

## A Robust Acoustic Echo Canceller for Noisy Environment<sup>1</sup>

Shenghao Qin, Sha Meng, and Jia Liu

Department of Electronic Engineering, Tsinghua University, Beijing 100084  
{qinsh99, mings04}@mails.tsinghua.edu.cn, liuj@tsinghua.edu.cn

**Abstract.** A new acoustic echo cancellation method for noisy environment is proposed and a real-time system is built. Different from the conventional acoustic echo cancellers, the proposed has a novel structure and can remove the echo effectively at low SNR. In our system, the speech with stationary noise is processed in order to separate the echo from the noise before updating the adaptive filter which is used for echo cancellation. Robust speech detector is another crucial component in the system. A module called noise generator is used for the comfort and continuity of the speech in real applications. With efficient organization and combination, the system gets excellent performance and is evaluated by statistical experiments

### 1 Introduction

Acoustic echo canceller (AEC) has been a valuable topic for a long time. With the development of communication technology, acoustic echo cancellation became much more urgent. As one of the important parts of AEC, adaptive filter algorithm has been developed for decades. It is used to design a FIR filter to estimate the acoustic echo channel. Then the AEC system removes the estimation of the echo according to the FIR filter and input signal of the echo channel from the speech collected by microphone.

In the last few years, the adaptive filter in frequency domain [1][9][10] became a technical focus. Comparing with the algorithms in time domain, Partial Rank Algorithm (PRA) [6] for instance, filters in frequency domain reduce computational complexity using FFT, IFFT and multiplication in frequency domain called Overlap-Save method [9] instead of time-domain convolution and gradient correlation in PRA.

A basic frequency-domain adaptive filter algorithm called Frequency-Domain Adaptive Filter (FDAF) [9] was introduced by John J. Shynk in 1992. But the FDAF is not suitable for real-time echo cancellation systems. It appears that the FDAF is computationally attractive only if the block length has the same order of magnitude as the filter length. In practice however, this leads to unacceptable input/output delays. The filter length is always very large in order to cover long acoustic echo delay. The delay is twice as long as the acoustic impulse response, i.e., 128ms when sampled at 8

---

<sup>1</sup> Supported by the National Natural Science Foundation of China (NSFC) (60572083).

kHz. Another frequency-domain algorithm called the Partitioned Block Frequency-Domain Adaptive Filter (PBFDAF) [1] introduced by Koen Eneman, Marc Moonen in 2003 is applied in our system by splitting the adaptive filter coefficients into several parts. The  $N$ -tap adaptive filter is partitioned into  $N/P$  segments. Each of segments of length  $P$  is transformed to frequency domain. The *Overlap-Save* method in PBFDAF is the same as in FDAF, but the block size can be shorter than the length of filter which means shorter delay.

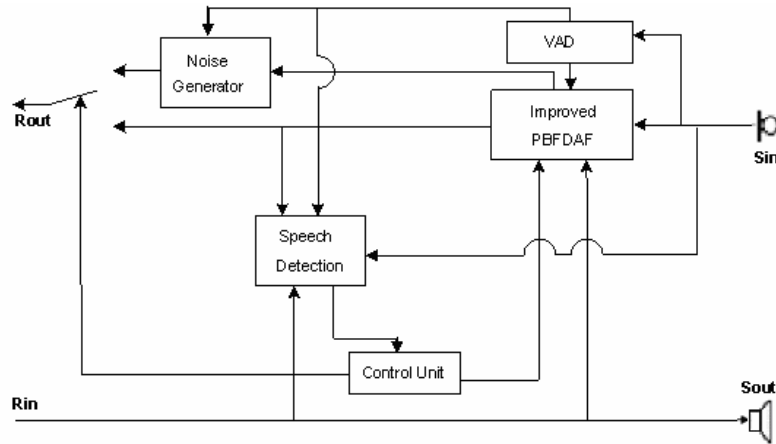
In recent years, wireless, hands-free, Internet protocol (IP) phones and vehicle communication systems become popular in which noise and echo exist simultaneously. The conventional adaptive filter has already got good performance in quiet environment. But in the noisy environment, the filter always can not estimate the echo channel accurately.

