A2.5

A MICROPHONE ARRAY WITH ADAPTIVE POST-FILTERING FOR NOISE REDUCTION IN REVERBERANT ROOMS

Rainer Zelinski

Deutsche Bundespost, Research Institute Berlin
P.O. Box 420200, 1000 Berlin 42, West Germany

ABSTRACT

Speech communication is often disturbed by acoustic room noise in the environment of the speaker. This paper presents a self-adapting noise reduction system which is based on a 4-microphone array combined with an adaptive post-filtering scheme. Noise reduction is achieved by utilizing the directivity gain of the array and by reducing the residual noise through post-filtering of the received microphone signals. The post-filtering scheme depends on a Wiener filter estimating the desired speech signal and is computed from short-term measurements of the autocorrelation and cross-correlation functions of the microphone signals. The performance of the filtering scheme is increased substantially by additional post-processing of the cross-correlation measurements. The noise reduction system has been tested experimentally in a typical office room. The system produces an enhanced speech signal with barely noticeable residual noise if the input SNR is greater than 0 dB. The received noise power - measured in the absence of the speech signal - can be reduced by 28 dB.

I. INTRODUCTION

In a speech communication system the speech signal received by the microphone is often disturbed by an additive noise component which is caused by acoustic noise sources in the environment of the speaker. Several algorithms for noise reduction have been proposed; however, most of them fail in practical applications due to the complexity of real noise sources. For instance, noise filtering schemes with single-microphone input which are based on measurements of noise statistics in speech pauses [1] can hardly be applied for non-stationary noise sources. Noise cancellation systems with primary and reference microphone input [2] show a strong decrease in performance if the noise source cannot be modelled as a single point-source [3,4]. Microphone arrays with a null-steering post-processor (adaptive beam forming [5]) do not perform well in reverberant rooms, because the interfering noise signals arrive from nearly all directions due to multipath room reflections.

In this paper, a noise reduction system with multiple microphone input is presented which reduces the received room noise in two different steps. The system is based on a two-dimensional microphone array with 4 microphones, which is steered towards the direction of the desired sound source (speaker). In the first step, the directivity gain of the microphone array is utilized for noise reduction. In the second step, the microphone signals are post-filtered using an adaptive Wiener filter which estimates the desired speech signal.

II. NOISE REDUCTION SYSTEM

Fig. 1 shows the noise reduction system. The signals received by microphones $M_1...M_4$ pass the beam steering unit with delays $T_1...T_4$, which are adjusted such that the desired speech signal s arrives simultaneously in the four signals

$$x_i = s + n_i \quad ; \; i = 1...4 \quad . \qquad (1)$$

In the first step, the effect of the noise components n_i is reduced by computing the averaged signal

$$x_S = \frac{1}{4} \sum_{i=1}^{4} x_i \quad . \qquad (2)$$

In the second step, the residual noise is decreased additionally by post-filtering x_S, yielding the speech signal estimate $\hat{s}$. The adaptation of the post-filtering scheme is based on the well-known fact that the correlation between two received microphone signals in a reverberant room decreases with increasing distance between the two microphones and also with increasing distance between microphones and sound source. Hence, to discriminate between wanted signal s and interfering noise components n_i, it is necessary that the desired speaker is relatively close to the microphone array (the direct sound dominates), the noise sources are more distant to the array (the reflected sounds dominate), and the distance between adjacent microphones is not too small. On these assumptions, the received noise components can be regarded as being mutually uncorrelated, and the complete system can be adapted automatically. The adaptation is based on short-term measurements of autocorrelation and cross-correlation of the microphone signals $x_1...x_4$, which are evaluated in the frequency domain. A particular post-processing algorithm (described below) increases the performance of the noise reduction system substantially. The adaptive Wiener filter, which is implemented in the time domain, is finally computed from the processed correlation measurements.

A similar approach for noise reduction has been investigated by Kaneda and Tohyama [6]. However, the scheme in [6] is based on a 2-microphone system and does not include post-processing of the cross-correlation measurement. A 4-microphone system with adaptive post-filtering, where the LMS algorithm is used for adaptation of the Wiener filter, has been presented by the author [7]. Compared with the system in Fig. 1, the LMS-adapted scheme [7] is easier to implement, but is inferior in noise reduction performance.

