TREIZIÈME COLLOQUE GRETSI - JUAN-LES-PINS DU 16 AU 20 SEPTEMBRE 1991

Analysis and Comparison of Systems for Adaptive Array Processing of Speech Signals in a Noisy Environment

Dr. K. U. Simmer and Prof. Dr. A. Wasiljeff

University of Bremen, Department of Physics and Electrical Engineering,
P.O. Box 330 440, D-2800 Bremen 33, FRG

RÉSUMÉ

On présente ici une étude des différents systèmes adaptifs pour améliorer la qualité de signaux de paroles en utilisant un réseau de quatre microphones. Parmi les systèmes que nous avons étudiés sont ceux de Griffiths-Jim et Duvall beamformer, la méthode de Compernolle et le système pour la réduction de bruit sur la parole proposé par Zelinski. On montre en particulier que la suppression du bruit dépend de la cohérence spatiale des signaux de bruit reçus sur les microphones. On discute la qualité d'amélioration du rapport signal sur bruit obtenues avec les différents procédés.

ABSTRACT

In this paper we present a study of various adaptive systems for noise reduction of noisy speech using an array of four microphones. These systems are Griffiths-Jim and Duvall beamformers, van Compernolle's method and Zelinski's noise reduction system. We have shown that noise suppression depends on the spatial coherence of the noise signals received at the microphones of the array. We discuss the quality of Signal-to-Noise Ratio improvement achieved with the different methods.

1. INTRODUCTION

Speech recognition in man-machine communication systems is hindered by nonstationary noise from unwanted sources such as, in a factory, different types of machines. Adaptive noise reduction systems are often restricted to one or two microphones. In this paper we analyse and compare various multichannel systems for noise reduction by means of adaptive microphone arrays. The results may be applied to the enhancement of noisy speech in teleconferencing and free-hand telephoning and to the above mentioned man-machine communication.

2. ADAPTIVE MULTICHANNEL SIGNAL PROCESSING

Adaptive array processing algorithms are known from seismic, sonar and radar applications. However when applying them to speech, additional problems arise which reduce the performance of the algorithms. Speech and noise interference are both nonstationary and broadband signals, and often have similar spectral characteristics. In an office environment reverberation and echoes often predominate over direct path signals. To investigate these effects we have used a linear array of four omnidirectional microphones with different spacing and a planar two-dimensional array, where the four microphones are placed at the corners of a square. The signals received by the four microphones pass a beam steering unit with four delays which are adjusted such that the desired speech signal arrives simultaneously in the four receivers.

In order to evaluate the SNR improvement of the different systems speech signal and disturbing noise were recorded successively. These signals were multiplied by different gain factors and added in the computer. Thus it was possible to determine the noise suppression of various adaptive systems as a function of frequency and time.

3. PRESENTATION OF THE METHODS STUDIED

The first three methods are based on O. L. Frost's linearly constrained adaptive array processing algorithm [1].

The K-dimensional vector of data $\mathbf{x}[n]$ observed at the output of the steering delays at the n-th data sample is:

$$\mathbf{x}[n] = [x_1(n), x_2(n), \ldots x_K(n)]^T \tag{1}$$

In our case K, the number of microphones, equals four.
This vector is termed the array vector and a KL-dimensional stacked vector $\mathbf{z}[n]$ containing L delayed array vectors is formed as

$$\mathbf{z}[n] = [\mathbf{x}^T(n), \ldots, \mathbf{x}^T(n-L+1)]. \tag{2}$$

The output signal y(n) at the n-th sample is given by the inner product of the stacked vector $\mathbf{z}[n]$ and a KL-dimensional weight vector $\mathbf{w}$

$$y[n] = \mathbf{w}^T \mathbf{z}[n] \tag{3}$$

The linearly constrained minimization problem is defined as that of finding the weight vector $\mathbf{w}$ which satisfies

$$\mathbf{w}^T \mathbf{R}_{xx} \mathbf{w} \tag{4}$$

subject to $$\mathbf{C}^T \mathbf{w} \tag{5}$$

with the covariance matrix

$$\mathbf{R}_{xx} = E\{ \mathbf{x}(n)\, \mathbf{x}^T(n) \}. \tag{6}$$

The L-dimensional gain vector is defined as

$$\mathbf{f} = (f_0, \ldots, f_{L-1}) \tag{7}$$

and the KL dimensional constraint matrix C

$$\mathbf{C} = \begin{pmatrix} 1_k & 0_k & \ldots & 0_k \\ 0_k & 1_k & \ldots & 0_k \\ . & . & . & . \\ 0_k & 0_k & \ldots & 1_k \end{pmatrix} \tag{8}$$

