

A STEREO ECHO CANCELLER IMPLEMENTED USING A STEREO SHAKER AND A DUO-FILTER CONTROL SYSTEM

Suehiro Shimauchi Shoji Makino Yoichi Haneda Akira Nakagawa Sumitaka Sakauchi

NTT Human Interface Laboratories
3-9-11, Midori-cho, Musashino-shi, Tokyo, 180-8585 Japan
Tel: +81 422 59 3028, Fax: +81 422 60 7811
Email: sima@splab.hil.ntt.co.jp

ABSTRACT

Stereo echo cancellation has been achieved and used in daily teleconferencing. To overcome the non-uniqueness problem, a stereo shaker is introduced in eight frequency bands and adjusted so as to be inaudible and not affect stereo perception. A duo-filter control system including a continually running adaptive filter and a fixed filter is used for double-talk control. A second-order stereo projection algorithm is used in the adaptive filter. A stereo voice switch is also included. This stereo echo canceller was tested in two-way conversation in a conference room, and the strength of the stereo shaker was subjectively adjusted. A misalignment of 20 dB was obtained in the teleconferencing environment, and changing the talker's position in the transmission room did not affect the cancellation. This echo canceller is now used daily in a high-presence teleconferencing system and has been demonstrated to more than 300 attendees.

1. INTRODUCTION

A stereo teleconferencing system provides a more realistic presence in teleconferencing compared to monaural systems. It helps listeners distinguish who is talking at the other end by means of spatial information. The most significant problem with stereo echo cancellation using the conventional linear combiner structure is that the adaptive filter often misconverges or, if not, its convergence speed is very slow because of the cross-correlation between the stereo signals [1]. As a result, the conventional stereo echo canceller suffers from variation in both the near-end echo path and the far-end transmission path. There is a strong need for a stereo echo canceller that can identify the true echo path impulse response quickly with low computational complexity.

Several methods for overcoming this problem have been proposed. Some utilize the uncorrelated elements of stereo signals [2]-[5] and some the variations in the cross-correlation between stereo signals [6]-[9].

We have developed a stereo echo canceller that uses a stereo shaker and stereo projection algorithm. This echo canceller has an adaptive filter and a fixed filter (duo-filter structure), and a stereo voice switch.

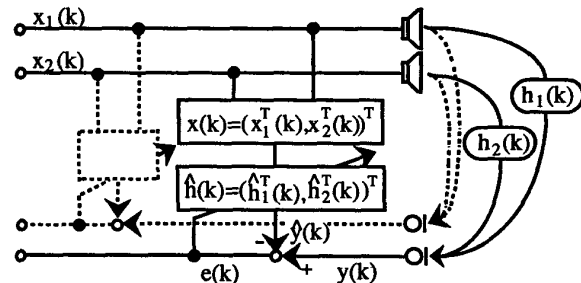


Figure 1: Configuration of stereo echo canceller.

2. STEREO ECHO CANCELLATION

2.1. Fundamental problem of stereo echo cancellation

A stereo (two-channel) telecommunication system is shown in Fig. 1. Input signals $x_1(k)$ and $x_2(k)$ and filter coefficients $\hat{h}_1(k)$ and $\hat{h}_2(k)$ are combined as $\mathbf{x}(k) = [x_1^T(k), x_2^T(k)]^T$ and $\hat{\mathbf{h}}(k) = [\hat{h}_1^T(k), \hat{h}_2^T(k)]^T$. Thus, stereo echo cancellation is achieved by linearly combining two monaural echo cancellers.

Unlike monaural echo cancellation, stereo echo cancellation has the specific problem of non-uniqueness. This problem is that the adaptive filter often misconverges or, even when it converges, its convergence speed is very slow because of cross-correlation between the stereo signals [1]. As a result, the conventional stereo echo canceller suffers from variation in both the near-end echo path and the far-end transmission path. Consequently, talker movement or changes in the transmission room are considered to be variations in the echo path in the receiving room. Accordingly, the performance of the stereo echo canceller degrades at the instant of abrupt changes in the environment in the transmission room.

2.2. Recent solutions

A clue to solving the non-uniqueness problem can be found in practical teleconferencing situations.

