A Method to Analyze the Spatial Response of Informed Spatial Filters

Soumitro Chakrabarty, Oliver Thiergart and Emanuël A. P. Habets

International Audio Laboratories Erlangen, 91058 Erlangen, Germany
Email: {soumitro.chakrabarty,oliver.thiergart,emanuel.habets}@audiolabs-erlangen.de
Web: www.audiolabs-erlangen.de

Abstract

Informed spatial filters (ISF) aim to capture multiple sound sources with a desired spatial response while attenuating the undesired signals. The desired spatial response is an arbitrary function, based on which directional gains at each time-frequency instant are computed. In this work, we propose a method to analyze the obtained spatial response at the output of the ISF and the influence of direction-of-arrival (DOA) estimation errors. The proposed method considers two simultaneously active sound sources, where one source is kept static at a reference position while the other source is moved across the DOA space by placing it at discrete DOA points. For each position of the moving source, we compute the average directional array gain. Through analysis with simulated speech signals, we show that with perfect knowledge of the source DOAs the obtained spatial response matches the desired one, and also demonstrate the adverse effects of DOA estimation errors.

1 Introduction

In modern hands-free communication systems, with the presence of multiple microphones, microphone array processing techniques, known as spatial filtering [1], have become a common approach to extract a desired speech signal while suppressing undesired signal components. In recent decades, a variety of spatial filtering techniques were developed [2–8].

For acoustic scenarios with directional undesired signal components such as interfering speech, different methods were developed for extracting the desired signal by computing directional gains at each time-frequency (TF) instant [9–11]. These methods use a phase-based TF masking method to compute the directional gain. However, they employ the common single plane wave signal model which is easily violated when multiple sources are simultaneously active [12].

Recently in [6–8], a spatial filter, known as the *informed* spatial filter (ISF), was proposed to capture at most L sound sources with a desired, arbitrary spatial response at each TF instant while attenuating the undesired signal components due to the acoustic environment. The underlying signal model considers L plane waves per TF instant, as well as diffuse sound and noise, which makes the possibility of model violations less likely. The arbitrary spatial response enables the employment of ISF for different applications such as source extraction or spatial sound reproduction.

In the ISF framework, the desired spatial response is used to compute the directional gain for each of the L plane waves at each TF instant by evaluating the spatial response function at the directions-of-arrival (DOAs) of the L plane waves. For quick adaptability to the changes in the acoustic scene, L DOA estimates, along with other parametric information, is computed at each TF instant. The DOA estimates are used to compute both array propagation vectors as well as the directional gains.

It should be noted that the desired spatial response function is different from the directivity pattern of the spatial filter [8]. For varying L and DOAs of the direct sound, the directivity pattern also varies. The ISF aims to provide a desired spatial response only for the L DOAs rather than resample the response function for all angles in the DOA space. Though the performance evaluation of the ISF in practical acoustic scenarios was presented in [8], a comparative analysis of the *obtained* and the desired spatial response of the ISF has not yet been performed. The main motivation behind such an analysis is to gain insight into the obtained spatial response at the output of the ISF and also provide an objective motivation for the necessity of robustness to DOA estimation errors, in the ISF framework.

Since the ISF considers a multiple plane wave signal model, to properly analyze the obtained spatial response for the ISF, it is not enough to place a sound source at sampled points of the response function and compute the gain at the output. Therefore, we propose an analysis method in this paper that considers two simultaneously active sources, with one of the sources kept static at a reference direction, while changing the position of the other source by placing it at discrete points throughout the DOA space. As an objective measure, for each discrete position of the second source, we compute the average directional array gain to get the obtained spatial response. Through analysis with simulated signals, we show that a perfect knowledge of the source DOAs is required to obtain the desired spatial response at the output. We also show that DOA estimation error leads to the deviation of the obtained spatial response from the desired one, irrespective of the angular distance between the sources.

