IMPLEMENTATION OF A BASIC ACOUSTIC ECHO CANCELLER

Duong Quang K. Ngoc* and Seung-Hyon Nam**
Department of Electronic Engineering, Paichai University
Daejeon, Korea
*dqkngoc@gmail.com, **shnam@pcu.ac.kr

ABSTRACT

Acoustic echo cancellation (AEC) is an essential part of a hands-free system to eliminate acoustic echo received by a microphone. This paper presents an implementation of a basic acoustic echo canceller consisting of the blocks of band-pass filter, de-correlation filter, voice activity detection, double talk detection, normalized least mean square (NLMS) adaptive filter, non-linear processor, and optional auto gain control. In addition, a modified open-loop correlation method is proposed for reliable double talk detection. Simulation results using real audio recordings in a car confirm the expectation.

I. INTRODUCTION

Acoustic echo cancellation reduces the acoustic coupling between a loudspeaker and a microphone by estimating the impulse response of the loudspeaker-enclosure-microphone (LEM) system as shown in Fig. 1.

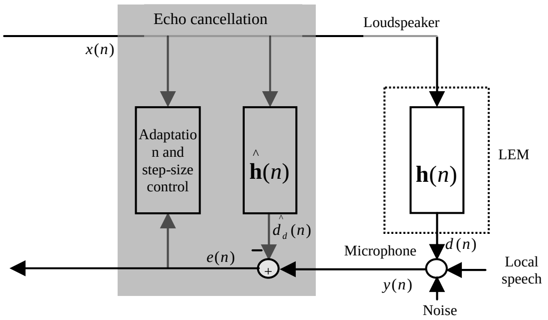

Fig. 1. Principle of acoustic echo cancellation.

The echo signal $d(n)$ which originates from the far-end speech $x(n)$ can be estimated by the output signal of the adaptive filter $\hat{d}(n)$ given by

$$\hat{d}(n) = \mathbf{x}(n) * \hat{\mathbf{h}}(n) = \sum_{i=0}^{M-1} x(n-i)\hat{h}_i(n) \tag{1}$$

where $\hat{h}_i(n)$ denotes the i^{th} coefficient of the tap-weight vector at time n of the adaptive filter and M is the filter length. Local speech can be obtained by subtracting this estimated echo from the microphone signal.

Due to the environmental changes such as the movement of local speaker, background noise, or temperature variance, the impulse response $\mathbf{h}(n)$ of the LEM system is time varying. Thus the impulse response of adaptive filter $\hat{\mathbf{h}}(n)$ should be estimated adaptively as close to $\mathbf{h}(n)$ as possible to reduce error. In this paper, the NLMS algorithm is adopted for this purpose.

Double talk detection is a typical problem that is deeply concerned in our AEC implementation to detect the situation in which both sides talk simultaneously. During the double talk period, the residual error increases due to local speech so that the stability bound decreases and the algorithm may start to diverge. This situation must be prevented and the adaptive filter coefficients must be frozen during the double talk. The correlation-based double-talk detection method is discussed and a new modified one is proposed in this paper.

Besides, for robustness, the implementation of AEC needs a band-pass filter (BPF) to remove unwanted low frequency and high frequency components of loud-speaker and microphone signals; a decorrelation filter (DCF) for faster convergence of NLMS adaptive algorithm; a voice activity detector (VAD) to detect the presence of far-end speech and a non-linear processor (NLP) to suppress the residual echo level after the NLMS adaptive filter.

II. AEC IMPLEMENTATION

The conceptual block diagram illustrating the operation of the AEC algorithm is shown in Fig. 2.

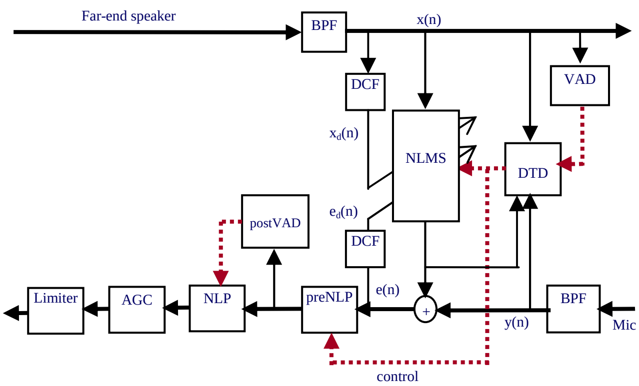

Fig. 2. Block diagram of AEC algorithm.

