SPATIAL FILTERING USING DIRECTIONAL AUDIO CODING PARAMETERS

Markus Kallinger, Giovanni Del Galdo, Fabian Kuech, Dirk Mahne, Richard Schultz-Amling

Fraunhofer Institute for Integrated Circuits IIS, Am Wolfsmantel 33, 91058 Erlangen, Germany

markus.kallinger@iis.fraunhofer.de

ABSTRACT

Directional audio coding (DirAC) is a recent method for spatial audio processing, based on a perceptually motivated representation of spatial sound. Due to its efficiency, DirAC has already been proposed for spatial audio teleconferencing scenarios. Modern hands-free communication systems usually include beamforming techniques to improve speech intelligibility by suppressing diffuse background noise and interfering sources. In this paper, we propose a novel spatial filtering method which can be integrated into the DirAC spatial codec. It uses a spectral weighting of the recorded audio signal, where the design of the corresponding spatial filter transfer function is based on the DirAC parameters, i. e., direction-of-arrival and diffuseness of the sound field. Simulation results show that compared to a standard beamformer the novel technique offers significantly higher interference attenuation, while introducing similar distortion of the desired signal.

Index Terms— Spatial filters, beamforming, spatial audio coding

1. INTRODUCTION

Directional Audio Coding (DirAC) represents an efficient technique to capture and reproduce spatial sound. In the analysis part, the spatial sound is expressed by a DirAC stream, comprising an omnidirectional microphone pressure signal together with direction-of-arrival (DOA) and diffuseness of the sound field expressed in time-frequency domain [1]. This parametric representation captures all information that is relevant for human perception of spatial sound. On the reproduction side, the loudspeaker signals are determined based on the DirAC stream and the specific loudspeaker configuration used for playback. As discussed in [1], the efficient representation of spatial sound makes DirAC especially suitable for teleconferencing applications. Spatial audio not only provides a realistic sound perception but also improves intelligibility of speech [2].

In hands-free speech communication, diffuse background noise and interfering sources impede speech intelligibility and quality, making conversation more exhausting. Standard approaches to this problem apply beamformers with fixed directional filtering characteristics and steerable look-direction [3]. These approaches are known to minimize the distortion of the desired signal, but, on the other hand, usually fail to provide sufficient interference suppression. To further increase the attenuation of interferences and diffuse noise, spectral weighting by post-filters is commonly employed [4].

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. ICT-2007-214793.

The authors gratefully acknowledge Ville Pulkki and his colleagues from the Laboratory of Acoustics and Signal Processing, Helsinki University of Technology, TKK, for the helpful comments and discussions on DirAC.

In this contribution we propose a novel method to suppress diffuse noise and interfering sources which is based on a direct modification of the DirAC stream. Similar to the aforementioned post-filters, the method applies a spectral weighting to the recorded audio signal. This spatial filtering transfer function is determined using the estimated DirAC parameters, i. e., the DOA of sound and the diffuseness of the observed sound field. It is important to note that this speech enhancement method does not affect the spatial distribution of sound sources after the DirAC synthesis at the reproduction side and, thus, preserves the advantages of spatial audio communication.

The remainder of the paper is organized as follows: the novel filtering technique is described in Section 2. Since the optimization criterion is focused on keeping speech distortion at a minimum, we compare the proposed method with a standard minimum variance distortionless response (MVDR) beamformer [3] in Section 3. Section 4 concludes the paper.

Notation Vectors are printed in boldface. The superscripts ^{T} and * denote transposition and complex conjugation, respectively. The operator $\|\cdot\|$ returns the ℓ_2-norm.

2. A NOVEL SPATIAL FILTERING METHOD

An overview of the proposed spatial filtering structure is depicted in Fig. 1: the basic Directional Audio Coding (DirAC) parameters, namely the omnidirectional signal $W(k,\omega)$, the direction, i. e., the azimuth angle $\phi(k,\omega)$, and the diffuseness $\Psi(k,\omega)$ serve as an input to the spatial filtering signal processing block (depicted as a gray box). Note that we use a notation in the discrete-time short-time Fourier transform (STFT) domain with a temporal block index k and the angular frequency ω. Even though DirAC enables spatial coding of three-dimensional sound fields, we only consider the two-dimensional case, here. The subscript 'spatial filtering', or in short 'sf', denotes signals which have been spatially filtered. We obtain $W_{\mathrm{sf}}(k,\omega)$ and $\Psi_{\mathrm{sf}}(k,\omega)$ at the output, respectively. Direction remains unchanged. The block denoted as 'magnitude trans-

