ERROR ALLOCATIONS IN STEREO ECHO CANCELLATION ALGORITHMS

*[1] *Suehiro Shimauchi*, [1] *Yoichi Haneda, and* [1] *Akitoshi Kataoka*

*shimauchi.suehiro@lab.ntt.co.jp
[1] NTT Cyber Space Laboratories, NTT Corporation
3-9-11, Midori-cho, Musashino-shi, Tokyo, 180-8585 Japan

ABSTRACT

We discuss stereo echo cancellation in terms of error allocation fed back to the echo path estimate of each channel. In stereo or multi-channel echo cancellation, the output errors between multi-channel echo paths and their corresponding estimates cannot be obtained separately but are observed only in a mixed form at each microphone. Thus, the choice of rule for allocating the mixed error to the channels is important to achieve good echo path estimation. We propose a new error allocation strategy, which corresponds to the combination of two allocation rules: stereo-input-power-balance-based allocation and division-by-two-based allocation. To control the contribution of these two rules appropriately, we also introduce statistical amplitude estimation of the unobserved error component.

1. INTRODUCTION

In teleconferencing with someone at the far-end, stereo (or more generally multi-channel) audio provides a more realistic experience than monaural audio. Stereo echo cancellation is an important issue to achieve such communication with high quality. The non-uniqueness of stereo echo path estimates is a well known problem [1, 2, 3] and widely understood to be caused by inter-channel correlation of the loudspeaker input signals. Many approaches to overcome this problem are based on linear combined multi-channel adaptive filtering. It straightforwardly allocates the mixed error, which is observed in a mixed form of output errors of the echo path estimates at each microphone, to each echo path channel based on Wiener estimation [4]. In this paper, we focus on error allocation strategies, which should not necessarily be limited to the above straightforward one. In fact, some conventional algorithms [5, 6, 7] can be regarded as being based on different error allocations. We review some of their features from the aspect of error allocation and propose a new allocation rule for the stereo echo cancellation algorithm.

2. STEREO ECHO CANCELLATION

In a stereo teleconferencing system, the echoes from two loudspeakers are picked up in a mixed form by each microphone. A mixed echo observed as one microphone output is

$$y(n) = \sum_{m=1}^{2} d_m(n), \tag{1}$$

where $d_m(n) = \mathbf{x}_m^T(n)\mathbf{h}_m$ for the loudspeaker channel $m(=1,2)$, where $\mathbf{x}_m(n) = [x_m(n), \ldots, x_m(n-L+1)]^T$ is the vector of the stereo input signal $x_m(n)$, where L is the effective length of the echo path $\mathbf{h}_m$ whose elements are the impulse response between the m-th loudspeaker and the microphone truncated by L samples. For simplicity, we assume below that the microphone output is composed of only the mixed echo. As an example of a linear combined adaptive filter, the stereo normalized least-mean-squares (NLMS) algorithm updates the coefficients as

$$\begin{bmatrix}\hat{\mathbf{h}}_1(n+1)\\ \hat{\mathbf{h}}_2(n+1)\end{bmatrix} = \begin{bmatrix}\hat{\mathbf{h}}_1(n)\\ \hat{\mathbf{h}}_2(n)\end{bmatrix} + \frac{\mu \cdot e(n)}{\sum_{k=1}^{2} \|\mathbf{x}_k(n)\|^2} \begin{bmatrix}\mathbf{x}_1(n)\\ \mathbf{x}_2(n)\end{bmatrix}, \tag{2}$$

where $\hat{\mathbf{h}}_m(n)$ denotes an estimate of $\mathbf{h}_m$, the mixed error $e(n) = y(n) - \hat{d}_1(n) - \hat{d}_2(n)$, $\hat{d}_m(n) = \mathbf{x}_m^T(n)\hat{\mathbf{h}}_m(n)$, μ is the step-size, $\|\cdot\|$ denotes a vector norm, and T indicates a transposition.

