Design and Implementation on a Sub-band based Acoustic Echo Cancellation Approach

Zhixin Chen
ILX Lightwave Corporation
Bozeman, Montana, USA
chen.zhixin.mt@gmail.com

***Abstract*— This paper describes the design and implementation of a sub-band based acoustic echo cancellation approach, which incorporates the normalized least mean square algorithm and the double talk detection algorithm. According to the simulation, the proposed approach works well in the modest linear noisy environment. Since the proposed approach is implemented in fix-pointed C, it can be easily ported into fixed-point DSPs to cancel acoustic echo in real systems.**

Keywords - Digital Signal Processing; Adaptive Filter; Acoustic Echo Cancellation.

I. INTRODUCTION

People have been using telephones as a way of distant voice communication for more than one century now. In a phone conversation, echo is the sound of one's own voice being played back to him after a delay. There are two types of echo presenting in typical communication networks: hybrid echo and acoustic echo. Hybrid echo is the outcome of impedance mismatches along a telephone line, and is relatively easy to identify and cancel. Acoustic echo is due to the leakage from loudspeaker to microphone in a hands-free telephone, and is much more difficult to tackle [1].

Strong and long delayed acoustic echo signals can be very annoying, and in some cases, make conversation impossible. The effective removal of acoustic echo is thus the key to maintaining and improving voice quality on a phone call. This has led to intensive research into the area of echo cancellation, with the aim of providing solutions that can significantly reduce or even remove acoustic echo [1-3].

Good acoustic echo cancellation system can greatly enhance the audio quality of a communication system by allowing conversation to progress more smoothly and naturally, keeping the participants more comfortable, and preventing listener fatigue. However, a poorly designed acoustic echo cancellation will not provide these benefits and can even degrade audio quality significantly. This paper describes the design and simulation of an acoustic echo cancellation (AEC) approach, which is based on the sub-band method and the normalized least mean square (NLMS) algorithm, and includes the double talk detection. According to the simulation, the proposed acoustic echo cancellation approach works well in the modest linear noisy environment. Since the approach is implemented in fixed-point C language, it can be easily ported into fixed-point DSPs and used in real systems to cancel acoustic echo.

The remaining part of this paper is organized as follows. First, the principle of acoustic echo cancellation is reviewed. Next, the implementation of the proposed acoustic echo cancellation approach is described. Then, the simulation results of the proposed approach are presented. Finally, the paper is concluded with a summary of results.

II. PRINCIPLE OF ACOUSTIC ECHO CANCELLATION

The way how acoustic echo canceller works is shown in Figure 1. In general, the near end microphone signal d(n) consists of signals from near end microphone, near end background noise and echo of the far-end audio signal originating from the loudspeaker [1]. The last signal is undesired. The objective of the echo canceller is thus to form a replica of the acoustic echo signal picked up by the terminal microphone by inserting an adaptive filter h(n) parallel to the signal path through loudspeaker, room, and microphone [4]. The objective of the adaptive filter is to provide a response as equal as possible to that of the acoustic signal path. The far end speech signal x(n) is then fed through the adaptive filter h(n) to generate the signal y(n) to resemble the echo part of the microphone signal. The echo y(n) is then subtracted from the microphone signal d(n) to get the desire speech signal e(n).

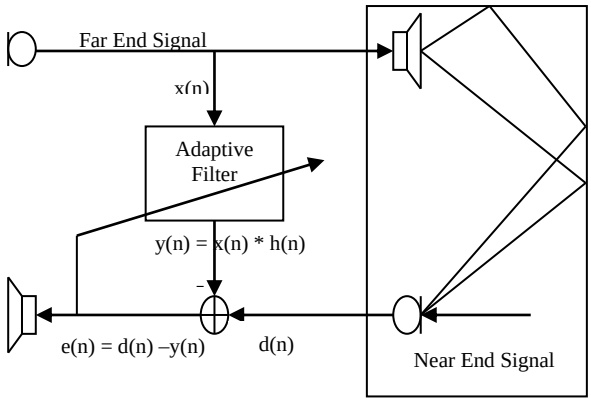

Figure 1: Principle of the acoustic echo cancellation system.

