Correspondence

Suppression of Acoustic Noise in Speech Using Two Microphone Adaptive Noise Cancellation

STEVEN F. BOLL AND DENNIS C. PULSIPHER

Abstract—Acoustic noise with energy greater or equal to the speech can be suppressed by adaptively filtering a separately recorded correlated version of the noise signal and subtracting it from the speech waveform. It is shown that for this application of adaptive noise cancellation, large filter lengths are required to account for a highly reverberant recording environment and that there is a direct relation between filter misadjustment and induced echo in the output speech. The second reference noise signal is adaptively filtered using the least mean squares, LMS, and the lattice gradient algorithms. These two approaches are compared in terms of degree of noise power reduction, algorithm convergence time, and degree of speech enhancement. Both methods were shown to reduce ambient noise power by at least 20 dB with minimal speech distortion and thus to be potentially powerful as noise suppression preprocessors for voice communication in severe noise environments.

I. Introduction

It has been shown that there is a significant reduction in measured speech intelligibility and quality due to the ambient background noise generated in many operating environments [1]. A number of single microphone approaches for reducing the background noise added to speech have been developed [2]. However, these methods become ineffective when the noise power is equal to or greater than the signal power or when the noise spectral characteristics change rapidly in time. This correspondence describes an alternative approach to noise suppression in which a second correlated noise source is adaptively filtered to minimize the output power between the two microphone signals. Two adaptive algorithm implementations were investigated: the LMS approach [3] and the lattice gradient approach [4]. Each approach was compared in terms of degree of noise power reduction, algorithm settling time, and degree of speech enhancement.

II. Implementation Considerations

The estimated adaptive filter in the absence of uncorrelated noise represents a transfer function equal to the product of the transfer function from the noise source to the speaker multiplied by the inverse of the transfer function from the noise source to the reference microphone. Based on simulation studies [5], approximating this inverse transfer function adequately requires using an all-zero filter having 1500 tap weights. Such a large filter in turn increases misadjustment (the ratio of excess to minimum mean-square error, [3]). As is discussed below, the amount of misadjustment is an important design factor in noise suppression since large misadjustment manifests itself as a pronounced echo in the speech waveform. Echo can be present in the output speech since the output is continually fed back when estimating the tap weights. The echo is removed by reducing the adaptation step size, and thus the misadjustment. This reduction, of course, conflicts with the requirement of quick settling time. The tradeoff between misadjustment and settling time is discussed below.

Another issue is filter causalty. In general, a noncausal filter is required if the noise reaches the speaker before reaching the reference microphone. Noncausal adaptive filters are easily generated by placing a delay into the primary channel. However, more tap weights are then required with the accompanying misadjustment problems described above. For the experiments described here, the reference microphone was placed next to the noise source, eliminating the need for delay.

III. Experimentation and Results

An analog white noise generator was played out through a loud speaker into a hard-walled room. The reference signal microphone was placed next to the loud speaker, while the primary microphone was placed 12 ft away next to the control terminal. The author (D.P.) spoke into the primary microphone while controlling the stereo recording program. The noise power was amplified to such a level that the recorded speech was completely masked. The signals were filtered at 3.2 kHz, sampled at 6.67 kHz, and quantized to 15 bits. Recordings were made with and without speech present, each lasting 23.4 s.

Each algorithm's performance is measured in terms of the degree of steady-state noise power reduction during nonspeech activity, the time it takes to reach this steady-state value (algorithm settling time), and the amount of echo induced when speech is present. Three experiments were conducted to measure algorithm settling time and induced echo as a function of specified misadjustment. Step sizes were used corresponding to misadjustments of 1, 5, and 10 percent. The results showed that both algorithms converge to a steady-state noise power reduction of -20 dB in approximately 15 s for 10 percent misadjustment and 21 s for 5 percent misadjustment. At 1 percent misadjustment the step size for the LMS algorithm was so small that the noise power was reduced by only -10 dB before the data ran out. For the lattice algorithm, at 1 percent misadjustment, essentially no convergence was measured. In listening to the output during speech activity, it was judged that at 10 percent misadjustment an unacceptable amount of echo was present and that at 5 percent misadjustment the echo was just noticeable.

To illustrate this noise suppression capability, isometric plots of the short-time magnitude spectra with and without noise suppression are shown in Figs. 1 and 2. A description of the plot construction is described in [2]. Fig. 1 corresponds to the short-time spectrum of the unprocessed speech signal: "the pipe began to." Fig. 2 corresponds to the processed speech signal using the 5 percent misadjustment after the filter has converged. Since the noise was acoustically added, no underlying clean speech spectrum was available for comparison. However, it was judged that the intelligibility of the processed speech had clearly improved. This was based

Manuscript received September 5, 1979; revised February 8, 1980. This work was supported by the Information Processing Techniques Branch of the Defense Advanced Research Projects Agency, monitored by the Naval Research Laboratory under Contract N00173-79-C-0045.

S. F. Boll is with the Department of Computer Science, University of Utah, Salt Lake City, UT 84112.

D. C. Pulsipher is with Sandia Laboratories, Livermore, CA 94550.

0096-3518/80/1200-0752$00.75 © 1980 IEEE

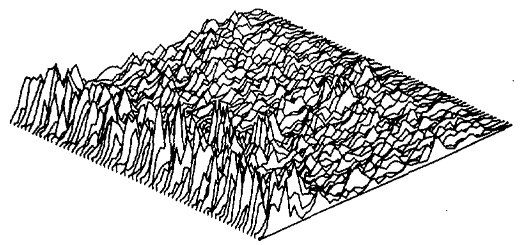

Fig. 1. Short-time spectrum of the unprocessed speech: "the pipe began to."

