Influence of Reflections on Crosstalk Cancelled Playback of Binaural Sound

Doctoral Dissertation

Asbjørn Sæbø

Department of Telecommunications
Norwegian University of Science and Technology

N-7491 Trondheim, NORWAY

December, 2001

Abstract

This work has investigated how crosstalk cancelled loudspeaker reproduction of binaural sound intended for anechoic conditions is influenced by taking place under non-anechoic conditions. The scope has been limited to an investigation of the effect of early reflections on localisation for target directions in the horizontal plane.

Crosstalk cancellation was implemented in software, with routines for the inversion of transfer functions, the design of crosstalk cancelling filters, filtering of sound and playback simulation. The system for crosstalk cancellation was found to give a realistic, natural and convincing reproduction of binaural signals.

The effect of a single reflecting surface on a given setup for crosstalk cancelled loudspeaker playback was studied using a simplified model and playback simulation. It was found that the crosstalk cancelling filtering designed for the direct sound would not give crosstalk cancellation for the reflected sound, that the reflections in many cases were not geometrically plausible for a real source in the intended source directions, and that the interaural time difference for the reflections was nearly equal to that for a source placed at the reflecting surface.

The effect of reflections delayed 5 ms and 10 ms with respect to the direct sound was investigated through listening tests. The playback setup was two Bowers & Wilkins 801 loudspeakers placed next to each other in front of the listener, at two meters distance, in an anechoic room. The reflections came from front left and front right, and were produced using reflecting surfaces. Five setups were used: Anechoic, close wall, close wall with reflection cancelling, far wall and two walls. Crosstalk cancelled binaural recordings of male and female speech from sixteen directions were used as source signals. A total of 5180 answers were collected from nine participants.

The listening test data were bimodally distributed due to front/back reversals. A model for statistical characterisation of such data was developed. The model describes reversed and unreversed answers independently, and may be used to find estimates for reversal rate, average perceived directions and spread for reversed and unreversed answers. The model was used to describe the listening test data.

The deviation of the perceived directions from the presented directions was studied, and used to compare the setups with reflections to the anechoic setup. The results showed that compared to anechoic reproduction, localisation was altered for reproduction taking place under non-anechoic conditions. It was found that the deviation of the average perceived directions from the presented directions was higher for the setups with reflections than for the anechoic setup. The setups with reflections also had relatively more answers localised towards the direction of the reflection(s). The non-anechoic setups showed reversal patterns differing from those for the anechoic setup, reversal rates 11 to 30 percent higher than for the anechoic setup, and a larger fraction of answers localised to the front half plane. Generally it was found that the setups that differed most from the anechoic setup (close wall, two walls) also gave the localisation that differed most from localisation in the anechoic setup. Reflection cancelling was found to give results closer to the anechoic setup than the identical setup without reflection cancelling did.

Preface

This dissertation is written for the degree of *Doktor Ingeniør*, and summarises most of my work on crosstalk cancelled binaural reproduction and reflections. The work has been carried out at the Acoustics Group of the Department of Telecommunications at the Norwegian University of Science and Technology (NTNU), in the years 1995 to 2001.

Undertaking this study has been interesting, fun, demanding and time-consuming. It has been a valuable experience from which I have learnt much; about acoustics and signal processing, about research, and about myself.

Acknowledgements

Firstly, I thank my teacher and supervisor Professor Asbjørn Krokstad of the Acoustics group and Research Scientist Svein Sørsdal of SINTEF, co-supervisor during the last part of the work. Their help and advice have been of great value, both to my work and to me personally. They have complemented each other well, and have always been willing to answer questions, discuss problems and spend time on me and my work. If I have managed to pick up some of their ways of thinking, that might be one of the most valuable results of this work.

I am grateful to Laboratory Engineer Øyvind Lervik, Åse Vikdal, secretary of the Acoustics Group until recently and Inghild Torgersen, the current secretary of the Acoustics Group. Generally acknowledged as the most important persons for the day to day business of the Acoustics Group, they have been very helpful in solving all kinds of practical problems.

Olav Arntzen and Martin Teschl are thanked for their cooperation and help in parts of the experimental work. Likewise, I thank the participants in the listening tests and those who assisted in the making of the binaural recordings. Associate Professor Magne Kringlebotn of the Department of Physics is thanked for making available the necessary equipment for hearing tests, and for his advice in this respect. I am also grateful to Works Manager Leif Malvik and his colleagues of the department workshop who, among other things, have produced the reflecting walls for the experiments.

Associate Professor Magne Halstein Johnsen of the Department of Telecommunications is thanked for his guidance on MSE FIR-filters and inversion. Professor Henrik Møller of Aalborg University and D.Sc.(Tech) Jyri Huopaniemi of Nokia are thanked for advice and valuable discussion in email conversations. Professor Philip A. Nelson of Institute of Sound and Vibration Research is acknowledged for providing a copy of an article he has co-authored.

The technical division at NTNU has managed to teach me a few things on the topics of structure-borne sound and the practical consequences of noise. In doing this, they have also reminded me that knowledge comes at a cost.

The Acoustics Group and the Department of Telecommunications have financed the work, provided necessary equipment and generally arranged things well for my work. Associate Professor Ragnar Hergum, administrative head of the department, have been helpful in many ways. I am also grateful to Silence International, my employer since March 2000, whose financial support was a great help in finishing this dissertation. My colleagues there, Managing Director Peter Molthe and Development Manager Johan L. Nielsen are thanked for pleasant company and their continuing interest in the progress of my work.

My colleagues, friends and fellow students at the Acoustics Groups of NTNU and SINTEF have provided good company both at work and privately, and a friendly and stimulating working atmosphere which I have greatly appreciated. Rolf Tore Randeberg, with whom I have shared office, is especially thanked.

My family and friends are thanked for their continuing support and help in so many ways during these years. Synnøve and Sigve in particular are thanked for their help during rough periods. My wife Kirsti is especially thanked for her encouragement and patience, and for taking care of me, our sons and our lives in general during the periods where I have spent most of my time at work and little of it at home.

Asbjørn Sæbø
Trondheim, 2001-12-12

... doch möcht' ich alles wissen
(Johann Wolfgang von Goethe)

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"When I sit in an acoustically perfect hall (full of people), in the best seat for hearing, and listen to an orchestra, I hear such and such sounds. I want to hear precisely this effect from a recording, if that be possible."

(A 1928 music critic, cited in [Torick, 1998].)

"[...] we heard [...] a womans laughter, so clear and distinct as if she were in the same room, or at least in the kitchen near by."

(Listening to a radio for the first time, as described in a humorous parody of a small-town newspaper [Hilditch, 1970].)

1.1 Background

The possibilities for recording and subsequent playback of sound came with Thomas Alva Edisons invention of the phonograph in 1877 [Hugonnet and Walder, 1998]. Following this, the development of audio technology has progressed through a series of innovations and improvements [Hugonnet and Walder, 1998, Audio Eng. Soc., 1998]. To achieve a reproduction of sound that is true to the original has always been one of the main goals. Ideally, for this the whole transmission chain, including recording, mixing, storage and playback, should be transparent and nothing but a "hole through" to the original sound event, allowing the listener to access it or (passively) participate in it as if he was present. An important aspect of this is to preserve the spatial aspects of the sound event.

Several recording and playback techniques have been developed with the aim of preserving or creating spatial auditory impressions for the listener. The first stereophonic transmission of sound, Clement Aders *theatrephone*, took place in Paris as early as 1881 [Hugonnet and Walder, 1998, Torick, 1998]. Now, various forms of 3-D

(three-dimensional) and spatial audio exist. This area sees much interest, and a rapid development is taking place [Begault, 1994, AES 16th Int. Conf., Gilkey and Anderson, 1997]. The applications are many. Examples include recording and playback of sound events with improved reproduction of spatiality, virtualisation and auralisation for such diverse uses as games, development and research [e.g. Kleiner et al., 1993, Huopaniemi, 1999], telecommunications [West et al., 1992, Rimmel, 1999, Begault, 1999b] and user interfaces [Shinn-Cunningham et al., 1997].

Binaural technology and binaural reproduction have a central place among these technologies [Møller, 1992, Blauert, 1997a]. Binaural technology utilises the natural human spatial hearing and is based upon the use of the acoustic signals at the two ears, binaural signals. Binaural reproduction deals with the playback of such signals. The strength of this technology lies in its ability to deliver very natural and convincing 3-D auditory images to a listener, for instance for high quality authentic reproduction of sound events.

Loudspeakers may be used for binaural reproduction. This introduces *crosstalk*, where the left channel signal intended for the left ear will also be heard by the right ear and vice versa. This unwanted effect may be eliminated by prefiltering the binaural signal with an inverse system called a *crosstalk cancelling filter*. The resulting playback signals will partly cancel each other, leaving only the intended binaural signals at the ears of the listener.

Loudspeaker playback of audio usually takes place in rooms. A room is characterised acoustically by the presence of reflections and reverberation [Kuttruff, 2000, Cremer et al., 1982a,b]. Sound radiated from the loudspeakers will be reflected and scattered by the room boundary surfaces and anything placed in the room, and this reflected sound will interfere with the direct and intended transmission from loudspeaker to listener.

Crosstalk cancelled loudspeaker reproduction of binaural signals is typically designed for anechoic conditions. Due to the precise and accurate nature of the cancellation process, it might be expected that the effects of the room will degrade the relative quality of binaural reproduction more than for other playback methods, like e.g. stereophony. But if such playback systems can be used in more ordinary rooms, with reflections and reverberation, it will be of great importance for their practical use, opening up for new applications and a far more widespread use.

How non-anechoic conditions affect the performance of crosstalk cancelled playback has been the matter of some discussion, but has, with a few exceptions, not been assessed directly. A review of earlier work related to the topic is given in chapter 2.4. Mainly, two opinions are present, one being that such systems requires an anechoic or mostly anechoic space to function correctly, and the other that the systems may function well under more or less reverberant conditions, at least as long as care is taken to avoid early reflections. The topic has not been thoroughly investigated before, though.

1.2 Objective and method

The purpose of this work has been to investigate how crosstalk cancelled loudspeaker reproduction of binaural sound intended for anechoic conditions is influenced by taking place under non-anechoic conditions. It has been the aim to try to reach more specific conclusions about how this affects the performance of such playback, and to identify and quantify the effects this give rise to. The scope has been limited to an investigation of the effect of early reflections for target directions in the horizontal plane.

The approach chosen has been mainly experimental. Crosstalk cancelling has been implemented, binaural signals prepared from recordings of sound sources, and a series of listening test experiments with crosstalk cancelled binaural playback under anechoic and non-anechoic conditions has been undertaken. A statistical model suitable for describing the results from the listening tests has been developed and used extensively.

As a measure of performance, localisation has been chosen. Deviation of localised directions from the original directions of the recorded sources has been computed and used to compare non-anechoic reproduction to reproduction under anechoic conditions.

1.3 Overview of the dissertation

Chapter 2: A short review of reflections in rooms is given, and directional hearing is covered in some detail. Binaural reproduction and the principle of crosstalk cancellation is presented. Earlier work related to the effect of reflections on crosstalk cancelled binaural reproduction is reviewed.

Chapter 3: The implementation of crosstalk cancellation developed during the course of this work is presented. This includes routines for data treatment, design of inverse filters for single responses, design of crosstalk cancelling filters, crosstalk cancelling filtering, and playback simulation.

Chapter 4: Binaural reproduction over loudspeakers with a reflection present is discussed and simulated using a simplified model.

Chapter 5: The experimental work, consisting of a series of listening tests of crosstalk cancelled reproduction under different playback conditions, is presented. Design considerations and the experimental conditions are described.

Chapter 6: Criteria for analysis of the listening test data are presented. The statistical problems associated with reversals are discussed. A statistical model for bimodally distributed data is developed and applied to the data from the listening tests.

Chapter 7: Results are computed from the listening test data according to the analysis criteria given in the previous chapter.

Chapter 8: The results are discussed to evaluate the effect of reflections on several aspects of localisation: Deviation of average perceived direction, spread of answers, reversals and localisation towards the reflections.

Chapter 9: The work and the conclusions are summarised and suggestions for further work are given.

1.4 Contributions

During the course of this work many persons have contributed help and advice, as acknowledged in the preface. Two of them, Olav Arntzen and Martin Teschl, have cooperated with the author during their term projects, for which the author acted as an advisor. Arntzen carried out the preliminary experiment reported in chapter 5.1, with a crosstalk cancelling filter designed by the author. Teschl took hand of most of the production of binaural signals to be filtered for the listening tests, and also conducted much of the first half of the tests (chapters 5.4.2 and 5.4.4). Their work is also reported in [Arntzen, 1998] and [Teschl, 1999]. Except for this, and with the acknowledgements in the preface noted, the work reported here is carried out by the author.

Parts of the work reported here has been previously presented in [Sæbø, 1998, 1999]. Also, a slightly rewritten version of chapter 2.3, "Fundamentals of binaural reproduction", is being used as course material at the Acoustics Group.

During the work reported in this dissertation, the author has also acted as an advisor for some student works on related topics. In addition to the term projects of Arntzen and Teschl mentioned above, this is the main theses (similar to masters theses) of Strømsvåg [1996] and Larsen [1997], and the term projects of Holmefjord [1997] and Woje [1997].

With the exception of the article by Takeuchi et al. [1997a], this is, to the best of the authors knowledge, the first work that specifically investigates the effects of reflections on crosstalk cancelled binaural reproduction, and the first to quantify the effects upon localisation performance. The results obtained should be useful for further research in this area, and for the continued application of this technology.

The software implementation for the design of crosstalk cancelling filters and crosstalk cancelling playback is easily accessible and usable for others, as it is written in a common high-level language for mathematics and signal processing. In addition to the current work, it has already been used in several student works at the Acoustics Group [Arntzen, 1998, Teschl, 1999, Kløften, 2001, Anmarkrud, 2001].

The statistical method developed consists of a combination of a *dual-normal* statistical model and associated curve-fitting techniques for the extraction of statistical parameters. This model gives the ability to accurately describe listening test results containing reversed data points, something which up to now in many cases has been problematic to the degree that statistical description has been avoided. This model should therefore prove very useful in these situations.

Chapter 2

Binaural reproduction and reflections

Binaural means *"Of or pertaining to, or used by, both ears"* [Webster (1913), 2000]. A good definition of binaural technology is given by Blauert [1997a]: *"Binaural technology is a body of methods that involves the acoustic input signals to both ears of the listener for achieving practical purposes, for example, by recording, analyzing, synthesizing, processing, presenting, and evaluating such signals."* Binaural technology utilises acoustic ear signals, and is therefore closely tied to natural human hearing and the way this works. The ear signals are usually called *binaural signals*. Such signals may be natural ear signals of origin or imitate such natural ear signals.

Playback of binaural signals in a way that may reproduce these signals at the ears of a listener is called *binaural reproduction*, and is an important part of binaural technology. A key application is as the playback part of a complete binaural transmission chain. Such a transmission chain is capable of faithful reproduction of sound events, with the attractive property that also the auditory spatial aspects are preserved.

Binaural reproduction may be done using loudspeakers, using a technique called *crosstalk cancellation*. This involves, among other things, the inversion of transfer functions. The room in which crosstalk cancelled reproduction takes place will influence the playback, interfering with the intended reproduction to some degree. But the effects of such room interaction are not clear. Playback under anechoic conditions has been the normal approach, and is often taken as a necessity. However, there are also claims that more or less reverberant spaces may be used for such reproduction, as long as strong early reflections are avoided [e.g. Bauck and Cooper, 1996].

2.1 Reflections in rooms

Audio playback systems are usually placed in rooms. Acoustically, rooms are characterised by beeing volumes enclosed by surrounding surfaces; walls, ceiling and floor.

These surfaces, and also furniture and other objects placed in the room, reflect and scatter sound, setting up a complicated sound field and changing the transmission of sound from what it would have been in a free field.

The transmission of sound from a source to a receiver is given, in the time domain, by a *room impulse response* (RIR). This response completely describes the sound transmission, including all effects of reflection and scattering. In addition to the room itself, the RIR is a function of the source and receiver positions and their directivities [e.g. Mourjopoulos, 1985]. The corresponding frequency domain response is called a *room transfer function* (RTF).

The temporal pattern of reflections, which may be seen directly from the RIR, typically follows a general structure: First comes the direct sound, corresponding to sound transmission in a free field. Then there is a series of single reflections, the result of sound beeing reflected once or a few times. The reflection density increases rapidly (it is proportional to time squared), and at some point in time individual reflections may no longer be distinguished; there is a general reverberation. There is a decay of sound, later reflections generally beeing weaker than earlier ones [Kuttruff, 2000].

The phase of RTFs will turn out to be of interest. The work by Neely and Allen [1979] show that RTFs can not generally be expected to be minimum phase.

A typical living-room might have a rectangular floor with an area of 20 to 50 square meters, and a height of about 2.5 meters. For an audio playback system in such a room, the loudspeakers placed not to close to the walls, the first reflection to reach the listener will typically be the floor reflection, which will be delayed 1 to 4 milliseconds with respect to the direct sound. Further first order reflections (sound reflected once) will come from the wall behind the loudspeakers, the ceiling and the side-walls, with delays typically up to ten milliseconds.

2.2 Directional hearing

Directional hearing is a large subject. Only some aspects relevant to the topic of this thesis will be discussed here. More comprehensive and detailed treatments can be found in [Blauert, 1997b] and [Gilkey and Anderson, 1997]. A tutorial overview can be found in [Hartmann, 1999].

2.2.1 Introduction

Hearing may be defined as the "physiological process of perceiving sound" [Encyclopædia Britannica Online, Accessed 20 January 2000]. It is one of several senses, or *modalities* [1], we use to experience our environment and ourselves. In importance it may be ranged as second, after vision. Hearing does not provide us with as good a

[1] A group of sensory impressions that are similar (or of the "same kind") and mediated by a given (particular) sense organ, is called a modality, or, less precisely, a sense. [Schmidt, 1978, p. 2]

resolution or precise gathering of spatial information as the visual or tactile systems may. But it is, as the only one of our senses, capable of giving us a complete three-dimensional impression of the world, and is therefore a good overview and warning sense.

Stimuli and perceptions

In general, a sensing process may be described like this: *Stimuli*, physical environmental influences, are picked up by *sense organs*, like e.g. the ears or the eyes. Information that is in some way corresponding to the stimuli is passed on via the sensory nerves to the central nervous system, where it is represented as *sensory impressions*. The sensory impressions are combined into *sensations* and interpreted, giving *perceptions* [Schmidt, 1978, p 2-5]. For the sense of hearing, *sound* is the main stimulus. For the associated perceptions, the term *"auditory"* is used, like in "auditory event" or "auditory object" [Blauert, 1997b].

A clear distinction must be made between the physical world and our perception of it. Specifically, the stimuli exciting our senses, and the sensations and perceptions arising from these stimuli should not be confused, as they are not the same. There is an association between stimuli and perceptions. But this correspondence is not a causality, but rather a *mapping* [Blauert, 1997b, Schmidt, 1978, Zwicker and Fastl, 1999].

Reference coordinate system

When working with hearing it is common that directions and positions are given with reference to a *head-related*, or *listener-centric*, coordinate system, as shown in figure 2.1 [Blauert, 1997b, p 14]. Such a coordinate system will be assumed throughout this work. Spherical coordinates (azimuth, elevation and distance) are used, and the origin is taken to be in the middle of the head, halfway on a straight line connecting the upper margins of the ear channel entrances. Terms like "forward", "backward", "up", "down", "left" and "right" have their natural meaning. Three planes are defined, which intersect at right angles to each other at the origin: The *horizontal plane*, the *median plane* (or *median saggital plane*) and the *frontal plane*. The median plane divides between the left half-space and the right half-space, the frontal plane between the forward and backward half-spaces. If symmetry is assumed, it is mirror symmetry of the left and right half-spaces about the median plane.

2.2.2 Localisation and localisation cues

Hearing is inherently spatial. Auditory events are taking place in an auditory space at specific times, and their spatial and temporal aspects correspond naturally to spatial and temporal aspects of the physical stimuli [Blauert, 1997b, Schmidt, 1978]. The term *localisation* is used for the law or rule by which the location of an auditory event

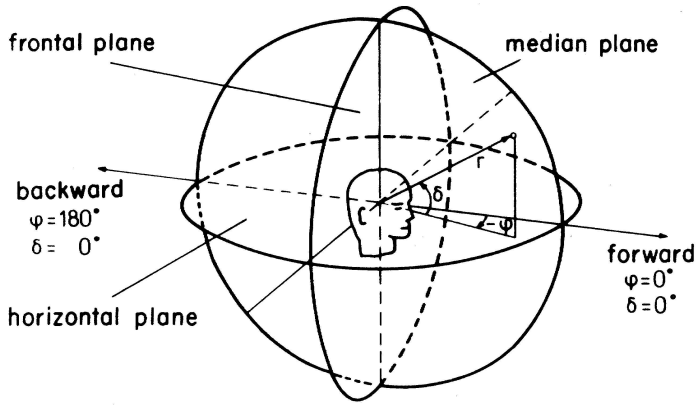


Figure 2.1: A head-related coordinate system. Figure copied from [Blauert, 1997b, figure 1.4].

in auditory space is related to a specific attribute or attributes of a sound event, or another event that is in some way correlated with the auditory event [Blauert, 1997b]. This localisation is the topic of directional hearing.

Localisation is guided by, or determined by, what is called *localisation cues*. These cues are properties of the stimuli presented to us. Most cues are acoustical, connected to physical properties of the sound signal. Other cues may be non-acoustical. The totality of cues is weighted and interpreted to form auditory events.

All acoustical cues for directional hearing are present in the sound received by the ears/listener. The sound, on its way from a source to a listener's ears, undergo linear distortions due to the presence of the listener's outer ears, head and body. These distortions are dependent upon, and characteristic for, the position of the sound source relative to the listener [e.g. Blauert, 1997b, Zwicker and Fastl, 1999]. These distortions form the basis for most acoustical cues.

Any physical aspect of the acoustical waveforms reaching a listener's ears that is altered by changes in the position of the sound source may be considered a potential cue [Wightman and Kistler, 1997]. A possible grouping of these cues is into binaural, or interaural, cues, where information from both ears is used, and monaural cues, where information from each ear is used separately.

Binaural cues

Binaural cues are derived from differences in the signals at the two ears [e.g. Wightman and Kistler, 1997, Blauert, 1997b]. The two binaural cues, which are the two main acoustical cues in general, are the *interaural time difference* (ITD) and the *interaural level difference* (ILD) [e.g. Wightman and Kistler, 1997, Hartmann, 1999, Wightman and Kistler, 1992, Bernstein, 1997]. In the horizontal plane, these cues have been found to be dominant for localisation [Middlebrooks, 1997, Makous and Middlebrooks, 1990].

The ITD for a given signal is the difference in arrival time between the two ears. It varies with the horizontal and vertical angle of the source, from zero in the median plane to about 600 to 800 microseconds for directions straight to the sides [e.g. Møller et al., 1995, Wightman and Kistler, 1997, Huopaniemi, 1999]. It does not vary much with frequency, and may to a first approximation be considered constant in this respect [Wightman and Kistler, 1997]. In the horizontal plane, ITD at low frequencies has been found to be 1.5 times larger than at high frequencies [Kuhn, 1977].

For low frequency signals, up to around 1.5kHz, ITD cues are derived from the relative phase, or the fine structure, of the signals. For higher frequency signals, the size of the head is comparable to the wavelength of sound, turning phase differences ambiguous, but ITD cues may still be extracted from the envelope of the signals [e.g. Blauert, 1997b, Shinn-Cunningham et al., 1997].

ITD is generally accepted as the dominating cue for localisation [Gilkey and Anderson, 1997, p. xvii]. Wightman and Kistler [1992] conducted listening tests with artificial signals, wideband stimuli with conflicting ITD and ILD cues. Their results showed that ITD almost totally dominated localisation for stimuli containing low frequencies. For high-pass filtered stimuli, this dominance disappeared. In [Wightman and Kistler, 1997, p. 12] they argue that ITD should be considered the most reliable cue: It is not dependent upon the characteristics of the source, it gives nearly the same information in all frequency bands, and the relation between ITD and source position is not very dependent upon the individual listener.

The ILD is the level difference between the signals at the ears. Sound from a source will be scattered by the head, causing the sound pressures at the two ears to differ for sources not in the median plane. The hearing is actually sensitive to ILD over the whole frequency range [Hartmann, 1999, Blauert, 1997b]. But physically, ILDs exist only for wavelengths where the head is large enough to actually cast a shadow or cause interference patterns [Morse and Ingard, 1986, Hartmann, 1999], and they are therefore small at low frequencies. For sources close to the head, ILDs exist also for low frequencies [Huopaniemi, 1999]. The domain of ILD cues is often taken to be from around 1500Hz or higher and upwards through the range of hearing [Wightman and Kistler, 1997, Blauert, 1997b, Middlebrooks, 1997]. The ILD varies with both the horizontal and vertical angle of the source and with frequency, as can be seen from e.g. [Wightman and Kistler, 1997, fig. 6], [Duda, 1997, fig. 3].

Monaural spectral cues

Monaural cues are also important for localisation. Unlike binaural cues, monaural cues are derived from the signal itself. They are therefore dependent upon, and may require knowledge of, the characteristics of the sound source [e.g. Wightman and Kistler, 1997]. The main monaural cues are spectral cues. (For a discussion of why monaural temporal cues may be considered unimportant, see [Wightman and Kistler, 1997].)

Due to scattering and reflections, the frequency spectrum of the sound arriving at the ears is dependent upon the direction of the source. Therefore, cues may be extracted from these spectra. These cues are believed to aid in front-back localisation and elevation determination [Blauert, 1997b, Hartmann, 1999, Wightman and Kistler, 1997, Makous and Middlebrooks, 1990]. But they may also aid in horizontal localisation when binaural cues are not present [Middlebrooks, 1997].

In general, accurate localisation requires broad band sources of sound [e.g. Wightman and Kistler, 1997]. This is especially true for sources in the median plane, where the position of the auditory event of narrow band sources is mostly dependent upon the frequency content of the signal [Blauert, 1997b]. Similarly, one would believe broad band sources to be necessary to resolve front/back ambiguities with the help of monaural spectral cues.

Non-acoustical cues

Localisation may be influenced by non-acoustical cues as well. An important case is *multi-modality*, where information from more than one sense interact to form the final perception(s). The interaction between vision and hearing may be strong, with visual cues affecting auditory localisation. What a listener sees, and where he sees it, during presentation of a sound may influence the position of the auditory event [Blauert, 1997b]. For correlated visual and auditory information, the modalities will often reinforce each other, giving a single perceptual event [Shinn-Cunningham et al., 1997]. However, for conflicting cues, the modalities will bias each other, one modality often beeing dominant. For auditory localisation, vision is often dominating, some times to the degree that the auditory event is perceived at the position of the visual event, [Begault, 1999a, Shinn-Cunningham et al., 1997, Blauert, 1997b, p 19], although this is not always the case [Perrot, 1993].

Source familiarity is another cue, as mentioned in [Wightman and Kistler, 1992]. Knowledge of the specter of the sound source may allow the separation of this from the spectral modifications done to the sound on its way to the ears, thereby influencing the use of monaural spectral cues. It may also play a role for resolving front/back confusion [Wightman and Kistler, 1997].

Head movements, or listener-controlled movement of the relative position of the source, may also be used as a cue, and may be useful to resolve front-back confusions [Blauert, 1997b, Shinn-Cunningham et al., 1997, Wightman and Kistler, 1999]

Reversals, the cone of confusion

Reversals may occur [e.g. Wightman and Kistler, 1989b, Makous and Middlebrooks, 1990, Wenzel et al., 1993, Blauert, 1997b, Hammershøi, 1995, Shinn-Cunningham et al., 1997]. This is confusions between front and back,[2] auditory events appearing not in the direction of the sound source, but in a direction that is reversed, or mirrored, with respect to the frontal plane, or, for directions in the horizontal plane, reversed with respect to the interaural axis [e.g. Blauert, 1997b, p 43].

The ears and head may be approximated as two points at opposite sides of a sphere. The locus of points having a constant distance difference to the two ears is then a shell of conical shape, termed the *cone of confusion* by Mills [1972] (see figure 2.2). For a given distance difference and a given distance, this is a circle parallel to the median plane. All sound source positions on this circle will cause equal ITDs and equal ILDs, and can not be distinguished between on basis of these cues. Similar cones of confusion exist also for real heads. These are approximately cone-shaped shells where the variation in ITD and ILD is small [Wightman and Kistler, 1999, figs. 1 and 2]. Localisation based on these cues may therefore be ambiguous, and have to be aided by other cues. If this fails, reversals may occur.

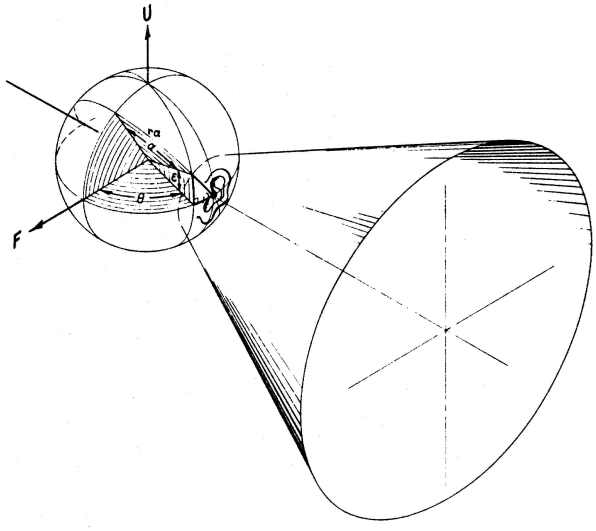


Figure 2.2: A cone of confusion for a spherical head. "U" is up, "F" is forward. The cone is centered at the interaural axis. Taken from [Mills, 1972, figure 13].

[2]Up/down confusions are also noted [Wenzel et al., 1993, Gardner, 1998, Shinn-Cunningham et al., 1997], but will not be treated here.

2.2.3 The precedence effect

The term *precedence effect* is here used a bit imprecisely for a set of effects related to what happens when correlated signals from two spatially separated sound sources reach the ears at the same or nearly the same time, i.e. within up to a few tens of milleseconds of each other. Direct and reflected sound from a sound event, where the reflections are regarded as a separate sources, is a typical example of this.

Dependent upon the arrival time difference between the signals and their relative levels, one or two auditory events will occur, the positions of which will also be dependent upon the properties of the signals [Blauert, 1997b].

Summing localisation

Summing localisation is the name used for the effect where the perceptions of the two sound signals fuse into one auditory event, the position of which is influenced by both sound signals [Blauert, 1997b, Litovsky et al., 1999]. The range of arrival time differences for which summing localisation occurs is commonly taken to be up to one millisecond [Blauert, 1997b, Litovsky et al., 1999, Hartmann, 1997], although Hartmann [1997] mentions that the upper limit may be as low as 0.5 milliseconds. According to Blauert [1997b, fig. 3.5 and p. 325], signals with larger arrival time differences may also produce summing localisation for low levels of the lagging signal.

Summing localisation is the foundation of conventional stereo playback, which is also a typical example of this effect. Two loudspeakers (placed in a normal symmetrical setup with respect to the listener) radiating equal signals will cause an auditory event to appear midway between the speakers. If the signal from one of the speakers is leading the other in time it will dominate localisation, and the auditory event will be displaced towards this speaker [Blauert, 1997b, Litovsky et al., 1999].

The *level* of the sound events also plays a major role in this localisation process [Blauert, 1997b]. For two simultaneous sound events, the stronger one will dominate localisation. Arrival time differences and level differences may reinforce each other (leading signal stronger), or they may counteract each other (lagging signal stronger). They may be "traded" for each other, so similar localisation may be achieved for various combinations of level difference and delay.

Localisation dominance and echo threshold

As the arrival time difference between the two sound signals increase, there is a transition from summing localisation to *localisation dominance* [Litovsky et al., 1999]. This effect is also called *the law of the first wavefront*, [Blauert, 1997b], *echo suppression* [e.g. Clifton and Freyman, 1997] and the *Haas effect* [Haas, 1951], and is the effect for which the term "precedence effect" originally was coined by Wallach, Newman and Rosenzweig [Hartmann, 1997].

As for summing localisation, a single auditory event arises. But in this case, localisation is dominated almost totally by the first sound, so that the position of the auditory event corresponds to the position of the source radiating the first arriving sound [e.g. Blauert, 1997b]. The direct (first) sound takes presedence over later arriving replicas of the same sound. In the words of Hartmann [1997]:"The direct sound wins the perceptual competition over later arriving sound". It is necessary for this effect to occur that the sounds are sufficiently broad band [Blauert, 1997b].

The later arriving sound is not heard as a separate auditory event, it is suppressed. But it may still be noticeable as changes in loudness, timbre and spaciousness of the single auditory event [Clifton and Freyman, 1997, Blauert, 1997b]. It also influences localisation to a certain degree [Blauert, 1997b, Litovsky et al., 1999, Hartmann, 1997]. Litovsky et al. [1999] cite experiments by Shinn-Cunningham et al. [1993], where 80 to 90 percent weight to the leading sound and 10 to 20 percent to the lagging sound is found as typical values (for headphone listening). The lagging sound also increases localisation blur somewhat [Hartmann, 1997].

As the arrival time increases even further, the sounds are heard as separate auditory events, the positions of which are corresponding to the positions of the two sound sources, respectively [e.g. Blauert, 1997b]. The arrival time difference where this happens is called the *echo threshold* [Blauert, 1997b]. This threshold is dependent upon many factors, like e.g. the relative levels of the two sounds and the kind of signals used. It is also subject to individual variations [Litovsky et al., 1999], and it may change over time as a function of the auditory context [Clifton and Freyman, 1997]. Estimates of the echo threshold vary from 2 to 50 milliseconds. For short "clicks", it is somewhere between 5 and 10 milliseconds, while for more complex sounds it is generally higher [Litovsky et al., 1999]. Blauert [1997b] gives a value of 20 milliseconds for speech, while Litovsky et al. [1999] cite even longer delays.

The precedence effects, and binaural hearing in general, are now believed to be cognitive processes that involve evaluation and decisions in the higher and central layers of the nervous system, where auditory cues, cues from other modalities and the listener's expectations and previous knowledge of the situation are taken into account with the purpose of organising the information in a sensible way [Clifton and Freyman, 1997, Blauert, 1997b,a, Moore, 1999]. An example of this is Hartmanns plausibility hypothesis, from his work on localisation in rooms, where ITD cues are weighted according to how plausible they are [Rakerd and Hartmann, 1985, Hartmann, 1993, 1997].

2.3 Fundamentals of binaural reproduction

Binaural signals are signals related to the two ears, corresponding to sound pressures that exist or have existed at a listener's ears, or are intended to be reproduced at a listener's ears. Binaural reproduction deals with binaural signals and their playback. It is a technique for 3-D sound where binaural signals are used and presented to one or more listeners with the purpose of creating auditory illusions with given spatial

and other properties. This way, auditory images consisting of virtual sources of real or synthetic origin may be presented to a listener.

Here, binaural reproduction is taken to include the whole chain of events from the acquisition of binaural signals through the playback, possibly also involving other aspects of binaural technology. For later parts of this work, a stricter definition is used, where binaural reproduction is taken to be the specific process of reproducing binaural signals at the ears of a listener.

2.3.1 Spatial reproduction through natural hearing

The main advantage of binaural reproduction, compared to other kinds of sound reproduction, is the very convincing and natural spatial effects that can be achieved. Full three-dimensional immersive sound images with true reproduction of direction and distance may be obtained [Hammershøi, 1995, Blauert, 1997b]. Several sound sources can be reproduced simultaneously [Schroeder, 1973], and phenomena like the "cocktail party effect", the ability of a listener to focus his attention to one of several concurrent speakers or sound sources at will works as normally [e.g. Sæbø, 1998]. This is achieved by utilising the natural human hearing and its normal workings.

As discussed in the previous section, the sound pressures at the eardrums are the main, and by far the most important, physical inputs to the auditory sense. Many authors have pointed out that these signals therefore necessarily carry the information upon which audition is based. If they can be captured and recreated exactly, the complete auditive experience corresponding to them is assumed to be reproduced, including directional and spatial aspects, reverberation and more. See e.g. Damaske and Mellert [1969/70], Schroeder [1970], Møller [1989, 1992], Hammershøi [1995], Blauert [1997a,b] and Shinn-Cunningham et al. [1997]. From this follows the fundamental idea of binaural reproduction, to present to the ears of the listeners signals that either are original ear signals of origin or imitate such signals, and let audition work as normal.

An important implication of this is that with a proper choice of signals, i.e. the ear signals, only two channels of transmission or storage are necessary to preserve and recreate all spatial and other information necessary to recreate the perception of a given sound event.

2.3.2 Head-related transfer functions

Head-related transfer functions (HRTFs) are fundamental to binaural technology. For a sound source and a listener in free-field, the HRTF is formally defined as the sound pressure at the ear divided by the sound pressure at the position of the center of the listener's head without the listener present [Blauert, 1997b]:[3]

$$\text{HRTF} = \frac{\text{sound pressure at ear of listener}}{\text{sound pressure at head center position (listener absent)}} \tag{2.1}$$

[3]Blauert uses the term *free-field transfer function*.

Other, similar definitions are also used. It may be suitable to leave out the normalisation to the sound pressure without the listener present, or to measure the HRTF under non-anechoic conditions, obtaining a combined RTF and HRTF. The crucial and common point is that the HRTF is a measurement of the sound transmission from a source to a listener's ears and thereby also of the effect of the listener being present.

The HRTF is a function of the position of the sound source with respect to the listener. It is dependent upon both azimuth, elevation and distance. As a transfer function, it is also a function of frequency [e.g. Møller et al., 1995, Huopaniemi and Riederer, 1998, Huopaniemi, 1999]. A complete HRTF set for a given source direction and distance consists of two transfer functions, one for each ear [e.g. Møller, 1992, Hammershøi, 1995]. The time domain equivalent of an HRTF is usually called a *head-related impulse response* (HRIR). For simplicity, the term HRTF will be used here for both cases. The *interaural transfer function* (ITF) is defined as the ratio of the two HRTFs, the HRTF for the ear closer to the source being the reference [Blauert, 1997b]. This transfer function describes the difference in sound transmission from the source to the two ears.

According to the definition, HRTFs should be measured at the eardrum, as it is the signals here that are the input to the hearing mechanism. It has however been shown that transmission along the ear canal and maybe even a few mm outside its entrance is independent of the direction to the source. So measurement points anywhere along this path may be used, while still giving full spatial information [Middlebrooks et al., 1989, Hammershøi, 1995, Hammershøi and Møller, 1996, Burkhard, 1997]. HRTFs measured at the blocked ear canal entrance are favourable because they show smaller inter-individual variation than HRTFs measured at the eardrum or at the open ear canal. They can therefore be expected to be more representative of a typical population [Hammershøi and Møller, 1996, Hammershøi, 1995].

To obtain the correct sound pressure at the eardrum from a measurement done elsewhere, appropriate equalizing must be applied [e.g. Hammershøi, 1995]. But often the sound pressure at the eardrum is not of interest. For purposes of reproduction, the reference sound pressure may be measured anywhere along the path where transmission is independent of source direction, as long as the same reference is used for the playback.

An HRTF set contains a complete description of the sound transmission from the source to the listener [e.g. Hammershøi, 1996, Lehnert, 1993]. Acoustical cues for directional hearing are a result of the filtering of sound resulting from this transmission (section 2.2), and the HRTFs therefore contain all information giving these cues. In the same way, the ITF contains all information for forming interaural difference cues.

