DSP IMPLEMENTATION AND PERFORMANCE EVALUATION OF A STEREO ECHO CANCELER WITH PRE-PROCESSING

Yann Joncour, Akihiko Sugiyama and Akihiro Hirano

C&C Media Research Laboratories, NEC Corporation
1-1, Miyazaki 4-Chome, Miyamae-ku, Kawasaki 216-8555, Japan
Email: ken-nec@dsp.cl.nec.co.jp

ABSTRACT

This paper presents implementation and performance evaluation of a stereo echo canceler with pre-processing. A two-tap time-varying filter located in one of the two channels periodically delays the input signal by one sample. By this pre-processing, the correct echo-path identification is achieved. This stereo echo canceler is implemented by four 32-bit floating-point digital signal processors (ADSP-21062). Experimental results show that the implemented echo canceler can reduce the echo by approximately 25 dB for a white Gaussian signal and by 23 dB on average for a speech signal. The ERLE is not degraded by talker movements in the remote room. The Mean Opinion Score for the implemented echo canceler is 0.55-point and 0.48-point higher than that for the echo canceler based on linear combination for round-trip delays of 100 ms and 600 ms, respectively.

1 INTRODUCTION

High-quality teleconferencing provides a good localization of participants by means of stereo systems. Teleconference systems should include a stereo echo canceler to remove undesired acoustic echoes. With the structure based on linear combination, the echo paths are not correctly identified for strongly cross-correlated input signals, like stereo speech signals [1]. Actually, the filter coefficients misconverge to a final value which depends on the acoustic environment in the remote room [1]. Thus, any acoustic change in the remote room, like a talker movement, seriously degrades the ERLE (Echo Return Loss Enhancement).

In practical situations, the input signals have low-level components which are not cross-correlated [2], and there are slight variations in the interchannel correlation [3]. The correct echo-path identification can be achieved providing that the adaptation algorithm makes use of these uncorrelated low-level components or small variations in the interchannel correlation. Therefore, a fast-convergence algorithm is required to obtain good performance. Such an algorithm requires much computation which makes its implementation difficult.

Another solution consists in a partial decorrelation of input signals by introducing a non-linearity in each channel [4]. However, the convergence of filter coefficients is still slow [5] unless a fast-convergence algorithm is used.

To obtain a satisfactory convergence speed with the normalized LMS algorithm, a stereo echo canceler with pre-processing of input signals has been proposed [5][6]. The normalized LMS algorithm requires less computation, making the implementation of the stereo echo canceler with pre-processing easier.

This paper presents DSP implementation and performance evaluation of a stereo echo canceler with pre-processing. Section 2 reviews the structure of the implemented stereo echo canceler. DSP implementation is described in Section 3. Experimental results and listening test results presented in the last section show the performance of the implemented stereo echo canceler.

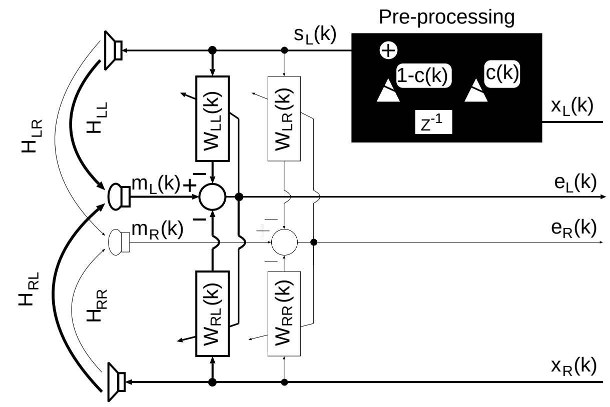

Fig. 1: Stereo echo canceler with pre-processing.

2 STEREO ECHO CANCELER WITH PRE-PROCESSING

Fig. 1 shows the stereo echo canceler with pre-processing [5][6]. It is assumed that the pre-processing unit is located in the left channel, as depicted in the shaded area in Fig. 1. It may be equipped with in the right channel instead. The pre-processing is performed by a two-tap filter whose time-varying coefficients are controlled by a periodic function $c(k)$ with a period Q. When $c(k) = 1$ for the first $Q/2$ iterations, the processed signal $s_L(k)$ and the input signal $x_L(k)$ are identical. When $c(k) = 0$ for the following $Q/2$ iterations, the input signal is simply delayed by one sample. These operations are repeated every Q iterations. In this way, the input signal in the left channel is periodically delayed by one sample.

