Multi-Microphone Noise Reduction - Theoretical Optimum and Practical Realization

K. Uwe Simmer*, Joerg Bitzer**
*ASP Acoustic Signal Processing GmbH, Wachtstr. 17, 28195 Bremen; Email: uwe.simmer@t-online.de
**Houpert Digital Audio, Anne-Conway-Str. 1, 28359 Bremen; Email: j.bitzer@hda.de

Introduction

In speech processing systems with large speaker to microphone distance such as hearing aids or hands-free telephony the recorded speech signal is often heavily corrupted by additive acoustic background noise. Microphone arrays can significantly improve the received speech quality by extracting the desired speech source and suppressing disturbing background noise and reverberation. High performance array processing algorithms exploit the spatial characteristics of the sound field as well as the time and frequency dependent SNR.

Optimum solution in the minimum mean-square error (MMSE) sense

In the sequel all multichannel signals and systems are presented in the frequency domain using complex vector notation for the sake of compactness. All equations represent just a single frequency bin. The frequency index is generally omitted. The superscript T denotes the transpose operator, the superscript * denotes the complex conjugation and the superscript H the conjugate transpose.

We assume that the input vector $\mathbf{x} = [x_0, x_1, \ldots, x_N]^T$ (frequency domain version of the input signal recorded by the N microphones) consists of a single desired source signal s which is transformed by the acoustic path $\mathbf{d}$ (propagation vector) and is corrupted by an additive multichannel noise source $\mathbf{v}$:

$$\mathbf{x} = s\mathbf{d} + \mathbf{v}. \qquad \text{eq. 1}$$

The output signal y of a general linear array processor is given by

$$y = \mathbf{w}^H \mathbf{x}. \qquad \text{eq. 2}$$

The optimum weight vector provides the estimate y of a the desired signal s by minimising the mean square of the error $e = s - y$. The best possible solution in the MMSE sense is given by the multichannel Wiener filter or Wiener-Hopf equation in matrix form:

$$\mathbf{w}_{\text{opt}} = \mathbf{\Phi}_{\text{xx}}^{-1} \boldsymbol{\phi}_{xs} \qquad \text{eq. 3}$$

where $\mathbf{\Phi}_{\text{xx}} = E\left[\mathbf{x}\,\mathbf{x}^H\right]$ is the correlation matrix of the input signal and $\boldsymbol{\phi}_{xs} = E\left[\mathbf{x}\, s^*\right]$ is the cross-correlation vector of the noisy input and the desired signal. Eq. 3 is also known from the context of optimum FIR filters in the time-domain. This is not really surprising since array processors and FIR filters are closely related: in both cases the output is a weighted sum of delayed samples.

Assuming that the speech and noise signals are uncorrelated, and inserting the propagation model (eq. 1), optimum weight becomes:

$$\mathbf{w}_{\text{opt}} = [\phi_{ss}\mathbf{d}\mathbf{d}^H + \mathbf{\Phi}_{vv}]^{-1} \phi_{ss}\mathbf{d}. \qquad \text{eq. 4}$$

By applying the matrix inversion lemma (see [3], [11] for details) the multi-channel Wiener filter can be written as a product of a spatial filter depending on the noise correlation matrix and a one-dimensional Wiener filter depending on the output SNR of the spatial filter:

$$\mathbf{w}_{opt} = \underbrace{\frac{\phi_{s_o s_o}}{\phi_{s_o s_o} + \phi_{v_o v_o}}}_{\text{Wiener Postfilter H}} \cdot \underbrace{\frac{\mathbf{\Phi}_{vv}^{-1}\mathbf{d}}{\mathbf{d}^H \mathbf{\Phi}_{vv}^{-1}\mathbf{d}}}_{\text{MVDR Beamformer }\mathbf{F}} \qquad \text{eq. 5}$$

The spatial weight vector $\mathbf{F}$ is the well-known Minimum Variance Distortionless Response (MVDR) beamformer. It provides an MMSE estimate of the desired signal under the constraint of a distortionless look direction response. The Frost beamformer and the Generalized Sidelobe Canceller (GSC) are adaptive realisations of the MVDR beamformer. A direct frequency domain realisation can be implemented by estimating the noise correlation matrix during speech pauses. Furthermore, the MVDR beamformer implements a superdirective array if a fixed diffuse noise correlation matrix is assumed. In any case, a practical design should be constrained to limit the sensitivity against sensor errors and spatial white noise [1].

