EXPERIMENTAL RESULTS OF SUBBAND ACOUSTIC ECHO CANCELERS UNDER SPHERICALLY INVARIANT RANDOM PROCESSES

Phillip L. De León II[1] *and Delores M. Etter*[2]

[1] Klipsch Department of Electrical and Computer Engineering
New Mexico State University
Las Cruces, New Mexico 88003
pdeleon@nmsu.edu

[2] Department of Electrical and Computer Engineering
University of Colorado
Boulder, Colorado 80309-0425
etter@boulder.colorado.edu

ABSTRACT

The subband adaptive filter system has been applied to the acoustic echo cancellation problem in order to overcome the problems of slow convergence due to spectrally dynamic input and high computational costs associated with a single, long adaptive filter. Research into the convergence characteristics of subband acoustic echo cancelers (AECs) have either used Gaussian white noise, USASI noise (as an approximation for speech signals), or short term speech signals. More recently, research in the statistical modeling of speech signals using spherically invariant random processes (SIRPs) and analyses of LMS and NLMS algorithms under SIRPs has led to a better understanding of adaptive filter performance under more realistic input in the context of AEC. In this paper, we present experimental results using a subband AEC under SIRP input. These results yield a better understanding of the performance in actual acoustic echo cancellation applications and highlight the benefits of subband techniques.

1. INTRODUCTION

The motivation for adaptive filtering in subbands stems from two well-known problems in fullband adaptive filtering. First, the convergence and tracking of a normalized least-mean square (NLMS) adaptive filter can be very slow if the input correlation matrix is ill-conditioned as in the case with speech input. Second, very high order adaptive filters are computationally expensive. One application that has both speech input and typically needs a very high order adaptive filter is acoustic echo cancellation.

Research in subband acoustic echo cancelers (AECs) (Figure 1) is on-going and notable progress has been made [1], [2]. Research into these systems have either used Gaussian white noise, USASI noise (as an approximation for speech signals) [3], or short term speech signals. More recently, research in the statistical modeling of speech signals using spherically invariant random processes (SIRPs) [4] and analyses of LMS and NLMS algorithms under SIRPs [5] has led to a better understanding of the convergence characteristics of adaptive filters under more realistic input in the context of AEC.

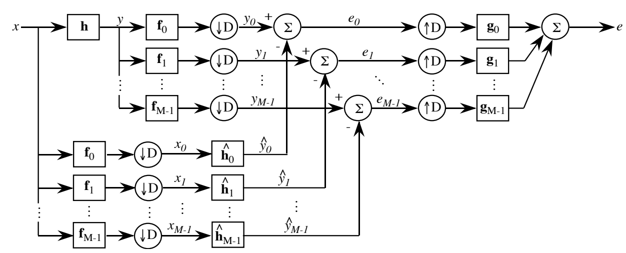

Figure 1: Subband acoustic echo canceler.

In this paper, we present experimental results using a subband AEC under SIRP input. These results are then evaluated using the talker-echo model so as to give an indication of the subjective quality rating of the echo cancellation. This leads to a better understanding of the performance in actual AEC applications and highlights the benefits of subband techniques.

2. SPHERICALLY INVARIANT RANDOM PROCESSES

In many DSP applications that process speech signals, it is desirable to have a statistical model for speech in order to predict system performance. One important aspect of the statistical model for speech is the probability density function (PDF). Typically, the Laplace or Gamma PDFs are used to model speech; however, both are approximations that can vary substantially from speaker to speaker and furthermore these may not properly model the corresponding bivariate speech PDF (taken from samples of speech signals spaced less than 5 ms apart) [4]. Toward this goal of a more refined statistical model for speech, SIRPs have been proposed to model speech sampled at 8 kHz and bandlimited from 300 – 3400 Hz. This particular speech signal characterizes the basic signal in telephone channels. The proposal to model the speech PDF with a SIRP is based on the facts that many random

This work was supported by the AT&T Cooperative Research Fellowship Program.

processes (including the Laplace and Gamma PDFs) are SIRPs and that actual speech PDFs have been shown to exhibit SIRP-like qualities [4].

In [4], G-functions are used to express SIRPs and simplify calculations involving SIRPs. Based on this work, the $G_{0\,2}^{2\,0}\left(\lambda x^2 \,|\, b_1, b_2\right)$ class of functions which includes the Gaussian, Laplace, Gamma, and K_0 PDFs, was found to be especially suited to modeling the measured densities of speech signals.

