

Article

Stereophonic Microphone Array for the Recording of the Direct Sound Field in a Reverberant Environment

Jonathan Albert Gößwein *, Julian Grosse and Steven van de Par

Acoustics Group, Cluster of Excellence "Hearing4All", Carl von Ossietzky University, 26111 Oldenburg, Germany; julian.grosse@uni-oldenburg.de (J.G.); steven.van.de.par@uni-oldenburg.de (S.v.d.P.)

* Correspondence: jonathan.goesswein@uni-oldenburg.de; Tel.: +49-441-798-3248

Academic Editors: Woon-Seng Gan and Jung-Woo Choi

Received: 15 March 2017; Accepted: 17 May 2017; Published: 24 May 2017

Abstract: State-of-the-art stereo recording techniques using two microphones have two main disadvantages: first, a limited reduction of the reverberation in the direct sound component, and second, compression or expansion of the angular position of sound sources. To address these disadvantages, the aim of this study is the development of a true stereo recording microphone array that aims to record the direct and reverberant sound field separately. This array can be used within the recording and playback configuration developed in Grosse and van de Par, 2015. Instead of using only two microphones, the proposed method combines two logarithmically-spaced microphone arrays, whose directivity patterns are optimized with a superdirective beamforming algorithm. The optimization allows us to have a better control of the overall beam pattern and of interchannel level differences. A comparison between the newly-proposed system and existing microphone techniques shows a lower percentage of the recorded reverberance within the sound field.

Keywords: intensity-based stereo-recording; convex optimization; superdirective beamformer; white noise gain; logarithmic array design; spatial audio

1. Introduction

Sound reproduction systems play an important role in our everyday life. They allow us to listen to recordings from a different place and a past time. Many different methods for the recording and playback of sound exist, utilizing different combinations of microphone and loudspeaker setups. The most common one is a simple stereo reproduction, but there are more complex reproduction techniques, such as wave field synthesis [1] or ambisonics [2]. Even though the state-of-the-art methods achieve a very good accuracy in reproducing sound fields, they do not consider the interaction between the acoustics of the recording and playback environment. In particular, extra reverberation is created by the playback environment, and in addition, there is no control over the spatial distribution of the reverberant sound field, which may influence the apparent source width and perceived listener envelopment. For this reason, ongoing investigations aim to improve the performance of these methods.

In particular, Grosse and van de Par proposed a new way of recording and playing back sound fields [3]. The main idea behind their research was to record the direct and reverberant sound field separately in order to be able to render it in a playback room while optimizing certain perceptually-motivated criteria for the authentic audio reproduction. These criteria aim for recreating the reverberant sound field in the playback environment as faithfully as possible by optimizing the amount and spectral shape of the reverberation, as well as the interaural cross-correlation created by the reproduced reverberant sound field, such as it is created in the reproduction room, including its added reverberant effect. In their paper, Grosse and van de Par assumed that optimizing these perceptual criteria is sufficient for an authentic reproduction of the sound field present in the recording room,

which is created by a single source. This claim was supported by subjective evaluations. The playback and recording configuration can be seen in Figure 1. In addition to the two basic stereo loudspeakers, the proposed approach used two dipole loudspeakers to excite and equalize the reverberant sound field. For the optimized rendering, the system relies on the presence of a relatively dry direct signal to be rendered on the frontal loudspeakers and a reverberant signal to be optimized and rendered on the dipole loudspeakers. To record the direct sound, a microphone (C) was positioned close to the sound source. This also avoided early reflections, which could cause a change in coloration [4,5]. For recording the reverberant sound field, two microphones (B^l, B^r) were placed at two distant positions in the diffuse field.

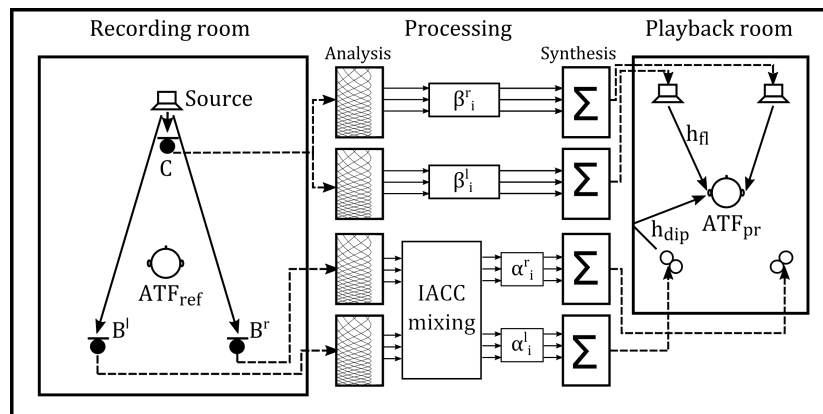


Figure 1. Recording and playback configuration with a processing stage in between to maintain the acoustical perception of a recording room. The microphone (C) records the direct sound, which is played later by two conventional loudspeakers, whereas the two microphones (B^l) and (B^r) record the reverberant sound field, which is played later by two dipole loudspeakers. Figure reproduced with permission from [3], Copyright IEEE, 2015.

Since the method of Grosse and van de Par [3] until now is limited to a single source and only records the direct sound field with one microphone, an extension is needed to also represent the spatial distribution of sources within the direct sound field signals as perceived at the listener position. Although this could in principle be achieved by using multiple close microphones and an appropriate mixing scheme, in this contribution, we want to provide a method with only a single ‘true-stereo’ microphone setup that is placed at the intended listener position within the recording room. Particular attention has to be paid to reduce the reverberant sound field in the direct sound field signals to be able to separately optimize the rendering of the direct and reverberant sound fields according to perceptual criteria within the playback room [3].

Although the specific design criteria for the proposed microphone array are envisioned to be used in the audio reproduction system of Grosse and van de Par [3], it can also be considered to use the proposed microphone array to record a relatively dry spatial image of the sound sources on stage to be combined with a reverberant track that can be mixed at a level that the recording engineer deems suitable. In this case, however, it will not necessarily fulfill the optimization criteria as formulated in Grosse and van de Par [3] that create a faithful audio reproduction.

The state-of-the-art true stereo systems combine two microphones with a characteristic directivity pattern, placed at different distances and under different angles relative to one another. Depending on these parameters, a deviating spatial rendering of the distributed sources can be observed [6]. Despite this, for use in the method proposed by Grosse and van de Par [3], these systems have some disadvantages that make them unsuitable to be implemented in this specific sound reproduction system because there is a high percentage of recorded reverberant sound, which should be avoided in the system of [3].

We overcome these disadvantages with the development of a new method of a true stereo microphone array, using a superdirective beamforming algorithm that is applied on two logarithmically-spaced microphone arrays. Correct, frequency-dependent interchannel level differences are captured by optimizing the shape of the two main lobes of the arrays. Together, they create the proper interchannel level difference required for an accurate spatial reproduction of the sound field while ensuring that no interchannel phase differences occur that can result in unintended changes in the perceived location of sound sources. Additionally, an optimal side lobe suppression is applied to reduce the influence of the reverberant sound field on the recording of the direct sound. This proposed stereo microphone array is compared to the state-of-the-art stereo microphone configurations mentioned earlier that shows a clearly reduced level of the reverberant sound field.

2. Methods

The following section is divided into five parts. The first Section 2.1 gives a brief introduction to the most relevant theory on beamforming needed for our proposed method. Section 2.2 focuses on the issue of the robustness of beamforming algorithms. The desired directivity pattern is specified in Section 2.3, which is based on a stereo intensity-panning rule related to the auditory processing of the interaural level differences. Section 2.4 introduces an optimal array design to suppress side lobes and, in this way, reduce the influence of the reverberant sound field on the recording of the direct sound. Further, a specific filter design is proposed in Section 2.5, which will be used and evaluated throughout this study. The design is based on a superdirective beamforming algorithm and describes how the directivity pattern that is specified in Section 2.3 can be used for the optimization.

2.1. Beamforming

Beamforming describes the process of forming the directivity pattern of several microphones, which are arranged into an array, with signal processing techniques to obtain a specific, frequency-dependent directivity pattern. The directivity pattern $b(f, \phi)$ of a linear discrete microphone array, consisting of N microphones, is calculated as follows [7]:

$$b(f, \phi) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f) G_n(f, \phi) \tag{1}$$

where ϕ denotes the angle ranging from $-\pi$ to π , f the frequency, $w_n(f)$ the frequency-dependent complex weighting filtering applied to microphone n and $G_n(f, \phi)$ the steering vector denoting the direction and frequency-dependent transfer function from the sound source to microphone n . Such a microphone array is illustrated in Figure 2.