Residual noise and reverberation that is not cancelled by the MVDR beamformer can be suppressed by the single channel Wiener filter H. Therefore, the multichannel Wiener filter provides a significantly higher output SNR than the MVDR beamformer alone. The inevitable linear distortion introduced by the Wiener filter can be minimized by carefully designed postfilter algorithms that implement a reasonable compromise between signal distortion and noise suppression.

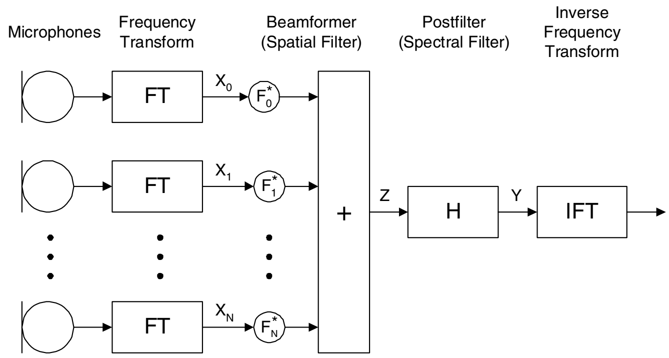

Fig. 1: Frequency domain implementation of a microphone array with beamforming an postfiltering.

Postfilter Estimation

Many postfilter algorithms ([4], [5], [6], [7], [8], [9], [10]) are based on the assumption of spatially uncorrelated noise. The postfilter is estimated by computing some sort of normalized cross-correlation coefficient (coherence) as a function of frequency.

Measurements in a normal office room (fig. 2 and 3.) have shown that the noise field in a reverberant room can be approximated by a diffuse sound field. The coherence function of the 3-dimensional diffuse sound field [2] is given by $\Gamma(f) = \mathrm{sinc}(2\pi f l / c)$, where sinc(x)=sin(x)/x, l is the microphone distance, c is the sound speed, and f is the frequency.

In a diffuse noise field correlation-based postfilters have several drawbacks that significantly reduce the speech quality [11]:

1. Negative coherence at certain frequencies leads to a negative postfilter transfer function [11]
2. The transfer function can become unstable if superdirective coefficients are used [10].
3. A large microphone distance $l > c/(2f_{\min})$ is required for suppression of low frequencies [8] which leads to a violation of the spatial sampling theorem $l < c/(2f_{\max})$.

An improved postfilter algorithm $H = \phi_{zz} / \phi_{xx}$ proposed in [11] evaluates the ratio of the output power and the input power of the beamformer. This approach avoids negative postfilters can be used with small size arrays and superdirective coefficients and outperforms correlation based algorithms.

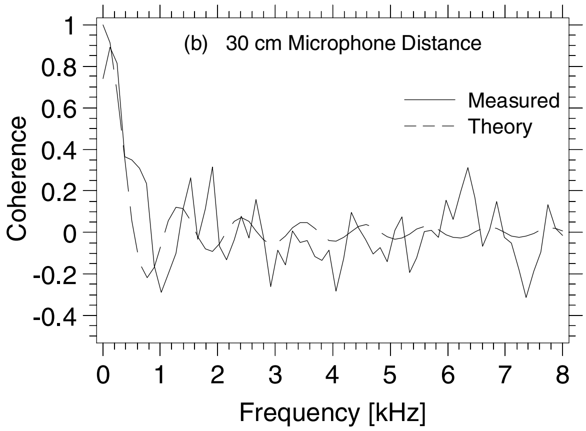

Fig. 2 Real part of the complex coherence function (30 cm mic. distance) Solid line: measured in an office room. Dashed line: theoretical result for a 3-dimensional diffuse sound field.

