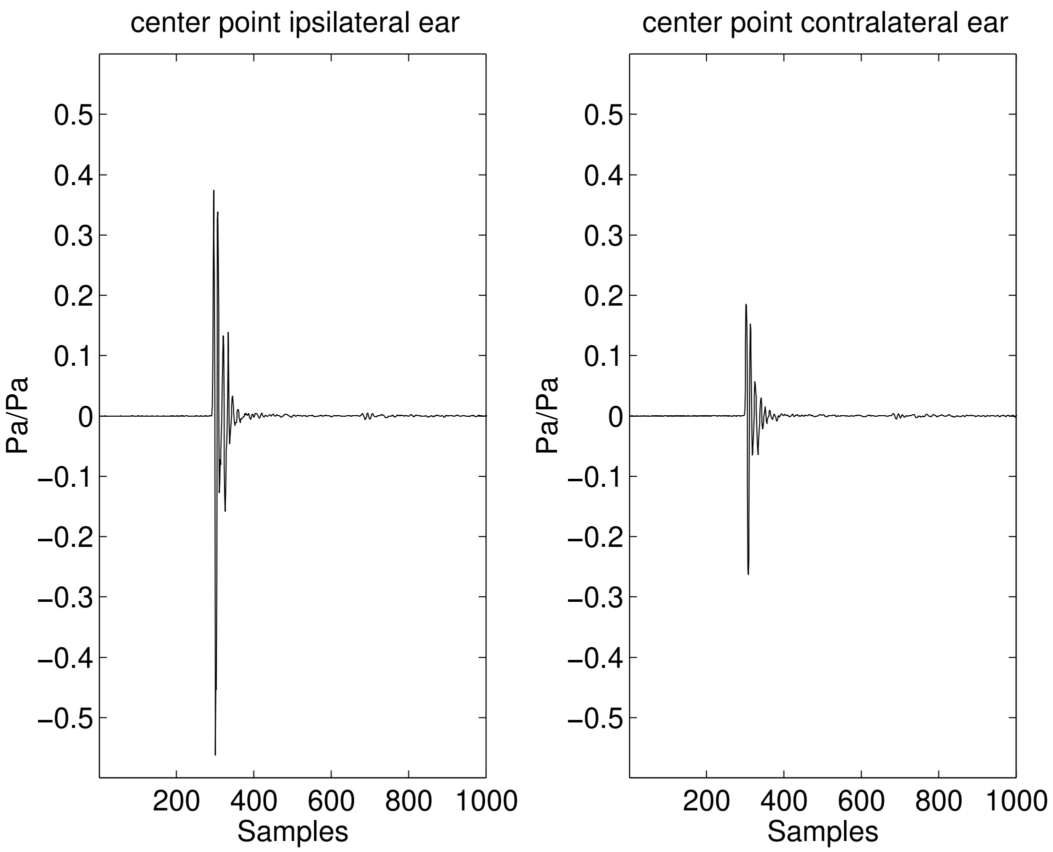

Figure 5.2: Loaded impulse and frequency responses at the center position of the grid for both contra- and ipsilateral ear and the free field correction of the head (P).

3. 3. Perform the deconvolution by a division of $H_L L$ and $H_L R$ with P in the frequency domain, by taking a long FFT of the impulse responses. Eight times the length of the impulse response is used for the length of the FFT to avoid aliasing of the circular deconvolution. After the division the result is again transformed back to time domain and truncated.

4. 4.Reinsert the delay which was removed in 2.

The impulse- and frequency responses of the HRTFs for the 2 by 2 system are illustrated in figure 5.3

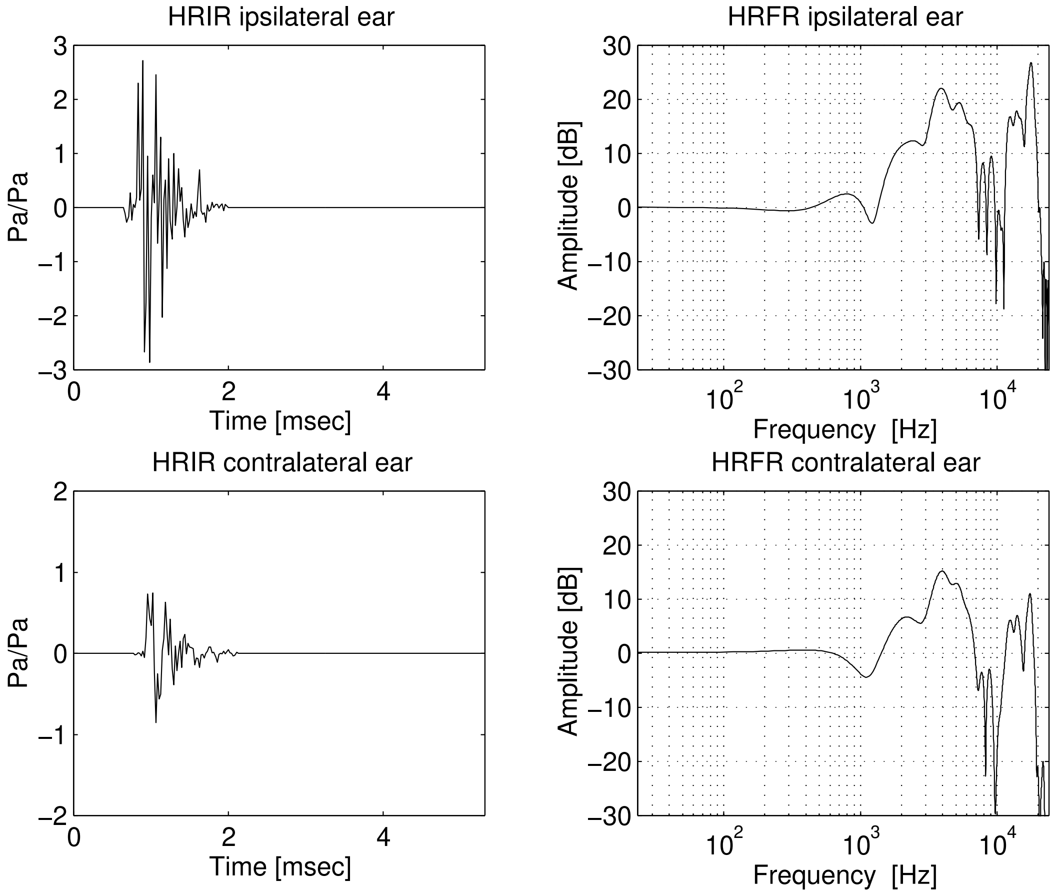

Figure 5.3: The impulse and frequency responses of the true HRTFs for both the ipsi- and contralateral ear.

Measurement of transfer functions with the MLSSA system, which is used in this project, is not very precise in the measurement of the DC sound pressure level. Therefore the DC values for the different positions in the measured grid should have been considered. The centre position should have a DC component of 0dB and the DC value of the other points should be referenced to this position. This is however not performed since the signal is bandpass filtered from 150 Hz to 19 kHz.

5.1.3 Applying FFT to the HRTFs

Since the calculation of the crosstalk cancellation filters is carried out in the frequency domain. A long FFT is again applied in order to avoid the aliasing. Again the length of the FFT is eight times the length of the HRTFs.

5.1.4 Generation of the C matrix

The matrix $\underline{\underline{C}}$ of transfer functions is given from the Cooper and Bauck algorithm in [2]. It is composed of the previously calculated HRTFs represented in the frequency domain.
For a 2 by 2 system the matrix $\underline{\underline{C}}$ has the following form:

$$\underline{\underline{C}} = \begin{bmatrix} C_1 & C_2 \\ C_2 & C_1 \end{bmatrix}$$

For a 2 by N system the matrix $\underline{\underline{C}}$ will be a have the dimension 2 by N:

$$\underline{\underline{C}} = \begin{bmatrix} C_1 & C_2 \\ \vdots & \vdots \\ C_{1N} & C_{2N} \\ C_2 & C_1 \\ \vdots & \vdots \\ C_{2N} & C_{1N} \end{bmatrix}$$

This C matrix now needs to be inverted to obtain the crosstalk cancellation system. The C matrix is visualised in figure 5.4.

5.1.5 Calculation of the inverted C matrix, the H matrix

The impulse responses of the transfer functions that compose the $\underline{\underline{C}}^{-1}$ matrix,$\underline{\underline{H}}$, must be stable and causal. Kirkeby [10] has proposed a good algorithm which can perform this invertion and which is able to take care of non minimum phase problems, if any. Consequently the matrix $\underline{\underline{H}}$ is calculated by applying this method. Applying to the system the matrix $\underline{\underline{C}}$ is inverted using the following formula:

$$\underline{\underline{H}}(k) = [\underline{\underline{C}}^H(k)\underline{\underline{C}}(k) + \beta\underline{\underline{I}}]^{-1}\underline{\underline{C}}^H(k)$$

with β regularization factor,
$\underline{\underline{I}}$ identity matrix.

To optimise this algorithm, the invertion is done frequency by frequency. The k in the formula denotes the kth frequency index.

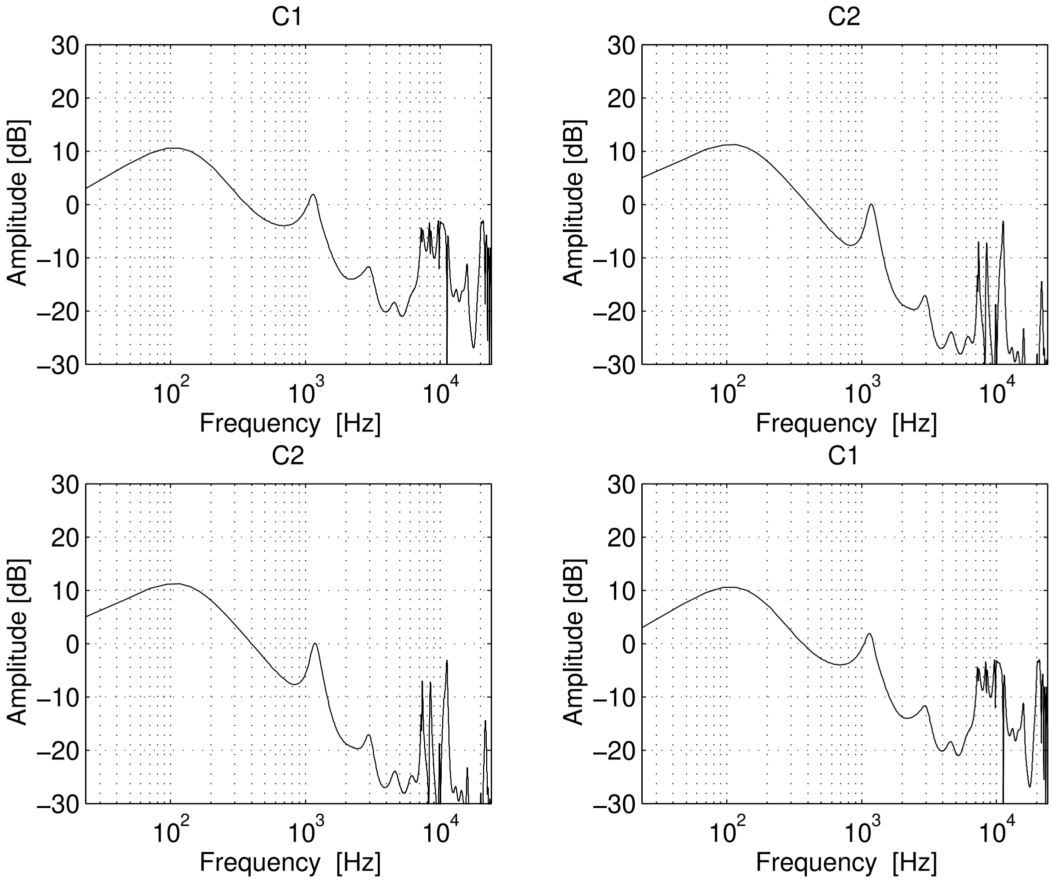

Figure 5.4: The C matrix which should be inverted to obtain the crosstalk filters.