By this pre-processing, the condition for which the echo is canceled, or the residual echo becomes zero, leads to two equations [5]. The first equation is obtained for $c(k) = 1$, and the second one for $c(k) = 0$. The common solution to these equations is unique and corresponds to the true echo-path impulse responses [7].

The processed signal which is heard by listeners should not be degraded. Audible aliasing by pre-processing is removed by increasing the period Q for $c(k)$ [6]. A noise perceived as "clicks" is caused by a sudden phase change for the processed signal at each alternation between the input signal and its one-sample delayed version. To avoid these audible "clicks", $c(k)$ varies smoothly over L iterations between zero and one, as shown in Fig. 2.

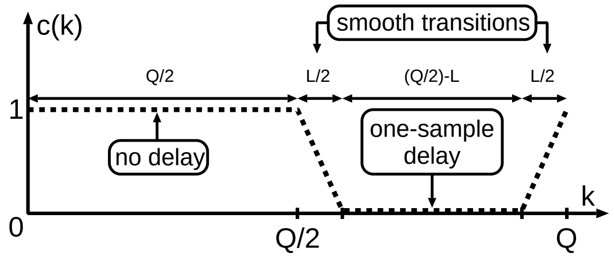

Fig. 2: c(k) with a period of Q.

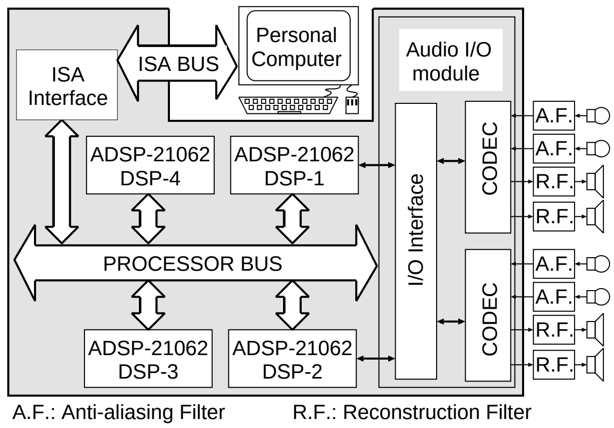

Fig. 3: Blockdiagram of DSP implementation.

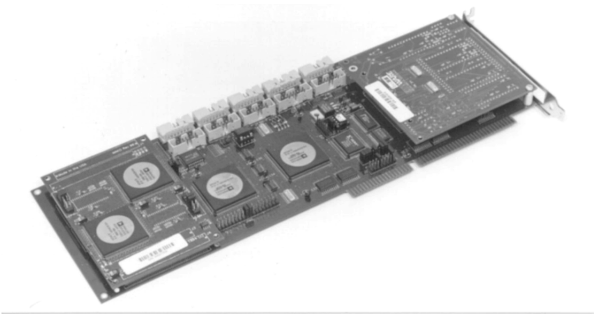
Fig. 4: Implemented stereo echo canceler.

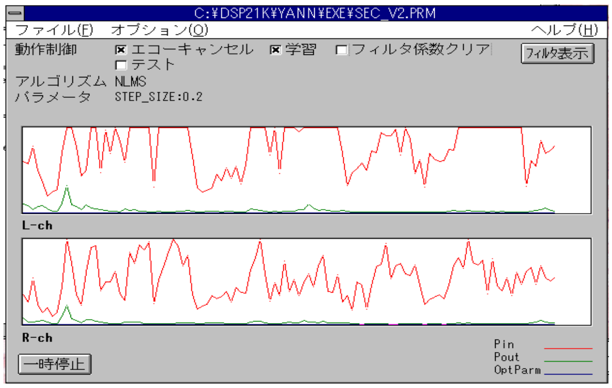

Fig. 5: Display of the power of input signals and echo powers.

