IOP Conf. Series: Materials Science and Engineering 466 (2018) 012040 doi:10.1088/1757-899X/466/1/012040

An Improved Double-talk Detection Algorithm for Echo Cancellation in Teleconference System

Ruxue Guo and Li Zhao

School of Information Science and Engineering, Southeast University, No.2 Sipailou, Nanjing, 210096, China
Email: grx0904@sina.com

Abstract. Echo cancellation commonly employs a double-talk detector (DTD), which is essential to keep the adaptive filter from diverging in the presence of near-end speech. An improved double-talk detection algorithm is proposed to promote the performance of the echo cancellation in teleconference system in the presence of speech issued from the near-end speaker (double-talk). Based on the partitioned block frequency domain adaptive filter algorithm (PBFDAF), the advanced double-talk detection method is presented in this paper constructs a joint correlation coefficient variable to judge the communication state and the variable is applied as a smoothing parameter to the error signal and then get the final output signal. Simulation results show that the improved algorithm can effectively detect the current voice state, eliminate the echo signal, and keep the near-end speaker's voice well. The performance of the echo cancellation algorithm is improved when both echo and near end signals are simultaneously active by the proposed technique.

1. Introduction

With the rapid development of communication technology, people demand higher speech quality in speech communication system, such as teleconference system, etc. [1]. However, it is not easy to acquire satisfactory speech quality due to the presence of inevitable echo. Therefore, in order to improve communication quality, an echo canceller must be integrated in the communication device to eliminate or suppress the echo [2].

Nowadays, the adaptive filtering technique is the most effective way to remove echo [3]. The block diagram of the implementation of echo cancellation system is shown in figure.1. The far-end signal $x(n)$ passes through the actual echo path $\omega(n)$ to produce the echo $y(n)$. The desired input signal from the near end speaker is called the near-end signal $v(n)$. The two signals combine to form the microphone input signal $d(n)$ and the aim of an echo canceller is to remove the echo involved in $d(n)$. The adaptive filter approximately models the echo path $\hat{\omega}(n)$ and this approximated output of the adaptive filter $\hat{y}(n)$ is subtracted from the $d(n)$. This resulting error signal $e(n)$ contains the near-end signal with residual echo which is used to update the filter coefficients.

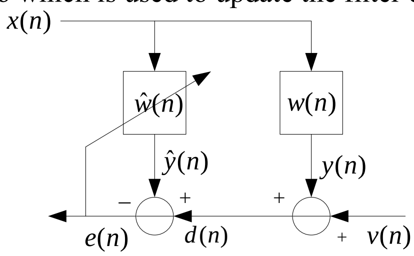

Figure 1. The block diagram of echo cancellation system.

When the near-end talker is silent, i.e. $v(n) = 0$, the existing adaptive algorithms can successfully cancel the echo and the adaptive filter can converge to a good estimate of the acoustic echo path [4-6]. However, when signals from both ends of an echo cancellation system are simultaneously active, near-end signal $v(n)$ which acts as an uncorrelated noise to the adaptive algorithm can cause the adaptive filter to diverge. Therefore, detecting the presence of near-end signal effectively plays a vital role in the practical application of the echo cancellation and this is commonly referred to as Double-Talk Detection (DTD) [7-8].

In the current the echo cancellation system for teleconference, the double-talk detection and the adaptive echo cancellation technique must collaborate effectively to eliminate acoustic echo. An improved double-talk detection algorithm for echo cancellation in teleconference system is proposed in this paper, using the partitioned block frequency domain adaptive filter algorithm for basic echo cancellation. A joint correlation coefficient variable is constructed to judge the communication state. The advanced algorithm can effectively detect the current voice state and perform adaptive processing, which improves the performance of the echo cancellation algorithm in double talk situation.

2. Partitioned block frequency domain adaptive filter algorithm

It is necessary to suppress echo, especially the long-time delay echo in some telecommunication systems to obtain a good performance of acoustic echo cancellation. However, the classical LMS-based adaptive filter algorithms will bring a large and complex computation which is linear in the filter length and the convergence is slow [9-10]. In order to handle such a problem, the partitioned block frequency domain adaptive filter (PBFDAF) algorithm is applied to cancel long-time delay echo. The input speech signals are transformed to frequency domain, and the FFT, as well as the block processing technique to process those signals. So, the PBFDAF algorithm will obtain considerable computational amount and better convergence [11].

