FREQUENCY-DOMAIN CRITERION FOR SPEECH DISTORTION WEIGHTED MULTI-CHANNEL WIENER FILTERING FOR ROBUST NOISE REDUCTION

Simon Doclo[1], *Ann Spriet*[1,2], *Marc Moonen*[1] *Jan Wouters*[2]

[1]Katholieke Universiteit Leuven
Dept. of Electrical Engineering (ESAT-SCD)
Kasteelpark Arenberg 10, 3001 Leuven, Belgium
Email: {doclo,spriet,moonen}@esat.kuleuven.ac.be

[2]Katholieke Universiteit Leuven
Laboratory for Exp. ORL
Kapucijnenvoer 33, 3000 Leuven, Belgium
Email: jan.wouters@uz.kuleuven.ac.be

1. INTRODUCTION

In [1], a generalised noise reduction scheme, called the spatially pre-processed speech distortion weighted multichannel Wiener filter (SP-SDW-MWF), has been presented, which consists of a fixed spatial pre-processor and an adaptive SDW-MWF stage. By taking speech distortion explicitly into account in the design criterion of the adaptive stage, the SP-SDW-MWF adds robustness to the standard generalised sidelobe canceller (GSC). In [1], it has been shown that, compared to the GSC with quadratic inequality constraint [4], the SP-SDW-MWF achieves a better noise reduction performance for a given maximum speech distortion level. In [2] cheap (time-domain and frequency-domain) stochastic gradient algorithms have been presented, which however require large data buffers. In [3] an adaptive frequency-domain algorithm for the SDW-MWF has been presented using diagonal correlation matrices, reducing the memory usage and the computational complexity.
In this contribution, we present a novel *frequency-domain criterion* for the SDW-MWF, trading off noise reduction and speech distortion. This frequency-domain criterion for multichannel speech enhancement is an extension of the criterion used in [5] for multichannel echo cancellation. Using the proposed criterion, existing and novel adaptive frequency-domain algorithms for the SDW-MWF can be derived. The noise reduction performance and the robustness against microphone mismatch of the proposed algorithms are illustrated using experimental results for a small-sized microphone array in a hearing aid.

2. SPATIALLY PRE-PROCESSED SDW-MWF (SP-SDW-MWF)

The SP-SDW-MWF [1] is depicted in Fig. 1a. It consists of a fixed spatial pre-processor, i.e. a fixed beamformer and a blocking matrix, and an adaptive stage. Note that the structure of the SP-SDW-MWF strongly resembles the GSC [6, 7], where the difference lies in the fact that an adaptive SDW-MWF is used in the adaptive stage and that it is possible to include an extra filter $\mathbf{w}_0$ on the speech reference. The SDW-MWF takes speech distortion due to speech leakage into the noise references explicitly into account in the design criterion of the filter $\mathbf{w}[k]$ and minimises the weighted sum of the residual noise energy $\varepsilon_v^2[k]$ and the speech distortion energy $\varepsilon_x^2[k]$, i.e.

$$J(\mathbf{w}[k]) = \varepsilon_v^2[k] + \frac{1}{\mu}\varepsilon_x^2[k] = \mathcal{E}\Big\{\big|v_0[k-\Delta] - \mathbf{w}^T[k]\mathbf{v}[k]\big|^2\Big\} + \frac{1}{\mu}\mathcal{E}\Big\{\big|\mathbf{w}^T[k]\mathbf{x}[k]\big|^2\Big\}, \tag{1}$$

where $v_0[k]$ is the noise component of the speech reference, $\mathbf{v}[k]$ is the stacked noise component vector, and $\mu \in [0, \infty]$ provides a trade-off between noise reduction and speech distortion [1, 8]. Depending on the setting of μ and the presence/absence of the filter $\mathbf{w}_0$, different algorithms are obtained: (1) without filter $\mathbf{w}_0$, we obtain the *speech distortion regularised GSC* (SDR-GSC), where the optimisation criterion of the GSC is supplemented with a regularisation term $1/\mu\,\varepsilon_x^2$. For $\mu = \infty$, speech distortion is completely ignored, which corresponds to the standard GSC; (2) with filter $\mathbf{w}_0$, we obtain the SP-SDW-MWF. For $\mu = 1$, the output signal is the MMSE estimate of the delayed speech component $x_0[k-\Delta]$ in the speech reference.

3. FREQUENCY-DOMAIN CRITERION FOR SDW-MWF

By using a block notation, we can define a *frequency-domain criterion* similar to (1),

$$J_f[m] = (1-\lambda_v)\sum_{i=0}^{m}\lambda_v^{m-i}\underline{\mathbf{e}}_v^H[i]\,\underline{\mathbf{e}}_v[i] + \frac{1}{\mu}(1-\lambda_x)\sum_{i=0}^{m}\lambda_x^{m-i}\underline{\mathbf{e}}_x^H[i]\,\underline{\mathbf{e}}_x[i], \tag{2}$$

where $\underline{\mathbf{e}}_v[i]$ represents the residual noise in the frequency-domain, $\underline{\mathbf{e}}_x[i]$ represents the speech distortion in the frequency-domain, and λ_v and λ_x are exponential forgetting factors respectively for noise and speech. By using a similar derivation as in [5], we obtain an update formula for the filter $\underline{\mathbf{w}}_{2NL}[m]$ in the frequency-domain, making use of frequency-domain speech and noise correlation matrices. By approximating these correlation matrices by (block-)diagonal matrices, and by either using