2 Informed Spatial Filter: Review

In this section, a simplified signal model for the ISF framework, with the diffuse sound component omitted, is first presented, followed by the informed LCMV filter [6], for which we present the analysis in the work.

2.1 Signal model

Let us consider a uniform linear array (ULA) of M microphones located at $\mathbf{d}_{1\dots M}$. For each TF instant we assume that the sound field is composed of $L < M$ plane waves. The $M \times 1$ vector of received microphone signals, $\mathbf{y}(n,k) = [Y(n,k,\mathbf{d}_1)\dots Y(n,k,\mathbf{d}_M)]^T$, at time frame n and frequency bin k is given by

$$\mathbf{y}(n,k) = \underbrace{\sum_{l=1}^{L} \mathbf{x}_l(n,k)}_{\mathbf{x}(n,k)} + \mathbf{x}_{\mathrm{n}}(n,k), \tag{1}$$

where $\mathbf{x}_l(n,k) = [X_l(n,k,\mathbf{d}_1)\dots X_l(n,k,\mathbf{d}_M)]^T$ contains the microphone signals for the l-th plane wave, and $\mathbf{x}_{\mathrm{n}}(n,k)$

ISBN 978-3-8007-4275-2

© 2016 VDE VERLAG GMBH · Berlin · Offenbach

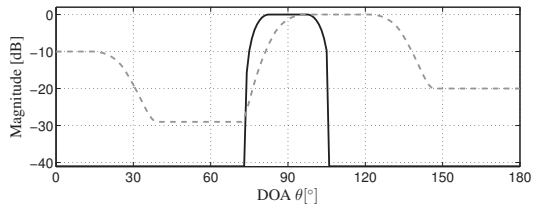


Figure 1: Examples of spatial response functions.

is the spatially uncorrelated and stationary microphone self-noise. The sound pressure corresponding to the l-th plane wave, i.e., the directional sound $\mathbf{x}_l(n,k)$ is given by

$$\mathbf{x}_l(n,k) = \mathbf{a}(\theta_l,k)X_l(n,k,\mathbf{d}_1), \tag{2}$$

where $\theta_l(n,k)$ is the DOA of the l-th plane wave ($\theta = 90^\circ$ denotes the array broadside). For a ULA with omnidirectional microphones, the m-th element of the steering vector $\mathbf{a}(\theta_l,k)$ is given by

$$a_m(\theta_l,k) = \exp\{-j\kappa r_m \cos\theta_l(n,k)\}, \tag{3}$$

where r_m is the distance between the first and the m-th microphone, and κ denotes the wavenumber. Assuming the signal components in (1) to be mutually uncorrelated, the power spectral density (PSD) matrix of the microphone signals can be expressed as

$$\begin{aligned}\mathbf{\Phi}_\mathbf{y}(n,k) &= \mathrm{E}\{\mathbf{y}(n,k)\mathbf{y}^\mathrm{H}(n,k)\} &(4)\\ &= \mathbf{\Phi}_\mathbf{x}(n,k) + \mathbf{\Phi}_\mathrm{n}(n,k), &(5)\end{aligned}$$

with

$$\mathbf{\Phi}_\mathbf{x}(n,k) = \sum_{l=1}^{L} \phi_l(n,k)\mathbf{a}(\theta_l,k)\mathbf{a}^\mathrm{H}(\theta_l,k), \tag{6}$$

$$\mathbf{\Phi}_\mathrm{n}(n,k) = \phi_\mathrm{n}(k)\,\mathbf{I}, \forall n \tag{7}$$

where $\mathbf{I}$ is an identity matrix, $\phi_\mathrm{n}(k)$ denotes the expected spatially white stationary noise power, which is assumed to be identical for all microphones, and $\phi_l(n,k)$ denotes the expected power of the l-th plane wave.

In the ISF framework, the aim is to capture the directional sounds from a specific direction with a specific gain while attenuating the noise. Then, the desired signal can be expressed as

$$Z(n,k) = \sum_{l=1}^{L} G(\theta_l,k)X_l(n,k,\mathbf{d}_1), \tag{8}$$

where $G(\theta_l,k)$ is the directional gain corresponding to the l-th plane wave, whose value is determined based on an arbitrary desired spatial response.

