International Workshop on Acoustic Echo and Noise Control (IWAENC2003), Sept. 2003, Kyoto, Japan

STEREOPHONIC ACOUSTIC ECHO CANCELLER WITH PRE-PROCESSING — SECOND-ORDER PRE-PROCESSING FILTER AND ITS CONVERGENCE —

Akihiro Hirano, Kenji Nakayama, Kazue Takebe

Faculty of Engineering, Kanazawa University
40-20, Kodatsuno 2-Chome, Kanazawa, 920-8667, JAPAN
e-mail:{hirano,nakayama}@t.kanazawa-u.ac.jp

ABSTRACT

This paper proposes a stereophonic acoustic echo canceller with randomly time-varying second-order pre-processing filters. A quantitative approach for the convergence speed assessment by using time-averaged correlation matrix is also introduced. Simulation results show that the convergence speed of the proposed algorithm is more than twice as fast as that of a conventional pre-processor. The quantitative approach provides us with a good overview for the convergence speed assessments.

1. INTRODUCTION

Echo cancellers are used to reduce echoes in a wide range of applications, such as TV conference systems and hands-free telephones. To realistic TV conferencing, multi-channel audio, at least stereophonic, is essential. For stereophonic teleconferencing, stereophonic acoustic echo cancellers (SAEC's) [1–4] have been studied.

SAEC's have a fundamental problem in which their filter coefficients cannot have an unique solution [1]. SAEC's with pre-processing [2, 3] are good candidates for solving this problem. As pre-processing filters, periodically time-varying two-tap FIR Filter [2] and randomly time-varying first-order all-pass filter (APF) [3] are proposed.

For periodically time-varying pre-processors, analytical framework was introduced which results in better pre-processor with second-order APF [4]. For randomly time-varying pre-processors, however, neither higher-order pre-processors nor analyses are known because of its complex non-stationary behavior. Many computer simulations for the convergence speed assessment and subjective tests for the sound quality evaluation make the development of randomly time-varying pre-processors difficult.

This paper proposes a stereophonic acoustic echo canceller with second-order pre-processing filters. Section 2 reviews the conventional first-order pre-processor, followed by second-order pre-processing filters. A quantitative approach for the convergence speed assessment is also introduced in Section 4. Simulation results show its performance and validate analyses.

2. SAEC WITH PRE-PROCESSING

Figure 1 shows a teleconferencing using an SAEC with pre-processing. This echo canceller consists of four adaptive filters corresponding to four echo paths from two loudspeakers to two microphones. Each adaptive filter estimates the corresponding echo path.

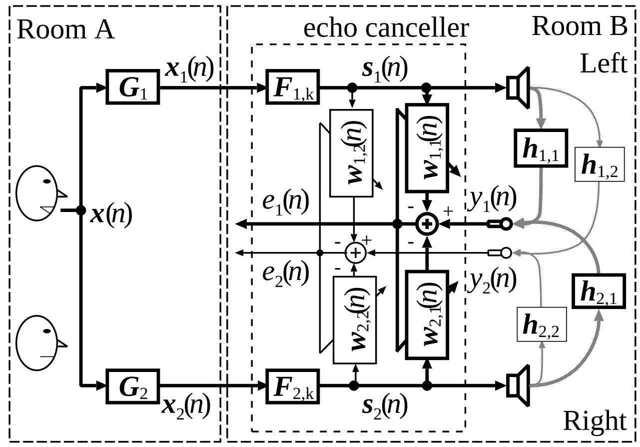

Fig. 1. SAEC with pre-processing

The far-end signal $\mathbf{x}_i(n)$ in the i-th channel at time index n is generated from a talker speech $\mathbf{x}(n)$ by passing RoomA impulse response $\mathbf{G}_i$. Pre-processing filters $\mathbf{F}_{i,k}$ in channel i at k-th state alter far-end signals $\mathbf{x}_i(n)$ to generate reference signals $\mathbf{s}_i(n)$. $\mathbf{s}_i(n)$ passes an echo path $\mathbf{h}_{i,j}$ from the i-th loudspeaker to the j-th microphone and become an echo $y_j(n)$. Similarly, adaptive filters $\mathbf{w}_{i,j}(n)$ generates echo replicas.