3. ERROR ALLOCATIONS

In an ideal situation where the output error, $e_m(n) = d_m(n) - \hat{d}_m(n)$, could be obtained separately, the estimate $\hat{\mathbf{h}}_m(n)$ could be updated separately as

$$\hat{\mathbf{h}}_m(n+1) = \hat{\mathbf{h}}_m(n) + \mu \frac{\mathbf{x}_m(n)}{\|\mathbf{x}_m(n)\|^2} e_m(n), \tag{3}$$

and it would be free from multi-channel problems such as non-uniqueness. However, it is impossible to obtain $e_m(n)$ separately, so a practical alternative is to obtain an estimated error $\hat{e}_m(n)$. Below we investigate how $e_m(n)$ can be estimated by using observable information. The output errors can be written as

$$e_1(n) = e_c(n) + e_a(n), \tag{4}$$

$$e_2(n) = e_c(n) - e_a(n), \tag{5}$$

where

$$e_c(n) = [\mathbf{\Delta}_c^T(n), \mathbf{\Delta}_a^T(n)] \begin{bmatrix} \mathbf{x}_c(n) \\ \mathbf{x}_a(n) \end{bmatrix}, \quad (6)$$

$$e_a(n) = [\mathbf{\Delta}_c^T(n), \mathbf{\Delta}_a^T(n)] \begin{bmatrix} \mathbf{x}_a(n) \\ \mathbf{x}_c(n) \end{bmatrix}, \quad (7)$$

$$\mathbf{\Delta}_c(n) = \frac{\mathbf{\Delta}_1(n) + \mathbf{\Delta}_2(n)}{2}, \quad (8)$$

$$\mathbf{\Delta}_a(n) = \frac{\mathbf{\Delta}_1(n) - \mathbf{\Delta}_2(n)}{2}, \quad (9)$$

$$\mathbf{x}_c(n) = \frac{\mathbf{x}_1(n) + \mathbf{x}_2(n)}{2}, \quad (10)$$

$$\mathbf{x}_a(n) = \frac{\mathbf{x}_1(n) - \mathbf{x}_2(n)}{2}, \quad (11)$$

$$\mathbf{\Delta}_1(n) = \mathbf{h}_1 - \hat{\mathbf{h}}_1(n), \quad (12)$$

$$\mathbf{\Delta}_2(n) = \mathbf{h}_2 - \hat{\mathbf{h}}_2(n). \quad (13)$$

By substituting (8)-(11) into (6) and (7) and using orthogonal decomposition of

$$\begin{bmatrix} \mathbf{x}_a(n) \\ \mathbf{x}_c(n) \end{bmatrix} = \alpha \begin{bmatrix} \mathbf{x}_c(n) \\ \mathbf{x}_a(n) \end{bmatrix} + \begin{bmatrix} \mathbf{w}(n) \\ \mathbf{z}(n) \end{bmatrix}, \quad (14)$$

we can rewrite $e_c(n)$ and $e_a(n)$ as

$$e_c(n) = \frac{1}{2} e(n), \quad (15)$$

$$e_a(n) = \frac{\alpha}{2} e(n) + e_u(n), \quad (16)$$

where

$$\alpha = \frac{2\mathbf{x}_c^T(n)\mathbf{x}_a(n)}{||\mathbf{x}_c(n)||^2 + ||\mathbf{x}_a(n)||^2} = \frac{||\mathbf{x}_1(n)||^2 - ||\mathbf{x}_2(n)||^2}{||\mathbf{x}_1(n)||^2 + ||\mathbf{x}_2(n)||^2}, \quad (17)$$

$$e_u(n) = [\mathbf{\Delta}_c^T(n), \mathbf{\Delta}_a^T(n)] \begin{bmatrix} \mathbf{w}(n) \\ \mathbf{z}(n) \end{bmatrix}, \quad (18)$$

$$\begin{bmatrix} \mathbf{w}(n) \\ \mathbf{z}(n) \end{bmatrix} = \begin{bmatrix} \mathbf{x}_a(n) \\ \mathbf{x}_c(n) \end{bmatrix} - \alpha \begin{bmatrix} \mathbf{x}_c(n) \\ \mathbf{x}_a(n) \end{bmatrix}. \quad (19)$$

According to (15) and (16), the estimates of $e_1(n)$ and $e_2(n)$ can be written in terms of the known $e_c(n)$ and an estimate of the unknown $e_a(n)$ as

$$\hat{e}_1(n) = e_c(n) + \hat{e}_a(n), \quad (20)$$

$$\hat{e}_2(n) = e_c(n) - \hat{e}_a(n). \quad (21)$$

To estimate $e_1(n)$ and $e_2(n)$, the stereo NLMS is regarded as using the stereo-input-power-balance-based allocation. The estimates are given by:

$$\hat{e}_1(n) = \frac{1}{2}e(n) + \frac{\alpha}{2}e(n) = \frac{||\mathbf{x}_1(n)||^2}{||\mathbf{x}_1(n)||^2 + ||\mathbf{x}_2(n)||^2} e(n), \quad (22)$$

$$\hat{e}_2(n) = \frac{1}{2}e(n) - \frac{\alpha}{2}e(n) = \frac{||\mathbf{x}_2(n)||^2}{||\mathbf{x}_1(n)||^2 + ||\mathbf{x}_2(n)||^2} e(n), \quad (23)$$

