ELectronics EXpress

IEICE Electronics Express, Vol.4, No.24, 762–767

Residual echo reduction based on MMSE estimator in acoustic echo canceller

Joon-Hyuk Chang[1a)], **Hyoung-Gon Kim**[2], **and Sangki Kang**[3]

[1] *School of Electronic and Electrical Engineering, Inha University, Incheon, 402–751, Korea*

[2] *Image Media Research Center, Korea Institute of Science Technology (KIST), Seoul, 136–791, Korea*

[3] *Telecommunication R&D Center, Samsung Electronics, Suwon, P.O. Box 105, Korea*

a) *changjh@inha.ac.kr*

Abstract: In this paper, a residual echo cancellation method is proposed that uses an estimation of the minimum mean-square error (MMSE) based on a statistical model of a speech signal and an echo signal. After the suppression of the echo signal based on the adaptive filter, residual echo is further reduced by the proposed MMSE estimator and the results are compared with the conventional Wiener filter based method.

Keywords: acoustic echo cancellation, residual echo, MMSE, wiener filtering

Classification: Science and engineering for electronics

References

[1] S. J. Park, C. G. Cho, C. Lee, and D. H. Youn, "Integrated echo and noise canceler for hands-free applications," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 49, no. 3, March 2002.

[2] S. J. Park, C. Lee, and D. H. Youn, "A residual echo cancellation scheme for hands-free telephony," *IEEE Signal Process. Lett.*, vol. 9, no. 12, Dec. 2002.

[3] V. Turbin, A. Gilloire, and P. Scalart, "Comparison of three post-filtering algorithms for residual echo reduction," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 307–310, 1997.

[4] F. Basbug, K. Swaminathan, and S. Nandkumar, "Integrated noise reduction and echo cancellation for IS-136 systems," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 1863–1866, 2000.

[5] J.-H. Chang and N. S. Kim, "Speech enhancement: new approaches to soft decision," *IEICE Trans. Inf. & Syst.*, vol. E84-D, no. 9, Sept. 2001.

[6] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

DOI: 10.1587/elex.4.762
Received October 18, 2007
Accepted November 20, 2007
Published December 25, 2007

[7] O. Cappé, "Elimination of musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 2, pp. 345–349, April 1994.
[8] R. J. McAulary and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, pp. 137–145, April 1980.
[9] J.-H. Chang, S. Gazor, N. S. Kim, and S. K. Mitra, "Multiple statistical models for soft decision in noisy speech enhancement," *Pattern Recognition*, vol. 40, no. 3, pp. 1123–1134, March 2007.
[10] S. Gustafsson, R. Martin, P. Jax, and P. Vary, "A pyschoacoustic approach to combined acoustic echo cancellation and noise reduction," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 245–256, July 2002.

1 Introduction

An acoustic echo canceller (AEC) is an essential component in a variety of commercial signal processing systems such as the hands-free mobile communication system and the PC messenger system [1, 2, 3, 4]. The AEC system involves generating the sum from reflected echoes of the original speech, then subtracting this from any signal the microphone picks up. However, in many practical situations, residual echo still exists at the output of the adaptive echo canceller as well due to the misadjustment of the adaptive filter and the constraint of the filter length on the actual processor. Additionally, performance degradation exists in the form of a perceptually residual echo background noise exists in the near-end talker speech signal [5, 6, 7, 8, 9].

Much work has been dedicated to the problem of overcoming this artifact. One relevant approach now popular is a postfilter such as Wiener filtering coupled with spectral subtraction [3]. In this technique, a postfilter is designed to reduce the size of the echo canceller while preserving a high amount of echo attenuation. One of the major degradation issues present in a postfilter employing the Wiener filter is "musical noise" which is the temporal discontinuity in the noise suppression filter in the near-end talker signal [8, 9]. Another important approach involves a whitening process based on autoregressive analysis. Specifically, in [1], it was shown that a noise reduction system followed by a whitening process could successfully reduce residual echo as well as background noise.

In this paper, a minimum mean-square error (MMSE) short time spectral amplitude estimator is proposed for residual echo reduction. It is noted that our approach is based on a statistical model of the residual echo signal. By assigning the statistical model to the microphone input signal and echo signal, the MMSE estimator of the postfilter is derived for the residual echo reduction. One advantage of this method is the smoothness of the main parameter in the MMSE parameter that helps reduce the musical noise.

