AMBISONICS SYMPOSIUM 2009
June 25-27, Graz

APPLICATION OF MVDR BEAMFORMING TO SPHERICAL ARRAYS

Anton Schlesinger[1], Marinus M. Boone[2]

[1] University of Technology Delft, The Netherlands (a.schlesinger@tudelft.nl)
[2] University of Technology Delft, The Netherlands (m.m.boone@tudelft.nl)

Abstract: *The minimum variance distortionless response (MVDR) beamforming technique is applied to a spherical microphone array. Therewith optimal spatial filters are calculated. Besides, a freely chosen measure of stability in the calculation facilitates a tradeoff between directivity and noise sensitivity. The beamforming method is compared to the highly directive method of phase-mode processing and the highly robust method of delay-and-sum beamforming. Using MVDR beamfoming, a stable and practical implementation of optimal spatial filters is found in between theses extremes. This finding suggests the consideration of MVDR beamforming in the analysis of sound fields and speech enhancement when applying spherical arrays.*

Key words: beamforming, minimum variance distortionless response beamforming, phase mode processing, delay-and-sum beamforming, sound field analysis and spherical arrays

1 INTRODUCTION

Spherical arrays are today's answer to a 3D analysis of the sound field. In addition to the volumetric reproduction of the sound field quantities in the interior of a sampled spherical surface, the investigation of the spatial density of wave amplitudes, which impinge onto the spherical surface, augments the room acoustical analysis and beamforming enables the remote sound pickup in adverse acoustical situations.

The spherical geometry of the array is accompanied by the potential virtue of an orthonormal decomposition of the sound field into spherical harmonics. The term phase-mode processing is widely used for the spatial, here spherical, Fourier transforms. Based on a spherical Fourier transform of the sound field, the spatial analysis of wave amplitudes is performed by a plane wave decomposition. A high and frequency independent spatial resolution is achieved. This advantage is however limited in practical applications by a high sensitivity to internal microphone noise, positioning errors as well as phase errors of the transducers and the signal acquisition chain.

Since the delay-and-sum beamformer adds correlated signals in phase, uncorrelated random errors vanish the more transducers are employed. The delay-and-sum technique is therefore a robust beamforming technique. However it does not establish sufficient gain, as long as the dimension of the array is small in comparison with the wave length. Many decades ago this drawback motivated the derivation of superdirective arrays, the MVDR beamformers, that allow for a broad-band gain at the expense of a high noise sensitivity. As a result, the calculation of optimal filters was extended by an extra constraint on robustness. Consequently, it became possible to trade directivity for sensitivity and vice versa.

Rafaely [1] compared the two opposite extremes of phase-mode processing and delay-and-sum beamforming. In order to realize a balance between the two desirable but incompatible features, Meyer and Elko [2] introduced MVDR processing to the weighting of spherical harmonics. Based on a combination of phase-mode and MVDR processing, Meyer and Elko [2] built a flexible spherical beamformer, with alterable directivity and noise sensitivty.

In this contribution, we pursue a similar trade-off by applying MVDR processing directly to the array output. In this manner, more degrees of freedom are maintained in the formulation of optimal filters and less computational complexity is required.

First, the key points of the three presented approaches of directional sound field analysis are outlined and in a second part, these are contrasted in a simulated application to spherical arrays.

2 ALGORITHMS

The array performance is often assessed with the directivity index (DI) and the robustness with the white noise gain (WNG). The DI is a dB measure of the directivty factor $Q(k)$ and k is the wave number. The directivity factor is described as the ratio of the maximum squared array response $|\Gamma(\theta,\phi,k)|^2$ with respect to the angles θ,ϕ to the averaged squared array response $|\Gamma(\theta,\phi,k)|^2$ due to diffuse sound incidence from all directions. For discrete arrays, the array response can be written as the inner product of the frequency dependent weights or filter vector $\mathbf{F}(k)$ and the propagation delay vector $\mathbf{W}(\theta,\phi,k)$:

$$\Gamma(\theta,\phi,k) = \mathbf{F}^{\mathrm{T}}(\omega)\mathbf{W}(\theta,\phi,k), \qquad (1)$$

