TELKOMNIKA, Vol.15, No.1, March 2017, pp. 172~178
ISSN: 1693-6930, accredited **A** by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v15i1.4053

Research on VoIP Acoustic Echo Cancelation Algorithm Based on Speex

Liu hua-zhu[1], Zhao xiao-fang*[2], Lin sheng-xin[3]
[1,2,3]School of Electrical Engineering, Dongguan University of Technology, Dongguan, 523808, China
[1]School of Electronics and Information, South China University of Technology, Guangzhou, 510640, China
*Corresponding author, e-mail: aozhy119@126.com

Abstract

Echo cancellation has been a major problem to be solved in VoIP, although the integrated echo cancellation module in Speex, it does not consider thread synchronization issues. The frequency domain echo cancellation algorithm MDF of speex is analyzed, and then a synchronization method of playing thread and recording thread is proposed. The results show that the acoustic echo canceller which achieved by the proposed method meet the requirements of voice communication, implementation is easier and therefore provides a reference for the VoIP voice communication and mobile communication terminal.

Keywords*: acoustic echo cancelation, MDF algorithm, synchronization, speex*

1. Introduction

VoIP (Over Internet Protocol Voice) is a kind of new network technology, which is used to realize the integrated transmission of voice and data services. Compared with traditional telephone, IP telephone highly favored by the market for its network bandwidth utilization, low cost, and can provide a variety of multimedia services. Recently, the traditional telecom services caused an impact because of the rapid development of VoIP technology [1, 2]. However, since when the VoIP voice transmission together with other data in the network to go through compression, coding, packaging and a series of treatment, resulting in a larger echo path delay, which seriously affecting the voice quality and hindering the development of VoIP markets.

Therefore, an increase in the VoIP terminal echo cancellation algorithm has become necessarily. Speex is an open source encoding algorithm in VoIP, which based on CELP (Code Excited Linear Prediction) encoding algorithm principle [3]. The encoder includes echo cancellation module, and gives the estimate of the best step, and is widely used. However, the echo cancellation algorithm in Speex requires the far-end signal and near-end signal holding consistency. And the play sounds and voices are often enrolled in two different threads is done in practical application, so the thread synchronization problem is there. This problem can cause the echo cancellation is not effective. Based on this, a thread synchronization method is proposed, which can effectively eliminate echo.

In VoIP phone, a general method is using the acoustic echo canceller to offset the echo in call, so that it can improve the quality of voice. A basic echo canceller includes two parts: the adaptive filter and doubletalk detector. Adaptive filter can eliminate echo by adaptively simulate echo path, its performance determines the effect of echo cancellation. In practice, the design of adaptive filter is usually with NLMS algorithm or some the improved NLMS algorithms. The reason is that the improved NLMS algorithm is concise, low complexity [4]. However, NLMS algorithm has a fatal weakness that when the input signal is correlation, such as voice signal, the algorithm convergence speed will be significantly reduced, thus affecting the quality of the echo cancellation. Therefore, various algorithms about decorrelation process is also put forward accordingly [5], so that the convergence rate of the NLMS algorithm is improved effectively. Doubletalk detector controls the update of the adaptive filter coefficient by judging the existence of the proximal voice to refrain from the divergence of the adaptive filter in Double-talk. The current Double - talk detection algorithm can be divided into three categories: based on energy detection, based on correlation detection and based on the echo path detection. The algorithm

Received September 26, 2016; Revised December 21, 2016; Accepted January 7, 2017

based on energy detection is simple and low complexity, but in the case of low SNR misjudgment rate is high. The judgment based on correlation detection is relatively accurate, but its computational complexity is too high. The algorithm complexity based on the echo path detection is a bit low, but the misjudge is easy to happen when the echo path has changed.

2. Principle of Speex Echo Generation

According to the causes of the echo, it can be divided into two categories: acoustic echo and electrical echo. Electrical echo is caused by the mismatch of circuit impedance, electrical echo are usually less affected. With the development of eliminate echo technology, the study of echo cancellation get to be a key point recently, they are already by the elimination of "electrical echo", turned to eliminate acoustic echo.

Acoustic echo is refers to the equipment of a portion of the sound signal back to the same device in the receiver and its principle as shown in Figure 1. Proximal microphone collected the voice of far end speech from loudspeaker, and back to the far end speaker, so the remote users will be able to hear their own voice, which is a great influence on call quality, necessary for echo cancellation. Basically there are the following two ways to eliminate:

(1) Echo suppressor

The echo suppressor is a kind of way based on nonlinear echo cancellation. It compared the voice received the preparation by loudspeaker with current sound level picked up from microphones through a simple comparator, to achieve the purpose of echo cancellation. With the emergence of high performance of the echo canceller, echo suppressor has rarely used.