III. IMPLEMENTATION OF ACOUSTIC ECHO CANCELLATION APPROACH

From the previous description, the acoustic echo cancellation system appears as a straight-forward realization of an adaptive filtering algorithm. Some algorithms, such as LMS or NLMS, can be used in the full speech band to implement it directly [5]. However, one issue for the full-band

implementation is the high computational complexity. For example, in a typical conference room, the environment echo delay varies from 100 to 500 milliseconds. For a sampling rate of 8 kHz, this means about 800 to 4000 samples and thus requires a very long adaptive filter length. Moreover, if the full band method is used, the spectrum may vary dramatically within the full band of 4 KHz [6]. Consequently, the sub-band method was used in the proposed acoustic echo cancellation approach because the spectrum will not vary a lot within each sub-band.

The LMS algorithm is the most popular algorithm in adaptive signal processing due to its simplicity and reliability. Its main drawback is that the speed of convergence gets very slow if there is a big spread among the eigenvalues [6]. Thus, the normalized NLMS algorithm is used in the proposed approach because it is insensitive to the amplitude variations of the inputs signal so that it can achieve a robust performance.

Double talk is another important issue related to acoustic echo cancellation. If the state of double talk cannot be detected accuarately, it can cause divergence of the adaptive algorithm. Consequently, a robust double-talk and single-talk detection algorithm was designed in the proposed AEC approach.

In the following sections, the sub-band approach, the LMS and NLMS algorithm, and the double talk detection algorithm will be presented.

A. Sub-band Approach

As mentioned previously, the sub-band implementation has two advantages. First, the complexity is reduced by dividing the signal into sub-bands and applying adaptive filters to a decimated signal in each sub-band. Second, the spectral variability within a sub-band is reduced compared with the full band signal [1]. To maintain the high quality of speech, the analysis and synthesis filterbanks in the sub-band realization must be designed to provide perfect reconstruction, which means that the signal fed through the analysis and synthesis filterbank system shall be an exact but delayed copy of the input signal. In the proposed AEC approach, the Quadrature Mirror Filters (QMF) was used to decompose the full band into four sub-bands with an identical bandwidth of 1 kHz. The reason that the full band is divided into four sub-bands is that the energy for speech focuses on the lowest 1 kHz range and the spectrum will be almost constant within the lowest 1 kHz sub-band. The basic architecture for a two-channel QMF system is shown in Figure 2.

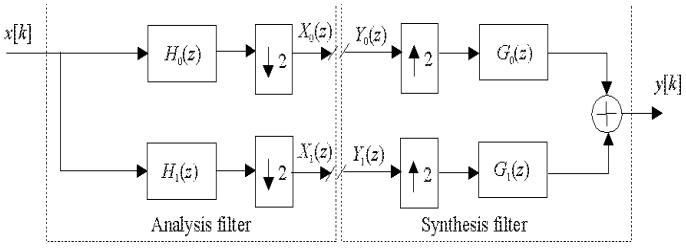

Figure 2: A block diagram of the QMF filter architecture

In Figure 2, the down arrows represent decimation by 2 and the up arrows denote interpolation by 2. The two-channel QMF system consists of two input-output paths, each of which has a bandwidth requirement that is half the original bandwidth requirements. In Figure 2, the top path contains a low-pass filter while the bottom path has a high-pass filter. Since the FIR filter is used here, a very steep filter cannot be achieved without a very large filter tap. The frequency response of the top path will have some overlap with that of the bottom. The overlap that exists beyond one-fourth the sampling frequency in the top channel, when re-sampled at half the sampling frequency, introduces aliasing errors. The bottom path also introduces aliasing errors. According to [1], the aliasing errors in the top and bottom path are self-cancelling if the following condition is met.

$$(H_0(-z)G_0(z) + H_1(-z)G_1(z))X(-z) = 0, \quad (1)$$

which is satisfied if $G_0(z) = H_1(-z)$ and $G_1(z) = H_0(-z)$. A special case assumes that $H_0(z)$ and $H_1(z)$ are sub-band filters satisfying the mirror filter relationship $H_1(z) = H_0(-z)$ [1, 6].