DOI: 10.1587/elex.4.762
Received October 18, 2007
Accepted November 20, 2007
Published December 25, 2007

2 Review of AEC

To begin this review, an echo signal $d(k)$ and a background noise signal $n(k)$ are added to the near-end talker signal $s(k)$, and their sum is being denoted by $z(k)$, as follows [1]:

$$z(k) = s(k) + d(k) + n(k). \tag{1}$$

Acoustic echo cancellation is performed using an FIR adaptive filter. Fig. 2 shows a basic diagram of the proposed technique including the double talk detection (DTD) routine. The adaptive filter creates a replica $\hat{d}(k)$ of the echo signal $d(k)$. When this replica is subtracted from the overall near-end signal $z(k)$, the echo is suppressed such that

$$e(k) = z(k) - \hat{d}(k). \tag{2}$$

The output of the echo canceller $e(k)$ is used to adjust the coefficients $\hat{\mathbf{w}}(k)$ of the adaptive filter using an adaptation algorithm so that the filter coefficients converge to a close representation of the echo path $\mathbf{w}(k)$. For the adaptation algorithm, the normalized least mean-square (NLMS) algorithm is used in this study. This adaptive algorithm is widely accepted adaptive algorithm due to its relative simplicity and overall performance as follows:

$$\hat{\mathbf{w}}(k+1) = \hat{\mathbf{w}}(k) + \frac{\mu}{P_x(k) + \beta}\mathbf{x}(k)e(k), \tag{3}$$

where $\hat{\mathbf{w}}(k) = [\hat{w}_0(k), \hat{w}_1(k), \cdots, \hat{w}_{N-1}(k)]^T$ is an $(N \times 1)$ weight vector, and $\hat{x}(k) = [x(k), x(k-1), \cdots, x(k-N+1)]^T$ is an $(N \times 1)$ input vector. Additionally, T denotes the matrix transpose, μ is a constant controlling the convergence and $P_x(k)$ is the power of the input signal. Particularly, β is the stabilization factor adaptive echo cancellation is accomplished by automatically estimating a replica of the echo path response $w(k)$, synthesizing an estimated echo $\hat{d}(k)$ and subtracting the estimated echo from the echo path output $z(k)$ such that $e(k) = z(k) - \hat{z}(k)$.

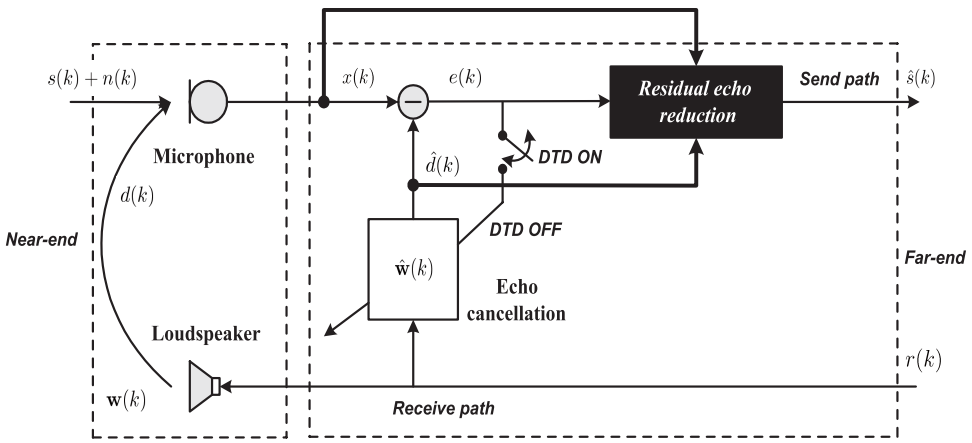

Fig. 1. Block diagram of the proposed acoustic echo canceller system.

DOI: 10.1587/elex.4.762
Received October 18, 2007
Accepted November 20, 2007
Published December 25, 2007

3 Residual Echo Reduction using MMSE Estimator

To achieve the reduction of the residual echo that exists at the output of the echo canceller, many postprocessors based on short-time spectral noise attenuation system have been proposed. Essentially, however, postprocessors based on the noise attenuation routines such as the Wiener filter rarely reduces the residual echo without distorting the near-end talker speech [8]. After the application of the spectral attenuation, the short-time magnitude spectrum in the frequency bands that originally contained noise now consists of a succession of randomly spaced spectral peaks. This artifact is known as the "musical phenomenon" which is created by temporal discontinuity in the noise suppression filter [9].