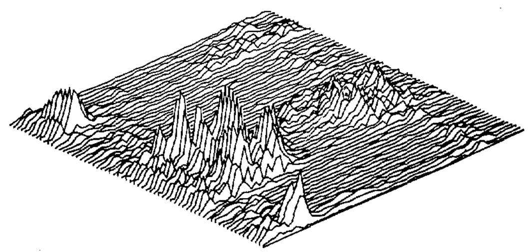

Fig. 2. Short-time spectrum of the noise cancelled output.

upon the fact that before processing it was difficult to even detect that there was speech present in the noise, while after processing the speech was understandable.

IV. Conclusions

In terms of noise power reduction and amount of echo present, both approaches can be adjusted to give equivalent results. Using step sizes corresponding to approximately 5 percent misadjustment, each algorithm converges (noise power down 20 dB) after 20 s of input, with a just noticeable amount of echo. For this white noise environment, the orthogonalization and energy normalization provided in the gradient lattice approach offered no advantage.

In summary, although this two microphone approach to noise suppression requires a second signal and possibly excessive computation due to long filter lengths, it offers a potentially powerful approach for speech enhancement in severe noise environments.

Acknowledgment

The authors wish to thank J. Makhoul, L. J. Griffiths, and E.Satorius, for their helpful discussions. Also they are grateful to R. Power and G. Randall for their assistance in implementing the algorithms on the FPS-120B array processor, to M. Milochik for preparation of the photographs, and to E. Collins for preparation of the manuscript.

References

[1] C. F. Teacher and D. Coulter, "Performance of LPC vocoders in a noisy environment," in *Proc. IEEE Conf. Acoust., Speech, Signal Processing*, Washington, DC, Apr. 1979, pp. 216–219.

[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, Apr. 1979.

[3] B. Widrow, J. McCool, M. Larimore, and C. R. Johnson, Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, pp. 1151–1162, Aug. 1976.

[4] L. Griffiths, "An adaptive lattice structure used for noise-cancelling applications," in *Proc. IEEE Conf. Acoust., Speech, Signal Processing*, Tulsa, OK, Apr. 1978, pp. 87–90.

[5] D. Pulsipher, S. F. Boll, C. K. Rushforth, and L. K. Timothy, "Reduction of nonstationary acoustic noise in speech using LMS adaptive noise cancelling," in *Proc. IEEE Conf. Acoust., Speech, Signal Processing*, Washington, DC, Apr. 1979, pp. 204–208.

On Some Suboptimum ARMA Spectral Estimators

S. BRUZZONE AND M. KAVEH

Abstract—**This correspondence describes some suboptimum schemes for ARMA spectral estimation. A least squares method is presented and compared to the method based on the modified Yule-Walker equations. A modification of the latter method is also given that improves its behavior in estimating spectra with narrow peaks. Examples are then shown that compare the suboptimum methods to the maximum likelihood one.**

I. Introduction

In this correspondence a new suboptimum scheme is reported for estimating the power spectral density (PSD) of an autoregressive moving-average (ARMA) process of known orders. After a preliminary data reduction, this scheme, called the least squares (LS) estimator, minimizes a sum of squared quadratic functions of the autoregressive (AR) coefficients using a nonlinear least squares algorithm. The poles of the estimated PSD are found from the minimizing AR coefficients, and zeros are found from quadratic functions of these coefficients.

The general idea of least squares fitting the ARMA parameters is not new, and various other approaches have been suggested (see, e.g., [3] and [7]). The scheme discussed here is, however, analogous to a minimum mean-squared error estimation of the parameters appearing in the estimator discussed in [2] and [5]. A modification of the latter estimator that is based on the modified Yule-Walker equations (the MYW estimator) is also presented, in which the problem of negative excursions of the estimated PSD is corrected by tapering the estimated moving-average autocorrelation function. Examples are shown that compare the performance of these ad hoc techniques to an approximate maximum likelihood one, the unconditional least squares method of Box and Jenkins [3, pp. 231–235].

II. The Spectrum of an ARMA Process

Assume that we observe x_t, $t = 1, \cdots, N$ where x_t is stationary and Gaussian of mean zero, and that x_t fits an ARMA (L, M) model. Then we can write

$$x_t - \sum_{i=1}^{L} a_i x_{t-1} = u_t \tag{1}$$

where a_i are the AR parameters for the ARMA model and u_t is the MA residual sequence given by

$$u_t = \epsilon_t - \sum_{i=1}^{M} b_i \epsilon_{t-i} \tag{2}$$

where b_i are the MA parameters and ϵ_t is a zero-mean uncorrelated normal sequence of variance σ_ϵ^2. Define

$$A^T = [1, -a_1, \cdots, -a_L].$$

Then the PSD of x_t is given by

$$S_x(z) = S_u(z) \left| A^T Z_L \right|^{-2} \tag{3}$$

where $Z_k^T = [1, z^{-1}, \cdots, z^{-k}]$, z being the z-transfer operator

Manuscript received March 14, 1980; revised June 26, 1980. This work was supported by the Air Force Office of Scientific Research under Grant AFOSR-78-3628.

The authors are with the Department of Electrical Engineering, University of Minnesota, Minneapolis, MN 55455.