where $(\circ)^{\mathrm{T}}$ is the transpose. Therewith the directivity factor becomes:

$$Q(k) = \frac{\max_{\theta,\phi}\{\mathbf{F}^{\mathrm{H}}(k)\mathbf{W}^{*}(\theta,\phi,k)\mathbf{W}^{\mathrm{T}}(\theta,\phi,k)\mathbf{F}(k)\}}{\mathbf{F}^{\mathrm{H}}(k)\mathbf{S}^{\mathrm{T}}(k)\mathbf{F}(k)}, \tag{2}$$

where $(\circ)^{*}$ represents the complex conjugate and $(\circ)^{\mathrm{H}}$ represents the Hermitian. In (2) $\mathbf{S}(k)$ is the cross-spectral density matrix. Under the assumption of a diffuse sound field and discrete sampling, $\mathbf{S}(k)$ is equivalent to:

$$S_{mn} = si(k(M_m - M_{m'})), \tag{3}$$

and M_m is the location of the m-th microphone. The DI is obtained as:

$$DI(k) = 10\log_{10}(Q(k)). \tag{4}$$

To simplify the calculation of the WNG, the array steering direction and the propagation direction of a plane wave are confined to the z-axis. The propagation delay vector $\mathbf{W}(\theta,\phi,k)$ becomes:

$$\mathbf{W}(0,0,k) = [e^{jkz_1} \dots e^{jkz_O}]^{\mathrm{T}}, \tag{5}$$

where O is the number of microphones. To calculate the robustness of the array with the WNG, the array target response is referred to the mutually uncorrelated noise sources of the sensors self noise and here expressed in dB:

$$WNG(k) = 10\log_{10}\left(\frac{\mathbf{F}^{\mathrm{H}}(k)\mathbf{W}^{*}(k)\mathbf{W}^{\mathrm{T}}(k)\mathbf{F}(k)}{\mathbf{F}^{\mathrm{H}}(k)\mathbf{F}(k)}\right). \tag{6}$$

2.1. Delay-And-Sum Beamforming

The filters of a delay-and-sum method are calculated with [3]:

$$\mathbf{F}(k) = \frac{\mathbf{W}^{*}(k)}{\mathbf{W}^{\mathrm{T}}(k)\mathbf{W}(k)}. \tag{7}$$

The resulting filters are:

$$\mathbf{F}(k) = \frac{1}{O}[1\ e^{-jk(z_2-z_1)} \dots e^{-jk(z_O-z_1)}]^{\mathrm{T}}. \tag{8}$$

The benefit of the delay-and-sum beamforming is the maximization of the WNG. As can be seen, the delay-and-sum processing improves the SNR (related to internal noise sources) of the output signal by $10\log_{10}(O)$ with respect to the SNR of one sensor.

2.2. Phase-Mode Processing

Since the purpose of this contribution is simply to compare MVDR beamforming with the performance of phase-mode processing, a simulation of the elaborate plane wave decomposition is not included (the reader is referred to Rafaely [4]). As an alternative, an analytic specification of the operation measures is given.
To begin, the array output of an order n limited plane-wave decomposition array ($n \leq N$) can be written as [1]:

$$\Gamma(\theta,\phi) = \sum_{n=0}^{N} \frac{2n+1}{4\pi} P_n(\cos\Theta), \tag{9}$$

where P_n is the Legendre polynominal. The angle Θ defines the angle between the arriving wave (θ_0,ϕ_0) and the array look direction (θ_l,ϕ_l) [4]:

$$\cos\Theta = \cos\theta_0\cos\theta_l + \cos(\phi_0-\phi_l)\sin\theta_0\sin\theta_l. \tag{10}$$

