An Adaptive Microphone Array for Hands–Free Communication

Sven Fischer[1] *Klaus Uwe Simmer*[2]

[1] University of Bremen
Department of Physics and Electrical Engineering
P.O. Box 330 440, D–28334 Bremen, Germany
e-mail: fischer@comm.uni-bremen.de

[2] Houpert Digital Audio
Wiener Str. 5, D–28359 Bremen, Germany, Fax: +49 421 705675

ABSTRACT

In this paper we present an adaptive microphone array to suppress coherent as well as incoherent noise in disturbed speech signals. We use a generalized sidelobe cancelling (GSC) structure implemented in the frequency domain, because it allows a separate handling of determining the adaptive look direction response to suppress incoherent noise and adjusting the adaptive filters for cancellation of coherent noise. The transfer function in the look direction is an adaptive Wiener–Filter which is estimated using the short time Fourier–Transform and the Nuttall/Carter method for spectrum estimation.

1. INTRODUCTION

Various methods for noise reduction and speech enhancement with the aid of microphone arrays have previously been described in the literature. The different approaches can be classified into three main categories:

- conventional beamforming [1], [2], [3]
- adaptive beamforming [4], [5],
- microphone arrays with adaptive postfiltering [6] [7].

The performance of these array techniques for noise reduction depends on the acoustical environment in which they have to operate. For example, adaptive beamforming works well if the number of point noise sources is smaller than the number of sensors. However, in closed environments, noise is influenced by multipath propagation and reverberation which yields a multi–source noise field. In such noise fields the third method yields a much better noise reduction performance, but theoretically requires a completely incoherent noise field.

Realistic noise fields are neither perfectly diffuse nor do they consist of direct–path noise only. The reflection coefficients of the walls as well as the distance between the noise sources and the array determine the ratio of coherent and incoherent noise components received by a microphone array. Therefore, a practical system for noise reduction must operate independently of the correlation properties of the noise field.

The method presented here is able to suppress coherent (i.e. direct path) noise and incoherent (i.e. diffuse) noise and can be conceived as a unification of the above mentioned three array techniques for noise reduction.

2. CONSTRAINED ADAPTIVE BEAMFORMING WITH ADAPTIVE LOOK–DIRECTION RESPONSE

The task of constraint minimum variance beamforming is to minimize the total output power of the array subject to the constraint of preventing an a priori specified impulse response in the look direction[4]. In beamforming for speech application however, the performance of the array can be improved by allowing the look direction impulse response to vary with time. We use as adaptive look direction response the impulse response of a non–causal Wiener filter. The implementation is straightforward, if we choose a generalized sidelobe cancelling structure as shown in figure 1. The constraints are included in the signal blocking matrix and an unconstrained update algorithm can be used [5]. In our application we use the short time Fourier transform and the overlap–add method to estimate the transfer functions. The transfer functions H_i of the sidelobe cancelling part (lower signal track in figure 1) are given by [8]:

$$H_i(z) = \frac{\Phi_{\delta_i y_w}(z)}{\Phi_{\delta_i \delta_i}(z)} \quad , i = 1, \ldots, L \quad , \tag{1}$$

where the number L depends on the blocking matrix structure (L can be $M-1$ or less ($M \widehat{=}$ number of microphones)). The cross power density spectrum $\Phi_{\delta_i y_w}$ and the auto power density spectrum $\Phi_{\delta_i \delta_i}$ of equation (1) are estimated using the recursive update formulas:

$$\widehat{\Phi}^{(l)}_{\delta_i y_w}[k] = \alpha \widehat{\Phi}^{(l-1)}_{\delta_i y_w}[k] + \delta^{*}_{i,l}[k] Y_{w,l}[k] \tag{2}$$

$$\widehat{\Phi}^{(l)}_{\delta_i \delta_i}[k] = \alpha \widehat{\Phi}^{(l-1)}_{\delta_i \delta_i}[k] + |\delta_{i,l}[k]|^2 \quad , \tag{3}$$

where k is the frequency index, l is the time segment index, $\delta_{i,l}[k]$ is the short time spectrum at the output of the signal blocking unit and $Y_{w,l}[k]$ is the postfiltered output spectrum of the conventional beamformer (see also figure 1). In equations (2) and (3), α is a number close to one and defines the average time. The GSC approach for noise reduction is closely related to adaptive noise cancelling proposed by Widrow *et al.* [8]. The noise reduction which can be achieved by this type of processor is completely specified by the spatial coherence of the noise field. A reasonable noise reduction can be achieved, if the noise signals between adjacent microphones are highly correlated [9]. Therefore, this part of our noise reduction system is able to suppress