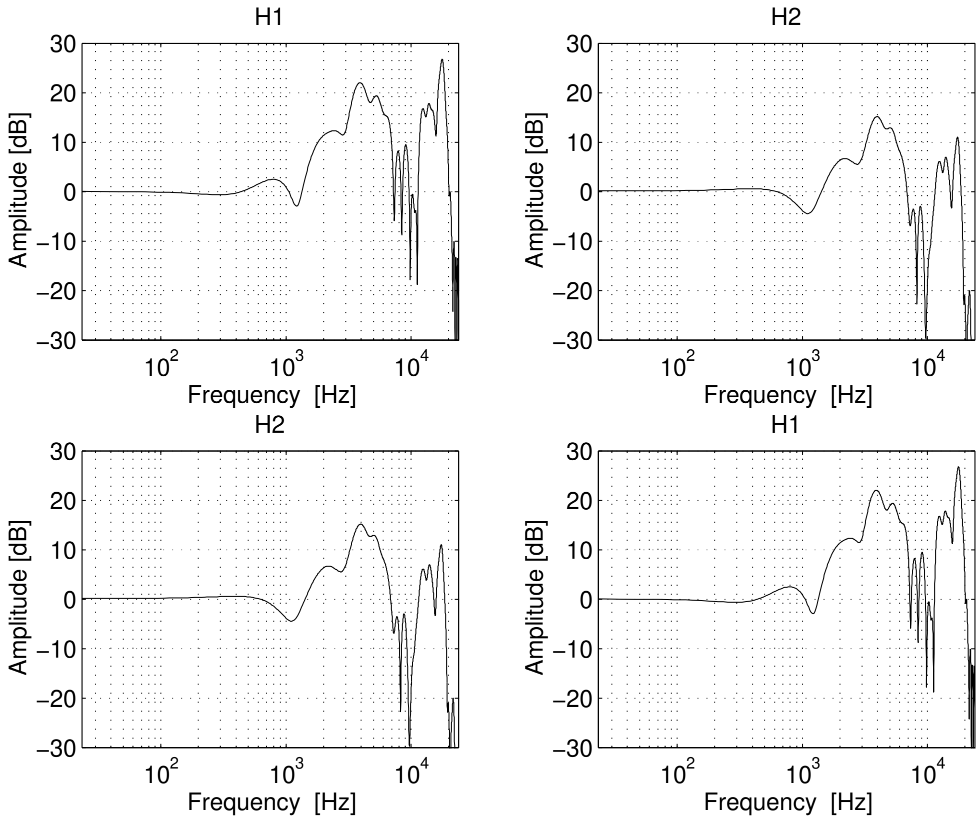

Figure 5.5: $\underline{\underline{H}}$ - the inverted system.

The regularization factor β guarantees the stability of the system. since it is not necessary to regulate equally over the whole frequency range it is chosen to make β frequency dependent. This frequency dependent β is shown in figure 5.6. Furthermore, the causality is assured in the time domain by doing a FFT shift.

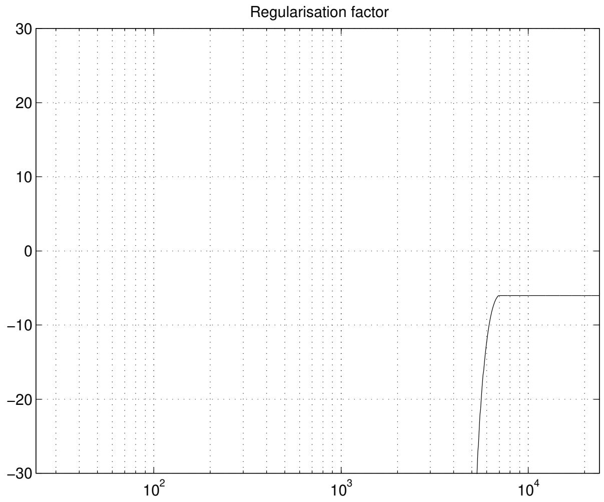

Figure 5.6: The frequency dependent regularisation factor β.

5.1.6 Multiply with the target matrix Z

The target matrix $\underline{\underline{Z}}$ displays the perfect crosstalk cancellation. This is the same matrix as described in section 3.2.1.

$$Z = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}$$

When this $\underline{\underline{Z}}$ matrix is multiplied with $\underline{\underline{H}}$, the resulting matrix is $\underline{\underline{Y}} = \underline{\underline{H}} \times \underline{\underline{Z}}$. This is a 2 by 2 matrix. This matrix is the frequency domain representation of the filters, which should be implemented in the crosstalk cancellation system. The $\underline{\underline{Y}}$ for the 2 by 2 example is seen in figure 5.7.

In order to save computation time the three last operations (5.1.4, 5.1.5 and 5.1.6) are performed only on the first half of the spectra. The first half is then mirrored to the second half to have the complete length.

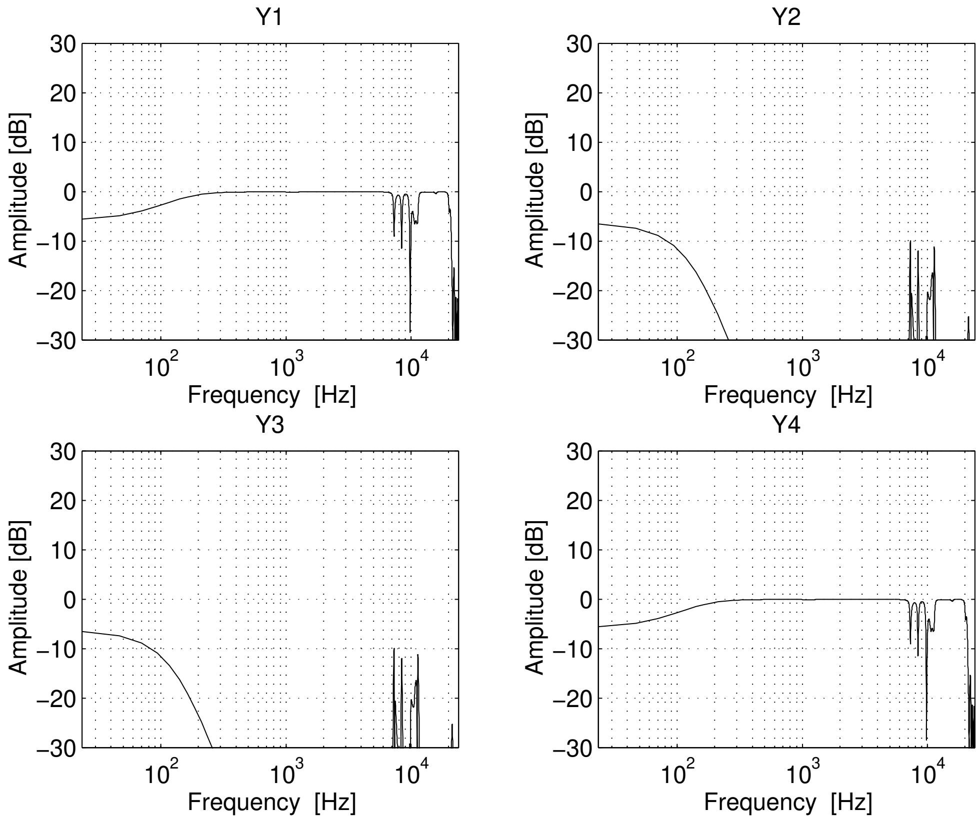

Figure 5.7: The $\underline{\underline{Y}}$ matrix in the frequency domain representation of the crosstalk cancellation filters.

5.1.7 Returning to the time domain using IFFT

The impulse responses are then converted to the time domain trought a IFFT. In order ensure causality an FFT-shift is carried out in time domain. In figure 5.8 a time representation of the crosstalk cancellation filters is seen.

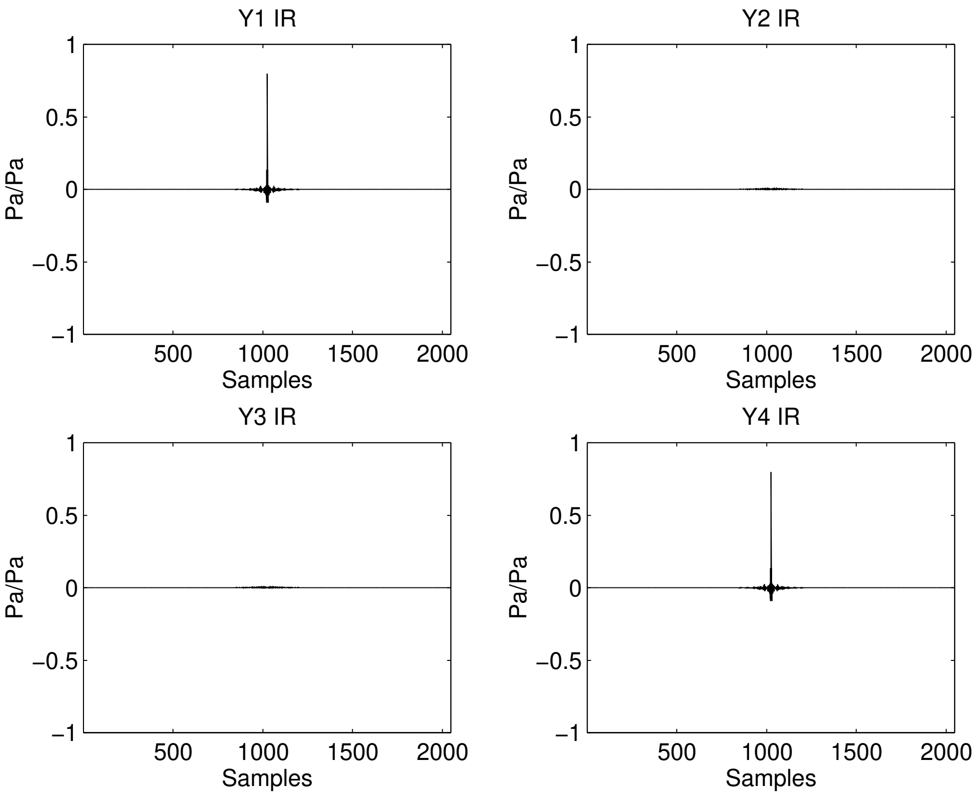

Figure 5.8: The crosstalk cancellation filters in the time domain.

5.1.8 Bandpass filtering the responses

The impulse responses are filtered to match the system requirements which was to crosstalk the signal in the middle and high frequency. This requirement is justified regarding the characteristic of the used loudspeakers which have a flat response between 150Hz and 19kHz.

5.1.9 Saving crosstalk filters

In order to use the crosstalk filters with a real setup the filter coefficients are saved in a mat format.

5.2 Performing the crosstalk cancellation

The actual crosstalk cancellation is performed with the block diagram in figure **??**. The crosstalk filters are loaded from the files described above. A simulation of how well the crosstalk cancellation works, can be made by multiplying the $\underline{\underline{C}}$ and the $\underline{\underline{H}}$ matrices. In this case, where the example is a 2 by 2 system, the result of the crosstalk cancellation is the same as the $\underline{\underline{Y}}$ matrix. Therefore the result of the crosstalk cancellation is seen in figure 5.8 and figure 5.7.

Chapter 6

Comparing the 2 by N and the 2 by 2 system

In order to find out whether the computed 2 by N crosstalk cancellation systems perform better than the conventional 2 by 2 system, objective and subjective tests were carried out. The objective of all the tests were to test how robust the different systems are to head movements.
This can be tested by placing the test object/subject in one position and allow continuous head movements. However it is difficult to control that the test object/subjects always move in the same way and in the same area. Therefore it was chosen to perform the tests by moving the test object/subject and then keep the head in a fixed position. Rotating movements are not considered in the system and was therefore not tested.

In this chapter it is described how these objective and subjective tests were carried out. At first, objective tests were performed to compare how well crosstalk is cancelled for the conventional 2 by 2 system and the three 2 by N systems described in section 4.7.
The results of these objective tests led to a choice of a system and setup, which then was tested for its ability to reproduce binaural sound in a subjective localisation test.

Common for both tests were that they took place in the VR-room at the Department of Acoustics at Aalborg University. Also the stimuli for the tests were the same, namely pink noise. Pink noise was chosen because it has constant energy in a broad frequency range in contrast to other signals like speech or music.

6.1 Objective Test

The purpose of the objective test was to measure how well the crosstalk cancellation works. This was done by presenting a 2-channel signal to the crosstalk canceller with pink noise in the left channel and no signal in the right channel. If the signal is perceived at the left ear and no signal at the right ear, the system works perfectly at the position of the listener. Due to symmetry between the loudspeakers and the listening position it was not necessary to test the system with signal in the right channel and no signal in the left channel.

The measurements were carried out with the artificial head Valdemar, so the signals at the ears of the listener, could be recorded for later analysis.