Based on the above-mentioned theory, a 45th order linear phase filter that is shown in Figure 3 was used in this paper. It is shown that the low-pass and high-pass filter are mirror symmetry. This QMF was used twice to get four sub-bands.

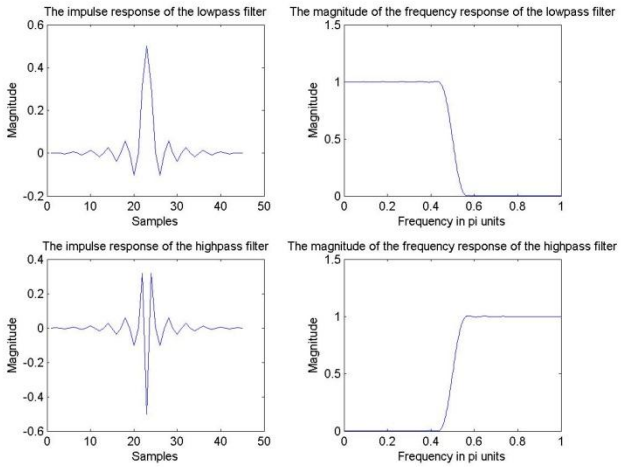

Figure 3: The impulse and frequency response of the QMF used in the proposed AEC approach

B. NLMS Algorithm

It was described previously that the goal of the AEC system is to identify an unknown system, the loudspeaker-to-microphone transfer function, with an adaptive filter. That is, the input signal needs to be processed so that it matches the desired signal $d(n)$. To measure how well the cancellation is done, a performance criterion $J(e)$ to the error is used here. According to [4], the performance criterion is chosen to be a squared error measure and it can be defined as follows.

$$J(e) = \frac{1}{2}E(e^2(n)) = \frac{1}{2}E[(d(n) - y(n))^2]. \quad (2)$$

Here the room transfer function is assumed to be linear and wide-sense stationary and the desired and input signals are assumed to be jointly wide sense stationary and zero mean. The

filter is composed by a certain number of parameters $h_n(k)$ that are set to minimize the error $J(e)$. For FIR filters, the output of the filter can be written in the form $y(n) = \sum_{k=0}^{N-1} h_n(k)x(n-k)$, (3)

where N is the filter order and $x(n)$ is the input signal. The optimal $h_n(k)$ parameters are found by solving

$$\nabla_h(J) = \nabla_h(E(e^2(n))) = 0, \quad (4)$$

which leads to the Yule-Walker equation [4, 5, 7].

There are many algorithms for this equation, among which the most popular ones are LMS and NLMS. In the LMS algorithm, the adaptation rules are shown in the following equation:

$$\hat{h}_{n+1}(k) = \hat{h}_n(k) + \mu e(n)x(n-k), \quad (5)$$

where $k = 1,2,....,N$ and $e(n) = d(n) - y(n)$. Here $\hat{h}_{n+1}(k)$ are the filter parameters at time $n+1$ and μ is a constant adaptation step. The algorithm is proven to converge if the error-performance surface is quadratic and if the adaptation step satisfies $\mu < \frac{1}{\lambda_{\max}}$, where $\lambda_{\max}$ is the biggest eigenvalue of input correlation matrix R. Usually, the following condition is taken

$$\mu < \frac{2}{input\ \ power}. \quad (6)$$

The main drawback of the LMS algorithm is the speed of convergence gets very slow if there is a big spread among the eigenvalues of R. Different from LMS, the NLMS algorithm is insensitive to the amplitude variations of the input signal. The NLMS algorithm is given by

$$h_{n+1} = h_n + \beta \frac{x^*(n)}{||x(n)||^2} e(n). \quad (7)$$

According to [5], the NLMS algorithm converges in the mean-square if $0 < \beta < 2$. In the LMS algorithm, the correction that is applied to $h_n(k+1)$ is proportional to the input vector $x(n)$. Therefore, when $x(n)$ is large, the LMS algorithm experiences a problem with gradient noise amplification. With the normalization of the LMS step size by $||x(n)||^2$ in the NLMS algorithm, however, this noise amplification problem is diminished. However, a similar problem occurs when $x(n)$ becomes too small. An alternative way is therefore to use the following modification to the NLMS algorithm:

$$h_{n+1} = h_n + \beta \frac{x^*(n)}{\varepsilon + ||x(n)||^2} e(n), \quad (8)$$

where ε is some small positive number [4, 5].

