Audio Teleconferencing Set with Multipath Echo Canceller

*Hiroshi OIKAWA**, *Nobuo KOIZUMI*** and *Shoji MAKINO***

An audio teleconferencing set with an acoustic echo canceller has been developed which eliminates acoustic coupling and echo feedback, and allows full open and life-like conversation. It is designed to be used with a multiple microphone system containing gain switching.

The echo canceller is provided with memories for multiple echo paths so that filter coefficients may be updated when an echo path is switched due to a change of actuated microphone when speaker alternation occurs. Delay processing time of 0.5 sec. for 7 kHz band signal is achieved by using digital signal processor chips.

1 Introduction

Recently, the introduction of teleconferencing systems has been promoted for efficient office management as well as reducing travel time and expense. In such a system, it is desired to provide acoustic feedback compensation and room echo suppression so that participants can engage in natural two-way conversations.

Usually, either the voice switching method [1] or the echo cancelling method [2] is used for realizing hands-free telephony using microphones and loudspeakers. While the former method often causes the chopping of speech when the send/receive status alternates, a latter method keeps full open communication by subtracting simulated echo path filter output from the acoustically returned signal. Due to its excellent characteristics, the echo cancelling method has been introduced in teleconferencing systems.

The maximum delay processing capacity of current systems is about 0.25 sec., however, they require some acoustic treatment of the room. Also, when many microphones are used, direct mixing of multiple microphone outputs will decrease the signal-to-noise ratio, and will cause the degradation of speech quality. To avoid these effects, it is preferable to control the microphone gains individually [3]. This implies that there exist multiple acoustic echo paths, a condition in which a conventional canceller configuration experiences unstable operation.

The proposed configuration achieves stable performance during gain switching of multiple microphones, and delay processing time is extended to 0.5 sec., which will alleviate the need for room acoustics treatment.

* NTT Customer Equipment Division
** NTT Human Interface Laboratories

2 Design of the System

In comparison with current systems, the following methods are adopted to achieve stable performance during microphone switching, alleviation of room acoustic conditions, and convenient use during teleconferencing.

2.1 Realization of Stable Performance during Microphone Switching

When using multiple microphones, it is preferable for speech quality to actuate only one or two microphones rather than simply mixing all the microphone outputs directly [3]. In such a case, cancellation may be reduced due to a sudden change of acoustic echo path when speakers at the same site alternate. To avoid this effect, a new method has been developed which generates a current total acoustic echo path by a simple summation of impulse responses, each of which corresponds to a given microphone.

2.2 Alleviation of Room Acoustic Conditions

Reverberation time is from 0.2 to 0.5 sec. in conference rooms with excellent acoustics, and in some case, it reaches almost 1 second in ordinary rooms. The applicable reverberation time range for the system can be extended by expanding the delay processing time of the echo cancellation filter.

In use in an ordinary room environment, the maximum amount of cancellation will be approximately 30 dB due to the effect of ambient noise. Therefore, if the delay processing time of the echo cancelling filter is increased to 0.5 sec., the performance will be satisfactory even in a room with a reverberation time of 1 sec. Considering that the speech spectrum is about 10 dB lower in the high frequency range (4—7 kHz) than in the low range (0.2—4 kHz), and that room reverberation characteristics are lower in the high range, the delay processing time is set to 0.5 sec. at the lower band and 0.2 sec. at the higher band.

2.3 Function of Data Communications during Audio Conferencing

As it is convenient to use facsimile and electronic blackboards to send or represent documents efficiently during conferencing, a function is supplied to use one audio line for data communications by detecting the status of the peripheral equipment. To achieve this function, simultaneous voice and data communications are available in 2-line communications, while voice and data are alternated in single-line communications.

3 Technical Features

3.1 Control Technique for Multipath Echo Canceller

To keep speech quality high, microphone gains are switched to a high level, as we say "actuated", when the participant addresses the corresponding or nearest microphone, while the gains of microphones not been used at that time are attenuated. If an echo canceller is applied to such a microphone system, the amount of cancellation will decrease abruptly as the microphone gain switches, and the recovery should be delayed until a convergence of the adaptation scheme. Therefore, a system is proposed, where acoustic echo path transfer function memories are provided for each microphone (only one microphone can be actuated at a time), and the canceller filter coefficients for the total acoustic echo path are replaced by a linear combination of the values stored in the memories. The system configuration is shown in Fig. 1. As gain switching is distinguished from environmental variations in the acoustic echo path, cancellation degradation due to a change of speaker can be avoided by this method.