- (1) The stereo signals $x_1(k)$ and $x_2(k)$ contain independent noise.
- (2) The length of the adaptive filter $\hat{h}_1(k)$ and $\hat{h}_2(k)$ is shorter than that of the impulse response in the transmission room. These truncated components act as in-

dependent noise.

(3) The cross-correlation between the stereo signals $\mathbf{x}_1(k)$ and $\mathbf{x}_2(k)$ varies slightly even when the talker does not move his body or head while speaking.

Several methods for overcoming the non-uniqueness problem have been proposed.

Regarding items (1) and (2), several functions (stereo shakers) were successfully applied to create an independent component in the stereo input signals [2]-[5]. The important point is that the noise generated should not be audible and should not degrade stereo perception.

Regarding item (3), one might think that the change in the variation causes another misconvergence and hence would not suppress the non-uniqueness problem. Fortunately, however, a "new" convergence process starts from the "old" misconverged solution. Thus, after many variations in the cross-correlation, the stereo echo canceller can converge to the "true" solution [6].

To emphasize the slightly varying cross-correlation between stereo signals in actual teleconferencing situations, a stereo projection algorithm [6], [10] and sub-band processing [11] were successfully introduced, and several functions (stereo shakers) were successfully applied to create an effective variation in the stereo input signals [6]-[9]. Thus convergence speed to the true echo path impulse response was significantly improved.

3. NEW STEREO ECHO CANCELLER

3.1. Stereo shaker for true echo path estimation

The configuration of the developed stereo echo canceller is shown in Fig. 2. The frequency characteristics of the human auditory system are not flat, and speech signals mask the distortion. Therefore, if we adjust these stereo shakers in the frequency band, we can make the processed distortion less audible and make the stereo echo canceller converge faster. With this aim, we divided an 8-kHz frequency band into eight frequency bands and adjusted the strength of the stereo shaker in each band. Also, we used different kinds of stereo shakers in different frequency bands as suggested in [11]. The shakers used have the following formula and are as listed in Table 2.

$$\mathbf{x}'(k) = \mathbf{x}(k) + \alpha|\mathbf{x}(k)| \quad (\text{Non-linear}) \quad (1)$$

$$\mathbf{x}'(k) = \mathbf{x}(k) + \alpha\mathbf{n}(k)\mathbf{x}(k) \quad (\text{Additive noise}) \quad (2)$$

3.2. Stereo projection algorithm

The stereo projection algorithm has been shown to be effective in stereo echo cancellation [6], [10]. This algorithm emphasizes the variation in cross-correlation in the stereo signals and also de-autocorrelates the two input signals, to improve the convergence speed to the true echo path impulse response.

The p -th order projection algorithm, or affine projection algorithm, updates filter coefficient vector $\hat{\mathbf{h}}(k)$ as follows [12].

$$\begin{aligned} \hat{\mathbf{h}}(k+1) &= \hat{\mathbf{h}}(k) + \mu\mathbf{X}(k)[\mathbf{X}(k)^T\mathbf{X}(k) + \delta\mathbf{I}]^{-1}\mathbf{e}(k) \\ &= \hat{\mathbf{h}}(k) + \mu[\beta_1(k)\mathbf{x}(k) + \beta_2(k)\mathbf{x}(k-1) \\ &\quad + \dots + \beta_p(k)\mathbf{x}(k-p+1)] \end{aligned} \quad (3)$$

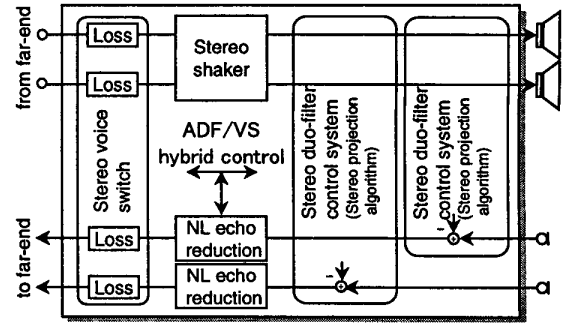


Figure 2: Developed stereo echo canceller.