To avoid the musical phenomenon described above and obtain a significant residual echo reduction, an MMSE estimator for near-end signal that assigns a statistical model for the near-end signal and echo signal is proposed. Assuming that the background noise is not taken into account, the output of the microphone through the Fourier transform can be given as

$$X_f(t) = S_f(t) + D_f(t), \quad f = 1, 2, \cdots, M, \tag{4}$$

where f denotes the fth frequency bin, M is the total number of frequency components, and t is the frame index in the time domain. If $S_f = A_f \exp(j\phi_f)$ and $X_f = B_f \exp(j\psi_f)$, the MMSE estimator $\hat{A}_f$ of A_f is then obtained as follows:

$$\hat{A}_f = E\{A_f|X_f\} \tag{5}$$

$$= \frac{\int_0^\infty \int_0^{2\pi} a_f p(X_f|a_f,\phi_f) p(a_f,\phi_f) da_f d\phi_f}{\int_0^\infty \int_0^{2\pi} p(X_f|a_f,\phi_f) p(a_f,\phi_f) da_f d\phi_f} \tag{6}$$

Here, $E(\cdot)$ denotes the expectation operator and $p(\cdot)$ denotes a probability density function (pdf). With the complex Gaussian and Rayleigh pdf assumption, $p(X_f|a_f,\pi_f)$ and $p(a_f,\pi_f)$ are given by

$$p(X_f|a_f,\phi_f) = \frac{1}{\pi\lambda_{d,f}} \exp\Big\{-\frac{1}{\lambda_{d,f}}|X_f - a_f e^{j\phi_f}|^2\Big\}$$

$$p(a_f,\phi_f) = \frac{a_f}{\pi\lambda_{a,f}} \exp\Big\{-\frac{a_f^2}{\lambda_{a,f}}\Big\} \tag{7}$$

in which $\lambda_{a,f} = E\{|X_f|^2\}$ and $\lambda_{d,f} = E\{|D_f|^2\}$ are the variances the fth spectral component of the near-end speech and the echo signal, respectively. Substituting (7) into (6) results in the reduction gain for the residual echo based on the MMSE estimator as follows:

$$\hat{A}_f = \frac{\sqrt{\pi}}{2}\sqrt{\frac{\eta_f}{\gamma_f(1+\eta_f)}} M\Big[\frac{\gamma_f\eta_f}{1+\eta_f}\Big]|E_f|, \tag{8}$$

where $|E_f|$ is the magnitude of the Fourier transform of $e(k)$ and

$$M[\theta] = \exp\Big(-\frac{\theta}{2}\Big)\Big[(1+\theta)I_0\Big(\frac{\theta}{2}\Big) + \theta I_1\Big(\frac{\theta}{2}\Big)\Big] \tag{9}$$

DOI: 10.1587/elex.4.762
Received October 18, 2007
Accepted November 20, 2007
Published December 25, 2007

with I_0 and I_1 representing the modified Bessel function of the zero- and first-order, respectively. In addition, the *a posteriori* signal-to-echo ratio (SER) γ_f is given by

$$\gamma_f(t) = \frac{|X_f(t)|^2}{\lambda_{d,f}(t)} \tag{10}$$

The *a priori* SER η_f is computed by the well-known decision-directed method as follows [6]:

$$\eta_f(t) \quad = \quad \alpha \frac{|\hat{S}_f(t-1)|^2}{\lambda_{d,f}(t)} + (1-\alpha)u\Big[\hat{\gamma}_f(t) - 1\Big] \tag{11}$$

where α is the weighting term, e.g., 0.98, and $\lambda_{d,f}(t)$ is estimated by

$$\lambda_{d,f}(t) \quad = \quad \zeta_d \lambda_{d,f}(t-1) + (1-\zeta_d)|\hat{D}_f(t)|^2 \tag{12}$$

Here, ζ_d is the long-term smoothing factor and $\hat{D}_f(t)$ is obtained from the Fourier transform of the output of the adaptive filter.