3 DSP IMPLEMENTATION

3.1 Hardware configuration

A stereo echo canceler for teleconferencing requires much computation since four adaptive filters with a large number of coefficients are necessary to identify the four echo paths. To match the processing requirement, the stereo echo canceler with pre-processing has been implemented with four ADSP-21062s by Analog Devices [8]. ADSP-21062 is a 32-bit floating-point processor with a 2-megabit on-chip SRAM. Fig. 3 illustrates a blockdiagram of the DSP implementation. An audio input/output (I/O) module provides two stereo CODECs. Each CODEC consists of two digital-to-analog converters and two analog-to-digital converters with 16-bit precision. The sampling rate is 8 kHz and is software-programmed. Anti-aliasing filters are used to filter the microphone signals before the sampling operation. Reconstruction filters at the output of the digital-to-analog converters are necessary to recover the analog signals which are sent to the loudspeakers. An ISA interface allows communication between the four processors and a personal computer with an ISA bus.

3.2 Implementation of the stereo echo canceler

Each DSP is used as an adaptive filter. They perform 1750-tap FIR filtering and the coefficient adaptation with the LMS or the normalized LMS algorithm. DSP-1 and DSP-2 communicate with the audio I/O module to transmit or to receive audio data. In addition, as the main processor, DSP-1 synchronizes the tasks performed by other processors. It calculates the power of input signals with 64-bit precision and subtracts the filter outputs from the microphone signals to cancel the echo. DSP-1 also performs the pre-processing. The stereo echo canceler is implemented on a single board with a dimension of 235 mm $\times$ 120 mm, as shown in Fig. 4.

A personal computer downloads the program into the processors. A graphic interface has been developed to facilitate the control of the implemented stereo echo canceler and to display its performance in real-time. The graphic interface allows the selection of the adaptation algorithm and its step size. It also displays the power of input signals, the echo powers and the filter coefficients, as shown in Figs. 5 and 6.

4 PERFORMANCE EVALUATION

4.1 Experimental results

Performance evaluation was carried out for the echo canceler based on linear combination (LC), and the implemented echo canceler with pre-processing. Note that the same hardware configuration was used for both echo cancelers. The LC-based echo canceler was obtained by removing the program dedicated to the pre-processing. The echo was generated in a large room with a dimension of 7.5 m $\times$ 5 m $\times$ 3 m. The coefficient adaptation was performed by the normalized LMS algorithm with a step size μ of 0.5. For the new stereo echo canceler, $Q = 2000$ and $L = 200$ were selected to minimize the audio degradation of the pre-processed signal [6]. The power spectrum of the echo was computed by an FFT analyzer. To show the convergence characteristics of the ERLE, the echo before and after cancellation were recorded in a DAT (Digital Audio Tape).

Fig. 7 exhibits the power spectrum of the residual echo by the LC-based echo canceler with a white Gaussian signal as the input. The echo power is reduced by 25 dB on average. However, when a talker movement occurs in the remote room, the power of the residual echo becomes almost as large as the power of the echo without cancellation. For the implemented echo canceler, the echo power is decreased by 25 dB, as shown in Fig. 8. When there is a talker movement, the power of the residual echo remains low compared to the echo power without echo cancellation. These results confirm that the implemented echo canceler is robust against acoustic changes in the remote room.

Fig. 9 compares the convergence characteristics of the ERLE for a white Gaussian signal. A talker movement oc-

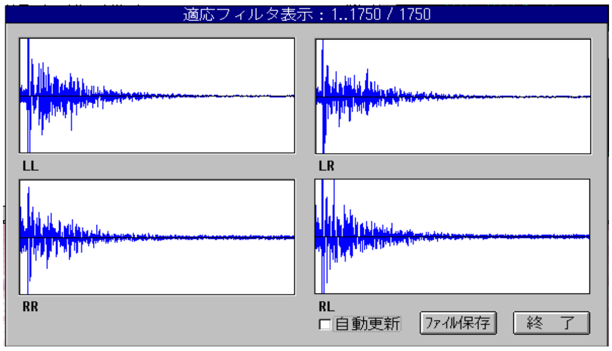

Fig. 6: Display of filter coefficients.

