Friedrich-Alexander-Universität Erlangen-Nürnberg

Lehrstuhl für Multimediakommunikation und Signalverarbeitung

Prof. Dr.-Ing. Walter Kellermann

SIM Project

Sectorial Optimization of Robust Polynomial Beamformers for Uniformly Spaced Arrays

Michael Bürger

29.09.2011

Supervisors: Prof. Dr.-Ing. Walter Kellermann
M.Sc. Edwin Mabande

Erklärung

Ich versichere, dass ich die vorliegende Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe, und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Ort, Datum	Unterschrift

Contents

Abstract

Beamforming is a spatial filtering technique which aims at extracting desired signals arriving from certain directions while signals from other directions are attenuated. This is done by exploiting spatial properties of the acoustic field which is spatially sampled using a microphone array. Each microphone signal is then processed by a digital filter. Polynomial beamformers are especially suited for dynamic scenes as they enable an easy, efficient, and flexible steering of the beam to any desired direction. Such beamformers can be designed using a numerical design procedure where the responses of the polynomial beamformer are simultaneously optimized for different predefined prototype look directions. Steering the beam to intermediate directions is achieved by interpolation. To minimize the distortion of the desired signal, a distortionless response constraint can be added at each prototype look direction. Also, the white noise gain is typically constrained in order to obtain a robust design. Even though these constraints are necessary they reduce the number of available degrees of freedom. The aim of this work is to investigate the possibility of improving the performance of polynomial beamformers by exploiting the structure of symmetric microphone arrays. For this, we seize on the idea of designing the polynomial beamformer in a limited angular region only, as proposed by Lai et al. [LNL10] for uniformly spaced spiral arrays. Thereby, all available degrees of freedom can be used to optimize the responses for only a part of the entire angular region. Steering the beam to directions outside this area is achieved by rotating the filters to the neighboring microphone channel. In this thesis we extend the proposed method and present a more general technique which can be applied to any type of symmetric array. Hence, this method includes the special case of uniformly spaced spiral arrays. Simulation results confirm the benefit of the presented optimization scheme.

Kurzfassung

Beamforming ist ein Verfahren zur räumlichen Filterung von akustischen Szenen, mit dem Ziel, gewünschte Signale aus bestimmten Richtungen zu verstärken und gleichzeitig störende Signale aus anderen Richtungen zu unterdrücken. Dies wird mit Hilfe von Mikrofonanordnungen erreicht, die das Schallfeld räumlich abtasten und so dessen räumliche Eigenschaften erfassen, wobei jedes Mikrofonsignal mit einem digitalen Filter verarbeitet wird. Für dynamische Szenen ist ein *Polynomial Beamformer* besonders geeignet, da sich dessen *Beam* einfach und effektiv in beliebige Richtungen lenken lässt. Solche *Beamformer* werden üblicherweise mittels eines numerischen Verfahrens erstellt, bei dem die *Beamformer*-Übertragungsfunktionen gleichzeitig für einzelne, ausgewählte Richtungen optimiert werden. Soll der *Beam* später in dazwischenliegende Richtungen bewegt werden, so geschieht dies mittels Interpolation. Häufig muss beim Entwurf eines *Beamformers* die Randbedingung eingehalten werden, dass gewünschte Signale aus den Richtungen, die für die Optimierung gewählten werden, nicht verzerrt werden dürfen. Zusätzlich ist in den meisten Fällen auch eine Begrenzung des *White Noise Gains* erforderlich, um ein robustes System zu erhalten. Diese Randbedingungen reduzieren jedoch die für die Optimierung zur Verfügung stehenden Freiheitsgrade. Ziel der vorliegenden Arbeit ist es, zu überprüfen, ob die Leistungsfähigkeit eines *Polynomial Beamformers* verbessert werden kann, wenn man die Struktur symmetrischer Mikrofonanordnungen ausnutzt. Hierbei wird der in [LNL10] präsentierte Ansatz von Lai et al. aufgegriffen. Die Autoren zeigten, dass es für den Fall einer gleichmäßigen, spiralförmigen Mikrofonanordnung ausreicht, den *Beamformer* nur für einen begrenzten Winkelbereich zu entwerfen. Demnach kann dieselbe Anzahl an verfügbaren Freiheitsgraden verwendet werden, um die Übertragungsfunktionen für einen kleineren Winkelbereich zu optimieren. Der *Beam* kann anschließend in andere Bereiche gelenkt werden, indem man die jeweiligen Filter rotiert, d.h. auf die benachbarten Mikrofonsignale anwendet. In dieser Arbeit erweitern wir die von Lai et al. präsentierte Methode und stellen ein allgemeineres Verfahren vor, das bei allen symmetrischen Mikrofonanordnungen angewandt werden kann und somit den Sonderfall einer gleichmäßigen, spiralförmigen Anordnung beinhaltet. Anhand von Simulationen wird der Nutzen des vorgestellten Verfahrens aufgezeigt.

1. Introduction

In recent years the importance of hands-free acoustic human-machine interfaces has steadily increased. Much effort has been made to improve existing telecommunication services such as audio-/video-conferencing. In many applications the desired sound source is degraded by background noise or other undesired sources. For example, an air conditioning system may decrease the speech intelligibility during a teleconference. Or, an automatic speech recognition system in a distant-talking speech interface can be negatively affected if another person is talking at the same time. In order to enhance the desired signal there are several different digital signal processing techniques available.

A widely used spatial filtering approach termed beamforming enhances the signals arriving from desired look directions while suppressing interfering signals from other directions. This is achieved by spatially sampling the sound field using a microphone array, and thereby exploiting its spatial properties. Classical filter-and-sum beamformers have a fixed look direction and are therefore not well suited if more than one desired sound sources are present, or, if moving sound sources are to be tracked. Polynomial beamformers, however, allow for an efficient, easy, and flexible steering of the beam to any desired direction. This type of beamformer consists of a set of filter-and-sum units and a polynomial postfilter. Typically, the filter coefficients of the system are obtained by a numerical design procedure, where several desired responses for different predefined prototype look directions are simultaneously approximated by the responses of the polynomial beamformer. Steering the beam to intermediate directions is then achieved by interpolation, where the prototype look directions act as sampling points. The optimization problem is usually overdetermined and the resulting set of equations is therefore minimized in the least-squares sense. In order to minimize the distortion of the desired signal, a distortionless response constraint is often added to each prototype look direction. Also, the white noise gain must be constrained in the majority of cases to obtain a robust design. Although these constraints are necessary they reduce the number of available degrees of freedoms.

The motivation for this work is to investigate the possibility of improving the performance of polynomial beamformers by exploiting certain structures of microphone arrays—similar to what Lai et al. already proposed in [LNL10]. The authors showed that, for uniformly spaced spiral arrays, it is sufficient to design the polynomial beamformer in a restricted angular region only. Then, the same number of degrees of freedom can be used to optimize the responses for only a part of the entire angular region. During operation, the beam can be steered out of this area by rotating the set of filters to the neighboring microphones. In this thesis we propose a method which can be applied

to any type of symmetric array. The set of filters is then to be mirrored to the opposite microphone instead of rotated to the neighboring one. Since symmetric arrays include uniformly spaced spiral arrays as a special case our proposal is more general compared to [LNL10].

This thesis is structured as follows: In Chapter 2 we make some general assumptions regarding wave propagation which are valid throughout the entire document. After introducing microphone arrays we explain filter-and-sum beamforming and polynomial beamforming. A design method for robust beamformers is presented.

In Chapter 3 we propose a method which exploits the structure of symmetric arrays, and thereby allows for a reduction of the area in which the prototype look directions are distributed. Hence, the polynomial beamformer is designed for a limited angular region only. The principle is explained by way of example for symmetric linear and symmetric circular arrays.

Simulation results for different microphone arrays are shown in Chapter 4. The proposed design method is evaluated in terms of the mean square error between the desired response and the actual beamformer response. Beampatterns of the different polynomial beamformer designs are compared. Also, the impact of the choice of the prototype look directions is investigated.

Finally, the thesis is summarized and concluded in chapter 5. Suggestions for future work are presented.

2. Fundamentals

In this chapter some general assumptions regarding wave propagation are made, which are valid for the whole document. After a short introduction to microphone arrays two types of beamforming schemes are explained, namely filter-and-sum beamforming and polynomial beamforming. Finally, a design method for robust (polynomial) beamformers is presented.

2.1 Wave Propagation

Throughout this work it is assumed that the local pressure changes caused by sound waves are small relative to the static atmospheric pressure. The propagation medium, air, is considered to be an ideal fluid with zero viscosity, i.e., there is no dissipation of acoustic energy. Then, the propagation of sound waves can be reasonably well approximated as being linear [ZZ98, LSW08].

The sound pressure p at position $\mathbf{r} = [x, y, z]^{\mathrm{T}}$ and time t is given by the wave equation [LSW08],

$$\nabla^2 p(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} p(\mathbf{r}, t) = 0, \tag{2.1}$$

where c is the speed of sound, ∇ is the nabla operator, and the superscript $(\cdot)^{\mathrm{T}}$ denotes the transpose. A solution to this differential equation is, e.g., a spherical wave caused by a point source [ZZ98],

$$p(t, \mathbf{r}) = \widehat{p} \cdot \frac{R_0}{R} \, \mathrm{e}^{j(\omega t - \mathbf{k}^{\mathrm{T}} \mathbf{r})}, \tag{2.2}$$

where $R = \|\mathbf{r}\|_2$ is the distance to the point source, $\widehat{p}$ is the amplitude of the sound pressure at reference distance R_0, and $\omega = 2\pi f$ is the angular frequency. The wave vector $\mathbf{k} = [k_x, k_y, k_z]^{\mathrm{T}}$ points to the direction of propagation and describes the relation between angular frequency ω, speed of sound c, and wavelength λ by its magnitude

$$k = \|\mathbf{k}\|_2 = \frac{2\pi}{\lambda} = \frac{\omega}{c}, \tag{2.3}$$

which is called wave number [McC01].

Here, we assume that all sound sources are located in the far-field of the microphone array. A source is considered to be in the far field if [McC01]

$$R > \frac{2L^2}{\lambda} \tag{2.4}$$

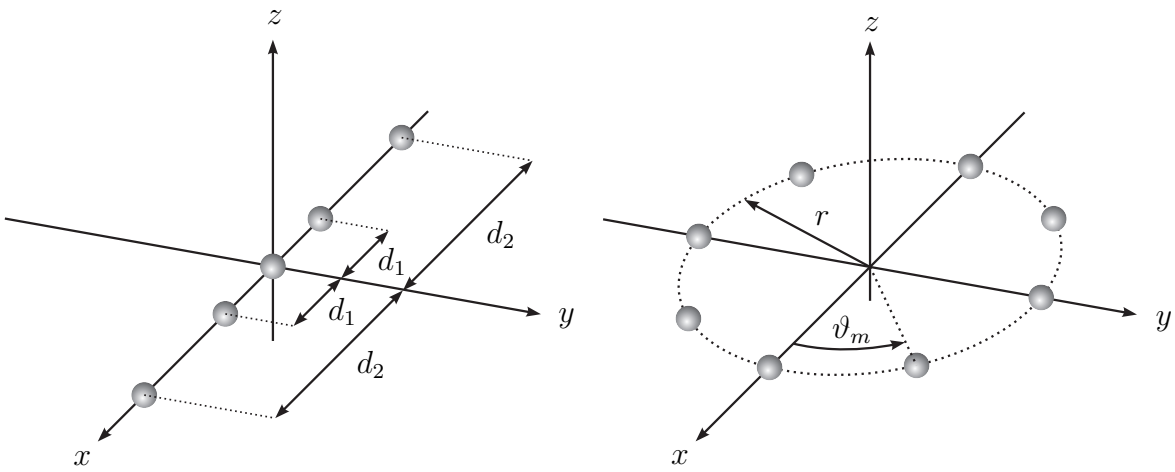
holds, where L is the longest dimension of the array. In this case, the sound wave arriving at the microphone array can be approximated as a plane wave [ZZ98],

$$p(t, \mathbf{r}) = \widehat{p} \cdot \mathrm{e}^{j(\omega t - \mathbf{k}^{\mathrm{T}}\mathbf{r})} \,. \tag{2.5}$$

Note that the magnitude of the pressure is then no longer dependent on the distance R. However, the position $\mathbf{r}$ has still an impact on the phase.

2.2 Microphone Arrays

The task of a microphone array is to pick up sound waves and spatially sample an acoustic field. Depending on the application, the geometry of an array and also the type of sensors varies [Ear04]. In some cases it is necessary to have a 3-dimensional array, i.e., sensors along the Cartesian x-, y-, and z-axis. However, a planar array (2-D) or even a linear array (1-D) consisting of omnidirectional sensors is often sufficient, e.g., in a typical distant-talking speech interface. This work focuses on symmetric linear and symmetric circular arrays (with omnidirectional sensors) as depicted in Fig. 2.1.



(a) Linear (b) Circular

Fig. 2.1: Examples of a symmetric linear array (a) and a symmetric circular array with radius r and uniform spacing (b).

Linear Array Let us consider a monochromatic plane wave $p(t, \mathbf{r})$ traveling in the x-y-plane incident on a linear array at angle $\vartheta < 90°$ as can be seen in Fig. 2.2. In this case the sound pressure is independent of the z-coordinate and the wavevector reads $\mathbf{k} = [-k\cos(\vartheta), -k\sin(\vartheta), 0]^{\mathrm{T}}$ (see [VT02]). Since all points of interest (sensors) lie on the x-axis, i.e., $\mathbf{r} = [x_m, 0, 0]^{\mathrm{T}}$, where m is the index of the respective sensor, we can simplify Eq. 2.5 to

$$p(t, x_m) = \widehat{p}\,\mathrm{e}^{j(\omega t + k\cos(\vartheta)x_m)}, \tag{2.6}$$

and thereby reduce the problem to one dimension.

First, we look at a particular wavefront which is at time $t = 0$ at a reference position $x_{\mathrm{ref}} = 0$, i.e., the center of the array. The argument (phase) of the complex harmonic is in this case equal to zero. We are then interested in the point in time τ_m which results in the same relative phase (zero) at sensor position x_m. In other words, we look for the time it takes the wave to travel from the array center to the m-th sensor. This can easily be done by plugging $t = \tau_m$ into Eq. (2.6) and equating the phase to zero, i.e., $\omega\tau_m + k\cos(\vartheta)x_m \overset{!}{=} 0$. Replacing k by Eq. (2.3) yields for the desired time

$$\tau_m(\vartheta) = -\frac{x_m}{c}\cos(\vartheta), \tag{2.7}$$

which is called propagation delay. That means, the sound pressure $p(t, x_m)$ at position x_m is equal to the pressure at the array center delayed by $\tau_m(\vartheta)$, $p(t - \tau_m(\vartheta), x_{\mathrm{ref}})$. More precisely, it is only a delay if $\tau_m(\vartheta) > 0$. Otherwise $p(t, x_{\mathrm{ref}})$ is delayed relative to $p(t, x_m)$.

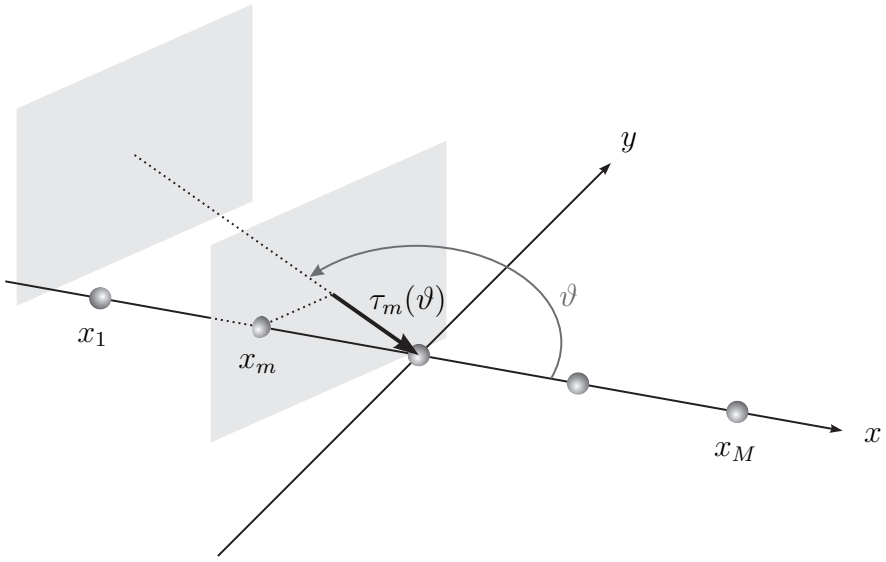


Fig. 2.2: Illustration of a plane wave arriving at a linear microphone array at incident angle ϑ.

In the following, all sensors are assumed to be ideal, i.e., the sensitivity is constant for all directions, the frequency response is flat, of equal phase, and there is no inherent sensor self-noise. Then, all sensor signals are delayed versions of each other and can be expressed in terms of a reference signal at the array center, $s_{\mathrm{ref}}(t) = p(t, x_{\mathrm{ref}} = 0)$. The time signal of the m-th sensor can thus be written as

$$s_m(t) = s_{\mathrm{ref}}(t - \tau_m(\vartheta)), \tag{2.8}$$

where the dependency on ϑ is omitted on the left side. Note that another plane wave incident from $\vartheta^* = -\vartheta$ would cause exactly the same delays and sensor signals, respectively, which follows directly from Eq. (2.7) and the symmetry of the cos-function. This is referred to as the forward-backward ambiguity [Teu07].

Circular Array Analogous to the scenario above, we consider ideal sensors and a monochromatic plane wave $p(t, \mathbf{r})$, which propagates in the x-y-plane and impinges on a circular array at incident angle ϑ. The sensor signals are then again delayed versions of each other and can be written with respect to a (now virtual) reference signal, $s_{\mathrm{ref}}(t) = p(t, \mathbf{r}_{\mathrm{ref}})$. The reference position $\mathbf{r}_{\mathrm{ref}} = \mathbf{0}$ is again chosen to be the array center.

In case of circular arrays the propagation delays do not merely depend on the direction of arrival ϑ, but rather on the difference between ϑ and the respective angular position ϑ_m of the sensor. For the m-th sensor the relative signal delay reads

$$\tau_m(\vartheta) = -\frac{r}{c}\cos(\vartheta - \vartheta_m), \tag{2.9}$$

where r is the array radius.

It should be mentioned that knowledge about τ_m is of fundamental importance for this work, especially for section 2.3 and chapter 3.

In the explanation above an isolated sound source and ideal sensors are assumed. In practice, however, there is always inherent sensor self-noise and/or other interfering sources are present. These undesired signal components are then supposed to be suppressed. One way of doing this is to time align the desired signal component and to properly add the sensor signals. This leads us directly to beamforming schemes which are discussed in the next section.

2.3 Beamforming

Beamforming is a widely used spatial filtering technique in digital signal processing [VVB88, VT02, McC01, MSK09]. The goal is to enhance the desired signal incident from a particular direction and suppress noise/interfering sound sources located at

other angles. This can be achieved by spatially sampling the sound field using a microphone array, and thereby exploiting its spatial properties. The basic idea of many beamforming schemes is to make use of the frequency- and space-dependent delays $\tau_m(\vartheta)$ between the microphone signals. More precisely, the goal is to properly time align the desired signal in each microphone channel in order to allow for a coherent, constructive summation of the desired components, whereas noise and interfering sources from other directions are combined incoherently/destructive.

2.3.1 Filter-and-Sum Beamforming

In the following, we want to discuss the filter-and-sum beamformer (FSB) and analyze the response of such a system. As depicted in Fig. 2.3, each sensor signal is processed by a linear, time-invariant FIR filter of length-L with impulse response coefficients $\mathbf{w}_m = [w_m[0], \ldots, w_m[L-1]]^{\mathrm{T}}$. The signals $s_m[n]$ are discrete-time representations of the sound pressure $p(t, \mathbf{r}_m)$ at microphone position $\mathbf{r}_m$. Then, the output $y[n]$ of the beamformer is given by a convolution sum,

$$y[n] = \sum_{m=1}^{M} \sum_{l=0}^{L-1} w_m[l] s_m[n-l], \tag{2.10}$$

where M is the number of microphones. In the frequency domain, this reads

$$Y(\omega) = \sum_{m=1}^{M} W_m(\omega) S_m(\omega), \tag{2.11}$$

where

$$W_m(\omega) = \sum_{l=0}^{L-1} w_m[l] \, \mathrm{e}^{-j\omega/f_{\mathrm{s}} l} \tag{2.12}$$

and

$$S_m(\omega) = \sum_{n=-\infty}^{\infty} s_m[n] \, \mathrm{e}^{-j\omega/f_{\mathrm{s}} n} \tag{2.13}$$

are discrete-time Fourier transforms (DTFTs) of $\mathbf{w}_m$ and $s_m[n]$, respectively, and f_{s} denotes the sampling frequency.