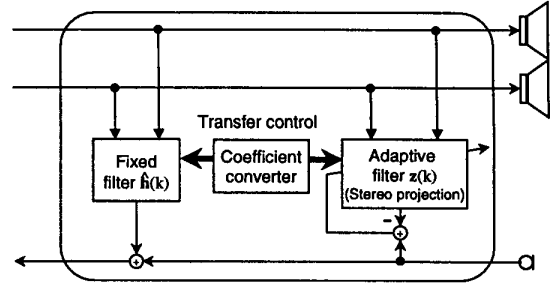


Figure 3: Stereo Duo-filter structure.

$$[\beta_1(k), \beta_2(k), \dots, \beta_p(k)]^T = [\mathbf{X}(k)^T\mathbf{X}(k) + \delta\mathbf{I}]^{-1}\mathbf{e}(k) \quad (4)$$

$$\mathbf{X}(k) = [\mathbf{x}(k), \mathbf{x}(k-1), \dots, \mathbf{x}(k-p+1)] \quad (5)$$

$$\mathbf{e}(k) = \mathbf{y}(k) - \mathbf{X}(k)^T\hat{\mathbf{h}}(k) + \mathbf{n}(k) \quad (6)$$

$$= [\mathbf{e}(k), (1-\mu)\mathbf{e}(k-1), \dots, (1-\mu)^{p-1}\mathbf{e}(k-p+1)]^T \quad (7)$$

$$\mathbf{x}(k) = [x(k), x(k-1), \dots, x(k-L+1)]^T \quad (8)$$

$$\mathbf{y}(k) = [y(k), y(k-1), \dots, y(k-p+1)]^T \quad (9)$$

$$\mathbf{n}(k) = [n(k), n(k-1), \dots, n(k-p+1)]^T, \quad (9)$$

where

μ : scalar stepsize ($0 < \mu < 2$),

δ : small positive constant.

By introducing an intermediate variable $\mathbf{z}(k)$, the computational complexity for the second order ($p=2$) is $4L + 40$ multiply-add operations per microphone.

3.3. Duo-filter control system

To determine whether to freeze or update the coefficients, an echo canceller has to distinguish between double-talk and a change in echo path. If the echo canceller fails to realize effective double-talk control, the adaptive filter will be degraded, resulting in annoying echo and howling. The echo canceller must detect the double-talk before the adaptive filter becomes degraded. To handle this requirement, we use a structure composed of an adaptive filter and a fixed filter. The adaptive filter estimates the room impulse response, and the fixed filter reduces the echo. The coefficients for the adaptive filter are transferred to the fixed filter

when the adaptive filter reduces the echo more than the fixed filter. The convergence of the adaptive filter is judged by monitoring the far-end speech signal, the microphone input signal, and the errors of the adaptive filter and the fixed filter. During double-talk, the coefficients of the adaptive filter may become degraded, increasing the error in the adaptive filter. If this occurs, the filter coefficients of the adaptive filter are not transferred to the fixed filter. The filter coefficients of the fixed filter are, therefore, not updated during double-talk, and the echo canceling level before double-talk is maintained. We use the projection algorithm for the adaptive filter; the adaptive filter has the intermediate variable $z(k)$, and the fixed filter uses the impulse response replica $\hat{h}(k)$. Hence we call this system the "Duo-filter" system [13]. A block diagram of this system is shown in Fig. 3. The intermediate variable $z(k)$ is converted to the impulse response replica $\hat{h}(k)$ before the filter coefficients are transferred.

3.4. Variable voice switch

While adaptive filters can achieve full-duplex communication, they are limited by their convergence speed and the echo canceling level. Therefore, other echo-suppression methods, such as a variable voice switch, are used to enhance echo canceling. When the acoustic coupling level is above 0 dB, the echo canceller may begin howling immediately after it is turned on. It may also begin howling when the echo path changes rapidly. To prevent howling, variable losses are inserted into the system, as shown in Fig. 2. If the loss levels are too large, they may produce subjective dissatisfaction, so they are kept as small as the howling margin level. The howling margin level is determined from the multichannel acoustic coupling level, which should be measured only when far-end speech exists. The convergence of the continually working adaptive filter ensures single-talk. Our echo canceller determines the loss level only when far-end speech is active and the adaptive filter has converged. Once the duo-filter convergence level is brought below the howling margin level, the variable loss levels are set to 0 dB.