III. ADAPTIVE POST-FILTERING

For determining the optimum Wiener filter, let us consider the time domain sequence

$$x(m) = s(m) + n(m) \quad , \qquad (3)$$

CH2561-9/88/0000-2578 $1.00 © 1988 IEEE

where m is the sampling index, s(m) the speech signal, and n(m) an additive noise signal being independent of s(m). The Wiener filter with coefficients w(j), defined in the index range I := {J ≤ j ≤ K} , yields the signal estimate

$$\hat{s}(m) = \sum_{j\in I} w(j)\, x(m-j) \quad . \qquad (4)$$

Minimization of the mean-squared error $E[(\, s(m) - \hat{s}(m)\,)^2]$ leads to the well-known discrete Wiener-Hopf equation, which can be formulated here as

$$\sum_{j\in I} w(j)\, R_{xx}(l - j) = R_{ss}(l) \quad ; \quad l \in I \quad . \qquad (5)$$

Functions $R_{xx}(\cdot)$ and $R_{ss}(\cdot)$ are the autocorrelation functions of signals x(m) and s(m), respectively.

For application of Eq. 5 in the post-filtering scheme of Fig. 1, functions $R_{xx}(\cdot)$ and $R_{ss}(\cdot)$ have to be estimated from the observed microphone signals $x_i(m)$; i=1...4. $R_{xx}(\cdot)$ can be estimated immediately from each of the four microphone signals $x_i(m)$. $R_{ss}(\cdot)$ can be estimated from the cross correlation of two microphone signals $x_i(m)$ and $x_j(m)$ if noise components $n_1(m)...n_4(m)$ are mutually uncorrelated and independent of s(m) :

$$E[x_i(m)\, x_j(m+l)] = E[(s(m)+n_i(m))\, (s(m+l)+n_j(m+l))]$$
$$= R_{ss}(l) \qquad \text{for } i \neq j \qquad (6)$$

The convolutional computations for estimating $R_{xx}(\cdot)$ and $R_{ss}(\cdot)$ are carried out in the frequency domain using the discrete Fourier transform (DFT) with block length L . Each block of L/2 consecutive samples $\{x_i(m)\}$ is appended by L/2 zeros and then transformed into the frequency domain yielding the DFT coefficients

$$\{Y_i(k)\} \quad ; \; k = 0...L-1 \quad \text{and} \quad i = 1...4 \quad . \qquad (7)$$

From $Y_1(k)...Y_4(k)$ we compute the auto-spectral density

$$A(k) = \frac{1}{4} \sum_{i=1}^{4} |Y_i(k)|^2 \quad ; \; k = 0...L-1 \qquad (8)$$

and the cross-spectral density

$$C(k) = \frac{1}{6} \sum_{i=1}^{3} \sum_{j=i+1}^{4} Y_i(k)\, Y_j^*(k) \quad ; \; k = 0...L-1 \quad , \qquad (9)$$

where * denotes the conjugate complex value.

The inverse DFTs of A(k) and C(k) lead to the time domain functions a(m) and c(m) (see also Fig. 1), which are estimates of autocorrelation functions $R_{xx}(\cdot)$ and $R_{ss}(\cdot)$, respectively. Finally, coefficients w(j) of the Wiener filter are computed according to Eq. (5) .

IV. POST-PROCESSING OF CROSS-CORRELATION MEASUREMENTS

Since the adaptive filtering scheme has to track the time-varying statistics of the desired speech signal, the block length L is restricted to a relatively small value. Due to this restriction, cross-spectral density C(k) contains an estimation error which causes an audible residual noise in output signal $\hat{s}$. This residual noise can be reduced by implementing a cross-correlation post-processing algorithm, which is described in the following.

Let us assume that the cross-spectral component $Y_i(k)\, Y_j^*(k)$ in C(k) can be modelled as

$$Y_i(k)\, Y_j^*(k) = S(k) + N_{ij}(k) \quad , \qquad (10)$$

where S(k) is the auto-spectral density of speech signal s(m) (real-valued) and $N_{ij}(k)$ is an additive zero-mean estimation error (complex-valued) with a phase angle distributed uniformly over [0 , 2π] ; $N_{ij}(k)$ being independent of S(k).

The variances of real and imaginary parts of $N_{ij}(k)$ are defined as

$$E[\mathrm{Re}^2\{N_{ij}(k)\}] = E[\mathrm{Im}^2\{N_{ij}(k)\}] = V(k) \quad . \qquad (11)$$

Note that in the absence of the speech signal s the term $N_{ij}(k)$ is identical to the measured cross-spectral density of the two noise signals $n_i(m)$ and $n_j(m)$. Variance V(k) then simply indicates the power of the received noise at frequency-index k . The model assumption in Eq. (10) provides a useful tool for developing the post-processing algorithm, and its applicability is confirmed by the performance improvements achieved in the experiments described below.