The three different 2 by N systems and the conventional 2 by 2 system have been tested for two loudspeaker elevation angles, which means that eight different systems were tested:

- 2 by 2, 0° elevation
- 2 by 10-5cm, 0° elevation
- 2 by 18-5cm, 0° elevation
- 2 by 18-10cm, 0° elevation
- 2 by 2, 90° elevation
- 2 by 10-5cm, 90° elevation
- 2 by 18-5cm, 90° elevation
- 2 by 18-10cm, 90° elevation

To test how the crosstalk cancellation performs for small head movements, recordings of the signal in Valdemar's ears were made in different positions around the centre position. The different positions can be seen in figure 6.1.

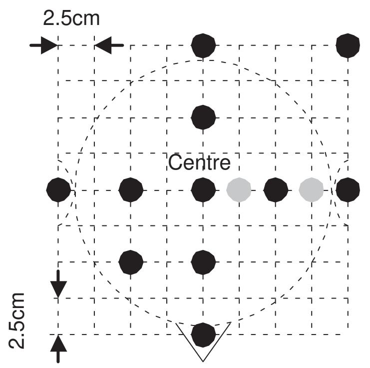

Figure 6.1: The positions where recordings are made with Valdemar for the objective test.

Most of the positions in figure 6.1 are included as HRTFs in the calculation of the crosstalk filters for the 2 by N system and therefore the performance at these points are predictible. There are two points however, which lie between the measured HRTFs. They are pointed out with a grey color in figure 6.1 and are included because it is not known what happens between the measured HRTFs.

More details about the measurement setup can be seen in A.

6.1.1 Expected results

From the measurements it was expected that the conventional 2 by 2 system would perform better than the other systems in the centre position because it should be able to make perfect crosstalk cancellation in this point.
The 2 by 18-5cm should perform well in the area of $10 \times 10cm$ around the centre which it is calculated for.
The 2 by 18-10cm should perform well in the area of $20 \times 20cm$, but since larger deviations are expected in the HRTFs, the crosstalk cancellation should be less than the 2 by 18-5cm.
The 2 by 10 system is expected to perform better in the five calculated crosstalk cancelled points than the 2 by 18-5cm system, because less points are included. But less crosstalk cancellation was expected in the positions outside these points.
In the simulations the 90° elevation of the loudspeakers showed better results than 0°, so the measurements for this angle were expected to perform better.

6.1.2 Results

As mentioned the two signals at Valdemar's ears were recorded. The recordings were imported into Matlab, where they were transformed to the frequency domain so the two sound pressure levels could be evaluated in dependency of the frequency.
The difference in sound pressure level at the two ears represents the quality of the system, the larger the difference between the sound pressure at the two ears, the better the system performs.
The level differences of the recordings were evaluated in two ways:

1. Observing the difference in sound pressure level frequency by frequency, to evaluate frequency dependency of the crosstalk cancellation.

2. Because there are so many systems and positions to compare, one RMS level difference is calculated for each.

For the first evaluation method two examples are shown in figure 6.2. The spectral difference between the two sound pressure levels for the 2 by 2 and the 2 by 18-5cm system in the centre position are plotted.

Similar figures have been investigated for the remaining systems. In general all the spectra show that the crosstalk cancellation is very dependent on the frequency.
For the centre position of Valdemar and for the front and back positions the crosstalk cancellation is best from approximately 400Hz up to 8 - 10kHz depending on the system.
In general the level difference for the 2 by 18-10cm system is lower compared to the other systems.
For the points 5cm to the sides of Valdemar and 45° in front, the results are similar to the points on the centre line, except the crosstalk cancellation is not good from around 2 - 5kHz.
The measured points outside the 5cm square look acceptable in the intervals 400 - 1kHz and 5 -

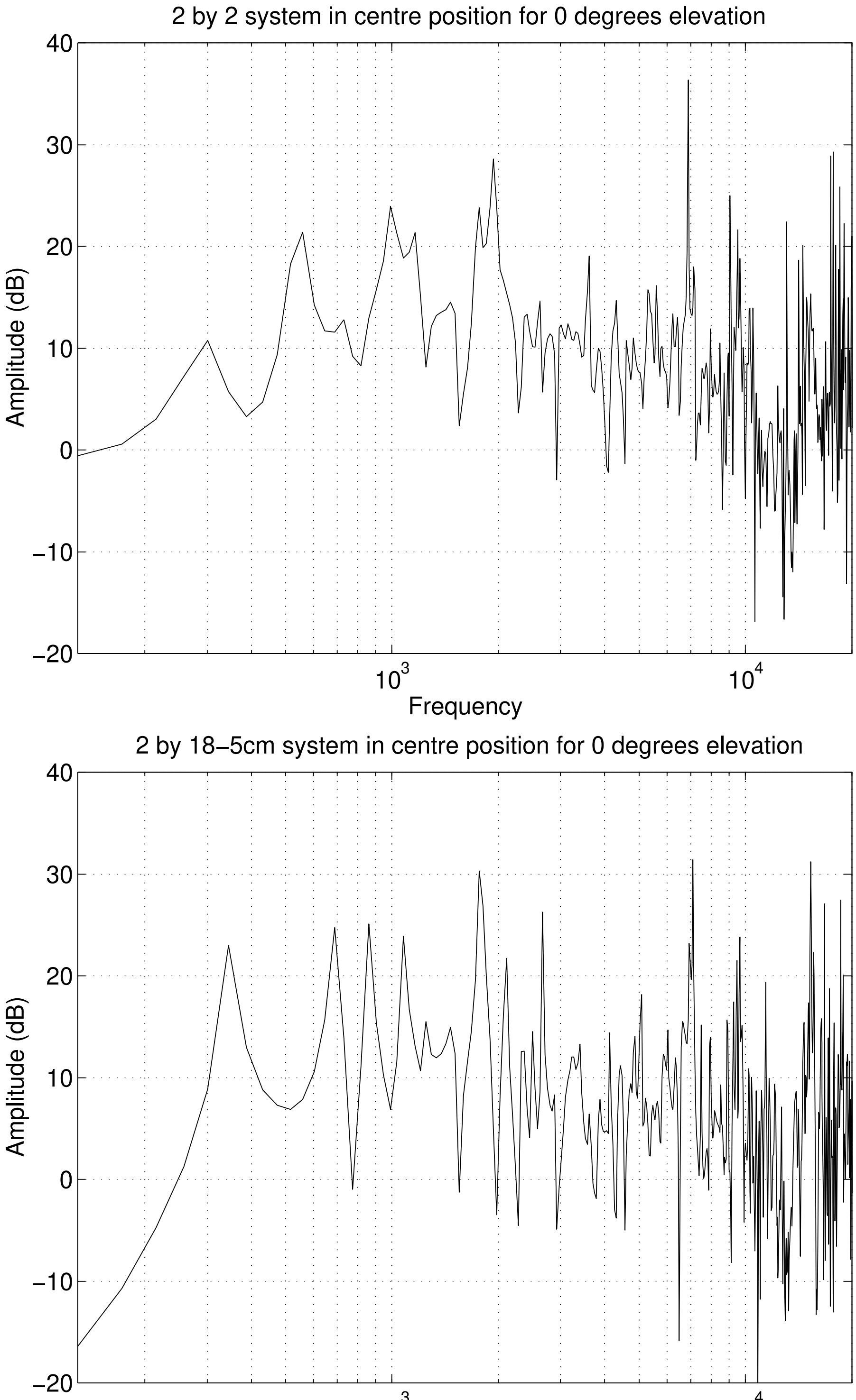

Figure 6.2: Spectral level difference for the 2 by 2 and 2 by 18-5cm system, recorded in centre position

8kHz. In general the 2 by 18-10cm system shows the best crosstalk cancellation at these positions, which were expected.

Another way of evaluating the large amount of measurements is to group the results for each system. The RMS value of the level difference from the two ears is calculated and then plotted in one figure, which reflects the recording position. Figure 6.3 and figure 6.4 show the result of this evaluation.
For all the systems it can be seen that the crosstalk cancellation is better for front-back listener displacements than for sideways displacements. This is because there are no interaural differences in this plane, since front-back displacements only cause a change in distance. The results for the measurements for 0° elevation are shown in figure 6.3.

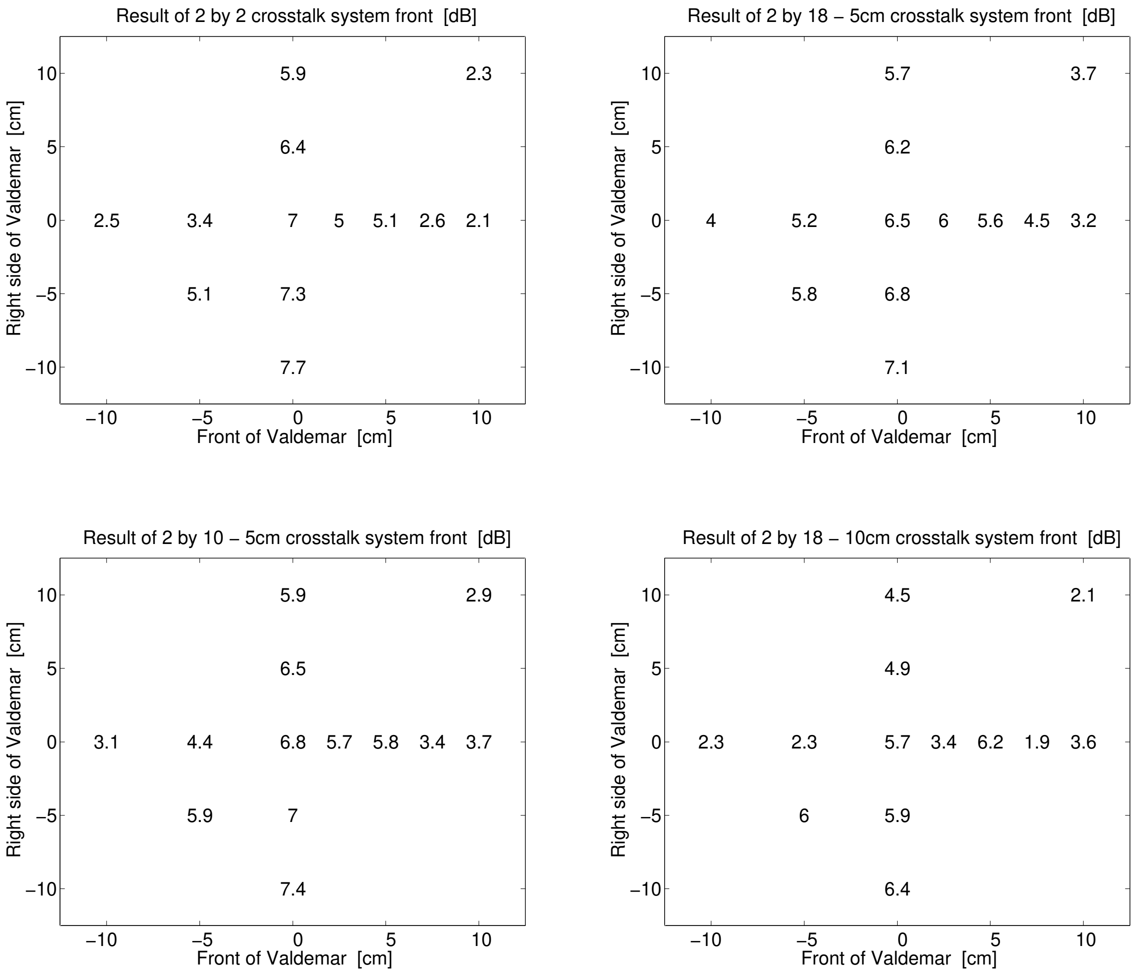

Figure 6.3: The values indicate the difference in sound pressure level between left and right ear. Results are shown for four different systems in all of the measured points. The results are for 0° elevation.

From figure 6.3 the values show that the level difference for each system in the centre position is almost similar. The lowest level difference is reached for the 2 by 18-10cm system, which was expected.
When Valdemar is moved from the centre closer to the loudspeakers, the level difference increases. It decreases when Valdemar is moved further away from the loudspeakers. The reason for this could be, if Valdemar was placed closer to the loudspeakers during the HRTF measurements.
As expected the level difference deteriorates for all systems for the points further away from the centre. However, the 2 by 18-10cm system the level difference of the points to the left varies a lot. For 45° off centre to the front, the results are quite similar compared to the centre, but as expected the 2 by 2 system has the smallest level difference. The value for the point 45° to the back is significantly lower compared to the centre, which was also expected.
In the two points 2.5cm and 7.5cm off centre to the left, which are not included in the calculation on the 2 ny N system, the crosstalk cancellation does not seem to deteriorate.