1.1. Band-pass filter

The band-pass filter is implemented by cascading a 2^{nd} order low-pass filter (LPF) and a 2^{nd} order high-pass filter (HPF) to provide the pass band of 100 Hz-3400 Hz. The Butterworth filter is used to provide the flat frequency response at band-pass region.

The transfer function of the LPF and HPF are given by

$$H_l = \frac{0.7157 + 1.4315z^{-1} + 0.7157z^{-2}}{1 + 1.3490z^{-1} + 0.5140z^{-2}} \tag{2}$$

and

$$H_h = \frac{0.9460 - 1.8920z^{-1} + 0.9460z^{-2}}{1 - 1.8890z^{-1} + 0.8949z^{-2}}, \tag{3}$$

respectively.

1.2. Decorrelation filter

It is known that a fixed first order decorrelation filter provides remarkable result [2]. The first order filter can be a simple high-pass filter since speech signals have low-pass characteristics. The formula to decorrelate the far-end and the error signals for NLMS tap-weight update are:

$$x_d(n) = x(n) - ax(n-1) \tag{4}$$

$$e_d(n) = e(n) - ae(n-1) \tag{5}$$

where $a = 0.85$.

Two decorrelation filters are necessary to decorrelate two signals that are fed into the NLMS adaptive filter as shown in Figure 2. This structure is very simple and efficient since there is no need to use inverse decorrelation filters.

1.3. Voice activity detection

A short-term magnitude estimation of input signal is computed as [2]

$$\overline{|x(n)|} = (1-\gamma(n))|x(n)| + \gamma(n)\overline{|x(n-1)|} \tag{6}$$

where

$$\gamma(n) = \begin{cases} \gamma_r, & \text{if } |x(n)| > \overline{|x(n-1)|} \\ \gamma_f, & \text{else.} \end{cases} \tag{7}$$

For 8 kHz sampling rate, we choose $\gamma_r = 0.995$ and $\gamma_f = 0.997$.

Voice activity of a far-end signal is detected if the short-term magnitude exceeds a predefined noise threshold of 35 dB. A hangover time of 250 ms is specified after the end of voice activity.

1.4. Double talk detection

To detect the double talk, correlation-based methods use degree of similarity between microphone signal $y(n)$ and the output signal $\hat{d}(n)$ of adaptive filter or the loudspeaker signal $x(n)$ in term of correlation. The detail of these methods is described in session 3.

During the double-talk periods, the AEC algorithm processes as follows:

+ The adaptive filter coefficients are frozen (no adaptation).

+ The far-end signal $x(n)$ is still filtered to get $\hat{d}(n)$.

+ The error signal is given by $e(n) = y(n) - \hat{d}(n)$.

1.5. NLMS adaptive filter

The tap-weight is updated during the non double-talk period by the rule of the NLMS algorithm given by [1]:

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu(n)\frac{e_d(n)\mathbf{x}_d(n)}{\|\mathbf{x}_d(n)\|^2} \tag{8}$$

where $\mu(n)$ is the step-size, $\mathbf{x}_d(n)$ and $e_d(n)$ are the loud-speaker and error signals after decorrelation, respectively, and $\hat{\mathbf{h}}(n) = (\hat{h}_o(n), \hat{h}_1(n), ..., \hat{h}_{M-1}(n))^T$ is the tap-weight vector of adaptive filter at time n.

The step-size governs the convergence rate and the misadjustment of the adaptive filter. In this paper, the step-size is chosen appropriately as follows [5]

$$\mu = \begin{cases} \mu_h & \text{if } E_x(n) \geq E_{threshold} \\ \mu_l & \text{if } E_x(n) < E_{threshold} \text{ and } \mu_l < \mu_h \\ 0 & \text{if double-talk is detected} \end{cases} \tag{9}$$

where $\mu_l = \mu_h/2$ and $E_{threshold} = 32768$.

1.6. Non-linear processor

A simple pre-NLP attenuates the error signal by 6 dB only if double talk is not detected. The main NPL attenuates remaining error signal by another 18 dB if post-VAD determines remaining error as noise. This *double NLP* structure has prominent advantages compared to single NLP in case of the delayed and wrong double-talk detection.