The effect of estimation error $N_{ij}(k)$ is reduced in two steps:

step 1: symmetrical cross-correlation function

Since C(k) represents an estimation of an auto-spectral density, it has to be real-valued (analogously, the time domain function c(m) has to be symmetrical). Thus we define a modified estimation $\tilde{C}(k)$, where

$$\mathrm{Re}\{\tilde{C}(k)\} = \mathrm{Re}\{C(k)\} \quad \text{and} \quad \mathrm{Im}\{\tilde{C}(k)\} = 0 \quad . \qquad (12)$$

In comparison with the estimation C(k), the estimation error variance in $\tilde{C}(k)$ is cut in half, since only the real parts of $\{N_{ij}(k)\}$ have an effect on $\tilde{C}(k)$.

step 2: decreasing the amplitudes of cross-spectrum estimation in noisy frequency regions

Each of the spectral density coefficients {S(k); k=0...L-1} can be estimated from 6 basic measurements of cross-spectral density, which can be written according to Eqs. (10) and (12) as

$$\tilde{C}_{ij}(k) = \mathrm{Re}\{Y_i(k)\, Y_j^*(k)\}$$
$$= S(k) + \mathrm{Re}\{N_{ij}(k)\} \quad , \qquad (13)$$

where ij is an element of the amount of index pairs IP := {12, 13, 14, 23, 24, 34} . Now we replace estimation C(k) in Eq. (9) by the post-processed estimation

$$P(k) = \alpha(k)\, \frac{1}{6} \sum_{ij\in IP} \tilde{C}_{ij}(k) \quad , \qquad (14)$$

where α(k) represents a frequency-dependent reduction factor. Factor α(k) is determined by minimizing the mean-squared estimation error $E[(S(k)-P(k))^2]$. If the six components $\{N_{ij}(k);\ ij \in IP\}$ are mutually uncorrelated and are independent of S(k), it is easy to show that α(k) is given by

$$\alpha(k) = S^2(k) \;/\; (S^2(k) + \frac{1}{6}\, V(k)) \quad . \qquad (15)$$

The higher the estimation error variance V(k), the smaller is the reduction factor α(k).

For evaluating Eq. (15), we have to estimate $S^2(k)$ and V(k). Both terms are estimated solely from the observed 6 measurement values $\{\tilde{C}_{ij}(k) \; ; \; ij \in IP\}$.

a) estimation of $S^2(k)$:

We start with the cross-spectral density $\tilde{C}(k)$, set negative values of $\tilde{C}(k)$ to zero (since S(k) has to be non-negative), and square the resulting terms. This squared spectrum is then smoothed by averaging over neighboured indexes of k and used as the estimation of $S^2(k)$ in Eq. (15).

b) estimation of V(k):

Since S(k) is non-negative, a negative value of $\tilde{C}_{ij}(k)$ can be caused only by the effect of $N_{ij}(k)$ (see Eq. (13)). Hence, we can use the observed negative values of $\{\tilde{C}_{ij}(k) \; ; \; ij \in IP\}$ for estimating V(k). Let us denote the number of negative values by M (M ≤ 6). We compute the average

$$\tilde{V}(k) = \frac{1}{M} \sum_{ij} \tilde{C}^2_{ij}(k) \qquad (16)$$

for those ij where $ij \in IP$ and $\tilde{C}_{ij}(k) < 0$.
Finally, $\{\tilde{V}(k)\ ,\ k=0...L-1\}$ is smoothed by averaging over neighboured values of k and the smoothed version is used as the estimation of V(k) in Eq. (15).

Equation (16) produces an underestimation of variance V(k) if S(k) is not equal to zero. This behaviour is not critical, however, since there is no need for reducing spectral density estimation P(k) in those frequency regions where $S^2(k)$ is essentially larger than V(k). On the other hand, the procedure of evaluating only the negative terms of $\{\tilde{C}_{ij}(k)\}$ has the advantage of being robust against small misadjustments of the beam steering unit. For small misadjustments , the speech component in the cross-spectral density is complex-valued instead of real-valued, but its real part will hardly be negative. Hence, an overestimation of V(k) is avoided in such cases.

The post-processing procedure described above reduces particularly the residual noise in the inter-formant frequency regions. The output speech signal then sounds nearly noise-free even for high noise levels at the system input.

