ON INTEGRATING ACOUSTIC ECHO AND NOISE CANCELLATION SYSTEMS FOR HANDS-FREE TELEPHONY

Seon Joon Park, Chum Gun Cho, Chungyong Lee, and Dae Hee Youn

ASSP Lab., Dept. of Electrical and Electronic Eng., Yonsei University
134 Shinchon-dong, Sudaemoon-ku, Seoul 120-749, Korea
E-mail: [seonjoon, chronos, clee, dhyoun]@assp.yonsei.ac.kr

ABSTRACT

An integrated acoustic echo and noise cancellation system for hands-free telephony is presented. The proposed system includes a new residual echo cancellation scheme based on spectral analysis and a new double-talk detector suitable for real-time implementation. Residual echo is whitened via AR analysis during no near-end-talk period and is cancelled by noise reduction. Removing speech characteristics of the residual echo signal, noise reduction successfully reduces the power of the residual echo as well as the ambient noise. For further integration with commercial low-bit rate speech coders, noise reduction in IS-127 (EVRC) was considered. For the hands-free situation in the moving car, the proposed system attenuated the interferences more than 30 dB at a constant speed of 90 km/h. The proposed system was implemented on a low-cost DSP with 16-bit fixed-point arithmetic.

1. INTRODUCTION

For hands-free mobile communications, the problem of combining acoustic echo canceller (AEC) and noise reduction (NR) has been considered to achieve sufficient quality of the transmitted speech signals [1]. In practice, residual echo remains at the output of an adaptive echo canceller. This phenomenon is due to the misadjustment of the adaptive algorithm and constraint of short filter length imposed on the actual processors [2]. A postprocessor, e.g., NR, has been used for reducing the residual echo as well as the ambient noise [3, 4]. However, the postprocessors based on NR rarely reduce the residual echo without distortion in the near-end talker (NET) speech, since the dominant characteristics of the residual echo is also speech. Although the power of the residual echo is very small, a VAD used in NR may misjudge residual echo as transmitting speech, which results in transmission of the residual echo to a far-end talker (FET).

This work was supported by Hyundai Electronics Industries Co., Ltd.

In this paper, an integrated acoustic echo and noise cancellation system with a computationally efficient double-talk detector (DTD) and a new residual echo cancellation (REC) method is proposed. In the proposed REC method, a whitening process based on autoregressive (AR) analysis of the speech signal is applied to the residual echo signal during no NET period. Then NR regards the whitened residual echo as the ambient noise, and tries to reduce it. Considering a real-time implementation of the proposed system using a low-cost DSP with 16-bit fixed-point arithmetic, a new double-talk detection (DTD) algorithm using two average cross-correlations is also proposed.

2. STRUCTURE OF THE PROPOSED SYSTEM

In the proposed system, we perform echo cancellation followed by noise reduction. Fig. 1 shows the integrated system structure including the proposed DTD and REC scheme, where $\hat{\mathbf{H}}(k)$ and $\mathbf{G}(k)$ indicate the AEC and the NR system, respectively. It is well-known that this not only prevents the echo canceller's performance from degrading due to nonlinearities caused by NR, but this also has advantage that the noise reduction may reduce the residual echo [5].

Echo cancellation is performed by using an adaptive filter. Referring to Fig. 1, 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 $y(k)$, the echo is

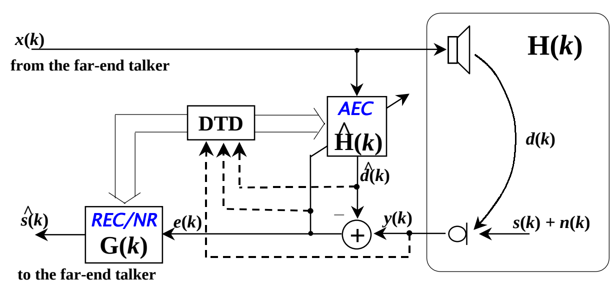


Fig. 1. Combined structure of acoustic echo cancellation and noise reduction.

eliminated.