3.5. DSP implementation

We implemented our system with a frequency range from 100 Hz to 20 kHz on DSPs. The number of taps in the filter and the filter length are listed in Table 1. In the range from 8 kHz to 20 kHz, the stereo voice switch alone was used.

4. REAL-TIME EXPERIMENTS

Real-time experiments were performed with the hardware in two teleconferencing rooms. Each room had a volume of 150 m³ and a reverberation time of 300 ms. Ambient noise was about 25 dB in SNR. The stepsize μ of the projection algorithm was set to be 0.5. The loudspeakers, microphones, and talkers were placed as shown in Fig. 4.

Table 1 Specifications of the developed stereo echo canceller.

Frequency band (kHz)	0.1 - 4	4 - 8	8 - 20
Number of taps	1200	800	0 (voice switch)
Filter length (ms)	150	100	0 (voice switch)

Table 2 Inaudible stereo shaker level.

Frequency band (kHz)	0.1 1	1 2	2 3	3 4	4 5	5 6	6 7	7 8
Stereo shaker	NL	NL	NL	NL	N	N	N	N
Strength α	0.2	0.4	0.4	0.4	0.5	0.5	0.5	0.5

NL: Non-linear N: Additive Noise

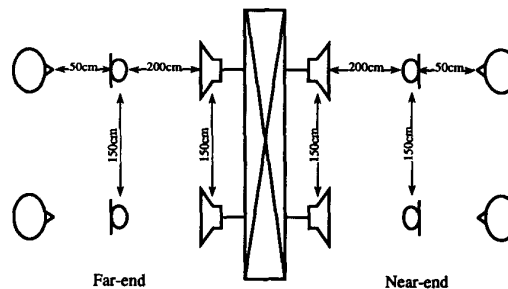


Figure 4: Positions of loudspeakers, microphones, and talkers.

4.1. Listening test in two-way conversation

We tested the echo canceller in two-way conversation. Four speakers, who remained in their places, spoke alternately. Speech quality was subjectively evaluated and the stereo shaker level was adjusted so as to be inaudible and not affect stereo perception. Table 2 lists the stereo shaker level in each frequency band. The lower the frequency band, the smaller the shaker level should be.

4.2. Convergence at change of far-end talker position

Figure 5 shows the error level of (a) right channel S_{out} , (b) left channel S_{out} . Only the echo canceller was active. The far-end stereo signals were white Gaussian noise with fixed cross-correlation. After far-end right

“talker” talked 90 s, the “talker” changed positions in the transmission room from right to left. Without the stereo shaker, the error was degraded 20 dB when the talker changed positions, since the echo path was not correctly identified. With the stereo shaker, the error was not degraded by the change in talker positions. This error difference of 20 dB indicates that the misalignment of our echo canceller was about 20 dB.

5. DAILY USE IN HIGH-PRESENCE TELECONFERENCING SYSTEM

Our stereo echo canceller has been used daily in our high-presence teleconferencing system. Our echo canceller gained a sound level of 10 dB compared to a conventional system. Howling and echo were eliminated and speech quality was improved. Combined with a 2.4-m(W) \times 1.3-m(H) screen, the sound localization brought high presence to the teleconferencing. We have demonstrated the echo canceller to more than 300 guests.

6. CONCLUSIONS

A stereo echo canceller has been developed and used daily in teleconferencing. To achieve true echo path estimation, a stereo shaker was introduced in eight frequency bands and adjusted subjectively so as to be inaudible and not affect stereo localization in two-way conversations in teleconferencing rooms. A misalignment of 20 dB was obtained and no echo was fed back even when the far-end talker changed positions. A duo-filter control system with a stereo projection adaptive filter and fixed filter, and a stereo voice switch were also implemented. This echo canceller is now used daily in our teleconferencing and more than 300 guests have used it.

7. ACKNOWLEDGEMENTS

We would like to thank Dr. Yutaka Kaneda, Mr. Masashi Tanaka, and Mr. Kuniyasu Suzuki for many fruitful discussions.

