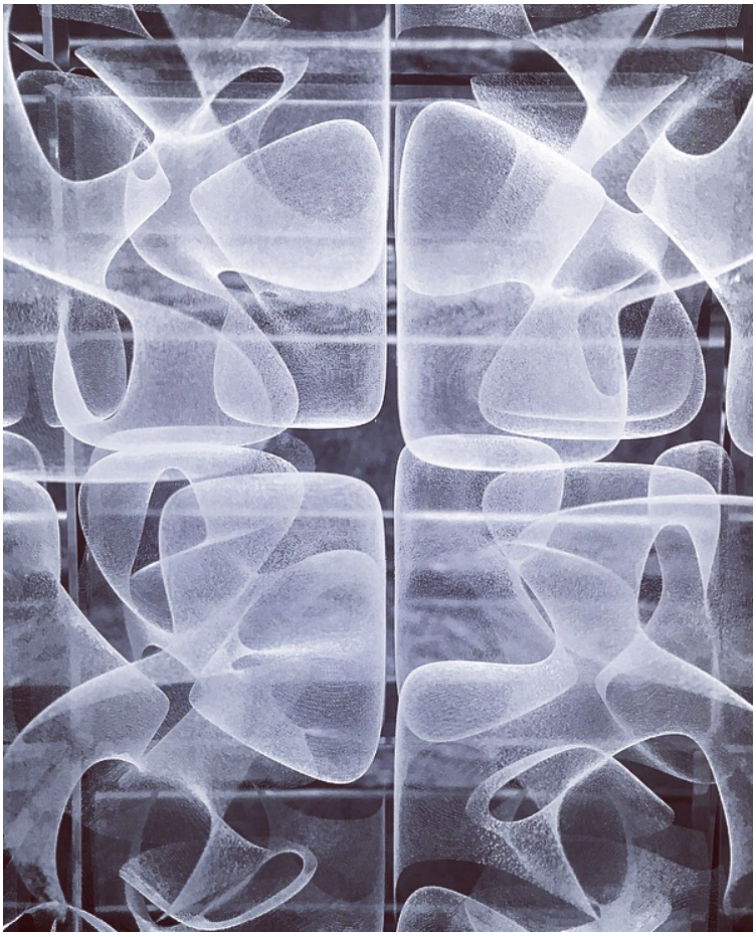
CHALMERS
UNIVERSITY OF TECHNOLOGY



Loudspeaker Directivity and Playback Environment in Acoustic Crosstalk Cancelation

Master's thesis in Master Program Sound and Vibration

KARIM BAHRI

Division of Applied Acoustics
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2019

Master's thesis ACEX30-19-102

Loudspeaker Directivity and Playback Environment in Acoustic Crosstalk Cancelation

KARIM BAHRI

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Architecture and Civil Engineering
Division of Applied Acoustics
Audio Technology Group
Chalmers University of Technology
Gothenburg, Sweden 2019

Loudspeaker Directivity and Playback Environment in Acoustic Crosstalk Cancelation

KARIM BAHRI

Supervisor: Jens Ahrens, Department of Architecture and Civil Engineering
Examiner: Jens Ahrens, Department of Architecture and Civil Engineering

Master's Thesis ACEX30-19-102
Department of Architecture and Civil Engineering
Division of Applied Acoustics
Audio Technology Group
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Standing Waveforms 2.0 (2019), courtesy of Ricardo Mondragon.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2019

Loudspeaker Directivity and Playback Environement in Acoustic Crosstalk Cancelation

KARIM BAHRI
Division of Applied Acoustics
Chalmers University of Technology

Abstract

Audiovisual immersive interfaces are growing in popularity in today's world. For the audio part, the "immersive experience" can be accomplished by means of systems that are able to surround a listener with sounds coming from arbitrary locations. When that is done with a set of loudspeakers, the performance of the acoustic crosstalk cancelation process is essential for achieving a convincing immersion. This thesis aims to improve a previously proposed beamforming-based crosstalk cancelation system that uses a linear array of loudspeakers and to verify its performance through simulation.

As the original beamformer employed a point-source model for the loudspeakers, here we investigate the effect of loudspeaker radiation properties on the performance of the system and how this contribution departs significantly from that of the point source model. It is demonstrated that the measured channel separation between the listener's ears increases when the actual loudspeaker directivities are taken into consideration in the beamformer design. The improvement is mainly noticeable in the frequency range of 1-2 kHz and is globally approximated to 3 dB over the frequency range where beamforming is applied.

This thesis also investigates the perceptual effect of different reflecting surfaces that are apparent in the reproduction environment on binaural audio content that is presented through that system. A user study shows that as reverberation from the playback environment increases, the general perception is more pleasant, the impression of space is expanded and feels more real, the front-back confusion is mitigated and even a strong lateral reflection does not weaken the localization cues in a significant way.

Keywords: beamforming, binaural audio, crosstalk cancelation, linear loudspeaker array, loudspeaker directivity, room acoustics

Acknowledgements

I would like to deeply thank Dr. Jens Ahrens for his dedicated supervision on this project, for the extraordinarily patient guidance and for all the help provided in order to carry out this project.
I am also very thankful to Georgios Zachos for building the loudspeaker array, as well as to Christoph Hohnerlein and Xiaohui Ma for the highly useful previous related work.
I would like to acknowledge Niklas Zeidler, Carl Andersson and Hannes Helmholz for the many advices and the time spent with me discussing practical signal processing issues.
Many thanks to the wonderful people of the Division of Applied Acoustics for providing an enjoyable and stimulating environment.
Dr. Hached Gaubi and Dr. Dominique Chéenne deserve my sincere gratitude for bringing me into the brilliant field of acoustics.
Last but not least, I would like to express my deepest gratitude to my family and closest friends for their support throughout the journey of my studies.

Karim Bahri, Gothenburg, August 2019

Contents

List of Figures

1
Introduction

A 3-D audio experience can be brought to a listener by means of binaural audio reproduction, which independently provides each ear with signals on which head-related transfer functions (HRTFs) are encoded. Due to its natural channel separation, headphone-based reproduction of binaural audio content is commonly used. However, traditional headphones reproduction has some drawbacks as it suffers from head internalization of sound [1], it can cause social isolation and finally it is excluded from a range of situations where wearing headphones might be unacceptable. That is why loudspeaker binaural rendering could be a sound alternative.
In this case, a fundamental challenge to overcome is to achieve the maximum of separation between the channels reaching each of the listener's ears, i.e. acoustic crosstalk must be meticulously eliminated. To preserve the localization cues encoded in the binaural audio content and without introducing any change in amplitude of phase to the original channels, an ideal acoustic crosstalk cancelation system would provide a signal intended for the ipsi-lateral ear while that will not be received by the contra-lateral ear.
In the field of acoustics, Crosstalk Cancelation (CTC) has been pursued for more than 50 years [2]. Early implementations employed a pair of loudspeakers and CTC was achieved by filter inversion of the transmission matrix between the ears [2, 3]. Back then, the desired result would easily break down in the presence of even small deviations from the assumptions. Through the years, there have been many attempts to develop more robust systems either by means of keeping track of the listener's head and adjusting the filter inversion respectively [4], or by optimizing the position of the loudspeakers set [5]. In an alternative fashion, a technique that stood apart from the rest was the Recursive Ambiophonic Crosstalk Elimination (RACE) [6] which provides simple means of CTC for a two loudspeakers symmetric setup. This latter one is robust with respect to head movement, although the system performance strongly depends on the loudspeaker position and even the loudspeaker model. In the 1990s, the capabilities of loudspeaker arrays to perform acoustic CTC have been evaluated [7] and recent research using more than two loudspeakers have introduced interesting results.
Lately, Hohnerlein, Ahrens and Ma [8, 9, 10] proposed a superdirective beamforming-based acoustic CTC system employing a linear equispaced 8-channel loudspeaker array. This approach is the basis for the work presented through this document.
While the original beamformer makes the assumption that the used loudspeakers emit ideal spherical waves, this master thesis investigates the effect of actual loudspeaker radiation properties on the CTC performance by measuring the sources directivities and introducing them into the beamformer.

CTC systems are usually designed assuming free-field conditions. The impact that reverberation has on channel separation was studied instrumentally in [11], while the effect of a low number of lateral early reflections on localization of sound sources in binaural audio content was studied in [12]. This latter study used a setup composed of two loudspeakers, initially operating under anechoic conditions to which controlled reflections were introduced by means of flat rigid lateral surfaces. Among the main findings were that such isolated early reflections can cause a localization bias towards the direction where the reflection is coming from. Also, front-back confusions were higher in the non-anechoic conditions with a bias towards localization in the front hemisphere where the pair of loudspeakers were located. As specified above, the experiments were conducted with reflections produced by lateral surfaces, so the effect of floor reflection was not explored.
In another study [13], this one based on a simulation with an image-source model, it was found that the room response did not affect localization in a significant manner. It is important to note that such simple room simulations do not account for all relevant acoustic effects and can sound artificial when rendered spatially, while the physical manipulation of reflective surfaces, as in [12], limits the range of acoustical conditions that can be experimented. Therefore, in this master thesis, impulse responses from the loudspeakers to the ears of a dummy head were measured in different acoustic environments and then used to auralize the CTC system through a set of headphones. This approach allows for switching between environments by the click of a button and reduces the limits on the range of acoustical conditions that can be covered.
As a second topic, the present thesis investigates the perceptual effects that various apparent reflecting surfaces in the playback room have on audio binaural content presented through the CTC system.

This document is structured in the following way. Chapter 2 briefly presents the theory regarding sound localization cues and spatial sound perception, audio transmission properties, loudspeaker directivity, acoustic crosstalk cancelation and finally beamforming. Chapter 3 deals with the description of the linear loudspeaker array used in this work, the performed measurements and the implementation of the processed data. Chapter 4 presents and discusses the obtained results from the incorporation of the actual source directivities into the beamformer, as well as the results of the perceptual evaluation from the auralization of the CTC system in different playback environments. Finally, chapter 5 draws conclusions and suggests future research directions.

2
Theory

The following chapter gives a short overview on the topics and techniques used in this thesis.

2.1 Binaural Hearing and Localization Cues

Binaural means listening with two ears, as compared to monaural, which means listening with one ear. Listening with two ears gives the listener accurate information about sound location and leads to a clarity and subjective perception of auditory space. The localization ability is the the result of the ability of hearing to perform cross-correlation analysis of the signals from the two ears. Spatial hearing is the capacity of a listener to analyze and process a real or virtual auditory scene to localize a sound in space. A special case of binaural hearing is virtual spatial hearing, which refers to the formation of synthetic spatial acoustic imagery using a binaural playback system (headphones or loudspeakers with the respective crosstalk cancelation filter).

The main cues for localizing both real and virtual sound sources are interaural cues: Interaural Level Difference (ILD)and Interaural Time Difference (ITD), and spectral cues. In addition to the above, there are dynamic cues evaluated in specific cases.
Any sound source that is not positioned on the vertical plane directly between the listener's ears, called the median plane, will result in a difference in sound pressure reaching both ears. This is called Interaural Level Difference.
The same way, pressure waves propagating from sound sources positioned outside the median plane will reach the contra-lateral ear slightly later. The time difference corresponds to the travel time across the listener's head at the propagation speed of sound in air. This is called Interaural Time Difference.
Binaural cues are not sufficient to account for all aspects of sound localization. For example, an ILD or ITD will not indicate whether a sound is coming from in front or behind, or above or below, but such evaluations can clearly be made. The shape of the pinna (outer ear) and the geometry of the listener's head play an important role in sound localization. One will perceive a difference between a sound source straight in front of the listener and an identical source straight behind as the spectra of sounds entering the ear are modified due to the reflection and shadow action of the head. This direction-dependent filtering provides cues for sound source location.
For each ITD, there is a cone of possible sound source locations, extending from the side of the head, that will produce that time difference. Locations on such a

"cone of confusion" may also produce similar ILDs. In that case, the listener has to exclusively rely on the spectral cues for correct localization. It is a common problem in binaural synthesis, called front-back confusion.