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

Here, $s(k)$ is the NET signal and $n(k)$ is the ambient noise signal. The output of the AEC, or the error signal, $e(k)$ is used to adjust the coefficients $\hat{\mathbf{H}}(k)$ of the adaptive filter by using an adaptation algorithm so that the coefficients converge to a close representation of the echo path $\mathbf{H}(k)$.

$$\begin{aligned} e(k) &= s(k) + n(k) + d(k) - \hat{d}(k) \\ &= s(k) + n(k) + r(k) \end{aligned} \tag{2}$$

where $r(k) = d(k) - \hat{d}(k)$ indicates residual echo.

An adaptive FIR filter with 256 coefficients is applied to the AEC, and the NLMS algorithm is used for adaptation. The coefficients of the adaptive filter $\hat{\mathbf{H}}(k)$ are updated according to:

$$\hat{\mathbf{H}}(k+1) = \hat{\mathbf{H}}(k) + \frac{\mu}{||\mathbf{X}(k)||^2} \cdot e(k)\mathbf{X}(k), \tag{3}$$

where $\mathbf{X}(k)$ is the excitation vector and μ is a step-size.

For stable operation of the AEC, the proposed DTD algorithm controls the adaptive algorithm. For canceling the residual echo $r(k)$ and the ambient noise $n(k)$, NR with the proposed REC scheme is employed. Considering the further integration with commercial low-bit rate speech coders, NR in IS-127 (EVRC) is used [6].

3. DOUBLE-TALK DETECTION AND RESIDUAL ECHO CANCELLATION

3.1. Double-talk detection

To stabilize the operation of the AEC, the DTD is used to freeze the adaptation of the filter coefficients $\hat{\mathbf{H}}(k)$ when the NET exists. Several authors have proposed to use the cross-correlation vector for the DTD, which improves the DTD performance [7]. Those methods are quite efficient for the DTD with appropriate thresholds, which are hardly available in real-environment. Furthermore, those are not appropriate for a real-time implementation with a limited processing power and 16-bit processing precision.

Based on those cross-correlation methods, we propose a DTD suitable for the real-time implementation using a low-cost DSP. The proposed DTD uses two average cross-correlations: i) the average cross-correlation $\rho_{\hat{d},y}(k)$ between the estimated echo $\hat{d}(k)$ and the microphone input $y(k)$ and ii) the average cross-correlation $\rho_{e,y}(k)$ between the AEC error $e(k)$ and the microphone input $y(k)$. $\rho_{\hat{d},y}(k)$ and $\rho_{e,y}(k)$ are defined as follows:

$$\rho_{\hat{d},y}(k) = \frac{P_{\hat{d},y}(k)}{\sqrt{P_{\hat{d}}(k)P_y(k)}} \tag{4}$$

where

$$\begin{aligned} P_{\hat{d},y}(k) &= (1-\lambda)P_{\hat{d},y}(k-1) + \lambda\hat{d}(k)y(k) \\ P_{\hat{d}}(k) &= (1-\lambda)P_{\hat{d}}(k-1) + \lambda\hat{d}^2(k) \\ P_y(k) &= (1-\lambda)P_y(k-1) + \lambda y^2(k) \end{aligned}$$

and

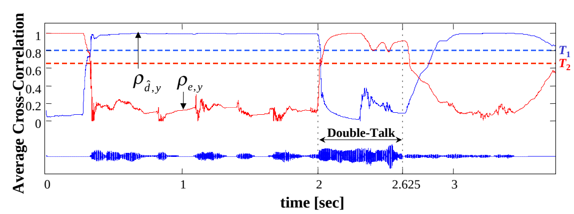
$$\rho_{e,y}(k) = \frac{P_{e,y}(k)}{\sqrt{P_e(k)P_y(k)}} \tag{5}$$

where

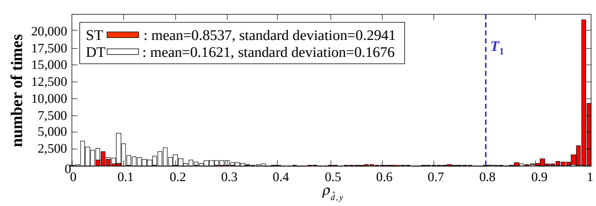
$$\begin{aligned} P_{e,y}(k) &= (1-\lambda)P_{e,y}(k-1) + \lambda e(k)y(k) \\ P_e(k) &= (1-\lambda)P_e(k-1) + \lambda e^2(k). \end{aligned}$$