2.2 Informed LCMV filter

An estimate of the desired signal $Z(n,k)$ can be given as a linear combination of the microphone signals $\mathbf{y}(n,k)$ at each TF instant. This estimate, $\hat{Z}(n,k)$, can be written as

$$\hat{Z}(n,k) = \mathbf{w}^\mathrm{H}(n,k)\mathbf{y}(n,k), \tag{9}$$

where $\mathbf{w}$ is a complex weight vector of length M. As in [6], using the LCMV criterion, the weights can be found by minimizing the power of the stationary noise at the output, i.e.,

$$\mathbf{w}(n,k) = \arg\min_\mathbf{w} \mathbf{w}^\mathrm{H}\mathbf{\Phi}_\mathrm{n}(n,k)\mathbf{w} \tag{10}$$

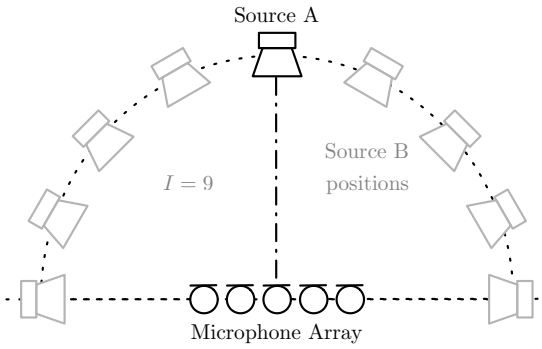


Figure 2: Illustrative figure for the analysis setup.

subject to

$$\mathbf{w}^\mathrm{H}(n,k)\mathbf{a}(\theta_l,k) = G(\theta_l,k)\ \forall\ l \in \{1,\ldots,L\}. \tag{11}$$

The solution is given by

$$\mathbf{w}(n,k) = \mathbf{\Phi}_\mathrm{n}^{-1}\mathbf{A}\left[\mathbf{A}^\mathrm{H}\mathbf{\Phi}_\mathrm{n}^{-1}\mathbf{A}\right]^{-1}\mathbf{g}. \tag{12}$$

where $\mathbf{A} = [\mathbf{a}(\theta_1,k),\ldots,\mathbf{a}(\theta_L,k)]$ contains the propagation vectors corresponding to the L source DOAs. The directional gains are given by $\mathbf{g} = [G(\theta_1,k),\ldots,G(\theta_L,k)]^\mathrm{H}$.

The directional gains $G(\theta_l,k)\ \forall\ l \in \{1,\ldots,L\}$, presented in (8) and (11) correspond to the value of an arbitrary spatial response function, denoted by $\boldsymbol{g}(\boldsymbol{\theta},k)$, evaluated at the DOA of the l-th plane wave. In the following section, a brief explanation regarding the spatial response function is presented.

3 Desired spatial response

In the informed spatial filtering framework, the spatial response is an arbitrary, user-defined function that can be potentially complex valued and frequency dependent. The design of the desired spatial response function is also dependent on the application. In general, one can design any arbitrary spatial response function, e.g., the one presented in Fig. 1 (dashed grey line), which can be a user-defined function to attenuate plane waves from 60° by 29 dB while capturing plane waves from 110° with unit gain.

In source extraction applications, it is desirable that a sound source originating from a specific direction is captured without distortion, while attenuating the direct sound sources from all other angular regions. For example, if the desired source is located in the broadside region of the array, then a spatial response function as shown in Fig. 1 (solid black line) can be employed. This function extracts all plane waves arriving from the angular region close to 90°, while attenuating plane waves from other directions by 41 dB. In some applications, e.g., smart TV, it can be desirable to capture the sound sources from a specific direction of interest irrespective of the location of the active sound sources in the acoustic environment.