We now examine the scenario of section 2.2, i.e., a unit plane wave of angular frequency ω traveling in the x-y-plane with incident angle ϑ. Using relation (2.8) with $p(t, \mathbf{r}_{\mathrm{ref}} = \mathbf{0}) = \mathrm{e}^{j\omega t}$, the m-th sensor signal reads in discrete-time domain

$$s_m[n] = s_{\mathrm{ref}}[n - f_{\mathrm{s}} \tau_m(\vartheta)] = \mathrm{e}^{j\omega(n/f_{\mathrm{s}} - \tau_m(\vartheta))} \tag{2.14}$$

and Eq. (2.10) becomes

$$y[n] = \sum_{m=1}^{M} \sum_{l=0}^{L-1} w_m[l]\, \mathrm{e}^{j\omega((n-l)/f_\mathrm{s} - \tau_m(\vartheta))}\,. \tag{2.15}$$

We can then easily obtain the transfer function of the filter-and-sum beamformer,

$$B\left(\omega, \vartheta\right) := \sum_{m=1}^{M} W_m\left(\omega\right) \mathrm{e}^{-j\omega\tau_m(\vartheta)}, \tag{2.16}$$

which is called **beamformer response** [VVB88]. Note that it depends on both angle and frequency. In a more compact matrix notation, the beamformer response reads

$$B\left(\omega, \vartheta\right) = \mathbf{w}(\omega)^\mathrm{T}\mathbf{a}(\omega), \tag{2.17}$$

where $\mathbf{w}(\omega) = [W_1(\omega), \ldots, W_M(\omega)]^\mathrm{T}$ is a vector containing the frequency responses of the FIR filters $\mathbf{w}_m$, and $\mathbf{a}(\omega) = [\exp\left(-j\omega\tau_1(\vartheta)\right), \ldots, \exp\left(-j\omega\tau_M(\vartheta)\right)]^\mathrm{T}$ is the **steering vector**. The magnitude square of the beamformer response is known as the **beampattern** [BW01].

The simplest way to obtain the reference signal $s_\mathrm{ref}[n]$ at the output is to compensate for the relative delays $\tau_m(\vartheta)$ between the microphone signals, i.e., to apply an additional delay $\tilde{\tau}_m(\vartheta) = -\tau_m(\vartheta)$ to the m-th channel in order to allow for a coherent summation. In matrix notation this reads $\mathbf{w}(\omega) = \frac{1}{M}\mathbf{a}^\mathrm{H}(\omega)$, where $\frac{1}{M}$ is a scaling factor and $(\cdot)^\mathrm{H}$ denotes conjugate complex and transpose (Hermitian operator). Note that a further delay needs to be included in order to have a causal system. Merely delaying and adding all microphone signals with equal weight is referred to as delay-and-sum beamforming (DSB), which is a subclass of filter-and-sum beamforming.

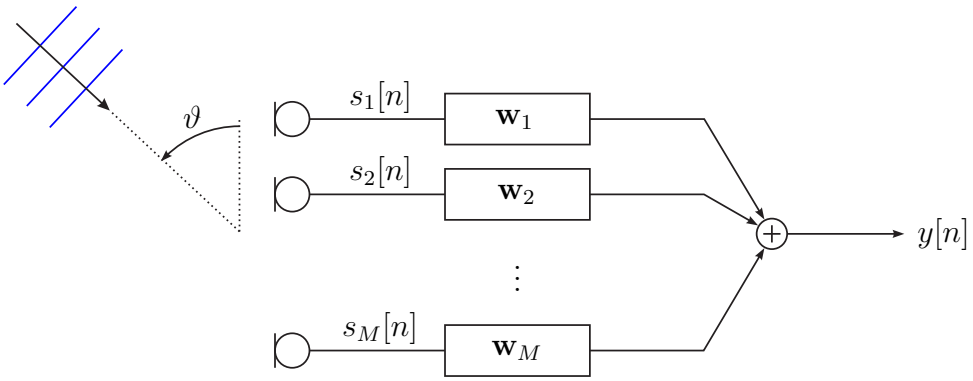
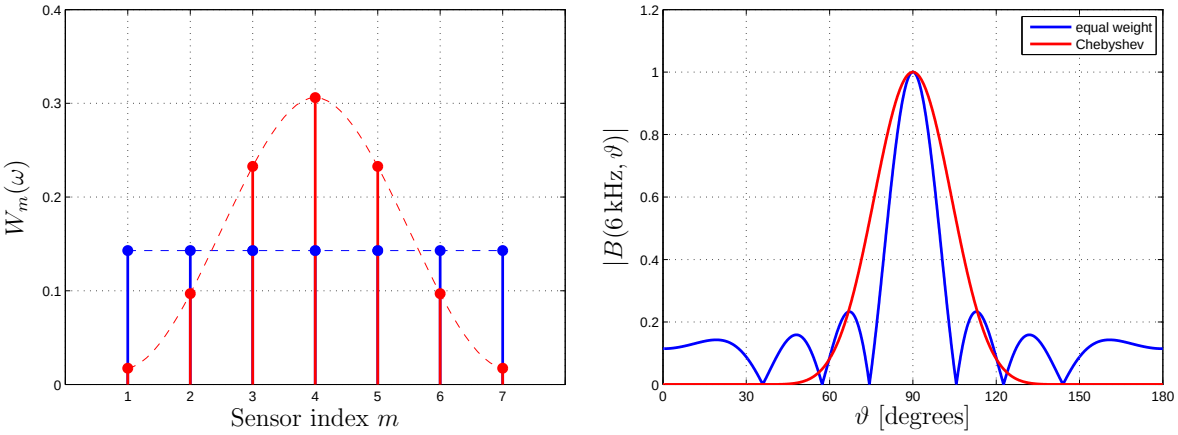


Fig. 2.3: Filter-and-sum beamformer consisting of M microphones with a plane wave incident from ϑ.

In general, however, the respective signals are weighted by different frequency-dependent gain factors. This offers more flexibility when it comes to designing a certain beamformer response. The effect of differently weighting the signals is shown in the following. As the beamformer response from Eq. (2.16) can be interpreted as a spatial DTFT of $W_m(\omega)$[1], we can apply the knowledge of the relations between time domain and frequency domain in order to influence the spatial characteristics of beamformers. Hence, it is also reasonable to weight the channels according to spatial window functions.

As an example, we want to compare equal weights $W_m(\omega) = 1/M$ on the one hand, with a Chebyshev window [VT02] on the other hand, as depicted in Fig. 2.4(a). The desired sound source is located at $\vartheta_{\text{des}} = 90°$ which means that all relative delays are equal to zero according to Eq. (2.7). Fig. 2.4(b) shows the corresponding magnitude responses for a linear array with $M = 7$ microphones, spacing $d = 3\,\text{cm}$, and $f = 6\,\text{kHz}$. In both cases the desired signal incident from $\vartheta = 90°$ remains undistorted whereas other directions are attenuated. However, applying weights according to a Chebyshev window significantly reduces the side lobes compared to equal gain factors, which is achieved at the cost of a wider main lobe. That is, not only the delays, but also the gain factors highly affect the behavior of a beamformer.



(a) Array weights (b) Magnitude responses

Fig. 2.4: Beamformer responses for a linear array ($M = 7$ microphones, spacing $d = 3\,\text{cm}$), frequency $f = 6\,\text{kHz}$, and different array weights $W_m(\omega)$.

[1] The sum is not computed along discrete-time index n, but space index m.

2.3.2 Polynomial Beamforming

The filter-and-sum beamforming scheme described in the previous section allows for enhancement of signals arriving from a desired, but fixed direction. In many applications, however, a flexible look direction is desirable, e.g., when tracking moving sound sources. Another example is teleconferencing, where several speakers of interest are located at distinct positions and talk at different times.

Changing the response of an FSB requires an update of all filter coefficients. Adapting the filters in real-time is done in adaptive optimal beamforming schemes (see, e.g., [VT02]). Using data-independent beamformers, as it is discussed here, an instantaneous computation of new filters is not feasible in most real-time applications as this is relatively time-consuming. A possible alternative is to precompute an extensive set of filters for all desired look directions in advance and use look-up tables during operation. However, this would require a huge memory and is inflexible. A different approach to this issue is polynomial beamforming [MK10].

A polynomial beamformer comprises $P+1$ fixed filter-and-sum units (FSUs) and a polynomial postfilter (PPF) of order P, as can be seen in Fig. 2.5. The impulse response coefficients of the FIR filter of the m-th microphone channel in the p-th filter-and-sum unit are denoted as $\mathbf{w}_{m,p} = [w_{m,p}[0], \ldots, w_{m,p}[L-1]]^{\mathrm{T}}$. The beamformer response is obtained by a straightforward extension of Eq. (2.16),

$$B_D(\omega, \vartheta) = \sum_{p=0}^{P} D^p \sum_{m=1}^{M} W_{m,p}(\omega) e^{-j\omega\tau_m(\vartheta)}, \tag{2.18}$$

where

$$W_{m,p}(w) = \sum_{l=0}^{L-1} w_{m,p}[l]\, \mathrm{e}^{-j\omega/f_{\mathrm{s}}l} \tag{2.19}$$

is the frequency response of $\mathbf{w}_{m,p}$, D denotes the steering direction and $\tau_m(\vartheta)$ is the relative delay between the microphone signals given by Eqs. (2.7) and (2.9) for linear and circular arrays, respectively. Note that for the special case $P = 0$ the polynomial beamformer degenerates to a filter-and-sum beamformer.

One can interpret the PPF as an interpolator between different FSU output signals $y_p[n]$, and hence between different FSU responses. The idea of polynomial beamforming is to keep all filter coefficients constant and to obtain a different look direction by interpolation, i.e., by simply changing the scalar value D. This parameter depends on the maximum steering angle, $\vartheta_{\max}$, which also defines the steering range $\vartheta_{\mathrm{SR}} = [0°, \vartheta_{\max}]$. In order to cover all possible look directions, $\vartheta_{\mathrm{SR}} = [0°, 360°[$ in case of a circular array. For a linear array the steering range is limited to $\vartheta_{\mathrm{SR}} = [0°, 180°]$ due to the forward-backward ambiguity explained in section 2.2.

The steering direction is defined as

$$D = \frac{\vartheta_{\text{des}} - \vartheta_{\max}/2}{\vartheta_{\max}/2}, \quad 0^\circ \leq \vartheta_{\text{des}} \leq \vartheta_{\max}, \tag{2.20}$$

and therefore limited to $-1 \leq D \leq 1$. For a proper choice of filters $W_{m,p}(w)$ it is then possible to flexibly steer the beam to different directions by varying D, which can easily be done in real-time applications. A method of how to design such a beamformer will be discussed subsequently.

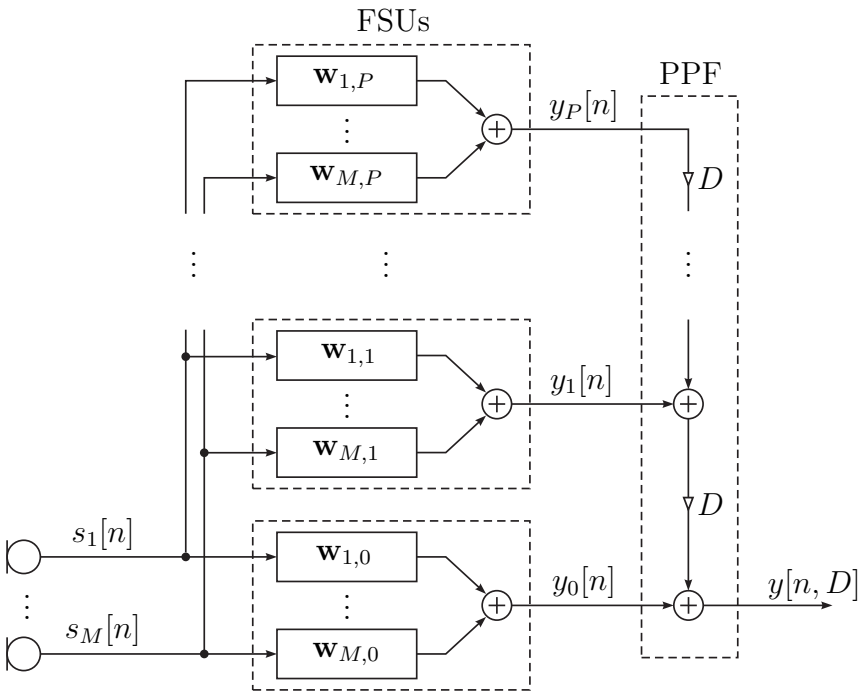


Fig. 2.5: Polynomial beamformer of order P with M microphones.

2.3.3 Design of Robust Beamformers

In this section a design procedure of robust beamformers is described. To start with, we have to define some performance measures and what "robust" actually means. Until now, the microphones have been assumed to be ideal, i.e., the frequency response is flat and identical for all directions, and there is no inherent self-noise. In practice, however, the magnitude and phase responses among the microphones differ, and electrical noise is present. Furthermore, the microphones are typically placed slightly off the desired position, which means that the sound field is not sampled exactly at the intended spot.

Especially in the case of small arrays and low frequencies[2] the microphones should be located as accurately as possible.

As a first performance measure we introduce the **directivity**. So far we have only considered the special case of planar arrays and waves traveling parallel to this plane, i.e., the elevation angle $\phi = 90°$. Then, the direction of arrival of a wave is completely defined by the azimuth angle ϑ. However, waves can incident from any direction in general and the comprehensive definition of the directivity reads [VT02]

$$D(\omega) = \frac{|B(\omega, \vartheta_{\text{des}}, \phi_{\text{des}})|^2}{\frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi} |B(\omega, \vartheta, \phi)|^2 \sin(\vartheta) \mathrm{d}\phi \mathrm{d}\vartheta}, \tag{2.21}$$

where the numerator represents the power of a plane wave arriving from the desired direction $(\vartheta_{\text{des}}, \phi_{\text{des}})$ and the denominator corresponds to the average output power for the entire angular range. A directivity of $D(\omega) = 1$ means that the system response is independent of direction. On the other hand, high directivity values imply a high angular selectivity. That is, all signals are strongly attenuated except for the ones incident from a small angular region around $(\vartheta_{\text{des}}, \phi_{\text{des}})$[3].

Another measure deduced from the directivity is the **white noise gain** (WNG). It expresses the ability of suppressing spatially white noise and reads for a filter-and-sum beamformer [Her05]

$$A(\omega) = \frac{\left|\mathbf{w}^{\mathrm{T}}(\omega)\mathbf{a}_{\text{des}}(\omega)\right|^2}{\mathbf{w}^{\mathrm{H}}(\omega)\mathbf{w}(\omega)}, \tag{2.22}$$

where $\mathbf{w}(\omega) = [W_1(\omega), \ldots, W_M(\omega)]^{\mathrm{T}}$ contains the frequency responses of the filters $\mathbf{w}_m$ and $\mathbf{a}_{\text{des}}(\omega) = [\exp(-j\omega\tau_1(\vartheta_{\text{des}})), \ldots, \exp(-j\omega\tau_M(\vartheta_{\text{des}}))]^{\mathrm{T}}$. Even though this measure is called "gain", it actually describes how much white noise gets attenuated relative to the response in the desired look direction, i.e., only negative values of $A(\omega)$ correspond to a true amplification of white noise. The WNG is an important measure as it defines the amplification of inherent microphone self-noise, which is assumed to be uncorrelated or spatially white, respectively. Also, misplacement and errors in the magnitude and/or phase responses are nearly uncorrelated from sensor to sensor and can therefore be modeled as white noise. A beamformer may perform very well in theory, but may be useless in practice if its white noise gain is very small, and hence the output signal is too noisy. A beamformer is called robust if its white noise gain is high and the mentioned irregularities have only small impact on the beamformer response.

We now start with the actual design scheme of robust beamformers as described in [MSK09, MK10]. This is done for a polynomial beamformer here, as it includes the filter-and-sum beamformer as a special case. First, we define multiple desired frequency-invariant responses $B_{\text{des},i}(\vartheta, \vartheta_{\text{des}_i})$, $i = 1, \ldots, I$, with beams pointing to dif-

[2] Typically, the array weights are large for low frequencies.

[3] For all subsequent considerations we exclusively consider the special case $\phi = 90°$ again.

ferent so-called prototype look directions (PLDs), $\vartheta_{\mathrm{des}_i}$. We term the angular region comprising the prototype look directions **PLD range**. The goal is to simultaneously approximate the desired responses for the PLDs by the responses of a polynomial beamformer, $B_{D_i}(\omega, \vartheta)$. For a numerical solution we discretize the frequency axis into Q frequencies ω_q, $q = 0, \ldots, Q-1$, and the angular axis into K angles ϑ_k, $k = 0, \ldots, K-1$. The resulting set of linear equations is then typically resolved numerically. Formally, this reads

$$B_{\mathrm{des},i}(\vartheta_k, \vartheta_{\mathrm{des}_i}) \stackrel{!}{=} \sum_{p=0}^{P} D_i^p \sum_{m=1}^{M} W_{m,p}(\omega_q) e^{-j\omega_q \tau_m(\vartheta_k)}, \quad \forall i = 0, \ldots, I-1. \tag{2.23}$$

In matrix notation this becomes

$$\mathbf{b}_{\mathrm{des}_i} \stackrel{!}{=} \mathbf{A}(\omega_q)\mathbf{W}(\omega_q)\mathbf{d}_i, \quad \forall i = 0, \ldots, I-1, \tag{2.24}$$

where $\mathbf{b}_{\mathrm{des}_i} = [B_{\mathrm{des},i}(\vartheta_0, \vartheta_{\mathrm{des}_i}), \ldots, B_{\mathrm{des},i}(\vartheta_{K-1}, \vartheta_{\mathrm{des}_i})]^{\mathrm{T}}$, $[\mathbf{A}(\omega_q)]_{km} = \exp\left(-j\omega_q \tau_m(\vartheta_k)\right)$, $[\mathbf{W}(\omega_q)]_{mp} = W_{m,p}(\omega_q)$, and $\mathbf{d}_i = [D_i^0, \ldots, D_i^P]^{\mathrm{T}}$. If the number of discretized angles K is greater than the product of the number of microphones M and the number of FSUs $P+1$, i.e., $K > M(P+1)$, the problem is overdetermined. This is typically the case and the beamformer design problem reads

$$\min_{\mathbf{W}(\omega_q)} \sum_{i=0}^{I-1} \left\| \mathbf{A}(\omega_q)\mathbf{W}(\omega_q)\mathbf{d}_i - \mathbf{b}_{\mathrm{des}_i} \right\|_2^2, \quad \forall q = 0, \ldots, Q-1. \tag{2.25}$$

We now formulate a vector $\mathbf{w}(\omega_q) = [W_{1,0}(\omega_q), \ldots, W_{1,P}(\omega_q), W_{2,0}(\omega_q), \ldots, W_{M,P}(\omega_q)]^{\mathrm{T}}$ consisting of the entries of matrix $\mathbf{W}(\omega_q)$. With an identity matrix $\mathbf{I}_M$ of size $M \times M$ we can write

$$\begin{aligned} \mathbf{W}(\omega_q)\mathbf{d}_i &= \mathbf{I}_M \otimes \mathbf{d}_i^{\mathrm{T}} \mathbf{w}(\omega_q) \\ &= \mathcal{D}_i \mathbf{w}(\omega_q), \end{aligned} \tag{2.26}$$

where $\otimes$ denotes the Kronecker product and $\mathcal{D}_i$ is of size $M \times M(P+1)$. Then, Eq. (2.25) becomes

$$\min_{\mathbf{w}(\omega_q)} \sum_{i=0}^{I-1} \left\| \mathbf{A}(\omega_q)\mathcal{D}_i\mathbf{w}(\omega_q) - \mathbf{b}_{\mathrm{des}_i} \right\|_2^2, \quad \forall q = 0, \ldots, Q-1. \tag{2.27}$$

Introducing $\mathcal{A}(\omega_q) = [\mathbf{A}(\omega_q)\mathcal{D}_0, \ldots, \mathbf{A}(\omega_q)\mathcal{D}_{I-1}]^{\mathrm{T}}$ and $\mathcal{B}_{\mathrm{des}} = [\mathbf{b}_{\mathrm{des}_0}^{\mathrm{T}}, \ldots, \mathbf{b}_{\mathrm{des}_{I-1}}^{\mathrm{T}}]^{\mathrm{T}}$ we finally obtain the classical form of a minimization problem in matrix notation,

$$\min_{\mathbf{w}(\omega_q)} \left\| \mathcal{A}(\omega_q)\mathbf{w}(\omega_q) - \mathcal{B}_{\mathrm{des}} \right\|_2^2, \quad \forall q = 0, \ldots, Q-1. \tag{2.28}$$

If some prototype look directions or angular regions are more important than others, one can add weights to Eq. (2.28) and obtain a weighted least squares minimization problem,

$$\min_{\mathbf{w}(\omega_q)} \left\| \mathcal{W} \left(\mathcal{A}(\omega_q)\mathbf{w}(\omega_q) - \mathcal{B}_{\text{des}} \right) \right\|_2^2, \quad \forall q = 0, \ldots, Q-1, \tag{2.29}$$

where $\mathcal{W}$ is a diagonal matrix of size $KI \times KI$ containing the respective weights. If $\mathcal{W}$ is an identity matrix all angles are weighted equally.