Let the gain for attenuated microphones be normalized to unity and then represent the gain of the actuated microphone when a single microphone is actuated by a_1, and the gain when two microphones are actuated simultaneously by a_2. Then by representing the total number of microphones as N, and the impulse response of the acoustic echo path to the i-th microphone as H_i, the impulse response when the j-th microphone is actuated can be described as

$$H_j^* = (a_1 - 1)H_j + \sum_{i=1}^{N} H_i. \tag{1}$$

On the other hand, if impulse response for both the j-th and k-th microphones are actuated, the equation is

$$H_{jk}^* = (a_2 - 1)\,(H_j + H_k) + \sum_{i=1}^{N} H_i. \tag{2}$$

If the relation

$$a_2 = (a_1 + 1)/2 \tag{3}$$

holds, the impulse response for when two microphones are actuated is obtained from that for when one is actuated by following equation.

$$H_{jk}^* = (H_j^* + H_k^*)/2 \tag{4}$$

This result implies that impulse responses for when a single microphone is actuated should be stored in memory.

If we assume that there are no correlations between microphone signals, the condition for a total mixed signal to be maintained at a constant level is

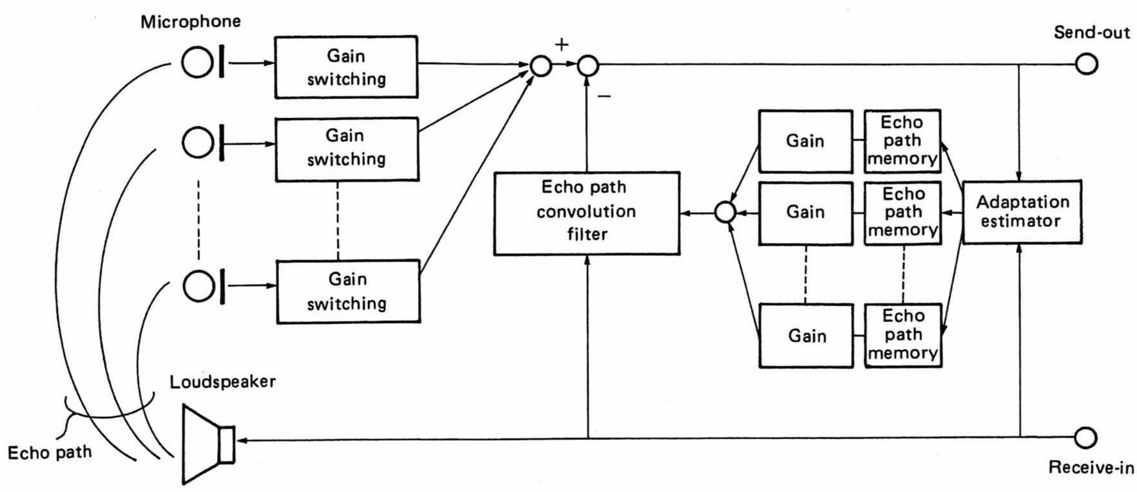

Fig. 1—Configuration of multipath echo canceller.

$$a_1^2+(N-1)=2a_2^2+(N-2)$$

or

$$a_2=\sqrt{\frac{a_1^2+1}{2}}\ . \tag{5}$$

Equation (3) does not satisfy the constant level condition in Eq. (5), however, the condition will be interpreted as a constant if the differences are about 1 dB. Set the gain for a single actuated microphone to 3, which means that off microphone gains are attenuated 9.5 dB, then the gain for when two microphones are actuated becomes 2 (6 dB), from the condition of the linear combination, where the constant level condition is 2.2 (7 dB).

Variation of echo cancellation over a period of time is simulated using real impulse response data from a room where the reverberation time was 0.5 sec. The results are shown in Fig. 2. While cancellation is greatly reduced after microphone switching by a conventional adaptation method, the proposed method maintains a stable cancellation rate during a change of speakers. The unique features of this technique are the new configuration for multiple echo paths and the condition for gain setting. The hardware can be implemented by simply adding of memories.

