CHALMERS
UNIVERSITY OF TECHNOLOGY

The B-Format – Recording, Auralization, and Absorption Measurements

Master's thesis in the program Sound and Vibration

KILIAN SCHULZE-FORSTER

Division of Applied Acoustics
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2018

Master's thesis ACEX30-18-64

The B-Format – Recording, Auralization, and Absorption Measurements

KILIAN SCHULZE-FORSTER

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Architecture and Civil Engineering
Division of Applied Acoustics
Audio Technology Group
Chalmers University of Technology
Gothenburg, Sweden 2018

The B-Format – Recording, Auralization, and Absorption Measurements

KILIAN SCHULZE-FORSTER

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

Master's Thesis ACEX30-18-64
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: Photo of a B-format microphone (Sennheiser AMBEO VR).

Typeset in LaTeX

Gothenburg, Sweden 2018

The B-Format – Recording, Auralization, and Absorption Measurements

KILIAN SCHULZE-FORSTER
Division of Applied Acoustics
Chalmers University of Technology

Abstract

B-format microphones are composed of four nearly coincident cardioid capsules arranged at the corners of a tetrahedron. They are capable of capturing the pressure signal as well as the particle velocity of the impinging sound field in all three Cartesian dimensions. This information can be used to auralize the captured sound field via headphones or loudspeaker arrays or to retrieve some of its physical properties. Commercial renderers for B-format sound files in 360° videos such as Facebook360 and Youtube are available but they do not disclose details about their methods. Consequently, they are inapplicable for research in auralization of 3D-sound. Moreover, B-format recordings are mainly used for entertainment by the Ambisonics community, while little attention has been paid to their potential in room acoustical analysis, especially for in-situ absorption measurements.

This thesis deals with three different issues regarding B-format microphones, namely the computation of the so-called B-format signals, the auralization of B-format recordings and the applicability of B-format microphones for in-situ absorption measurements. In all three areas the basis for further research is provided.

At first, filters to improve the B-format signals are compared by simulations and measurements. Then, a real-time binaural renderer including head tracking is built, whose methods are completely transparent. Finally, a new absorption measurement method, which extracts the sound field impedance from a B-format recording, is proposed and compared to the standard impedance tube method.

Results indicate that simple filters designed based on measurements can improve the B-format signals in the same way as theoretical filters proposed in literature, which are less practical. The binaural renderer works as basis for research related to auralization but needs further improvement and to be tested against its commercial counterparts. Finally, it was confirmed that the measurement of absorption with a B-format microphone works and suggestions to advance this method are derived.

Keywords: B-format, Ambisonics, auralization, binaural renderer, digital signal processing, room acoustics, absorption measurement

Acknowledgements

First of all, I would like to thank Jens Ahrens for his exceptionally dedicated supervision of my work on this thesis. He always took the time to patiently discuss some issues or the next steps with me and I got valuable advice whenever I had questions.
I am also very thankful for the support of Börje Wijk, who helped me with the impedance tube absorption measurement and everything related to the electrical equipment I needed for my work.
Carl Andersson had many pieces of advice for me regarding different practical issues with the measurements, thank you for that.
Finally, I would like to thank the company Artifon, in particular Johan Andersson and Alf Berntson, for finding a B-format microphone for me, on which I could do the measurements. Special thanks also to Sennheiser Sweden for providing the microphone.

Kilian Schulze-Forster, Gothenburg, June 2018

Contents

1
Introduction

The B-format microphone – also called soundfield or Ambisonics microphone – consists of four microphone capsules arranged on the surface of a sphere at the corners of a tetrahedron. It was invented in the 1970th by Craven and Gerzon [1] to make recordings with full 3-dimensional spatial information. A picture of one example of such a microphone is shown in Figure 1.1.

Figure 1.1: Picture of a B-format microphone

The idea is to construct four desired microphone signals by combining the capsule signals in different ways. This way one signal corresponding to an omnidirectional microphone can be obtained, which contains the pressure in the centre of the capsule array. It is denoted W. Moreover, three pressure gradient signals (X, Y, Z) can be obtained having figure-of-eight polar patterns along the three Cartesian axes. The four signals are called B-format signals. With a B-format microphone it is possible, to a certain extend, to obtain these four signals for the exact same location in

space. This would not be possible if the four signals were to be recorded by four different microphones with the respective characteristics due to their finite size. The information captured by the B-format signals satisfy the requirements of human directional hearing [2].
The four capsules of the soundfield microphone are closely spaced but cannot be considered coincident at higher frequencies, where the wave length is in the order of the spacing. Consequences are gain, directional response and phase errors in the B-format signals. Different filters has been suggested to correct these errors [3, 4]. The first topic of this thesis is those filters. Their effects are investigated by means of simulation and measurements and the achievable quality of B-format signals is discussed.
3-dimensional sound recordings became recently even more relevant thanks to new emerging technologies and gadgets linked to smartphones, virtual reality glasses, and 360° cameras. In this context the most relevant question is how to play back 3-dimensional recordings to the user providing them the best possible spatial impression. The process of calculating for example two ear signals from a B-format sound file is called encoding or rendering. There are commercial renderers like facebook360 [5] available, that render 360° sound in connection with 360° videos. However, they do not disclose details about their implementation. Thus, the transformation of the signals from the B-format to the ear signal cannot be comprehended entirely making these renderers inapplicable for research in this field, where one needs control of every processing step. Consequently, the second topic of this thesis is the encoding of B-format recordings. In particular, a real-time binaural renderer including head tracking will be implemented in the programming language Python. It should serve as a basis for the development of more advanced solutions and to make further research about auralization of 3-dimensional sound fields.
The coincident pressure and pressure gradient signals of the B-format microphone can also be of interest beyond audio applications. Especially because the particle velocity can be obtained from the pressure gradient and pressure and particle velocity are required to calculate the sound field impedance. This makes, in theory, a B-format microphone able to measure the absorption coefficient of materials in-situ. The standard methods for measurements of the absorption of a material are impedance tube methods and the reverberant room method. However, for those measurements a sample of the material of interest needs to be cut out and put into the tube or the room [6]. This is not always possible in practice. You cannot cut out a piece of the wall in a building or the seat of an old concert hall. Also, the properties of surfaces might differ between being in the laboratory and being in the actual room, where they belong. Therefore, in-situ measurements of absorption would greatly facilitate the work in room acoustics. No standard method for in-situ measurements exists to date. However, a hand-held device has recently attracted attention, that can measure pressure and particle velocity in two close points. It is called Microflown [7] and measures the particle velocity directly from the temperature difference between two close points [8]. A similar device could be constructed with a B-format microphone instead of the Microflown sensor. Therefore, it is interesting to investigate to which extend absorption can be measured in-situ with a B-format microphone. This is the third topic of this thesis. The question will be

examined theoretically by means of simulation and practically by means of measurements.
The rest of this thesis is structured in the following way. In chapter 2, the theory regarding the B-format microphone and its signals is summarized by a literature review. Chapter 3 deals with the implementation of non-coincidence correction filters. Applications of the B-format microphone such as Auralization of 3D sound recordings and in-situ absorption measurements are addressed in chapter 4. All results are presented and discussed in chapter 5 and the thesis is finished by a conclusion in chapter 6.

2
Theory

In this chapter the theory about B-format microphones is summarized. At first, the microphone array signal processing is explained. Then, two Python scripts that were developed in the course of this thesis to simulate B-format microphones are explained.

2.1 Microphone Array Signal Processing

Some signal processing is required in order to obtain the desired signals from a recording with a B-format microphone. In this section, the set-up of a B-format microphone is explained followed by an overview of the required signal processing.

2.1.1 The B-Format Microphone

A B-Format microphone consists of four microphone capsules which are arranged in the shape of a tetrahedron. The capsules can be interpreted to be at the vertices [2, 9, 10], the edges [11] or the faces [12, 13] of the tetrahedron, which can be confusing when comparing the work of different authors. In this thesis the capsules are assumed to be at the vertices of the tetrahedron. Furthermore, the microphone's looking direction is assumed to be towards the positive x-axis. This implies that the four capsules do not point in the directions of the Cartesian axes, they are arranged with a 45 degree offset relative to the axes as shown in Figures 2.1 and 2.2. Each capsule can then be labeled by three letters indicating its position relative to the look direction of the microphone. This leads to the labels LFU (left-front-up), RFD (right-front-down), LBD (left-back down) and RBU (right-back-up), which are used in this thesis and were originally introduced by [1]. They are also used by most soundfield microphone manufacturers [4].
The individual capsules have the general polar pattern

$$(\alpha + (1 - \alpha)\cos(\phi)) \tag{2.1}$$

In most papers they are assumed to be sub-cardioids with $\alpha = \frac{2}{3}$ [2, 10, 11], but there are also microphones with cardioid capsules with $\alpha = \frac{1}{2}$, for example the Core Sound TetraMic [14] or the Sennheiser AMBEO VR Mic [15].
The output vector that contains the four capsule signals is called A-format and is denoted by $s_A = [s_{LFU}, s_{RFD}, s_{LBD}, s_{RBU}]$ [11]. However, the signal of interest is the so-called B-format, which contains a signal with omnidirectional characteristics (W) and three signals with figure-of-eight characteristics (X, Y, Z) pointing towards the

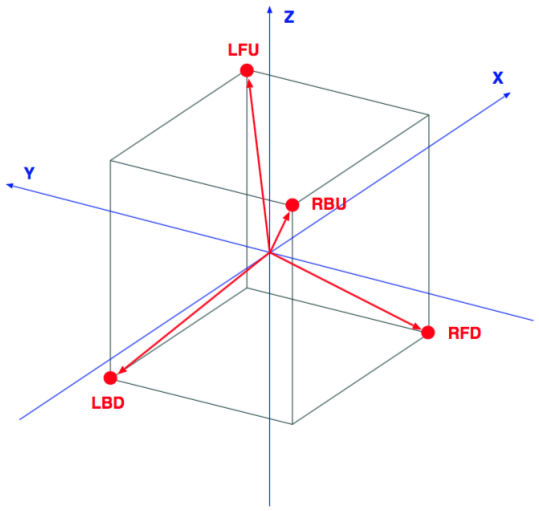


Figure 2.1: 3-dimensional view on capsule positions and labels. Source: [9]

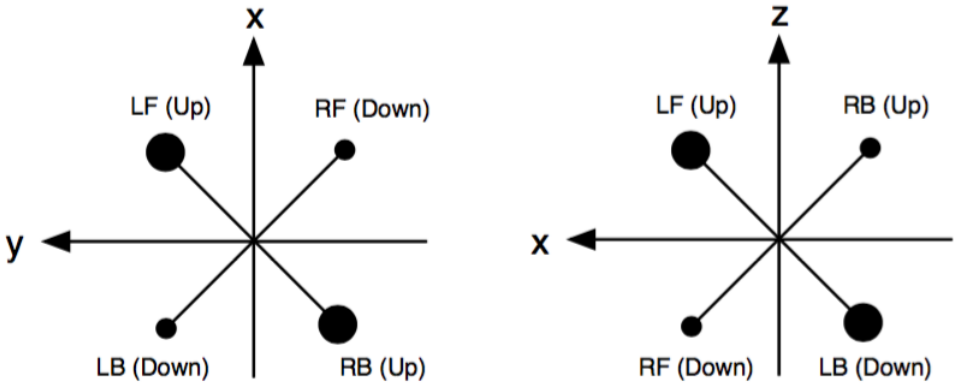


Figure 2.2: 2-dimensional view on capsule positions and labels. Source: [10]

positive x-, y- and z-axis respectively [10]. How to obtain the B-format is explained in the next section.

2.1.2 A-Format to B-Format Conversion

Figure 2.3 shows the directivity of the four microphone capsules in the x-y-plane. It can be seen that by adding or subtracting the signals from the four capsules one can obtain a figure-of-eight or an omnidirectional polar pattern. The signals W, Y, X, Z are constructed as originally introduced by [1] and explained in more detail by [2] according to the following rules.

$$\begin{aligned} W &= s_{LFU} + s_{RFD} + s_{LBD} + s_{RBU} \\ X &= s_{LFU} + s_{RFD} - s_{LBD} - s_{RBU} \\ Y &= s_{LFU} - s_{RFD} + s_{LBD} - s_{RBU} \\ Z &= s_{LFU} - s_{RFD} - s_{LBD} + s_{RBU} \end{aligned} \tag{2.2}$$

To obtain the omnidirectional signal W, all capsule signals are added. The X signal is obtained by adding the signals of forward pointing capsules and subtracting the signals of backwards point capsules. Y and Z are obtained by adding respectively subtracting left and right pointing or up and down pointing capsule signals.

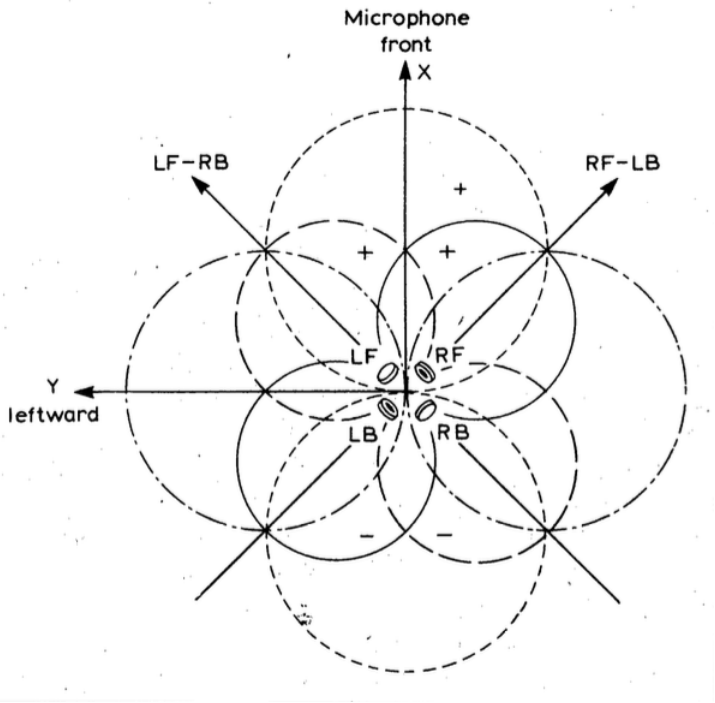


Figure 2.3: Directivity pattern of capsules in x-y-plane. Source: [2]

It should be noted that the sensitivity of the signals obtained by adding the signals of the capsules is increased by approximately 3 dB [2]. Furthermore, some more signal processing is required to keep the desired characteristics of the B-format signals also at higher frequencies, where the capsules can no longer be considered as coincident. This will be elaborated in the section 2.1.4.

2.1.3 Spherical Coordinates, Ambisonics, and Spherical Harmonics

Often 3-dimensional soundfields – and this is what a B-format microphone is supposed to capture – are described using spherical harmonics decomposition, also called Higher Order Ambisonics. From this point of view a B-format signal contains an Ambisonics signal of first order [11]. Therefore, a brief introduction to the underlying theory is given in the following.
At first, to avoid confusion, it is preferable to give a definition of spherical coordinates, which will be consistently used in this thesis. A point in the 3-dimensional space is defined by (r, θ, ϕ), where r is its distance to the origin, ϕ with $0 \leq \phi < 2\pi$ describes the azimuthal angle in the $z = 0$ plane originating at the x-axis and θ with $0 \leq \theta < \pi$ describes the polar angle originating at the z axis, which is called colatitude in this case [16]. A visualization of this definition is shown in Figure 2.4.

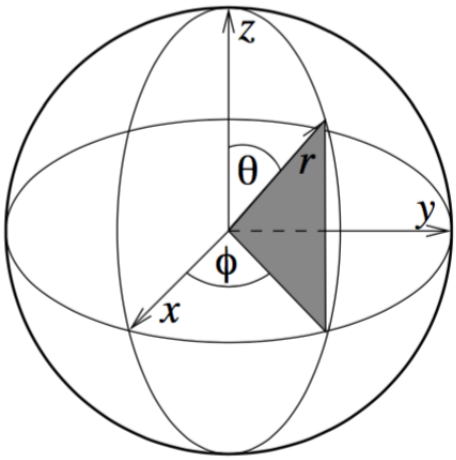


Figure 2.4: Spherical coordinate system. Source: [17]

Ambisonics describe a soundfield in one location $x = (r, \theta, \phi)$ by the following approximation [11].

$$p(x,k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_n^m(k) j_n(kr) Y_n^m(\theta, \phi) \tag{2.3}$$

where $j_n(kr)$ is the spherical Bessel function of first kind, k denotes the wave number and $Y_n^m(\theta, \phi)$ are the spherical harmonics explained below. The so-called Ambisonics coefficients $A_n^m(k)$ contain the information about the soundfield and can be interpreted as the W, X, Y and Z signals when Ambisonics theory is applied to the B-format microphone as shown further below. In this case n only runs to the order $N = 1$ [11], as mentioned above, the B-format holds an Ambisonics signal of first order.
Spherical harmonics describe functions on the surface of a sphere. For a detailed derivation, which is beyond the scope of this thesis, the reader is referred to [17]. They can be expressed as

$$Y_n^m(\theta, \phi) = N_n^m P_n^m(\cos(\theta)) e^{jm\phi} \tag{2.4}$$

where $n \in \mathbb{N}_0$ is the order, m is the degree with $-n \leq m \leq n$, N_n^m is a normalization term (will be discussed in section 2.1.5 "B-Format Audio Formats"), $P_n^m(\cos(\theta))$ is the Legendre function describing the dependence on the colatitude and $e^{jm\phi}$ is the azimuth function [11]. The angles (θ, ϕ) describe a position on the unit sphere following the definition of spherical coordinates given above.
The shape of the in the context of the soundfield microphone relevant zeroth and first order spherical harmonics is shown in Figure 2.5. Note that the shapes are identical with the directivity patterns of the ideal B-format signals.
Spherical harmonics are – among many other applications – used to compute the Ambisonics coefficients. When applied to a B-format microphone, which uses $L = 4$ microphone capsules for spatial sampling on a sphere, the coefficients are calculated as follows [11].

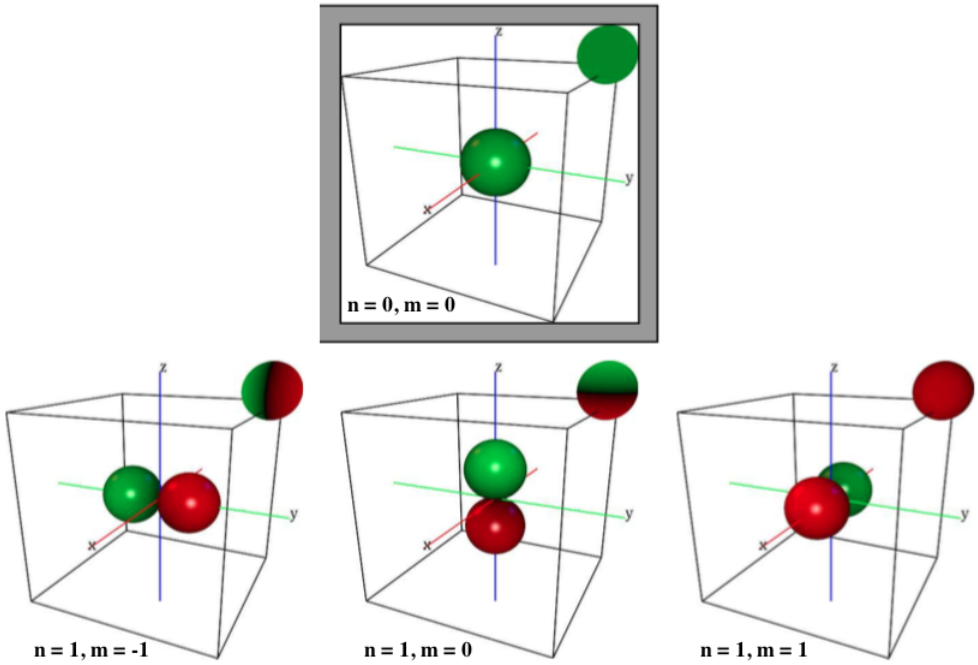


Figure 2.5: Zeroth and first order spherical harmonics, green parts indicate positive and red parts negative extends. The sphere on the top right shows the distribution on the unit sphere. Source: [17], edited.

$$A_n^m(k) = V_{n,\alpha}(kR) \sum_{l=1}^{L} s_l(x_l, k) Y_n^m(\theta_l, \phi_l) \tag{2.5}$$

where $V_{n,\alpha}(kR)$ is a so-called non-coincidence correction filter, which is discussed in the section 2.1.4 "Non-Coincidence Correction", R is the distance from the centre of the capsule's diaphragm to the origin of the coordinate system, which lies in the centre of the microphone array and s_l is the signal of capsule number $l = 1, 2, 3, 4$. It can be seen that for the B-format microphone case, where n runs to $N = 1$, four Ambisonics coefficients must be calculated, namely A_0^0, A_1^{-1}, A_1^0 and A_1^1. Thus, equation (2.5) can be written in matrix form [11]:

$$\begin{bmatrix} A_0^0 \\ A_1^{-1} \\ A_1^0 \\ A_1^1 \end{bmatrix} = \begin{bmatrix} V_{0,\alpha} & 0 & 0 & 0 \\ 0 & V_{1,\alpha} & 0 & 0 \\ 0 & 0 & V_{1,\alpha} & 0 \\ 0 & 0 & 0 & V_{1,\alpha} \end{bmatrix} \begin{bmatrix} Y_{1,0}^0 & Y_{2,0}^0 & Y_{3,0}^0 & Y_{4,0}^0 \\ Y_{1,1}^{-1} & Y_{2,1}^{-1} & Y_{3,1}^{-1} & Y_{4,1}^{-1} \\ Y_{1,1}^0 & Y_{2,1}^0 & Y_{3,1}^0 & Y_{4,1}^0 \\ Y_{1,1}^1 & Y_{2,1}^1 & Y_{3,1}^1 & Y_{4,1}^1 \end{bmatrix} \begin{bmatrix} s_{LFU} \\ s_{RFD} \\ s_{LBD} \\ s_{RBU} \end{bmatrix} \tag{2.6}$$

where the spherical harmonics are expressed in the form $Y_{l,n}^m$, where l is the number of the respective microphone capsule. In [11] it is derived that the matrix containing the spherical harmonics eventually can be written as

$$\frac{1}{\sqrt{4\pi}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \end{bmatrix} \tag{2.7}$$

which results in the same combination of the four capsule signals as outlined in the previous section. The Ambisonics coefficient vector contains the B-format signals as $\left[A_0^0 A_1^{-1} A_1^0 A_1^1\right] = [WYZX]$. It can be seen that this results in a different order of the B-format signals following the order and degree succession of the corresponding Ambisonics coefficients as opposed to the alphabetic order of the signals' names.