As our speech model, we have used a SIRP generator [6] implemented according to the algorithm described in [4] and selected the PDF to be the G-function $G_{0\,2}^{2\,0}\left(\lambda x^2 \,|\, -\frac{1}{3}, 0.3\right)$. The AR model, used in correlating the random process, is chosen to match Speaker 1 in [4]. The correlations from this model are illustrated in Figure 2.

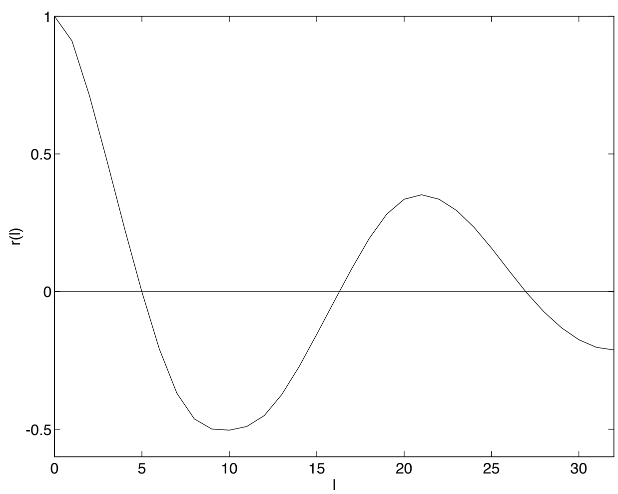

Figure 2: Speech signal correlations.

The G-function (with $b_1 = -\frac{1}{3}$ and $b_2 = 0.3$) along with the AR model yield an excellent statistical model (both univariate and bivariate PDFs match actual speech samples) for Speaker 1 of [4]. A single realization of this SIRP PDF is shown in Figure 3 and the long-term spectrum is given in Figure 4 .

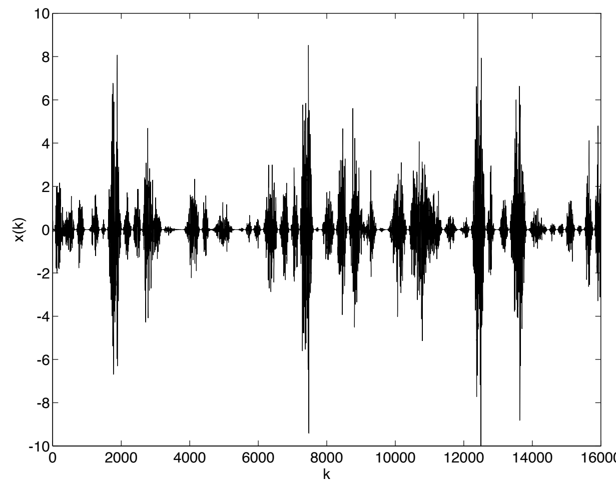

Figure 3: $G_{0\,2}^{2\,0}\left(\lambda x^2 \,|\, -\frac{1}{3}, 0.3\right)$ SIRP realization.

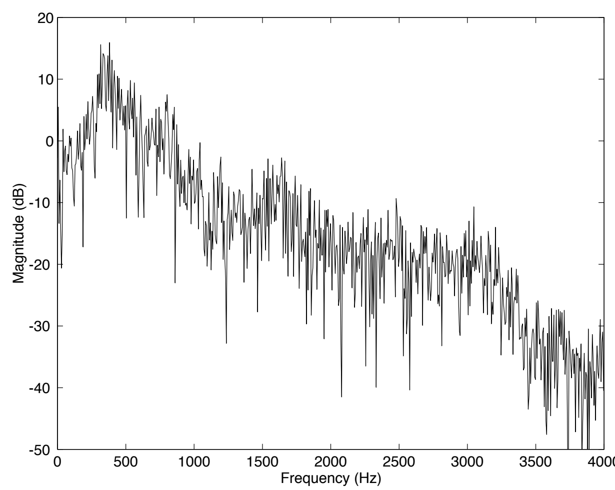

Figure 4: $G_{0\,2}^{2\,0}\left(\lambda x^2 \,|\, -\frac{1}{3}, 0.3\right)$ SIRP long-term spectrum.

3. EXPERIMENTAL RESULTS

Figure 1 illustrates the M / D oversampled, M-subband AEC. When $D < M$, this system suffers from slow asymptotic convergence of the subband MSE and consequently a slow asymptotic convergence of the fullband MSE due to small eigenvalues generated by the band edge of the subband input power spectrum [7]. It has been shown [2] that by increasing the bandwidth of the analysis filters $(\mathbf{f}_0, \mathbf{f}_1, \cdots, \mathbf{f}_{M-1})$ relative to the synthesis filters $(\mathbf{g}_0, \mathbf{g}_1, \cdots, \mathbf{g}_{M-1})$ the slowly converging spectral components of the subband error signals are shifted further out in the spectrum. Subsequent synthesis filtering removes these components thereby increasing the convergence of the MSE.

