HYBRID OF ACOUSTIC ECHO CANCELLERS AND VOICE SWITCHING CONTROL FOR MULTI-CHANNEL APPLICATIONS

Suehiro Shimauchi, Akira Nakagawa, Yoichi Haneda, and Shoji Makino

NTT Cyber Space Laboratories
3-9-11, Midori-cho, Musashino-shi, Tokyo, 180-8585 Japan
Email: sima@splab.hil.ntt.co.jp

ABSTRACT

We discuss an extension of a hybrid of an acoustic echo canceller (AEC) and voice switching control (VSC) for multi-channel applications. In the single-channel case, the hybrid has already become a practical method of making a hands-free teleconferencing system stable. To extend it to the multi-channel case, we took into account the stability of a multi-channel acoustic system. We also considered the loss insertion operation for multi-channel voice switching, which should be done more carefully than for the single-channel case.

1. INTRODUCTION

An acoustic echo canceller (AEC) cannot always cancel an echo, because of echo path changes, system initialization, and so on. When echo cancellation fails, the acoustical loop gain increases and howling occurs. To avoid this, in the single-channel case, a hybrid of an AEC and voice switching control (VSC) has been widely used for actual implementation to make a hands-free teleconferencing system stable. The switching loss is inserted into the reception or transmission line by judging the reception/transmission status. The insertion loss level is adequately set to achieve full-duplex communication and to avoid howling by measuring the echo cancellation level and the acoustic coupling level. As a result, the system is kept stable.

In multi-channel applications such as stereophonic teleconferencing and multi-point conferencing, the multi-channel AEC is indispensable. Since the echo path tracking speed of the multi-channel AEC is slower than that of the single-channel AEC, affected by the interchannel correlation[1], the echo cancellation performance may not be good enough. So hybrid operation is more helpful in the multi-channel case. Here we present an extension of the hybrid operation for multi-channel acoustic systems.

2. REVIEW OF SINGLE CHANNEL CASE

In the single-channel case, a hybrid system [2] is constructed as shown in Fig. 1. The VSC inserts the switching loss into the reception or transmission line. Its timing is controlled by judging whether the current speech is far-end speech (reception) or near-end speech (transmission). In some applications, the double-talk situation is also judged as either reception or transmission. The insertion loss level is adaptively determined from the echo cancellation level and the acoustic coupling level between the loudspeaker and the microphone. The loss insertion does not normally affect the full-duplex communication, since the AEC usually performs well and the loss level is set close to 0 dB. The echo cancellation level and the acoustic coupling level can be measured in the reception situation. In the case of the duo-filter structure AEC, the reception situation is guaranteed when the foreground echo cancellation filter coefficients are copied from those of the background adaptive echo path estimation filter. This is because coefficient coping is permitted only when the background echo cancellation is better to some extent than the foreground one. In the case of transmission or double-talk, the background adaptive filter cannot reduce the error output and coefficient coping is not permitted.

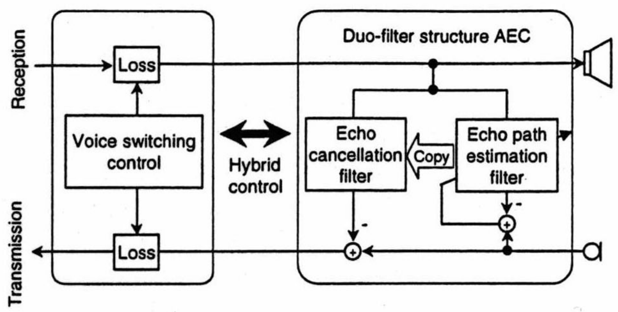

Figure 1: Single-channel hybrid system.

3. STABILITY ANALYSIS OF MULTI-CHANNEL ACOUSTIC SYSTEM

Before discussing the extension of the above hybrid system for the multi-channel case, we consider the stability of a multi-channel acoustic system [3]. Figure 2 shows an example of an M-channel hands-free teleconferencing system. The relationship between the microphone inputs $U_1(z), U_2(z), \ldots, U_M(z)$ at point A and the loudspeaker outputs $V_1(z), V_2(z), \ldots, V_M(z)$ at point B is described by

$$\begin{bmatrix} V_1(z) \\ V_2(z) \\ \vdots \\ V_M(z) \end{bmatrix} = \left[\boldsymbol{I} - z^{-1} \boldsymbol{G}_a^T(z) \boldsymbol{G}_b^T(z) \right]^{-1} \begin{bmatrix} U_1(z) \\ U_2(z) \\ \vdots \\ U_M(z) \end{bmatrix}, \quad (1)$$

