

AN ADAPTIVE MICROPHONE ARRAY FOR OPTIMUM BEAMFORMING AND NOISE REDUCTION

Gerhard Doblinger

Institute of Communications and Radio-Frequency Engineering
Vienna University of Technology
Gusshausstr. 25/389, A-1040 Vienna, Austria
phone: + (43) 1 58801 38927, fax: + (43) 1 58801 38999, email: gerhard.doblinger@tuwien.ac.at,
web: www.nt.tuwien.ac.at/about-us/staff/gerhard-doblinger/

ABSTRACT

We present a new adaptive microphone array efficiently implemented as a multi-channel FFT-filterbank. The array design is based on a minimum variance distortionless response (MVDR) optimization criterion. MVDR beamformer weights are updated for each signal frame using an estimated spatio-spectral correlation matrix of the environmental noise field. We avoid matrix inversion by means of an iterative algorithm for weight vector computation. The beamformer performance is superior to designs based on an assumed homogeneous diffuse noise field. The new design also outperforms LMS-adaptive beamformers at the expense of a higher computational load. Additional noise reduction is achieved with the well-known beamformer/postfilter combination of the optimum multi-channel filter. An Ephraim-Malah spectral amplitude modification with minimum statistics noise estimation is employed as a postfilter. Experimental results are presented using sound recordings in a reverberant noisy room.

1. INTRODUCTION

Suppression of noise and reverberation is needed for many sound capturing applications. Multi-channel interference suppression algorithms are superior to single-channel systems since they incorporate both spatial and temporal information of the sound field. Microphone arrays with a beamformer/postfilter combination for noise reduction are highly efficient. Based on a multi-channel Wiener optimum filter, the beamformer/postfilter technique is widely used for speech enhancement purposes (see e.g. [1, 2]). Normally, the generalized side-lobe canceler (GSC) is used as an adaptive beamforming device. It is more efficient than the classical Frost beamformer [3] but in general tends to suppress the desired speech signal.

A robust GSC beamformer with an adaptive blocking matrix is presented in [4], and an efficient implementation is reported in [5]. The main advantages are a flat top main lobe, and reduced desired signal suppression resulting in an improved array pattern as compared with the standard GSC beamformer. However, two adaptive algorithms (one for blocking matrix update, the other for noise cancellation) must cooperate in order to achieve the desired behavior. A long convergence time of the adaptive algorithms is needed, especially in acoustic environments with strong reverberation and echoes. A further extension of the classical GSC beamformer is presented in [6, 7]. A fixed blocking matrix is used, but the actual steering vector is included in the design by acoustic channel estimation. Nevertheless, convergence

speed is limited by the LMS-adaptive algorithms and by the time constant of channel transfer function estimation.

GSC-based beamforming algorithms are capable to incorporate an estimated noise spatio-spectral correlation matrix into the update of the beamformer weight vector. Because this update is carried out on a frame-by-frame basis, optimum weight vectors are approximated during a relative large number of frames. In this paper, the actual spatio-spectral correlation matrix is estimated too. However, beamformer weight vectors are optimized for each signal frame. Therefore, a significantly improved array pattern and noise reduction, and faster tracking of time-variant noise fields can be achieved.

We first derive an iterative minimum variance distortionless response (MVDR) beamforming algorithm. Afterwards the optimum beamformer is used as a pre-processing device to a single channel noise reduction system. This approach is motivated by an efficient representation of a frequency-domain multichannel filter. Experimental results are presented to justify the proposed technique.

2. MVDR BEAMFORMER WITH ITERATIVE WEIGHT VECTOR COMPUTATION

We consider a sound capture situation as sketched in Fig. 1. The channel impulse responses $h_i(\vec{r}, t)$ describe sound propagation from the source to the individual microphones and include not only the direct paths but also echoes and reverberation.