Interindividual differences in HRTFs

There are interindividual differences in HRTFs [Wightman and Kistler, 1989a, Hammershøi and Møller, 1996]. These differences increase with frequency. The main

directional-dependent features (peaks and notches in the HRTF amplitude frequency response) follow a common pattern, though, where interindividual variation is small up to 8kHz [Middlebrooks et al., 1989, Møller et al., 1995]. Møller et al. [1995] conclude that in spite of the interindividual differences, their data suggest an objective basis for the binaural technique.

Listening tests using individualised (the listener's own) HRTFs have shown localisation performance equally good as in reality [Wightman and Kistler, 1989b, Møller et al., 1996, Hammershøi, 1995]. Listening tests with non-individualised HRTFs give poorer results, and can in general not be expected to give localisation performance quite as good as in reality [Hammershøi, 1996, 1995]. Typically, the number of errors in the median plane and the number of front-back confusions increase. It is suggested that interaural difference cues are retained, while spectral details are distorted in this case [Wenzel et al., 1993, Møller et al., 1996]).

In [Møller et al., 1999] eleven artificial heads and variants of artificial heads were tested. It was concluded that artificial head recordings do no result in the same localisation as observed in real life. Performance was similar or poorer to the use of HRTFs for a random (other than the listener) person.

Phase of HRTFs

The phase of HRTFs has been a topic of discussion. Mehrgardt and Mellert [1977] found that the transfer functions of the external ear are (nearly) minimum phase up to 10kHz. The data of Wightman and Kistler [1989a] support this conclusion. In [Møller et al., 1995], however, it is found that HRTFs are generally not minimum phase. Also, Kulkarni et al. [1999] claim that the phase spectrum for measured HRTFs can be described as minimum-phase for most (but not all) positions of the source. Shaw [1997] claims that while the HRTF in principle is not minimum-phase, it may be useful to treat it as such. Experiments suggests that HRTF phase can be adequately represented by a combination of minimum-phase functions and a pure time delay (for the ITD) [Kistler and Wightman, 1992, Plogsties et al., 2000], and that listeners are insensitive to details of the interaural phase spectrum as long as the low frequency ITD is maintained [Kulkarni et al., 1999].

Reversals

As noted in chapter 2.2.1, reversals may occur in the localisation process. For binaural reproduction, the reversal rates are usually higher than for real-life listening [e.g. Begault, 1994], in some cases much higher. Several explanations have been offered for this, having in common that simplifications usually done are the cause.

As said above, there are interindividual differences in HRTFs, and it is believed that these details may be utilised to advantage by the listener. In binaural reproduction, though, the same set of HRTFs is often used for all listeners. Wenzel et al. [1993] have found that the use of non-individualised HRTFs in binaural reproduction will

increase the number of front/back reversals, probably due to the distortion of spectral cues. Similar conclusions may be drawn from the results presented in [Hammershøi, 1995].

Head movements may provide information that can be used to resolve the ambiguity inherent in ILD and ITD cues, thereby preventing reversals [Blauert, 1997b, Møller, 1992, Wightman and Kistler, 1999]. However, for binaural reproduction where head-tracking or other dynamic or adaptive techniques are not used, this will not work.

2.3.3 Binaural signals

Binaural signals may be original ear signals, recorded at or in the ears during a sound event. It is now customary to use an *artificial head* [4] for this purpose. An artificial head is a mannequin shaped with the main features like a human head, and may also be equipped with shoulders and a torso. Artificial heads are equipped with microphone capsules in the ears to facilitate recording of binaural signals and measurement of HRTFs.[5]

Binaural signals may also be produced by filtering sound through a set of HRTFs. By choosing appropriate HRTFs (which may include room responses), the sound sources may be placed at will in a virtual environment. One option is to build a virtual sound environment from (typically anechoic) monophonic recordings of sound, or to simulate performances in a given venue, another to use this technique to allow playback of sound intended for one loudspeaker setup to take place over another one [e.g. Schroeder, 1973, Kleiner et al., 1993, Bauck and Cooper, 1996, McKeag and McGrath, 1997, Horbach and Pellegrini, 1998]. HRTFs and binaural signals may also be created synthetically, by modeling transmission to the ears or from knowledge of the acoustical cues and their dependence upon source position [e.g. Kahana et al., 1998, Kistler and Wightman, 1992, Duda and Martens, 1998].

One of the most common and important applications of binaural reproduction is as a means of authentic reproduction of sound events. It is then used as part of a binaural transmission chain, with the purpose of reproducing at a listener's ears the acoustical signals that would have been there if the listener had been present at the sound event that is transmitted. In this case, there lies a certain elegance in the fact that one is, in principle, not concerned with audition as such. Efforts are made to preserve and recreate exactly the physical stimuli that are input to the hearing, not worrying about their interpretation, in the belief that identical input to the hearing will lead to identical sensations. In practice, the mechanisms of audition must of course be considered. Practical limitations enforce simplifications of the process, and knowledge of how the hearing works must be applied to evaluate what simplifications can be done while still producing stimuli that are sufficiently accurate to obtain acceptable results.

[4] Also called *dummy head* or *Kunstkopf*

[5] The first sound recording carried out with an artificial head took place as early as 1927 [Hugonnet and Walder, 1998].

2.3.4 Crosstalk cancelled playback

For the actual playback of binaural signals, headsets or loudspeakers may be used. Both methods are common. Playback with headphones is an intuitive approach, with the advantage that the binaural signals are delivered directly to the ears, with good channel separation. Proper equalisation must be ensured, though [e.g. Hammershøi, 1995, Larcher, 1998]. A drawback of this method is that listeners may experience "inside-the-head locatedness", the auditory events appearing inside the head. Proper equalisation and reverberation may help to remedy this [Begault, 1994]. Another drawback is that listeners may prefer not to wear headphones for reasons of comfort.

In many situations, loudspeaker reproduction may be preferred. There is reason to believe that proper externalisation will be easier to attain using external sources. Gardner [1998] has used the same set of non-individualised HRTFs for reproduction over both headset and loudspeakers, and found, among other things, that loudspeaker reproduction gave better externalisation. Griesinger [1989] claims that (crosstalk cancelled) loudspeaker reproduction in some ways is better than headphone reproduction, because the sound images always are outside of the head. The main limitation of loudspeaker reproduction is that the listener should be in a given position with the head pointing in a given direction [e.g. Schroeder, 1973, Møller, 1992, Blauert, 1997b], unless head-tracking techniques [e.g. Gardner, 1998] are applied.

Many variations of loudspeaker configuration and loudspeaker placement are possible [e.g. Bauck and Cooper, 1996, Foo et al., 1999]. In this work the most common setup, two loudspeakers placed symmetrically in front of one listener, is considered.

Crosstalk

The basis for binaural reproduction is that the binaural signal should be delivered to the ears of the listener, the left ear signal to the left ear, and the right ear signal to the right ear. However, straight forward loudspeaker playback, where the left ear signal is fed to the left loudspeaker and the right ear signal to the right loudspeaker does not lead to the desired result.

The problem is that sound radiated from each of the loudspeakers will reach both ears. This phenomena is called *crosstalk* [e.g. Schroeder, 1970, Damaske, 1971, Møller, 1989], and is illustrated in figure 2.3. The signal reaching the left ear is not the left ear signal alone, but a combination of the left ear signal and the right ear signal, and vice versa. The actual crosstalk is these unwanted interfering signals intended for the opposite ear. Also, there is the complication that the transmission from loudspeaker to ears will filter the sound with the corresponding head related transfer functions.

By suitable equalising, not caring for the crosstalk, good stereophonic reproduction can be achieved [e.g. Griesinger, 1989, Gierlich and Genuit, 1989]. However, due to the crosstalk this technique does not deliver the binaural signals at the ears as intended. It does therefore not qualify for the definition of binaural reproduction used here.

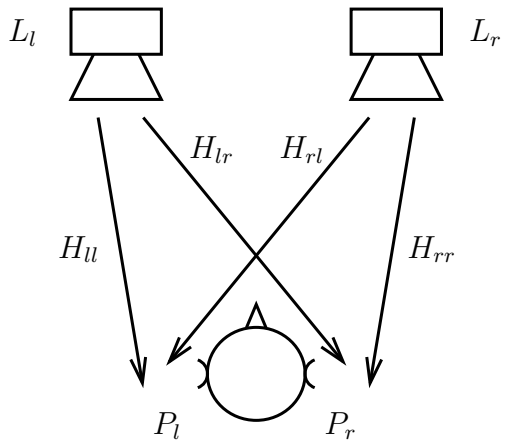


Figure 2.3: Crosstalk. Two loudspeakers and a listener, sound radiated from each of the loudspeakers will reach both ears. L_l and L_r are the loudspeaker input signals. P_l and P_r are the sound pressures at the ears. H_{ll}, H_{lr}, H_{rl} and H_{rr} are the transfer functions from the left and right loudspeakers to the left and right ears, including the voltage to sound pressure transfer functions of the loudspeakers, head related transfer functions and possibly room transfer functions also.

Crosstalk cancellation

To achieve the intended effect; presentation of the ear signals at the ears, various *crosstalk cancellation* techniques have been developed. The purpose of these techniques is to eliminate the crosstalk and the colouration of sound given by the crosstalk and the playback. This is done by designing filtering that is an *inverse system* of the playback situation.

Bauer [1961] launched the idea that binaural recordings might be played back over a sterosystem by eliminating crosstalk. Atal and Schroeder are generally acknowledged as beeing the first implementing such a system [Atal and Schroeder, 1962, Schroeder et al., 1962, Schroeder and Atal, 1963]. Early work was also contributed by Damaske, who introduced the terms *head-related stereophony* and *TRADIS (True Reproduction of All Directional Information by Stereophony)* [Damaske and Wagener, 1969, Damaske and Mellert, 1969/70, Damaske, 1971]. In the following development, Cooper and Bauck [1989] proposed a simplified filter structure called a *shuffler*, and later came up with a generalised approach allowing for more than two loudspeakers and more than one listener [Bauck and Cooper, 1996]. Among the recent contributions are the work by Gardner [1998] on head-tracked crosstalk cancellation, the "Stereo Dipole" with its closely spaced loudspeakers [Kirkeby et al., 1997, Takeuchi and Nel-

son, 1999, Kirkeby and Nelson, 1997, Kirkeby et al., 1998], adaptive inversion filters [e.g. Nelson et al., 1992, Nelson and Orduña Bustamante, 1996, Garas and Sommen, 1998] and the use of *warped filters* where the frequency resolution of the ear is taken into account [e.g. Kirkeby et al., 1999, Huopaniemi, 1999, Huopaniemi et al., 1999].

Crosstalk cancelling filter

(The following development follows Møller [1992] quite closely.) From figure 2.3 it can be seen that the signals at the listener's ears are given as

$$P_l = L_l H_{ll} + L_r H_{rl} \qquad \text{and} \qquad P_r = L_r H_{rr} + L_l H_{lr}, \tag{2.2}$$

Given a binaural signal B, consisting of channels B_l intended for the left ear and B_r intended for the right ear, the goal is that P_l should equal B_l and P_r should equal B_r. Substituting this into equation 2.2 the corresponding loudspeaker signals may be found:

$$L_l = \frac{B_l - B_r \frac{H_{rl}}{H_{rr}}}{H_{ll}\left(1 - \frac{H_{lr}H_{rl}}{H_{ll}H_{rr}}\right)} \qquad \text{and} \qquad L_r = \frac{B_r - B_l \frac{H_{lr}}{H_{ll}}}{H_{rr}\left(1 - \frac{H_{lr}H_{rl}}{H_{ll}H_{rr}}\right)} \tag{2.3}$$

A filter structure implementing these equations is shown in figure 2.4. As can be seen, the inverse system has the same basic structure as the setup giving the crosstalk. It should be noted that the terms H_{lr}/H_{ll} and H_{rl}/H_{rr} are the interaural transfer functions mentioned in chapter 2.3.2, the transfer functions from one ear to the other, corresponding to the difference between transmission on the same side and transmission to the other side, i.e. crosstalk.

The numerators consist of two parts, where the first part is simply the intended signal. The second part is the crosstalk cancellation, where the signal intended for the other ear, multiplied by the interaural transfer function, is subtracted. The denominators also consist of two parts. The first part is the transfer function from the ear to the loudspeaker to the same side ear. The function of this part is to invert this transmission. The second part, one minus the product of the two interaural transfer functions, equalizes a colouration that is due to the crosstalk.

In some situations, symmetry may be assumed, with H_{ll} equalling H_{rr} and H_{lr} equaling H_{rl}. Equation 2.3 may then be simplified accordingly. As an alternative, the "shuffler" filter structure introduced by Cooper and Bauck [1989] may be then used. This filter utilises the symmetry of the setup to achieve less complex filtering.

Loudspeaker spacing

How the loudspeakers are placed is of importance for the crosstalk cancelling process. Damaske and Mellert [1969/70] mentions that a favourable angle between the loudspeakers may simplify the cancellation process. It has been argued and shown that it is advantageous to place the two loudspeakers close together, as this gives a larger zone of cancellation and a more robust system [Bauck and Cooper, 1996,

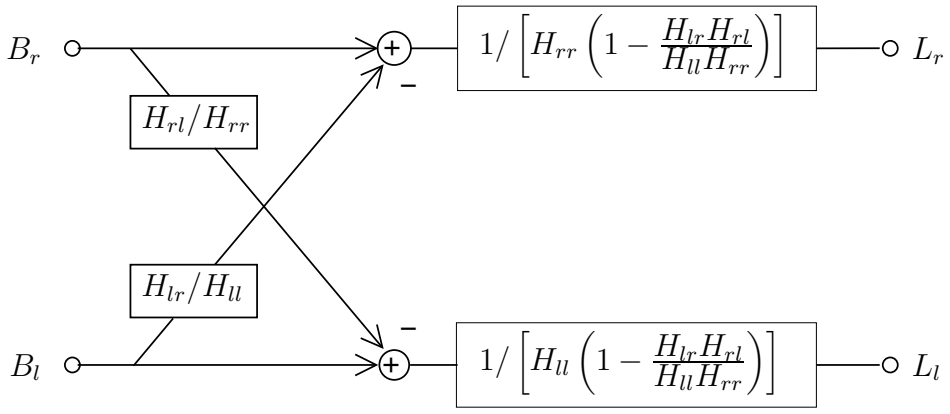


Figure 2.4: Crosstalk cancelling filter, the inverse system of the setup shown in figure 2.3. B_l and B_r are the input signals. H_{ll}, H_{lr}, H_{rl} and H_{rr} are transfer functions and L_l and L_r the loudspeaker input signals, as in figure 2.3.

Takeuchi et al., 1997b, Kirkeby et al., 1998]. Also, under non-anechoic conditions, the close spacing of the loudspeakers helps to avoid early and strong reflections from the sidewalls [Bauck and Cooper, 1996, Nelson et al., publishing date not given].

Ward and Elko [1998] have calculated the optimum loudspeaker spacing for frequencies above 600 Hz, given as $d_s = 2\lambda d_H$, where d_s is the distance between the loudspeakers, λ the wavelength and d_H the distance from the listener's position to a line connecting the loudspeakers. Spacings less than the optimum will also provide good robustness.

The technique of closely spaced loudspeakers has been termed "stereo dipole" [e.g. Kirkeby et al., 1997]. It has been shown that for such a setup, and for virtual sources outside of the angle spanned by the loudspeakers, the two loudspeakers must radiate signals that are out of phase, completely so at the low frequency limit, and therefore function as a dipole [Kirkeby and Nelson, 1997]. An implication of this is that the reproduction of low frequencies necessitates the use of large amplitude input signals [Morse and Ingard, 1986]. An angle of 10 degrees has been suggested as a suitable compromise [e.g. Kirkeby et al., 1998].

2.4 Crosstalk cancelled playback and reflections

When using loudspeakers for the playback of sound, the transmission from loudspeaker to the listeners' ears is influenced by the room in which the playback takes place.

Reflected and reverberant sound caused by surrounding surfaces add to the direct sound, as discussed in chapter 2.1.

Crosstalk cancelling filters for loudspeaker playback of binaural signals are usually designed for anechoic conditions. The reproduction may however take place under non-anechoic conditions. How this affects the playback is not clear. Previous research indicates that the performance obtained may be sensitive to reflections, dependent upon their strength and how early they arrive, but it seems that the effects of the room and its reflections have not been fully investigated.

It is commonly believed that anechoic or largely anechoic conditions are necessary for successful reproduction of crosstalk cancelled sound over loudspeakers. See e.g. Blauert [1997b, p 287, p 360]: *"[...] the subject must take a precisely determined position in a largely anechoic environment"*, *"[...] these processes function correctly only if the room in which the sound is played back has highly sound-absorbing walls [...]"*. But it has also been claimed that this is not critical: *"It is, in fact, a misunderstanding that an anechoic space need be used."* [Cooper and Bauck, 1989]. In this section, a review of earlier work relevant to this topic is given.

Atal and Schroeder are usually credited as the inventors of crosstalk cancelling [Atal and Schroeder, 1962]. Their technique is presented in several articles [Schroeder et al., 1962, Schroeder and Atal, 1963, Schroeder, 1969, 1970, 1973]. In all of these articles it is said that crosstalk cancellation works in free-field, and that playback should take place in anechoic chambers. (It has however been claimed that the main point of using anechoic listening spaces was to exclude the listening room reverberation that might otherwise have influenced their concert hall research results [Cooper and Bauck, 1989].)

Damaske [1971] tried playback in an anechoic room, and in rooms with reverberation times of 0.5 seconds and 1.4 seconds. Localisation in the anechoic room was good. In the reverberant rooms back to front reversals occured. For the room with the longer reverberation time half of the answers for backwards directions were mirrored to the front. Damaske also tested sideways displacement of the listener head, and concluded that to obtain good results it is more important that the listener is correctly placed than that the reverberation time in the reproduction room is short. For reproduction in reverberant rooms, he would however have preferred to use more directional loudspeakers.

Damaske and Mellert [1969/70] reported that in addition to experiments under free-field conditions, four persons had listened to crosstalk cancelled sound in rooms with reverberation times of 1.3 seconds and 3 seconds. Detailed results from this part were not given, but it was said that a somewhat correct imaging of directions might be achieved also in these cases. One gets the impression, however, that the results were not as good as for the anechoic case.

Møller [1989] reports on a digital crosstalk cancelling system: *"The system is shown to work in an anechoic room, but is not formally limited to this."* In [Møller, 1992], however, he says that *"Basically,* [crosstalk cancelling systems] *only work in a free field, which means that an anechoic room is needed."*

Nelson et al. [1992] cite the acoustic response of the listening room as one of three main sources of imperfections in playback of (ordinary) stereo sound, crosstalk and uneven loudspeaker responses being the two others. It was therefore found necessary to introduce inverse filters that would compensate for both the loudspeaker and the room responses, and also remove crosstalk. An adaptive crosstalk cancelling scheme was used. The experiments reported were undertaken in an anechoic chamber, but "*Other experiments were also undertaken under reverberant conditions and successful results were produced.*".

Experiments from inversion of the sound field in a car are reported by Kahana et al. [1999]. The inversion of a "free-field" response containing only the direct sound and the first few reflections was preferred to the inversion of a complete response from the car, which gave a highly coloured impression.

Playback of crosstalk cancelled sound with the stereo dipole system (closely spaced loudspeakers) is mentioned in [Kirkeby et al., 1999]: *"Under non-anechoic conditions, it is crucial that the direct-to-reverberant sound ratio is quite high. The system is generally quite sensitive to room reflections, particularly from the side."*

Institute of Sound and Vibration Research, Southampton, has produced an audio CD with examples of HRTF filtering and crosstalk cancellation [Nelson et al., publishing date not given]. It is intended to be played back also in non-anechoic spaces, but *"[...] it is recommended that you avoid placing the loudspeakers close to reflecting surfaces such as walls and heavy furniture."*

Cooper and Bauck [1989] claim that it is not necessary for playback to take place under anechoic conditions: *"The impulse response [...] is of short duration, which shows that the crosstalk cancellation is speedily completed, requiring the listening space to be anechoic for only the first few milliseconds."* Strong reflections arriving less than one to two milliseconds after the direct sound may degrade side imaging and cause front-back ambiguity, and should be avoided. Later reflections, however, will be perceived as reverberance and attributed to the recording. Even with early reflections present, the quality of the playback will still be *"excellent"* (compared to normal stereo). In a later article [Bauck and Cooper, 1996] they mention that room responses may be a problem in some applications, but maintain that *"[...] as long as strong reflections arriving somewhat sooner than 2-3ms after the direct sound are avoided, the room does not substantially spoil the imaging. Even moderately strong reflections in this epoch are often tolerable."*

Griesinger [1989] claims *"startlingly realistic"* results if, among other things, the room is non-reflective. Gierlich [1992] mentions an anechoic or semi-anechoic chamber as part of ideal conditions. Pulkki et al. [1999] give *"critical listening room conditions"* as one of two main limitations of loudspeaker based HRTF systems.

Begault [1994] claims that 3-D sound over loudspeakers is problematic, due to reflections in the listening space. Crosstalk is mentioned as a main problem, which can be solved by means of crosstalk cancellation *"If one can ignore the effects of the listening environment [...]"*. Gardner [1998] conducted listening experiments in a sound

studio with a 500Hz reverberation time of approximately 230 ms. Room equalization techniques were not used in his study.

Lopez et al. [1999] have investigated a crosstalk canceller in a room with a reverberation time of 200 milliseconds. Their results tell that the equalization zone is smaller in the reverberant room than in free field. The zone is also circular instead of elongated along the speaker/listener axis.

Foo et al. [1999] present an experiment where HRTFs recorded under anechoic conditions are used for crosstalk cancellation in a setup imitating real conditions in a workspace environment with reflecting surfaces. Loudspeakers were placed 70 centimeters from the center of the listening position. With the loudspeakers in front, localisation to the back was difficult to achieve, and with loudspeakers at the sides and back (90 degrees and 110 degrees) localisation to the front was difficult.

Takeuchi et al. [1997a] have specifically studied the effect of reflections on the performance of a crosstalk cancelling system. They simulated a single reflection from an infinite wall perpendicular to the interaural axis, and studied its effect both theoretically on crosstalk cancellation and subjectively on localisation performance. They concluded from their theoretical studies that a reflecting surface placed at the same side as the ear which is to receive a smaller signal causes more problems than the opposite case. They found that localisation based on time cues will probably not suffer from the reflection. However, even for small values of the reflection coefficient, large deviations in the frequency response of the transmission to the ears occured, severely degrading interaural level differences. Their subjective experiments showed that localisation performance was degraded for higher reflection coefficients. Overall, the ability to localise was retained in spite of the presence of a single reflection, and the performance was *"not degraded severely"*, suggesting that the system might work reasonably well in acoustically relatively dead environments.

Chapter 3

Implementation of crosstalk cancellation

This chapter presents and describes the development and implementation of the crosstalk cancellation system used in this work, a software system for the design of filters for crosstalk cancellation and for filtering sound with these filters.

The purpose of this work is to investigate the effects of reflections on crosstalk cancelled reproduction of binaural sound. To be able to do so, a means of obtaining crosstalk cancelled sound was needed. Such systems were commerically available at the time [e.g. Farina and Righini, 1997, Farina, Accessed 8th August 2000]. It was however found advantageous to design and implement such a system specifically for this task, as it was found that this would yield a more flexible system better suited to the needs at hand, while providing an opportunity to gain better insight into the problems and issues of the topic.

A system for the production of crosstalk-cancelled binaural sound was therefore developed, comprising software routines for the handling of signals and responses, design of single inverse filters, design of complete crosstalk cancelling filters, filtering with crosstalk cancelling filters and playback simulation.

Here, an initial overview is given, presenting design choices and the overall structure of the system. Inversion of transfer functions is discussed, and the algorithm chosen for this presented. Filter design and filtering of sound is presented in some detail, and an example of the software in use is given.

3.1 Overview

3.1.1 Requirements and choices

Based on the intended use of the crosstalk cancellation system, some requirements for its design and implementation were laid out. These requirements are discussed here, along with the choices made based on them.

First and foremost the system should be able to produce crosstalk cancelled sound for loudspeaker playback, in order to enable the presentation of binaural signals at the ears of a listener and give a faithful reproduction of binaurally recorded sound events.

Crosstalk cancellation was primarily a means, not a goal in itself, and perfection of crosstalk cancellation was not the purpose of this work. The crosstalk cancellation should perform at least so well that any effects due to imperfections in the implementation should be dominated by, or at least not mask, the effects of reflections and non-anechoic playback conditions.

The system was intended for situations where the listeners would be standing still in a given position. Thus head-tracking or other dynamic or adaptive features were not considered necessary.

The intended use of the system was for playback of previously recorded sound. Therefore, real-time capabilities were not considered necessary. This allows for simplifications. The filtering and the playback can be separated, and the filtering can be done on ordinary computers instead of dedicated digital signal processors (DSPs). Furthermore, as there are no absolute requirements on how fast the filtering has to be, more complex filtering and longer filters can be applied.

Matlab was chosen as the implementation language. Being a high level language allowing for quick and efficient formulation of complex algorithms and well suited for signal processing, it was well suited for the task at hand.

As the central point of the work was to study reflections, it was apparent that non-symmetrical setups might be encountered. This ruled out the possibility of using the "shuffler" filter structure proposed by Cooper and Bauck [1989]. An implementation along the lines given by Møller [1989], following equation 2.3 and figure 2.4 (in this work) was therefore chosen.

The system was intended for setups where the transfer functions would be combinations of head-related transfer functions (HRTFs) and room transfer functions (RTFs) with one or a few single reflections. It might be of interest to also cancel the reflections. These responses may be of considerable length, and filters of similar lengths are needed in order to invert them to handle the reflections properly. Based on this, it was decided that all filters should be FIR-filters. While IIR-filters do in general give higher performance for a given computational load [Proakis and Manolakis, 1992, Huopaniemi et al., 1999], they were not considered necessary for this situation, as FIR-filters of the actual lengths were expected to give satisfactory performance.

3.1.2 Data structures and data flow

The crosstalk cancellation process may be divided into two main parts; design of crosstalk cancelling filters, and filtering of sound with crosstalk cancelling filters. For these and associated purposes, the software developed defines routines and data structures for the following topics:

- Binaural impulse response data sets
- Inversion of transfer functions
- Crosstalk cancelling filter data sets
- Filtering of sound
- Playback simulation

A listing of the main routines and data structures involved is given in appendix A. The general structure of the process is illustrated in figure 3.1.

The filter design process may be described in terms of its input and output data. The design deals with producing the latter from the former in such a way that the goal, proper crosstalk cancellation, will be achieved.

The necessary input data is a description of the setup for which the crosstalk cancelling filter is to be designed. This information is gathered in a data structure called a *response structure*. It contains information of what kind of setup this is, the *binaural impulse responses* associated with the setup, and additional information such as date, identification and comments. In this work, only setups with two loudspeakers and one listener have been considered. The definition of the data structure, taking a general approach, also allows for other kinds of setups. For the two loudspeaker/one listener case, the response may be either symmetrical or non-symmetrical. For the non-symmetrical case, which has been used exclusively in this work, the impulse responses corresponding to the four transfer functions shown in figure 2.3 are included. For symmetrical cases, there are only two distinct transfer functions, and these two are included.

The output data is called a *filter structure*. Similarly to the response structure, it includes identifying information, both for itself and for the response structure it is based upon. Also here, there are possibilities for various kinds of setups, and for symmetrical and non-symmetrical two-loudspeaker/one-listener setups. The *filters* included are inverses of responses and combinations of responses, the interaural transfer functions, according to equations 2.3 and 3.1, and extra information needed to apply these.

The filter design process is further described in section 3.3. Applying the filters for filtering of sound is described in section 3.4.

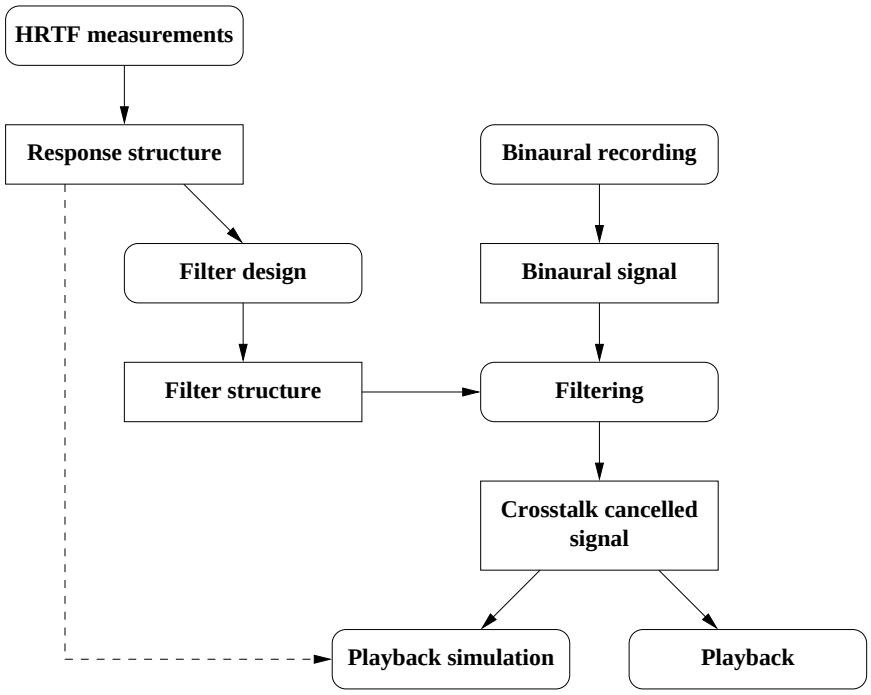


Figure 3.1: Design of crosstalk cancelling filters and crosstalk cancelling filtering of binaural signals schematically pictured. Boxes with rounded corners symbolise actions, while boxes with square corners symbolise data. The arrows show how the actions and data are connected. The arrow from the response structure to playback simulation is dashed to indicate that another response structure (for another setup) may be substituted.
The flow shown is one-way only. In practice, there is also feedback in the design process, e.g. from playback simulation to filter design.

3.2 Inversion of transfer functions

In order to implement crosstalk cancellation as given in equation 2.3 and figure 2.4, several transfer functions and combinations of transfer functions have to be inverted. This can be more easily seen by rewriting equation 2.3 as

$$\begin{aligned} L_l &= \left(B_l - B_r H_{rl} H_{rr}^{-1}\right) H_{ll}^{-1} \left(1 - H_{lr} H_{rl} H_{ll}^{-1} H_{rr}^{-1}\right)^{-1} \\ L_r &= \left(B_r - B_l H_{lr} H_{ll}^{-1}\right) H_{rr}^{-1} \left(1 - H_{lr} H_{rl} H_{ll}^{-1} H_{rr}^{-1}\right)^{-1} \end{aligned} \tag{3.1}$$

As said, in our case the transfer functions H_{xx} are combined HRTFs and RTFs, as they may contain reflections in addition to the effects caused by the listener being present. Consequently, they may not be considered minimum phase (chapters 2.1 and 2.3.2).

A transfer function of an linear, time-invariant (LTI) system is said to be minimum phase if it has the minimum net phase change of all possible transfer functions having the same amplitude frequency response [Proakis and Manolakis, 1992]. This translates into a requirement on the placement of the poles and zeros of the transfer function, and is closely connected to its invertibility. A discrete-time transfer function in the z-domain is minimum phase if all of its zeros are placed inside the unit circle. A stable and minimum phase transfer function has a causal, stable and minimum phase inverse. Transfer functions that are not minimum phase have not. Inverses may be found, but they will be either unstable, which is not very useful, or non-causal.

Transfer functions that are not minimum phase may be divided into a minimum phase part and a maximum phase part, or into a minimum phase part and an all-pass part [Mourjopoulos, 1994, Proakis and Manolakis, 1992]. In the first case, there is the choice between a causal and unstable or an anticausal and stable inverse for the maximum phase part. In the latter case, the minimum phase part may be inverted, giving a total inverse response that correctly equalises the amplitude, but smears the energy out in time. Due to our intent to cancel out reflections, this is not a suitable approach here.

For the playback of stored sound, non-causality of the inverse may be circumvented by allowing an extra modeling delay in the total response of transfer function and inverse [e.g. Clarkson et al., 1985]. This is equivalent to shifting the inverse in time, after truncating its anticausal part. The delay added results in additional linear phase, and does not have any degrading effects on the signal.

Instead of inverting the transfer functions directly, it has been shown that better results can be achieved using least squares methods [Mourjopoulos, 1994, Neely and Allen, 1979]. The method chosen in this work is to compute approximate inverses as Wiener FIR filters [Proakis and Manolakis, 1992]. These are filters that minimise the squared error with respect to a desired total response. The computation of the filter is done in the time domain, and the result is FIR filters, as wanted.

Given a system with an impulse-response $h(n), n \geq 0$, we wish to find a (FIR) filter $b(k), k \in 0 - M$ such that $y(n)$, the convolution of $h(n)$ with $b(k)$, approximates a desired total impulse-response $d(n)$ as closely as possible. The coefficients of $b(k)$ are found by minimising the squared error between $y(n)$ and the desired response $d(n)$.

The expression for the total squared error E is:

$$E = \sum_{n=-\infty}^{\infty} [d(n) - y(n)]^2$$

$$
\begin{aligned}
E &= \sum_{n=-\infty}^{\infty}\left[d(n)-\sum_{k=0}^{M}b(k)h(n-k)\right]^2 \\
E &= \sum_{n=-\infty}^{\infty}d^2(n)-\sum_{n=-\infty}^{\infty}2d(n)\sum_{k=0}^{M}b(k)h(n-k) \\
&\quad +\sum_{n=-\infty}^{\infty}\left[\sum_{k=0}^{M}b(k)h(n-k)\sum_{l=0}^{M}b(l)h(n-l)\right] \\
E &= \sum_{n=-\infty}^{\infty}d^2(n)-2\sum_{k=0}^{M}b(k)\sum_{n=-\infty}^{\infty}d(n)h(n-k) \\
&\quad +\sum_{k=0}^{M}\sum_{l=0}^{M}b(k)b(l)\sum_{n=-\infty}^{\infty}h(n-k)h(n-l) \\
E &= E_d-2\sum_{k=0}^{M}b(k)r_{dh}(k)+\sum_{k=0}^{M}\sum_{l=0}^{M}b(k)b(l)r_{hh}(k-l) \\
E &= E_d-2\vec{b}^T\vec{r}_{dh}+\vec{b}^T\mathbf{r_{hh}}\vec{b} \qquad (3.2)
\end{aligned}
$$

Here $\vec{b}$ is a column vector of filter coefficients, $\mathbf{r_{hh}}$ is the auto-correlation matrix of the system we want to invert, and $\vec{r}_{dh}$ is the correlation between b and d.

To minimise the total error we differentiate 3.2 with respect to $\vec{b}$, equal the result to zero, and obtain:

$$
\begin{aligned}
\nabla E=-2\vec{r}_{dh}+2\mathbf{r_{hh}}\vec{b} &= \vec{0} \\
\mathbf{r_{hh}}\vec{b} &= \vec{r}_{dh} \\
\vec{b} &= \mathbf{r_{hh}^{-1}}\vec{r}_{dh} \qquad (3.3)
\end{aligned}
$$

This result (eq. 3.3) tells us that the best (in the squared error sense, and for this filter length, M) filter $b(k)$ is found by multiplying the inverse of the autocorrelation matrix of the system we want to invert with the correlation vector between the impulse response of the system and the desired total impulse response.

This may be used to approximate any desired total impulse response $d(n)$. A common special case is to compute $b(k)$ as an approximate inverse of $h(k)$. This is done by choosing $d(n)$ as a delta-function (unit sample function). In this case $\vec{r}_{dh}$ reduces to a vector with $h(0)$ in the first element and remaining elements zero.

If $h(k)$ should happen to be non-minimum phase, a better approximation of the inverse may be achieved by allowing a delay in the desired total response, letting $d(n)=\delta(n-delay)$. In this case, $\vec{r}_{dh}$ is composed of the reversed first part of $h(n)$, the length of this part given from $delay$, followed by zeros.

3.3 Filter-design

The design of crosstalk cancelling filters starts from equation 3.1, and consists of computing inverses for the transfer functions H_{ll} and H_{rr}, computing the interaural transfer functions $H_{lr}H_{ll}^{-1}$ and $H_{rl}H_{rr}^{-1}$, and the expression $(1 - H_{lr}H_{rl}H_{ll}^{-1}H_{rr}^{-1})^{-1}$. For this, the algorithm of equation 3.3 is used, using a unit impulse, possibly delayed, as the desired total response. Before this can be done, two parameters, the length of the inverse filter and the modeling delay to apply, must be chosen. Routines have been written to aid in exploring the effects of these parameters in each case.

The approach taken to find suitable values for the length of the inverse and the modeling delay is to compute a series of inverses for a set of combinations of possible filter lengths and delays. For each inverse, the convolution of the proposed inverse and the response to be inverted is found. An error sum is then computed as the deviation from the desired target response. These results, combined with the user's experience and knowledge of the setup at hand, may then be used to choose values that will give satisfactory performance.

The summed squared difference between the desired total response and the obtained total response was chosen as the error metric used as an aid to distinguish between the various possible inverses. Inverses of different lengths were to be compared to each other, and this posed the question of what would give the most correct comparison, total squared error, or squared error normalised to the filter length. The latter was believed to be more suitable, and was chosen. As a test, both alternatives were tried. The results are shown in figure 3.2. As can be seen, there is a difference, but it does not influence the relative ranking of the various combinations, and therefore does not have any practical consequences.

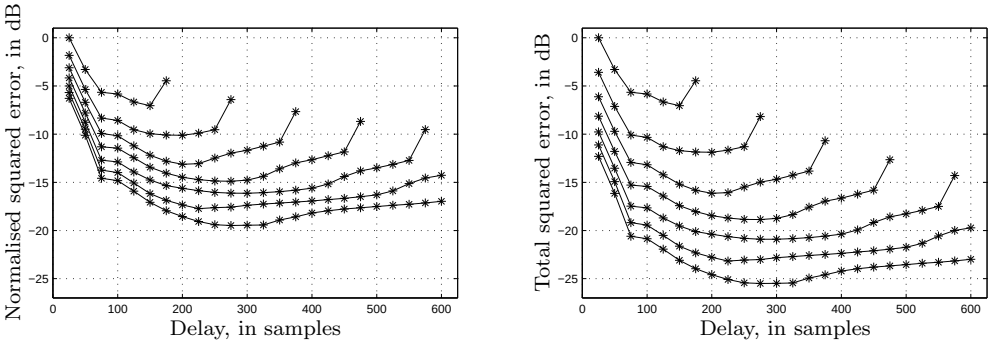


Figure 3.2: Summed squared error as a function of delay, for filters of different lengths. Left: Squared error normalized to filter length. Right: Total squared error. In both cases, the error is normalised to the maximum error. Filter lengths from 100 to 800 samples, delay from 25 samples up to the length of the filter, maximum 600 samples. (Longer filters towards the bottom of the figure.)

Figure 3.2 demonstrates a general trend, that longer filters give better results. For each filter length, there is an optimal delay, that gives the lowest error. This optimal

delay typically increases with increasing filter length, a typical value being somewhat less than half the length of the filter for the longer filter lengths tried in this work.

3.3.1 An example

To demonstrate the filter-design, an example is presented. The task is to design filters for crosstalk-cancellation for a loudspeaker setup in an anechoic room. The setup is two high-quality loudspeakers placed close together at two meters distance in front of the listener.

The starting point is a set of impulse responses measured in the actual setup. The early part of these responses are shown in figure 3.3.