As the ISF aims to adapt quickly to any change in the sound scene, at every TF instant, L instantaneous DOA estimates are obtained, based on which the corresponding directional gains are computed by evaluating the response function at the estimated DOA values. Therefore, the obtained spatial response at the output of the filter can potentially vary from the desired one due to DOA estimation errors. To analyze whether the obtained spatial response

ISBN 978-3-8007-4275-2

© 2016 VDE VERLAG GMBH · Berlin · Offenbach

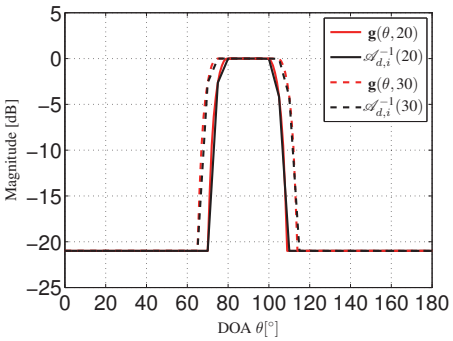


Figure 3: Obtained vs desired spatial response with known source DOAs for different passband widths.

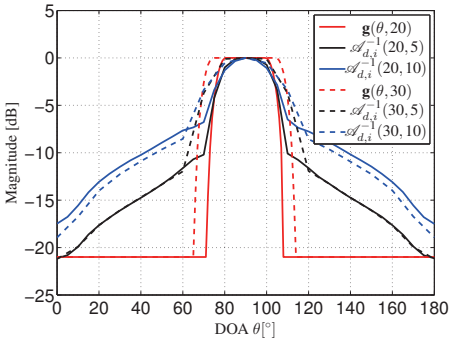


Figure 4: Obtained vs desired spatial response for varying DOA error and passband width.

at the output of the ISF matches the desired spatial response, and also provide an objective analysis of the influence of the DOA estimation errors on the obtained spatial response, we propose an analysis method, described in the following section.

4 Proposed analysis method

The analysis method presented here considers two simultaneously active sound sources with $L = 2$ in the signal model presented in Sec. 2. First, the complete DOA range ($[0, 180]$ for a ULA) is sampled at I discrete points. Then, keeping the position of one of the sources (Source A) constant at one of the I discrete points where the desired response corresponds to unit gain, the other source (Source B) is moved across the whole DOA range, placing it at each of the I discrete points. An illustrative figure for the analysis setup is provided in Fig. 2, where Source A is kept static for the whole experiment at 90°, and Source B is moved through $I = 9$ discrete points.

For each of the I distinct positions of Source B, we compute an objective measure, termed as average directional array gain. To avoid the influence of the spectral characteristics of the source signals, the average directional array gain $\mathscr{A}_{\mathrm{d},i}$ for the i-th position of Source B is given by

$$\mathscr{A}_{\mathrm{d},i} = \frac{1}{\mathrm{card}(\mathscr{T}_{\mathrm{i}})} \sum_{(n,k)\in\mathscr{T}_{\mathrm{i}}} \frac{G_{\mathrm{A}}(n,k)}{G_{\mathrm{B,i}}(n,k)}, \tag{13}$$

where $G_{\mathrm{A}}(n,k)$ and $G_{\mathrm{B,i}}(n,k)$ are the narrowband gains associated with Source A and B, respectively, $\mathscr{T}_{\mathrm{i}}$ represents the set of TF bins where both the sources are simultaneously active, and $\mathrm{card}(\cdot)$ denotes the cardinality operator. The average directional array gain can be interpreted as the obtained spatial response when the directional gain corresponding to Source A is 0 dB.

