Chapter 8
Room in a Room: A Neglected Concept for Auralization

Markus Christoph

Abstract The term *Room in a Room* describes an old technique which can be utilized for auralization purposes. Thereby the auralization procedure can be divided into an analysis- and a synthesis task, usually dependent of each other. The analysis recordings in the source room have to be made exactly at those directions where loudspeakers are physically placed in the target room. In doing so, the spatial separation, respectively, filtering will be realized by beamforming. The outputs of the different, fixed beamformers provide the signals for the loudspeakers in the target room, representing the synthesis of this auralization technique. Different methods of how such a beamformer can robustly be designed will be presented in this chapter.

Keywords Auralization • Beamforming • 3D-beamformer • Modal beamforming

8.1 Introduction

Auralization techniques have been an object of interest since a long time. They can be used to create an illusion of being acoustically in a desired room, whilst, in reality, sitting in a completely different one. If this illusion becomes authentic, such techniques can be utilized, for example, for acoustical documentation or virtual tuning purposes. Especially in automotive acoustics, such a tool is highly desired, since up to now the majority of prototype cars are still individually tuned by an acoustician, mostly by hand, which is very time consuming. At the same time, automotive companies constantly cut time frames for such tasks, since working time on the prototypes is expensive, as many people from different disciplines have to work on them. By using auralization techniques one could measure the car once

M. Christoph (✉)
Harman/Becker Automotive Systems GmbH, Schlesische Str. 135, Straubing, Germany
e-mail: markus.christoph@harman.com

and work, from this moment on, solely on this data, to get the car tuned. In doing so, valuable working time on a prototype can be reduced to a minimum, whilst at the same time; tuning time can be increased, eventually leading to a better acoustical result.

8.1.1 Background

There is a whole number of possibilities how auralization can be achieved, which can roughly be divided into two categories: headphone- and loudspeaker-based solutions. As representatives of the headphone solutions, a pure binaural recording and reproduction solution, as is more deeply described in [1], can be mentioned. Thereby the idea of the system, presented in [1], is to duplicate the binaural signal, as recorded directly at the eardrums within the source room by a headphone reproduction system. Thereby the headphones have to be adequately compensated, ideally with the inverse of the headphone transfer function, which can only be realized approximately, since this transfer function is usually not minimum phase. Despite its simplicity, this method is able to deliver a very authentic room impression. The BRS (Binaural Room Scanning) principle [2] is also based on binaural recordings, which are, in contrast to the before mentioned principle, not individualized. Due to the fact that BRS utilizes general HRTFs (Head Related Transfer Function) data, a head tracking system has to be applied in order to externalize the acoustic expression, or in other words to avoid in-the-head localizations. Thereby, the head tracker is able to measure the current head rotation. This information is then used to pick the most adequate, i.e., the two, in 2D applications, neighboring HRTFs out of a library, interpolate in between them and insert the result into a convolution machine. As a consequence an impression can be achieved that the stage does not move with the head during head rotations.

The topic of this chapter belongs to the category of loudspeaker-based auralization methods. There exist a number of different methods, from which only a few, which are considered to be the most promising ones, will briefly be mentioned in the following. In [3] the author utilized the *inverse filter theory* as three dimensional (=3D) sound reproduction technique. The aim of this technique is to compensate for any disturbing effects of the target room, by utilizing matrix inversion to calculate the necessary inverse filter matrix such, that in connection with the target room, a Dirac impulse, at the desired location within the target room, will ideally result. As soon as this task is successfully accomplished, it is easy to create any desired room impression by additionally inserting the room transfer matrix of the desired room into the system, which could, above all, be efficiently integrated into the already inevitable, inverse filter matrix. Despite its mathematical correctness, this principle suffers from diverse practical problems, such that the results are only valid at the location in the target room where the measurement had been applied but not at its close proximity. Furthermore, the acoustics are adversely affected by the inverse filter, as they usually show a great deal of *pre-ringing*.

The term *Ambisonics*, as described, e.g., in [4], stands for an analysis and synthesis method able to measure and reproduce spherical harmonics of a sound field. Thereby, the main task consists of creating a gain or filter matrix, jointly maximizing the energy as well as the particle velocity vector. Since Ambisonic is based on free-field conditions, it can only auralize the desired wave field as expected, if the target room is free of reverberation and/or early reflections, which is the case if the target room is an anechoic chamber. This problem is tackled by the *DirAC* (Directional Audio Coding) system, as introduced in [5], which can be looked upon as an extension of the previously mentioned Ambisonic system. Here, in addition to the extraction of the azimuth and elevation information from the B-Format signal, which act as input, as is the case in the Ambisonic system, spatial parameters, like the diffuseness are extracted during analysis and used in synthesis, with the objective of creating a realistic spatial impression. This of course can only approximate spatial impressions. If one wants to replicate the "real" sound field by loudspeakers, probably the most accurate way is to utilize Wave Field Synthesis (WFS), as disclosed, e.g., in [6], or High Order Ambisonics (HOA) as introduced, e.g., in [7], despite the fact that here one has to deal with still unsolved problems too, such as spatial aliasing effects. Another negative aspect in this concern is the enormous effort necessary to successfully run a WFS or HOA system, since a great number of (closely spaced) loudspeakers and the accompanying signal processing are necessary to create the desired effect.

The aim of the *Room in a Room* method, as introduced in this chapter, is to create a realistic sound field in the target room, in an easy and efficient way, thereby circumventing diverse problems inherent in some of the above-mentioned principles.

This chapter is structured as follows. In Sect. 8.2 the principles of the Room in a Room method is presented. Then, in Sect. 8.3 the first utilized beamforming technique is reviewed. Thereby, the underlying 3D microphone array is presented, as well as a new concept of how superdirective beamformers can effectively be combined by utilization of the presented 3D microphone array, which forms the "heart" of the whole method. Section 8.4 presents simulations of the novel beamforming technique and reconsiders its outputs. In Sect. 8.5 measurements of the novel beamformer are discussed. Finally, Sect. 8.6 summarizes this chapter.