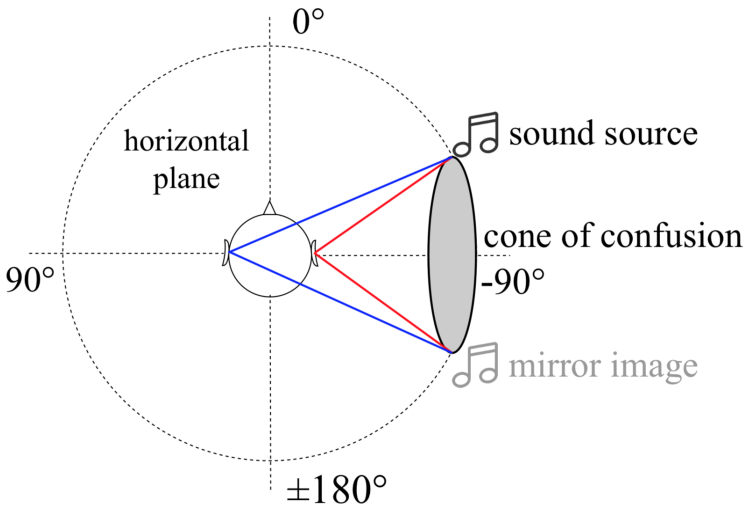


Figure 2.1: Cone of confusion as an open cone of points that all invoke the same ITD and ILD cues

If a sound source is directly in front of the listener, then turning the head to the left will decrease the pressure level in the left ear and cause the propagating wave to reach the right ear before the left ear. Therefore, much of the ambiguity about sound localization of static sound sources can be resolved by slight movements of the head. This is called dynamic cues. In binaural synthesis, this can be recreated by encouraging natural head movements while moving the source in relation to the listener.

2.2 Room Acoustics Impact on Spatial Sound Perception

Sound waves radiating from any kind of source go through various states before reaching the listener's tympanic membranes (eardrums) or any other observation point in the room.

There are two types of sound waves that can radiate from a source: plane waves and the spherical waves. On one side, the specification of plane waves is that they have constant amplitude and phase on any perpendicular plane to their direction of propagation, meaning that their wave fronts can be considered as plane. On the other side, spherical waves are produced by the so-called point sources which can be thought of as really small sources placed in the centre of concentric spheres. In the case of spherical waves, the wave fronts are concentric and the acoustic variables are not constant but a function of radial distance. Generally, it is more convenient to work with plane waves instead of spherical, when examining the behaviour of

sound. However, at distances long enough from a point source the wave fronts can be regarded as plane.

In the case of multiple sound sources, the signal reaching any observation point is the superposition of all the signals originating from each individual sound source. In a given room different situations can be encountered, such as when the sound propagation is disturbed by obstacles and surfaces (walls, floor, ceiling) that would produce reflections when the sound wave impinges on them. In other terms, a listener placed in a typical sound environment receives signals that contain both the direct and the reflected sound waves. The sound propagation in a room can be influenced by the following parameters: dimensions of the room, obstacles (screens, columns), types of surfaces (hard, porous), as well as the presence of people within the room. All these parameters must be considered when aiming for specific sound reproduction conditions.

In order for a listener to approximate the physics of the sound, the superposition of the signals might be enough, however, the same cannot be said for the actual perception of the auditory event. As a matter of fact, the psychoacoustical phenomena cannot be described as linear systems and therefore, the reproduction of virtual sources in a reverberant environment has unpredictable results in terms of perception and localization. The following points can be stated:
- Since the loudspeakers aiming to reproduce a virtual source are usually placed in a different position from the simulated source, the reflections of the waves radiating from the loudspeakers are different from the reflections that occur with the wave from the position of a real sound source.
- In a case where the position of a virtual source and the reflecting waves of the real source correspond with the placement of the loudspeakers, the resulting auditory event can still differ because there are multiple sources (two loudspeakers or more) that are radiating signals and causing multiple reflections instead of one (real source). In addition to that, the sound reproduction setup does not change, meaning that the time delay of the reflections is always the same and different from the real one, regardless of the virtual source position.

When it comes to perception, a listener is familiar to the localization cues produced by a reverberant environment. The decrease in sound pressure level due to propagation attenuation in the air is used as an indicator of the distance from the source. Similarly, the ratio of direct to reverberant energy is a useful cue for approximating that distance.
A reverberant environment has also an impact on the interaural cues (ITD and ILD) that the listener perceives and it has been shown that the hearing system is tuned to the ITDs outside anechoic conditions. It can be said that the interaural cues from reverberant environments are indeed expected and used by the localization mechanism of hearing. It also appears that with reflections the range of both interaural cues extend further than when in anechoic conditions.
It is essential to note that sound reflections can be either helpful or detrimental in the localization process. In the simple situation of a single specular reflection from

a horizontal surface (floor, ceiling), the perceived image source can be assumed as having the same azimuth as the original sound source. That reflection can then be thought of as helping in the localization of the original source in the horizontal plane. However, situations that give rise to lateral reflections (side walls), will have image sources with different azimuth, and therefore their localization cues will not correspond to the ones from the original source.

In order to study the effect of various sound field parameters in a controlled and detailed way, it is necessary to move to the laboratory. For convenience, in psychoacoustic tests in the laboratory, a technique known as sound field synthesis is used. This technique involves the change of different sound field components, sound source radiation patterns, reflections, noise, so that the parameters that are thought to impact the spatial sound perception can be studied.
To have a complete control on the sound field to which the test person is exposed, usually sound field synthesis is applied by means of either binaural sound reproduction or sound reproduction over loudspeakers in an anechoic chamber. If placed in a regular room, the test person would be subject to uncontrolled sound field components due to sound reflections or other uncontrolled sound propagation phenomena. Some types of sound field synthesis use loudspeaker equivalent binaural sound reproduction, using either two loudspeaker or a loudspeaker array, like in this thesis work, and an associated technique known as acoustic crosstalk cancelation. In that case, the signals fed to the loudspeakers are processed by digital filtering in such a way that, at the listener ears, they correspond to those signals that would be obtained by conventional binaural listening using a set of headphones.

2.3 Precedence Effect

In an environment in which there are reflective surfaces, such as a room, the sound waves reaching a listener's ears are a complex combination of the sound coming directly from the source and the one reflected by nearby surfaces, as illustrated by Figure 2.2. Despite the fact that the reflected sound provides directional information that conflicts with that from the direct sound, the human auditory system can fairly well overcome this ambiguity and therefore locate sound sources in reverberant environments.

The precedence effect, also called the law of the first wave front, is a psychoacoustic effect that refers to our capacity to locate a source based on the dominance of information from the first arriving sound, while ignoring the late (delayed or reflected) sound information. The precedence effect is signal dependent and it only works for short reflections that correspond to path length differences of approximately 10 meters. For complex sounds, such as speech or music, the effect appears for time lags in the range of about 1 - 30 ms, while for clicks the upper limit of the precedence effect can be approximated to 5 ms. In simple terms, when the effect occurs, a single auditory event is perceived and the location of the source is determined by the location of the leading sound.

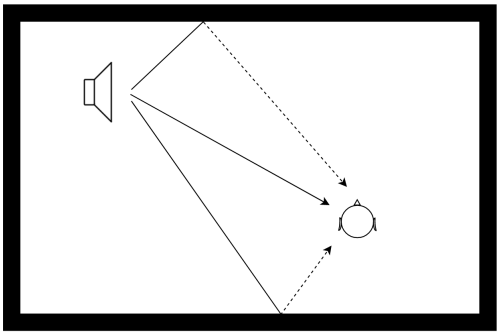


Figure 2.2: To localize a sound source in a reflective room, the listener must ignore sound waves that appear to originate from the direction nearby reflective surfaces

When the time lag between two coherent sounds is beyond the echo threshold, the precedence effect disappears and the percept breaks down into two sounds, the second one is then heard as an echo. In that case, both sounds are perceived to be coming from their respective direction of arrival to the listener's ears.

It is important to note that although the precedence effect shows that short-delay reflected sounds can be ignored for the purpose of localization, the information of the reflected sound still provide the listener with valuable information about the environment, such as the volume of the room and the distance from the sound source.

2.4 Auralization

In analogy with visualization by computer aided techniques, in order to achieve audible room soundfield simulation, a technique called auralization is used. It is defined as the process of rendering audible the soundfield of a source in a virtualized space by means of mathematical or physical modeling. The purpose of this method is to simulate the binaural listening experience at any given position in the virtualized space [14].
Today, auralization is a technique employed to generate coherent sound environments within virtual immersion systems, as well as a tool commonly used by acousticians to predict and simulate the performance of critical listening rooms so that their soundscapes can be adjusted.

2.5 HRIR, HRTF and BRIR

All transformations of a Linear, Time-Invariant (LTI) system are encoded in its response to a Dirac impulse $\delta(t)$. This response is called the Impulse Response (IR) $h(t)$ of that system. Its Fourier transform in the frequency domain is the Transfer Function (TF) $H(\omega)$ of that system.

Sound transmission through a linear medium, such as air, is considered to be a LTI system. Thus, all changes to an audio signal traveling from a certain point in space to the entrance of the ear canal can be captured in the Head-Related Impulse Responses (HRIR) as they describe the effects our outer ears, head and torso have on sound waves. HRIRs encode the psychoacoustic cues on source localization in space, such as level and time differences between the two ears, as well as the properties of the transmission room. The frequency domain counterpart to the HRIR is the Head-Related Transfer Function (HRTF). This can be seen as a filter that our ears, head, and torso apply to all sounds we hear.

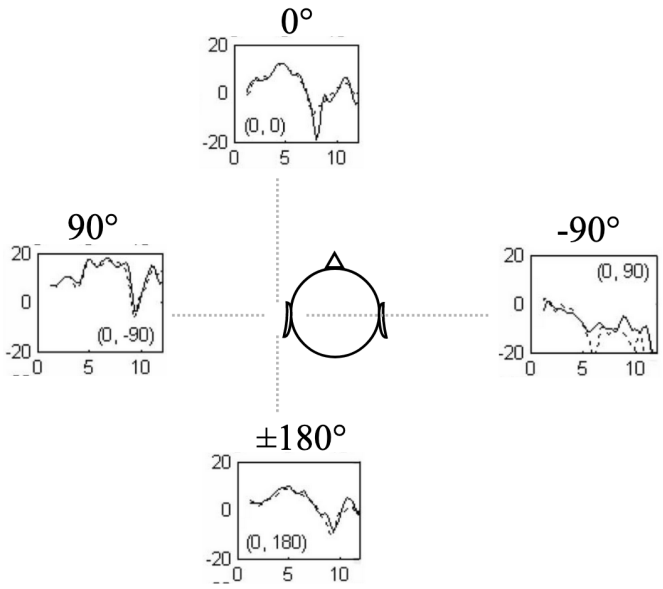


Figure 2.3: Horizontal plane HRTF examples, after [15]

In practice, such transfer functions can be measured with small microphones placed in the ear canals of a test subject. The test subject in question can be either a human person, or, for more repeatable results, a manikin that resembles average human proportions. If these HRTFs are measured for all directions, one at a time, the measurement set describes the information necessary for our hearing system to localize a sound source in space. These obtained filters (measured HRTFs) can be applied to audio signals that lack the aural localization cues. This will create an audio signal with a virtual position relative to the listener, the same position as the source position used in the recording of the HRTF. This method is called binaural synthesis.

To be independent of the room acoustic properties such as absorption and reverberation, HRIRs are measured under anechoic conditions. The acoustic properties of a room can be added later by convolving with the IR of that room measured with a soundfield microphone at the listener desired location in the room. Such a superposition of localization cues and room IR is referred to as Binaural Room Impulse Response (BRIR). Alternatively, a full set of HRIRs can be measured directly in the desired environment. This BRIR can then be convolved with an anechoic sound

signal in the time domain to generate a binaural audio signal that corresponds to the sound signal playing in the room described by the IR.

2.6 Radiation and Loudspeaker Directivity

To analyze the radiation by a loudspeaker, one can use an approximation, where the diaphragm is assumed to be a vibrating planar piston.
Realizing that a loudspeaker is omnidirectional at low frequencies, in order to find the pressure obtained in front of it, the expression of the pressure obtained at a distance r from an ideal point source (monopole) can be used. According to [16], the expression of the pressure is as follows:

$$p(r) = j\omega\rho_0 Q \frac{e^{-jkr}}{4\pi r} \tag{2.1}$$

where $\omega = 2\pi f$, ρ_0 is the air density and Q is the source strength ($Q = U$; U is the volume velocity).

In the literature, the expression of the pressure at a distance r from a small pulsating sphere (monopole) is also commonly given as:

$$p(r) = A \frac{e^{-jkr}}{r} \tag{2.2}$$

where A is the pressure amplitude.

