

Acoustic Echo Cancellation for the Telepresence System

Inseon Jang, Hyun Woo Kim, Do Young Kim

Future Internet Research Department
ETRI (Electronics and Telecommunications Research Institute), KOREA
{jinsn, kimhw, dyk}@etri.re.kr

Abstract— This paper presents an implementation of an acoustic echo cancellation (AEC) system for the Telepresence environment where high quality speech communication is required. Implemented AEC system consists of the blocks related to pre/post-processing, the partitioned block frequency domain adaptive filter (PBFDAF) with a variable step size. Simulation results using real audio recording with the Telepresence system confirms superiority of the Implemented AEC.

Keywords— Acoustic Echo cancellation, Telepresence, Video Conference, Smart Work

I. INTRODUCTION

Recently the Telepresence is getting spotlight in the services related to the Smart Work. It can achieve cost cutting and production growth innovatively so that it is expected to contribute fast decision making, growth of production and balance between work and private life of employees as well as reduction of greenhouse gas emissions. [1] The Telepresence provides two or more locations to communicate by simultaneous two-way video and audio transmissions. Also it allows users to feel as if they were present, to give the appearance of being present, or to have an effect at a place other than their true location. It is an integrated solution with high quality video and audio, user convenient tools (e.g. white board and document conferencing) and other interior including furniture, lighting and so on. Regarding the audio, it is required to provide high quality speech communication and spatial sound.

In the Telepresence environments, the hands-free speech communication is usually used to provide not only convenience but also naturalness so that acoustic echo is inevitably produced by the acoustic coupling between the loudspeaker and the microphone. It affects the conversation quality significantly - most people find it difficult to carry on a normal conversation when the delay is over 250ms - so that the effective acoustic echo cancellation (AEC) is required for the hands-free communication, especially the Telepresence.

In this paper, we describe the implementation of AEC for the Telepresence system, which consists of several blocks which are a DC blocking filter, a decorrelation filter, an adaptive filter based on PBFLMS (Partitioned Block Frequency domain Least Mean Square) and a post-processing.

The PBFLMS based adaptive filter supports faster convergence and lower time delay than the other adaptive filter algorithms. [2], [3], [4] In this AEC system, it is combined with the variable step size algorithm. [5], [6] It supports to control the convergence parameter of the adaptive filter so that the adaptive filter coefficients are slowly adapted while the near-end talker is active, such as the double talk, whereas they are fast adapted while only far-end talker is active. Besides, the DC blocking filter and the decorrelation filter provide robust performance and fast convergence of the adaptive algorithm, respectively. Also the post-processing suppresses the residual echo level after adaptive filtering.

In the following, section II briefly introduces the Telepresence and requirements related to its audio. Then, section III gives the description of the implemented AEC system. In addition, simulation results of the implemented AEC system are shown in section IV. Finally, section V concludes the works.

II. TELEPRESENCE AND ITS AUDIO

The Telepresence is an advanced concept of conventional teleconference (or a video conference) which provides two-way, interactive audio and video communications between two or more locations.

The Telepresence solutions use video and audio conferencing components as well as other technologies to create a two-way immersive communications experience that simulates an in-person, interactive encounter so that it provides user the same room experience. For the same-room illusion, Telepresence solutions combine many kinds of technology elements such as high quality video/audio and environmental design, and consequently high level of the Telepresence service is accomplished, for example, Cisco Telepresence System 3000 [7] and RealPresence Experience (RPX) of the Polycom [8].

Especially the audio, it is required to be clear, without noise, echo-free and intelligible. Wider band audio, e.g. SWB (Super Wide Band) audio whose bandwidth is 50~14kHz, and related speech processing technologies such as noise suppression, AEC and AGC (Automatic Gain Control) are able to support high quality voice communication. Also spatial audio can be achieved by directional cues as regards who is speaking, even though several people are sitting around a table.

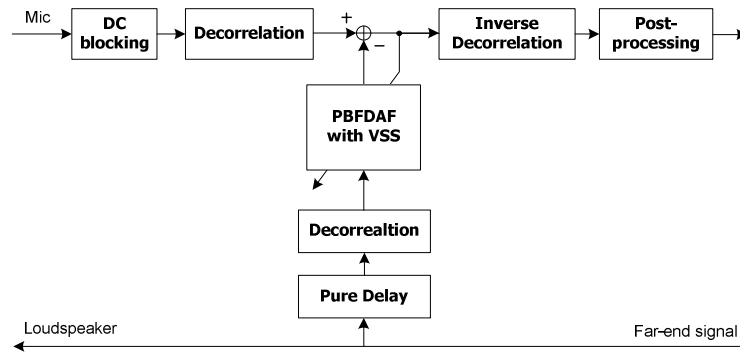


Figure 1. Block diagram of AEC implementation

III. ACOUSTIC ECHO CANCELLATION

AEC achieves the echo removal by estimating the echo path impulse response of the loudspeaker-enclosure-microphone (LEM) system.