2.1.4 Non-Coincidence Correction

When converting the A-format to the B-format signal it is assumed that the four microphone capsules are coincident. This is a good assumption at low frequencies but as the frequency increases so that the wave length is in the order of the capsule spacing it is clear that the capsules cannot be considered as coincident anymore. The finite capsule spacing causes phase differences between the A-format signals, which result in gain, phase and directional response errors when converting them to the B-format. Different solutions how to correct these errors can be found the literature. Three of them, which are based on the idea of correcting the responses by filters, are introduced in this section.
The first is developed by Gerzon [3], who showed that for a hypothetical ideal spherical microphone uniformly covered with several capsules all outputs corresponding to spherical harmonics of the same order need exactly the same frequency response correction. Consequently, only two different correction terms are necessary if only omnidirectional and figure-of-eight responses corresponding to zeroth and first order spherical harmonics (fc. figure 2.5) are of interest. Since the B-format microphone array is the best approximation to a uniform covering of a sphere with only four capsules, the same applies to the B-format signals. This is the reason why in equation (2.6) the matrix with the non-coincidence correction filters only contains two different filters, namely $V_{0,\alpha}$ and $V_{1,\alpha}$ on the diagonal and zeros everywhere else. These filters are based on Gerzon's suggestions in [3] and are derived in [10] and expressed in a different way in [11] to work with the Ambisonics way of obtaining and correcting the B-format shown in the previous section. In the latter case they are computed as follows.

$$V_{n,\alpha}(kR) = \frac{1}{\alpha j_n(kR) - j(1-\alpha) j_n'(kR)} \tag{2.8}$$

where n is the spherical harmonics order, α is defining the polar pattern of the capsules according to equation (2.1), k is the wave number, R is the distance of the capsules' centers to the array center, j is the imaginary unit and j_n is the spherical Bessel function of first kind, the apostrophe denotes the derivative. The filters are basically the inverse of the array responses [11]. Gerzon also notes that there is an upper frequency limit for this non-coincidence correction given by $f = 10.8/R$ kHz, where R is given in cm. Above that limit, higher order spherical harmonics dominate the B-format signals and the correction becomes less effective. He suggests diffuse field compensation for the capsules above that limit [3]. For $R = 1.47$ cm, which is a typical value for a B-format microphone [2], this limit becomes $f = 7.35$ kHz [10].

In [10] Faller and Kolunzija show that this filter is not optimal and suggest that an optimal filter would minimize the mean square error between the desired and

the actual directional responses. They formulate them as follows in the frequency domain. Note that in the original paper another definition of spherical coordinates was used. The notation below is adapted to the definition used in this thesis.

$$H_W(\omega) = \frac{\int_{-\pi}^{\pi}\int_{0}^{\pi} d_W(\theta,\phi)D_W^*(\theta,\phi,\omega)\cos(\theta)d\theta d\phi}{\int_{-\pi}^{\pi}\int_{0}^{\pi} D_W(\theta,\phi,\omega)D_W^*(\theta,\phi,\omega)\cos(\theta)d\theta d\phi} \tag{2.9}$$

$$H_X(\omega) = \frac{\int_{-\pi}^{\pi}\int_{0}^{\pi} d_X(\theta,\phi)D_X^*(\theta,\phi,\omega)\cos(\theta)d\theta d\phi}{\int_{-\pi}^{\pi}\int_{0}^{\pi} D_X(\theta,\phi,\omega)D_X^*(\theta,\phi,\omega)\cos(\theta)d\theta d\phi} \tag{2.10}$$

where $d_W(\theta,\phi)$ and $d_X(\theta,\phi)$ are the desired responses for the W respectively X signals

$$d_W(\theta,\phi) = 1 \tag{2.11}$$

$$d_X(\theta,\phi) = \cos(\phi)\sin(\theta) \tag{2.12}$$

and $D_W(\theta,\phi,\omega)$ and $D_X(\theta,\phi,\omega)$ are the actual directional responses, the "*" indicates the complex conjugate. Due to symmetry the filter for Y and Z look basically the same.
The two introduced non-coincidence correction filters are based on theory. In practice one also has to deal with differences among the microphone capsules, limited positioning precision and diffraction effects [10] when designing such a filter. For the filters proposed by Faller and Kolunzija one could actually determine $D_W(\theta,\phi,\omega)$ and $D_X(\theta,\phi,\omega)$ but it would be necessary to measure in many positions and also every measurement has a limited precision.
To tackle these problems and to design practical non-coincidence correction filters Farina suggests a filter design based on one measurement per B-format signal [4]. In detail, he proposes to measure the magnitude and phase response of the B-format microphone to a sound impinging from the positive x-, y- and z-axis and compare it with a reference microphone with a flat frequency response. The inverse of the transfer function between the two microphone responses should then be applied as a filter to the corresponding B-format signals, where the x-axis measurement is used for the W and X signal and the y- and z-axis measurement for Y and Z respectively. Furthermore, he suggests to calibrate the four capsules as a first step by comparing the frequency response of each capsule with the one of the reference microphone to a sound impinging frontal. Again, the inverse transfer function between the capsule and the reference microphone can be applied as an equalization filter. This way differences between the capsules can be evened out [4]. The non-coincidence correction filters obtained that way are very similar to the least mean square error filters up to 7 kHz. However, the magnitude of the thus obtained W correction filter should be reduced above 7 kHz because it is much higher than that of the theoretically optimal filter [10].

2.1.5 B-Format Audio Formats

Two different standards regarding how to normalize, store and exchange Ambisonics data have arisen. They are called FuMa and AmbiX format and introduced in this

section.
The FuMa format – named after its inventors Furse and Malham – stores the four B-format channels in alphabetic order (W, X, Y, Z). The normalization term N_n^m in equation (2.4) becomes $1/\sqrt{2}$ for the W signal. The other three signals have weighting factors applied so that the maximum value each can take is $|1|$. These factors can be obtained by inspection up to order $n = 3$. This is called MaxN normalization [19]. The weighting of the channels has the advantage that approximately equal signal levels are produced in the four channels [1]. A disadvantage is that this format is limited to 16 channels and a file size of 4 GB [18], which might be enough for the first order Ambisonics signal of a B-format microphone but not if one wants to use Higher Order Ambisonics.
Therefore, Nachbar et al. proposed a new format called AmbiX in [18]. The first difference to the FuMa format is the channel order. It uses the Ambisonics Channel Numbers (ACN), where

$$ACN = n^2 + n + m \tag{2.13}$$

assigns a channel number to each Ambisonics signal depending on its order n and degree m. Consequently, the order of the signals is $\left[A_0^0 A_1^{-1} A_1^0 A_1^1\right] = [WYZX]$ (cf. section 2.1.3). As a second difference the AmbiX signal applies SN3D normalization, where

$$N_n^m = \sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}} \tag{2.14}$$

which can be found in [11, 17, 18]. For the purpose of normalizing and storing recordings of a B-format microphone these are the essential differences between the formats to be considered. More details can be found in [18]. The FuMa format was the standard for a long time but recently the application of the AmbiX format is more and more preferred.

2.2 Simulation of the B-Format Signal Directivity

This script was developed to test the expected change in the B-format signals' directivity with increasing frequency and the influence of the theoretical non-coincidence correction filters introduced above. It can be found in the GitHub repository `https://github.com/kilian-schufo/B-Format-masters-thesis` under the name Directivity_Simulation. The script is available in two formats. The ".py" file can be used for an immediate online preview of the script while the ".ipynb" file can be downloaded and then opened and edited as notebook with Jupyter. The different functions implemented in the script are explained in the following.
The function "distance" returns the distance in meters between two points given in spherical coordinates. It is used to calculate the distance of sound sources to the microphone capsules. It transforms the given coordinates to carthesian coordinates and then calculates the distance d by

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \tag{2.15}$$

The function "angle" returns the angle in radiants between the two vectors connecting two given points in spherical coordinates respectively with the origin of the coordinate system. It is applied to compute the angle with which sound waves impinge on a microphone capsule (first point) from a certain source (second point). Since the capsules are assumed to be cardioids the direction of incidence has an influence on their sensitivity. The angle Φ between the vectors $\vec{v}_1$ and $\vec{v}_2$ is calculated as follows.

$$\Phi = \arccos\left(\frac{\vec{v}_1 \cdot \vec{v}_2}{|\vec{v}_1||\vec{v}_2|}\right) \tag{2.16}$$

where $\vec{v}_1 \cdot \vec{v}_2$ is the dot product of the vectors.
The function "bformat" returns the B-format signals W, X, Y, Z for one given frequency f and capsule directivity constant a. It computes the signals for many source positions everywhere around the microphone and returns them as a 200×200 matrix. Along the rows the colatitude angle θ takes values from 0 to π and along the columns the azimuth angle ϕ takes values from 0 to 2π. This way a whole sphere of possible source positions around the B-format microphone is covered.
The microphone array's centre is assumed to be in the origin of the spherical coordinate system. The capsules are modelled as points with the following coordinates in accordance with [2, 3].

$$\begin{aligned} LFU &= (R, \tfrac{\pi}{2} - \theta_{\text{tilt}}, \tfrac{\pi}{4}) \\ LBD &= (R, \tfrac{\pi}{2} + \theta_{\text{tilt}}, \tfrac{3\pi}{4}) \\ RBU &= (R, \tfrac{\pi}{2} - \theta_{\text{tilt}}, \tfrac{5\pi}{4}) \\ RFD &= (R, \tfrac{\pi}{2} + \theta_{\text{tilt}}, \tfrac{7\pi}{4}) \end{aligned} \tag{2.17}$$

where the array radius is $R = 0.0147$ m and $\theta_{\text{tilt}} = \arctan\frac{1}{\sqrt{2}}$. The pressure signal $p(k)$ at a capsule given one source radiating plane waves with amplitude $A = 1$ at distance d is then

$$p(k) = e^{-jkd} \tag{2.18}$$

where $k = \frac{\omega}{c}$ is the wave number, ω is the angular frequency and $c = 340$ m/s the speed of sound. Depending on the angle of incidence Φ the signal of a capsule is then

$$s(k, \Phi) = (a + a\cos(\Phi))p(k) \tag{2.19}$$

where $a = 0.5$ for cardioid capsules (cf. equation (2.1). By combining the capsule signals as explained in equation (2.2) the signals W, X, Y, Z are obtained for all possible angles of incidence. They can then be used to visualize the three-dimensional signal directivities or the directivities in the horizontal plane.
Two more functions are comprised in this script, which apply the non-coincidence correction filters to the signals. The first one is called "corr_batke". It is an implementation of Gerzon's suggestion for such filters as derived in [11] by Batke et al. The general form was given in equation (2.8) and in the implementation of the relevant filter looks as follows for the omnidirectional signal (Ambisonics order $n = 0$).

$$V_{0,0.5}(kR) = \frac{1}{0.5 \cdot \frac{\sin(kR)}{kR} - j(1-0.5)(kR \cdot \cos(kR) - \frac{\sin(kR)}{(kR)^2})} \tag{2.20}$$

And for the pressure gradient signals with Ambisonics order $n = 1$ the filter becomes

$$V_{1,0.5}(kR) = \frac{1}{0.5 \cdot (\frac{\sin(kR)}{(kR)^2} - \frac{\cos(kR)}{kR}) - j(1-0.5)(((kR)^2 - 2) \cdot \sin(kR) + \frac{2kR\cos(kR)}{(kR)^3})} \tag{2.21}$$

The function multiplies the raw B-format signal matrices with the respective filter for the given frequency. As mentioned in [11] the filter for the first order Ambisonics signals introduces a 90° phase shift. This is compensated in the implementation by multiplication with the factor $-j$. Otherwise the obtained signals would carry falsified spacial information.
The second filter function called "corr_faller" implements the corrections as proposed by Faller et al. in [10] according to the equations (2.9) and (2.10). For the W signal the desired response is 1 for all directions of incidence. Consequently, in the numerator the matrix for the W signal obtained from the "bformat" function is just element-wise conjugated and then all the elements are summed up, which is the discrete form of the integration over all colatitude and azimuth angles. In the denominator the W signal matrix is multiplied element-wise with its conjugated version and then also summed up. The obtained filter $H_W(\omega)$ for the considered frequency is then multiplied with the original W signal matrix to obtain the corrected signal.
The procedure for the correction of the X, Y and Z signals is almost the same, only the desired response is of course different. Therefore, the desired figure-of-eight responses along the respective axes needed to be computed for all the angles of incidence present in the 200×200 matrices for the X, Y and Z signals and also be written in such a matrix with corresponding angles θ and ϕ along the rows and columns respectively. This was done by calculating the angle Φ between a vector pointing along the respective axis and all considered source positions in the "bformat" function using the "angle" function. The matrix with the desired response is then obtained by calculating $\cos(\Phi)$ for all the source positions. By looking into the actual scrip found in the appendix this will become more clear to the reader.
Once the desired response matrix is computed, it is element-wise multiplied with the conjugate of the signal matrix and the resulting matrix is again summed up to obtain the numerator of equation (2.10). The denominator is obtained as for the W signal and then the filters $H_X(\omega)$, $H_Y(\omega)$, and $H_Z(\omega)$ are multiplied with the signal matrices X, Y, Z respectively.
Both filter functions also attenuate the W signal by 3 dB by multiplying the signal with $1/\sqrt{2}$ to comply with the FuMa standard.
Results from simulations with this script are presented and discussed in chapter 4, section 5.1.

2.3 Simulation of the Microphone Impulse Response

In order to test potential applications of B-format microphones it is useful to be able to simulate a recording by such a microphone. Therefore, a script was developed which computes the impulse responses of the B-format signals W, X, Y, Z to a source placed at any desired position relative to the microphone. These impulse responses can then be convoluted with arbitrary signals to obtain a simulation of a recording with a B-format microphone. It can be found in the GitHub repository `https://github.com/kilian-schufo/B-Format-masters-thesis` under the name IR_Simulation. The script is available in two formats. The ".py" file can be used for an immediate online preview of the script while the ".ipynb" file can be downloaded and then opened and edited as notebook with Jupyter. The script uses similar functions as the above explained simulation script with some adaptions.
The "distance" and "angle" functions are identical with the ones from the previous section. The function "bformat" follows the same principle as above. However, in this script it calculates the B-format signals only for one given source position and for a given frequency vector (instead of many source positions for one frequency). Thus, the output of this function is the frequency response of the four B-format signals W, X, Y, Z to a source at the given position.
The script also includes the "corr_batke" function for the theoretical non-coincidence correction. The second non-coincidence correction filter introduced in 2.1.4 as suggested in [10] by Faller et al. was not used in this implementation because it only yields reasonable results if sources from many different angles are considered at the same time, which is not the case here.
The last step to obtain the impulse responses of the B-format signals is to take the inverse Fourier transform of the frequency domain signals. However, these frequency domain signals are to this point single-sided containing only the frequencies given to the "bformat" function. To get reasonable impulse responses from the inverse Fourier transform we need double-sided spectra. Therefore, the function "doubside" is implemented. It takes the single-sided spectrum as input and adds a reversed, conjugated version of itself (without the first ($n = 0$) and last ($n = N/2$) value) yielding a double sided spectrum with an even number of N frequency points. All points except at $n = 0$ and $n = N/2$ are also scaled by $\frac{1}{\sqrt{2}}$ to split the energy between the "two sides" of the spectrum. Then, the real parts of the inverse Fourier transforms of the spectra are computed. These are the impulse responses. Any .wav-soundfile can be loaded in the script and convolved with the W, X, Y, Z impulse responses to place the sounds from the soundfile at the source position given to the "bformat" function. This yields the four signals as separate arrays. They can then be saved as .wav-files and combined to a four channel .wav-file using for example the free software Audacity. Depending on the channel order and whether the W signal is attenuated by 3 dB or not both the FuMa and the AmbiX format can be created this way.

3
Filter Implementation

This chapter deals with the implementation of the practical non-coincidence correction filters as proposed by Farina [4]. First, the measurement setup and procedure are described. Then, the filter design is explained.

3.1 Microphone Measurement Setup

The B-format microphone under test was a Sennheiser AMBEO VR Mic [15]. The measurements were carried out under anechoic conditions in the anechoic chamber of the Division of Applied Acoustics at Chalmers University of Technology. A test signal was generated and played from a loudspeaker and then recorded with the B-format microphone and a reference microphone. A sketch of the setup is shown in Figure 3.1, a picture of the setup is shown in Figure 3.2 and a list of equipment is provided below. In Figure 3.2 one can see that there was a wooden plate on the floor between loudspeaker and microphones, which was covered with absorptive materials in order to avoid reflections.

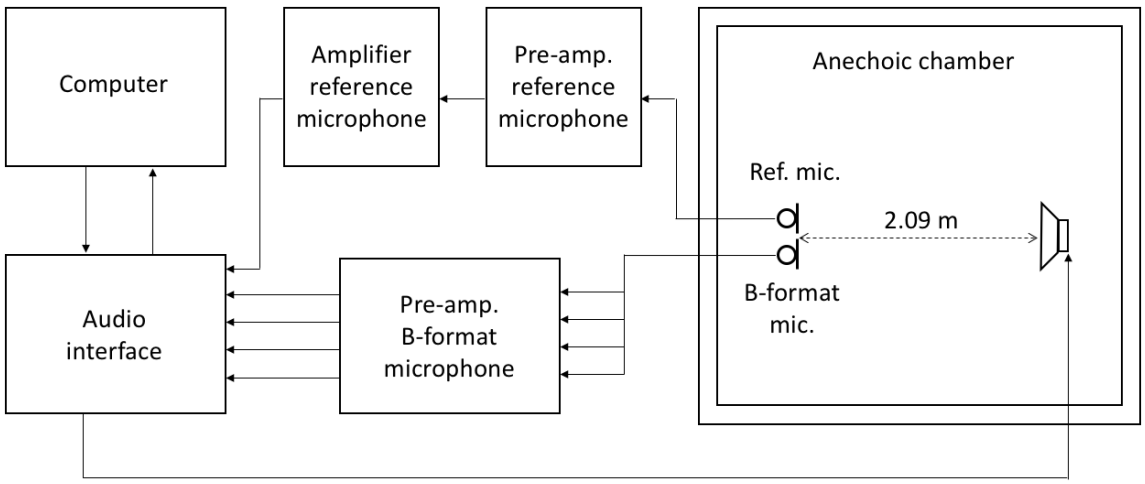


Figure 3.1: Sketch of the measurement setup

The following equipment was used in the measurements.