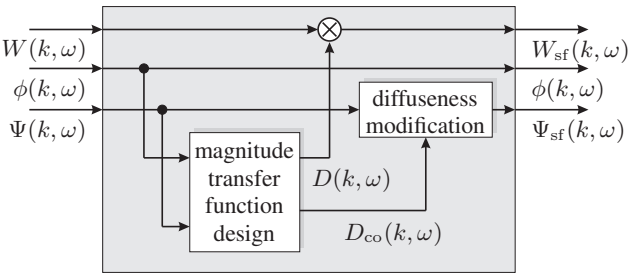

Fig. 1. Block diagram of proposed spatial filtering structure. The technique works on the parametric signal representation of DirAC.

fer function design' makes use of direction and diffuseness. Details about this block and the calculation of the spatial filtering magnitude transfer functions $D(k,\omega)$ and $D_{\mathrm{co}}(k,\omega)$ are treated in Section 2.1. Spatial filtering involves direction- and frequency-dependent attenuation of the recorded signal. If this signal was spatially rendered and analyzed again, it would generate a different diffuseness than an unfiltered signal. Processing inside the block 'diffuseness modification' is described in Section 2.2.

2.1. Magnitude transfer function design

The proposed spatial filtering technique employs a short-time spectral attenuation (STSA) processing block, which is depicted as a multiplication in Fig. 1. The design of the zero-phase magnitude transfer function $D(k,\omega)$, which carries out the spatial filtering, is based on the assumption that the sound field is composed of a plane wave (direct sound) and an ideal diffuse field. Let W and $\boldsymbol{V}$ denote the pressure and particle velocity vector, respectively, and let further the subscripts 'co' and 'di' stand for *coherent*, i. e., non-diffuse and *diffuse*, respectively. Due to the linearity of the medium, we can write

$$W(k,\omega) = W_{\mathrm{co}}(k,\omega) + W_{\mathrm{di}}(k,\omega), \tag{1}$$
$$\boldsymbol{V}(k,\omega) = \boldsymbol{V}_{\mathrm{co}}(k,\omega) + \boldsymbol{V}_{\mathrm{di}}(k,\omega), \tag{2}$$

where the pressures and the particle velocity vectors of the diffuse and non diffuse (i.e., of the plane wave) part are explicitly given.

In DirAC [5], the diffuseness parameter $\Psi(k,\omega)$ is defined as

$$\Psi(k,\omega) = 1 - \frac{\|\mathrm{E}\{\boldsymbol{I}_{\mathrm{a}}(k,\omega)\}\|}{c\,\mathrm{E}\{E(k,\omega)\}}, \tag{3}$$

where the energy density $E(k,\omega)$ is defined as

$$E(k,\omega) = \frac{1}{2\rho_0 c^2}\left(|W(k,\omega)|^2 + \frac{1}{2}\|\boldsymbol{V}(k,\omega)\|^2\right), \tag{4}$$

and the active intensity $\boldsymbol{I}_{\mathrm{a}}(k,\omega)$ as

$$\boldsymbol{I}_{\mathrm{a}}(k,\omega) = \frac{1}{\sqrt{2}\,\rho_0 c}\mathrm{Re}\{W^*(k,\omega)\boldsymbol{V}(k,\omega)\}. \tag{5}$$

Given the implicit statistical independence of the coherent and the diffuse parts it is possible to rewrite (3) as

$$\Psi(k,\omega) = 1 - \frac{W_{\mathrm{co},0}^2}{W_0^2} = 1 - \frac{W_{\mathrm{co},0}^2}{W_{\mathrm{co},0}^2 + W_{\mathrm{di},0}^2}. \tag{6}$$

where the terms W_0, $W_{\mathrm{co},0}$, and $W_{\mathrm{di},0}$ are defined as

$$W_0^2 = \mathrm{E}\{|W(k,\omega)|^2\} \tag{7}$$
$$W_{\mathrm{co},0}^2 = \mathrm{E}\{|W_{\mathrm{co}}(k,\omega)|^2\} \tag{8}$$
$$W_{\mathrm{di},0}^2 = \mathrm{E}\{|W_{\mathrm{di}}(k,\omega)|^2\}. \tag{9}$$