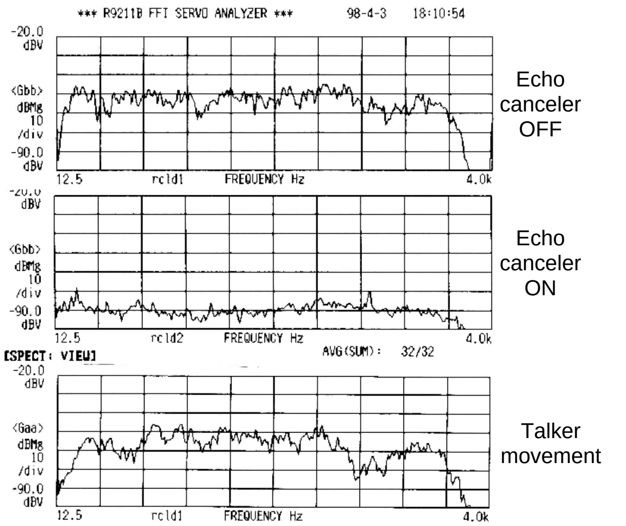

Fig. 7: Power spectrum of the echo for the LC-based echo canceler and a white Gaussian signal.

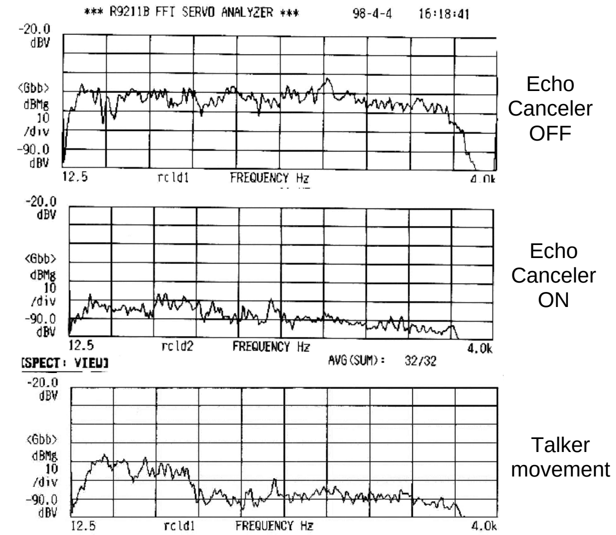

Fig. 8: Power spectrum of the echo for the implemented echo canceler and a white Gaussian signal.

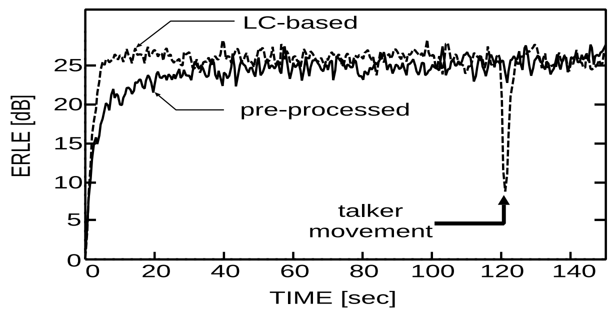

Fig. 9: Measured ERLE for a white Gaussian signal.

curs after 120 seconds (2 minutes). The echo to noise ratio is 25.5 dB. Although the convergence of the ERLE is slightly faster for the LC-based echo canceler, the ERLE is degraded when there is a talker movement. For the implemented echo canceler, the ERLE reaches 25 dB and is not degraded by a talker movement.

Fig. 10 exhibits the convergence characteristic of the ERLE for the implemented echo canceler with a speech signal and a talker movement after 300 seconds (5 minutes). The ERLE reaches 23 dB on average and is not degraded by the talker movement. The implemented echo canceler successfully cancels the echo for a speech signal.

4.2 Listening test results

A listening test has been carried out to compare the subjective performance of the implemented echo canceler with the conventional one based on linear combination. In a teleconference system, the original speech is reflected by walls and objects in the local room. Afterwards, the speech is sent back to the remote room as an echo with some delay due to the transmission lines between the two teleconference rooms. For the listening test, the transmission line was simulated by a flat-delay generator. The echo was delayed by 100 ms and 600 ms which are typical round-trip delays for transmissions over terrestrial circuits and satellite circuits, respectively [9]. The configuration of the listening room is depicted in Fig. 11. The original speech came from the loudspeaker in front of the listener and the echo came from the two loudspeakers which were facing the listener. In this way, the original speech and the echo were mixed in the acoustic environment as if it were a real teleconference system.