where ^{T} is a transpose and $\boldsymbol{I}$ is a unit matrix. $\boldsymbol{G}_a(z)$ and $\boldsymbol{G}_b(z)$ are the transfer function matrices:

$$\boldsymbol{G}_a(z) = \begin{bmatrix} G_{a11}(z) & \ldots & G_{a1M}(z) \\ \vdots & \ddots & \vdots \\ G_{aM1}(z) & \ldots & G_{aMM}(z) \end{bmatrix} \quad (2)$$

and

$$\boldsymbol{G}_b(z) = \begin{bmatrix} G_{b11}(z) & \ldots & G_{b1M}(z) \\ \vdots & \ddots & \vdots \\ G_{bM1}(z) & \ldots & G_{bMM}(z) \end{bmatrix}, \quad (3)$$

where $G_{amn}(z)$ and $G_{bmn}(z)$ are the transfer functions between the m-th ($m = 1, 2, \ldots, M$) loudspeaker and the n-th ($n = 1, 2, \ldots, M$) microphone at points A and B respectively. The multi-channel system corresponding to Eq. (1) is stable if and only if the following equation is satisfied.

$$\| \boldsymbol{G}_a(e^{j\omega}) \boldsymbol{G}_b(e^{j\omega}) \| < 1 \quad \text{for } \forall \omega, \quad (4)$$

where the norm of matrix $\boldsymbol{A}$, $\| \boldsymbol{A} \|$, is defined as the square root of the maximum eigenvalue of $\boldsymbol{A}^H \boldsymbol{A}$, where ^{H} denotes a conjugate transpose. Assuming that the acoustical conditions at points A and B are similar, Eq. (4) can be approximated as

$$\| \boldsymbol{G}_a(e^{j\omega}) \| < 1, \quad \| \boldsymbol{G}_b(e^{j\omega}) \| < 1 \quad \text{for } \forall \omega, \quad (5)$$

where $\| \boldsymbol{G}_a(e^{j\omega}) \|$ and $\| \boldsymbol{G}_b(e^{j\omega}) \|$ in dB correspond to the multi-channel acoustic coupling levels at points A and B, respectively.

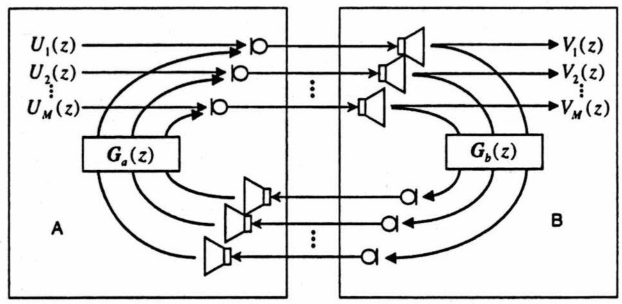

Figure 2: Multi-channel teleconferencing system.

4. EXTENDED HYBRID SYSTEM FOR MULTI-CHANNEL CASE

An extended hybrid system for the multi-channel case is shown in Fig. 3. It uses the multi-channel AEC described in Ref. [4], which has the duo-filter structure. Here, we mainly discuss the VSC. A simple loss insertion assignment is to switch between all the transmission channel losses and all the reception channel losses simultaneously by judging the conversation status from only two categories, reception or transmission. This assignment is reasonable for single-talk situations, in which only a single talker on one side is speaking. However, in the double-talk case, especially when several talkers are speaking on both sides, we need a wider range of the conversation statuses so that each channel can have an independent status.

4.1. Two-channel VSC

Below, we discuss in more detail, considering a two-channel VSC as an example.

4.1.1. Loss insertion assignment

Here we consider how the loss insertion or bypass is assigned to each reception or transmission line. The loss bypass should not be assigned to both the reception and transmission lines simultaneously,

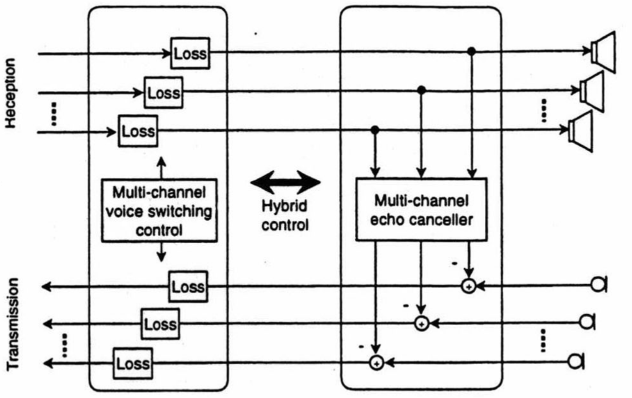

Figure 3: Multi-channel hybrid system.