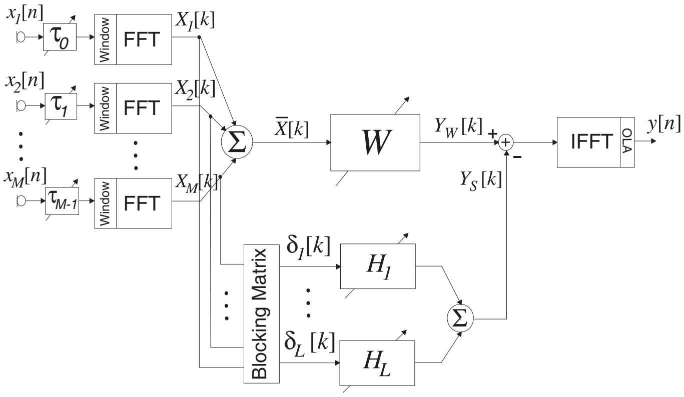

Fig. 1. Block diagram of the noise reduction system.

the coherent direct path noise, but is inefficient for incoherent noise.

2.1. Transfer Function in Look Direction

The transfer function W in look direction contains the constraint values and is designed to suppress spatially incoherent noise signals only. It is based on the assumption of a spatially white noise field, where in the ideal case of uncorrelated speech and noise the spatial cross power density spectrum of the received signals $\Phi_{x_i x_j}(z)$ equals the auto power density spectrum of the desired speech signal $\Phi_{ss}(z)$ [6]:

$$\Phi_{x_i x_j}(z) = \Phi_{ss}(z) \quad . \tag{4}$$

This fact is utilized to estimate the Wiener Filter W in look direction. The transfer function $\widehat{W}$ is given in our application by [10]:

$$\widehat{W}(z) = \frac{\left| \dfrac{2}{M(M-1)} \displaystyle\sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \Phi_{x_i x_j}(z) \right|}{\Phi_{\bar{x}\bar{x}}(z)} = \frac{\left|\widehat{\Phi}_{ss}(z)\right|}{\Phi_{\bar{x}\bar{x}}(z)} . \tag{5}$$

$\Phi_{\bar{x}\bar{x}}(z)$ is the auto power density spectrum of the output signal of the conventional beamformer $\bar{x}$. It can be shown that this transfer function $\widehat{W}$ is identical to the transfer function of a non causal Wiener Filter in the case of zero spatial correlation of the noise signals [10]. In the case of a completely coherent noise field the transfer function $\widehat{W}$ equals one and the noise reduction is only due to the sidelobe cancelling path of the system shown in figure 1. The power density spectra in the numerator and denominator of equation (5) can be estimated in a manner similar to equations (2) and (3) from the short time spectra:

$$\begin{aligned} \widehat{\Phi}_{ss}^{(l)}[k] &= \alpha \widehat{\Phi}_{ss}^{(l-1)}[k] + \qquad (6) \\ & \frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} X_{i,l}^{*}[k] X_{j,l}[k] \\ \widehat{\Phi}_{\bar{x}\bar{x}}^{(l)}[k] &= \alpha \widehat{\Phi}_{\bar{x}\bar{x}}^{(l-1)}[k] + |\overline{X}_l[k]|^2 \quad . \qquad (7) \end{aligned}$$

The transfer functions $\widehat{W}$ and H_i are determined as the data arrives at the input microphones. Thus, the adaption of all the transfer functions take place simultaneously.

2.2. Improvement of the Transfer Function Estimate $\widehat{W}$

The identity in equation (4) holds only in a statistical sense. In practice, only estimates of the spatial cross power densities $\widehat{\Phi}_{x_i x_j}$ are available, and validation of the identity in equation (4) requires infinitely long periodogram time–averaging. Due to the nonstationary nature of the speech signals, only short time intervals are available for spectrum estimation. Therefore, the transfer function $\widehat{W}$ is only a rough estimate for the true Wiener Filter.

