A SYMMETRIC TWO MICROPHONE SPEECH ENHANCEMENT SYSTEM — THEORETICAL LIMITS AND APPLICATION IN A CAR ENVIRONMENT

Rainer Martin, Peter Vary

Institute for Communication Systems and Data Processing, Aachen University of Technology,
Templergraben 55, 5100 Aachen, Germany, Phone: +49 241 806984, Fax: +49 241 806985

Abstract

This contribution presents an LMS-driven two microphone speech enhancement system which effectively takes advantage of coherence properties of speech and noise in car environments. The adaptive algorithm computes a 65–tap linear phase FIR filter which enhances the coherent speech signal while suppressing uncoherent noise. The theoretical limit for the noise suppression of the algorithm is derived and its performance under ideal and realistic conditions investigated.

I. Introduction

The use of a hands-free telephone in a car requires a noise suppression system to cancel out environmental noise. During the past 20–some years several methods were proposed which suffer from residual noise [2] or requirements for microphone placement which are hard to fulfill [1], [3]. Microphone arrays with adaptive postprocessing, however, appear to be very promising candidates for speech enhancement in an automobile environment. Zelinski recently proposed a four microphone rectangular array with good noise suppression properties [4]. Since a four microphone solution is not practical in the restricted spatial environment of a car, we here present a symmetric two microphone speech enhancement system with a LMS driven postprocessing filter and low computational complexity. We derive a theoretical bound on its performance and demonstrate its use in a car environment.

II. Speech and Noise in Car Environments

The performance of the speech enhancement system depends on the spectral and correlation characteristics of the speech and the noise signals. Car noise typically has its peak power between 100–800 Hz, depending on driving conditions and the car. For higher frequencies the noise power spectrum decreases with approximately 10 dB/1000 Hz. Unfortunately the power spectrum of speech exhibits a very similar behavior. Hence, a complete separation of noise and speech may not always achieved.

The magnitude squared coherence (MSC) of the recorded speech signal is above 0.9 for microphones distances up to 40 cm. Our measurements confirm that the MSC of noise fields in a car is well described by the MSC of the ideal diffuse sound field.

III. The Symmetric Two Microphone Speech Enhancement System

The noise suppression in the proposed system is achieved by coherently adding the input signals and by employing an adaptive FIR filter which is estimated using the two noisy input signals. The input signals are highpass filtered (f_c=240 Hz) and shifted for minimum time delay difference by means of an LMS algorithm (which will not be treated in this paper). The coherently added signals are then processed by the adaptive filter. The adaptive filter is estimated using the symmetric structure shown in Figure 1 which features two independent linear phase LMS-driven FIR filters. The i-th coefficient $w_i(k)$ of a linear phase filter is adjusted according to equation 1. x(k) is the input signal, e(k) is the adaptation error, and $\mu(k)$ is a variable step size parameter.

$$w_i(k) = w_i(k-1)+\mu(k)e(k)[x(k-i+1)+x(k-N+i)] \tag{1}$$

The three-tap FIR preemphasis filters were optimized in an off line experiment. One of the key features of our system is its relatively white residual noise.

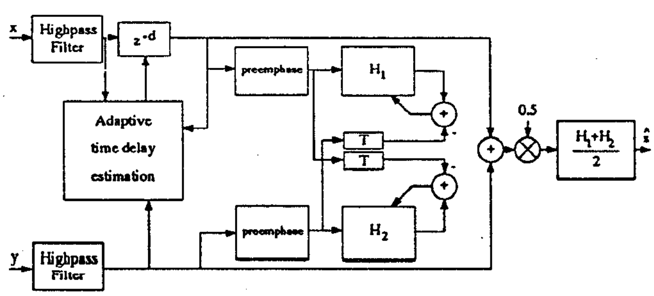

Figure 1: The speech enhancement system

IV. SNR Improvement

Array Gain

The gain obtained by coherently adding two (or N) microphone channels clearly depends on the coherence properties of speech and noise and on their respective spectral distribution. In case of uncorrelated noise and totally correlated speech the maximum gain of $10 \times \log_{10}(N)$ is achieved. However, in a diffuse noise field the gain can be considerably less, depending on the microphone distance and the spectral distribution of noise power.

