Abstract

A new post-processing method for acoustic echo canceller is proposed to reduce the residual echo and ambient noise. The method is based on the correlation of the desired signal and the estimation error signal of adaptive filter. The residual echoes are attenuated as proportional to the correlation of the desired signal and estimation error signal plays a role as Wiener filter for residual echo. Through computer simulations, it is shown that the performance of noise reduction and echo cancellation is dramatically improved.

1. Introduction

Multi-channel sound teleconferencing systems provide a real existence that could not be offered by actual mono-channel systems. Hence one of the promising applications in modern communications is desktop conferencing, which can involve several participants over a widely distributed area. This kind of conferencing with stereo or multi-channel sound will likely grow rapidly in the near future, especially over the Internet[1]. Multi-channel acoustic echo cancellers(AEC) become one of the keys in the successful realizations of conference systems, such as teleconferencing and desktop conferencing. However, the performance of multi-channel AEC is worse than single-channel AEC system. The residual echoes are always remained because the coefficients are not exactly matched echo paths due to ambient noise and its slow convergence speed as well as inherent problem.

To reduce the residual echoes, a post-processing method, which is co-operated with the noise-robust adaptive algorithm, is proposed in this paper. This method is based on the correlation of the desired signal and the estimation error signal. The residual echoes are attenuated as proportional to the correlation normalized with the power of desired signals. The normalized correlation of the desired signal and estimation error signal plays a role as Wiener filter for residual echoes. As results of simulations, it is shown that the proposed stereophonic acoustic echo cancellation schemes are well performed.

2. Stereophonic Acoustic Echo Cancellation

Configuration of the stereo echo canceller is shown in Fig. 1. To avoid clutter, we show the echo paths corresponding to only one of the two channels. In reality, similar paths exist in the other channel as well. Let us denote the impulse responses from the speech source in the far-end room to the right and left microphones as $g_1$ and $g_2$, respectively. The impulse responses of the echo-paths from the left and right speakers to the left microphone in the near-end room are assumed to be $h_1$ and $h_2$, respectively. Let $d_n$ be the echo received by the left microphone.

The acoustic echo canceller is modeled using FIR filters, generates an estimate $\hat{d}_n$ for the echo, which is subtracted from the true echo to form the error signal $e_n$. The left-channel echo canceller weight vectors $w_1$, $w_2$ are adapted with the objective of minimizing the mean squared left-channel residual echo, $E(e_n^2)$, in the absence of near-end speech. The dimensions of the adaptive filter length are the estimated lengths of the left and right echo paths, respectively.

A serious problem encountered in multi-channel AEC is that coefficients of the echo canceller do not converge to the true impulse response of the echo path. Due to the
cross-correlation between the left and right channel
signals, the weight estimate that minimizes the error
between echo $\hat{n}$ and estimated echo $\hat{n}$ is not unique.

To match the echo path is to de-correlate partially or
totally input signals. Some de-correlation methods are
proposed to improve the performance of multi-channel
AEC[3-12].

3. Proposed Acoustic Echo Cancellation
   Scheme with Post-processing

The performance of multi-channel AEC is worse than
single-channel AEC system. In multi-channel AEC case,
the residual echoes are always remained because the
coefficients are not exactly matched echo paths due to
ambient noise and its slow convergence speed as well as
the inherent problem.

The proposed multi-channel AEC scheme is depicted in
Fig. 2. We use the half-wave rectifier method for de-
correlating each channel signals because it is the
simplest method.