In the above equations, $0 < \lambda < 1$ indicates a forgetting factor for average.

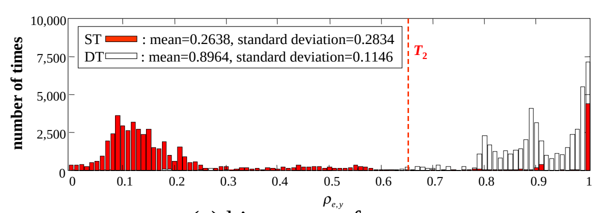
Fig. 2 shows values and histograms of $\rho_{\hat{d},y}(k)$ and $\rho_{e,y}(k)$ for hands-free telephony data measured in a car cabin. As shown in Fig. 2 (a), $\rho_{\hat{d},y}(k)$ has a value close to 1 at single-talk (ST) period. Thus, $\rho_{\hat{d},y}(k)$ is a good indicator for single-talk period. However, $\rho_{\hat{d},y}(k)$ has a long transient time when the situation changes from double-talk (DT) to single-talk. Therefore, it is hard to select an appropriate threshold for DTD in real environment. In contrast to $\rho_{\hat{d},y}(k)$, $\rho_{e,y}(k)$ has relatively a short transient time although it has large variance in single-talk period. Thus, it can be used as a complementing parameter.



(a) $\rho_{\hat{d},y}(k)$ and $\rho_{e,y}(k)$ measured in a car



(b) histogram of $\rho_{\hat{d},y}$



(c) histogram of $\rho_{e,y}$

Fig. 2. Average cross-correlations and their histograms.

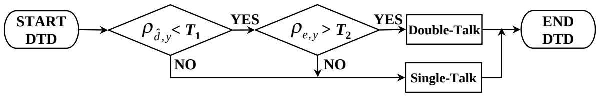


Fig. 3. The proposed double-talk detection algorithm.

Fig. 3 shows the decision flow of the proposed DTD. $\rho_{\hat{d},y}(k)$ is first calculated and compared with a threshold T_1. If $\rho_{\hat{d},y}(k)$ is less than or equal to T_1, single-talk situation is declared. Otherwise, the proposed DTD calculates $\rho_{e,y}(k)$ and compares it with another threshold T_2, since it possibly indicates double-talk situation. If $\rho_{e,y}(k)$ is less than T_2, double-talk situation is declared and the system freezes the filter coefficients updates. The thresholds T_1 and T_2 are decided based on the histograms of $\rho_{\hat{d},y}(k)$ and $\rho_{e,y}(k)$, which makes the probability of false alarm less than 0.1 except the initial period (see Fig. 2 (b) and (c)) [7]. Using the proposed DTD, start and end points of double-talk situation are detected fast in real situations.

3.2. Residual echo cancellation

As mentioned in the previous sections, the residual echo remains after AEC. Since the source of the echo signal is the FET speech, the residual echo also has speech characteristics. In the proposed residual echo cancellation scheme, a whitening process based on AR spectral analysis removes speech characteristics of the residual echo. Based on the well-known fact that the speech signal can be modeled as an AR model, the AR coefficients of the residual echo are estimated. Using the estimated AR coefficients, the residual echo is inverse filtered and is applied to NR. Therefore, in NR, the VAD regards the whitened residual echo signal as the ambient noise, and reflects to noise power updates.