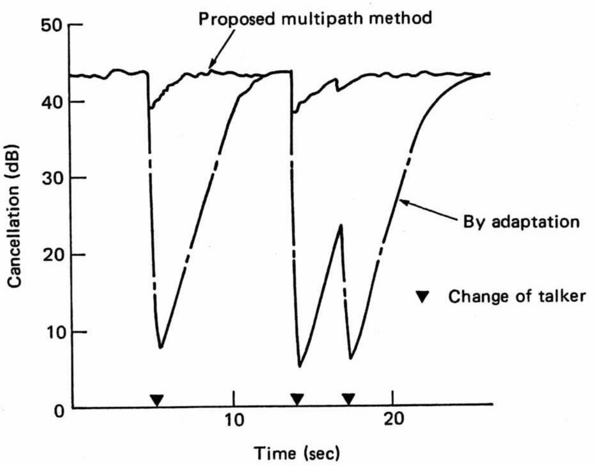

Fig. 2—Variation of echo cancellation over time.

3.2 Configuration of Digital Signal Processing Unit

The number of filter taps to execute delay processing of 0.5 sec. may become 4 thousand, at a sampling rate of 8 kHz. As it is impossible to implement the whole filter in a single LSI chip, even using state-of-the-art LSI technology, multiple chips must be used. For realtime operation, the digital signal processing unit is required to execute the whole process in a sampling period of 125 μsec. When the learning algorithm [4] is used for the adaptation of convolution filter coefficients, approximately 8000 (2N) multiply-and-add operations are required. Therefore, parallel processing is used to decrease the operation rate per chip and thus achieve realtime operation.

The digital signal processing unit consists of a combination of speech signal processing chips comprising a DSSP (Digital Speech Signal Processor, a 256×2 words RAM, with a machine cycle time of 100 nsec., and a 4 k words external expansion RAM for filter coefficients.) [5]. The block diagram is shown in Fig. 3. As the DSSP chip allows 256 tap convolution filtering in a single chip, the total hardware can be organized compactly.

Furthermore, a band splitting method is used [6] for 7 kHz band signals, where low (0.15–4 kHz) and high (4–7 kHz) range components of the speech signals are split into sub-band signals which are processed in parallel at an 8 kHz sampling rate. In the 3.4 kHz band

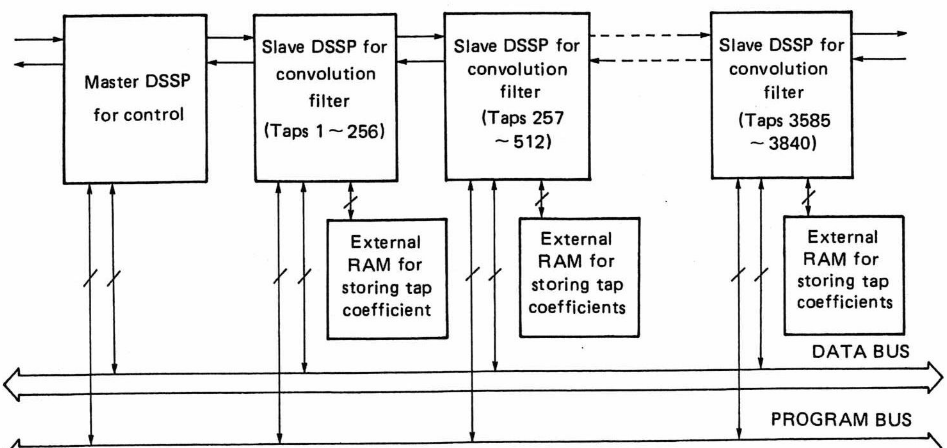

Fig. 3—Block diagram of digital signal processing unit.