There are other problems for echo cancellation under noisy environment. The noise makes the performance of another component of AEC, the speech detector which is used to detect which side of conversation is speaking degrade. Meanwhile, the discomfort and uncontinuity of speech processed by conventional AEC systems become more visible because of the noisy background. In order to design a robust AEC for noisy environment, it is important to solve these problems as well as the robust adaptive filter.

In this paper, an integrated speech signal processing system is proposed based on the speech signal from one microphone. The base of this system is the improved frequency-domain adaptive filter. Additional, a noise generator (NG) module, a voice activity detection module and a robust speech detection module are necessary for different purposes.

The rest of the paper is organized as follows. In Section 2, an overview of the proposed system is illustrated and introduced. In Section 3, a novel adaptive filter algorithm is proposed. Section 4 explains speech detector and noise generator which are another two crucial components of the proposed system. The experiment result is presented in Section 5 and we will summarize our work in the last section.

## 2 System Overview



**Fig. 1.** Block scheme of the acoustic echo canceller

Fig.1. illustrates the block scheme of the proposed. *Sin* is the near-end input speech, collected through a microphone and *Sout* is from far-end speaker and played by loudspeaker. *Sin* contains not only the near-end speech, but also the echo and the environment noise such like car-noise in vehicle phone system.

In this system, the signal *Sin* is firstly been processed by the improved adaptive filter. Inside the adaptive filter proposed, echo estimation could make the adaptive filter have good performance in nosy environment. The noise generator is used to send a stationary noise which is similar to the background for comfort and continuity of the speech. Speech detection [2][3][11] is another important technique which decides whether the adaptive filter should be used or the coefficients should be updated at the moment. The result of the detector is sent to the control unit which controls the behavior of both the adaptive filter and the noise generator. These three main parts will be explained in more details in the following sections.

Voice Activity Detection (VAD) is used in the system for 3 purposes. Firstly, the result of VAD is used to estimate the noise spectrum for echo estimation in the adaptive filter. Secondly, it is used for the estimation of the noise energy in speech detector. It also helps the noise generator choose correct moment to do the LPC analysis of the noise. In our system, the sliding filter endpoint detection algorithm [4] is used because of its excellent performance at various noise levels. To embed the algorithm into real-time system, some modification is needed.

### 3 Improved PBFDAF

Basic variables are assumed as following:

$x$  Far-end speech, sent to the near-end speaker.

$d$  Near-end speech, obtained from microphone.

$w$  The coefficient of the FIR filter.

$N$  The length of the FIR filter.

Lower letters are used for time-domain samples, lower bold face letters for time-domain column vectors, upper letters for matrices, and upper bold face letters for frequency-domain column vectors.

Traditional adaptive filters assume that the signal collected from the microphone only includes the near-end speech and the echo of the far-end speech. Therefore they can not converge and follow the change of echo channel under noisy environment. Signal collected from the microphone under noisy environment does not equal to the convolution of far-end speech and the echo channel because of the noise, which prevents the adaptive algorithm from estimating the channel accurately. The residual echo signal is always very large and may even diverge which is quite severe when the system is running in real conditions.

For dealing with this problem, the real echo should be extracted from the noise firstly and then used to update the adaptive filter. We use the estimation of the priori SNR to estimate the echo. The priori SNR is estimated using the prior SNR of last frame and the posteriori SNR of current frame with different weights as:

$$\xi_k(n) = \alpha \frac{\hat{A}_k^2(n-1)}{\hat{\lambda}_k(n-1)} + (1-\alpha) \left( \frac{D_k^2(n)}{\hat{\lambda}_k(n)} - 1 \right) \quad (1)$$

where  $\alpha$  is the weight of the prior SNR,  $\hat{A}_k^2(n-1)$  is the echo estimation in frequency domain of the (n-1)th frame and  $\hat{\lambda}_k(n-1)$  is the spectrum of noise of the (n-1)th frame.

We can get the estimation of echo in current frame through a filter designed according to the estimation of priori SNR as equation 2. The time-domain echo estimation is calculated through IFFT as shown in equation 3.

$$\hat{A}_k(n) = \frac{\xi_k(n)}{1 + \xi_k(n)} D_k(n) \quad (2)$$

$$\hat{\mathbf{a}}(n) = F^{-1}[\hat{\mathbf{A}}(n)] \quad (3)$$

The estimation of noise is updated when the state of VAD is *Silence*. The detail of the algorithm is explained in Table 1. The improved PBFDAF does not only cancel echo, but also remove the background noise. The performance will be evaluated in Section 5.

