A PRACTICAL BEAMFORMER-POSTFILTER SYSTEM FOR ADAPTIVE SPEECH ENHANCEMENT IN NON-STATIONARY NOISE ENVIRONMENTS

Tobias Wolff, Markus Buck

Nuance Communications Aachen GmbH, D-89077 Ulm, Germany.
Acoustic Speech Enhancement Research
Email: {tobias.wolff, markus.buck}@nuance.com

ABSTRACT

In this contribution we present an adaptive beamformer-postfilter system which can be used to suppress non-stationary noises. The emphasis lies on the spatial filtering property of the proposed postfilter. Due to the spatial characteristic of the postfilter, an interfering speaker in background noise can be suppressed effectively while maintaining the quality of the desired speech using only two microphones.

Index Terms— Spatial Postfilter, Beamforming, Multi-Channel Speech Enhancement, Non-Stationary Noise.

1. INTRODUCTION

In [1] we have introduced a postfilter transfer function that generalizes some known postfilters such as those proposed by Zelinski, Simmer, McCowan, Leukimmiatis etc.. The generalized postfilter that is optimal in the MMSE-sense can be written as:

$$H_{\mathsf{pf}} = \max\left[\frac{\Phi_{xx}}{\Phi_{aa}}\left(1 - \frac{\mathrm{tr}\{\mathbf{B}\mathbf{J}_{xx}\mathbf{B}^H\}}{\mathrm{tr}\{\mathbf{B}\mathbf{J}_{vv}\mathbf{B}^H\}}\right)\frac{\mathsf{G}_{\mathsf{bm}}}{\mathsf{G}_{\mathsf{bm}} - 1}, 0\right]. \qquad (1)$$

Here, Φ_{xx} denotes power spectral density (PSD) at the microphones and Φ_{aa} is the PSD at the beamformer output. The term in the middle determines the spatial characteristic of the filter as it depends on the coherence matrix of the entire sound field $\mathbf{J}_{xx}$ as well as the one of the noise $\mathbf{J}_{vv}$. The matrix $\mathbf{B}$ can be designed to match the postfilter to any given beamformer: If $\mathbf{B}$ is orthogonal to the LCMV constraint matrix $\mathbf{C}_{\mathsf{bf}}$, hence $\mathbf{B}\mathbf{C}_{\mathsf{bf}} = \mathbf{0}$, then $\mathbf{B}$ becomes a blocking matrix and the postfilter implements the same constraints as the beamformer. The third part of the transfer function denotes the influence of the blocking matrix gain G_{bm}. Further details can be found in [1, 2].

2. SPATIAL CHARACTERISTIC

The maximum operator in Eq. 1 is motivated by the fact that its first argument may become negative, which is not a reasonable solution. In particular, this plays a role once $\mathbf{J}_{xx}$ does not match the assumed noise coherence $\mathbf{J}_{vv}$. With the coherence function of a plane wave $J_{\shortparallel}(e^{j\Omega}, \theta) := \exp(-j\Omega f_{\mathsf{s}} d_{\mathsf{mic}} \cos(\theta)/c)$, where θ is the angle of incidence, c is the speed of sound, and f_{s} denotes the sampling frequency, the effect can be visualized using the postfilter beampattern

$$\Upsilon_{\mathsf{pf}}(\Omega, \theta) := \left|H_{\mathsf{pf}}(e^{j\Omega})\right|^2\Big|_{\mathbf{J}_{xx} = \mathbf{J}_{\shortparallel}(e^{j\Omega}, \theta)}. \qquad (2)$$

In Fig. 1 two examples of this function are depicted for $M = 4$ microphones, $\mathbf{B} = \mathbf{I}_{M\times M} - \frac{1}{M}\mathbf{1}_{M\times M}$ and an assumed diffuse noise field, hence $\mathbf{J}_{vv} = \mathrm{si}(\Omega f_{\mathsf{s}} d_{\mathsf{mic}}/c)$. The upper plot corresponds to the transfer function without the maximum operator, whereas the beampattern in the lower plot includes the maximum operator. As can be seen from this example, due to the maximum operator, the postfilter fully suppresses coherent sounds which have not been protected by the respective $M \times K$ constraint matrix $\mathbf{C}_{\mathsf{pf}}$ (preferably $\mathbf{C}_{\mathsf{pf}} = \mathbf{C}_{\mathsf{bf}}$). In this example $\mathbf{C}_{\mathsf{pf}} = [1\ 1\ 1\ 1]^T$ is used for simplicity ($K = 1$ spatial constraints). This spatial filtering property is an important feature of the postfilter presented here as it enables the suppression of non-stationary interferences such as undesired speech while preserving the quality of the desired signal.