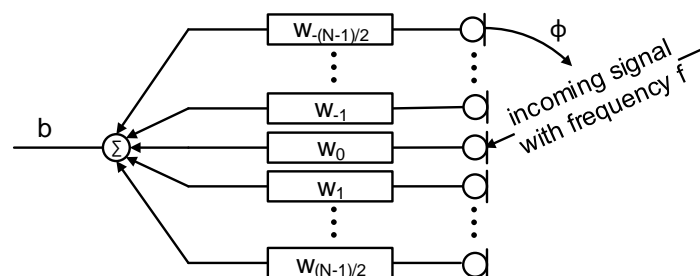


Figure 2. Microphone array receiving a signal with frequency f and angle of incidence ϕ . The incoming wavefront is captured with a microphone n , modified with the respective filter w_n and, at the end, summed up to form the directivity pattern $b(f, \phi)$.

Assuming a far field condition with the microphones that have an omnidirectional directivity pattern, the transfer function states:

$$G_n(f, \phi) = e^{-i\frac{2\pi f}{c}x_n \cos(\phi)} \tag{2}$$

where c is the speed of sound and x_n represents the distance of the n -th microphone to the center of the array [7].

The influence on the directivity patterns of the microphones in the array can be taken into account by changing the transfer function G_n . The filter optimization used to match the directivity pattern of the array with a desired one is called beamforming. The look direction of the microphone array is defined as the angle of the main lobe of the desired directivity pattern, which is also called the steering angle.

There are several beamforming algorithms based on an analytic solution for the optimal filter $w_n(f)$ and some others on a numerical approximation. Analytic solutions allow us to set N constraints on the directivity pattern for a finite number of frequencies, as for example described in [8]. Since we have a higher number of constraints in our problem, we will use numerical methods that allow accommodating a higher number of constraints to control the directivity pattern.

Equation (1) will be solved numerically, and for this purpose, the frequency range is discretized into P frequencies $f_p, p = 0, \dots, P - 1$ and the angular range into M angles $\phi_m, m = 0, \dots, M - 1$:

$$b(f_p, \phi_m) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f_p)G_n(f_p, \phi_m) \tag{3}$$

Equation (3) is reformulated in matrix notation as:

$$\mathbf{b}_m(f_p) = \mathbf{G}_{mn}(f_p)\mathbf{w}_n(f_p) \tag{4}$$

where the directivity pattern is an $M \times 1$ vector $\mathbf{b}_m^T(f_p) = [b(f_p, \phi_0), b(f_p, \phi_1), \dots, b(f_p, \phi_{M-1})]$, the transfer function an $M \times N$ matrix $[\mathbf{G}(f_p)]_{mn} = e^{-i\frac{2\pi f_p}{c}x_n \cos(\phi_m)}$ and the filter a $N \times 1$ vector $\mathbf{w}_n(f_p) = [w_{-\frac{N-1}{2}}(f_p), w_{-\frac{N-3}{2}}(f_p), \dots, w_{\frac{N-1}{2}}(f_p)]^T$ [7]. All bold variables are either vectors or matrices in the remainder of this manuscript.

2.2. Robustness and White Noise Gain

One of the problems that beamforming algorithms often have is their lack of robustness. This property is related to a resistance to the presence of spatially white noise and can be impaired by deviations from the specified microphone characteristics and microphone position errors. These imperfections affect the beamformer in a manner similar to a recorded spatially white noise that is amplified. Hence, the White Noise Gain (WNG) is a measure commonly used for quantifying the robustness of a beamformer design. The WNG shows the ability of a beamformer to suppress spatial white noise, because it expresses the gain of the beamformer in the desired look direction relative to the amplification of spatially white noise.

The WNG $A(f_p)$ is defined as follows:

$$A(f_p) = \frac{|b_{steer}(f_p)|^2}{\mathbf{w}_n^H(f_p)\mathbf{w}_n(f_p)} \tag{5}$$

where $b_{steer}(f_p)$ denotes the value of the directivity pattern in steer direction [7]. A high value of the WNG $A(f_p) > 1$ corresponds to a robust beamforming design, whereas a small value $A(f_p) < 1$

effectively corresponds to an amplification of spatial white noise [7]. The maximum possible value of the WNG is equal to the number of microphones used:

$$\max(A(f_p)) = N \tag{6}$$

which corresponds to a uniform filter [7]:

$$|w_n(f_p)| = \frac{1}{N} \tag{7}$$

2.3. Desired Directivity Pattern

The playback of the recorded signals should be in a stereophonic configuration, as mentioned in Section 1 and illustrated in Figure 3a.

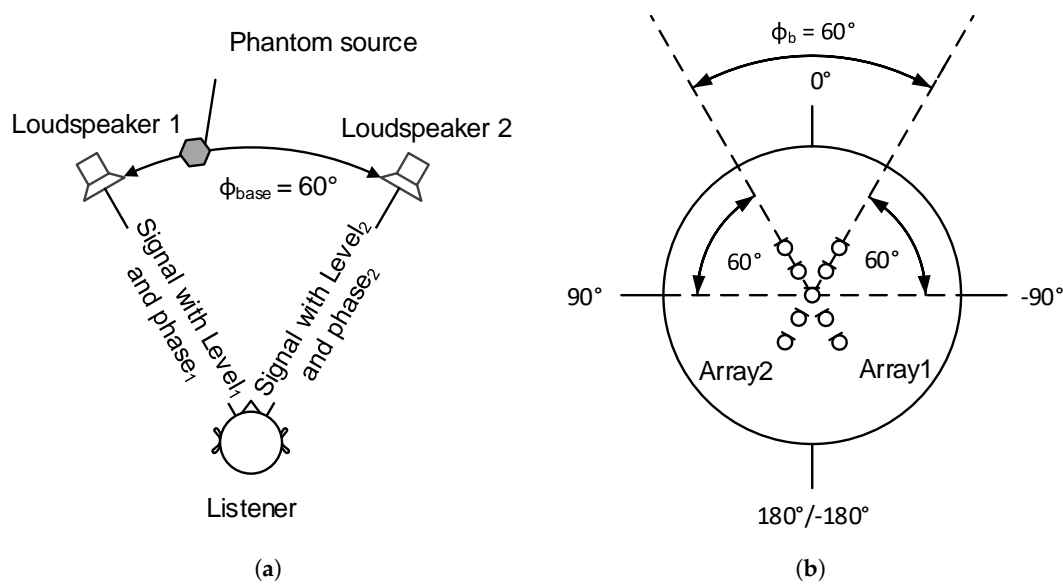


Figure 3. The stereophonic recording configuration is based on the playback one. Recorded level and phase differences with the two end-fire microphone arrays generate a phantom source between the two loudspeakers in the playback configuration. The signal emitted from Loudspeaker 1 has the level $Level_1$ and the phase $phase_1$. The signal emitted from Loudspeaker 2 has the level $Level_2$ and the phase $phase_2$. (a) Typical stereophonic playback configuration [9]; (b) proposed stereophonic recording configuration with sketched microphone positions. The absolute microphone positions are shown in Section 3.

The playback approach proposed by Grosse and van de Par [3] uses two loudspeakers for the direct sound reproduction with a typical base angle of $\phi_{base} = 60^\circ$ relative to the listener’s position [9]. There are several approaches to shift a phantom source from one loudspeaker to the other, utilizing phase differences $\Delta phase = phase_1 - phase_2$ and/or level differences (amplitude panning) $\Delta Level = Level_1 - Level_2$ applied on the two loudspeaker signals.

Based on this playback configuration, the recording configuration presented in this paper consists of two crossed end-fire microphone arrays with a 60° opening angle, sharing one center microphone and using omnidirectional microphones, illustrated in Figure 3b. The microphone positions in this figure can only be considered as a sketch, the absolute positions can be found in Section 3. The phantom-source shifting approaches of the playback configuration can be used to formulate either the correct phase and/or level differences between the two arrays. In this way, the perceived location of the sound source in the playback situation is identical to the one of the

recording provided that the distribution of recorded sound sources does not span more than 60° of angle. Although not evaluated here, in principle, a different opening angle could be used for the microphone arrays, thus effectively compressing or expanding the reproduced sound stage. We restrict our proposed method to have only level differences, and for this reason, the desired directivity pattern \hat{b} is purely real valued. With this desired directivity pattern, the phase of the directivity pattern is mainly controlled by the array design, which will be explained in Section 2.4.