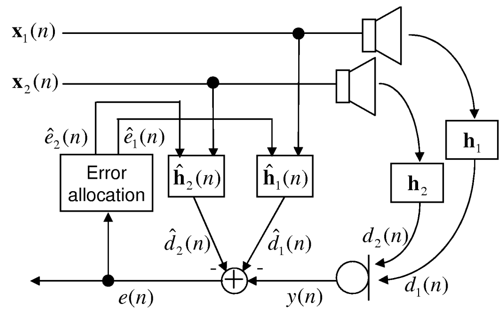


Figure 1: Configuration of stereo echo canceller.

by setting $\hat{e}_a(n) = \alpha e(n)/2$. In particular, when $|e_u(n)| \ll |\alpha e(n)/2|$, they are reliable candidates of estimates, because $\hat{e}_a(n) \approx e_a(n)$ holds. Other candidates given by $\hat{e}_a(n) = 0$ are obtained by the division-by-two-based allocation:

$$\hat{e}_m(n) = \frac{1}{2} e(n), \quad (24)$$

which correspond to the estimates used in an algorithm proposed by Fujii et al. [5]. They are also reasonable when $|e_u(n)| \gg |\alpha e(n)/2|$ and $e_a(n)$ can hardly be estimated.

Besides the above allocation rules, some statistical knowledge is helpful to derive better allocation. By assuming $||\mathbf{\Delta}_1(n)|| \approx ||\mathbf{\Delta}_2(n)||$, Nakagawa et al. derived another kind of algorithm [6] which uses the stereo-input-amplitude-balance-based allocation:

$$\hat{e}_m(n) = \frac{||\mathbf{x}_m(n)||}{||\mathbf{x}_1(n)|| + ||\mathbf{x}_2(n)||} e(n). \quad (25)$$

However, the assumption may not always hold properly in all the situations where it is implemented.

Figure 1 shows the configuration model of the stereo echo canceller that we are discussing.

4. PROPOSED ALGORITHM

We apply more general statistical properties to error allocation. As mentioned in the previous section, the accuracy of the estimate of $e_a(n)$ in the stereo NLMS and the algorithm using (24) depends on the amount of $|e_u(n)|$, so, we consider combining these two algorithms by weighting their contribution according to the statistically estimated amount of $|e_u(n)|$. From (6) and (18), the following relationship holds approximately for any $\mathbf{\Delta}_c(n)$ and $\mathbf{\Delta}_a(n)$.

$$\frac{E[|e_u(n)|^2]}{E[|e_c(n)|^2]} \approx \frac{[\mathbf{w}^T(n), \mathbf{z}^T(n)] \begin{bmatrix} \mathbf{w}(n) \\ \mathbf{z}(n) \end{bmatrix}}{[\mathbf{x}_c^T(n), \mathbf{x}_a^T(n)] \begin{bmatrix} \mathbf{x}_c(n) \\ \mathbf{x}_a(n) \end{bmatrix}}. \quad (26)$$

Table 1: Error allocations

	$\hat{e}_m(n), m = 1, 2$
Stereo NLMS	$\frac{\|\|\mathbf{x}_m(n)\|\|^2}{\|\|\mathbf{x}_1(n)\|\|^2+\|\|\mathbf{x}_2(n)\|\|^2}e(n)$
Fujii [5]	$\frac{1}{2}e(n)$
Nakagawa [6]	$\frac{\|\|\mathbf{x}_m(n)\|\|}{\|\|\mathbf{x}_1(n)\|\|+\|\|\mathbf{x}_2(n)\|\|}e(n)$
Proposed	$[\frac{\gamma_m}{2} + \frac{(1-\gamma_m)\|\|\mathbf{x}_m(n)\|\|^2}{\|\|\mathbf{x}_1(n)\|\|^2+\|\|\mathbf{x}_2(n)\|\|^2}]e(n)$