8.2 Room in a Room

As one can perceive in Fig. 8.1, a microphone array is placed in the source room at a desired position, which acoustic should eventually be reproduced at a certain position within the target room. Thereby two types of recordings are feasible. Firstly it is always possible to directly record the desired sound, picked up by all microphones, which we will subsequently refer to as *signal-dependent recording*. This is necessary if one wants to "document" the acoustics, e.g., of an opus at a specific location within an opera. Secondly, if the sound which should be
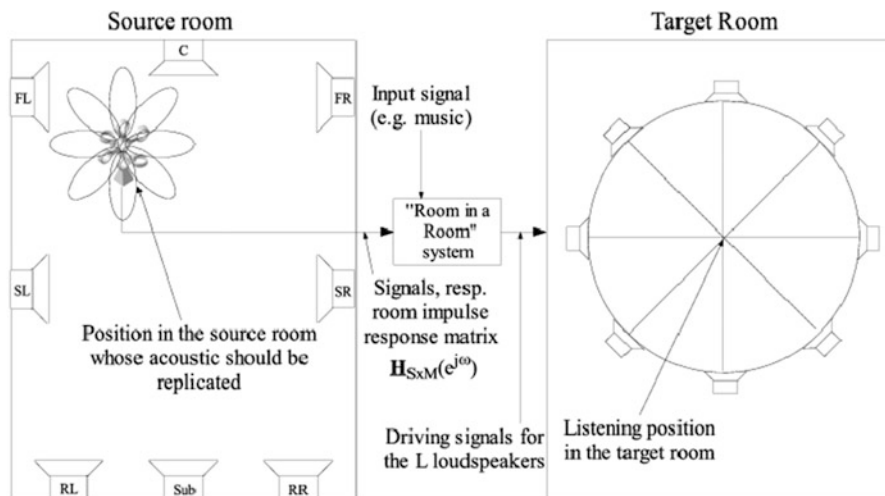
Fig. 8.1 Principle of the *Room in a Room* auralization method

reproduced stems from a sound system with a defined number and location of loudspeakers, as is the case in an automobile, it is reasonable to pick up the room impulse responses (RIR) from all S speakers to all M microphones of the microphone array in order to eventually create a signal-independent auralization system, which we will subsequently call *room-dependent recording*.

Independent of the type of recording, successively, based on the recorded data, beamforming filter will be applied, such that L beams, pointing exactly to the positions of the L loudspeakers, as located in the target room, will be created. Regarding a signal-dependent recording, further processing for the synthesis is not necessary. Figure 8.2 depicts all steps of the synthesis procedure, necessary for the room-dependent recording. In this process, based on the N input signals x, the driving signals for the S speakers (located in the target room) have to be calculated, by replication of the whole signal processing chain of the sound system, utilized in the target room. This may simply consist of a pure passive upmixing matrix $\underline{\mathbf{M}}_{N \times S}$ ($e^{j\omega}$), broadcasting parts, combinations or originals of the N input signals to the S speakers. Afterwards, the influence of the transfer functions of the target room, stored in the RIR matrix $\underline{\mathbf{H}}_{S \times M}(e^{j\omega})$, will be considered during the course to create the M virtual microphone signals, as would be picked up by the microphone array if one were to play the desired input signal in the target room, utilizing its sound system and directly record these signals at a desired location by the M microphones of the microphone array. Finally, on grounds of these "virtual" microphone signals, L beams, pointing to the L loudspeakers, as located in the target room will be designed, whose output signals form their driving signals, the same as in the signal-dependent recording.
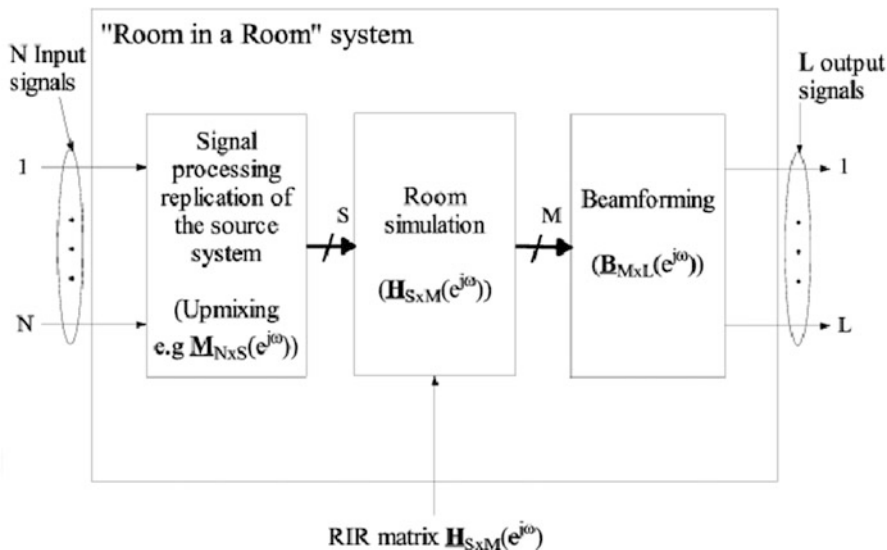
Fig. 8.2 Synthesis of room-dependent recordings

8.3   Beamforming

A microphone array consisting at least of two microphones, from which, at least one signal is filtered, with a successive filter, calculated such that a desired spatial filtering will eventually arise by combination of the processed microphone signals, is called a *beamformer*.

In the coordinate system, utilized for the design of a beamformer, as shown in Fig. 8.3, it is expected that all microphones are aligned along the *x*-axis. Furthermore, to design a beamformer with a beam pointing to a desired direction, e.g., as depicted by the vector **u** in Fig. 8.3, the direction of the beam will be assigned by its corresponding horizontal- (azimuth Θ) and vertical angle (elevation φ).

As revealed by Fig. 8.4, showing a beamformer, implemented in the spectral domain, the required signal processing can be divided into two parts, that is to say the so-called beamsteering on the one hand, which stands for a time delay compensation, necessary to ensure coherent, i.e., phase aligned addition of the microphone signals by $e^{j\omega\tau_i}$, and the filtering on the other hand by $\mathbf{A}(\omega)$, which performs the intrinsic spatial filtering. It should be noted that in Fig. 8.4, it is expected that free-field conditions be met, i.e., signals picked up by the microphones differ in their phasing but not in their amplitude.
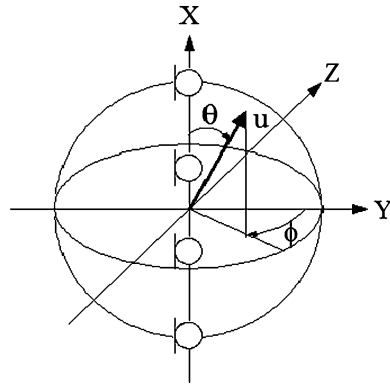
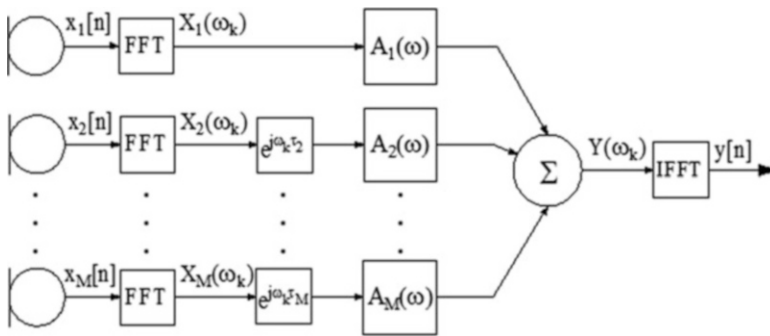**Fig. 8.3** Coordinate system utilized for the design of a beamformer