The measurements from 90° elevation of the loudspeakers is seen in figure 6.4.

One significant observation for the 90° elevation setup is that the level difference is lower for all the observations compared to the 0° elevation. Especially the centre positions all have low values. Generally the results do not make sence for this elevation angle. For the positions in front of and behind Valdemar the level differences increase in both directions compared to the centre, but they remain quite similar. Two points in particular show good results, the point 45° to the front and the point 5cm to the left. The rest of the positions to the left and to the right show very poor results. The project group has unfortunately not found a reasonable explanation for these results.

A comparison between the crosstalk cancellation systems for 0° elevation and 90° elevation shows, that the level differences are higher for the 0° measurements. This indicates that the crosstalk cancellation is better for the 0° loudspeaker setup.

6.2 Subjective tests

Since the idea of binaural recordings is a good reproduction of the spatial impression, subjective localisation tests have to be carried out to verify if the implemented system fulfill these requirements.
For a sound localisation test a binaural recording with sound sources at different positions are played back through the crosstalk cancellation system. During the playback the test subject has to define a virtual sound source for each presented signal. If the subjectively defined sound source positions match the positions of the loudspeakers during the recording, the spatial information of the signal has been reproduced. The crosstalk system can be compared to the setup, used for the recording of the binaural signals, also called the *real life system*.
Before the test, binaural recordings with different sound source positions had to take place. And therefore it was also needed to define positions of these sources.

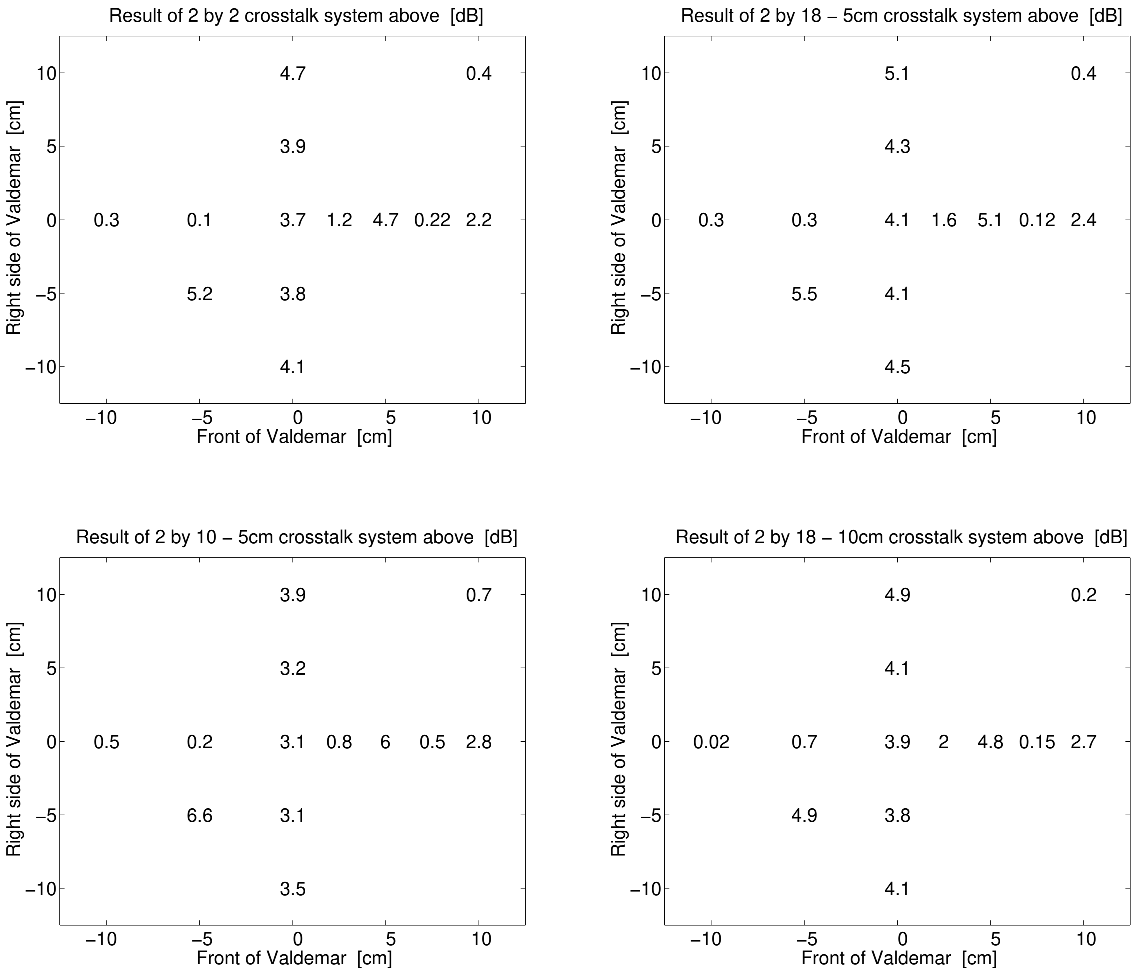

Figure 6.4: The values indicate the difference in sound pressure level between left and right ear. Results are shown for four different systems in all of the measured points. The results are for 90° elevation.

6.2.1 Loudspeaker setup and binaural recordings

A setup of 17 loudspeakers, as shown in figure 6.5, placed in a sphere with a 1.5m radius around the centre of Valdemar's head, was used. This setup is similar to the setup used in [7]. Eight

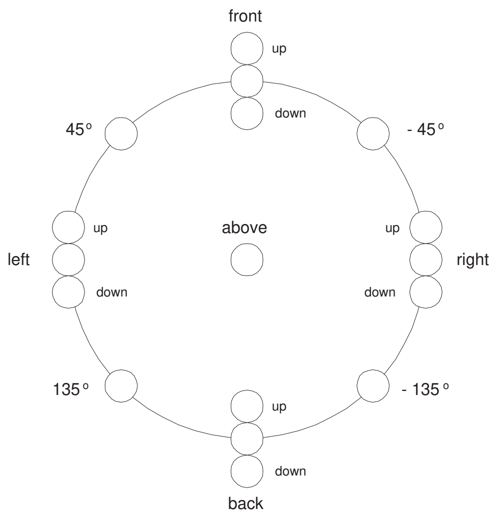

Figure 6.5: Loudspeaker positions used for binaural recordings and the setup used for the subjective localisation test.

loudspeakers are placed in the horizontal plane of the listener's ears. They are defined by their direction front, back, left, right or by their azimuth angle. For the positions in front, in the back and to both sides of the listener two more loudspeakers are placed with an angle of 45˚elevation above and below the middle loudspeaker (displayed by up and down). All loudspeakers in this chapter are called with their azimuth angle first and with up and down when they are positioned outside the azimuth plane. The last loudspeaker is placed above the listener's head and is called above.

An advantage of the setup is that it is easy for the test subject to remember. Futhermore several sound sources lie on the cones of confusion. There is a left cone of confusion consisting of four loudspeakers: The 45˚ and the 135˚ loudspeakers in the horisontal plane and the left up and down loudspeakers. The right cone of confusion consists of the corresponding loudspeakers on the right side. Seven loudspeakers lie on the median plane as described in 2 and shown in figure 6.6. These are the three loudspeakers in front, the one above and the three loudspeakers behind.

The binaural recordings for these 17 loudspeaker positions were made with the artificial head Valdemar. They were carried out in the same room where the subjective test was performed, the VR-room at the Department of Acoustics, Aalborg University. In order to avoid clicks in the beginning and at the end of the signal, which could give information about the sound source

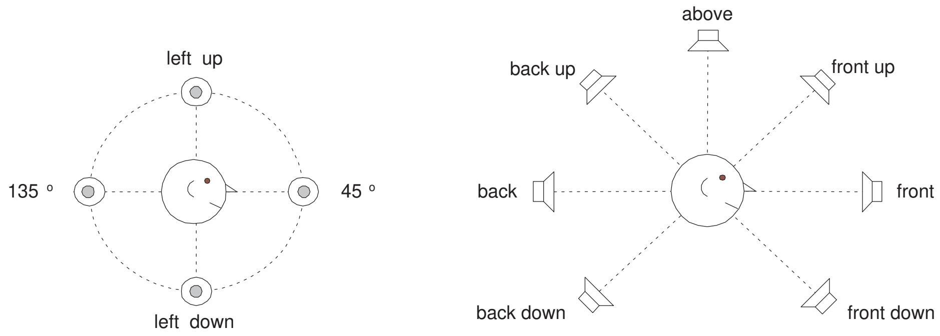

Figure 6.6: Placement of the loudspeakers lying on: a) The left cone of confusion, b) The median plane.

position, a linear fade-in and fade-out of 1ms was added to the signal.

6.2.2 Purpose of the subjective test

According to the result of the objective test the 0° loudspeaker elevation turned out to give the best results. Therefore only this loudspeaker setup was tested. Furthermore, it was decided to compare the 2 by 2 system to one of the 2 by N systems. In general the 2 by 18-5cm and 2 by 10-5cm performed the best of the 2 by N systems. The 2 by 18-5cm was used, and it was the purpose to test how the system calculated for the most points performs against a 2 by 2 system. These systems were both tested in the centre position of the loudspeaker setup and in two positions to the side of the centre to test the systems according to lateral head movements.

Also the real life system was included in the test, both as a reference, but also to check if the test subjects would make significant errors. In the real life system the signals are played back through single loudspeakers placed in the positions shown in figure 6.5. The real life system was only tested with the test subject in centre of the setup, because it was not expected the results to change much, even if the test subject moved out of the centre position. The real life system was tested as a reference because this system would perform better, since the crosstalk cancellation systems are only an approximation of the real life system. If some mistakes occured during the sound localisation test for the real life system like front-back confusion and side confusion, these errors should also be present in the reproduction systems.

For the crosstalk cancellation systems the binaurally recorded signal was reproduced by the 2 by 2 and 2 by 18-5cm crosstalk cancellation systems. This recording was played back through two loudspeakers, placed independently of the real life setup.

According to the purpose of the test the results were recorded for different head positions. The three systems were compared to each other to find how large the difference between the real life situation and the binaural reproduction would be, and if the 2 by 2 system performed better than the 2 by 18-5cm system. The conventional system was expected to perform better in the centre position because the 2 by 18-5cm system is also calculated for other points around the centre. Outside the centre the two crosstalk cancellation systems were tested in order to compare each system according to robustness of head movements. Since the purpose was to realise a system, which is robust to head movements, the 2 by 18-5cm system should give better results than the conventional system for positions out of the center position.

6.2.3 Method

For the subjective test a setup of 19 loudspeakers was needed: 17 ball-loudspeakers formed the loudspeaker setup and two loudspeakers were used for the playback of the crosstalk cancelled signals. The 17 ball-loudspeakers were also used by the test subjects to point out where the sound was coming from in the crosstalk cancellation systems. The real life system was controlled from a switching box from which the same signal was directed to one loudspeaker at a time. For the crosstalk cancellation systems the binaurally recorded signals were processed in Matlab and played back through the two separate loudspeakers.

The test consisted of four sessions: One familiarization session and three other sessions, each for a different listening position. Three positions of the test subject were tested, the centre position and two positions to the left of the centre (seen from the test subject), with a distance of 5cm and 10cm.
One session is split into two or three sequences, each sequence relating to one system. The composition of each session is visualised in table 6.1.

	Familiarization	Session 1	Session 2	Session 3
Position	Centre	Centre	5cm left	10cm left
System	Real life 2 by 18	Real life 2 by 2 2 by 18	 2 by 2 2 by 18	 2 by 2 2 by 18

Table 6.1: Overview of the different sessions of the subjective localisation test.

Each source position was presented twice for each system, giving a total of 34 sounds per system. The order of the presented signals were randomly chosen. The order of the systems was also randomly chosen to avoid bias from the fact that each test subject has a learning phase, followed by a constant performance and a decreasing performance in the end, because the test subject will get tired.

6.2.4 Procedure of subjective localisation test

Ten subjects, 8 male and 2 female, all between 20 and 30 years of age, performed the subjective test. The test was 45 minutes long with breaks.