The conceptual block diagram which is shown in Figure 1 illustrates the operation of the AEC system. It consists of the blocks related to the pre-processing, the PBFLMS based adaptive filter and the post-processing.

A. Pre-processing

The DC blocking filter and the decorrelation filter are operated prior to the adaptive filter in order to improve the echo path modelling.

The notch filter is used to eliminate the DC offset which is present in the microphone signal. In this system, it is adopted from Speex AEC [9] and its cut-off frequency is around 50Hz for SWB audio.

The decorrelation filtering transforms speech into more white noise like signal, so that the energy of high frequency signals is similar to the energy of low frequency signals. In result, the performance of the adaptive filter is more improved. In this system, a fixed first order decorrelation filter with $\alpha = 0.9$ in [2] is used. Two decorrelation filters are applied to far-end signal and microphone signal respectively. In order to recover true error signal, the inverse decorrelation filter is applied to difference signal between the decorrelated far-end signal and the estimated echo of the adaptive filter.

B. Adaptive Filtering

The adaptive filter plays a central role in the AEC system. It attempts to dynamically identify the acoustic echo path. As long as the room impulse response of the echo path can be quickly and accurately determined, it is then to generate a good estimate of the echo and to subtract it from the microphone signal.

In this system, PBFLMS (Partitioned Block Frequency Domain Least Mean Square) algorithm is adopted. Generally, the frequency domain adaptive filter provides lower

complexity and faster convergence than the time domain adaptive filter. But it has unavoidable additional time delay because of the block processing. PBFLMS makes a sequential partition of the impulse response in the time domain prior to the filtering operation in the frequency domain and the time segmentation allows the adaptive filter to set up individual coefficient updating strategies concerning to different sections of the adaptive filter. Hence it support low time delay with still lower complexity. Further detailed information and formula related to PBFLMS, please see [3], [4].

Detecting the double talk is inevitably required to prevent the adaptive filter coefficients from diverging when both a near-end speech and a far-end speech are present. In order to achieve both double talk detection and fast convergence of the adaptive filtering in the proposed, the variable step size (VSS) algorithm is combined with PBFLMS. The adopted VSS algorithm was derived by taking into account the existence of the near-end speech and the nonstationarity of speech signal as well as the undermodeling noise. [5], [6] Its remarkable results under near-end signal variations were reported.

C. Post-processing

In the real Telepresence environments, acoustic echo cannot be sufficiently eliminated by an adaptive filter alone. The adaptive filter must be followed by a residual echo suppressor as the post-processing, which attenuates the residual echo to a level where it is unnoticeable to the person on the far side. Residual echo suppressor is inherently a nonlinear processor as most of the remaining echo arises due to the echo path non-stationary nature and the system (i.e. loudspeaker and microphone) nonlinearities which are not linearly related to the reference signal.

In this AEC system, the post-processing is adopted from Speex AEC [9], which is based on not only signal processing techniques such as OM-LSA (Optimally-Modified Log-Spectral Amplitude) speech estimation and MCRA (Minima Controlled Recursive Averaging) noise estimation but also psychoacoustic approaches considering the human voice perception. [10], [11]

IV. SIMULATION

A. Setup

Echo signals are recorded using the ETRI Telepresence system in the multimedia room whose reverberation time is approximately 150ms. The audio specification of the Telepresence system is followed:

- Speech Codec: G.711.1 (at 128kbps)
- Loudspeaker: BOSE Companion 2 multimedia speaker
- Microphone: ETM-003 of Edutige
 - Condenser / Omnidirectional / Boundary microphone
 - Sensitivity: -23dB
 - Frequency response: 50~18kHz

A test sequence consists of a single talk, a double talk, a near-end talk only and a silence, in order. The microphone signal comprises 3 test sequences. In this simulation, four kinds of the microphone signal are used, which are recorded with various volumes.

AEC implementation is setup with PBFLMS filter of length 2560 with 4 partitions.

B. Results

Figure 2 shows the far-end speech, the microphone signal and its AEC result signal with the step size value change. As shown in bottom of the figure, the step size of the adaptive filter is close to zero during the double talk, the near-end talk only and the silence whereas it has larger value during the single talk. In results, the echo is remarkably suppressed after AEC and near-end speech survived well without divergence.

To evaluate the performance of AEC, the error return loss enhancement (ERLE) and the convergence speed are calculated. The ERLE measures how much echo attenuation removed from the microphone signal so that it is defined as the ratio of the microphone signal's power and the residual error signal's power. The convergence speed measures the time interval which the attenuation of the echo reaches 20dB. [12]

One result of the implemented AEC system is shown. The average ERLE in single talk period is about 37.3dB and the convergence time is about 492ms.