In the proposed AEC approach, the echo delay that this system can handle is set to 360 ms, which means that the taps of the adaptive FIR filters will be up to 720. Since the speech energy is different for different sub-bands, the taps of the filter are different for different sub-bands. The taps are 720, 540, 360, and 240 for the sub-bands from low to high respectively. The step size was also chosen to be larger for higher sub-bands, which mean that fast convergence is expected when the echo signal has shorter delay and has higher frequency components. This is because the speech energy focuses on the lower frequency and so does the echo.

For the updating of the filter coefficient, Eq. (8) was used. Here ε was chosen to be 10^{-8} and the step size was chosen from 10^{-5} to 10^{-4}, which is much larger than the traditional value of 10^{-7} but still gets a good result.

C. Double Talk Detection

If the coefficients of the adaptive filters are updated during double talk, the filter may be divergent. A double talk detection algorithm was thus included in the proposed AEC approach to detect the double talk and is described as follows. For both far end and near end signal, a long-term energy and a short-term energy were used to judge whether it is silent or not. The coefficients of the adaptive filters will only be updated when the far end talks and the near end does not talk because the transfer function of the room acoustics is expected in this case. Three convergence states, low, medium, and deep convergence states are defined here and shown in Figure 4. Initially, the system is in low convergence state. When neither far end nor near end talks, the system will keep the convergence state. When the far end talks and the power of the echo reduced signal is less than that of the near end signal for at least two seconds, the system will transit into medium convergence state. The system will transit into deep convergence state if this condition continues for two more seconds. If double talk stays for one second, the system will transit into the low convergence state no matter what state it is in.

If the system is in deep convergence state, the filter coefficients will not be updated since in this case the room transfer function will not change. If the system is in low or medium convergence state, the coefficients will be updated when the far end talks. The step size for low convergence state is larger than that for medium convergence state because the system is expected to converge faster initially but converge slower when it is already in the modest convergence.

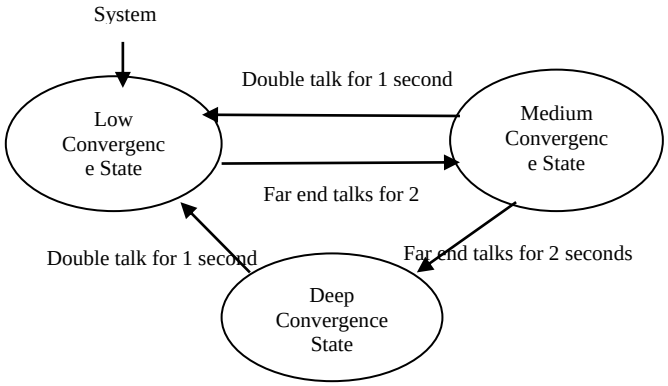

Figure 4: Transition among the three convergence states

With the single talk and double talk detection and the flexible transition of the convergence states presented above, the system performs very well in the modest noise environment.

IV. Simulation of the Acoustic Echo Cancellation Approach

The proposed acoustic echo cancellation approach was implemented using fix-point C language. The simulation results for several different scenarios are presented in the following subsections.

A. Only Near End Talks

In the first scenario, the far end is set to silent. This simulates the case that the microphone is turned off in the far end. The simulation is to test that the quality of the near end signal does not degrade. The system performs as expected in this case, which can be seen in Figure 5. In the figure, R_{in} means the far end signal, S_{in} denotes the near end signal, and S_{out} represents the desired near end output signal.