**Table 1.** The novel BPFDAF

<p><i>For each frame of <math>L</math> samples</i></p> <p><math>\mathbf{X}_r(n) = \text{diag}(F[x[(n+1)L - rP - M + 1], \dots, x[(n+1)L - rP]]^T)</math></p> <p><math>\mathbf{y}(n) = \text{last } L \text{ components of } F^{-1}[\sum_{r=0}^{(N/P)-1} \mathbf{X}_r(n)\mathbf{W}_r(n)]</math></p> <p><math>\mathbf{d}(n) = [d(nL), d(nL+1), \dots, d(nL+L-1)]^T</math></p> <p><math>\mathbf{D}(n) = F[\mathbf{d}(n)]</math></p> <p><math>\xi_k(n) = \alpha \frac{\hat{A}_k^2(n-1)}{\hat{\lambda}_k(n-1)} + (1-\alpha)(\frac{D_k^2(n)}{\hat{\lambda}_k(n)} - 1)</math></p> <p><math>\hat{A}_k(n) = \frac{\xi_k(n)}{1 + \xi_k(n)} D_k(n)</math></p> <p><math>\hat{\mathbf{a}}(n) = F^{-1}[\hat{\mathbf{A}}(n)]</math></p> <p><math>\mathbf{e}(n) = \hat{\mathbf{a}}(n) - \mathbf{y}(n)</math></p> <p><math>\mathbf{E}(n) = F[0, \dots, 0, \mathbf{e}^T(n)]^T</math></p> <p><math>\mathbf{W}_r(n+1) = \mathbf{W}_r(n) + 2FGF^{-1}\Pi_M \mathbf{X}_r^H(n)\mathbf{E}(n)</math></p> <p style="margin-left: 40px;"><math>\alpha</math> weight factor</p> <p style="margin-left: 40px;"><math>\hat{\lambda}_k(n)</math> the estimation of noise spectrum</p> <p style="margin-left: 40px;"><math>G = \begin{bmatrix} I_P &amp; 0 \\ 0 &amp; 0_{M-P} \end{bmatrix}</math>      <math>\Pi_M = \mu(\sum_{r=0}^{\frac{N}{P}-1} \mathbf{X}_r(n)\mathbf{X}_r^*(n))^{-1}</math></p>	
---	--

#### 4 Speech Detector and Noise Generator

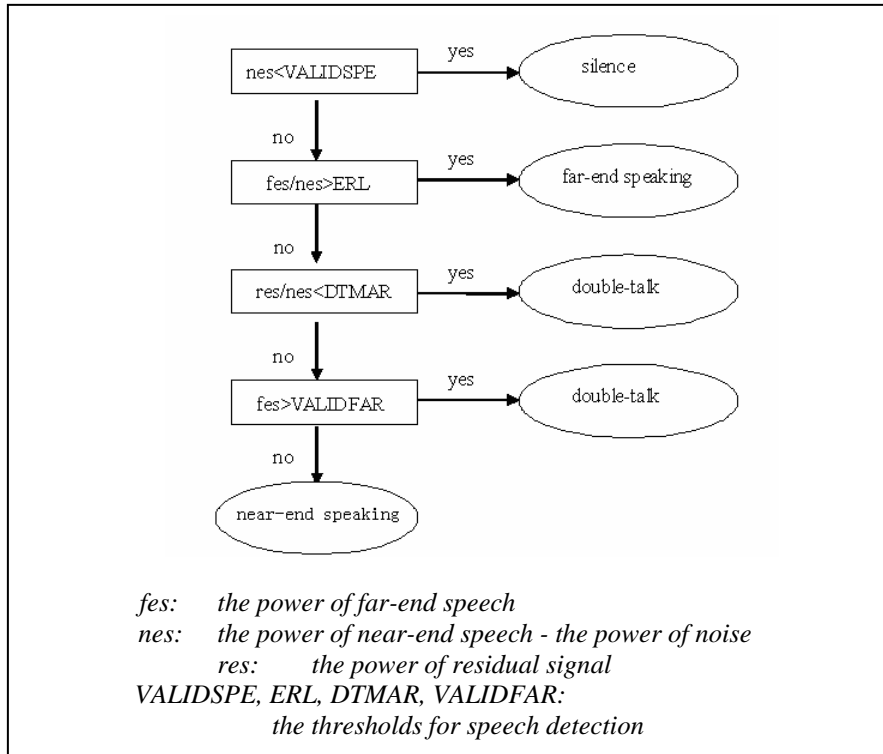
Speech detector is crucial in an acoustic echo cancellation system. Using the power of far-end speech  $x(n)$ , near-end speech  $d(n)$  and residual signal  $e(n)$  after adaptive filtering, it decides which side of the conversation is speaking. The speech detectors have 4 states: silence, far-end speaking, near-end speaking and double-talk. The coefficients of the adaptive filter are updated during far-end speaking and the filter is used during double-talk to cancel the echo.