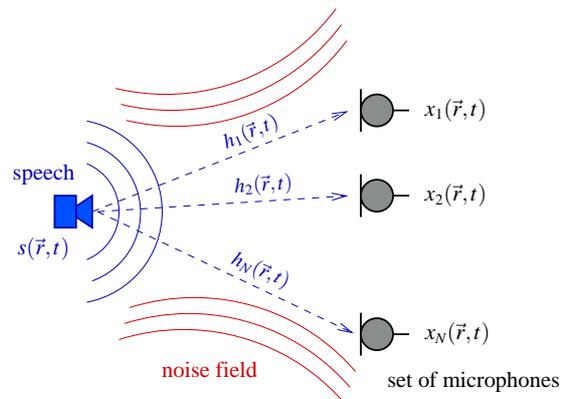


Figure 1: Sound capture in a noisy acoustical environment with N microphones (acoustic channels modeled by impulse responses $h_i(\vec{r}, t)$).

It is assumed that $h_i(\vec{r}, t)$ represents a time-invariant system. In our practical implementation of the microphone array, we estimate speaker location by time-delay estimation. Thus, $h_i(\vec{r}, t)$ is approximated by a signal delay which may vary according to speaker movements.

The discrete-time beamformer is realized with an FFT overlap-add filterbank. Therefore, we derive the MVDR beamformer algorithm based on the frequency domain multi-channel system as shown in Fig. 2.

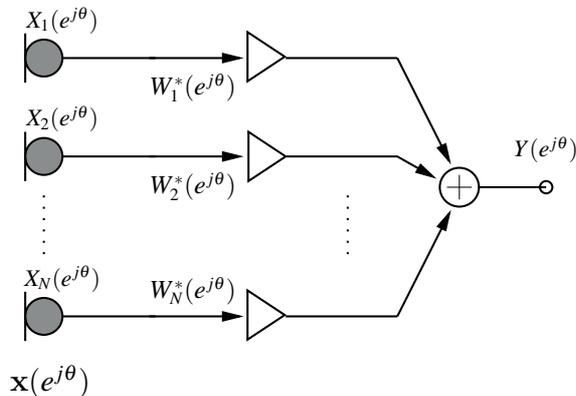


Figure 2: Beamformer in frequency domain (* denotes conjugate complex, $\theta = 2\pi\frac{f}{f_s}$ is the frequency variable).

The microphone signal spectra $X_i(e^{j\theta})$ are organized as $N \times 1$ vector $\mathbf{x}(e^{j\theta}) = [X_1(e^{j\theta}), X_2(e^{j\theta}), \dots, X_N(e^{j\theta})]^T$. Using the signal model

$$X_i(e^{j\theta}) = H_i(e^{j\theta})S(e^{j\theta}) + V_i(e^{j\theta}), \quad i = 1 \dots N, \quad (1)$$

the $N \times N$ spatio-spectral correlation matrix of the microphone signals is given by

$$\begin{aligned} \mathbf{S}_{\mathbf{xx}}(e^{j\theta}) &= E\{\mathbf{x}(e^{j\theta})\mathbf{x}^H(e^{j\theta})\} \\ &= P_s(e^{j\theta})\mathbf{h}(e^{j\theta})\mathbf{h}^H(e^{j\theta}) + \mathbf{S}_{\mathbf{vv}}(e^{j\theta}) \end{aligned} \quad (2)$$

(provided that speech is not correlated with noise). $P_s(e^{j\theta})$ is the speech power spectral density, \mathbf{h} the channel transfer function vector, and \mathbf{v} the vector of the noise spectra at the microphone inputs. Superscript H denotes conjugate-transpose, and $E\{\cdot\}$ is the expectation operation. We assume a time-stationary environment. In our practical implementation, however, $\mathbf{S}_{\mathbf{xx}}$ is estimated on a frame-by-frame basis allowing for slowly time-varying acoustical environments.