For a spherical array steering in the direction of an arriving plane wave with O microphones and a maximal order N, where $O \geq (N+1)^2$, the WNG is calculated with [1]:

$$WNG(kr) = 10\log_{10}\left(\frac{O}{(4\pi)^2}\frac{\left|\sum_{n=0}^{N}(2n+1)\right|^2}{\sum_{n=0}^{N}\frac{1}{|b_n(kr)|^2}(2n+1)}\right), \tag{11}$$

where b_n is the modal coefficient and for an open sphere defined as:

$$b_n(kr) = 4\pi i^n j_n(kr). \tag{12}$$

In (12), r is the radius of the array, $i = \sqrt{-1}$ and j_n is a spherical Bessel function of first kind. In case of $(\theta_l,\phi_l) = (0,0)$ target direction, i.e., the z-axis, the DI can be written as:

$$DI = 10\log_{10}\left(\frac{\left|\sum_{n=0}^{N}(2n+1)\right|^2}{\sum_{n=0}^{N}(2n+1)}\right). \tag{13}$$

As apparent from these formulas, the array output Γ as well as the DI are frequency independent (apart from aliasing, which is not considered here), while the WNG is a function of kr.

2.3. MVDR Beamforming

The optimal filters of the MVDR beamformer result from the optimization task to minimize the total output power of the array (i.e., the signal variance, assuming zero mean) while preserving unity gain (the distortionless response) in the look direction. The optimal filters are given by [5]:

$$\mathbf{F}^{\mathrm{T}}(k) = \frac{\mathbf{W}^{\mathrm{H}}(k)\mathbf{S}^{-1}(k)}{\mathbf{W}^{\mathrm{H}}(k)\mathbf{S}^{-1}(k)\mathbf{W}(k)}, \tag{14}$$

where $(\circ)^{-1}$ is the matrix inverse. When these filters are applied to arrays with a microphone spacing of $\Delta d \leq \lambda/2$ a superdirective beam is established. However, superdirective arrays show a large noise sensitivity. Therefore the optimization is extended by a second constraint on the WNG. Using the stability factor $\beta(k)$, the optimal filter can be rephrased as [5]:

$$\mathbf{F}^{\mathrm{T}}(k) = \frac{\mathbf{W}^{\mathrm{H}}(k)(\mathbf{S}^{-1}(k)+\beta(k)\mathbf{I})^{-1}}{\mathbf{W}^{\mathrm{H}}(k)(\mathbf{S}^{-1}(k)+\beta(k)\mathbf{I})^{-1}\mathbf{W}(k)}, \tag{15}$$

in here $\mathbf{I}$ is the identity matrix. With $\beta(k)\mathbf{I}$ a relative level of sensor noise is injected in each microphone. This uncorrelated noise adds to the isotropic noise of the cross-spectral density matrix $\mathbf{S}$ and the robustness of the optimal filter can be varied.