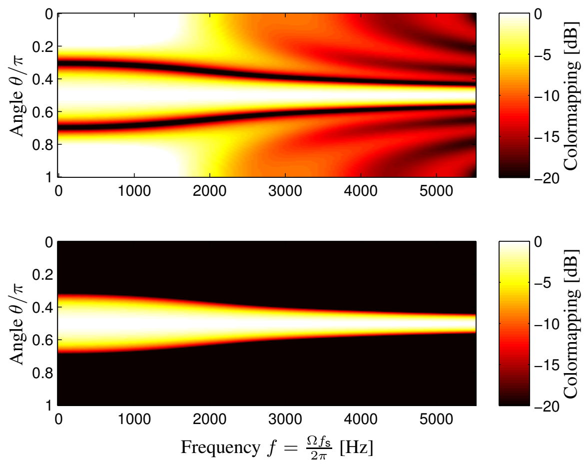

Fig. 1. Postfilter beampattern $\Upsilon_{\mathsf{pf}}(\Omega, \theta)$ with (bottom) - and without (top) the max-operator for $M = 4$ and $d_{\mathsf{mic}} = 5$ cm.

3. IMPLEMENTATION

The presented beamformer-postfilter system consists of an adaptive GSC-type beamformer and a fully adaptive postfilter. The beamformer and the postfilter share the blocking matrix, meaning that the postfilter is ideally matched to the beamformer ($\mathbf{C}_{\mathsf{pf}} = \mathbf{C}_{\mathsf{bf}}$). To minimize speech distortions, we use an NLMS-adaptive subband implementation of the blocking matrix proposed in [3]. In particular, four-tap subband filters are used to minimize speech leakage. The blocking matrix gain G_{bm} is then considered to be infinite, so there is no need to estimate it (see also [2]). As a consequence, the residual speech that practically passes the blocking matrix is also treated as interference which results in a dereverberating effect.

Eq. 1 states that the sum of all PSDs at the blocking matrix output $\Phi^{+}_{uu} = \mathrm{tr}\{\mathbf{B}\mathbf{\Phi}_{xx}\mathbf{B}^H\}$ has to be equalized by

$$\mathsf{G}_{\mathsf{eq}} = \mathrm{tr}\left\{\mathbf{B}\mathbf{J}_{vv}\mathbf{B}^H\right\}^{-1} = \Phi_{vv}/\Phi^{+}_{uu}. \quad (3)$$

to make up for the coloration introduced through $\mathbf{B}$. Here, Φ_{vv} is the noise-PSD at the microphones. The equalizer G_{eq}, however, is not known in general and must therefore be estimated. In our implementation we do so by temporal averaging of Φ_{vv}/Φ^{+}_{uu} during speech pauses. It is thus necessary to use voice activity detection (VAD), which is implemented similar to the *spatial* VAD proposed in [3]. A block diagram of the presented system is depicted in Fig. 2.

As long as the noise coherence changes sufficiently slow, it is possible to obtain unbiased estimates for the PSD of non-stationary interferences by $\hat{\Phi}_{vv} = \hat{\mathsf{G}}_{\mathsf{eq}} \cdot \hat{\Phi}^{+}_{uu}$ which can then be subtracted from $\hat{\Phi}_{xx}$ to obtain $\hat{\Phi}_{ss}$. For mixed noise fields the noise is likely to be overestimated leading to increased speech distortions. Experiments have shown, however, that this method allows to suppress at least a single interfering speaker in background noise without introducing large signal distortions. Fig. 3 finally shows an example of this spatial filtering property with only $M = 2$ microphones with a spacing of $d_{\mathsf{mic}} = 10.5$ cm. The signals have been recorded in a car environment, whereas the microphones were integrated in the rear-view mirror. The filter attenuation has been limited to 10 dB.

The described beamformer-postfilter system can be demonstrated in a real-time demo that also uses two microphones with $d_{\mathsf{mic}} = 6$ cm spacing. As it is often encountered in hands-free systems, we use cardioid microphones in a broadside configuration. The sampling frequency is $f_{\mathsf{s}} = 11025$ Hz and a Hann-windowed 256-point FFT is used to produce the subband signals with a frameshift of 64 samples.