The coherent and diffuse portions can now be expressed with respect to the overall power spectral density (PSD) $\mathrm{E}\{|W(k,\omega)|^2\}$ as follows

$$\mathrm{E}\{|W_{\mathrm{co}}(k,\omega)|^2\} = (1-\Psi(k,\omega))\cdot\mathrm{E}\{|W(k,\omega)|^2\}, \tag{10}$$
$$\mathrm{E}\{|W_{\mathrm{di}}(k,\omega)|^2\} = \Psi(k,\omega)\cdot\mathrm{E}\{|W(k,\omega)|^2\}. \tag{11}$$

The idea behind the filtering approach proposed in this contribution is to apply a different filter to each of the two parts just mentioned. Let these two filters be denoted by $D_{\mathrm{co}}(k,\omega)$ and $D_{\mathrm{di}}(k,\omega)$. The overall filtered PSD

$$\mathrm{E}\{|W_{\mathrm{sf}}|^2\} = D_{\mathrm{co}}^2\,\mathrm{E}\{|W_{\mathrm{co}}|^2\} + D_{\mathrm{di}}^2\,\mathrm{E}\{|W_{\mathrm{di}}|^2\}$$

can be expressed using the coherend and diffuse portions defined above as

$$\mathrm{E}\{|W_{\mathrm{sf}}|^2\} = \left((1-\Psi)\,D_{\mathrm{co}}^2 + \Psi\,D_{\mathrm{di}}^2\right)\,\mathrm{E}\{|W|^2\}, \tag{12}$$

where we have omitted the dependency from (k,ω) to maintain a good overview.

$D_{\mathrm{co}}(k,\omega)$ and $D_{\mathrm{di}}(k,\omega)$ are obtained from frequency-independent *directional patterns* $D_{\mathrm{co,dp}}(\phi)$ and $D_{\mathrm{di,dp}}(\phi)$, which are functions of the DirAC parameter DOA, $\phi(k,\omega)$. Since we want to design spatial filtering with an undistorted look direction we choose

$$D_{\mathrm{co,dp}}(\phi_{\mathrm{d}}) = D_{\mathrm{di,dp}}(\phi_{\mathrm{d}}) = 1. \tag{13}$$

The direction in which the desired source is located at is denoted by ϕ_{d}. We would only need a discrete pulse as a directional pattern for the coherent part, if the direction was known perfectly. However, considering that these assumptions hardly hold in practice, we choose a directional pattern shaped like a cosine between -90° and 90°; the total width of this window is compressed from 180° to 60°. If the desired and the interfering signal were both completely diffuse, separation on the basis of spatial cues would not be possible at all. However, our observations have shown that even a sound field, which is analyzed to be completely diffuse, has a certain coherent part. Therefore, we choose a very smooth directional pattern which is steered towards the desired source for the diffuse part of the sound field. The pattern is denoted as 'sub-cardioid' and its directional pattern is described by

$$D_{\mathrm{di,dp}}(\phi) = \alpha + (1-\alpha)\cos\phi, \tag{14}$$

if $0.5 < \alpha < 1$ is set [6]. In our case, we chose $\alpha = 0.55$. Fig. 2 shows exemplary directional patterns as polar plots. The sub-cardioid is plotted with a solid line. For comparison, a cardioid is obtained at $\alpha = 0.5$ and is plotted with a dotted line in Fig. 2.

The directional patterns $D_{\mathrm{co,dp}}(\phi)$ and $D_{\mathrm{di,dp}}(\phi)$ have a similar meaning as the square root of a beam pattern [3] – they specify the attenuation of a coherent signal as a function of DOA. In [7], directional patterns are designed according to the *von Mises* probability density functions (pdfs) of the observed instantaneous directions. However, informal listening tests and experimental setups have shown that the shape of directional responses is not as important as their width.

The design procedure is summarized in Fig. 3. The boxes denoted by 'smoothing' contain temporal and spectral smoothing. These processing blocks are designed in the same way as in [1]: temporal smoothing is carried out by means of one-pole low-pass filters with an adaptive time constant. Spectral smoothing is performed by weighted averaging of frequency bins in ERB-bands.