Fig. 8.4 Signal flow diagram of a beamformer, realized in the spectral domain

8.3.1 Design of Optimal Beamforming Filter

Optimal beamforming filter $\mathbf{A}(\omega)$ can generally be calculated as follow:

$$\mathbf{A}(\omega) = \frac{\underline{\varphi}_{NN}^{-1}(\omega)\mathbf{d}(\omega)}{\mathbf{d}^{H}(\omega)\underline{\varphi}_{NN}^{-1}(\omega)\mathbf{d}(\omega)} \tag{8.1}$$

where

$\mathbf{A}(\omega)$=Vector holding the beamforming filter

$\mathbf{A}(\omega) = [A_1(\omega), \ldots, A_M(\omega)]^{T}$, where $\mathbf{X}^{T}$ denotes the transpose of $\mathbf{X}$,

$\underline{\varphi}_{NN}(\omega)$=Power spectral density (PSD) matrix of the background noise N,

$\mathbf{X}^{H}$=Hermitian (conjugate transpose) of $\mathbf{X}$,

ω=Angular frequency in $\left[\frac{1}{s}\right]$ ($\omega = 2\pi f$),
$\mathbf{d}(\omega)$=Steering vector $\mathbf{d}(\omega) = [d_1(\omega), \ldots, d_M(\omega)]^T$,
with

$$\underline{\varphi}_{NN}(\omega) = \begin{pmatrix} \varphi_{N_1N_1}(\omega) & \varphi_{N_1N_2}(\omega) & \cdots & \varphi_{N_1N_M}(\omega) \\ \varphi_{N_2N_1}(\omega) & \varphi_{N_2N_2}(\omega) & \cdots & \varphi_{N_2N_M}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{N_MN_1}(\omega) & \varphi_{N_MN_2}(\omega) & \cdots & \varphi_{N_MN_M}(\omega) \end{pmatrix} \tag{8.2}$$

where
M=Number of microphones,
and

$$d(\omega) = e^{-j\frac{\omega\left(\frac{M+1}{2}-n\right) d\cos(\Theta_0)}{c}} \tag{8.3}$$

where
$n \in [1, \ldots, M]$,
c=Speed of sound in $\left[\frac{m}{s}\right]$ $\left(c = 343\left[\frac{m}{s}\right]@\vartheta = 20°C\right)$,
Θ_0=Main receive direction, respectively direction, where the beam points in [rad].

In case, the sound source resides in the near field, the beam steering vector $\mathbf{d}(\omega)$ calculates to:

$$d(\omega) = \left[a_0 e^{-j(2\pi f\, \tau_0)}, a_1 e^{-j(2\pi f\, \tau_1)}, \ldots, a_{M-1}e^{-j(2\pi f\, \tau_{M-1})}\right] \tag{8.4}$$

where
a_0=Amplitude compensation value of the nth microphone signal,
τ_n=Time compensation value of the nth microphone signal,
    with

$$a_n = \frac{\left\|q - p_{ref}\right\|}{\left\|q - p_n\right\|},$$

where
$\left\|q - p_{ref}\right\|$=Distance between the sound source q and the reference microphone p_{ref} in [m],
$\left\|q - p_n\right\|$=Distance between the sound source q and nth microphone p_n in [m].

Regarding a rule of thumb, one talk about far-field conditions if the sound source is located at a distance from the microphone array, exceeding twice its dimension, which is usually always the case, hence, (8.4) usually does not apply in practical applications.

8.3.2   Practical Modifications

According to Fig. 8.4 one usually excludes the beam steering vector $\mathbf{d}(\omega)$ from the design of the beamforming filter $\mathbf{A}(\omega)$. The beam steering is usually applied upstream the actual beamforming filter, i.e., first one calculates the delays and phase shifts, necessary for all microphones of the array, combined in the beam steering vector $\mathbf{d}(\omega)$, in order to let the resulting beam point to the desired direction, before the beamforming filter $\mathbf{A}(\omega)$ takes place. Thus the steering vector $\mathbf{d}(\omega)$ within (8.1) reduces to $\mathbf{d}(\omega) = \mathbf{1} = [1,1, \ldots,1]^{T}$.

In a further step, the cross-correlation matrix of the background noise $\underline{\varphi}_{NN}(\omega)$, which usually has to be measured continuously, or at least in situ, will be replaced by the complex coherence matrix of a diffuse noise field $\underline{\mathbf{\Gamma}}(\omega)$, for which a closed solution exists. This modification can be conducted, since measurements showed, that spatially homogeneous noise fields, as approximately apparent in automobiles, closely resemble a diffuse noise field.

Taking all these modifications into account, the design of the beamforming filter converts to:

$$\mathbf{A}(\omega) = \frac{\underline{\mathbf{\Gamma}}^{-1}(\omega)\mathbf{1}}{\mathbf{1}^{\mathrm{H}}\underline{\mathbf{\Gamma}}^{-1}(\omega)\mathbf{1}} \tag{8.5}$$

where
$\underline{\mathbf{\Gamma}}(\omega)$=Complex coherence of a diffuse noise field,
$\mathbf{1}$=Residual, respectively neutral steering vector $\mathbf{1} = M[1, 1, \ldots, 1]^{T} M$.

After the beam steering has been carried out, the complex coherence matrix of the diffuse noise field $\underline{\mathbf{\Gamma}}(\omega)$ calculates to:

$$\underline{\mathbf{\Gamma}}(\omega) = \begin{pmatrix} 1 & \Gamma_{X_1X_2}(\omega) & \cdots & \Gamma_{X_1X_M}(\omega) \\ \Gamma_{X_2X_1}(\omega) & 1 & \cdots & \Gamma_{X_2X_M}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{X_MX_1}(\omega) & \Gamma_{X_MX_2}(\omega) & \cdots & 1 \end{pmatrix} \tag{8.6}$$

with

$$\Gamma_{X_iX_j}(\omega) = \mathrm{sinc}\left(\frac{\omega\,d_{ij}}{c}\right) e^{-j\frac{\omega\,d_{ij}\cos(\Theta_0)}{c}} \tag{8.7}$$

where
$i, j \in [1, \ldots,M]$,
$\mathrm{sinc}(x)$=Sinc function $\left(\frac{\sin(x)}{x}\right)$,
d_{ij}=Element located at the ith row and jth column of the distance matrix $\underline{\mathbf{d}}$,

and

$$\underline{\mathbf{d}} = \begin{pmatrix} 0 & d & \cdots & (M-1)\,d \\ -d & 0 & \cdots & (M-2)\,d \\ \vdots & \vdots & \ddots & \vdots \\ -(M-1)\,d & -(M-2)\,d & \cdots & 0 \end{pmatrix} \qquad (8.8)$$

where
d=Inter-microphone distance in [m] of the equidistant microphone array.