When a monopole is on top of a hard surface, the resulting sound at the observation point will be the linear sum of the sound from the monopole and its mirror image, which is also a monopole. Therefore, each element on the surface area of a vibrating planar piston can be thought of as an individual, in-phase monopole, generating volume velocity [17]. Each monopole contributes to the total volume velocity due to the vibrational velocity of each diaphragm u and the total size of the surface element S, thus $U = uS$. The total pressure is then obtained by summing the sound pressure contributions from each surface element of the entire piston surface. For points far away from the piston, the following expression can be obtained:

$$p(r, \theta) = j\omega\rho_0 U \left[\frac{2J_1(ka\ \sin(\theta))}{ka\ \sin(\theta)} \right] \frac{e^{-jkr}}{2\pi r} \tag{2.3}$$

where a is the radius of the piston, θ is the angle off-axis and J_1 is the Bessel function of the first kind.

The radiated acoustic power is estimated by integration of the sound intensity for the far-field of the vibrating planar piston, where sound pressure and particle velocity are in phase. In the far-field, the sum pressure decreases with distance as $1/r$. The sound pressure in the far-field depends on the radiated power from the sound source, its directional characteristics and the sound field it radiates into.

The directivity of a source (loudspeaker) is specified with the Directivity Factor (DF), which is described as the intensity of the sound in the direction of maximum radiation relative to the intensity at the same distance from a monopole having the same radiated power [17]. Considering W_{tot} the total radiated power from a source, DF is given by the ratio [16]:

$$DF = \frac{I_{max}(r)}{I_{avg}(r)} = \frac{I_{max}(r)4\pi r^2}{W_{tot}} \tag{2.4}$$

where I_{max} is the intensity of the sound in the direction of maximum radiation and I_{avg} is the average sound intensity for a source in a 4π space, with $I_{avg}(r) = W_{tot}/4\pi r^2$.

As a metric for the directionality of a loudspeaker radiation, it is more common to express the directivity in decibels, then called the Directivity Index (DI), defined as:

$$DI = 10\log_{10}(DF) \tag{2.5}$$

A planar source radiates sound uniformly for low frequencies' wavelengths longer than the dimensions of the planar source, and as frequency increases, the sound from such a source becomes highly directional and focuses into an increasingly narrower angle. In fact, for $ka > 1$ the piston becomes large compared to the wavelength and the directivity increases proportional to ω^2. To illustrate that, Figure 2.4 shows the far-field directivity patterns for a piston having radius a in a baffle for three values of ka as a function of the angle θ relative to the z-axis.

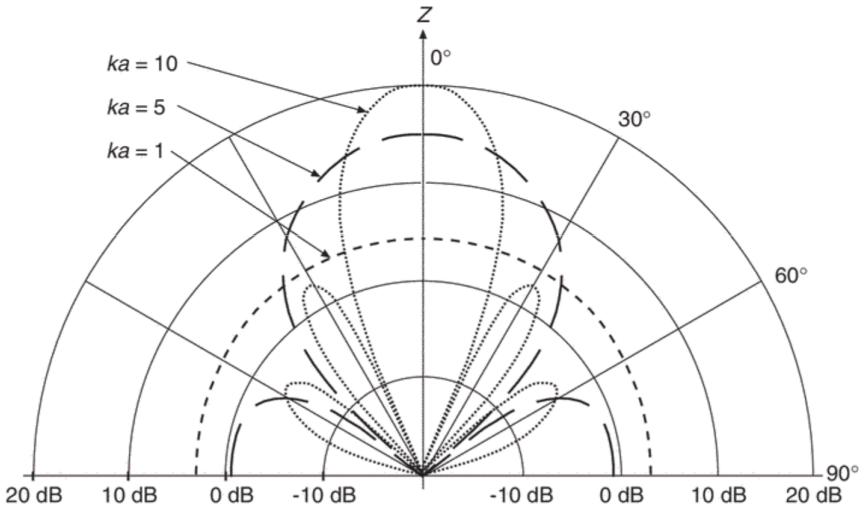


Figure 2.4: Directivity pattern: the curves show gain in dB over that of a free monopole having the same radiated power as the piston [17]

The on-axis sound pressure is the pressure along the axis normal to the vibrating planar piston, going through its center. Loudspeakers with a rapidly increasing directivity at high frequencies can give the impression that there is too much high frequency content if the listener is on-axis, or too little if the listener is off-axis. It is important to note that on-axis frequency response measurement is not a complete

characterization of the sound radiation of a loudspeaker.

In real life, individual loudspeaker drivers are complex three-dimensional shapes such as cones and domes, commonly placed on a baffle for various reasons. Therefore, in practice, the directivity of a loudspeaker is affected by how it is mounted, how many loudspeaker drivers are being used and how the signal is distributed between them. A loudspeaker directivity is also influenced by the diffraction of sound by the edges of the loudspeaker box. The directivity is an important issue to consider as the directional behavior of a loudspeaker determines the effective area of good listening in terms of frequency balance and the interaction of the sound system with the room and its contents.

2.7 Acoustic Crosstalk Cancelation

Crosstalk is defined as any phenomenon by which a signal transmitted on one transmission system channel creates an unwanted effect on another channel. It is a significant issue in electronics and communications systems. Acoustic crosstalk refers to the incomplete isolation of the left and right audio channels so that one leaks into the other, as illustrated in Figure 2.5.

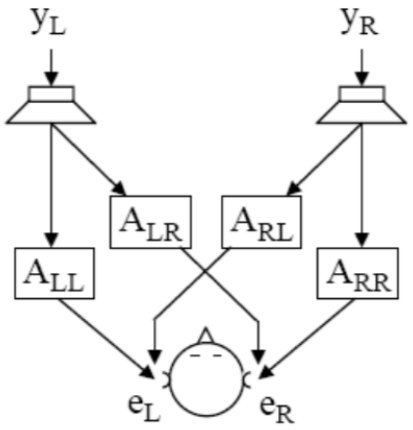


Figure 2.5: Acoustic transfer functions between a set of two loudspeakers and the listener's ears [18]

For binaural audio contents, any change of the ear signals will alter the perception in unpredictable ways. Acoustic crosstalk cancelation (CTC) is a technique that consists in sending independent signals to the ears of the listener while canceling the second arrival at each ear or in other terms canceling the crosstalk that transits to the opposite ear. The technique was first introduced by Bauer in 1961 and put into practice in 1963 by Schroeder and Atal.

Several methods have been tried since the 1960s. The usual method of creating crosstalk canceling filters is to invert head responses obtained by direct measurement or modelling, but it has been shown that CTC using inverse transfer functions

is not a very robust method.

Alternatively, a pragmatic and successful approach was proposed by Glasgal [6], where the path from a loudspeaker to the contra-lateral ear is estimated as a simple delayed attenuator. Thus, it can be canceled out by the inverse of that attenuated and delayed signal on the opposite channel. This cancelation signal, on its turn, also needs to be compensated on the original ear. This leads to the recursive filter at the heart of the proposed RACE system.

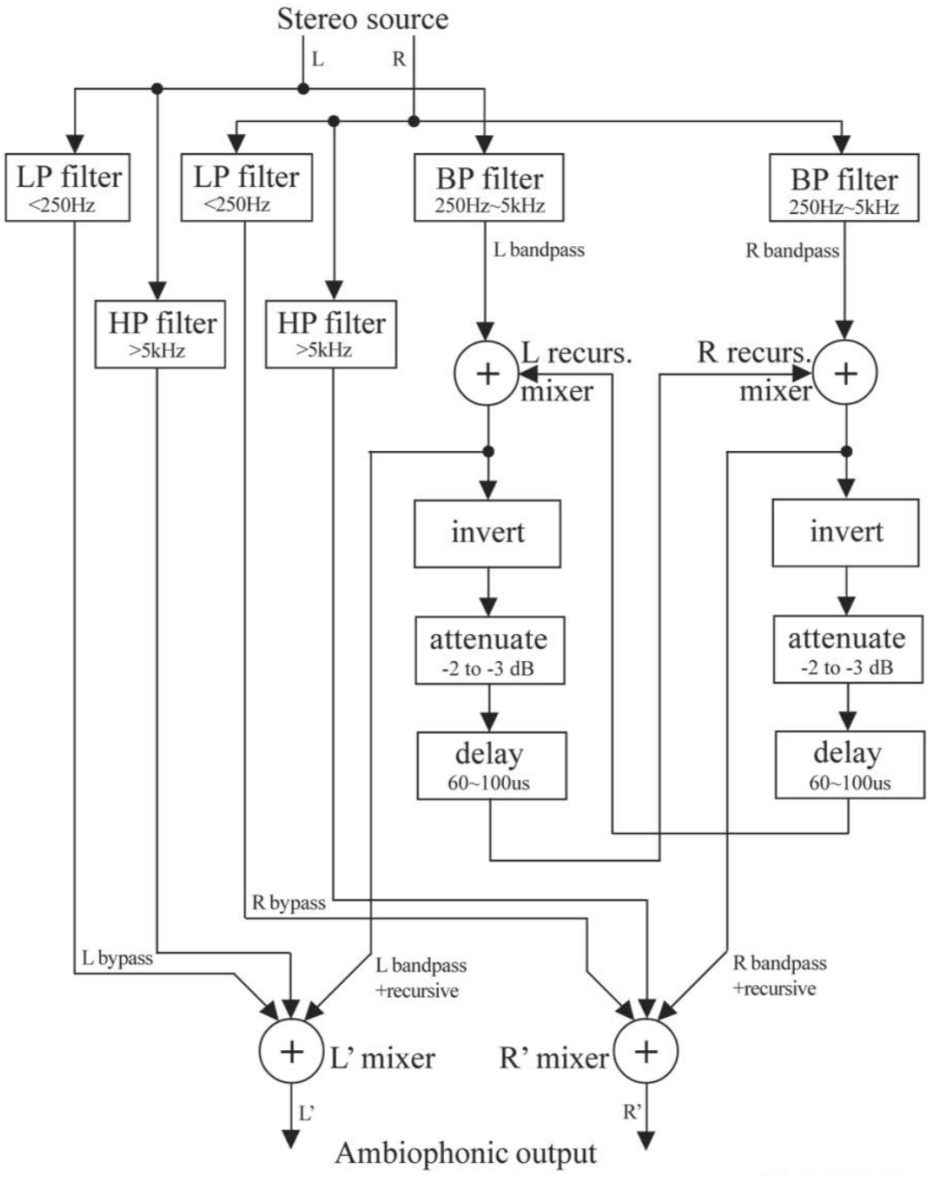


Figure 2.6: Block diagram of the RACE Processor [6]

More recently, other approaches leverage the optimization of loudspeaker arrays, similar to the system used in this thesis. Generally, they all have much improved robustness against displaced listeners and they tend to fail more gently than the previous technologies.

Hohnerlein, Ahrens and Ma [8, 9, 10] proposed a superdirective beamforming-based acoustic CTC using a linear equispaced 8-channel loudspeaker array. Furthermore, it introduces the idea of additionally using the RACE method to increase channel separation at lower frequencies.

2.8 Beamforming

This section on beamforming follows the literature in taking the perspective of an array of receiving sensors (microphones). Based on the Helmholtz Reciprocity, a source-receiver pair can be swapped and therefore, the same principles and design methods of beamforming hold for an array of emitting sources (loudspeakers arrays).

Beamforming, or spatial filtering, is a method for discriminating between different signals based on the physical location of the sources. Beamforming uses an array of microphones as an approach for creating a focused beam-like sensitivity pattern. In its simplest form, the beamformer is a single array of equally spaced sensors, of which each input may be independently delayed and weighted with a complex factor. The system output is then the simple summation of all individually delayed and weighted inputs. This is called a Delay-Sum beamformer.

$$y(k) = \sum_{n=1}^{N} w_n^* x_n(k) \tag{2.6}$$

where N is the number of sensors, w_n is the n^{th} complex weighting factor and x_n is the n^{th} input signal.

In order to increase the frequency range, more frequency dependent weights are needed, which are then multiplied with the respective incoming signal and summed to form the system output.

$$y(k) = \sum_{n=1}^{N} \sum_{p=0}^{P-1} w_{n,p}^* x_n(k-p) \tag{2.7}$$

where P is the delayed complex weight for each sensor.

At low frequencies, the main lobe tend to broaden up and the beam-like shape cannot be achieved. At high frequencies, the limiting mechanism is the spatial aliasing that depends on the distance separating the elements of the array. In fact, spatial aliasing first occurs at the frequency corresponding to the wavelength that equals to twice the distance between the array sensors.