As in Fig. 4, the proposed system whitens the residual echo signal $e(k)$ during echo-and-noise only period (i.e., no NET period) not to distort the transmitting NET signal. No NET period is determined by the proposed DTD. For the whitening process, Pth order AR analysis for $e(k)$ is performed and inverse filtering is applied by using the estimated AR coefficients:

$$w_e(k) = -\sum_{i=0}^{P} \hat{a}(i)e(k-i), \quad \hat{a}(0) = 1 \qquad (6)$$

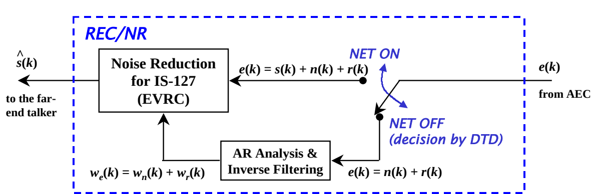


Fig. 4. The proposed residual echo cancellation scheme.

where $\hat{a}(k)$ is the estimated AR coefficients, and $w_e(k) = w_n(k) + w_r(k)$ is the whitened residual echo signal, which is applied as an input of the noise reduction system only for no NET period.

4. EXPERIMENTAL RESULTS

For instrumental evaluation, the speech data for the FET signal were constructed, which consisted of twelve phonetically balanced Korean sentences. Six of them were spoken by a man and the other six by a woman. For the NET signals, six sentences pronounced by male and female speakers were used. Echo-and-noise data were measured in a middle-sized car for stopped and moving at a constant speed of about 90 km/h. All data were sampled at 8 kHz sampling rate and stored in a 16-bit integer format. Prior to usage, all input signals were filtered with a band pass filter with the telephone bandwidth (300 Hz to 3400 Hz).

4.1. Double-talk detection

The proposed DTD was applied to the integrated acoustic echo and noise cancellation system. For the performance comparison, the conventional DTD using the normalized cross-correlation between the FET signal vector and microphone input signal was investigated [7]. Probability of false alarm was set to 0.1 for the conventional method. In Fig. 5, double-talk detection results for both methods are shown. With small computational complexity, the proposed method shows comparable performance to the conventional method and shows fast detection when double-talk occurs.

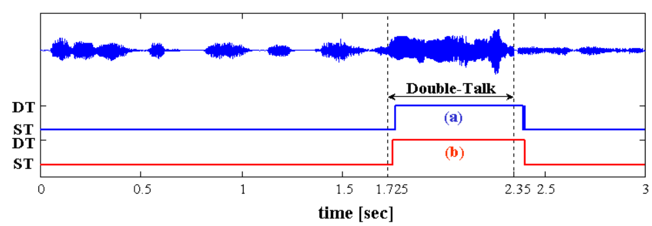


Fig. 5. Results of the DTD in the stopped car using (a) the conventional DTD [7] and (b) the proposed DTD. Double-talk occurred from 1.725 to 2.35 sec.

4.2. Interference cancellation

We define a parameter IC (interference cancellation) for the performance evaluation of the proposed system as follows:

$$IC = 10log_{10}\frac{y^2(k)}{\hat{s}^2(k)} = 10log_{10}\frac{y^2(k)}{\{y(k)-\hat{i}(k)\}^2} \; [dB] \quad (7)$$

where $\hat{s}(k)$ is the transmitted signal to the FET and $\hat{i}(k) = \hat{d}(k) + \hat{r}(k) + \hat{n}(k)$ is the estimated interferences, i.e., echo,

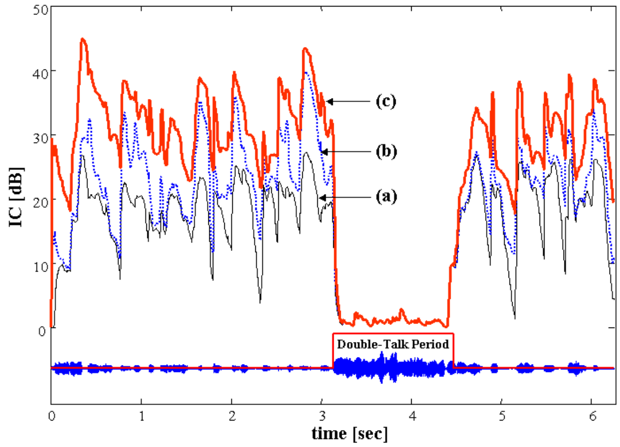


Fig. 6. Interference cancellation (IC) in the stopped car. (a) AEC only, (b) AEC+NR, (c) AEC+REC+NR