1.7. Auto gain control and limiter

AGC is an optional part of an AEC and is used to adjust local signal level before it is transmitted to the far-end side. Low level speech will be amplified while high level speech is attenuated (up to a certain level). This AEC is implemented by combining the slowly-changing gain factor and the rapidly-changing short-term signal power.

After processing, the outgoing signal might exceed the range and need to be limited by a limiter. A detail description for these blocks is presented in [2].

III. DOUBLE TALK DETECTION

In this paper, two structures are considered for double talk detection; open-loop and close-loop structures as illustrated in Fig. 3. Open-loop structure calculates the normalized correlation between the far-end signal $x(n)$ and the microphone signal $y(n)$ (Eq. (10)), it is independent of the adaptive filter and therefore called the open-loop structure [3]. The close-loop structure calculates the correlation between the microphone signal $y(n)$ and the output of the adaptive filter $\hat{d}(n)$ (Eq. (11)). In this case, estimation depends on the adjustment of the adaptive filter, and is called the closed-loop structure. If the adaptive is adjusted sufficiently, this structure provides accurate double-talk detection due to high similarity between the real echo $d(n)$ and the estimated echo $\hat{d}(n)$.

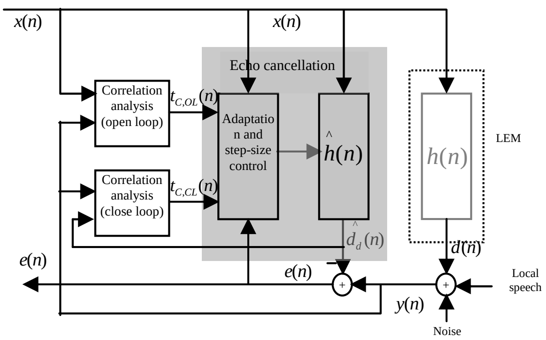

Fig. 3. Structure of correlation analysis.

The open-loop correlation factor is given by

$$\rho_{OL}(n) = \max_{l \in [0, L_C]} \frac{\left|\sum_{k=0}^{N_C-1} x(n-k-l)y(n-k)\right|}{\sum_{k=0}^{N_C-1} |x(n-k-l)y(n-k)|} \tag{10}$$

This is calculated for different delay l corresponding to the time delay of loudspeaker-microphone path. The parameter L_c is chosen to be greater than the time delay of the direct path between the loudspeaker and the microphone. In contrast, no delay has to be considered for the closed-loop correlation factor because both signals are synchronous through the adaptive filter. The close-loop correlation factor is given by:

$$\rho_{CL}(n) = \max_{l\in[0,L_C]} \frac{\left|\sum_{k=0}^{N_C-1} \hat{d}(n-k)y(n-k)\right|}{\sum_{k=0}^{N_C-1}\left|\hat{d}(n-k)y(n-k)\right|} \quad (11)$$

The closed-loop correlation factor depends on the working of the adaptive filter. The close-loop correlation factor can not be used to detect double talk when the adaptive filter does not adapt well enough due to LEM changes. For these reasons open-loop factor plays a key role since its value only depend on the loudspeaker and microphone signals that are independent of adaptive mechanism.

In this paper, a modified formula that calculates the open-loop factor between $x(n)$ and $x(n)-y(n)$ is proposed. The modified open-loop factor is given by:

$$\rho'_{OL}(n) = \max_{l\in[0,L_C]} \frac{\left|\sum_{k=0}^{N_C-1} x(n-k-l)[x(n-k-l)-y(n-k)]\right|}{\sum_{k=0}^{N_C-1}\left|x(n-k-l)[x(n-k-l)-y(n-k)]\right|} \quad (12)$$

Both correlation factors are calculated using N_C samples, where a larger number assures better estimation quality but may lead to delayed decision and instability. Since the correlation factor that varies over a wide range may cause wrong double-talk detection, we need to smooth out these values by the smoothing coefficient γ ranging from 0.950 to 0.995 as following:

$$\rho'_{OL}(n) = \gamma.\rho'_{OL}(n) + (1-\gamma).\rho'_{OL}(n-1) \quad (13)$$

$$\rho_{CL}(n) = \gamma.\rho_{CL}(n) + (1-\gamma).\rho_{CL}(n-1) \quad (14)$$

The factors $\rho'_{OL}(n)$ and $\rho_{CL}(n)$ are calculated at each block of 10 ms. A hang-over time of one 10 ms block is specified to detect non double talk situation from double talk. Double talk is determined if the calculated correlation value is smaller than a predetermined threshold.