Thus, according to (15), (17), and (19),

$$E[|e_u(n)|^2] \approx \frac{1-\alpha^2}{4}E[|e(n)|^2]. \quad (27)$$

We define a value for each channel to evaluate the ratio between the observable and unobservable components in $e_m(n)$:

$$\mathrm{SNR}_1 = \frac{E[|e_1(n) - e_u(n)|^2]}{E[|e_u(n)|^2]} \approx \frac{1+\alpha}{1-\alpha}, \quad (28)$$

$$\mathrm{SNR}_2 = \frac{E[|e_2(n) + e_u(n)|^2]}{E[|e_u(n)|^2]} \approx \frac{1-\alpha}{1+\alpha}. \quad (29)$$

By denoting

$$\gamma_m = \frac{1}{1 + \mathrm{SNR}_m}, \quad (30)$$

we propose the following error allocation to apply to (3):

$$\hat{e}_1(n) = \left[\frac{1}{2}\gamma_1 + \frac{1+\alpha}{2}(1-\gamma_1)\right]e(n), \quad (31)$$

$$\hat{e}_2(n) = \left[\frac{1}{2}\gamma_2 + \frac{1-\alpha}{2}(1-\gamma_2)\right]e(n). \quad (32)$$

By incorporating the statistical amplitude estimate of the unknown component of $e_m(n)$, we combined the two conventional algorithms appropriately. When SNR_m is large, the estimated error $\hat{e}_m(n)$ becomes close to that of the stereo NLMS. When SNR_m is small, it becomes close to that of the algorithm using (24).

Table 1 summarizes the error allocations mentioned above, including the proposed one. From the above derivation, the proposed algorithm is based on the control based on the value of α in (17), which deals with the power difference of the stereo input. It does not essentially solve the non-uniqueness problem, but it can reduce degradation caused by the stereo input signal power difference using a different formulation from the algorithms proposed by Fujii et al. [5] and Nakagawa et al. [6].

5. SIMULATIONS

We evaluated the performance of the proposed algorithm by simulation. The conventional stereo NLMS, the algorithm using (24), and the algorithm using (25) were also simulated for comparison. All algorithms were implemented based on (3) and their own error allocation rules. The sampling frequency was 8 kHz in all the simulations. The stereo echo paths $\mathbf{h}_m$ were impulse responses truncated by 500 samples after being measured in a room with a reverberation time of 200 ms, where $||\mathbf{h}_1|| \approx 1.2||\mathbf{h}_2||$. For the stereo input signals $x_m(n)$, a male talker's voice (Fig. 2) was recorded in the same room using two microphones, both of which were located about 80 cm away from the talker and 130 cm away from each other. Three sets of stereo input signals $x_m(n)$ were made from the voice signals by changing their level balance as 1:1, 1:2, and 1:5. Since the voice signals included slight uncorrelated ambient noise and were also convolved by actual room impulse responses sufficiently longer than 500 samples, they did not cause the non-uniqueness problem, but they were still highly correlated. As a microphone output $y(n)$, $x_m(n)$ and $\mathbf{h}_m$ were convolved and white noise was added to give an echo-to-noise ratio of 45 dB. To relax the influence of the additive noise and avoid division by zero, we added $\delta = 1.5 \times 10^5$ to $||\mathbf{x}_m(n)||^2$ in (3).

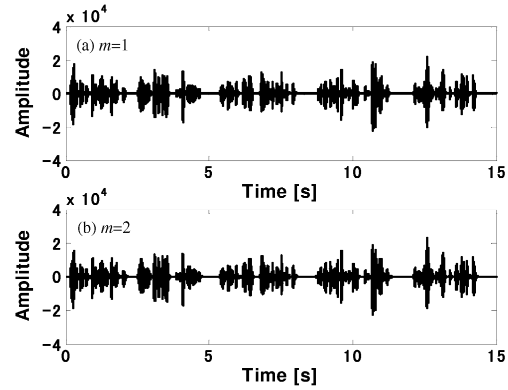


Figure 2: Stereo voice signals.

Figures 3(a), (b), and (c) show the results of the simulations. The performances were evaluated using the normalized coefficient error

$$10\log_{10}\frac{||\mathbf{h}_1 - \hat{\mathbf{h}}_1(n)||^2 + ||\mathbf{h}_2 - \hat{\mathbf{h}}_2(n)||^2}{||\mathbf{h}_1||^2 + ||\mathbf{h}_2||^2}[\mathrm{dB}]. \quad (33)$$

Figures 3(a), (b), and (c) respectively correspond to the cases of $||\mathbf{x}_1(n)|| \approx ||\mathbf{x}_2(n)||$, $2||\mathbf{x}_1(n)|| \approx ||\mathbf{x}_2(n)||$, and $5||\mathbf{x}_1(n)|| \approx ||\mathbf{x}_2(n)||$. When the stereo signal had similar power in each channel, all the algorithms gave similar results (Fig. 3(a)). As the power difference was increased, the performance of the stereo NLMS showed the worst degradation among them. The others had less degradation. In particular, the proposed one exhibited the best performance in this set of simulation conditions.

