CHANNEL-NUMBER-COMPRESSED MULTI-CHANNEL ACOUSTIC ECHO CANCELLER FOR HIGH-PRESENCE TELECONFERENCING SYSTEM WITH LARGE DISPLAY

Akira Nakagawa, Suehiro Shimauchi,
Yoichi Haneda, Shigeaki Aoki, and Shoji Makino
NTT Cyber Space Laboratories
3-9-11, Midori-cho, Musashino-shi, Tokyo, 180 Japan
Tel: +81 422 59 3546, Fax: +81 422 60 7811, Email: nakagawa.akira@lab.ntt.co.jp

ABSTRACT

Sound localization is important to make conversation easy between local and remote sites in a teleconference. This requires a multi-channel sound system having a multi-channel acoustic echo canceller (MAEC). The appropriate number of channels is determined from a trade-off between high presence and MAEC performance, so it is not possible to increase the channel number by much.

We propose a channel-number-compressed MAEC to provide teleconferencing systems that exhibit high presence. The channel number of the MAEC inputs is compressed and that of its outputs is expanded.

1. INTRODUCTION

Large-scale teleconferences are now possible because of the lower cost of improved network environments and display devices. Teleconferencing systems using a large display may be able to provide high presence. The key to making communication realistic is sound localization. This helps participants distinguish who is talking at the remote site and allows them to face and listen to the talker easily. A high-presence teleconferencing system requires at least two-channel sound.

Acoustic problems such as echoes and howling occur in a hands-free environment because microphones pick up echo signals that come out from loudspeakers. Simple solutions to avoid these acoustic problems are to set the loudspeaker volume low or to implement a half-duplex system. However, these solutions make it difficult for the participants to continue the conference smoothly. Acoustic echo cancellers (AECs) are therefore widely used to overcome these acoustic problems. AECs have an adaptive digital filter (ADF) to reduce the acoustic feedback and achieve a full-duplex system. A multi-channel acoustic echo canceller (MAEC) can be viewed as a simple generalization of the single-channel AEC principle. However, the multi-channel case gives rise to a non-uniqueness problem that does not arise in the single-channel case [1]. Several methods for overcoming this problem have been proposed. Some utilize the uncorrelated elements of stereo signals [2] while others utilize variations in the cross-correlation between stereo signals [3].

However, the trade-off between performance and presence presents another problem for the MAEC. When the number of channels is increased to realize higher presence, the computational load on the MAEC increases and its performance degrades. Some way of reconciling this trade-off is necessary.

Our solution is to use a channel-number-compressed MAEC which includes a channel number compressor and expander to achieve a low computational load and high presence. Such a system provides a high-presence teleconferencing system with large displays.

2. HIGH-PRESENCE TELECONFERENCING SYSTEM

2.1. Features

We are developing a high-presence teleconferencing system with large displays as shown in Fig. 1. The figure shows that three separate sites are connected. The left and right displays present the other two sites. All participants at the three separate sites can feel as if they are sitting around a large table holding a conference in the same room. The features of this high-presence teleconferencing system are summarized as fol-

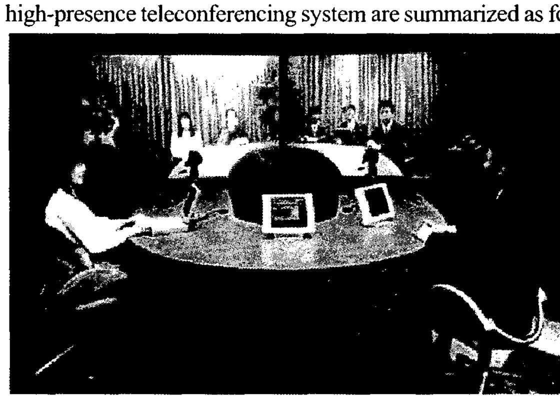

Fig. 1 High-presence conference using large displays.

0-7803-6293-4/00/$10.00 ©2000 IEEE.

lows.

(i) Super-high-resolution large displays each having a diagonal length of 110 inches and a resolution (2880 x 2048) four times that of high-definition television (HDTV).

(ii) HDTV coding and transmission for high-definition.

(iii) A highly effective collaborative tasking environment.

This system uses three loudspeakers for each display because more than two-channel sound is necessary to localize the sound source even for the participants placed off-center [4].