An instruction paper, see the enclosures, was given to each test subject before the participation in the test. The setup and the purpose of the test were explained to the subject. The subject was then seated on a chair placed in the centre position of the setup. The seat was adjusted in order to place the ears in the centre position of the setup (the ears should be in the same height as the azimuth loudspeakers). The test subject was familiarized with the loudspeaker setup: The positions and the names of the different loudspeakers were explained. Because not all loudspeakers could be seen from the chair position a paper displaying the loudspeakers, see figure 6.5, was placed in front of the test subject. The loudspeakers were distinguished by numbers (1-8) and vertical direction if necessary (up, down and above). Before the familiarization, the order and duration of the different sessions was explained to the test subject.

During the test, a loudspeaker had to be defined among the 17 loudspeakers of the setup for each played sound. This means that the test is a 17-alternative forced choice test. Each sound had a duration of 3.5sec followed by 3.5sec of silence. During the time interval between each sound the test subject had to answer the name of the located sound source to one of the group members, who was sitting behind the listener, writing down the answers for later analysis. The test subject was asked not to move the head during the listening session.

After the explanation a test was performed to familiarize the person with the test procedure. The familiarization introduced the real life signal and the crosstalk cancelled signal to the test subject. The person had to answer, but these answers were not taken into account in the evaluation of the results. The purpose of the familiarization was to give the test subject an idea of what could be expected during the test and for the test subject to get familiarized with the rhythm of sound representation and answering time. After the familiarization the test subject participated in the three different sessions corresponding to the three chair positions. Between each session, the subject had some minutes to rest. In the same time, the chair was moved to another position.

6.2.5 Evaluation

For the evaluation of the subjective tests the response of sound source positions for all 10 test subjects are plotted related to the position of the presented stimulus. In the real life system the stimulus is the position of the loudspeaker from which the signal is represented. For the different crosstalk cancellation systems the stimulus is given by the loudspeaker position during the recording of the signal.
The loudspeaker positions on the axes of the figures are organised as it can be seen from figure 6.7. This figure also shows how the answers will be grouped in the analysis of them.

From figure 6.7 is is seen that correct answers lie on a diagonal from the lower left corner to the upper right corner. All error categories, explained in chapter 2, have certain sections in the graph. The loudspeaker positions on both axes are organized “cone wise”, so that all sources providing the same ITD are next to each other. In this way, all wrong answers resulting from cone of confusion errors build squares, lying around the diagonal of correct answers, in the graph.
All right cone of confusion errors are plotted in a square in the upper right corner, all left cone of confusion errors are plotted in a square in the lower left corner, and all median plane errors are

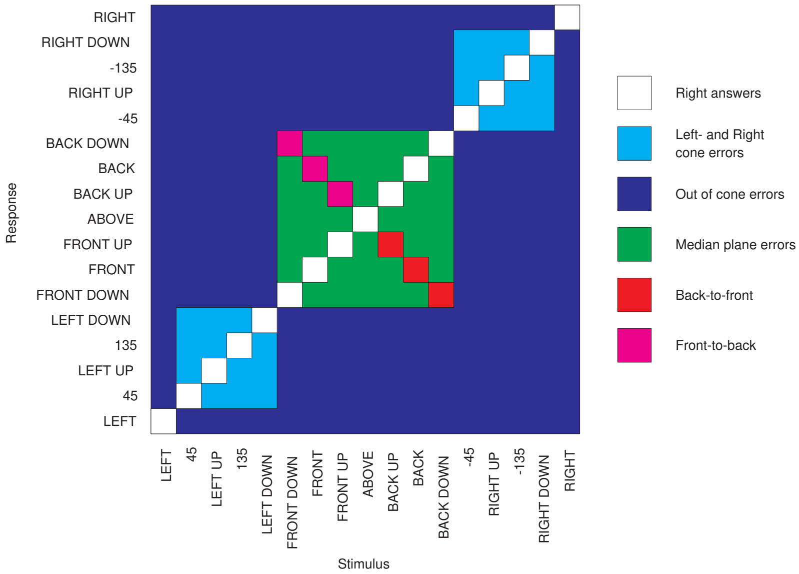

Figure 6.7: Error categories of the subjective localisation test.

plotted in the middle square. In the median plane errors the front-to-back and back-to-front errors are included. All errors lying outside this area are called out of cone errors.

6.2.6 Results

The results for each tested system and for each position of the test subjects are shown in the figures 6.8-6.14. The area of each circle plotted in the graphs is proportional to the number of answers it holds. Each graph contains 340 answers. 10 test subjects participated, and each of 17 stimuli was presented twice.

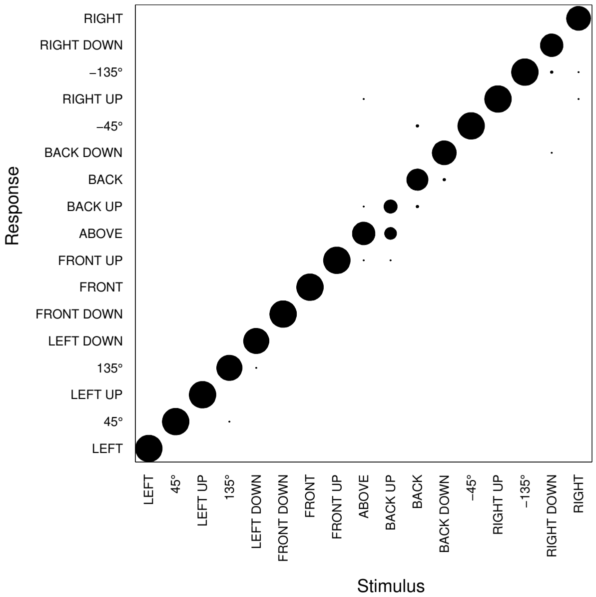

Figure 6.8: Results for the real life system with the test subject seated in the centre position.

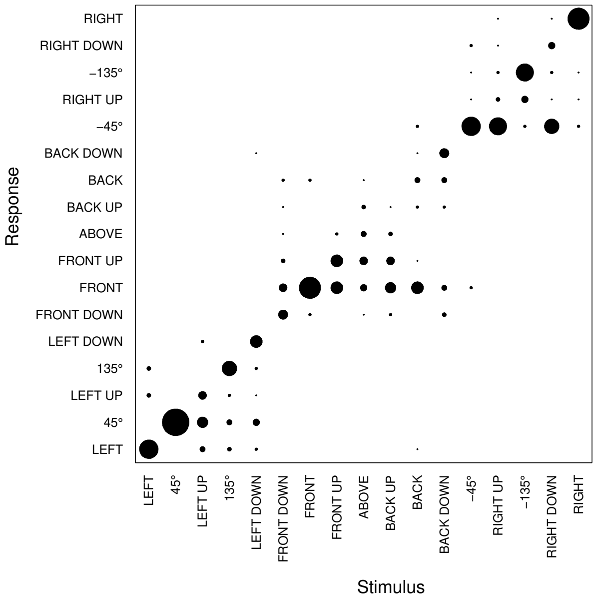

Figure 6.9: Results for the 2 by 2 system with the test subject seated in the centre position.

From figure 6.8-6.14 it can be seen that the real life system performs best as it was expected. Only

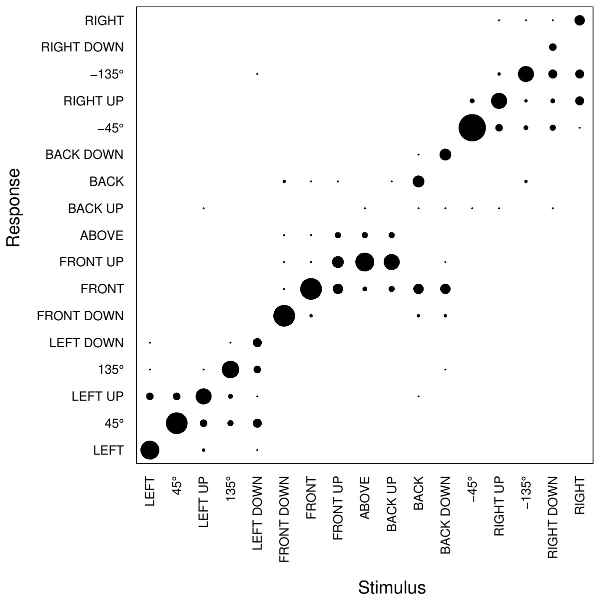

Figure 6.10: Results for the 2 by 18 system with the test subject seated in the centre position.

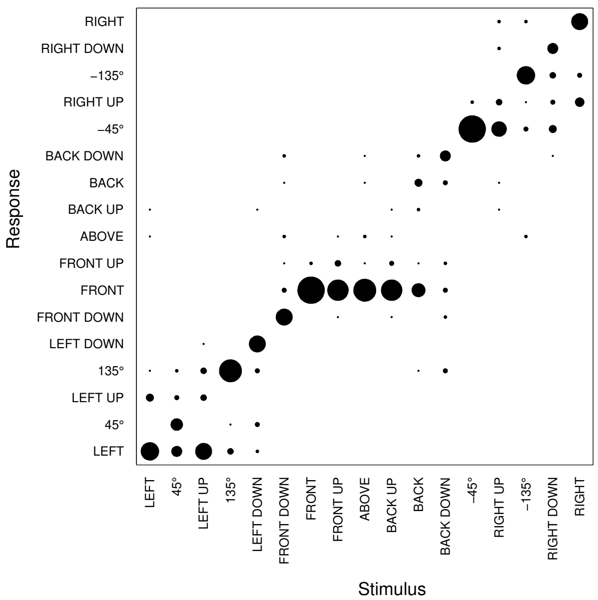

Figure 6.11: Results for the 2 by 2 system with the test subject seated 5cm left from the centre position.

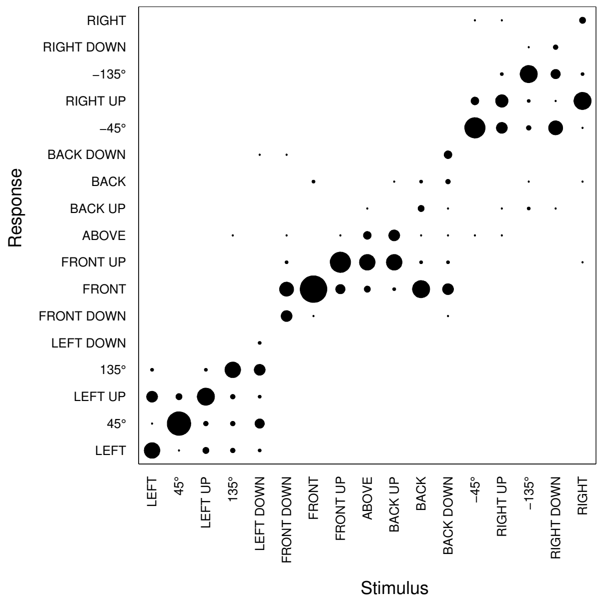

Figure 6.12: Results for the 2 by 18 system with the test subject seated 5cm left from the centre position.

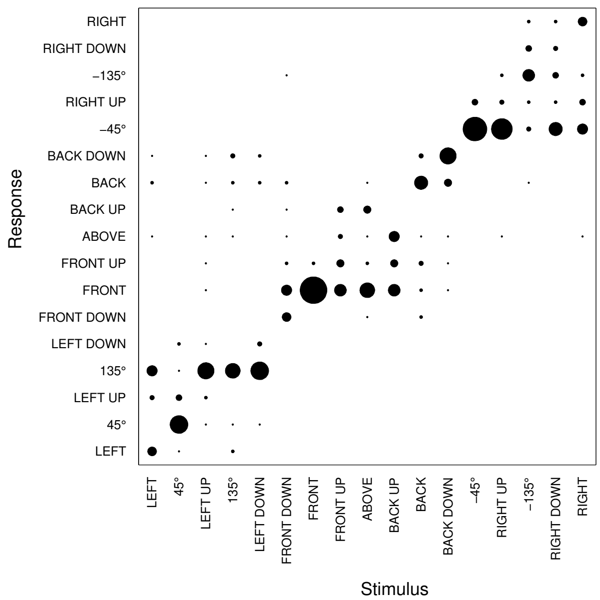

Figure 6.13: Results for the 2 by 2 system with the test subject seated 10cm left from the centre position.

