

ADAPTIVE FILTER-AND-SUM BEAMFORMING IN SPATIALLY CORRELATED NOISE

Reinhold Haeb-Umbach and Ernst Warsitz

{haeb, warsitz}@nt.uni-paderborn.de

University of Paderborn, Dept. of Communications Engineering, 33098 Paderborn, Germany

ABSTRACT

In this paper we propose a novel adaptation algorithm for Filter-and-Sum beamforming in spatially correlated noise. Deterministic and stochastic gradient ascent algorithms are derived from a constrained optimization problem, which iteratively estimate the principal eigenvector of a generalized eigenvalue problem. The method does not require an explicit estimation of the speaker location. It is shown that the well-known Delay-and-Sum beamformer and the previously introduced Filter-and-Sum beamformer in spatially white noise are obtained as special cases. Further, bounds on the maximally achievable SNR gains are derived and it is shown that the proposed adaptation algorithm is able to approach these performance bounds.

1. INTRODUCTION

Microphone arrays are often employed for hands-free speech communication or recognition. With multi-channel signal processing a beam of increased sensitivity is directed towards a possibly moving speaker, whose position is not known a priori.

A popular solution is the use of a Delay-and-Sum (DSB) beamformer, which compensates propagation path length differences of the direct ("line-of-sight") path from the source to each sensor, to obtain a properly aligned direct path signal at the output. However, there are various shortcomings of this approach: First, source localization techniques are required to estimate the speaker position or at least the direction-of-arrival (DOA), which are problematic on their own right in reverberant environments [1]. Further, the position of the microphones has to be exactly known to correctly transform DOA-estimates into time delays in each microphone path. Since the DSB aligns only the direct path signal, it does not take into account reflections, and the effect of spatially correlated noise is not addressed.

Recently we have proposed an adaptive Filter-and-Sum beamformer (FSB) which overcomes many of the shortcomings of the DSB [2]: The adaptation works blindly, i.e. no explicit source localization is required. Further, the beamformer explicitly addresses reverberant environments by aligning the direct path signal and, in addition, early reflections. The FSB adaptation algorithm has been derived from a constraint optimization problem, the solution of which is the principal eigenvector of the cross power spectral density matrix of the microphone signals. Deterministic and stochastic gradient ascent algorithms have been derived assuming spatially white noise. Simulations have revealed high robustness of the scheme to fast changes of speaker position.

The purpose of this paper is to extend our previous work to the practically more realistic case of spatially correlated noise. An

algorithm is derived to adaptively estimate the principal eigenvector of a generalized eigenvalue problem. Further, it is shown that the FSB reduces to a DSB if certain simplifying assumptions are made about the acoustic environment. We derive performance bounds and show that the proposed adaptation scheme comes close to these bounds.

2. SIGNAL MODEL

We are given an array of M microphones. Each microphone signal $x_i(n)$, $i = 1, \dots, M$, where n denotes the discrete-time index, is assumed to consist of two components: A signal component $s_i(n)$, which results from the convolution of the desired (speech) source signal $u(n)$ with the room impulse response $h_i(n)$ from the source position to the i -th sensor, and the noise term $n_i(n)$:

$$\begin{aligned} x_i(n) &= s_i(n) + n_i(n) \\ &= h_i(n) * u(n) + n_i(n); \quad i = 1, \dots, M. \end{aligned} \quad (1)$$

The goal of beamforming is to obtain an estimate of $u(n)$ by filtering and then summing the microphone signals:

$$y(n) = \sum_{i=1}^M \tilde{f}_i(n) * x_i(n). \quad (2)$$

Here, $\tilde{f}_i(n) = f_i(-n)$ is the impulse response of the filter of the i -th microphone signal.

The filtering operation is preferably done in the frequency domain employing a Discrete Fourier Transform:

$$Y(k) = \sum_{i=1}^M F_i^*(k) \cdot X_i(k) = \sum_{i=1}^M F_i^*(k) \cdot (S_i(k) + N_i(k)). \quad (3)$$

Here, k denotes the frequency bin: $k = 0, \dots, L - 1$ (L : DFT-length). The frame index has been omitted for ease of notation. $F_i^*(k)$, $X_i(k)$, $S_i(k)$, and $N_i(k)$ are the DFTs of $\tilde{f}_i(n)$, $x_i(n)$, $s_i(n)$, and $n_i(n)$, respectively. In the following we will use the vector notation, i.e. $\mathbf{X}(k) = (X_1(k), \dots, X_M(k))^T$ and $\mathbf{F}(k) = (F_1(k), \dots, F_M(k))^T$, such that (3) can be written as:

$$Y(k) = \mathbf{F}^H(k) \cdot \mathbf{X}(k); \quad k = 0, \dots, L - 1, \quad (4)$$

where $(\cdot)^H$ denotes Hermitian transpose.