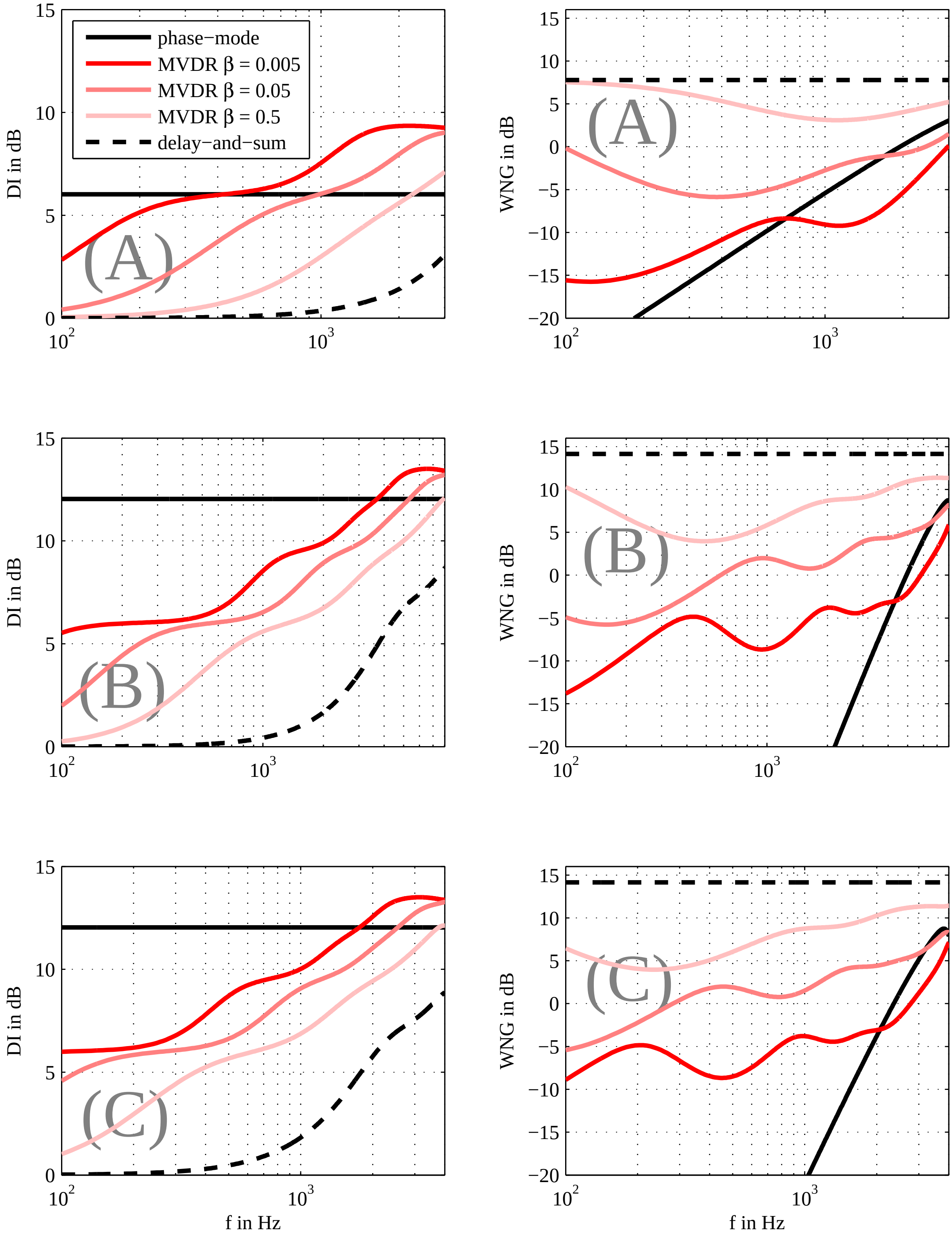

Figure 1: (A) shows the performance with the directivity index (DI) and the white noise gain (WNG) of the presented beamforming techniques applied to spherical array of $O = 6$ microphones and $r = 2$ cm radius (for the phase-mode processing the order is $N = 1$), (B) $O = 26$, $r = 2$ cm, $N = 3$, (C) $O = 26$, $r = 4$ cm, $N = 3$