4. CONCLUSIONS

We present a fully adaptive beamformer-postfilter system which can be used to suppress non-stationary noises. The postfilter is combined with an adaptive GSC-type beamformer in a highly efficient manner. The basic spatial filtering property of the postfilter has been pointed out with the help of the postfilter beampattern and can be demonstrated in a real-time demo. It cannot be claimed that the proposed systems works robustly in arbitrary noise-fields with rapidly changing coherence functions. For the practically important case of one interfering speaker in background noise, the proposed implementation does produce good results.

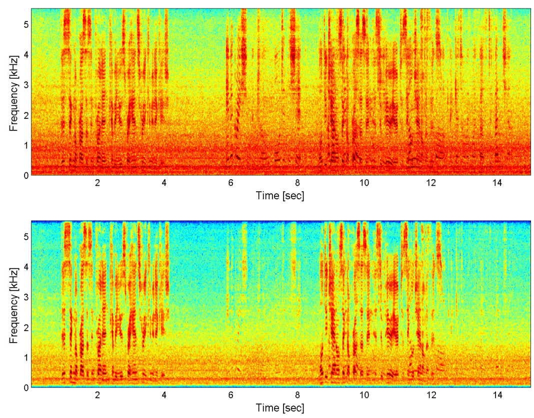

Fig. 3. Top: spectrogram of a microphone signal recorded in car environment. From 6 to 14 sec the co-driver interferes with the driver, to whom the system is steered to. Bottom: Output signal of the proposed beamformer-postfilter system.

5. REFERENCES

[1] T. Wolff and M. Buck: "A generalized view on microphone array postfilters", *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, 2010.

[2] T. Wolff and M. Buck: "Influence of blocking matrix design on microphone array postfilters", *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, 2010.

[3] O. Hoshuyama and A. Sugiyama: "Robust Adaptive Beamforming", in: *Microphone Arrays*, Springer, Berlin, Heidelberg, New York, 2001.

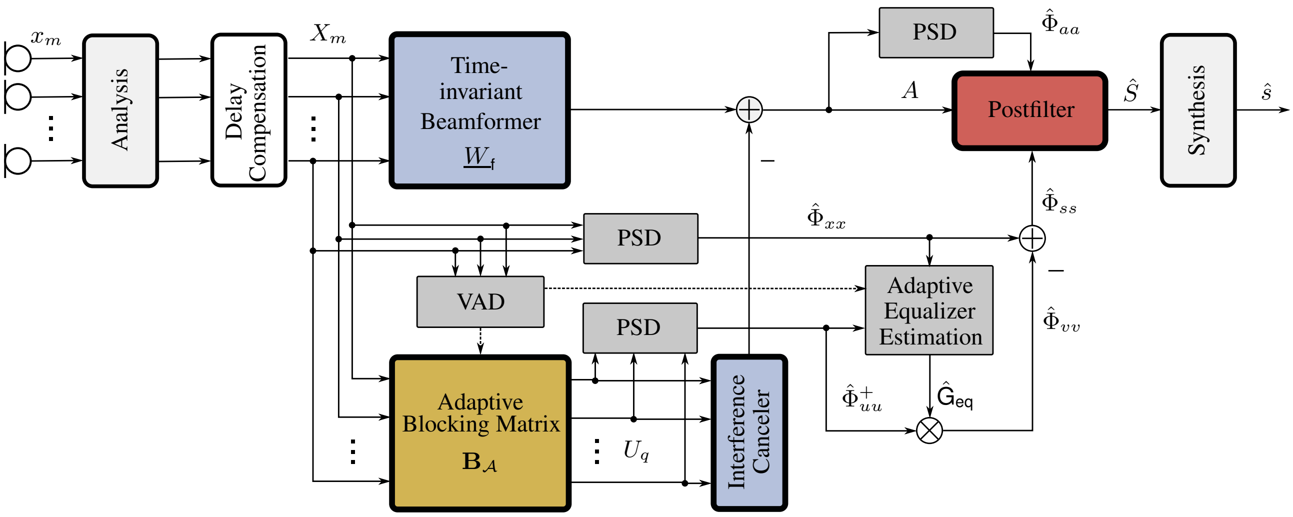

Fig. 2. Blockdiagram of the adaptive beamformer-postfilter system.