2.2. Problems with the high-presence teleconferencing system

Figure 2 shows the acoustic structure of our teleconferencing system. Each site has three loudspeakers and microphones for each remote site, so there are six loudspeakers and microphones per site. This produces thirty-six combinations of loudspeakers and microphones, so there are thirty-six echo paths at one site. The computational load of the MAEC increases in proportion to the square of the channel number. So when the six-channel MAEC is introduced, the computational load is thirty-six times that of a single-channel AEC. The convergence speed degrades in proportion to the channel number, so the echo cancelling performance also degrades. The computational load and performance problems can be avoided by decreasing the channel number for the MAEC. However, it is also degrades the presence, so we need a way to increase the presence with a small channel number.

3. NEW MULTI-CHANNEL ACOUSTIC ECHO CANCELLER

We propose a channel-number-compressed MAEC, which has a channel number compressor and expander. This improves the presence of teleconferencing systems by compressing the channel number for microphones and expanding the channel number for loudspeakers.

3.1. Channel number compressor and expander

3.1.1. Outline of compressor and expander

Here, we discuss one connection between two sites, each having three loudspeakers and microphones. Figure 3 shows a diagram of the channel number compressor and expander of the proposed system, where $a_1(k)$, $a_2(k)$, and $a_3(k)$ represent the microphone outputs. The compressor and expander should have linear and time-invariant operations because the MAEC estimates the channel number compressor and expander as part of the echo paths.

The compressor has gain factor f, which can control the mixing ratio. The compressed sounds, $b_1(k)$ and $b_2(k)$, are obtained by following equations.

$$b_1(k) = a_1(k) + f a_2(k) \quad (1)$$

$$b_2(k) = a_3(k) + f a_2(k), \quad (2)$$

where k : time index

The expander has filters $\mathbf{g}_1$ and $\mathbf{g}_2$, which can control the frequency characteristics, time delay, and gain level. The expanded sounds, $c_1(k)$, $c_2(k)$, and $c_3(k)$, are obtained by following equations.

$$c_1(k) = b_1(k) * \mathbf{g}_1 \quad (3)$$

$$c_2(k) = \{b_1(k) + b_2(k)\} * \mathbf{g}_2 \quad (4)$$

$$c_3(k) = b_2(k) * \mathbf{g}_1, \quad (5)$$

where * denotes convolution

3.1.2. Design of compressor and expander

To design the gain factor f, filter $\mathbf{g}_1$, and filter $\mathbf{g}_2$, we conducted tests to evaluate how well the sound was localized from the loudspeaker by using the mean opinion score (MOS).

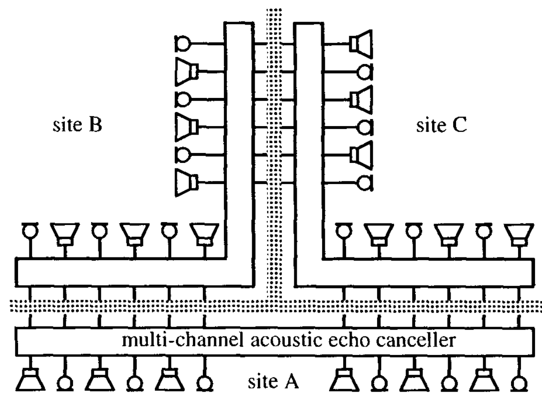

Fig. 2 Acoustic structure of the high-presence teleconferencing system.

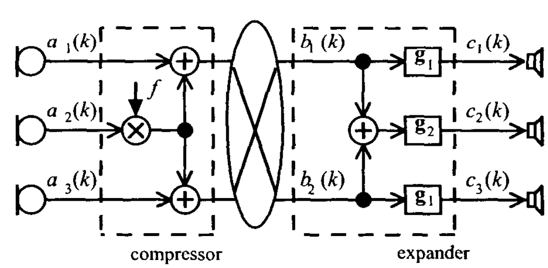

Fig. 3 Block diagram of a channel number compressor and expander.

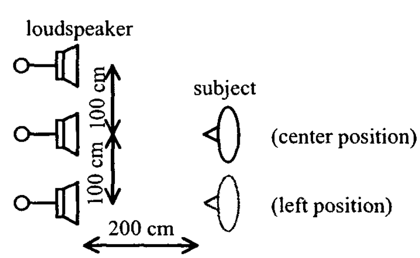

Fig. 4 Configuration of the trial system.