There are different options for $\tilde{f}_i(n)$: In the case of a *Delay-and-Sum Beamformer* (DSB) the filter is a discrete-time interpolator to realize integer or fractional delays. The delays are chosen to compensate for the path length differences of the direct ("line-of-sight") propagation path from the source to the individual sensors.

In this paper we consider the more general case of a *Filter-and-Sum Beamformer* (FSB), where $\tilde{f}_i(n)$ is an arbitrary FIR filter. The filter coefficients are chosen such, that in addition to the line-of-sight signal component, also early reflections are aligned in order to add constructively at the output of the beamformer. In the next section we show how optimal FSB coefficients can be obtained.

3. OPTIMAL FSB

3.1. Constrained Optimization Problem

If the desired signal $s_i(n)$ and the noise $n_i(n)$ are uncorrelated the power spectral density (PSD) of the FSB output can be written as

$$\begin{aligned}\Phi_{YY}(k) &= \mathbf{F}^H(k)\Phi_{XX}(k)\mathbf{F}(k) \\ &= \mathbf{F}^H(k)\Phi_{SS}(k)\mathbf{F}(k) + \mathbf{F}^H(k)\Phi_{NN}(k)\mathbf{F}(k),\end{aligned}\quad (5)$$

where $\Phi_{XX}(k)$, $\Phi_{SS}(k)$ and $\Phi_{NN}(k)$ are the cross power spectral density matrices of the microphone signals, the speech and noise terms, respectively.

Our goal is to determine a vector of filter coefficients $\mathbf{F}(k)$ such that the Signal-to-Noise ratio

$$\text{SNR}(k) = \frac{\mathbf{F}^H(k)\Phi_{XX}(k)\mathbf{F}(k)}{\mathbf{F}^H(k)\Phi_{NN}(k)\mathbf{F}(k)} - 1 \quad (6)$$

of the output signal $Y(k)$ is maximized.

In light of eq. (6), maximizing the SNR of the beamformer output is then equivalent to the following constrained optimization problem:

$$\max_{\mathbf{F}^H(k)} \mathbf{F}^H(k)\Phi_{XX}(k)\mathbf{F}(k) \text{ subj. to } \mathbf{F}^H(k)\Phi_{NN}(k)\mathbf{F}(k) = C, \quad (7)$$

where $C \in \mathbf{R}$ is an arbitrary real constant, which is set to $C = 1$ in the following.

Introducing a Lagrange multiplier β we arrive at the optimization function

$$J(\mathbf{F}, \beta) = \mathbf{F}^H(k)\Phi_{XX}(k)\mathbf{F}(k) + \beta(\mathbf{F}^H(k)\Phi_{NN}(k)\mathbf{F}(k) - 1). \quad (8)$$

Computing the gradient $\nabla_{\mathbf{F}} J(\mathbf{F}, \beta)$ w.r.t. \mathbf{F} and setting it to zero results in

$$\nabla_{\mathbf{F}} J(\mathbf{F}, \beta) = 2\Phi_{XX}(k)\mathbf{F}(k) + 2\beta\Phi_{NN}(k)\mathbf{F}(k) = 0, \quad (9)$$

which demonstrates that the optimal filter coefficient vector is an eigenvector of a generalized eigenvalue decomposition, a well-known result in the array processing literature. It is also well-known that other optimization criteria, e.g. MMSE, ML or MVDR result in a weight vector $\mathbf{F}'(k)$ which is the same as the above weight vector, up to a scalar constant [3]:

$$\mathbf{F}'(k) = w(k)\mathbf{F}(k). \quad (10)$$

Thus any optimization criterion can be realized by an FSB with filter coefficients as derived above and a single-channel post-filter, which implements $w(k)$.