In this paper, the phantom source shifting approach of amplitude panning is used for formulating the desired directivity pattern of Array 1 \hat{b}_{array1} and Array 2 \hat{b}_{array2} [9]:

$$\begin{aligned} \hat{b}_{array1}(\phi_\delta) &= \sqrt{\left(1 + \left(\frac{\tan(\phi_\delta) - \tan(\phi_b/2)}{\tan(\phi_\delta) + \tan(\phi_b/2)}\right)^2\right)^{-1}} \\ \hat{b}_{array2}(\phi_\delta) &= \sqrt{\left(1 + \left(\frac{\tan(\phi_\delta) + \tan(\phi_b/2)}{\tan(\phi_\delta) - \tan(\phi_b/2)}\right)^2\right)^{-1}} \end{aligned} \tag{8}$$

The angle area ϕ_δ between both arrays is defined by:

$$\phi_\delta = \{\phi_m | -\phi_b/2 \leq \phi_m \leq \phi_b/2\} \tag{9}$$

with the constant $\phi_b = \phi_{base} = 60^\circ$. The derivation of the desired directivity patterns according to [9] gives two possible recording room assumptions: an anechoic chamber or a real room. The latter one is chosen for Equation (8) since the microphone array configuration will be used in real rooms, such as concert halls.

The desired directivity pattern of the one array is the mirror-flipped version of the other array. This symmetry of the recording configuration makes it possible to formulate one desired directivity pattern, which is the same for both arrays. The following parts of the desired directivity pattern, the first \hat{b}_{beam} valid for the beam area and the second \hat{b}_{steer} valid for the steering angle, consider a microphone array aligned on the 0° axis corresponding to the steering angle $\phi_{steer} = 0^\circ$:

$$\hat{b}_{beam} = \begin{cases} \sqrt{\left(1 + \left(\frac{\tan(\phi + \phi_b/2) - \tan(\phi_b/2)}{\tan(\phi + \phi_b/2) + \tan(\phi_b/2)}\right)^2\right)^{-1}} & \text{for } -\phi_b \leq \phi < 0^\circ \\ \sqrt{\left(1 + \left(\frac{\tan(\phi - \phi_b/2) + \tan(\phi_b/2)}{\tan(\phi - \phi_b/2) - \tan(\phi_b/2)}\right)^2\right)^{-1}} & \text{for } 0^\circ < \phi \leq \phi_b \end{cases} \tag{10}$$

$$\hat{b}_{steer}(\phi_{steer} = 0^\circ) = 1 \tag{11}$$

In the following subsections, an optimal array design in terms of optimal microphone positions and an optimal filter design is proposed to achieve the desired directivity pattern.

2.4. Array Design

The positions of the microphones have an influence both on the filter $\mathbf{w}_n(f_p)$ and the transfer function $\mathbf{G}_{mn}(f_p)$, and thus, on the directivity pattern itself. The optimal microphone positions selected for this paper maximize the spatial aliasing frequency and, at the same time, minimize the frequency from which beamforming is effectively possible. The spatial aliasing frequency describes the lowest frequency f_{al} for which aliasing effects occur, which is caused by a spatial undersampling of the array for sound waves at high frequencies. The aliasing leads to side lobes with the same amplitude as the main lobe. The spatial aliasing frequency of an array with linear microphone spacing is usually given in the literature as:

$$f_{al} = \frac{c}{2\Delta x} \tag{12}$$

with Δx as the space between the microphones [10].

A small microphone spacing sets an upper limit to the spatial aliasing frequency. In contrast, a large microphone spacing sets a lower limit to the frequency from which beamforming is effectively possible. In order to have good directional properties of the microphone array across a wide frequency range, an irregularly-spaced microphone array is used in which both kinds of spacing can occur. A linear-shaped, logarithmically-spaced, to the reference microphone ($n = 0$), symmetrical array is used in this paper. Consequentially, the number of the used microphones N has to be uneven ($N \in \mathbb{N}_U$). The symmetry around one central microphone ensures a purely real directivity. The microphone positions are calculated as follows [11].

$$\begin{aligned} (x_{n+1} - x_n) &= (x_n - x_{n-1})\xi & \text{if } n > 0 \\ (x_{n-1} - x_n) &= (x_n - x_{n+1})\xi & \text{if } n < 0 \end{aligned} \tag{13}$$

with:

$$\begin{aligned} x_0 &= 0 \\ \xi &= \left(l_{spread}\right)^{\frac{2}{N-3}} \\ (x_1 - x_0) &= (x_0 - x_{-1}) = \frac{Length}{2 \sum_{n=1}^{\frac{N-1}{2}} \xi^{n-1}} \end{aligned}$$

where $Length$ is the total length of the array. The array parameter $l_{spread} \in \mathbb{R}^{>0}$ is a free variable describing the ratio between the spacing of the microphones at the extremities of the array and the spacing of the microphones at the center of the array. Linear microphone spacings are archived with $l_{spread} = 1$. If $l_{spread} < 1$, the spacing of the microphones at the extremities of the array is smaller than the one at the center of the array. In the case of $l_{spread} > 1$, it is the opposite.

2.5. Filter Design

In this section, an optimal filter design is proposed to fit the directivity pattern of the array, whose design was specified in Section 2.4, to the desired directivity pattern specified in Section 2.3. The following filter design is based on numerical convex optimization and has the advantage that only one global minimum exists. In general, this end-fire design can also be used with different desired directivity patterns and array designs. In Section 3, we indicate the ideal values of the constants for the desired directivity pattern and array design proposed in this study.

The aim of this algorithm is to minimize the quadratic error \mathbf{error}_m between the directivity pattern obtained by a microphone array $\mathbf{b}_m(f_p)$ and a desired frequency independent directivity pattern $\hat{\mathbf{b}}_m$ [7]:

$$\begin{aligned} \mathbf{error}_m &= \mathbf{G}_{mn}(f_p)\mathbf{w}_n(f_p) - \hat{\mathbf{b}}_m = \mathbf{b}_m(f_p) - \hat{\mathbf{b}}_m \\ \min_{\mathbf{w}_n(f_p)} &\|\mathbf{error}_m\|_2^2 \end{aligned} \tag{14}$$

This minimization task will be subjected to additional constraints, and therefore, the beamformer will be termed the Constrained Least-Squares Beamformer (CLSB).

In the following subsections, the main minimization task and the used constraints will be explained paying particularly attention to the WNG and different spatial areas. These areas are shown in Figure 4.

Additionally, this optimization process is placed within an optimization loop in order to optimize several important constants. This optimization procedure will be explained in the last subsection of this section.

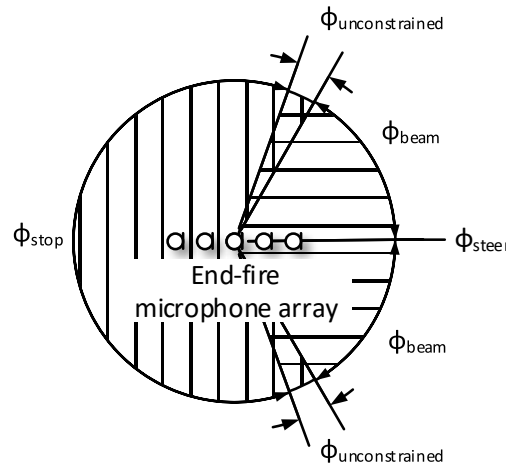


Figure 4. Different spatial areas in the directivity pattern optimization problem. The steering angle ϕ_{steer} , the beam area ϕ_{beam} (indicated by horizontal hash lines), an area without any constraints $\phi_{unconstrained}$ (indicated by crossed hash lines) and the stop area ϕ_{stop} (indicated by vertical hash lines).

2.5.1. White Noise Gain

Such a convex optimization procedure allows including a frequency-dependent lower bound $\gamma(f_p)$ for the WNG when optimizing the filters $w_n(f_p)$ [7]:

$$A(f_p) = \frac{|b_{steer}(f_p)|^2}{\mathbf{w}_n^H(f_p) \mathbf{w}_n(f_p)} \geq \gamma(f_p) \tag{15}$$

with $\gamma(f_p) \in \mathbb{R}^{\geq 0}$

This constraint has a direct influence on the robustness and on how well the desired directivity pattern can be achieved. A high value for the lower bound reduces the accuracy of forming the directivity pattern because the filter is too restricted by this constraint, whereas a low value leads to a not robust filter. In Section 3, an optimal value for this lower bound will be discussed.