(2) The acoustic echo canceller

Acoustic Echo canceller (AEC) is to use the loudspeaker signal and multipath echo correlation produced by the signal to build the speech model about the echo path, thus estimating the acoustic echo, then subtracting the estimated value of the echo from sampling signal of the proximal voice, and according to the feedback signal to adaptively modify coefficient to achieve more and more close to the real echo effect, so as to achieve the aim of eliminate echo.

The method of echo cancellation is to estimate the filter parameters for generating echo, then simulate the echo signal, and then subtract the analog echo signal from the proximal end to achieve the purpose of eliminating echo. In the adaptive echo cancellation, the key point is still adaptive algorithm, which directly affects the overall performance of the echo canceller. Therefore, in actually, the adaptive algorithms have to achieve faster convergence speed, small residual echo, tracking ability and so on, also it is necessary to consider the complexity of the algorithm. Speex is an open source encoding algorithm. The encoder includes echo cancellation module, which uses MDF (MultiDelay block Frequency-domain) algorithm and NLMS (Normalized Least Mean Square) algorithm [3]. Next, the principle of Speex echo cancellation algorithm is analyzed.

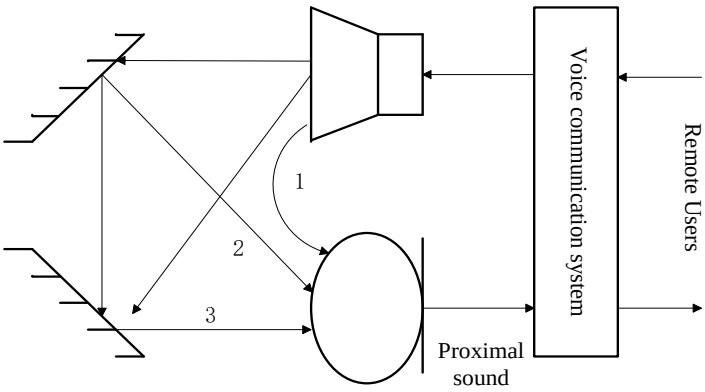

Figure 1. Acoustic Echo Generation Principle

3. Principle of Speex Echo Cancellation Algorithm

The block diagram of the Speex echo cancellation algorithm shown in Figure 2 [4, 6], the basic principle is to estimate filter parameters, simulation results echo signal , the analog signal is then subtracted from the proximal end of the recording signal, to achieve the purpose of eliminating echo.

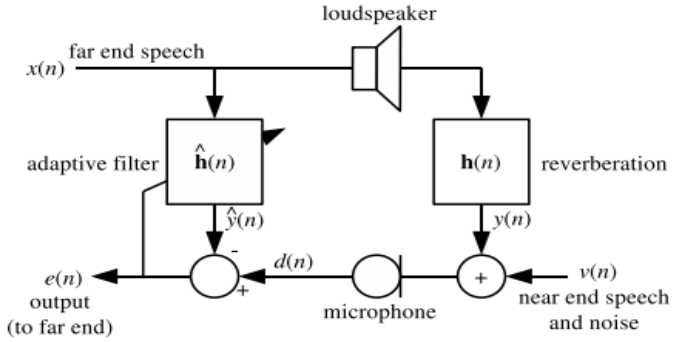

Figure 2. The Block Diagram of the Speex Echo Cancellation Algorithm

Among them, $x(n)$ is the speech signal from the far end, $v(n)$ is the proximal input speech, and $y(n)$ is the speech signal obtained by the $x(n)$ through the speaker and the echo path, $d(n)$ is $y(n)$ plus $v(n)$. As can be seen from Figure 2, the basic principle of the algorithm is to simulate the echo signal $y(n)$ by the adaptive filter, echo cancellation of the speech signal can be obtained with $d(n)$ minus $y(n)$. The parameters of the adaptive filter are adjusted by $x(n)$ and the output error signal $e(n)$. Background noise writing w(n), when the value of $v(n)$ is zero and proximal background noise ignores w(n), the echo path can be well estimated by adaptive filter, the larger components of the echo can be eliminated. When the value of $v(n)$ is not zero and is in bilingual environment, coefficient of filter can divergence, this kind of situation is caused by the proximal voice signal $v(n)$, most of the echo signal will be transmitted to the remote, In order to prevent the divergence of coefficient, once the proximal voice signal exist, through slow down or stop completely the update of adaptive filter coefficient value, the bilingual detector can accurately detect the activity status of the doubletalk voice. It will directly decide the effect of echo cancellation.