6. CONCLUSIONS

We discussed stereo echo cancellation in terms of the choice of rule for allocating the mixed error to the channels. By analyzing some conventional algorithms in that sense, we devised a new error allocation rule, which corresponds to the combination of two allocation rules: the stereo-input-power-balance-based rule and the division-by-two-based rule. In order to control the contribution of these two rules appropriately, we also introduced the statistical amplitude estimation of the unobserved error component. We confirmed the effectiveness of the proposed algorithm by simulation.

7. ACKNOWLEDGEMENT

We would like to thank Mr. A. Imamura for fruitful discussions and helpful suggestions.

8. REFERENCES

[1] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic acoustic echo cancellation – an overview of the fundamental problem," *IEEE SP Letters*, vol. 2, no. 8, pp. 148-151, 1995.

[2] S. Shimauchi and S. Makino, "Stereo projection echo canceller with true echo path estimation," *Proc. IEEE ICASSP '95*, vol. 5, pp. 3059–3062, 1995.

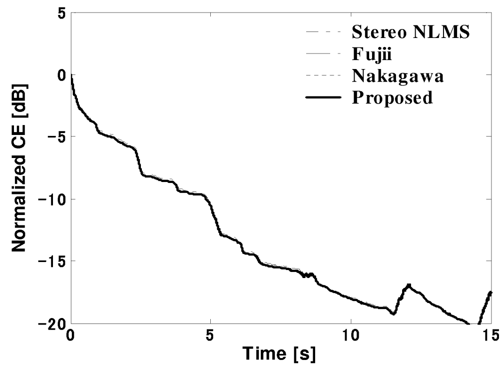
[3] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the specific problem of stereophonic echo cancellation," *IEEE Trans. SAP*, vol. 6, no. 2, pp. 156–165, 1998.

[4] H. L. V. Trees, *Detection, estimation, and modulation theory, Part I*, John Wiley & Sons, Inc. 1968.

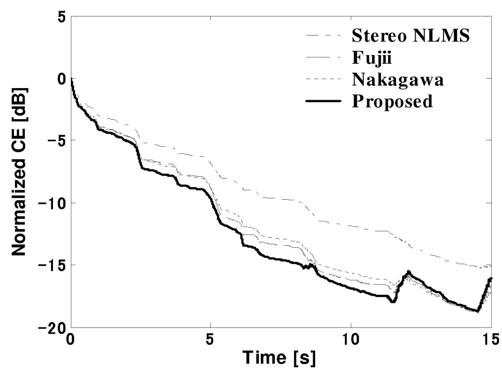
[5] T. Fujii and S. Shimada, "Linear-combined multichannel adaptive digital filters," *IEICE Trans.*, vol. J69-A, no. 10, pp. 1226–1233, 1986.

[6] A. Nakagawa, Y. Haneda, and S. Makino, "A study of an adaptive algorithm for stereo signals with a power difference," *Proc. IEEE ICASSP '02*, vol. 2, pp. 1913–1916, 2002.

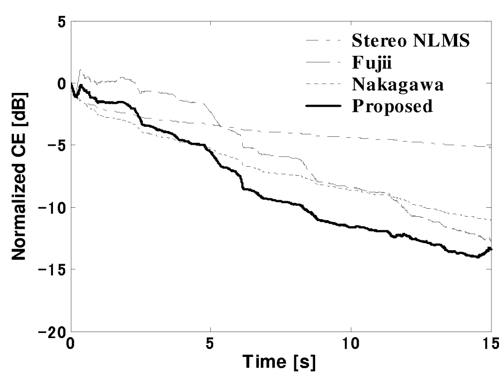
[7] S. Emura, Y. Haneda, and A. Kataoka, "A solution to echo path imbalance problem in stereo echo cancellation," *Proc. IEEE ICASSP '04*, vol. 4, pp. 129–132, 2004.



(a) $||\mathbf{x}_1(n)|| \approx ||\mathbf{x}_2(n)||$.



(b) $2||\mathbf{x}_1(n)|| \approx ||\mathbf{x}_2(n)||$.



(c) $5||\mathbf{x}_1(n)|| \approx ||\mathbf{x}_2(n)||$.

Figure 3: Comparisons of normalized coefficient error convergence.