In the conventional SAEC with randomly time-varying pre-processor [3], a first-order APF

$$A(z, n) = \frac{z^{-1} - r(n)}{1 - r(n)z^{-1}} \tag{1}$$

is used. A random variable $r(n)$ is so updated by

$$r(n+1) = r(n) + random(n) \tag{2}$$

$$r(n+1) = r_{max} \quad if \;\; r(n+1) > r_{max} \tag{3}$$

$$r(n+1) = r_{min} \quad if \;\; r(n+1) < r_{min} \tag{4}$$

as to satisfy $r_{min} \le r(n) \le r_{max}$. Because of the sound quality, (r_{min}, r_{max}) is chosen as a non-positive value, e.g. $(-0.9, 0)$. Though larger $random(n)$ results in faster convergence, the range of $random(n)$ should be small for negligible sound distortion.

Figure 2 shows the group delay of the pre-processor APF. Curves correspond to several $r(n)$'s between $r_{min} = -0.9$ and $r_{max} = 0$. The group delay in lower frequency is small for the

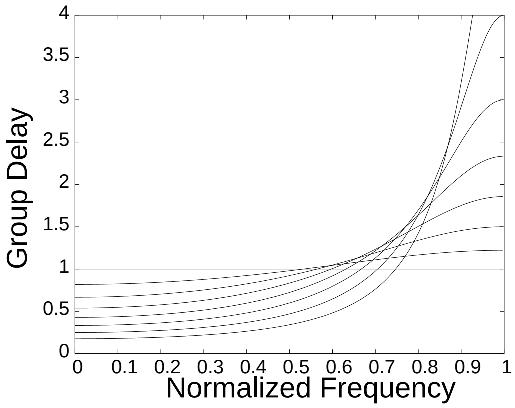

Fig. 2. Group delay.

sound quality. For the convergence speed, however, larger group delay in lower frequency is preferable. By using 1st-order APF, larger group delay without noticeable sound distortion is impossible.

3. PRE-PROCESSOR WITH SECOND-ORDER APF

To overcome a trade-off between the convergence speed and the sound quality [4], a second-order APF

$$A(z) = \frac{r^2(n) - 2r(n)cos\theta(n)z^{-1} + Z^{-2}}{1 - 2r(n)cos\theta(n)z^{-1} + r^2(n)z^{-2}} \tag{5}$$

is used as a pre-processor. There are two parameters, $r(n)$ and $\theta(n)$, which control the pole location of the APF. For simple control and easy tuning, the random parameters $(r(n), \theta(n))$ are controlled by one of the following manners.

- 2APF(r): Variable $r(n)$, fixed $\theta(n)$

$$r(n+1) = r(n) + random(n) \tag{6}$$
$$r(n+1) = r_{max} \quad if \;\; r(n+1) > r_{max} \tag{7}$$
$$r(n+1) = r_{min} \quad if \;\; r(n+1) < r_{min} \tag{8}$$
$$\theta(n) = \theta_0 \tag{9}$$

- 2APF(θ): Variable $\theta(n)$, fixed $r(n)$

$$\theta(n+1) = \theta(n) + \pi random(n) \tag{10}$$
$$\theta(n+1) = \theta_{max} \quad if \;\; \theta(n+1) > \theta_{max} \tag{11}$$
$$\theta(n+1) = \theta_{min} \quad if \;\; \theta(n+1) < \theta_{min} \tag{12}$$
$$r(n) = r_0 \tag{13}$$

- 2APF(r, θ): Variable $r(n), \theta(n)$.

$$u(n+1) = u(n) + random(n) \tag{14}$$
$$u(n+1) = u_{max} \quad if \;\; u(n+1) > u_{max} \tag{15}$$
$$u(n+1) = u_{min} \quad if \;\; u(n+1) < u_{min} \tag{16}$$
$$\theta(n+1) = \pi u(n+1) \tag{17}$$
$$r(n+1) = Ru(n+1) \tag{18}$$
$$\theta_{max} = \pi u_{max}, \theta_{min} = \pi u_{min} \tag{19}$$
$$r_{max} = Ru_{max}, r_{min} = Ru_{min} \tag{20}$$