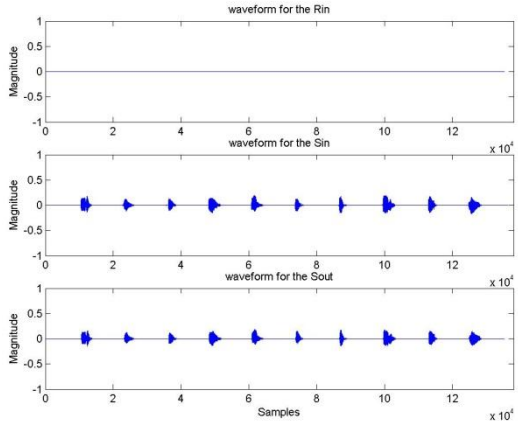

Figure 5 : Far end does not talk

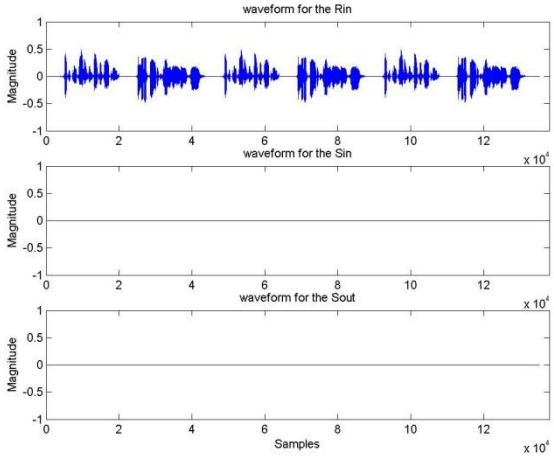

Figure 6: Microphone is turned off in the near end

B. Only Far End Talks

In the second scenario, the near end was set to silent, which simulates the case that the microphone is turned off in the near end. The goal of this simulation is to verify that there is no output signal. Figure 6 shows that the system performs as expected.

In the third scenario, the microphone in the near end is set to open. This means that an echo might present in the near end. The goal of this simulation is to test how well the system deletes the echo. The simulated echo signal is a 20% attenuation and 300 ms delay version of the original signal. As shown in Figure 7, the system completely removes the acoustic echo.

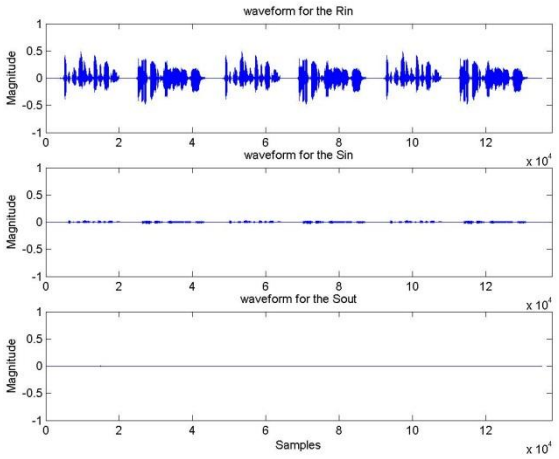

Figure 7: Near end does not talk but the microphone is open

In the following scenario, the effect of the number the sub-bands was tested. One to four sub-bands were used for the signal and it is shown that the simulation result is almost the same. It takes about 20000 samples (2.5 seconds) for the system to converge and remove the acoustic echo. The similar results for different number of used subbands are because the energy of the speech focuses on the low frequency range and so does that of the echo. Consequently, it does not matter whether the higher sub-bands were processed or not. More details can be seen in Figure 8.

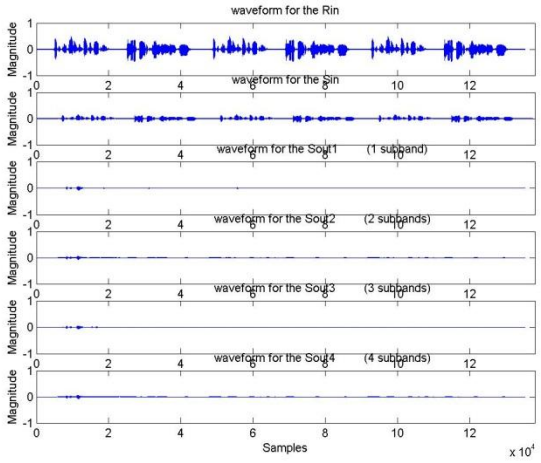

Figure 8: Effect of the number of sub-bands on the AEC system performance

In the former scenarios, the delay is constant for the entire signal. In this scenario, the delay is not constant, which means that the delay for the first sentence can be 150 ms while that for

the second sentence is 200 ms. The system also works well in this case, which is shown Figure 9.