otherwise the coupling between the bypassed reception line and the bypassed transmission line cannot be controlled. So in this example, only the transmission signal of each channel is checked and the conversation status is categorized as shown in Table 1. In the case of status 1, the losses are inserted in all transmission lines and are bypassed in all reception lines. In the case of status 2 or 3, the loss is bypassed only in the transmission line in which transmission speech is detected. In the case of status 4, the losses are inserted in all reception lines and are bypassed in all transmission lines.

Table 1: Conversation statuses for two-channel case.

Status	Transmission speech	
	Channel 1	Channel 2
1	×	×
2	○	×
3	×	○
4	○	○

○ : Detected × : Not detected

4.1.2. Insertion loss level

To determine the insertion loss level for each line, ideally we should measure the echo cancellation rates $E_1(z)$ and $E_2(z)$ and the acoustic coupling transfer function matrix:

$$\boldsymbol{G}_a(z) = \begin{bmatrix} G_{a11}(z) & G_{a12}(z) \\ G_{a21}(z) & G_{a22}(z) \end{bmatrix}. \qquad (6)$$

However, they may be estimated as scalar values in the time domain, since the loss levels are scalar values here. The echo cancellation rates E_1 and E_2 are very easy to estimate by comparing signals before and after the AEC. However, $\boldsymbol{G}_a(z)$ is indeterminate in general. So we assume $G_{a11}(z) = G_{a22}(z) = A$ and $G_{a12}(z) = G_{a21}(z) = B$. Then we can obtain the estimated $\boldsymbol{G}_a(z)$ from the relationship between the loudspeaker input signals $[X_1, X_2]$ and the microphone output signals $[Y_1, Y_2]$, unless $A = B$. Figure 4 shows a block diagram of this system, where LR_1 and LR_2 are the reception loss rates, and LT_1 and LT_2 are the transmission loss rates. The system matrix $\boldsymbol{M}$ becomes

$$\boldsymbol{M} = \begin{bmatrix} LT_1 & 0 \\ 0 & LT_2 \end{bmatrix} \begin{bmatrix} E_1 & 0 \\ 0 & E_2 \end{bmatrix} \begin{bmatrix} A & B \\ B & A \end{bmatrix} \begin{bmatrix} LR_1 & 0 \\ 0 & LR_2 \end{bmatrix}. \qquad (7)$$

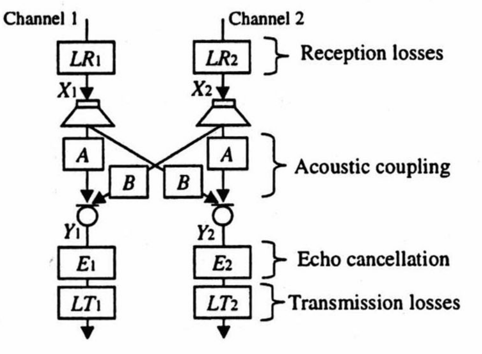

Figure 4: Two-channel acoustic control system.

Status 1

In the case of status 1, where $LR_1 = 1$ and $LR_2 = 1$ (both of the reception losses are bypassed), the system stability can be evaluated by analyzing the following matrix:

$$\boldsymbol{M}_1 = \begin{bmatrix} LT_1 & 0 \\ 0 & LT_2 \end{bmatrix} \begin{bmatrix} E_1 & 0 \\ 0 & E_2 \end{bmatrix} \begin{bmatrix} A & B \\ B & A \end{bmatrix}. \qquad (8)$$

In this case, by setting

$$LT_1 < \frac{1}{|E_1|(|A| + |B|)} \qquad (9)$$

and

$$LT_2 < \frac{1}{|E_2|(|A| + |B|)}, \qquad (10)$$

$\| \boldsymbol{M}_1 \| < 1$ is satisfied.

Status 4

In the case of status 4, where $LT_1 = 1$ and $LT_2 = 1$, the system stability can be evaluated by analyzing the following matrix:

$$M_4 = \begin{bmatrix} E_1 & 0 \\ 0 & E_2 \end{bmatrix} \begin{bmatrix} A & B \\ B & A \end{bmatrix} \begin{bmatrix} LR_1 & 0 \\ 0 & LR_2 \end{bmatrix}. \tag{11}$$

In this case, the critical solution is rather complicated. So we simply set

$$LR_1 < \frac{1}{|E_1||A| + |E_2||B|} \tag{12}$$

and

$$LR_2 < \frac{1}{|E_2||A| + |E_1||B|}. \tag{13}$$

Then $\| M_4 \| < 2$. However, if some attenuation margin is given, the system can be controlled to be stable.