To improve the estimate $\widehat{W}$ we use the combined time and lag weighting technique for periodogram smoothing as introduced by Nuttall and Carter [11], which was adapted to our application. The starting point are equations (6) and (7), which describes a short time Weighted Overlapped Segment Averaging (WOSA) method (excluding constant factors). These estimates are subjected to an inverse Fourier Transform to yield the correlation function estimates $\widehat{R}_{ss}$ and $\widehat{R}_{\bar{x}\bar{x}}$ respectively. In a next step, the correlation estimates are multiplied by a symmetric real lag weighting function w_{lag}, which takes into account the windowing of the input data prior to the computation of the FFT's, and is calculated according to the following expression [11]:

$$w_{lag}[n] = \frac{w_d[n] R_{ww}[0]}{R_{ww}[n]} \quad . \tag{8}$$

In equation (8), w_d is the desired lag window (in our case a Hanning window of one fourth the FFT length to perform the desired smoothing), R_{ww} is the auto correlation function of the data window, and w_{lag} is the reshaped lag window. In a final step the weighted correlation estimates are transformed back into the frequency domain to yield the desired power density spectra used in determining $\widehat{W}$ according to equation (5).

It should be noted that, the lag multiplication and Fourier transform can be replaced with frequency domain convolution. But as stated in [11], the unusual lag weighting is important for achieving good mean behaviour and is unlikely to be apparent by considering a frequency domain convolution approach.

Finally, because a conventional beamformer perfectly cancels single frequencies in several directions, $\widehat{W}$ is bounded between values of zero and one to avoid poles in the transfer function estimate.

3. SPATIOTEMPORAL BLOCKING MATRIX

In the Generalized Sidelobe Canceller the constraints are included in the signal blocking matrix. In its simplest form, the signal blocking is realized by taking the difference between adjacent sensor signals to yield (in the ideal case) noise only reference signals. The major drawback of this approach is that, if the desired speech signal is not blocked out completely at the input of the

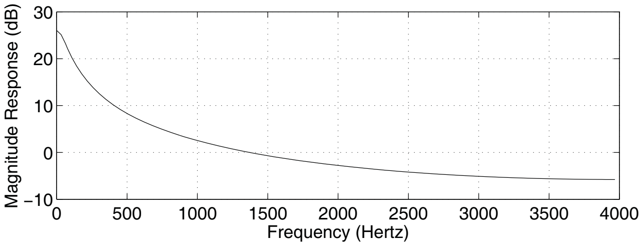

Fig. 2. Magnitude response of the lowpass filter for the signals δ_i to increase the noise reduction performance in the low temporal frequency region.

noise cancelling filters H_i, the filters will adapt to the desired speech signal and as a consequence the latter will be partially cancelled in the output signal y. In beamforming for speech application, this signal cancellation is often observed.

The rows of the blocking matrix can be interpreted as fixed beamformers, each of them forming a spatial null in the look direction [5]. Based on this interpretation, a blocking matrix using a spatial filtering technique was proposed in [12] to broaden the look direction and therefore prevent the adaptive filters H_i from cancelling signals coming from an area around the focal point. This approach can reduce the signal cancellation due to steering delay errors or widespread signal sources, but in general requires many sensors.

To increase the overall performance of the noise reduction system shown in figure 1, we propose in addition to the spatial filter approach a temporal filtering in the blocking matrix. The motivation behind this is as follows:
The low frequency components of the noise field can neither be suppressed by the conventional beamformer nor by the postfilter $\widehat{W}$, because of the good spatial correlation in this frequency region. The summation of the array signals has the effect of a temporal lowpass filter on the noise signals. On the other hand, the low temporal frequencies of the noise field will be attenuated at the output of the blocking matrix. Therefore, the signal blocking has the effect of a temporal highpass filter on the array signals. The transformation filters H_i have to compensate for this opposed behaviour to form a proper cancelling signal $\hat{Y}_S[k]$. The transformation filters are theoretically given by equation (1) for time stationary signals. But in practice, there are only estimates available and the filter order is limited. There exists always a potential for mismatch in the transfer functions $\hat{H}_i$, and because these filters have to operate over a large range of gain values, a mismatch can result in a very distorted output signal. Therefore, we include a fixed temporal lowpass filter in the blocking matrix, with the effect that the low frequency components in the signals δ_i will be emphasized. The used lowpass filter, whose magnitude response is shown in figure 2, is a one pole filter with transfer function $G(z) = 1/(1 - az^{-1})$ and $a = 0.95$. To avoid poles in the estimated transfer functions H_i due to zero power densities, their magnitudes were in addition constrained between values of zero and one.