The speech detection algorithm used in this system is base on the energy. Different with other energy-based algorithms, it has more branches and robust performance in noisy environment. The flow is shown in Table 2.

The noise energy is removed from the near-end speech energy firstly in order to keep the accuracy of detector at various SNR. The experiment result using this detector at different SNRs will be shown in Section 5

When the output of the speech detector is far-end speaking which means the signal received from microphone includes only the echo and noise, the canceller does not need to send the residual signal. It could send silence under quiet environment. But in the noisy environment, for the comfort and continuity of the speech, far-end speaker may prefer the noise similar to the background rather than silence. Noise generator is used for this purpose. According to the principle of linear prediction, a full-pole filter is designed to generate the target noise [12-14]. In noisy segment of the signal, we calculate the autocorrelation coefficients of the target noise. The Levinson-Durbin algorithm is then used to get the coefficients of the full-pole filter. After inputting Gaussian White with certain gain which is still computed from the LPC analysis, we get the simulative noise from the output.

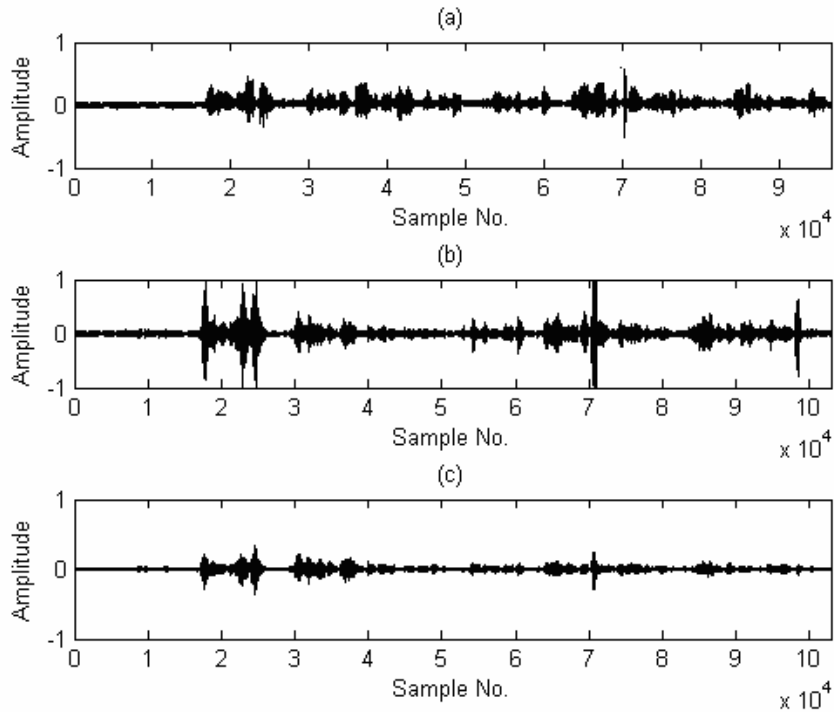
**Table 2.** The Flow of Speech Detection



## 5 Performance Evaluation

The performance of the proposed algorithm and system has been tested by a group of experiments.