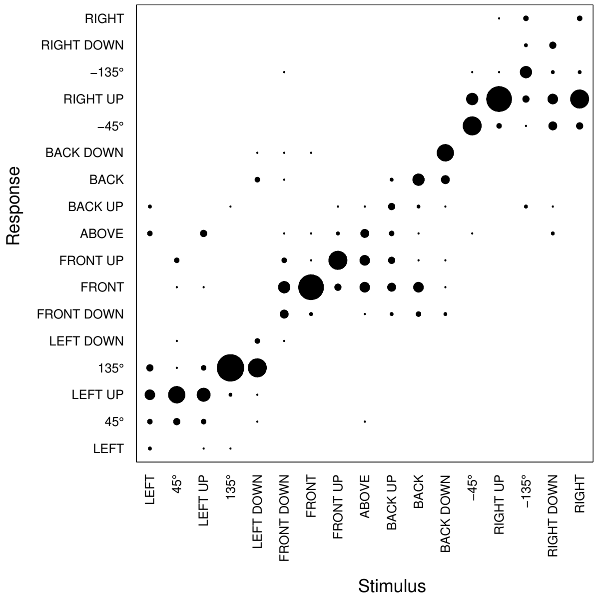

Figure 6.14: Results for the 2 by 18 system with the test subject seated 10cm left from the centre position.

a few errors were made, almost all of them are median plane errors.

In the centre position the different error categories (see figure 6.7) can clearly be seen for both crosstalk cancellation systems. When a long vertical line of circles is present in the figures, this means that one stimulus has been localised randomly. A horisontal line of circles means that several sound source positions are perceived to come from the same position.

When the test subjects move out of the centre position, the out of cone errors increase. Sound from both sides move to the median plane. With a large displacement of the head, the out of cone errors appear asymmetrical. When the listener moves to left, the stimuli from the left, are perceived in the median plane, while there are less out of cone errors for stimuli presented from the right.

Signicant from the graphs is that the front loudspeaker has been localised most often while the back loudspeakers are rarely localised. The stimuli from the back are mainly localised in the front. This means that the placement of the crosstalk cancellation loudspeakers in the front, could be the cause of these errors.

From the graphs it is not possible to see significant differences between the two crosstalk cancellation systems. To see if more significant differences could be found, the number of correct and wrong answers for the different error categories were counted. The percentages of this count are listed in table 6.2.

Errors in %	Real life Centre	2 by 2 Centre	2 by 18 Centre	2 by 2 5cm	2 by 18 5cm	2 by 2 10cm	2 by 18 10cm
Median plane	4.4	26.2	23.8	25.9	26.8	25.9	22.9
Right cone	0.6	12.1	8.2	8.8	10.6	12.9	9.4
Left cone	0.6	7.1	9.1	5.3	8.8	10.0	9.7
Back-to-front	0.3	5.3	5.9	3.8	6.5	2.1	3.5
Front-to-back	0	0.6	0.3	0.9	0.9	1.2	0.6
Out of cone	1.5	7.9	9.4	16.5	13.8	15.9	18.8
Correct answers	92.4	46.8	49.4	43.2	40.0	35.0	38.2

Table 6.2: Categorised errors and correct answers in percent.

From the results of table 6.2 some more conclusions could be made.
The numbers of correct answers, median plane errors and out of cone errors are approximately the same for both crosstalk cancellation systems in each test subject position. This was not what was expected. It was expected the 2 by 2 system to perform best in the centre position and the 2 by 18 system to perform best in the positions out of the centre. But this tendency could not be seen in the number of correct answers in the tests because they are almost equal for both systems in each position. When the test subject moves out of the centre the number of correct answers decreases, and the number of out of cone errors increases for both systems. A difference between the two systems is that the number of right and left cone errors is approximately the same for the 2 by 18 system while the number of right cone errors is always higher than the number of left cone errors for the 2 by 2 system. This phenomenon could not be logically explained.
The median plane errors do not change with a change of test subject position. They are almost equal for all tests except the real life test.
Significant from the results is that much more back-to-front than front-to-back errors are made. This was also expected because the two crosstalk cancellation system loudspeakers were positioned in front of the listener, and in this way the sound moved to the front.

Chapter 7

Conclusion

The aim of this project work was to reproduce binaural sound via loudspeakers in a way that small head movements of the listener can be allowed, without losing the spatial impression. In a conventional crosstalk cancellation system, a 2 by 2 system, binaural sound is only reproduced correctly in one listening position. However, as soon as the listener moves out of the sweet spot, the spatial impression deteriorates.

In order to develop a crosstalk cancellation system robust to head movements, a different approach has been investigated and realised. This approach is a further development from the algorithm proposed by [2]. This algorithm has been used to develop a 2 by N system. 2 by N means that two loudspeakers are used to improve the spatial impression for N microphone points (referring to the two ears so representing one head position). This system includes N points, lying in a square of $(20 \cdot 20cm)$ around the sweet spot, in the calculation of the crosstalk cancellation filters. Even though crosstalk is cancelled for N microphone points, only two loudspeakers are used in the binaural sound reproduction.
One 2 by 2 system and three 2 by N systems were realised with singular specifications. A 2 by 10-5cm system represents two loudspeakers crosstalk cancelling for five head positions, four position spaced 5cm from the sweet spot and one sweet spot position. A 2 by 18-5cm system represents two loudspeakers crosstalk cancelling for nine head positions, eight positions evenly spaced in a square$(10 \cdot 10cm)$ around the sweet spot and the sweet spot position. Finally a 2 by 18-10cm system represents two loudspeakers crosstalk cancelling for nine head positions, eight positions evenly spaced in a square$(20 \cdot 20cm)$ around the sweet spot and the sweet spot.
Since lower frequencies up to 150Hz do not contain any spatial information [4] and because of the limitation of the loudspeakers, it is decided to filter the signal in a frequency range of 150Hz to 19kHz.
The setup consisted of two ball loudspeakers, placed with a distance of 1.5m and an azimuth angle of $\pm$ 15° to the listener. Furthermore, two elevation angles of 0° and 90° were chosen from measurements of different transfer functions, so two loudspeaker setups were made. In total, the four systems are implemented for each elevation angle.

These systems were compared to each other in an objective test. For this test the artificial head Valdemar was placed in the sweet spot and in several positions around it. The test signals at both ears were recorded. From these recordings the sound pressure level difference between the ears were evaluated. The larger the level difference between the two ears, the better the system cancels crosstalk.
The results of the objective test showed that the systems for the 0° elevation performed better crosstalk cancellation than the ones for the 90° elevation. Also the results from the objective tests were used to select which system should be used in a subjective test. From the results one of the 2 by N systems with 0° elevation angle was chosen. In the subjective test the 2 by 18-5cm system was compared to the 2 by 2 system and a real life system in order to test how good the systems reproduce binaural sound. The subjective sound localisation test was evaluated by counting correct and wrong answers for different sound source positions. The results of the subjective test showed no significant differences between the 2 by 2 and the 2 by 18-5cm system.

In general, we can confirm that the 2 by N algorithm can be used for crosstalk cancellation in a transaural system, but it does not show any significant improvement to a 2 by 2 system. However, the system has not been tested for a real life binaural sound recording e.g people walking around and speaking. So we can not give final conclusions of how well the system reproduces a real life binaural recording.

The system can be improved. By adding a subwoofer it is possible to cover the lower frequency range without crosstalk cancelling these frequencies. Also a multichannel system could be made by splitting up the mid- and high frequencies. Furthermore, more loudspeakers could be added to the setup to better control the crosstalk cancellation of the N points.

Further investigation could be focused on weighting the N points individually, instead of giving them equal importance to obtain a better spatial impression. Also, the algorithm used has been implemented in off-line processing. An evolution could be an implementation of the system in real time. If the crosstalk cancellation filters are preprocessed and stored, the deconvolution could be performed in real time as well.

Bibliography

[1] 00gr963. *Measurement of sound insulation for a window facade and background noise level in VR-room*. 2000.

[2] J. Bauck and D. H. Cooper. Generalized transaural stereo and applications. *J. Audio Eng. Soc.*, 44(9):683–705, 1996.

[3] J. Blauert. *Spatial Hearing : The psychophysics of human sound localization*. The MIT Press, 1997, Revised edition
ISBN: 0-262-02413-6.

[4] P. Damaske. Head-related two-channel stereophony with loudspeaker reproduction. *J. Acoust. Soc. Am.*, 50(4):1109–1115, 1971.

[5] R. O. Duda and W. L. Martens. Range dependence of the response of a spherical head model. *J. Acoust. Soc. Am.*, 104(5):3048–3058, 1998.

[6] B. et. al. *3-D sound presented through loudspeakers in a multimedia setup*. 1998.

[7] D. Hammershøj. *Binaural Technique - a method of true 3D sound reproduction*. 1995
ISBN: 87-7307-516-7.

[8] J. Jakeuchi and P. A. Nelson. *Optimal Source Distribution for Virtual Acoustic Imaging*. 2000, ISVR Technical Report no. 288.

[9] C. Kinsler, Frey and Sanders. *Fundamentals of Acoustics*. John Wiley & Sons Inc., 2000, 4. edition
ISBN: 0-471-84789-5.

[10] O. Kirkeby. Fast deconvolution of multichannel systems using regularization. *J. Acoust. Soc. Am.*, 6(2):189–194, 1998.

[11] P. Költzsch and G. Pfeifer. *Akustik*. 1998, 2. edition.

[12] H. Møller. Reproduction of artificial-head recordings through loudspeakers. *J. Audio Eng. Soc.*, 37(1/2):30–33, 1989.

[13] H. Møller. Fundamentals of Binaural Technology. *Applied Acoustics*, 36:171–218, 1992.

[14] B. C. Moore. *An Introduction to the Psychology of Hearing*. Academic Press, 1997, 4. edition
ISBN: 0-12-505627-3.

[15] J. Nilsson and U. Koch-Henriksen. *Role of Loudspeaker Angle in a Crosstalk Cancellation System*. 2000.

[16] I. of Sound and S. Vibration Research. *Stereo Dipole System*. Soundvision Ltd., Japan.

[17] P. A. N. Ole Kirkeby and H. Hamada. The "stereo dipole" - a virtual source imaging system using two closely spaced loudspeakers. *J. Audio Eng. Soc.*, 46(5):387–395, 1998.

Appendix A

Measurements

This appendix contains a description of the measurements and recordings made for this project. A list of the used equipment is found at the end of the appendix. The setup of the MLSSA-program is found in the enclosure.

A.1 Loudspeaker frequency responses

Frequency responses of four types of loudspeakers have been measured in order to find loudspeakers that could be used for the crosstalk cancellation system. The measurements have taken place under free field conditions in the anechoic chamber at the Department of Acoustics at Aalborg University. The frequency responses have been measured using the MLSSA system. The measured loudspeakers were two standard PC loudspeakers, two BOSS MA-12V, two Jolly FPT, all active loudspeakers and four passive ball-shaped loudspeakers with Vifa MD10 units.

A.2 Transfer function measurements for different elevation angles

The purpose of measuring the transfer functions was to obtain the sound pressures at each eardrum. The frequency responses were measured at the ears with an artificial head, "Valdemar". The output from the microphones in Valdemar's ears were recorded by the MLSSA system. This was done in the anechoic chamber. Because of the symmetry of the two loudspeakers and the head, only one loudspeaker was used for the measurements. It was placed in an azimuth angle of 15°, with a distance of 1.50m to the center of the head. Seven different elevation angles from 0° to 90° in steps of 15° were chosen. For practical reasons Valdemar was turned and the loudspeakers were fixed. The measurement setup can be seen in figure A.1.

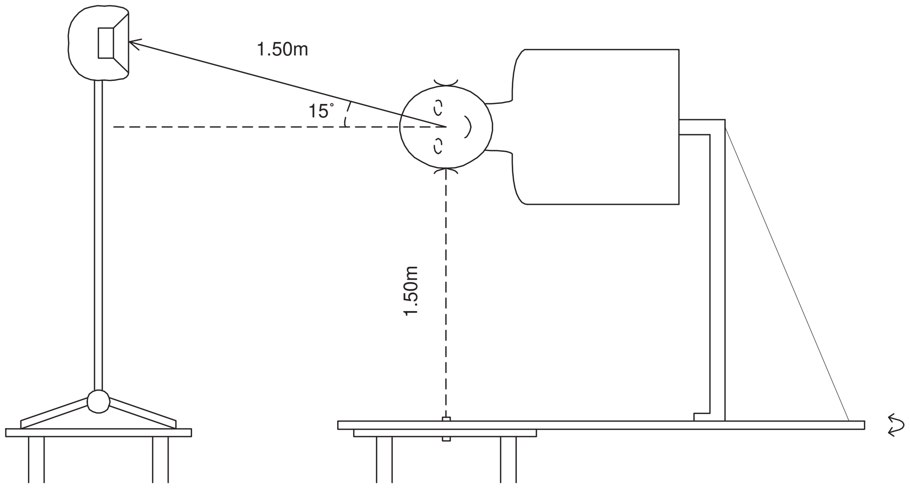

Figure A.1: The setup for measuring transfer functions in the vertical plane on Valdemar

The transfer functions for the different elevation angles can be seen in figure A.2, A.3 and A.4.