Next, we add I constraints to the optimization problem in order to ensure that the signals arriving from the I desired angles ϑ_{des_i} remain undistorted. That is,

$$\mathbf{a}_{\text{des}_i}^{\text{T}}(\omega_q)\mathcal{D}_i\mathbf{w}(\omega_q) \overset{!}{=} 1, \quad \forall i = 0, \ldots, I-1, \tag{2.30}$$

must be satisfied, where $\mathbf{a}_{\text{des}_i}(\omega_q) = \left[\exp\left(-j\omega_q\tau_1(\vartheta_{\text{des}_i})\right), \ldots, \exp\left(-j\omega_q\tau_M(\vartheta_{\text{des}_i})\right)\right]^{\text{T}}$.

In order to make the beamformer robust, a lower bound γ of the WNG can be added as a second set of constraints,

$$\frac{\left|\mathbf{a}_{\text{des}_i}^{\text{T}}(\omega_q)\mathcal{D}_i\mathbf{w}(\omega_q)\right|^2}{\left\|\mathcal{D}_i\mathbf{w}(\omega_q)\right\|_2^2} \overset{!}{\geq} \gamma, \quad \forall i = 0, \ldots, I-1. \tag{2.31}$$

The parameter γ can be defined by the user according to the predefined robustness requirement which depends, e.g., on the quality of the applied microphones. Adjusting γ enables a tradeoff between robustness and spatial selectivity. This implies that the more restrictive the constraints are, the higher are the deviations of the obtained design from the desired solution. Note that $\gamma \leq M$ must hold as this is the maximum achievable value of the WNG [VT02]. Otherwise the optimization problem is infeasible and a solution cannot be obtained.

After minimizing Eq. (2.29) subject to Eqs. (2.30) and (2.31) the resulting frequency responses $W_{m,p}(\omega)$ need to be approximated by FIR filters as a last step.

Note that the optimization is done exclusively for I prototype look directions. The resulting filters are then fixed and steering the beam to intermediate directions is achieved by interpolation, where the PLDs act as sampling points. Thus, the performance of a polynomial beamformer strongly depends on the choice of the position and number of PLDs, as will be shown in section 4.3.

3. Sectorial Optimization applied to Polynomial Beamforming

In polynomial beamforming the prototype look directions (PLDs) are typically distributed in the entire angular region in order to be able to steer the beam to any desired direction. The following chapter seizes on the idea of limiting the PLD range to only a part of the whole angular region, as proposed by Lai et al. [LNL10] for uniformly spaced spiral arrays. The same number of prototype look directions can then be used to cover a smaller angular region. As a consequence, the angular distance between these prototype look directions, which act as sampling points for interpolation, is decreased. Also, all available degrees of freedom are then used to optimize the responses for a smaller angular region. This concept is referred to as sectorial optimization. Simply adding more PLDs in order to have a finer sampling grid is often not desirable due to the increase in complexity when designing the beamformer (see section 2.3.3). Also, the number of PLDs is related to the order of the polynomial postfilter, which should be as small as possible if, for example, an acoustic echo canceller is applied to the beamformer [HM07]. In this thesis we extend the method proposed by Lai et al. and present a more general optimization technique, which can be applied to any type of symmetric array. It also covers the special case of uniformly spaced spiral arrays and allows for a further reduction of the PLD range compared to [LNL10]. The presented method is explained by way of example for symmetric linear and symmetric circular arrays.

3.1 Symmetric Linear Array

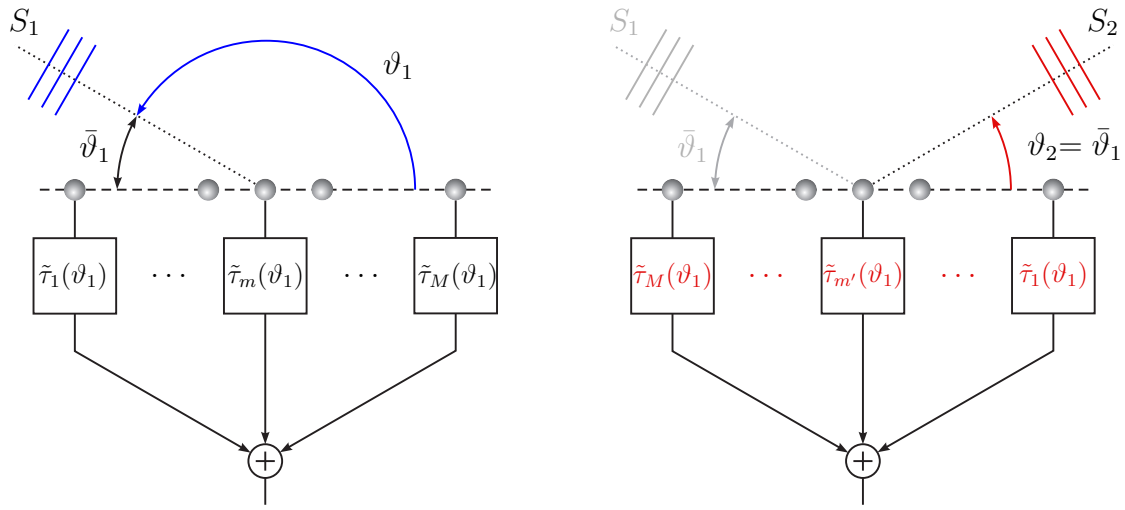
First, we examine a monochromatic plane wave S_1 incident on a symmetric linear array with microphones placed along the x-axis, as depicted in Fig. 3.1(a). The wave arrives at an angle ϑ_1, where $90° < \vartheta_1 \leq 180°$. Assuming ideal microphones, the microphone signals are delayed versions of each other, as shown in section 2.2. S_1 is considered as a desired source signal, which is supposed to be enhanced. In order to do this, we need to time-align the signals and apply an additional delay $\tilde{\tau}_m(\vartheta_1) = -\tau_m(\vartheta_1)$ to the m-th channel according to section 2.3.1. This basically creates a beam pointing towards the desired direction ϑ_1.

Now, we look at another plane wave S_2 of the same frequency arriving from $\vartheta_2 = \bar{\vartheta}_1 = 180° - \vartheta_1$, where $\bar{\vartheta}_1$ is the complementary angle of ϑ_1 (see Fig. 3.1(b)).

Introducing $m' = M - m + 1$, $m = 1, \ldots, M$, which reverses the order of m, and $x_{m'} = -x_m$, which denotes the distance from the m'-th microphone to the array center, the relative delay of the m-th microphone signal reads with Eq. (2.7)

$$\begin{aligned}\tau_m(\vartheta_2) &= -\frac{x_m}{c}\cos\left(180^\circ - \vartheta_1\right)\\ &= -\frac{x_m}{c}\left(\cos\left(180^\circ\right)\cos\left(\vartheta_1\right) + \sin\left(180^\circ\right)\sin\left(\vartheta_1\right)\right)\\ &= \frac{x_m}{c}\cos\left(\vartheta_1\right) = -\frac{x_{m'}}{c}\cos\left(\vartheta_1\right)\\ &= \tau_{m'}(\vartheta_1). \end{aligned} \tag{3.1}$$

That is, the relative delays $\tau_m(\vartheta_2)$ caused by a plane wave arriving from ϑ_2 are mirrored compared to the case of a wave incident from ϑ_1. If we also mirror the additional delays $\tilde{\tau}_m(\vartheta_1)$ in the same way, which we apply in order to time-align the signal S_1, i.e., $\tilde{\tau}_m(\vartheta_2) = \tilde{\tau}_{m'}(\vartheta_1)$, the output signal S_2 can be time-aligned. In other words, the beam can be steered from ϑ_1 to ϑ_2 when swapping the delays applied to the channels. By doing this, we can limit the design of the polynomial beamformer to the angular region $[0^\circ, 90^\circ]$, which means that the PLD range ϑ_{PLD} is reduced by a factor of two.



(a) Real source position (b) Corresponding source in mirrored setup

Fig. 3.1: Illustration of time-aligning signals arriving from different directions by applying proper delays to the microphone channels. If mirroring the delays $\tilde{\tau}_m(\vartheta_1)$ to the opposite microphone channel the beam pointing to direction ϑ_1 can be steered to $\vartheta_2 = \bar{\vartheta}_1$, where $0^\circ \leq \vartheta_2 < 90^\circ$. This allows for a reduction of the PLD range to one single quadrant.

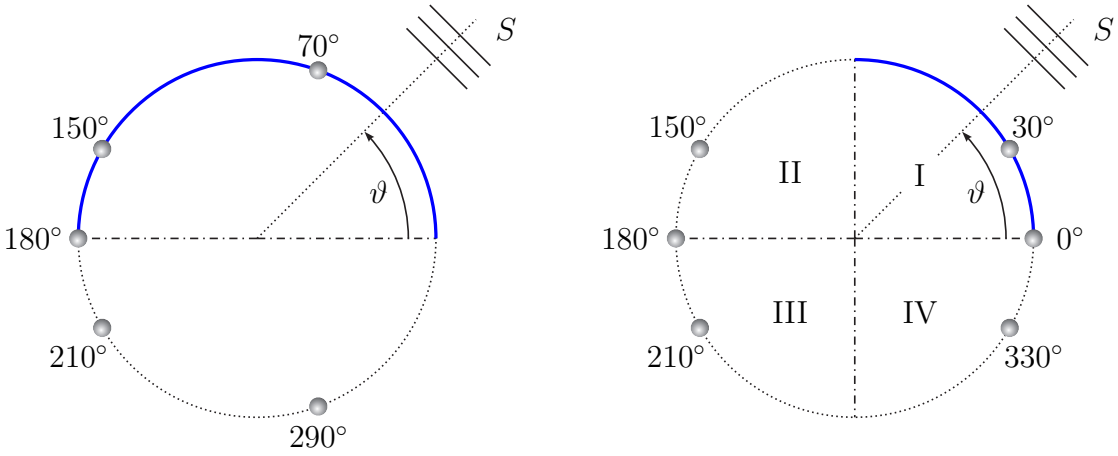
For sake of simplicity the idea of exploiting the symmetry of an array is explained for a delay-and-sum beamformer. However, it also works for broadband signals, i.e., instead of mirroring mere delays, one has to swap FIR filters applied to the microphone signals.

Note that this method is not restricted to symmetric linear arrays, but it can be applied to any type of symmetric arrays. As another example, this will be shown for symmetric circular arrays in the next section.

3.2 Symmetric Circular Array

The idea of exploiting the structure of an array in order to reduce the PLD range can also be applied to symmetric circular arrays. Fig. 3.2 shows two examples of such arrays, with the dash-dotted lines indicating the symmetry axes.

First, we investigate the structure depicted in Fig. 3.2(a), which has one symmetry axis. This array can be decomposed into a single microphone at $\vartheta_m = 180°$ and two symmetric linear arrays consisting of two microphones each: {70°, 290°} and {150°, 210°}. As shown in the previous section, when mirroring the set of filters applied to a linear symmetric array the beam can be steered to the opposite (w.r.t. the symmetry axis) angle. This also holds for a symmetric circular array, i.e., a beam pointing to ϑ_1, $0 < \vartheta_1 < 180°$, can be steered to $\vartheta_2 = 360° - \vartheta_1$ by pairwise mirroring the filters to the opposite microphone[1]. Hence, designing the polynomial beamformer for the angular region between 0° and 180° is sufficient, as indicated by the blue semicircle in Fig. 3.2(a). That is, the PLD range can be reduced from $\vartheta_{\mathrm{PLD}} = [0°, 360°[$ to $[0°, 180°]$.



(a) One symmetry axis

(b) Two symmetry axes

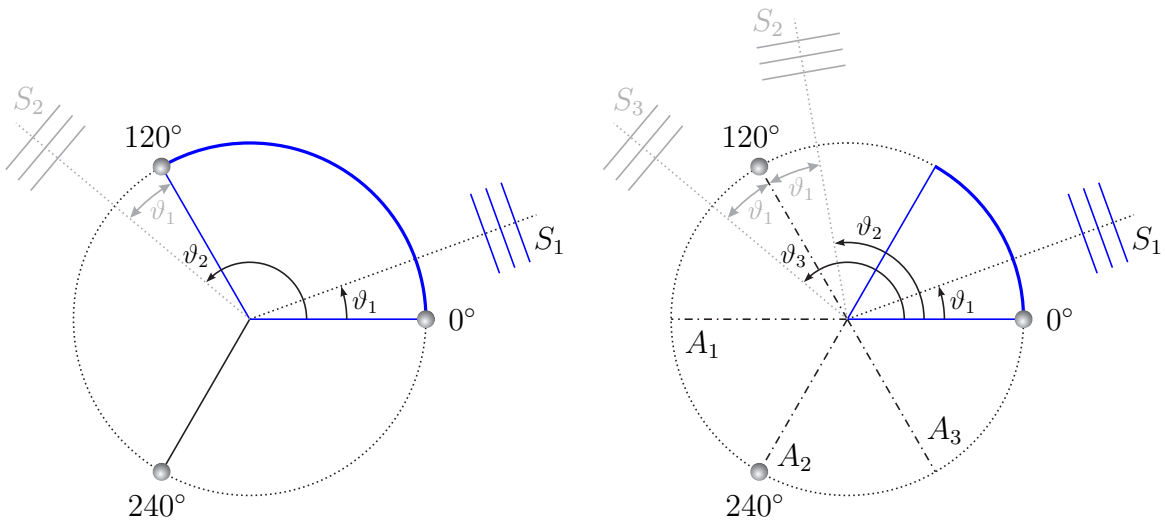
Fig. 3.2: Examples of symmetric circular arrays.

[1] The single microphone is not affected by this procedure.

In case of a second symmetry axis (see Fig. 3.2(b)) there is the possibility to either mirror the filters horizontally, or vertically, or even both. Then, the PLD range can be reduced even further to a single quadrant, i.e., $\vartheta_{\text{PLD}} = [0°, 90°]$ (see blue indication). If the beam is to be steered from $0° \leq \vartheta \leq 90°$ to the second quadrant the filters need to be mirrored horizontally, i.e., at the vertical symmetry axis. In case of quadrant IV the filters are mirrored in vertical direction, and in order to steer the beam to the third quadrant the filters need to be consecutively mirrored at both axes.

Uniformly Spaced Circular Array

A special case of a symmetric circular array is an array with uniformly spaced microphones. These arrays exhibit as many symmetry axes as number of microphones M. Lai et al. already exploited this particular structure and proposed in [LNL10] to limit the PLD range to a single sector, which is defined as the angular area between two microphones, i.e., $\vartheta_{\text{PLD}} = [0°, 360°/M]$. Steering the beam to other sectors is achieved by rotating the filters. For example, assuming an array with $M = 3$ microphones, as illustrated in Fig. 3.3(a), a beam pointing to ϑ_1, $0° \leq \vartheta_1 \leq 120°$, can be steered to $\vartheta_2 = \vartheta_1 + 120°$ by rotating the filters to the neighboring microphone. More generally, steering the beam to any angle $\vartheta_{j+1} = \vartheta_1 + j \cdot 360°/M$, $j = 1, \ldots, M-1$, $0° \leq \vartheta_1 \leq 360°/M$, can be obtained by rotating the set of filters by consecutive j steps in the anti-clockwise direction.



(a) Proposal by Lai et al.

(b) Exploiting all symmetry axes

Fig. 3.3: Reduction of PLD range (blue indication) according to the proposal of Lai et al. [LNL10] (a) and by taking all symmetry axes A_m, $m = 1, \ldots, M$, into account (b).

Taking all symmetry axes A_m, $m = 1, \ldots, M$, into account the PLD range can be reduced even further to $\vartheta_{\mathrm{PLD}} = [0°, 360°/(2M)]$. Steering the beam to any desired direction outside this area can be achieved by (consecutively) mirroring the filters. For example, if the beam is to be steered to $120° \leq \vartheta \leq 180°$ the filters need to be mirrored at A_2 followed by A_3, as illustrated in Fig. 3.3(b). It should be mentioned that all angles can be obtained by mirroring two times or less.

Note that in Fig. 3.3 we only illustrated steering angles lying in the upper half-plane for reasons of clarity and comprehensibility. Steering the beam to the lower half-plane can be obtained by an additional rotation of the filters in Fig. 3.3(a) and by mirroring the filters at A_1 in Fig. 3.3(b).

3.3 Summary

In this chapter we showed that by exploiting the symmetry of microphone arrays the PLD range which is used for designing a polynomial beamformer can be reduced, and hence the distance between the prototype look directions can be decreased. Using a symmetric linear array the PLD range can be reduced from $\vartheta_{\mathrm{PLD}} = [0°, 180°]$ to $[0°, 90°]$. In case of a symmetric circular array we can reduce ϑ_{PLD} from $[0°, 360°[$ to $[0°, 360°/(2\beta)]$, where β is the number of symmetry axes. For the special case of a uniformly spaced circular array $\beta = M$, where $M \geq 3$ in order to have a circular array. That is, the number of symmetry axes is given by the number of microphones. Accordingly, the PLD range can be limited to $\vartheta_{\mathrm{PLD}} = [0°, 360°/(2M)]$ which is a significant reduction. In comparison, the method proposed by Lai et al. [LNL10] allows for a reduction to $[0°, 360°/M]$ only.

4. Evaluation

In this chapter we evaluate the potential benefit of the presented optimization technique having telephone speech signal capture in mind as a possible application of beamforming. In the following, the frequency range of interest lies therefore between 0.3 kHz and 3.4 kHz, and a sampling frequency of 8 kHz is used. For all simulations the desired response $B_{\mathrm{des}}(\vartheta, \vartheta_{\mathrm{des}})$ has an angular passband of 10° and the stopbands start at 30° off the center, as depicted in Fig. 4.1 for $\vartheta_{\mathrm{des}} = 60°$. The order of the evaluated robust polynomial beamformers (RPBs) and robust polynomial beamformers with sectorial optimization (RPBSOs) is chosen as $P = 3$, and the white noise gain is limited to $\gamma_{\mathrm{dB}} = 10\log_{10}(\gamma) = -30\,\mathrm{dB}$ (i.e., $\gamma = 0.001$ in linear scale), unless stated otherwise. In order to approximate the frequency response vectors $[W_{p,m}(\omega_0), \ldots, W_{p,m}(\omega_{Q-1})]$ resulting from the design procedure, FIR filters of length $L = 512$ are utilized.

4.1 Symmetric Linear Array

The proposed optimization scheme is first evaluated for a symmetric linear array with $M = 7$ microphones and spacing $d = 2.5\,\mathrm{cm}$. Note that a regular spacing is not required as long as the arrangement is symmetric. The polynomial beamformer is jointly optimized for $I = 6$ uniformly distributed prototype look directions (PLDs). In

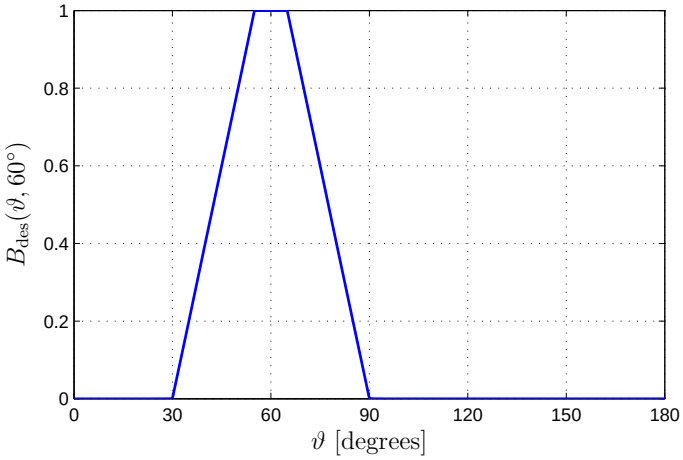


Fig. 4.1: Desired beamformer response $B_{\mathrm{des}}(\vartheta, 60°)$.

case of an RPB the PLD range is $\vartheta_{\mathrm{PLD}} = [0°, 180°]$, i.e., the prototype look directions are located at {0°, 36°, 72°, 108°, 144°, 180°}. According to chapter 3.1 we can exploit the array symmetry and reduce ϑ_{PLD} to $[0°, 90°]$. For the resulting RPBSO the PLDs are then located closer to each other, namely at {0°, 18°, 36°, 54°, 72°, 90°}.

As a first evaluation criterion of the beamformers, the mean square error (MSE) between the desired response $B_{\mathrm{des}}(\vartheta_k, \vartheta_{\mathrm{des}})$ and the actual response $B_D(\omega_q, \vartheta_k)$ is calculated for the steering range $\vartheta_{\mathrm{SR}} = [0°, 180°]$ in 5° steps. The MSE is given by

$$\mathrm{MSE}(\vartheta_{\mathrm{des}}) = \frac{1}{QK} \sum_{q=0}^{Q-1} \sum_{k=0}^{K-1} \left(|B_D(\omega_q, \vartheta_k)| - |B_{\mathrm{des}}(\vartheta_k, \vartheta_{\mathrm{des}})| \right)^2, \tag{4.1}$$

where ω_q and ϑ_k are discretized frequency and angle, respectively.