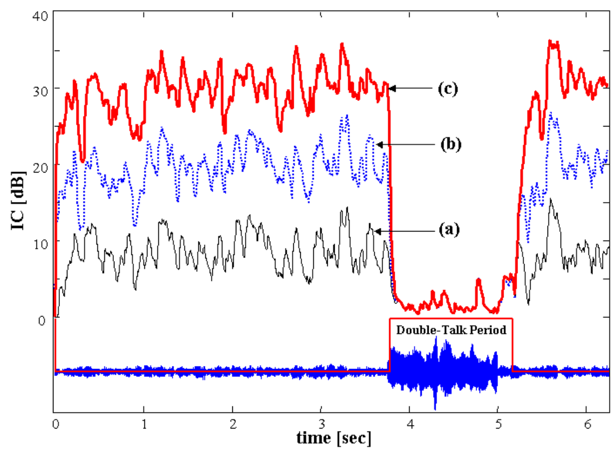


Fig. 7. Interference cancellation (IC) in the moving car (90km/h). (a) AEC only, (b) AEC+NR, (c) AEC+REC+NR

residual echo, and ambient noise.

Performance of the proposed system was compared with the AEC only system and the conventional AEC+NR system for stopped and moving situation. In the stopped car, the proposed system improves the IC performance more than average 12.6 dB compared to the AEC only system and average 7.4 dB compared to the conventional AEC+NR system (see Fig. 6). When the car is moving at a constant speed of 90 km/h in a highway, the proposed system improves the IC performance more than average 21.4 dB and 10.1 dB, respectively, compared to the conventional systems (see Fig. 7). The results indicates that the proposed system is robust to the noisy environment.

4.3. Implementation

The proposed system was implemented on the EVM board with OakDSPCore, which is a 16-bit general purpose low-power, low-voltage and high speed DSP core designed for speech/audio processing, telecommunications, digital cellular, and embedded control applications. For the overall system, 20.28 MIPS of computational power are required. This figure decomposes into 0.35 MIPS for the I/O control, 13.36 MIPS for AEC, 0.96 MIPS for DTD, 0.92 MIPS for REC and 4.69 MIPS for NR. We used only internal program and data memory in OAKDSPCore.

5. CONCLUSION

We presented the integrated system comprising echo cancellation and noise reduction. The proposed system includes the computationally efficient and robust double-talk detection algorithm and residual echo cancellation algorithm. The proposed system attenuated the interferences more than 30 dB in the moving car cabin. The algorithms were integrated in a real-time system using a low-cost DSP with 16-bit fixed-point arithmetic.

6. REFERENCES

[1] H. Puder and P. Dreiseitel, "Implementation of a hands-free car phone with echo cancellation and noise-dependent loss control," in *Proc. IEEE Intl. Conf. Acoust., Speech, Sig. Proc.*, 2000, pp. 3622–3627.

[2] S. M. Kuo and J. Chen, "Analysis of finite length acoustic echo cancellation system," *Speech Communication*, vol. 16, pp. 255–260, 1995.

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

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

[5] C. Beaugeant, V. Turvin, P. Scalart, and A. Gilloire, "New optimal filtering approaches for hands-free telecommunication terminals," *Signal Processing*, vol. 64, pp. 33–47, 1998.

[6] TIA/EIA/IS-127, *Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems*, 1997.

[7] J. Benesty, D. R. Morgan, and J. H., Cho, "A new class of doubletalk detectors based on cross-correlation," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 2, pp. 168–172, Mar. 2000.