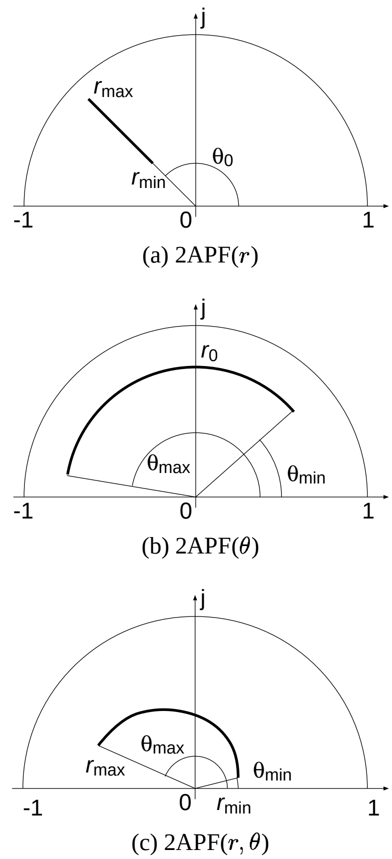

Fig. 3. Pole location for 2nd-order APF.

Figure 3 depicts the pole locations for these pre-processors; the pole randomly works on the bold lines. The group delay for the proposed pre-processors are shown in Fig. 4. Pre-processors 2APF(r) and 2APF(r, θ) provide us with larger group delay in lower frequency. The group delay change by 2APF(θ) is almost same in both lower and higher frequency ranges, while it is small around normalized frequency 0.5.

4. CONVERGENCE ANALYSIS

Assuming the LMS algorithm [5], a merged coefficient vector $\mathbf{w}_i(n)$ is updated by

$$\mathbf{w}_i(n+1) = \mathbf{w}_i(n) + \mu e_i(n)\mathbf{s}(n) \tag{21}$$

where $\mathbf{s}(n)$ and $\mathbf{w}_i(n)$ are defined as

$$\mathbf{s}(n) = \left[\mathbf{s}_1^T(n)\mathbf{s}_2^T(n)\right]^T \tag{22}$$

$$\mathbf{w}_i(n) = \left[\mathbf{w}_{1,i}^T(n)\mathbf{w}_{2,i}^T(n)\right]^T. \tag{23}$$

Introducing a coefficient error vector $\mathbf{c}_i(n)$ and an echo path vector $\mathbf{h}_i$ defined by

$$\mathbf{c}_i(n) = \mathbf{w}_i(n) - \mathbf{h}_i \tag{24}$$

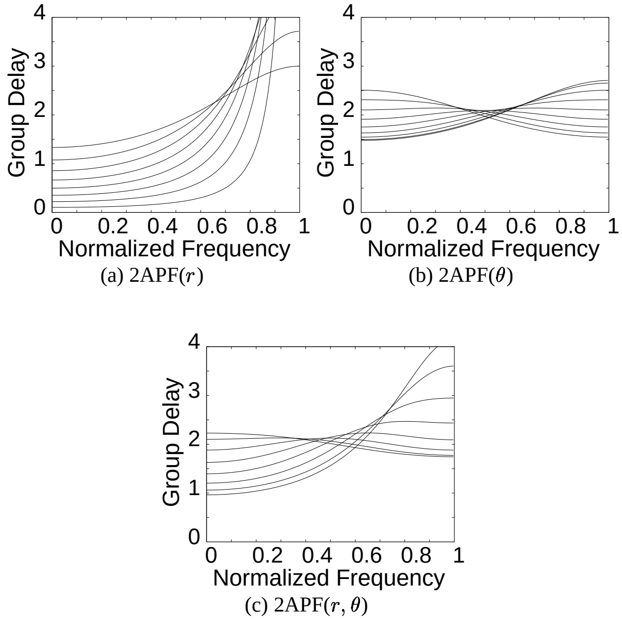

Fig. 4. Group delay for 2nd-order APF.

Table 1. Parameters for pre-processors.

1APF	$r_{min} = -0.9$	$r_{max} = 0$	
2APF (r)	$r_{min} = 0.2$	$r_{max} = 0.9$	$\theta_0 = \pi$
2APF (θ)	$\theta_{min} = 0.2\pi$	$\theta_{max} = \pi$	$r_0 = 0.2$
2AFP (r, θ)	$v_{min} = 0.2$	$v_{max} = 1.0$	$R = 0.35$