The basic idea of PBFDAF algorithm is dividing the length of N-order filter into N/P sections in time domain before filtering. Each segment contains P coefficients:

$$w_p(n) = \begin{bmatrix} w[pP] \\ \dots \\ w[(p+1)P-1] \end{bmatrix}, p = 0,1,\dots,\frac{N}{P}-1 \tag{2.1}$$

Where the p represents the sequence number of every section. After zero-padding, the $\omega_p(n)$ is transformed to frequency domain:

$$W_p(n) = F\begin{bmatrix} w_p(n) \\ 0_{M-P} \end{bmatrix} \tag{2.2}$$

Where the F represents the M-order Fourier transformation matrix, and n represents the serial number of iteration. An iterative process of PBFDAF includes the following steps:

1) Far end reference signal x is divided into sections and the partitioned signal is transformed to frequency domain by FFT

$$X_p(n) = diag\left\{F\begin{bmatrix} x[(n+1)L-pP-M] \\ \dots \\ x[(n+1)L-pP-1] \end{bmatrix}\right\}, p = 0,1,\dots,\frac{N}{P}-1 \tag{2.3}$$

2) After performing filtering in frequency domain, an accumulative signal can be obtains by accumulating all the segment signals. Then then inverse Fourier transformation is carried on to get the estimated signal $y(n)$. Only L points in the latter half of the signal are the effective results of linear convolution.

$$y(n) = \begin{pmatrix} 0_{M-L} & 0 \\ 0 & I_L \end{pmatrix} F^{-1} \sum_{p=0}^{N/P-1} X_p(n) W_p(n) \tag{2.4}$$

3) The estimated echo signal $y(n)$ is subtracted from the microphone input signal $d(n)$ and error signal $e(n)$ is obtained:

$$e(n) = d(n) - y(n) \tag{2.5}$$

4) Calculating the step factor in frequency domain and Updating the adaptive filter coefficients in accordance with the error signal $e(n)$. μ is a positive constant called the step-size parameter, which influences the stability performance of the algorithm and γ is a small number which is used to avoid data overflow.

$$\prod(n) = diag[\mu(n)] = \mu\left(\sum_{p=0}^{N/P-1} X_p^T(n) X_p(n) + \gamma I_M\right)^{-1} \tag{2.6}$$

$$W_p(n+1) = W_p(n) + FGF^{-1}\prod(n) X_p^*(n) Fe(n),\ p = 0,1,\ldots,\frac{N}{P}-1 \tag{2.7}$$

3. Double-talk detection algorithm based on joint correlation coefficient

Double-talk is the situation where signals from both ends of an echo cancellation system are simultaneously active. The algorithmic difficulty of dealing with such problem is the presence of signal from the near-end speaker. A common solution to this problem is to use a double-talk detector to detect the voice state, and stop the adaptive update of the filter coefficients in the presence of the near-end user voice [12].

A method based on correlation coefficient between certain signals to distinguish single-end and double-talk vocalization is presented in this part. The output signal of PBFDAF is further processed by the advanced double-talk detection algorithm. A joint correlation coefficient variable is constructed by the cross-correlation value between far-end block signal, near-end block signal and the error block signal in frequency domain. The variable represents the current communication state and is applied as a parameter in smoothing the error signal to get final output signal. The algorithm is implemented as follows:

1) First, calculate and smooth the power spectral density of the near-end block signal $d(k)$, the far-end block signal $x(k)$ and the error block signal $e(k)$, respectively.

$$\begin{cases} S_d(k) = \lambda S_d(k) + (1-\lambda) d(k)\cdot d(k)^* \\ S_x(k) = \lambda S_x(k) + (1-\lambda) x(k)\cdot x(k)^* \\ S_e(k) = \lambda S_e(k) + (1-\lambda) e(k)\cdot e(k)^* \end{cases} \tag{3.1}$$

Where $S_d(k)$, $S_x(k)$ and $S_e(k)$ represents the power spectral density of those signals respectively, λ is the smoothing factor.