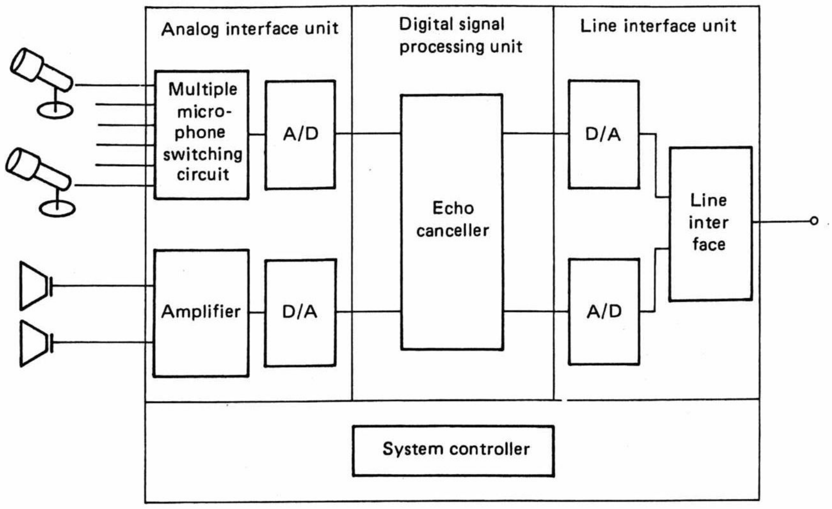

Fig. 4—System block diagram of audio teleconferencing set.

telephone line mode, and in 2-wire/1-line operation, the high range convolution filters are converted to the cancellation of the returned echo signal in the hybrid. This configuration economizes the hardware.

4 Hardware Implementation of the Set

The system block diagram of the set is shown in Fig. 4. This set is composed of an analog interface unit, a digital signal processing unit, line interface unit, and the system controller. The analog interface unit consists of a multiple microphone switching unit, audio signal amplifier, and A/D and D/A converters. The digital signal processing unit has two cancellation filter units for low and high frequency ranges, where the high range unit can also work as a hybrid echo canceller. The line interface unit provides network interface for 2-wire/1-line or 4-wire/2-line operations. The system controller manages the units and the panel is controlled by microprocessor.

5 Specifications

A summary of the set specification is shown in Table 1. The design parameters of the echo canceller unit are shown in Table 2. As up to 6 switching microphones can be connected, the transmitted speech quality can be maintained for teleconferencing with many pariticipants. Since the delay processing time is 0.5 seconds, the set can be used in the ordinary rooms, without special acoustic modifications. An example of the stationary cancellation characteristics measured in a room environment is shown in Fig. 5. The main functions of the set are summarized in Table 3.

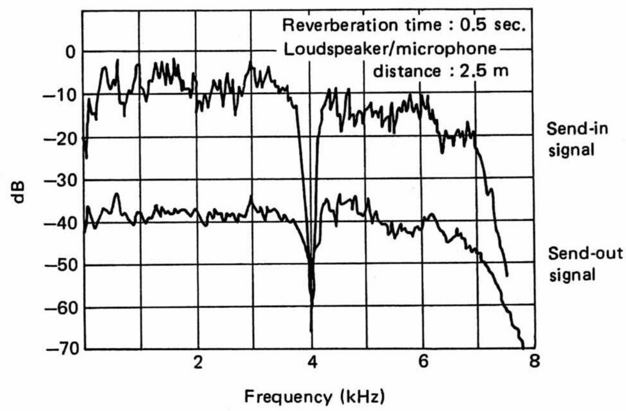

Fig. 5—Spectrum of echo cancellation.

6 Conclusion

An audio teleconferencing set has been developed using echo cancelling techniques. A design for hands-free telephony which has consistent performance for multiple microphones is described. It achieves stable performance for teleconferencing with many participants, while alleviating requirements for special room acoustics. Following are the key techniques of this development.

(1) The echo canceller design which maintains echo cancellation during the abrupt change of echo path when the microphone is switched due to speaker alternation.

(2) The circuit configuration of the speech signal processing chips to realize the echo cancelling of broadband (7 kHz) audio signals with a long convolution delay time (0.5 sec.).

Table 1 AUDIO TELECONFERENCING SET SPECIFICATIONS.

Items		Specifications
Frequency Band Width		0.3 ~ 3.4 kHz or 0.15 ~ 7 kHz
Line	Leased Circuit	2-wire or 4-wire
	Public Line	1-line or 2-line
Basic Telephone Features		External
Recommended Acoustic Conditions	Reverberation Time	0.8 sec. max.
	Room Noise	45 dB (A) max.
Number of Microphones		6 microphones are standard
Number of Loudspeakers		2 loudspeakers are standard
Multiple Microphone Switching Contorl		1 or 2 microphones are actuated
Acoustic Coupling Control		Acoustic echo canceller
Hybrid Coupling Control		Line echo canceller
Line Interface	Leased Circuit Interface	—10 dBm 600 Ω balance
	Public Line Interface	—10 dBm 600 Ω balance
	Dialing	External telephone set
Analog Interface	Microphone Sensitivity Directionality	—75 dBv/0.1 Pa unidirectional
	Loudspeaker Sensitivity	90 dBspl/w/m
	Recording Output Level	—26 dBm 600 Ω unbalance
	Volume Range	30 dB
	A/D, D/A Converter	16-bit linear

Table 2 ACOUSTIC AND LINE ECHO CANCELLER SPECIFICATIONS.