4. EXPERIMENTAL RESULTS

4.1. Simulation Description

To test the noise reduction performance of the described system, a computer program has been developed which allows easy changing of the acoustical properties of the enclosure. The input signals were generated by convolving one channel anechoic recordings of speech and noise with the source–to–microphone impulse responses. These room impulse responses were simulated using the image method described by Allen and Berkley [13]. The room dimensions were $3.50 \times 7.10 \times 2.96$ m and the wall reflection coefficients were varied to simulate different reverberation times, i.e. different ratios of direct path noise and diffuse noise. The desired speaker was positioned 50 cm in front of the array and as noise source we used a hair–drier positioned 4.3 m away from the array center. The input SNR was 3 dB.

4.2. Choosing the Array Aperture

An array of discrete sensors can be conceived as a sampled continuous aperture. If the sampling period is not chosen appropriately, this sampling introduces spatial aliasing in form of grating lobes [14] [15]. On the other hand, the estimation of the transfer function in look direction W assumes a spatially white noise field. In practice the noise field can be at best diffuse with a spatial coherence function given by a sinc function. To yield spatially uncorrelated noise signals, undersampling the continuous aperture is usually performed. This works well for pure diffuse noise fields and if the desired speaker is close to the array. Our proposed system for noise reduction is in principle an adaptive beamformer which includes the method proposed in [7] as special case. An undersampled aperture yields a poor system performance in the case of direct path noise.

We used a seven element linear, equally spaced array with 5 cm inter–element spacing and total aperture length of 35 cm to avoid spatial aliasing in the frequency band below 3400 Hz. Experiments with various sensor configurations led us to revert to the linear array, which yielded the best performance under the constraint of a maximal number of seven sensors.

A better performance is expected by splitting the array in subarrays [3] and performing the noise reduction in each subarray with the system shown in figure 1, and combining the outputs. However, this will increase the cost.

4.3. Performance Measure

In speech communications, the ultimate recipient of information is the human being. The artefacts generated by many speech enhancement techniques decrease the user acceptance for voice communication systems. Therefore, for performance evaluation, subjective listening tests are absolutely necessary. But because this is time and cost intensive, we used for performance evaluation the Log Area Ratio (LAR) Distance (L_1 norm without energy weighting) as objective measure for speech quality which is found to correlate well with the subjective sensation [16].

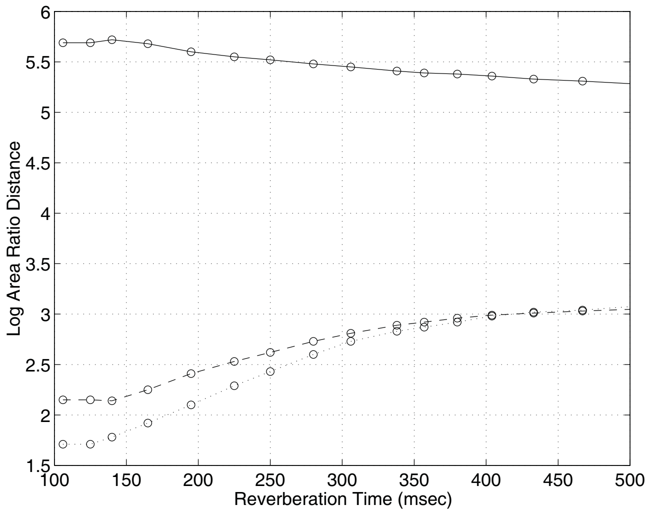

Fig. 3. Log Area Ratio Distance as function of reverberation time of the enclosure. Solid line: input LAR, dotted line: output LAR, dashed line: output LAR without temporal filtering in the blocking matrix.