The simulations use a $4/3$ oversampled, 4-subband adaptive filter system and an identical one with 100% increased bandwidth analysis filters [2] under SIRP input. A single (fullband) adaptive filter system is also used for comparison. All systems use a simulated acoustic echo path (64 ms in duration) and the NLMS algorithm ($\tilde{\mu} = 1$) for adaptive filter update. Figure 5 compares the MSE (averaged over 100 runs) of the various systems under SIRP input. From this, convergence rates (after the initial fast convergence period) over the observation period for the three systems are calculated in Table 1.

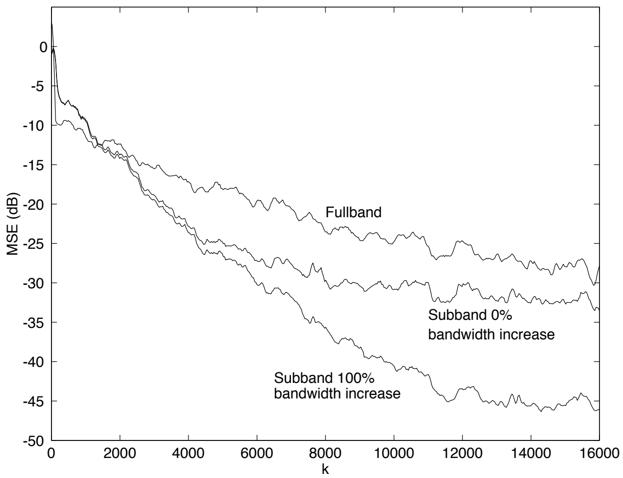

Figure 5: MSE for the subband adaptive filter systems with 0% and 100% analysis filter bandwidth increase and the fullband adaptive filter system under SIRP input.

Table 1: Steady-state convergence rates (in dB per 1000 samples) for acoustic echo cancelers under speech input modeled by SIRPs.

	Acoustic Echo Canceler		
	Fullband	**Subband**	**Subband (100% BW Increase)**
Convergence Rate	–1.1 dB/1000 samples	–1.5 dB/1000 samples	–2.4 dB/1000 samples

It is clear from Table 1, the convergence advantage subband systems have over the equivalent fullband system under speech input. Furthermore, slow band edge convergence under this input can be largely eliminated by increasing the bandwidth of the analysis filters relative to the synthesis filters. This was previously demonstrated for white noise [2].

4. SUBJECTIVE EVALUATION

Subjective studies on the effects of echo in telephone transmission quality were conducted during the 1950's and 1970's by the Bell Telephone Laboratories [8]. The source of echo was the impedance mismatch at the 2-to-4 wire hybrid in the telephone network. These subjective studies led to the talker-echo model which gives a transmission rating, R_E as a function of talker-echo path loss, E, in dB and delay, D, in milliseconds [8]:

$$R_E = 95.01 - 53.45\log_{10}\left(\frac{1+D}{\sqrt{1+(D/480)^2}}\right) + 2.277E. \quad (1)$$

The talker-echo path loss represents the amount by which the talker's speech signal is attenuated when traversing the path between the talker's lips to the point of impedance mismatch and then back to the talker's ears. In acoustic echo cancellation, the MSE of the AEC measures of the attenuation of the return echo power or the attenuation of the talker's speech signal through the echo path (in the acoustic echo cancellation application, the hybrid is replaced by the room). Thus the talker-echo path loss is the negative of the MSE, in dB. In the acoustic echo case as simulated, the paths between the talker's lips and the remote loudspeaker and the remote microphone and the local loudspeaker are neglected and thus the echo path is simply the room reverberation characteristic.

Assuming a normal population, the proportion of subjects rating the transmission as good or better, β, can be computed from R_E by

$$\beta = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\gamma} \exp\left(-\frac{t^2}{2}\right)dt \quad (2)$$

where γ is derived from opinion distributions for a particular study and expressed as a function of R_E in Table 2.