8. REFERENCES

- [1] M. Sondhi, D. Morgan, and J. Hall, “Stereo acoustic echo cancellation - an overview of the fundamental problem,” *IEEE Signal Processing Letters*, vol. 2, no. 8, pp. 148-151, Aug. 1995.
- [2] J. Benesty, D. Morgan, and M. Sondhi, “A better understanding and an improved solution to the problems of stereophonic acoustic echo cancellation,” *Proc. ICASSP97*, pp. 303-306, Apr. 1997.
- [3] S. Shimauchi, Y. Haneda, S. Makino, and Y. Kaneda, “New configuration for a stereo echo canceller with nonlinear pre-processing,” *Proc. ICASSP98*, pp. 3685-3688, May 1998.
- [4] T. Gansler and P. Eneroth, “Influence of audio coding on stereophonic acoustic echo cancellation,” *Proc. ICASSP98*, pp. 3649-3652, May 1998.
- [5] A. Gilloire and V. Turbin, “Using auditory properties to improve the behaviour of stereophonic acoustic echo cancellers,” *Proc. ICASSP98*, pp. 3681-3684, May 1998.
- [6] S. Shimauchi and S. Makino, “Stereo projection echo canceller with true echo path estimation,” *Proc. ICASSP95*, pp. 3059-3062, May 1995.
- [7] S. Shimauchi and S. Makino, “Stereo echo cancellation algorithm using imaginary input-output relationships,” *Proc. ICASSP96*, pp. 941-944, May 1996.
- [8] M. Ali, “Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation,” *Proc. ICASSP98*, pp. 3689-3692, May 1998.
- [9] Y. Joncour and A. Sugiyama, “A Stereo echo canceller with pre-processing for correct echo-path identification,” *Proc. ICASSP98*, pp. 3677-3680, May 1998.
- [10] J. Benesty, P. Duhamel, and Y. Grenier, “A multi-channel affine projection algorithm with applications to multichannel acoustic echo cancellation,” *IEEE Signal Processing Letters*, vol. 3, no. 2, pp. 35-37, Feb. 1996.
- [11] S. Makino, K. Strauss, S. Shimauchi, Y. Haneda, and A. Nakagawa, “Subband stereo echo canceller using the projection algorithm with fast convergence to the true echo path,” *Proc. ICASSP97*, pp. 299-302, Apr. 1997.
- [12] S. Makino and Y. Kaneda, “Exponentially weighted stepsize projection algorithm for acoustic echo cancellers,” *Trans. IEICE Japan*, vol. E75-A, no. 11, pp. 1500-1508, Nov. 1992.
- [13] Y. Haneda, S. Makino, J. Kojima, and S. Shimauchi, “Implementation and evaluation of an acoustic echo canceller using duo-filter control system,” *Proc. EU-SIPCO96*, pp. 1115-1118, Sept. 1996.

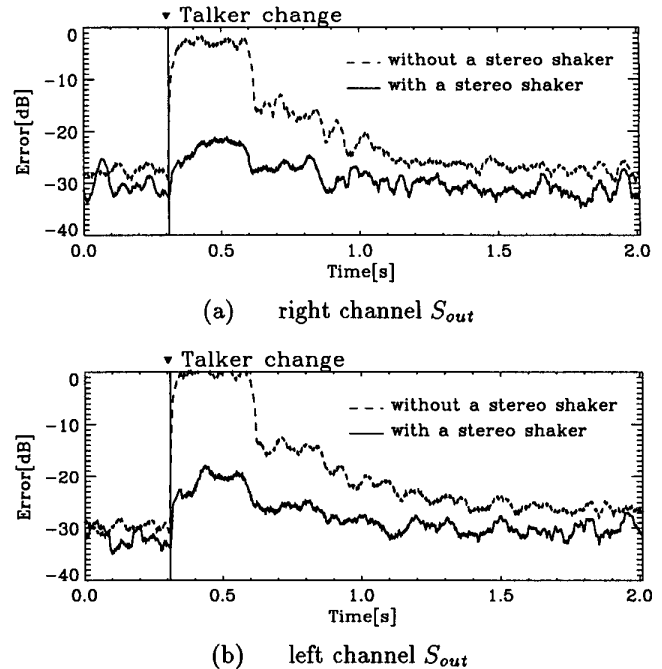


Figure 5: Error level of (a) right channel S_{out} , (b) left channel S_{out} for white Gaussian input with fixed cross-correlation. After 90 s, the far-end “talker” changed positions in the transmission room from right to left.