2.2. Diffuseness modification

Spatial filtering acts as attenuation of sources from certain directions. Accordingly, we can expect a modified diffuseness after spatial filtering. Because spatial filtering is located inbetween DirAC analysis and rendering units, we need the modified diffuseness parameter for a proper balance of coherent and diffuse signal portions in the rendering unit. To determine quantitatively the correct value of the diffuseness we model the coherent portion of the sound field as a

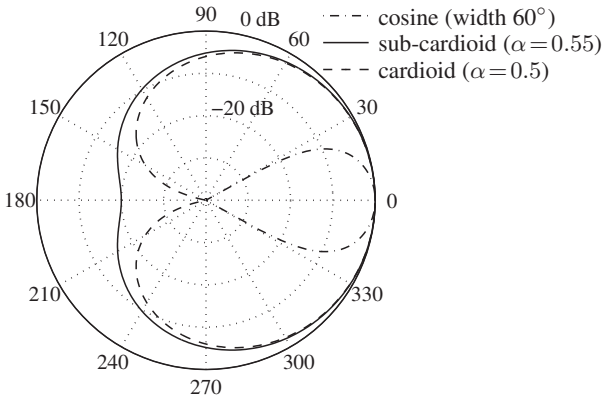

Fig. 2. Exemplary polar plots of spatial filtering directional patterns.

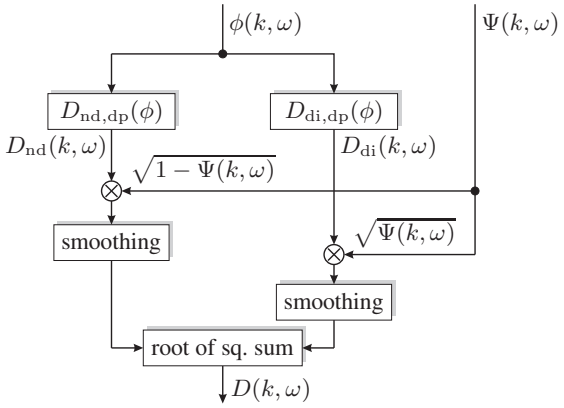

Fig. 3. Overview of the spatial filter's magnitude transfer function design.

plane wave with the amplitude $W_{\mathrm{co},0}$ and a random phase $\theta_{\mathrm{co}}[k]$. The modification of the coherent part of the sound pressure is easily expressed by

$$W_{\mathrm{sf,co}}(k,\omega) = D_{\mathrm{co}}(k,\omega)\, W_{\mathrm{co},0}\, e^{j\omega\theta_{\mathrm{co}}[k]}. \tag{15}$$

The diffuse portion is assumed to result from a sum of an infinite number of plane waves. Their DOAs are uniformly distributed and their phases are random and mutually uncorrelated as well as uncorrelated to the coherent portion's phase. This leads to a power spectrum which is spatially white and equal to $W_{\mathrm{di},0}^2$. The diffuse part is processed with $D_{\mathrm{di,dp}}(\phi)$ as desribed by equation (14) using $\alpha = 0.55$. We obtain the overall sound pressure of a filtered diffuse field as

$$\begin{aligned}\mathrm{E}\{|W_{\mathrm{sf,di}}(k,\omega)|^2\} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} D_{\mathrm{di,dp}}^2(\phi)\, W_{\mathrm{di},0}^2\ \mathrm{d}\phi \\ &= \left(\frac{3}{2}\alpha^2 - \alpha + \frac{1}{2}\right) W_{\mathrm{di},0}^2 \end{aligned} \tag{16}$$

after a straightforward calculation. The directional pattern $D_{\mathrm{di,dp}}(\phi)$ does not only have an effect on the overall sound pressure but also the mean active intensity vector. We can predict that it will not be a null vector anymore – in unprocessed diffuse sound fields the mean active intensity vector is known to equal a null vector [5]. Note that the decomposition of an intensity vector into its coherent and its diffuse portion is only possible by investigating its mean, and if both portions are mutually uncorrelated. Using $D_{\mathrm{di,dp}}(\phi)$ we obtain the mean active intensity vector of a spatially filtered diffuse sound field:

$$\mathrm{E}\{\boldsymbol{I}_{\mathrm{a,sf,di}}(k,\omega)\} = \frac{1}{2\pi}\, W_{\mathrm{di},0}^2 \int_{-\pi}^{\pi} D_{\mathrm{di,dp}}^2(\phi) \begin{bmatrix}\cos\phi \\ \sin\phi\end{bmatrix} \mathrm{d}\phi. \tag{17}$$

Exploiting the symmetries exhibited by $D_{\mathrm{di}}(\phi)$ we can expect the y-element of the mean active intensity vector to equal zero. We carry on with the x-element, which then equals the norm of the mean active intensity vector:

$$\|\mathrm{E}\{\boldsymbol{I}_{\mathrm{a,sf,di}}(k,\omega)\}\| = \left(\alpha - \alpha^2\right)\, W_{\mathrm{di},0}^2. \tag{18}$$