- Edirol UA-101 Audio interface
- RME Quadmic II 4 Channel Mic to Line preamplifier for the B-format microphone

- Brüel & Kjær preamplifier, type 2669-C-001, serial number: 3084952 for the reference microphone
- Brüel & Kjær amplifier, type 1708 for the reference microphone
- Brüel & Kjær condenser microphone (omnidirectional), cartridge type 4166, serial number: 583733 (reference microphone)
- Sennheiser AMBEO VR B-format microphone
- Genelec 8030A Active Monitoring System

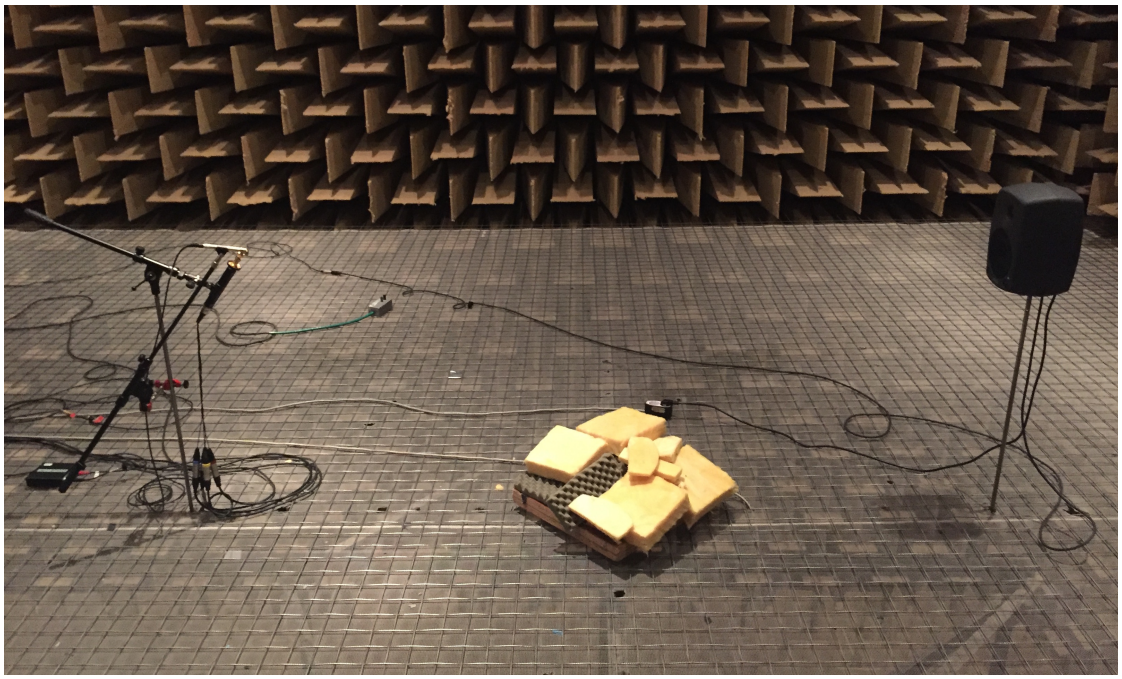


Figure 3.2: Photo of the measurement setup

The frequency response of the loudspeaker is shown in Figure 3.3 and the frequency response of the reference microphone is presented in Figure 3.4.

The test signal used in all measurements was a logarithmic sweep from 100 Hz to 21000 Hz in 5 seconds with an initial phase shift of 90° to avoid an impulse at the beginning of the signal. It had an maximal amplitude of 0.6 and was generated with Python using the Jupyter Notebook. The sampling frequency of the signal and the recordings was $f_s = 44100$ Hz. 22050 samples (= 0.5 seconds) with the value zero were added to the test signal because the recording and playing time of the test signal are equally long in the applied "Sounddevice" Python module and the last sound waves need to propagate the distance between loudspeaker and microphones after the sweep was played.

At first, the four different microphone capsules were compared to the reference microphone. Therefore, the microphone was mounted in a way so that the capsule under test pointed directly towards the loudspeaker and the capsule's and driver's centers were on one horizontal line. The reference microphone was placed next to the capsule under test.

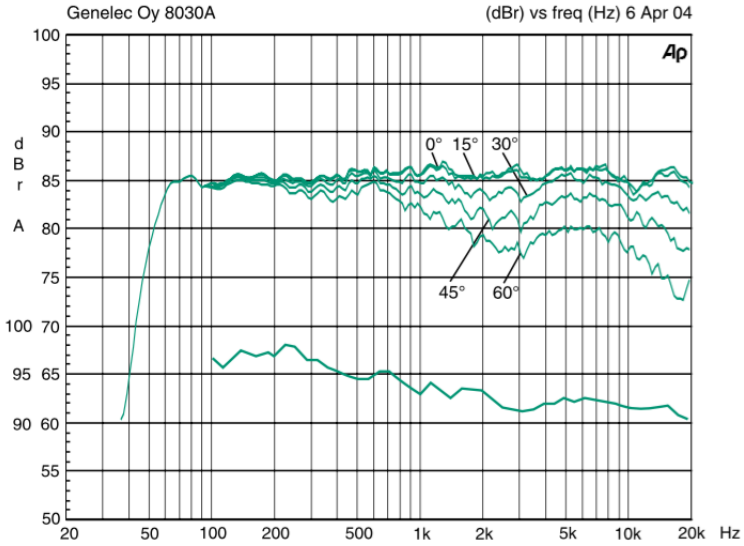


Figure 3.3: Frequency response of the loudspeaker. The upper curve group shows the horizontal directivity characteristics measured at 1 m. The lower curve shows the system's power response. Source: [34]

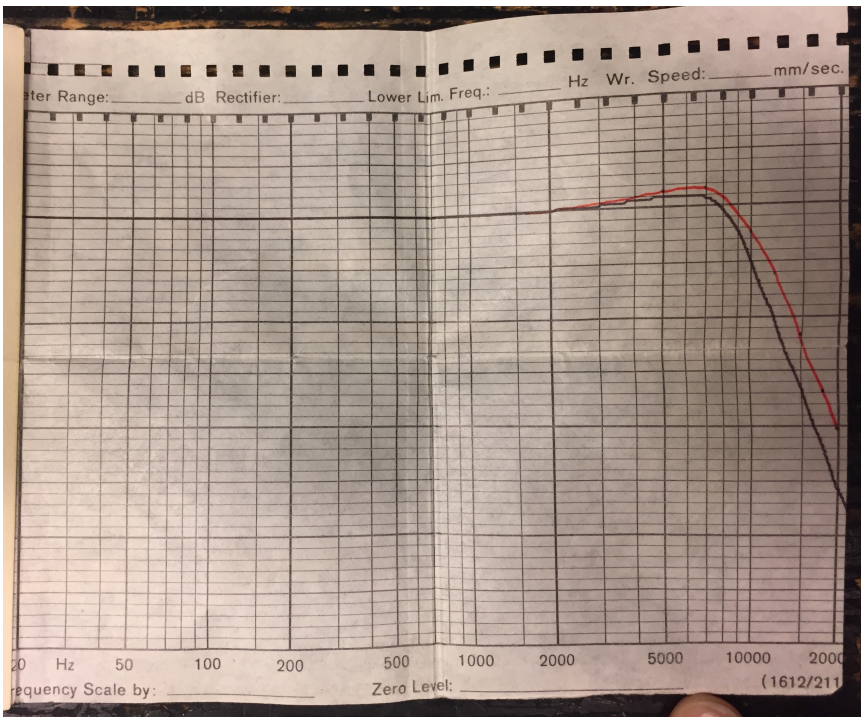


Figure 3.4: Frequency response of the reference microphone

The microphone positions are shown in Figure 3.5 for upwards and downwards pointing capsules. In order to position both microphones the best way, the distance between the microphones and the loudspeaker had to vary a few centimeters between the different capsule measurements. Also, the reference microphone was moved between the different measurements. All individual capsule measurements were

done with the same channel of the pre-amplifier for the B-format microphone so that they all had the exactly same gain.

(a) Capsule pointing up **(b)** Capsule pointing down

Figure 3.5: Microphone orientation at the capsule measurements

In a second measurement the influence of the capsule's position and look direction on its frequency response in the measurement setup was investigated. Therefore, the microphone was mounted vertically and the capsule signals were recorded for the front side (looking direction) of the microphone (cf. section 2.1.1) looking towards the loudspeaker and then looking away from the loudspeaker (turned by 180°). This way, each capsule was in the second measurement at the same position in which another capsule was in the first measurement. Thus, it could be investigated if the measured frequency response of a capsule depends on its position. A sketch of the two microphone orientations is shown in Figure 3.6 .

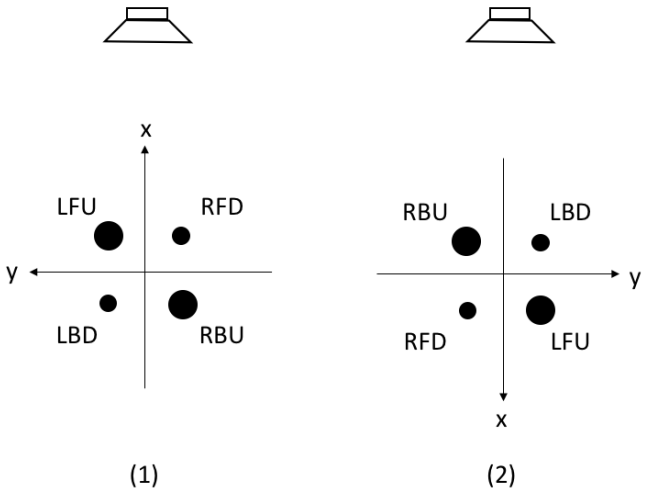


Figure 3.6: Setup of the second measurement in the horizontal plane

In a third measurement, the four B-format signals were compared to a reference microphone. Therefore, the B-format microphone was mounted in a way that the B-format signal under test was pointing towards the loudspeaker. The reference microphone was again placed close to the B-format microphone. The positions are shown in Figure 3.7 for the measurement of the X and the Z signal.

(a) Microphone positions of X measurement

(b) Microphone positions for Z measurement

Figure 3.7: Microphone orientation at the B-format signals measurements

For these measurements all 4 channels of the pre-amplifier were used and the gain was set manually to as equal as possible values. However, with the knobs on the pre-amp only a limited precision can be achieved but equal gains are important in order to obtain the regular B-format signal polar patterns. Therefore, the gain of each channel was measured by sending a low amplitude version of the test signal to each channel and recording the output. The measured gains (transfer functions between input and output) are shown in Figure 3.8. It can be seen that the gains are fairly similar but there are still differences. All channel gains fluctuate around a mean value probably because a very low amplitude input signal was chosen for this measurement. Thus, the mean value for each channel gain was calculated and then used to equalize the channels by dividing the corresponding capsule signals by this value. The gains are $g_{ch1} = 34.96$, $g_{ch2} = 31.56$, $g_{ch3} = 29.85$, and $g_{ch4} = 32.31$ (the values are not in dB but are the amplitude multiplication factors). The signals from all channels were then multiplied by 30 to have equal gain in each channel.

3.2 Non-Coincidence Correction Filter Design

The design of the practical non-coincidence correction filters is explained in this section using the example of the filter for the W signal. The filters for the three other signals were designed with the exact same procedure.

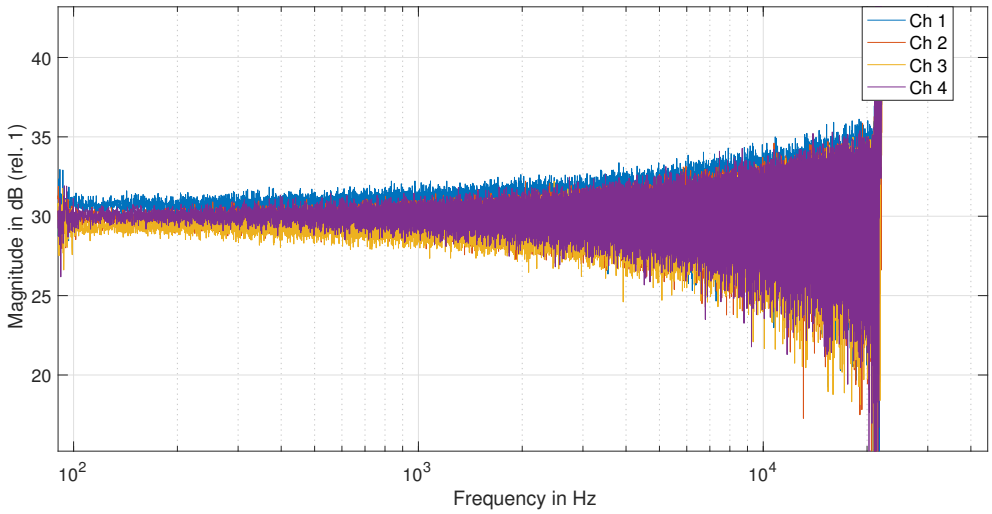


Figure 3.8: Gains of the four pre-amp channels

The W signal $s_W(t)$ was obtained from the above described measurement – where the B-format signals were compared to the reference microphone – by adding the gain corrected signals of all four capsules. Then, a Fast Fourier Transform (FFT) was done to obtain the signal in the frequency domain, which is denoted $S_W(\omega)$. Also the reference microphone signal was transformed to the frequency domain giving $S_{ref}(\omega)$. Then, the transfer function between the signals was computed as

$$H_W(\omega) = \frac{S_W(\omega)}{S_{ref}(\omega)} \tag{3.1}$$

An Inverse Fourier Transform was then performed on the inverse of the transfer function leading to the impulse response that is supposed to be the filter.

$$h_W(t) = \mathscr{F}^{-1}(\frac{1}{H_W(\omega)}) \tag{3.2}$$

However, since the sweep signal in the measurements started at 100 Hz and went up until 21000 Hz, the signals do not contain useful information outside of this range. Thus, a fourth order bandpass butterworth filter with the cutoff frequencies 100 Hz and 19000 Hz was applied to the impulse response. The resulting impulse response was cut off after 512 samples and this is the final non-coincidence correction filter for the W signal.
The resulting filter impulse responses and frequency responses are presented in section 5.2. They are also applied to the B-format signals and their effect is discussed in the same section.

4
Applications

In this chapter two applications of the B-format respectively the B-format microphone are introduced and investigated. At first, the auralization of B-format recordings on head phones is addressed. Then, in-situ absorption measurements with a B-format microphone are investigated.

4.1 Auralization

This section deals with the auralization of B-format recordings. After a summary of the underlying theory, the binaural real-time renderer, which was developed in the course of this thesis, is introduced.

4.1.1 Theory of Auralization of B-Format Recordings on Headphones

To make use of recordings with a B-format microphone, one needs to be able to play back the files leveraging its spatial information. This can be done in several ways on loudspeaker arrays or headphones. In the following, one way to play back the B-Format on headphones – which is also called binaural rendering – is explained.
As explained before, the B-format contains the Ambisonics coefficients $A_n^m(\omega)$, which can be used to describe the sound field in the middle of the tetrahedral microphone array according to equation (2.3). In other words, they tell us what the listener would have heard if he was in the position of the microphone during the recording and his or her both ears were located in the mid-point of the array. This pressure signal can be reproduced by a spherical loudspeaker array and one can calculate the driver signal for each loudspeaker. In a second step, one can calculate the ear signals of a listener in the middle of this loudspeaker array by convolving the loudspeaker signals with Head Related Impulse Responses (HRIR) corresponding to the respective positions of the loudspeakers relative to the listener's ears.
According to equation (3.21) in [20] the driver signals can be calculated as follows (the notation is adapted to the conventions in this thesis):

$$D(\theta, \phi, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{1}{2\pi R^2} \sqrt{\frac{2n+1}{4\pi}} \frac{A_n^m(\omega)}{G_n^0(\omega)} Y_n^m(\theta, \phi) \tag{4.1}$$

where θ and ϕ are colatitude and azimuth of the position of the considered loudspeaker, R is the radius of the loudspeaker array, $G_n^0(\omega)$ are the coefficients of the

spherical harmonics decomposition of the loudspeaker directivity and $Y_n^m(\theta, \phi)$ are the spherical harmonics.
Using the definition of real spherical harmonics in [21], which is

$$Y_n^m(\theta, \phi) = N_n^{|m|} P_n^{|m|}(\cos\theta) \begin{cases} \sin(m\phi) & \text{for } m < 0 \\ \cos(m\phi) & \text{for } m \geq 0 \end{cases} \tag{4.2}$$

with the normalization

$$N_n^m = (-1)^m \sqrt{\frac{(2n+1)(2-\delta_m)}{4\pi} \frac{(n-m)!}{(n+m)!}} \tag{4.3}$$

where ()! is the factorial and $\delta_m = 1$ for $m = 0$ and is zero otherwise and the recurrence relations for the associated Legendre functions P_n^m given in [21] , the zeroth and first order spherical harmonics look as follows.

$$\begin{aligned} Y_0^0 &= \sqrt{\frac{1}{4\pi}} \\ Y_1^{-1} &= \sqrt{\frac{3}{4\pi}} \sin\theta \sin\phi \\ Y_1^0 &= \sqrt{\frac{3}{4\pi}} \cos\theta \\ Y_1^1 &= \sqrt{\frac{3}{4\pi}} \sin\theta \cos\phi \end{aligned} \tag{4.4}$$

The relevant coefficients G_n^0 are

$$\begin{aligned} G_0^0 &= 4\pi\sqrt{\frac{1}{4\pi}} \\ G_1^0 &= -4\pi(j)^-1\sqrt{\frac{3}{4\pi}} \end{aligned} \tag{4.5}$$

With these equations and a set of appropriate HRIRs one can produce binaural renderings of a B-format signal, which give the listener a 3D-sound perception. However, to improve the spatial perception of sound one might want to use a head tracker, which measures the head position of the listener, because it is easier to localize a sound source if we can turn the head and analyze sound differences. This implies that the reproduced sound scene must move according to the listeners head position inside the virtual loudspeaker array.
This could be realized by turning the loudspeaker array while keeping the driver signals constant. This way one has to have data of the HRIRs for many source positions, because for each head position the angles to the loudspeakers is different. Another possibility, which does not require updating the HRIRs, is to turn the sound scene (by changing the driver signals) while keeping the loudspeaker positions constant. From the two figure-of-eight signals X and Y in the horizontal plane one can calculate new figure-of-eight signals X' and Y', which are pointing to a desired

direction with an offset of ϕ from the look direction of the B-format microphone during the recording. The new signals can be calculated by

$$\begin{aligned} X' &= \sin\phi Y + \cos\phi X \\ Y' &= \cos\phi Y - \sin\phi X \end{aligned} \tag{4.6}$$

4.1.2 Implementaton of a Real-Time Binaural Renderer

The real-time binaural renderer reads B-format sound files stored as .wav-files with the FuMa format and plays them back on headphones. In doing so, the spatial information of the B-format file is preserved and audible for the listener. The Python script of this renderer implementation is available in the GitHub repository `https://github.com/kilian-schufo/B-Format-masters-thesis` under the name $\text{Binaural}_R enderer.Itisavailableintwoformats.The".py"filecanbeusedforanimmediateonlinep$ The listener's head rotation is measured with a head tracker so that the sound scene seems to stand still around the listener even if they turn their head. Consequently, the position of sound sources in the scene is easier to determine. The general principle of the renderer is depicted in Figure 4.1.