As the most crucial part of AEC, we firstly evaluated the performance of the proposed improved PBFDAF. Fig 2 shows the comparison result with the conventional ones. Fig. 2(a) shows the signal collected by one microphone in traffic environment. The noise of the car and the echo are both inputted into the system. It leads to the poor performance of conventional PBFDAF as shown in Fig. 2(b). Fig. 2(c) shows the result of echo cancellation applying improved PBFDAF. The result shows that the proposed algorithm is much more effective under noise environment.



**Fig. 2.** Comparison of conventional PBFDAF and the proposed improved PBFDAF under noisy environment. (a) The signal ( $d(n)$ ) in traffic environment, includes the echo and cars' noise (SNR=7dB). (b) The result using the conventional PBFDAF. (c) The result using the novel adaptive filter algorithm.

In Table 3, we calculate the accuracy of the speech detection algorithm mentioned in Section 4 at different SNRs and we find that it has robust performance.

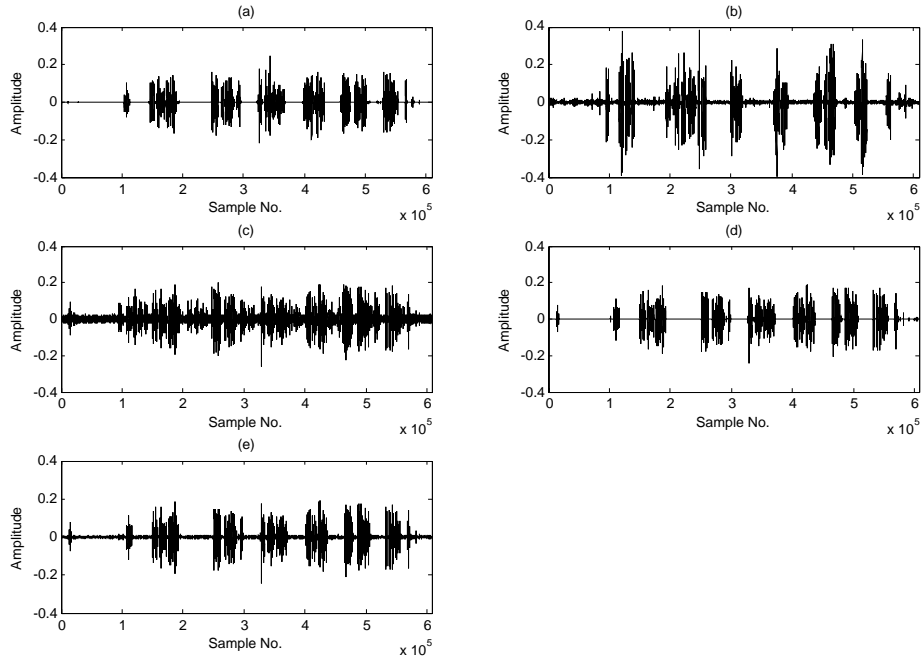
**Table 3.** The statistic experiment result of the speech detector proposed at various SNR

SNR=5dB			
The result of detection	Far-end	Near-end	double-talk
The answer			
Far-end (100)	99.14	0.84	0.02
Near-end (100)	2.06	96.09	1.84
Double-talk (100)	3.61	17.80	78.59
SNR=7dB			
The result of detection	Far-end	Near-end	double-talk
The answer			
Far-end (100)	97.54	2.38	0.07
Near-end (100)	1.79	96.32	1.89
Double-talk (100)	12.06	7.14	80.79
SNR=10dB			
The result of detection	Far-end	Near-end	double-talk
The answer			
Far-end (100)	99.20	0.70	0.10
Near-end (100)	2.10	96.02	1.88
Double-talk (100)	5.26	10.65	84.09

To test the performance of the system, 520 sentences of speech from 863 speech library are used and the noise is the car noise. All of the speech signal is played and recorded using the TI DSP platform. The SNR and ERLE (Echo Return Loss Enhancement) at various SNRs is shown in Table 4. The experiment result shows that the proposed system has robust performance in noisy environment. An example for the proposed system is shown in Fig.3. We may find that the signal shown in Fig 3(d) which is the result of the echo cancellation becomes a little distorted from the signal shown in Fig 3(a) which is the clean speech of the near-end. However it can be ignored for listening.