2.5.2. Steering Angle

In the direction of the steering angle ϕ_{steer} , representing the direction of the main lobe of the microphone array, the directivity pattern obtained by the array is constrained to the value of the desired directivity pattern [7]:

$$\mathbf{G}_{steer,n}(f_p) \mathbf{w}_n(f_p) = b_{steer}(f_p) \stackrel{!}{=} \hat{b}_{steer} \tag{16}$$

In this way, the directivity pattern is normalized to \hat{b}_{steer} . The steering angle is limited to the array-axis, since the goal is an end-fire array.

2.5.3. Beam Area

The area around the steering angle is the beam area, which defines the main lobe of the directivity pattern:

$$\phi_{beam} = \{ \phi_m | \phi_{steer} - \phi_b \leq \phi_m \leq \phi_{steer-1} \wedge \phi_{steer+1} \leq \phi_m \leq \phi_{steer} + \phi_b \} \tag{17}$$

with $\phi_b \in \mathbb{R}^{\geq 0}$

$\phi_{steer-1}$ and $\phi_{steer+1}$ indicate one discrete angle before and after the steering angle, respectively. The constant ϕ_b can be chosen freely and defines the width of the beam area. Fitting the directivity pattern to the desired one, an angle-dependent upper bound ϵ_{beam} is set to the error (cf. Equation (14)) in this area:

$$\begin{aligned} \text{abs}(\mathbf{error}_{beam}) &\leq \epsilon_{beam} \\ \text{with } \epsilon_{beam} &\in \mathbb{R}^{\geq 0} \end{aligned} \tag{18}$$

where $\text{abs}()$ denotes the absolute value of every entry of the vector argument. In this case, ϵ_{beam} is a column vector with as many entries as the directivity pattern in the beam area.

2.5.4. Unconstrained Area

An angle area without any constraints is defined to avoid an effective discontinuity in the intermediate zone between the beam and the stop area, which would have a negative impact on the optimized solution that would be obtained:

$$\begin{aligned} \phi_{unconstrained} &= \{ \phi_m | \phi_{steer} - \phi_b - \phi_u \leq \phi_m < \phi_{steer} - \phi_b \wedge \phi_{steer} + \phi_b < \phi_m \leq \phi_{steer} + \phi_b + \phi_u \} \\ \text{with } \phi_u &\in \mathbb{R}^{\geq 0} \end{aligned} \tag{19}$$

The constant ϕ_u can be chosen freely and defines the width of the unconstrained area.

2.5.5. Stop Area

The remaining area is called the stop area:

$$\phi_{stop} = \{ \phi_m | \phi_{steer} + \phi_b + \phi_u < \phi_m < \phi_{steer} - \phi_b - \phi_u \} \tag{20}$$

The main optimization task is applied to this area. In the context of this work, the sound from this direction can be assumed to be mainly reverberant sound that does not belong to the direct sound and is therefore undesired. For this reason, the desired directivity pattern in this area is set to zero to suppress sound coming from this area as much as possible [7]:

$$\begin{aligned} \min_{\mathbf{w}_n(f_p)} &\| \mathbf{error}_{stop} \|_2^2 \\ \text{with } \hat{\mathbf{b}}_{stop} &= 0 \end{aligned} \tag{21}$$

In addition to this optimization, an upper bound ϵ_{stop} is set to the uniform norm of the directivity pattern:

$$\begin{aligned} \| \mathbf{error}_{stop} \|_{\infty} &\leq \epsilon_{stop} \\ \text{with } \epsilon_{stop} &\in \mathbb{R}^{\geq 0} \end{aligned} \tag{22}$$

This upper bound is not angle-dependent, but restricted to the stop area because of the uniform norm and will play an important role in the following loop design.

2.5.6. Loop Design

Choosing the correct upper bound for the beam area is difficult: on the one hand, a low upper bound for the beam area leads to a good fit in this area (low \mathbf{error}_{beam} values), but to undesired side lobes in the stop area (high \mathbf{error}_{stop} values). Consequential, the direct sound will be recorded correctly, but is mixed with the undesired reverberant sound field, which should be ideally suppressed. On the other hand, a high upper bound for the beam area leads to the opposite, a bad fit in the beam area (high \mathbf{error}_{beam} values), but low undesired side lobes (low \mathbf{error}_{stop} values). The following loop

design finds a frequency-dependent optimal upper bound for the beam area, which is a compromise between a good fit in the beam area and only small side-lobes in the stop area.

As a first step in the loop design, the upper bound of the beam area is initialized in matrix notation:

$$\epsilon_{beam}^k = \begin{matrix} & k = 1 & k = 2 & \dots & k = K \\ \phi_{steer} - \phi_b & \left(\begin{matrix} 0 & \alpha & \dots & \hat{b}_{steer} - \hat{b}(\phi_{steer} - \phi_b) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{steer} + \phi_b & 0 & \alpha & \dots & \hat{b}_{steer} - \hat{b}(\phi_{steer} + \phi_b) \end{matrix} \right) \end{matrix} \quad (23)$$

with $K \in \mathbb{N}^{>1}$, $k \in \mathbb{N}^{\leq K}$ and $\alpha = [\hat{b}_{steer} - \hat{b}(\phi_{steer} \pm \phi_b)] / K$

The rows cover the beam area, whereas the columns cover the different iterations of the following loops with k as the counter, where $k = K$ indicates the last iteration. The upper bound starts in the first iteration with $\epsilon_{beam}^{k=1} = 0$ and continues linearly spaced with step size α . The step size is designed in such a way that the maximum value of the upper bound of the beam area $\hat{b}_{steer} - \hat{b}(\phi_{steer} \pm \phi_b)$ is reached in overall K steps. Either $\hat{b}(\phi_{steer} - \phi_b)$ or $\hat{b}(\phi_{steer} + \phi_b)$ can be chosen to calculate α , since they are equal according to the symmetry of the desired directivity pattern. The upper bound then ends with the difference between \hat{b}_{steer} and \hat{b}_{beam} at the row specific angle. If this difference is reached before the last iteration ($k < K$), this value will stay till this iteration is reached. This will be the case for every row, except the first and the last one. This procedure ensures that \hat{b}_{steer} stays the maximum value of the directivity pattern.

In contrast to the upper bound of the beam area, the bound of the stop area is initialized as a vector, since there is no angle dependency:

$$\epsilon_{stop}^l = \begin{matrix} l = 1 & \dots & l = L \\ \left(\hat{b}_{steer} \cdot b_{stop}^{first} & \dots & \hat{b}_{steer} \right) \end{matrix} \quad (24)$$

with $L \in \mathbb{N}^{>1}$, $l \in \mathbb{N}^{\leq L}$ and $b_{stop}^{first} \in \mathbb{R}^{\geq 0, \leq 1}$

The entries with the counter l , where $l = L$ indicates the last iteration, correspond to the iterations of the following loops and are linearly spaced. The constant b_{stop}^{first} controls the maximum allowed value of the directivity pattern in the stop area for the first iteration.

The loop design itself can be seen in Figure 5 and is repeated for every frequency f_p , where the constants K_{temp} and K_{step} can be chosen freely so that $K / K_{temp} \in \mathbb{N}$ and $K_{temp} / K_{step} \in \mathbb{N}$, respectively. These two constants regulate the part of the upper bound of the beam area, which is used in the looped optimization process.