Compared to a standard Delay-Sum beamformer, a superdirective beamformer achieves a higher directivity by means of numerical optimization methods in order to find the optimal amplitude and phase shifts, assigned to each sensor, for a prescribed directivity.

Constraints may be introduced to strongly reduce the gain for a certain angle of arrival. This is called null-steering and it is employed to attenuate the array sensitivity towards particular directions. It is a statistically optimized design pattern, as it relies on statistical properties to optimize the array response. Figure 2.7 shows the directivity obtained from a linear loudspeaker array that has a target angle at $-6°$ and null-steering applied at an angle of $6°$ with a null width of $9°$.

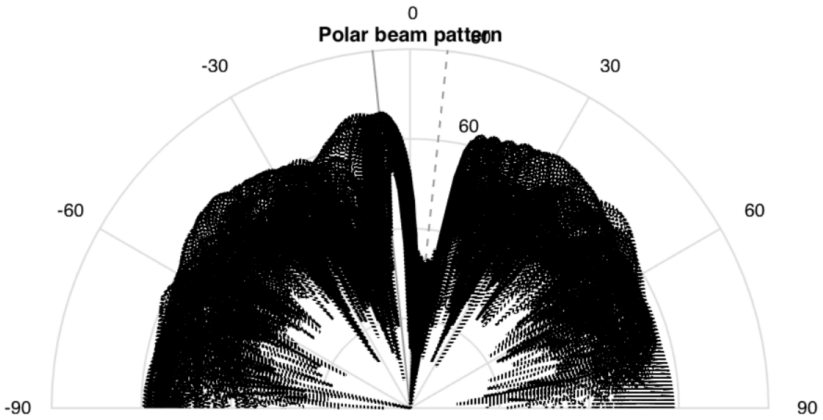


Figure 2.7: Polar beam-pattern of 8 speaker array with 14.4 cm spacing. Target angle is $-6°$ (solid line), stop angle is $6°$ (dashed line) with a null width of $9°$. Frequency range of optimization is [1 kHz - 9 kHz] with L = 1024 points [8]

3
Methods

This chapter deals with the description of the acoustic crosstalk cancelation system used in this project, the followed procedure for the loudspeaker directivity measurements as well as the HRIRs measurements in different playback environments, the presentation of the processing methodology of the gathered data and finally the description of the experiment for the perceptual evaluation.

3.1 CTC System Prototype

This master thesis work employed the linear equispaced 8-channel loudspeaker array presented in [9, 10], as depicted in Figure 3.1.

Figure 3.1: Example prototype using 8 *Neumann KH* 80 *DSP* loudspeakers with a spacing of 154 mm [19]

The core of the CTC system is a convex superdirective nearfield beamformer that directs a beam to one of the ears of the listener and produces a null at the opposite (contra-lateral) ear.
This acoustic crosstalk cancelation based on least-squares frequency-invariant beamforming (LSFIB) is limited at low frequencies as the beams tend to broaden up causing more crosstalk. A hybrid solution is proposed in [9], where the low and mid

frequencies in the range [250 Hz 1000 Hz] are rendered through a RACE processor using a pair of loudspeakers [6] while the high frequency components (above 10 kHz), are rendered as classic stereo through the two loudspeakers at the ends of the array to evoke natural shadowing due to the listener's head.

Under the assumption that the loudspeakers composing the array have radiation properties similar to the ones of an ideal point source (spherical waves), by means of a simulation the obtained transfer functions from one of the two input channels of the system to the ears of a listener, placed centrally at 1 m distance from the array, show that there is a channel separation, between the ipsi-lateral and the contra-lateral ears, of at least 20 dB over the vast part of the audible frequency range, as presented in Figure 3.2.

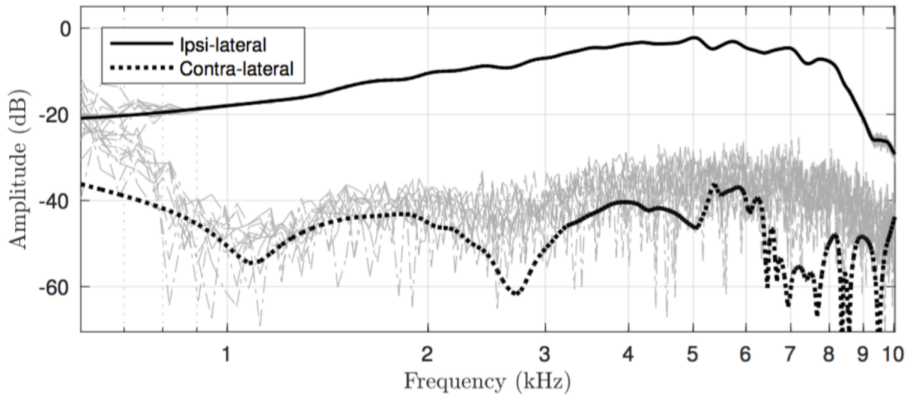


Figure 3.2: System transfer function to the two ears of the listener under ideal conditions (black lines) as well as with simulated loudspeaker mismatch (gray lines) [9]

Figure 3.2 also reveals that the channel separation drops significantly below 1 kHz as soon as some amount of mismatch of the sensitivity of the loudspeakers (0.3 dB) and uncertainties in the loudspeaker placement (1 mm) are included in the simulation. This reduction of the channel separation is expected since in the frequency range below, roughly, 1 kHz the two control points (the ears) are separated by less than a wavelength.

Taking into account that the spacing between the ideal sound sources (point sources) is 154 mm, the spatial aliasing occurs above 1.1 kHz. As presented in [8], even if grating lobes appear within the beamforming frequency range because of spatial aliasing, the space gap between consecutive beams remains large enough at 1 m distance from the center of the array as depicted in Figure 3.3.
Although the sound pressure level difference between the listener's ears slightly decreases as the frequency increases, the crosstalk cancelation stays almost completely unaffected by the spatial aliasing. The phenomenon should not be considered to be a limiting factor for the intended purpose.

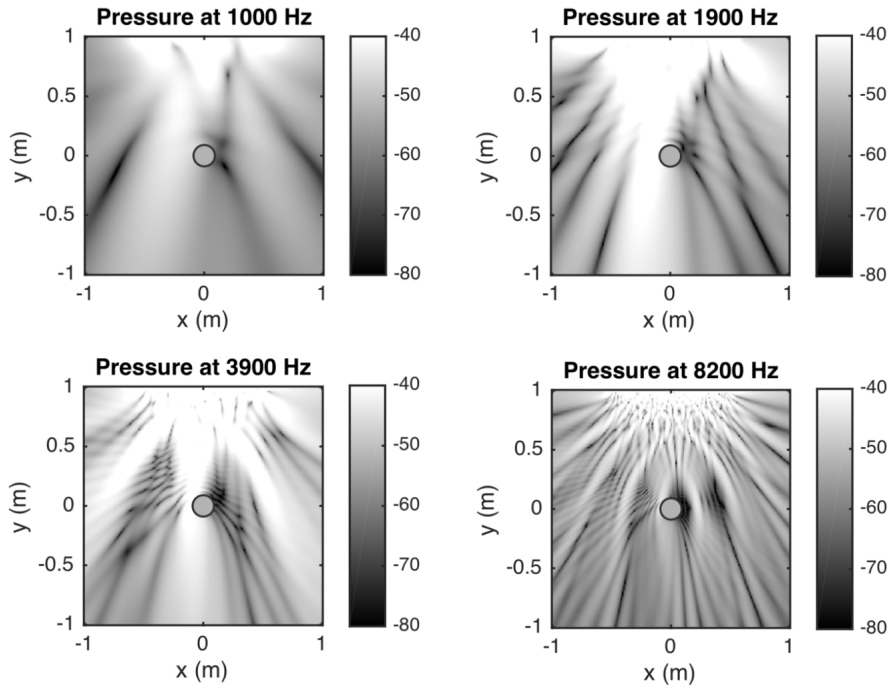


Figure 3.3: Sound pressure level at various frequencies over a [2 m× 2 m] area. The gray dot at [0, 0] represents the listeners head, modeled as an acoustically hard sphere with a diameter of 18 cm [8]

An array prototype was built, composed of 8 *Neumann KH* 80 *DSP* loudspeakers, as shown in Figure 3.1. The loudspeakers were linearly arranged with a constant distance of 154 mm between the center of their respective drivers. This employed loudspeaker model consists of two drivers of 100 mm and 25 mm diameter, respectively, in a vented box.

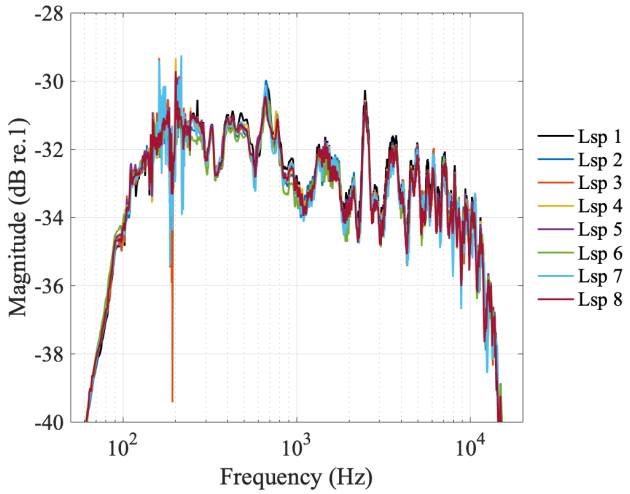


Figure 3.4: Transfer functions of the 8 loudspeakers composing the linear array

In order to evaluate the robustness of the array with respect to mismatches that can be found between its loudspeakers, the transfer function of each of the 8 loud-

speakers were compared, as depicted in Figure 3.4. The transfer functions follow the same general trend and the magnitude mismatches are in the range of 0.1 to 0.5 dB. On the plot, the sudden drop in the magnitude of "lsp 3" at a single frequency band ($\approx$ 190 Hz) can be assumed to be a measurement error and should be neglected.

The employed loudspeaker model has an integrated DSP and different ways to adjust the output and input gains. To verify the repeatability of the array performance regarding the available loudspeaker settings, the TF of two different loudspeakers composing the array were measured twice for each of 5 different combinations of output and input gain settings (the same gain is set again after being changed for a second measurement series). For the two tested loudspeakers, Figure 3.5 depicts how the transfer functions follow the same general trend at different settings.

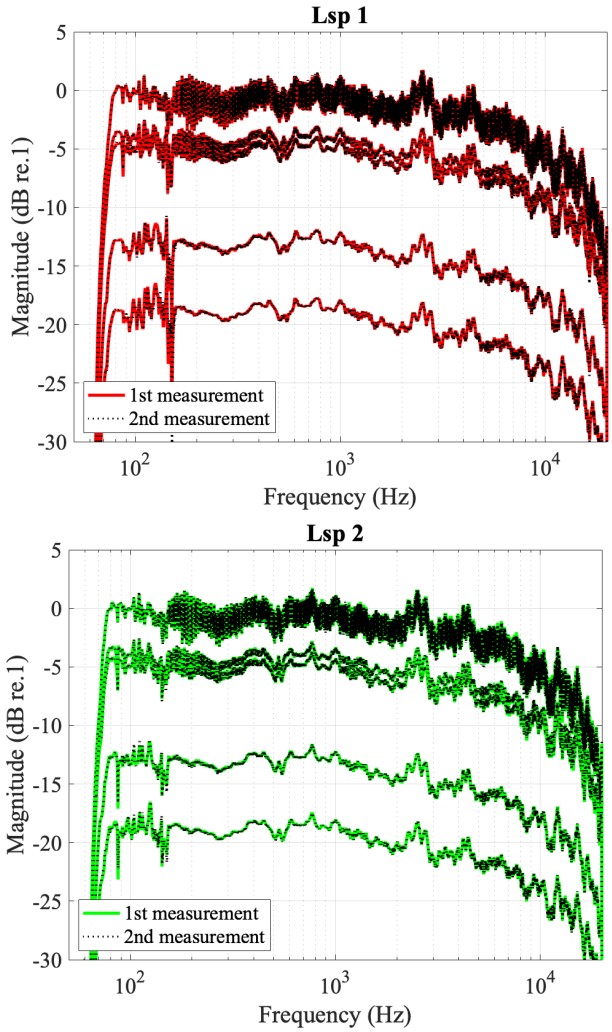


Figure 3.5: Transfer functions of two different loudspeakers measured twice at different available input and output gain settings

A more convenient way to illustrate the stability of this loudspeaker model, while set at different input and output gain settings, is depicted in Figure 3.6. The gathered data of each setting were arbitrarily normalized with respect to the magnitude at

1 kHz. As the curves overlay each other, it is clearly shown that for all available settings the loudspeakers exhibit almost the same transfer functions and the maximum deviation from the mean is found to be around 1.5 dB.