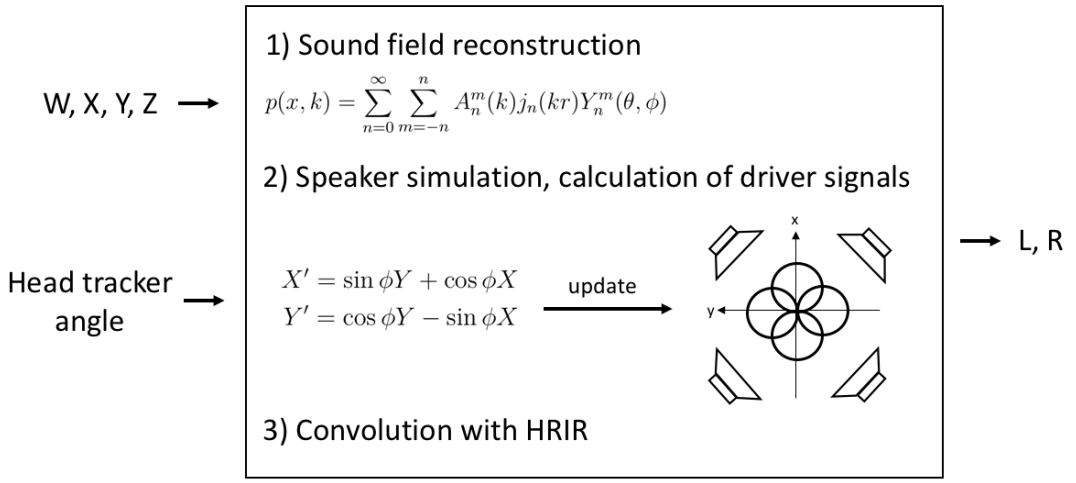


Figure 4.1: General principle of the binaural real-time renderer

The four channels of the sound file corresponding to the W, X, Y, Z signals are interpreted as the Ambisonics coefficients A_n^m as mentioned in section 2.1.3. Then, the sound field at the centre of the microphone array can be reconstructed by the given equation in Figure 4.1. The listener is assumed to be in the centre of the coordinate system looking towards the positive x-axis. It is also assumed that he is surrounded by a spherical loudspeaker array. The ideal loudspeaker positions would have been at the same angles as the microphone capsules with a bigger radius so that they are equally distributed on the sphere. However, also the availability of head related transfer functions (HRTFs) is required for every loudspeaker position. Taking these two conditions into account the actual loudspeaker positions are $(2, 70°, 45°)$, $(2, 115°, 135°)$, $(2, 65°, 225°)$, and $(2, 110°, 315°)$.
The signals for the drivers in order to produce the reconstructed sound field in the middle of the array can be calculated with equation (4.1). The head tracker inputs

the angle of the listener's current look direction. This angle is used to update the sound scene accordingly. Finally, the four driver signals are convolved with the head related impulse responses (HRIR) corresponding to their respective position and the four resulting signals per ear are summed to generate the binaural signals L and R. A more detailed description of the renderer is given in the following four subsections. First, the "play_file.py" script [29] by spatialaudio.net, which is the basis for the renderer, is briefly introduced. Then, necessary extensions in order to play B-format sound files on headphones are explained. The last two subsections deal with the integration of the head tracker and some details about the used HRTFs respectively HRIRs in the renderer.

4.1.2.1 The Basis Script

The binaural renderer is based on a Python script, which was made to play any mono or stereo .wav-soundfile blockwise by connecting the script to the so-called JackPilot of the JACK Audio Connection Kit [30]. It is provided by spatialaudio.net in [29]. The script loads the specified sound file, opens a Jack client with two channels and plays back the sound file via these two channels. The blocklength with which the files are processed is determined in the JackPilot settings. For the binaural renderer the minimum block length is 2048 samples. The JackPilot connects the two channels received from the script to the output channels of the device on which it is running, for example, built-in loudspeakers or connected headphones on a computer.
The script divides the sound file into blocks and writes every block into a matrix called "data", which has as many rows as the file has channels and as many columns as the number of samples per block. The most important function is the "process" function. It is called repetitively for every block of the sound file that is processed and sends the currently processed block to the JackPilot. Inside this function some processing can be implemented to manipulate the signals within each block. To play the sound file in real-time, the processing time may not exceed the time it takes to play one block. In order to build the binaural renderer, this script was extended. The extensions are explained in the next subsection.

4.1.2.2 Extension of the Basis Script

Most of the extensions are written into the "process" function and are explained in the following. At first, every time the "process" function is called, the current angle of the head tracker is read and written into the variable "rad" in radiants. It describes the azimuth angle ϕ of the look direction of the listener, where $\phi = 0$, if the listener looks towards the positive x-axis. This means that the head position is updated every block and its accuracy depends on the blocklength. More details about the integration of the head tracker are given in section 4.1.2.3. Depending on the look direction of the listener, new signals for the horizontal plane X' and Y' (in the script called XX and YY) are then computed from the original signals X and Y according to equation (4.6). For FuMa files the X and Y signals are found in the second and third row of the "data" matrix. Since we are now loading files with four channels and we use a blocklength of 2048 samples, this "data" matrix has the shape 4×2048 and the rows correspond to W, X, Y, Z. As a next step, the driver signals

are computed for the current block, which is done by the function "driver_signal". It takes the position of a loudspeaker in spherical coordinates as well as the four signals W, XX, YY, and Z, which are interpreted as Ambisonics coefficients A_n^m, as input. It then computes and returns the driver signal for the corresponding loudspeaker according to equation (4.1). Next in the "process" function, each driver signal is convolved with a HRIR for the left and the right ear. The length of the HRIRs is 236 samples and in section 4.1.2.4 more details about them are provided. This yields four signals for each ear and the signals for each ear are then summed and result in the final ear signals that can be played back on headphones. However, due to the convolution the signals become longer than the blocks. This problem is handled with the overlap-add method. Therefore, an object based buffer is implemented. The class "buff" is defined and an instance called "my_buffer" is created. Each of the eight ear signals gets its own buffer as an attribute of "my_buffer". They are initialized with arrays containing only zeros having the length as the HRIR minus one sample. The last step in the "process" function is then to update these buffers with the samples that do not fit in the block size anymore to save them for the next block.

4.1.2.3 Integration of Head Tracking

The head tracker used for the binaural renderer is based on the SparkFun "9DOF Razor IMU". Detailed information about its hardware and software are found in [31]. It is connected via USB to the computer and read as serial port by the Python script. The baud rate must be set to 57600 so that the head tracker is running in text output mode. If the baud rate is different from that, the head tracker switches to binary output mode and the output is not readable with the Python module "pySerial". In text mode it outputs the angles yaw, pitch and roll as strings in a format that looks like this: "#YPR=11.71,-0.34,34.12". This piece of information is also called a line. For the binaural renderer only the head rotation in the horizontal plane is considered, which is defined by the yaw. It describes the azimuth angle of the listener's look direction.
The output rate of such strings is independent of the call of the "process" function and all strings are stored in a buffer. Consequently, it has to be made sure that each time the "process" function is called – meaning that one audio block is prepared to be played back – the most recent string is accessed and read. This is not as straightforward as one could expect. The part of the buffer that is accessible is limited to 1020 bytes. Once this is filled, the tracker keeps sending strings, that are stored as well but cannot be read directly. This causes problems when some time goes by between the activation of the tracker reading and the start of the audio file play back. Thus, a while loop was implemented that reads the whole accessible buffer as long as it is 1020 bytes long. Once a string is read it is deleted from the buffer. As soon as the buffer contains less than 1020 bytes, the script allows the "process" function to be called. Inside the function all lines which are at this time in the buffer are read and stored in a list. Then, the last and therefore most up-to-date yaw angle is taken from that list, which is either in the last or second last line in the list. The head tracker outputs the information byte by byte and not line by line. Therefore the last line can be incomplete at the moment the buffer is read and look

for example like "#YPR=2". That is why the second last line must be considered as well. Another aspect to be considered is that the coordinate system of the head tracker is most likely not aligned with the coordinate system of the listener's head. However, we want the yaw to be zero when the listener looks straight ahead, this is in positive x-direction. This is achieved in the following way. At the moment the script starts running the listener should look straight ahead. The very first yaw ϕ_0 that is read inside the "process" function does then correspond to looking straight ahead but its actual value might not be zero. This value is stored in a buffer and then used to calibrate the head tracker's coordinate system according to the following rule, where ϕ_{az} is the desired azimuth angle corresponding to the definition of spherical coordinates used in this work and ϕ_{yaw} is the value that the head tracker outputs.

$$\phi_{az} = -(\phi_{yaw} - \phi_0) \tag{4.7}$$

The first minus on the right hand side of the equation is necessary to change the orientation of the positive and negative sense of orientation to adapt it as well to the considered definition of spherical coordinates. In Figure 4.2 the head tracker's and the desired coordinate system are depicted to clarify this issue. Note that the coordinate system on the left is just one example of a random orientation of the not calibrated coordinate system of the head tracker. Furthermore, in the calibrated system the angles will not necessarily be expressed as numbers between 180° and −180°, depending on the value of ϕ_0. However, the sense of rotation will follow the right system in Figure 4.2 regarding positive and negative angles, which is important for equations (4.6) for the updated X and Y signals.

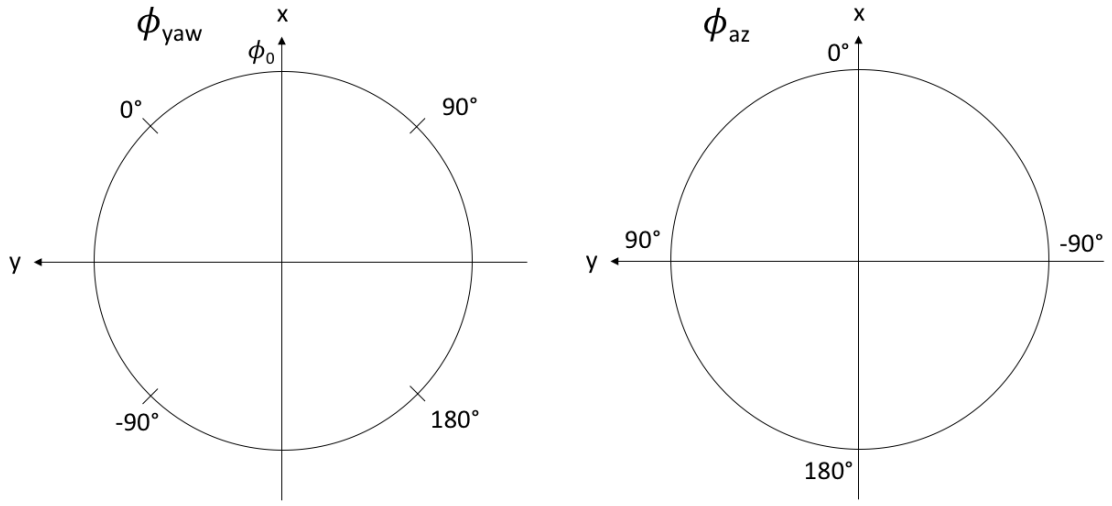


Figure 4.2: Left: coordinate system with random orientation of the head tracker. Right: calibrated coordinate system

4.1.2.4 Head Related Transfer Functions

The head related transfer functions (HRTFs) used in the binaural renderer are taken from the data base of the "Acoustics Research Institute" of the "Austrian Academy of Science". Details about the measurement method and other information can be

found in [32]. They are in-the-ear HRTFs downloaded in the SOFA-format 1.0. The data base comprises measured HRTFs of hundreds of persons. The specific HRTFs used in this work are found under the name "dtf_nh163.sofa" in [33]. The HRTFs were measured with a sampling frequency of 48000 Hz. For this work they were resampled to 44100 Hz because most music is recorded with that sampling rate.
The HRTF for the left ear for a source coming from the front of the test person, this is $(r, \theta, \phi) = (1.2, 90°, 0°)$, is shown in Figure 4.3. It can be seen that it has a peak at 3800 Hz and is attenuating high and low frequencies. This reflects our human auditory system but is also due to the limited power of loudspeakers at low frequencies used in the measurements. As a result, sounds convolved with the corresponding impulse response will be strongly manipulated and have a lower sound quality.

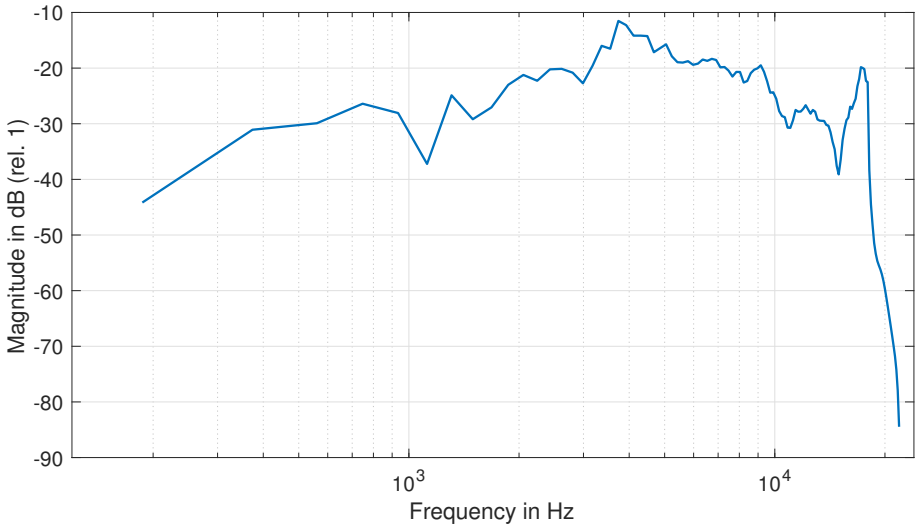


Figure 4.3: Head related transfer function for the left ear for source position $(r, \theta, \phi) = (1.2, 90°, 0°)$

This must be equalized if the binaural renderer is supposed to produce good sound quality. Therefore, the HRTFs used in the rendrer are equalized with the following procedure.

- A music file was convolved with the HRIR corresponding to the HRTF shown in the Figure above and then equalized with first-order low-frequency shelving and peak filters in Matlab. The parameters of the filters were adapted until the music sounded again acceptable.
- The song used was "Cold Hard Bitch" by "Jet" and the final filters are a low-frequency shelving filter with cut-off frequency $f_c = 100$ Hz and gain $G = 16$ dB, and five peak filters with the parameters centre frequency, bandwidth, gain (f_c, f_w, G) as follows: (200 Hz, 200 Hz, 13 dB), (400 Hz, 300 Hz, 2 dB), (800 Hz, 460 Hz, 4 dB), (1600 Hz, 1000 Hz, 3 dB), and (4000 Hz, 2000 Hz, -5 dB).
- The HRTF for $(r, \theta, \phi) = (1.2, 90°, 0°)$ was equalized with these filters and the resulting curve was taken as a reference curve. It its shown in Figure 4.4.

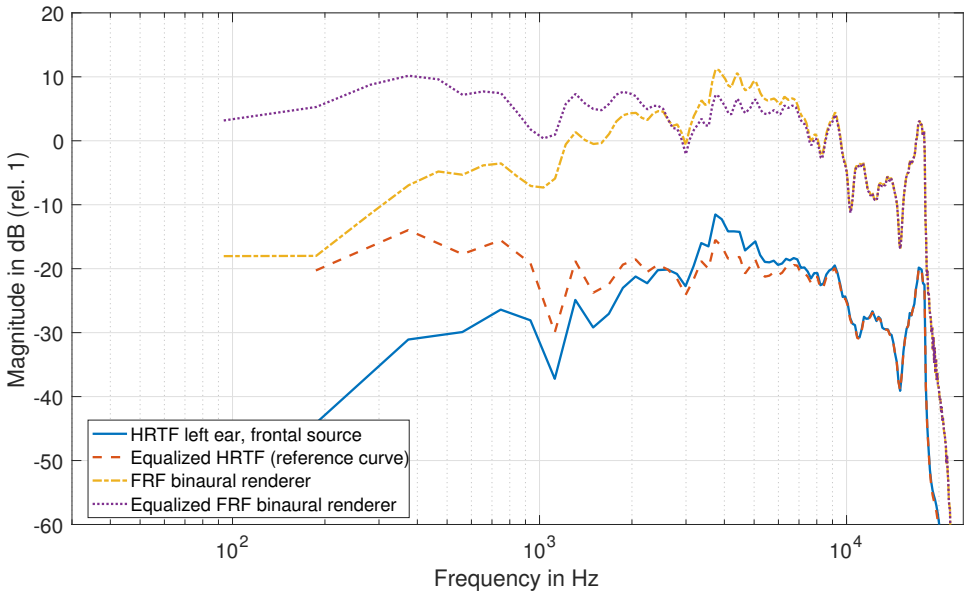


Figure 4.4: All relevant curves for the equalization for the left ear

- In order to obtain the frequency response function of the binaural renderer, the impulse response of a B-format microphone to a source at the frontal position $(r, \theta, \phi) = (1.2, 90°, 0°)$ was created with the microphone simulation script from section 2.3, saved as B-format FuMa file and then sent through the renderer. The FFTs of the two ear signals produced by the renderer are its frequency response functions. They are the test curves and the curve for the left ear is shown in Figure 4.4.
- The test curve needed to be adapted to the reference curve as close as possible. Experiments showed that this could be achieved with the same filters as used to produce the reference curve. The equalized test curve is shown in Figure 4.4.
- All HRIRs were equalized with the above described filter bank and the resulting impulses were then applied in the binaural renderer.

In Figure 4.4 it is also visible that the HRTFs reduce the amplitude of any signal that is convolved with them in the time domain. However, the renderer needs to produce outputs with high enough amplitudes so that the signals are still audible. Therefore, the driver signals are multiplied by 20000 in the binaural renderer (also the calculation of the driver signals itself reduces the amplitudes significantly) and that is why the FRF of the renderer has a higher magnitude than the original HRTFs.

4.2 In-situ Measurement of Absorption Coefficients

This section deals with in-situ absorption measurements using a B-format microphone. First, the theory behind this measurement is explained. Then, a simulation of the measurement as well as an actual measurement are addressed.

4.2.1 Theory of In-situ Measurement of Absorption Coefficients

The absorption coefficient α of a material describes the ratio of energy that is lost inside the material (e.g. transformed into heat) to the energy of the impinging sound wave. It can be seen as the part of the energy that is not reflected by the surface and thus calculated the following way from the reflection factor $r = |r|e^{j\phi_r}$.

$$\alpha = 1 - |r|^2 \tag{4.8}$$

A sound wave gets reflected when it encounters an impedance change as for example on the transition between air and another medium such as the wall of a room. The reflection factor is then obtained from the impedances Z_1 and Z_2 of the two media.

$$r = \frac{Z_2 - Z_1}{Z_2 + Z_1} \tag{4.9}$$

The specific acoustic impedance of a material is defined as the ratio of pressure p and particle velocity u.

$$Z = \frac{p}{u} \tag{4.10}$$

For wave propagation in air this becomes $Z_0 = \rho_0 c$, where ρ is the density of air and c is the speed of sound. Consequently, by measuring the pressure and particle velocity one can obtain the impedance of a surface and finally its absorption coefficient [23]. Both the pressure and particle velocity can be measured with a B-format microphone. The W signal corresponds to a signal of an omnidirectional microphone (pressure transducer) and is thus proportional to the pressure [26]. The X signal can be interpreted as a signal from a figure-of-eight microphone, which measures the pressure gradient $\frac{\Delta p}{\Delta x}$. With Newton's equation of motion it can be shown that the pressure gradient is proportional to the particle velocity in x-direction u_x [26].

$$\frac{\partial p}{\partial x} = -j\omega\rho_0 u_x \approx \frac{\Delta p}{\Delta x} \tag{4.11}$$

A possible setup of an in-situ measurement to obtain the absorption coefficient of a surface in a room with a B-format microphone is sketched in Figure 4.5.
However, in [24] it is shown that the impedance of a surface cannot be directly obtained by simply measuring p and u very close to the surface to approximate the actual values inside the material. One needs to model the soundfield at the microphone position to obtain reasonable values for Z. Three models of the soundfield in front of a sample material are proposed in [24]. The most appropriate one is the image source model with a plane wave reflection factor, which is also applied in [23]. Then the reflection factor r can be found as follows.

$$r = \frac{\frac{Z_{meas}}{Z_{free}} - 1}{\frac{Z_{meas}}{Z_{free}} + 1} \frac{h_s + h}{h_s - h} e^{jk2h} \tag{4.12}$$

where Z_{meas} is the quotient of pressure and velocity obtained by the measurement as sketched in Figure 4.5 and Z_{free} is the quotient of pressure and velocity obtained

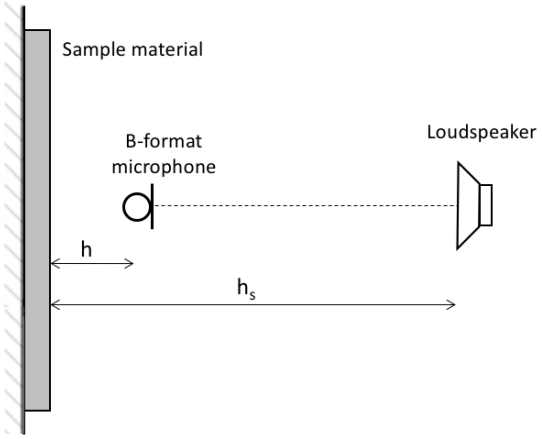


Figure 4.5: Setup of the in-situ absorption measurement. Based on [23].

under free field conditions. Z_{free} needs to be measured before a sample material can be tested with the described method. It is used in equation (4.12) to account for the fact that both the incident and reflected wave are recorded by the microphone [23]. The factor $\frac{h_s+h}{h_s-h}$ equalizes the attenuation of spherical waves with distance, it is the quotient of the reflected and direct path length. The factor e^{jk2h} accounts for the phase shift between direct and reflected wave at the microphone position due to the different path lengths.
One might think that a major challenge in using the B-format microphone for the described measurement is the calibration to obtain correct numerical values and the calibration of p-u probes is not at all straightforward [27, 28]. But as explained in [25], calibration is not necessary. We assume that our W and X signal have an arbitrary frequency dependent gain change ϵ_W and ϵ_X. The measured impedance Z' is then the actual impedance Z times a gain factor ϵ_Z

$$Z' = \frac{W\epsilon_W}{X\epsilon_X} = \epsilon_Z Z \tag{4.13}$$

Since in equation (4.12) the quotient of two impedance measurements with the same microphone are used, the factor ϵ_Z disappears and we can obtain the actual impedance ratio [25].
It would also be possible to use a spherical reflection factor for the soundfield model which led to equation (4.12). However, Alvarez and Jacobsen show in [24] that this does not yield better results in practice even though it is a more complex model. They also state that one can assume plane wave incidence without compromising on the quality of the results. In [23] in-situ absorption measurements with a p-u probe hand-held device using the Microflown [8] to measure particle velocity are tested against the conventional method in a Kundt's tube. Their results suggest that the method works in practice but has some limitations. For example, highly reflective surfaces cause some problems due to the low particle velocity close to the surface. The method gets also disturbed by reflections from sidewalls, which limits its usability in small rooms. Also a strong curvature – convex or concave – of the surface under test can cause significant deviations.