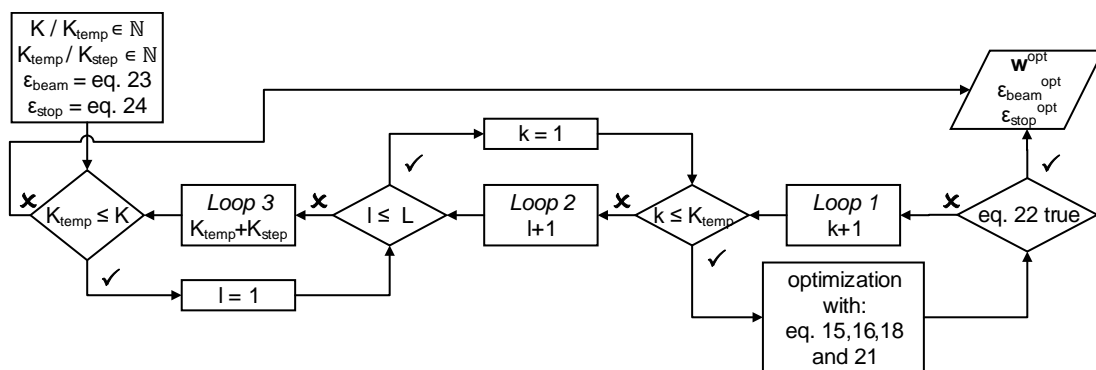


Figure 5. Loop design to determine the optimal filter, as well as the optimal upper bound for the beam and the stop area.

The first loop repeats the optimization with the first part of the upper bound of the beam area (from $\epsilon_{beam}^{k=1}$ to $\epsilon_{beam}^{k=K_{temp} \leq K}$) till Equation (22) with ϵ_{stop}^1 is true. A result of the optimization, fulfilling Equation (22), is denoted as valid. If this is not the case, Loop 2 repeats Loop 1 with different upper bounds of the stop area (from ϵ_{stop}^2 to ϵ_{stop}^L). If still no valid result is found, Loop 3 increases K_{temp} with the step width of K_{step} . The upper bounds, for which the loop design finds a valid solution, are denoted as optimal ϵ_{beam}^{opt} and ϵ_{stop}^{opt} . The filter \mathbf{w} , which corresponds to these upper bounds, is also denoted as optimal \mathbf{w}^{opt} . For the case that K_{step} increases K_{temp} over K ($K_{temp} + K_{step} > K$), the last $k = K$ calculated result of the optimization is taken as a valid solution.

3. Setup

The following setup is used for the numerical simulations, whose results are described in Sections 4 and 5. The angular range is discretized into $M = 360$ linearly-spaced angles $\{\phi_0 = 0^\circ, \phi_1 = 1^\circ, \dots, \phi_{359} = 360^\circ\}$. The frequency range covers the range of $f_{p=0} = 0$ Hz to $f_{p=256} = 24$ kHz generated at a sampling rate of $f_s = 48$ kHz using a filter length of 512 samples. This results in $P = 257$ linear spaced frequency bins. This frequency range covers the spectral content of music [12] that is to be recorded by these microphone arrays. To obtain impulse responses of the filters, the complex spectrum was mirrored, conjugated and transformed towards the time domain via an ifft.

The microphone array consists of $N = 9$ omnidirectional microphones and has a total length of $Length = 1$ m. The array design is done with $l_{spread} \approx 35$, so that the smallest microphone spacing (s) in the center of the array is $s = 0.01$ m. Following that, the spatial aliasing frequency can be maximized to a frequency of $f_{al} \approx 17,000$ Hz. For practical reasons, the limitation is set to $s = 0.01$ m to ensure enough space for the microphones. The absolute microphone positions are set as follows (displayed in millimeter precision): $x_{n=-4} = -0.500$ m, $x_{n=-3} = -0.150$ m, $x_{n=-2} = -0.043$ m, $x_{n=-1} = -0.010$ m, $x_{n=0} = 0$ m, $x_{n=1} = 0.010$ m, $x_{n=2} = 0.043$ m, $x_{n=3} = 0.150$ m, $x_{n=4} = 0.500$ m.

After having specified the microphone positions, the convex functions of the CLSB, shown in Section 2.5, are solved utilizing CVX, a package for specifying and solving convex programs [13,14]. Parts of these convex functions are the WNG constraint and the loop design.

For the WNG constraint, the lower bound γ for the WNG $A(f_p)$ is set up as follows:

$$\gamma(f_p) = \begin{cases} 5 & \text{for } f_p = 0 \text{ Hz} \\ \text{CSI} & \text{for } 0 \text{ Hz} < f_p < 187.5 \text{ Hz} \\ 1 & \text{for } 187.5 \text{ Hz} \leq f_p \leq f_s/2 \text{ Hz} \end{cases} \quad (25)$$

The lower bound starts with $\gamma(f_p = 0 \text{ Hz}) = 5$ and ends with $\gamma(f_p \geq 187.5 \text{ Hz}) = 1$. In the intermediate zone, a Cubic Spline Interpolation (CSI) connects both points. The CSI in the intermediate zone avoids rapid changes of the directivity pattern across frequency below ($f_p < 187.5$ Hz). In the high frequency range ($f_p \geq 187.5$ Hz), a lower bound of $\gamma = 1$ ensures a robust beamforming design.

For the loop design, the constants are set up as follows:

$$\begin{aligned} K &= 100, K_{temp} = K_{step} = 10, \alpha = 0.01 \text{ cf. Equation (23)} \\ L &= 9, b_{stop}^{first} = 0.2 \\ \phi_u &= 10^\circ \end{aligned} \quad (26)$$

The constants ϕ_b and ϕ_{steer} , as well as the parts of the desired directivity pattern \hat{b}_{beam} and \hat{b}_{steer} are set up according to Section 2.3.

The values of the constants K , K_{temp} and K_{step} are chosen in such a way that Loop 1 scans the beam area from $\epsilon_{beam}^{k=1} = 0$ in steps of $\alpha = 0.01$ till $\epsilon_{beam}^{k=K_{temp}=10} = K_{temp} \cdot \alpha = 0.1$. If necessary, Loop 3 increases the value of the upper bound of the beam area according to the value of the constant K_{step} (cf. Section 2.5).

An increase of the value of the constant K leads to an improvement in the beam area (lower $error_{beam}$ values), because the step size α is smaller. The validity (cf. Section 2.5) of more possible directivity patterns with small $error_{beam}$ values is checked by the loop design. In fact, to find a valid solution, Loop 2 has to increase ϵ_{stop} further than before, which leads to a worsening in the stop area (higher $error_{stop}$ values). A decrease of the value of the constant K leads consequently to the opposite effect.

An increase of the values of the constants K_{temp} and K_{step} leads to a worsening in the beam area (higher $error_{beam}$ values), because the first end point of Loop 1 $\epsilon_{beam}^{k=K_{temp}}$, as well as all of the other ones $\epsilon_{beam}^{k=K_{temp}+K_{step}+K_{step}+\dots}$ is now higher. More possible directivity patterns with high $error_{beam}$ values are checked by the loop design: Loop 2 does not have to increase ϵ_{stop} so much than before, because these directivity patterns are in general more likely to be valid. This leads then to an improvement in the stop area (lower $error_{stop}$ values). A decrease of the values of the constants K_{temp} and K_{step} leads consequently to the opposite effect.

The values of the constants L and b_{stop}^{first} are chosen in such a way that Loop 2 scans the stop area from $\epsilon_{stop}^{l=1} = 0.2$ in steps of $(\hat{b}_{steer} - b_{stop}^{first} \cdot \hat{b}_{steer}) / (L - 1) = 0.1$ till $\epsilon_{stop}^{l=L} = \hat{b}_{steer} = 1$.

An increase of the value of the constant b_{stop}^{first} and at the same time a decrease of the value of the constant L , preserving the step width of 0.1 as mentioned earlier, lead to a worsening in the stop area. The start point of Loop 2 is now higher, allowing higher $error_{stop}$ values from the beginning. It is now easier for Loop 1 to find a valid solution, which leads to an improvement in the beam area. A decrease of the value of the constant b_{stop}^{first} and a coherent increase of the value of the constant L lead to the opposite effect.

Overall, it can be said that a variation of the values of the constants K , K_{temp} , K_{step} , L and b_{stop}^{first} leads to a changed balance, fulfilling the constraints between the beam and the stop area. For every desired directivity pattern and intended purpose of the microphone array has to be found separately optimal values.

A variation of the value of the constant ϕ_u does not significantly change the results in terms of the error in the beam and the stop area. Nevertheless, the value should not be chosen too big to avoid undesired results (very big differences between the obtained and the desired directivity pattern), since there is no control over the directivity pattern in the unconstrained area. The maximum value of ϕ_u till there are no undesired results depends in a complex manner on the number of used microphones and the desired directivity pattern.

With the setup shown in Equation (26), we achieved best results in fitting the directivity pattern to the desired one. Different initializations of the constants are also possible, as mentioned before (a detailed analysis of the effect on the results regarding the variation of the constants' values given in Equation (26) is beyond the scope of this article). Our results are, however, discussed in the following Sections 4 and 5.