**Table 4.** The statistic experiment result of the proposed system

SNR of Sin (dB)	The ERLE of PBFDAF (dB)	SNR after processed by proposed system (dB)	ERLE of proposed algorithm (dB)
3.60	-13.90	6.14	8.47
5.82	5.08	7.57	10.02
8.81	10.17	9.67	12.83



**Fig. 3.** (a) near-end speech (b) The signal  $S_{out}/R_{in}$  , far-end speech (c) The signal  $S_{in}$ , including near-end speech, echo and background noise (SNR=6) (d) The result of echo cancellation without NG (e) The result of system proposed with NG which will be sent to far-end as  $R_{out}$  (SNR=10dB).

## 6 Conclusion

In this paper, a robust system for echo cancellation under noisy environment is proposed. Improved PBFDAF is proposed for estimating and canceling the echo and noise. The module called noise generator makes the result of the cancellation sound more comfortable. By combining the improved adaptive filter and the improved robust speech detector, together with the effective VAD, the system is effective in various noisy environments. The experiments show that system proposed could remove most of echo and noise.

## Reference

- 1 Koen Eneman, Marc Moonen "Iterated Partitioned Block Frequency-Domain Adaptive Filtering for Acoustic Echo Cancellation" IEEE Transactions on Speech and Audio Processing, Volume: 11 , Issue: 2 , March 2003

- 2 Si Ho Kim, Hong Seok Kwon, Keun Sung Boe, Kyung Jin Byun, Kyung Soo Kim  
“Performance Improvement of Double-talk Detection Algorithm in the Acoustic Echo Cancellation” Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on , Volume: 5 , 7-11 May 2001
- 3 Filippo S. Hallack, Mariane R. Petraglia “Performance Comparison of Adaptive Algorithms Applied to Acoustic Echo Cancelling” Industrial Electronics, 2003. ISIE '03. 2003 IEEE International Symposium on , Volume: 2 , 9-11 June 2003
- 4 Qi Li, Jinsong Zheng, Augustine Tsai, Qiru Zhou “Robust Endpoint Detection and Energy Normalization for Real-Time Speech and Speaker Recognition” **IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 10, NO. 3, MARCH 2002**
- 5 F. Beritelli, S. Casale, G. Ruggeri, and S. Serrano “Performance Evaluation and Comparison of G.729/AMR/Fuzzy Voice Activity Detectors” IEEE SIGNAL PROCESSING LETTERS, VOL. 9, NO. 3, MARCH 2002 85
- 6 Koen Eneman , Simon Doclo , Marc Moonen. “A Comparison between Iterative block\_LMS and Partial Rank Algorithm” March 1999 Internal Report ESAT-SISTA
- 7 Dennis R. Morgan, Steven G. Kratzer “On a Class of Computationally Efficient, Rapidly Converging, Generalized NLMS Algorithms” IEEE Signal Processing Letters Vol.3 NO.8 August 1996
- 8 Justin Song, Jian Li, Yen-Kuang Chen “Quality-delay-and-computation Trade-off Analysis of Acoustic Echo Cancellation on General-purpose CPU” Microprocesser Research Laboratories, Intel Corporation
- 9 John J. Shynk “Frequency-Domain and Multirate Adaptive Filtering” Signal Processing Magazine, IEEE , Volume: 9 , Issue: 1 , Jan. 1992
- 10 Schobben D.W.E., Egelmeers G.P.M., Sommen P.C.W “Efficient Realization of the Block Frequency Domain Adaptive Filter” Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on , Volume: 3 , 21-24 April 1997
- 11 Hua Ye, Bo-Xiu Wu “A New Double-Talk Detection Algorithm Based on the Orthogonality Theorem” IEEE Transactions on communications Vol 39, NO.11, November 1991
- 12 G.729 “coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)” telecommunication standardization sector of ITU
- 13 G. 729 Annex B “A silence compression scheme for G.729 optimized for terminals conforming to Recommendation V.70” telecommunication standardization sector of ITU
- 14 G. 723 Annex A “Silence compression scheme” telecommunication standardization sector of ITU
- 15 G.165 “Echo cancellers” telecommunication standardization sector of ITU
- 16 G.167 “Acoustic echo controllers” telecommunication standardization sector of ITU