4.2.2 Simulation of the Absorption Measurement

To test if the planned absorption measurement can work in theory, a simulation of the procedure was implemented based on the image source model described in section 4.2.1. The corresponding Python script of this simulation is available in the GitHub repository `https://github.com/kilian-schufo/B-Format-masters-thesis` under the name Absorption_Measurement_Simulation. It is available in two formats. The ".py" file can be used for an immediate online preview of the script while the ".ipynb" file can be downloaded and then opened and edited as notebook with Jupyter.
The previously introduced function "bformat" (cf. section 2.2) was used to simulate the B-format signals for the setup of the absorption measurement shown in Figure 4.5. It was extended by an input parameter which is the amplitude of the simulated sound source. The microphone, the loudspeaker and the image source emitting the reflection on the test material are assumed to be in the horizontal plane with the microphone's looking direction pointing towards the sample material. The B-format signals W_d and X_d for the direct sound from the loudspeaker are calculated with the input parameters amplitude $A = 1$, a frequency vector, and the source position in spherical coordinates ($h_s - h$, $\pi/2$, π). The input parameters for the B-format signals W_r and X_r for for reflected sound generated by the image source are the amplitude, which in this case is the assumed reflection factor of the material (absorber) under test $A = r_{abs}$, the same frequency vector, and the source position ($h_s + h$, $\pi/2$, 0). All signals are non-coincidence corrected by the filter proposed by Gerzon [3]. Since in the measurement both the direct and reflected sound will be recorded, the signals are obtained as follows.

$$W_{meas} = W_d + W_r \tag{4.14}$$

$$X_{meas} = X_d + X_r \tag{4.15}$$

The free field impedance and the impedance of the measurement are then

$$Z_{frf} \cdot \epsilon_Z = \frac{W_d}{X_d} \tag{4.16}$$

$$Z_{meas} \cdot \epsilon_Z = \frac{W_{meas}}{X_{meas}} \tag{4.17}$$

Since plane waves are assumed, the complex reflection factor r is calculated the following equation, where the amplitude correction for spherical waves is neglected. The frequency dependent gain change ϵ_Z in the obtained impedances will disappear in this equation as explained in section 4.2.1.

$$r = \frac{\frac{Z_{frf}}{Z_{meas}} - 1}{\frac{Z_{frf}}{Z_{meas}} + 1} \cdot e^{jk2h} \tag{4.18}$$

Different values for the reflection coefficient of the test material r_{abs} or the distances h_s and h can then be defined to test the performance of the measurement procedure. This has been done and the results are presented and discussed in section 5.3.1. The

actual value of interest will be the absorption coefficient of the material, that can be obtained from the reflection coefficient (cf. section 4.2.1). However, to understand the performance in detail it is beneficial to look at the results for the complex reflection coefficient in the simulation.

4.2.3 Absorption Measurement Setup

The measurement of the absorption coefficient of a test material was carried out as introduced in the Theory chapter with the B-format microphone. A second measurement of the same material was then done in an impedance tube to have a reference value of its absorption coefficient.
For the measurement with the B-format microphone the same equipment was used as for the microphone measurements described in the section 3.1. However, no reference microphone and corresponding amplifiers were used. A sketch of the setup is shown in Figure 4.6 and a photo of the setup is presented in Figure 4.7.

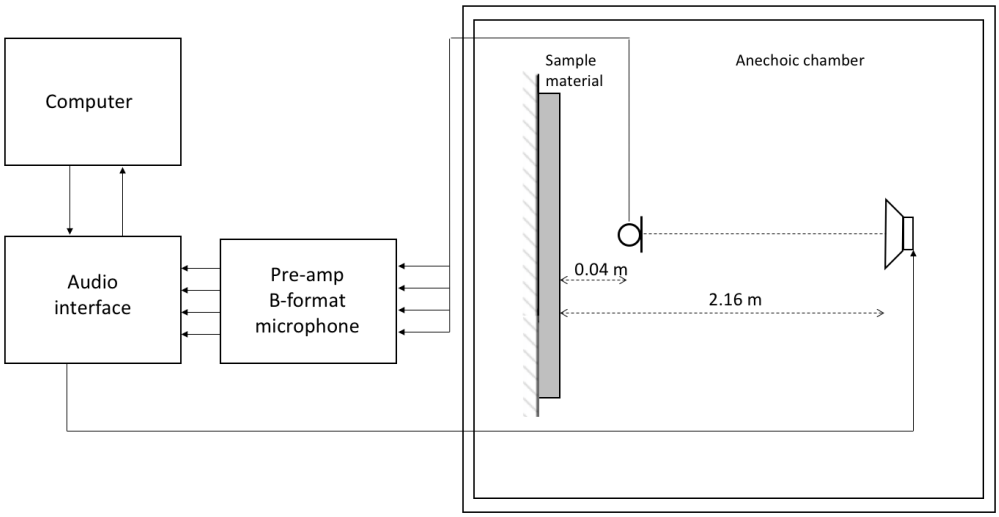


Figure 4.6: Setup of the absorption coefficient measurement

The test material was a sheet of glass fibre with a thickness of 5.5 cm. It was put on a 1.5 cm thick wood plate as can be seen in Figure 4.8. The same test signal as in the previous section was used.
At first, the test signal was recorded with the B-format microphone without the test material and wood plate present to obtain the W and X signal to compute the impedance of the sound field under free field conditions Z_{frf}. Then, the test material was mounted without changing the microphone position and another recording was done. From this second measurement Z_{meas} can be computed. The value for the absorption coefficient obtained from the measurements calculated with equations (4.8) and (4.12) is presented in section 5.3.
The measurement in the impedance tube was done with the well known transfer function method as for example explained in [35] with the small impedance tube at

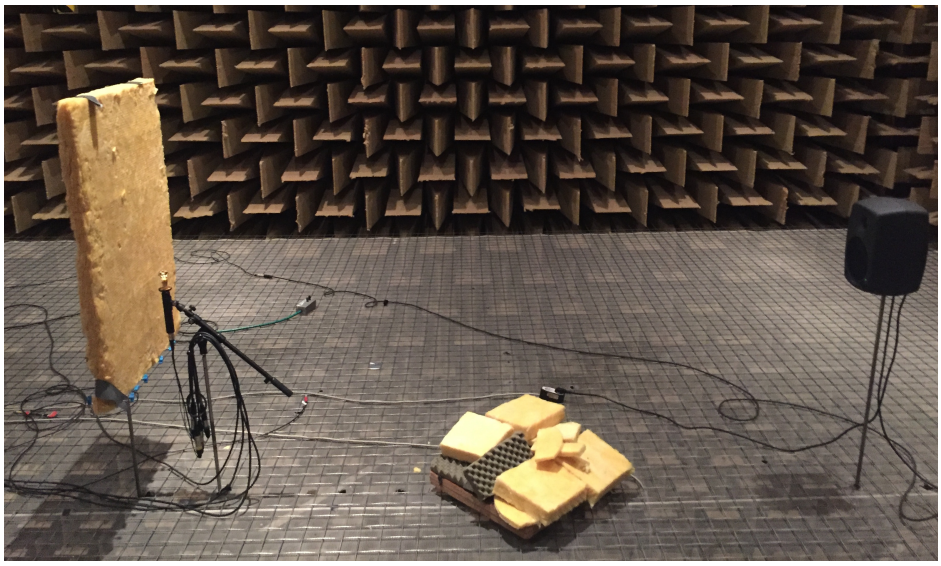


Figure 4.7: Photo of the absorption coefficient measurement

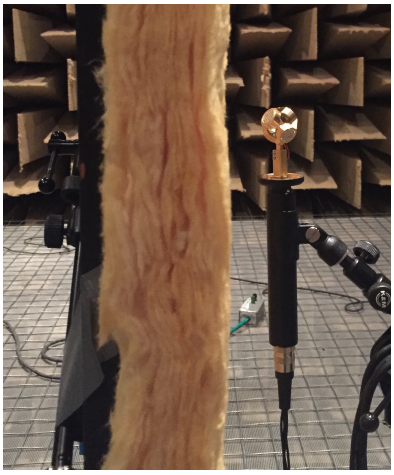


Figure 4.8: Detail of the mounting of the test material and microphone position

the Divison of Applied Acoustics at Chalmers University of Technology. It has an upper frequency limit of 2000 Hz and the results are shwown in section 5.3 as well.

5
Results

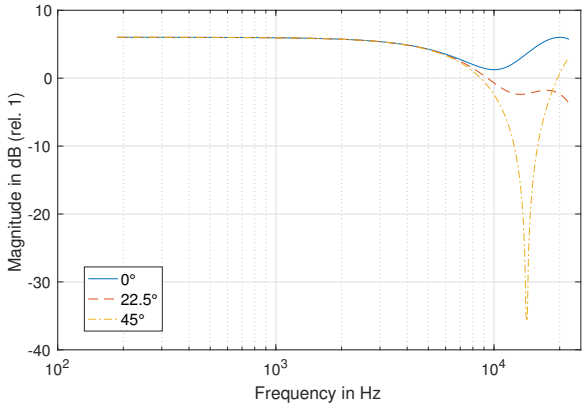
In this chapter the results are presented and discussed. First, the results of the microphone simulations are shown followed by the results of the microphone measurements including the practical non-coincidence correction filters. Then, the results of the absorption measurement simulation are shown as well as the results from the actual measurements.

5.1 Microphone Simulation Results

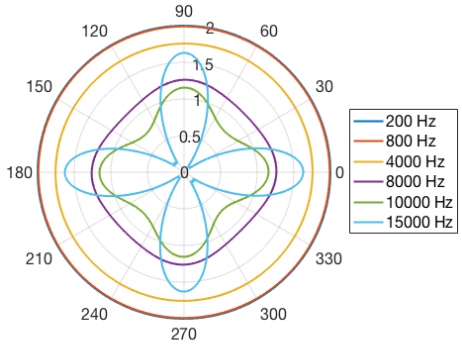
Simulations of the B-format signals' responses to sources at different positions at different frequencies were done with the simulation scripts introduced in sections 2.2 and 2.3 in order to know, what can be expected from a B-format microphone under perfectly controllable conditions. Then, the effect of the theoretical non-coincidence correction filters was simulated as well with these scripts.
As mentioned earlier, the ideal behaviour of the microphone would be to produce perfectly shaped omnidirectional (W) and figure-of-eight (X, Y, Z) patterns at all frequencies with a maximal gain of $1/\sqrt{2}$ for the W signal and unity of the other signals. There are of course limitations to these ideal responses, which are shown and discussed with the simulation results. The results are shown for the W and X signal. Due to symmetry of the microphone array, the Y and Z signals have the same behaviour as the X signal in the respective planes [10]. At first, the magnitude of the W and the X signal are discussed. Then, the phase difference between the signals is addressed.

5.1.1 Magnitude of the Omnidirectional Signal

In Figure 5.1 the frequency response at different angles in the horizontal plane and the polar pattern for different frequencies are shown for the W signal without any correction. Note that due to symmetry it is sufficient to only consider these three angles presented in the frequency response function. It can be seen that the signal has a flat frequency response up to 2000 Hz for all source positions. In the range of this flat response it has a very high gain of 6dB, which is due to the fact that the signals of all four capsules are added to obtain this signal and the capsule signals are in phase at these low frequencies. In the polar plot it can be seen that the shape of the signal stays rather omnidirectional up to 8000 Hz and derivations from this shape are most prominent at 45° (plus multiples of 90° due to symmetry) while at $0° + n \cdot 90°$ the response does not change as much with increasing frequency.



(a) Frequency response at different azimuth angles



(b) Polar pattern

Figure 5.1: W signal in the $z = 0$ plane, no correction

It can also be seen in the the frequency response function that above 2000 Hz the gain decreases with increasing frequency for all angles of incidence leading to dips at 9900 Hz for 0° and at 14200 Hz for 45° incidence. For the latter case the dip is also much stronger than for 0°. This needs further explanation especially because the directional signals have an increasing gain with increasing frequency (cf. Figure 5.9). The explanation for the dips can be found by looking at the signals of the four cardioid capsules for the different source angles. In Figure 5.2 the signals are shown for a source at 0° azimuth in the horizontal plane for one time instant.

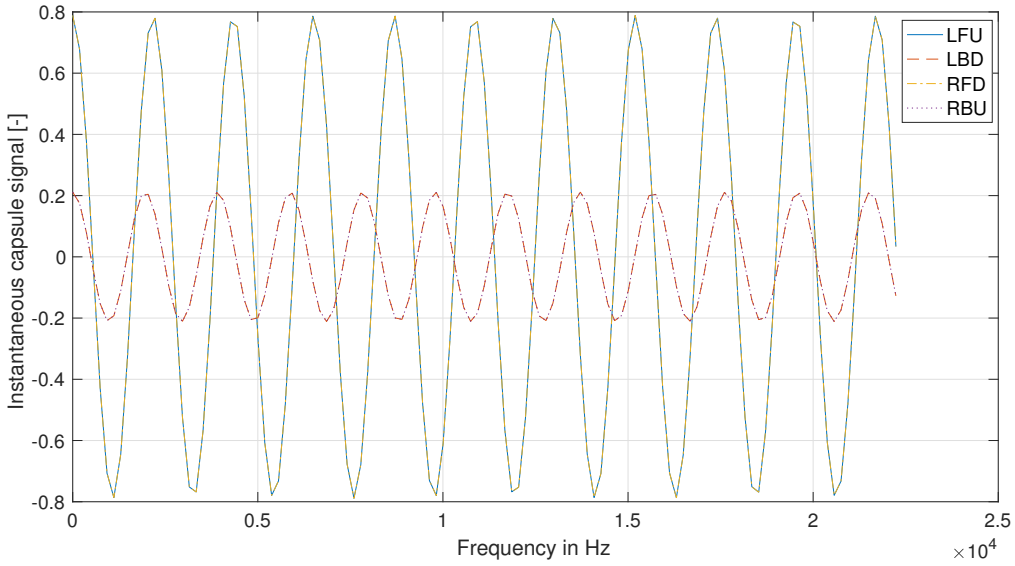


Figure 5.2: Instantaneous capsule signals at different frequencies for a source at 0° azimuth angle in the horizontal plane emitting all frequencies in phase

It can be seen that the phase relation between the signals changes with increasing frequency. The signals for the two front capsules (LFU and RFD) are identical and

the signals of the two back capsules (LBD and RBU) as well, because they have the same distance to the source. Since the W signal is obtained by adding the four capsule signals, the phase relation between the front and back capsule pair influences the signal. It can be seen that the pairs are 180° out of phase at 9900 Hz, which explains the dip in the frequency response. It can also be seen the the front capsule signals are stronger than the back capsule signals. Consequently, they cannot cancel out each other completely and the dip in the frequency response is "only" a few dB strong. This is not the case for a source at 45° azimuth, which can be explained by looking at the capsule signals for this case shown in Figure 5.3. The same phase shift between the capsule signals with increasing frequency is visible. However, since the source is at 45° azimuth now, the amplitudes of the capsules are different from the previous case. The LFU capsule pointing towards the source has the strongest amplitude, the LBD and RFD capsules have the same distance to the source and the same relative orientation (+/- 90°rotation) to it, so their signals are identical. The RBU capsule has the lowest amplitude since it is pointing away from the source. Keep in mind that the capsules are cardioid microphones. At 14200 Hz the relation of the capsule signals' amplitude and phase is such that by adding them up, they result in almost zero amplitude, which explains the very strong dip for a source at 45° azimuth in the horizontal plane as seen the the frequency response function in Figure 5.1.

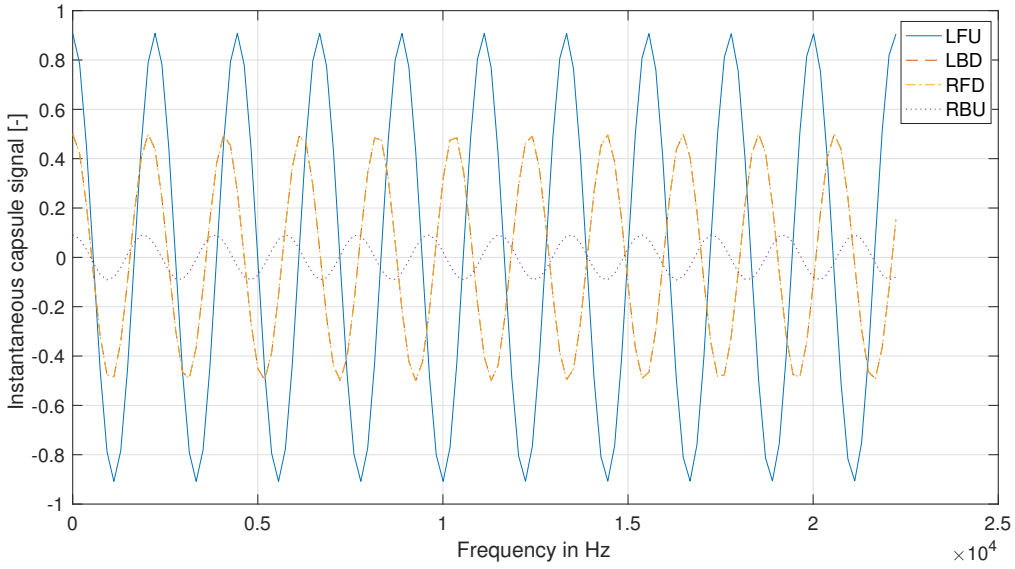
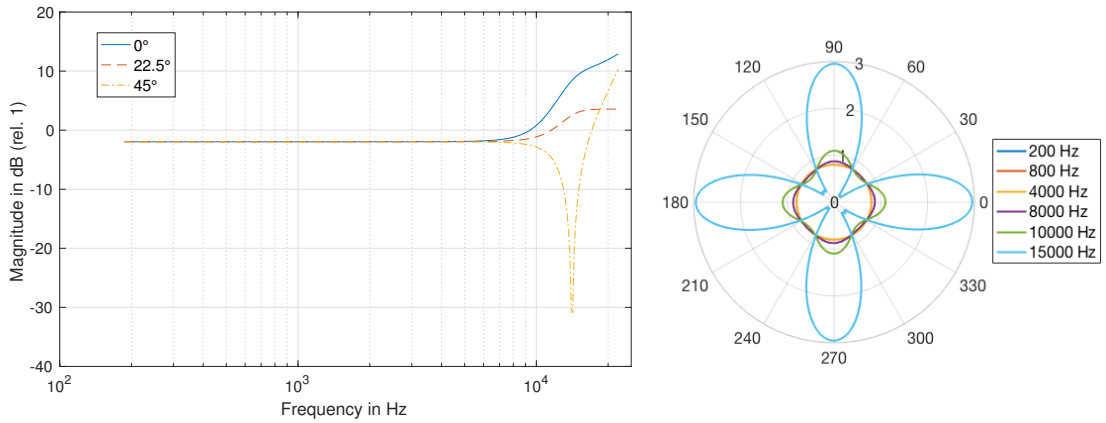


Figure 5.3: Instantaneous capsule signals at different frequencies for a source at 45° azimuth angle in the horizontal plane emitting all frequencies in phase

In Figure 5.4 the frequency response at different angles in the horizontal plane and the polar pattern for different frequencies are shown for the W signal corrected with the filter suggested by Gerzon [3]. It can be seen that the flat frequency response is extended up to 6000 Hz and the gain is reduced to -2 dB which is much closer to the desired gain of -3 dB than the gain without correction. Above the frequency range of the flat response the signal is amplified significantly by the filter for sources

at 0°. Since the unfiltered signal goes down to zero amplitude for sources at 45° at 14200 Hz, the filter does not have an influence around this frequency range. In the polar plot it can be seen that the omnidirectional pattern is still kept until 8000 Hz but derivations at higher frequencies are more extreme.