3.2. Adaptation Algorithm

To iteratively solve eq. (9) we developed a deterministic gradient ascent scheme:

$$\mathbf{F}(\kappa + 1) = \mathbf{F}(\kappa) + \frac{\mu}{2} \nabla_{\mathbf{F}} J(\mathbf{F}, \beta) \Big|_{\mathbf{F}=\mathbf{F}(\kappa)}, \quad (11)$$

where κ counts the iterations, and μ is the step size parameter. The Lagrange multiplier β is computed by postulating that the constraint is to be met at the next iteration step:

$$\mathbf{F}^H(\kappa + 1)\Phi_{NN}\mathbf{F}(\kappa + 1) \stackrel{!}{=} C. \quad (12)$$

Using (11) in (12), neglecting terms of order $\mathcal{O}(\mu^2)$ and solving for β we obtain

$$\beta \approx \frac{C - \mathbf{F}^H(\kappa)\Phi_{NN}\mathbf{F}(\kappa) - \mu\mathbf{F}^H(\kappa)\Phi^{(XN)}\mathbf{F}(\kappa)}{2\mu\mathbf{F}^H(\kappa)\Phi_{NN}\mathbf{F}(\kappa)}, \quad (13)$$

where

$$\Phi^{(XN)} = \Phi_{XX}\Phi_{NN} + \Phi_{NN}\Phi_{XX}. \quad (14)$$

Using this in (11) we obtain the following deterministic gradient ascent algorithm for solving the GEVD:

$$\begin{aligned}\mathbf{F}(\kappa + 1) &= \mathbf{F}(\kappa) + \frac{C - \mathbf{F}^H(\kappa)\Phi_{NN}\mathbf{F}(\kappa)}{2\mathbf{F}^H(\kappa)\Phi_{NN}\mathbf{F}(\kappa)}\Phi_{NN}\mathbf{F}(\kappa) \\ &+ \mu \left[\Phi_{XX}\mathbf{F}(\kappa) - \frac{\mathbf{F}^H(\kappa)\Phi^{(XN)}\mathbf{F}(\kappa)}{2\mathbf{F}^H(\kappa)\Phi_{NN}\mathbf{F}(\kappa)}\Phi_{NN}\mathbf{F}(\kappa) \right].\end{aligned}\quad (15)$$

Since the cross power spectral density matrices are not known in practice, a stochastic gradient ascent algorithm is used by approximating:

$$\phi_{XX} \approx \mathbf{X}(m)\mathbf{X}^H(m), \quad (16)$$

where m denotes the frame index. The cross power spectral density matrix of the noise is also not known a priori. It can be replaced by an estimate, which we obtain at the sensors during speech pauses

$$\begin{aligned}\Phi_{NN} &\approx \hat{\Phi}_{NN}(m) = \epsilon \cdot \hat{\Phi}_{NN}(m-1) \\ &+ (1 - \epsilon) \cdot \mathbf{X}(m)\mathbf{X}^H(m) \Big|_{\mathbf{X}(m)=\mathbf{N}(m)}\end{aligned}\quad (17)$$

with some smoothing constant ϵ , $0 < \epsilon < 1$.

Using (16) and (17) in (15) we obtain the stochastic version of the iterative generalized eigenvalue decomposition:

$$\begin{aligned}\mathbf{F}(m + 1) &= \mathbf{F}(m) + \frac{1 - \mathbf{F}^H(m)\mathbf{G}(m)}{2\mathbf{G}^H(m)\mathbf{G}(m)}\mathbf{G}(m) \\ &+ \mu Y^*(m) \left[\mathbf{X}(m) - \frac{\mathbf{G}(m)A(m)}{2\mathbf{G}^H(m)\mathbf{G}(m)} \right]\end{aligned}\quad (18)$$

where

$$\begin{aligned}\mathbf{G}(m) &= \hat{\Phi}_{NN}(m)\mathbf{F}(m), \\ A(m) &= Y(m)/Y^*(m)\mathbf{X}^H(m)\mathbf{G}(m) + \mathbf{G}^H(m)\mathbf{X}(m).\end{aligned}$$

In the special case of spatially white noise, see eq. (24) later on, (18) reduces to the adaptation rule presented in [2].

In this section we omitted the frequency bin index k . However it has to be kept in mind that the aforementioned iterations have to be carried out for every frequency bin separately.

A convergence and stability analysis showed that the rule is very robust to speaker movements. The step size parameter must be chosen from the interval $0 < \mu < 2/\lambda_{max}$ to ensure stability [5].

4. PERFORMANCE BOUNDS

We have seen in Section 3.1 that the vector of optimal filter coefficients is the principal eigenvector of the generalized eigenvalue problem (9). Here we consider special cases where the eigenvector and corresponding eigenvalue can be easily computed.

4.1. FSB and DSB

Computing the DFT of (1) the vector of microphone signals can be written as

$$\mathbf{X}(k) = \mathbf{H}(k)U(k) + \mathbf{N}(k). \quad (19)$$

Here, $U(k)$ is the DFT of $u(n)$. In the following we assume for convenience: $|U(k)|^2 = 1$.