4. Objective Evaluation

The following section is divided into four parts. In Section 4.1, two array designs are compared to each other to show the improvement of the spatial aliasing of a logarithmically-spaced array over a linearly-spaced one. In the second Section 4.2, the new stereo system proposed in this study is compared to the state-of-the-art ones, which utilize two microphones. In the third Section 4.3, the WNG constraint and the frequency response are analyzed. Finally, in the last Section 4.4, the angular constraints, as well as the phase of the directivity pattern are investigated.

4.1. Directivity Index Comparison

The directivity pattern of the logarithmically-spaced array ($l_{spread} \approx 35, s = 0.01$ m) is more directive for high frequencies than the one of a linearly-spaced array ($l_{spread} = 1, s = 0.125$ m) having the same total length of $Length = 1$ m. Less reverberant sound is recorded by the first type of array

than by the latter one. As a measure, we choose the directivity index DI , which is the logarithm of the directivity D [15]:

$$D(f_p) = \frac{\sum_{m=0}^{M-1} \max_{\phi_m} (|b(f_p, \phi_m)|^2)}{\sum_{m=0}^{M-1} |b(f_p, \phi_m)|^2} \tag{27}$$

$$DI(f_p) = 10 \log_{10}(D(f_p))$$

In fact, Figure 6 shows that the linearly-spaced array has lower DI values for high frequencies ($f_p > 1200$ Hz) than the logarithmically-spaced one. This is caused by aliasing effects, as the aliasing frequency for the linearly-spaced array is $f_{al} \approx 1460$ Hz. There is a big drop of the DI values ($DI < 7$ dB) for the logarithmically-spaced array for very high frequencies ($f_p > 10,500$ Hz), which is also caused by aliasing effects. The lowest values of the DI for the logarithmically-spaced array are located around the aliasing frequency $f_{al}(\Delta x = s) \approx 17,000$ Hz.

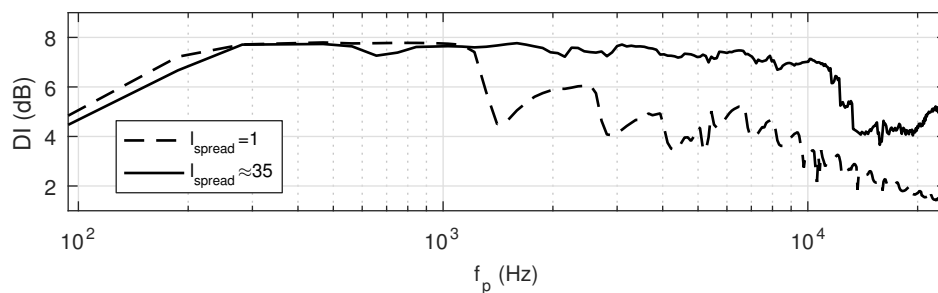


Figure 6. Directivity index $DI(f_p)$ of a linearly-spaced array ($l_{spread} = 1, s = 0.125$ m) (dashed line) and the logarithmically-spaced one ($l_{spread} \approx 35, s = 0.01$ m) (solid line) with the same total length of $Length = 1$ m.

4.2. Comparison Stereo Systems

The necessary phase and/or level differences for a stereophonic recording as mentioned in Section 2.3 can also be obtained by only two microphones. Different angles and distances between these two microphones, as well as different microphone directivity patterns are possible, as described, for example, by the A-B or the X-Y technique [12]. A unified theory of these two-microphone systems for stereophonic sound recording can be found in [6].

Assuming no phase differences, this theory states that a level difference of $\Delta Level = \pm 15$ dB determines the left or right lateral shift towards the loudspeakers of a phantom sound source in the playback situation. This level difference is achieved in the recording situation with different angles between two microphones with specific directivity patterns. The angle covering this level difference is called recording angle ϕ_{rec} . If $\phi_{rec} > \phi_{base}$, the recorded sound scene is compressed in the playback configuration, whereas $\phi_{rec} < \phi_{base}$, the recorded sound scene is expanded [6]. Therefore, we can assume that if we have $\phi_{rec} = \phi_{base}$, the recorded spatial properties are the same after playback. Table 1 shows the possible microphone directivities and base angles between the microphone pairs.

The microphone array stereo system described in this study records less reverberant sound than these state-of-the-art two-microphone stereo systems. As a measure, we choose a modified definition of the directivity index DI_{mod} , which is the logarithm of a modified directivity D_{mod} , mentioned in Section 4.1:

$$D_{mod} = \frac{\sum_{m=0}^{M-1} 2 \max_{\phi_m} (b_{mic1}(\phi_m)^2)}{\sum_{m=0}^{M-1} b_{mic1}(\phi_m)^2 + b_{mic2}(\phi_m)^2} \tag{28}$$

$$DI_{mod} = 10 \log_{10}(D_{mod})$$

where $b_{mic1}(\phi_m)$ and $b_{mic2}(\phi_m)$ are the directivity patterns of the first and the second microphone, respectively. The modified directivity index includes the sum of the directivity patterns of the two microphones. The modified directivity index considers the angle between these two directivity patterns, which determines the percentage of recorded reverberant sound in addition to the directivity pattern itself. As shown in Table 1, the proposed microphone array stereo system is, in fact, more directive than the two-microphone stereo ones, taking also into account the angle between the two microphone arrays.

Table 1. The modified directivity index DI_{mod} of the state-of-the-art two-microphone stereo systems and the microphone array stereo system described in this study. For the latter one, the desired directivity patterns are used. Only stereo systems with $\phi_{rec} = \phi_{base}$ are displayed. This angle constraint avoids angular compression or angular expansion in the playback situation.

Two-Microphone Stereo Systems		
Microphone Directivity	Angle between the Microphones (°)	DI_{mod}
Figure of Eight	101	5.95
Hypercardioid (back attenuation = -6 dB)	136	8.29
Hypercardioid (back attenuation = -10 dB)	156	8.7
Microphone Array Stereo System		
$DI_{mod} = 11.29$ with $b_{mic1}(\phi_m) = \hat{b}_{array1}(\phi_m)$ and $b_{mic2}(\phi_m) = \hat{b}_{array2}(\phi_m)$		

4.3. WNG and Frequency Response

The algorithm successfully fits the WNG $A(f_p)$ to the lower bound $\gamma(f_p)$ specified in Section 3, as shown in Figure 7a.

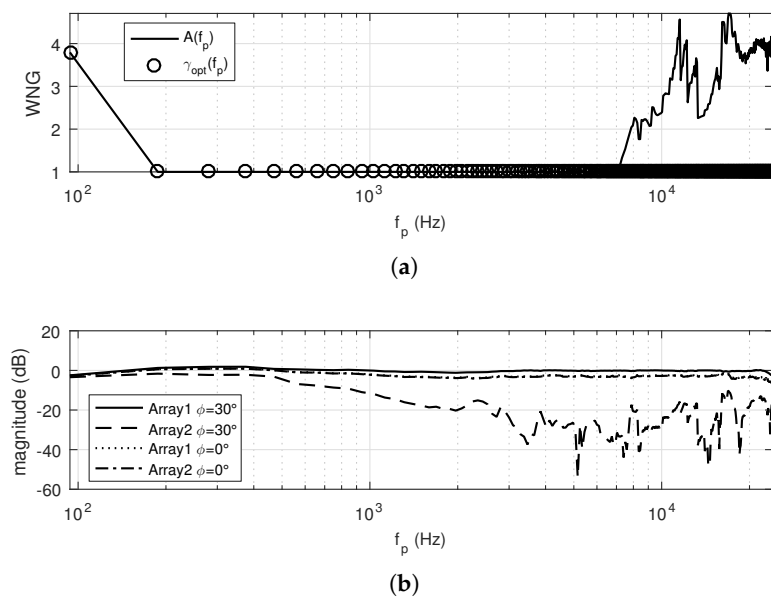


Figure 7. (a) White Noise Gain (WNG) $A(f_p)$, as well as the lower bound for the WNG $\gamma(f_p)$ across frequency; (b) shown are frequency responses of both arrays for two sound sources emanating from $\phi = 30^\circ$ and $\phi = 0^\circ$ according to the configuration illustrated in Figure 3.

This ensures a robust beamforming design. For high frequencies $f_p \geq 7031$ Hz, the algorithm finds even higher WNG values than the lower bound.