The column vectors 1_k and 0_K contain K ones and K zeros respectively. The set of weights which satisfies (4) and (5) is

$$\mathbf{w}_{opt} = \mathbf{R}_{xx}^{-1} \mathbf{C} (\mathbf{C}^T \mathbf{R}_{xx}^{-1} \mathbf{C})^{-1} \mathbf{f} \tag{9}$$

Frost's algorithm is given by

$$\mathbf{w}[n+1] = \mathbf{P}\, [\mathbf{w}[n] - \mu\, y[n]\, \mathbf{z}[n]] + \mathbf{F} \tag{10}$$

where $\mathbf{P}$ is a projection operator

$$\mathbf{P} = [\, \mathbf{I}_{KL} - \mathbf{C} (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T] \tag{11}$$

and $\mathbf{F}$ a quiescent solution vector

$$\mathbf{F} = \mathbf{C} (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{f} \tag{12}$$

$\mathbf{w}$ converges to this value under conditions of spatially uncorrelated white noise at the steered outputs. μ is the adaptive step size parameter which controls the algorithm convergency.

The first method studied was Frost's oringinal algorithm as described above. The second method due to Griffiths and Jim is an alternative structure which has characteristics similar to those of Frost's algorithm. It converges to the same optimal weight vector.

The Duvall beamformer [2] is based on two signal processing systems, one to perform the adaptation according to Frost and a slaved beamformer to generate the system output signal. The Frost adaptive beamformer is connected to the sensors through a subtractive preprocessor which excludes the look-direction signal from the beamformer.

The fourth method, proposed by van Compernolle, avoids the problem of signal cancellation by switching off the adaptation in segments which contain speech [7].

The fifth method, suggested by Zelinski, is an open loop noise suppression system. The output signals x_i of the K microphones are used to estimate the weighting function $\mathbf{W}[n]$ of a Wiener filter . With $\Phi_{ss}(f)$ and $\Phi_{nn}(f)$ the power spectrum of the signal and the signal and noise, respectively, we obtain the following transfer function of the Wiener Filter

$$W(f) = \frac{\Phi_{ss}(f)}{\Phi_{xx}(f)} \approx \frac{\frac{2}{K \cdot K\text{-}1} \sum_{i=0}^{K-2} \sum_{j=i+1}^{K-1} \mathrm{Re}\{ X_i(f) \cdot X_j^*(f)\}}{\frac{1}{K} \sum_{i=0}^{K-1} |X(f)|^2} \tag{13}$$

According to Zelinski the cross spectral component can be modelled as

$$X_i(f)\, X^*_j(f) = S(f) + N_{ij}(f) \tag{14}$$

S(f) is the autospectral density of the speech signal and $N_{ij}(f)$ is an additive zero-mean estimation error with uniform distribution of the phase angle; $N_{ij}(f)$ beeing independent of S(f). The variances of the real and imaginary parts of $N_{ij}(f)$ are given as

$$E\{ \mathrm{Re}^2\{N_{ij}(f)\} = E\{ \mathrm{Im}^2\{N_{ij}(f)\} \} \tag{15}$$

Since the dominator in (13) represents an estimation of an autospectral density it has to be realvalued with

$$\mathrm{Im}\{ X_i(f)\, X^*_i(f) \} = 0 \tag{16}$$

From (14) we get

$$\mathrm{Re}\{ X_i(f)\, X^*_i(f) \} = S(f) + \mathrm{Re}\{ N_{ij}(f)\} \tag{17}$$

Thus the estimation error variance is cut in half, since only the real parts of $N_{ij}(f)$ are taken into account. The estimation error decreases with decreasing coherency of the noise received at the microphones. The estimation error of the cross-spectral density in the denominator of (13) is further reduced by applying a frequency dependent reduction factor which is described in [8].