Fig. 4.2 shows the MSE-curves for a robust polynomial beamformer and a robust polynomial beamformer with sectorial optimization. For comparison, the MSE is also shown for a robust filter-and-sum beamformer (RFSB) with the same WNG constraint value. This beamformer is designed by setting $P = 0$ and optimizing a single response for the PLD $\vartheta_{\mathrm{des}} = 120°$. Using sectorial optimization the average MSE can be reduced from 0.10 to 0.03 in the current example and almost approaches the value 0.02 of the RFSB. In chapter 4.4 we will see that the (average) MSE alone is not always the best measure to judge the performance of a beamformer. In the current scenario, however, the MSE curves correspond well to the beampatterns, which are depicted in Fig. 4.3(a) for an RPB and in Fig. 4.3(b) for an RPBSO. Each sub-figure consists of three plots: the beampattern on the left side, the magnitude response for the desired look direction

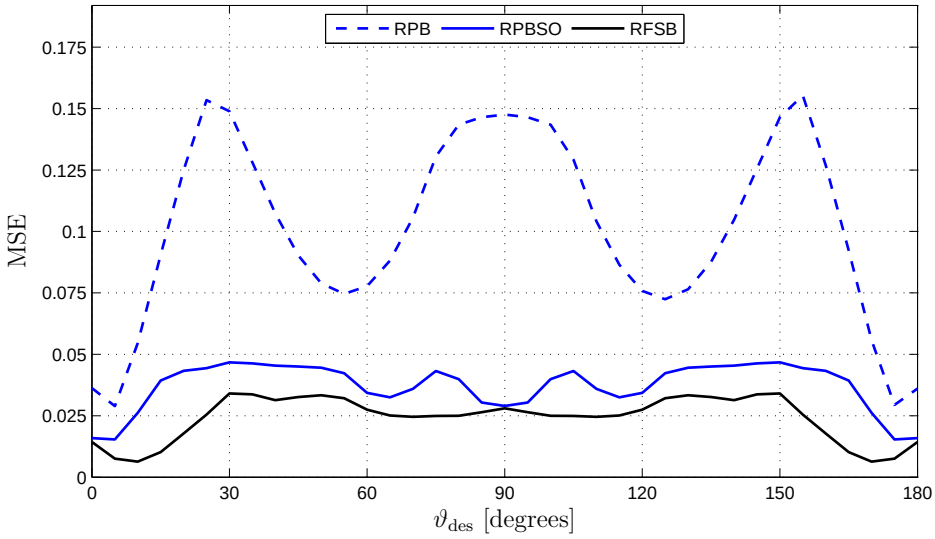
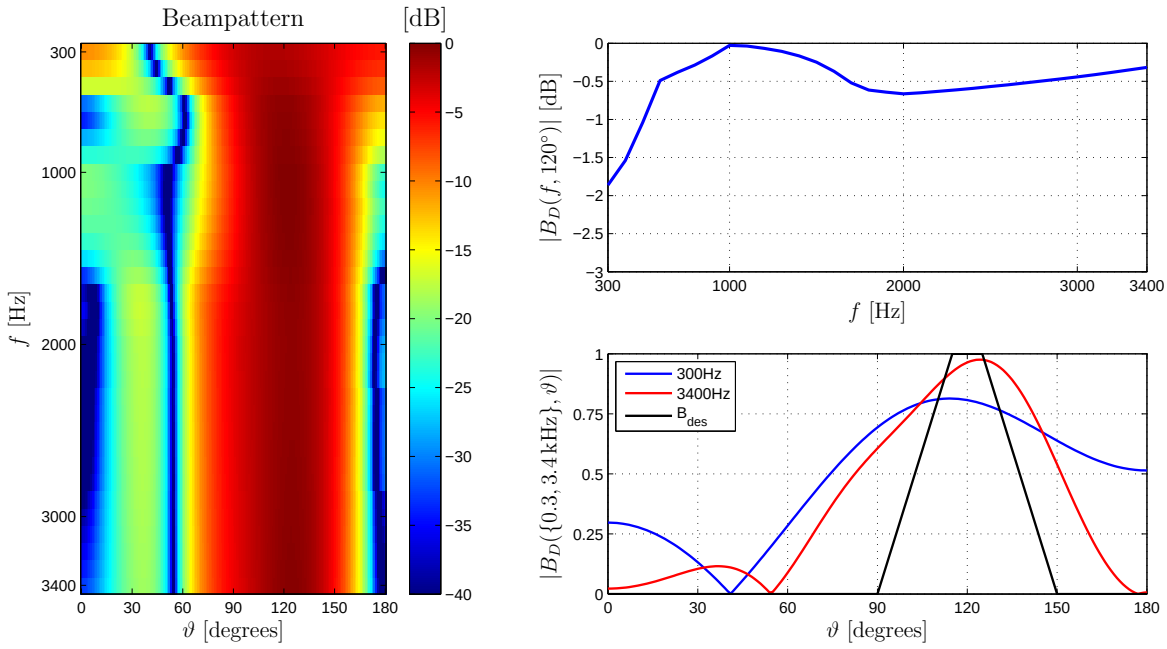
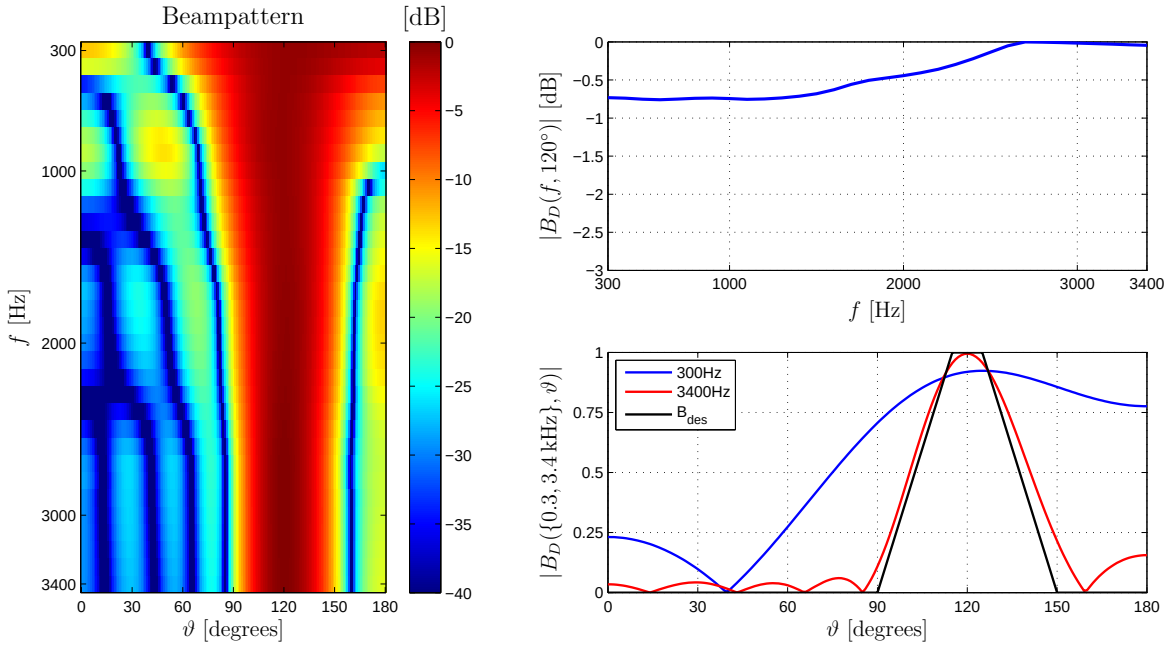


Fig. 4.2: MSE-curves for an RPB, RPBSO, and RFSB applied to a linear array with $M = 7$ microphones and uniform spacing $d = 2.5\,\mathrm{cm}$.



(a) Robust polynomial beamformer



(b) Robust polynomial beamformer with sectorial optimization

Fig. 4.3: Comparison of an RPB (a) and RPBSO (b) applied to a linear array ($M = 7$ microphones, uniform spacing $d = 2.5\,\mathrm{cm}$). Each sub-figure shows the beampattern together with the magnitude responses for the desired look direction $\vartheta_{\mathrm{des}} = 120°$ and for $f = \{0.3,\ 3.4\}\,\mathrm{kHz}$.

$\vartheta_{\text{des}} = 120°$ on the upper right, and the magnitude response for the cutoff frequencies $f_c = \{300, 3400\}$ Hz in comparison with the desired response $B_{\text{des}}(\vartheta, 120°)$ on the bottom right. The angle $\vartheta_{\text{des}} = 120°$ is chosen as its corresponding angle in the PLD range (obtained by mirroring) lies approximately halfway between two PLDs of the RPBSO. Hence, the beamformers are not optimized for this direction and the comparison is fair. Looking at the beampatterns one can see that the beam widens for low frequencies in both cases, and thus the directivity is low. Beyond $f \approx 1.5\,\text{kHz}$ both beamformers have approximately a frequency-independent main beam. However, using sectorial optimization the beam is significantly narrower and approaches the desired response much better as can be seen in the plots on the lower right (red curve). Also, the maximum deviation of the magnitude response for the desired direction is smaller. In case of an RPBSO $|B_D(f, 120°)|$ deviates less than 0.7 dB, whereas for an RPB the highest deviation is 1.8 dB.

The resulting white noise gain $A(\omega)$ of the three different beamformer designs is depicted in Fig. 4.4 for $\vartheta_{\text{des}} = 120°$. One can see that the WNG is above $-30\,\text{dB}$ in the entire frequency range for all beamformer designs. That is, the specified constraint $A(\omega) \geq \gamma_{\text{dB}} = -30\,\text{dB}$ is met even though the polynomial beamformers are not optimized for this look direction. In the following we will abandon the comparison of the polynomial beamformers with the corresponding RFSB in terms of MSE and WNG.

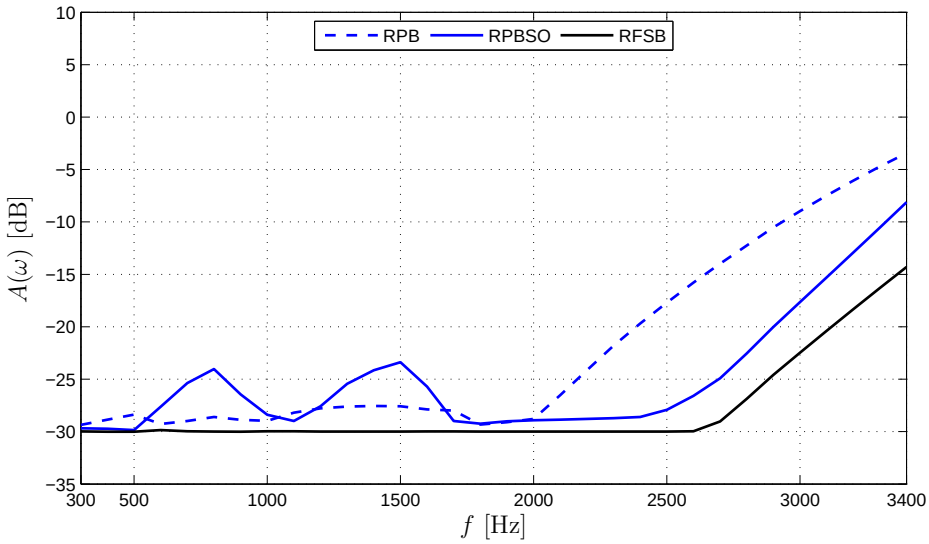


Fig. 4.4: White noise gain of an RPB, RPBSO, and RFSB applied to a symmetric linear array with $M = 7$ microphones and uniform spacing $d = 2.5\,\text{cm}$.

4.2 Symmetric Circular Arrays

4.2.1 Symmetric Circular Array with One Symmetry Axis

Next, we consider a symmetric circular array with $M = 6$ non-uniformly spaced microphones and radius $r = 3\,\text{cm}$. This is a reasonable array size if using small microphones, e.g., microphones with a diameter of 0.5 in. The microphones are located at angles $\vartheta_m = \{0°, 50°, 140°, 180°, 220°, 310°\}$, as depicted in Fig. 4.5, with the dash-dotted line indicating the symmetry axis. We use again $I = 6$ uniformly distributed prototype look directions, which are in case of an RPB located at $\{0°, 72°, 144°, 215°, 287°, 359°\}$. That is, they are spread within the entire angular region in order to allow for steering the beam to all possible directions. Note that we assume an angular resolution of 1° here, i.e., the maximum steering angle is $\vartheta_{\max} = 359°$ and the PLDs are rounded to integer values. As shown in section 3.2, we can make use of the array symmetry and reduce the PLD range from $[0°, 359°]$ to $[0°, 180°]$. The resulting prototype look directions for an RPBSO are then at $\{0°, 36°, 72°, 108°, 144°, 180°\}$.

As a first performance measure we compute again the mean square error in 5° steps using Eq. (4.1), where the steering range ϑ_{SR} is now $[0°, 359°]$ as we consider a circular array here. The MSE-curves in Fig. 4.6 show that using sectorial optimization the average MSE can be reduced from 0.21 to 0.07 in the current scenario. This is also reflected in the beampatterns, which can be seen in Fig. 4.7(a) for the RPB and in Fig. 4.7(b) for the RPBSO. Each sub-figure consists of three plots: the beampattern on the left side, the magnitude response for the desired angle $\vartheta_{\text{des}} = 72°$ on the upper right, and the magnitude response for the cutoff frequencies $f_{\text{c}} = \{300, 3400\}\,\text{Hz}$ in comparison with the desired response $B_{\text{des}}(\vartheta, 72°)$ on the bottom right. Note that the chosen angle of 72° coincides with a PLD for both the RPB and the RPBSO. That is,

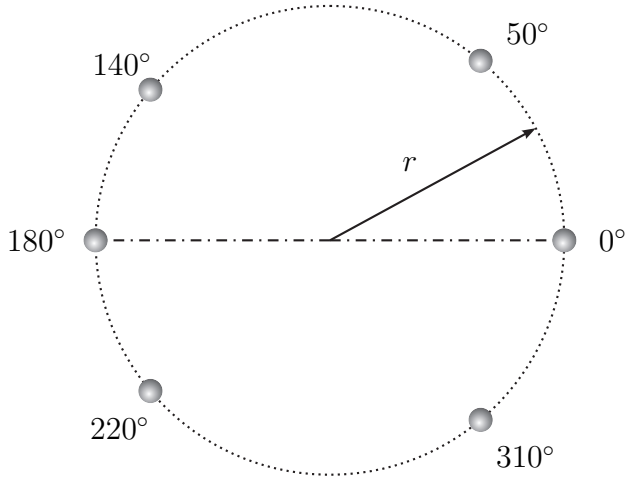


Fig. 4.5: Symmetric circular array with non-uniform spacing.

both beamformers are optimized for this specific direction. Nevertheless, the differences between the beamformers' performance are significant. Although both designs result in almost frequency-independent beampatterns the RPB has a much wider main lobe as well as higher side lobes compared to the RPBSO. Also, the position of the peak of the main beam of the RPB is about 30° off the desired look direction $\vartheta_{\text{des}} = 72°$. Furthermore, the magnitude response of the RPBSO has negligible deviations of less than 0.2 dB for the desired direction, whereas the magnitude response of the RPB deviates up to 0.9 dB.

Even though there is still a clear gap between the desired response B_{des} and the response of the RPBSO, for the chosen setup one cannot achieve a much better performance with an RFSB either. In fact, an RFSB design which is exclusively optimized for this particular look direction performs similar for the given array geometry and limitations on the white noise gain (see Fig. C.1 in Appendix C). The main beam of the RFSB is slightly narrower compared to the RPBSO, which has lower side lobes in return.

The white noise gain $A(\omega)$ of both polynomial beamformers is depicted in Fig. 4.8. We can see that the WNG of the RPBSO is in a wide frequency range smaller compared to the RPB design. However, the constraint on the WNG is still satisfied since $A(\omega) \geq \gamma_{\text{dB}} = -30\,\text{dB}$ for all frequencies.

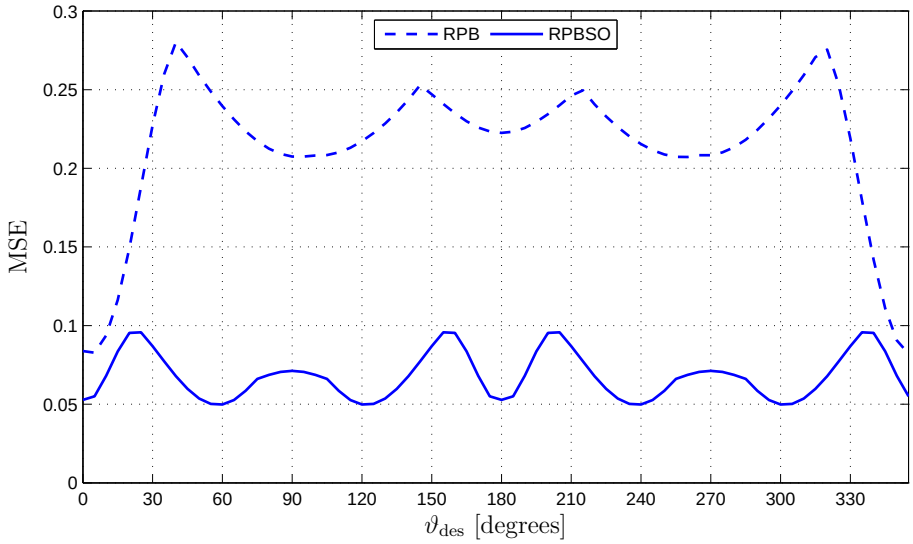
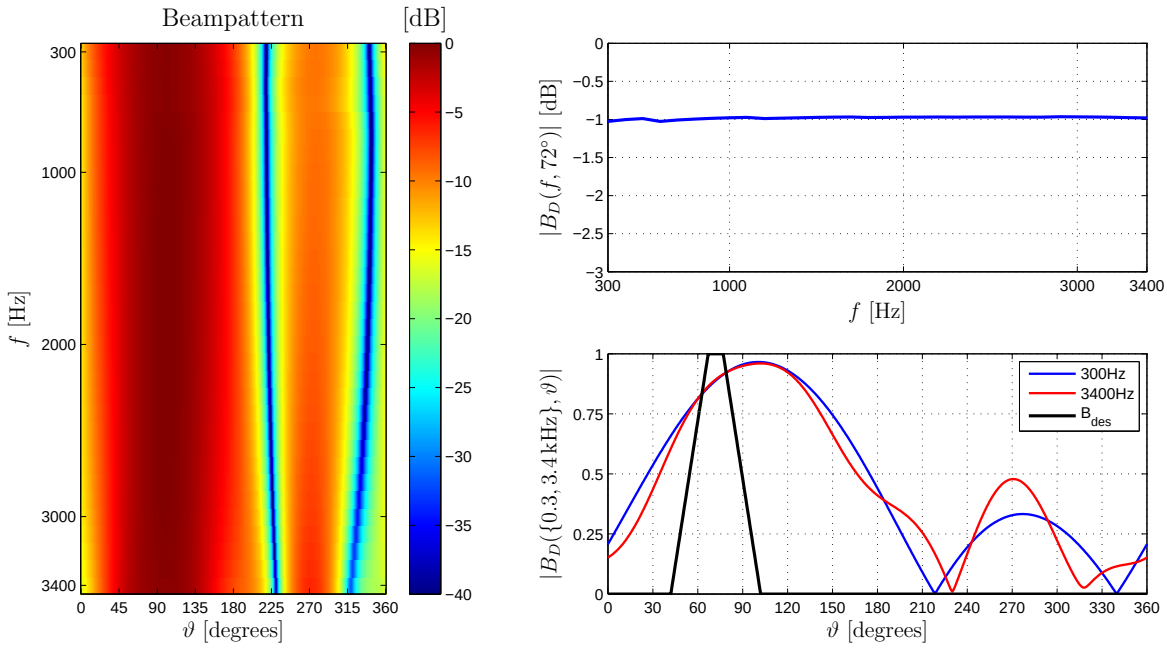
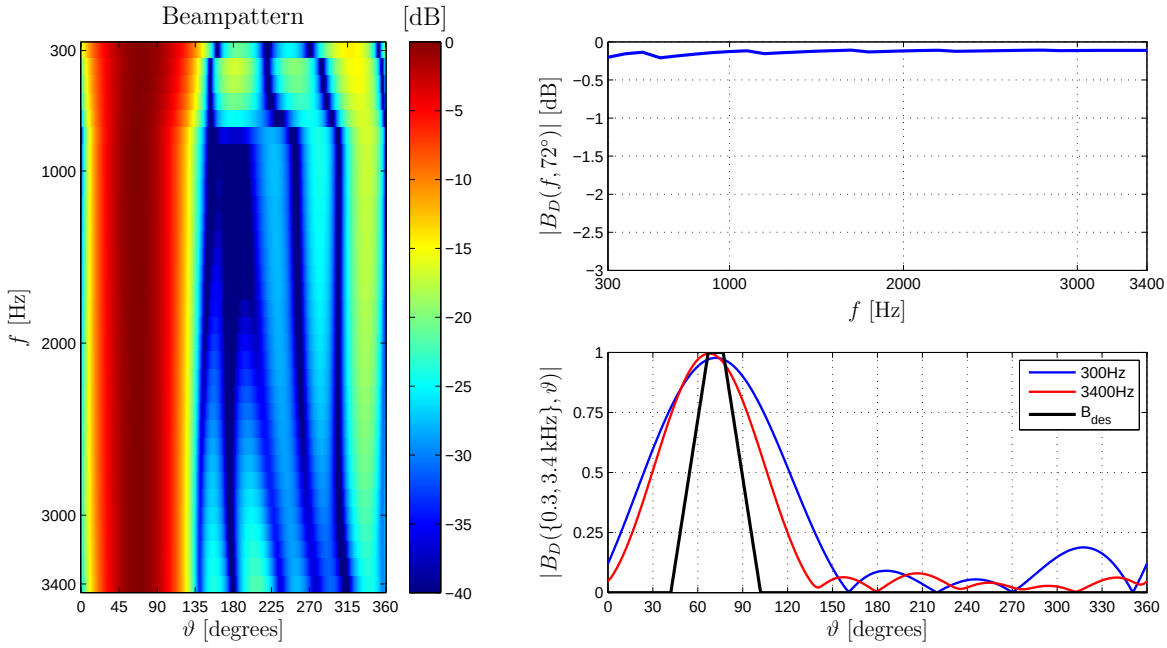


Fig. 4.6: Comparison of the MSE-curves for an RPB and RPBSO applied to a circular array of radius $r = 3\,\text{cm}$ with $M = 6$ microphones and non-uniform spacing.