Figure 7b shows the frequency response of both arrays according to the configuration that is shown in Figure 3. The responses for both arrays were calculated for a sound source emanating from $\phi = 30^\circ$ (resulting in a sound source perceived at the location of the left loudspeaker, solid and dashed line)

according to Figure 3 and $\phi = 0^\circ$ (resulting in a phantom source between both speakers, dotted and dash-dotted line). It can be seen that for $\phi = 0^\circ$, the responses of both arrays show a high similarity in terms of level differences and have only minor fluctuations of approximately ± 2 dB above 1000 Hz. Below 1000 Hz, it can be observed that there is a boost of approximately 3 dB, which might be attributed to a violation of a constraint at low frequencies. When the sound source is emanating from $\phi = 30^\circ$, a flat frequency response can be observed for Array 1 (on axis) with minor fluctuations of approximately 1 dB across frequency. Array 2 shows a considerably lower level, but larger fluctuations. It can be assumed that these fluctuations will not be perceivable because the location of the sound source will be determined by Array 1.

4.4. Beam and Stop Area Constraints

The results of the loop design mentioned in Section 2.5 are shown in Figure 8. This loop design finds a compromise between a good fit in the beam area and low directivity pattern values in the stop area.

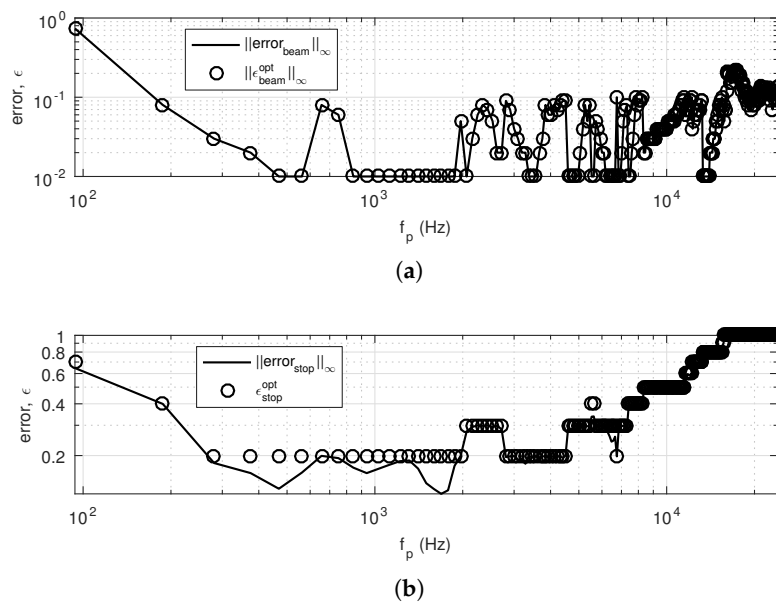


Figure 8. The difference between the simulated directivity pattern and the desired one (*error*) in the beam (a) and the stop (b) area, as well as the corresponding upper bounds of both areas as function of the frequency.

For low frequencies $f_p < 187.5$ Hz, the directivity pattern is quite omnidirectional ($\|\mathbf{error}_{stop}\|_\infty > 0.2$ and $\|\mathbf{error}_{beam}\|_\infty > 0.1$), so that Loop 3 has to increase ϵ_{beam} to $\|\epsilon_{beam}^{opt}\|_\infty > 0.1$. For higher frequencies $f_p \geq 187.5$ Hz, there is a good fit in the beam area $\|\mathbf{error}_{beam}\|_\infty \leq 0.1$ so that Loop 1 and Loop 2 find the ideal upper bound for the beam and the stop area. Overall, it can be said that the best result is found in the frequency range of $281.3 \text{ Hz} \leq f_p \leq 1969 \text{ Hz}$: a good fit in the beam area combined with low directivity pattern values in the stop area $\|\mathbf{error}_{stop}\|_\infty \leq 0.2$. At high frequencies ($f_p \geq 16,690 \text{ Hz}$), Figure 8b shows aliasing effects ($\|\mathbf{error}_{stop}\|_\infty = 1$), which are expected, since the aliasing frequency of the logarithmically-spaced array is $f_{al}(\Delta x = s) \approx 17,000 \text{ Hz}$.

Figure 9 shows the polar plot of the desired directivity pattern in addition to the absolute value of the directivity patterns of the frequencies $f_p = 250 \text{ Hz}$, $f_p = 1000 \text{ Hz}$, $f_p = 4000 \text{ Hz}$ and $f_p = 8000 \text{ Hz}$. For all frequencies, there is a good fit (a small difference between desired and obtained directivity pattern) in the beam area, as already quantified by Figure 8a. Comparing the side-lobe-levels of the different frequencies, the following can be stated: the side-lobe-level decreases from $f_p = 250 \text{ Hz}$ to 1000 Hz ; there is no big difference in side-lobe-level between $f_p = 1000 \text{ Hz}$ and $f_p = 4000 \text{ Hz}$;

the side-lobe-level increases from $f_p = 4000\text{ Hz}$ to $f_p = 8000\text{ Hz}$. This analysis is described in a quantified matter in Figure 8b.

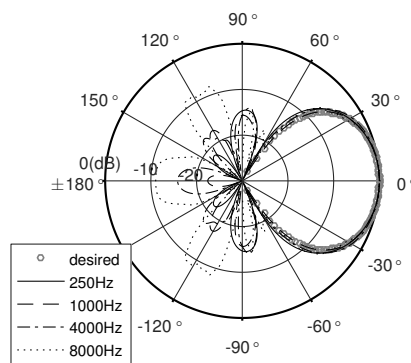


Figure 9. Polar plot of the desired directivity pattern (grey markers) and the absolute value of the obtained directivity patterns of the frequencies $f_p = 250\text{ Hz}$ (solid line), $f_p = 1000\text{ Hz}$ (dashed line), $f_p = 4000\text{ Hz}$ (dashed-dotted line) and $f_p = 8000\text{ Hz}$ (dotted line).

Figure 10a allows for a more detailed analysis, as it shows the absolute value of the difference between the directivity pattern and the desired one in the whole angular range $|error(\phi_m, f_p)|$. The omnidirectional behavior of the directivity pattern up to $f_p = 187.5\text{ Hz}$ can be also seen there. For higher frequencies, side lobes appear at $\phi_m = \pm 180^\circ$ and move with increasing frequency into the direction of the beam $-60^\circ \leq \phi_m \leq 60^\circ$. Aliasing effects can be seen in Figure 10a, like in Figure 8b.

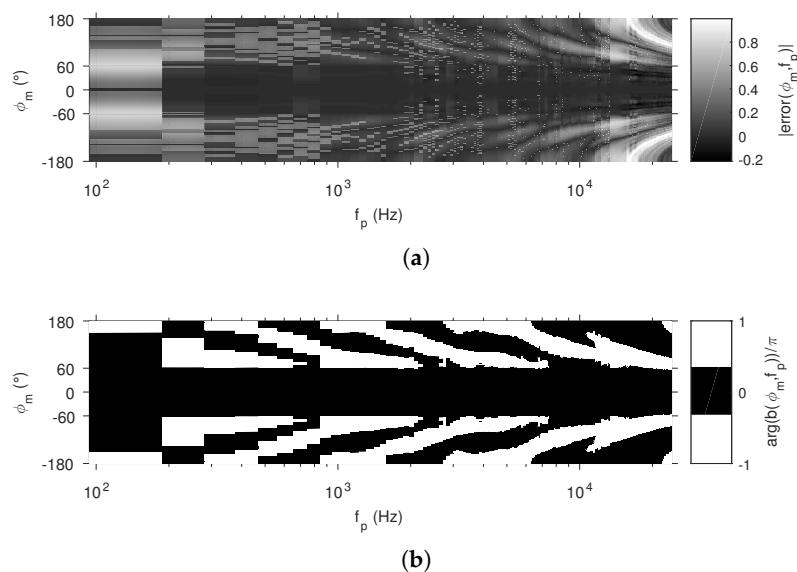


Figure 10. The difference between the directivity pattern and the desired one $|error(f_p, \phi_m)|$ (a), as well as the phase of the directivity pattern $\arg(b(f_p, \phi_m))$ (b).

In addition to the absolute value of the directivity pattern, the phase $\arg(b(f_p, \phi_m))$ is represented in Figure 10b.