V. ADDITIONAL DETAILS OF THE NOISE REDUCTION SYSTEM

The residual noise in output signal $\hat{s}$, which is still audible in speech pauses, can be reduced additionally by implementing a "coherence detector" which decides whether there is a coherent signal (e.g. the desired speech signal) or pure noise in the received microphone signals $x_1...x_4$. The detector decision controls an attenuation factor for reducing the amplitude of the output signal. The detector decision is a soft decision, i.e. the more the received signals resemble pure noise signals, the stronger is the attenuation. The decision is based on a comparison of amplitudes of positive and negative values in $\{\tilde{C}(k);\ k=0...L-1\}$:

$$\beta = \Big[\sum_{k'} \tilde{C}(k') \Big|_{\tilde{C}(k') \geqslant 0} \Big] \Big/ \Big[\sum_{k''} -\tilde{C}(k'') \Big|_{\tilde{C}(k'') < 0} \Big] \qquad (17)$$

If signals $x_1...x_4$ are pure noise signals, β is about one; for coherent signals $x_1...x_4$ the value of β is considerably larger than one.

The post-processing scheme described in Section IV allows an efficient noise reduction also in the low-frequency region. The lower the frequency is, the more two received noise signals $n_i(m)$ and $n_j(m)$ resemble each other, i.e. the more they seem to be correlated. This effect decreases the noise reduction performance of the Wiener filter for low frequencies. It can partly be compensated by increasing the values of variance estimation $\tilde{V}(k)$ in the low-frequency region. This frequency-dependent increasing factor (which also depends on the dimensions of the microphone array) can be determined from statistical measurements with noise signals.

It has to be mentioned that the noise reduction system is robust against some residual correlation between the received noise signals also in the other frequency regions. This property is achieved by estimating V(k) according to Eq. (16), since it is highly probable even for (slightly) correlated noise signals that at least one of the six terms $\{\tilde{C}_{ij}(k)\}$ is negative. Such a negative value indicates a noise signal in that frequency region, which can then be attenuated by applying the post-processing algorithm.

VI. EXPERIMENTAL RESULTS

The noise reduction system described above has been tested experimentally in a typical office room (25 m^2 area; 0.8 s reverberation time). The acoustic noise field is generated by four real noise sources (e.g. a drilling machine), which are spatially distributed. The microphone array is configured as a 2-dimensional array, where the microphones are placed in the edges of a square of 0.6 m x 0.6 m . The distance between microphone array and desired speaker is 0.6 m .

First of all, the coherence function has been measured for two noise signals $n_i(m)$ and $n_j(m)$ in adjacent microphones of the array (Fig. 2). The measured correlation is relatively small, with the exception of low frequencies and certain discrete frequencies which correspond to the harmonics of the machine rotations. Hence, we can regard the basic assumption of mutually uncorrelated noise signals $\{n_i(m)\}$ as being fulfilled.

The parameters of the noise reduction system are as follows: the sampling frequency is 8 kHz, the data length L/2 is 256 samples (data segment of 32 ms), 33 coefficients are used for post-filtering (J=-16, K=16). Computation of filter coefficients is done every 16 ms ; in between the filter coefficients are interpolated linearly. The automatic adjustment of the beam-steering delays has not been realized yet; the delays have been set to fixed values corresponding to the direction of the desired speaker.

Fig. 3 shows the noise power spectra measured at single-microphone input x_i and at output $\hat{s}$ of the system (measured in the absence of the speech signal s). Note that this measurement is based on a system simulation, where the coherence detector is not implemented (otherwise a measurement of noise power at signal $\hat{s}$ would be senseless). The noise power can be reduced by 28 dB for frequencies greater than 0.3 kHz . In this figure of 28 dB, a portion of 10 dB is due to the additional cross-correlation post-processing implemented in the system.

Adaptation of the Wiener filter to the time-varying speech spectrum is demonstrated in Fig. 4 for a speech syllable (time increment: 16 ms). The transfer functions of the Wiener filter reflect precisely the formant structure of the speech signal, even for input signal-to-noise ratios as low as 0 dB (as in Fig. 4b).

The noise reduction system produces an enhanced speech signal with barely noticeable residual noise if the input signal-to-noise ratio is greater than 0 dB. The output speech signal is free of "musical tones" or other extra noise, which can often be observed with noise filtering schemes based on single-microphone input.

VII. CONCLUSIONS

A self-adapting noise reduction system for application in reverberant rooms has been presented, which offers the following advantages:

* A priori knowledge about statistics of speech or noise signals is not necessary; the system does work reliably even if the noise itself consists of speech signals .
* Power density spectra and room positions of the noise sources may vary arbitrarily with time.
* There is no limit on the number of noise sources which can be tolerated by the system (the system performs even the better the more noise sources are present).
* A slight residual correlation in the received noise signals does not decrease the performance of the system.
* The speech output signal of the noise reduction system is free of musical tones or other extra noise.

REFERENCES

[1] J.S. Lim and A.V. Oppenheim,"Enhancement and bandwidth compression of noisy speech," Proc. IEEE, Vol. 67, pp. 1586-1604, 1979.