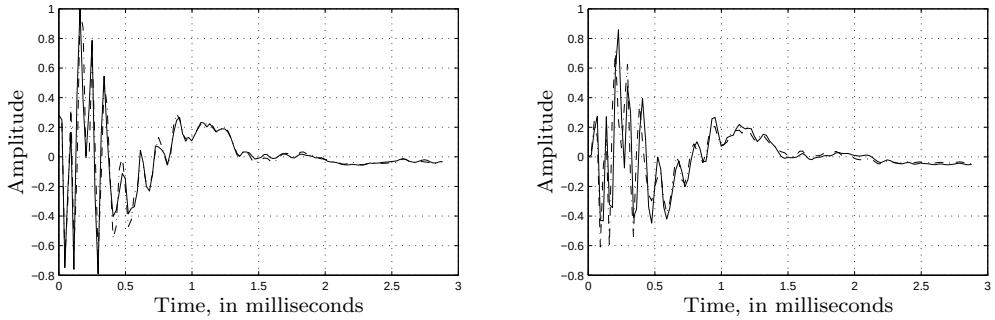


Figure 3.3: Impulse responses for setup in anechoic room. Left: Responses from loudspeakers to ear at same side, H_{ll} and H_{rr}. Right: Responses from loudspeakers to ears at opposite side, H_{lr} and H_{rl} (the crosstalk). Responses for left loudspeaker drawn with solid lines, responses for right loudspeaker drawn with dashed lines.

The first task is to compute inverses for the transfer functions from the loudspeakers to the ears at the same side as the loudspeaker, the responses called H_{ll} and H_{rr}. This involves choosing the length of the inverse filters and the delay they should introduce. It is practical to have equal filter lengths for these two. To keep the left and right channels in the crosstalk cancelled sound synchronised, they should also introduce the same delay. As can be seen from figure 3.3 this setup is very close to symmetric, with very similar responses for the two sides, so these requirements can easily be fulfilled. The fact that the responses are so alike also means that it is sufficient to investigate one of them to find parameters for computing suitable inverses. It is randomly chosen to use H_{ll} for this.

Based on the plot of the responses, knowledge of the setup and experience, a set of parameters is chosen with filter lengths from 100 to 800 taps in steps of 100, and modeling delays from 25 to 600 samples, in steps of 25. Inverses for the combinations of these parameters are then computed. The resulting error energies are those shown to the left in figure 3.2. This plot is a starting point for choosing the filter length

and modeling delay for the inverse. For this example, an inverse filter with a length of 500 taps and a delay of 275 samples is chosen as a trade-off between quality and computational load. To evaluate the inverse filter, the total response of the convolution of the response and the inverse is plotted, as shown in figure 3.4. Using the same parameters for filter length and modelling delay a corresponding inverse for H_{rr} is also computed.

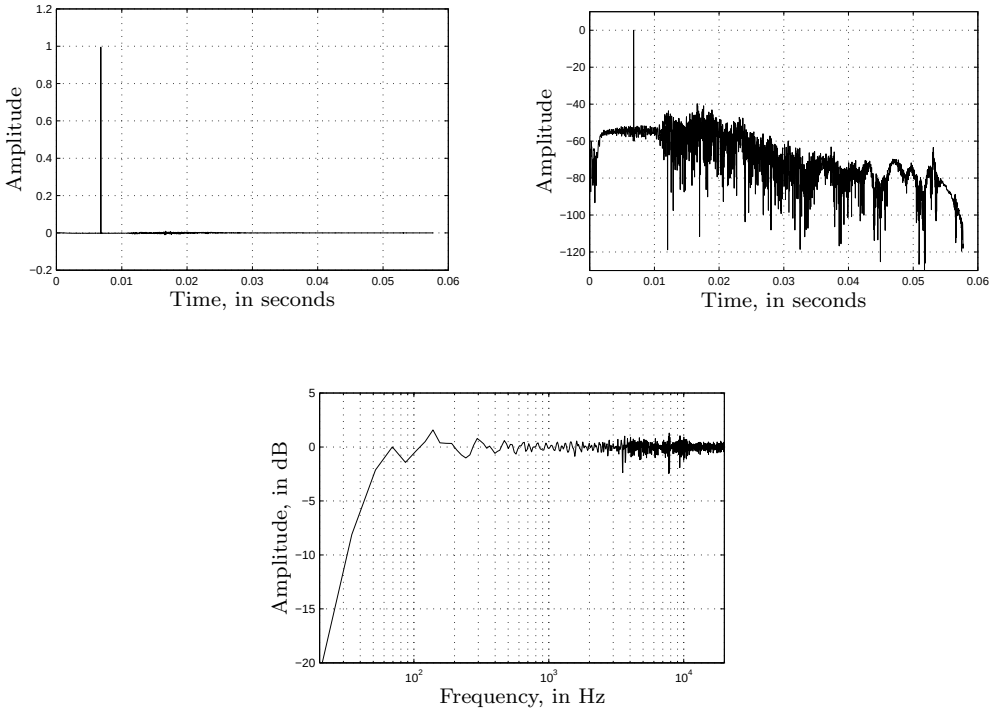


Figure 3.4: Total response of H_{ll} convolved with its approximate inverse. Upper left: Time domain, linear amplitude. Upper right: Time domain, amplitude in dB. The response is an impulse, as intended, but there is also an "error floor" due to the fact that the inverse is only approximate. Lower: Frequency domain amplitude. The target function is an impulse in the time domain, corresponding to a constant amplitude of 0dB in the frequency domain. As can be seen, the response has a high-pass characteristic. A longer inverse filter would have given a lower cut-off frequency and a flatter frequency response.

Next, the interaural transfer functions are computed. These are given as $Hia_l = H_{lr}H_{ll}^{-1}$ and $Hia_r = H_{rl}H_{rr}^{-1}$. They are computed by convolving the response from loudspeaker to the ear at the opposite side with the inverse (which we have just computed) of the response from (the same) loudspeaker to the ear at the same side. These interaural functions get an inherent delay equal to the one in the inverses, which has to be stored as it will be used later. The interaural transfer functions may often

be truncated after inspection, as the tails go quickly towards zero. This truncation is a tradeoff of accuracy for a decrease in computation time. In this case, the responses are truncated to 1024 samples.

The factor $1 - Hia_l Hia_r$ must then be computed. Here, the "1" must be delayed twice the inherent delay of the interaural transfer functions for a correct subtraction. The resulting response may, after inspection, be truncated both at the start and at the end. There are zero-samples at the start, due to the delay inherent in Hia_l and Hia_r, and the tail of the response goes towards zero. In this example, 550 samples at the start of the compound response are cut, and the next 800 samples kept. Lastly, this factor has to be inverted, which is done as for the other inverses. The parameters used for the inversion are a filter length of 600 samples and a delay of 100 samples.

To verify the results, an artificial test signal, consisting of a positive impulse in the left channel and a negative impulse in the right channel is used. The signals are shown to the left in figure 3.5. The filter-designed above is applied to this signal, producing the loudspeaker signals shown to the right in figure 3.5. As can be seen from the figure, both loudspeaker channels contribute to both the left channel test signal and the right channel test signal.

To verify the proper functioning of the filter (and the design), playback in the setup for which the filter is designed is simulated. The resulting ear signals are shown in figures 3.6 and 3.7. As can be seen from these figures, the signals are indeed as expected in the time domain, like the original test signal, but delayed. The "error floor" due to the approximations and simplifications done is at -40dB relative to the level of the impulse.

As the test signals are impulses in the time domain, the amplitude of the ear signals should ideally be constant 0dB in the frequency domain. The major deviations from this are found below 300Hz, and are due to the quite short filters and the heavy-handed truncation of filter responses used for this example.

3.4 Filtering of sound

For the filtering of sound, a binaural recording and a crosstalk cancelling filter are needed. The filtering implements equations 2.3/3.1, using the filters and responses from the crosstalk cancelling filter (section 3.1.2).

Short pieces of sound may be imported into Matlab and filtered as a whole. But this is impractical for sound-files of practical length, due to the sheer amount of data involved. Stereo sound sampled at 44.1kHz with 16 bit samples has a data rate of 176.4 kilobytes per second. Matlab uses eight-byte floats[1] as its internal representation, giving a data rate of four times this, 705.6 kilobytes per second. This rapidly leads to exhaustion of computer memory, causing the filtering operation to slow down considerably and effectively prohibiting any kind of useful processing on such files as a whole.

[1] The *long* format specified by the IEEE floating point standard [Mat, 1996]

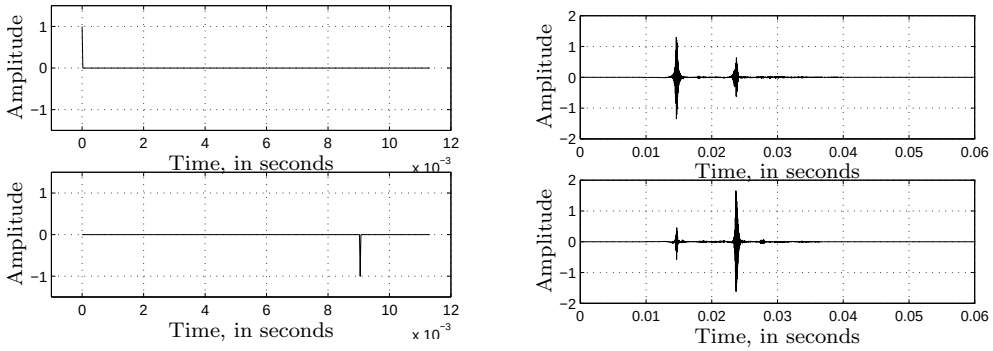


Figure 3.5: Left: Test signal, what is to be produced at the listener's ears. Right: Loudspeaker signals, produced by filtering the test signal with the crosstalk cancelling filter. Left channel signals in the upper row, right channel signals in the lower row. Note that the time scales for the two plots differ.

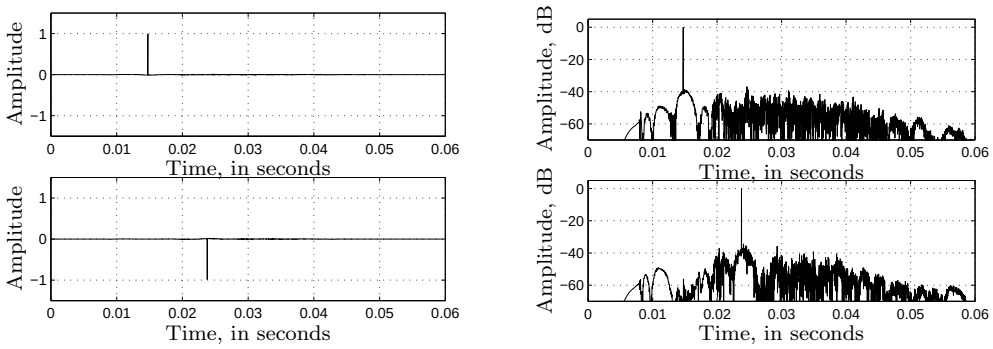


Figure 3.6: Ear signals, produced by simulating playback of the loudspeaker signals shown in figure 3.5. Left: Linear amplitude. Right: Amplitude in dB. Left channel signals in the upper row, right channel signals in the lower row. Time scale as for the right plot in figure 3.5.

To avoid this problem, larger sound files are read into Matlab in smaller parts, each part being filtered and written out before the next part is read. Care is taken to handle correctly the overlaps between parts that results from the filtering. To illustrate the performance difference between these two methods, it can be mentioned that an attempt to filter a sound file of about two and a half minutes length as a whole took more than three hours on a given computer (a 200MHz PC with 48MB of internal memory), while filtering it in parts took about ten minutes.

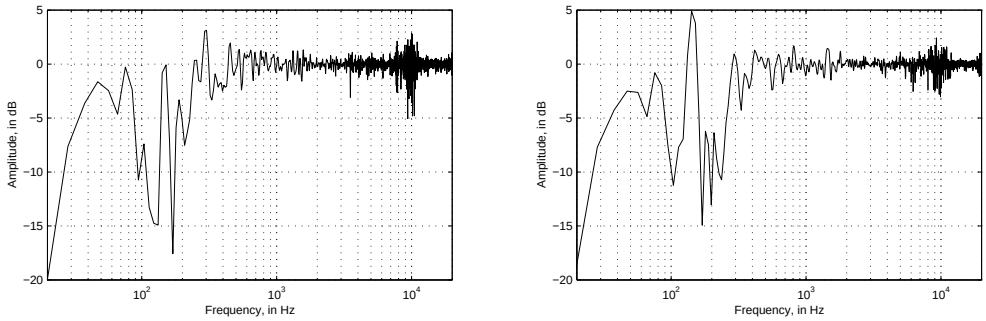


Figure 3.7: The ear signals of figure 3.6, shown in the frequency domain. Left: Left channel amplitude. Right: Right channel amplitude.

Using a PC with a 166MHz Pentium processor and 64MB of internal memory, a test run where a 135 second sound file was filtered with a typical filter took 742 seconds (when filtered in parts). On current computers (e.g. Pentium III, 700Mhz), the filtering operation typically takes less time than playback of the file. Real-time operation should therefore be possible.

As FIR filters are being used, all filterings are conceptually convolutions. This is sped up by the implementation of an overlap-add algorithm doing the filtering in the frequency domain [Proakis and Manolakis, 1992]. (Matlab does have a similar function, *fftfilt()*, but this one can not be used, as it does not provide for the overlap necessary when concatenating the results of such filtering operations.) In a test run, convolution of a signal of length 65024 samples with a filter of length 2048 taps took 26 seconds using the Matlab *conv()* function. Using the overlap-add function implemented, it took a little more than 2 seconds, an improvement with a factor of more than 10.

Doing the filtering on the computer instead of doing it in real-time on a DSP has one major drawback. It makes it nearly impossible to measure through the system to assess and verify its performance in the actual setup. No solution was found for this problem.

3.5 Evaluation of the implementation

The crosstalk cancellation system was evaluated using listening tests. Binaural recordings were reproduced and compared to the original situations. Listening to recordings of moving and talking persons, sound could be localised to all directions, both in front and behind. Distance hearing also worked well, with the difference between

close sources and sources further away being clearly evident. Listening to recordings of several groups of persons talking, each group could be localised separately. The conversations could easily be kept apart, and any of them could be followed upon will. Similarly, listening to recordings of three persons talking, the position of any and each of them could easily be perceived, and any of the three different messages be singled out. All the recordings gave auditory impressions that corresponded well with the original situations.

The "sweet spot" was found to be quite large, allowing for some movement without destroying the function altogether. The listener's head could be moved a couple of decimeters sideways, from knee-height to head-height vertically, and along the whole axis between the loudspeakers, while still receiving a sound image with some of the three-dimensionality retained.

It was concluded that the system performed the intended function well and produced three-dimensional sound of good quality, and that it could be used for the experimental work regarding the effects of reflections.

Chapter 4

Crosstalk cancelling with a single reflection

In this chapter, the effect of a single reflection on crosstalk cancelled playback of binaural sound is studied. This is done theoretically, using numerical computations and simulations.

It is assumed that the crosstalk cancelling filtering is designed for anechoic playback, but that the playback takes place with a reflection present.

A simplified model of the playback situation is used, as shown in figure 4.1. The two loudspeakers are shown as the sources SL and SR, and the listener is modeled as two receivers, RL and RR, representing the ears. The setup is symmetrical, with the sources and the receivers placed at distances dR and dS to the sides of the symmetry line. The wall is placed at a distance dW to one side of the symmetry line. The distance from the line connecting the sources to the line connecting the receivers is d.

Three sets of transfer functions are shown in the figure: For the direct sound from the sources to the receivers, for the reflected sound from the sources to the left receiver, and for the reflected sound from the sources to the right receiver. These are called HS, HO, HLL, HLR, HRL and HRR. (Note that these are not the same as the similarly named transfer functions in chapter 2.3.4.)

4.1 Timing of reflections and crosstalk cancellation

Crosstalk cancellation requires accurate timing of the signals that should cancel each other. It is therefore of interest to investigate the timing of the reflected signals.

The expressions for the path lengths for the various transmission routes are given in table 4.1. Substituting values corresponding to setups used in this work, $d = 2.0$m,

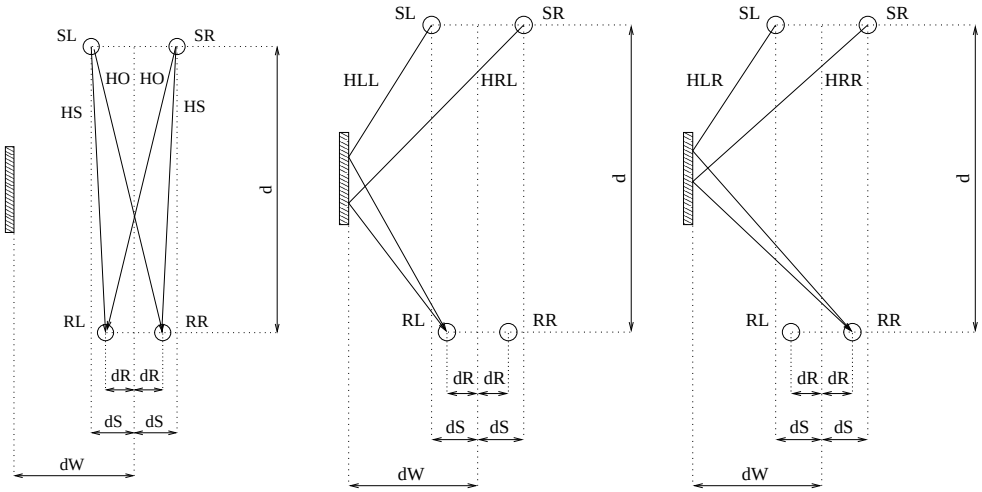


Figure 4.1: Playback with reflection, simplified. Two sources, two receivers and a reflecting wall. Sound transmission for direct sound is shown in the left figure, for reflected sound to the left ear in the middle, and for reflected sound to the right ear in the figure to the right.

Table 4.1: Expressions for path lengths

d_{HO}, d_{HS}	=	$\sqrt{d^2 + (d_S \pm d_R)^2}$
d_{HRL}, d_{HLL}	=	$\sqrt{d^2 + (2d_W \pm d_S - d_R)^2}$
d_{HRR}, d_{HLR}	=	$\sqrt{d^2 + (2d_W \pm d_S + d_R)^2}$

$d_W = 1.5$m, $d_S = 0.21$m and $d_R = 0.085$m (typical radius of human head), the results shown in table 4.2 are obtained.

As can be seen from table 4.2, the difference in path length for the direct and reflected sound is about 1.7 meters. This corresponds to a difference in arrival times of about five milliseconds. With respect to the precedence effect, this is well outside of the domain for summing localisation, but not so large a difference that the reflected sound should be heard separately (section 2.2.3).

Signals prefiltered with a crosstalk cancelling filter will have a timing that produces cancellation for the direct sound. I.e., they are designed for a *path length difference* for the paths from the two loudspeakers to one ear corresponding to the direct trans-

Table 4.2: Path lengths and path length differences

Ear	Route	Path	Length	Difference	Difference change
Left	Direct	d_{HS}	2.0039 m	- 0.018 m	- 0.328 m
		d_{HO}	2.0216 m		
	Reflected	d_{HLL}	3.3641 m	- 0.346 m	
		d_{HRL}	3.7102 m		
Right	Direct	d_{HO}	2.0216 m	0.018 m	- 0.370 m
		d_{HS}	2.0039 m		
	Reflected	d_{HLR}	3.5022 m	- 0.352 m	
		d_{HRR}	3.8545 m		

mission, HS and HO. This path length difference may be different for the reflected sound. If it is, the signals that were designed to cancel each other will arrive at different times, and cancellation will not occur for the reflected sound. (Note that it is not the path lengths for the direct sound that are compared to the path lengths for reflected sound here. The *differences* in path length for sound from the two loudspeakers to one ear are computed, and these differences for direct and reflected sound are compared.)

As can be seen from table 4.2, the distance differences for the reflected sound differ from those for the direct sound with 33 and 37 cm. This corresponds to an arrival time error of about one millisecond. This timing error is so large that the reflected loudspeaker signals will not cancel each other as they will for the direct sound. It must therefore be concluded that the reflected parts of the sound will not combine in a way that gives rise to crosstalk cancellation.

4.2 Unnatural reflections

Binaural reproduction may be viewed as a way of creating virtual sound sources. These sources may be placed anywhere with respect to the listener. But, given a setup like the one used here (figure 4.1), there is, for most placements of a virtual source, no natural correspondence between the position of the source and the sound of the reflection. I.e., the reflection that arises from the playback of the binaural sound is different from the reflection that would have been caused by a real source in the position of the virtual source.

First, a real source placed in positions around the listener will in many cases not give rise to reflections that can be heard by the listener at all, at least so for specular reflections. But the loudspeakers producing the signals corresponding to the virtual source will always cause reflections that will reach the listener. Second, even for virtual sources in positions that might have given reflections for real sources, there is not necessarily any natural correspondence between the reflection and the position of the source.

For the setup above, and for specular reflection, the only positions where a real source would have caused reflections are those in the area from the wall and towards the front right, centered on the mirror line of the line from the receivers to the wall with respect to a line perpendicular to the wall. For virtual sources positioned in this sector, the reflections may sound natural, as it is at least geometrically plausible that the source should cause a reflection. Real sources outside of this area, e.g. in the back half plane would have given no reflection at all, whereas virtual sources in these positions would.

But even for a virtual source placed so that a reflection from the wall is plausible, the reflection will not equal the reflection caused by a real source in this position. There are *two* actual sources causing the reflection, instead of one, so there will be two sets of reflections, instead of one. Further, the arrival times for the reflections will not be correct. The reflections will come at fixed delays with respect to the direct sound, while for a real source the delays would vary with the source position. Also the reflected signals do not correspond to those caused by a real source, as the binaural signals are designed to be played back with a given crosstalk, which the transmission via the reflecting wall does not give.

4.3 A source in the wall

The reflected sound may be compared to the sound coming from a source placed at the position of the reflecting wall. To accomplish this, a source SW is added to the model (figure 4.2). The transfer functions from this source to the receivers are called HWL and HWR.

For a source placed at the position of the wall, the path lengths would be as shown in table 4.3. The differences computed here are *interaural* differences, i.e. differences in the transmission from one source to the two receivers. (The differences in section 4.1 was for transmission from two sources to one ear.) The path length values for the reflected sound are filled in from table 4.2.

As can be seen from table 4.3, the interaural time difference (ITD) for the reflected sound is within three percent of the ITD for sound from the source in the wall. The 10 microseconds difference is small compared to the maximum physical ITD value of 600 to 800 microseconds (section 2.2.2). So the ITDs of the reflected sound from the loudspeakers correspond well to those for a source in the wall. It is therefore possible that a listener might perceive the reflected sound as secondary source at the position of the wall, possibly radiating some distorted version of the signal from the virtual

source. (But due to the precedence effect, it is probable that it will not be heard as a separate source (section 2.2.3)).

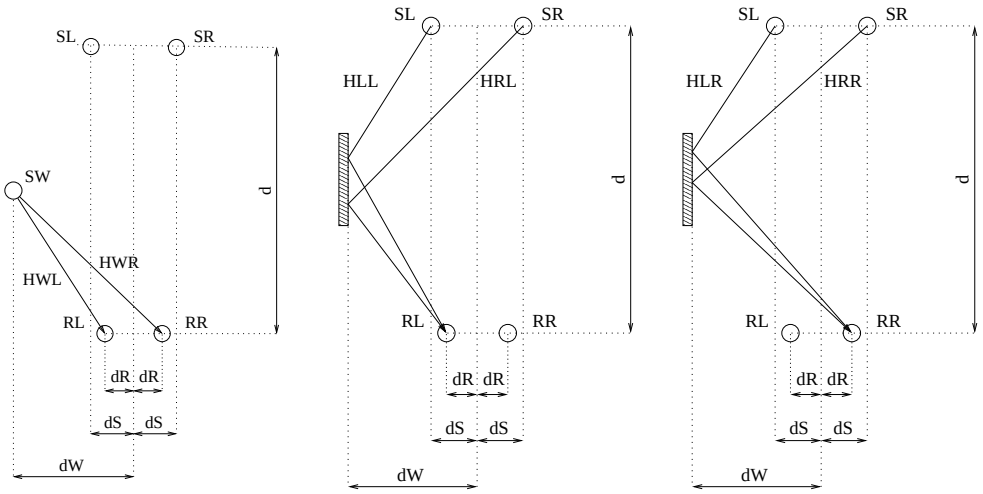


Figure 4.2: Model for the comparison of reflected sound to the sound from a source placed at the position of the reflecting wall. Sound transmission from the source at the wall position to the ears is shown in the left figure, for reflected sound to the left ear in the middle, and for reflected sound to the right ear in the right figure.

Table 4.3: Interaural path length differences and time differences (ITD) for a source placed at the position of the wall and for reflected sound from the actual sources.

	Path	Length	Difference	ITD
Wall to left ear	d_{HWL}	1.7327 m	- 0.1414 m	- 0.41 ms
Wall to right ear	d_{HWR}	1.8741 m		
Left loudspeaker	d_{HLL}	3.3641 m	- 0.1381 m	- 0.40 ms
	d_{HLR}	3.5022 m		
Right loudspeaker	d_{HRL}	3.7102 m	- 0.1443 m	- 0.42 ms
	d_{HRR}	3.8545 m		

4.4 Simulated playback with reflection

For the simplified model in figure 4.1 it is further assumed that all transfer functions are pure delays. The path length of the reflected sound is nearly twice that of the direct sound, corresponding to a relative level loss of about 6 dB. The reflection coefficient of the wall is taken to be 0.7 (-3dB), giving a total level of the reflected sound at the receivers that is 9 dB below that of the direct sound. From these assumptions, a crosstalk cancelling scheme for the anechoic case (without the reflecting wall present) and playback for anechoic conditions and for the setup with the reflections present may be simulated. For this, the crosstalk implementation presented in chapter 3 is used. Due to the gross simplifications done in the model, the details of the results will not be valid, but a representative insight into what happens can be obtained.

In figure 4.3, the signals to be presented at the ears and the corresponding loudspeaker signals resulting from the crosstalk cancelling filtering are shown. The ear signals resulting from anechoic playback, for which the filters were designed, are shown in figure 4.4, and the ear signals resulting from playback with the reflections from the wall present are shown in figure 4.5. The same signals, transformed to the frequency domain, are shown in figure 4.6.

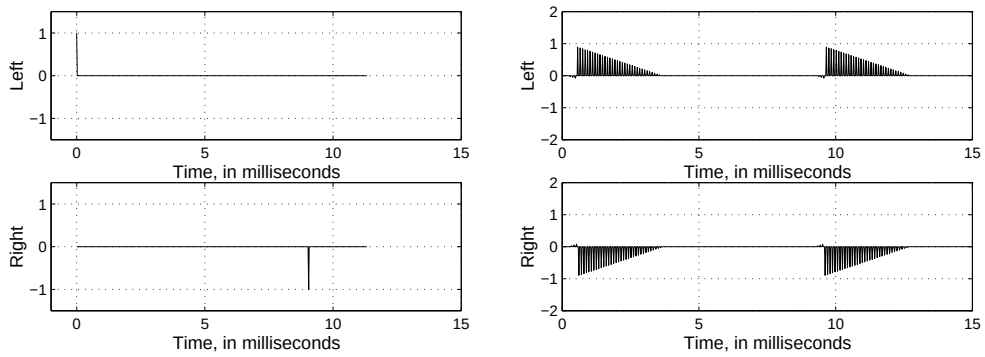


Figure 4.3: Simulated binaural reproduction. Left: Input signals, intended for presentation at the ears. Right: Loudspeakers signals, obtained by filtering the input signals with a crosstalk cancelling filter.

The input signal chosen consists of two unit impulses, one for each ear, with different signs and at different times. As can be seen from figure 4.4, the design for crosstalk cancellation produces loudspeaker signals where the left and right channels contribute to produce the correct left ear and right ear signals. When the left loudspeaker sends a positive pulse intended for the left ear, the right loudspeaker must send a negative and somewhat weaker pulse to cancel the first pulse at the right ear. This must again be cancelled at the left ear by sending a positive pulse from the left loudspeaker, and so on. In this case, the loudspeaker signals have a very regular structure, with gradually decreasing impulses. This is due to the simplification done, where the head-related transfer functions are replaced with pure delays.

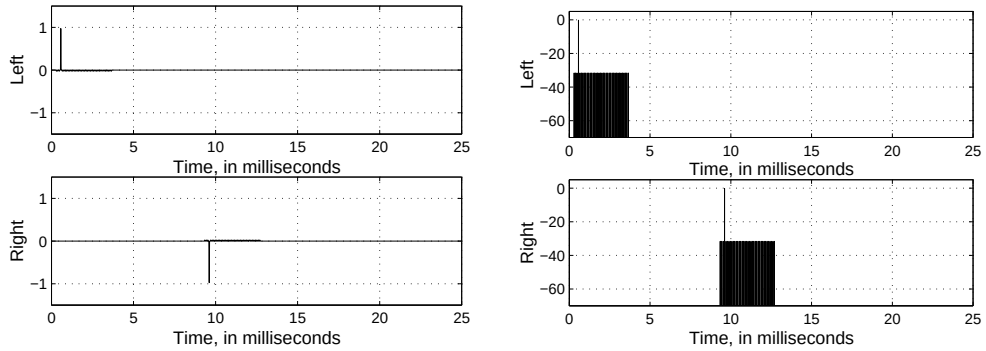


Figure 4.4: Simulated binaural reproduction: Playback under anechoic conditions. Left: Ear signals, with linear amplitude axis. Right: Ear signals, with logarithmic amplitude axis.

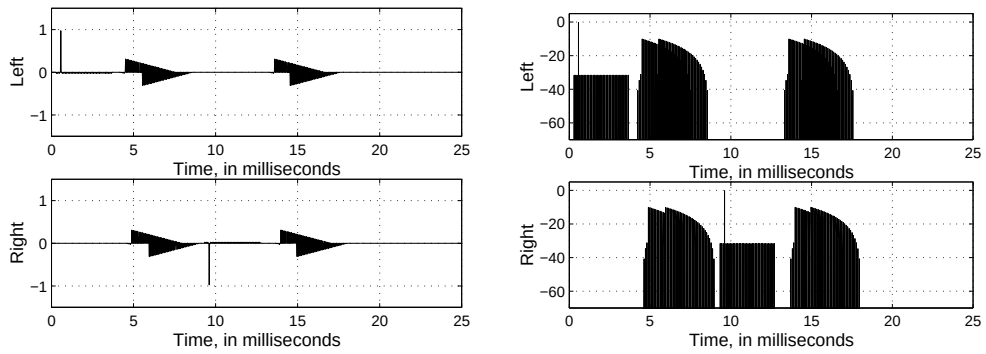


Figure 4.5: Simulated binaural reproduction: Playback with reflections. Left: Ear signals, with linear amplitude axis. Right: Ear signals, with logarithmic amplitude axis.

For the playback under anechoic conditions, for which the crosstalk cancelling was designed, the crosstalk is below -70 dB (figure 4.4). The filtering produces a "noise floor" at -30 dB which clearly shows the length of the total response of the crosstalk cancelling filter. For the playback with reflections present, the situation is quite different (figure 4.5). The crosstalk cancelling of the direct sound still works, as it should, with crosstalk below -70 dB. The "noise floor" is also still present. But replicas of the left and right loudspeaker signals show up in the ear signals, due to the reflection. The reflection of the right loudspeaker signal arrives about a millisecond later than the reflection of the left loudspeaker signal, due to the different path lengths.

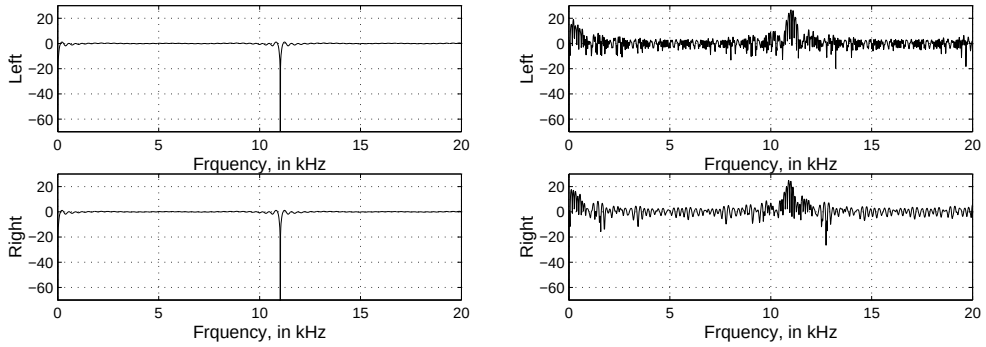


Figure 4.6: Simulated binaural reproduction: Ear signals in the frequency domain. Left: Amplitude of ear signal for anechoic playback, in dB. Right: Amplitude of ear signal for playback with reflections, in dB.

For the same reason, the reflections arrive a little later at the right ear than at the left ear. Due to the length of the impulse responses of the crosstalk cancelling filters used, the reflected parts of the sound contribute a large part of the energy delivered at the ears.

As can be seen, there is no cancellation for the reflections, so they arrive at the ears at their -9 dB level. For a different input signal, for example a -10 dB impulse in the right channel, delayed by 5 ms with respect to the left channel impulse, the right ear signal would have been swamped by the reflected sound, and possibly masked by it.

The input signal for each channel is an impulse, and has a flat frequency spectrum. As can be seen from figure 4.6, the resulting ear signals for the anechoic case also has a very flat spectrum. (The dip at 11kHz is an artifact caused by the grossly simplified transfer functions used.) The frequency spectrum for the setup with reflections, on the other hand, shows large amounts of ripple.

4.5 Conclusions

A simplified model of a setup for binaural reproduction has been investigated to study the effect of a single reflection on the playback. For this setup, the following conclusions may be drawn:

It has been shown that crosstalk cancelling designed for the direct sound will not give crosstalk cancellation for the reflected sound.

Binaural reproduction is used to place virtual source around a listener. But the reflections resulting from playback of the binaural sound corresponding to the virtual

source will differ from the reflections caused by a real source. Reflections may be heard when a real source would not have caused reflections, or the reflections may be different from those caused by a real source. These unnatural reflections may influence the listeners perception of the binaural reproduction.

The reflected sound gives an interaural time difference (ITD) that is nearly equal to the ITD for sound from a source located at position of the reflecting wall. It is possible that the listener will perceive the reflections as a secondary source (which may not be heard separately, due to the precedence effect.)

The reflected sound will show up at the ears of the listener and contribute much of the energy of the ear signals. Depending upon the input signals, it is possible that the intended ear signals will be partially masked by the reflections.

It has been shown that the reflection will influence the playback of binaural sound. It may be expected that this will influence the listener's perception of the binaural reproduction, but from the simplified model used, it is difficult to predict to which degree.

Chapter 5

Experiments with reflections

Listening tests was a natural choice of method for the current investigation. While time-consuming and requiring a lot of resources, this is a very practical and direct approach. The results obtained should be highly relevant, as we are investigating the effects as experienced by the listener, who after all, is the actual reason for applying a reproduction system at all.

This chapter present the experimental work done on these grounds. A preliminary experiment was first carried out, to get some indications of the results to be expected and to try out experimental methods and techniques. Based on the experiences gathered, a series of listening tests for combinations of playback setups and crosstalk cancelling filtering was undertaken. The design of the tests is presented, and the experimental conditions and the analysis of the resulting data is described.

5.1 A preliminary experiment

A preliminary experiment was carried out. This was a listening test where a playback system based on binaural recordings and crosstalk cancellation was compared to the same system with a reflecting wall added.

The primary purpose of the test was to investigate what kind of effects to expect when adding reflections to such a system, and how grave these effects would be. Additional goals were to further test and evaluate the implementation of crosstalk cancellation (chapter 3), and to try out experimental methods.

The experiment was carried out as a term project by Olav Mellum Arntzen, in co-operation with the author, who also acted as project advisor. Further details of the experiment are given by Arntzen [1998].

5.1.1 Preparations for the tests

The source material used for the test was binaural recordings of a talking person, made with an artificial head (Neumann KU81i). The recordings were made in a normal (i.e. not anechoic) room of medium size.[1] The dummy head was placed in the center of the room, and the talking person recorded while placed at directions every 45th degree on a circle around it. Two distances were used, 1.5 meters and close to the artificial head. Recordings were also made of a talking person walking in a circle around the dummy head, both in this room, and in a large auditorium.

The recordings were filtered with a set of crosstalk cancelling filters before playback, designed from impulse response measurements of the anechoic playback setup. The design of the filters and the filtering of the recordings were done with the software presented in chapter 3.

From the recordings, two sets of ten signals were chosen for the tests. The first set included recordings of directions in the frontal half plane at 1.5 meters distance. In the second set, half of the signals were recorded at close distance, and a backward direction was included. The sets of directions are shown in figure 5.1.

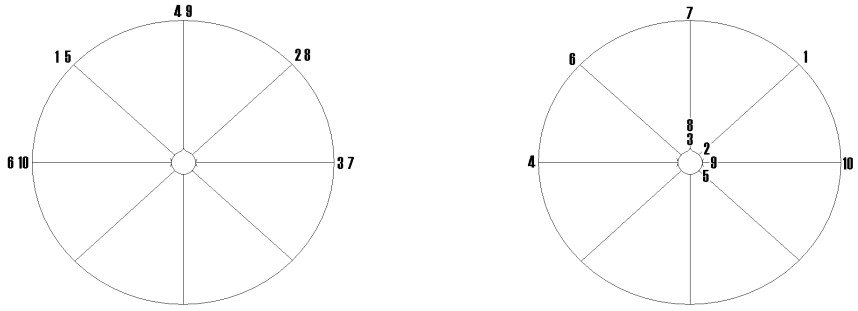


Figure 5.1: Presented source positions. Signal set one to the left, signal set two to the right. The "forward" direction for the listener is upwards in the figure. The numbers indicate the order in which the signals were presented.

[1] Ca 28.5 m^2 floor area and 85 m^3 room volume.

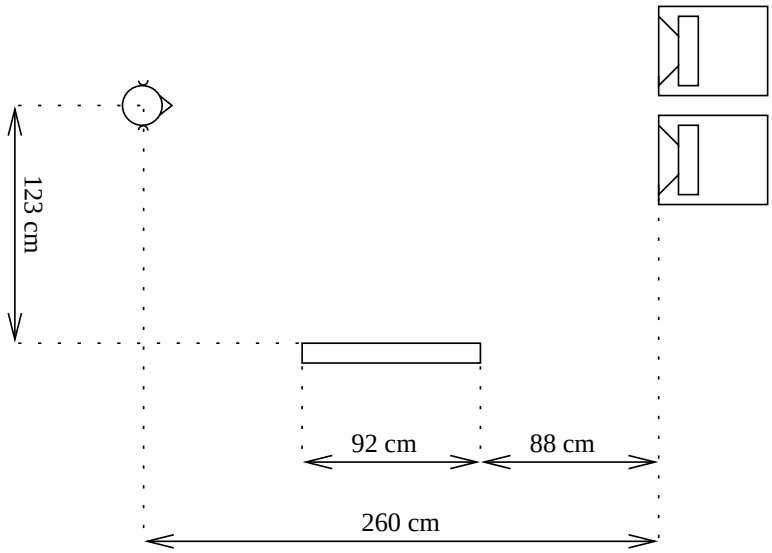


Figure 5.2: Setup in the anechoic room for the preliminary experiment. The loudspeakers, the listener position and the wall are shown. The height of the wall was 200 cm.

5.1.2 The tests

Five persons volunteered as listeners. They were 23 to 26 years old, and of both sexes. None of them had any previous experience with listening tests.

The tests were carried out in an anechoic room. Two different setups were tried, one anechoic and one with a wall placed to produce a reflection. A sketch of the setup, showing the loudspeakers, the listener position and the wall is given in figure 5.2.