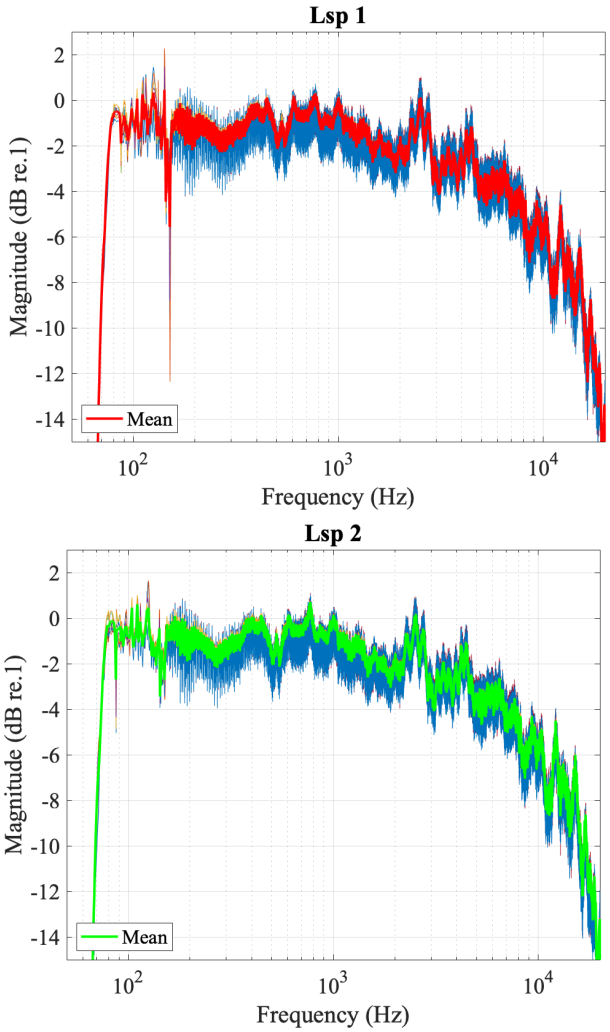


Figure 3.6: Normalized transfer functions of two different loudspeakers measured twice at different input and output gain settings

The loudspeakers employed in the array prototype turned out to be practical as they exhibit various reproducible gain settings that are digitally matched by the manufacturer to avoid mismatch. This behaviour match is a guarantee of reproducible measurements with the array.

3.2 Measurement Procedure

All the measurements took place in the facilities of the Division of Applied Acoustics, at Chalmers.
A Python script was used for the acquisition of the measurements, while the data processing was performed on Matlab. The excitation signal consisted of a 3 seconds

long logarithmic sine sweep. All automatic rotations were performed using the scanning array system VariSphear [20].

The measurements were conducted with the following equipment:

- Antelope Orion 32 audio interface (SN: 1000218340117)
- RME QuadMic preamplifier
- B&K type 1708 signal conditioner (SN: 100102)
- B&K type 4190 free-field 1/2" microphone (SN: 2455390)
- G.R.A.S. KEMAR type 45BB Head and Torso (SN: 250201)
- G.R.A.S. KEMAR type KB0091 large left ear (SN: 225384)
- G.R.A.S. KEMAR type KB0090 large right ear (SN: 231543)
- G.R.A.S. type 12AL CCP supply (SN: 272718 ; 279640)
- VariSphear scanning array system
- Neumann KH 80 DSP loudspeakers (SN: 506834-3297192147 ; 506834-3248339459 ; 506834-3297192124 ; 506834-3297189689 ; 506834-3297192146 ; 506834-3297191080 ; 506834-3297189744 ; 506834-3297189787)

3.2.1 Loudspeaker Directivity Measurement

As stated in the previous chapter, the directivity of a loudspeaker can be highly affected by its mounting type, as well as by the diffraction of the emitted sound over the edges of the loudspeaker box. Therefore, the directivities were measured while each loudspeaker was mounted within the array.
Taking into consideration the main interest of this project, the directivities were not measured along 360° around the loudspeakers but only along a semicircle around the center of the loudspeaker array, as illustrated in Figure 3.7. The on-axis measurements were conducted under anechoic conditions, along an arc of 1.00 m radius with its center at the center of the array and a spacing of 1°.

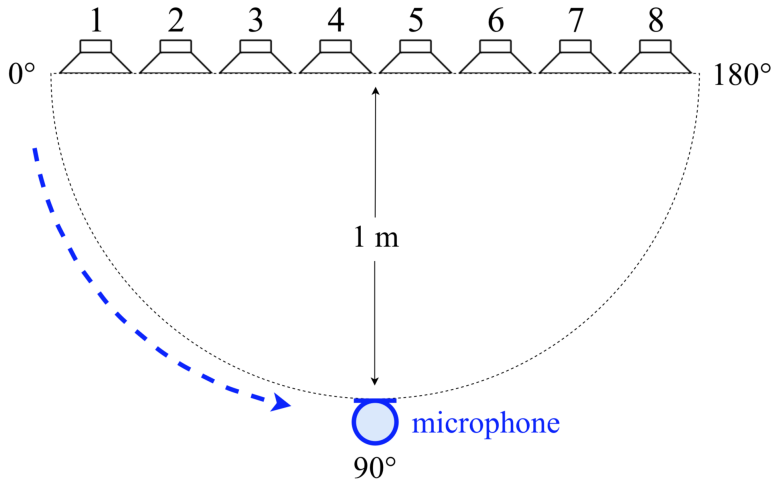


Figure 3.7: Sketch of the loudspeaker directivity measurement setup

Figure 3.8: Photograph of the loudspeaker directivity measurement setup in the anechoic chamber

3.2.2 HRIR Measurement

To evaluate the perceptual influence of different playback rooms on crosstalk canceled binaural content, one can auralize the array response in different environments, through a set of headphones.
IRs from each of the loudspeakers of the array to a KEMAR dummy head (DH) were measured in 6 different environments and for different head orientations in steps of 1°. During the measurements, the KEMAR DH was 1 m away from the center of the array.

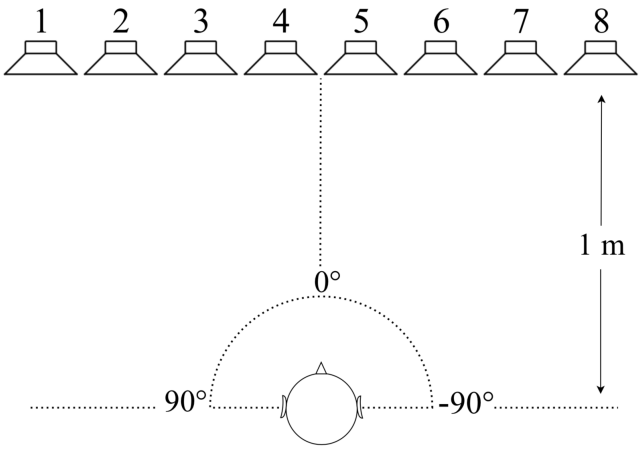


Figure 3.9: Sketch of the HRIRs measurement setup

The measurements took place in the following environments:

1. Anechoic chamber
2. Anechoic chamber with a floor reflection (DH ears at 1.15 m from a hard floor)
3. Small dry laboratory (DH ears at 1.38 m from a hard floor)
4. Small dry laboratory room with a reflective side wall and the DH located in the center of the room (DH ears at 1.10 m from a carpeted floor)

5. Small dry laboratory room with one reflective side and rear wall and the DH located in the center of the room
6. Small dry laboratory room with a reflective rear wall and with the DH located 1 m away from a highly reflective lateral wall (to the right of the DH) side wall and the DH located in the center of the room

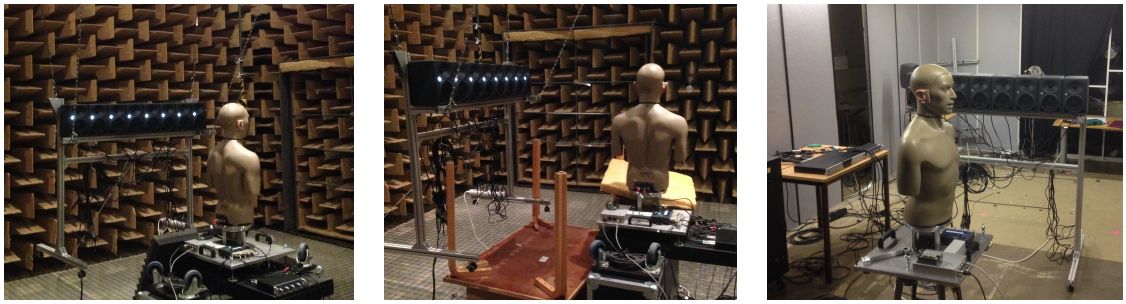


Figure 3.10: Photographs of HRIRs measurement setups: environment 1 (left), environment 2 (center), environment 3 (right)

Figure 3.11: Photographs of HRIRs measurement setups: environment 4 (top left), environment 5 (top right), environment 6 (bottom)

3.3 Data Processing

First, the processing of the collected data consisted in applying deconvolution in order to get the IR of each loudspeaker of the array, as perceived by each DH ear, at every head rotation step angle, in each of the 6 environments. Figure 3.12 presents the followed steps in the data processing to obtain the impulse responses.

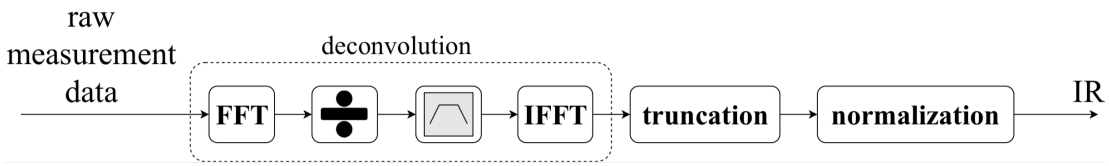


Figure 3.12: Block diagram of the deconvolution process

To focus only on the the frequency range of interest and attenuate the influence of the surrounding noise, a window in frequency domain was applied. A 4096 samples long band-pass filter [70 Hz - 20 kHz] was designed using the Filter Designer App in Matlab. Since the processed data were to be employed for auralization, the choice of a FIR filter was made in order to guarantee a linear phase.

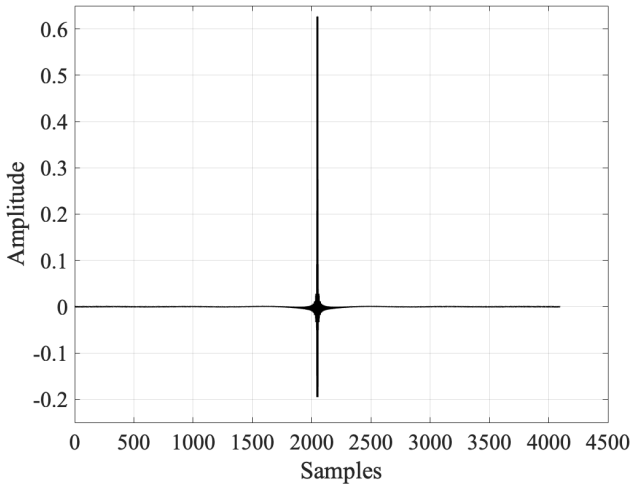


Figure 3.13: IR of the designed band-pass filter

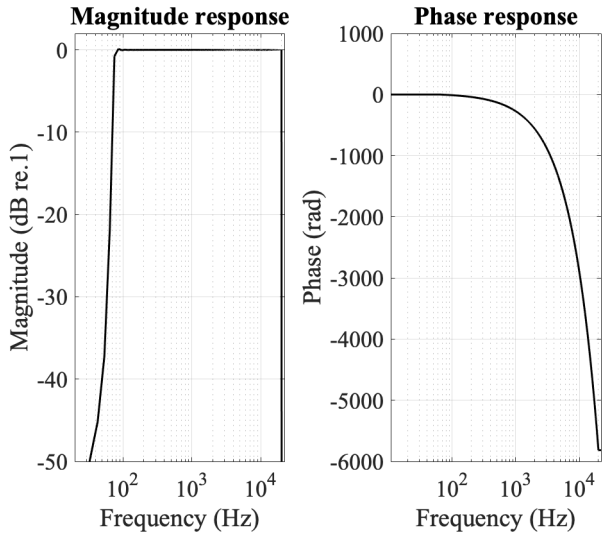


Figure 3.14: Frequency response of the designed band-pass filter

Note that for the purpose for auralization, an IIR filter would have worked correctly too, as long as the phase distortion in the left and right ear signals is the same.