In [7], the *a priori* signal-to-noise ratio (SNR) is known as the dominant parameter for noise suppression. Strong reductions are obtained only if η_f is low while low reductions are made only if η_f is high. In addition, when γ_f is sufficiently low, η_f corresponds to a highly smoothed version of the *a posteriori* SNR over successive short-time frames. Consequently, the attenuation itself does not show large variations over successive frames, as the *a priori SNR* has a significantly reduced variance. For that reason, the musical noise (sinusoidal components appearing and disappearing rapidly over successive frames) is significantly reduced. In a similar reason, we present the MMSE estimator as in (8) is utilized for effective residual echo reduction incorporating the *a priori* SER in (11). Fig. 2 shows a typical example of

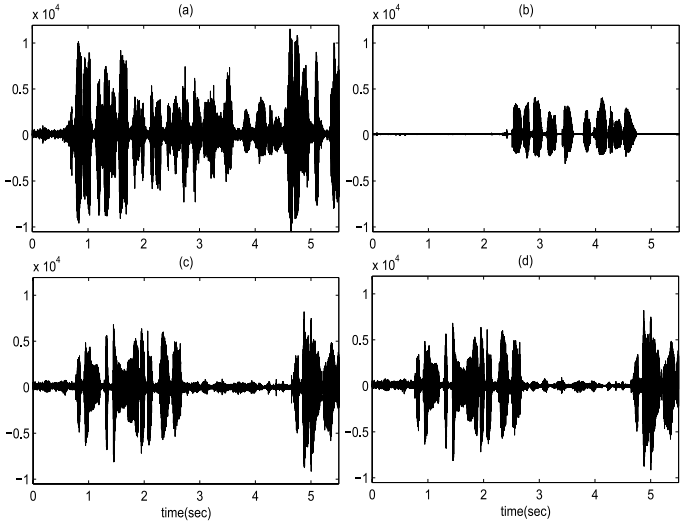

Fig. 2. Comparison of the speech waveform under the babble noise at SNR = 20 dB (a) Near-end speech (b) Reference speech (c) Output speech with residual echo reduced by a Wiener filter (d) Output speech with residual echo reduced by the MMSE.

DOI: 10.1587/elex.4.762
Received October 18, 2007
Accepted November 20, 2007
Published December 25, 2007

Table I. Result of the ERLE scores (dB).

SNR (dB)	Wiener	Gustafsson [10]	MMSE
30	26.2	25.9	24.2
20	25.4	24.5	20.3
10	21.9	20.3	17.9

a speech segment with reduced echo by the conventional Wiener filter and the proposed MMSE estimator. Also depicted is the corresponding near-end speech and reference speech. It should be noted that the residual echo is strongly suppressed while inaudible musical noise is maintained. As shown in the figure, this technique has clear advantages in terms of echo reduction conditions that are free from musical noise at the AEC output.

4 Experimental Results

In order to evaluate the performance of the proposed approach, a set of P.862 tests which have been widely adopted in the field of speech enhancement [9]. Twenty test sentences, ten of which were generated by a male speaker and ten by a female speaker, were used for the near-end signal $s(k)$, while a sentence spoken by a male was used for the far-end signal $r(k)$. Both signals were sampled at 8 kHz with a frame size of 10 ms. In order to create noisy condition, babble, vehicular noises from the NOISEX-92 database were added to clean near-end speech signals while the SNR was varied. For the purpose of comparison, we evaluated the performance of the Wiener filter-based system and the one with the residual echo suppression algorithm proposed by Gustafsson *et al.* [10]. The performance of the approach was measured in terms of echo return loss enhancement (ERLE) [10]. Overall ERLE results are shown in Table I. From the ERLE results, it is evident that in most noisy conditions, the proposed MMSE estimator yielded a higher score compared to the previous techniques. In addition, we can observe that the audible "musical noise" is significantly reduced using the proposed approach compared to the previous works [3, 10], as the undesirable fluctuation of the *a priori* SER is diminished when we the MMSE estimator based on the DD method is used.

5 Conclusions

A residual echo reduction scheme based on the MMSE estimation is proposed. The audible musical noise of the residual echo with the proposed approach is significantly reduced as we employ the *a priori SER* based on the MMSE estimator of which the attenuation itself does not show much variation over successive frames.

Acknowledgments

This work was supported by the IT R&D program of MIC&IITA, 2007-S001-01, Development of an Intelligent Service technology based on the PLL.

DOI: 10.1587/elex.4.762
Received October 18, 2007
Accepted November 20, 2007
Published December 25, 2007