By arranging the beamformer weights as an $N \times 1$ vector $\mathbf{w}(e^{j\theta}) = [W_1(e^{j\theta}), W_2(e^{j\theta}), \dots, W_N(e^{j\theta})]^T$, the output spectrum $Y(e^{j\theta})$ can be written as

$$Y(e^{j\theta}) = \mathbf{w}^H(e^{j\theta})\mathbf{x}(e^{j\theta}). \quad (3)$$

An MVDR beamformer minimizes the output signal power under the constraint that signals from the desired direction are maintained [8]:

$$\mathbf{w}_o = \arg \min_{\mathbf{w}} \mathbf{w}^H \mathbf{S}_{\mathbf{xx}} \mathbf{w}, \quad \text{with } \mathbf{w}^H \mathbf{h} = 1. \quad (4)$$

(Frequency variable θ is omitted for clarity.) The constraint minimization (4) can be solved using Lagrange's method:

$$\nabla_{\mathbf{w}}[\mathbf{w}^H \mathbf{S}_{\mathbf{xx}} \mathbf{w} + \lambda(\mathbf{w}^H \mathbf{h} - 1)] = \mathbf{S}_{\mathbf{vv}} \mathbf{w} + \lambda \mathbf{h} = 0, \quad (5)$$

where $\nabla_{\mathbf{w}}$ is the gradient with respect to the weight vector. Note that (2) and the constraint imply $\mathbf{w}^H \mathbf{S}_{\mathbf{xx}} \mathbf{w} = P_s + \mathbf{w}^H \mathbf{S}_{\mathbf{vv}} \mathbf{w}$. Combining the constraint equation from (4) with (5) leads to the well-known solution for the optimum weight vector

$$\mathbf{w}_o = \frac{\mathbf{S}_{\mathbf{vv}}^{-1} \mathbf{h}}{\mathbf{h}^H \mathbf{S}_{\mathbf{vv}}^{-1} \mathbf{h}}. \quad (6)$$

This solution must be computed at each frequency point of the FFT filter bank. In a conventional MVDR beamformer design, a homogeneous diffuse noise field is assumed. Therefore, $\mathbf{S}_{\mathbf{vv}}^{-1}$ can be pre-computed for a given array geometry at each frequency point, and thus no matrix inversion is needed for such a noise field model. When incorporating the actual noise field in the beamformer design, $\mathbf{S}_{\mathbf{vv}}$ must be estimated resulting in a complexity $\mathcal{O}(N^3)$ of the optimum weight vector computation at each frequency point. If we estimate $\mathbf{S}_{\mathbf{vv}}$ for each signal frame with index m by

$$\begin{aligned} \mathbf{S}_{\mathbf{vv}}(e^{j\theta}, m) &= \alpha \mathbf{S}_{\mathbf{vv}}(e^{j\theta}, m-1) \\ &\quad + (1-\alpha) \mathbf{v}(e^{j\theta}, m) \mathbf{v}^H(e^{j\theta}, m), \end{aligned} \quad (7)$$

($\alpha \approx 0.8$), then $\mathbf{S}_{\mathbf{vv}}^{-1}$ could – in principle – be calculated using the matrix inversion lemma with a computational complexity of $\mathcal{O}(N^2)$. However, the matrix inversion lemma is prone to roundoff errors, especially if the matrix is ill-conditioned. Unfortunately, $\mathbf{S}_{\mathbf{vv}}$ is ill-conditioned in the low frequency range where the microphone signals are highly correlated. As a consequence, diagonal loading (regularization) of $\mathbf{S}_{\mathbf{vv}}$ is mandatory in order to get a robust beamformer. Diagonal loading, however, prevents an easy application of the matrix inversion lemma.

As an alternative, we can compute the MVDR beamformer weight vector by means of an iterative procedure. Such an algorithm has been proposed in [9]. Since the derivation presented in [9] is rather involved due to the optimization criteria used, we show that this iterative algorithm is an improved version of the classical Frost beamformer [3]. The optimum weight vector can be found iteratively with a steepest descent algorithm expressed by

$$\begin{aligned} \mathbf{w}_{k+1} &= \mathbf{w}_k - \mu \nabla_{\mathbf{w}}[\mathbf{w}_k^H \mathbf{S}_{\mathbf{xx}} \mathbf{w}_k + \lambda(\mathbf{w}_k^H \mathbf{h} - 1)] \\ &= \mathbf{w}_k - \mu (\mathbf{S}_{\mathbf{vv}} \mathbf{w}_k + \lambda \mathbf{h}), \end{aligned} \quad (8)$$

where we have used the cost function gradient from (5). Lagrange multiplier λ is obtained by substituting (8) in the constraint equation $\mathbf{h}^H \mathbf{w}_{k+1} = 1$. By eliminating λ from (8), we finally get the update equation