The power spectral density matrix of the sensor signals is given by

$$\Phi_{\mathbf{X}\mathbf{X}} = \mathbf{H}\mathbf{H}^H + \Phi_{\mathbf{N}\mathbf{N}} \quad (20)$$

where we have omitted the frequency index k for ease of notation. Since $\mathbf{H}^H \Phi_{\mathbf{N}\mathbf{N}}^{-1} \mathbf{H}$ is a scalar, the following holds

$$\Phi_{\mathbf{N}\mathbf{N}}^{-1} (\mathbf{H}\mathbf{H}^H + \Phi_{\mathbf{N}\mathbf{N}}) \Phi_{\mathbf{N}\mathbf{N}}^{-1} \mathbf{H} = (\mathbf{H}^H \Phi_{\mathbf{N}\mathbf{N}}^{-1} \mathbf{H} + 1) \Phi_{\mathbf{N}\mathbf{N}}^{-1} \mathbf{H}. \quad (21)$$

Obviously, $\Phi_{\mathbf{N}\mathbf{N}}^{-1} \mathbf{H}$ is a generalized eigenvector of $\Phi_{\mathbf{N}\mathbf{N}}^{-1} \Phi_{\mathbf{X}\mathbf{X}}$ = $\Phi_{\mathbf{N}\mathbf{N}}^{-1} (\mathbf{H}\mathbf{H}^H + \Phi_{\mathbf{N}\mathbf{N}})$ with eigenvalue $(\mathbf{H}^H \Phi_{\mathbf{N}\mathbf{N}}^{-1} \mathbf{H} + 1)$.

If we assume that the source is in the far field of a uniform linear microphone array, then the time which a plane wave, arriving from a source in direction θ relative to broadside, takes to travel from the first sensor to the i -th is easily computed as

$$\tau_i(\theta) = \frac{d}{c}(i-1) \sin \theta, \quad (22)$$

where d denotes the inter-element spacing and c is the propagation velocity. In the absence of reverberation the vector of transfer functions has the form

$$\mathbf{H}(k) = (1 e^{-j\omega_k \tau_2(\theta)}, \dots, e^{-j\omega_k \tau_M(\theta)})^T. \quad (23)$$

Here, $\omega_k = 2\pi k/(LT)$ is the frequency variable (T : sampling period).

In the case of spatially uncorrelated noise we further have

$$\Phi_{\mathbf{N}\mathbf{N}}(k) = \sigma_N^2(k) \cdot \mathbf{I}_M, \quad (24)$$

where \mathbf{I}_M is the identity matrix of dimension $M \times M$, and $\sigma_{N_1}^2(k) = \dots = \sigma_{N_M}^2(k) = \sigma_N^2(k)$ is the variance of the noise term at a microphone, which is assumed to be equal for all microphones, but which can depend on the frequency index.

As a result, $\Phi_{\mathbf{N}\mathbf{N}}^{-1} \Phi_{\mathbf{X}\mathbf{X}}$ has only one eigenvector with eigenvalue larger than zero, and this eigenvector is equal to $\mathbf{H}(k)$. Therefore the optimal beamformer weights are $\mathbf{F}(k) = \mathbf{H}(k)$. But these are exactly the weights of a Delay-and-Sum beamformer. We conclude that optimization of an FSB in the case of a source in the far-field, absence of reverberation and spatially white noise leads to the DSB solution.

4.2. SNR Gain

Next we compute the signal-to-noise ratio at the beamformer output, if the optimal filter coefficients are used.

From (6) and (9) we obtain

$$\begin{aligned} \text{SNR}(k) &= \frac{\mathbf{F}^H(k) \lambda_{max} \Phi_{\mathbf{N}\mathbf{N}}(k) \mathbf{F}(k)}{\mathbf{F}^H(k) \Phi_{\mathbf{N}\mathbf{N}}(k) \mathbf{F}(k)} - 1 \quad (25) \\ &= \lambda_{max} - 1 = \tilde{\lambda}_{max}. \quad (26) \end{aligned}$$

Where λ_{max} is the largest eigenvalue of $\Phi_{\mathbf{N}\mathbf{N}}^{-1} \Phi_{\mathbf{X}\mathbf{X}}$, and $\tilde{\lambda}_{max}$ is the largest eigenvalue of $\Phi_{\mathbf{N}\mathbf{N}}^{-1} \Phi_{\mathbf{S}\mathbf{S}}$. The maximally achievable SNR gain, which is realized if the filter coefficients are equal to the principal eigenvector, is given by the largest eigenvalue of the generalized eigenvalue problem!