(a) Frequency response at different azimuth angles

(b) Polar pattern

Figure 5.4: W signal in the $z = 0$ plane, non-coincidence correction as suggested by Gerzon [3]

The above made observations on the filtered W signal can be explained by the frequency response of the applied filter for the W signal shown in Figure 5.5. The attenuation at lower frequencies and the rapidly increasing magnitude response with increasing frequency are visible.

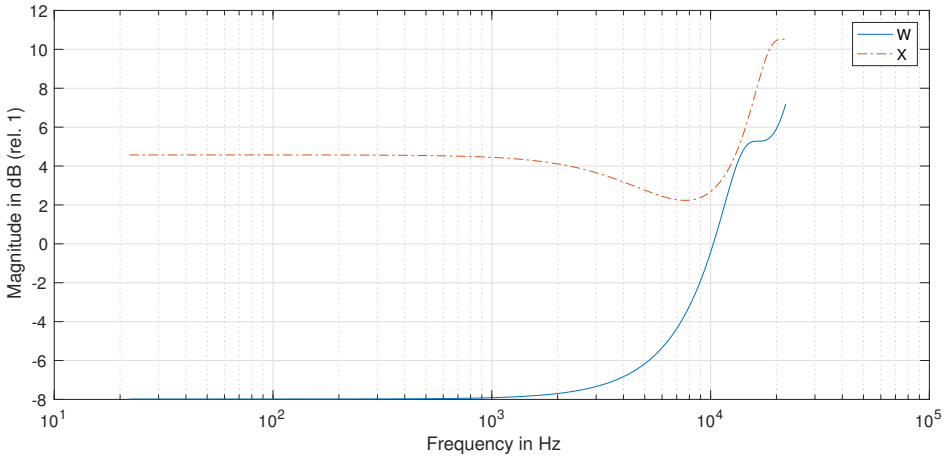
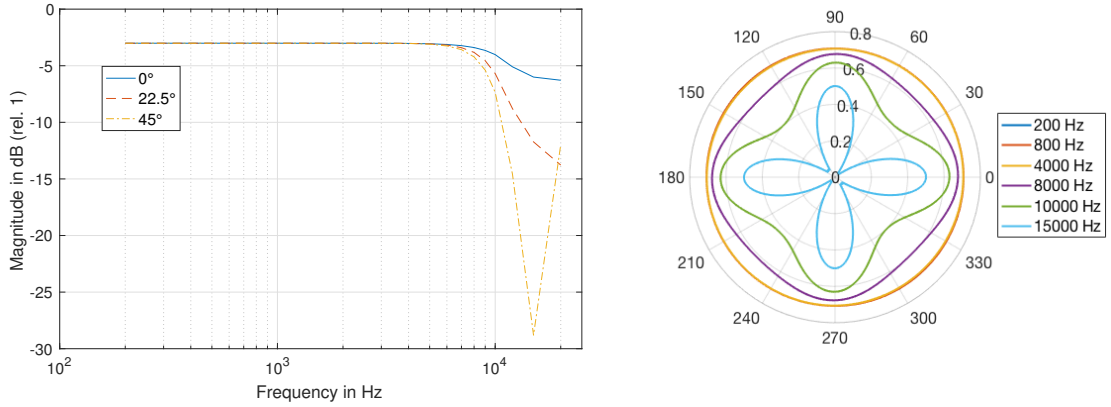


Figure 5.5: Frequency response of the filter suggested by Gerzon [3] for the omnidirectional and figure-of-eight signals

In Figure 5.6 the frequency response at different angles in the horizontal plane and the polar pattern for different frequencies are shown for the W signal corrected with the filter suggested by Faller et al. [10]. Also this filter extends the flat frequency

response to 6000 Hz. Furthermore, the gain is at the desired -3 dB in the flat frequency range. Above 6000 Hz the gain is decreasing less strongly than for the unfiltered signal. For the 45° source the dip at 14200 Hz cannot be mitigated by this filter either, for the same reason as for the other filter. In the polar plot we see again the omnidirectional shape up to 8000 Hz. The derivations above this frequency are less extreme than for the other filter. The observations can be explained by the frequency response of the filter presented in Figure 5.7.



(a) Frequency response at different azimuth angles

(b) Polar pattern

Figure 5.6: W signal in the $z = 0$ plane, non-coincidence correction as suggested by Faller et al. [10]

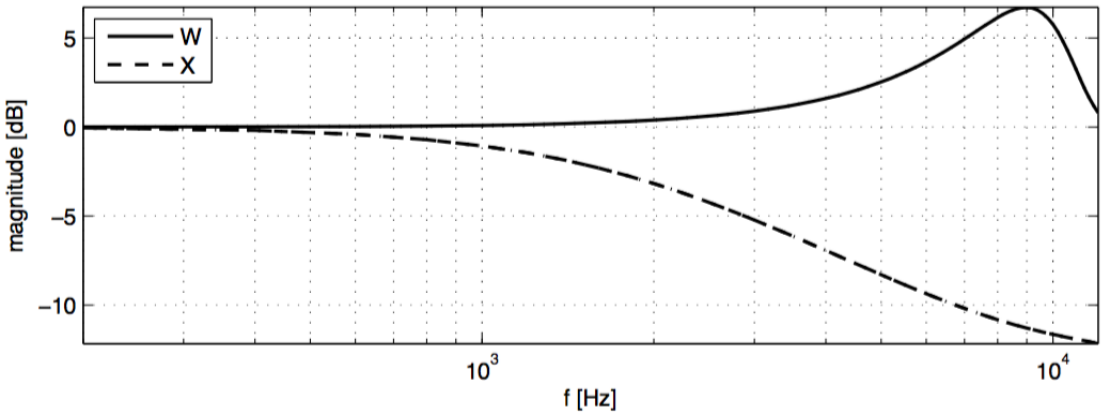


Figure 5.7: Frequency response of the filter suggested by Faller et al. [10] for the omnidirectional and figure-of-eight signals. Normalized gain. Source: [10]

In Figure 5.8 the 3-dimensional polar shapes of the W signal are shown for different frequencies in order to give an impression of what happens outside the horizontal plane, which was discussed above. It can be seen that the shapes differ already at 4000 Hz from the ideal shape, while in the $z = 0$, $y = 0$, and $x = 0$ plane the polar pattern stays close to the ideal shape. The shapes look the same for the unfiltered and filtered signals. The filters cannot correct the deviations of the three dimensional

shapes. When due to phase differences between the capsules, the W signal becomes very low in amplitude, the filter does not have much influence anymore.

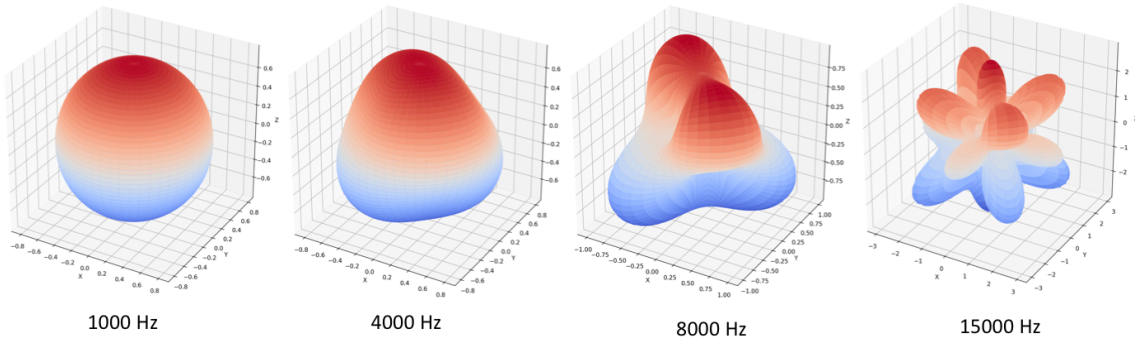
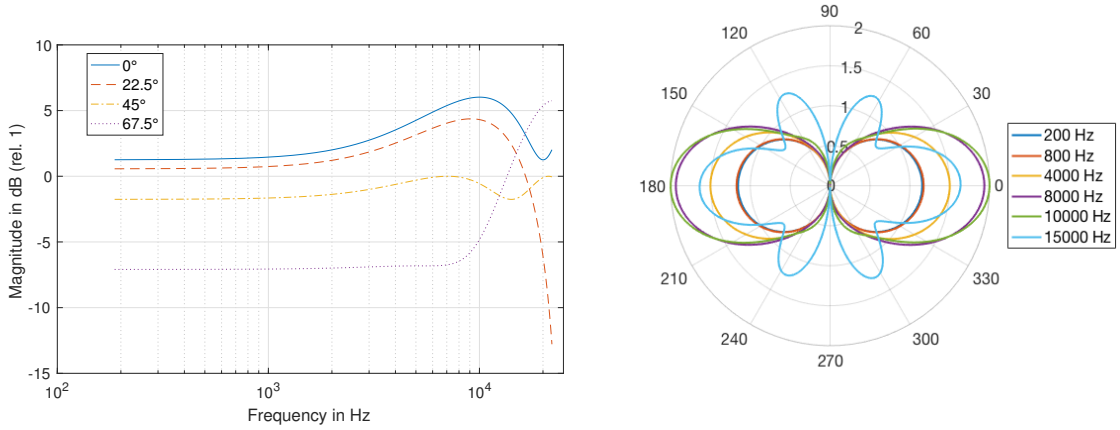


Figure 5.8: 3-dimensional polar pattern of the W signal for different frequencies

5.1.2 Magnitude of the Figure-Of-Eight Signals

In Figure 5.9 the frequency response and polar pattern for the X signal without any correction are shown. The frequency range with a flat response is smaller than for the W signal and the gain at low frequencies is 1 dB. Above 1000 Hz the gain increases, reaching a peak at 9900 Hz, which is where W has a dip. This can again be explained with the capsule signals and the way they are combined to obtain the X signal.



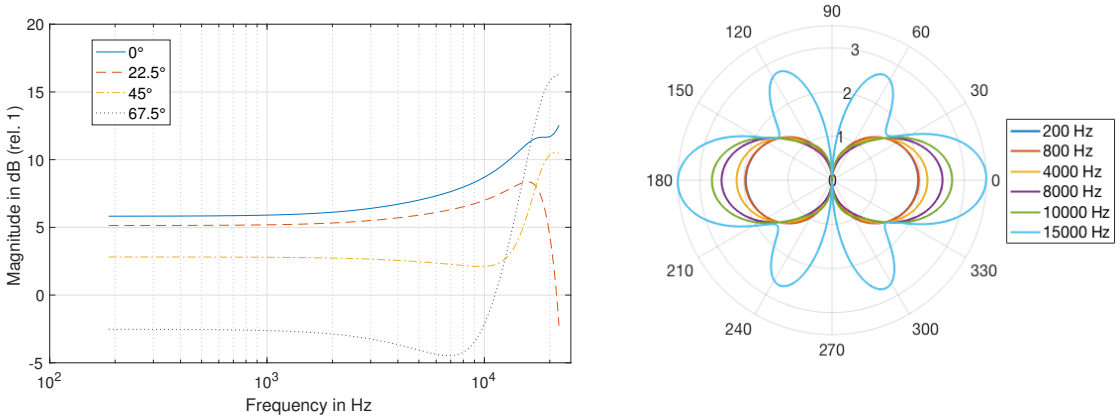
(a) Frequency response at different azimuth angles

(b) Polar pattern

Figure 5.9: X signal in the $z = 0$ plane, no correction

In Figure 5.2 it can be seen that the signals are in phase at low frequencies. Since the capsule signals LFU and RFD ("front" pair) are added and LBD and RBU ("back" pair) are subtracted, the resulting amplitude for X is lower than for W. The more the capsule pairs' signals "front" and "back" get out of phase, the higher the resulting gain for X because the "back" signals are subtracted from the "front" signals. The gain increase with increasing frequency is also visible in the polar plot in Figure

5.9. This increase is most prominent for sources at 0° azimuth which leads to a deformation of the figure-of-eight pattern. However, up to 10 kHz the directional information can still be captured very well in the horizontal plane because the gain decreases with increasing azimuth angle and becomes zero for 90° and 270°.
In Figure 5.10 the frequency response and polar pattern of the X signal corrected with the filter proposed by Gerzon [3] are shown. The flat frequency range is extended to approximately 2000 Hz and the peak which was present for all angles of incidence except 67.5ř was moved to higher frequencies. The signal level is much higher than for the unfiltered signal. The polar pattern has the same characteristics as before, but the derivations are more proportional to the frequency increase. At very high frequencies like 15000 Hz and above the gain becomes extremely high. These results are in line with the frequency response of the filter shown in Figure 5.5.

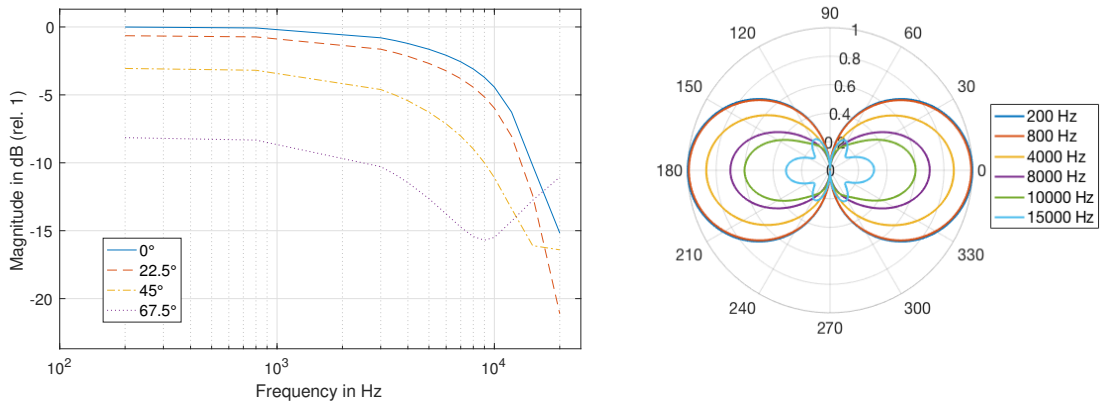


(a) Frequency response at different azimuth angles

(b) Polar pattern

Figure 5.10: X signal in the $z = 0$ plane, non-coincidence correction as suggested by Gerzon [3]

In Figure 5.11 the frequency response and polar pattern of the X signal corrected with the filter proposed by Faller et al. [10] are shown. The signal level has the desired 0 dB at low frequencies. The flat frequency response is not extended. Instead, the gain decreases with increasing frequency above 1000 Hz. However, in contrast to the raw and the Gerzon filtered X signals, the level decreases almost equally for all angles of sound incidence. This means that a frequency dependent gain correction can be applied to the signal and almost perfect figure-of-eight shapes with the same level can be obtained up to 10 kHz. This makes this filter the preferred theoretical choice as for the case of the W signal. The results are again in line with the frequency response of the applied filter depicted in Figure 5.7.



(a) Frequency response at different azimuth angles

(b) Polar pattern

Figure 5.11: X signal in the $z = 0$ plane, non-coincidence correction as suggested by Faller et al. [10]

The change of the 3-dimensional polar shapes of the X signal is similar to the W signal. It is shown in Figure 5.12. Again, up to 8000 Hz the shapes in the $z = 0$, $y = 0$, and $x = 0$ planes are rather stable while in other directions deviations from the ideal shape start already at 4000 Hz.

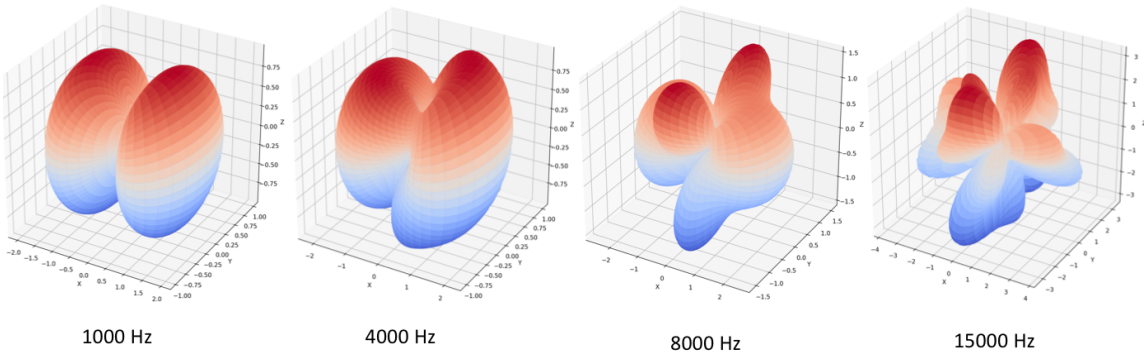
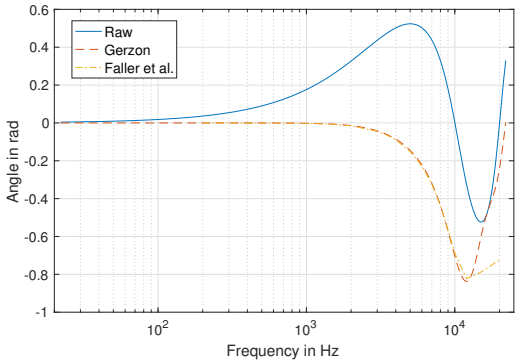


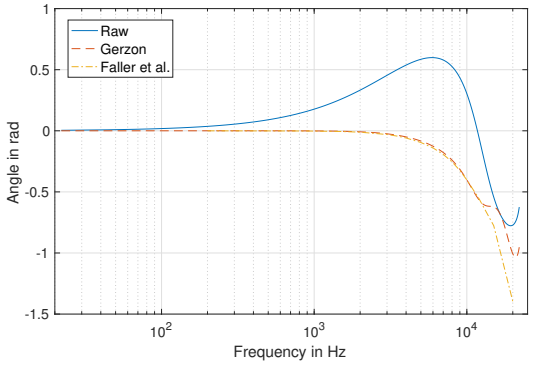
Figure 5.12: 3-dimensional polar pattern of the X signal for different frequencies

5.1.3 Phase Difference between the Signals

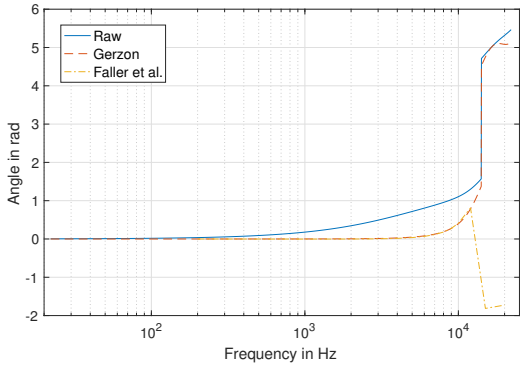
When B-format signals are decoded for example to auralize a recording, not only the magnitude but also the phase of the B-format signals is important in order to obtain the correct spatial impression. The ideal case would be that the B-format signals are in-phase. The phase difference between the W and the X signal for different angles of sound incidence is shown in Figure 5.13.