Figure 4 shows the locations of the loudspeakers and a subject with respect to each other. First, each subject was placed in front of the center loudspeaker and listened to loudspeaker outputs. Next each subject was placed in front of the left loudspeaker and listened to loudspeaker outputs. Test signals were Japanese female and male speech, and these test signals were input to sound channels $a_1(k)$, $a_2(k)$, and $a_3(k)$ in Fig. 3 in that order in three-second intervals. The subjects evaluated the level of sound localization using the following scale: 4 (Excellent), 3 (good), 2 (fair), 1 (poor), 0 (bad). The number of subjects was twelve.

• Gain level

First, we investigated the gain level of gain factor f, filter $\mathbf{g}_1$, and filter $\mathbf{g}_2$. Several combinations of gain level were evaluated. The combinations of gain level listed in Table 1 gave the best results for both positions (set 1), for the center position (set 2), and for the left position (set 3) (Fig. 5(a)). Set 2 achieved the highest average, so we decided to use its combination of gain level.

Table 1: Sets of the gain level.

	f	$\mathbf{g}_1$	$\mathbf{g}_2$
set 1	-6.0	-1.6	-4.9
set 2	-4.2	-1.6	-5.2
set 3	-4.6	-2.5	-3.6

[dB]

• Frequency characteristics

Even though the optimum gain level was determined, the MOS evaluation results were not good enough. Some subjects commented that the sound source was not clear. So we investigated the possibility of improving the sound localization by changing the frequency characteristics of the center loudspeaker output. We evaluated the frequency band of the center loudspeaker output, which was fullband, below 1 kHz, and below 2 kHz. Figure 5(b) shows the MOS results. The band-limited sound received a higher average than the fullband sound. So we decided to use the gain level combination of set 2 listed in Table 1 and band-limited sound for the center loudspeaker output.

• Time difference

Although the optimum gain level and frequency characteristics were determined, the MOS results for the left-side participant were not good enough, so the sound localization for the left-side participant must be improved. Even if only $a_2(k)$ has a signal in Fig. 3 (only the center microphone picks up the signal), signals come out from all loudspeakers. When all loudspeakers emit sound, the left-side participant localizes the sound source to the left side because he / she hears the sound coming out of the left loudspeaker first. This is caused by the precedence effect. To avoid this effect, we considered the arrival time difference. The difference between the arrival time from center and left loudspeakers to the left-side participant could be set to zero by introducing a time delay to filter $\mathbf{g}_1$. So the left-side participant could hear the sounds coming out of the center and left loudspeakers simultaneously. We evaluated the effect of this compensation by MOS. Figure 5(c) shows the results. They show that sound localization was improved by compensating for the arrival time difference for the left-side participant.

Table 2 lists the gain level of gain factor f, and the gain level, frequency characteristics, and time delay of filters $\mathbf{g}_1$ and $\mathbf{g}_2$. Figure 5(d) shows the MOS results when the channel number is two; expanded to three by using gain factor f, filter $\mathbf{g}_1$, and $\mathbf{g}_2$ listed in Table 2; and three. The MOS results for the proposed system were 3.0 for both positions. These results show that the expanded-three channel sound had higher presence than the two-channel sound.

Table 2: Characteristics of each filter.

	f	$\mathbf{g}_1$	$\mathbf{g}_2$
gain level [dB]	-4.2	-1.6	-5.2
frequency characteristics	—	all pass	low pass (1kHz)
time delay [ms]	—	2.875	0

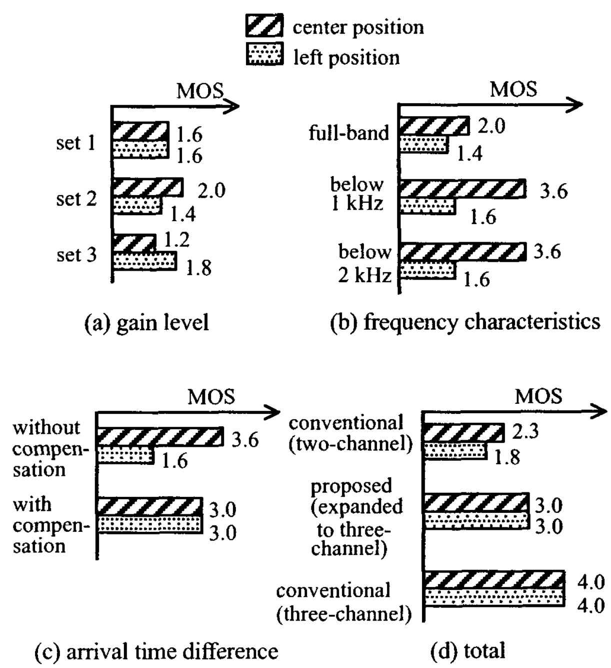

Fig. 5 Results of MOS evaluation.