(a) Robust polynomial beamformer



(b) Robust polynomial beamformer with sectorial optimization

Fig. 4.7: Shown are the beampatterns together with the magnitude responses in look direction $\vartheta_{\text{des}} = 72°$ and for $f = \{0.3,\ 3.4\}\,\text{kHz}$ obtained by an RPB (a) and RPBSO (b). The beamformers are applied to a symmetric circular array consisting of $M = 6$ microphones with radius $r = 3\,\text{cm}$.

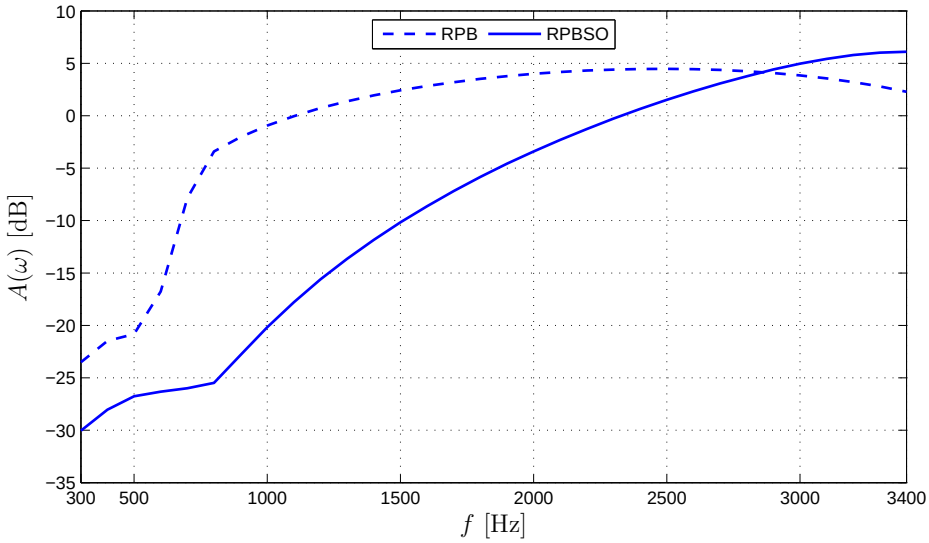


Fig. 4.8: White noise gain of an RPB and RPBSO applied to a symmetric circular array with $M = 6$ non-uniformly spaced microphones and radius $d = 3\,\text{cm}$.

4.2.2 Symmetric Circular Array with Two Symmetry Axes

We now slightly modify the array discussed in the previous section (see Fig. 4.5) and move the two microphones located at 50° and 310° to angles 40° and 320°, respectively. Then, the array is symmetric with respect to two axes, as indicated in Fig. 4.9, and the PLD range can be reduced even further to $\vartheta_{\text{PLD}} = [0°, 90°]$ as discussed in section 3.2. Again $I = 6$ uniformly distributed prototype look directions are used and we denote the resulting polynomial beamformer exploiting both symmetry axes as RPBSO_2, where the PLDs are located at {0°, 18°, 36°, 54°, 72°, 90°}. In case of the RPB the PLDs remain unchanged at angles {0°, 72°, 144°, 215°, 287°, 359°} in order to cover the entire steering range. We also want to compare these two beamformers with another RPBSO design, which exploits only one symmetry axis and is denoted as RPBSO_1. For this beamformer the PLDs are at {0°, 36°, 72°, 108°, 144°, 180°} (cf. section 4.2.1).

The mean square error between the desired response and the magnitude response of the beamformer is computed in 5° steps using Eq. (4.1) for the steering range $\vartheta_{\text{SR}} = [0°, 359°]$ and is depicted in Fig. 4.10. Exploiting the symmetrical structure of the array the average MSE can be significantly reduced from 0.21 (RPB) to 0.07 for both the RPBSO_1 and RPBSO_2. The MSE curves of both RPBSO designs almost coincide which suggests a very similar performance of both beamformers. The benefit of designing a beamformer in a reduced PLD range becomes also obvious when comparing the

beampatterns, which are depicted in Fig. 4.11(a) for the RPB, in Fig. 4.11(b) for the RPBSO_1, and in Fig. 4.11(c) for the RPBSO_2. Each sub-figure shows the beampattern on the left, the magnitude response for the desired angle $\vartheta_{\mathrm{des}} = 155°$ on the upper right, and the magnitude response for the cutoff frequencies $f_{\mathrm{c}} = \{300, 3400\}\,\mathrm{Hz}$ in comparison with the desired response $B_{\mathrm{des}}(\vartheta, 155°)$ on the bottom right. Note that none of the three beamformer designs is optimized for this look direction. Looking at the

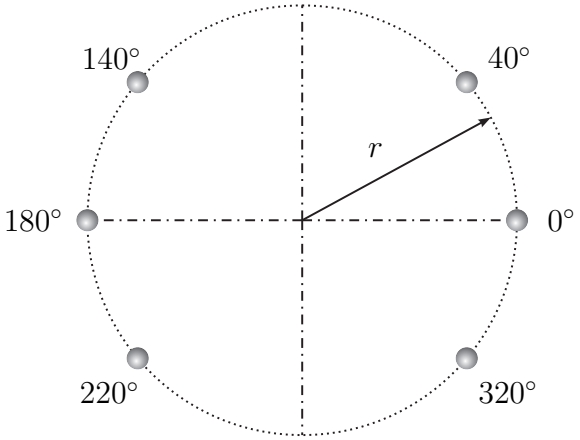


Fig. 4.9: Circular array with non-uniform spacing and two symmetry axis.

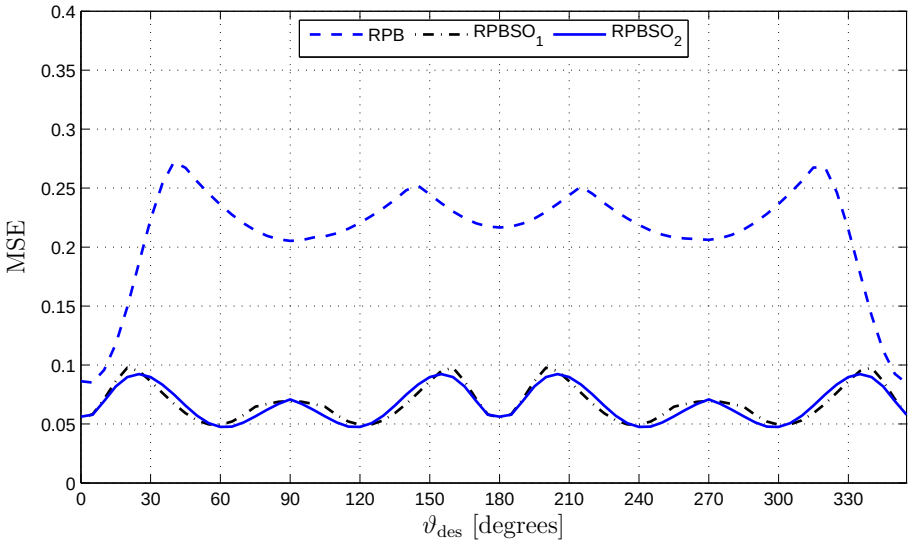
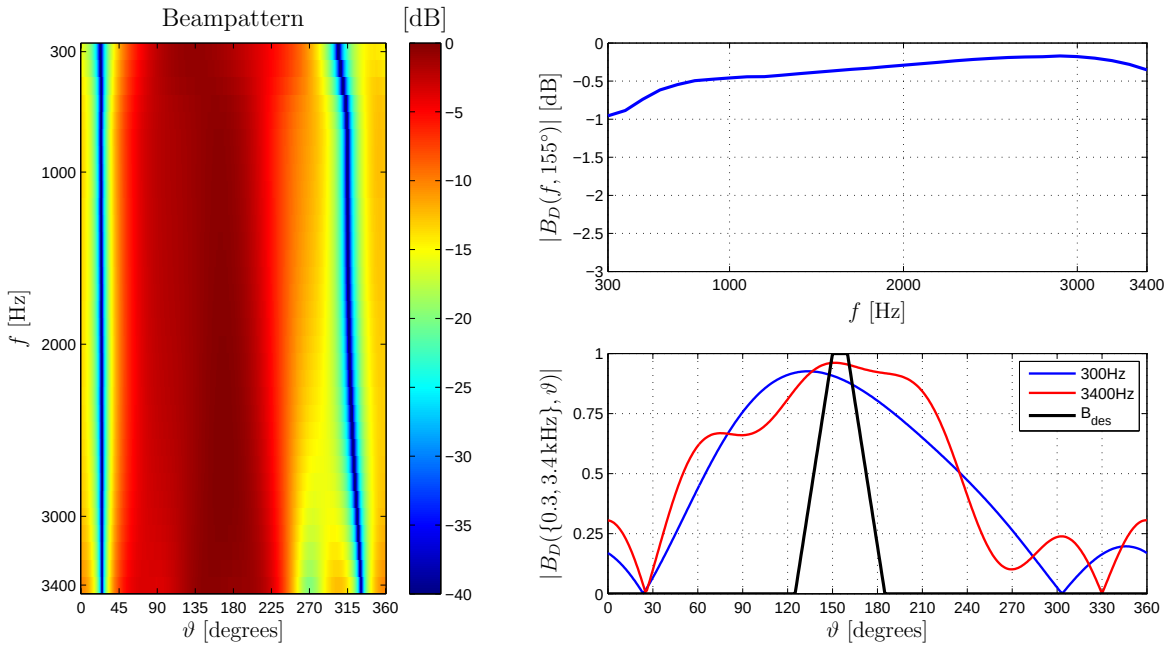


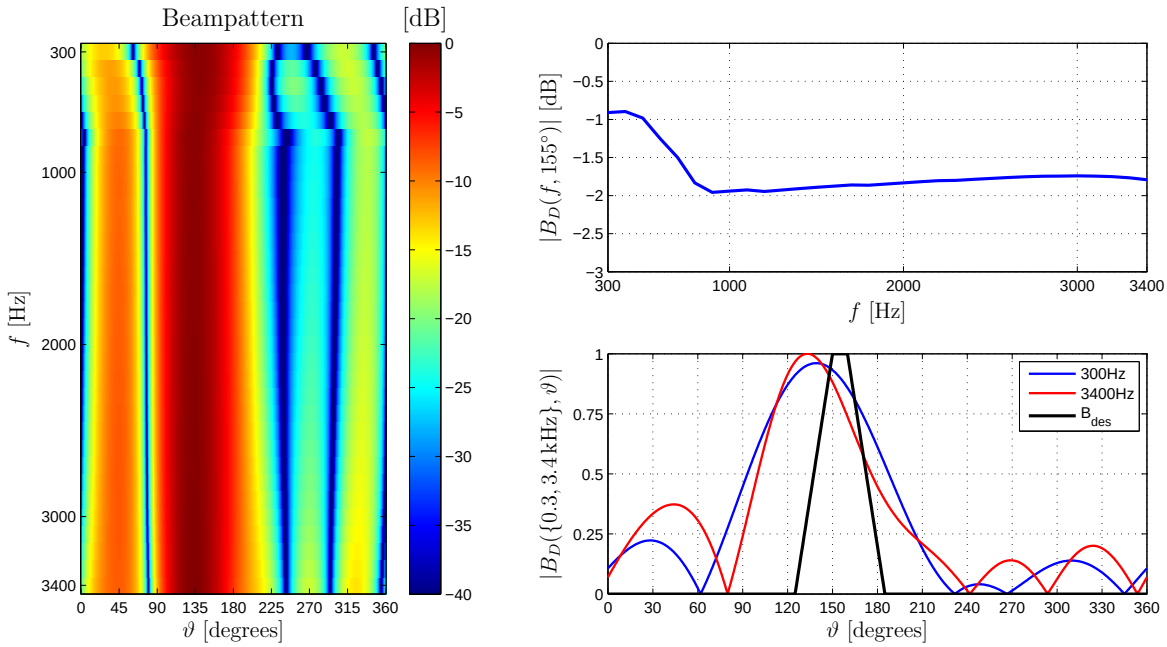
Fig. 4.10: MSE-curves for an RPB, RPBSO_1 exploiting one symmetry axis, and RPBSO_2 exploiting both symmetry axes. The circular array is of radius $r = 3\,\mathrm{cm}$ and consists of $M = 6$ microphones.

magnitude responses $|B_D(f, 155°)|$, the RPB seems to outperform the RPBSO$_1$ at first glance, since the deviations are less than 1 dB as opposed to almost 2 dB in case of the RPBSO$_1$. However, the main beam of the RPB is very wide which means that the directivity is much lower, and hence potential interfering sound sources cannot be attenuated well in a large angular region. In case of the RPBSO$_1$, the peak of the main beam is about 20° off the desired angle resulting in a slightly more distorted desired signal, but the spatial selectivity is much higher. If the second symmetry axis is also exploited, as it is done in case of the RPBSO$_2$, the performance is superior. Then, the deviations of the magnitude response in the desired look direction are less than 0.3 dB and the angular offset of the peak of the main beam is negligible. However, the side lobes of the RPBSO$_2$ are slightly increased.

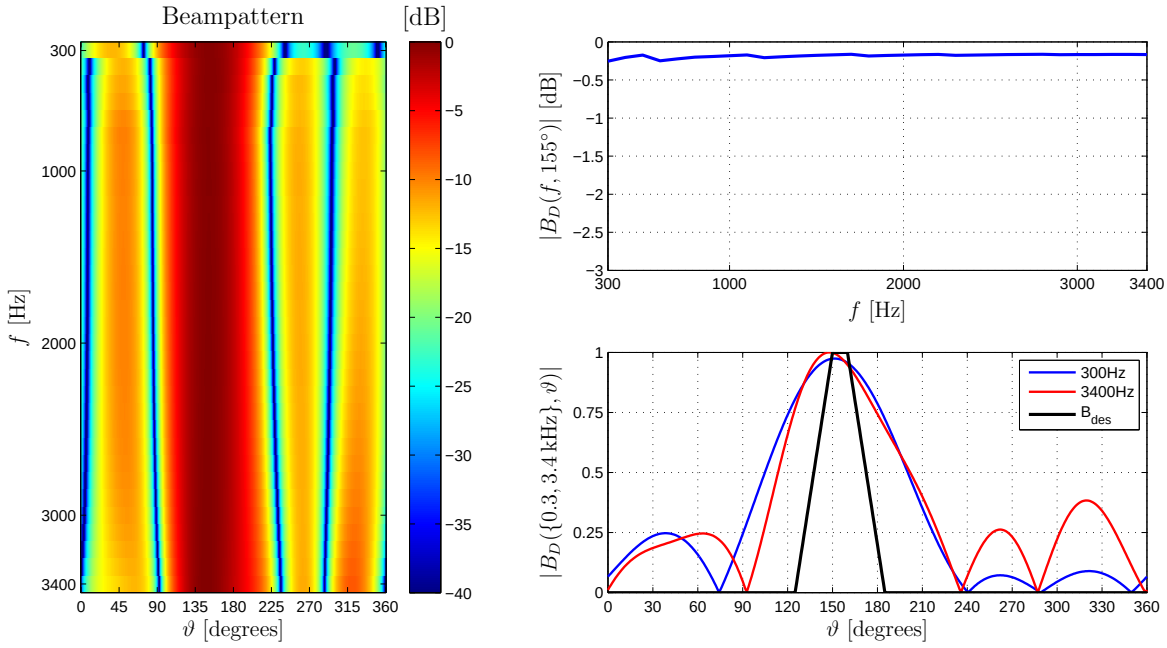
Fig. 4.12 shows the white noise gain $A(\omega)$ of the three different beamformer designs. The WNG of the RPB and RPBSO$_1$ fall somewhat below the threshold of $\gamma_{\text{dB}} = -30\,\text{dB}$. This is possible since the beamformers are not designed for this look direction and it is only assured that the WNG constraint is fulfilled at the prototype look directions themselves. In case of the RPBSO$_2$ $A(\omega) > -30\,\text{dB}$ for all frequencies, i.e., the WNG constraint is satisfied. A possible explanation for this is that the angular distance between the desired look direction and the PLDs, at which the constraint is really met, is smaller compared to the other designs.



(a) Robust polynomial beamformer



(b) Robust polynomial beamformer exploiting one symmetry axis



(c) Robust polynomial beamformer exploiting two symmetry axes

Fig. 4.11: Shown are the beampatterns together with the magnitude responses in look direction $\vartheta_{\text{des}} = 155°$ and for $f = \{0.3,\ 3.4\}\,\text{kHz}$ obtained by an RPB (a), RPBSO_1 (b), and RPBSO_2 (c). The beamformers are applied to a circular array ($M = 6$ microphones, radius $r = 3\,\text{cm}$) with two symmetry axes.

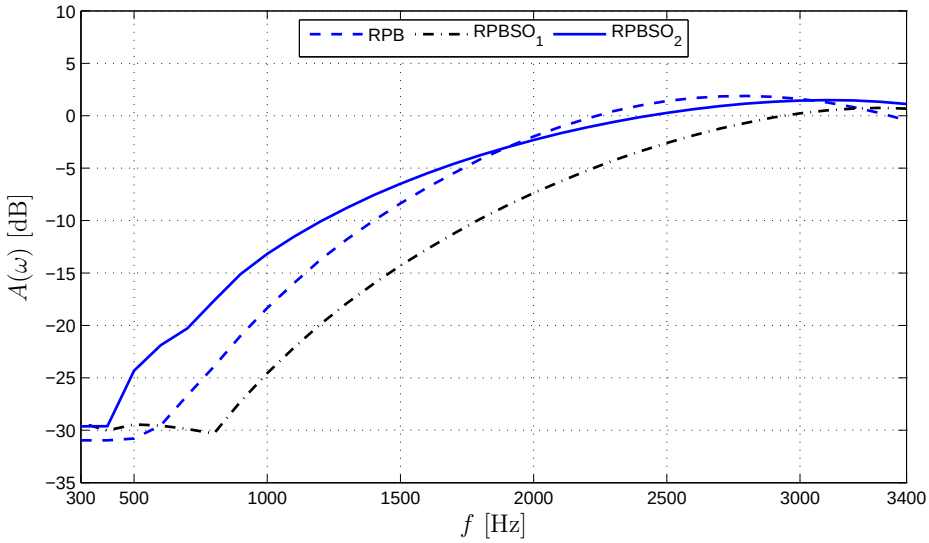


Fig. 4.12: White noise gain of an RPB, RPBSO_1, and RPBSO_2 applied to a circular array of radius $r = 3\,\text{cm}$ with two symmetry axes and $M = 6$ non-uniformly spaced microphones.

4.2.3 Uniformly Spaced Circular Array

As a special case of symmetric arrays we examine a uniformly spaced circular array as a last example. The array is of radius $r = 3\,\text{cm}$ and consists of $M = 6$ microphones, i.e., the microphones are located at $\vartheta_m = \{0°, 60°, 120°, 180°, 240, 300°\}$. We choose again $I = 6$ uniformly distributed PLDs which are located at {0°, 72°, 144°, 216°, 288°, 359°} for the RPB. Using the method proposed by Lai et al. the PLD range can be limited to $\vartheta_{\text{PLD}} = [0°, 360°/M]$ (cf. section 3.2), which is $[0°, 60°]$ in the current example. Then, the prototype look directions are {0°, 12°, 24°, 36°, 48°, 60°} and we term the resulting beamformer RPB_{Lai}. Exploiting all M symmetry axes we can further reduce the PLD range to $\vartheta_{\text{PLD}} = [0°, 360°/(2M)]$, i.e., to $[0°, 30°]$. For this RPBSO design the PLDs are {0°, 6°, 12°, 18°, 24°, 30°}.