The diffuseness corrected with respect to the modified amplitudes can be expressed as

$$\Psi_{\mathrm{sf}}(k,\omega) = 1 - \frac{|D_{\mathrm{co}}(\phi)|^2\, W_{\mathrm{co},0}^2 + (\alpha - \alpha^2)\, W_{\mathrm{di},0}^2}{|D_{\mathrm{co}}(\phi)|^2\, W_{\mathrm{co},0}^2 + (\frac{3}{2}\alpha^2 - \alpha + \frac{1}{2})\, W_{\mathrm{di},0}^2}. \tag{19}$$

By substituting equation (6) into (19) we obtain the estimated diffuseness of the spatially filtered and newly rendered sound field by

$$\Psi_{\mathrm{sf}}(k,\omega) = \frac{\frac{5}{2}\alpha^2 - 2\alpha + \frac{1}{2}}{D_{\mathrm{co}}(\phi)^2\, (\Psi^{-1}(k,\omega) - 1) + (\frac{3}{2}\alpha^2 - \alpha + \frac{1}{2})}. \tag{20}$$

To assess the extent of diffuseness modification, Fig. 4 shows the diffuseness $\Psi_{\mathrm{sf}}(k,\omega)$ after spatial filtering as a function of the diffuseness prior to filtering, $\Psi(k,\omega)$. We show three exemplary cases of the spatial filter for the coherent part, $D_{\mathrm{co}}(\phi)$: $D_{\mathrm{co}} = 1$ occurs, if the current active intensity vector points at the desired source, which is then not attenuated. By contrast, $D_{\mathrm{co}} = 0.1$ is typical, if a coherent source is undesired and attenuated. $D_{\mathrm{co}} = 0.5$ is chosen as an example for a mixture of desired and interfering sources. We can

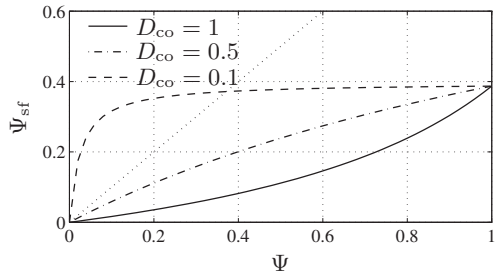

Fig. 4. Diffuseness after spatial filtering as a function of diffuseness prior to filtering. The function depends on the spatial filter for the coherent part of the processed signal, D_{co}.

observe that there is an upper limit which indicates that even a completely diffuse sound field would be spatially filtered by a certain amount and its diffuseness cannot reach one. By leaving the coherent part untouched ($D_{\mathrm{co}} = 1$) the diffuseness is clearly reduced. By attenuating the coherent part, e. g. an undesired source, we increase diffuseness (see dashed curve for $D_{\mathrm{co}} = 0.1$ in Fig. 4).

3. SIMULATION RESULTS

The spatial filtering technique described in the previous section is now compared to a conventional MVDR beamformer. We choose

a line array of four microphones in endfire steering. The spacing is 2.3 cm to achieve a spatial aliasing frequency of 7.5 kHz. Using DirAC, we need a spacing of 3.2 cm of opposing microphones in a square-shaped grid to reach the same spatial aliasing limit (for an explanation of this limit refer to [8]). The MVDR beamformer is designed for a three-dimensional diffuse noise field with an assumed acoustic signal-to-microphone self-noise ratio of 30 dB (see [3]). We simulated a room using the well-known image method with varying reverberation time. There are two speech sources with a radius of 1 m around the array's center located at 0° and 60°. The novel spatial filtering approach first processes the mixture and its time-variant transfer function is stored. Afterwards it is applied as a determined filter to the spectra of the single signals.

Fig. 5 shows the interference reduction (IR) in dB. The mean power is equally weighted in each critical band (see, e. g., [9]). Especially, at low and medium reverberation times spatial filtering clearly outperforms the MVDR beamformer by achieving higher IR.

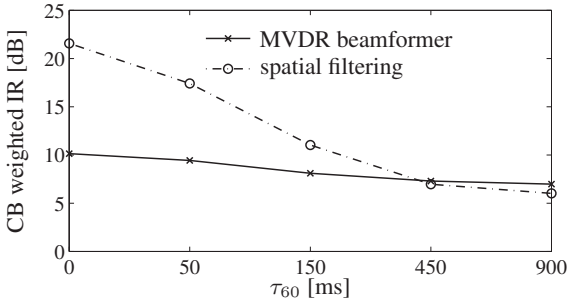

Fig. 5. Critical band weighted interference reduction as a function of reverberation time.