3.2. Channel-number-compressed MAEC

There was very little influence on the computational load and performance of the MAEC when the channel number compressor and expander were introduced.

The features of the proposed MAEC are as follows.

(i) A stereo projection algorithm produces faster convergence [3].

(ii) A post filter reduces the residual echo after ADFs.

(iii) A stereo voice switch [5] prevents howling and stabilizes the MAEC.

(iv) A stereo shaker [6] estimates the true echo path.

(v) A channel number compressor and expander improve the presence at low computational load.

Figure 6 shows a block diagram of our proposed MAEC. The MAEC handles two channels while the loudspeakers and microphones systems employ three-channel sounds. Therefore, the channel number compressor is set between the microphones and ADF, and the channel number expander is set between the ADF and the loudspeakers.

4. REAL-TIME EXPERIMENTS

Real-time experiments were performed with the hardware in a teleconferencing room. The room had a reverberation time of 300 ms. Ambient noise made the signal-to-noise ratio about 20 dB. The stepsize of the projection algorithm was set to be 0.5. The microphones were located 150 cm in front of the loudspeakers and the distance between each loudspeaker was 75 cm. Figure 7 shows the convergence behavior of the proposed MAEC and a conventional two-channel AEC. Pseudo speech was used as the received signal. The acoustic echo could be reduced almost to the noise level by using the proposed MAEC. This system achieved almost the same performance as the two-channel AEC even when the compressor and expander were introduced.

CONCLUSION

We have proposed a channel-number-compressed multi-channel acoustic echo canceller to provide high-presence teleconferencing systems with large displays. This system has fewer channels for the multi-channel acoustic echo canceller than the actual channel number for loudspeakers and microphones as a result of using an expander and a compressor. The presence of the expanded-channel-number sound from the proposed system could be better than that of the original-channel-number sound according to mean opinion score tests. We implemented this system and verified its performance by actual measurement.

ACKNOWLEDGMENTS

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

REFERENCES

[1] M. Sondhi, D. Morgan, and J. Hall, "Stereophonic 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 and S. Makino, "Stereo projection echo canceller with true echo path estimation," *Proc. ICASSP95*, pp. 3059-3062, May 1995.

[4] S. Aoki and M. Okamoto, "Audio teleconferencing system with sound localization effect," *Proc. Joint Meeting ASA / EAA Berlin 99*, 2PSPA_5, Mar. 1999.

[5] S. Shimauchi, A. Nakagawa, Y. Haneda, and S. Makino, "Hybrid of acoustic echo canceller and voice switching control for multi-channel applications," *IWAENC99*, pp. 48-51, Sep. 1999.

[6] S. Shimauchi, S. Makino, Y. Haneda, A. Nakagawa, and S. Sakauchi, "A stereo echo canceller implemented using a stereo shaker and a duo-filter control system," *Proc. ICASSP99*, pp. 857-860, Mar. 1999.

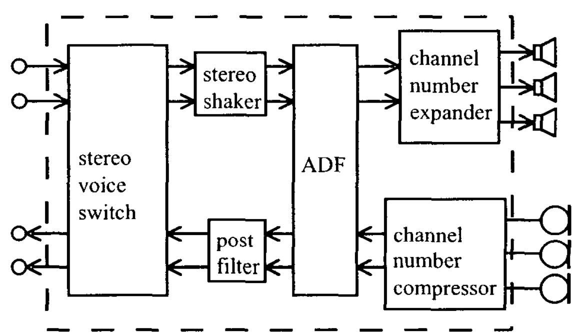

Fig. 6 Block diagram of our proposed multi-channel acoustic echo canceller.

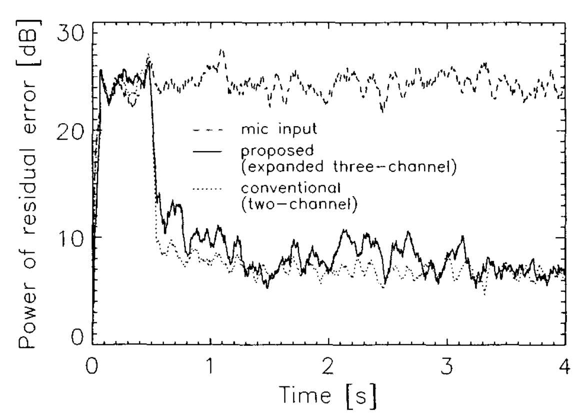

Fig. 7 Convergence behavior of actual measurements.