Fig. 4.13 shows the MSE of the three different designs which is calculated in 5° steps using Eq. (4.1) for the steering range $\vartheta_{\text{SR}} = [0°, 359°]$. Exploiting the uniform spacing of the array the average MSE can be reduced from 0.22 (RPB) to 0.07 (RPB_{Lai}) and 0.06 (RPBSO), respectively. The shape of the MSE curve for both optimization schemes is very similar and so are the beampatterns. In Fig. 4.14(a) the beampattern of the RPB is depicted in the left plot, Fig. 4.14(b) and Fig. 4.14(c) show the results for RPB_{Lai} and RPBSO. The magnitude response for the desired angle $\vartheta_{\text{des}} = 130°$ can be seen on the upper right of each sub-figure[1], and the magnitude response for the cutoff frequencies $f_{\text{c}} = \{300, 3400\}\,\text{Hz}$ in comparison with the desired response

[1] The ordinate is now scaled differently and only covers the range between $-1\,\text{dB} \ldots 0\,\text{dB}$ in order to bring out the small deviations.

$B_{\text{des}}(\vartheta, 130°)$ is on the bottom right. Note that none of the beamformers is optimized for the look direction $\vartheta_{\text{des}} = 130°$. The beampattern of the RPB is almost frequency-independent and signals arriving from the desired direction are attenuated less than 0.4 dB. However, the spatial selectivity of this design is rather poor since the main beam is very wide. In contrast, the RPBSO as well as RPB_{Lai} have a much higher directivity and distort the desired signals less. The maximum deviations of the magnitude response for the desired look direction are smaller than 0.2 dB (RPB_{Lai}) and 0.3 dB (RPBSO), respectively. Also, these two designs are very similar in terms of the beampattern. The RPBSO has a main lobe which is slightly narrower compared to RPB_{Lai}, which has lower side lobes in return.

Looking at the white noise gain which is depicted in Fig. 4.15 the differences between the designs become clearer. The RPBSO is the only design which truly satisfies the constraint on the WNG, $A(\omega) \geq \gamma_{\text{dB}} = -30\,\text{dB}$, for the given look direction[2]. Furthermore, the WNG of the RPBSO is approximately 5 dB above the curve of RPB_{Lai} in about half of the frequency range of interest.

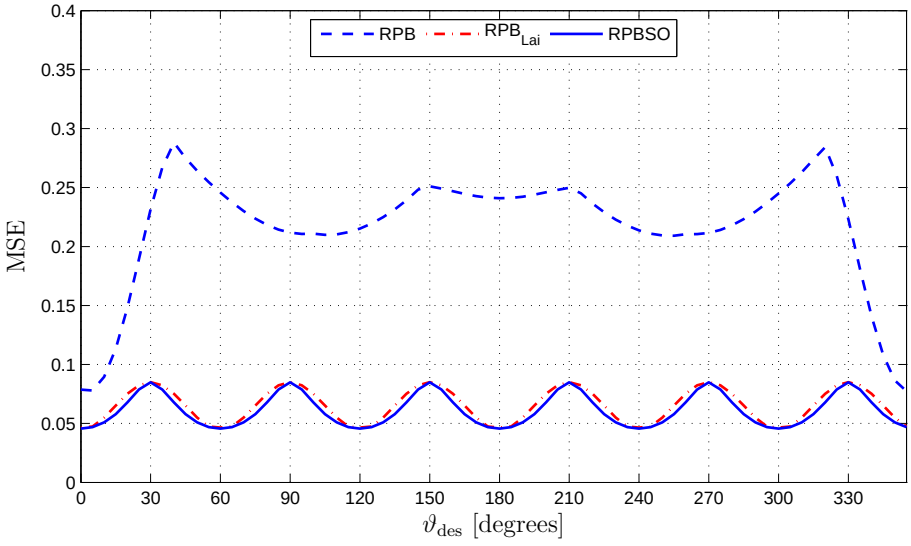
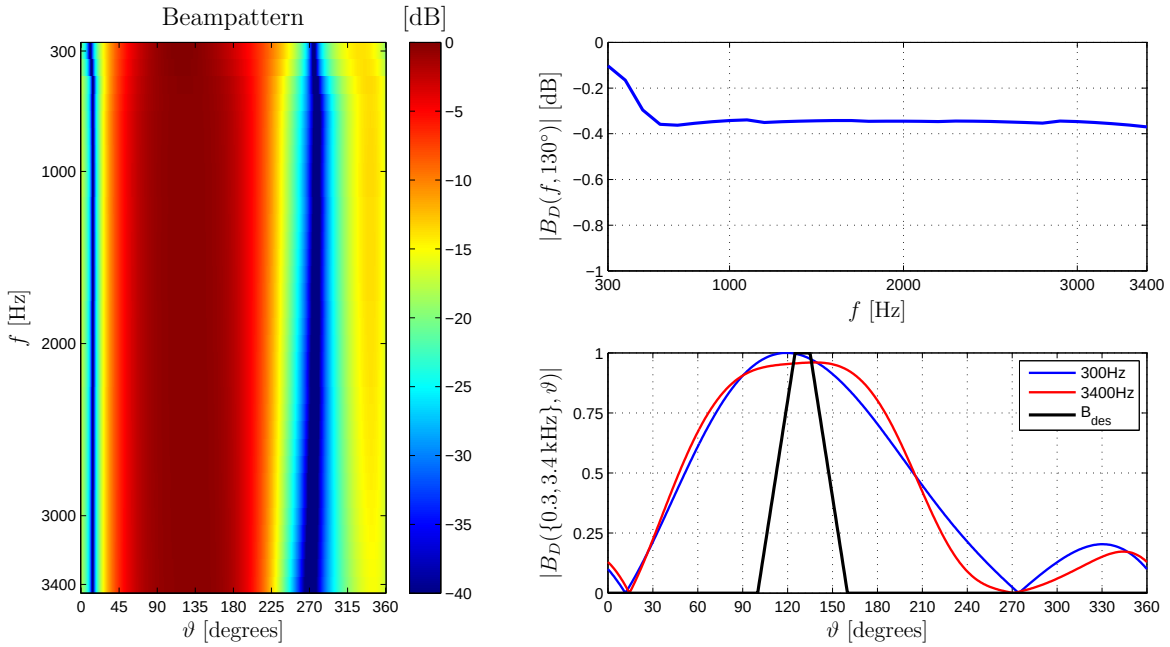
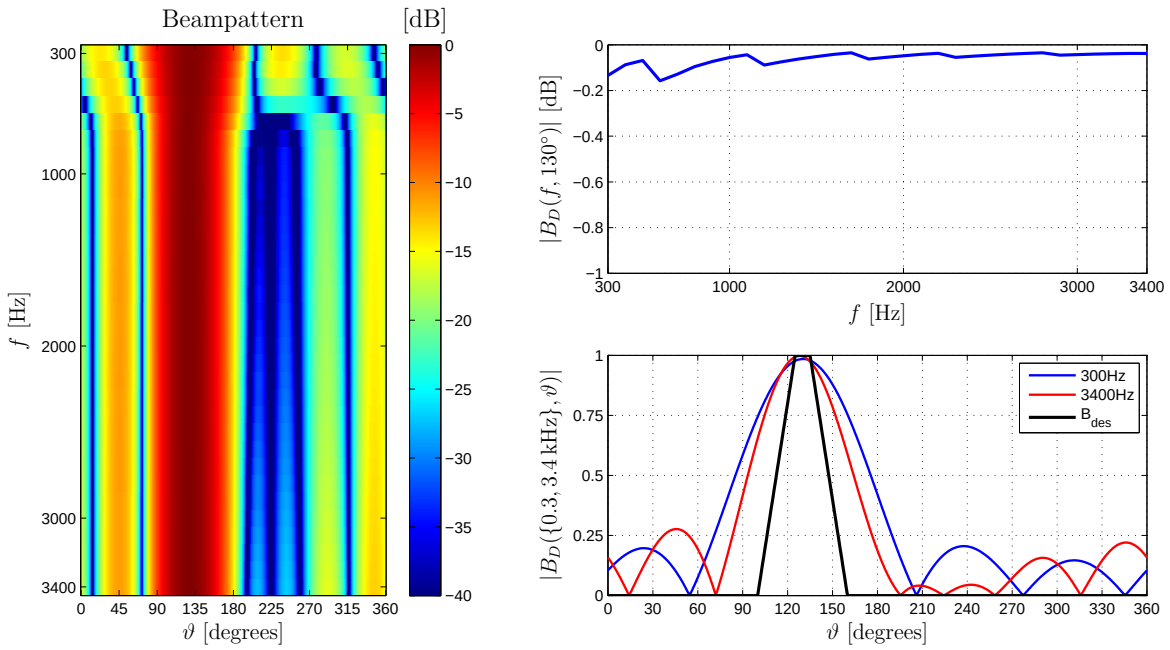


Fig. 4.13: MSE curves for an RPB, an RPBSO exploiting all symmetry axes, and according to the proposal of Lai et al. (RPB_{Lai}). The beamformers are applied to a uniformly spaced circular array with $M = 6$ microphones and radius $r = 3\,\text{cm}$.

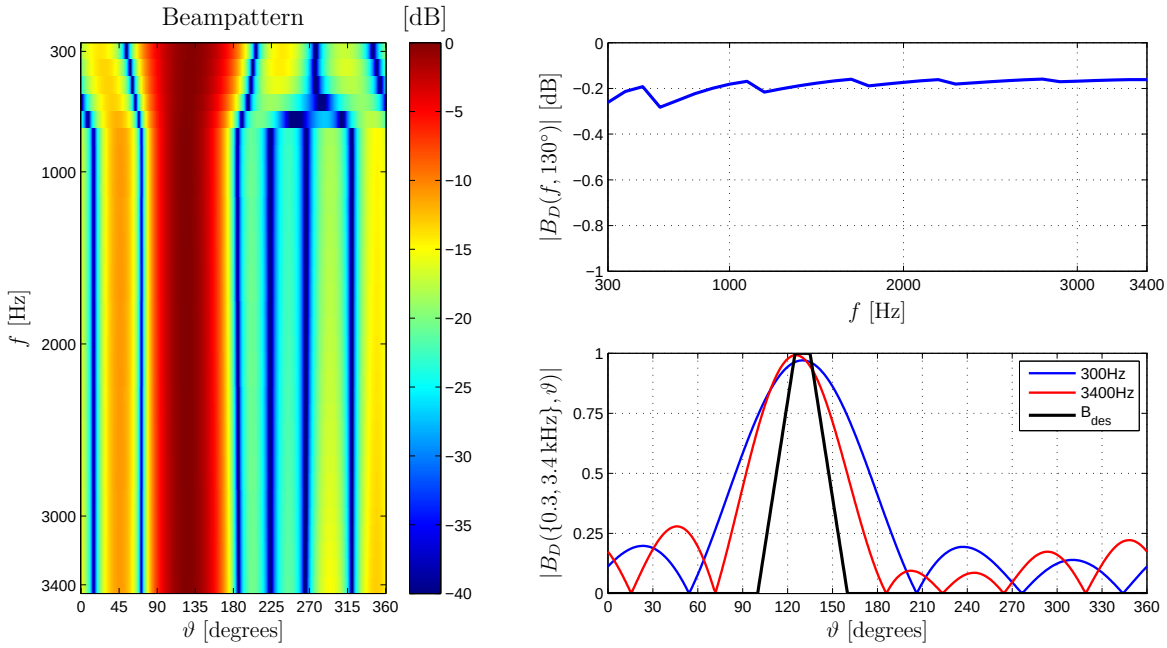
[2] Indeed, the violations of the WNG constraint are only minor.



(a) Robust polynomial beamformer



(b) Robust polynomial beamformer according to Lai et al.



(c) Robust polynomial beamformer exploiting all symmetry axes

Fig. 4.14: Comparison of the beampatterns and the magnitude responses for $\vartheta_{\text{des}} = 130°$ and $f = \{0.3, 3.4\}$ kHz, respectively, obtained by an RPB (a), a design according to Lai et al. (RPB_{Lai}) (b), and an RPBSO (c). The uniformly spaced circular array consists of $M = 6$ microphones and is of radius $r = 3\,\text{cm}$.

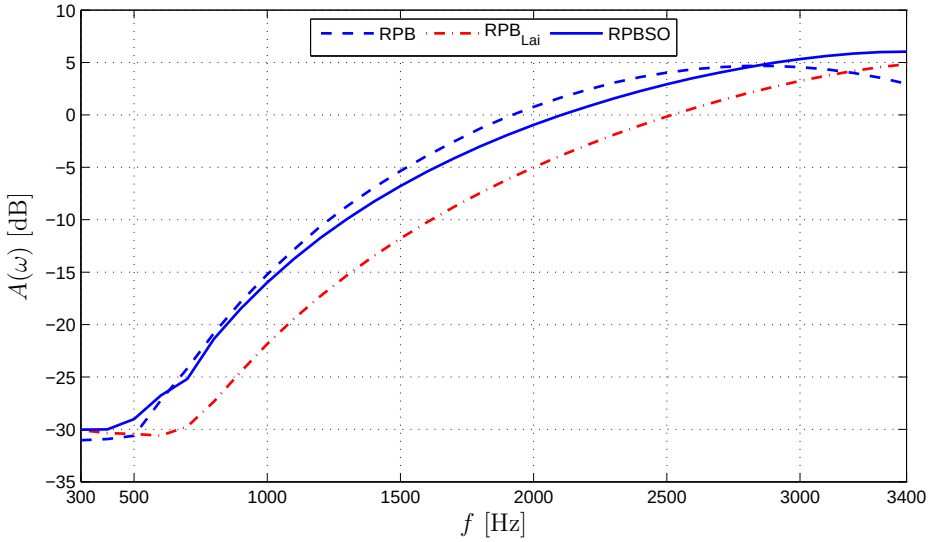
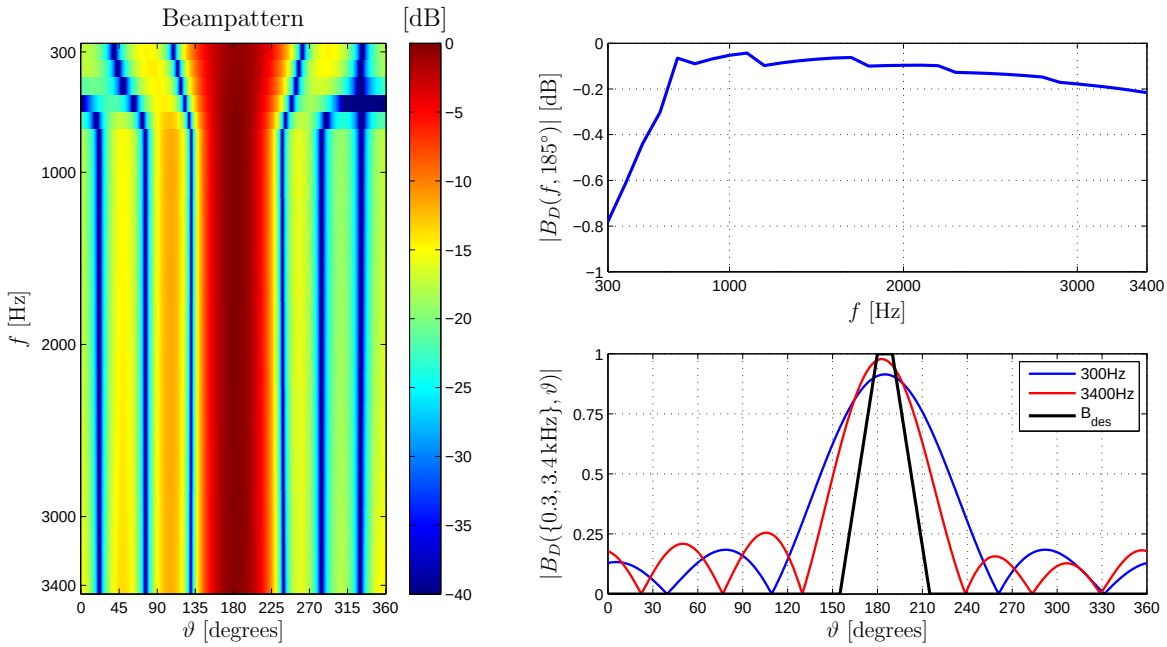


Fig. 4.15: WNG for an RPB, a design according to Lai et al. (RPB_{Lai}), and an RPBSO applied to a circular array of radius $r = 3\,\text{cm}$ with $M = 6$ uniformly spaced microphones.

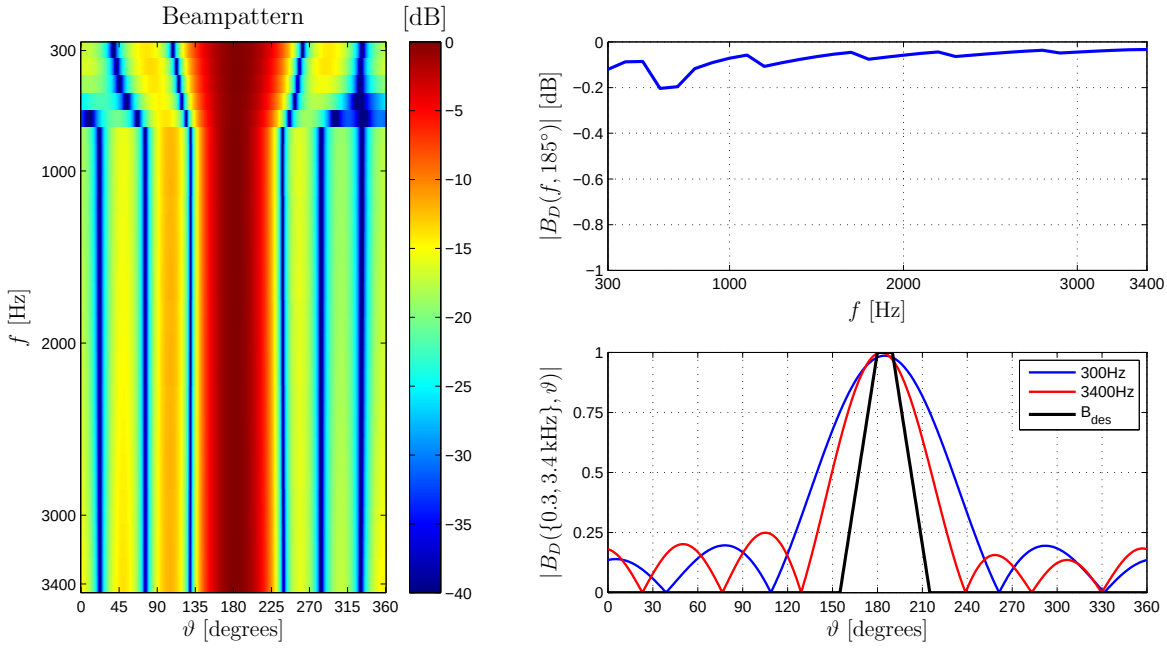
In the following we reduce the number of PLDs to $I = 4$ for the same scenario as before. This is reasonable since the angular distance between the sampling points (PLDs) is very small for $I = 6$. Now, the PLDs are at $\{0°, 20°, 40°, 60°\}$ in case of the $\mathrm{RPB}_{\mathrm{Lai}}$ and at $\{0°, 10°, 20°, 30°\}$ for an RPBSO design. The comparison with a classical RPB design is omitted here as its performance is much worse compared to the results shown in Fig. 4.14(a). Also, we do not show the MSE curves for the different beamformer designs as the results are very similar to Fig. 4.13 and do not give new insights.

In Fig. 4.16(a) the beampattern of the $\mathrm{RPB}_{\mathrm{Lai}}$ design is depicted in the left plot, Fig. 4.16(b) shows the result for the RPBSO. The magnitude response for the desired look direction $\vartheta_{\mathrm{des}} = 185°$ can be seen on the upper right of each sub-figure, and the magnitude response for the cutoff frequencies $f_{\mathrm{c}} = \{300, 3400\}\,\mathrm{Hz}$ in comparison with the desired response $B_{\mathrm{des}}(\vartheta, 185°)$ is on the bottom right. Note that both beamformers are not optimized for $\vartheta_{\mathrm{des}} = 185°$ or the corresponding angle in the PLD range, respectively. Comparing the beampatterns the differences between the two designs are difficult to determine. When looking at the magnitude responses one can see that the performance of the RPBSO is slightly superior as the maximum deviations are only about 0.2 dB as opposed to almost 0.8 dB in case of the $\mathrm{RPB}_{\mathrm{Lai}}$. Especially below $f = 500\,\mathrm{Hz}$ the desired signal is attenuated more when using the beamformer design proposed by Lai et al.