4.4. Results

Figure 3 shows the LAR improvement as a function of Sabine's reverberation time T_{60} (low LAR $\widehat{=}$ high speech quality). The solid line shows the input LAR and the dotted Line shows the output LAR of the proposed noise reduction system. We can deduce from figure 3 that the proposed method works well for a large range of reverberation times and is therefore able to operate independently of the acoustical properties of the enclosure. The speech quality is considerably increased at the output of the noise reduction system. The dashed line in figure 3 shows the LAR at the output of the noise reduction system without the temporal lowpass filter in the blocking matrix, thus the overall performance can be increased by the proposed spatiotemporal signal blocking matrix, especially for reverberation times below 400 msec.

5. CONCLUSION

In this paper we proposed a noise reduction system for suppression of coherent and incoherent noise in disturbed speech signals which is based on a Generalized Sidelobe Canceller with two adaptive portions. Improvements of the estimation of the adaptive transfer function in look direction and the design of the signal blocking matrix were given. The experimental results demonstrated that the proposed method works well for a large range of reverberation times and is therefore able to operate independently of the acoustical properties of the enclosure.

ACKNOWLEDGEMENT

The authors would like to thank Mr. E. Ochieng–Ogolla of the University of Bremen for his helpful advice and suggestions that contributed to the improvement of this paper.

REFERENCES

[1] J.L. Flanagan, J.D. Johnston, R. Zahn, and G.W. Elko, "Computer–steered microphone arrays for sound transduction in large rooms," *J. Acoust. Soc. Amer.*, vol. 78, no. 5, pp. 1508–1518, Nov. 1985.

[2] W. Kellermann, "A self–steering digital microphone array," in *Proc. of the Internat. Conference on Acoustics, Speech and Signal Processing ICASSP–91*, pp. 3581–3584, 1991.

[3] Y. Mahieux, G. Le Tourneur, A. Gilloire, A. Saliou, and J.P. Jullien, "A microphone array for multimedia workstations," in *Proc. of the 3rd International Workshop on Acoustic Echo Control*, (Plestin les Grèves, France), pp. 145–149, Sep. 1993.

[4] O.L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[5] L.J. Griffiths and C.W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propagat.*, vol. AP-30, no. 1, pp. 27–34, Jan. 1982.

[6] R. Zelinski, "A microphone array with adaptive post–filtering for noise reduction in reverberant rooms," in *Proc. of the Internat. Conference on Acoustics, Speech and Signal Processing ICASSP–88*, (New York), pp. 2578–2581, Apr. 1988.

[7] K.U. Simmer and A. Wasiljeff, "Adaptive microphone arrays for noise suppression in the frequency domain," in *Second Cost 229 Workshop on Adaptive Algorithms in Communications*, (Bordeaux, France), pp. 185–194, 30.9.–2.10 1992.

[8] B. Widrow, J.R. Glover, J.M. McCool, J. Kaunitz, Ch.S. Williams, R.H. Hearn, J.R. Zeidler, E. Dong, and R.C. Goodlin, "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. 63, no. 12, pp. 1692–1975, Dec. 1975.

[9] W. Armbrüster, R. Czarnach, and P. Vary, "Adaptive noise cancellation with reference input – possible applications and theoretical limits," in *Proc. European Signal Processing Conf. EUSIPCO–86*, (The Hague), pp. 391–394, Sep. 1986.

[10] K.U. Simmer, S. Fischer, and A. Wasiljeff, "Suppression of coherent and incoherent noise using a microphone array," *Annals of telecommunications*, vol. 49, no. 7/8, no. 7/8, pp. 439–446, 1994.

[11] A. H. Nuttall and G.C. Carter, "Spectral estimation using combined time and lag weighting," *Proc. IEEE*, vol. 70, no. 9, pp. 1115–1125, Sep. 1982.

[12] I. Claesson and S. Nordholm, "A spatial filtering approach to robust adaptive beamforming," *IEEE Trans. Antennas Propagat.*, vol. 40, no. 9, pp. 1093–1096, Sep. 1992.

[13] J.B. Allen and D.A. Berkley, "Image method for efficiently simulating small–room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[14] Don H. Johnson and Dan E. Dudgeon, *Array Signal Processing — Concepts and Techniques*. Englewood Cliffs: Prentice Hall, 1993.

[15] L.J. Ziomek, *Fundamentals of Acoustic Field Theory and Space–Time Signal Processing*. Boca Raton: CRC Press, 1995.

[16] S.R Quackenbush, T.P. Barnwell, and M.A. Clements, *Objective Measures of Speech Quality*. Englewood Cliffs: Prentice Hall, 1988.