The directivity pattern is purely real: the phase shows only three possible values $\arg(b) = \{-\pi, 0, \pi\}$ as mentioned in Section 2.3. In the beam area, the phase has, in fact, only values $\arg(b) = 0$, which leads to no phase differences between the two arrays in the recording configuration mentioned in Section 2.3.

5. Subjective Evaluation

In this section, the proposed microphone array is subjectively evaluated. For this purpose, a listening experiment was performed, whose results are shown.

5.1. Subjective Evaluation: Localization Accuracy

In order to evaluate the proposed stereophonic-microphone array in terms of localization accuracy when simulating spatially-distributed sound sources, subjective data were obtained in a localization experiment within a real room from listeners. The loudspeaker signals were generated using a single sound source and by simulating the delays between the microphones and the sound source. The optimized filters w^{opt} were applied on each microphone signal to obtain the output signal for the left and right array, which was then played back via the two loudspeakers during the listening experiment. The loudspeaker and array configurations are shown in Figure 3.

The sound sources were placed on virtual locations between -30° and $+30^\circ$ in a five degree resolution, resulting in a phantom source stereo image based on intensity-panning between the left and the right loudspeakers. The evaluation took place in a reverberant room with the dimensions (7.5, 7.1, 2.97) m with a reverberation time of $T_{60} = 0.45$ s. The distance between the loudspeakers was 3 m, and the listeners were seated at the position that created a 60° stereo triangle with the loudspeakers (cf. Figure 3). As a source signal, three short pink noise bursts with a total length of 1.1 s were presented to the listeners. The noise covered a frequency rang from 100 Hz to $f_s/2$ covering the spectral content of musical signals. Data were obtained from seven listeners, and the 13 source position angles were presented in random order. For each subject, the experiment covered one training session and three measurement sessions. The task of the participants was to indicate the perceived direction between the loudspeaker using indicators placed between the loudspeakers in five degree steps.

5.2. Subjective Evaluation: Results

Figure 11 shows the perceived directions of the subjective evaluation. The dotted line indicates perfect correspondence between the true source location and the perceived location. Circles show the average perceived location in dependence of the simulated source location. As can be seen, there is a rather linear behavior on localization, indicating a mostly precise representation of the presented directions. Exceptions can be observed around ± 20 degrees at which the presented source is perceived more lateral than the simulated source location. The maximum localization error of ≈ 6 degrees that can be observed can probably be attributed to the target functions that were used to optimize the directivity pattern, which may cause too high level differences when both arrays are used in combination.

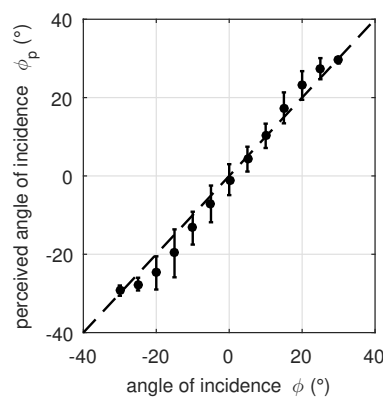


Figure 11. Illustrated are the mean-values of the perceived angle of incidence with the standard deviation across seven participants' means. The x -axis represents the simulated angle of incidence ϕ of the presented noise sources. The dotted line indicates a perfect match between simulated and perceived localization.

6. Discussion and Conclusions

In this study, a new approach for intensity stereophonic recording has been investigated. Guided by the playback situation and its auditory requirements, we decided to postulate a setup consisting of two crossed end-fire microphone arrays and a fitting desired directivity pattern. The difference between the directivity pattern obtained and the one desired was minimized by a superdirective beamforming algorithm. It was based on convex numeric optimization and also contains a frequency-dependent WNG constraint to ensure a robust beamforming design.

In addition to designing the filters of the microphones via beamforming algorithms, we found an ideal array design. This design maximizes the spatial aliasing frequency and also takes practical issues into account, which will appear in an actualization of the arrays. The extent of the microphones demands a particular spacing, also to avoid interferences between them.

A comparison between the new stereo system and the state-of-the-art ones, which use two microphones, has shown that the former has the advantage of less recorded reverberant sound, as it is more directive in the look direction than the latter are. This matches the requirements posed by the recording method proposed in Grosse and van de Par [3], which requires separate dry and reverberated representations of the audio signal. The reverberated sound field can be taken from single microphone signals.

Future research could develop a method to optimize the directivity pattern of both arrays as one system rather than handling them separately. Furthermore, two additional beams pointing into the diffuse field could be introduced for optimization to replace the two microphones placed in that field and to use only the array system.

A final assessment of the proposed recording and playback system needs to run listening tests and investigate the perception of the recording and playback room.

Acknowledgments: We would like to thank the Deutsche Forschungsgemeinschaft for supporting this work as part of the Forschergruppe Individualisierte Hoerakustik (FOR-1732). We also would like to thank the reviewers for their helpful and insightful comments.

Author Contributions: Steven van de Par and Julian Grosse formulated the constraints for the true stereo microphone array. Jonathan Albert Gößwein developed and evaluated the methods for optimizing the true stereo microphone array. Julian Grosse planned and performed the localization experiment.

Conflicts of Interest: The authors declare no conflict of interest

Abbreviations

The following abbreviations are used in this manuscript:

WNG	White Noise Gain
CLSB	Constrained Least-Squares Beamformer
CSI	Cubic Spline Interpolation

References

1. Berkhout, A.J. A holographic approach to acoustic control. *J. Audio Eng. Soc.* **1988**, *36*, 977–995.
2. Gerzon, M.A. Periphony: With-Height sound reproduction. *J. Audio Eng. Soc.* **1973**, *21*, 2–10.
3. Grosse, J.; van de Par, S. Perceptually accurate reproduction of recorded sound fields in a reverberant room using spatially distributed loudspeakers. *IEEE J. Sel. Top. Signal Process.* **2015**, *9*, 867–880.
4. Schroeder, M.R. Statistical parameters of the frequency response curves of large rooms. *J. Audio Eng. Soc.* **1987**, *35*, 299–306.
5. Haeussler, A.; van de Par, S. Theoretischer und subjektiver Einfluss des Aufnahmeraumes auf den Wiedergaberaum. In Proceedings of the 40th DAGA'14 Jahrestagung fuer Akustik, Oldenburg, Germany, 10–13 March 2014.
6. Williams, M. Unified theory of microphone systems for stereophonic sound recording. In Proceedings of the 82th Audio Engineering Society Convention, London, UK, 10–13 March 1987.

7. Mabande, E.; Schad, A.; Kellermann, W. Design of robust superdirective beamformers as a convex optimization problem. In Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 19–24 April 2009; pp. 77–80.
8. Frost, O.L. An algorithm for linearly constrained adaptive array processing. *Proc. IEEE* **1972**, *60*, 926–935.
9. Pulkki, V. Compensating displacement of amplitude-panned virtual sources. In Proceedings of the Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio, Espoo, Finland, 15–17 June 2002.
10. McCowan, I.A. Robust Speech Recognition using Microphone Arrays. Ph.D. Thesis, Queensland University of Technology, Brisbane City, QLD, Australia, 2001.
11. Corteel, E. On the use of irregularly spaced loudspeaker arrays for wave field synthesis, potential impact in spatial aliasing frequency. In Proceedings of the 9th international conference on Digital Audio Effects (DAFx'06), Montreal, QC, Canada, 18–20 September 2006; pp. 209–214.
12. Dickreiter, M.; Dittel, V.; Hoeg, W.; Woehr, M. *Handbuch der Tonstudioteknik*, 7th ed.; K. G. Sauer Verlag: München, Germany, 2008; Volume 1.
13. Grant, M.; Boyd, S. CVX: Matlab Software for Disciplined Convex Programming, version 2.1. 2014. Available online: <http://cvxr.com/cvx> (accessed on 18 May 2017).
14. Grant, M.; Boyd, S. Graph implementations for nonsmooth convex programs. In *Recent Advances in Learning and Control: Lecture Notes in Control and Information Sciences*; Blondel, V., Boyd, S., Kimura, H., Eds.; Springer: New York, NY, USA, 2008; pp. 95–110. Available online: http://stanford.edu/~boyd/graph_dcp.html (accessed on 18 May 2017).
15. Kinsler, L.; Frey, A.; Coppers, A.; Sanders, J. *Fundamentals of Acoustics*; John Wiley and Sons, Inc.: New York, NY, USA, 2000.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).