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu \underbrace{\left(\mathbf{I} - \frac{\mathbf{h} \mathbf{h}^H}{\|\mathbf{h}\|^2} \right)}_{\mathbf{g}_k} \mathbf{S}_{\mathbf{vv}} \mathbf{w}_k, \quad (9)$$

with $N \times N$ identity matrix \mathbf{I} . The LMS-type Frost adaptive beamformer is related to (9) if $\mathbf{S}_{\mathbf{vv}}$ is replaced by its instantaneous estimate $\mathbf{S}_{\mathbf{vv}} = \mathbf{v} \mathbf{v}^H$ and the update is carried out on a frame-by-frame basis (thus k is the frame index). In contrast, in our weight vector update we use $\mathbf{S}_{\mathbf{vv}}$ estimated with (7) and iterate in each frame (so k is not the frame index). Furthermore, convergence speed is improved by computing an optimum step size factor μ . According to [9], we choose the

step size that minimizes the noise power at the beamformer output for each iteration:

$$\frac{\partial(\mathbf{w}_{k+1}^H \mathbf{S}_{\mathbf{v}\mathbf{v}} \mathbf{w}_{k+1})}{\partial \mu^*} = 0, \quad (10)$$

(* means conjugate complex). Combining (9) and (10) results in

$$\mu_k = \frac{\mathbf{g}_k^H \mathbf{S}_{\mathbf{v}\mathbf{v}} \mathbf{w}_k}{\mathbf{g}_k^H \mathbf{S}_{\mathbf{v}\mathbf{v}} \mathbf{g}_k}. \quad (11)$$

The complete iterative beamformer weight vector algorithm is listed in Tab. 1. Although this algorithm requires a higher

- | |
|---|
| <ol style="list-style-type: none"> 1. update $\mathbf{S}_{\mathbf{v}\mathbf{v}}$ using (7) 2. apply diagonal loading $\tilde{\mathbf{S}}_{\mathbf{v}\mathbf{v}} = \mathbf{S}_{\mathbf{v}\mathbf{v}} + \varepsilon \mathbf{I}$ 3. starting solution: $\mathbf{w}_0 = \frac{\mathbf{h}}{\ \mathbf{h}\ ^2}$ 4. for each $k = 0, 1, 2, \dots$ $\mathbf{g}_k = \left(\mathbf{I} - \frac{\mathbf{h}\mathbf{h}^H}{\ \mathbf{h}\ ^2} \right) \tilde{\mathbf{S}}_{\mathbf{v}\mathbf{v}} \mathbf{w}_k$ $\mu_k = \frac{\mathbf{g}_k^H \tilde{\mathbf{S}}_{\mathbf{v}\mathbf{v}} \mathbf{w}_k}{\mathbf{g}_k^H \tilde{\mathbf{S}}_{\mathbf{v}\mathbf{v}} \mathbf{g}_k}$ $\mathbf{w}_{k+1} = \mathbf{w}_k - \mu_k \mathbf{g}_k$ 5. terminate, if $\mathbf{g}_k \approx \mathbf{0}$ |
|---|

Table 1: Iterative weight vector computation for each frame and each frequency point.

computational load than LMS-type adaptive beamformer algorithms, it offers faster convergence and an improved beam pattern since we optimize the beamformer weights for each frame. As shown by our experimental results, only a few iterations (3...6, typically) are needed to significantly improve the beam pattern, and thus the noise reduction behavior of the adaptive array. Compared to other adaptive beamformers like the GSC beamformer, the improvements are especially notable in the low frequency range.