A.2.1 HRTFs for 41 Valdemar positions

HRTFs for the grid in 4.3 for both elevation loudspeaker angles of 0˚ and 90˚ have been measured in the anechoic chamber, using the MLSSA system. 41 different head positions were measured. For the centre position the sound pressure at the centre of the head with the head absent, has been measured, to obtain the free-field correction of the head. Because of the symmetry of the two loudspeakers and the head, only one loudspeaker was used for the measurements for each of the elevation angles.

A.3 Equipment list

For all these measurements different equipment was used and it is listed below.

- Microphone: B&*K* 4133 LBNR 06546
- Microphone: B&*K* 4166 LBNR 08601
- Preamplifier: B&*K* 2619 LBNR 06560
- Microphone amplifier: B&*K* 2636 LBNR 08022
- Measuring system: MLSSA PC LBNR 31269

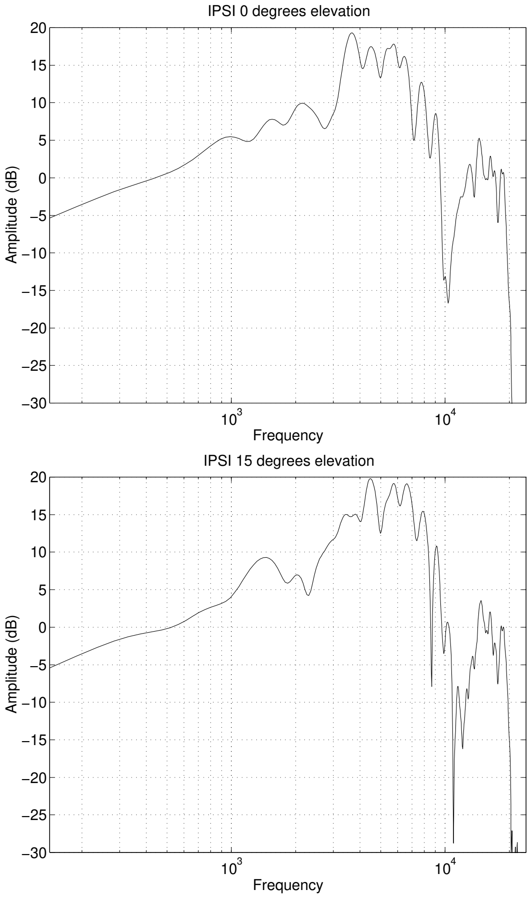

Figure A.2: The transfer functions for the vertical plane from 0° and 15°.

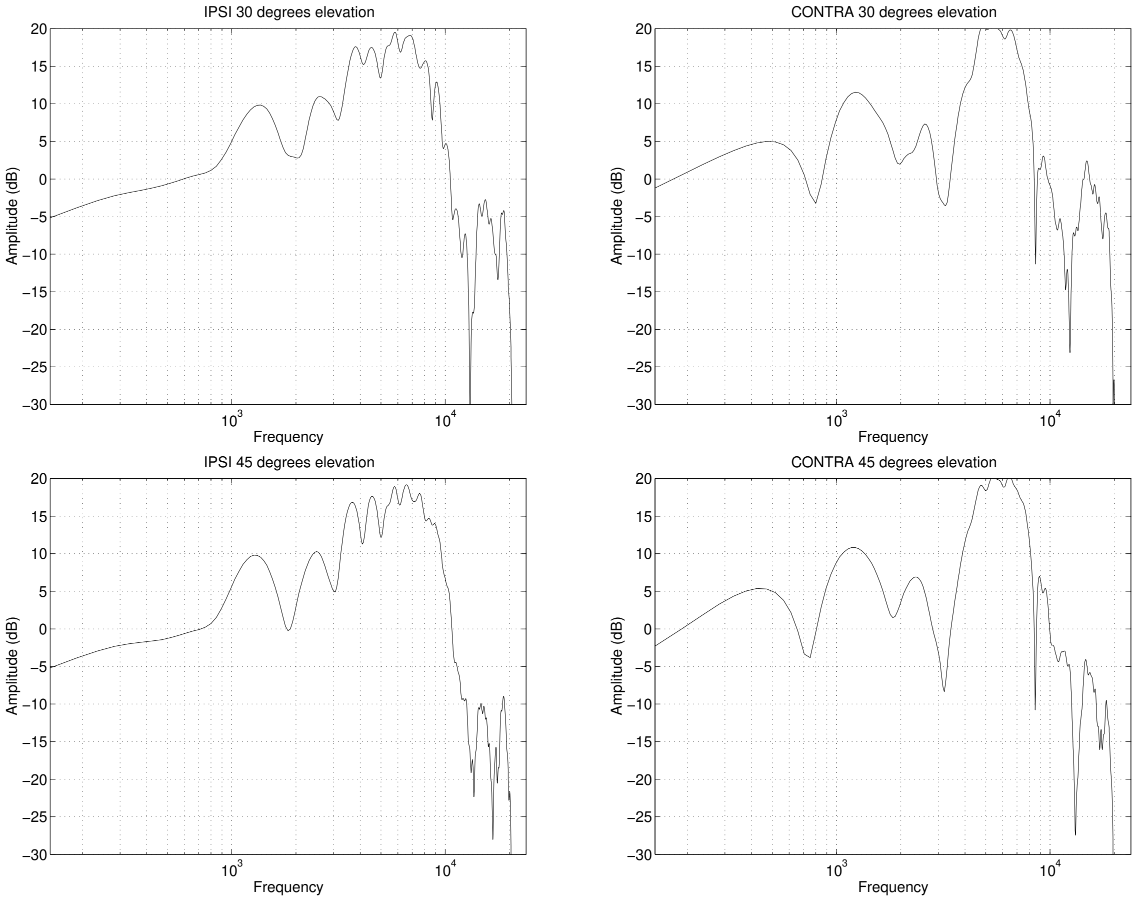

Figure A.3: The transfer functions for the vertical plane from 30° and 45°.

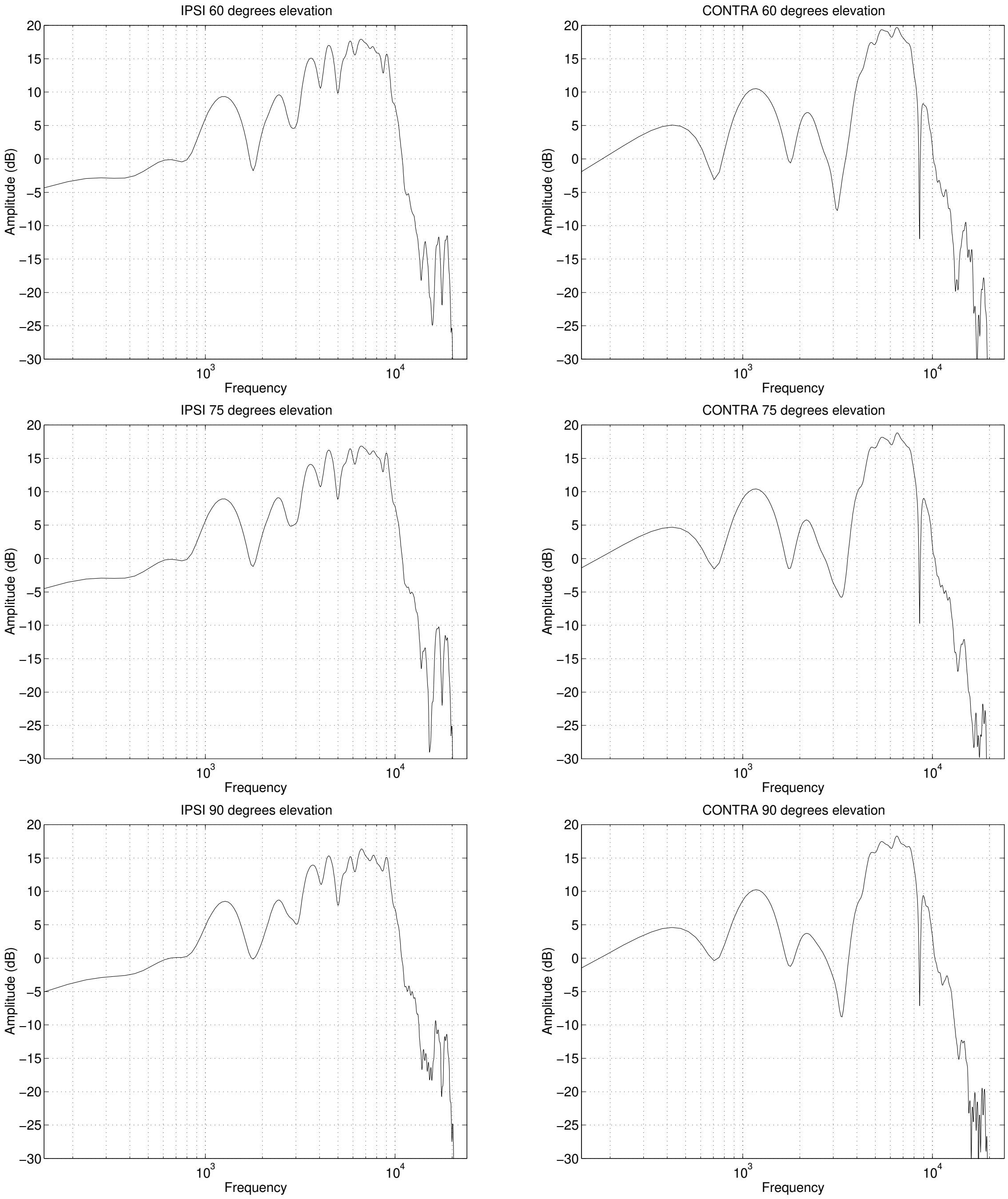

Figure A.4: The transfer functions for the vertical plane from 60°, 75° and 90°.

- Amplifiers: Pioneer SA330 LBNR 07925
- Loudspeakers: Vifa MD10 units mounted in a ball
- Noise generator: B&*K* 1049 LBNR 08718
- Microphone preamplifier: TC Gold channel LBNR 33748
- Artificial head: ‘Valdemar’