Back in the time domain, the signals were also truncated to 1024 samples and normalized with respect to the maximum amplitude recorded in each measurement series.

Expecting to only get a delay corresponding to the sound propagation from the center of the linear array to the KEMAR DH (1 m distance), an additional delay of 2 ms, corresponding to 88 samples (f_s = 44.1 kHz), was clearly noticed on all the measurements. It was assumed that this additional delay was due to the integrated DSP of the loudspeaker model. As depicted in Figure 3.15, a backward shift of 88 samples was applied to the signals to compensate for that extra delay.

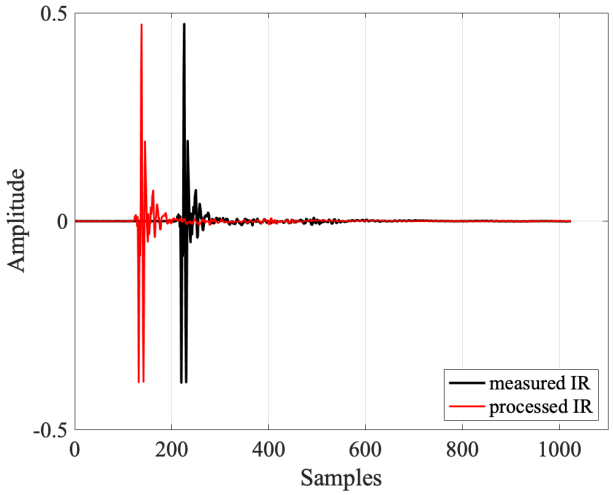


Figure 3.15: Backward shift of 88 samples to compensate for the delay introduced by the loudspeaker DSP

To include the actual loudspeakers directivities in the simulation, the part of the CTC algorithm describing each source radiation property as given by Equation 2.2 (ideal point source) was replaced by a matrix containing the obtained data from the directivity measurement, for each loudspeaker at 1 m in front of the center of the array (the 90° microphone position in Figure 3.7).

Considering the auralization of the CTC system based on the HRIRs measurements, the process was as follows:
16 optimized IRs were obtained from the beamformer algorithm (for each ear, 8 IRs corresponding to each loudspeaker of the array). Each of these loudspeaker impulse responses was convolved with the recorded HRIRs. Because of linearity, the principle of superposition of pressure to determine total sound pressure at the observation point can be applied, thus the simple sum of those described IRs over all the loudspeakers leads to the impulse response of the entire array in the playback environment.

The impulse response of the linear 8-channel loudspeaker array is given by:

$$Array\ IR_i = \sum_{n=1}^{N} IR_{i,\theta} \tag{3.1}$$

where IR is the impulse response obtained by the convolution of the beamformer IRs with the HRIRs, i is the ear (left or right), N is the number of loudspeakers in the array and θ is the angle of the head orientation.

The obtained left and right IRs of the CTC system can then be used to auralize it through a set of headphones. Figure 3.16 depicts the process followed for the case of a single head orientation, in a given playback environment.

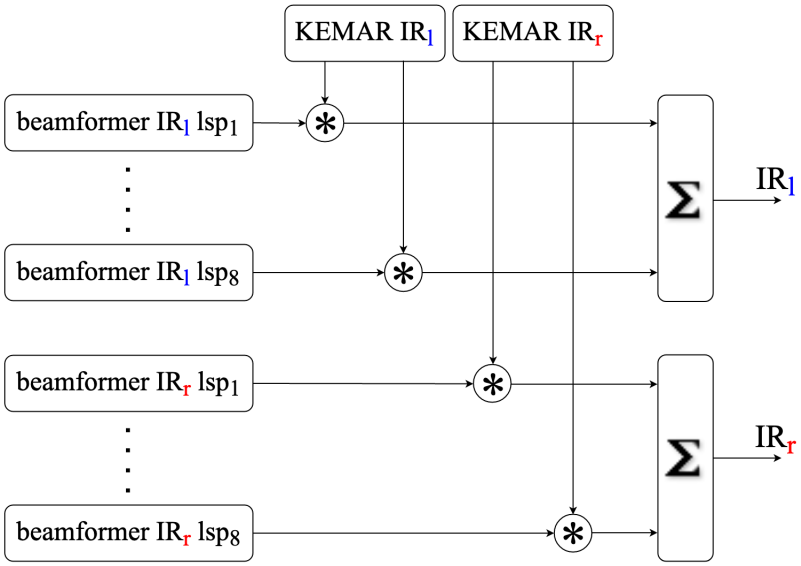


Figure 3.16: Block diagram of the data processing for auralization

This entire process was performed for each measured head rotation angle and in each of the 6 playback environments.

3.4 Perceptual Study: Experiment Design

As presented above, the HRIR measurements for different head orientations in steps of 1° allow for performing binaural head-tracked auralization of the virtual array through headphones as well as switching between different environments by the click of a button.

Following the process described in the previous section, the array was auralized in all 6 environments by means of head-tracked binaural synthesis of the KEMAR dummy head measurement data using the *SoundScape Renderer* (SSR) [21, 22] to which the signals were routed via *Jackaudio*. Finally, the audio signals were played through a pair of *Sennheiser HD* − 650 headphones.
The subjects, were instructed not to move their head excessively as the CTC method

does not account for this. Head-tracking was nevertheless employed so that the auralization accounts for small head movements, which may reduce distortion of the spatial perception [23].

The virtual loudspeaker array played a binaural audio content, which consisted of anechoic male speech spatialized, by means of KEMAR HRTFs, in 3 different directions: straight ahead (0°), 30° and 90° to the left. The listener was positioned symmetrically with respect to the array and at 1 m distance from it.

The subjects were provided with a graphical interface with which they were able to switch seamlessly between the conditions while the speech signal was playing continuously. A total of 18 different conditions (3 virtual source positions in 6 environments) were presented to the test subjects. Figure 3.17 presents the graphical interface used by the subjects in the perceptual evaluation.

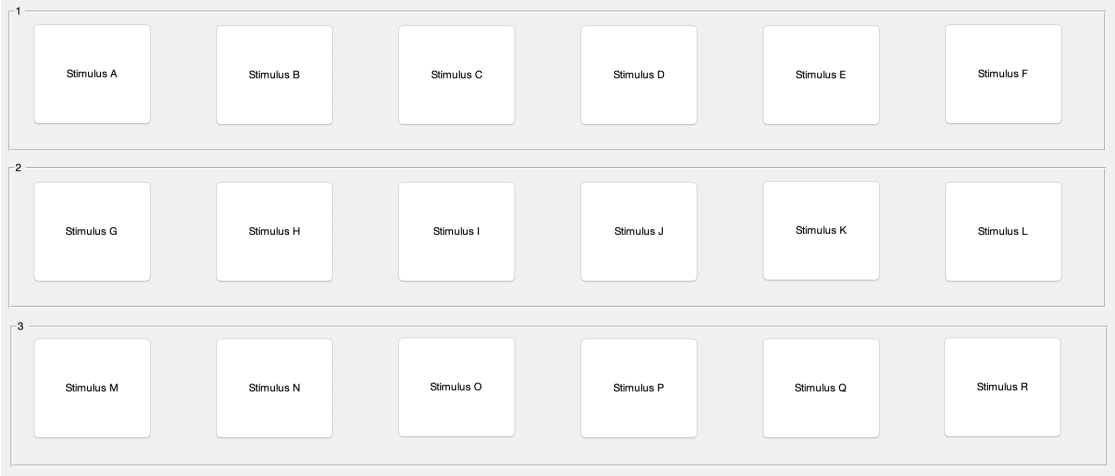


Figure 3.17: The Matlab graphical interface employed for the perceptual study

In a pilot study conducted by the experimenters, it was found that the perceptual differences between the auralizations of the different room conditions can span a broad range from hardly or not perceptible to clear multidimensional differences. Thus, it was chosen to run the study as an interview.
The subjects were comparing the 6 different conditions for each of the 3 virtual source positions separately and reported in free speech to the experimenter what differences they were hearing. First, for each virtual source position the subjects were asked to localize the source. Then, for a given source position, while changing the playback environments the subjects were asked to localize the source once again, as well as to describe their perceptions in terms of similarity between the audio signals, externalization of the sound, locatedness, source width and plausibility or realism of what they were listening to. The test subjects did not have further information on what it was that they were listening to.

9 grown-up subjects with self-reported normal hearing participated in the evaluation. In average, every test/interview lasted 23 minutes.

4

Results and Discussion

In this chapter the obtained results are presented and discussed. First, some adjusting of the mismatch introduced in the simulation, next, the incorporation of the obtained data from the HRIRs measurement in the anechoic chamber and the measured loudspeaker directivities and their impact on the simulated crosstalk cancelation. Then, the data collected from all 6 playback environments are presented and the results of the perceptual evaluation are given. Finally minor experimental errors are discussed and ideas for further research are suggested.

4.1 Simulation Mismatch Adjusting

In the previous chapter, it was shown that introducing a certain amount of mismatch in either the sensitivity or the placement of the loudspeakers can affect the channel separation. Therefore, in the simulation different trials were done with various mismatch combinations. Taking into consideration the measured loudspeaker directivities, Figures 4.1 and 4.2 show the obtained results over the frequency range where beamforming is applied [1 kHz - 9 kHz]. It was clearly noticed that among these two parameters, the mismatch in loudspeaker placement was the most prominent one on the channel separation.

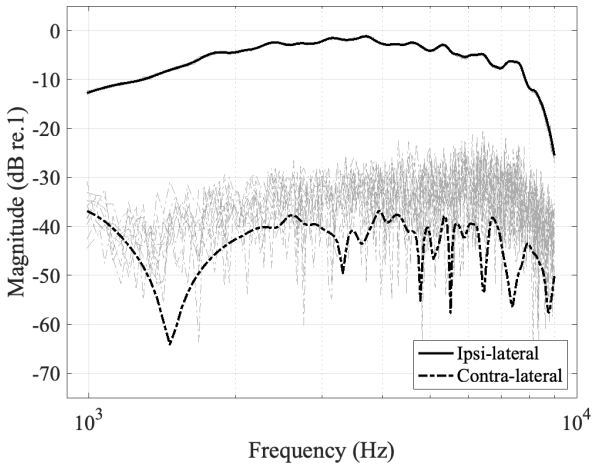


Figure 4.1: System transfer function to the two ears of the listener under ideal conditions (black lines) as well as with simulated loudspeaker mismatch (gray lines) ; Gain mismatch = 0.3 dB ; Placement mismatch = 1 mm

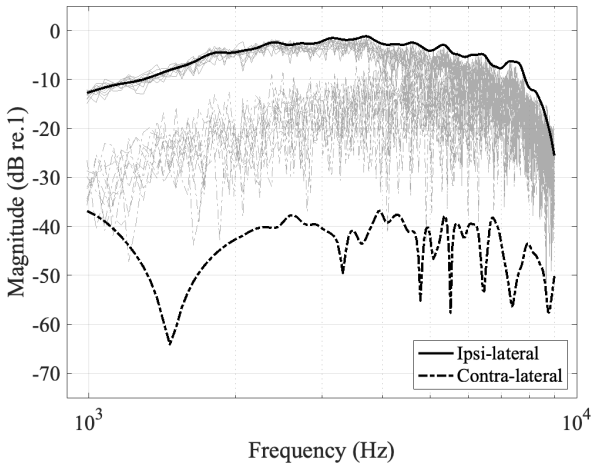


Figure 4.2: System transfer function to the two ears of the listener under ideal conditions (black lines) as well as with simulated loudspeaker mismatch (gray lines) ; Gain mismatch = 0.3 dB ; Placement mismatch = 10 mm

A more realistic mismatch is presented in Figure 4.3, including 1 dB gain and a 5 mm placement mismatches. Note that even with these mismatches taken into account in the simulation, a channel separation greater than 15 dB is still maintained especially in the lower part of the beamformer frequency range (below 4 kHz).