Taking the discrete Fourier transform $\mathbf{w}[n]$ of (13) we get the output signal as the convolution of the average sensor signal

$$\overline{x(n)} = \frac{1}{K}\sum_{i=0}^{K} x_i(n) \qquad (18)$$

with the weighting function of the Wiener filter

$$y = \sum_{i=0}^{N} w(i) \cdot \overline{x(n-i)} \qquad (19)$$

4. EXPERIMENTAL RESULTS

To verify the various adaptive algorithms and to study their performance several experiments were conducted in an office room (20 m^2 area and approximately 1 second reverberation time) and in an anechoic chamber. White random noise was emitted from a loudspeaker in the first case and a hair dryer was used as noise source in the anechoic chamber.

The recorded signals were subject to an endpoint analysis to determine the start time T_s and the end time T_e of the interval within which the speech utterance of an isolated word occured. Fig. 1 shows a typical office recording with 0 dB average input SNR (1b). Fig. 1a displays the speech signal of one channel within the interval $[T_s, T_e]$. Fig. 1c to 1f show the respective noise reduction as a function of time for various algorithms.

Figures 2 and 3 show the noise reduction as a function of SNR for different algorithms. In the interval between the start time T_s and end time T_e of the speech signal, the noise reduction NR was calculated according to formula (20).

$$NR = {}^{10}\log \frac{\sum_{t_i=T_s}^{T_e} n^2(t_i)}{\sum_{t_i=T_s}^{T_e} e^2(t_i)} \qquad (20)$$

where $e^2(t)$ is the square of the difference between the desired and filtered signal and $n^2(t)$ is the input noise power at discrete time points t_i. Thus we eliminated the influence of the segments that do not contain speech. Figure 4 shows the noise coherence $\Gamma_{ij}^2(f)$ between two adjjacent microphones i, j as a function of frequency, measured in the office room and Figure 5 gives $\Gamma_{ij}^2(f)$ for the anechoic chamber.

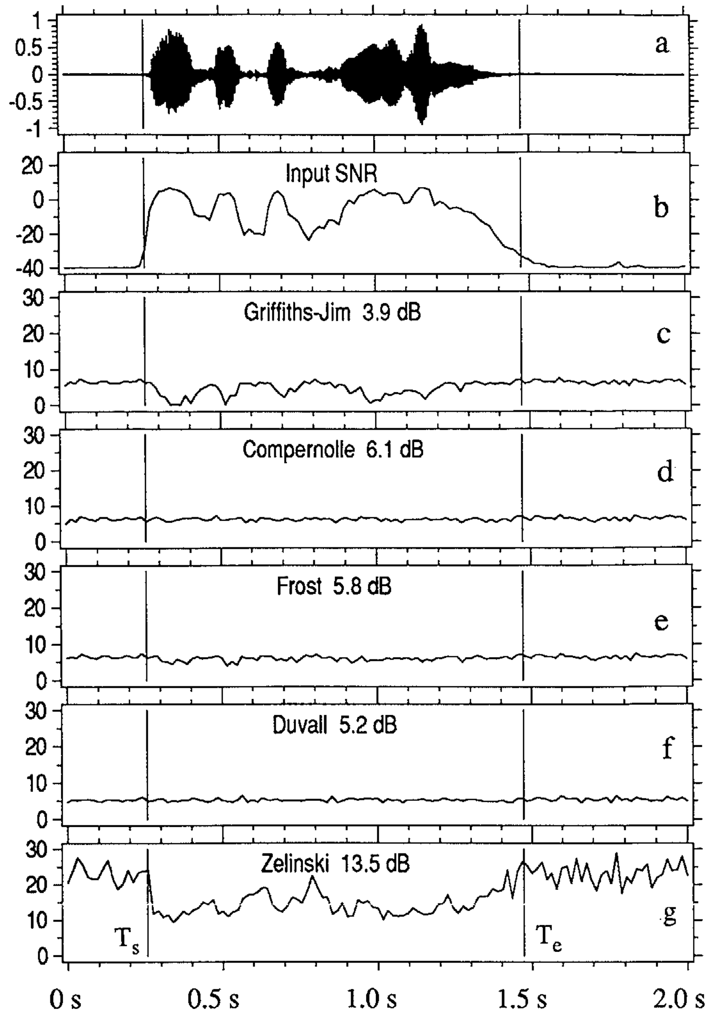

Fig. 1: Noise Reduction of Different Systems as a Function of Time in an Office Room.