8.3.3 Constrained Design

Both design rules, shown in (8.1) and (8.5) deliver the same, optimal beamforming filter $\mathbf{A}(\omega)$, in a diffuse noise field. Unfortunately neither the one nor the other can be applied without any further modifications, considering inevitable, practical limits, such as manufacturing tolerances, or variations in the placement of the microphones. These incertitudes have been considered in [8] by the addition of a small scalar μ to the elements at the main diagonal of the cross-correlation matrix $\underline{\varphi}_{NN}(\omega)$, or as proposed in [9] to the coherence matrix of a diffuse noise field $\underline{\mathbf{\Gamma}}(\omega)$. Another version, disclosed in [10], directly considers the inaccuracies in the design of the beamforming filter, leading to a constrained filter design as follows:

$$\mathbf{A}(\omega) = \frac{\left(\underline{\mathbf{\Gamma}}(\omega) + \mu(\omega)\underline{\mathbf{I}}\right)^{-1} \mathbf{d}(\omega)}{\mathbf{d}(\omega)^{\mathrm{H}} \left(\underline{\mathbf{\Gamma}}(\omega) + \mu(\omega)\underline{\mathbf{I}}\right)^{-1} \mathbf{d}(\omega)} \qquad (8.9)$$

where
$\mathbf{d}(\omega)$=Steering vector (=$\mathbf{1}$, if previously applied),
$\underline{\mathbf{I}}$=Identity matrix in the size of $\underline{\mathbf{\Gamma}}(\omega)$,
$\mu(\omega)$=Regularization parameter.

The value of the regularization parameter $\mu(\omega)$, which is now frequency dependent, and not a scalar as, e.g., in [8], depends on the MSE[1] of the imprecision of the placement of the microphones ($=\delta(\omega)^2$) within the array, but mainly on the MSE of the inter-microphone tolerances ($=\varepsilon(\omega,\Theta)^2$). The higher the quality of the microphone array, i.e., the lower the complete MSE ($=\Delta(\omega,\Theta)^2$), the smaller the value for the regularization parameter $\mu(\omega)$ can be. Practical values reside within a range of: $\mu(\omega) = [-40, \ldots, 40]$ [dB].

---

[1] MSE = Mean Squared Error.

The susceptibility $K(\omega)$ of a beamformer, given as:

$$K(\omega) = \frac{\mathbf{A}(\omega)^H \mathbf{A}(\omega)}{\left| \mathbf{A}(\omega)^H \mathbf{d}(\omega) \right|}, \tag{8.10}$$

describes the sensitivity of a beamformer regarding tolerances of the corresponding microphone array. Aim of the constraint algorithm is to design a robust beamformer by limiting the susceptibility to a maximal value $K_{\text{Max}}(\omega)$. After [10], this upper limit $K_{\text{Max}}(\omega)$ directly results from the total MSE of the microphone array $\Delta(\omega,\Theta)^2$ and the maximum tolerable deviations of the directional diagram $\Delta\Psi(\omega,\Theta)$, with the directional diagram $\Psi(\omega,\Theta)$ given as:

$$\Psi(\omega,\Theta) = \frac{\varphi_{y,y}(\omega,\Theta)}{\varphi_{x_{ref},x_{ref}}(\omega,\Theta)} = \left| \sum_{n=1}^{M} \mathbf{A}(\omega)\, e^{j\frac{\omega d \left(\frac{(M+1)}{2} - n \right) \left(\cos(\Theta_0) - \cos(\Theta) \right)}{c}} \right|^2 \tag{8.11}$$

where
$\varphi_{y,y}(\omega,\Theta)$=Auto power spectral density of the beamformer output signal y,
$\varphi_{x_{ref},x_{ref}}(\omega,\Theta)$=Auto power spectral density of the reference microphone signal x_{ref}.
   For the total MSE of the microphone array holds:

$$\Delta(\omega,\Theta)^2 = \varepsilon(\omega,\Theta)^2 + \delta(\omega)^2 \tag{8.12}$$

with

$$\varepsilon(\omega,\Theta)^2 = E\left\{ \frac{\left| \Delta H_n^M(\omega,\Theta) \right|^2}{\left| H_0^M(\omega,\Theta) \right|^2} \right\} \tag{8.13}$$

where
$|H_0^M(\omega,\Theta)|^2$=Nominal, respectively mean transfer function of all microphones,
$|\Delta H_n^M(\omega,\Theta)|^2$=Deviation of the transfer function of the nth microphone from the
   nominal transfer function,
$E\{.\}$=Expectation operator,
and

$$\delta(\omega)^2 = \left(\frac{\omega}{c} \right)^2 \frac{\sigma^2}{3} \tag{8.14}$$

where
σ=Variance of the zero-mean, normally distributed positioning error of the
   microphone, equal for each dimension, hence the scaling by $\frac{1}{3}$.

Tolerances in the microphone array can be considered in the directional diagram $\Psi(\omega,\Theta)$ by addition of an error term, represented by $\Delta\Psi(\omega,\Theta)$ to it, resulting in:

$$E\left\{\widetilde{\Psi}(\omega,\Theta)\right\} = \Psi(\omega,\Theta) + \Delta\Psi(\omega,\Theta) \qquad (8.15)$$

with

$$\Delta\Psi(\omega,\Theta) = \Delta(\omega,\Theta)^2 K(\omega) \leq \Delta\Psi_{\mathrm{Max}}(\omega,\Theta), \qquad (8.16)$$

which must not exceed a certain threshold, provided by $\Delta\Psi_{\mathrm{Max}}(\omega,\Theta)$.

By inserting (8.12) and (8.14) in (8.16), it follows a maximally tolerable susceptibility $K_{\mathrm{Max}}(\omega,\Theta)$ of:

$$K_{\mathrm{Max}}(\omega,\Theta) = \frac{\Delta\Psi_{\mathrm{Max}}(\omega,\Theta)}{\varepsilon(\omega,\Theta)^2 + \left(\frac{\omega}{c}\right)^2 \frac{\sigma^2}{3}} \qquad (8.17)$$

The following practical simplifications can be applied to (8.17):

- Due to the fact that $\varepsilon(\omega,\Theta)^2$ hardly varies with Θ it suffices to determine $\varepsilon(\omega,\Theta)^2$ at a certain receive direction. Thereby the main receive direction Θ_0 is usually selected, which is $\Theta_0 = 90°$ in broadside and $\Theta_0 = 0°$ in endfire alignment of the beamformer.
- Inaccuracies in microphone placements, represented by $\delta(\omega)^2$, are much less probable then variations in the transfer functions of the array microphones, provided by $\varepsilon(\omega,\Theta)^2$. As such it suffices to consider mechanical deviations by a general value of, e.g., $\delta(\omega)^2 = 1\%$.
- A dependency on Θ of $\Delta\Psi_{\mathrm{Max}}(\omega,\Theta)$ only makes sense, if one is interested in an exact reconstruction of the whole directional pattern, i.e., also of all side lobes, which is usually not the case. By taking a maximal, Θ-independent value $\Delta\Psi_{\mathrm{Max}}(\omega)$, an almost perfect replication of the directional pattern in the main direction can still be obtained. Thereby, $\Delta\Psi_{\mathrm{Max}}(\omega)$ can be determined by taking the maximum side lobe value of the ideal directional pattern. Furthermore, dependent on the use case, $\Delta\Psi_{\mathrm{Max}}(\omega)$ could also be utilized as a frequency-independent threshold, e.g., $\Delta\Psi_{\mathrm{Max}} = 15[\mathrm{dB}]$.