V. CONCLUSIONS

In this paper, we have described an implementation of the AEC. The implemented AEC system combines the PBFLMS based adaptive filtering with the variable step size as well as several pre/post-processing. Draft simulation results shows that proposed AEC system can enhance echo cancellation performance during both a single talk and a double talk period in a high quality speech communication system such as the Telepresence, and its performance meets the standard requirement for convergence time. [12] Future work will focus on the stereo and multichannel setup for

loudspeakers/microphone, which is required for spatial audio for the Telepresence.

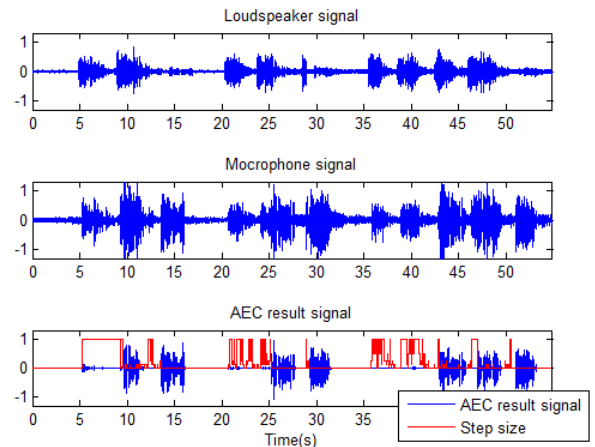


Figure 2. The far-end speech signal, microphone signal (including 3 test sequences with single talk, double talk, NET signal only and silence) and the AEC result with the step size value change.

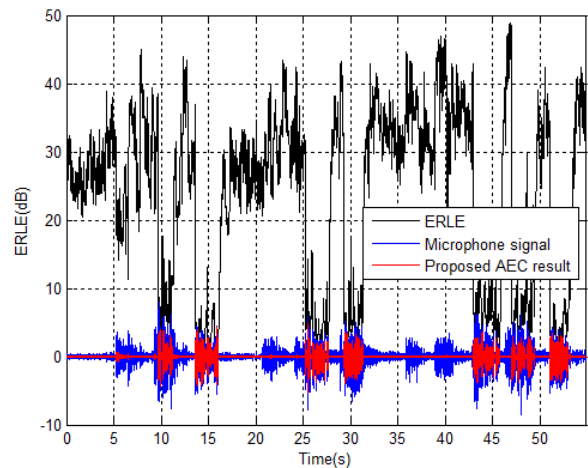


Figure 3. Performance of the implemented AEC

ACKNOWLEDGMENT

This research was supported by the KCC (Korea Communications Commission), Korea, under the ETRI R&D program supervised by the KCA (Korea Communications Agency). (KCA-2012-11921-05001)

REFERENCES

- [1] E. J. Lee, S. J. Yeon and W. G. Ha, "Market trends and current status on enterprise telepresence," *Electronics and Telecommunication Trends*, vol. 26, no. 3, pp.122~131, Jun. 2012.
- [2] E. Hansler and G. Schmidt, *Acoustic Echo and Noise Control-A Practical Approach*, Wiley, 2004.
- [3] J. M. P. Borrallo and M. G. Otero, "On the implementation of a partitioned block frequency domain adaptive filter (PBFDAF) for long

- acoustic echo cancellation,” *Signal Process*, vol. 27, pp. 301–315, Jun. 1992.
- [4] K. S. Chan and B. Farhang-Boroujeny, “Analysis of the partitioned frequency-domain block LMS (PFBLMS) algorithm,” *IEEE Trans. Signal Process*, vol. 49, no. 9, pp. 1860–1874, Sep. 2001.
 - [5] F. Albu, C. Paleologu, The Variable Step-Size Gauss-Seidel Pseudo Affine Projection Algorithm, pp. 642-645, in *Proc. ICCSE*, Dubai, E.A.U, 2009.
 - [6] C. Paleologu, S. Ciochina, and J. Benesty, “Variable step-size NLMS algorithm for under-modelling acoustic echo cancellation”, *IEEE Signal Processing Letter*. vol.15, pp.5-8, 2008.
 - [7] <http://www.cisco.com>
 - [8] <http://www.polycom.com>
 - [9] <http://www.speex.org/>
 - [10] I. Cohen and B. Berdugo, “Speech enhancement for non-stationary noise environments,” *Signal Processing*, vol. 81, no. 2, pp. 2403-2418, 2001.
 - [11] S. Gustafsson, R. Martin, P. Jax and P. Vary, “A psychoacoustic approach to combined acoustic echo cancellation and noise reduction,” *IEEE Trans. on Speech and Audio Processing*, 2002.
 - [12] ITU-T Recommendation G.167 – Acoustic Echo Controllers.