$\Gamma_{ij}^2(f)$ is defined as

$$\Gamma_{ij}^2(f) = \frac{|\Phi_{ij}(f)|^2}{\Phi_{ii}(f)\,\Phi_{jj}(f)} \qquad (21)$$

The noise coherence is very low above 600 Hz in Figure 4 as compared to Figure 5. Zelinski's noise suppression algorithm is based on the assumption of incoherent noise [8] .

Our experiments show that this assumption is justified in a typical office room (Figure 4.) and yields a better noise reduction (solid line of Figure 2.) than the investigated noise cancellation algorithms (dashed lines of Figure 2).

Under the artificial condition of coherent noise (Figure 5.) we see from Figure 3. for input SNR above 10 dB an excellent performance of the noise cancellation algorithms (dashed lines), particularly of Compernolle's system. This system avoids the signal cancellation effect of the Griffiths-Jim beam-former by switching adaptive filters according to presence and absence of speech [7].

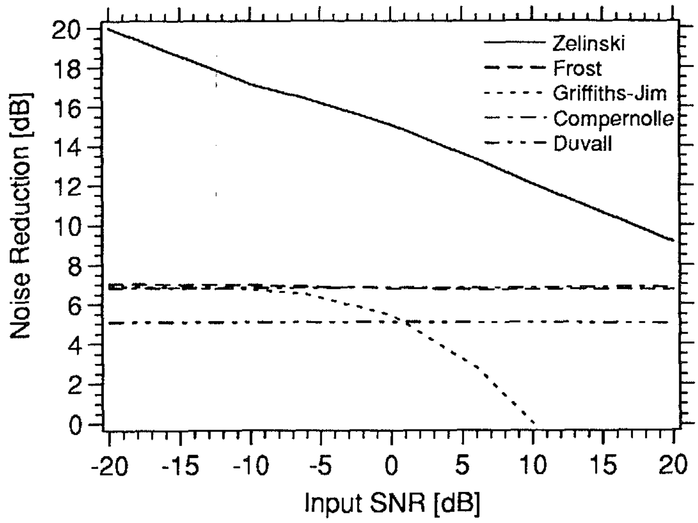

Fig. 2: Noise Reduction as a Function of SNR in Office Room

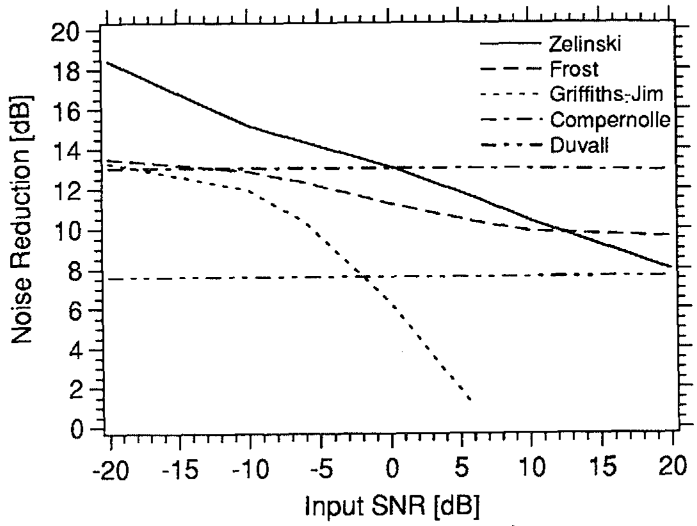

Fig. 3: Noise Reduction as a Function of SNR in Anechoic Chamber

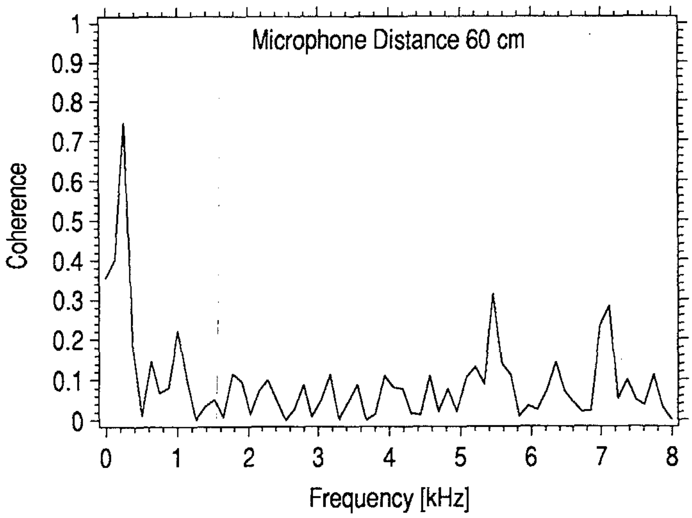

Fig. 4: Noise Coherence as a Function of Frequency in Office Room

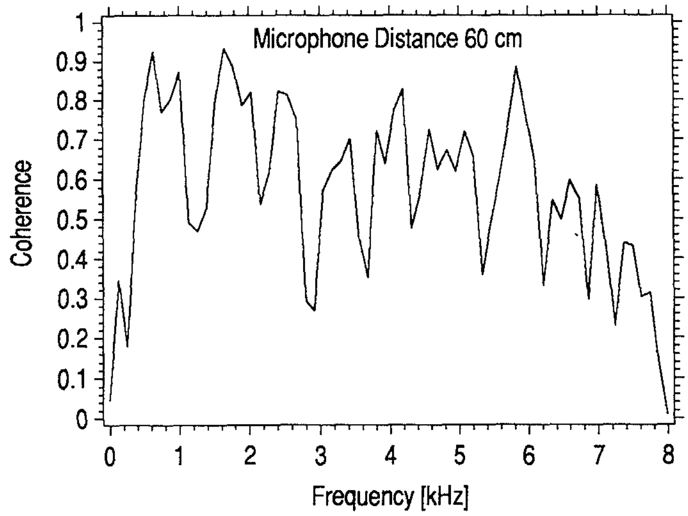

Fig. 5: Noise Coherence as a Function of Frequency in Anechoic Chamber

CONCLUSIONS

Several adaptive systems for noise reduction of noisy speech signals have been tested in a typical office and in an anechoic chamber. Noise suppression highly depends on the spatial coherence of the noise signals received at the microphones of the array.