3.1. MDF Algorithm Principle

Echo cancellation theory is difficult to estimate the synchronization between the proximal input signal and the echo, and the problem how to process the speak to the double end, If these two problems are not solved well, it will cause the divergence of filter. not only can not eliminate echo, it will introduce more annoying noise. In the NLMS algorithm, it was assumed that the input proximal background noise and remote signal are white noise, so for it is the uncorrelated about time between the two signals, so we can obtain the most optimal step-length factor. But when use LMS/NLMS algorithm for speech signal to achieve the acoustic echo cancellation, the assumption that two signal are the uncorrelated about time is not established completely, so only by using frequency domain method.

Frequency domain adaptive algorithm has the advantages of low computational complexity, but have larger delay, is not conducive to real-time communication. In order to banlance the computational complexity and system time delay, it can consider using multiple delay frequency domain algorithms (MDF). MDF is an adaptive filtering method, which divides the original multiple filters into multiple sub blocks, and then performs adaptive filtering in each sub block. Set the filter length L, each read data frame length is N, and $L = MN$, wherein M is an integer (the number of filter delay block) [7, 8].

Algorithm works as follows:

$$X(M, j) = FFT(x(j-1), x(j)) \quad (1)$$

$$X(m, j) = X(m+1, j-1) \ , m = 1, 2, 3, \cdots, M \quad (2)$$

Where, $x(j)$ is the j-th frame of the far end speech, m represents the number of the block delay filter.

Then the echo signal output from the filter expression $y(j)$ as follows:

$$y(j) = \text{ The latter part of } FFT^{-1}[\sum_{m=1}^{M} X(m, j)W(m, j)] \quad (3)$$

Error signals in the frequency domain expression are:

$$E(j) = FFT\{0, \ldots, 0, [d(j) - y(j)]^T\}^T \quad (4)$$

Where, $d(j)$ is the sum of echo signal and the near end speech signal, $W(m, j)$ is the filter coefficient. And,

$$W(m, j+1) = W(m, j) + \mu\phi(m, j) \quad (5)$$

$$\phi(m, j) = FFT[\varphi(m, j), 0, 0, \ldots, 0]^T \quad (6)$$

$$\varphi(m, j) = \text{The first half of } FFT^{-1}[X^*(m, j)E(j)] \quad (7)$$

The calculated of the step length μ is the focus of algorithm, Speex calculate μ by adaptive algorithm.

3.2. Best-Step Calculation

The correlation of the input signal in the time domain is stronger than in the frequency domain. The learning rate $\mu(k,l)$ can be associated with frequency. Speex echo cancellation algorithm is independently using the NLMS algorithm to each frequency in the frequency domain [9, 10]. NLMS algorithm is a kind of improvement based on LMS algorithm, which inherits the most of the advantages and disadvantages of LMS algorithm. When the step factor is certain, the convergence rate of the NLMS algorithm is affected by the size of the input signal. When the stability of the input signal is relatively poor, to the rate of convergence of filter will be reduced. NLMS algorithm complexity does not increase and the algorithm is simple and has robustness.

Best step can be expressed as:

$$\mu(k,l) \approx \frac{\sigma_r(k,l)}{\sigma_e(k,l)} \quad (8)$$

Where, k is the sampling point number, l is a speech frame number, $\sigma_r(k,l)$ is the energy of frequency domain echo residual signal, and $\sigma_e(k,l)$ is the energy of output signal $e(n)$ after the echo cancellation in frequency domain. Wherein, $\sigma_r(k,l)$ readily obtained by estimating. The $\sigma_e(k,l)$ can be calculated by the following formula:

$$\sigma_r(k,l) = \eta(l)\sigma_{\hat{y}}(k,l) \quad (9)$$

Where, $\eta(l)$ is the leakage factor independent of frequency, change slowly and are difficult to estimate, and $\sigma_{\hat{y}}(k,l)$ is the energy of an analog echo in frequency domain, rapidly changing but easy to get. The estimation formula of $\eta(l)$ is as follows:

$$\hat{\eta}(l) = \frac{\sum_k R_{EY}(k,l)}{\sum_k R_{YY}(k,l)} \quad (10)$$

Among them:

$$R_{YY}(k,l) = (1-\beta(l))R_{YY}(k,l) + \beta(l)(P_Y(k,l))^2 \quad (11)$$

$$R_{EY}(k,l) = (1-\beta(l))R_{EY}(k,l) + \beta(l)P_Y(k,l)P_E(k,l) \quad (12)$$

$$P_Y(k,l) = (1-\gamma)P_Y(k,l-1) + \gamma(|\hat{Y}(k,l)|^2 - |\hat{Y}(k,l-1)|^2) \quad (13)$$

$$P_E(k,l) = (1-\gamma)P_E(k,l-1) + \gamma(|\hat{E}(k,l)|^2 - |\hat{E}(k,l-1)|^2) \quad (14)$$

$P_Y(k,l)$ is the energy of the k-th frequency point in the echo signal frame l, $P_E(k,l)$ the energy of the k-th frequency point in the frame l of the output signal after eliminate the echo signal. In equation (10), the molecule is on behalf of the cross-correlation values between estimate echo and error signal, the denominator is the autocorrelation value of estimate echo. If there is a doubletalk speech, the error will be very big, so the step-length factor will become very small, it won't make the filter coefficient change too big. If the background noise exists, because in the step-length factor formula, the molecular and denominator has been affected by noise, therefore, after the offset among each other, the influence of noise will become very small.

4. Thread Synchronization Method

Through the analysis of the principle of algorithm we known that algorithm is to use reference echo and the relationship between the real echo to work, namely remote data and proximal data should be consistent. But the play sounds and voices are often enrolled in two different threads is done in practical application, so there is a thread synchronization issues, and the propagation of echo in the true path will cause delay. These two issues will cause the distal and proximal signal are not synchronized, not only echo cancellation algorithm may lead to ineffective but also have the desired impact on sound data [11].

4.1. Thread Synchronization Step

Using the following method for synchronization:

1. When the voice signals are received and decoded in the proximal end, the voice signals are stored in a buffer queue for echo cancellation reference signals, and then through the speakers. Recording thread will do echo cancellation processing between the recording data and the corresponding data in buffer queue. Due to the network time delay, the influence of many factors such as sound equipment, audio input will delay for a period of time, the recording data and the data in the buffer queue does not match the wrong. Approved in actual system, through the estimate, the delay time is about 15 frames, so the tape input 15 frames do not echo cancellation, and then do synchronized echo cancellations with the data in the buffer queue.

2. Because the network delay is not stable, so using fixed delay time in the first step is not enough. So set two variables to tag the number of speech frame in the recording thread and playing thread. Variable mic_count tags the serial number of recording frame, and play count marks the serial number of far end signals in the buffer queue. When mic_count > paly_count, playing thread is slower than the recording thread, should increase the delay time frames in the first step; when mic_count < paly count, recording thread is slower than the playing thread, should reduce the delay time frames in the first step.

4.2. The Test Results

Speex is an open source, free, voice codec specifically for VoIP and it use a kind of dynamic bit rate coding way, this would mean it can dynamically modify the bitrates according to the change of network environment, it provides the corresponding version in narrowband and broadband. Echo cancellation has been urgent main problem to solve in VoIP, so in recent years integrating the echo cancellation in Speex module. Because Speex algorithm, AEC did not consider thread synchronization problems, it puts forward a kind of method solving thread synchronization between the playing thread and the recording thread.

The echo cancellation effect of the above solution was tested in a real environment. The voice from far end was saved as speak.pcm in playing thread, and saved the recording data as mic.pcm in the recording thread. Finally, the residual data was saved as echo.pcm after performed the echo cancellation algorithm. Converted three pcm files to wav files and displayed their waveforms by Matlab, which were shown in Figure 3.

In the Figure 3, the signal far (far-end signal) is the waveform of speak.pcm, signal near (near-end signal) is the waveform of mic.pcm, and the signal error (residual signal) is the waveform of echo.pcm. As can be seen from the figure, the near-end signal and the far-end signal are not aligned, and near-end signal has delay. From the time domain waveform of signal error can be seen that most echoes have been eliminated, and have good effect. And by a lot of subjective test, the remote user could not hear the echo or the sound is very small. The results show than the above method can eliminate echo well and make the voice quality to meet the requirements.