Table 2: Parameter γ formulas for two echo studies

Study	γ
1	$\frac{R_E - 75.05}{14.30}$
2	$\frac{R_E - 66.66}{11.84}$

Figure 6 contains a plot of transmission rating, R_E as a function of talker-echo path loss, E for various echo path delays. The delay times are chosen to be representative of typical reverberation times in rooms. Included in this figure are transmission rating conversions to the proportion of subjects commenting the transmission quality is good or excellent as averaged from both studies.

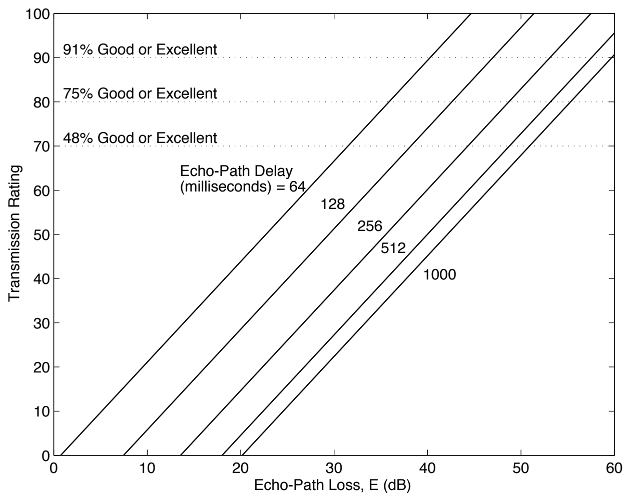

Figure 6: Transmission rating for talker echo.

Using the simulation results in Figure 5 and the transmission rating in Figure 6 for a 64 ms echo path (as used in simulations), the conclusions in Table 3 can be inferred using the average γ for the two studies in Table 2.

Table 3: Subjective quality of echo cancelers.

	Proportion of Subjects Rating Quality as Good or Excellent at Elapsed Time		
Acoustic Echo Canceler	**0.5 s**	**1 s**	**2 s**
Fullband	3%	21%	38%
Subband (0% Increase)	9%	38%	71%
Subband (100% Increase)	13%	80%	99%

These results show that the subband adaptive filter system under speech input [modeled by the correlated $G_{0\,2}^{2\,0}\left(\lambda x^2 \middle| -\frac{1}{3}, 0.3\right)$ SIRP] shows substantial improvement in subjective quality of echo cancellation when compared to the fullband system. Furthermore, the subband adaptive filter system with increased bandwidth analysis filters shows an almost equal improvement over the conventional subband system.

5. CONCLUSIONS

In this paper we have presented experimental results of subband adaptive filter systems under SIRP input. By using SIRPs to closely model speech signals, these results give an indication of the expected performance under actual speech input. These results are evaluated using the talker-echo model to indicate the subjective approval rating. Our results verify the performance advantage subband systems have over fullband systems in applications where the input correlation matrix is ill-conditioned as in the case with speech. Finally, we have demonstrated the advantage of increased bandwidth analysis filters at reducing the slow asymptotic convergence effects found in oversampled subband systems under SIRP input.

REFERENCES

[1] Morgan, D.R. and Thi, J.C., “A delayless subband adaptive filter achitecture,” *IEEE Transactions on Signal Processing*, vol. 43, no. 8, 1819-1830, Aug. 1995.

[2] DeLeón, P.L. and Etter, D.M., “Experimental results with increased bandwidth analysis filters in oversampled, subband acoustic echo cancelers,” *IEEE Signal Processing Letters*, vol. 2, no. 1, 1-3, Jan. 1995.

[3] Gillore, A. and Vetterli, M., “Adaptive filtering in subbands with critical sampling: analysis, experiments, and application to acoustic echo cancellation,” *IEEE Transactions on Signal Processing*, vol. 40, no. 8, 1862-1875, Aug. 1992.

[4] Brehm, H. and Stammler, W., “Description and generation of spherically invariant speech-model signals,” *Signal Processing*, vol. 12, no. 2, 119-141, Mar. 1987.

[5] Rupp, M., “The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes,” *IEEE Transactions on Signal Processing*, vol. 41, no. 3, 1149-1160, Mar. 1993.

[6] Werner, M., *A sirp generator for MATLAB*. http://www.mathworks.com, Jul. 91.

[7] Morgan, D.R., “Slow asymptotic convergence of LMS acoustic echo cancelers,” *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 2, 126-136, Mar. 1995.

[8] Cavanaugh, J.R., Hatch, R.W., and Sullivan, J.L., “Models for the subjective effects of loss, noise, and talker echo on telephone connections,” *The Bell System Technical Journal*, vol. 55, no. 9, 1319-1371, Nov. 1976.