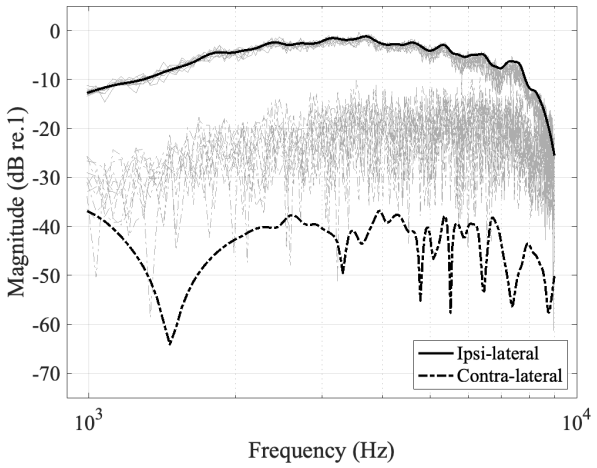


Figure 4.3: System transfer function to the two ears of the listener under ideal conditions (black lines) as well as with simulated loudspeaker mismatch (gray lines) ; Gain mismatch = 1 dB ; Placement mismatch = 5 mm

4.2 Incorporation of Anechoic HRIRs and Loudspeaker Directivities

The original simulation assumed scattering over a spherical shape placed at 1 m from the array and compared the sound pressure at 2 receiving points on each side

of the sphere (representing the ears). In order to make the simulation more realistic, this assumption can be modified by including the KEMAR HRTFs measured under anechoic conditions (environment 1).

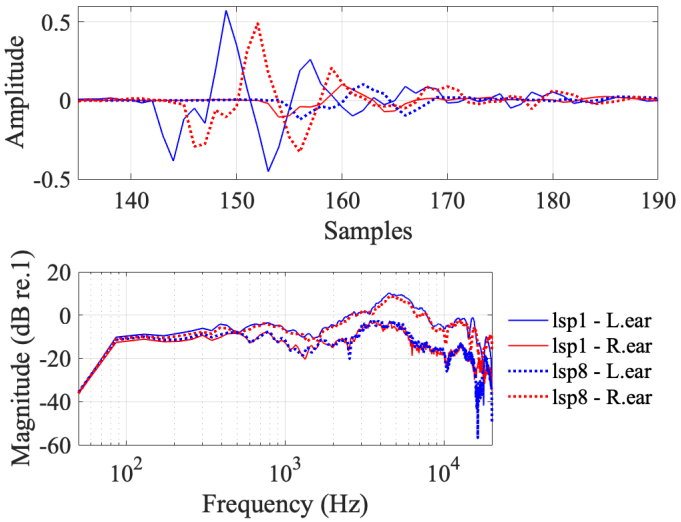


Figure 4.4: KEMAR HRIRs and HRTFs for loudspeakers 1 and 8

Figure 4.4 shows the HRIRs and HRTFs when running each of loudspeakers 1 and 8 while the dummy head faces the array. Accounting for an ipsi-lateral contra-lateral inversion in this case, one would expect to observe a perfect superposition of the HRTFs. The slight spectral deviation is explained by the 3 samples deviation of the measured IRs. Considering the used sampling frequency of 44.1 kHz, one can conclude that the dummy head was placed approximately 23 mm off center, towards loudspeaker 1.

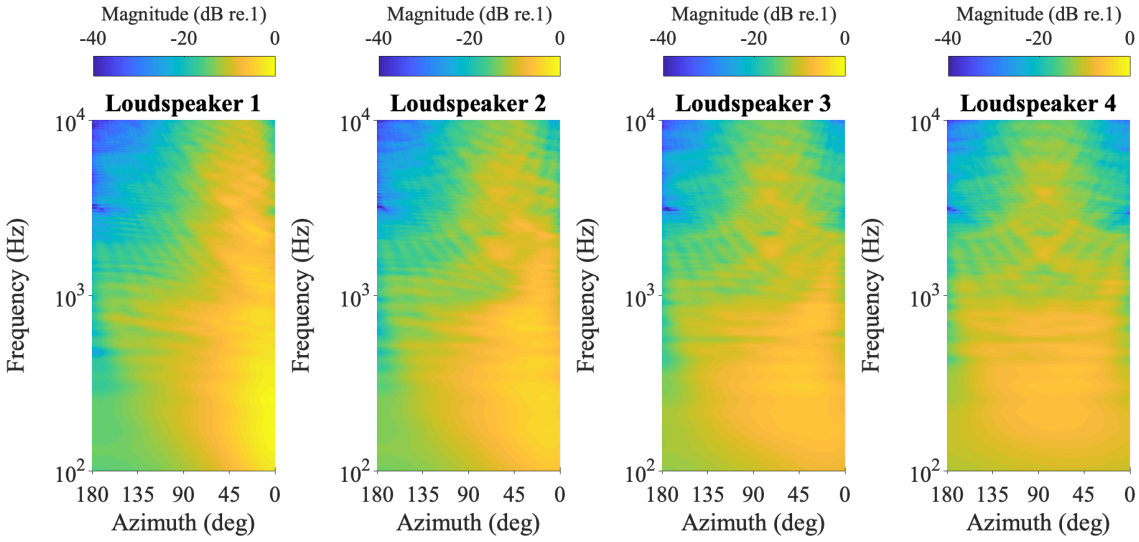


Figure 4.5: Measured directivities of the first half of the array, from loudspeaker 1 to 4; the abscissa specifies the azimuth of the measurement locations along a semicircle around the center of the loudspeaker array

The on-axis directivity of each loudspeaker of the array was measured over 180° with a step of 1°. Figure 4.5 presents the directivities of the first half of the array, from loudspeaker 1 to 4. It is seen that the actual directivities depart significantly from the assumed spherical wave.

By computing the beamformer weights based on the measured directivities instead of on the point source model, as well as by including the KEMAR HRTFs in the simulation, the channel separation between the illuminated ear of the listener and the opposite ear is depicted in Figure 4.6.

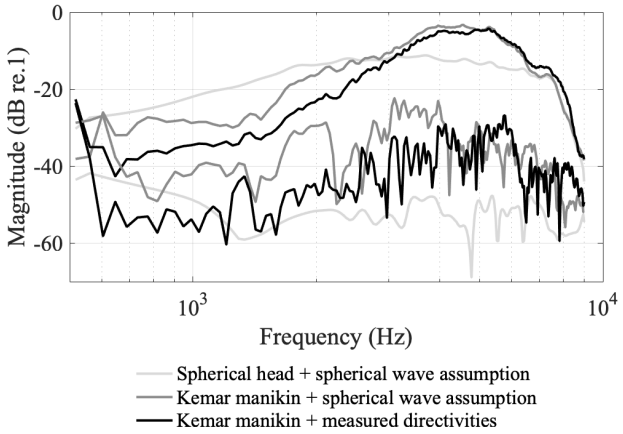


Figure 4.6: Channel separation between ipsi-lateral and contra-lateral ear

The absolute difference of pressure level between ipsi-lateral and contra-lateral ear is depicted in Figure 4.7.

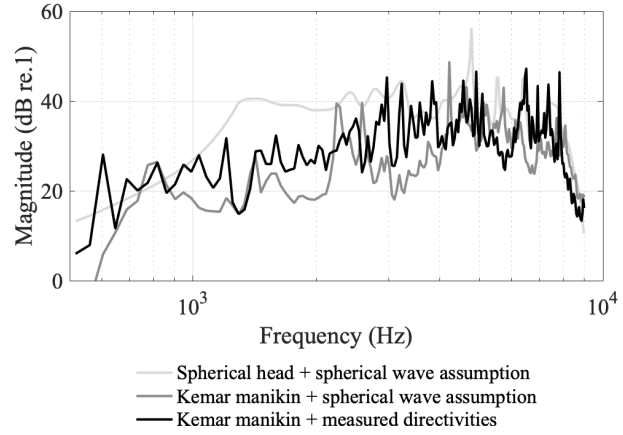


Figure 4.7: Absolute difference between ipsi-lateral and contra-lateral ear

Over the frequency range [1 kHz - 9 kHz] on which beamforming is operated , the general channel separation went from 36 dB in the initial simulation to 22 dB

after considering a more realistic approach with the KEMAR HRTFs, to finally 25 dB after including the actual loudspeaker directivities. Besides the 3 dB general increase, the improvements are mainly observed below 700 Hz and in the frequency ranges [1 kHz - 2 kHz] and [3 kHz - 4 kHz]. Note that the frequency ranges in which the improvements occur are where the channel separation, without considering the directivities, was at its lowest.

4.3 HRIRs in Tested Playback Environments

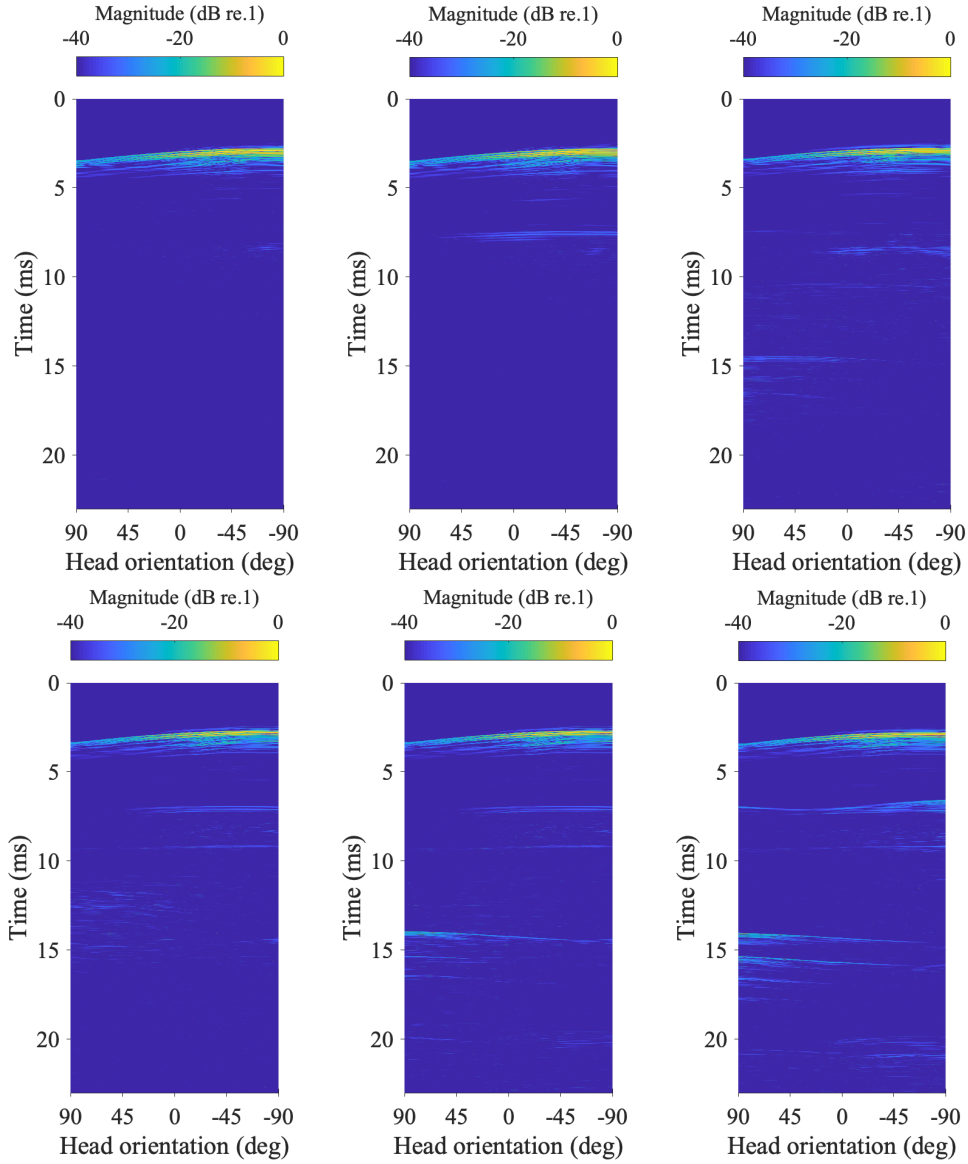


Figure 4.8: Impulse responses from loudspeaker 4 to the left ear of the KEMAR dummy head: environment 1 (top left), environment 2 (top center), environment 3 (top right), environment 4 (bottom left), environment 5 (bottom center), environment 6 (bottom right)

Figure 4.8 depicts the impulse responses measured in the 6 different playback environments from loudspeaker 4 to the left ear of the dummy head. The energy brought by the multiple strong early reflections can be clearly observed. The first reflected energy to reach the dummy head ears corresponds to the floor reflection. For the case of loudspeaker 4, the first reflection arrives 4.6 ms after the energy coming directly from the sound source.