To introduce the listeners to the playback system and their task, they got to listen to some recordings before the tests. This was a piece of music and the two recordings of a person walking around them.

Both sets of signals were tried both with and without the reflecting wall present. To gather the answers, forms like the ones used for showing the source directions in figure 5.1 were used. For the first signal set, the listeners were instructed to mark the direction of the sound source. For the second set of signals, they were instructed to mark the direction, and also whether the sound was “close” or not. After the tests, the listeners were asked for their impressions during the tests.

5.1.3 Results and conclusions

The listeners' answers were compared to the directions used for the recordings. If they were equal, the answer was noted as correct. The results are summarised in table 5.1. As an example of the detailed results, one listener's answers are shown in figure 5.3. These answers are not typical, this person was the one with fewest correct answers in all cases.

As can be seen from table 5.1, the addition of a wall lead to fewer correct answers for the first set of test signals. For the second set of test signals, the number of correct answers is the same for the two setups.

From the detailed results and the interviews it was clear that the presence of the wall indeed made a difference upon localisation and how the auditory images were perceived. This can be seen in figure 5.3. In the anechoic setup, the answers correspond quite well with the source directions for the first signal set, and to some degree for the second set of signals. With the reflecting wall present, the imaging has nearly totally collapsed. Most of the presented sound has been localised either to the front or forward left.

Some conclusions could be drawn: There seem to be individual differences both in how the listeners perceived the playback system and in how well they localised sound within it. It also seems clear that the reflection caused by the wall influenced the crosstalk cancellation and the playback in such a way that a distinctly different auditory image was perceived by the listener than under pure anechoic conditions.

	Signal set 1		Signal set 2	
Listener	W.o.wall	With wall	W.o.wall	With wall
No. 1	60%	40%	50%	20%
No. 2	100%	90%	70%	70%
No. 3	90%	90%	80%	80%
No. 4	80%	70%	50%	60%
No. 5	80%	90%	70%	90%
Mean	82%	76%	64%	64%

Table 5.1: Percentage of correct answers in localisation tests

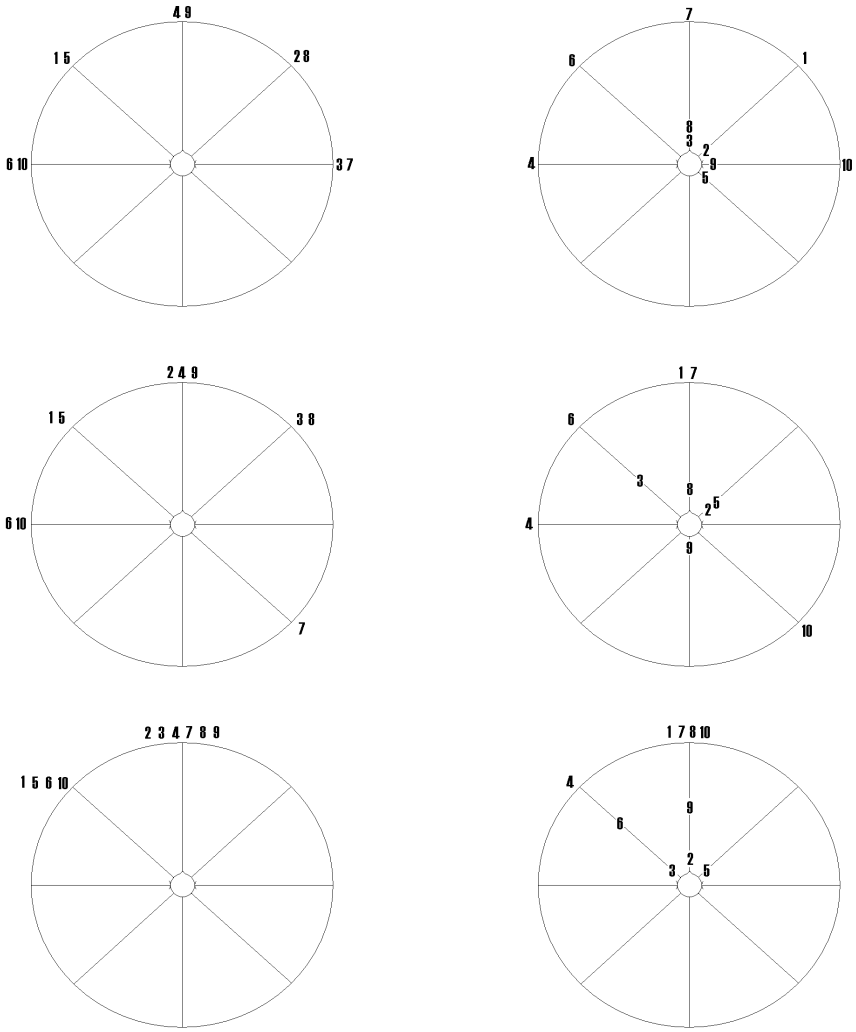


Figure 5.3: One listener's answers. Upper row: The presented source positions. (Also shown in figure 5.3, repeated her to ease comparison to the answers.) Middle row: Answers for the setup without reflections. Lower row: Answers for the setup with the wall present.

5.2 Experimental design

Sound in reverberant rooms sets up complex sound fields with a large number of reflections (section 2.1). Studying such rooms and sound-fields in order to find the characteristics that affect loudspeaker playback of binaural sound would be difficult. It was instead decided to limit the scope of the investigation to a simpler case. The simplest case possible that is not anechoic, is a setup with a single reflection. This can easily be achieved by adding a reflecting surface to a free-field setup. As a further simplification, it was decided that only sources and directions in the horizontal plane were to be considered.

It was chosen to let the reflection come from the listener's side. This is preferable for several reasons. Sound from the sides will give strong interaural cues, important for localisation (chapter 2.2.2). It is also known from previous work that reflections from the sides are the ones most likely to cause performance degradation (chapter 2.4).

Binaural recordings of speech were chosen as the test signal. Speech is sufficiently broad band, contains transients, is a very natural signal and should be familiar to the listeners (cfr. section 2.2.2).

The timing of the reflection should correspond to what one might get in a normal room (section 2.1). Two to three milliseconds has been cited as a limit above which reflections do no harm (section 2.4), so longer delays are interesting. The chosen delays, five and ten milliseconds, fulfill these conditions, and are also suitable middle values with respect to the precedence effect. For a speech signal, these delays will be in the domain of localisation dominance, below the echo threshold (section 2.2.3).

Localisation (section 2.2.2) was chosen as a measure of performance. It is both a salient feature of this kind of sound reproduction, has a clear reference, is easily measured and is commonly used as a performance parameter. Playback under anechoic conditions was chosen as a reference setup to which the setups with reflections could be compared.

Before presentation, the binaural recordings would have to be prefiltered with a crosstalk cancelling filter. In addition to filtering for the anechoic setup, it was decided to also produce a set of filters for the cancellation of both crosstalk and reflection.

It was chosen not to individualise the binaural recordings and the crosstalk cancelling filtering. Not doing this would probably lead to poorer performance of the reproduction system (section 2.3.2, page 15). But for many applications of binaural reproduction, individualisation is not an option. It was also believed, based on the experiences from the preliminary experiment (section 5.1) that the effects of the reflections would be clearly evident even without individualisation.

It was decided to do the tests without trying to hide the reflecting walls from the listeners. The motivation of the work is the potential application of crosstalk cancelled binaural reproduction in ordinary rooms, where the walls most definitely are visible. So it was believed that setup with visible walls would correspond more closely to this situation than would artificial reflections produced by signal processing (as done by

Takeuchi and Nelson [1999]). Also, as discussed in section 2.2.3, the inner workings of the sense of hearing does a good job of "making sense" of the input in order to present an auditory image that is coherent and plausible. A reflection appearing from nowhere, with no apparent cause, is far less plausible than a reflection coming from a nearby wall, and might have been considered unnatural.

5.3 Coordinate system

A coordinate system was defined, which is used throughout the rest of this work. It is a variation of the type mentioned in section 2.2.1, and is illustrated in figure 5.4. There is one coordinate, namely direction in the horizontal plane. Directions are given as angles in degrees, in the interval $-180°$ to $180°$, increasing clockwise. Negative angles are to the left, positive angles to the right. The direction "forward" is at zero degrees.

The coordinate system has a discontinuity at $\pm 180°$ which is not representative of the physical setup. In situations where it is necessary to present data situated near or at this discontinuity the coordinate system is extended beyond $\pm 180°$, to include more negative or more positive angles. This is illustrated to the right in figure 5.4. The total angular interval used will not in any case be larger than $360°$, in order to maintain a unique representation of directions. As an example, the angular range may be $-270°$ to $90°$, with the direction "Left" $(-90°)$ as the midpoint.

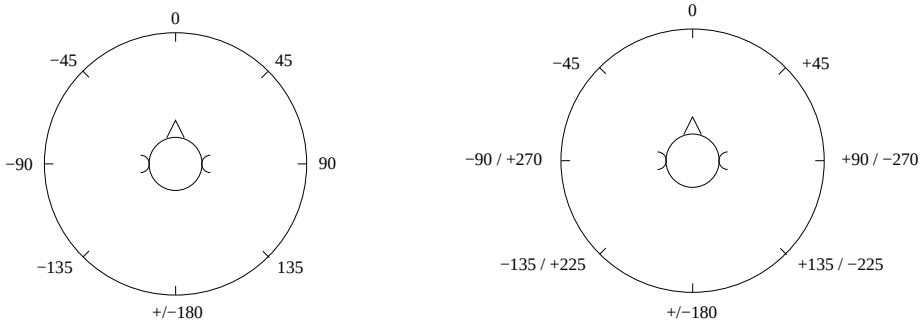


Figure 5.4: Left: The directional reference system used. The origin is taken to be at the middle of the listener's head. Directions are given as angles, in degrees. Right: To avoid discontinuity problems, the coordinate system is in some cases extended beyond ± 180 degrees, towards more positive or more negative angles. The total angular interval used is never more than $360°$.

5.4 Experimental conditions

5.4.1 Setups

To be able to have free-field conditions, an anechoic room [Ustad, 1985] was chosen as the test venue. Reflections were produced by adding free-standing partition walls. To give strong reflections, these partitions were hard, made from 12mm chipboard mounted on a stiff wooden framework. Up to two of these walls were used for the tests. A floor plan of the setup is shown in figure 5.5. As can be seen, three positions were used for the walls, causing (first) reflections which were delayed approximately 5 milliseconds and 10 milliseconds with respect to the direct sound from the loudspeakers. The direction towards the center of the walls were $\pm 56.3°$ and $-68.2°$. In the close positions, the wall(s) filled an angle of $24.4°$, from $\pm 45.8°$ to $\pm 70.2°$, and in the far position an angle of $18.1°$, from $-59.7°$ to $-77.8°$.

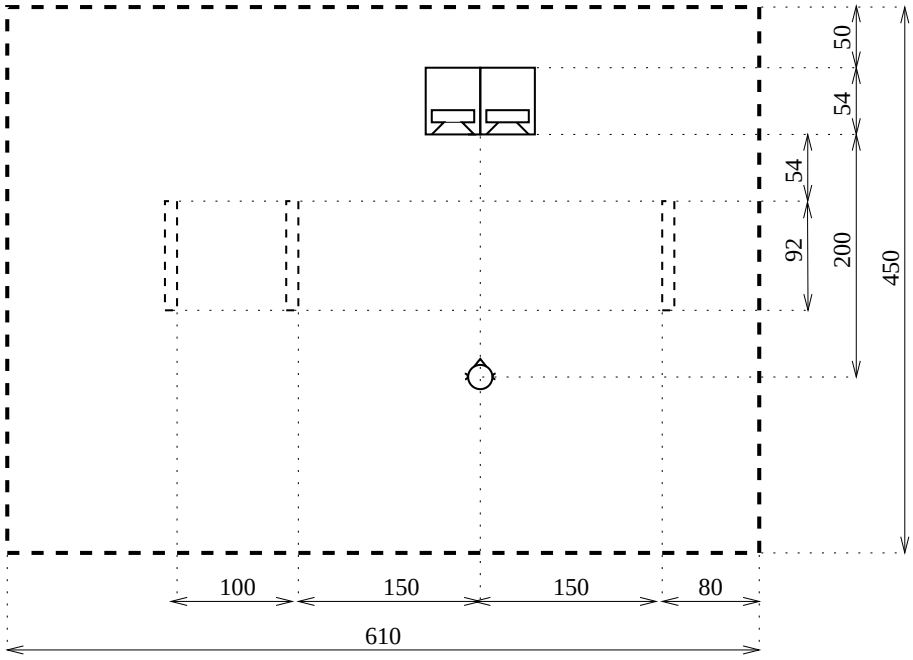
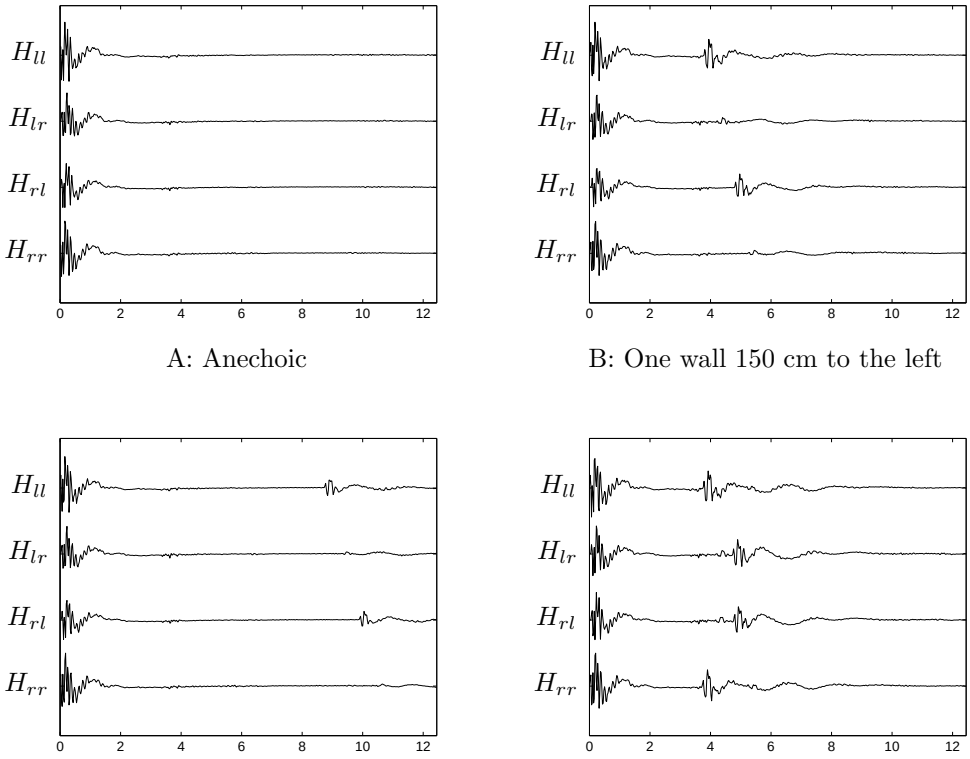


Figure 5.5: Experimental setup in the anechoic room. The loudspeakers and the listener position are shown. The locations used for the reflecting wall(s) are shown with dashed lines. All measures are in centimetres. The height of the room was 450 cm, the height of the reflecting walls was 200 cm. The thick dashed line marks the inner boundary of the room.

Four different physical setups were used: Anechoic, a wall in the close left position, a wall in the far left position, and walls both left and right (close positions). The

binaural impulse responses for the setups are shown in figure 5.6. These were measured using an artificial head, Neumann KU81i, equipped with shoulders made at the Acoustics Group. From the figure it is evident that the reflections give large ILDs for the one wall setups. The reflections are clearly visible for the left ear responses, and hardly visible at all for the right ear responses. The ITDs for the reflection are not illustrated, but for the wall in the close position it is ca 0.4 ms, corresponding well to the values found in section 4.3.



A: Anechoic B: One wall 150 cm to the left

C: One wall 250 cm to the left D: Two walls, 150 cm to each side

Figure 5.6: Binaural impulse responses for the various setups for the listening tests. Amplitude along the vertical axes, time in milliseconds along the horizontal axes. For each setup the four impulse responses H_{ll}, H_{lr}, H_{rl} and H_{lr} (cfr. figure 2.3) are shown, shifted along the vertical axis. For the non-anechoic cases, the reflections produced by the walls can be seen, approximately 5 ms (B, D) and 10 ms delayed (C).

The room was equipped with a playback system for audio, consisting of two loudspeakers and a power amplifier, connected to a pre-amplifier and a CD-player outside the room. As close loudspeaker placement is preferable (section 2.3.4, page 20), the loudspeakers were placed next to each other, spanning an angle (center to center) of

12 degrees when seen from the listening position. The loudspeakers and the listener position were placed asymmetrically in the room. This was done to be able to place reflecting surfaces at a greater distance (at the left side) than what would otherwise have been possible (see figure 5.5).

The loudspeakers used were Bowers & Wilkins Matrix 801, series 3. These are high-quality three-way systems, with crossover frequencies 380Hz and 3kHz. The free-field frequency response for one of them is shown to the left in figure 5.7. To evaluate the directional characteristics of the loudspeakers, the total radiated power was measured under diffuse-field conditions, as shown to the right in the figure.

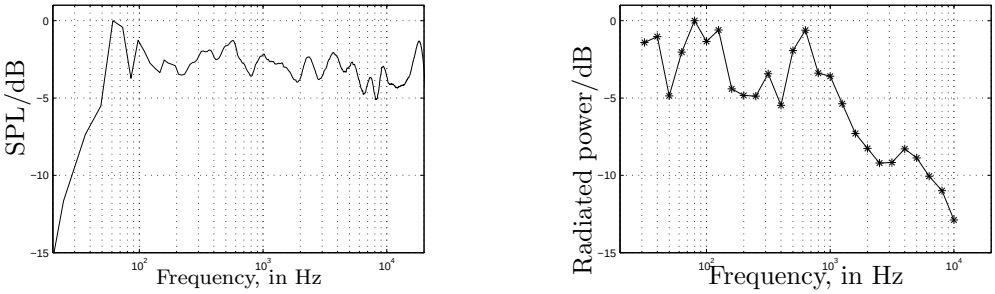


Figure 5.7: Frequency responses for one of the two loudspeakers used for playback. Left: Sound pressure level, measured on-axis under free-field conditions. Right: Radiated power measured in 1/3 octaves under diffuse-field conditions. The responses are normalised to the maximum values.

5.4.2 Source signals

For use as source signals, twenty binaural signal sequences were recorded, ten with a male speaker as a source, and ten with a female speaker. The artificial head used for this was the same as the one used for the HRTF measurements of the playback setups (Neumann KU81i, see above). The head was placed in the middle of an acoustically well damped room at a height of 180 cm.

Sixteen source positions were used, evenly distributed on a circle of two meters radius centered around the artificial head, giving source directions 0°, ±22.5°, ±45° ... ±157.5°, 180° (cfr. section 5.3). Each signal sequence consisted of twelve source positions chosen randomly from the sixteen, with the added restriction that no position could be chosen more than twice for any given sequence. For each sequence, the speaker would first stand directly in front of the artificial head (at 0°) and say a few sentences. Thereafter, the speaker would move to each of the sixteen positions in turn, face the artificial head and say a short sentence. The recorded sequences were later edited to retain only the parts where the speaker was standing still talking. Gaps of silence (all of equal length) were inserted instead of the parts removed. The final sequences lasted from two to three minutes each.

5.4.3 Filters

Two sets of crosstalk cancelling filters were made, using the software routines presented in chapter 3. One set was based upon the HRTFs measured in the anechoic setup, the other upon the HRTFs for the setup with the wall in the close left position. The purpose of the latter was to compensate for the playback situation by cancelling the reflection in addition to the crosstalk.

The filter lengths and delays were chosen so the same values could be used in both cases. Compared to the example given in chapter 3.3.1, longer filters and responses were used, to improve the filtering. The inverses of H_{ll} and H_{rr} were chosen to be 800 taps long, with a delay of 300 samples. The total response of H_{ll} convolved with its inverse is shown in figure 5.8. The interaural transfer functions were computed and truncated to 1024 samples, as in the example. The factor $1 - Hia_l Hia_r$ was computed, 299 samples at the start were cut and the next 1501 samples retained. The inverse of this factor was then computed, with a filter length of 800 taps and a delay of 400 samples.

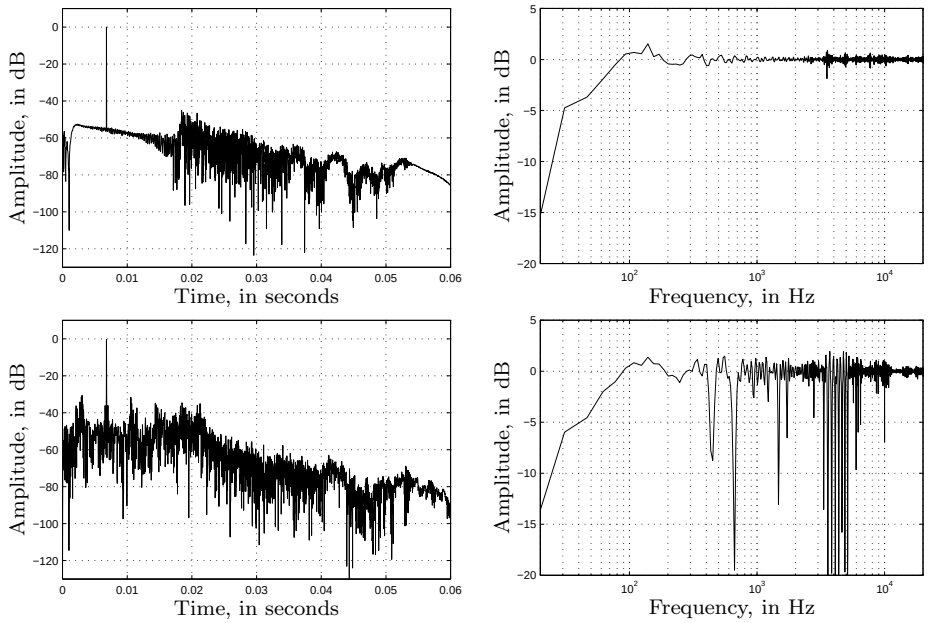


Figure 5.8: Total response of H_{ll} convolved with its approximate inverse. Upper row: Filter for the anechoic setup. Lower row: Filter for the setup with a wall in the close left position. Left: Time domain, amplitude in dB. Right: Frequency domain amplitude. (Comp. fig. 3.4.)

The cancellation of both crosstalk and the presence of the wall is a more complicated operation than crosstalk cancellation alone. Firstly, some of the filter resources will be allocated to the cancellation of the wall, placing relatively less weight upon the

crosstalk cancellation of the direct sound. Secondly, as the cancellation depends upon more sound transmission paths, this setup is more prone to errors due to variations in the setup.

As discussed in chapter 3.4, the filtering was not done in real time, and this made it difficult to evaluate the performance of the filters by measurements. Loudspeaker signals and ear signals resulting from simulated reproduction of a test signal in the setups for which the two filters were designed are shown in figures 5.9 and 5.10. As can be seen from the time domain ear signals, the crosstalk is cancelled in both cases, with a level below the error floor, which is at -45 dB and -30 dB. For the setup with a wall present, the reflection from the wall is also cancelled. Its influence can be seen in the loudspeaker signals, most notably so in the left channel.

The frequency domain amplitudes of the ear signals for the anechoic setup are generally flat down to 60 Hz, with the exception of a 4 dB peak and 10 – 15 dB dips between 100 Hz and 200 Hz. The fundamental frequency of speech is 120 Hz for males and 220 Hz for females (middle values) [Krokstad, 1994], with only the fundamental frequency of the male speech falling into the frequency range with the irregular response. The frequency response was therefore found acceptable.

The extra filtering "expense" incurred by also cancelling the reflection for the setup with a wall shows up as a higher error floor in the time-domain ear signals (figure 5.10). In the frequency domain, the response is 5 – 10 dB lower than for the anechoic setup for frequencies from 200Hz up to about 1 kHz, and has a number of sharp dips for higher frequencies.

The crosstalk cancelling filter designed for the anechoic setup was also used for reproduction in the setups with walls present. Ear signals simulated for these situations are shown in figure 5.11. For the setup with a wall in the close left position, the reflections can be clearly seen in the left channel, delayed about 5 ms with respect to the intended impulses. There are four of them, grouped two and two, resulting from the left and right speaker radiating the signals necessary for the left channel impulse and the right channel impulse. The signals for the setup with a wall in the far left position are similar, but the reflections are delayed five milliseconds more and are weaker. The setup with two walls has the response that differs most from the one for the anechoic setup, showing strong reflections in both the left ear and right ear signals.

For the two one-wall setups, the level of the crosstalk is about -20 dB, as can be seen from the figure. For the two-wall setup, the crosstalk is at ca -10 dB, which is still lower than the level of the reflections by a few dB. In all these cases, the crosstalk should ideally have been at the levels obtained for the anechoic case. The fact that it is not is probably due to slightly different positioning of the dummy head for the HRTF measurements of the various setups.

For all these three setups, the frequency domain amplitudes of the ear signals are generally irregular. The amplitude in the frequency range from 200 Hz to 1 kHz is higher than for the anechoic setup, and a comb filter structure can be seen.

5.4.4 Experimental procedure

Setups

The four physical setups and the two crosstalk cancelling filters were combined. In each of the four physical setups, signals filtered for anechoic reproduction were used (section 5.4.3). For the setup with a wall in the close left position, signals filtered specifically for this very situation was also used (section 5.4.3), giving a total of five situations:

- No walls (anechoic)
- One wall, close left position
- One wall, far left position
- Two walls, close left and right positions
- One wall, close left position, reflection corrected for in the filtering

Listeners

Nine paid volunteers participated as listeners in this tests. They were all male, 22 to 26 years old. Six of them were students in the Acoustics group, in their fourth year of study. The other three were third year students, also at the faculty of electrical engineering and telecommunications. All of them had normal hearing, as showed by audiometric testing. None of them had participated in listening tests before.

Tests

Totally, eight tests were carried out over a period of six weeks. The first two tests were intended to familiarise the listeners with the playback system and their task. The first of these was announced as a "training session", and differed somewhat from the other tests. The second one was announced and carried out as a real test, but the results were not used.

At the end of the first training session the listeners were subjected to a test of their ability to localise the presented directions, combined with an interview afterwards. The results of the test indicated that some of them localised less well than others, and seemed to have problems localising to the rear. But from the interviews it was clear that all of them at least to some degree had experienced sound from all half planes in question; front, back, left and right. Based on this, it was found that they all localised well enough to participate in the rest of the tests.

In the remaining six tests, localisation in the five setups were tested. The anechoic, chosen as the reference situation, was tested twice. For each test, eight signal sequences were chosen, four from the ten sequences recorded with a male speaker and four from those recorded with a female speaker. Different combinations of sequences

were used for consecutive tests, to prevent the listeners from learning and remembering the sequences. The sequences were filtered with the crosstalk cancelling filter corresponding to the setup. Playback levels were set to correspond to those during the recording of the source signals (section 5.4.2).

The listeners were given written and oral instructions. They were told that they would hear a person talking from various positions. Their task was to mark the perceived direction of the talker on a form (one form for each sequence). The form is shown in figure B.1, page 125. The listener position was marked on the floor, and the listeners were instructed to stand in this position and to face the loudspeakers. Each test took nearly half an hour, including a five minute break, to work through for a listener, except for the first training session which was a few minutes shorter.

5.4.5 Gathering of data

The data were gathered from the forms filled in by the listeners during the listening tests. The answers were read from the forms and transfered to a computer. This was mostly done with the help of a digitising board connected to the computer, but some forms were treated manually. In this process all answers were treated as angles. All directions were rounded to the nearest integer degree value, giving 360 possible values for the perceived directions. (As noted in section 5.4.2, the presented direction could take on 16 possible values.)

Two values were missing, and two rejected due to improper filling out of the form. These were removed from the final dataset, so that totally for the five setups there were 5180 answers, 1728 for the anechoic setup and 862 to 864 for each of the other setups.

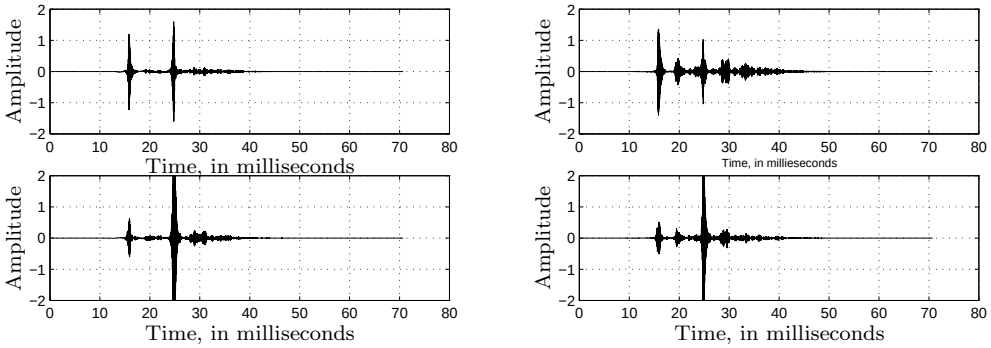


Figure 5.9: Loudspeaker signals for the two crosstalk cancelling filters. Signals for the anechoic setup to the left, signals for the setup with a wall in the close left position to the right. The input signal is the the same as in figure 4.3.

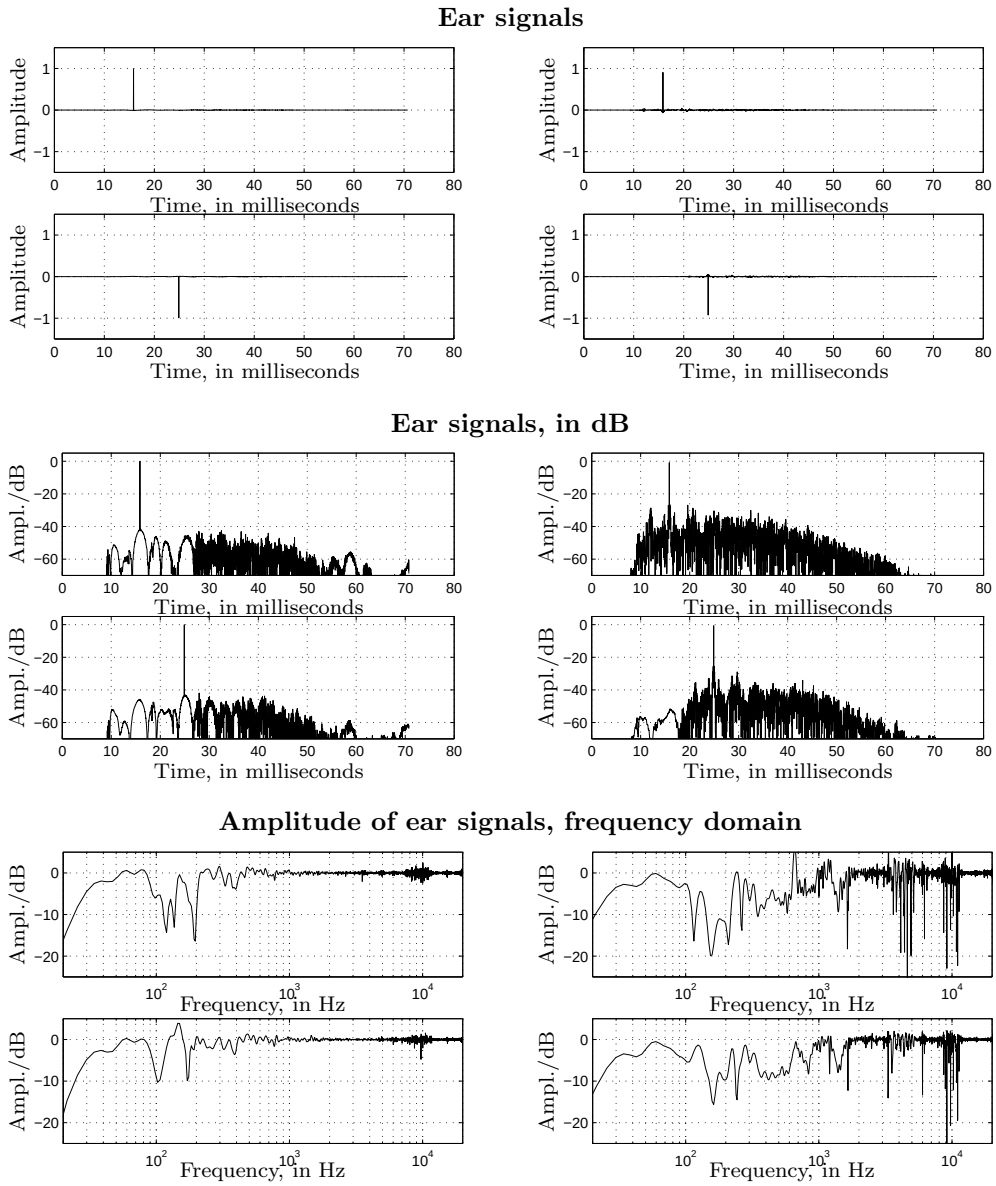


Figure 5.10: Ear signals, simulated for playback in the setups for which the filters were designed. Signals for the anechoic setup in the left column, signals for the setup with a wall in the close left position in the right column. Left channel signals in the upper rows, right channel signals in the lower rows. The input signal is the the same as in figure 4.3, and the corresponding loudspeaker signals are those shown in figure 5.9.

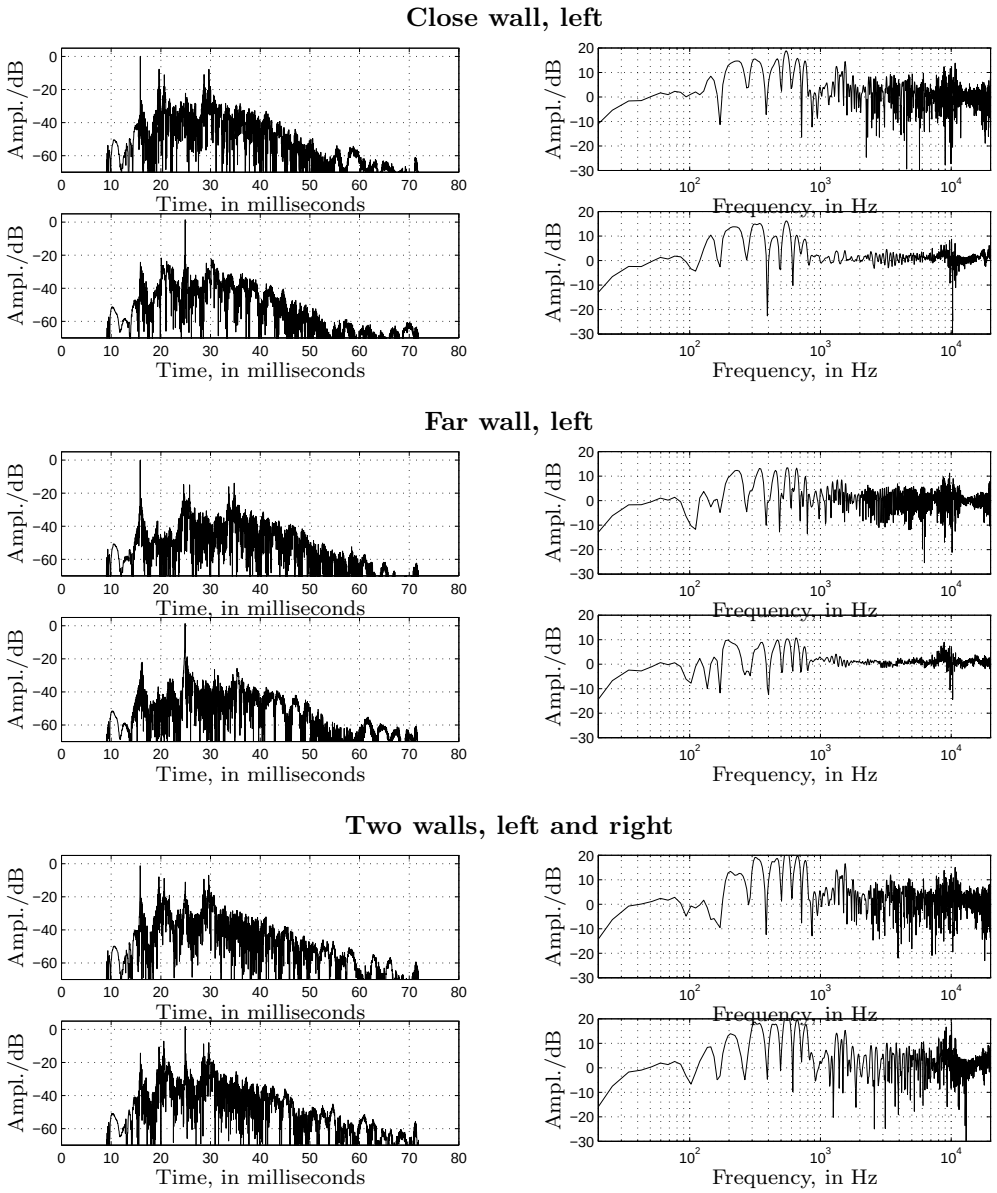


Figure 5.11: Simulated ear signals when using the crosstalk cancelling filter designed for the anechoic setup for setups with walls present. Time domain signals with the amplitude expressed in decibels in the left column, amplitude of signals in frequency domain in the right column. Left channel signals in the upper rows, right channel signals in the lower rows. Input signal as in figure 4.3. See also figure 5.10.

Chapter 6

Statistical analysis of listening test data

Large amounts of data were produced from the listening tests (section 5.4.5). These data must be interpreted to find the effects of the reflections. A statistical description of the data may be useful for this, but such characterisation is complicated by the presence of reversals (section 2.2.2), which causes bimodally distributed data

This chapter first defines a few terms and introduces the criteria used for the analysis and interpretation of the data. The statistical diffuculties due to the presence of reversed answers are then discussed. A dual-normal statistical model developed during this work is proposed as a solution, and several variants of this model are applied to the data from the listening tests.

6.1 Analysis criteria

Localisation was chosen as the performance measure for the tests (section 5.2). As discussed, localisation relates the location of an auditory event to attributes of, or correlated to, a sound event (section 2.2.2). The aspect of localisation studied here is the correspondence between the *presented direction* and the *perceived direction*. The presented direction is defined as the direction towards the sound source when the binaural signals for the tests were recorded (section 5.4.2). The perceived directions are the answers from the listeners, as read from the forms filled in during the tests (sections 5.4.4 and 5.4.5).

For the purpose of this work, *correct localisation* is defined as localisation where the perceived direction equals the presented direction.[1] To achieve this is the goal of binaural reproduction as it is defined and used here (sections 2.3 and 3.1.1). Localisation

[1] This term is defined for practical reasons. It does not imply that this localisation necessarily is "better" in any way than other possible localisations.

can be more or less correct, for perceived directions corresponding more or less well with the presented direction.

In general, totally correct localisation will not be achieved. The perceived directions for a given presented direction will typically be more or less randomly distributed, differing more or less from the presented direction. It was chosen to study these deviations from correct localisation according to the following criteria:

- Deviation of average perceived direction from the presented direction
- Spread of answers around the average perceived direction
- Reversals – front/back errors
- Localisation towards the reflections

For a set of perceived directions an average value may be found. This average perceived direction will describe the average localisation. It may differ more or less from the corresponding presented direction, giving a more or less correct average localisation.

The answers will be distributed in some way around their average value. This distribution may be described by computing a measure of the spread of the answers with respect to the average.[2]

Reversals is a deviation type distinct from the others. Reversals are front/back confusions, leading to auditory events being perceived in the opposite half plane of the presented direction, mirrored around the interaural axis. This effect can take place for normal hearing, but is often more common for binaural reproduction (sections 2.2.2 and 2.3.2). To distinguish between answers that have been subjected to reversal and those that have not, the terms *reversed* and *unreversed* are used. The *reversal rate* is defined as the ratio of the number of reversed answers to the total number of answers.