3. BEAMFORMER/POSTFILTER COMBINATION

An MVDR beamformer as designed in the previous section reduces noise from all but the desired direction. In order to achieve an additional suppression of noise from the desired direction, we must use a different optimization criterion to calculate the optimum weight vector of Fig. 2. In frequency domain, this criterion minimizes the mean-squared error magnitude between the beamformer output spectrum $Y(e^{j\theta})$, and the desired speech spectrum $S(e^{j\theta})$:

$$\mathbf{w}_o = \arg \min_{\mathbf{w}} E \{ |(S(e^{j\theta}) - \mathbf{w}^H(e^{j\theta}) \mathbf{x}(e^{j\theta}))|^2 \}. \quad (12)$$

Minimization of this error cost function leads to the Wiener solution

$$\mathbf{w}_o(e^{j\theta}) = \underbrace{E \{ \mathbf{x}(e^{j\theta}) \mathbf{x}^H(e^{j\theta}) \}^{-1}}_{\mathbf{S}_{\mathbf{x}\mathbf{x}}^{-1}(e^{j\theta})} \underbrace{E \{ \mathbf{x}(e^{j\theta}) S^*(e^{j\theta}) \}}_{\mathbf{S}_{\mathbf{x}s}(e^{j\theta})}. \quad (13)$$

Using (2) and $\mathbf{S}_{\mathbf{x}s} = P_s \mathbf{h}$, this solution can be written as

$$\mathbf{w}_o = \mathbf{S}_{\mathbf{x}\mathbf{x}}^{-1} \mathbf{S}_{\mathbf{x}s} = (P_s \mathbf{h}\mathbf{h}^H + \mathbf{S}_{\mathbf{v}\mathbf{v}})^{-1} P_s \mathbf{h}, \quad (14)$$

where the frequency variable θ is omitted again for clarity. Application of the matrix inversion lemma results in the factorization

$$\mathbf{w}_o = \underbrace{\frac{\mathbf{S}_{\mathbf{v}\mathbf{v}}^{-1} \mathbf{h}}{\mathbf{h}^H \mathbf{S}_{\mathbf{v}\mathbf{v}}^{-1} \mathbf{h}}}_{\mathbf{w}_{bf}, \text{ beamformer}} \underbrace{\frac{P'_s}{P'_s + P'_v}}_{\mathbf{w}_p, \text{ postfilter}}, \quad (15)$$

with P'_s (P'_v) denoting the spectral power density of speech (noise) at the beamformer output [1]. The beamformer/postfilter combination of the optimum multi-channel noise reduction system is shown in Fig. 3.

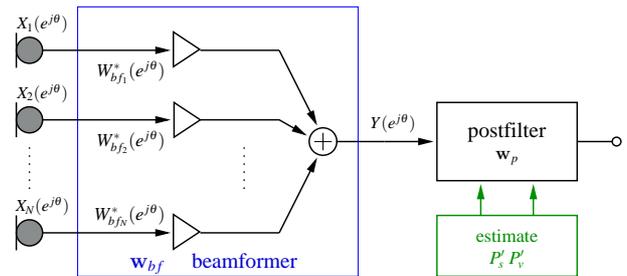


Figure 3: Beamformer/postfilter in frequency domain.

The cascade connection of a beamformer, and a single-channel noise reduction system is very attractive. If we are able to design a good beamformer matched to the noise field properties, then the postfilter operates at a much higher input signal-to-noise ratio (SNR). Consequently, speech distortion and residual noise (musical tones) are definitely less pronounced compared to the case with no beamformer pre-processing. In addition, we can apply highly advanced noise suppression algorithms for the post-filter.

In our adaptive beamformer, the iterative algorithm of Tab. 1 is used because we want to match the beamformer behavior to the actual noise field and to avoid a sub-optimal solution by assuming a diffuse noise field. For the selection of the postfilter there are several choices [1, 2]. The majority of algorithms is based on the postfilter expression in (15)

$$\mathbf{w}_p(e^{j\theta}) = \frac{P'_s(e^{j\theta})}{P'_s(e^{j\theta}) + P'_v(e^{j\theta})} = \frac{P'_s(e^{j\theta})}{P_y(e^{j\theta})} \quad (16)$$

and on estimation of the spectral power density $P'_s(e^{j\theta})$ by using $\mathbf{S}_{\mathbf{x}\mathbf{x}}(e^{j\theta})$, $\mathbf{S}_{\mathbf{v}\mathbf{v}}(e^{j\theta})$, and (2). Since (2) exhibits $\frac{(N-1)N}{2}$ equations to calculate P'_s from $S_{x_i x_j}$, averaging can be applied to obtain a smooth least-squares estimate

$$\hat{P}'_s = \Re e \left\{ \frac{\sum_{i=1}^N \sum_{j=i+1}^N (\hat{S}_{x_i x_j} - \hat{S}_{v_i v_j})}{\sum_{i=1}^N \sum_{j=i+1}^N H_i H_j^*} \right\} \quad (17)$$