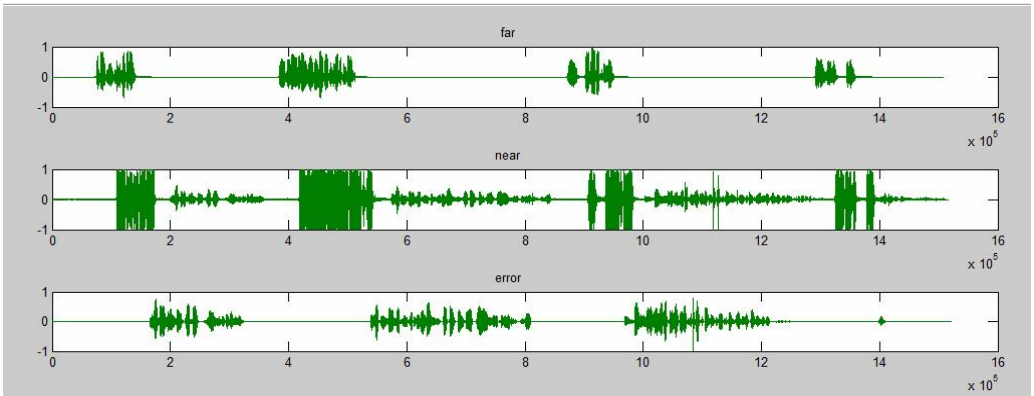

Figure 3. The Effect of Echo Cancellation

5. Conclusion

IP phone has a certain technical advantages and the characteristics of the times, once solved the problem of voice quality, IP phone will be widely used in the new generation of telecom network. This requires the new algorithm to eliminate echo, and further improve the convergence speed and reduce computational complexity, so that the voice quality of VOIP system as soon as possible to catch up with the traditional phone, so as to win more customers. Based on this, the echo cancelation theory and bosical structure of adaptive echo cancelation are introduced, and the multi-delay block frequency domain algorithm of speex is analyzed. Then a way of synchronization between the referenced echo and the record input signal is

proposed. The computer tests based on the speex open source project show that it performs better than current algorithm and is simple to implement.

Acknowledgements

This work was supported by the Science and technology project of Guangdong Province under Grand No. 2014A050503068 and No. 2015A010103019.

References

[1] Harjit Pal Singh, Sarabjeet Singh, J Singh, SA Khan. VoIP: State of art for global connectivity-A critical review. *Journal of Network and Computer Applications*. 2014; 37: 365-379.
[2] Septama Hery Dian. High available VoIP server failover mechanism in Wide Area Network. *TELKOMNIKA.* 2015; 13(2): 739-744.
[3] Valin JM. *On adjusting the learning rate in frequency domain echo cancellation with double-talk*. Audio, Speech, and Language Processing, IEEE Transactions on. 2007; 15(3): 1030-1034.
[4] Valin JM. Speex-manual [EB/OL]. http://www.speex.org/docs/manual/speex-manual.pdf 2007.
[5] Krishna EH, Raghuram M, Madhav KV, Reddy KA. *Acoustic echo cancellation using a computationally efficient transform domain LMS adaptive filter*. 10th International Conference on Information sciences signal processing and their applications (ISSPA). 2010: 409-412.
[6] Breining Christina. Robust fuzzy logic-based step-gain control for adaptive filters in acoustic echo cancellation. *IEEE Transactions on Speech and Audio Processing*. 2001; 9(2): 162-167.
[7] Nguyen-Ky T, Leis J, Xiang W. An improved new error estimation algorithm for optimal filter lengths for stereophonic acoustic echo cancellation. *Computers and Electrical Engineering*. 2010; 36(4): 664-675.
[8] Wada Ted S, Juang Biing-Hwang. Enhancement of residual echo for robust acoustic echo cancellation. *IEEE Transactions on Audio, Speech and Language Processing*. 2012; 20(1): 163-177.
[9] Zhanjun Liu, Yue Shen, Zhonghua Yu, Fengxie Qin, Qianbin Chen. Adaptive Resource Allocation Algorithm in Wireless Access Network. *TELKOMNIKA*. 2016; 14(3).
[10] Zoran M Šarić, Istvan I Papp, Dragan D Kukolj, Ivan Velikić, Gordana Velikić. Partitioned block frequency domain acoustic echo canceller with fast multiple iterations. *Digital Signal Processing.* 2014; 27: 119-128.
[11] Contan Cristian, Zeller Marcus, Kellermann Walter, Topa Marina. *Excitation-dependent stepsize control of adaptive volterra filters for acoustic echo cancellation*. Proceedings of the 20th European Signal Processing Conference. 2012: 604-608.