Taking all above-mentioned items into consideration, (8.17) simplifies to:

$$K_{\mathrm{Max}}(\omega) = \frac{\Delta\Psi_{\mathrm{Max}}(\omega)}{\varepsilon(\omega)^2 + 1\%} \qquad (8.18)$$

with

$$\varepsilon(\omega)^2 \geq 1\% \qquad (8.19)$$

Based on the previous findings, the following, iterative, constrained algorithm for the design of the beamforming filter $\mathbf{A}(\omega)$, eventually leading to a robust, *superdirective beamformer* can be derived:

1. Preliminaries:

    (a) Determine $\Delta\Psi_{\text{Max}}(\omega)$, based on the maximum values of the side lobes of the desired, ideal beamformer, over frequency (for $M = 3$: $\Delta\Psi_{\text{Max}}(\omega) \approx -9.5$ [dB]).
    (b) Measure all transfer functions of the microphones $H_n^M(\omega)$ at the desired main direction Θ_0. Afterwards, use (8.13) to calculate $\varepsilon(\omega)^2$, thereby taking (8.19) into account.

2. Calculation of the maximum allowable susceptibility $K_{\text{Max}}(\omega)$, utilizing (8.18).
3. As initialization for the iteration, use $\mu(\omega) = 1$.
4. Calculation of the beamforming filter $\mathbf{A}(\omega)$, utilizing (8.9).
5. Based on the beamforming filter $\mathbf{A}(\omega)$, calculated in step 4, calculate the current susceptibility $K(\omega)$, utilizing (8.10).
6. Increase the regularization parameter $\mu(\omega)$, if $K(\omega) > K_{\text{Max}}(\omega)$, otherwise decrease $\mu(\omega)$, e.g., by $\Delta\mu = 10^{-5}$.
7. Repeat steps 4–6 until $K(\omega)$ approaches $K_{\text{Max}}(\omega)$ as close as possible or if $\mu(\omega)$ drops below a certain lower threshold, given, e.g., by $\mu_{\text{Min}} = 10^{-8}$, which is usually the case at higher frequencies $f \geq \frac{c}{2d}$.

8.4 Microphone Array

The susceptibility $K(\omega)$ of a beamformer mainly depends on the deviations of the inter-microphone transfer functions $\varepsilon(\omega)^2$, as discussed in Sect. 8.3.3. To enhance the quality of the beamformer, these differences have to be kept as small as possible. Therefore, so called *matched* or *paired microphones* have been used during the construction of the microphone array, which frame, by the way, is shown in Fig. 8.1. For this purpose the transfer functions of 100 microphone capsules (Panasonic WM-62a) have been measured at $\Theta_0 = 0°$ in an anechoic chamber, from which the 7, best matching capsules have been chosen, as shown in Fig. 8.5.

Since the sensitivity of a beamformer against tolerances decreases with increasing frequency, a frequency-dependent weighting function, provided by a nonlinear smoothing filter (e.g., $\frac{1}{3}$ octave filter) has been applied during the selection process, prior to the calculation of the difference matrices.

An analysis of the inter-microphone differences revealed, that due to the selection process, the deviation could be decreased from primarily ± 3 [dB], as provided by the data sheet of the manufacturer, to ± 0.5 [dB], as shown in Fig. 8.6, corresponding to an value of $\varepsilon(\omega)^2 < 0.7\%$, which is already below the lower limit as noted in (8.19).
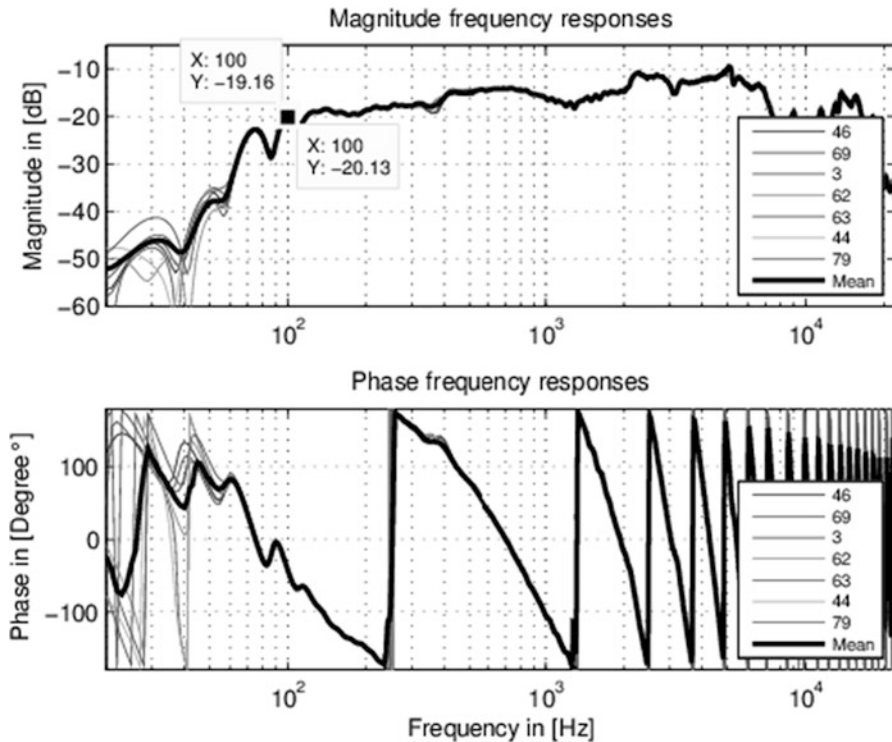
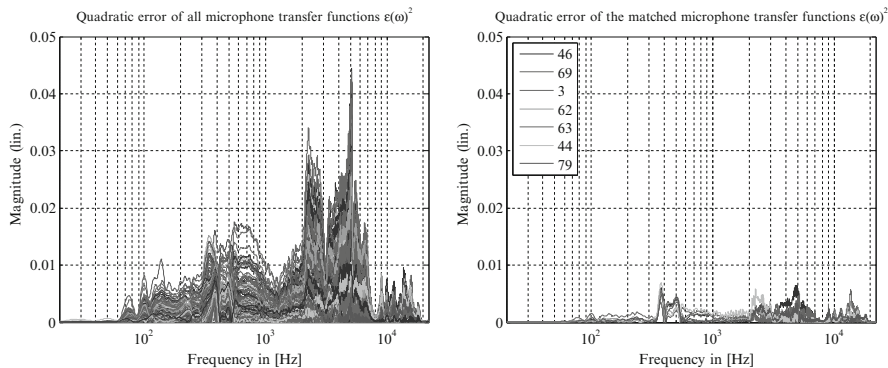**Fig. 8.5** Bode diagram of the 7 best matching microphone capsules