Localisation towards the directions of the reflections is another possible deviation type. It is closely related to the precedence effect (section 2.2.3), and very directly connected to the topic of the work, the effect of reflections. If present, this deviation type will show up as a change of average perceived direction, possibly also as a change of spread. But the cause of the deviations is distinct, and this deviation type is therefore treated separately.

[2]This is related to the term *localisation blur*, which describes the spatial resolution of localisation [Blauert, 1997b, Zwicker and Fastl, 1999]. Localisation blur is however defined as the smallest change of an attribute related to a sound event that is sufficient to produce a change of location for the corresponding auditory event [Blauert, 1997b, p 37]. The distribution of answers studied here is not caused by varying sound attributes, and it has therefore been chosen to use the term *spread* instead.

6.2 Reversals and statistical characterisation

Large amounts of data were produced from the listening tests (section 5.4.5). To be able to present these datasets in a well arranged and comprehensible manner, it is beneficial to reduce them to smaller sets of parameters describing and characterising them. This may typically be achieved by fitting a parametric statistical model to the data. Such a parametric description allows for easy comparison of datasets from different cases, and may also make it possible to test hypotheses and draw conclusions concerning the results.

The statistical characterisation should make it possible to describe the data according to the criteria presented in the previous section. Average value and spread are common statistical parameters. But reversals complicate this statistical characterisation, as the distribution of answers caused by this effect is not well described with common statistical models.

6.2.1 Bimodal distributions

Reversals cause data that are to some degree bimodally distributed. For a presented direction, there will be two clusters of answers, one mainly consisting of answers where reversal has occurred, and the other one mainly consisting of the unreversed answers. This is also the case for the data gathered from the tests done in this work. Two examples, where this effect can be clearly seen, are shown in figure 6.1.

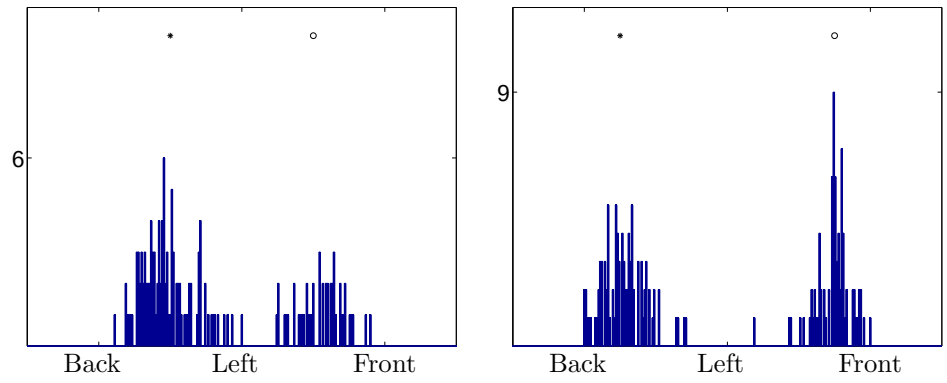


Figure 6.1: Histograms of answers from localisation test using crosstalk cancelled binaural reproduction, showing bimodal distribution caused by reversals. Direction in the horizontal plane along the horizontal axes, the number of answers along the vertical axes. The presented direction is marked with an asterisk, the mirrored direction with a circle.

These bimodally distributed localisation data are not well described by usual descriptive statistics and common statistical models. Average value computed as mean value

and spread computed as variance are the primary parameters of such descriptions, and the models are typically unimodal [Høyland, 1989a]. A characterisation using these methods will not describe the data according to the criteria outlined in the previous chapter, and may also lead to an interpretation of the data that does not correspond to the actual localisation process.

Firstly, the reversals effect is not present in these models. And there are no parameters that can be used to describe or estimate it, at least not directly. Therefore, the reversals and their effect will be hidden in the other results found from the model.

Further, the presence of reversals may cause a poor correspondence of the other statistical parameters of the model to the effects that should be described. The mean value will generally not correspond neither to the unreversed perceived directions nor to the reversed perceived directions, but to a direction somewhere in between those, dependent upon the reversal rate. Spread, computed as variance, will not distinguish between "normal" spread and deviations from the mean resulting from reversal, giving an incorrect, or at least incomplete, description of the data. Neither will the spread estimates be unique (one to one): A low spread and a large rate of reversals might lead to the same variance value as a large spread and few reversals, indicating falsely that these two situations were equal.

Finally, if these parameters (mean value and spread) are still computed, they may not be used for hypothesis testing without difficulties. The usual hypothesis testing formulas are based upon assumptions regarding the distribution of the stochastic variable that do not hold true for these bimodally distributed data [Høyland, 1989b].

It seems clear that in order to obtain a valid and useful statistical description of the data, the reversal effect must first be separated out.

6.2.2 Treatment of reversals

In order to obtain a meaningful statistical characterisation of results from listening tests, reversals have been treated in various ways in previous studies. One possibility is to treat reversed and unreversed perceived directions separately, as done by Damaske and Mellert [1969/70]. A similar approach is taken by Makous and Middlebrooks [1990]. Here, answers that are considered reversals are removed from the main dataset, and are commented on separately. To resolve apparently reversed answers is also common practice. This is done as a *remirroring* around the interaural axis: Answers are moved to the opposite (rear or frontal) hemisphere if this gives a direction that is closer to the direction of the sound source [Wightman and Kistler, 1989b, Wenzel et al., 1993].

Another approach, justified by the difficulties of treating apparently reversed data statistically correct, is to report raw data, as done by Wightman and Kistler [1997, 1999]. Here directions are decomposed into right/left, front/back and up/down coordinates. In [Wightman and Kistler, 1992], a mixed approach is taken. Centroids are calculated without resolving front/back confusions. Presentation of raw data is mostly avoided, but is used to emphasize important points. Neither of these two last

methods give a description according to the criteria for statistical description chosen for this work (section 6.1).

6.2.3 Resolving of reversals by remirroring

A problem when working with data sets containing reversals is to distinguish between reversed and unreversed answers. This is necessary both in the case of treating reversed answers separately and when resolving them by remirroring.

For small spreads and source directions not too close to the left and right directions, the answers may cluster in two distinct groups, as seen in figure 6.1. In cases like this it will be clearly evident whether a given answer is reversed or not. The reversal rate can be computed directly from the number of answers in the two clusters, and the reversals can easily be resolved by remirroring, as seen in in figure 6.2.

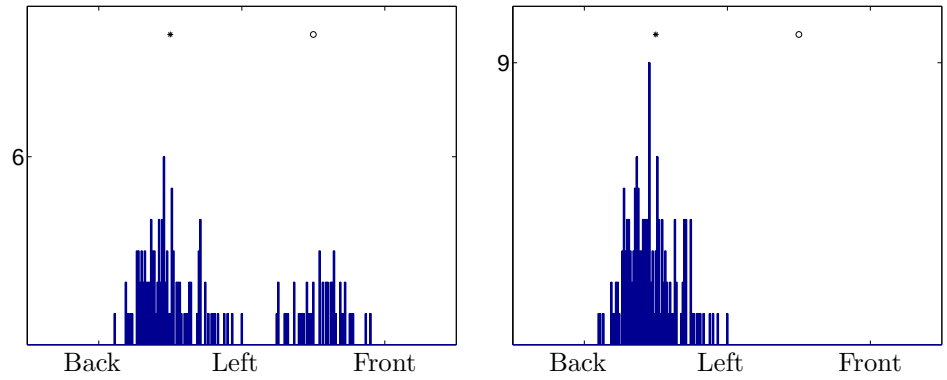


Figure 6.2: Histograms of answers from localisation test using crosstalk cancelled binaural reproduction. Direction in the horizontal plane along the horizontal axes, the presented direction is marked with an asterisk, the mirrored direction with a circle. The number of answers along the vertical axes. Left: The original histogram. The cluster of unreversed answers and the cluster of reversed answers are not overlapping. Right: Histogram after resolving of reversals by remirroring. The resulting distribution of answers is unimodal and symmetrical.

For larger spreads and for source directions close to the reversal axis, the clusters of unreversed and reversed answers may overlap, as seen in the left plot of figure 6.3. In cases like this, some (principally) reversed answers will end up in the same half plane as the source, and some (principally) unreversed answers will end up in the opposite half plane of the source direction. It is then generally not possible to tell whether a given single answer is principally unreversed, but has a large error, or whether it is reversed but otherwise has a low error.

Resolving the reversals by remirroring in cases like this will produce skewed distributions, as can be appreciated from the right plot in figure 6.3. As the clusters of

answers overlap at the reversal axis, the (principally) reversed answers that happen to be in the source half plane will not be remirrored, and the (principally) unreversed answers in the opposite half plane of the source will be wrongly moved to the source half plane. Using the distribution of answers found from this remirroring for a statistical description of the answers will give a wrong average value for the perceived direction and too low values for spread. The reversal rate computed from the number of remirrored answers may also be wrong, dependent upon, amongst other factors, the real reversal rate.

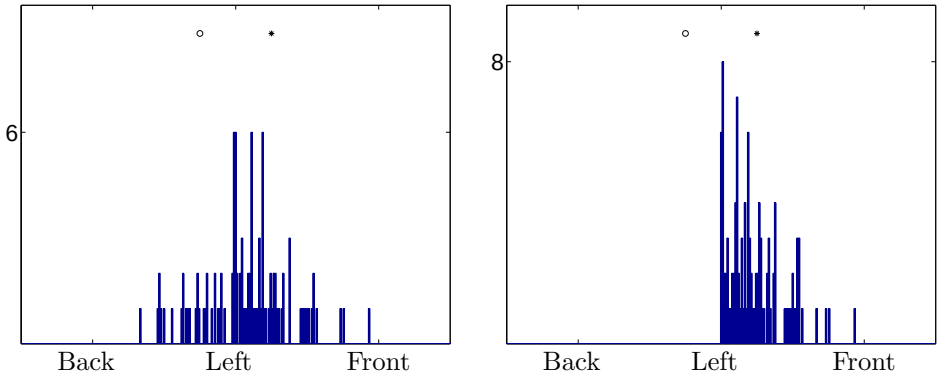


Figure 6.3: Histogram of answers from localisation test using crosstalk cancelled binaural reproduction. Direction in the horizontal plane along the horizontal axes, the presented direction is marked with an asterisk, the mirrored direction with a circle. The number of answers along the vertical axes. Left: The original histogram. The cluster of unreversed answers and the cluster of reversed answers overlap at the reversal axis. Right: Histogram after resolving of reversals by remirroring.

6.2.4 Remirroring applied to listening test data

The answers from the listening tests were generally bimodally distributed. An example of this is shown in figure 6.4. As can be seen, for presented directions towards the front and rear, the two peaks of the bimodal distribution are clearly separated. But for presented directions closer to the interaural axis, the peaks are more or less overlapping, sometimes to the extent that a seemingly single peak is formed.

Resolving the reversals by remirroring was tried as a first attempt to simplify the datasets from the tests to something statistically tractable. For some cases this gave useful results. For other cases, there were difficulties due to the clusters of reversed and unreversed answers overlapping, as discussed above (section 6.2.3). Generally it was found that resolving of reversed answers by remirroring did not give results that could be used to give useful and valid descriptions corresponding to the analysis criteria chosen.

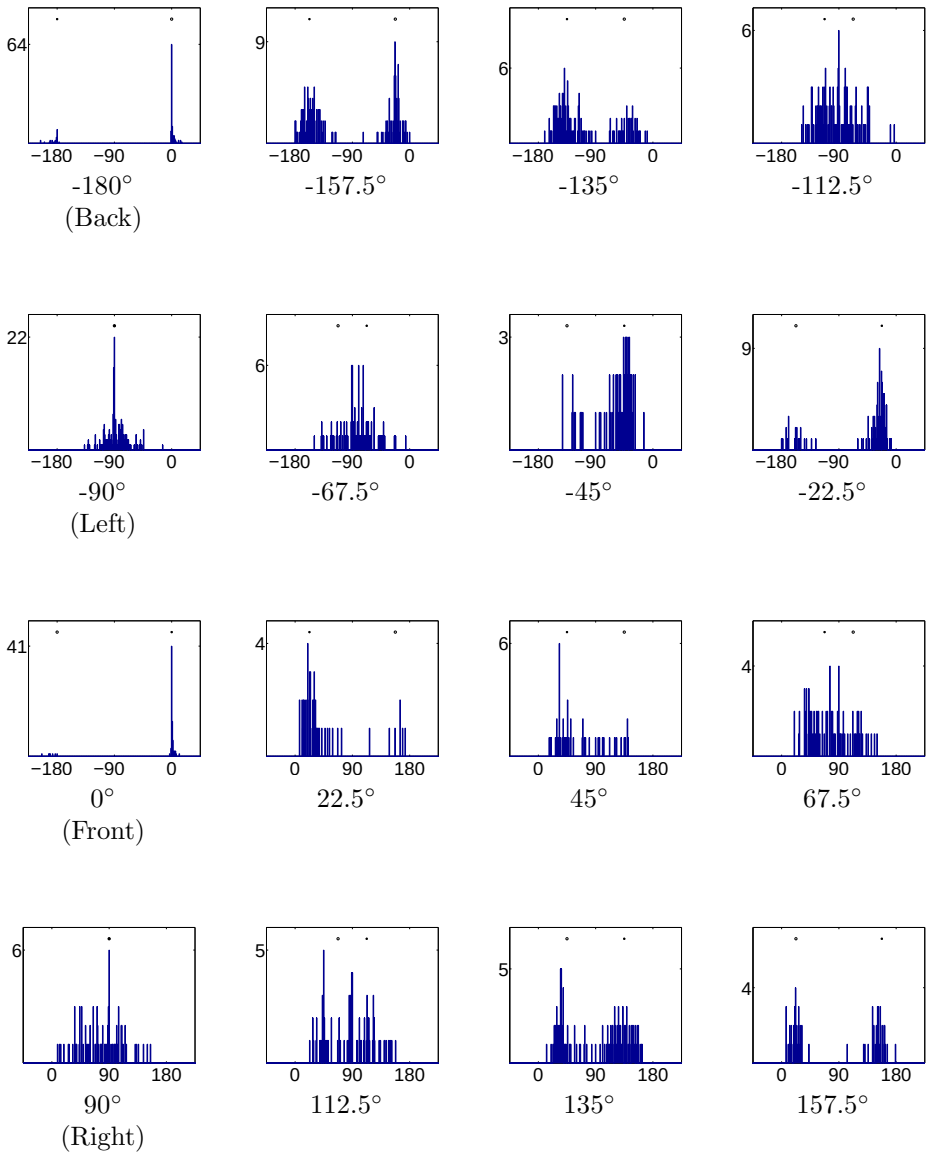


Figure 6.4: Histograms of answers for one of the two tests done using the anechoic setup (section 5.4.4). The histograms are labelled with the corresponding presented direction. Perceived direction along the horizontal axes. The presented direction and its reversal are marked with dots in the upper part of the histograms. Note that the y-axes have different scalings, in order to ease comparison of the shapes of the histograms. Maximum values are indicated on the y-axis to the left of each histogram.

6.3 Dual-normal probability distribution

As discussed in the previous section (6.2), previously used methods for statistical description of listening test data are not satisfactory for the needs of this work. Another, and more suitable, method for the statistical description and treatment of the data is therefore needed. This model should be targeted to the localisation process and the resulting data, and should give a small number of parameters that characterises the dataset well. Such a model, developed during this work, is proposed and described here.

6.3.1 Modeling localisation with reversals

When the hearing is presented a sound stimulus from a sound source in a given direction in the horizontal plane, a corresponding auditory event arises (section 2.2.1). The direction of the auditory event is called the perceived direction (section 6.1). Chances are that the perceived direction will differ from the presented direction. The fundamental properties of the localisation process connecting the presented and perceived directions seem to be that two main mechanisms are at work causing this difference. The first is a *spread of the perceived directions around an average direction*, where the average direction typically is displaced from, but close to, the presented direction. The other effect, which is separate from and additional to the first, is *reversals*, which has been discussed in sections 2.2.2, 2.3.2 and 6.1. Together these two stochastic processes cause the bimodal distributions seen in the data from the listening tests (e.g. figures 6.1 and 6.4).

Spread and localisation uncertainty are inherent properties of the human hearing [Blauert, 1997b]. When binaural reproduction is used, additional error sources may be introduced that may also contribute to this effect. It is assumed here that this effect has a probability distribution density (pdf) that may be approximated well with a Gaussian (normal) distribution, described with a mean value and a variance (figure 6.5).

The reversal mechanism is a mirroring with respect to the interaural axis (section 2.2.2). It may be appropriately described with a binomial probability distribution, where the outcome "reversal" happens with a given, but unknown probability [Høyland, 1989a].

Both of these mechanisms are supposed to be dependent upon the direction of the presented sound. It is known that our localisation ability in the horizontal plane varies with the position of the sound source [Blauert, 1997b], and it is reasonable to believe that this also may be the case for the reversal effect. It is therefore necessary to treat each presented direction separately.

Whether the mirroring should be considered to happen "before" or "after" the distributive errors is not known. In the latter case, the spread of the answers would be the same for the reversed and the non-reversed clusters of answers. In the first case, the spreads do not necessarily have to be equal, but they can be.

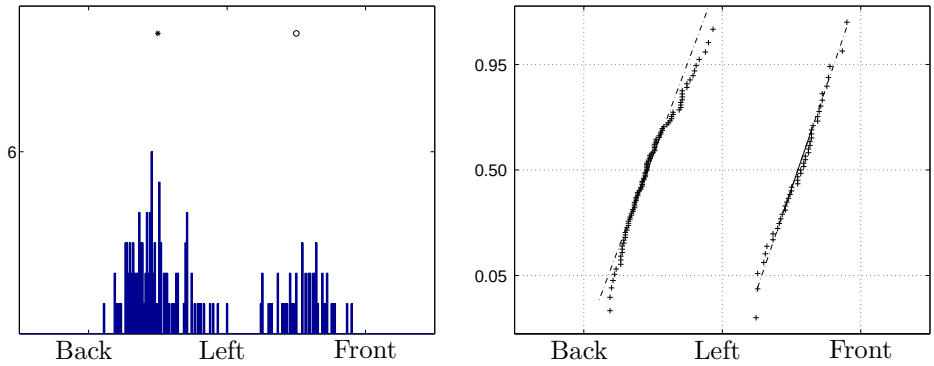


Figure 6.5: Bimodally distributed data from localisation test and separate normal probability plots for the answers from the two peaks of the data distribution. Left: Histogram of data. Right: Normal probability plots. Direction in the horizontal plane along the horizontal axes. The number of answers along the vertical axis of the histogram, percentiles along the vertical axis of the normal probability plot. The data from the two clusters of answers show up as nearly straight lines in the normal probability plot, showing that the two clusters of answers may each be suitably modelled by a Gaussian distribution [Høyland, 1989b].

A model is proposed where the localisation process and the resulting data are described with a *dual-normal probability distribution*,[3] composed as a weighted sum of two normal distributions. One normal distribution represents the unmirrored part, the other the mirrored part, and the weight the probability for reversal.

The probability density function (*pdf*) for a normally distributed variable with mean μ and standard deviation σ is given as [Crow et al., 1960]:

$$f_n(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma^2}$$

The corresponding cumulative distribution function (*cdf*) F_n is given as the integral of f_n:

$$F_n(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(x-\mu)^2/2\sigma^2}\, dx.$$

A general dual-normal probability density function may then be given as a weighted mean of two normal distributions:

$$f_{dn}(x, w_1, \mu_1, \sigma_1, \mu_2, \sigma_2) \;=\; w_1 f_n(x, \mu_1, \sigma_1) + (1 - w_1) f_n(x, \mu_2, \sigma_2), \qquad (6.1)$$

[3]The term *dual-normal* used for this distribution should not be confused with the *binormal* distribution, which is a two-dimensional probability distribution [Høyland, 1989b].

with a corresponding cdf:

$$F_{dn}(x, w_1, \mu_1, \sigma_1, \mu_2, \sigma_2) \;=\; w_1 f_n(x, \mu_1, \sigma_1) + (1 - w_1) f_n(x, \mu_2, \sigma_2). \tag{6.2}$$

Here the factors w_1 and $(1 - w_1)$, $w_1 \in [0, 1]$, are the relative weights of the two parts of the distribution. The weight of the second part is expressed in terms of the first to emphasize the fact that the two weights must add up to one in order for the function to be a valid pdf. This model describes a situation where a random variable is distributed according to a combination of two normal distributions.

Applying this to localisation with reversals, a probability density function describing the localisation process in the horizontal plane may be given as

$$f_{loc}(x, p_r, \mu_{nr}, \sigma_{nr}, \mu_r, \sigma_r) \;=\; (1 - p_r) f_n(x, \mu_{nr}, \sigma_{nr}) + p_r f_n(x, \mu_r, \sigma_r). \tag{6.3}$$

Here $p_r \in [0, 1]$, the weight of the second part, is the probability of obtaining a reversed answer. The probability for non-reversal is then $(1 - p_r)$. The parameter μ_{nr} represents the perceived direction as it would be if there were no spread and reversal. The value of this parameter can be different from the presented direction. Similarly, μ_r represents the reversed perceived direction. The parameters σ_{nr} and σ_r describes the spread mechanism for unreversed and reversed answers, respectively.

Applying the model not to the localisation process, but to the resulting data, p_r describes the reversal rate of the answers (section 6.1). The average perceived directions, unreversed and reversed, are given by μ_{nr} and μ_r, and the spreads by the standard deviations σ_{nr} and σ_r. In this case the parameters describe the data, and may also be taken as estimates of the real parameters of the localisation process; reversal probability, reversed and unreversed perceived directions and spread.

Examples of plots of the pdf and cdf for dual-normal distributions are shown in figure 6.6. Note from the left plots in the figure that the value of the reversal probability can be read from the height of the horizontal middle part of the cdf. It can also be seen that the cdf has its steepest slopes at the mean values. Steeper slopes correspond to lower standard deviations (lower left plot). Unimodal behavior of the distribution may be achieved, either by setting the reversal probability to zero or one, or by choosing means close to each other and large standard deviations, causing the two peaks to overlap to a high degree (lower right plot).

6.3.2 Choice of free parameters

The model developed has five parameters: A reversal probability, two means and two standard deviations. When applying this model for the description of listening test results containing reversals, it must be judged which parameters it is appropriate to choose as free parameters, and which, if any, to fix or to place restrictions on. Several possibilities exist.

The reversal rate should be free. To fix it, its value would have to be determined by other means. This is, as discussed (sections 6.2.2 and 6.2.3), difficult in many cases, and actually a part of the problem this model is intended to solve.

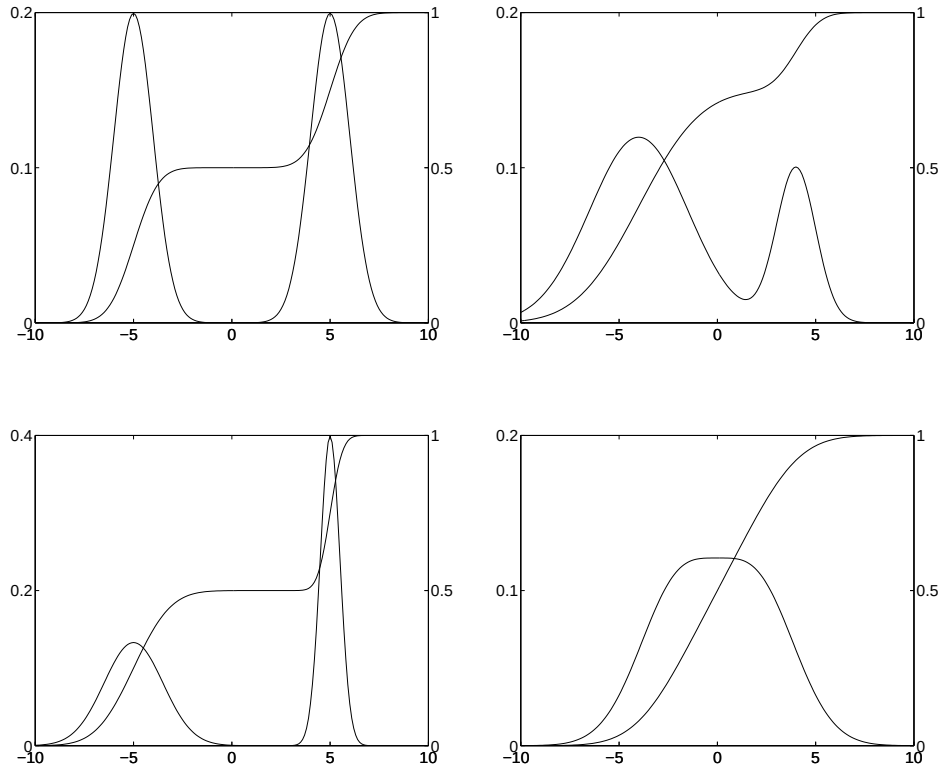


Figure 6.6: Dual-normal probability functions. Probability density function (left axis) and cumulative distribution function (right axis).
Upper left plot: The reversal probability p_r is set to 0.5, mean values to ± 5 and both standard deviations to 1.
Lower left plot: p_r is set to 0.5, mean values to ± 5, σ_{nr} to 1.5 and σ_r to 0.5.
Upper right plot: p_r is set to 0.25, mean values to ± 4, σ_{nr} to 2.5 and σ_r to 1.
Lower right plot: p_r is set to 0.5, mean values to ± 2 and standard deviations to 2.

Reversals are usually taken to be confusions of front and back, giving answers that are symmetrically placed around the interaural axis. According to this, the means could be fixed to the presented angle and its reversal, or they could be chosen as free variables with the restriction that they should mirror each other, (i.e. be placed symmetrically around the interaural axis). They could also be chosen to be independent of each other, but restricted to be one on each side of the interaural axis, or to be completely free.

The standard deviations could be chosen as free variables, or they could be restricted

to be equal. They can not be fixed, as unlike the means, there is no a priori information available on what they might be. Additionally, measuring the spread is one of the things the statistical description is intended for. The choice of equal variances or not corresponds to the question of which error mechanism works first, spread or reversal. If the spread is first, the variances should be expected to be equal. Otherwise, there is no reason to believe that they necessarily should be equal.

The most restrictive variant that could be useful is to choose the reversal rate as a free parameter, fix the means at the presented angle and its reversal, and restrict the variances to be equal. The least restrictive variant is, of course, to choose all parameters as free. The first of these cases leads to a pdf

$$f(x, p_r, a, a_r, \sigma) \;=\; (1 - p_r) f_n(x, a, \sigma) + p_r f_n(x, a_r, \sigma) \tag{6.4}$$

where p_r is the reversal rate, a is the angle of the presented direction and a_r its reversal, and σ the common standard deviation. The second case, all parameters free, give the pdf given in equation 6.3.

6.3.3 Tri-normal statistical model

For binaural reproduction with reflections present, it is possible that the answers will cluster also in the direction of a dominant reflection, in addition to the presented direction and its reversal, giving a three-peaked distribution. In cases like this, the dual-normal model may easily be extended to a *tri-normal* model, composed of three Gaussian distributions along the lines described above.

6.3.4 Parameter estimation

For the dual-normal statistical model to be useful as a tool for the characterisation of listening test results, the values of the statistical parameters that will fit the model to the data have to be found. For many distributions, like the Gaussian, estimates for the parameters, the mean and the standard deviation, can be computed from the data by means of simple formulas [Høyland, 1989b]. Such formulas have not been developed for the dual-normal distribution,[4] and estimates of the model parameters must therefore be found by other methods.

The approach chosen here was to adjust curves representing the statistical model to curves representing the data until a best fit is found. The parameters were then chosen as the model values corresponding to the best fit. Minimum squared error was chosen as the criterion for best fit.

The curve-fitting was done with help of the Matlab function *curvefit()*, which solves non-linear least squares problems. It was used in conjunction with functions implementing the pdf and cdf of the dual-normal distribution. A choice was made to fit the cdf of the model to the cumulative histogram of the data. These are monotonically

[4] It is not clear that such formulas can be found, either.

increasing curves, and it was believed that this might ease the fitting process and make it more robust.

The parameters in the model have finite or semi-finite domains. The reversal rate $p_r \in [0, 1]$. In practice, it should not reach the endpoints of this interval, because if this would happen, the effects of the mean and standard deviation in one part of the distribution functions of the model would be "invisible", and this may cause convergence problems for the curve fitting algorithm. The standard deviations should be positive, $\sigma \in (0, \infty)$. The means were limited to an interval of $\pm 180°$ from the interaural axis (centered on the direction $-90°$ for presented directions in the left half plane, on $90°$ for presented directions in the right half plane). In order to impose these constraints, the curve-fitting algorithm was used through wrappers that implemented suitable mappings from $(-\infty, \infty)$ to the respective correct intervals. The mapping functions were implemented using exponential and trigonometric functions [Edwards and Penney, 1990], and were chosen to be smooth, without discontinuities that might have disturbed the curve-fitting.

6.4 Applying the dual-normal model to the data

The statistical model discussed in the previous section (6.3) was applied to the data from the listening tests. Several variants of the model were tried.

6.4.1 Variants of the statistical model tried

The dual-normal statistical model has five parameters: Reversal rate, two mean values and two standard deviations (equation 6.3). As discussed in section 6.3.2, restrictions may be placed on these parameters. This may be done if there exists a priori information on what values these parameters should have, or if special approaches to the data are needed.

For the characterisation of the listening test data, three variants of the dual-normal model were tried. A tri-normal model (section 6.3.3) was also tried.

Model A: Reversal rate and common variance free: The mean values were restricted to lie at the presented direction and its reversal, and the variance was restricted to be common for the reversed and unreversed answers.

Model B: Symmetric means: The mean values were restricted to be symmetrically placed with respect to the interaural axis. Except for that, all parameters were free.

Model C: All parameters free: No restrictions on any of the statistical parameters

Model D: Tri-normal model: A tri-normal model, with the three mean values restricted to the presented direction, its reversal and the direction towards the reflecting wall.

All these model variants were fitted to the data and parameters estimated as described in section 6.3.4. The fitting of models C and D are further commented below.

Model C: All parameters free

For ten cases (of eighty) the model fitting with all parameters free was manually aided by imposing extra restrictions on the parameters. An example of this is shown in figure 6.7. This was cases where the two clusters of answers were clearly separated and where the reversal rate found by the automated fitting did not correspond to the reversal rate evident from the data. The reversal rate was then locked to the value found from the relative frequency cumulative histogram, and the fitting rerun. In many cases this locking caused convergence problems for the fitting. These were solved by restraining the other parameters to suitable intervals around the values found for them from the first fitting of the model. It is not known why the reversal rate found by the parameter estimation did not equal the one found by inspection of the cumulative histogram.

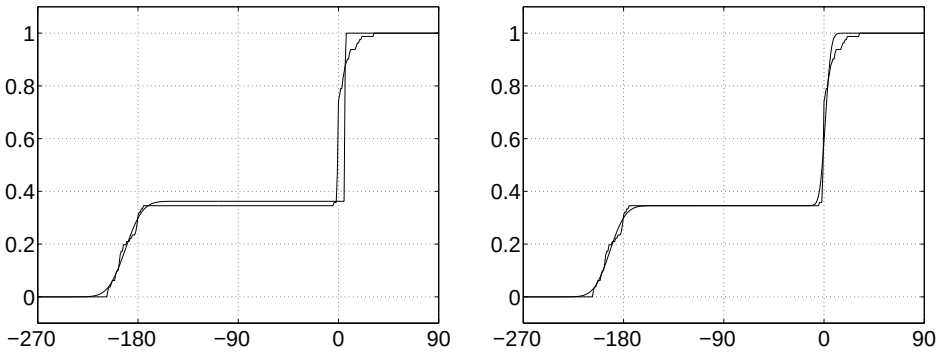


Figure 6.7: Fitting of statistical model to data. Relative frequency histogram of data, and fitted cdf (smooth curve). Left: Automated fitting. The reversal rate is found from the height of the middle horizontal part of the fitted cdf (cfr. section 6.3 and figure 6.6). This value differs from that of the relative frequency cumulative histogram to which the cdf has been fitted. Right: Fitting with the reversal rate manually fixed to the value found from the relative frequency cumulative histogram.

For this model, with all parameters free, there were cases where both means ended up lying at the same side of the interaural axis. Half of these cases were answers for the direction $-90°$ (left). For these, no mirroring should be expected, as the presented direction was at the interaural axis, which is usually assumed to be the symmetry axis for reversals (section 2.2.2). But there were also four other such cases, one for the anechoic setup and three for the setup with two walls. Distributions for one of these, the presented direction $-45°$ in the setup with two walls, is shown in figure 6.8.

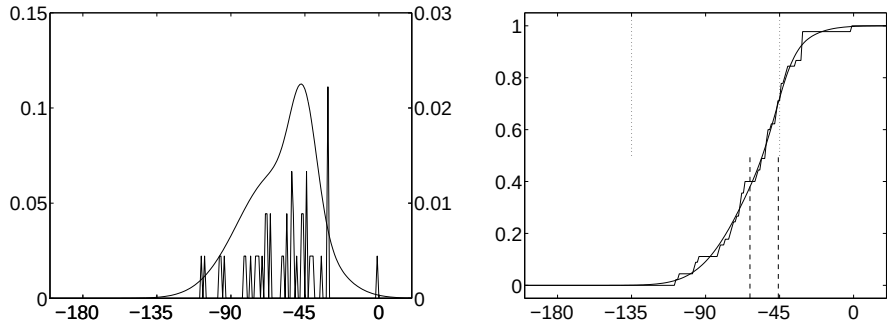


Figure 6.8: Both means from fitting of statistical model lying on the same side of the interaural axis. Setup with two reflecting walls, presented angle $-45°$. Left: Relative frequency histogram (left y-axis) and fitted dual-normal probability distribution (right y-axis). Right: Relative frequency cumulative histogram and fitted cumulative distribution. The presented direction and its mirror direction (around $-90°$) are marked by dotted lines (upper half of figure). The two means obtained from the fitting are marked by dashed lines (lower part of figure).

In all cases the best fit according to the error criterion (minimum squared error) is chosen. The parameters found are therefore, in this respect, those that most accurately describe the data. The fact that the two means describing the perceived directions end up on the same side of the interaural axis may reveal interesting effects. It does however violate the underlying model of reversals as symmetry around the interaural axis (section 6.1), and this may limit the usefulness of this choice of free parameters for some situations, e.g. computation of reversal rate.

Model D: The tri-normal model

The tri-normal model was applied to the results from the anechoic setup and from the two setups with a wall in the close left position. The maxima (mean values) of the distribution were restricted to be at the presented direction, the reversal of the presented direction and at the direction of the center of the reflecting wall. The purpose of using this model with these restrictions was to investigate whether localisation towards the reflections occured (section 6.1).

The fitting of this model had convergence problems, and did not give a good fit in some cases. Directions from front through the left side to rear was tried. Of these, the estimates for the presented directions $0°$, $-22.5°$ and $-180°$ were rejected due to convergence problems and ill fitting.

6.4.2 Comparison of the statistical models

Examples of fitted dual-normal and tri-normal distributions for four presented directions in the left half plane are shown in figure 6.9. Error sums (chapter 6.3.4) from the fittings are shown in figure 6.10.

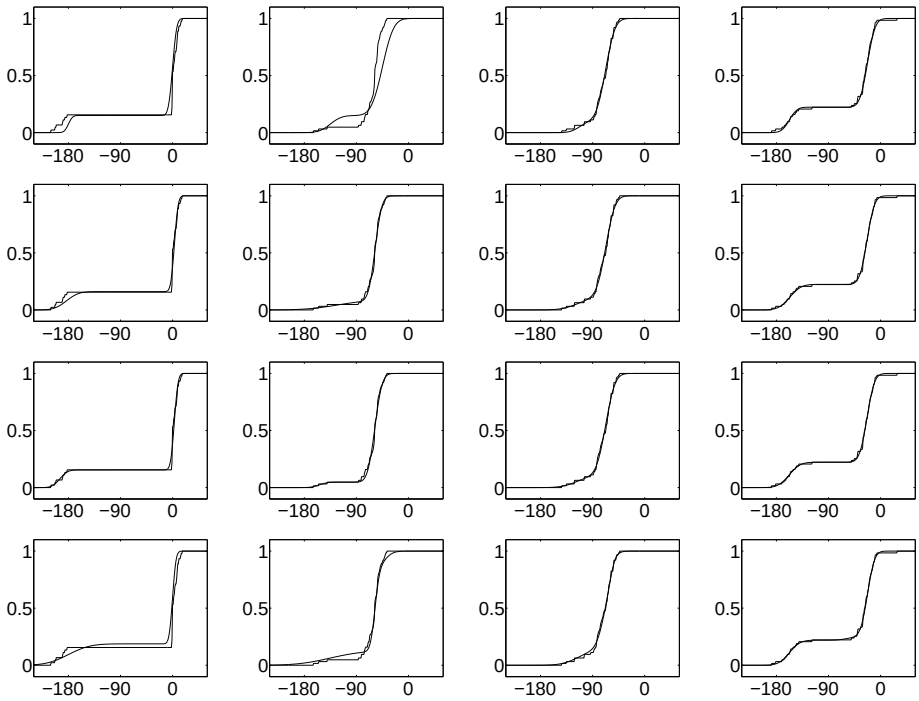


Figure 6.9: Comparison of statistical models fitted to data for the setup with one wall. Relative frequency cumulative histograms and fitted cumulative distributions are shown together.
Upper row: Model A, dual-normal model with means fixed to the presented direction and its reversal, both variances equal. Second row: Model B, dual-normal model with symmetry of means enforced. Third row: Model C, dual-normal model with all parameters free. Lower row: Model D, tri-normal model, with means fixed at the presented direction, its reversal and the direction towards the reflecting wall.
Presented direction, column-wise from left to right: 0°, −45°, −67.5° and −157.5°.

Model C, the dual-normal model with all parameters free, was found to fit the data best. This desirable property made it a natural choice for cases where its use was not affected by its sometimes poorer correspondence with the underlying reversal model. Several of the results given in the next chapter are based upon the parameters found using this model.