(channel transfer function H_i , omitting θ). Matrix elements $\hat{S}_{x_i x_j}$, $\hat{S}_{v_i v_j}$ may be estimated with a recursive algorithm like (7). A speech pause detection is needed to update noise estimates $\hat{S}_{v_i v_j}$. As shown in [2], speech pause detection can be avoided, if a homogeneous diffuse noise field is assumed. The estimation of P'_s with (17) is straight forward. However, due to fluctuations of the matrix elements, negative values of P'_s may occur and must be eliminated by introducing a lower

bound equal to zero or to some small spectral floor. Nonetheless, we may notice the typical musical noise and speech distortion of single-channel Wiener filters, if the input SNR is below approximately 6 dB. Less speech distortion and musical noise can be achieved with more advanced postfilters, like Ephraim-Malah spectral magnitude estimators [10], or recently published efficient variations thereof [11, 12].

4. EXPERIMENTAL RESULTS

The adaptive microphone array is implemented with an overlap-add multi-input 512 point FFT filterbank, and a sampling frequency $f_s = 16$ kHz. Signal frames are obtained by $L = 512$ point Hann windowing applied to the input signals. A frame hop size equal to $L/4 = 128$ results in a four times filterbank oversampling. For each FFT bin in the frequency range 200 Hz ... 6600 Hz, the optimum beamformer weight vector is computed by means of Tab. 1. The upper cut-off frequency is needed to avoid spatial aliasing of the $N = 8$ channel array with a geometry as shown in Fig. 4.

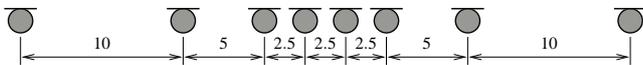


Figure 4: Microphone array geometry (dimensions in cm).

The channel impulse responses are approximated by delays τ_i matched to the desired speaker direction. Thus, the channel transfer functions are given by $H_i(e^{j\theta}) = e^{-j\theta f_s \tau_i}$, $i = 1, 2, \dots, N$. In our implementation, delays are either computed for a given direction of arrival or are estimated using the phase transform (PHAT) algorithm [13].

The postfilter is based on the Ephraim-Malah spectral amplitude modifiers. An improved minimum statistics algorithm [14] is employed for noise spectral density estimation. In addition, part of this algorithm is also used for robust speech pause detection needed to estimate S_{vv} for each signal frame. The minimum statistics algorithm requires a higher computational load as compared with basic speech activity detectors. However, it offers a significantly better performance at low input SNRs, and in case of nonstationary acoustical noise.

For evaluation of the proposed beamformer/postfilter combination, a test setup has been installed where the microphone array is placed in the middle of a large office room. This acoustical environment exhibits a measured frequency-averaged reverberation time of 0.84 seconds. Speaker direction is perpendicular to the array axis (broadside direction). A single noise source with approximate $1/f$ spectral power density is emitting at an angle of 25° (measured from the array axis). Due to the strong reverberation, there is also a diffuse noise component giving rise to a mixture of unidirectional and diffuse acoustical noise.