The performance differences between the two designs can also be seen in terms of the white noise gain $A(\omega)$ which is depicted in Fig. 4.17 for the desired look direction $\vartheta_{\mathrm{des}} = 185°$. In a wide frequency range the WNG of the RPBSO is approximately 5 dB higher compared to the $\mathrm{RPB}_{\mathrm{Lai}}$. Also, the RPBSO design satisfies the constraint on the WNG, $A(\omega) \geq \gamma_{\mathrm{dB}} = -30\,\mathrm{dB}$. In contrast, the $\mathrm{RPB}_{\mathrm{Lai}}$ violates the constraint as the WNG drops below $-32\,\mathrm{dB}$ for low frequencies.



(a) Robust polynomial beamformer according to Lai et al.



(b) Robust polynomial beamformer exploiting all symmetry axes

Fig. 4.16: Shown are the beampatterns and the magnitude responses for $\vartheta_{\text{des}} = 185°$ and $f = \{0.3,\ 3.4\}\,\text{kHz}$ of a design according to Lai et al. (RPB_{Lai}) (a) and an RPBSO (b). The beamformers are optimized for $I = 4$ PLDs and applied to a uniformly spaced circular array with $M = 6$ microphones and radius $r = 3\,\text{cm}$.

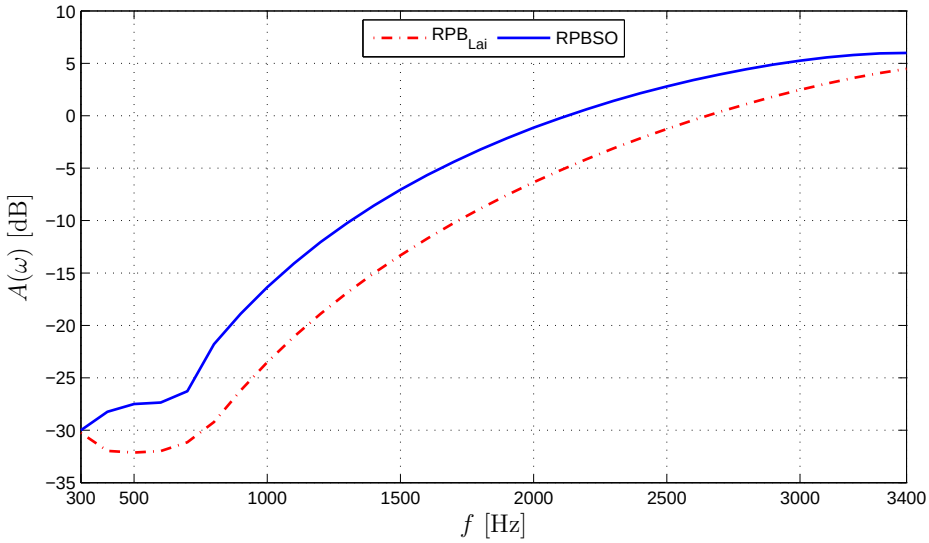


Fig. 4.17: WNG for a beamformer design according to Lai et al. ($\mathrm{RPB_{Lai}}$) and an RPBSO. The beamformers are optimized for $I = 4$ PLDs and applied to a circular array of radius $r = 3\,\mathrm{cm}$ with $M = 6$ uniformly spaced microphones.

4.3 Impact of Choice of Prototype Look Directions on Beamformer Performance

Typically, the prototype look directions are uniformly distributed within the whole PLD range, ϑ_{PLD}. This is done as the polynomial beamformer is then evenly optimized with respect to the angular axis, i.e., all possible look directions are considered to be equally important. In case of a non-uniform distribution there may be large angular regions not comprising PLDs, and hence the angular distance between some of the PLDs is also large. This should be avoided since the PLDs act as sampling points when steering the beam to intermediate directions.

There are different ways of how to distribute the PLDs uniformly, as will be shown by way of example for $\vartheta_{\mathrm{PLD}} = [0°, 359°]$[3]. One possibility is to simply divide the PLD range by the number of prototype look directions I, i.e.,

$$\vartheta_{\mathrm{des}_i} = i \cdot \frac{\vartheta_{\max}}{I}, \quad i = 0, \ldots, I-1. \tag{4.2}$$

Thus, the prototype look directions cover the angular region $[0°, (1-\frac{1}{I})\cdot 360°]$. According to Eq. (2.20) the associated values of D are then limited to $-1 \leq D \leq 2 \cdot \frac{I-1}{I} - 1$,

[3] The angular resolution is assumed to be 1°.

as illustrated in Fig. 4.18 for $I = 8$ prototype look directions. The figure shows that, in the example, there is no sampling point available for $D > 0.75$ or $\vartheta_{\text{des}} > 315°$, respectively. Indeed, the figure may suggest that there does exist a sampling point for $D > 0.75$. However, the sampling point at $\vartheta_{\text{des}} = 0°$ corresponds to $D = -1$. There is no PLD available for $D = +1$ which would be needed for interpolation. As a consequence, steering the beam to angles greater than 315° requires extrapolation, which should be avoided. The effect of extrapolation is shown in Fig. 4.19 for a uniformly spaced circular array with $M = 6$ microphones and radius $r = 3\,\text{cm}$. 4.19(a) shows the beampattern of an RPB on the left, its magnitude response for the desired angle $\vartheta_{\text{des}} = 20°$ on the upper right[4], and the magnitude response for $f = \{0.3,\ 3.4\}\,\text{kHz}$ on the lower right. The behavior for a second look direction $\vartheta_{\text{des}} = 340° \equiv -20°$ is shown in 4.19(b). Even though these two angles differ only in sign, which suggests a similar performance for both directions due to the symmetry of the array, the resulting beampattern for $\vartheta_{\text{des}} = 340°$ obtained by extrapolation is useless.

As a second option, an additional PLD can be appended for $\vartheta_{\text{des}} = 359°$ in order to circumvent the extrapolation issue. Then, there are sampling points at $\vartheta_{\text{des}} = 0°$ and $\vartheta_{\text{des}} = 359°$ corresponding to $D = -1$ and $+1$. That is, the beamformer is optimized for two PLDs pointing to approximately the same angular direction. Steering the beam from $\vartheta_{\text{des}} = 0°$ to $\vartheta_{\text{des}} = 359° \equiv -1°$ is only a slight change with respect to the angular axis. However, the associated values of the steering direction D, which determine the beamformer response, jump from -1 to $+1$. Hence, it cannot be expected that the response changes smoothly when steering the beam across these angles, as can be seen in Fig. 4.20. This can be a disadvantage if one wants to track a sound source moving

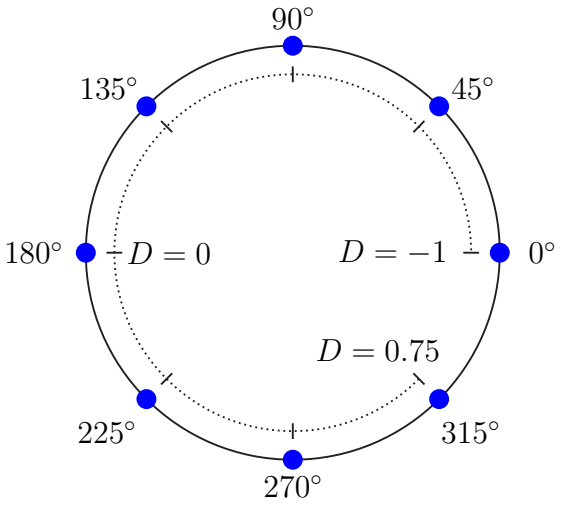
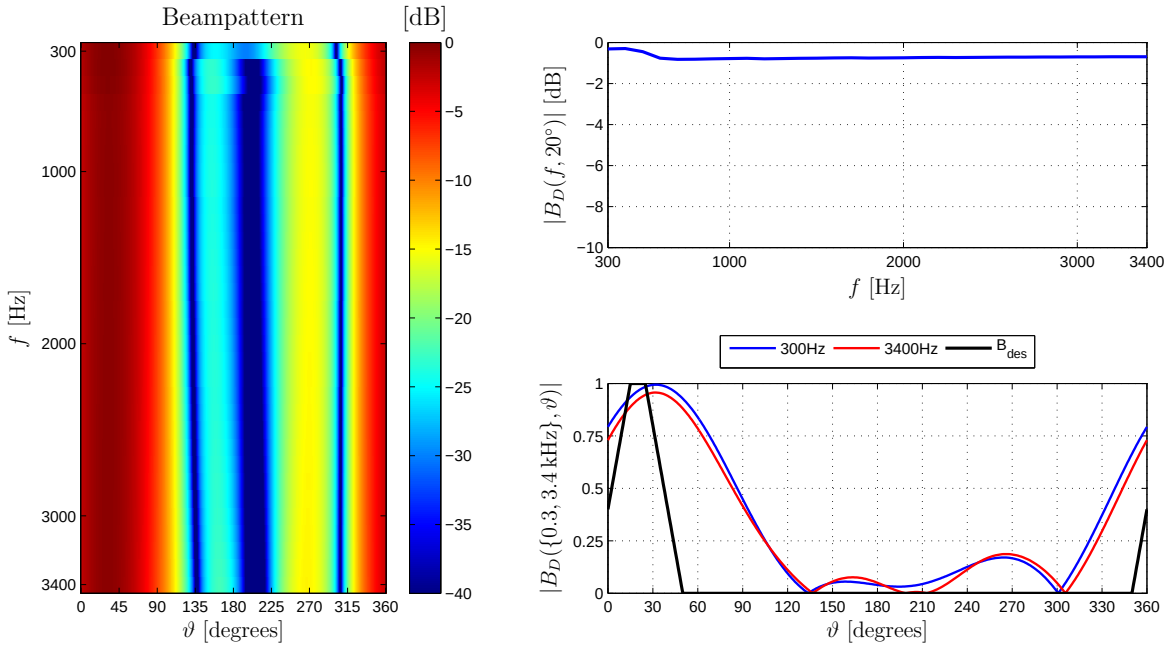
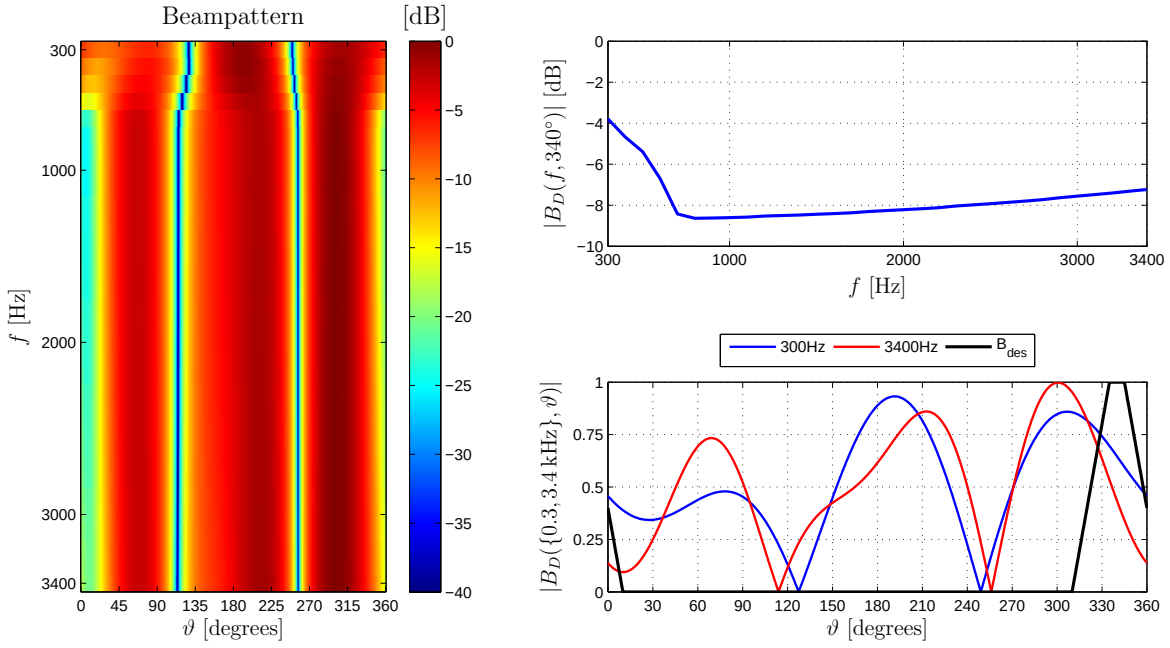


Fig. 4.18: Illustration of $I = 8$ prototype look directions and the corresponding values of D.

[4] Note that the ordinate already starts at $-10\,\text{dB}$.



(a) Look direction $\vartheta_{\text{des}} = 20°$ resulting from interpolation

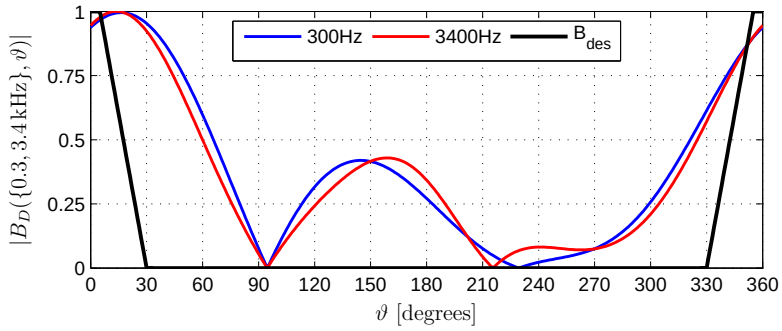


(b) Look direction $\vartheta_{\text{des}} = 340° \equiv -20°$ resulting from extrapolation

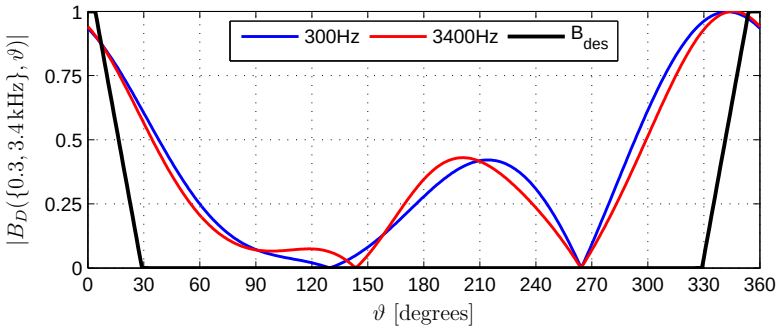
Fig. 4.19: Impact of extrapolation on the beamformer performance: The beampattern for the desired look direction $\vartheta_{\text{des}} = 20°$ obtained by interpolation is quite decent, whereas the beampattern for $\vartheta_{\text{des}} = 340° \equiv -20°$ is useless due to extrapolation.

around this point, since the responses can then change rapidly and sudden changes in signal amplification may occur, e.g., if an interfering source is located at $\vartheta \approx 215°$. The response obtained for $D = -1$ almost completely suppresses signals incident from this direction, whereas for $D = +1$ the attenuation is only around 7 dB. Also, due to the additional prototype look direction the complexity of the optimization problem increases. If I should be kept constant, the PLD initially located at $\vartheta_{\text{des}} = 315°$ can be shifted to 359°. This results, however, in an increase of the angular distance between all other sampling points. Note that for a linear array or when using sectorial optimization the PLD range is limited and the problem of having two sampling points for $\vartheta_{\text{des}} = 0°$ and $\vartheta_{\text{des}} = 359°$ does not occur.

An attempt to find a tradeoff between the two described methods is to split the extrapolation area and attach half of it each to the beginning and end of the steering range. That is, all prototype look directions are rotated in the anti-clockwise direction by half the angular distance between two PLDs, $\frac{1}{2} \cdot \frac{\vartheta_{\text{PLD}}}{I}$. However, this almost does not bring any improvement relative to the first case with one large extrapolation area at the end of the steering range. Therefore, we consider distributing the prototype look directions such that the entire D-axis from $-1, \ldots, +1$ is covered as the best choice as then extrapolation is not required.



(a) $\vartheta_{\text{des}} = 0°$ corresponding to $D = -1$



(b) $\vartheta_{\text{des}} = 359°$ corresponding to $D = +1$

Fig. 4.20: Magnitude responses for desired directions $\vartheta_{\text{des}} = 0°$ and $\vartheta_{\text{des}} = 359° \equiv -1°$ corresponding to $D = -1$ and $+1$, respectively.

4.4 Mean Square Error Discussion

In this section we want to show that the mean square error does not always reflect the performance of a beamformer completely and additional measures should be considered when evaluating the performance of the algorithms. For this we choose a smaller uniformly spaced circular array of radius $r = 2\,\text{cm}$ with $M = 4$ microphones located at $\vartheta_m = \{0°, 90°, 180°, 270°\}$ and apply two different beamformers with a constraint on the WNG, $A(\omega) \geq \gamma_{\text{dB}} = -20\,\text{dB}$. The first design is a robust polynomial beamformer designed for $I = 4$ prototype look directions at $\{0°, 90°, 180°, 270°\}$[5]. This system is compared to a robust polynomial beamformer with sectorial optimization, where the PLDs are at $\{0°, 30°, 60°, 90°\}$.

The resulting MSE is computed in 5° steps using Eq. (4.1) for $\vartheta_{\text{SR}} = [0°, 359°]$ and depicted in Fig. 4.21, where the large MSE-values of the RPB for $\vartheta_{\text{des}} > 270°$ are due to extrapolation (see chapter 4.3). The figure suggests that, e.g., for $\vartheta_{\text{des}} = 135°$, the RPB performs actually better than the RPBSO since its MSE-value is only about half as high. In fact, if one looks at the beampatterns shown in Fig. 4.22, one would surely prefer the RPBSO. This design is almost frequency-independent and the maximum deviation of the magnitude response for the desired look direction is less than 0.1 dB. In contrast, the RPB attenuates high frequencies up to 3.7 dB.

The reason why the MSE values for the RPB are for some angles smaller even though the beampattern is subjectively rated worse shall be explained by way of example. Let us imagine a fictional beamformer response which has an ideal passband

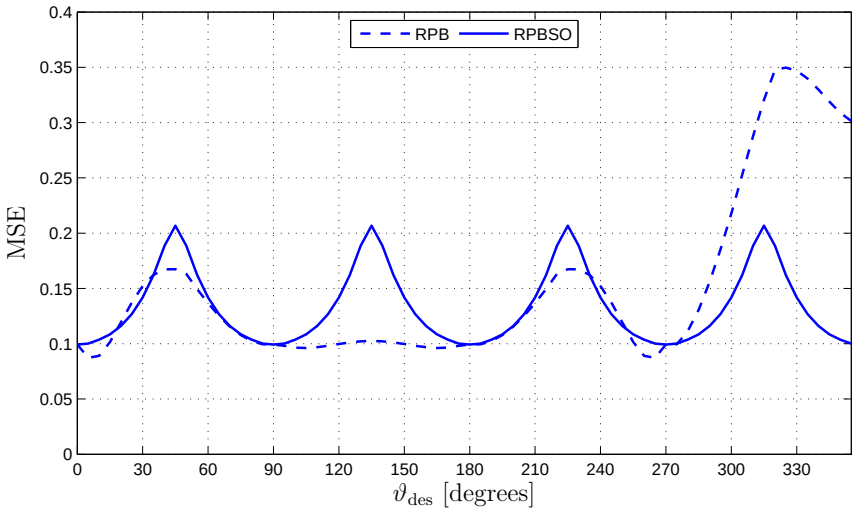
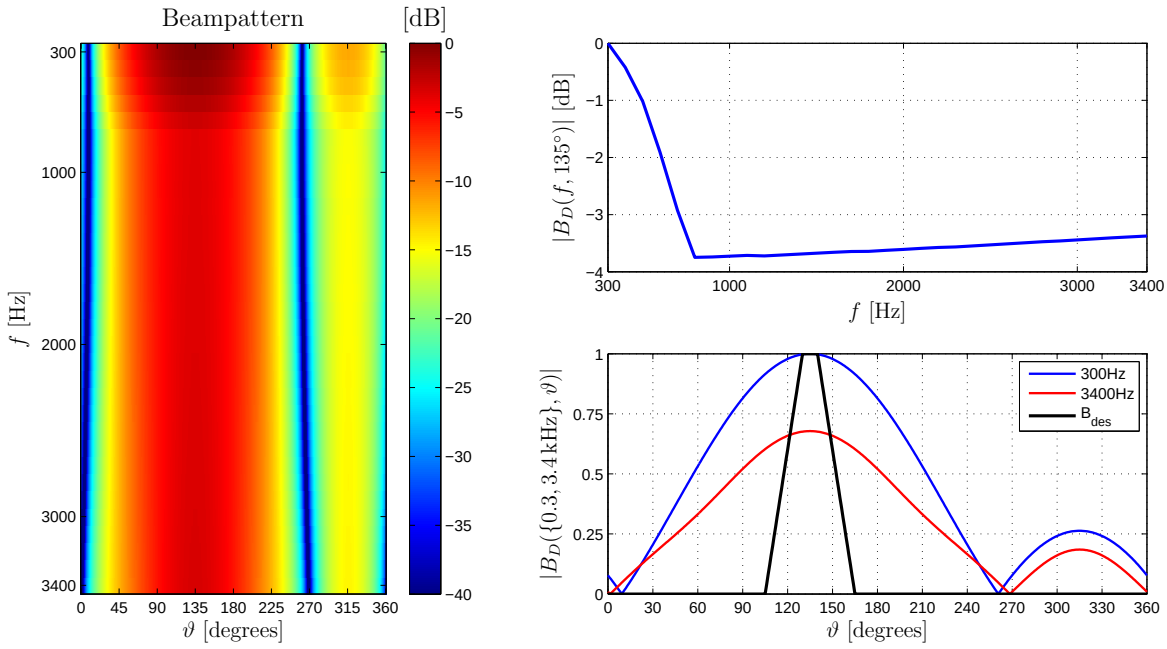
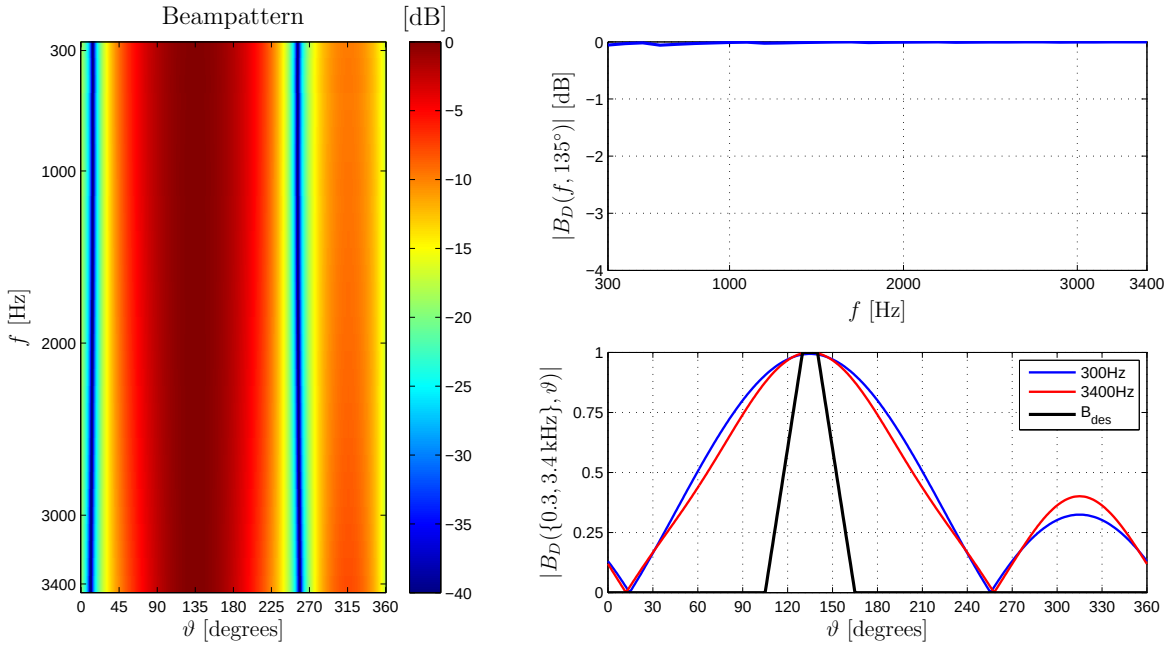


Fig. 4.21: MSE curves for an RPB and RPBSO applied to a uniformly spaced circular array with radius $r = 2\,\text{cm}$ and $M = 4$ microphones.