The gain associated with Source B for the i-th position is given by

$$G_{\mathrm{B,i}}(n,k) = \frac{|\tilde{X}_{\mathrm{B,i}}(n,k)|^2}{|X_{\mathrm{B,i}}(n,k,\mathbf{d}_1)|^2}, \tag{14}$$

where $|X_{\mathrm{B,i}}(n,k,\mathbf{d}_1)|^2$ is the instantaneous power of the signal corresponding to Source B for the i-th position at a reference microphone and $|\tilde{X}_{\mathrm{B,i}}(n,k)|^2$ denotes the instantaneous power of Source B for the i-th position at the filter output. The gain associated with Source A is defined similar to Source B, and is given by

$$G_{\mathrm{A}}(n,k) = \frac{|\tilde{X}_{\mathrm{A}}(n,k)|^2}{|X_{\mathrm{A}}(n,k,\mathbf{d}_1)|^2}, \tag{15}$$

where $|\tilde{X}_{\mathrm{A}}(n,k)|^2$ is the instantaneous power of Source A at the filter output and $|X_{\mathrm{A}}(n,k,\mathbf{d}_1)|^2$ is the instantaneous power of Source A at a reference microphone. It should be noted that only the gain associated with Source B varies with i, since the position of Source A is constant during the whole experiment.

Both sources are considered to be active at a certain TF bin if the dry speech signals corresponding to both sources are above a certain threshold. Mathematically, it is given by

$$\mathscr{T}_{\mathrm{i}} = \{(n,k) : |X_{\mathrm{A}}(n,k,\mathbf{d}_1)| \geq \delta \wedge |X_{\mathrm{B,i}}(n,k,\mathbf{d}_1)| \geq \epsilon\}, \tag{16}$$

where $X_{\mathrm{A}}(n,k,\mathbf{d}_1)$ and $X_{\mathrm{B,i}}(n,k,\mathbf{d}_1)$ are the speech signals corresponding to Source A and B as received by the reference microphone at $\mathbf{d}_1$, respectively, $|\cdot|$ represents the absolute magnitude, and δ and ϵ are the pre-defined thresholds for Source A and B, respectively. In this work, we consider the thresholds to be the average speech signal magnitude over the whole speech duration.

Though the proposed method is presented for the case of $L = 2$, the same method can also be used to analyze the desired response for spatial filtering algorithms that consider the common single plane wave signal model, i.e., $L = 1$. A generalization of the proposed method for $L > 2$ is also possible by considering the total power of all the interfering source signals. However, it should be noted that in this case, the visualization of the results need to be done in $L-1$ dimensions for a proper analysis. Finally, it is worthwhile to note that by varying the input signal-to-interference ratio (iSIR), it is also possible to identify the critical iSIR where the single plane wave model is violated.

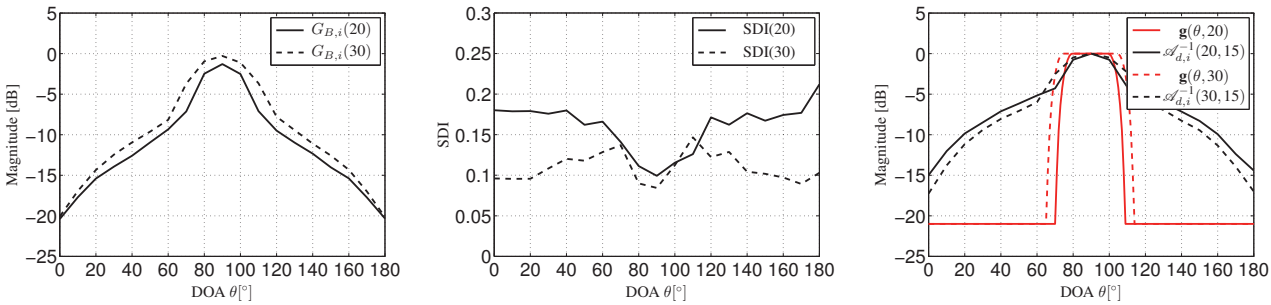
5 Experimental analysis

5.1 Simulation setup

We assume a ULA with $M = 5$ omnidirectional microphones where the inter-microphone distance is 3 cm. We consider an anechoic environment with no microphone self-noise, resulting in the choice of $\mathbf{\Phi}_{\mathrm{n}} = \mathbf{I}_{M\times M}$. For the proposed analysis method, to compute the positions for Source B, we sample the complete DOA range, i.e., $[0, 180]$, uniformly at $I = 37$ discrete points. Source A is kept static at 90°. The two sources are placed at a distance of 1.8 m from the array center, for all the possible positions of the Source B. The input signal consists of two simulated speech signals of 6 s duration, with a sampling rate of $F_s = 16$ kHz. It should be noted that with the varying position of Source B, the corresponding input speech signal also changes. A