Besides listening tests, we use the enhancement of the segmental SNR (SegSNRE) as a speech quality measure. The SegSNRE in dB is the difference in segmental SNR between the output signal of the beamformer/postfilter combination and the noisy microphone signals. A representative result is shown in Fig. 5. In the high noise region, the beamformer with S_{vv} estimation and iterative weight vector computation is clearly superior to a design based on a diffuse noise field. This is also true, if the proposed algorithm is compared to other beamformer/postfilter algorithms based

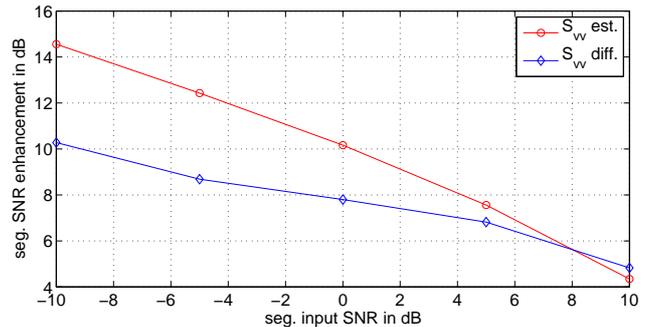


Figure 5: Enhancement of segmental SNR (SegSNRE) in dB of the proposed beamformer/postfilter combination (mixture of unidirectional and diffuse acoustical noise).

on the diffuse noise field assumption. In all experiments only 6 iteration are used for weight vector computation in each signal frame. A comprehensive comparison including formal listening tests of various known postfilter algorithms has been carried out in a diploma thesis [15]. In this study, the algorithm proposed in [2] performs best in case of diffuse noise fields and moderate noise (SNR > 5 dB). However, incorporating the estimated S_{vv} in the weight vector computation and using an Ephraim-Malah postfilter gives a better noise suppression and less speech distortion as compared to the beamformer/postfilter combination investigated in [2].

As an illustrative example, logarithmically scaled spectrograms are shown in Fig. 6, and Fig. 7, respectively.

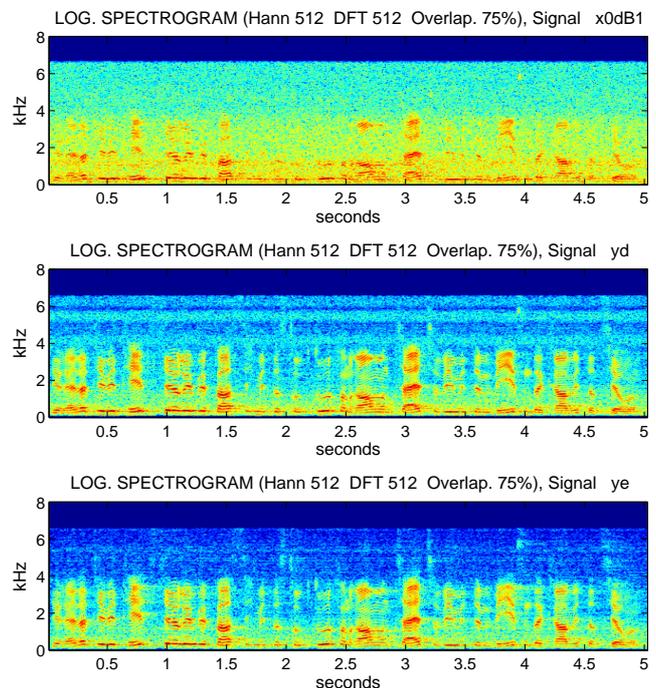


Figure 6: Log. spectrogram of noisy speech signal at microphone #1 (above), at beamformer output designed with diffuse noise assumption (middle), and at beamformer output with estimated S_{vv} (below), input seg. SNR = 0 dB.

The upper image in Fig. 6 is the spectrogram of the noisy

speech measured at microphone array channel #1. The spectrogram in the middle of Fig. 6 is obtained at the output of the beamformer designed with diffuse S_{VV} . There is substantially more noise as compared to a design with estimation of S_{VV} and application of Tab. 1 (lower picture in Fig. 6).

The effect of the postfilter is illustrated in Fig. 7. There is much less noise in case of the proposed beamformer design. A closer look to the lower spectrogram unveils virtually no musical noise phenomenon and only a slight speech distortion. Interested readers are invited to visit the author's homepage and listen to the particular signals of this example.