Model B, the dual-normal model with symmetry of means enforced, fitted the data

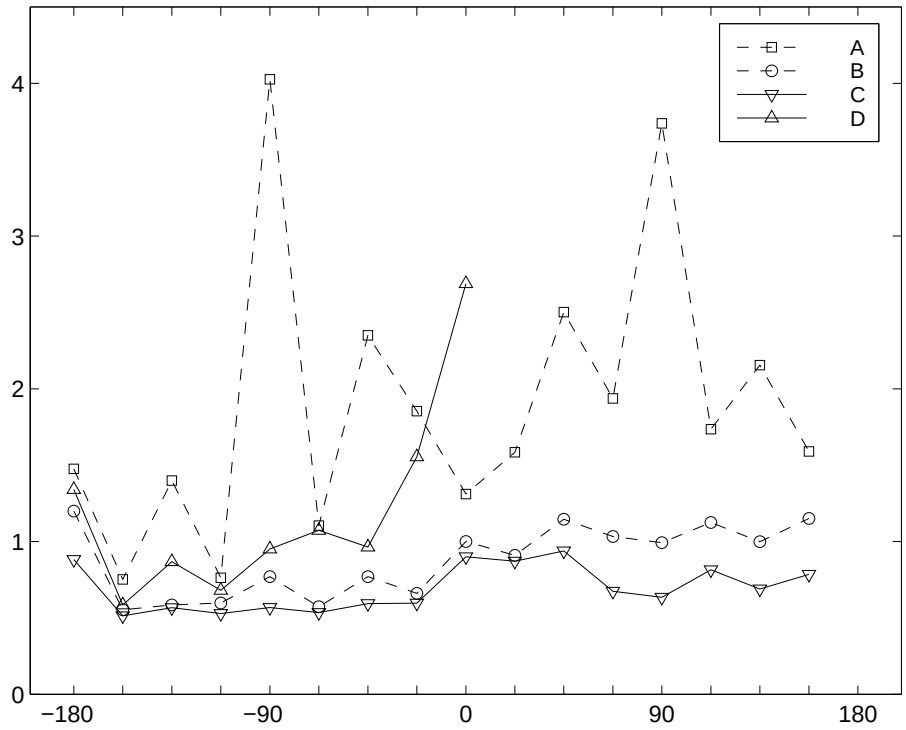


Figure 6.10: Error sums from the fitting of statistical models, mean values for all five setups. Values are normalised to the minimum error in the frontal direction. Presented direction along the horizontal axis, error along the vertical axis. Line "A": Model A, dual-normal model with means fixed to the presented direction and its reversal, both variances equal. Line "B": Model B, dual-normal model with symmetry of means enforced. Line "C": Model C, dual-normal model with all parameters free. Line "D": Model D, tri-normal model, with means fixed at the presented direction, its reversal and the direction towards the reflecting wall (three setups and nine presented directions).

nearly as well as model C, and gave generally similar results. This model has the advantage of corresponding to the reversal model used, due to the requirement of symmetry of the mean values around the interaural axis (section 6.1). Results found from this model are used to study reversal rates, and to complement the results found from model C.

In figure 6.11, the modeling error for the models B and C are shown. As can be seen, the fit is generally better for the latter, as it gives a lower error for all setups. It

should be noted that the relative ordering of the setups differs for the two models. For model B (symmetric means), the far wall setup has the lowest error, while for model C (all parameters free), the anechoic setup has the lowest error.

It should also be noted that the error is lower for the least complex setups, and higher for the more complex setups. The poorest fit is obtained for the two-wall setup, thereafter the one-wall setup with the wall close, the one-wall setup with corrective filtering, and last, with the lowest error, the far-wall setup and the anechoic setup. The exception to this is the interchange of the far wall setup and the anechoic setup for model B.

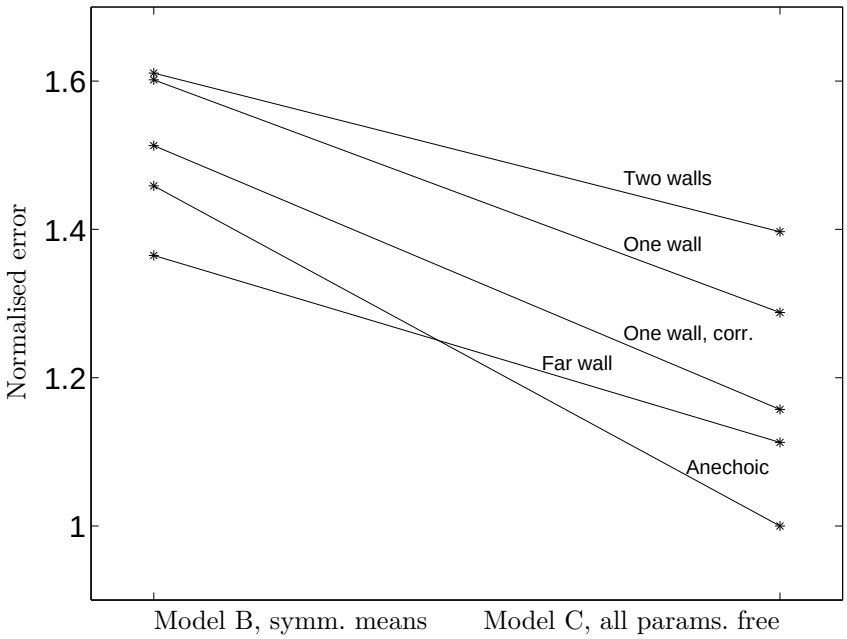


Figure 6.11: Normalised modelling error for the five setups, from the fitting of models B and C.

Model D, the tri-normal model, fitted quite well for most of the directions for which it was tried. (As mentioned earlier, the estimates found for the presented directions 0°, −22.5° and −180° were rejected due to convergence problems and ill fitting. These three directions also showed a larger error, as can be seen from figure 6.10.) This model has been used to investigate localisation towards the reflections.

Model A, with fixed means and common variance, was the model with the most restricted parameters. This model was not found to fit the data well, the results from fitting this model to the data showed that in many cases the fixed means and the equal standard deviations were not appropriate. An example, where this model is

compared to model C, is shown in figure 6.12. This model was not used for the rest of the work.

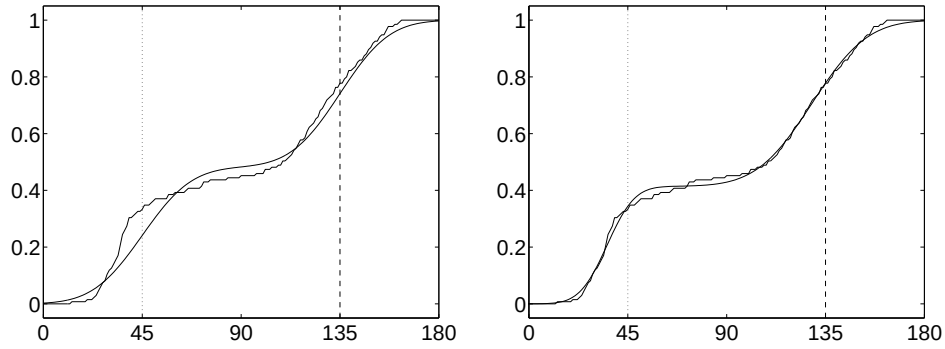


Figure 6.12: Statistical models A and C compared. Cumulative distribution function of dual-normal statistical model (smooth solid line) fitted to relative frequency histogram of results from listening test. Left: Model A, reversal probability and (common) variance as free parameters. Right: Model C, all parameters free. Presented angle marked with dashed lines, mirrored presented angle marked with dotted lines. It can be seen that the fit in the left plot (model A) is the worse. For model A, it is assumed that the variances of the reversed and unreversed answers are equal. But the different steepness of the two main slopes of the cumulative histogram shows that this is not the case. For model A it is further assumed that the unreversed and reversed mean values should correspond to the presented direction and its reversals. This is however not the case, as can be seen from the fact that the steepest parts of the cumulative histogram are not at these directions (the dashed and dotted lines) (cfr. section 6.3 and figure 6.6).

6.5 Discussion

A statistical model for the localisation process for directions in the horizontal plane has been developed. The model allows for the statistical description of localisation results containing reversals. It also makes it possible to describe the reversals properly, even independently of the non-reversed answers. The model has been applied to the data from the listening tests, and has been shown to fit the data well for suitable choices of free parameters in the fitting. The good fit obtained confirm the assumptions underlying the model; that localisation may be modeled as a combination of two Gaussian distributions.

A main feature of this statistical model is that it solves the problem of having to determine whether a given answer is reversed or not. It is no longer necessary to determine this on a per-answer basis in order to come forth to a tractable description of the data. Instead, the reversal rate may be determined for a set of answers as

a whole, as a statistical parameter, and the reversal effect separated from the other parameters of the answers. Resolving of reversals by remirroring is a process that may introduce errors, as it is in many cases not possible to know for certain whether an answer is reversed or not. It is not necessary to do this using this model, and the estimates of spread, mean and reversal rate produced should therefore be better.

The parameters obtained, mean values, spread and reversal rate, are useful descriptors of results from listening tests on localisation, and may be used for the comparison of different datasets.

Although the model may be used to describe data, it is not clear that the resulting parameters can be used for the computation of statistical significance measures, as for testing hypotheses regarding the underlying localisation process or regarding larger populations of which the gathered results are a sample. For distributions like the Gaussian, estimates for the mean and the spread may be directly computed from the data by means of given formulas. Given this computation, it is known that the estimates possess certain statistical properties that make it possible to draw conclusions that are certain within a given probability. The estimates of the statistical parameters of the dual-normal model are found with the help of least-squares fitting, and it is not known what statistical properties these estimates have. It does therefore not follow that the same formulas for hypotheses testing as for e.g. the Gaussian distribution may be applied to these estimates. This issue remains to be solved.

An alternative method for obtaining statistical significance measures might be to apply Monte Carlo simulations to the estimates of the statistical parameters of the dual-normal model. It is probable that this would lead to a usable method for hypotheses testing, but it has not been tried here.

Chapter 7

Results

The data from the listening tests have been analysed with respect to the criteria described in section 6.1 in order to produce results that may be used to evaluate the effects of the reflections. The influence of the reflections may then be assessed by comparing the results from the various playback setups. Reproduction under what is supposed to be ideal conditions, free field, is chosen as a reference condition (section 5.2).

7.1 Overview of the data

The answers gathered from the listening tests make up a large and complex dataset which is not easily characterised. Here, an overview of the data is given that makes it possible to appreciate the data as a whole and which gives a foundation for the more detailed analyses to follow.

In figure 7.1, the distribution of answers for the different setups are presented as scatter-plots. [1] For this figure, the data have been simplified. The perceived directions, originally given with an angular resolution of one degree, have been rounded to the nearest presented direction (multiples of $\pm 22.5°$, section 5.4.2).

In figure 7.2, the data are described by fitting statistical model C (section 6.4.1) to them.

7.2 Average perceived direction

The average perceived directions are taken to be the mean values found from the statistical description of the data (section 6.4). They are presented in figures 7.3 and 7.4. These figures also show the distribution of answers between the two perceived directions.

[1] Readers unfamiliar with this presentation form may find appendix C useful.

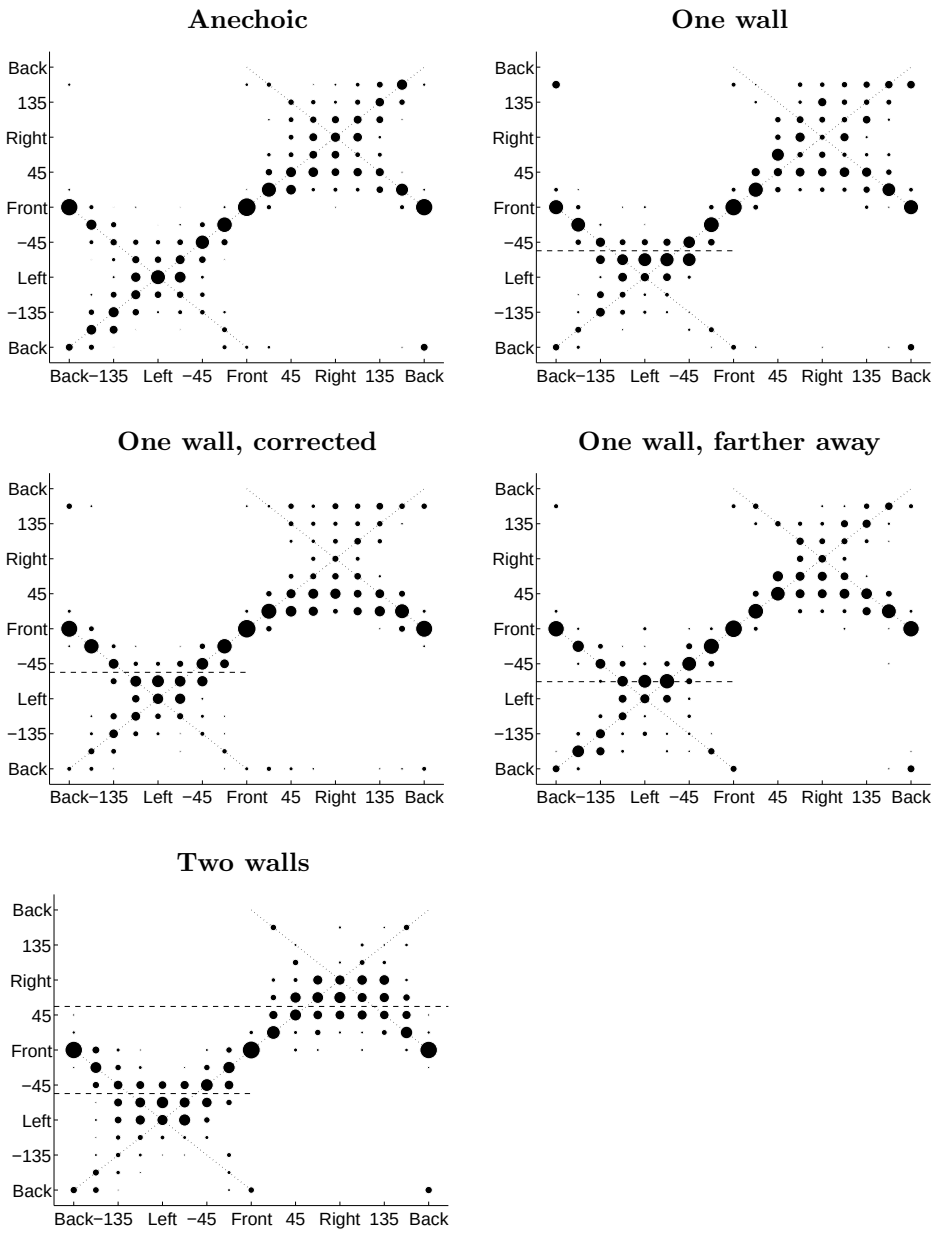


Figure 7.1: Answers from listening tests, grouped by test condition. Presented direction along the horizontal axes, perceived direction along the vertical axes. The answers are marked by circles, the area of which correspond to the number of answers normalised to the total number of answers for that presented direction. Perceived directions are rounded to the nearest presented direction. The positions of the walls are marked with dashed lines. Correct and reversed localisation are marked with dotted lines.

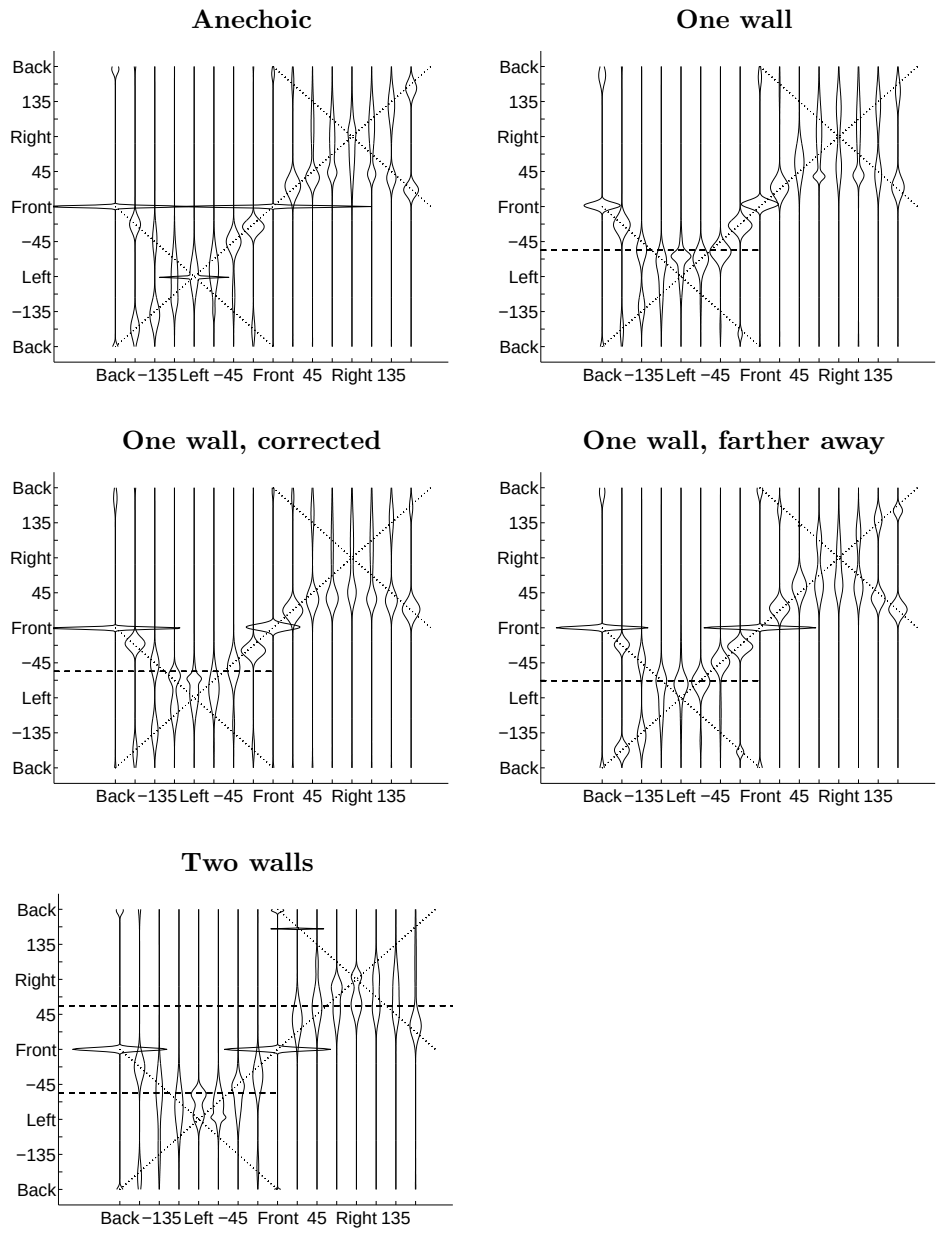


Figure 7.2: Statistical model C (all parameters free) fitted to the answers from the five different listening test setups. Presented directions along the horizontal axes, perceived direction along the vertical axes. For each presented direction, the width of a double vertical line shows the density of answers as a function of perceived direction. The density is estimated using the dual-normal probability distribution with all five parameters free. The positions of the walls are marked with dashed lines. Correct and reversed localisation are marked with dotted lines.

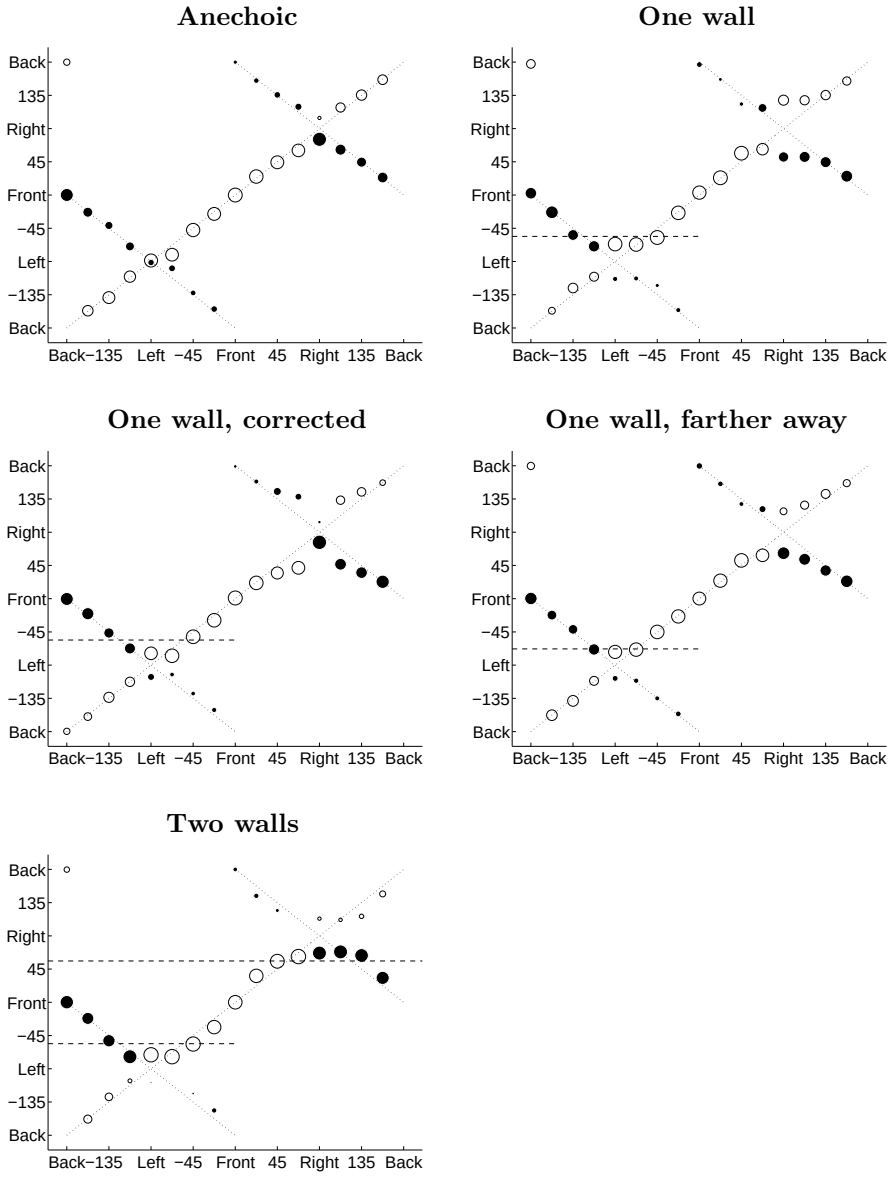


Figure 7.3: Perceived directions and reversal rates estimated by fitting model B (symmetric means) to the listening test answers. Presented direction along the horizontal axes, perceived direction along the vertical axes. Values for unreversed answers marked with open circles, values for reversed answers marked with filled circles. The distribution of answers among the two perceived directions is shown by letting the area of the circles correspond to their relative number of answers. Directions towards the center of the reflecting walls are marked with dashed lines.

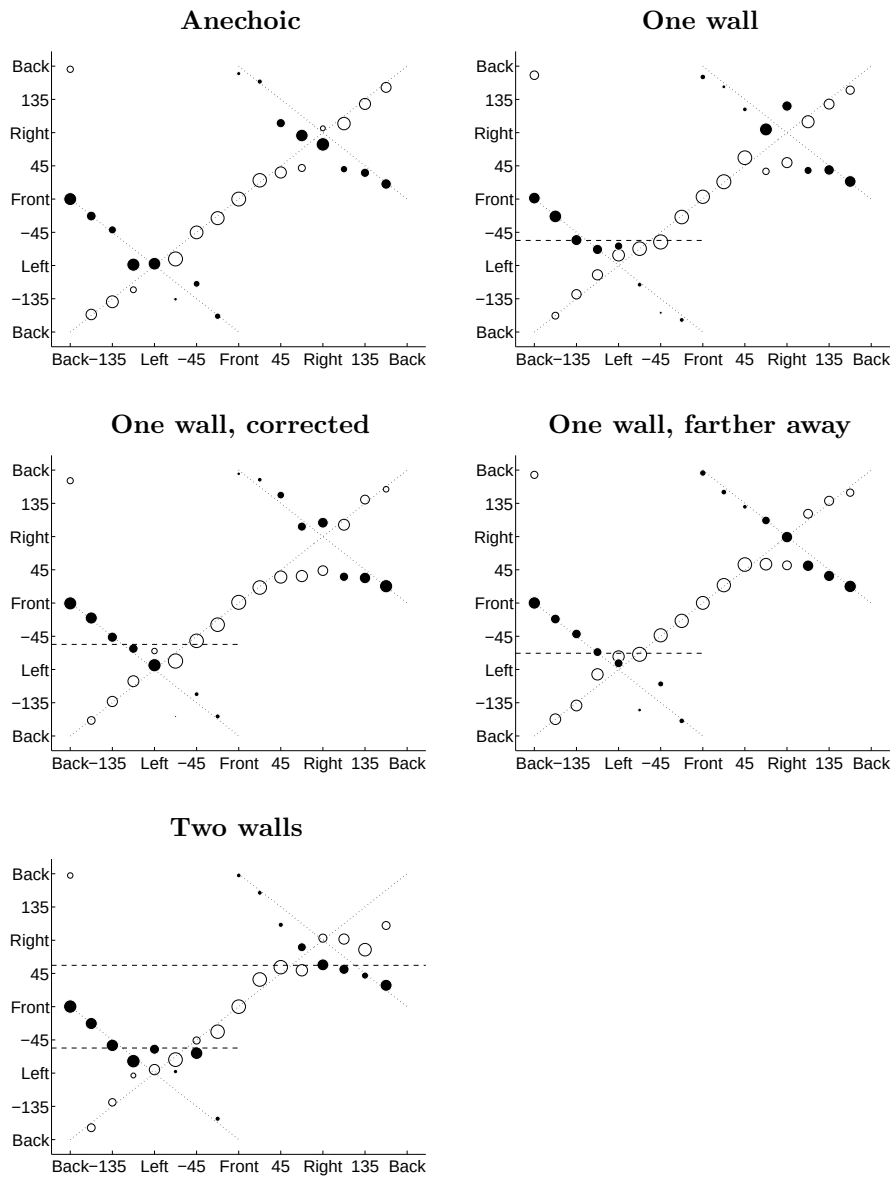


Figure 7.4: Perceived directions and reversal rates estimated by fitting model C (all parameters free) to the listening test answers. Presented direction along the horizontal axes, perceived direction along the vertical axes. Values for unreversed answers marked with open circles, values for reversed answers marked with filled circles. The distribution of answers among the two perceived directions is shown by letting the area of the circles correspond to their relative number of answers. Directions towards the center of the reflecting walls are marked with dashed lines. (Note that for the direction "left" in the anechoic setup the markers are overlapping.)

7.3 Deviation of average perceived direction

Deviation of the average perceived directions from reference directions may be used as a means for comparing the different setups. It is natural to use the angular difference between the directions as a measure of this deviation.

The computation of a signed deviation poses a problem, namely how to compute the sign in a meaningful way. Several approaches are possible. The alternative corresponding directly to the coordinate system used, positive sign for clockwise deviation and negative sign for counter-clockwise deviation may be rejected, as there is no general correspondence between this sign convention and the localisation mechanisms assumed to be at work.

Another alternative is to set the sign according to whether the deviation is towards or from the direction of a wall. But as the number of walls and their position vary between tests, this method will not allow the comparisons sought for. Other alternatives are to set the sign according to whether the deviation is towards the front or towards the rear, whether the deviation is towards or from the interaural axis, or, for completeness, whether the deviation is towards the left or towards the right. These sign conventions are conflicting. All of them may be suitable for some cases, but none is suitable for all.

The solution chosen here is to avoid the sign problem altogether by computing deviations as the unsigned angle (in degrees) between the perceived directions and the corresponding reference directions.

For each presented direction there are two average perceived directions, due to the reversal effect. The reversal effect is eliminated from the deviations by computing the deviations for the reversed and unreversed answers separately, and by computing the deviations for each of these with respect to the closer of the corresponding presented direction and its reversal. The deviations for the two average perceived directions are then summed according to their weights, as found from the statistical modelling.

As discussed in chapter 6.1, deviation of the average perceived direction from the presented direction is one of the criteria used to evaluate the data. Deviations from the presented directions were therefore computed for all five test setups. The deviations are given in tables 7.1 (average perceived directions found from model B) and 7.2 (average perceived directions found from model C).

A more direct comparison of the answers from setups with reflections to the answers from the anechoic setup is of interest. To allow for this, deviations were also computed with respect to the average perceived directions for the anechoic setup. In this case, the deviation of the unreversed perceived directions was computed with respect to the unreversed perceived directions for the anechoic setup, and similarly for the reversed answers. The results are given in tables 7.3 (model B) and 7.4 (model C).

Table 7.1: Deviation of average perceived directions from presented direction or reversed presented direction. Computed using model B (symmetric means). Deviation is measured as unsigned angular difference in degrees. Minimum values in bold, maximum values in italics. The total mean is the mean for all presented directions. The left and right means are means for presented directions in the left and right half planes, excluding the directions 0° and −180°. Similarly, the front and rear means are for presented directions in the frontal and rear half planes, excluding the directions 90° and −90°.

Presented direction	Anechoic	One wall	One wall, corrected for	One wall, farther away	Walls at both sides
−180.0	**0.10**	*2.44*	0.37	0.24	0.20
−157.5	0.82	0.82	*2.16*	**0.29**	0.72
−135.0	3.89	*9.10*	**1.45**	3.46	6.99
−112.5	2.00	1.91	**0.11**	1.27	*6.20*
−90.0	**1.37**	*23.59*	15.98	17.94	18.93
−67.5	*13.25*	**0.48**	9.76	1.45	6.15
−45.0	2.43	*12.56*	6.54	**0.01**	11.47
−22.5	3.06	1.66	6.78	**1.49**	*11.04*
0.0	**0.03**	*3.27*	0.93	0.21	0.07
22.5	2.53	**0.90**	1.16	1.94	*13.35*
45.0	**0.68**	*11.70*	10.19	6.80	10.59
67.5	7.05	**5.47**	*25.66*	8.74	5.53
90.0	14.47	*38.44*	**13.67**	28.37	23.33
112.5	6.02	15.84	*20.90*	14.15	**0.80**
135.0	**0.41**	**0.41**	9.69	6.89	*18.52*
157.5	1.27	3.08	**0.40**	1.03	*10.66*
Left mean	3.83	7.16	6.11	**3.70**	*8.78*
Right mean	**4.63**	10.83	11.67	9.70	*11.82*
Front mean	4.15	5.15	8.72	**2.95**	*8.31*
Rear mean	**2.07**	4.80	5.01	3.90	*6.30*
Total mean	**3.71**	8.23	7.86	5.89	*9.03*

Table 7.2: Deviation of average perceived directions from presented direction or reversed presented direction. Computed using model C (all parameters free). Deviation is measured as unsigned angular difference in degrees. Minimum values in bold, maximum values in italics. The total mean is the mean for all presented directions. The left and right means are means for presented directions in the left and right half planes, excluding the directions 0° and −180°. Similarly, the front and rear means are for presented directions in the frontal and rear half planes, excluding the directions 90° and −90°.

Presented direction	Anechoic	One wall	One wall, corrected for	One wall, farther away	Walls at both sides
−180.0	0.84	*5.10*	2.96	1.66	**0.48**
−157.5	0.86	0.79	2.06	**0.46**	*2.14*
−135.0	3.82	*8.67*	**1.48**	3.45	6.94
−112.5	*19.47*	**5.37**	6.27	10.15	7.79
−90.0	**2.15**	*18.30*	8.61	14.13	18.03
−67.5	*14.17*	**0.67**	11.66	4.03	5.77
−45.0	**4.67**	13.37	6.80	5.54	*13.73*
−22.5	2.93	2.34	6.49	**1.72**	*10.66*
0.0	0.89	*4.87*	1.18	0.91	**0.29**
22.5	2.78	**1.27**	2.34	2.58	*13.03*
45.0	*18.42*	11.26	10.17	**6.88**	10.52
67.5	20.20	20.50	*22.33*	**9.82**	16.24
90.0	**14.75**	*38.24*	31.02	14.98	23.59
112.5	14.70	14.32	17.20	**13.80**	*18.87*
135.0	7.49	**5.93**	8.82	6.35	*24.89*
157.5	4.04	4.19	**0.73**	2.06	*18.86*
Left mean	6.87	7.07	6.20	**5.64**	*9.29*
Right mean	11.77	13.67	13.23	**8.07**	*18.00*
Front mean	9.15	7.75	8.71	**4.50**	*10.03*
Rear mean	7.32	6.34	5.64	**5.42**	*11.43*
Total mean	8.26	9.70	8.76	**6.16**	*11.99*

Table 7.3: Deviation of average perceived directions from average perceived directions for the anechoic setup. Computed using model B (symmetric means). Deviation is measured as unsigned angular difference in degrees. Minimum values in bold, maximum values in italics. The total mean is the mean for all presented directions. The left and right means are means for presented directions in the left and right half planes, excluding the directions 0° and −180°. Similarly, the front and rear means are for presented directions in the frontal and rear half planes, excluding the directions 90° and −90°.

Presented direction	One wall	One wall, corrected for	One wall, farther away	Walls at both sides
−180.0	*2.34*	0.47	0.14	**0.10**
−157.5	**0.01**	*2.97*	1.11	1.54
−135.0	*12.99*	5.34	**0.43**	10.88
−112.5	**0.08**	2.10	0.73	*4.20*
−90.0	*22.22*	**14.60**	16.57	17.55
−67.5	*13.73*	**3.50**	11.81	7.10
−45.0	*10.13*	4.11	**2.43**	9.04
−22.5	**1.41**	3.71	1.58	*7.98*
0.0	*3.24*	0.89	0.18	**0.04**
22.5	1.64	3.69	**0.59**	*10.82*
45.0	*12.37*	9.52	**7.48**	11.27
67.5	1.58	*18.61*	1.69	**1.52**
90.0	*23.96*	**0.81**	13.89	8.86
112.5	9.82	*14.88*	8.13	**6.81**
135.0	**0.00**	9.27	6.48	*18.93*
157.5	1.81	0.87	**0.24**	*9.39*
Left mean	*8.65*	5.19	**4.95**	8.33
Right mean	7.31	8.23	**5.50**	*9.66*
Front mean	6.30	6.29	**3.68**	*6.82*
Rear mean	3.86	5.13	**2.47**	*7.41*
Total mean	7.33	5.96	**4.59**	*7.88*

Table 7.4: Deviation of average perceived directions from average perceived directions for the anechoic setup. Computed using model C (all parameters free). Deviation is measured as unsigned angular difference in degrees. Minimum values in bold, maximum values in italics. The total mean is the mean for all presented directions. The left and right means are means for presented directions in the left and right half planes, excluding the directions 0° and −180°. Similarly, the front and rear means are for presented directions in the frontal and rear half planes, excluding the directions 90° and −90°.

Presented direction	One wall	One wall, corrected for	One wall, farther away	Walls at both sides
−180.0	*3.61*	2.21	0.58	**0.39**
−157.5	**0.67**	*2.75*	1.15	2.38
−135.0	*12.28*	5.15	**0.26**	10.41
−112.5	17.14	*21.31*	**13.45**	13.78
−90.0	*15.77*	**6.08**	11.60	15.50
−67.5	*14.52*	**2.85**	11.47	9.08
−45.0	*14.19*	6.32	**2.09**	13.58
−22.5	1.92	3.69	**1.59**	*7.89*
0.0	*3.29*	1.17	1.33	**0.90**
22.5	**2.40**	4.78	2.45	*10.63*
45.0	*19.84*	**13.05**	17.42	16.16
67.5	7.52	10.26	*15.96*	**6.57**
90.0	*27.50*	20.63	**12.71**	12.83
112.5	**2.21**	4.30	*13.25*	10.41
135.0	**2.12**	5.22	4.46	*32.83*
157.5	3.52	2.42	**2.05**	*18.45*
Left mean	*10.93*	6.88	**5.94**	10.37
Right mean	9.30	**8.66**	9.76	*15.41*
Front mean	9.10	**6.01**	7.47	*9.26*
Rear mean	5.93	6.19	**5.03**	*12.66*
Total mean	9.28	7.01	**6.99**	*11.36*

7.4 Localisation towards the reflections

A possible effect of the reflections is to move the perceived directions closer to the directions towards the walls, due to the precedence effect (section 2.2.3). This question has been approached using three different methods.

For the anechoic setup and the two setups with a wall in the close left position, a tri-normal statistical model was applied (model D, section 6.4). The purpose of this was to find the distribution of answers among the direction towards the wall and the presented direction and its reversal. To obtain this, the three mean values of the model were restricted to these directions. The distribution of answers among these three directions, for most directions in the left half plane, is given in table 7.5.

The number of answers in the vicinity of the walls has been investigated. Figure 7.5 shows the number of answers lying in angular sectors centered at the walls, corrected for the number of presented stimuli in these sectors and normalised to the same results for the anechoic setup. (This is computed directly from the answers, without any statistical modelling.) Numbers larger than one show that relatively more answers are localised to the sector than for the anechoic setup.

The angular distance between the perceived directions and the direction of the walls has been computed for the setups with walls, and normalised to the corresponding values for the anechoic setup. The distances were found using statistical model C. Distances for the unreversed and reversed average perceived directions were computed separately and summed according to the weights found from the model. The results are shown in figure 7.6. These figures show how close to the wall the average perceived directions are, relative to the average perceived directions for the anechoic setup. Numbers smaller than one show that the distance is less than for the anechoic setup.

7.5 Localisation to front and rear half planes

The distribution of answers between the frontal and rear half planes has been studied. For each setup, the ratio of answers in the frontal half plane to the number of answers in the rear half plane has been computed. These results have been normalised to the equivalent ratio for the presented directions. The result is a measure of the front/back distribution of the answers relative to the front/back distribution of the presented directions.

The computation was done on the answers themselves, with no statistical modelling applied. Answers for the presented directions straight left and straight right ($\pm 90°$) were excluded from the data set.

For the anechoic setup, the value was 2.32, for the setup with one wall 3.29, for the setup with one wall and corrective filtering 2.86, for the far wall setup 3.20 and for the setup with two walls 5.69.

Table 7.5: Distribution of answers among the presented direction, its reversal and the direction towards the wall, as found by fitting model D (tri-normal).

Presented direction	Directions	Relative number of answers		
		Anechoic	One wall	One wall, corr.
	−45.0	0.51	0.17	0.34
−45.0	−56.3	0.36	0.70	0.57
	−135.0	0.13	0.14	0.09
	−67.5	0.62	0.74	0.72
−67.5	−56.3	0.00	0.10	0.00
	−112.5	0.38	0.15	0.28
	−90.0	0.18	0.36	0.02
−90.0	−56.3	0.07	0.54	0.27
	−90.0	0.75	0.09	0.71
	−112.5	0.60	0.40	0.44
−112.5	−56.3	0.00	0.00	0.06
	−67.5	0.40	0.60	0.51
	−135.0	0.69	0.39	0.48
−135.0	−56.3	0.03	0.53	0.16
	−45.0	0.29	0.08	0.36
	−157.5	0.52	0.22	0.27
−157.5	−56.3	0.06	0.03	0.03
	−22.5	0.43	0.75	0.71

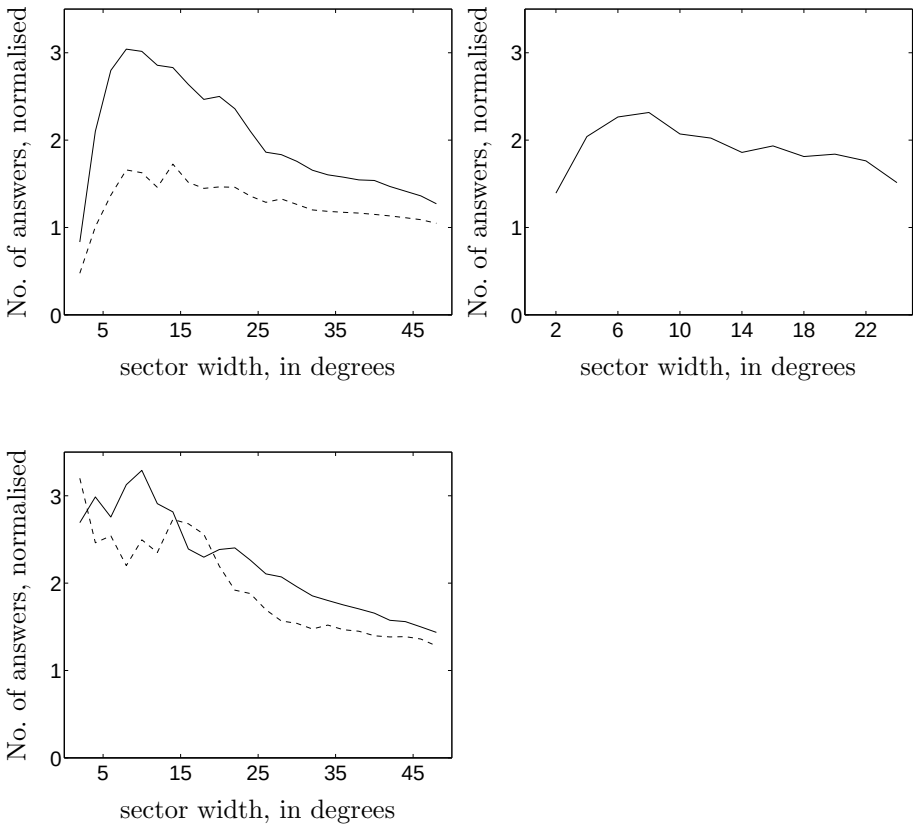


Figure 7.5: Relative number of answers in angular sectors centered at the reflecting walls. Upper left plot: Setup with one wall (solid line), and setup with one wall and corrective filtering (dashed). Upper right plot: Setup with a reflecting wall farther away. Lower left plot: Setup with two walls, left side (solid line) and right side (dashed line). Horizontal axes: Angular width of sector. Vertical axis: Relative number of answers, normalised to the anechoic setup. The angular width of the wall(s) is 24.4° for the close walls and 18.1° for the wall in the far position.