From (21) we obtain (21)

$$\text{SNR}(k) = \mathbf{H}^H(k) \Phi_{\mathbf{N}\mathbf{N}}^{-1}(k) \mathbf{H}(k) \quad (27)$$

In the special case of the far-field assumption, no reverberation and spatially uncorrelated noise we arrive at

$$\text{SNR}(k) = \frac{\mathbf{H}^H(k) \mathbf{H}(k)}{\sigma_N^2(k)} = \frac{M}{\sigma_N^2(k)}, \quad (28)$$

i.e. the SNR gain, which is the ratio of the SNR at the beamformer output to the SNR at the sensors and for spatially uncorrelated noise given by $10 \log_{10}(M)$ dB.

Another interesting case is diffuse noise. In this case the entry in the i -th row and j -th column of $\Phi_{\mathbf{N}\mathbf{N}}$ is given by

$$\{\Phi_{\mathbf{N}\mathbf{N}}\}_{i,j} = \sigma_N^2 \Gamma_{i,j}, \quad (29)$$

with

$$\Gamma_{i,j} = \text{si} \left(\frac{\omega_k d_{ij}}{c} \right), \quad (30)$$

where $\text{si}(x) = \sin(x)/x$. Here $d_{ij} = |i-j|d$ is the distance between the i -th and j -th sensor. Noise in the passenger cabin of a car can be well characterized as diffuse noise [4].

5. EXPERIMENTAL RESULTS

In the following experiments we are going to compare the performance of the adaptive FSB of the last section with the theoretical maximal achievable SNR gains under different reverberation and noise scenarios.

Figure 1 shows the SNR-gain obtained by the adaptive FSB as a function of room reverberation time RT_{60} in a spatially uncorrelated noise environment for different SNR values at the sensors. We used a 4-channel linear microphone array with a spacing of 5 cm at a sampling rate of 8 kHz and the speech source was positioned at 30 degrees relative to broadside. The FSB-filterlength was set to 64 and the DFT-length was 128. If it is known in advance that the noise is spatially uncorrelated, i.e. that (24) holds, eq. (18) can be considerably simplified [2]. However, here we used (18), where the cross power spectral density of the noise was estimated during speech pauses. The results demonstrate that the noise cross power spectral density matrix could be efficiently estimated during speech pauses. It can be seen that the adaptive FSB is able to achieve almost the complete maximally achievable SNR gain of 6 dB, both for an input SNR of 15 dB and 5 dB at the microphones.

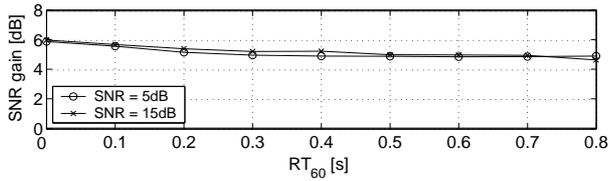


Figure 1: SNR gain achieved by the adaptive FSB as a function of room reverberation time RT_{60} (spatially uncorrelated noise).

In another set of experiments diffuse noise and absence of reverberation was considered. Note that the matrix (29) is singular for $\omega_k = 0$. A regularisation term was therefore included, i.e. $\Phi_{NN}(k)$ was replaced by $\Phi_{NN}(k) + \delta \mathbf{I}_M$. Assuming absence of reverberation the maximally achievable SNR gain as a function of frequency is displayed in Fig. 2 for different values of δ . Note, that the SNR gain also depends on the DOA θ , which is not surprising since \mathbf{H} has the form given in (23). In order to assess the performance of the adaptive FSB, diffuse

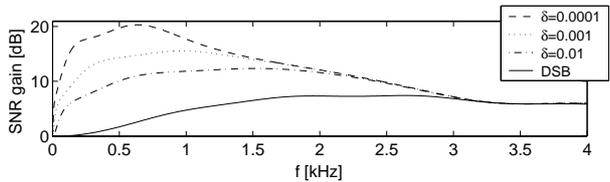


Figure 2: Maximally achievable SNR gain in diffuse noise as a function of frequency for different values of the regularization parameter δ and with DOA $\theta = 58^\circ$.

noise was simulated by placing two computer fans in a reverberant room without direct path to the microphones. Fig. 3 compares the measured coherence functions with that of an ideally diffuse noise field. The microphone signals were then computed