![Fig. 2. Block diagram of simulated stereo acoustic echo
canceller with post-processing](image)

The post-processing is based on the correlation of the
desired signal and the estimation error signal. The
residual echoes are attenuated as proportional to the
correlation normalized with the power of desired signals.
The output is yielded as follows

$$out_e = \rho_s e_n = \left( E[e_n d_n] / E[d_n^2] \right) e_n$$  \hspace{1cm} (1)

$\rho_s$ and $e_n$ is the output signal and time-varying gain
respectively. $s_n$ is near-end speech signal or ambient
noises. $\rho_s$ can be described as follows.

$$\rho_s = E[e_n d_n] / E[d_n^2] = E[e_n s_n] / E[y_n^2] + E[s_n^2]$$  \hspace{1cm} (2)

It can be assumed that ambient noises are uncorrelated
with echo signal and estimate of echo signal. The time-
varying gain can be reduced as (3).

$$\rho_s = \frac{E[y_n^2 - d_n y_n + s_n^2]}{E[y_n^2] + E[s_n^2]}$$  \hspace{1cm} (3)

As weights of adaptive filter are converged to echo
path, the gain $\rho_s$ becomes

$$\rho_s = \frac{E[s_n^2]}{E[y_n^2] + E[s_n^2]}$$  \hspace{1cm} (4)

From (4), we can see the normalized correlation of the
desired signal and estimation error signal plays a role as
Wiener filter for residual echoes. In the double-talk
situation, the estimation error signals that are residual
echoes dominantly include the near-end speech signal
and the normalized correlation close to unity. Therefore,
the residual echoes are hardly attenuated by the post-
processor and the near-end speech signal can be
transmitted without being attenuated. When the desired
signals consist of only acoustic echoes, the residual
echoes are mostly attenuated and canceled by the post-
processor. But this post-processing method impairs the
low power region because of its power estimation error.

![Fig. 3. Block Diagram of Hybrid Post-Processing
System](image)

To improve this problem, post-processor transfers sum
of post-processing signal and error signal with certain
proportion to far-end speaker. The algorithm is shown
in equation (5).

$$out_e = 0.2 + \frac{E[e_n d_n]}{E[d_n^2]} e_n$$  \hspace{1cm} (5)

Fig. 3 shows the block diagram of the Hybrid post-
processing system. This hybrid post-processing method
does not impair the near-end signal, so natural speech is transferred.

4. Simulation and Discussions

A stereophonic AEC’s are simulated in MATLAB. The impulse responses of receiving room and transmitting room are modeled using a 1024-th order FIR filter measured from two actual rooms in Fig. 4.

Computer simulations were performed using speech and music signals. First simulations used speech signal. A real speech signal, shown in Fig. 5, is used as the far-end and near-end inputs. The far-end speech signal is a 16 bit quantized, 8 KHz sampled, 50,000 sample-data, and female voiced Korean Sentence. And the near-end speech signal is a 16 bit quantized, 8 KHz sampled, 50,000 sample-data, and male voiced Korean word.

We simulated under two NER conditions. At first, when NER is 10 dB, the simulation results are Fig. 6 and Fig. 7. Fig. 6 shows ERLE of various algorithms. In Fig. 6, the solid line presents ERLEs before post-processing, and the dashed line represents ERLEs after post-processing. Over the first 40,000 samples, the ERLE of AP algorithm performs better than other algorithms, because of the fast convergence speed. But after 40,000 samples, the double-talk period starts for 5,000 samples.

In the double-talk region, the weight mismatches of AP and NLMS algorithms are abruptly increased. So the adaptive filter coefficients of AP and NLMS algorithms do not track the real echo paths.

But the mismatch of modified AP algorithm steadily decreased, it means that the adaptation of modified AP algorithm almost stopped, and the filter coefficients remained unchanged or small perturbation because of near-end signal, which made the step-size reduced. After double-talk region, ERLE of the modified AP algorithm is superior to AP, and NLMS algorithm over 10 dB, it means that the modified AP algorithm reduces the acoustic echoes well, but other algorithms do not reduce echoes for a while.
algorithm. Over the first 40,000 samples, the AP algorithm reduces echoes faster than other algorithms, but after the double-talk period, it amplifies echoes because of its wrong adaptation. The NLMS algorithm shows same results. But the modified AP algorithm still reduces echoes. After post-processing, the residual echoes are almost disappeared and only near-end signal remains.

5. Conclusions
The residual echoes are always remained because the coefficients are not exactly matched to echo path due to ambient noise. To reduce the residual echoes, a new post-processing method, which is co-operated with the proposed noise-robust adaptive algorithm, is proposed in this paper.

References