2) Calculate the cross-power spectral density $S_{xd}(k)$ of the far-end signal $x(k)$ and the near-end signal $d(k)$, as well as the cross-power spectral density $S_{ed}(k)$ of the error signal $e(k)$ and $d(k)$, respectively.

$$\begin{cases} S_{xd}(k) = \lambda S_{xd}(k) + (1-\lambda) S_x(k)\cdot S_d(k)^* \\ S_{ed}(k) = \lambda S_{ed}(k) + (1-\lambda) S_e(k)\cdot S_d(k)^* \end{cases} \tag{3.2}$$

3) Calculate the correlation coefficient $\mathrm{Coh}_{xd}(k)$ of the far-end signal $x(k)$ and the near-end signal $d(k)$, as well as the correlation coefficient $\mathrm{Coh}_{ed}(k)$ of the error signal $e(k)$ and $d(k)$, respectively.

$$\begin{cases} coh_{xd}(k) = \dfrac{S_{xd}(k)\cdot S_{xd}(k)^*}{S_x(k)\cdot S_d(k)} \\ coh_{ed}(k) = \dfrac{S_{ed}(k)\cdot S_{ed}(k)^*}{S_e(k)\cdot S_d(k)} \end{cases} \tag{3.3}$$

4) Take the smaller one of $1-\mathrm{Coh}_{xd}(k)$ and $\mathrm{Coh}_{ed}(k)$ to construct a joint correlation coefficient variable P_{coh} which ensures the maximum accuracy of detecting the current voice state by taking the two correlation variables into consideration.

$$P_{coh}(k) = \min(1-coh_{xd}(k), coh_{ed}(k)) \tag{3.4}$$

5) Construct a suppression function and function expression is given as follows

$$Supp(k) = P_{coh}(k)^{\beta(k)*curve(k)} \tag{3.5}$$

Where, $\beta(k)$ is a coefficient indicating the degree of suppression and $curve(k)$ is the pre-set suppression curve for each frequency point.

6) The error signal $e(k)$ is further processed by the suppression function $Supp(k)$ to obtain the final output signal:

$$ef(k) = e(k) * Supp(k) \tag{3.6}$$

By the introduction of algorithm implementation, it is found that the value of the suppression function is obtained by the joint correlation coefficient variable P_{coh}, which characterizes the degree of suppression of the error signal under different voice states to some extent. When both the far-end and the near-end speakers talk simultaneously, namely in double-talk situation, it needs to reduce the suppression depth and retain the near -end signals as much as possible, while the near-end speaker is silent, it need to increase the depth of the suppression to eliminate the possible residual echoes

4. Simulation results and analysis

The simulation experiment was conducted on the MATLAB in order to verify the performance of the proposed algorithm. The test speech used in experiment was selected from the mandarin level test reading works. The sampling rate of the test signal is 16 kHz and the speech duration is about 45 s. The time-domain waveform of the speech signals are shown in figure.2. Figure.2 (a) represents the signal from near end speaker $v(n)$, and the final output signal $ef(n)$ is shown in figure.2 (b). The red line in figure.2 (c) represents a detection variable indicating the existence of near-end speech signal and it is superimposed on the microphone input signal. It is observed that the improved algorithm can remove the echo in the near end microphone input signal and keep the near-end speech signal well. The results of subjective speech quality evaluation also shows that the output signal of the system in double-talk situation has a good audio effects, without any intermittent or words loss.

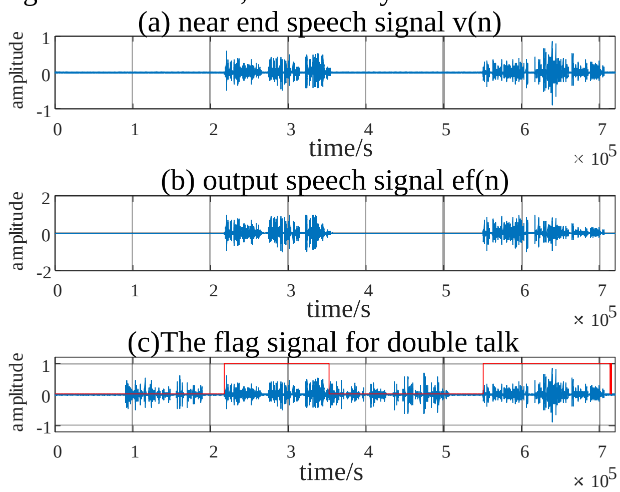

Figure 2. Time-domain waveforms of the speech signals. (a) near-end speech signal $v(n)$; (b) Final output signal $ef(n)$; (c) Detection variable indicating the existence of near-end signal.

To assess the performance of echo cancellation algorithms, the comparison of misalignment between the classical NLMS [4], FLMS [6] and the proposed algorithms. The misalignment comparison of three algorithms is shown in figure.3. It is observed that the improved algorithm proposed in this paper has a smaller misalignment and faster convergence rate compared with the NLMS and FLMS.