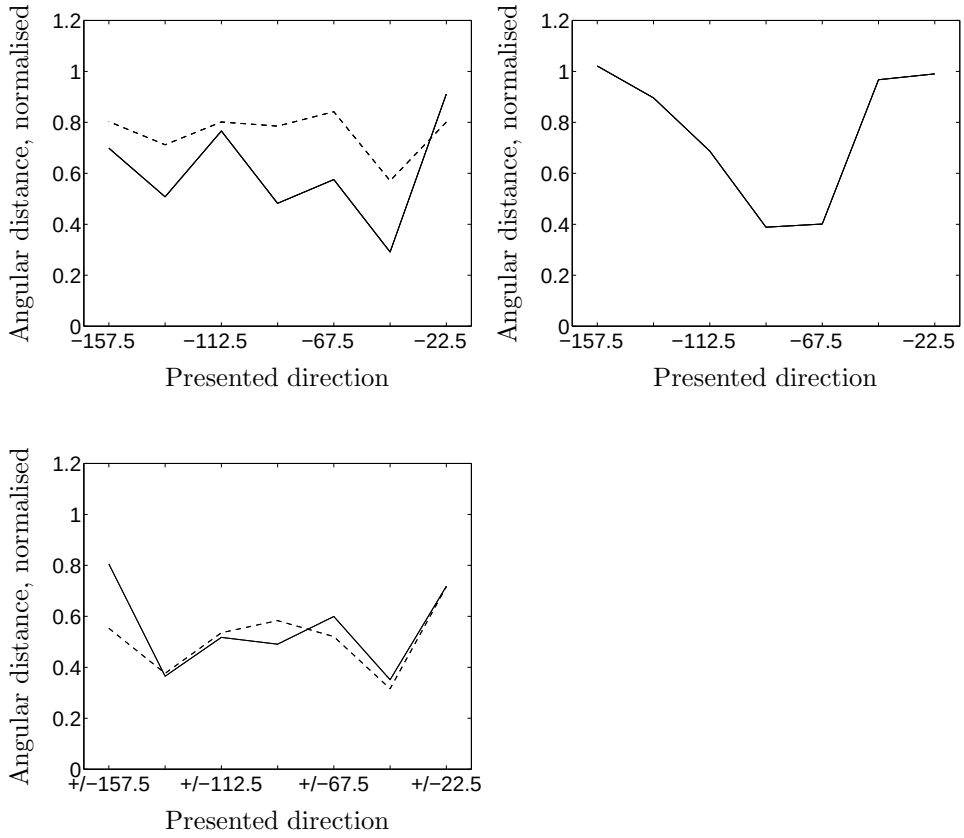


Figure 7.6: Distance between perceived directions and walls, normalised to the distances for the anechoic setup. Computed using statistical model C (all parameters free). Upper left plot: Setup with one wall (solid line), and setup with one wall and corrective filtering (dashed). Upper right plot: Setup with a reflecting wall farther away. Lower left plot: Setup with two walls, left side (solid line) and right side (dashed line). Horizontal axes: Presented direction. Vertical axes: Distance, normalised to the anechoic setup.

7.6 Reversals

The reversal rate has been assessed by statistical modelling using the dual-normal probability distribution. Model B (section 6.4.1) was used for this, as the requirement of symmetry of the means around the interaural axis used in this model corresponds with the reversal model used (section 6.4.2). The results are given in table 7.6.

Table 7.6: Reversal rates for all setups and presented directions. Minumum values for a presented direction are printed in bold, maximum values in italics. The lower four rows give the mean reversal rates for frontal directions and rear directions, mean reversal rates for all directions and normalised mean reversal rates for all directions.

Presented direction	Anechoic	One wall	One wall, corrected for	One wall, farther away	Walls at both sides
−180	0.81	**0.64**	0.81	0.75	*0.86*
−157.5	0.47	*0.78*	0.72	**0.46**	0.69
−135	**0.32**	0.55	0.50	0.44	*0.73*
−112.5	**0.39**	0.61	0.58	0.62	*0.93*
−90					
−67.5	*0.22*	0.12	0.11	0.14	**0.00**
−45	*0.16*	0.09	0.11	0.12	**0.03**
−22.5	*0.21*	**0.12**	0.13	0.16	0.14
0	0.09	*0.16*	**0.07**	0.20	0.11
22.5	*0.15*	**0.08**	0.13	*0.15*	0.14
45	0.21	0.10	*0.30*	0.12	**0.08**
67.5	0.25	*0.37*	0.22	0.23	**0.00**
90					
112.5	**0.58**	0.59	0.66	0.68	*0.94*
135	**0.49**	0.58	0.65	0.62	*0.90*
157.5	**0.55**	0.67	*0.85*	0.76	0.83
Frontal mean	*0.18*	0.15	0.15	0.16	**0.07**
Rear mean	**0.52**	0.63	0.68	0.62	*0.84*
Total mean	**0.35**	0.39	0.42	0.39	*0.46*
Normalised	**1.00**	1.12	1.19	1.11	*1.30*

7.7 Spread

As a measure of spread, the standard deviations of the perceived directions were used. The standard deviations were found using model C (all parameters free). For each presented direction a joint standard deviation for the unreversed and reversed answers was computed as the square root of the squared standard deviations summed according to their weights. The results are shown in figure 7.7.

Model D (tri-normal) and model B (symmetric means) were also tried. These gave results generally very similar those shown, although individual values and the relative ordering of the setups were somewhat different. The main difference was that model B did not give the low spread values at presented directions −90 degrees and 67.5 - 112.5 degrees for the two-wall setup, and that the setup with one wall and corrective filtering showed a peak value of 46 for the presented direction 90 degrees.

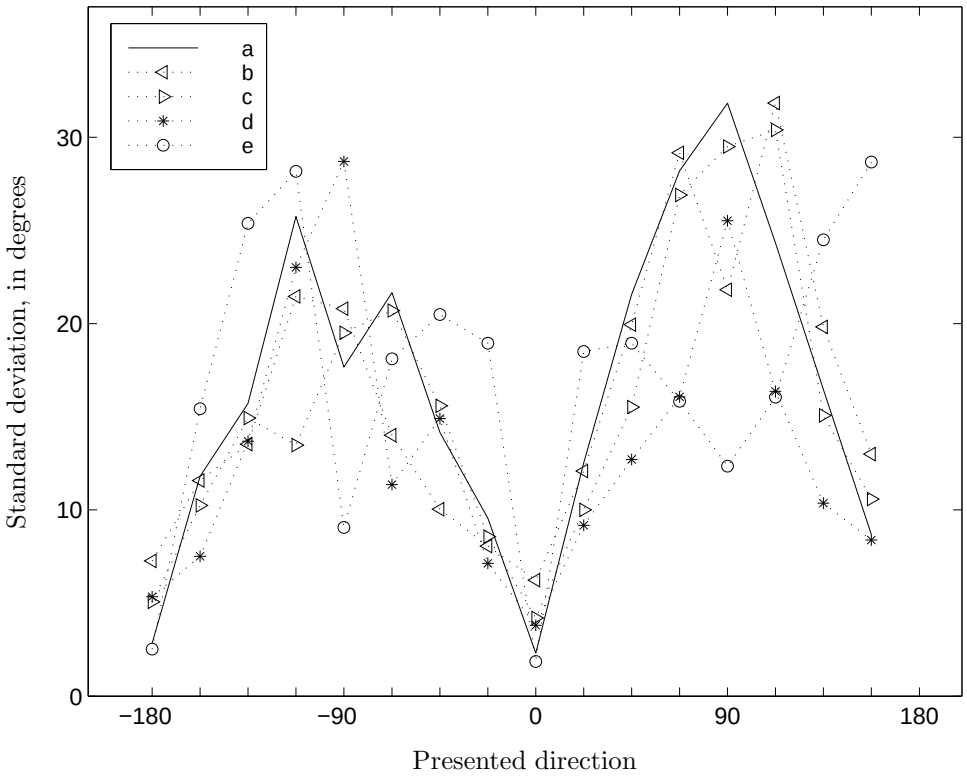


Figure 7.7: Spread of answers for the various setups, as computed by model C (all parameters free). Horizontal axis: Presented direction. Vertical axis: Joint standard deviation for the unreversed and reversed answers. Line types: a (solid line): Anechoic setup. b: Setup with one wall, c: setup with one wall and corrective filtering, d: Wall in far position, e: Two walls.

Chapter 8

Discussion

The results in chapter 7 describe listening test data from binaural reproduction under different conditions; in free field, with reflections present, and with a reflection present and reflection cancelling filtering applied (section 5.4.4). These results are discussed, to find and describe any differences in localisation caused by the various playback setups.

The criteria for analysis of the listening test data were outlined in section 6.1: Localisation, taken as the correspondence between the presented and perceived direction, is used as a performance measure. If these directions are equal, this is called correct localisation. Four aspects of deviation from correct localisation are studied: Deviation of the average perceived direction from the presented direction, spread of answers around the average perceived direction, reversals and localisation towards the reflections. The results given in chapter 7 were computed according to these criteria. Here, the effects of the reflections are assessed by comparing the results given in chapter 7 for the various playback conditions. Reproduction under what is supposed to be ideal conditions, free-field, is chosen as a reference condition (section 5.2).

After an overview (section 8.1), the deviation of the average perceived directions from the presented directions is discussed for all setups. Further, the average perceived directions for the non-anechoic setups are compared to those for the anechoic situation (section 8.2). The deviations may be explained as the result of localisation towards the reflections (section 8.3). Front/back localisation (section 8.4) and reversals (section 8.5) are investigated, and the spread of answers is briefly discussed (section 8.6). The effect of the reflection cancellation is evaluated (section 8.7). Finally, the influence of the experimental conditions are discussed (section 8.8).

8.1 Overview and general performance

Here, a few introductory remarks on the results as a whole will be given, before the effects of the walls are studied in the remaining sections.

From figures 7.1 and 7.2 it can be seen that the general appearance of the datasets are as expected. The perceived directions generally correspond to the presented directions or their reversals, distributed along the main diagonal and the two half-diagonals. Although there seems to be more perceived directions in the frontal half plane than in the rear half plane, answers are found for all directions for all setups.

The main mechanisms causing deviation from correct localisation seem to be reversal and spread. There is also some clustering of answers towards the directions of the walls.

As can be seen from these figures, and as will be pointed out later, the distribution of answers is different for the left side and the right side. This asymmetry can not be accounted for by the presence of the walls alone, as it is present also for the anechoic setup. For directions to the right, it seems that there is a general trend of localising more towards the frontal half plane than for the left side. The answers also seem to be less confined to the reversed/unreversed directions. Also, the spread seems to be somewhat larger at the right hand side. The cause of this effect is not known. A possible explanation might be the asymmetry of the setup in the anechoic room (figure 5.5). The distance to the room boundary was smaller on the right hand side. Also, the door through which the participants entered the room was placed at the right hand side. Another possible explanation may be asymmetry of the artificial head and its HRTFs, but this effect is largely countered by using the same head for both the binaural recordings of the source signals (section 5.4.2) and for HRTF measurements of the setups (section 5.4.1).

8.2 Deviation of average perceived direction

The deviation of the average perceived directions from the presented directions, and of the average perceived directions for the setups with reflections from the average perceived directions for the anechoic setup, are discussed. In addition to the results given in section 7.3, related information may also be found from figures 7.3 and 7.4, where the perceived directions are shown graphically.

8.2.1 Deviation from presented direction

Tables 7.1 (model B) and 7.2 (model C) show the deviation of the perceived directions from the presented directions.

Model B (symmetry of means enforced)

As seen from table 7.1, total mean deviations range from 3.71 degrees for the anechoic setup to 9.03 degrees for the two-wall setup. Of the setups with one wall, the far-wall setup shows the smallest deviation, and the one wall setup with corrective filtering shows a smaller deviation than the corresponding setup without such filtering.

The ordering of the setups can be explained by noting that larger deviations correspond to more complex playback situations. The anechoic setup is the setup for which the crosstalk cancelling is designed. It is also the simplest possible playback setup, where the intended playback is not disturbed in any way. So it is to be expected that this setup should give the localisation closest to the intended one.

The far wall setup shows the lowest deviation of the setups with reflections, with a mean deviation 58% larger than for the anechoic setup. As there is a reflection, the playback situation does not correspond to the one for which the filtering is designed, and this results in poorer performance. In this setup the wall is farther away than for the other setups with walls. This leads to reflections that are weaker and more delayed with respect to the direct sound compared to the close wall setups, as evident from figures 5.6 and 5.11. As discussed in section 2.2.3, later and weaker reflections influences localisation less than stronger and earlier reflections. This effect manifests itself here as a lower deviation for this setup than for the setups with close walls.

The two-wall setup shows the largest deviation. This is also to be expected, as this is the situation where the playback is most disturbed compared to the anechoic case for which the filtering was designed. Here there is not only one, but two early reflections, and the reflections come from both sides, possibly leading to even more conflicting cues.

For the setup with a close wall and corrective filtering, the playback situation is the one for which the filtering is designed, as it is for the anechoic setup. But the deviation is larger. This may be explained by noting that in this case the inverse filtering is a more complex operation, cancelling the reflection in addition to the crosstalk. As filters of the same length as for the anechoic setup are used to obtain a more complex filtering, the resulting filtering performs poorer. This can be observed from the impulse responses in figure 5.10, where it is seen that the "error floor" for the setup with wall is 10 – 15 dB higher than for the anechoic setup, and that the amplitude frequency spectrum for the former deviates more from the intended flat spectrum. Also, the setup with a wall gives a more complex playback situation. This increases the probability that the actual playback situation does not correspond exactly to the one for which the inverse filtering is designed. Small changes of the geometry of the playback situation, due e.g. to a misplaced listener, will have more severe consequences than for the anechoic setup.

Model C (all parameters free)

When the requirement of symmetrically placed reversals is removed, the results change somewhat, as can be seen by comparing table 7.2 to table 7.1. But in general, the results are similar for the two models.

The mean deviation is generally higher than that obtained from the symmetrical model, 6.16 degrees for the setup with the lowest deviation and 11.99 degrees for the setup with the highest deviation, compared to 3.71 and 9.03.

The ordering of the setups according to their mean deviation is the same as the one obtained with the symmetrical model, with one notable exception: In this case, the setup with a far wall has the lowest deviation, while the anechoic setup has the second lowest deviation, 34% higher. However, the far wall setup is the one least different from the anechoic setup. So also for this model, the general trend is that the setups that differs more from the intended playback situation show a larger deviation of the perceived directions from the presented ones.

While the deviation is higher for all setups using this model, the deviation for the anechoic setup has increased much more than the deviation for the other setups. This can also be seen comparing figures 7.3 and 7.4. When the statistical model is used with no requirement of reversal symmetry, the perceived directions for the anechoic setup are in many cases offset from the presented ones. The cause of this change is not clear. It may be related to the similar change seen in figure 6.11, where it is shown that the fitting error from the statistical modelling is smaller for the far wall for the symmetrical model, and smaller for the anechoic setup for the model with all parameters free.

For these two setups, anechoic and far wall, the main difference between the two variants of the statistical model seems to be at the angles 45° and 67.5°, where the deviation for the anechoic setup is much larger for the model with all parameters free. This is also the case, to a lesser degree, for the direction 112.5°. For the angle −112.5°, the deviation for both setups is larger for model C, but more so for the anechoic setup. And at 90° for the far wall setup, the deviation is less for this model than for the symmetrical model (model B).

For both variants of the statistical model, the deviation is small for the three most frontal directions and the three most rear directions. The exception to this is the two-wall setup, which has relatively large deviations for −22.5°, 22.5° and 157.5°.

A main difference between the anechoic setup and the other setups is for the direction −90°, where the deviation for the anechoic setup is far lower than the deviation for the other setups. A possible explanation is that this direction was marked on the form (appendix B). This mark may have influenced the localisation, while wall reflections may have dominated over it for the setups with walls. But this explanation is not satisfactory. The direction 90° was also marked, and for this direction all setups, including the anechoic one, show large deviations.

As discussed in section 8.1, there is a lack of symmetry between the left and right sides. As tables7.1 and 7.2 show, the right half plane has the highest mean deviation. This is common to all setups.

8.2.2 Deviation from the anechoic setup

When the average perceived directions for the setups with reflections are compared to the average perceived directions for the anechoic setup, the results in tables 7.3 and 7.4 are found. Also in this case, the two-wall setup gives the largest mean deviation. The far wall setup gives the smallest deviation, and the one-wall setup with reflection

cancelling gives smaller deviation than the corresponding setup without corrective filtering. This shows, as above, that more complex playback situations, differing more from the intended playback situation, give larger deviation of the perceived directions from the presented directions which the listener is intended to perceive.

The deviations follow the same pattern as when comparing the perceived directions to the presented ones, smaller in the front and rear, and larger at the sides. This corresponds well to what is known, that localisation is less precise at the sides [Blauert, 1997b].

The deviations computed here are generally lower than those computed with the presented directions as the reference. The largest decrease is for the one-wall setup with corrective filtering. The exception is the far wall setup, when reversal symmetry is not enforced in the modelling. This means that wall setup results are generally closer to the results for the anechoic setup than to the presented directions.

Here, the difference between the right side means and the left side means is less than when using the presented directions as reference. But the right side mean deviation is larger than the left side mean deviation for three of the four setups. This may indicate that the presence of the walls partly can explain this effect. It is however also probable that it is due to chance, as the one-wall setup, showing the second largest mean deviation in all cases, does not show this effect.

Also here, comparing the average perceived directions to those for the anechoic setup, the deviations are larger when computed without reversal symmetry in the statistical model.

8.2.3 Summary and conclusions

The deviation of the average perceived directions from the presented directions is larger for the setups with reflections than for the anechoic setup. Also, the deviation for the setup with reflection cancelling is less than for the identical setup without corrective filtering. I.e., the playback situations that differ from the one for which the crosstalk cancellation is intended give larger deviations than when the playback is as intended.

Both for deviation computed with respect to the presented directions and with respect to the perceived directions for the anechoic setup, it is found that the ordering of the setups after their deviation corresponds to an ordering after complexity of the playback situation. Reproduction under conditions that differs more from the intended ones gives larger deviation.

8.3 Localisation towards the reflections

A possible effect of the added reflections is that localisation is influenced so that the perceived direction is displaced towards the direction of the source of the reflections,

the walls (sections 2.2.3 and 4.3). This may explain the increased deviations for the non-anechoic setups seen in the previous section. The question of whether this effect is present is investigated using several approaches (section 8.3.1) and possible causes discussed (section 8.3.2).

8.3.1 Clustering towards the walls

From figure 7.1 it can be seen that for the setups with walls there is indeed a clustering of answers towards the directions of the walls. The answers for the presented directions close to the directions of the walls are displaced towards the wall, when compared to the answers for the anechoic setup. This effect can also be seen from the figures showing results of the statistical description of the answers, figures 7.2, 7.3 and 7.4.

Deviation patterns

Another indication of this effect can be seen when comparing the deviation pattern of the close-wall setup (without reflection cancelling) to that of the anechoic setup for directions close to the wall.

For the anechoic setup, the deviation is small for the direction $-90°$, large for $-67.5°$ and small for $-45°$ (tables 7.1 and 7.2). As can be seen from the corresponding figures (7.3 and 7.4), the mean value for $-67.5°$ is for some reason offset towards $-90°$, while for the two other directions, the mean values are placed close to the presented direction and its reversal.

For the close-wall setup, the situation is opposite, with a small deviation for $-67.5°$ and large deviations for the other two directions. The figures show that compared to the anechoic setup, the answers are displaced towards the direction of the wall. The perceived directions for $-90°$ and $-45°$ do no longer correspond to the presented direction and its reversal, but are offset towards the front and rear, respectively. For the presented direction $-67.5°$, the perceived direction has moved from its offset position and towards the wall, so that it now corresponds well to the presented direction.

Similarly, localisation towards the wall can also explain the deviation patterns for the far-wall setup (presented directions $-90°$ and $-67.5°$), and for the two-wall setup for the directions $-90°$ to $-45°$ and to some degree for the directions close to the right wall, $45°$ to $90°$.

Number of answers close to the walls

From figure 7.5 it can be seen that for the one-wall setup without corrective filtering, there are up to three times as many answers localised to the wall than there is for the anechoic setup. For sector widths equal to or less than the width of the wall, the ratio is between two and three, with an exception for the smallest sector widths. For sectors larger than the width of the wall, the ratio is between 2 and 1.3. (The

low ratio for small sector widths is most probably due to random variations. For so narrow a sector, there will be few answers lying within it, leading to large relative variations for small absolute changes in the number of answers.)

The one-wall setup with corrective filtering does not show the large ratios of the uncorrected setup. But there are still 1.5 times as many answers located to the wall as for the anechoic setup. In this case the curve is flatter, the difference between sectors wider and narrower than the wall is smaller than for the uncorrected setup.

The far wall setup shows ratios up to over 2 for directions within the angular width of the wall. Also for this situation, the curve is quite flat. For the two-wall setup, the ratio lies between 3 and 2 for sector widths up to the angular width of the wall, and between 2 and 1.5 for wider sectors.

Distance between answers and walls

Figure 7.6 shows a different way to measure focusing towards the walls. For several presented directions, the figure show how far from the walls the answers are lying. Values less than one show that the answers are closer to the wall than for the anechoic setup.

For the one-wall setup, the distance is less than for the anechoic setup for all directions measured. There is a minimum, with a relative value of about 0.3 for the direction $-45°$. This direction is close to the wall direction of $-56°$. The distance for the direction $-135°$, the reversal of $-45°$ is also smaller than for the two neighbouring directions. For the one-wall setup with corrective filtering the trend is similar, with the distance generally taking on values between those for the anechoic setup and the uncorrected setup. This setup also shows a minimum for the direction $-45°$.

For the far-wall setup, the presented directions $-157.5°$, $-45°$ and $22.5°$ have a normalised distance equal to one, i.e. the distance is equal to that computed for the anechoic setup. There is a minimum of 0.4 for the directions $-90°$ and $-67.5°$. This may be explained to some degree by noting that the wall-direction for this setup was $-68.2°$.

For the setup with two walls, the distance is less than one for all angles shown. There are minima, with a value of 0.4, at $\pm 45°$ and the corresponding reversals. For the angles between those, the relative distance is 0.5 to 0.6.

Distribution of answers between presented direction and wall direction

Table 7.5 shows the relative distribution of answers between the presented direction, its reversal and the direction of the wall. As can be seen, for four of the six directions shown, a significantly larger part of the answers are localised to the wall direction for the one wall setup than for the anechoic setup. For the two remaining directions, the difference is small.

For the direction $-45°$, the large difference (0.70 against 0.36) can be explained by noting that this direction is close to the wall direction. Small displacements of the answers are sufficient to make them belong to either one or the other direction.

The direction $-67.5°$ is also close to the wall direction. But here the difference is smaller, and there are far fever answers localised to the wall in general. This difference between the directions $-45°$ and $-67.5°$ can be explained by noting, as we did above, that the answers for the direction $-67.5°$ are offset towards $-90°$ for the anechoic setup, and are dragged to the presented direction with the wall present.

For $-90°$, the difference is 0.57 against 0.07. The direction $-112.5°$ is the reversal of $-67.5°$, and show the same low localisation towards the wall. The direction $-135°$ is the reversal of $-45°$, and shows and even larger difference than this one, with 0.53 against 0.03.

For the direction $-157.5°$, the difference is small. This is probably because this direction is so far from the wall direction that the influence from the reflection is small.

8.3.2 Possible causes

It is evident that for the setups with reflections added there are far more answers localised towards the walls and their vicinity than for the anechoic setup. There are several possible causes of this effect. Summing localisation may have influenced localisation, so that answers are displaced towards the reflections (section 2.2.3). It may also be the result of increased probability for localisation to the frontal half plane in general, not only towards the walls. And finally, as the walls were visible (section 5.2), it may be that the localisation has been influenced by the visual cues given by this, as mentioned in section 2.2.2.

Visual cues?

The influence of visual cues may be evaluated by comparing the two setups with a wall in the close left position. These setups were physically identical, but for one of them, corrective filtering was applied to cancel out the reflection from the wall. If localisation was based mainly on the visual cues, the results from these two situations should be similar. If, on the other hand, localisation was based mainly on acoustical cues, and, in addition, the reflection cancelling filtering worked perfectly, the results for the setup with corrective filtering applied should be similar to the results for the anechoic setup.

As can be seen from table 7.5, the results from the two close-wall setups differ. And they differ most for the directions where there is a large difference between the uncorrected one-wall setup and the anechoic setup. The relative number of answers localised towards the wall is significantly lower for the corrected setup than for the uncorrected one for these directions. This difference between these two setups can

also be seen from figures 7.5 and 7.6. In all these cases, the setup with corrective filtering shows results that are clearly more similar to those for the anechoic setup than those for the uncorrected setup are.

On the other hand, the results for the setup with corrective filtering also differs from the results for the anechoic setup. This may indicate that visual cues may indeed have influenced the localisation. It may also be the result of less than perfect cancellation of the reflection, a very probable situation.

General localisation to the front?

To find out whether the higher number of answers towards the walls is the result of localisation towards the walls or whether it is the result of a generally increased probability for localisation to the frontal half plane, figures 7.5 and 7.6 may be consulted. Front/back localisation will be discussed in section 8.4. Here, it should only be noted that a general tendency of localisation towards the frontal plane should result in answers more or less evenly distributed across all presented directions, while localisation towards the wall should manifest itself mostly for directions close to the wall.

As can be seen from figure 7.5, and as discussed above, the uncorrected one wall setup and the two wall setup show a marked difference in the relative number of answers for sector widths smaller and larger than the angular width of the wall. There are more answers localised to within the wall sectors than outside them. For the close wall setup with corrective filtering and the far wall setup the trend is similar, but not as marked. Figure 7.6 show similar results. All setups have minima for presented directions close to the wall position, showing that answers are more concentrated here than for other directions.

Conclusions

It may be concluded that there is localisation towards the walls. The effect is most pronounced for the close-wall and the two-wall setups, but is also at work for the far-wall setup and the close wall setup with reflection cancelling.

The localisation is to a large degree towards the walls specifically, and not only towards the frontal half plane. And while visual cues may have influenced the localisation to some degree, they do not dominate localisation. It must therefore be concluded that the localisation towards the wall is due to the acoustical and auditory effects of the reflections.

8.4 Localisation to front and rear half planes

From section 7.5 it can be seen that the distribution of answers between the front and rear half planes differs much between the various setups. The values range from

2.32 up to 5.69. The lowest value belongs to the anechoic setup, and the highest to the two-wall setup. The setup with corrective filtering shows a lower value than the one-wall setup; 2.86 against 3.29, while the far wall setup (3.20) is slightly better in this respect than the one-wall setup, while not as good as the setup with corrective filtering.

For perfect reproduction, this value should be 1. Values larger than this show that more answers are localised to the frontal half plane and fewer to the rear half plane than there should have been. This is the case for all setups, including the anechoic one. But the setups with walls have this effect to a larger degree than the anechoic setup, causing a larger part of the answers to be perceived towards frontal directions.

Several causes are possible. As found in the previous section (8.3), there is a localisation towards the walls. This may be the main reason for the higher rates of front-localisation for the setups with walls. There is also the possibility that the walls have given visible cues. But, as discussed in the previous section, this effect seems to be small. Also, this effect can not alone explain the large difference in front/back localisation between the different setups with walls.

There is also a general over-localisation towards the frontal half plane, as shown by the fact that also the anechoic setup has more perceived directions than presented directions in the frontal plane. While not a part of the topic of the work, this effect is still interesting. The most possible explanation is that the loudspeakers were placed in front of the listener. The sound actually coming from the front may give acoustical cues leading to frontal localisation. The fact that the loudspeakers were visible as sound sources may also have given strong visual cues towards the front. Non-individualised HRTFs were used, and this is known to increase the number of front/back localisation errors (section 2.3.2). Combined with this, these cues may have influenced more than they would otherwise have done.

There are more perceived directions in the frontal half plane than in the rear half plane. But answers are found for all directions for all setups, showing that the reproduction system is capable of reproducing sound from all directions for all conditions tried.

8.5 Reversals

The topic of this section, reversals and reversal rates, is closely related to the topics of the two previous sections, localisation towards the walls and front/back localisation. Reversal rates for each combination of setup and presented direction are given in table 7.6. These results corresponds to those illustrated in figure 7.3. Reversal rates and reversal patterns may also be studied from figures 7.1, 7.2 and 7.4.

The mean reversal rate varies from 35% for the anechoic setup to 46% for the setup with two walls. The setup with a wall in the far left position has 11% more reversals than the anechoic setup, the setup with a wall in the close left position 12% more,

the setup with a wall in the close left position and corrective filtering 19% more, and the two-wall setup 30% more.

As can be seen from the figures and the table, presented directions in the rear half plane are often perceived in the frontal half plane. The opposite, perception of frontal directions towards the rear happen far more seldom. Mean reversal rates for rear directions varies from 52% (anechoic setup) to 84% (two-wall setup), with 32% (anechoic setup, $-135°$) and 94% (two-wall setup, 112.5°) as the minimum and maximum single data values. The mean reversal rate for frontal directions varies from 7% (two-wall setup) to 18% (anechoic setup), with data points varying from 0% (two-wall setup, $-67.5°$) to 37% (anechoic setup, 67.5°).

For most directions (five of seven) in the rear half plane, the anechoic case has the lowest number of reversals. The exceptions are the direction straight backwards and the direction -157.5°. For this last one the anechoic setup is 1% from having the minimum value. The setup with two walls has the highest number of reversals for five of the seven directions in the rear.

For all three directions in the frontal left quarter plane, $-22.5°$ to $-67.5°$, the anechoic setup shows the highest reversal rate. The two-wall setup has the lowest value for two of these directions. The two-wall setup also has the lowest reversal rates for the directions 45° and 67.5°, the two directions closest to the right side wall. Especially for the latter direction, the difference to the other setups is large. For the reversals of these two directions, 112.5° and 13°, the two-wall setup has a very high reversal rate.

To sum this up, all setups with walls show increased mean reversal rates (11% – 30%) compared to the anechoic setup, with the largest increase for the two-wall setup. This rise is composed of an increase in reversal rates for rear directions and a decrease in reversal rates for frontal directions. Or, to put it another way, with walls present, more rear directions are perceived towards the front, and less frontal directions are perceived towards the rear.

This increased localisation towards the front is, at least partly, due to a displacement of answers towards the walls. With reversals, there are two possible perceived directions for each presented direction. With walls present, the perceived direction closer to the wall is preferred with respect to the other one, when compared to the anechoic setup.

8.6 Spread

The spread of the answers is shown in figure 7.7. The spread can also be seen more indirectly in the overview figures 7.1 and 7.2.

The spread is a function of the presented direction. It follows a pattern common to all setups, with lower values (around 5° for the presented directions 0° and $-180°$ (front and back), and higher values (10° to 30°) for presented directions towards the sides. This corresponds well with what is known about localisation blur for directions

in the horizontal plane, localisation is most precise for forward direction and least precise for directions towards the sides [Blauert, 1997b].

The spread is somewhat higher for directions towards the right than for directions towards the left. This is a parallel to the larger deviations for right-hand directions that was discussed in section 8.2.

These results, based on a statistical model where symmetry is not required, seems to indicate that the two-wall setup distinguishes itself by having a lower spread for the directions $-90°$ and $90°$. This is not the case when using the symmetrical model. The reason for this is that for the non-symmetrical case, one of the two mean values is placed at the presented direction, while the other is placed at the direction of the wall, as can be appreciated from figure 7.2. Together, these two means describe a distribution of answers with a large spread, and summing them, as is done in the figure shown, is not correct in this case.

The results are very similar for all setups, including the anechoic reference. The differences does not follow any regular pattern, with the possible exception that the two-wall setup has spread more than twice that of the others for the directions $-22.5°$ and $22.5°$.

The results have not been found to support any conclusions that the introduction of the reflecting walls has influenced the spread in general.

8.7 Reflection cancelling

The two setups with a wall in the close left position were physically identical. But the filtering was different; for one of the setups an attempt was made at compensating for the presence of the wall. In order to cancel out the reflection, the crosstalk cancelling filters were based upon HRTFs measured with the wall present (chapter 5.4.3). Here, the results obtained from these two tests will be compared, in order to evaluate the effect of the reflection cancelling. The results will also be compared to the results for the anechoic setup.

The deviation of the perceived directions from the presented directions for the setup with reflection cancellation is 90% of that of the setup without the cancellation, and 106% of that of the anechoic setup (table 7.1). For the statistical model with all parameters free (table 7.2), these numbers are 96% and 212%.

When the perceived directions from the anechoic setup are used as the reference (tables 7.3 and 7.4) the deviation of the setup with corrective filtering is found to be 76% of that for the other setup when using the symmetrical model, and 81% when using the other model.

The largest differences in deviation between the two setups are found for presented directions close to the direction towards the wall, $-45°$, $-67.5°$ and $-90°$, and also for the reversal of $-45°$, $-135°$. For deviations computed from the model with all parameters free, the setup with corrective filtering has the smaller deviation for all

these angles. When deviations are computed from the symmetrical model, the uncorrected setup has the lower deviation for the angle $-67.5°$, but the difference is still large. Also, for the model with all parameters free, there are large differences for some directions on the right hand side, 90°, where the setup with corrective filtering has the lower deviation, and the directions 67.5° and 135°, where the uncorrected setup has the lower deviation. The deviation pattern for the setup with crosstalk cancellation is generally more like the deviation pattern for the anechoic setup.

Table 7.5 shows that for four of six directions, the setup with reflection cancellation has a smaller part of the answers assigned to the direction of the reflection. Three of these directions are those closest to the direction of the reflection, the fourth the reversal of one of them. (For one of the remaining directions, the uncorrected setup has fewer answers assigned to the wall, and for the last direction, the two cases are equal in this respect.) Also, for four of six directions, the setup with corrective filtering has more answers assigned to the presented direction. The anechoic setup has the smallest part of answers assigned to the reflection direction for all directions but the one most backwards.

From figure 7.5, it can be seen that for the setup with corrective filtering, a much smaller part of the answers are localised towards the wall than what is the case for the other setup. The difference is a factor of about two for small angular sectors, about 1.5 for a sector as wide as the wall. But the values for the corrected setup are still higher than for the anechoic setup, with a factor of up to more than 1.5. Similarly, it can be seen from figure 7.6 that the summed distance between the answers and the wall is larger for the wall-cancelling setup than for the setup without cancelling, while not quite as large as for the anechoic setup.

From chapter 7.5, it can be seen that the setup with corrective filtering has a lower overweight of frontal localisation, 87% of that of the other. This is still 23% more than for the anechoic setup.

The setup with corrective filtering has a mean reversal rate that is 6% higher than the reversal rate for the one-wall setup, and 19% more than that of the anechoic setup (table 7.6). There are fewer reversals for rear left directions, more reversals for rear right directions. When the directions $-180°$ and $-157.5°$ are excluded, the mean reversal rates for these two setups are equal.

It is not possible to conclude that there is any general difference in spread between these two setups. However, the spread for the setup with corrective filtering is lower for the presented directions “front” and “rear”. (Figures 7.7, 7.1 and 7.2.)

To sum this up: The setup with reflection cancelling filtering has lower deviation than the other close-wall setup both with respect to the presented directions and with respect to the average perceived directions for the anechoic setup. Much of this difference is found for presented directions close to the direction of the reflection. The corrected setup also has a lower predominance of frontal localisation, but a somewhat higher reversal rate. A far smaller part of the answers are localised towards the wall than for the setup without corrective filtering.

In total, it may be concluded that the attempt to compensate for the playback setup with reflection cancelling has lead to a localisation that is more like the localisation for the anechoic setup.

As discussed in section 8.3, the differences between these two setups (with and without reflection cancelling) also show that the effects observed there (localisation towards the reflections) have acoustical causes, and are not the results of dominating visual cues.

8.8 Experimental conditions

8.8.1 The use of an artificial head

Individualised (the listeners own) HRTFS were not used in the listening tests done in this work. A Neumann KU81i artificial head was used both for the HRTF measurements of the setups for the listening tests (section 5.4.1) and for the recording of the binaural signals for the tests (section 5.4.2). This head originally comes without shoulders or a torso, but was equipped with shoulders (made at the Acoustics Group) when used in this work.

As discussed in section 2.3.2, listening with non-individualised HRTFs will generally give a localisation performance poorer than that obtained with individualised HRTFs. Typically, the number of front/back errors increase for non-individualised HRTFs [Wenzel et al., 1993, Møller et al., 1996, Hammershøi, 1995]. Although the works cited used headphone reproduction, their conclusions may be considered relevant for crosstalk cancelled loudspeaker reproduction also. Localisation using recordings made with artificial heads have, as mentioned in section 2.3.2, been found to give localisation performance similar or poorer to the performance obtained using recordings based on HRTFs from another person than the listener [Møller et al., 1999].

The Neumann KU81i (without shoulders) has been found to give a high number of reversals ("within-cone errors") compared to other artificial heads, but this may have been due to the lack of shoulder reflections [Møller et al., 1999]. As the head used in this work was equipped with shoulders, it should not be worse than other artificial heads in this respect.

The major influence of the artificial head and the non-individual treatment of the listeners is believed to be somewhat poorer localisation performance in general, and increased reversal rates especially (sections 7.6 and 8.5).

8.8.2 The gathering of the data

The answers were gathered by letting the listeners mark the perceived directions on a form (section 5.4.4, appendix B). This is believed to be an intuitive and robust method, especially as the listeners were accustomed to graphs and directions in polar coordinates through their education.

A method used successfully by Wightman and Kistler [1989b] and Wenzel et al. [1993] is to let the listeners call out numerical estimates of the coordinates of an apparent sound source. Marking the directions on a form should not be any more difficult, as no conversion to coordinates is necessary – the direction can be marked directly.

The marks on the forms for the directions 0°, −90°, 90° and −180° (figure B.1) were intended to improve the precision of the answers by giving the listeners some reference directions to relate their marks to. It is possible that these marks on the forms may have lead the listeners to place their marks at these directions. But looking at the data, this effect does not seem to be very pronounced.

For the direction −90° in the anechoic setup, there is a sharp peak of answers. The answers are unimodally distributed (figure 7.1), but are modelled by two peaks (section 6.4, figure 7.2). The two peaks have the same (or nearly the same) mean value, but one of them has a smaller spread, giving a sharp peak placed on a broader one. This may be due to the mark on the form. This is however a single occurrence, neither for the direction 90° in the anechoic setup nor for the directions ±90° in the other setups can any such effect be seen.