ISBN 978-3-8007-4275-2

© 2016 VDE VERLAG GMBH · Berlin · Offenbach



(a) $G_{\mathrm{B,i}}$ (b) SDI (c) Obtained vs desired spatial response

Figure 5: Experimental analysis with $\sigma_{\mathrm{DOA}} = 15^\circ$ for passband widths of 20° and 30°.

512 point short-time Fourier transform (STFT) with 50% overlap is used to transform the signals into the TF domain. For all analyses presented here, we consider the desired spatial response for the example application of source extraction, Fig. 1 (solid black line), except that we consider the level of attenuation as -21 dB. In all the following analysis results, we plot the inverse of $\mathscr{A}_{\mathrm{d},i}$ as the obtained spatial response.

5.2 With exact DOAs

We first consider the best case scenario and analyze the obtained spatial response for the case when the DOAs of both Source A and B are exactly known. For this case, we consider two different passband widths of the desired spatial response, 20° and 30°. In Fig. 3, it can be seen that the obtained spatial response, given by the inverse array gain, for both passband widths (solid and dashed black line) exactly matches the desired spatial responses (solid and dashed red line). Thus, with perfectly accurate DOAs, it is possible to obtain the desired spatial response at the filter output if the source DOAs are known.

5.3 With erroneous DOAs

To gain insight into the influence of the DOA estimation errors, we model them, at each TF instant, by a zero-mean Gaussian process with a standard deviation σ_{DOA}, added to the known DOAs of Source A and B. However, with the proposed method, it is also possible to do the same analysis for a specific DOA estimator. For Source A and the i-th position of Source B, the DOA estimates can be expressed as

$$\begin{aligned}\hat{\theta}_{\mathrm{A}}(n,k) &= \theta_{\mathrm{A}} + \Delta\theta_{\mathrm{A}}(n,k),\\ \hat{\theta}_{\mathrm{B,i}}(n,k) &= \theta_{\mathrm{B,i}} + \Delta\theta_{\mathrm{B,i}}(n,k),\end{aligned} \tag{17}$$

where $\Delta\theta_{\mathrm{A}}(n,k)$ and $\Delta\theta_{\mathrm{B}}(n,k)$ are the absolute DOA errors corresponding to Source A and B, respectively, with $\sigma^2_{\mathrm{DOA}} = \mathrm{E}\{\Delta\theta^2\}$ as the error variance. It should be noted that for a complete experiment, for each i, the error variance is kept constant.

In Fig. 4, the obtained spatial response with DOA errors of $\sigma_{\mathrm{DOA}} = 5^\circ, 10^\circ$, is shown for two different passband widths of 20° and 30°. It can be seen that with the introduction of DOA estimation errors, irrespective of the passband width, the obtained spatial response deviates from the desired response and is no longer able to achieve the desired level of suppression when Source B is outside the passband region. Also, as the DOA estimation error increases the deviation also increases. For low estimation error of $\sigma_{\mathrm{DOA}} = 5^\circ$ (black lines), the obtained spatial responses are similar for different passband widths, especially outside of the passband region. However, for the higher estimation error of $\sigma_{\mathrm{DOA}} = 10^\circ$ (blue lines), the obtained spatial response with 30° passband width is relatively closer to the corresponding desired spatial response, than the obtained spatial response with 20° passband width.