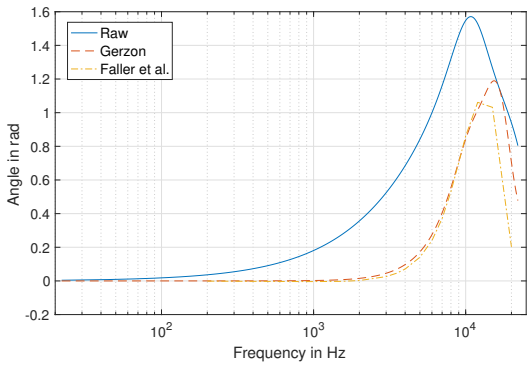
(a) 0° azimuth angle



(b) 22.5° azimuth angle



(c) 45° azimuth angle



(d) 67.5° azimuth angle

Figure 5.13: Phase difference between W and X for different angles of sound incidence in the horizontal plane

One can see that the filters reduce the phase difference between the signals and that both filters produce equally good results. However, for each angle of incidence there is a different frequency limit above which phase differences between the signals cannot be avoided by the filters.
The phase response of the filters suggested by Gerzon [3] and Faller et al. [10] are shown in Figures 5.14 and 5.15.

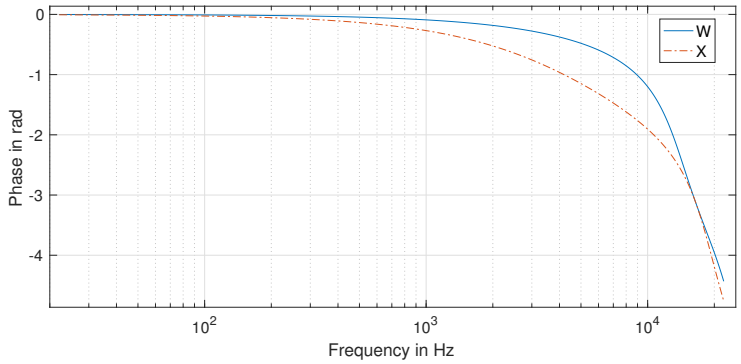


Figure 5.14: Phase response of the filter suggested by Gerzon [3] for W and X

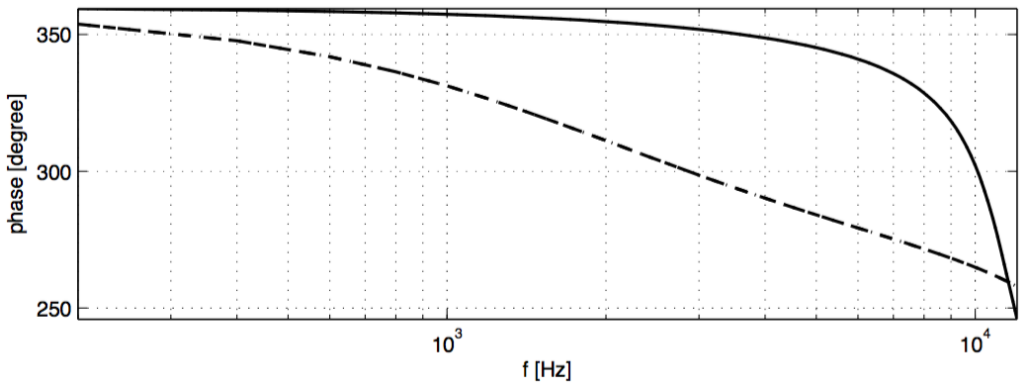


Figure 5.15: Phase response of the filter suggested by Faller et al. [10] for W and X. Source: [10]

5.1.4 Conclusion of the Simulation Results

It has been shown that the raw B-format signals deviate from the ideal characteristics and that the theoretical filters suggested in literature can correct the deviations in magnitude and phase to a certain extend. Due to its performance in the horizontal plane, which due to symmetry is the same in the $x = 0$ and $y = 0$ planes, the filter suggested by Faller et al. [10] yields the B-format signals, which are the closest to the desired ones. Phase differences between the signals cannot be avoided by the filters above approximately 8000 Hz. In general, the results suggest that recordings with a B-format microphone contain the best spatial information when the source is placed in the $z = 0$, $y = 0$, or $x = 0$ plane, where the polar shapes are more stable.

5.2 Microphone Measurement Results

The results of the performed measurements on the microphone are presented below. First, the results of the individual microphone capsules measurements are discussed. Then, the practical non-coincidence correction filters are presented and the unfiltered and filtered B-format signals are discussed.

5.2.1 Capsule Measurements

The recordings of the sweep signal done with the different capsules of the B-format microphone are shown in the frequency domain in Figure 5.16 (for details about the recording procedure see section 3.1). The figure shows that all the capsules already have a very similar response (the differences between 100 Hz and 300 Hz are discussed further below). The purpose of this first measurement was actually to test this and to develop and apply an equalization in case the capsule responses were not that equal. This is not necessary in this case, especially when considering the signals of the reference microphone shown in Figure 5.17, which would be used to design the correction filters. There, we see for the measurements of the capsules which are pointing downwards (RFD and LBD) an interference pattern that is stronger than for the measurements of the other capsules above 4000 Hz. This is most likely caused by

the body of the microphone which for the measurements of the downwards pointing capsules was positioned below the capsule under test pointing in the direction of the loudspeaker (cf. Figure 3.5(b)). Consequently, an equalization with these reference signals would not improve the response of the capsules.

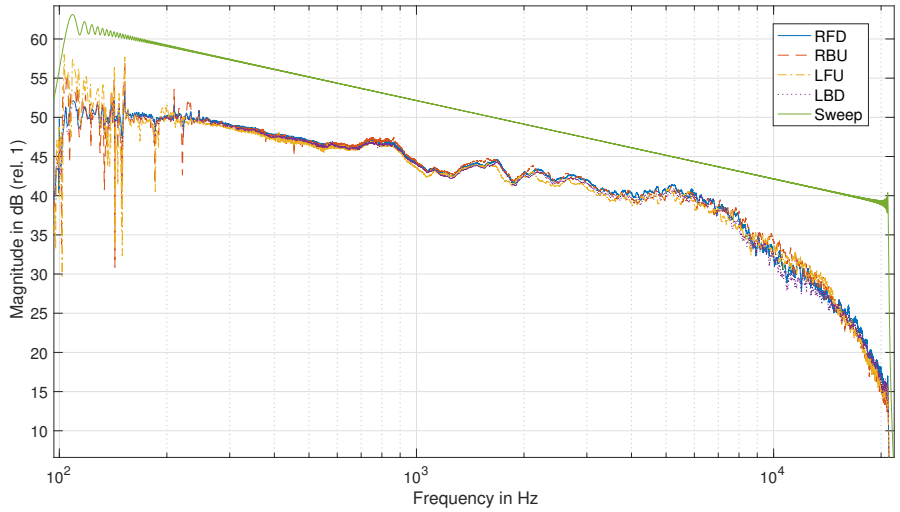


Figure 5.16: Recording of the sweep signal with the individual microphone capsules in the frequency domain

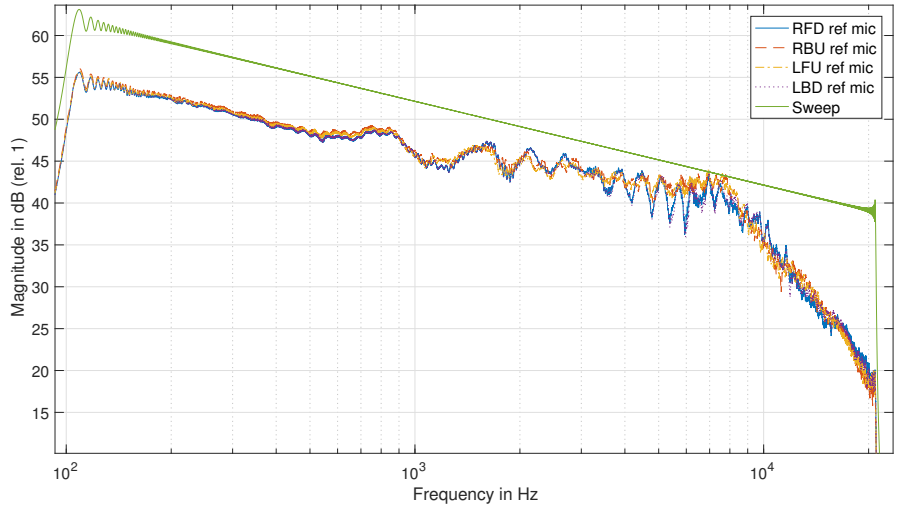
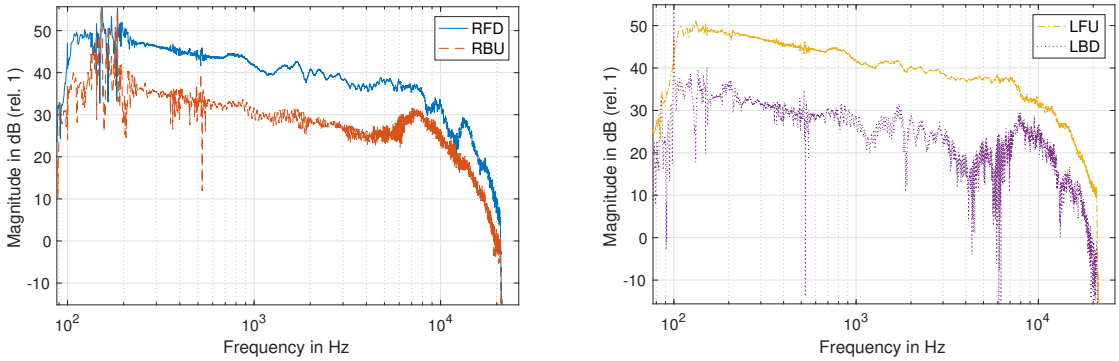


Figure 5.17: Recording of the sweep signal with the reference microphone in the frequency domain for the individual capsule measurements

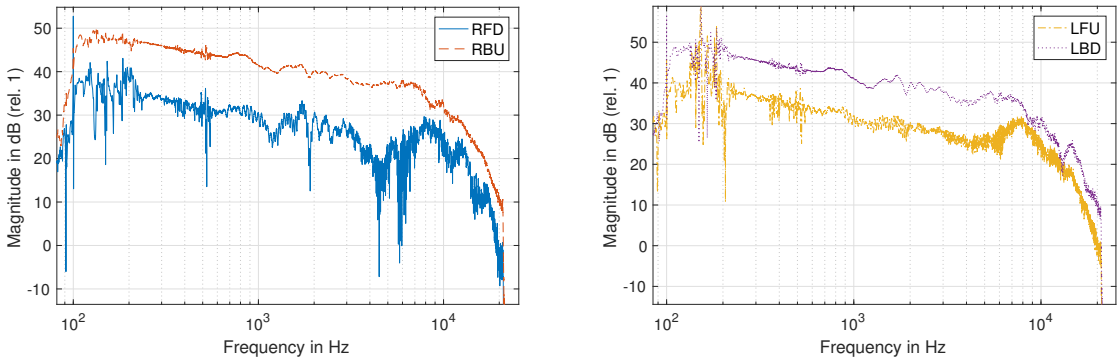
At low frequencies some irregularities are present in the capsule signals RBU and LFU shown in Figure 5.16. The question arises if this means that the capsules do not work as they are supposed to or if other factors are responsible for that. To test that, two more recordings have been done, where first the front and then the back of the

microphone faced the loudspeaker as explained in section 3.1 and sketched in Figure 3.6. The results shown in the Figures 5.18 and 5.19 indicate that the smoothness of a capsule signal in the frequency domain depends on its exact position during the recording and it is independent of the capsule itself. In the signals for the case "front towards loudspeaker" we see a rather smooth curve with some irregularities at low frequencies for the RFD capsule, the same but with a lower signal level for the RBU capsule since it looks away from the loudspeaker. The LFU signal is very smooth over the whole frequency range while the LBD signal has many dips over the whole range. Signals with the same characteristics are obtained for the case "back towards loudspeaker" but in this case not by the same capsules as before. Instead, the same signals are obtained in the same positions since each capsule is in the exact same position as another capsule was in before turning the microphone by 180°. The capsule pairs having the same signals before and after turning the microphone are RFD/LBD, RBU/LFU, LFU/RBU, and LBD/RFD.



(a) RFD and RBU **(b)** LFU and LBD

Figure 5.18: Capsule signals for the front of the B-format microphone pointing towards the loudspeaker

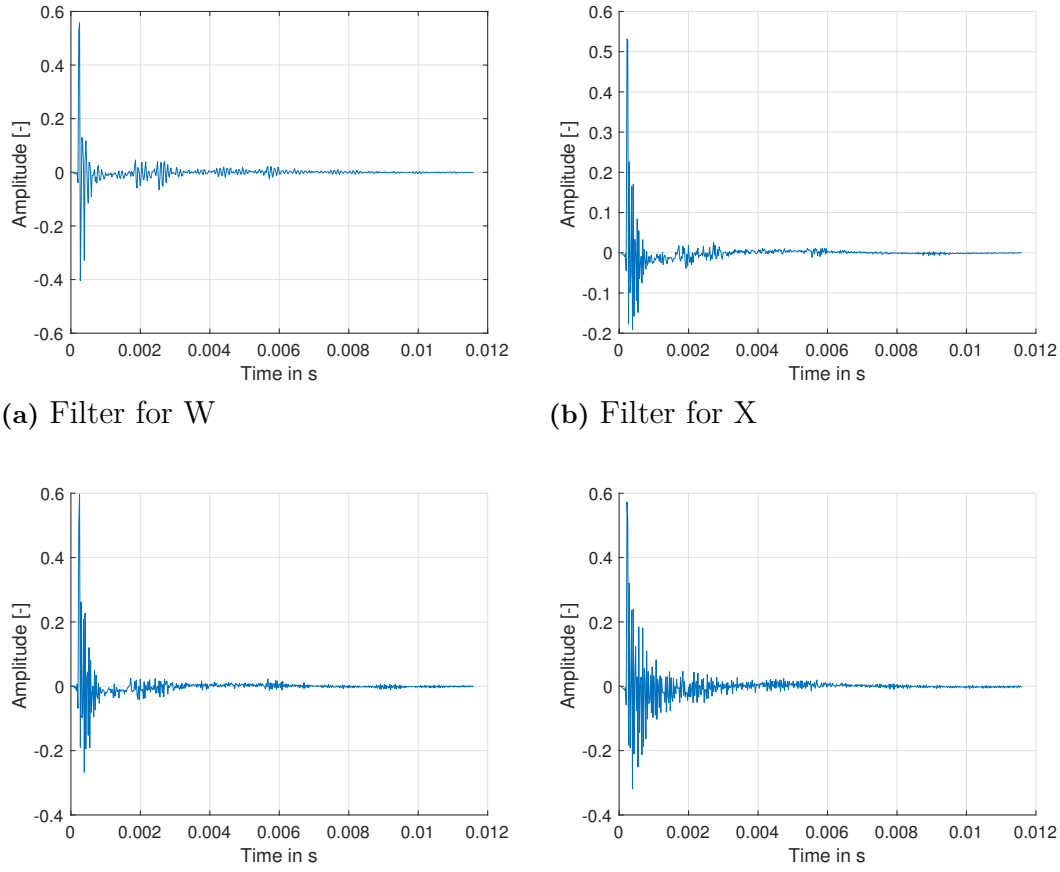


(a) RFD and RBU **(b)** LFU and LBD

Figure 5.19: Capsule signals for the back of the B-format microphone pointing towards the loudspeaker

5.2.2 Non-Coincidence Correction Filters

The non-coincidence correction filters obtained from the measurement data in the way described in section 3.2 are shown in the time domain in Figure 5.20. It can be seen that they are causal impulse responses. In Figure 5.21 all impulse responses are shown in the same plot so that one can see that all filters have the same delay. The frequency responses of the filters are shown in Figure 5.22. The influence of the bandpass filter is visible and that the filters do not alter the signals much between 200 Hz and 1000 Hz, which makes sense because in this frequency range the capsules can be considered coincident.



(a) Filter for W

(b) Filter for X

(c) Filter for Y

(d) Filter for Z

Figure 5.20: Impulse responses of the non-coincidence correction filters

In Figure 5.23 the transfer functions of the B-format signals, which are basically their frequency responses, are shown for the case with and without non-coincidence correction. We can see the dip at 10000 Hz in the not corrected W signal due to phase differences between the capsule signals, which was also visible in the simulation results. In the figure-of-eight signals we can see the expected peak around 10000 Hz. These irregularities are evened out by the applied filters as can be seen in the frequency responses after the correction. In the Z signal more peaks and dips than for the other signals are present in the uncorrected spectrum. They are most

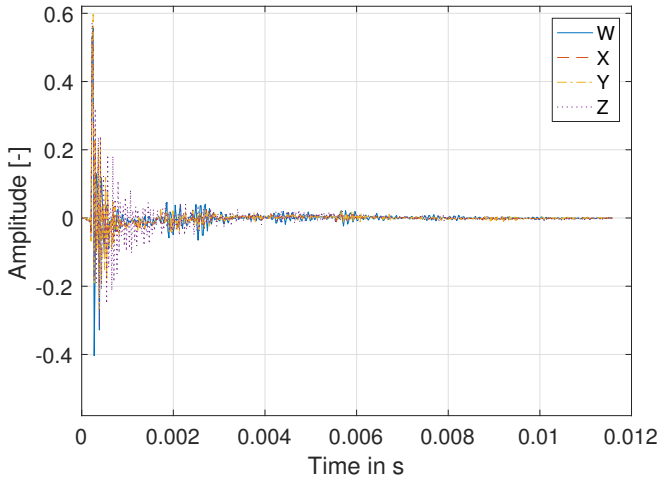


Figure 5.21: Impulse responses of all four non-coincidence correction filters in one plot

likely caused by reflections at the round bottom plate below the capsules, on which a protector can be screwed on. The microphone position for the corresponding measurement was shown in Figure 3.7(b). Also these disturbances are equalized by the filter, which is useful since they will always be present in the Z signal for sound waves impinging from above the microphone. However, since the same correction filter will always be applied on the Z signal, it might be altered unfavourably for source positions for which these reflections are not present.