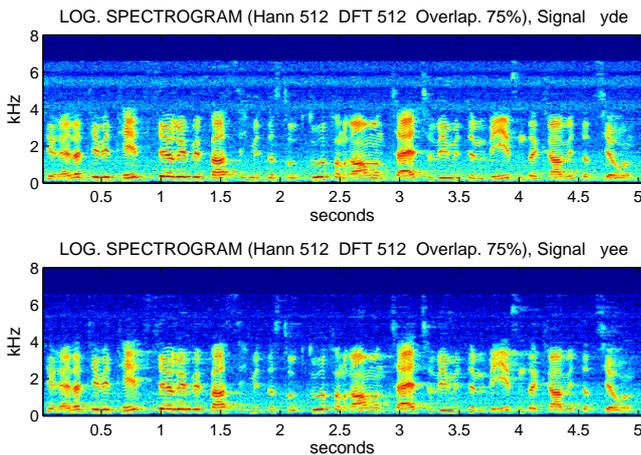


Figure 7: Log. spectrogram of beamformer + postfilter output with diffuse noise assumption (above), and with estimated S_{VV} (below), input seg. SNR = 0 dB.

5. CONCLUSIONS

We have presented an adaptive microphone array consisting of a beamformer/postfilter combination. The beamformer weights are optimized for each signal frame according to a spatio-spectral correlation matrix estimation of the disturbing noise field. Taking into account the actual noise field parameters results in an improved noise suppression of the beamformer as compared with beamformer designs assuming a diffuse noise field. Using the proposed beamformer as a pre-processor to a single-channel Ephraim-Malah noise reduction system yields an overall performance with negligible musical noise and speech distortion, even at segmental input SNRs less than 5 dB. The beamformer algorithm requires a higher computational load as compared to GSC beamformers. Nevertheless, the whole system is capable for real-time operation with 16 kHz sampling frequency on today's signal processing hardware.

Acknowledgement

The author would like to thank P. Fertl for supplying the sound recordings, and for the comprehensive study of various postfilter algorithms in his diploma thesis.

REFERENCES

[1] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in *Microphone arrays*, M. Brandstein, and

D. Ward (Eds.), Springer-Verlag, Berlin Heidelberg New York, 2001, ch. 3, pp. 39–60.

[2] I. A. McCowan, and H. Boullard, "Microphone array post-filter based on noise field coherence," *IEEE Trans. Speech Audio Processing*, vol. 11, pp. 709–716, Nov. 2003.

[3] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, pp. 926–935, Aug. 1972.

[4] 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 Processing*, vol. 47, pp. 2677–2684, Oct. 1999.

[5] W. Herboldt, W. Kellermann, "Efficient frequency-domain realization of robust generalized sidelobe cancellers," *IEEE Workshop on Multimedia Signal Processing*, pp. 377–382, Cannes, France, Oct. 2001.

[6] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Processing*, vol. 49, pp. 1614–1626, Oct. 2001.

[7] S. Gannot, and I. Cohen, "Speech Enhancement based on the general transfer function GSC and postfiltering," *IEEE Trans. Speech Audio Processing*, vol. 12, pp. 561–571, Nov. 2004.

[8] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, pp. 1365–1376, Oct. 1987.

[9] D. A. Pados, and G. N. Karystinos, "An iterative algorithm for the computation of the MVDR filter," *IEEE Trans. Signal Processing*, vol. 49, pp. 290–300, Feb. 2001.

[10] Y. Ephraim, and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, pp. 1109–1121, Dec. 1984.

[11] P. J. Wolfe, and S. J. Godsill, "Efficient alternatives to the Ephraim and Malah suppression rule for audio signal enhancement," *EURASIP Journal on Appl. Sig. Processing*, vol. 10, pp. 1043–1051, 2003.

[12] T. Lotter, and P. Vary, "Noise Reduction by Joint Maximum a Posteriori Spectral Amplitude and Phase Estimation with Super-Gaussian Speech Modelling," in *Proc. EUSIPCO 2004*, Vienna, Austria, Sept. 4-8, 2004, pp. 1447–1460.

[13] C. R. Knapp, and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 24, pp. 320–327, Aug. 1976.

[14] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Processing*, vol. 9, pp. 504–512, Jul. 2001.

[15] P. Fertl, "Mikrofonarray mit adaptivem Postfilter zur Sprachsignalentstörung," *Diploma Thesis*, Vienna University of Technology, Aug. 2005, (in German).