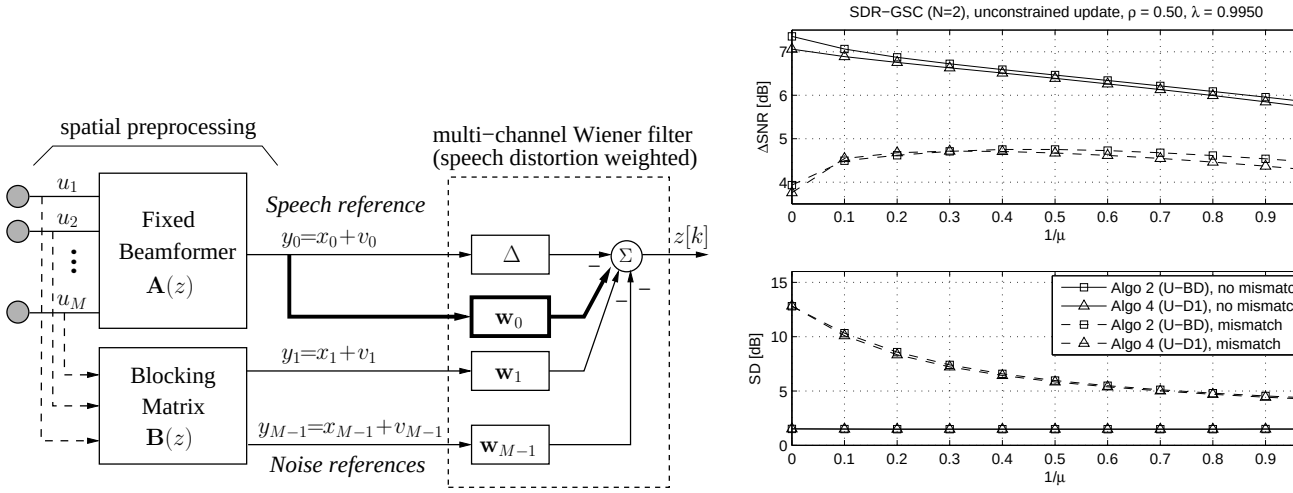


Figure 1: (a) Structure of the SP-SDW-MWF. (b) SNR improvement and speech distortion of SDR-GSC in function of trade-off parameter $1/\mu$, without and with gain mismatch (unconstrained update formula, $\rho = 0.5$, $\lambda = 0.995$).

a constrained or an unconstrained update formula, we obtain several adaptive algorithms for implementing the SDW-MWF in the frequency-domain, having a different performance and computational complexity.

4. EXPERIMENTAL RESULTS

We performed simulations using recordings with a small-sized microphone array mounted on a behind-the-ear hearing aid in an office room with reverberation time $T_{60\text{dB}} \approx 700\,\text{ms}$. The desired speech source at 0° consists of sentences, and a complex noise scenario consisting of 5 spectrally non-stationary multi-talker noise sources at 75°, 120°, 180°, 240° and 285° is present. The sampling frequency is $16\,\text{kHz}$ and the input SNR of the microphone signals is $0\,\text{dB}$. Figure 1b plots the SNR improvement and the speech distortion of the SDR-GSC, using the unconstrained update formula with block-diagonal (U-BD) and diagonal step size matrix (U-D1), as a function of the trade-off parameter $1/\mu$. This figure also depicts the effect of a gain mismatch of $4\,\text{dB}$ at the second microphone. In the absence of microphone mismatch, the amount of speech leakage into the noise references is limited, such that the speech distortion is small for all $1/\mu$. However, since there is some speech leakage present due to reverberation, the SNR improvement decreases for increasing $1/\mu$. In the presence of microphone mismatch, the amount of speech leakage into the noise references grows. For the standard GSC, i.e. $1/\mu = 0$, significant speech distortion now occurs and the SNR improvement is seriously degraded. Setting $1/\mu > 0$ improves the performance of the GSC in the presence of signal model errors, i.e. speech distortion decreases and the SNR degradation becomes smaller.

References

[1] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Processing*, vol. 84, pp. 2367–2387, Dec. 2004.

[2] A. Spriet, M. Moonen, and J. Wouters, "Stochastic gradient implementation of spatially pre-processed multi-channel Wiener filtering for noise reduction in hearing aids," in *Proc. IEEE ICASSP*, 2004, pp. 57–60.

[3] S. Doclo, A. Spriet, and M. Moonen, "Efficient frequency-domain implementation of speech distortion weighted multi-channel Wiener filtering for noise reduction," in *Proc. EUSIPCO*, 2004, pp. 2007–2010.

[4] 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.

[5] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, "General derivation of frequency-domain adaptive filtering," chapter 8 in *Advances in Network and Acoustic Echo Cancellation*, pp. 157–176, Springer-Verlag, 2001.

[6] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propagat.*, vol. 30, pp. 27–34, Jan. 1982.

[7] W. Herbordt and W. Kellermann, "Adaptive beamforming for audio signal acquisition," chapter 6 in *Adaptive Signal Processing: Applications to Real-World Problems*, J. Benesty and Y. Huang, Eds.), pp. 155–194, Springer-Verlag, 2003.

[8] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Processing*, vol. 50, pp. 2230–2244, Sept. 2002.