The "Absolute Category Rating" procedure [10] with a five-grade scale was used for the listening test. The minimum grade was 1.0 and corresponded to the echo when the echo canceler was turned off. The maximum grade was 5.0 and corresponded to the echo-free situation. The Mean Opinion Score (MOS) was calculated from grades given by a panel of twenty listeners. The test participants were asked to rate the quality of a female speech and a male speech after one, two and five minutes of adaptation. A speech sequence consisted of four sentences with a talker movement in the remote room after the first two sentences.

The MOS is represented with a horizontal bar and is written on the right side of the bar. The horizontal line centered around a mean value is the 90% confidence interval using a normal distribution. There is a significant difference in the MOS when the confidence intervals do not overlap. Fig. 12 shows results obtained for both round-trip delays.

With a round-trip delay of 100 ms, the MOS for the LC-based echo canceler was not significantly improved as the adaptation time became longer. The MOS ranged between 3.14 and 3.31. The filter coefficients had converged to a final value which does not correspond to the correct echo-path impulse response. As a consequence, the echo level was suddenly increased by the talker movement in the remote room. After one minute of adaptation, the MOS for the implemented echo canceler was equivalent to the one obtained for the LC-based echo canceler. The filter coefficients had not reached their convergence so that the talker movement could be perceived. As the adaptation continued, the MOS was improved since the filter coefficients were converging to the correct echo-path impulse response. After 2 minutes of

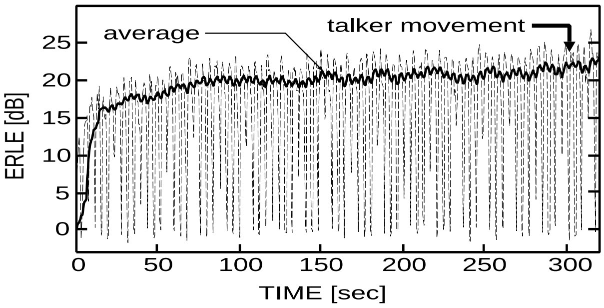

Fig. 10: Measured ERLE for a speech signal.

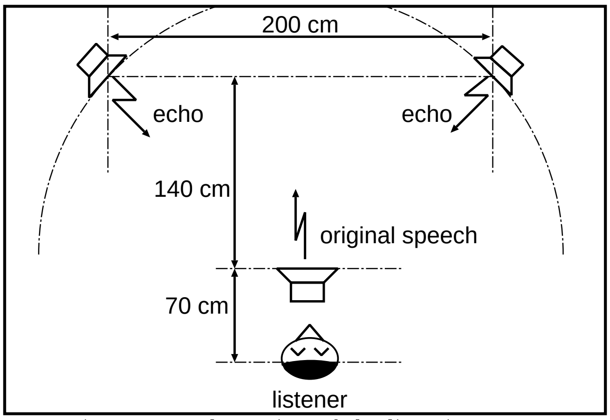

Fig. 11: Configuration of the listening room.

adaptation the MOS difference became favorable to the implemented echo canceler. After 5 minutes, the subjective quality of the implemented echo canceler surpassed that of the LC-based echo canceler. The MOS was 3.86 which was 0.55-point higher than for the conventional echo canceler.

With a round-trip delay of 600 ms, the implemented echo canceler after convergence of its filter coefficients obtained a higher MOS compared to the LC-based echo canceler. The MOS difference was 0.48. Comparing MOSs obtained for the two round-trip delays, it can been seen that the echo was more annoying for a round-trip delay of 600 ms. This result confirms that the longer the echo is delayed, the more it must be reduced. After convergence of filter coefficients, the reduction of the echo power is restricted by the echo to noise ratio. For the listening test, the measured echo to noise ratio was 24 dB. Subjective evaluation of echoes in telephone communications has shown that the echo must be reduced by at least 37 dB to be little sensible to the round-trip delay [11]. Therefore, it may be necessary to combine the stereo echo canceler with a system of noise reduction to achieve good performance for long transmission delays.