Fig. 8.6 Quadratic error $\varepsilon(\omega)^2$ of all microphones (*left* figure) and of the 7 best matching microphones (*right* figure)

Due to the fact, that the beamformer ought be used in an auralization method, it should ideally be able to work throughout the whole audio frequency range: $f \approx [20, \ldots, 20000]$ [Hz]. With only one compact microphone array, this task cannot be accomplished. The best probable compromise for this purpose, had been found by utilizing a superdirective beamformer, ideally showing a frequency-independent directivity pattern up to the *spatial aliasing frequency*, which calculates to:

$$f \leq \frac{c}{2d} \tag{8.20}$$

As can be seen in (8.20), the spatial aliasing frequency solely depends on the inter-microphone spacing d, thus the dimension of the microphone array should be kept small to enlarge the frequency range of operation. Since the utilized capsules already have a diameter of $\emptyset = 6$[mm] and the fact that the frame cannot be made too small, to ensure a minimum of mechanical robustness, a inter-microphone distance of $d = 1.25$[cm] has been applied for the microphone array, leading to a spatial aliasing frequency of $f = 13600$[Hz] for an array in endfire orientation, which can be considered as sufficient for our purpose. In order to let the beam point in any room direction, a 3D[2] arrangement of the microphones was mandatory. For that reason, three linear microphone arrays, each consisting of three microphones, were arranged along the *X*, *Y*, and *Z* axes, each sharing the center microphone, resulting in an array with 7 microphones. Depending on the direction where the beam shall point at, a beamformer for each of the three linear arrays will be calculated, either in endfire or broadside orientation, resulting in the desired beamformer by combination of the three individual beamformers. Hence, considering (8.3) and (8.5), the final superdirective beamformer calculates to:

$$\mathbf{A}_{\text{Out}}(\omega) = \frac{1}{3}\left(\mathbf{A}_x^T(\omega)\text{diag}\{\mathbf{d}_x(\omega)\} + \mathbf{A}_y^T(\omega)\text{diag}\{\mathbf{d}_y(\omega)\} + \mathbf{A}_z^T(\omega)\text{diag}\{\mathbf{d}_z(\omega)\}\right) \tag{8.21}$$
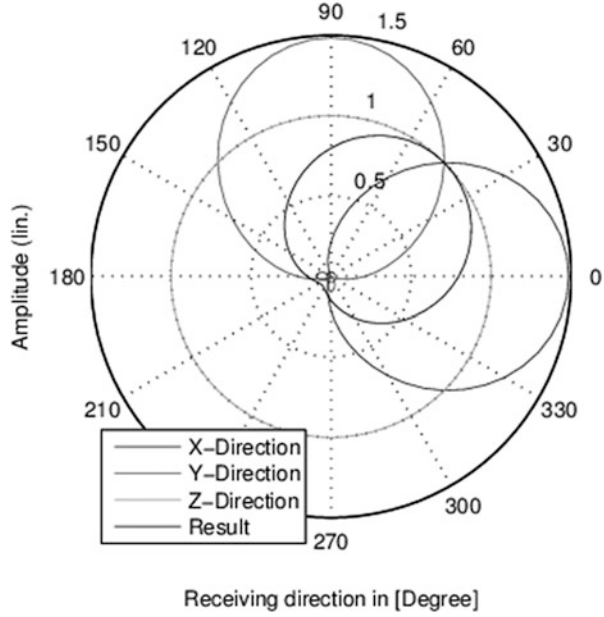
where
diag$\{\mathbf{X}\}$=Diagonal matrix of vector $\mathbf{X}$.

Because all beams of the three linear arrays, point as close as possible, to the desired direction, the resulting beam will completely point to this location, as depicted in Fig. 8.7.

Each of the three beamformers exhibit a different aliasing pattern, which, as a matter of fact will also be combined, resulting in the positive effect, that the combined beamformer shows much less disturbing aliasing effects compared to each of the three individual beamformer, on which it is based on, as can be seen in Fig. 8.8.

---

[2] 3D = Three dimensional.

Fig. 8.7 Polar diagram along the *X/Y*-plane of the three linear microphone arrays as well as of the resulting beamformer at $f = 1$[kHz], steered to $\varphi = 0°$ and $\Theta = 45°$