ACKNOWLEDGMENT

The authors wish to thank Dr. Rolf Wehrmann from Telecom Research Institute Berlin for his help and useful discussions.

REFERENCES

[1] O.L. Frost, III, "An Algorithm for Linearly- Constrained Adaptive Array Processing," Proc. IEEE, vol. 60, no.8, pp. 926-935, Aug. 1972.

[2] B. Widrow et al., "Signal Cancellation Phenomena in Adaptive Antennas: Causes and Cures," IEEE Trans. Antennas Propag., vol AP-30, p. 469, May 1982.

[3] K. M. Duvall, "Signal Cancellation in Adaptive Antennas: The Phenomenon and a Remedy," Stanford Univ., Stanford, Calif., Aug. 1983 (Ph.D. thesis).

[4] L. J. Griffiths, C. W. Jim, "An Alternative Approach to Linearly Constrained Adaptive Beamforming" IEEE Trans. Antennas Propagat. vol. 30, no.1, pp. 27-34, Jan. 1982.

[5] K. U. Simmer, A. Malunat, A. Wasiljeff: "Räumlich Systeme zur adaptiven digitalen Verarbeitung akustischer Signale", URSI - Kleinheubacher Berichte 33, 1989, pp. 263 - 271.

[6] D. Van Compernolle: "Adaptive Filter Structures for Enhancing Cocktail Party Speech", Twelfth GRETSI Symposium, Juan-Les-Pins, June 1989, pp 513 - 516.

[7] D. Van Compernolle: "Switching Adaptive Filters for Enhancing Noisy and Reverberant Speech from Microphone Array Recordings", Proc. Int. Conf. Acoustics, Speech and Signal Processing, ICASSP-90, Albuquerque, pp. 833-836.

[8] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms", Proc. Int. Conf. Acoustics, Speech and Signal Processing, ICASSP-88, New York, pp. 2578-2581, 1988.