The directions 0° (front) and −180° (rear) show sharp peaks for many of the setups. But these peaks are accompanied by low spreads (figure 7.7), there are no signs of the "dual-variance" distribution shown by the direction −90° for the anechoic setup. These sharp peaks may also be due to the more precise localisation towards front and rear [Blauert, 1997b].

8.8.3 The use of visible walls and visible loudspeakers

The reflecting walls were visible to the listeners (section 5.2), and it is known that visual cues may interact with acoustical cues to give a localisation different from that obtained from the acoustical cues alone (section 2.2.2). This is especially relevant as one of the results found is that there is a tendency of localisation towards the walls (section 8.3). But, as discussed in section 8.3.2, while the visual cues may have influenced localisation, their effect is minor compared to the acoustical effects. The visual cues are found to not be dominating localisation.

The visible loudspeakers in front of the listeners (figure 5.5) have probably influenced the localisation, as the listeners understood that these produced the sound that gave rise to the auditory impressions. This may explain to some degree the high rate of localisation to the frontal half plane (sections 7.5 and 7.6). On the other hand, it is has been found (section 8.3.2) that the localisation towards the frontal half plane to a large degree is a localisation towards the walls, not towards the frontal half plane in general.

8.8.4 Listener placement and filtering

The listeners were not fixed in any way in the listening position (section 5.4.4). This may have given room for misplacement and movement. It is reasonable to believe

that this may have been one of the main factors to determine the general localisation performance. As also discussed in chapter 3.5, displacement vertically and along the loudspeaker axis was not very critical for the imaging. Left/right displacement was more critical. (This is natural, considering the fact that movement along the latter axis will give larger differences in and between the actual HRTFs than movement along the other axes.)

Unfortunately, measuring the actual ear signals in the various setups was not found feasible (section 5.4.3). The simulations done show good crosstalk cancellation and a frequency response acceptable for the speech signals used. The filtering is believed to be adequate for the purpose, and to not introduce any errors that should influence the differences between the setups.

8.8.5 Conclusions

The method of the experimental work has been to investigate the differences between localisation in various setups. It is therefore believed to be generally robust against errors and imperfections that are constant between the setups. Although the factors discussed here may have influenced the reproduction and the general localisation performance some degree, they have not been found to do this in a way that is dependent upon the playback setup. It is therefore believed that the results give a valid description of the effects caused by the reflections.

Chapter 9

Summary and conclusions

9.1 Summary

The topic of this dissertation is how crosstalk cancelled reproduction of binaural signals intended for an anechoic playback venue is influenced by the presence of reflections during the playback. As a simplification, the scope of the work has been limited to an investigation of the effect of one or two early reflections upon localisation, for reproduction of source directions in the horizontal plane.

An overview of theory relevant to the topic has been given, presenting directional hearing, binaural reproduction and crosstalk cancellation. Earlier work regarding the effect of reflections on crosstalk cancelled playback of binaural signals has been reviewed.

Crosstalk cancellation has been implemented in software. Routines have been written for the inversion of transfer functions, the design of crosstalk cancelling filters, filtering of sound signals and playback simulation. The system for crosstalk cancellation has been found to give a realistic, natural and convincing reproduction of binaural signals, and has been used as a tool for the experimental work.

Crosstalk cancelled reproduction with a single reflection has been studied and simulated using a simplified model of such a setup. Timing, plausibility and crosstalk cancellation was discussed.

Binaural reproduction under different conditions has been investigated through listening tests. Binaural recordings (made with a Neumann KU81i artificial head) of male and female speech from sixteen different directions were filtered for crosstalk cancelled loudspeaker playback and used as signals. The playback system utilised two Bowers & Wilkins 801 loudspeakers placed next to each other, in front of the listener, in an anechoic room. Reflections, produced by real walls, came from the frontal half plane, and were delayed 5 ms and 10 ms with respect to the direct sound. Five setups were used: Anechoic, close wall, close wall with reflection cancelling, far

wall and two walls. A total of 5180 localisation responses were collected from nine listeners during six tests.

The problems associated with statistical characterisation of the bimodally distributed data from the listening tests have been discussed. A statistical model has been developed that may be used for such characterisation, giving estimates of reversal rate, average perceived directions and spread of reversed and unreversed answers. This *dual-normal* model is composed as a weighted sum of two normal (Gaussian) distributions. Several variants of it has been tried and used in this work. The model has been found to describe the data from the listening tests well, and has made quantitative comparisons of the results from different setups possible.

Perceived directions in the tests have been compared to the presented direction with respect to a set of analysis criteria that has been defined: Deviation of the average perceived direction from the presented direction, spread of answers around the average perceived direction, reversals and localisation towards the reflections. The influence of the reflections have been analysed by comparing the different setups with respect to these criteria. The analysis has mostly been done using the statistical model mentioned above. The results of these analyses have been presented and discussed.

9.2 Conclusions

From the simplified model of a playback setup with a reflection, it was concluded that the reflected sound would not combine in a way that would give rise to crosstalk cancellation. Neither would the reflections correspond to the reflections caused by a real source in the position of the presented virtual source. The interaural time difference (ITD) of the reflected signal was found to correspond well to the ITD for a signal radiated from a source placed at the reflecting wall.

The results from the listening tests show that the reflections do influence the playback in such a way that localisation is altered. The effect of the walls has been to give a localisation where the perceived directions differs more from the presented directions than what is the case for reproduction under anechoic conditions.

The deviation of the average perceived directions from the presented directions is larger for the setups with reflections than for the anechoic setup. The largest increase in mean deviation found was 140%. The deviation is closely connected to the complexity of the playback setup, with larger deviations for setups that differ more from the intended playback setup.

The same trend is found when comparing the average perceived directions for the setups with walls to the average perceived directions for the anechoic setup. The more complex setups, that differ more from the intended playback, show larger deviations.

There is localisation towards the reflections. For the non-anechoic setups, the perceived directions are displaced towards the direction of the source of the reflection. This effect is most pronounced for the setups differing most from the anechoic setup.

Compared to the anechoic setup, up to more than three times as many answers were located towards the direction of the source of the reflection.

The non-anechoic setups show a stronger tendency of localisation to the frontal half plane, more so for the more complex setups.

Mean reversal rates were found to be 11 to 30 percent higher for the non-anechoic setups than for the anechoic setup. The largest increase was found for the setup that differed most from free field, the smallest for the least different. The general increase is the result of a decrease in reversal rate for presented directions in the frontal half plane, and an increase in reversal rate for rear directions.

It was not found that the general spread of the answers was influenced by the playback situation.

Reflection cancelling lead to a localisation that was generally more like the localisation for the anechoic setup than the localisation for the physically identical setup without this cancellation was. The deviation of the perceived directions from the presented directions and from the perceived directions for the anechoic setups was less. There was less localisation towards source of the reflection. The reversal rate was somewhat higher, but the tendency of localisation towards the frontal half plane less.

The precedence effect is believed to be the main cause of the observed effects. For all cases where it is found that the reflections influence the results, the relative influence of the setups correspond to how much they differ from the anechoic setup. An earlier and stronger reflection changes the localisation more than a more delayed and weaker reflection. Also, two reflections change the localisation more than one. Corrective filtering cancelling the reflection in most cases gives localisation more like the anechoic setup than lack of such filtering does. This corresponds well with what is known about how the precedence effect works.

9.3 Discussion

In this work it has been found that reflections delayed 5 ms and 10 ms with respect to the direct sound has a clear effect upon crosstalk cancelled reproduction of binaural sound. This is contrary to what has been claimed by Cooper and Bauck [Cooper and Bauck, 1989, Bauck and Cooper, 1996], who cite $1 - 2$ ms and $2 - 3$ ms as limits above which reflections do little harm. It also contrasts somewhat with the results given by Takeuchi et al. [1997a], who found that the performance of the reproduction *"[...] was not degraded severely with existence of a single reflecting surface"*. (They have however not undertaken any detailed analysis of their results, but instead studied overview data plots similar to those given in figure 7.1.) On the other hand, the results found here correspond well with what is said by Kirkeby et al. [1999]: *"The system is generally quite sensitive to room reflections, particularly from the side."*, and to the common claim that anechoic or nearly anechoic locations should be used (section 2.4).

Generally, deviation of the perceived direction from the presented one is not necessarily a problem. E.g. for playback of music, it may be sufficient that the positions of the instruments in an orchestra are consistent with respect to each other, and that the envelopment of sound from the concert hall is reproduced. In this work, auditory events were, to lesser or greater degree, perceived at all presented directions, even for the setups with reflections (e.g. figure 7.1). This is important, both for the reproduction of localisable auditory events, but also to make it possible to achieve envelopment by reproducing non-localisable sound.

In other cases, the requirements to correct localisation may be strong, e.g. for kinds of auditory displays [McKinley and Ericson, 1997, Shinn-Cunningham et al., 1997]. Then deviation of the perceived direction from the intended and presented one may be intolerable.

Localisation towards the reflections may in many cases be a grave error. In this case, several directions deviate towards the same direction. This gives a presentation of directions that is "many-to-one", where a range of presented directions collapse into the same perceived direction. This may destroy an auditory image that might otherwise have been acceptable. It may also lead to localisation of sound that should be non-localisable.

Reversals may in many cases be a less serious error. This may be illustrated by the fact that the room acoustic criterion "Lateral Efficiency" is found using a pressure gradient microphone, which does not distinguish between sound from front and rear [Cremer et al., 1982a, p 447]. The auditory image may e.g. be restricted to one half plane only, and it may be clear from other cues, e.g. visual ones, that this is the case. If however, a completely true reproduction of all directions is required, reversals can not be tolerated.

A requirement of anechoic conditions will in many cases severely limit and restrict the application of binaural reproduction over loudspeakers. This work has shown that reflections like those used here should be avoided to obtain the best reproduction performance, and that reflections may give rise to reproduction difficulties. It has however not been shown that purely anechoic conditions are necessary. It may well be the case that playback under nearly "normal" conditions will be acceptable for many applications, but care and thoughtfulness should definitely be exercised.

This work has also shown that the effect of reflections can be reduced to a great degree by reflection cancelling filtering in addition to crosstalk cancelling. As done here, this is a more fragile process than crosstalk cancellation alone, and it is probably most suitable for one or a few reflections. But it may be a useful technique in some cases.

9.4 Further work

This work has given some insight into the topic at hand, and has brought forth some conclusions. But it has been limited in scope, and naturally does leave many questions unanswered. Also, the work has raised further questions and pointed to other topics

of interest. The directions for further work found most interesting are to extend the scope of the investigation, to develop the statistical model further, and to research the reversal process.

Extending the scope of the investigation

Extending the scope of the investigation may take several directions. Ideally, one would have liked to have a larger group of listening persons and more tests. Tests should have been done with the reflections originating from more and other directions than those used in these tests. First and foremost, reflections from the rear should have been included. It would also be of interest to experiment with a larger number of reflections and general reverberant conditions. An extension to three dimensions would have been natural. Also, other methods of reflection generation could be tried. Other placements and combinations of loudspeakers would have been of interest, e.g. the pair-wise loudspeaker paradigm proposed by Foo et al. [1998].

The statistical model

The statistical model has one major drawback as of yet; it has not been shown that, or how, the estimated parameters can be used for hypothesis testing. If this could be done, it would greatly improve the usefulness of the method. Monte Carlo simulations is a possible solution, but formulas based on the estimated parameters would be preferable. Another method of estimating the parameters might be necessary for this. Also, it should be looked into how the model behaves for presented directions left and right, where mirroring is not expected to occur. In these cases, the underlying model of reversals and spread does in principle not hold true any more. The model could also possibly be extended to three dimensions, where it would possibly have to be four-topped, to cater for both front/back and up/down confusions.

Reversals and the reversals process

The reversal process is interesting in and of itself. Of special interest pertaining to this work is the front/back symmetry of the reversals. Results from the statistical modelling with symmetry of the means enforced compared to the results from modelling with all parameters of the model free indicate that the reversal may not necessarily occur symmetrically around the left/right axis, at least when reflections are present. This is contrary to what is commonly assumed (section 2.2.2), and should have been investigated.

Appendix A

Overview of crosstalk cancellation software

Here, a short overview of the crosstalk cancellation software (chapter 3) is given, as a listing of the main routines and data structures along with brief comments.

Responses

***respstruct()*:** Documentation of the format of impulse response data structures

***chkresp()*:** Check the validity of an impulse response structure

Inversion

***iopstruct()*:** Values for combinations of filter length and delay

***set_iop()*:** Manually set the values of an *iopstruct*

***get_iop()*:** Set the values of an *iopstruct* based on the plot of an impulse response

***comp_inv_err()*:** Compute several inverses of a response for combinations of filter length and delay given by an *iopstruct*. Returns an *fpstruct*.

***fpstruct()*:** Filter parameter structure, containing filter lengths, delays and the associated error energies.

***choose_mse_inverse()*:** Choose parameters for inverse based upon a graphical presentation of the information given by an *fpstruct*

***comp_mse_inverse()*:** Compute an inverse with a given filter length and delay.

Filters

***filterstruct()*:** Documentation of filter data structure

***asym2symfilt()*:** Convert a non-symmetrical filter to a symmetrical one

***sym2asymfilt()*:** Convert a symmetrical filter to a non-symmetrical one

***chkfilt()*:** Check the validity of a filter data structure

Filtering

***crossfilt()*:** Filter with crosstalk cancelling filter

***crfltwav()*:** Filter .WAV file with crosstalk cancelling filter

Playback simulation

***pbsim()*:** Simulated playback with crosstalk

Additional routines

***oafilt()*:** (Overlap-Add-FILTering) Fast frequency domain computation for filtering with FIR-filters.

Appendix B

Form used in listening tests

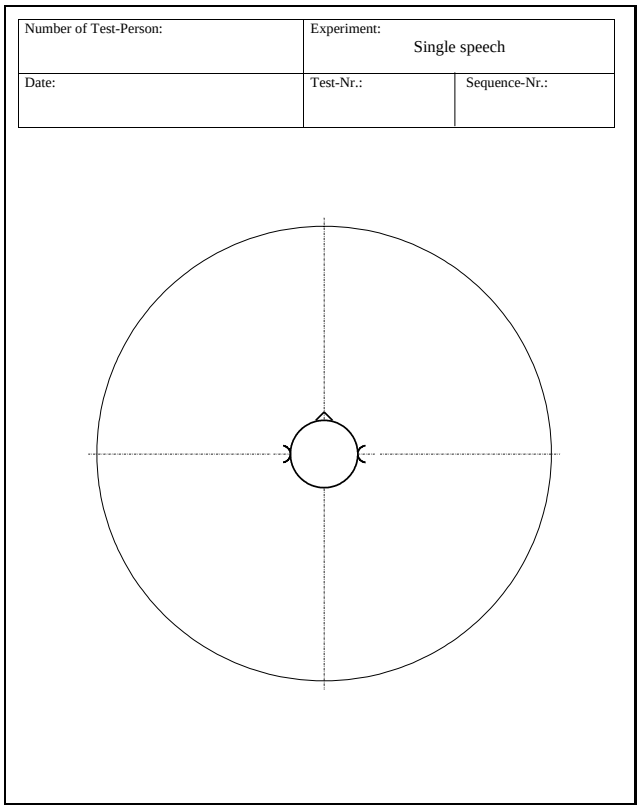


Figure B.1: The form used in the listening tests for marking directions (scaled down). One such form were used for each sequence of presented directions.

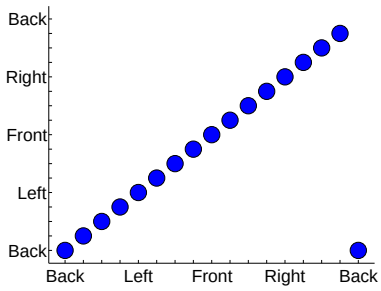
Appendix C

Scatter plots

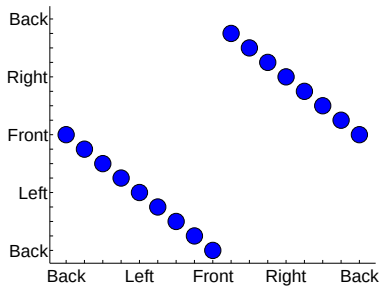
Data from listening tests of the kind used in this work may be usefully presented as some kind of *confusion matrix*, where perceived directions are plotted against presented directions [e.g. Damaske, 1971, Damaske and Mellert, 1969/70, Wightman and Kistler, 1989b, Hammershøi, 1995, Wightman and Kistler, 1997]. An example of one type used in this work is shown in figure C.1. Other kinds have similar properties.

The horizontal axis corresponds to the presented directions, the vertical axis to the perceived directions. The axes are marked according to the coordinate system used, either with angular values or terms like "Left". For a given combination of presented and perceived direction there may be a circle. The *area* of this circle is proportional to how often (in relative numbers) this presented direction has been perceived as this perceived direction.

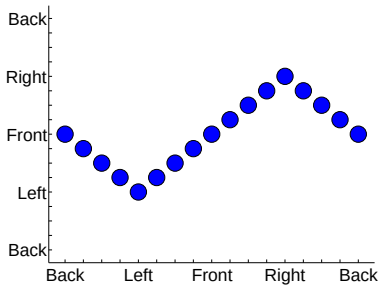
It is a property of these figures that all "correct" answers (i.e., answers where the perceived angle correspond to the presented angle) will be lying along the main diagonal from lower left to upper right. Front to back reversals and back to front reversals will end up on the two "half-diagonals" going from middle left to middle bottom, and from middle top to middle right, as shown in the figure.



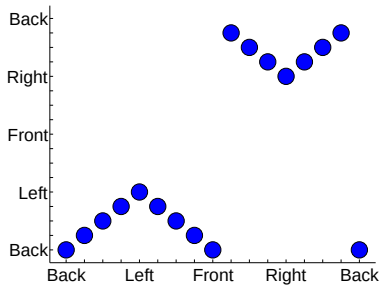
A: "Correct" answers



B: Complete front-back reversal



C: All directions perceived in front



D: All directions perceived in the back

Figure C.1: Examples of scatter plots, showing how various kinds of answers will show up. (Artificial data are used.) Presented direction along the horizontal axes, perceived direction along the vertical axes. A shows "correct" answers. B shows a situation where frontal and backward directions have been completely interchanged. C shows the pattern that will result when everything in the back has been perceived to come from the front, and D shows a situation where frontal sound have been perceived to come from the back.

Bibliography

V. Anmarkrud. Reproduction of spatial room acoustical attributes using cross-talk cancellation techniques. Master's thesis, NTNU, Norwegian University of Science and Technology, 2001.

O. M. Arntzen. Transaural stereo, the effect of early reflections. Term project, NTNU, Norwegian University of Science and Technology, 1998. (In Norwegian).

B. S. Atal and M. R. Schroeder. Apparent sound source translator. US Patent 3,236,949, Nov. 1962.

Audio Eng. Soc. 12th Int. Conf.: Perception of Reproduced Sound, June 1993. Audio Eng. Soc. ISBN No. 0-937803-19-7.

102nd Audio Eng. Soc. Convention Preprints, Mar. 1997a. Audio Eng. Soc.

103rd Audio Eng. Soc. Convention Preprints, Sept. 1997b. Audio Eng. Soc.

105th Audio Eng. Soc. Convention Preprints, Sept. 1998. Audio Eng. Soc.

Audio Eng. Soc. The AES: 50 years of contributions to audio engineering. *J. Audio Eng. Soc.*, 47(1/2), January/February 1998. Commemorative issue.

AES 16th Int. Conf. *Audio Eng. Soc. 16th Int. Conf.: Spatial Sound Reproduction*, Apr. 1999. Audio Eng. Soc.

S. Barnett and T. M. Cronin. *Mathematical Formulae*. Longman Scientific & Technical, Longman Group UK Limited, fourth edition, 1986.

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

B. Bauer. Stereophonic earphones and binaural loudspeakers. *J. Audio Eng. Soc.*, 9 (2):148–151, Apr. 1961.

D. R. Begault. *3-D sound for Virtual Reality and Multimedia*. AP Professional, 1994.

D. R. Begault. Auditory and non-auditory factors that potentially influence virtual acoustic imagery. In AES 16th Int. Conf. AES 16th Int. Conf., pages 13 – 26.

D. R. Begault. Virtual acoustic displays for teleconferencing: Intelligibility advantage for "telephone-grade" audio. *J. Audio Eng. Soc.*, 47(10):824 – 828, Oct. 1999b.

L. R. Bernstein. Detection and discrimination of interaural disparities: Modern earphone-based studies. In Gilkey and Anderson [1997].

J. Blauert. An introduction to binaural technology. In Gilkey and Anderson [1997].

J. Blauert. *Spatial Hearing*. The MIT Press, revised edition, 1997b.

M. D. Burkhard. Binaural measurements and applications. In Gilkey and Anderson [1997].

P. Clarkson, J. Mourjopoulos, and J. Hammond. Spectral, phase and transient equalization for audio systems. *J. Audio Eng. Soc.*, 33(3):127 – 132, Mar. 1985.

R. K. Clifton and R. L. Freyman. The presedence effect: Beyond echo suppression. In Gilkey and Anderson [1997].

D. H. Cooper and J. L. Bauck. Prospects for transaural recording. *J. Audio Eng. Soc.*, 37(1/2):3–19, January/February 1989.

L. Cremer, H. A. Müller, and T. J. Schultz. *Principles and Applications of Room Acoustics*, volume 1. Applied Science Publishers, 1982a.

L. Cremer, H. A. Müller, and T. J. Schultz. *Principles and Applications of Room Acoustics*, volume 2. Applied Science Publishers, 1982b.

E. L. Crow, F. A. Davis, and M. W. Maxfield. *Statistics Manual*. Dover Publications, Inc., 1960.

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

P. Damaske and V. Mellert. Ein verfahren zur richtungstreuen schallabbildung des oberen halbraumes über zwei lautsprecher. *Acustica*, 22:153 – 162, 1969/70.

P. Damaske and B. Wagener. Richtungshörversuche über einen nachgebildeten kopf. *Acustica*, 21:30 – 35, 1969.

R. O. Duda. Elevation dependence of the interaural transfer function. In Gilkey and Anderson [1997].

R. O. Duda and W. L. Martens. Range dependece of the response of a sperical head model. *J. Acoust. Soc. Am.*, 104(5):3048 – 3057, Nov. 1998.

C. H. Edwards and D. E. Penney. *Calculus and Analytic Geometry*. Prentice Hall International, Inc., third edition, 1990.

Encyclopædia Britannica Online. "hearing". `<URL: http://www.eb.com:180/bol/topic?eu=40542\&sctn=1>`, Accessed 20 January 2000.

A. Farina. Aurora 3.1 - home page. <URL: http://aurora.ramsete.com/aurora/>, Accessed 8th August 2000.

A. Farina and F. Righini. Software implementation of an MLS analyzer with tools for convolution, auralization and inverse filtering. In *103rd Audio Eng. Soc. Convention Preprints* Aud [1997b]. Preprint 4605.

K. Foo, M. Hawksford, and M. Hollier. Pair-wise loudspeaker paradigms for multi-channel audio in home theatre and virtual reality. In *105th Audio Eng. Soc. Convention Preprints* Aud [1998]. Preprint 4796.

K. C. K. Foo, M. O. J. Hawksford, and M. P. Hollier. Optimization of virtual sound reproduced using two loudspeakers. In AES 16th Int. Conf. AES 16th Int. Conf., pages 366 – 378.

J. Garas and P. C. Sommen. Improving virtual sound source robustness using multiresolution spectral analysis and synthesis. In *105th Audio Eng. Soc. Convention Preprints* Aud [1998]. Preprint 4824.

W. G. Gardner. *3-D Audio Using Loudspeakers*. Kluwer Academic Publishers, 1998.

H. W. Gierlich. The application of binaural technology. *Applied Acoustics*, 36:219–243, 1992. Elsevier Science Publishers Ltd, England.

H. W. Gierlich and K. Genuit. Processing artificial-head recordings. *J. Audio Eng. Soc.*, 37(1/2):34 – 39, January/February 1989.

R. H. Gilkey and T. R. Anderson, editors. *Binaural and Spatial Hearing in Real and Virtual Environments*. Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey, 1997.

D. Griesinger. Equalization and spatial equalization of dummy-head recordings for loudspeaker reproduction. *J. Audio Eng. Soc.*, 37(1/2):20 – 29, January/February 1989.

H. Haas. Über den einfluss eines einfachechos auf die hörsamkeit von sprache. *Acustica*, 1, 1951.

D. Hammershøi. *Binaural Technique - a method of true 3D sound reproduction*. PhD thesis, Aalborg University, 1995.

D. Hammershøi. Fundamental aspects of the binaural recording and synthesis techniques. In *100th Audio Eng. Soc. Convention Preprints*. Audio Eng. Soc., May 1996. Preprint 4155.

D. Hammershøi and H. Møller. Sound transmission to and within the human ear canal. *J. Acoust. Soc. Am.*, 100(1):408 – 427, July 1996.

W. M. Hartmann. Auditory localization in rooms. In *Audio Eng. Soc. 12th Int. Conf.: Perception of Reproduced Sound* Aud [1993]. ISBN No. 0-937803-19-7.

W. M. Hartmann. Listening in a room and the precedence effect. In Gilkey and Anderson [1997].

W. M. Hartmann. How we localize sound. *Physics Today*, pages 24 – 29, Nov. 1999.

J. Hilditch. *Trangviksposten*, volume 7. Grøndahl & Søns Forlag, Oslo, 1970. (Reprint as book).

T. Holmefjord. Subjective attributes and room-acoustical parameters in olavshallen. Term project, NTNU, Norwegian University of Science and Technology, 1997. (In Norwegian).

U. Horbach and R. Pellegrini. Design of positional filters for 3D audio rendering. In *105th Audio Eng. Soc. Convention Preprints* Aud [1998]. Preprint 4798.

A. Høyland. *Sannsynlighetsregning og statistisk metodelære*, volume 1, Sannsynlighetsregning. Tapir, Trondheim, Norway, 4. edition, 1989a.

A. Høyland. *Sannsynlighetsregning og statistisk metodelære*, volume 2, Statistisk metodelære. Tapir, Trondheim, Norway, 4. edition, 1989b.

C. Hugonnet and P. Walder. *Stereophonic Sound Recording.* John Wiley & Sons, 1998.

J. Huopaniemi. *Virtual Acoustics and 3-D Sound in Multimedia Signal Processing.* PhD thesis, Helsinki University of Technology, 1999.

J. Huopaniemi and K. A. J. Riederer. Measuring the effect of source distance in head-related transfer functions. In *16th Int. Congress on Acoustics*, 1998.

J. Huopaniemi, N. Zacharov, and M. Karjalainen. Objective and subjective evaluation of head-related transfer function filter design. *J. Audio Eng. Soc.*, 47(4):218 – 239, Apr. 1999.

R. R. Johnson. *Elementary statistics.* Duxbury Press, Belmont, California, 1992.

Y. Kahana, P. A. Nelson, M. Petyt, and S. Choi. Boundary element simulation of HRTFs and sound fields produced by virtual acoustic imaging. In *105th Audio Eng. Soc. Convention Preprints* Aud [1998]. Preprint 4817.

Y. Kahana, P. A. Nelson, and S. Yoon. Experiments on the synthesis of virtual acoustic sources in automotive interiors. In AES 16th Int. Conf. AES 16th Int. Conf., pages 218 – 232.

O. Kirkeby and P. A. Nelson. Virtual source imaging using the "stereo dipole". In *103rd Audio Eng. Soc. Convention Preprints* Aud [1997b]. Preprint 4574.

O. Kirkeby, P. A. Nelson, and H. Hamada. The "stereo dipole" - binaural sound reproduction using two closely spaced loudspeakers. In *102nd Audio Eng. Soc. Convention Preprints* Aud [1997a]. Preprint 4463.

O. Kirkeby, P. A. Nelson, and H. Hamada. Local sound field reproduction using two closely spaced loudspeakers. *J. Acoust. Soc. Am.*, 104(4):1973 – 1981, Oct. 1998.

O. Kirkeby, P. Rubak, L. G. Johansen, and P. A. Nelson. Implementation of cross-talk cancellation networks using warped FIR filters. In AES 16th Int. Conf. AES 16th Int. Conf., pages 358 – 365.

D. J. Kistler and F. L. Wightman. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Am.*, 91(3):1637 – 1647, Mar. 1992.

M. Kleiner, B.-I. Dalenbäck, and P. Svensson. Auralization - an overview. *J. Audio Eng. Soc.*, 41(11):861–875, Nov. 1993.

S. Kløften. Binaural sound reproduction for close loudspeaker positions. Master's thesis, NTNU, Norwegian University of Science and Technology, 2001.

A. Krokstad. Taleteknologi. Course material, Speech and Music Technology, 1994. (In norwegian).

G. F. Kuhn. Model for the interaural time difference in the azimuthal plane. *J. Acoust. Soc. Am.*, 62(1):157 – 167, July 1977.

A. Kulkarni, S. Isabelle, and H. Colburn. Sensitivity of human subjects to head-related transfer function phase spectra. *J. Acoust. Soc. Am.*, 105(5):2821 – 2840, May 1999.

H. Kuttruff. *Room Acoustics.* Spon Press, New York, fourth edition, 2000.

V. Larcher. Equalization methods in binaural technology. In *105th Audio Eng. Soc. Convention Preprints* Aud [1998]. Preprint 4858.

D. M. Larsen. Digital filtering for optimalisations of the low-frequency response of a stereo pairs of loudspeakers to a room. Master's thesis, Norwegian University of Science and Technology, NTNU, Dept. of Telecomm., Acoustics Group, 1997. (In norwegian).

H. Lehnert. Auditory spatial impression. In *Audio Eng. Soc. 12th Int. Conf.: Perception of Reproduced Sound* Aud [1993]. ISBN No. 0-937803-19-7.

R. Y. Litovsky, S. Colburn, W. A. Yost, and S. J. Guzman. The presedence effect. *J. Acoust. Soc. Am.*, 106(4):1633–1654, Oct. 1999.

J. J. Lopez, A. Gonzalez, and F. Orduña-Bustamante. Measurement of cross-talk cancellation and equalization zones in 3-D sound reproduction under real listening conditions. In AES 16th Int. Conf. AES 16th Int. Conf., pages 349 – 357.

J. C. Makous and J. C. Middlebrooks. Two-dimensional sound localization by human listeners. *J. Acoust. Soc. Am.*, 87(5):2188–2200, May 1990.

Getting Started with MATLAB, Version 5. The MathWorks Inc., Dec. 1996.

A. McKeag and D. S. McGrath. Using auralisation techniques to render 5.1 surround to binaural and transaural playback. In *102nd Audio Eng. Soc. Convention Preprints* Aud [1997a]. Preprint 4458.

R. L. McKinley and M. A. Ericson. Flight demonstration of a 3-d auditory display. In Gilkey and Anderson [1997].

S. Mehrgardt and V. Mellert. Transformation characteristics of the external human ear. *J. Acoust. Soc. Am.*, 61(6):1567 – 1576, June 1977.

J. C. Middlebrooks. Spectral shape cues for sound localization. In Gilkey and Anderson [1997].

J. C. Middlebrooks, J. C. Makous, and D. M. Green. Directional sensivity of sound-pressure levels in the human ear canal. *J. Acoust. Soc. Am.*, 86(1):89 – 108, July 1989.

A. W. Mills. Auditory localization. In J. V. Tobias, editor, *Foundations of Modern Auditory Theory*, volume II. Academic Press, 1972.

B. C. J. Moore. Controversies and mysteries in spatial hearing. In AES 16th Int. Conf. AES 16th Int. Conf., pages 249 – 256.

P. M. Morse and K. U. Ingard. *Theoretical Acoustics*. Princeton University Press, first princeton university press edition edition, 1986.

J. Mourjopoulos. On the variation and invertibility of room impulse response functions. *J. Sound and Vibration*, 102(2):217 – 228, 1985.

J. N. Mourjopoulos. Digital equalization of room acoustics. *J. Audio Eng. Soc.*, 42 (11):884–900, Nov. 1994.

H. Møller. Reproduction of artificial-head recordings trough loudspeakers. *J. Audio Eng. Soc.*, 37(1/2):30 – 33, January/February 1989.

H. Møller. Fundamentals of binaural technology. *Applied Acoustics*, 36:171–218, 1992. Elsevier Science Publishers Ltd, England.

H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen. Evaluation of artificial heads in listening tests. *J. Audio Eng. Soc.*, 47(3):83 – 100, Mar. 1999.

H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen. Head-related transfer functions of human subjects. *J. Audio Eng. Soc.*, 43(5):300 – 321, May 1995.

H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammersøi. Binaural technique: Do we need individual recordings? *J. Audio Eng. Soc.*, 44(6):451 – 469, June 1996.

S. B. Neely and J. B. Allen. Invertibility of a room impulse response. *J. Acoust. Soc. Am.*, 66(1):165–169, July 1979.

P. A. Nelson, H. Hamada, and S. J. Elliott. Adaptive inverse filters for stereophonic sound reproduction. *IEEE Transactions on Signal Processing*, 40(7):1621 – 1632, July 1992.

P. A. Nelson, O. Kirkeby, H. Hamada, et al. Virtual sound images / stereo dipole. Institute of Sound and Vibration Research, University of Southampton, UK, publishing date not given. Audio CD.

P. A. Nelson and F. Orduña Bustamante. Multichannel signal processing techniques in the reproduction of sound. *J. Audio Eng. Soc.*, 44(11):973–989, Nov. 1996.

D. R. Perrot. Auditory and visual localization: Two modalities, one world. In *Audio Eng. Soc. 12th Int. Conf.: Perception of Reproduced Sound* Aud [1993]. ISBN No. 0-937803-19-7.

J. Plogsties, S. K. Olesen, P. Minnaar, F. Christensen, and H. Møller. Audibility of all-pass components in head-related transfer functions. In *108th Audio Eng. Soc. Convention Preprints*. Audio Eng. Soc., Feb. 2000. Preprint 5132.

J. G. Proakis and D. G. Manolakis. *Digital Signal Processing: Principles, Algorithms, and Applications*. Macmillian Publishing Company, second edition, 1992.

V. Pulkki, M. Karjalainen, and J. Huopaniemi. Analyzing virtual sound source attributes using a binaural audiotory model. *J. Audio Eng. Soc.*, 47(4):203 – 217, Apr. 1999.

B. Rakerd and W. M. Hartmann. Localization of sound in rooms, II: The effects of a single reflecting surface. *J. Acoust. Soc. Am.*, 78(2):524 – 533, Aug. 1985.

A. Rimmel. Imersive spatial audio for telepresence applications: Sustems design and implementation. In AES 16th Int. Conf. AES 16th Int. Conf., pages 379 – 390.

R. F. Schmidt, editor. *Fundamentals of Sensory Physiology*. Springer-Verlag, 1978.

M. R. Schroeder. Computers in acoustics: Symbiosis of an old science and a new tool. *J. Acoust. Soc. Am.*, 45(5):1077 – 1088, 1969.

M. R. Schroeder. Digital simulation of sound transmission in reverberant spaces. *J. Acoust. Soc. Am.*, 47(2):424 – 431, 1970.

M. R. Schroeder. Computer models for concert hall acoustics. *Am. J. Phys.*, 41:461 – 471, Apr. 1973.

M. R. Schroeder and B. S. Atal. Computer simulation of sound transmission in rooms. In *IEEE International Convention Record*, number 7, pages 150 – 155. The Institute of Electrical and Electronics Engineers, May 1963.

M. R. Schroeder, B. S. Atal, and C. Bird. Digital computers in room acoustics (M21). In *Fourth Int. Congress on Acoustics*, 1962.

E. A. G. Shaw. Acoustical features of the human external ear. In Gilkey and Anderson [1997].

B. Shinn-Cunningham, H. Lehnert, G. Kramer, E. Wenzel, and N. Durlach. Auditory displays. In Gilkey and Anderson [1997].

B. G. Shinn-Cunningham, P. M. Zurek, and N. I. Durlach. Adjustment and discrimination measurements of the presedence effect. *J. Acoust. Soc. Am.*, 93(5):2923 – 2932, May 1993.

M. L. Strømsvåg. Auditive discrimination in virtual rooms. Master's thesis, Norwegian University of Science and Technology, NTNU, Dept. of Telecomm., Acoustics Group, 1996. (In norwegian).

A. Sæbø. Implementation of transaural systems in software on a pc. In *105th Audio Eng. Soc. Convention Preprints*. Audio Eng. Soc., Sept. 1998. Preprint 4799.

A. Sæbø. Effect of early reflections in binaural systems with loudspeaker reproduction. In *2nd COST-G6 Workshop on Digital Audio Effects (DAFx99)*, Dec. 1999.

T. Takeuchi, P. Nelson, O. Kirkeby, and H. Hamada. The effects of reflections on the performance of virtual acoustic imaging systems. In *The 1997 International Symposium on Active Control of Sound and Vibration (ACTIVE 97)*, pages 927 – 940, Budapest, Hungary, Aug. 1997a.

T. Takeuchi and P. A. Nelson. Robustness of the performance of the "stereo dipole" to head misalignment. Technical report, Institute of Sound and Vibration Research, University of Southampton, Southampton, England, Oct. 1999. ISVR Technical Report No. 285.

T. Takeuchi, P. A. Nelson, O. Kirkeby, and H. Hamada. Robustness of the performance of the "stereo dipole" to misalignment of head position. In *102nd Audio Eng. Soc. Convention Preprints* Aud [1997a]. Preprint 4464.

M. Teschl. Transaural stereo - influence of early reflection on localization. Term project, NTNU, Norwegian University of Science and Technology, Apr. 1999.

E. Torick. Highlights in the history of multichannel soound. *J. Audio Eng. Soc.*, 46 (1/2):27 – 31, January/February 1998.

A. Ustad. Målsatte skisser av lydrom, klangrom og ekkofritt rom ved akustisk laboratorium, med absorbenter og diffusorer. SINTEF ELAB: Project number 441900.13 (In norwegian), June 1985.

D. B. Ward and G. W. Elko. Optimum loudpsekaer spacing for robust crosstalk cancellation. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pages 3541–3544. IEEE, Signal Processing Soc., 1998.

Webster (1913). "binaural". `http://work.ucsd.edu:5141/cgi-bin/http_webster?binaural&method=exact`, 2000. Webster's Revised Unabridged Dictionary (1913). Accessed 25th August 2000.

E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman. Localization using noindividualized head-related transfer functions. *J. Acoust. Soc. Am.*, 94(1):111–123, July 1993.

J. E. West, J. Blauert, and D. MacLean. Technical note: Teleconferencing system using head-related signals. *Applied Acoustics*, 36:327–333, 1992. Elsevier Science Publishers Ltd, England.

F. L. Wightman and D. J. Kistler. Headphone simulation of free-field listening. I: stimulus synthesis. *J. Acoust. Soc. Am.*, 85(2):858–867, Feb. 1989a.

F. L. Wightman and D. J. Kistler. Headphone simulation of free-field listening. II: psychophysical validation. *J. Acoust. Soc. Am.*, 85(2):868–878, Feb. 1989b.

F. L. Wightman and D. J. Kistler. The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.*, 91(3):1648–1661, Mar. 1992.

F. L. Wightman and D. J. Kistler. Factors affecting the relative salience of sound localization cues. In Gilkey and Anderson [1997].

F. L. Wightman and D. J. Kistler. Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. Am.*, 105(5):2841 – 2853, May 1999.

G. Woje. 3D sound from two loudspeakers in a room. Term project, NTNU, Norwegian University of Science and Technology, 1997. (In Norwegian).

E. Zwicker and H. Fastl. *Psychoacoustics*. Springer Series in Information Sciences. Springer, second updated edition, 1999.

E. Zwicker and R. Feldtkeller. *Das Ohr als Nachrichtenempfänger*. Number XIX in Monographien der elektrischen Nachrichtentechnik. S. Hirzel Verlag, Stuttgart, 1967.