4.4 Perceptual Evaluation

From the perceptual evaluation conducted on 9 subjects, the following responses were extracted:

- The floor reflection has no audible influence (environment 1 vs. 2).
- The effect of the dry room (environment 3) is minor. The perception is very similar to environments 1 and 2.
- The lateral virtual source positions were perceived more spacious (30° and 90°).
- An increasing amount of reverberation increases externalization. Internalization can occur in environments 1 and 2.
- An increasing amount of reverberation makes spatial perception more plausible in general, for example, in terms of localization accuracy, locatedness, and source width.
- Front-back confusions occurred mostly in environments 1 and 2.
- Environments 4 and 5 evoked the most plausible and pleasant perception.
- Approximately half of the subjects perceived the virtual source as slightly elevated, but reverberation mitigated this effect.
- The strong lateral reflection in environment 6 was perceived disturbing by most of the subjects.
- No major effect of the strong lateral reflection in environment 6 on localization was reported.

These obtained results suggest that there is a range of room acoustic conditions that influence the presentation of binaural audio content only to a minor extent (environments 1-5). The results from [13], even if obtained based on a simple room simulation, can then be confirmed.

As the reflected sound energy came shortly after the direct sound, the test subjects could experience the precedence effect. In fact, the reflections coming from the floor or from nearby vertical surfaces did not add ambiguity to the localization of the virtual source.

In Sæbø's doctoral dissertation [12], it was found that strong and isolated lateral reflections caused serious localization impairments. However, this observation is not made in the results of the perceptual study of this thesis work. In this study, environment 6 comprised a strong lateral reflection (DH placed 1 m away from a hard reflective side wall) that was embedded in the room reverberation. Even if this reflection had a negative effect on the pleasantness as some participants compared

the tested environment to a small volume and described the soundscape as "boxy", the effect was rather on a general level than primarily on localization.
Globally, it seems that embedding a reflection in the reverberation reduces the perceptual deterioration.

During the experiment, the virtual orientation of the listener was normal towards the array even though it was shown in [10] that making the listener look away from the array in normal direction can significantly reduce front-back confusions. The phenomenon reported in [10] was observed with 4 prototypes that were all employing comparable parameters, however the phenomenon has not been decoded yet. The study suggests that neither listener awareness nor head rotations were a likely cause in the recurrence of front-back confusions. The authors suggest that the difference between the two listener orientations (facing towards/away the array) is the slightly different filtering introduced by the pinna (outer ear).

To verify this suggestion in this thesis work, the analysis of the HRTF imposed by the beamformer was made with the KEMAR dummy head facing towards and away the array. This was done using the HRTFs measured along 360° in environment 4, as this latter evoked to the test subjects the most plausible and pleasant perception. As shown in Figures 4.9 and 4.10, the two curves follow the relatively similar spectral trend until approximately 8 kHz where the filtering of the outer ear attenuates the signal by at least 15 dB. That difference might affect the observed amount of front-back confusions in [10].

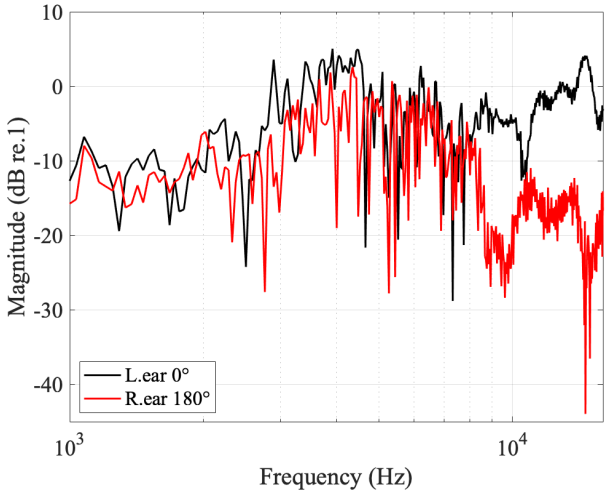


Figure 4.9: Imposed HRTF difference at head orientation angles of 0° and 180°

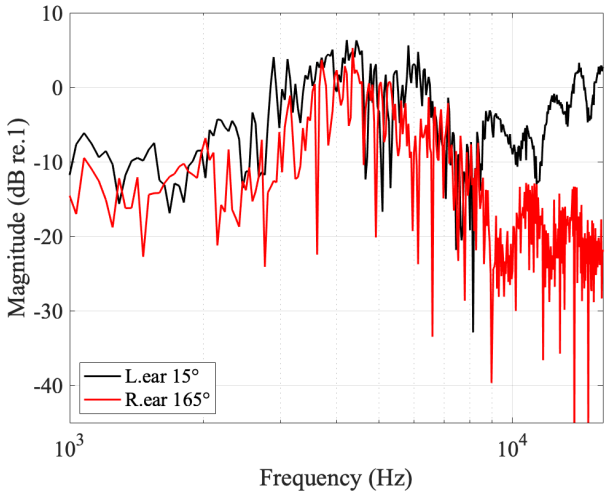


Figure 4.10: Imposed HRTF difference at head orientation angles of 15° and 165°

4.5 Experimental Conditions

It is important to specify that the test subjects did not have any bias regarding the experiment. As the CTC system was simulated in different playback environments through auralization using a set of headphones, the test subjects did not have any visual cues. They neither saw the loudspeaker array nor the playback environments. Prior to the experiment, the participants were only informed that the evaluation was on the topic of 3D audio reproduction and they were not told anything about acoustic crosstalk cancelation.

Even though it is believed that the perceptual evaluation method was robust against errors and imperfections, it is worthy to note some minor experimental conditions that could have affected the evaluation:
- The HRTFs used in the auralization were not individualized ones (the listeners own), so depending on the test subject the general plausibility and even the localization accuracy could have been affected.
- As stated in Chapter 2, in a case where the position of a virtual source and the reflecting waves of the real source correspond with the placement of the loudspeakers, the resulting auditory event can differ because there are multiple sources (the loudspeakers of the array) that are radiating signals and causing multiple reflections instead of one (real source). In addition to that, the position of the sound reproduction setup does not change in the playback room, meaning that the time delay of the reflections is always the same and different from the real one, regardless of the virtual source position.

4.6 Further Research

The perceptual evaluation took into consideration the incorporation of the actual loudspeaker directivity into the beamformer. Besides the general 3 dB improvement

observed in the channel separation, it would be interesting to run an evaluation to check if that has a significant impact on the perception.

In order to better understand the front-back confusion reported in [10], a perceptual evaluation could be conducted. The test could employ frequency equalization so that the magnitude of the signal could be attenuated in the range above 8 kHz while the listener is facing towards the array. This way the signals that arise at the listener's ears are similar to those that arise when the listener is facing away from the array. This might confirm the claim that the slightly different filtering of the pinna affects the observed amount of front-back confusions in [10].

5
Conclusion

The project initial objective was to study the effect of incorporating the actual loudspeaker radiation properties in the beamformer design of an array-based acoustic crosstalk cancelation system. On-axis loudspeaker directivity measurements were conducted and the data were integrated into the simulation. The improvement compared to the original beamformer employing a point-source model is limited and is approximated to 3 dB over the frequency range where beamforming is applied.
It is suggested to perceptually evaluate to significance of this improvement in a future work.

The project also contained a second phase. It consisted in the evaluation of the perceptual effect of reflective surfaces on that same acoustic crosstalk cancelation system. An auralization of the CTC array in 6 different playback environments was processed by means of head-tracked binaural synthesis based on manikin measured data. It was found that a single floor reflection does not have a noticeable influence and is often confused with the anechoic case. However, as reverberation from the playback environment increases, the general perception is more pleasant, the impression of space is expanded and feels more real, and even a strong lateral reflection does not weaken the localization cues in a significant way.
It was also noticed that an increasing amount of reverberation reduces the front-back confusion. This phenomenon should be explored in further research besides the future work on the recurrence of the front-back confusion depending on the listener's orientation towards the array.

Bibliography

[1] E. Choueiri, "Optimal Crosstalk Cancellation for Binaural Audio with Two Loudspeakers," presented at the Princeton University (2010)

[2] B.B. Bauer, "Stereophonic Earphones and Binaural Loudspeakers," J. Audio Eng. Soc, vol. 9, no. 2, pp. 148–151 (1961)

[3] O. Kirkeby, P.A. Nelson, "Digital Filter Design for Inversion Problems in Sound Reproduction," J. Audio Eng. Soc, vol. 47, no. 7/8, pp. 583–595 (1999)

[4] W.G. Gardner, "3-D audio using loudspeakers," Springer Science & Business Media (1998)

[5] D.B. Ward, G.W. Elko, "Effect of loudspeaker position on the robustness of acoustic crosstalk cancella- tion," IEEE Signal Process. Lett., vol. 6, no. 5, pp. 106–108 (1999)

[6] R. Glasgal, "360 Localization via 4.x RACE Processing," presented at the Audio Engineering Society Convention 123 (2007 October)

[7] J. Bauck, D.H. Cooper, "Generalized Transaural Stereo and Applications," in JAES 44(9) pp. 683-705 (1996)

[8] C. Hohnerlein, "Beamforming-based Acoustic Crosstalk Cancelation for Spatial Audio Presentation," MSc Thesis, Technical University of Berlin (2016 May)

[9] C. Hohnerlein, J. Ahrens, "Perceptual evaluation of a multiband acoustic crosstalk canceler using a linear loudspeaker array," presented at the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 96–100 (2017 March)

[10] X. Ma, C. Hohnerlein, J. Ahrens, "Concept and Perceptual Validation of Listener-Position Adaptive Superdirective Crosstalk Cancelation Using a Linear Loudspeaker Array," in JAES (2019)

[11] D.B. Ward, "On the performance of acoustic crosstalk cancellation in a reverberant environment," in JASA 110(2) (2001 August)

[12] A. Sæbø, "Influence of Reflections on Crosstalk Cancelled Playback of Binaural Sound," Doctoral Dissertation, NTNU Trondheim (2001)

[13] D. Kosmidis, Y. Lacouture-Parodi, and E.A.P. Habets, "The Influence of Low Order Reflections on the Interaural Time Differences in Crosstalk Cancellation Systems," in IEEE ICASSP, Florence, Italy (2014 May)

[14] F.A. Everest, K.C. Pohlmann, "Master Handbook of Acoustics," 5th edition, McGraw-Hill (2009)

[15] D.N. Zotkin, R. Duraiswami, E. Grassi and N. Gumerov, "Fast head-related transfer function measurement via reciprocity," Journal of Acoustical Society of America (2006)

[16] P. Sjösten, "Lecture notes in Electro-Acoustics," Chalmers University of Technology (2015)

[17] M. Kleiner, "Acoustics and Audio Technology," third edition, J. Ross Publishing (2012)

[18] W.G. Gardner "3-D Audio Using Loudspeakers," Ph.D. thesis, Dept. of Media Arts and Sciences, MIT. pp 52-118 (1997)

[19] J. Ahrens, "The Effect of Loudspeaker Radiation Properties on Acoustic Crosstalk Cancelation Using a Linear Loudspeaker Array," in Proc. of DAGA, Rostock, Germany (2019 March)

[20] B. Bernschutz, C. Porschmann, S. Spors, and S. Weinzierl, "Entwurf und Aufbau eines variablen spharischen Mikrofonarrays fur Forschungsanwendungen in Raumakustik und Virtual Audio," in Proc. of DAGA, Berlin, Germany (2010 March)

[21] M. Geier, S. Spors and J. Ahrens, "The SoundScape Renderer: A unified spatial audio reproduction framework for arbitrary rendering methods," in 124th Convention of the AES, p. 7330 (2008)

[22] M. Geier and S. Spors, "Conducting psychoacoustic experiments with the SoundScape Renderer," in 9. ITG Fachtagung Sprachkommunikation, Bochum, Germany, pp. 1–4 (2010)

[23] D.R. Begault, E.M. Wenzel and M.R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," in JAES 49(10), p. 904-16 (2001 October)