Thereby the top plots of Fig. 8.8 affirm a spatial aliasing frequency of $f = 13600$ [Hz] on the one hand and reveal that only at those regions in the spatial–spectral domain where the aliasing products of the individual beamformers overlap, the resulting beamformer shows aliasing too, which as a matter of fact appears enhanced. All other regions in the spatial–spectral domain have been suppressed during the overlapping process of the individual beamformer, as indicated by (8.20), leading to a final beamformer, exhibiting a higher spatial aliasing frequency, smaller aliasing regions within the spatial–spectral domain as well as a narrower beam width, as any of the underlying beamformers. With a mean squared error of the microphone transfer functions of $\varepsilon(\omega)^2 = 1\%$, which could, as previously shown, be achieved by sorting of the microphone capsules, a fix mean squared error of the microphone placement of $\delta(\omega)^2 = 1\%$ and a maximum deviation of the directivity pattern of $\Delta\Psi_{\text{Max}}(\omega) \approx -9.5$[dB], for $M = 3$, a maximum susceptibility of $K_{\text{Max}}(\omega) \approx 16.75$ results, leading to a frequency from which on the beamformer can be considered as superdirective, of $f \approx 150$[Hz], which is regarded as acceptable, since small rooms, such as the interiors of automobiles, behave more like a pressure chamber, leading to a distinct modal acoustical behavior up to a certain frequency. This transition frequency, known as *Schröder frequency*, given as:
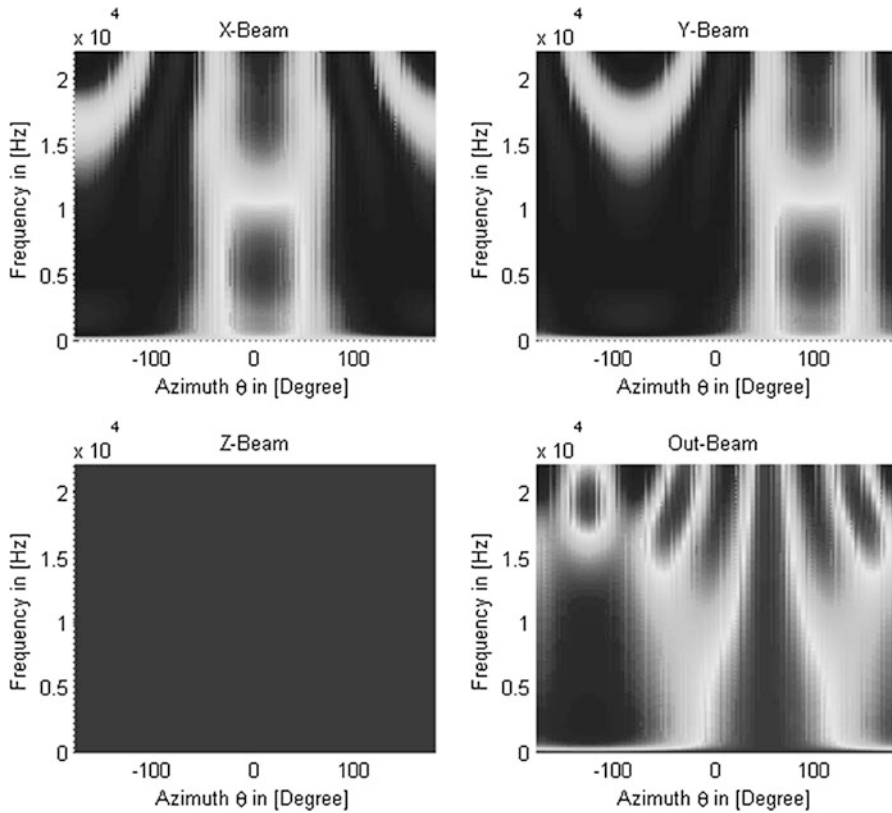
$$f_t = 2000\sqrt{\frac{T}{V}} \tag{8.22}$$

Fig. 8.8 Top view of the *X*-, *Y*-, and *Z*- as well as of the resulting, superdirective beamformer at $\varphi = 0°$ and $\Theta = 45°$

with

f_t=Schröder, or transition frequency in [Hz],

T=Reverberation time [s] (usually $T = T_{60}$),

V=Volume in the enclosure in [m³],

calculates for a typical medium-class car environment, with $V \approx 3.5$[m³] and $T_{60} \approx 0.08$[s] to:

$$f_t = 300[\text{Hz}] \qquad (8.23)$$

which is about twice the number of the previously determined, lower frequency bound of our final beamformer for superdirectivity. Hence it can be concluded that the novel, 3D microphone array, presents an adequate measuring device for broadband acoustical recordings.

8.5   Measurements

Measurements conducted in an anechoic chamber where used to verify the theory as described in the preceding chapters. Thereby, impulse responses from a broadband speaker with a membrane diameter of $\varnothing = 10$[cm], located 1[m] away from the center of the microphone array, to all 7 microphones had been gathered in the horizontal plane ($\varphi = 0°$) in 90° steps, i.e., for $\Theta = [0°, 90°, 180°, 270°]$, utilizing the exponential sine sweep technique as disclosed in [11].

In the first row of Fig. 8.9 one can see the behavior of the X beamformer, i.e., of a superdirective Beamformer in endfire direction, measured in four different orientations. At $\Theta = 0°$ its response should ideally follow the response of an omnidirectional microphone, represented by the reference microphone, located in the center of the microphone array, denoted as "RefMic" in Fig. 8.9, whereas at $\Theta = 180°$ the least amount of signal energy will be picked up. The second row shows the results of the Y beamformer, which forms a superdirective beamformer in broadside direction, measured at the same four orientations. Here one would expect equal responses with a maximum gain at $\Theta = [0°, 180°]$ and a minimum gain at $\Theta = [90°, 270°]$, with a gap, slowly increasing with frequency, which, in reality, is indeed the case.

The Z beamformer also shows this characteristics but along the vertical axis. Along the horizontal axis, ideally no deviation should occur, which again holds true, as the measurements show. In the last row the behavior of the novel beamformer, resulting out of the overlap of the X, Y, and Z beamformer is shown. In principle it exhibits a similar behavior to the X beamformer, but with much less directivity in the low- and mid-frequency region, relativizing the practicability of the new beamforming technique.

8.6   Conclusions

With the novel beamforming structure a beam, pointing at any desired position in a room, can easily be formed. This can solely be accomplished via software, i.e., by utilization of different beamforming filter. Following the preluding example, Fig. 8.10 shows the result of the novel beamforming technique.