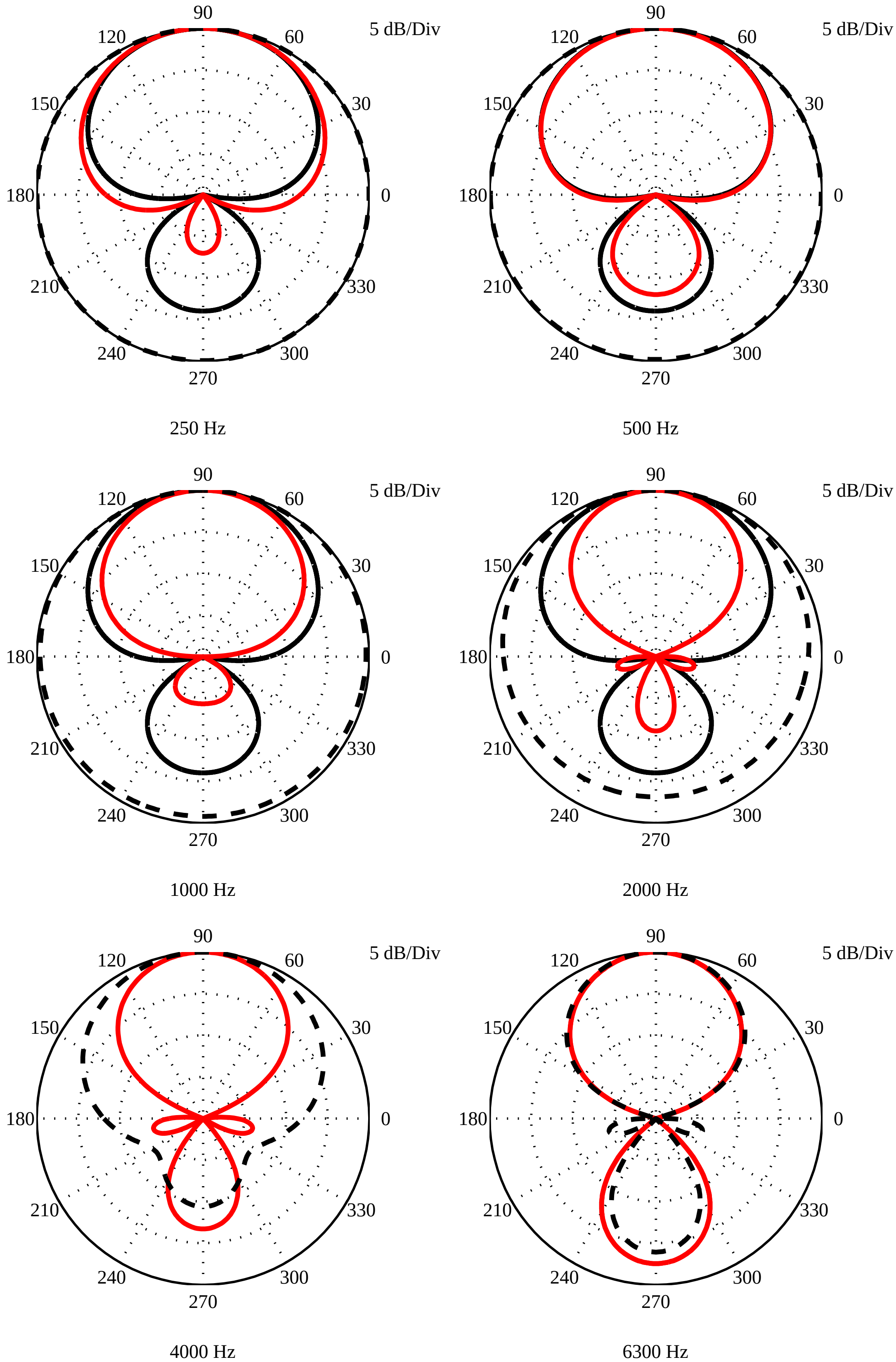

Figure 2: Directional response of the delay-and-sum beamformer (black), the MVDR beamformer (red) and the plane wave decomposition array (–, black) for $r = 2$ cm, $\beta = 0.005$, $O = 6$ and $N = 1$. Note that the directivity of the plane wave decompostion array is calculated analytically. As the spatial Nyquist frequency of the phase-mode array is at approximately 3 kHz, the phase-mode directivity is not included in the plots of 4 kHz and 6.3 kHz.

3 SIMULATION

To illustrate the performance of the MVDR beamformer in comparison with the delay-and-sum beamformer and the plane-wave decomposition arrays, three different spherical array configurations were analyzed (named A, B and C). An efficient spatial sampling scheme on the sphere, "Lebedev-Grids" [6], was chosen to enable a fair comparability between the beamforming methods and the phase mode processing. In all three simulations, no minimal WNG was determined in the optimization of the optimal MVDR filters. Instead, the optimal filters in (15) were calculated with a fixed set of β.

In situation (A) an open sphere with a radius of $r = 2$ cm and a first order sampling grid of the phase-mode processing (with Lebedev 6 quadrature points) is analyzed. Figure 1, (A) shows the resulting DI and the WNG for the delay-and-sum beamforming, the MVDR beamforming (with different stabilization factors β) and the plane wave decomposition array. The spatial Nyquist frequency is determined by the phase-mode processing with approximately 3 kHz. As expected, the delay-and-sum beamformer establishes hardly any directivity below 1 kHz and shows a constant WNG at about 8 dB. The phase-mode processing features a uniform directivity of 6 dB at the expense of amplifying noise at low frequencies, which is manifested in a low WNG. The MVDR beamformer reaches a compromise between DI and WNG depending on the stabilization factor β. When β increases, the optimal filters consequently converge to delay-and-sum filters. Moreover, the MVDR beamformer exposes a much higher gain at higher frequencies due to the fact that the analytic formulation of the phase-mode processing implies an order-limited signal to prevent spatial aliasing. In terms of DI and WNG, this behavior shows a feasible advantage of the MVDR beamforming over phase-mode processing when using spherical arrays that allow only for lower order phase-mode processing (see also Figure 2 for a comparison of the directivity plots).