IV. SIMULATIONS

1.8. Performance of AEC

Simulation setup: We measured impulse responses of an LEM in a car. The reverberation time is approximately 50 ms and the length is 512 samples at 8000 Hz sampling rate. AEC implementation is setup with NLMS adaptive filter of length 256. The step sizes are chosen as $\mu_h = 0.5$ and $\mu_l = 0.25$. Double talk is determined at each block of 80 samples with $N_C = 240$ and $L_C = 50$. The threshold values are 0.9 and 0.85 for open-loop and closed-loop, respectively.

Simulation result: Figure 4 shows the far-end speech and the microphone signals used in the simulation and the resulting error signal after AEC. To evaluate the performance of AEC, the system distance and the error return loss enhancement (ERLE) are calculated and shown in Fig. 5 and Fig. 6, respectively

The system distance measures the similarity between the estimated impulse response and the true LEM impulse response. It is calculated at time index n by equation:

$$D(n) = \sum_{i=0}^{M-1}\left(\hat{h}_i(n) - h_i\right)^2 \quad (15)$$

where $\hat{h}_i(n)$ and h_i denote the i^{th} coefficient of the adaptive filter and true LEM impulse response, respectively, at time n.

The ERLE measures the attenuation of the echo signal at the output of an AEC. Assuming that there is only a single far-end speech without any local noise, it is calculated at time index n as

$$ERLE(n) = \frac{\overline{y^2(n)}}{\overline{e^2(n)}} \quad (16)$$

where $\overline{y^2(n)}$ and $\overline{e^2(n)}$ are the power of the microphone signal and the error signals, respectively, averaged over a block.

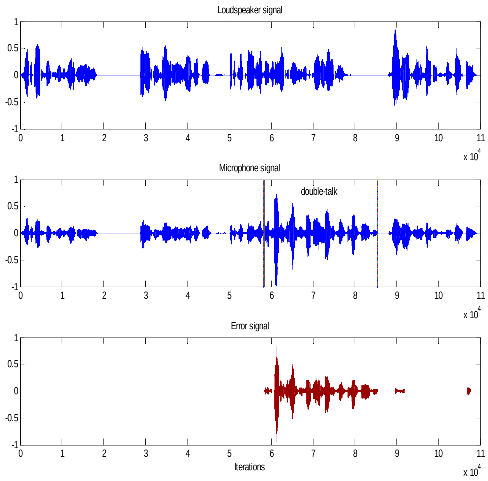

Fig. 4. The far-end speech, microphone, and error signals.

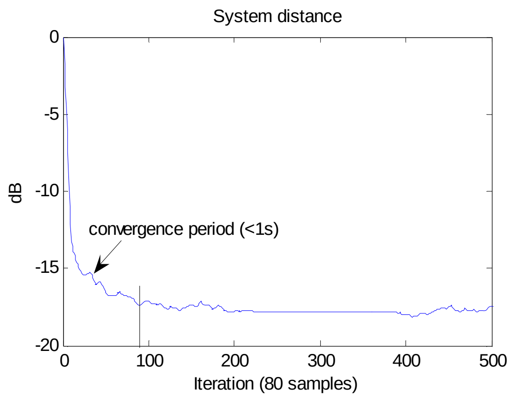

Fig. 5. System distance of the implemented AEC.

The convergence time shown in Fig. 5 is about 1 second and the average ERLE in single-talk period shown in Fig. 6 is higher than 40 dB that both meet the standard requirements for AEC.

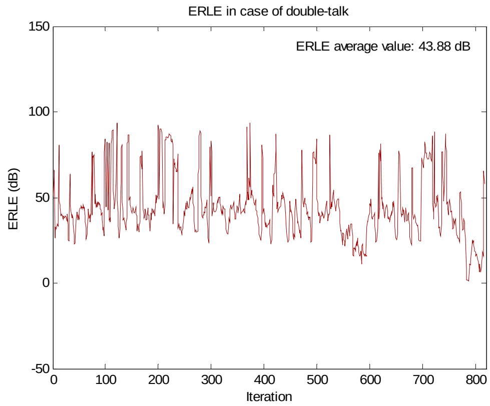

Fig. 6. The ERLE of the implemented AEC.