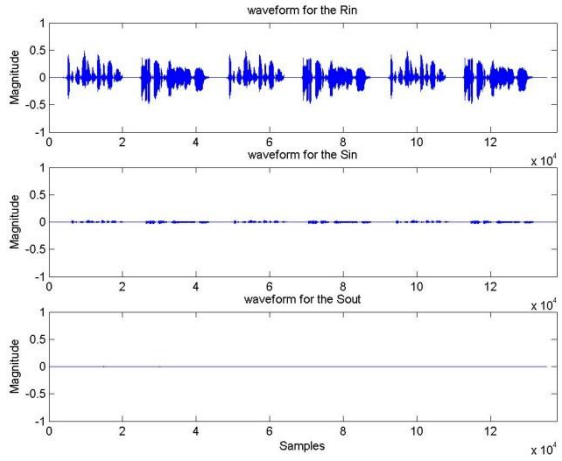

Figure 9: Simulation of irregular delay

Till now, only one echo component is considered in the simulations. In this scenario, the case that the echo includes many copies of the original signal with different delays and different gain is investigated. This means that the speech from the speaker can be reflected by the wall and then fed into the microphone in 10ms, 150ms or 200ms. It is shown in Figure 10 that the system also works well in this case.

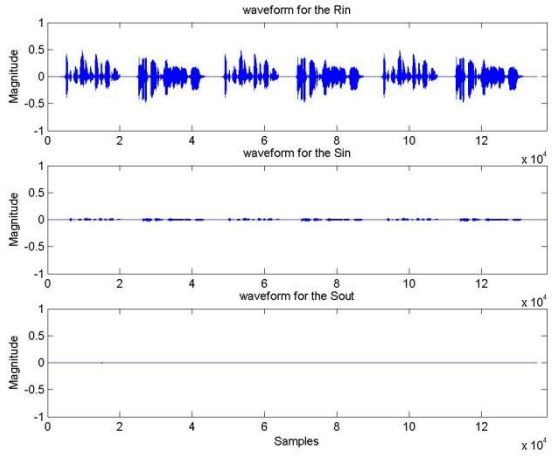

Figure 10: Simulation of many delays

C. Double Talk

In the following scenarios, both the far end and the near end are assumed to be active. A simulation was run to investigate the case that the far end and the near end talk in different time, which happens in most of the telephone communication systems. It is expected that if the far end and the near end do not talk at the same time, the performance should be as good as that of the single talk. The simulation result shown in Figure 11 verifies this.

In another simulation, some white noise was added in the near end. Since the noise is a part of the near end signal, it should not affect the performance. Figure 12 verifies the simulation result.