When a phase-mode processing of third order ($N = 3$) is employed, the slope of the WNG approaches 6 N dB per octave [2]. A suitable array with $O = 26$, $r = 2$ cm and a spatial Nyquist frequency of 8 kHz is analyzed in Figures 1, (B). Again, a high and constant directivity of the phase-mode processing is observed with DI = $20\log_{10}(N+1) \approx 12$ dB. For MVDR beamformers this DI is again exceeded for high frequencies. Note, for superdirective end-fire arrays the maximum DI is even approximated with $20\log_{10}(O)$. As can be seen, the delay-and-sum beamformer behaves similar to the spatial sampling with $O = 6$ in (A). In case the array size is stretched with respect to the wave length, the directivity is increased. This is found in Figure 1, (C), where the radius was changed to 4 cm. At the same time, the spatial Nyquist frequency of the phase-mode processing was lowered to 4 kHz. The DI of the phase-mode processing remains at 12 dB while the WNG reaches the maximum one octave earlier. The same effect is observed with the MVDR beamformer. Since the resolution of the spatial sampling is reduced by two upon the surface, the slopes of DI and WNG reach their maxima an octave earlier.

4 CONCLUSION

In this contribution three techniques of directional sound field analysis were investigated by means of spherical arrays. The presented techniques were phase-mode processing, delay-and-sum and MVDR beamforming. As a consequence of today's computational power, microphone arrays are often used with phase-mode processing. Phase-mode processing is flexible throughout the analysis and offers high spatial resolution as compared with classical delay-and-sum beamforming. This holds also for the application of phase-mode processing to spherical arrays. However, the robustness is poor, especially at $kx \ll N$. There are several ways to circumvent this drawback and to build more robust phase-mode arrays. This is possible by e.g., measuring the sound field at different radii [1] or by employing only optimal modes [2]. In here, it has been shown to negotiate a fair compromise between DI and WNG for spherical arrays by way of MVDR beamforming. The MVDR beamformer sets the limit of directivity at fixed grids without stabilization. When applying a stabilization, it has been shown that the MVDR beamforming can maintain this DI improvement over the phase-mode processing at reasonable noise sensitivity. Although MVDR beamforming lacks the potential to detach the 3D controlled beamforming from the sampling grid, as it is possible for the spherical coding of the phase-mode processing, it provides flexibility to establish a highly directional beamformer at a predefined robustness. Finally, fixed MVDR optimized filters are easily applicable in practical applications where high directionality and robustness are needed.

5 ACKNOWLEDGMENTS

REFERENCES

[1] B. Rafaely. Phase-mode versus delay-and-sum spherical microphone array processing. *IEEE Signal Processing Letters*, 12(10), October 2005.

[2] J. Meyer and G.W. Elko. *Spherical Microphone Arrays for 3D Sound Recording in Audio Signal Processing for Next-Generation Multimedia Communication*. Springer United States, engineering edition, 2004.

[3] Ivo Merks. *Binaural Application of Microphone Arrays for Improved Speech Intelligibility in a Noisy Environment*. PhD thesis, Technische Universiteit Delft, 2000.

[4] B. Rafaely. Plane-wave decomposition of the sound field on a sphere by spherical convolution. *J. Acoust. Soc. Am.*, 116(4):2149 – 2157, October 2004.

[5] H. Cox, R.M. Zeskind, and T. Kooij. Practical supergain. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(3), June 1986.

[6] C.v. Wuellen and D. Laikov. A fortran code to generate lebedev grids up to level l = 131. *Computational Chemistry List, ltd.*, 2000.