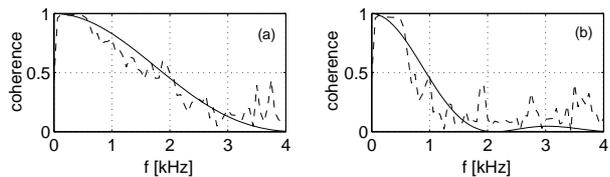


Figure 3: Coherence function of the ideally diffuse sound field (solid) and measured coherence (dotted) for microphone distance $d = 5\text{cm}$ in (a) and $d = 10\text{cm}$ in (b).

using the well-known image method. Fig. 4 shows the SNR gain obtained by the adaptive FSB and the SNR gain obtained by a "perfect" DSB, where the DOA is assumed to be perfectly known. Whereas a DSB can obtain at most $10 \log_{10}(M)$ dB gain, the FSB achieves much higher gains at low frequencies, which is very desirable in a car environment.

Finally Fig. 5 shows the SNR gain over the reverberation time RT_{60} achieved by the proposed FSB compared to a DSB for an SNR of about 5dB at the microphones and $\theta = 58^\circ$.

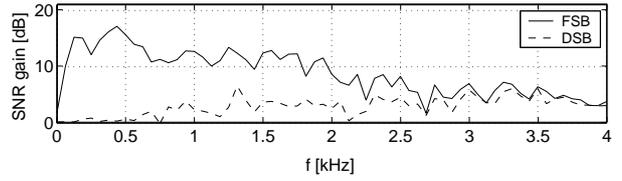


Figure 4: SNR gain achieved by the adaptive FSB and DSB as a function of frequency in a diffuse noise environment.

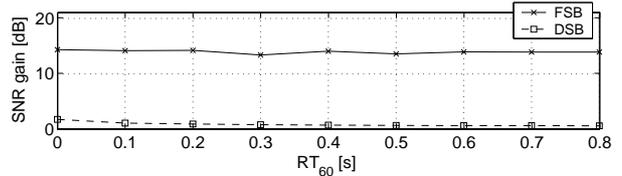


Figure 5: SNR gain achieved by the adaptive FSB and the DSB as a function of room reverberation time RT_{60} (diffuse noise).

6. CONCLUSIONS

In this paper we have presented an adaptive Filter-and-Sum (FSB) beamformer which is able to achieve close to the theoretically maximum SNR gain both in spatially correlated and uncorrelated noise, be it a reverberant or a non-reverberant environment. The well-known Delay-and-Sum (DSB) beamformer is obtained as a special case of the FSB, if absence of reverberation and spatially white noise is assumed. For the practically very relevant case of spatially correlated noise, the proposed FSB achieves considerably larger SNR gains than a conventional DSB. Note that the proposed FSB can be used as a steering component of a broadband beamformer such as the Generalized Sidelobe Canceller [6], where a blocking matrix and adaptive interference cancellers are used to suppress directional noise. Investigations of this setup are currently underway.

7. REFERENCES

- [1] J.H. DiBiase, H.F. Siverman and M.S. Brandstein, "Robust localization in reverberant rooms", in *Microphone Arrays: Signal Processing Techniques and Applications*, Springer Verlag, 2001.
- [2] E. Warsitz and R. Haeb-Umbach, "Acoustic Filter-and-Sum Beamforming by Adaptive Principal Component Analysis", in *Proc. ICASSP*, Philadelphia, USA, Mar. 2005.
- [3] L.C. Godara, "Applications of Antenna Arrays to Mobile Communications, Part II: Beam-Forming and Direction-of-Arrival Considerations", *Proceedings of the IEEE*, vol. 85, no. 8, pp 1195-1245, Aug. 1997.
- [4] J. Meyer and K.U. Simmer, "Multi-Channel Speech Enhancement in a Car Environment Using Wiener Filtering and Spectral Subtraction", in *Proc. ICASSP*, Munich, Apr. 1997.
- [5] R. Haeb-Umbach and E. Warsitz, "Adaptive Filter-and-Sum Beamforming by Generalized Eigenvalue Decomposition", Techn. Report 1/2005, University of Paderborn (submitted to *IEEE Transactions on Speech and Audio Processing*).
- [6] L.J. Griffiths and C.W. Jim, "An alternative approach to linearly constrained adaptive beamforming", *IEEE Trans. on Antennas and Propagation*, vol. 30, no. 1, pp. 27-34, Jan. 1982.