[2] B. Widrow et al.,"Adaptive noise cancelling: principles and applications," Proc. IEEE, Vol. 63, pp. 1692-1716, 1975.

[3] W. Armbrüster, R. Czarnach, P. Vary,"Adaptive noise cancellation with reference input - possible applications and theoretical limits," Proc. European Signal Process. Conf. EUSIPCO-86, The Hague, pp. 391-394, 1986.

[4] J.J. Rodriguez and J.S. Lim,"Adaptive noise reduction in aircraft communication systems," Proc. IEEE ICASSP-87, pp. 169-172, 1987.

[5] B. Widrow et al.,"Adaptive antenna systems," Proc. IEEE, Vol. 55, pp. 2143-2159, 1967.

[6] Y. Kaneda and M. Tohyama,"Noise suppression signal processing using 2-point received signals," Electronics and Commun. in Japan, Vol. 67-A, pp. 19-28, 1984.

[7] R. Zelinski,"A noise reduction system with two-dimensional microphone array and subsequent adaptive Wiener filtering," (in German), Proc. 6th Aachen Symposium on Signal Theory ASST'87, pp. 372-375, 1987.

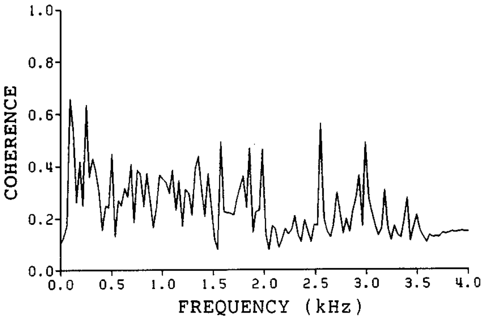

Fig. 2. Coherence function measured for two noise signals from adjacent microphones
(distance between microphones: 60 cm)

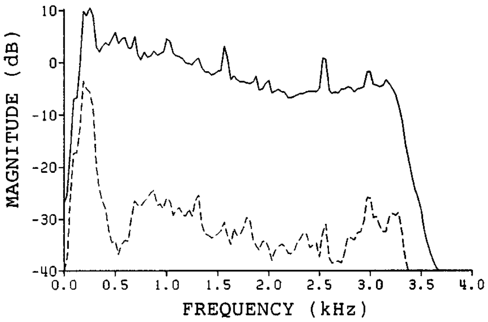

Fig. 3. Power density spectrum of noise signal (speech signal s = 0), measured in a system without coherence detector
solid line : spectrum at system input x_i
dotted line : spectrum at system output $\hat{s}$

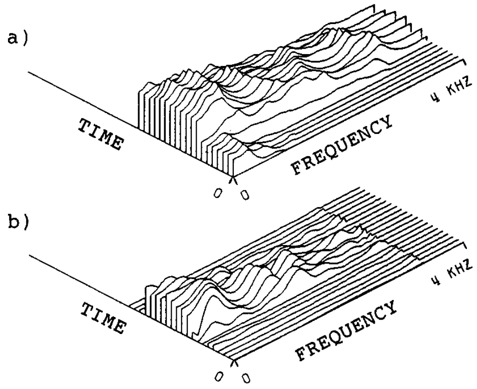

Fig. 4. Adaptation of Wiener filter to time-varying speech spectrum
a) short-term spectra of speech signal s
b) transfer functions of Wiener filter
(SNR at system input x_i : 0 dB)

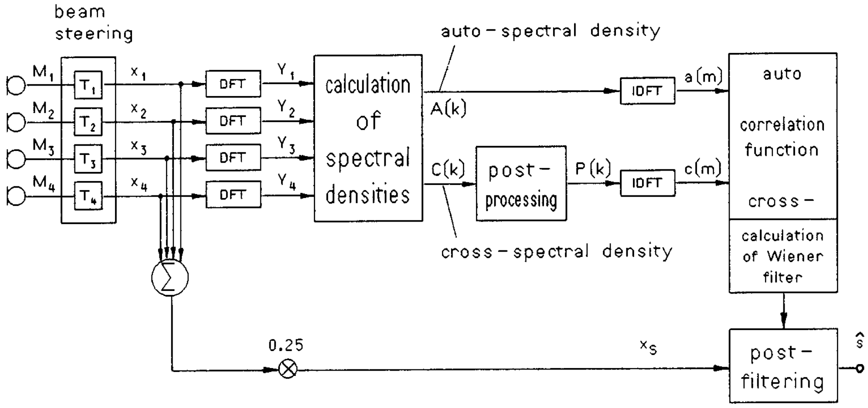

Fig. 1.

Block diagram of the noise reduction system