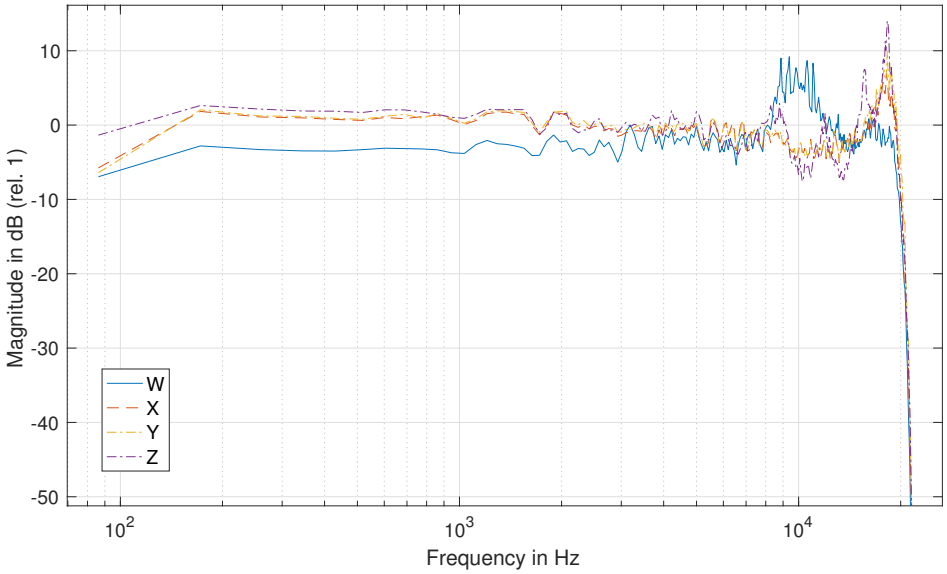
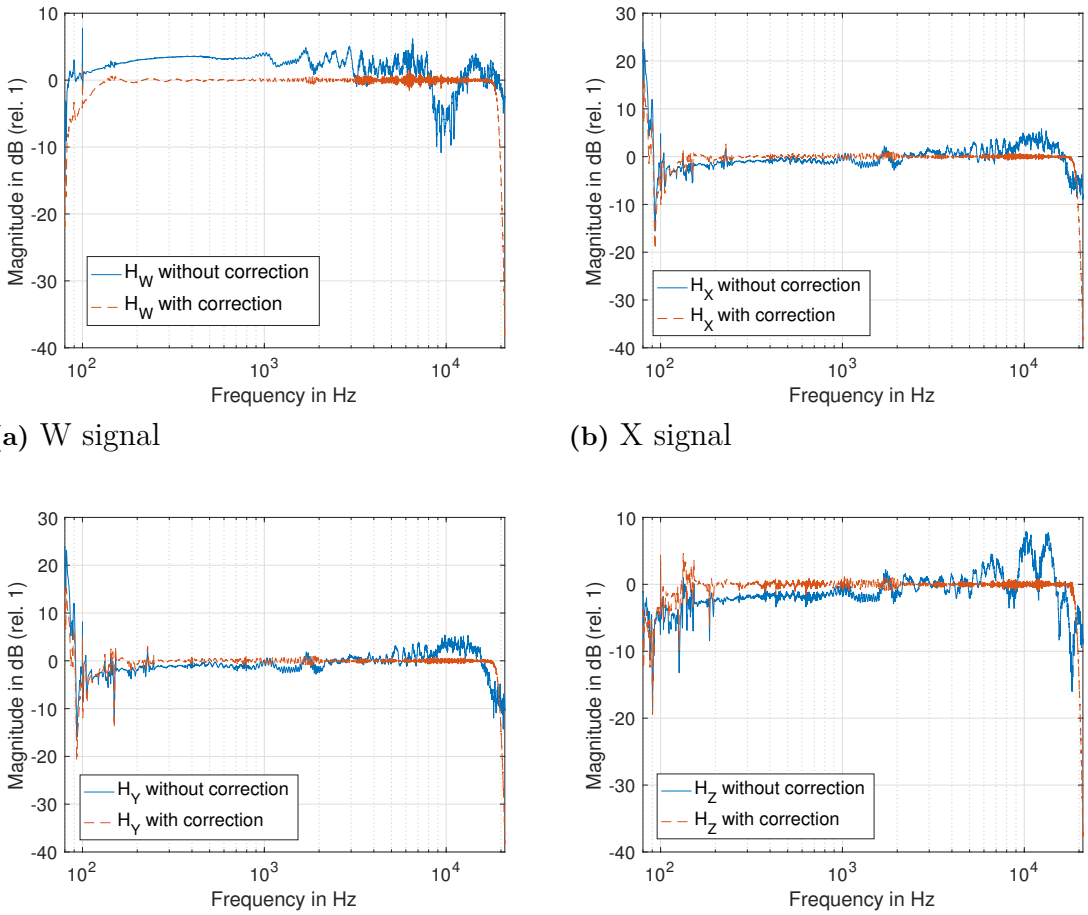


Figure 5.22: Frequency response of all four non-coincidence correction filters

It can also be seen that after the correction all signals have the same gain. Thus, one can use them directly to save the recording in the AmbiX-Format. If one wishes to use the FuMa format, the W signal needs to be attenuated by 3 dB.

After the correction, the B-format signals have a flat frequency response. However, common microphones for music recordings do not have an entirely flat frequency response. Some alterations are found to be favourable for a good sounding recording. Those alterations could be added with another filter if desired when the B-format microphone is used for audio recordings. If the microphone is used for a room acoustics measurement instead, the flat frequency response is desired.



(a) W signal

(b) X signal

(c) Y signal

(d) Z signal

Figure 5.23: Comparison of the frequency response of the B-format signals with and without non-coincidence correction

5.2.3 Conclusion of the Microphone Measurement Results

From the measurements we learned that for high quality B-format microphones the capsule equalization as suggested by Farina [4] might not be necessary. We also saw that reflections from the microphone body alter the capsule signals under anechoic conditions depending on the relative position of sound source and microphone. When recordings are done with a B-format microphone in enclosed spaces with reflective walls even more reflections will impinge on the body and alter the signals even more. This leads to a limited spatial resolutions of B-format recordings "in the field" for mid to high frequencies and should be kept in mind when decoding the signals.

Furthermore, it became clear that the capsule signals are very sensitive to their exact position relative to a sound source, also at low frequencies around 200 Hz. The frequency response of the B-format signals, that were simulated, has also been observed in the measured signals. Practically applicable non-coincidence correction filters have been designed and it was proven that they have the desired effect of flattening the frequency response of the B-format signals. It has also been shown that there are limitations to these filters since they are derived from only one measurement per B-format signal and may not suit all angles of sound incidence equally well.

5.3 Absorption Measurement Results

In this section, the results of the investigation to which extend a B-format microphone can be used for in-situ absorption measurements are presented. First, the results of the simulation are shown, then the results of the actual measurement are presented.

5.3.1 Simulation

At first, the influence of the distance of the microphone to the test material h was tested. In Figure 5.24 the results for the calculation of the reflection coefficient r for different values of h are presented. The true reflection coefficient was $0.3 - 0.3j$ and the distance from the source to the test material was $h_s = 1$.

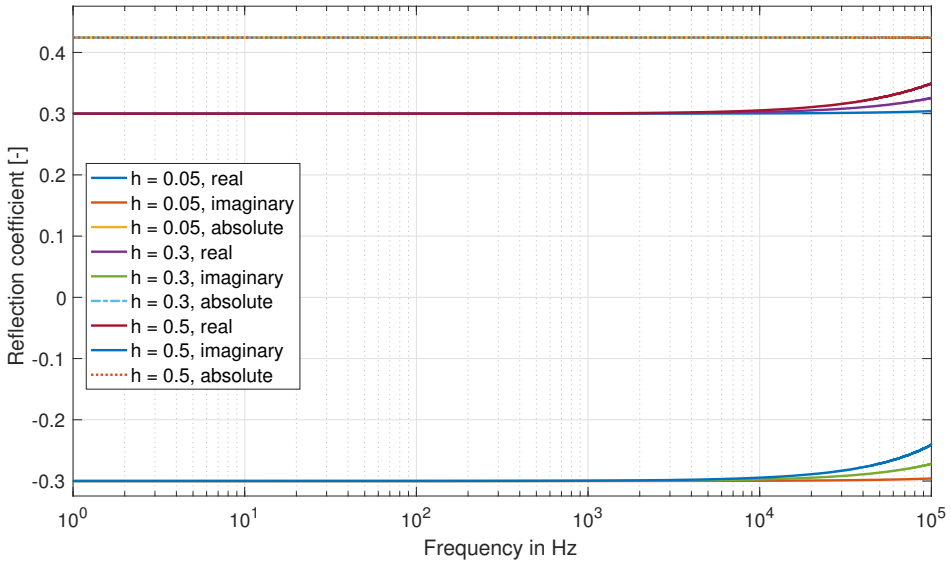


Figure 5.24: Simulation of absorption measurement for different distances h

We can see that the bigger the distance h, the more deviate the real and imaginary part of the obtained reflection coefficient from the true value at very high frequencies above 10 kHz. However, they deviate in a way so that the obtained absolute value

of r is correct for all frequencies. It is this absolute value we need to obtain the absorption coefficient, so there is basically no upper frequency limit for this method in theory. These results may be somewhat unexpected because it was shown before that the B-format signals become strongly altered at higher frequencies. However, it was also shown in the theory chapter that frequency dependent gain change factors for the W and X signals do not matter and are cancelled out in the calculation of r. This also holds for high frequencies with strong gain alterations.
To investigate if the measurement can nevertheless be disturbed, the influence of uneven capsule gains was investigated. It can be assumed that real world microphones are less perfect than the simulated one. Therefore, the capsule signals LFU, LBD, RBU, and RFD were multiplied with the factors 0.99, 1.1, 0.98, and 1.02 respectively. Then the reflection coefficient was calculated again. For this experiment the true reflection coefficient was again $0.3 - 0.3j$ and $h = 0.1$ and $h_s = 1$. The results are presented in Figure 5.25.

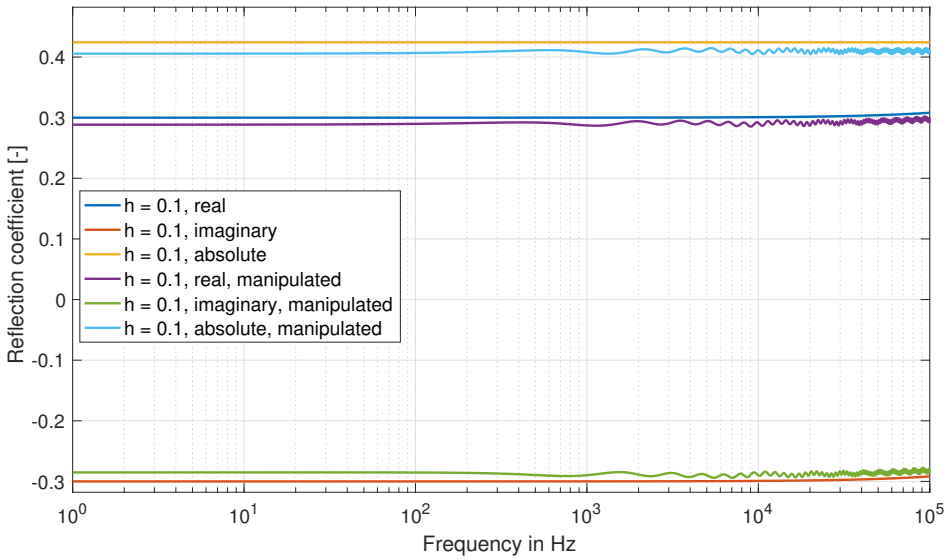


Figure 5.25: Simulation of absorption measurement manipulated capsule signals

It can be seen that the manipulations of capsules add ripples to the curves and make the obtained values deviate from the true one. This means, that it is important to have as equal as possible capsule gains for this measurement method. But we have also seen in the previous section that some capsule signals can easily get altered, for example by reflections from the microphone body.
Different values for the distance h_s and the true reflection coefficient did not have any influence on the results.

5.3.2 Measurement

In Figure 5.26 the results of the absorption measurements with the B-format microphone and the impedance tube are presented. It can be seen that both measurement methods give results that follow roughly the same trend.

Between 100 Hz and 200 Hz the results of the B-format method are not usable, the X signals are very noisy in this range and disturb the graph for alpha. Between 200 Hz and 300 Hz the methods give similar results having a difference of approximately 0.05. Between 300 Hz and 1200 Hz the absorption coefficient values of both methods have a small dip and then increase rapidly with increasing frequency. However, the dips are not at the same frequency and the slopes of the increasing curves are not identical either. Above 1200 Hz the curve of the impedance tube does align very well with the peaks of the B-format microphone curve. Thus, it can be concluded that above 1200 Hz the B-format microphone methods gives reliable results. Unfortunately, the impedance tube had an upper frequency limit of 2000 Hz. The results of the simulation suggest that the ripples in the curve of the B-format microphone result from slightly disturbed capsule signals. The gain of the pre-amp channels was corrected as explained above.
These results can be seen as a proof of concept for the idea of measuring absorption in-situ with a B-format microphone. However, more research has to be done on this topic to develop a more reliable and robust method. Measurements in different acoustical environments would give more insight into the applicability of the method. Also, new reference measurements in a larger frequency range should be done. We saw also that these measured results for the absorption coefficient are less accurate than the simulation results. This is due to limitations like slight differences between the capsules, not perfectly plane waves, and an environment that is not perfectly anechoic due to all the equipment in the room.

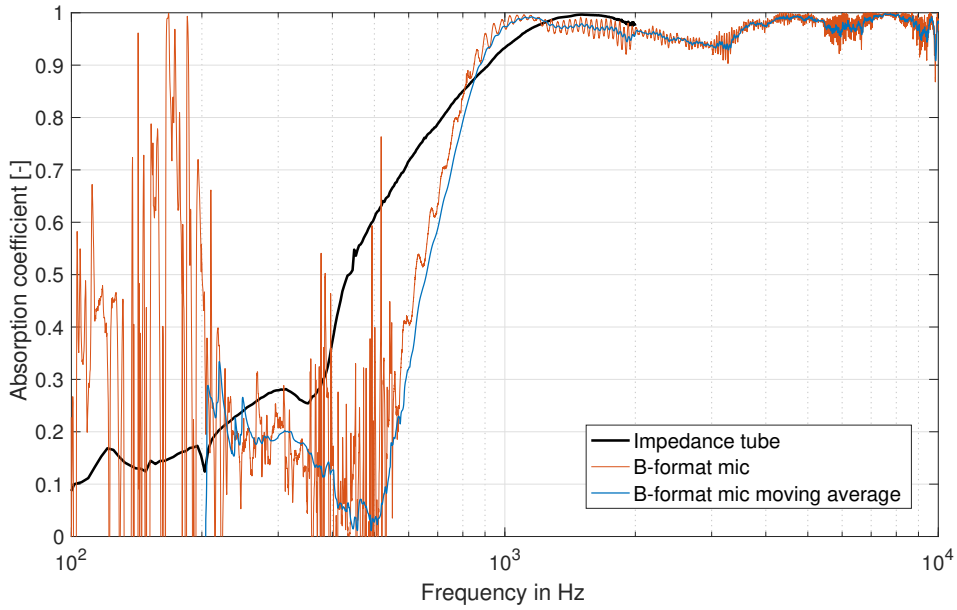


Figure 5.26: Comparison of the absorption coefficient measurement results of the B-format microphone measurement and the impedance tube measurement

6
Conclusion

This thesis has dealt with the three topics computation and filtering of B-format signals, auralization of B-format signals as well as absorption measurements with B-format microphones. In all of these three areas, the basis for further research has been provided. Thus, this work is rather broad, exploring different aspects of Ambisonics, than going in depth into one of these aspects. This is, however, recommended as a next step.
Regarding the B-format signals it has been shown that they are most robust in the three planes formed by the Cartesian axes. Their polar shapes will deviate from the ideal ones above 4000 Hz and phase differences between the signals cannot be avoided above 8000 Hz, not even in theory. In practice, simpler filters based on one measurement per B-format signal can be designed and they flatten the frequency response of the signals as desired. However, they are optimized for sound incidence along the axis on which the respective measurement was done. Furthermore, depending on the microphone design, reflections from the body and other parts will alter the capsule signals and thereby disturb the B-format signals. Further research could be done on the frequency limit up to which the B-format signals need to have a flat frequency response, be in phase or have the ideal polar shapes in order to provide enough spatial information for different use cases.
With respect to the issue of auralization of B-format recordings, a binaural real-time renderer including head tracking has been built. It is considered a first version on which future work can be based on at the Division of Applied Acoustics at Chalmers University of Technology. It works as it should but there are some parts that can be improved. For example, the reduction of the minimum block length, making the reading of the head tracker independent from the reading of the music file, improving the sound quality, and considering head rotations along all three axes. Finally, the renderer needs to be evaluated by listening tests.
It has been shown, that, in principle, the measurement of absorption with a B-format microphone is possible. More research can be done to improve the method. For example, testing the performance in reverberant environments, investigate the requirements on the loudspeaker and the distance between loudspeaker and microphone in order to create a portable device for this measurement method similar to the Microflown device.

Bibliography

[1] Craven, P., Gerzon, M. (1977). Coincident microphone simulation covering three dimensional space and yielding various directional outputs. *United States Patent*. US 4,042,779.

[2] Farrar, K. (1979). Soundfield microphone. *Wireless World*, 85(1526), 48-50.

[3] Gerzon, M. A. (1975). The design of precisely coincident microphone arrays for stereo and surround sound. In *Audio Engineering Society Convention 50*. Audio Engineering Society.

[4] Farina, A. (2006). A-format to B-format conversion. Retrieved from http://pcfarina.eng.unipr.it/Public/B-format/A2B-conversion/A2B.htm

[5] Facebook360 (2018). https://facebook360.fb.com

[6] Tijs, E. (2013). Study and development of an in situ acoustic absorption measurement method.

[7] Microflown Technologies (2018). In-situ absorption setup. Retrieved from http://www.microflown.com/products/solutions/in-situ-absorption-setup.html

[8] Woszczyk, W., Iwaki, M., Sugimoto, T., Ono, K., & de Bree, H. E. (2006). Anechoic measurements of particle-velocity probes compared to pressure gradient and pressure microphones. *The Journal of the Acoustical Society of America*, 120(5), 3233-3233.

[9] Adriaensen, F. (2007, March). A tetrahedral microphone processor for ambisonic recording. In *Proc. of the 5th Int. Linux Audio Conference (LAC07), Berlin, Germany* (pp. 64-69).

[10] Faller, C., & Kolundzija, M. (2009). Design and limitations of non-coincidence correction filters for soundfield microphones. In *Audio Engineering Society Convention 126.* Audio Engineering Society.

[11] Batke, J. M. (2009). The B-format microphone revised. *Ambisonics Symposium 2009, Graz*, 2(35.26), 9.

[12] Core Sound (2018). Core Sound TetraMic. Retrieved from http://core-sound.com/TetraMic/1.php

[13] The A-Format is the signals from four microphones on the faces of a tetrahedron.(2006) Retrieved from https://en.wikipedia.org/wiki/Soundfield_microphone#/media/File:Tetrahedron.png

[14] Core Sound (2018) TetraMic specifications. Retrieved from http://core-sound.com/TetraMic/2.php

[15] Sennheiser (2018). AMBEO VR MIC. Retrieved from https://en-uk.sennheiser.com/microphone-3d-audio-ambeo-vr-mic

[16] Weisstein, E. W. (2018). Colatitude. From *MathWorld – A Wolfram Web Resource*. Retrieved from http://mathworld.wolfram.com/Colatitude.html

[17] Schönefeld, V. (2005). Spherical harmonics.

[18] Nachbar, C., Zotter, F., Deleflie, E., & Sontacchi, A. (2011, June). Ambix-a suggested ambisonics format. In *Ambisonics Symposium, Lexington.*

[19] Malham, D. (2003). Higher order ambisonic systems. *abstracted from 'Space in Music-Music in Space' an Mphil thesis by Dave Malham, submitted to the University of York in April 2003.*

[20] Ahrens, J. (2012). Analytic methods of sound field synthesis. *Springer Science & Business Media.*

[21] Zotter, F. (2009). Analysis and synthesis of sound-radiation with spherical arrays.

[22] Farina, A., & Fausti, P. (1994). Standing wave tube techniques for measuring the normal incidence absorption coefficient: comparison of different experimental setups. In *Proc. of 11th. International FASE Symposium.*

[23] Jiříček, O., Švec, P., Jandák, V., & Brothánek, M. (2009). Comparison of sound absorption measurement methods. In *Proceedings of the Inter-Noise 2009, Ottawa.*

[24] Alvarez, J. D., & Jacobsen, F. (2008). An iterative method for determining the surface impedance of acoustic materials in situ. In *37th International Congress and Exposition on Noise Control Engineering.*

[25] Müller-Trapet, M., Dietrich, P., Aretz, M., van Gemmeren, J., & Vorländer, M. (2013). On the in situ impedance measurement with pu-probes—Simulation of the measurement setup. *The Journal of the Acoustical Society of America, 134(2)*, 1082-1089.

[26] Sjösten, P. (2015). Electroacoustics lecture notes. *Chalmers University of Technology*

[27] Errede, S. (2016). Absolute calibration of pressure and particle velocity microphones. *Dept. of Physics, The University of Illinois at Urbana-Champaign*

[28] Jacobsen, F., & De Bree, H. E. (2005). Measurement of sound intensity: pu probes versus pp probes. *Proceedings of NOVEM.*

[29] Spatialaudio (2016). Retrieved from https://github.com/spatialaudio/jackclient-python/blob/master/examples/play_file.py

[30] Jack Audio Connection Kit. http://jackaudio.org

[31] Bartz, P. (2016). Building an AHRS using the SparkFun "9DOF Razor IMU" or "9DOF Sensor Stick". Retrieved from https://github.com/Razor-AHRS/razor-9dof-ahrs/wiki/tutorial

[32] Institut für Schallforschung der Österreichischen Akademie der Wissenschaften (2014). HRTF-Datenbank des ISF. Retrieved from https://www.kfs.oeaw.ac.at/index.php?view=article&id=608lang=de

[33] Index of /data/database/ari (2014). Retrieved from http://sofacoustics.org/data/database/ari/

[34] Genelec (2007) Data Sheet Genelec 8030A Active Monitoring System. Retrieved from https://www.genelec.com/documents/datasheets/DS8030a.pdf

[35] Suhanek, M., Jambrosic, K., & Horvat, M. (2008, September). A comparison of two methods for measuring the sound absorption coefficient using impedance tubes. In *ELMAR, 2008. 50th International Symposium* (Vol. 1, pp. 321-324). IEEE.