Improvement by Adaptive Processing

Figure 2 shows the basic building block of our system: the adaptive filter which estimates the speech signal $\hat{s}$ while minimizing the error signal e. However, due to the noisy reference signal y the estimated speech signal $\hat{s}$ will also contain the correlated noise components. For this reason we use the mean square

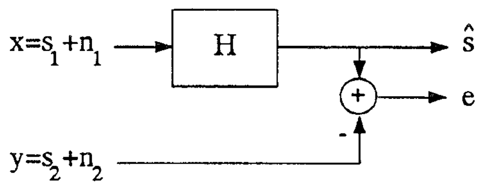

Figure 2: Error model with optimal filter H

of $d = s_2 - \hat{s}_2$ as a performance measure. s_2 is the undisturbed speech signal on the reference channel. The performance measure is computed under the assumption that error e is minimized. With the help of the orthogonality property of least squares estimates it follows that:

$$E\{d^2\} = E\{s_2(s_2 - \hat{s}_2)\} + E\{n_2\hat{n}_2\} \qquad (2)$$

where E{•} denotes the expected value. Assuming a noncausal, infinite length optimal filter we obtain the minimum value of $E\{d^2\}$in terms of (cross) power spectra S:

$$E\{d^2_{min}\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{s_2s_2} + \frac{|S_{n_1n_2}|^2 - |S_{s_1s_2}|^2}{S_{xx}} d\omega \qquad (3)$$

and in terms of MSC functions C we obtain:

$$E\{d^2_{min}\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{s_2s_2}\left(1 - C_{s_1s_2}\frac{SNR_x}{1+SNR_x}\right)d\omega + \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{n_2n_2}C_{n_1n_2}\frac{1}{1+SNR_x}d\omega \qquad (4)$$

where SNR_x denotes the signal to noise ratio of signal x. From equation 4 we find that a small estimation error requires strongly correlated speech signals and high SNR input signals. Under low SNR conditions low coherence of noise signals becomes important. In practice, the performance of equations 3 and 4 will not be achieved due to misadjustment and finite length of the filter and the nonstationary nature of signal and noise. However, the above formulae give a good illustration of how the system reduces noise and may be easily evaluated for special cases.

Performance in a Car Environment

Figure 3 shows the noise power spectra before and after the adaptive filter. An average gain of 12-15 dB can be obtained. A minimum microphone distance of 30-40 cm is necessary to restrict coherent noise to frequencies below the highpass cutoff frequency.

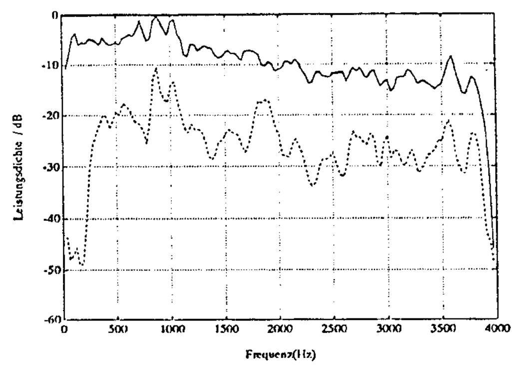

Figure 3: Noise power spectra before (solid) and after (dashed) the adaptive filter

V. Conclusion

The proposed algorithm is effective in suppressing noise over a large range of frequencies. The theoretical analysis gives a good interpretation of the noise suppression mechanism and its limits. It is easily extended to more than two channels. Microphone positions must be optimized for both low noise coherence and high SNR.

Acknowledgment

This work was supported by Philips Kommunikations Industrie AG, Germany.

References

[1] W. Armbrüster, R. Czarnach, and P. Vary. Adaptive Noise Cancellation with Reference Input - Possible Applications and Theoretical Limits. Signal Processing III: Theories and Applications, I.T. Young et al. (editors), Elsevier Science Publishers B.V. (North Holland), pp. 391-394, 1986.

[2] S.F. Boll. Suppression of Acoustic Noise in Speech Using Spectral Subtraction. IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-27, pp. 113-120, April 1979.

[3] B. Widrow, J.R. Glover, J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, Dong.E., and R.C. Goodlin. Adaptive Noise Cancelling: Principles and Applications. Proc. IEEE, Vol. 63, No. 12, pp. 1692-1716, December 1975.

[4] R. Zelinski. Noise Reduction Based on Microphone Array with LMS Adaptive Post-Filtering . Elect. Lett., Vol. 26, No.24, pp. 2036-2037, November 1990.