Statuses 2 and 3

In the case of status 2, where $LT_1 = 1$, the system stability can be evaluated by analyzing the following matrix:

$$M_2 = \begin{bmatrix} 1 & 0 \\ 0 & LT_2 \end{bmatrix} \begin{bmatrix} E_1 & 0 \\ 0 & E_2 \end{bmatrix} \begin{bmatrix} A & B \\ B & A \end{bmatrix} \begin{bmatrix} LR_1 & 0 \\ 0 & LR_2 \end{bmatrix}. \tag{14}$$

In this case, we set LT_2 to the same value as in status 1. Comparing to status 4, we can obtain

$$LR_1 < \frac{1}{|E_1||A| + \dfrac{|B|}{|A| + |B|}} \tag{15}$$

and

$$LR_2 < \frac{1}{\dfrac{|A|}{|A| + |B|} + |E_1||B|}. \tag{16}$$

Then, $\| M_2 \| < 2$ like status 4. If some more attenuation margin *Att* is given, simpler conditions:

$$LR_1 < \frac{Att}{|E_1||A|} \tag{17}$$

and

$$LR_2 < \frac{Att}{|E_2||A|} \tag{18}$$

can also make $\| M_2 \| < 1$.

For status 3, similar results can be derived.

The insertion loss levels, which are the loss rates in dB, can be obtained from Table 2. For all the loss rates, the same attenuation margin *Att* is applied. The loss rates are controlled not to be over 1.

Table 2: Loss insertion assignment (two-channel).

Status	Insertion loss rate			
	LT_1	LT_2	LR_1	LR_2
1	$\frac{Att}{\lvert E_1\rvert(\lvert A\rvert+\lvert B\rvert)}$	$\frac{Att}{\lvert E_2\rvert(\lvert A\rvert+\lvert B\rvert)}$	1	1
2	1	$\frac{Att}{\lvert E_2\rvert(\lvert A\rvert+\lvert B\rvert)}$	$\frac{Att}{\lvert E_1\rvert\lvert A\rvert}$	$\frac{Att}{\lvert E_1\rvert\lvert B\rvert}$
3	$\frac{Att}{\lvert E_1\rvert(\lvert A\rvert+\lvert B\rvert)}$	1	$\frac{Att}{\lvert E_2\rvert\lvert B\rvert}$	$\frac{Att}{\lvert E_2\rvert\lvert A\rvert}$
4	1	1	$\frac{Att}{\lvert E_1\rvert\lvert A\rvert+\lvert E_2\rvert\lvert B\rvert}$	$\frac{Att}{\lvert E_2\rvert\lvert A\rvert+\lvert E_1\rvert\lvert B\rvert}$

5. CONCLUSIONS

A hybrid system of an AEC and VSC has been extended for multi-channel applications. Focusing mainly on the VSC, the loss insertion assignment and the insertion loss levels were adequately derived. By using the hybrid control, the total performance of the multi-channel teleconferencing system can be made reliable.

ACKNOWLEDGMENTS

We would like to thank Mr. K. Yamamori, Dr. Y. Kaneda, and Dr. S. Aoki for fruitful discussions and helpful suggestions.

REFERENCES

[1] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic Acoustic Echo Cancellation -An Overview of the Fundamental Problem," *IEEE Signal Processing Letters*, vol. 2, no. 8, pp. 148-151 (1995).

[2] Y. Haneda, S. Makino, J. Kojima, and S. Shimauchi, "Implementation and evaluation of an acoustic echo canceller using duo-filter control system," *Proc. EUSIPCO96*, vol. 2, pp. 1115-1118 (1996).

[3] S. Shimauchi, Y. Haneda, and J. Kojima, "Study on acoustic coupling in multi-channel teleconferencing system", *Proc. Spring Meet. Acoust. Soc. Japan*, pp. 631-632 (1997) (in Japanese).

[4] S. Shimauchi, S. Makino, Y. Haneda, A. Nakagawa, and S. Sakauchi "A stereo echo canceller implemented using a stereo shaker and duo-filter control system", *Proc. ICASSP99*, vol. 2, pp. 857-860 (1999).