Items	Acoustic Echo Canceller	Line Echo Canceller
Tap Coefficient Replace Control	When actuated microphone changes	When voice or data are switched
Frequency Band Width	0.3 ~ 3.4 kHz or 0.15 ~ 7 kHz	0.3 ~ 3.4 kHz
Sampling Frequency	8 kHz or 16 kHz	8 kHz
Delay Processing Time	480 ms (Lower band) 224 ms (Higher band)	32 ms
Number of Taps	3840 Taps (Lower band) 1792 Taps (Higher band)	256 Taps
Adaptation Algorithm	Learning method	
Training	Chirp series	White noise

Table 3 AUDIO TELECONFERENCING SET FUNCTIONS.

Function	Contents
Prior Microphone	Microphone No. 1 is always actuated
External Equipment	Data communication during audio conferencing is available
Recording	Send and receive signal can be recorded
Remote Control	External control with video equipment is available
Microphone Mute	Send signal can be cut
Microphone Sensitivity Adjustment	Sensitivity adjustment is available for every type of microphone
Automatic Gain Control	Compensate line loss automatically
Training	Initial estimate of multiple echo path
Display	Reverberation time, Actuated microphone, Send level, Receive amplifier level, Frequency band width, etc.

This equipment is also expected to be installed in audio units for full-motion video and freeze-frame video conferencing systems.

Acknowledgment

The authors would like to thank Dr. Ryoichi Matsuda, Mr. Shinichi Yamazaki, and Mr. Isao Kawashima for their guidance, and Mr. Masakazu Nishino and Mr. Mizuhiro Tobita for fruitful discussions.

References

(1) D. Mitchell, et al.: General Transmission Considerations in Telephone Conference Systems, *IEEE Trans. on Communication Technology*, 16, 1, p. 163, 1968.

(2) Y. Itoh, Y. Maruyama, S. Furuya and T. Araseki: An Acoustic Echo Canceller for Teleconference, *ICC*, p. 1498, 1985.

(3) H. Oikawa, M. Nishino and M. Tobita: The Audio Teleconference System using Divided Speech-Band Voice-Switching Circuit, in this issue, p. 211.

(4) J. Nagumo and A. Noda: A Learning Method for System Identification, *IEEE Trans.*, AC12, 3, p. 282, 1967.

(5) H. Yamauchi, T. Kaneko and T. Kobayashi: An 18-bit Floating-point Signal Processor VLSI with an On-chip 512 W Dual-port RAM, *ICASSP*, p. 204, 1985.

(6) I. Furukawa: A Design of Canceller for Broad Band Acoustic Echo, *International Teleconference Symposium 5th*, p. 232, 1984.

The Authors

Hiroshi Oikawa
Senior Manager in the NTT Customer Equipment Division. There he is involved developing audio teleconference systems. Since joining the ECL in 1967, he has worked on developing audio teleconferencing systems, and on improving telephone equipment and audio-teleconferencing system and telephone speech quality.
Mr. Oikawa received a B.S. degree from Iwate University in 1967. He is a member of the Institute of Electronics, Information and Communication Engineers of Japan.

Nobuo Koizumi
Senior Research Engineer, Supervisor in the NTT Human Interface Laboratories. He is currently involved with the control of sound and vibration. Since joining the ECL system in 1975, he has been researching noise control, and more recently, teleconferencing equipment.
Mr. Koizumi received his B.E. in 1973 from Kyoto Institute of Technology and M.E. in 1975 from Kyoto University. He is a member of the Institute of Electronics, Information and Communication Engineers of Japan and the Acoustical Society of Japan.

Shoji Makino

Senior Research Engineer of the Speech and Acoustic Laboratory in the NTT Human Interface Laboratories. He is currently involved with the control of sound fields. Since joining the ECL system in 1981, he has been active in development of telephone sets, and more recently, teleconferencing equipment.

Mr. Makino received his B.E. in 1979 and M.E. in 1981 from Tohoku University. He is a member of the Institute of Electronics, Information and Communication Engineers of Japan and the Acoustical Society of Japan.

* * * *