To analyze this difference in the obtained spatial response for different passband widths with the same estimation error, we conducted the experiment with a higher DOA estimation error of $\sigma_{\mathrm{DOA}} = 15^\circ$. For further analysis, we also compute an additional objective measure, signal distortion index (SDI) [13], of Source A for each i. In Fig. 5a, the plot for the average gain associated with Source B for each i, i.e., $G_{\mathrm{B,i}} = \frac{1}{\mathrm{card}(\mathscr{T}_{\mathrm{i}})}\sum_{(n,k)\in\mathscr{T}_{\mathrm{i}}} G_{\mathrm{B,i}}(n,k)$, is shown, Fig. 5b shows the SDI for Source A and in Fig. 5c the obtained spatial response is shown along with the desired response. It can be observed that the average gain for Source B, $G_{\mathrm{B,i}}$, is closer to the desired spatial response compared to the obtained spatial response. However, due to the introduction of distortion to the signal of Source A, the obtained spatial response further deviates from the desired one. Also, for the wider passband of 30°, the obtained spatial response is closer to the desired response. This stems from the fact that for a wider passband, the signal distortion is lower, which leads to a better average directional array gain.

Based on the presented analysis, it can be seen that though ISF provides a flexible filtering framework to capture multiple sound sources with an arbitrary spatial response, DOA estimation errors severely affect its ability to achieve the desired spatial response. Therefore, robustness to DOA estimation errors is essential for its proper functioning.

6 Conclusions

A method to analyze the ability of ISF to obtain an arbitrary, user-defined, desired spatial response at the output of the filter was presented. It was shown that with perfect knowledge of the source DOAs the desired spatial response can be obtained at the output of the ISF. Through the analysis of the influence of DOA estimation errors on the obtained spatial response, it was shown that robustness against DOA estimation errors is important and essential for the ISF framework.

ISBN 978-3-8007-4275-2

© 2016 VDE VERLAG GMBH · Berlin · Offenbach

References

[1] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, Springer-Verlag, Berlin, Germany, 2008.

[2] S. Gannot and I. Cohen, "Adaptive Beamforming and Postfiltering," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds., chapter 48. Springer-Verlag, 2007.

[3] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2677–2684, Oct. 1999.

[4] S. Araki, H. Sawada, and S. Makino, "Blind speech separation in a meeting situation with maximum SNR beamformers," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2007.

[5] M. Kallinger, G. Del Galdo, F. Kuech, D. Mahne, and R. Schultz-Amling, "Spatial filtering using directional audio coding parameters," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2009.

[6] O. Thiergart and E.A.P. Habets, "An informed LCMV filter based on multiple instantaneous direction-of-arrival estimates," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 659–663.

[7] O. Thiergart, M. Taseska, and E.A.P. Habets, "An informed MMSE filter based on multiple instantaneous direction-of-arrival estimates," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Sept 2013, pp. 1–5.

[8] O. Thiergart, M. Taseska, and E.A.P. Habets, "An informed parametric spatial filter based on instantaneous direction-of-arrival estimates," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 22, no. 12, pp. 2182–2196, Dec 2014.

[9] P. Aarabi and G. Shi, "Phase-based dual-microphone robust speech enhancement," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 4, pp. 1763–1773, Aug 2004.

[10] A. Sugiyama and R. Miyahara, "A directional noise suppressor with a specified beamwidth," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 524–528.

[11] O. u. R. Qazi, B. van Dijk, M. Moonen, and J. Wouters, "Speech understanding performance of cochlear implant subjects using time frequency masking-based noise reduction," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1364–1373, May 2012.

[12] O. Thiergart and E. A. P. Habets, "Sound field model violations in parametric spatial sound processing," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, 2012.

[13] J. Benesty, J. Chen, and E. A. P. Habets, *Speech Enhancement in the STFT Domain*, Springer Briefs in Electrical and Computer Engineering. Springer, 2012.

ISBN 978-3-8007-4275-2

© 2016 VDE VERLAG GMBH · Berlin · Offenbach