Enclosure B

Matlab

Matlab program for 2 by 18-5cm front system

```
clear all

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Abbrevations: BK = Brüel & Kjaer
%               GC = Gold Channel
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Importing data from files
%%%%%%%%%%%%%%%%%%%%%%%%%%%
loadt

%general variables
L=256; % length of the signal (samples)
c=340; % velocity of the sound (m/s)

%%%%%%%%%%%%%%%%%%%%%%%%%
% Calibration definitions
%%%%%%%%%%%%%%%%%%%%%%%%%

% B&K 4166
Sens_BK = 52E-03; % Sensibility p1
Gain_BK1 = 10^(40/20); % Gain p1

Delay_BK = floor(1.48/c*Fs); % Initial delay in IR from BK

% Dummy head microphones
```

```
Sens_DUM_left = 26.71E-03; %Left ear sensibility
Sens_DUM_right = 28.11E-03; % Right ear sensibility
Gain_GC = 10^(25/20); % Gain gold channel

%%%%%%%%%%%%%%%%%%
% making true HRTF
%%%%%%%%%%%%%%%%%%

% resize P1
P1old_0 = -P1old_0/(Gain_BK1*Sens_BK); % taking sensibility & gain into account
P1 = P1old_0(Delay_BK:Delay_BK+(L-1)); %truncation for 256 samples

for position = 1:9,
% Convert to pressure values
  LLir(:,position) = LLir(:,position)/(Gain_GC*Sens_DUM_left);
  LRir(:,position) = LRir(:,position)/(Gain_GC*Sens_DUM_right);

% take out the delay
  LLir_1(:,position) = LLir(delay(position,'i',Fs):delay(position,'i',Fs)+
L-1,position);
  LRir_1(:,position) = LRir(delay(position,'c',Fs):delay(position,'c',Fs)+
L-1,position);
% true hrtf
  LLir_true(:,position) = hrtf(LLir_1(:,position),P1,Fs); %deconvolution of
the loudspeakers
  LRir_true(:,position) = hrtf(LRir_1(:,position),P1,Fs); %deconvolution of
the loudspeakers

% delay back
  LLir_d= [zeros(delay(position,'i',Fs)-delay(37,'i',Fs),1);
LLir_true(1:64,position);zeros(L,1)];
  LLir_delayed(:,position)=LLir_d(1:L);
  LRir_d = [zeros(delay(position,'c',Fs)-delay(37,'i',Fs),1);
LRir_true(1:64,position);zeros(L,1)];
  LRir_delayed(:,position)=LRir_d(1:L);
end

Tsec = 1000*[1/Fs:1/Fs:L/Fs]; % time axis in msec
Tmet = c*Tsec;                                   % time axis in meters
F = [0:Fs/L:Fs/2]';                    % Frequency axis for fft plots!

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Definitions
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
n = 2;              % Number of loudspeakers
m = 18;              % Number of points in space
Nc = L;  % Length of IRs
Nh = 8*Nc;          % Length of FFTs

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Generating a frequency dependent beta (beta_f)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
beta_low = 0.005;      % Regularization factor
beta_high = 10;
f1 = 4000;
f2 = 6000;
L1 = ceil((f2-f1)/(Fs/Nh));
L2 = ceil(f1/(Fs/Nh));
L3 = Nh/2-L2-L1;

up_slope = beta_low+(beta_high-beta_low)*(0.5*(cos(pi:pi/(L1-1):2*pi)+1));

beta_f = zeros(Nh,1);
beta_f(1:Nh/2)=[ones(1,L2)*beta_low up_slope ones(1,L3)*beta_high];
beta_f(Nh/2+1:Nh)=flipud(beta_f(1:Nh/2));
figure(7)
semiplot(fft((real(ifft(beta_f)))),Fs,20,'b')

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Fourier tranformation of IRs
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
LLir1(:,1:m/2) = LLir_delayed(:,1:m/2);
LRir1(:,1:m/2) = LRir_delayed(:,1:m/2);

LLfr = fft(LLir1,Nh).';
LRfr = fft(LRir1,Nh).';


RRir1 = zeros(L,m/2);
RLir1 = zeros(L,m/2);
RRir1(:,1:2) = LLir1(:,1:2);
RLir1(:,1:2) = LRir1(:,1:2);
if m>4,
% For 9 points in space
  for k = 3:m/2,
    RLir1(:,k) = LRir_delayed(:,12-k);
    RRir1(:,k) = LLir_delayed(:,12-k);
  end
```

```
end
RRfr = fft(RRir1,Nh).';
RLfr = fft(RLir1,Nh).';

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Calculation of inverse "matrix"
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
I = diag(ones(1,min(m,n)));     % Identity matrix for regularisation
Z = [ones(m/2,1) zeros(m/2,1);zeros(m/2,1) ones(m/2,1)];   % Target matrix for
Cooper and Bauck solution

for k= 1:Nh/2+1,
  C = [LLfr(:,k) LRfr(:,k);RLfr(:,k) RRfr(:,k)];

  H = (C'*C+beta_f(k)*I)\C';    % Inverse matrix of C

  for j=1:m,
    for l=1:n,
      eval(['H' num2str(l) num2str(j) '(k)= H(l,j);']);
    end
  end

  Y = H*Z;                      % Cooper & Bauck solution

  for j=1:n,
    for l=1:n,
      eval(['Y'num2str(j)num2str(l)'(k)= Y('num2str(j)','num2str(l)');']);
    end
  end
end

% Mirroring of result
for j=1:m,
  for l=1:n,
    eval(['H' num2str(l) num2str(j) '(Nh/2+2:Nh) = fliplr(conj(H' num2str(l)
num2str(j) '(2:Nh/2)));']);
  end
end
for j=1:n,
  for l=1:n,
    eval(['Y' num2str(j) num2str(l) '(Nh/2+2:Nh) = fliplr(conj(Y' num2str(j)
num2str(l) '(2:Nh/2)));']);
  end
end
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Going back to time domain and save result
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
[B,A] = butter(4,[150/24000 19000/24000]);

for j=1:n,
  for l=1:n,
    eval(['h' num2str(j) num2str(l) '=filter(B,A,fftshift(real(ifft(Y'
num2str(j) num2str(l) '))));']);
  end
end

save new_filters_2by18_5cm_front.mat h11 h12 h21 h22 Fs

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Control of result
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% W = H*C (Results in frequency domain)
for j=1:m/2,
    W11(j,:)=Y11.*LLfr(j,:) + Y21.*LRfr(j,:);
    W12(j,:)=Y12.*LLfr(j,:) + Y22.*LRfr(j,:);
    W21(j,:)=Y11.*RLfr(j,:) + Y21.*RRfr(j,:);
    W22(j,:)=Y12.*RLfr(j,:) + Y22.*RRfr(j,:);
end

% Results in time domain
for j=1:m/2,
    w11(j,:)=fftshift(real(ifft(W11(j,:))));
    w12(j,:)=fftshift(real(ifft(W12(j,:))));
    w21(j,:)=fftshift(real(ifft(W21(j,:))));
    w22(j,:)=fftshift(real(ifft(W22(j,:))));
end


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Figures
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Plot of the truncated impulse responses
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
time_axis = [0:1/44.1:(length(LLir1)-1)/44.1]';
```

```
figure(1)
i=1;
for j=1:m/2,
    subplot(m,n,i), plot(time_axis,LLir1(:,j));
%   title('IR - Left ear, left speaker')
    xlabel('Time [msec]')
    ylabel('Pa/Pa')
    axis([time_axis(1) max(time_axis) -5 5])
    i=i+1;

    subplot(m,n,i), plot(time_axis,LRir1(:,j));
%   title('IR - Left ear, left speaker')
    xlabel('Time [msec]')
    ylabel('Pa/Pa')
    axis([time_axis(1) max(time_axis) -5 5])
    i=i+1;
end

for j=1:m/2,
    subplot(m,n,i), plot(time_axis,RLir1(:,j));
%   title('IR - Left ear, left speaker')
    xlabel('Time [msec]')
    ylabel('Pa/Pa')
    axis([time_axis(1) max(time_axis) -5 5])
    i=i+1;

    subplot(m,n,i), plot(time_axis,RRir1(:,j));
%   title('IR - Left ear, left speaker')
    xlabel('Time [msec]')
    ylabel('Pa/Pa')
    axis([time_axis(1) max(time_axis) -5 5])
    i=i+1;
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Plot of frequency responses of IRs
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
figure(2)
i=1;
for j=1:m/2,
    subplot(m,n,i), semiplot(LLfr(j,:),Fs,20,'b');
%   title('FR - Left ear, left speaker')
    xlabel('Frequency  [Hz]')
```

```
    ylabel('Amplitude [dB]')
    i=i+1;

    subplot(m,n,i), semiplot(LRfr(j,:),Fs,20,'b');
%   title('FR - Left ear, left speaker')
    xlabel('Frequency  [Hz]')
    ylabel('Amplitude [dB]')
    i=i+1;
end

for j=1:m/2,
    subplot(m,n,i), semiplot(RLfr(j,:),Fs,20,'b');
%   title('FR - Left ear, left speaker')
    xlabel('Frequency  [Hz]')
    ylabel('Amplitude [dB]')
    i=i+1;

    subplot(m,n,i), semiplot(RRfr(j,:),Fs,20,'b');
%   title('FR - Left ear, left speaker')
    xlabel('Frequency  [Hz]')
    ylabel('Amplitude [dB]')
    i=i+1;
end


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Plot of IRs of the inverse system
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
figure(3)
i=1;
for j=1:n,
  for l=1:n,
    subplot(n,n,i), plot(eval(['h' num2str(j) num2str(l)]));
%   title('Inverse response - Left ear, left speaker')
    xlabel('Time [msec]')
    ylabel('Pa/Pa')
    axis([1 length(eval(['h' num2str(j) num2str(l)])) -1 1])
    i=i+1;
  end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Plot of FRs of the inverse system
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
figure(4)
i=1;
for j=1:n,
  for l=1:n,
    subplot(n,n,i), semiplot(eval(['fft(h' num2str(j) num2str(l) ')']),
Fs,20,'b');
    title('Inverse system - Left input, left speaker')
    xlabel('Frequency  [Hz]')
    ylabel('Amplitude [dB]')
    i=i+1;
  end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% IRs after convolution
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
figure(5)
i=1;
for j=1:m/2,
    subplot(m,n,i), plot(w11(j,:));
%   title('Resulting IR - Left ear 1, left speaker')
    xlabel('Time [msec]')
    ylabel('Pa/Pa')
    axis([1 length(w11) -1 1])
    i=i+1;

    subplot(m,n,i), plot(w12(j,:));
%   title('Resulting IR - Left ear 1, left speaker')
    xlabel('Time [msec]')
    ylabel('Pa/Pa')
    axis([1 length(w12) -1 1])
    i=i+1;
end

for j=1:m/2,
    subplot(m,n,i), plot(w21(j,:));
%   title('Resulting IR - Left ear 1, left speaker')
    xlabel('Time [msec]')
    ylabel('Pa/Pa')
    axis([1 length(w21) -1 1])
    i=i+1;

    subplot(m,n,i), plot(w22(j,:));
%   title('Resulting IR - Left ear 1, left speaker')
```

```
    xlabel('Time [msec]')
    ylabel('Pa/Pa')
    axis([1 length(w22) -1 1])
    i=i+1;
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% FRs after convolution
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
figure(6)
i=1;
for j=1:m/2,
    subplot(m,n,i), semiplot(W11(j,:),Fs,20,'b');
%   title('Resulting FR - Left ear, left speaker')
    xlabel('Frequency  [Hz]')
    ylabel('Amplitude [dB]')
    i=i+1;

    subplot(m,n,i), semiplot(W12(j,:),Fs,20,'b');
%   title('Resulting FR - Left ear, left speaker')
    xlabel('Frequency  [Hz]')
    ylabel('Amplitude [dB]')
    i=i+1;
end

for j=1:m/2,
    subplot(m,n,i), semiplot(W21(j,:),Fs,20,'b');
%   title('Resulting FR - Left ear, left speaker')
    xlabel('Frequency  [Hz]')
    ylabel('Amplitude [dB]')
    i=i+1;

    subplot(m,n,i), semiplot(W22(j,:),Fs,20,'b');
%   title('Resulting FR - Left ear, left speaker')
    xlabel('Frequency  [Hz]')
    ylabel('Amplitude [dB]')
    i=i+1;
end
```

Thank you for taking time to participate in this listening test.

This test is about the ability to localise sound around your head and is performed by pointing out the direction of the sound source. You will sit on a chair with loudspeakers placed in a circle around you.

In the first session you will be presented with some test sounds coming from the different directions. This will last for 10 minutes and is to familiarize you with the sounds in the test. You should answer according to the instructions, which will be given verbally. After this, the actual test begins. It consists of 3 sessions, the first taking around 15 minutes and the second 10 minutes. You will have a break between the sessions. The sound is coming from one of the loudspeakers specified in the figure. After each sound is played you should indicate the direction. The sounds

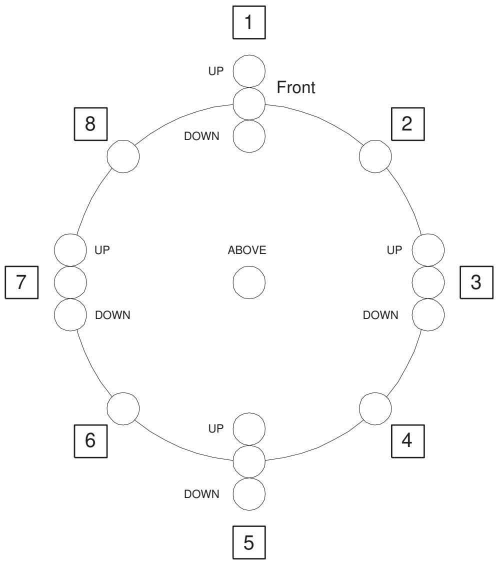

Figure B.1: Loudspeaker directions

and breaks in between have the same length during the session and you should give your answer verbally before the next sound is played. You have to give an answer, so if you are not sure of the direction, give your best possible guess. The sounds will be presented more than once in random order for each session. This means that two following sounds can come from the same direction.

If you have any questions please ask them and inform us about possible hearing disorders before the test.