We investigate the log-area-ratio (LAR) [9] to make sure that spatial filtering's increased interference reduction is not obtained at the cost of speech distortion. The LAR describes the degree of speech degradation by means of linear predictive coding (LPC) coefficients; positive values in dB describe the distance of processed to clean speech signals. In Fig. 6 we present mean values, which are collected during speech activity. The dotted curve denotes the LAR between the clean and the reverberated speech signal. We can see that spatial filtering adds some more distortions than the beamformer at very low reverberation. This results from the fact that the presented values are obtained during double-talk, i. e., the time-variant filter, which attenuates portions of the interferer in the mixture of both signals, slightly distorts the desired signal, if it is applied to it separately. Informal listening tests have shown that artifacts are almost completely masked by the remaining interference in the filtered mixture.

4. CONCLUSIONS

The present work contains a proposal for a novel spatial filtering method, which works in the coded domain of the Directional Audio Coding (DirAC) technique. It is based on the separate treatment of a sound field's coherent and diffuse parts. The two appropriate spatial filtering magnitude transfer functions are weighted depending on the diffuseness parameter of DirAC. Like a standard beamformer, the novel technique introduces no audible artifacts while offering significantly increased interference reduction. The efficient rendering capabilities of DirAC make the proposed approach very promising:

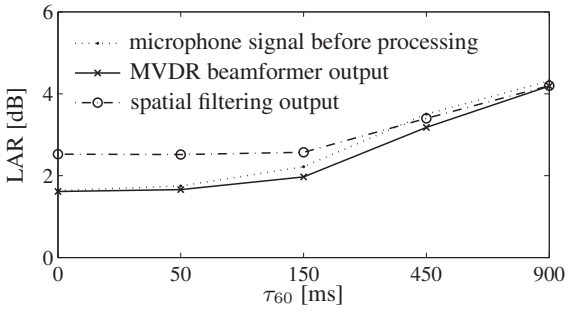

Fig. 6. This plot shows LAR enhancements as a function of reverberation time. Note that positive values indicate increased speech distortion.

intelligibility of speech and ease of communication is further improved, because interferences are not only attenuated as in standard approaches; they can be rendered to spatially separate positions, if a user selects DirAC's synthesis option for two or more loudspeakers. Intelligibility enhancement by these two aspects opens up opportunities for further promising techniques.

5. REFERENCES

[1] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *Journal of the Audio Engineering Society*, vol. 55, no. 6, pp. 503–516, June 2007.

[2] J. Peissig and B. Kollmeier, "Directivity of Binaural Noise Reduction in Spatial Multiple Noise-Source Arrangements for Normal and Impaired Listeners," *JASA*, vol. 101, pp. 1660–1670, 1997.

[3] J. Bitzer and K. U. Simmer, "Superdirective microphone arrays," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. S. Brandstein and D. Ward, Eds., chapter 2, pp. 19–38. Springer-Verlag, 2001.

[4] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. S. Brandstein and D. Ward, Eds., chapter 3, pp. 39–60. Springer-Verlag, 2001.

[5] J. Merimaa, *Analysis, Synthesis, and Perception of Spatial Sound – Binaural Localization Modeling and Multichannel Loudspeaker Reproduction*, Ph.D. thesis, Helsinki University of Technology, Aug. 2006.

[6] G. W. Elko, "Superdirectional microphone arrays," in *Acoustic Signal Processing for Telecommunication*, S. L. Gay and J. Benesty, Eds. Kluwer Academic Publishers, 2000.

[7] B. Gunel, H. Hacihabiboglu, and A. M. Kondoz, "Acoustic Source Separation of Convolutive Mixtures Based on Intensity Vector Statistics," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 16, no. 4, pp. 748–756, May 2008.

[8] M. Kallinger, F. Kuech, R. Schultz-Amling, G. Del Galdo, J. Ahonen, and V. Pulkki, "Analysis and Adjustment of Planar Microphone Arrays for Application in Directional Audio Coding," in *124th AES Convention,* Paper 7374, Amsterdam, the Netherlands, May 2008.

[9] T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Objective Perceptual Quality Measures for the Evaluation of Noise Reduction Schemes," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Eindhoven, The Netherlands, Sept. 2005.