Unfortunately the final beam shows a directivity factor, which is, compared to a superdirectional beamformer, directly pointing to a desired direction, inferior, which is true, especially at low and mid-frequencies. Doubtless, there will be applications for the novel beamforming structure, but for auralization purposes it appears logical to use robustly designed superdirectional beamformer for each speaker, as located in the target room, instead. In our example with 8 speakers,
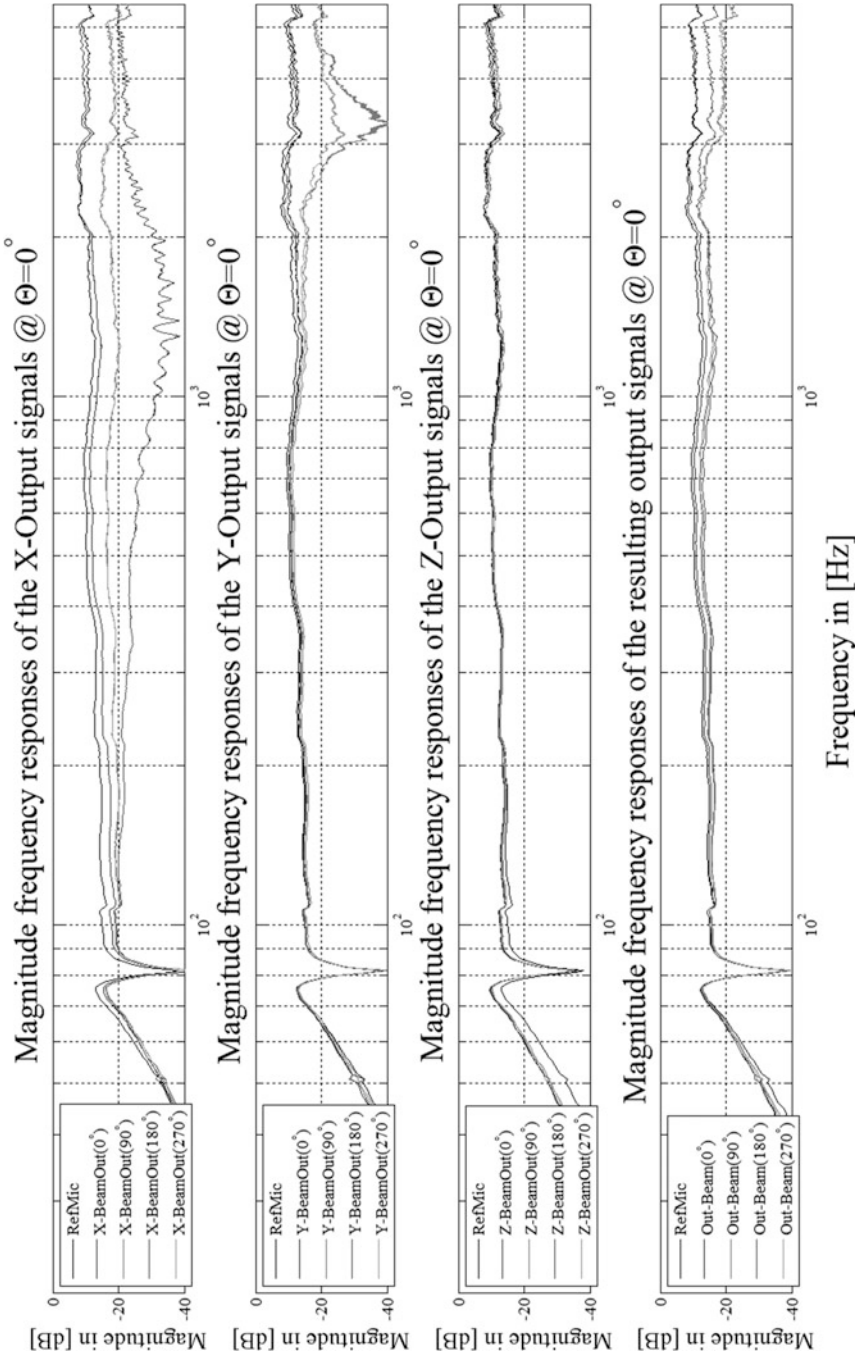
Fig. 8.9 Magnitude frequency responses of the X-, Y-, and Z- as well as of the resulting, beamformer at $\varphi = 0°$ and $\Theta = 0°$
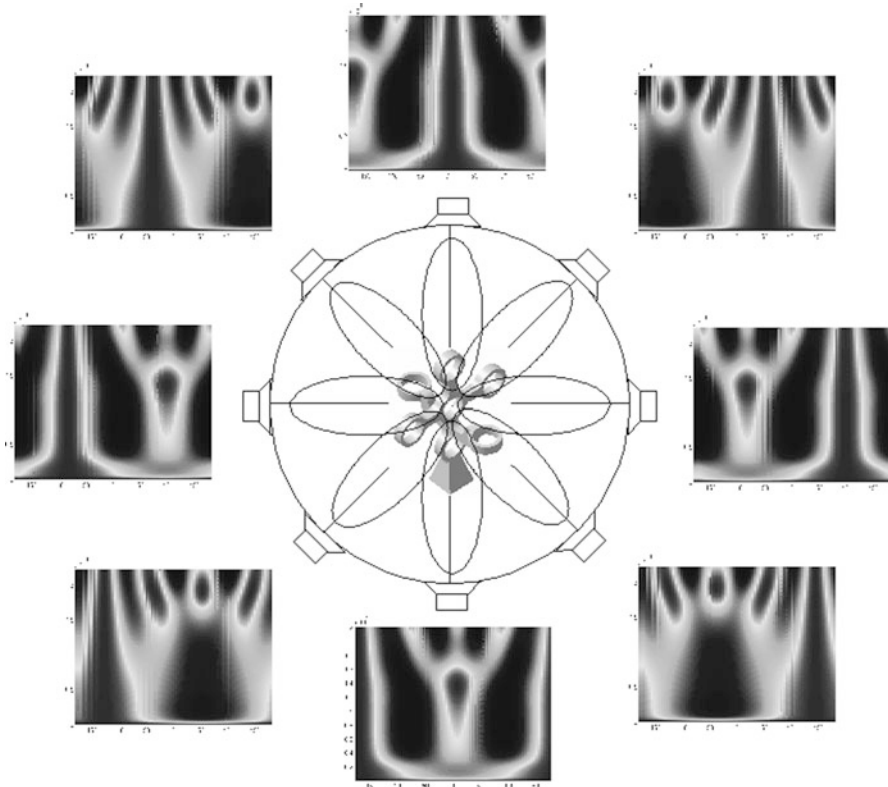
Fig. 8.10 Result of the novel beamforming technique, following the preluding example, i.e., beams pointing at $\varphi = 0°$ and $\Theta = [0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°]$

regularly arranged in the target room, this would mean to measure the horizontal plane in the source room twice—one time with the X beam pointing at $\Theta = 0°$ and the other time oriented at $\Theta = 45°$. Applicability of this method, especially regarding the introduced "Room in a Room" concept, remains a task for the future.

References

1. D. Griesinger, Binaural techniques for music reproduction, *8th International Converence of the AES*, May 1990
2. P. Mackensen, U. Felderhoff, G. Theile, U. Horbach, R. Pellegrini, Binaural room scanning—a new tool for acoustic and psychoacoustic research. *DAGA*. (1999)
3. J. Garas, *Adaptive 3D sound systems* (Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000)
4. B. Wiggins, An investigation into the real-time manipulation and control of three-dimensional sound fields, *Doctoral Thesis*, University of Derby, Derby, UK, 2004

5. J. Vilkamo, T. Lokki, V. Pulkki, Directional audio coding: virtual microphone-based synthesis and subjective evaluation. J. Audio Eng. Soc. **57**(9), 709–724 (2009)
6. S. Spors, H. Buchner, R. Rabenstein, W. Humboldt, Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering. J. Acoust. Soc. Am. **122**(1), 354–369 (2007)
7. J. Daniel, R. Nicol, S. Moreau, Further investigations of high order Ambisonics and wave field synthesis for holophonic sound imaging, *114th AES Convention*, Springer, Amsterdam, The Netherlands, Mar 2003
8. E. Gilbert, S. Morgan, Optimum design of directive antenna arrays subject to random variations. *Bell Syst. Tech. J.* 637–663 (1955)
9. J. Bitzer, K. D. Kammeyer, K. U. Simmer. An alternative implementation of the superdirective beamformer, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, Oct 1999
10. M. Dörbecker, Mehrkanalige Signalverarbeitung zur Verbesserung akustisch gestörter Sprachsignale am Beispiel elektronischer Hörhilfen, *Doctoral Thesis*, IND, RWTH Aachen, No. 10, ed. by P Vary, Aachener Beiträge zu Digitalen Nachrichtensystemen (ABDN), Mainz in Aachen, (1998)
11. A. Farina, Simultaneous measurement of impulse response and distortion with a swept-sine technique, *110th AES Convention*, Paris, France, Feb 2000