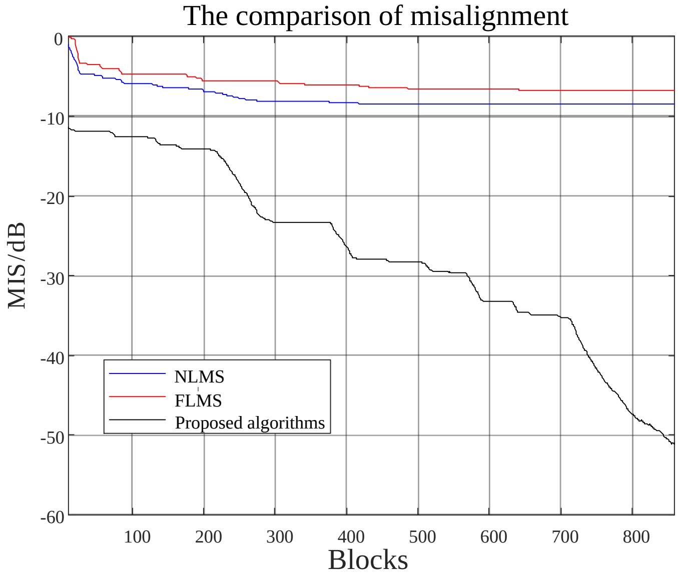

Figure 3. The misalignment comparison of three algorithms

5. Conclusion
This paper proposes a double-talk detection algorithm for echo cancellation in teleconferencing system. The new method combines the partitioned block frequency domain adaptive filter algorithm with the double-talk detection algorithm based on joint correlation coefficient variable. The current voice state can be detected effectively and perform adaptive processing according to the variable. The proposed algorithm is simulated in MATLAB, and the simulation results show that the improved algorithm can eliminate the echo effectively and voice quality is acceptable without any obvious intermittence, improving the performance of the echo cancellation algorithm in double-talk situation.

6. References
[1] Fukui M, Shimauchi S, Hioka Y, et al. Double-talk robust acoustic echo cancellation for CD-quality hands-free videoconferencing system[J]. IEEE Transactions on Consumer Electronics, 2014, 60(3):468-475.
[2] Wu C, Jiang K, Wang X, et al. A Robust Step-Size Control Technique Based on Proportionate Constraints on Filter Update for Acoustic Echo Cancellation [J]. Chinese Journal of Electronics, 2016, 25(4):692-699.
[3] Sondhi M M. An Adaptive Echo Canceller [J]. Bell Labs Technical Journal, 1967, 46(3):497-511.
[4] Tyagi R, Singh R, Tiwari R. The performance study of NLMS algorithm for acoustic echo cancellation[C]// International Conference on Information, Communication, Instrumentation and Control. 2017:1-5.
[5] Zheng Z, Liu Z, Zhao H, et al. Robust Set-Membership Normalized Subband Adaptive Filtering Algorithms and Their Application to Acoustic Echo Cancellation [J]. IEEE Transactions on Circuits & Systems I Regular Papers, 2017, PP (99):1-14.
[6] Huang Y, Skoglund J, Luebs A. Practically efficient nonlinear acoustic echo cancellers using cascaded block RLS and FLMS adaptive filters[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2017:596-600.
[7] Chien Y R, Jin L Y. Convex Combined Adaptive Filtering Algorithm for Acoustic Echo Cancellation in Hostile Environments [J]. IEEE Access, 2018, PP (99):1-1.
[8] Heo W, Kim T, Bae K. Robust double-talk detection in the acoustic echo canceller using normalized error signal power[C]// International Symposium on Signal Processing and ITS Applications. IEEE, 2008:1-4.

[9] Sristi P, Lu W S, Antoniou A. A new variable-step-size LMS algorithm and its application in subband adaptive filtering for echo cancellation[C]// IEEE International Symposium on Circuits and Systems. IEEE, 2001:721-724 vol. 2.
[10] Liu L G, Fukumoto M, Zhang S Y. A variable step-size proportionate NLMS adaptive filtering algorithm and its application in network echo cancellation[J]. Acta Electronica Sinica, 2010, 97(3):996-1001.
[11] Bernardi G, Waterschoot T V, Wouters J, et al. Adaptive Feedback Cancellation Using a Partitioned-Block Frequency-Domain Kalman Filter Approach with PEM-Based Signal Prewhitening[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2017, PP(99):1-1.
[12] Ikram M Z. Double-talk detection in acoustic echo cancellers using zero-crossings rate[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2015:1121-1125.