[5] Note that the PLDs do not cover the entire steering range, i.e., extrapolation is done for $\vartheta_{\text{des}} > 270°$.



(a) Robust polynomial beamformer



(b) Robust polynomial beamformer with sectorial optimization

Fig. 4.22: Depicted are the beampatterns and the magnitude responses in look direction $\vartheta_{\text{des}} = 135°$ and for $f = \{0.3,\ 3.4\}\,\text{kHz}$ obtained by an RPB (a) and RPBSO (b). The beamformers are applied to a uniformly spaced circular array with $M = 4$ microphones and radius $r = 2\,\text{cm}$.

of $\vartheta_{\text{des}} \pm 30°$, i.e., the response in this angular region is equal to one. In the stopband the signals are not completely suppressed, but weighted with 0.13, which is in logarithmic scale almost $-18\,\text{dB}$. This response would be fairly good and result in an MSE value of 0.075. Now, imagine a response which is equal to zero in the entire angular range, i.e., all signals are completely suppressed. Obviously, this is not at all a desirable response. However, it would lead to exactly the same MSE value of 0.075. This is due to the fact that deviations of the beamformer response from the desired response in the passband (a rather small angular region) have less impact on the overall MSE value than deviations in the wide stopband. Similarly, the (magnitude) response of the true RPB design shown on the lower right of Fig. 4.22(a) has a clear deviation in the passband, which is overcompensated by smaller errors in the stopband (relative to the RPBSO depicted in Fig. 4.22(b)) and finally results in a lower overall MSE value.

Since in some practical applications a distortionless response is more important than maximally suppressing interfering sources, the MSE does not always give a good subjective measure of the beamformer performance. A weighted mean square error measure could maybe account for this, but this is not in the scope of this thesis.

Here, we want to take a different aspect into account. Given a certain desired look direction ϑ_{des} the MSE-value obtained by evaluating Eq. (4.1) is just a single number. This number results from averaging the MSE along both frequency axis and angles. However, a beampattern has many different quality criteria and it is difficult to express the performance of a beamformer by just one single number. Therefore we want to examine the MSE as a function of angle, i.e., MSE(ϑ), and only take the average along frequency. Since the design goal of the beamformer is a frequency-independent beampattern this averaging will in general not degrade the information content in a great manner. The values of MSE(ϑ) are depicted in Fig. 4.23 and give more insights

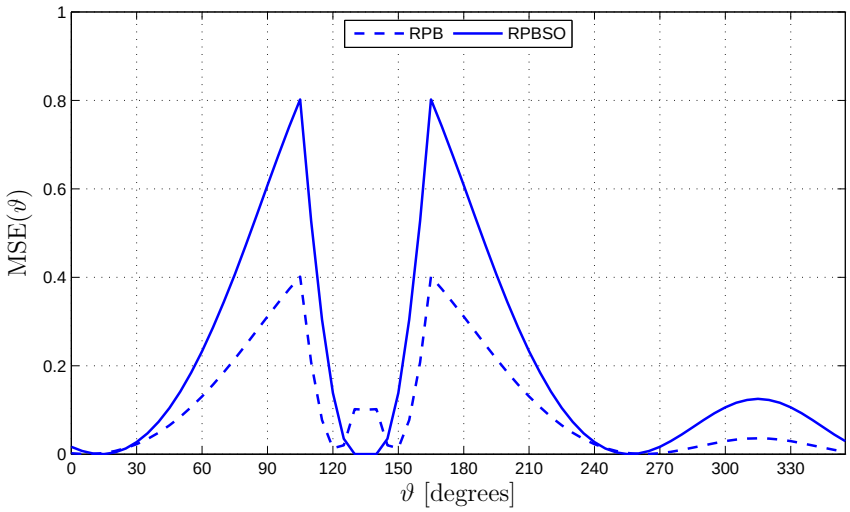


Fig. 4.23: MSE of an RPB and RPBSO as function of angle ϑ.

into the performance of a beamformer and the beampattern, respectively. We can see that MSE(ϑ) of the RPBSO is almost zero in the entire passband, $\vartheta = 130° \ldots 140°$. This implies that the magnitude response in the desired look direction $\vartheta_{\text{des}} = 135°$ is nearly distortionless. In case of the RPB MSE(ϑ) is about 0.1 in the passband, i.e., the magnitude response is not equal to one and the desired signal is attenuated. In the stopband the MSE values of the RPBSO are about twice as high as compared to the RPB. These results correspond well to magnitude responses and beampatterns shown in Fig. 4.22 and also explain which angular regions are responsible for the larger average MSE value of the RPBSO.

4.5 Conclusion

The simulation results confirm the benefit of exploiting the structure of symmetric arrays. By reducing the PLD range and designing the beamformer for only a limited angular region we can significantly increase the spatial selectivity which is shown by way of example for symmetric linear and symmetric circular arrays. Also, the deviations of the magnitude response for the desired look direction can be decreased. For the special case of uniformly spaced circular arrays the performance improvement obtained by the presented optimization scheme is comparable to the proposal of Lai et al. [LNL10]. However, it should be mentioned that the method presented in this work is more general and can be applied to any symmetric array, i.e., it contains the proposal of Lai et al. as a special case. The simulation results also show that the distribution of the PLDs has a large impact on the performance of polynomial beamformers. Another finding is that the MSE does not always reflect the actual performance of the beamformer completely.

5. Conclusions and Future Work

Polynomial beamforming is a versatile technique to perform spatial filtering of acoustic scenes and to enhance signals arriving from different desired directions. These beamformers are typically designed for several predefined prototype look directions covering the entire steering range. Steering the beam to intermediate directions is obtained by interpolation, where the prototype look directions act as sampling points. In order to obtain a robust beamformer besign the white noise gain is typically constrained. Also, a distortionless response constraint for the desired look direction can be added. Although these constraints are necessary, they reduce the number of available degrees of freedom.

In this thesis we presented a method which makes use of the structure of symmetric arrays. Thereby, the same number of prototype look directions can be distributed in only a part of the entire steering range, i.e., the angular distance between these sampling points is reduced and the polynomial beamformer is designed in a limited angular region only. All available degrees of freedom can then be used to optimize the beamformer responses for only a part of the entire angular region. The initial design region can be reduced by a factor of two for each symmetry axis which is exploited. During operation the beam can be steered to angles outside this design region by (consecutively) mirroring the set of filters to the opposite (w.r.t. the symmetry axes) microphones. Although we focus on symmetric linear and symmetric circular arrays in this work the proposed method can be applied to any type of symmetric array.

Simulation results for several designs show that the performance of polynomial beamformers can significantly be improved when applying the presented optimization scheme. The directivity can be clearly increased and the actual beamformer responses approximate the desired responses much better. Also, the deviations of the magnitude response for the desired look direction can be decreased. The benefit can as well be seen in terms of the white noise gain. When designing a robust polynomial beamformer a set of constraints assures that the white noise gain remains above a certain threshold for the prototype look directions. In between, however, the constraint may in general be violated since the beamformer response is then obtained by interpolation. Due to the fact that the sampling points move closer together when reducing the design region it is much more likely that the constraint on the WNG is also satisfied for interpolated look directions. In fact, all of the shown simulation results fulfill the specified constraint if the optimization is applied whereas the classical design fails in some cases. For the special case of uniformly spaced circular arrays the performance differences between the proposal of Lai et al. [LNL10] and the method introduced in this thesis are only

minor. If the array does not have symmetry axes the beamformer has to be designed for the entire steering range.

Another finding is that the MSE alone does not always give a good subjective measure of the performance of a beamformer. This is due to the fact that all angular regions are weighted equally when computing the MSE whereas in some practical applications an undistorted desired signal may be of higher priority than the suppression of interfering sources. Also, the MSE is just an averaged value which typically cannot express all different quality criteria of a beampattern. Therefore, additional measures should be considered when evaluating the performance of a beamformer.

In this work we only pointed out the possible improvements that can be achieved for a given scenario, i.e., for a fixed number of filter-and-sum units and order of polynomial postfilter, respectively. Another aim could be to obtain a certain performance with a reduced number of filter-and-sum units. This can be of special interest if the beamformer is combined with an acoustic echo cancellation system [HM07]. Then, each output signal of the filter-and-sum units requires one acoustic echo cancellation unit. In order to reduce the complexity a reduced number can therefore be desirable.

Also, we showed that the choice of the prototype look directions has a large impact on the performance of the beamformers. Another possible issue could therefore be the examination of non-uniformly distributed prototype look directions. For example, when using laptops with a built-in microphone array look directions perpendicular to the screen are very likely. Therefore, prototype look directions could be gathered around this angle in order to optimize the performance for this angular region.

Appendix

A. Notation

Abbreviations and Acronyms

DSB	delay-and-sum beamformer
DTFT	discrete-time Fourier transform
FSB	filter-and-sum beamformer
FSU	filter-and-sum unit
MSE	mean square error
PLD	prototype look direction
PPF	polynomial postfilter
RFSB	robust filter-and-sum beamformer
RPB	robust polynomial beamformer
RPBSO	robust polynomial beamformer with sectorial optimization
WNG	white noise gain

Mathematical Symbols

$(\cdot)^{\mathrm{H}}$	Hermitian operator
$\otimes$	Kronecker product
∇	nabla operator
$(\cdot)^{\mathrm{T}}$	transposition

Symbols

$\mathbf{a}(\omega)$	steering vector
$\mathbf{a}_{\mathrm{des}}(\omega)$	steering vector for desired look direction
$A(\omega)$	white noise gain

$\mathbf{A}(\omega_q)$	matrix containing K discretized steering vectors
$\mathcal{A}(\omega_q)$	matrix of size $K \times M(P+1)$ containing I products of $\mathbf{A}(\omega_q)\mathcal{D}_i$
$\mathbf{b}_{\mathrm{des}_i}$	vector containing K values of $B_{\mathrm{des},i}(\vartheta_k, \vartheta_{\mathrm{des}_i})$
$B(\omega, \vartheta)$	beamformer response
$B_D(\omega, \vartheta)$	response of a polynomial beamformer
$B_{\mathrm{des}}(\vartheta, \vartheta_{\mathrm{des}})$	desired response for look direction ϑ_{des}
$B_{\mathrm{des},i}(\vartheta_k, \vartheta_{\mathrm{des}_i})$	desired response at discretized angle ϑ_k for i-th look direction $\vartheta_{\mathrm{des}_i}$
$\mathcal{B}_{\mathrm{des}}$	vector containing all vectors $\mathbf{b}_{\mathrm{des}_i}^{\mathrm{T}}$
c	speed of sound
d	microphone spacing for a linear array
$\mathbf{d}_i$	vector containing steering direction D_i taken to the $(0, \ldots, P)$-th power
$D(\omega)$	directivity
D	steering direction
$\mathcal{D}_i$	block-diagonal matrix of size $M \times M(P+1)$ containing all vectors $\mathbf{d}_i^{\mathrm{T}}$
f	frequency
f_{s}	sampling frequency
I	number of prototype look directions
k	wave number: $k = \|\mathbf{k}\|_2$
$\mathbf{k}$	wave vector: $\mathbf{k} = [k_x, k_y, k_z]^{\mathrm{T}}$
K	number of discretized angles
L	longest dimension of an array
M	number of microphones
p	sound pressure
$\widehat{p}$	amplitude of the sound pressure
P	order of polynomial postfilter
Q	number of discretized angular frequencies
$\mathbf{r}$	position vector: $\mathbf{r} = [x, y, z]^{\mathrm{T}}$
$\mathbf{r}_{\mathrm{ref}}$	position vector to reference point
R	distance to a point source
R_0	reference distance

$s_m[n]$	signal of the m-th microphone
$s_{\text{ref}}[n]$	reference signal
$S_m(\omega)$	DTFT of $s_m[n]$
t	time
$\mathbf{w}(\omega)$	vector containing all $W_m(\omega)$
$\mathbf{w}_m$	vector containing L impulse response coefficients of m-th FIR filter
$\mathbf{w}_{m,p}$	vector containing L impulse response coefficients of m-th FIR filter in p-th FSU
$W_m(\omega)$	DTFT of $\mathbf{w}_m$
$W_{m,p}(\omega_q)$	sampled DTFT of $\mathbf{w}_{m,p}$
$\mathbf{W}(\omega_q)$	matrix containing all $W_{m,p}(\omega_q)$
$\mathcal{W}$	diagonal weighting matrix of size $KI \times KI$
x_m	x-coordinate of the m-th microphone
$y[n]$	output signal of a Beamformer
$y_p[n]$	output signal of p-th FSU
$y[n, D]$	output signal of a polynomial beamformer
$Y(\omega)$	DTFT of $y[n]$
ϕ	elevation angle
ϕ_{des}	elevation angle of desired sound source
ϑ	azimuth angle
ϑ_{des}	azimuth angle of desired sound source
ϑ_m	azimuth angle of m-th microphone
ϑ_{PLD}	PLD range
ϑ_{SR}	steering range
γ	threshold for white noise gain
γ_{dB}	threshold for white noise gain in logarithmic scale
λ	wavelength
τ_m	relative delay of the m-th microphone signal
$\tilde{\tau}_m$	delay applied to the m-th channel of a DSB
ω	angular frequency

B. List of Figures

C. Beampattern of an RFSB

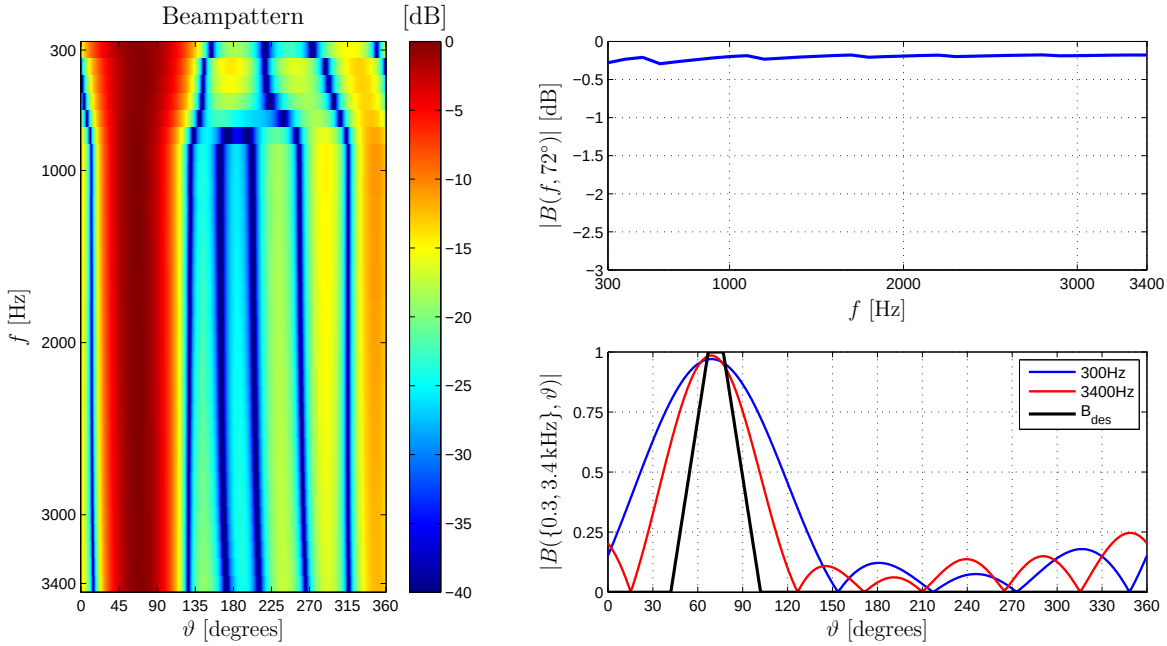


Fig. C.1: Shown is the beampattern, the magnitude responses in look direction $\vartheta_{\text{des}} = 72°$ and for $f = \{0.3,\ 3.4\}\,\text{kHz}$, respectively, obtained by an RFSB. The beamformer is applied to a symmetric circular array consisting of $M = 6$ microphones with radius $r = 3\,\text{cm}$.

Bibliography

[BW01] M. Brandstein and D. Ward. *Microphone Arrays: Signal Processing Techniques and Applications*. Springer Verlag, 2001.

[Ear04] J. Eargle. *The Microphone Book*. Focal Press, 2004.

[Her05] W. Herbordt. *Sound Capture for Human / Machine Interfaces: Practical Aspects of Microphone Array Signal Processing (Lecture Notes in Control and Information Sciences)*. Springer, 1st edition, March 2005.

[HM07] M. Hamalainen and V. Myllyla. Acoustic echo cancellation for dynamically steered microphone array systems. pages 58–61, Oct. 2007.

[LNL10] C. Lai, S. Nordholm, and Y. Leung. Design of Robust Steerable Beamformers with Spiral Arrays and the Farrow Filter Structure. *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, August 2010.

[LSW08] R. Lerch, G. Sessler, and D. Wolf. *Technische Akustik: Grundlagen und Anwendungen*. Springer, 2008.

[McC01] I. McCowan. Microphone arrays: A tutorial. *Queensland University, Australia*, pages 1–38, 2001.

[MK10] E. Mabande and W. Kellermann. Design of Robust Polynomial Beamformers as a Convex Optimization Problem. In *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel-Aviv, Isreal, August 2010.

[MSK09] E. Mabande, A. Schad, and W. Kellermann. Design of Robust Superdirective Beamformers as a Convex Optimization Problem. *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 77–80, April 2009.

[Teu07] H. Teutsch. *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition (Lecture Notes in Control and Information Sciences)*, volume 348. Springer Verlag, 2007.

[VT02] Harry L. Van Trees. *Optimum Array Processing (Detection, Estimation, and Modulation Theory, Part IV)*. Wiley-Interscience, 1st edition, April 2002.

[VVB88] B. D. Van Veen and K. M. Buckley. Beamforming: A versatile approach to spatial filtering. *ASSP Magazine, IEEE*, 5(2):4–24, 1988.

[ZZ98] M. Zollner and E. Zwicker. *Elektroakustik*. Springer, 1998.