1.9. Performance of the modified double-talk detection

Our simulation in two cases below will show effectiveness of the proposed open-loop correlation factor:

Case 1: The far-end speech is generated by an autoregressive model (AR) and convolved with the true car impulse response. The double talk period is simulated by adding another AR signal as local speech. Simulation result is shown in Fig. 7.

Case 2: Simulation results with real speech for both far-end and local signals are shown in Fig. 8.

In both cases, we can see that the correlation value calculated using the open-loop correlation factor (Eq.(10)) is quite “noisy” and it is difficult to determine a suitable threshold to detect double talk. However, the modified open-loop correlation factor (Eq.(12)) provides improved detection.

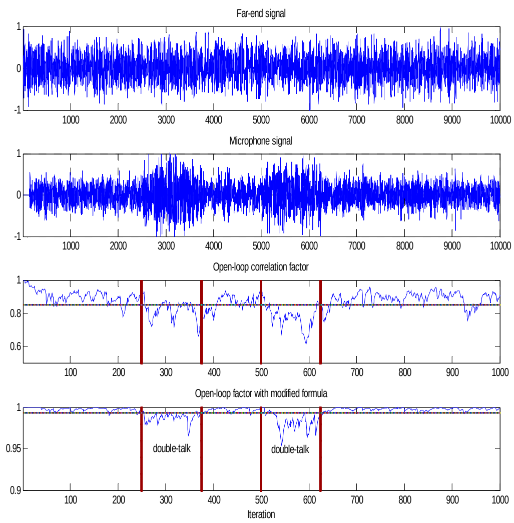

Fig. 7. The open-loop and the modified open loop correlation factors calculated with signals generated by AR models.

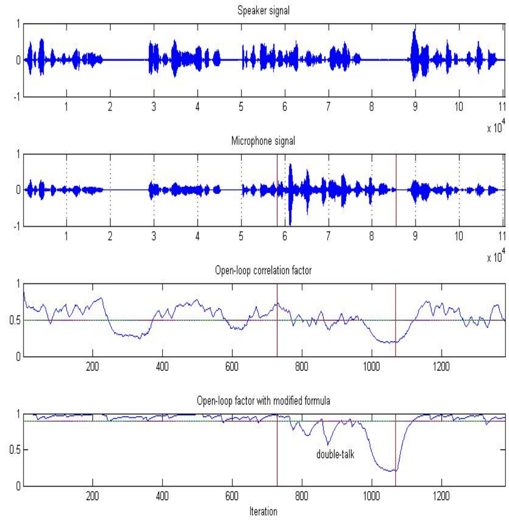

Fig. 8. The open-loop and the modified open loop correlation factors calculated with real speech signals.

V. CONCLUSIONS

In this paper, a basic AEC algorithm has been described in detail. Draft simulation results showed that its performance meets the standard requirement for convergence time and the ERLE. A modified open-loop correlation method to detect double talk is shown to be simple and efficient than the conventional open-loop method. Future work will be focused on the precise control of AEC such as step size and regularization parameters in the presence of an additional local disturbance.

ACKNOWLEDGEMENT

This work was supported by the Korea Research Foundation Grant funded by Korean Government (MOEHRD) (KRF-2005-041-D00603) and Basic Research Promotion Fund (KRF-2006-211-D00114).

REFERENCES

[1] Simon Haykin, “Adaptive filter theory,” 4th Edition
[2] Eberhard Hansler, Gerhard Schmidt, “Acoustic Echo and Noise Control-A Practical Approach”
[3] Adreas Mader, Henning Puder, Gerhard Uwe Schmidt, “Step-size control for acoustic echo cancellation filters – an overview”
[4] Ting Liu, Saeed Gazor, “De-correlated algorithms for faster adaptation”, *IEEE journal* 2004
[5] Microchip, “Acoustic Echo Cancellation Library,”-DS7013A.
[6] Andre Adrian, DFS Deutsche Flugsicherung – AEC open source code.