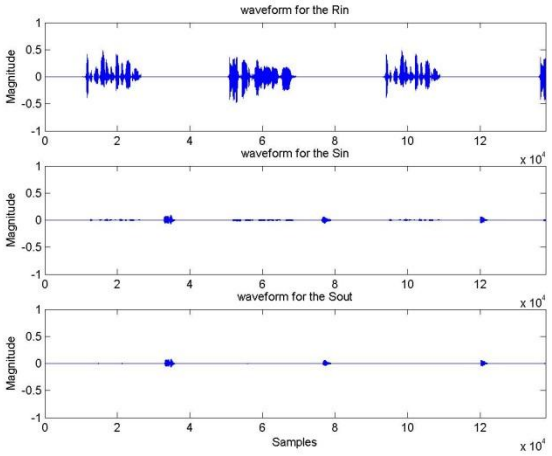

Figure 11: Far end and near end talk in different time

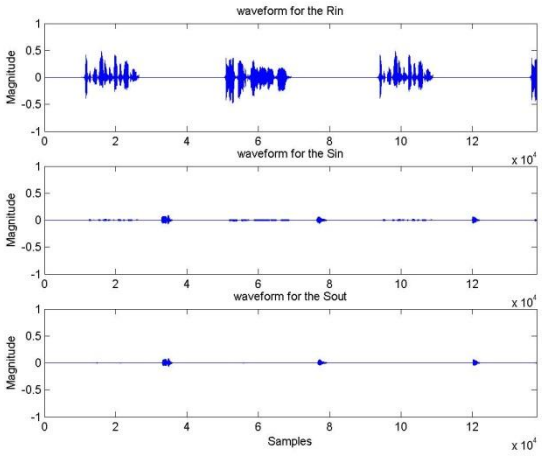

Figure 12: Noise shows up in the near end

It is assumed that the far end and the near end talk at the same time in the last scenario. Since the room transfer function changes from time to time, it is expected that the performance will degrade. Figure 13 shows the detailed simulation result.

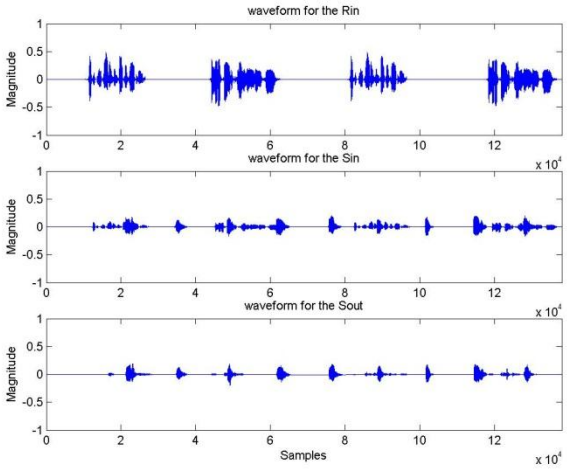

Figure 13: Far end and near end talk in the same time

V. CONCLUSION

In this paper, an acoustic echo cancellation system based on the sub-band approach and the NLMS algorithm is proposed. A double-talk detection algorithm is also included in the system. From the simulations, the system works well for the echo with 300 ms delay in modest noise environment. It works very well when only one end talks or when both end talk in the different time. But it does not work very well when both end talk in the same time. One thing deserving the attention is that this system is designed for the linear echo delay, so it is not appropriate for the environment that leads to echoes with new frequency components. An acoustic echo system which can eliminate the nonlinear echoes will be designed in the future.

REFERENCES

[1] A. Gilloire and M. Vetterli, "Adaptive filtering in subbands with critical sampling: analysis, experiments, and application to acoustic echo cancellation", IEEE Trans. Signal Process., vol. 40, p.1862-1875, 1992

[2] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, "Advances in network and acoustic echo cancellation", 2001, Springer.

[3] C. Antweiler, J. Grunwald and H. Quack, "Approximation for optimal stepsize control for acoustic echo cancellation," Proc. ICASSP-97, IEEE Conference on Acoustics Speech and Signal Processing, München, Germany, vol. 1, pp. 295-298, 1997

[4] Paolo Prandoni, "Acoustic echo cancellation" http://lcavwww.epfl.ch/~prandoni/dsp/echo/echo.html

[5] K. A. Lee and W. S. Gan "Improving convergence of the NLMS algorithm using constrained subband updates", IEEE Signal Process. Letter, vol. 11, p.736-739, 2004.

[6] G. Zelniker and F. Taylor, "Advanced digital signal processing", Marcel Dekker, 1994

[7] M. G. Siqueira, "Adaptive filtering algorithms in acoustic echo cancellation and feedback reduction", Ph.D. thesis, University of California, Los Angeles, 1998

AUTHORS PROFILE

Zhixin Chen is a Firmware Engineer in ILX Lightwave Corporation. He holds a BS degree from Xiamen University in China, a MS degree from Xiamen University in China, and a Ph.D. degree from Montana State University in USA, all in Electrical and Computer Engineering. His research interest and working experience are in the area of acoustics, audio, and speech processing, multimedia communication, and embedded system design for high power current source, temperature controller, and optical power meter.