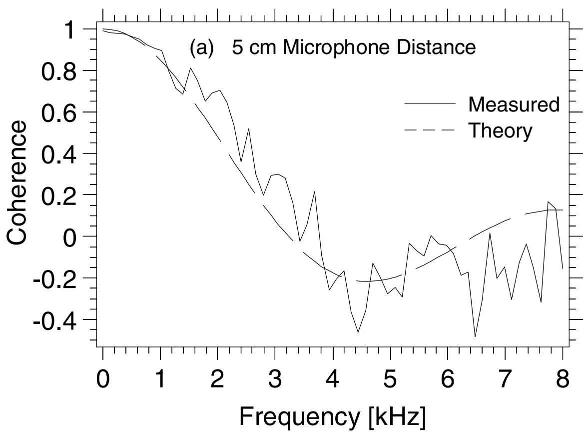

Fig. 3: Real part of the complex coherence function (5 cm mic. distance) Solid line: measured in an office room. Dashed line: theoretical result for a 3-dimensional diffuse sound field.

Conclusion

The multichannel Wiener filter can be decomposed into a product of a spatial (MVDR) filter depending on the noise correlation matrix and a one-dimensional Wiener postfilter depending on the output SNR of the spatial filter. Postfiltering algorithms based on the assumption of spatially uncorrelated noise provide suboptimal performance in diffuse noise field. However, appropriate postfilters applied to small size arrays provide high speech quality and are able to improve the output SNR of beamformers and superdirective arrays significantly.

References

[1] E. N. Gilbert and S. P. Morgan, "Optimum design of directive antenna arrays subject to random variations", Bell Syst. Tech. J., pp. 637-663, May 1955.

[2] R.K. Cook, R.V. Waterhouse, R.D. Berendt, S. Edelman, and M.C. Thompson, Jr., "Measurement of Correlation Coefficients in Reverberant Sound Fields," J. Acoust. Soc. Amer., vol. 27, no. 6, pp. 1072-1077. Nov. 1955.

[3] D. J. Edelblute, J. M. Fisk, and G. L. Kinnison, "Criteria for optimum-signal-detection theory for arrays", J. Acoust. Soc. Amer., vol. 41, no. 1, pp. 199-205, Jan. 1967.

[4] L. Danilenko, "Binaurales Hören im nichtstationären diffusen Schallfeld," PhD thesis, RWTH Aachen, Aachen, Germany, 1968.

[5] J. B. Allen, D. A. Berkley, and J. Blauert, "Multimicrophone signal-processing technique to remove room reverberation from speech signals", J. Acoust. Soc. Amer., vol. 62, no. 4, pp. 912-915, Oct. 1977.

[6] Y. Kaneda and M. Tohyama, "Noise suppression signal processing using 2-point received signals," Electron. and Communicat. Japan, vol. 67-A, no. 12, pp. 19-28, Apr. 1984.

[7] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Proc.(ICASSP), New York, USA, pp. 2578-2581, Apr. 1988.

[8] K. U. Simmer and A. Wasiljeff, "Adaptive microphone arrays for noise suppression in the frequency domain," in Second Cost 229 Workshop on Adapt. Alg. in Communicat., Bordeaux, France, pp. 185-194, Oct. 1992.

[9] M. Dörbecker and S. Ernst, "Combination of two-channel spectral subtraction and adaptive Wiener post-filtering for noise reduction and dereverberation," in European Signal Proc. Conf. (EUSIPCO), Trieste, Italy, Sept. 1996.

[10] C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," IEEE Trans. Speech and Audio Processing, vol. 6, no. 3, pp. 240-259, May 1998.

[11] K.U. Simmer, J. Bitzer, C. Marro: "Post-filtering Techniques", in: M. S. Brandstein and D. B. Ward (Eds.): "Microphone Arrays: Signal Processing Techniques and Applications", pp. 39-60, Springer-Verlag, Berlin, Heidelberg, New York, May 2001