5 CONCLUSION

DSP implementation and performance evaluation of a stereo echo canceler with pre-processing have been presented. The implemented echo canceler has reduced echoes by approximately 25 dB for a white signal, and 23 dB for a speech signal. Convergence characteristics of the ERLEs for both a white Gaussian signal and a speech signal have shown that the performance of the implemented echo canceler is not degraded by acoustic changes in the remote room. The subjective superiority of the implemented echo canceler compared to the LC-based one has been shown by the MOS that was 0.55-point and 0.48-point higher for round trip-delays of 100 ms and 600 ms, respectively.

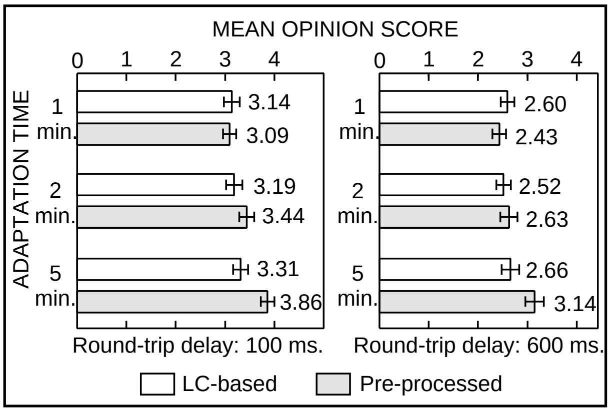

Fig. 12: Listening test results for both round-trip delays.

Acknowledgments

The authors would like to thank fellows of the Digital Signal Processing Technology Group, C&C Media Research Laboratories, NEC Corporation, for participating in the listening test. They are also indebted to Dr. Kazunori Ozawa, a Senior Manager of C&C Media Research Laboratories, NEC Corporation, for his guidance and encouragement.

References

[1] A. Hirano and A. Sugiyama, "Convergence Characteristics of a Multichannel Echo Canceller with Strongly Cross-Correlated Input Signals - Analytical Results-," Proc. of 6th DSP Symposium of IEICEJ, pp. 144-149, Nov. 1991.

[2] J. Benesty, F. Amand, A. Gilloire and Y. Grenier, "Adaptive Filtering Algorithms for Stereophonic Echo Cancellation," Proc. of ICASSP'95, pp. 3027-3030, May 1995.

[3] S. Shimauchi and S. Makino, "Stereo Projection Echo Canceller with True Echo Path Estimation," Proc. of ICASSP'95, pp. 3059-3062, May 1995.

[4] J. Benesty, D.R. Morgan and M.M. Sondhi, "A Better Understanding and an Improved Solution to the Problems of Stereophonic Acoustic Echo Cancellation," Proc. of ICASSP'97, pp. 303-306, Apr. 1997.

[5] Y. Joncour and A. Sugiyama, "A Stereo Echo Canceler with Pre-Processing for Correct Echo-Path Identification," to appear in Proc. of ICASSP'98.

[6] Y. Joncour and A. Sugiyama, "A Stereo Echo Canceler with Correct Echo-Path Identification," Technical Report of IEICE, DSP97-1, IE97-1, pp. 1-8, Apr. 1997.

[7] Y. Joncour and A. Sugiyama, "A Unique and Strict Identification of the Echo Path Impulse Response in Stereo Echo Cancellation," Technical Report of IEICE, DSP96-100, pp. 17-24, Dec. 1996.

[8] "ADSP-2106x SHARC User's Manual," Analog Devices, Mar. 1995.

[9] M.M. Sondhi and D.A. Berkley, "Silencing Echoes on the Telephone Network," Proc. of IEEE, vol. 68, no. 8, Aug. 1980.

[10] ITU-T Recommendation P.80, "Method for Subjective Determination of Transmission Quality", Mar. 93.

[11] G. Williams and L.S. Moye, "Subjective Evaluation of Unsuppressed Echo in Simulated Long-Delay Telephone Communications," Proc. of IEE, vol. 118, pp. 401-408, Mar./Apr. 1971.