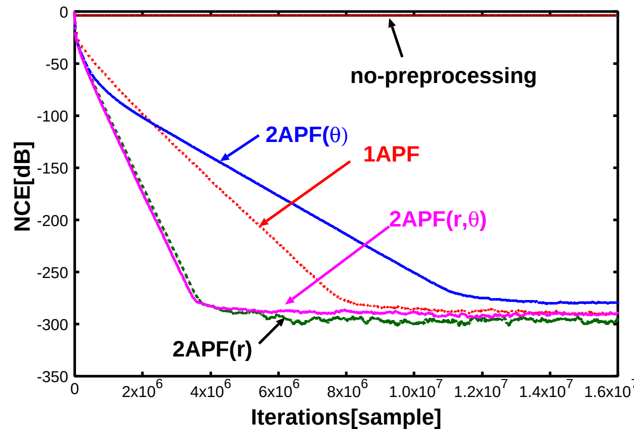

Fig. 5. Convergence of normalized coefficient error.

$$\mathbf{h}_i = \left[\mathbf{h}_{1,i}^T \mathbf{h}_{2,i}^T\right]^T \tag{25}$$

and taking ensemble average, (21) becomes

$$E[\mathbf{c}_i(n+1)] = (\mathbf{I} - \mu\mathbf{R}(n))E[\mathbf{c}_i(n)]. \tag{26}$$

Note that a correlation matrix $\mathbf{R}(n)$ defined by

$$\mathbf{R}(n) = E[\mathbf{s}(n)\mathbf{s}^T(n)] \tag{27}$$

depends on the time index n because of the time-varying pre-processor.

If $\mathbf{s}(n)$ is stationary, (26) can be simplified as

$$\mathbf{v}_i(n) = (\mathbf{I} - \mu\mathbf{\Lambda})^n \mathbf{v}(0) \tag{28}$$

$$\mathbf{v}_i(n) = \mathbf{Q}^T E[\mathbf{c}_i(n)] \tag{29}$$

$$\mathbf{\Lambda} = \mathbf{Q}^T \mathbf{R}\mathbf{Q}. \tag{30}$$

In this case, the convergence speed is evaluated by the eigenvalue spread $\frac{\lambda_{max}}{\lambda_{min}}$. For non-stationary signals, however, simplification as in (28) cannot be applied directly.

A time-averaged estimates of $\mathbf{R}$, though, leads us to a good approximation. The averaged version is calculated by

$$\hat{\mathbf{R}} = \frac{1}{M}\sum_{n=0}^{M-1} \mathbf{s}(n)\mathbf{s}^T(n). \tag{31}$$

Similar to a monaural and a stationary case, the eigenvalue spread suggests the convergence time to the final value. Examples and comparison with the ensemble average will be shown by the computer simulations.

For users of SAEC's, however, the convergence speed in an early stage, e.g. the convergence time until $-50dB$ of the normalized coefficient error (NCE), is more important. The eigenvalue spread is not suitable for this purpose. A numerically calculated version of $\|\mathbf{v}_i(n)\|^2$ with the time-averaged $\mathbf{R}$ gives us a good overview. $\|\mathbf{v}_i(n)\|^2$ can be calculated by (28) with (24) and the eigenvalues from (31). Though an estimated room impulse response is required, that from a similar-sized room will help the analyses.

5. COMPUTER SIMULATIONS

Simulations have been carried out to show the performance of the SAEC with the proposed pre-processors and to validate the analyses. The impulse responses for both Room A (far end) and Room B (near end) are 20-th order Butterworth low-pass filters. Different cut-off frequencies are used for each paths. The single-talk situation is assumed; only one talker is speaking in Room A. No additive noises are introduced. White Gaussian signals are used as the talker signals.

As adaptive filters, 64-tap FIR adaptive filters are used. The normalized LMS algorithm [6], which is identical to the LMS algorithm for stationary signals, is used as an adaptation algorithm. Though non-stationary pre-processor is used, the pre-processor output power $\sum_{i=1}^{2} |\mathbf{s}_i(n)|^2$ should be constant. The step-size is 0.5.

The parameters for pre-processors are so chosen as to achieve the same sound quality for real speech signals. Subjective tests have been performed in order to evaluate the sound quality. Table 1 shows the parameters. The range of a random generator is $-0.02 < random(n) < 0.02$.

Figure 5 compares the normalized coefficient error (NCE) de-

Table 2. Eigenvalue spread.

	TA	EA(1)	EA(2)
1APF	4871.7	3.4045×10^{13}	5545.7
2APF(r)	1786.9	3.4905×10^{13}	1707.2
2APF(θ)	1.0674×10^4	3.4613×10^{13}	1.3232×10^4
2APF(r, θ)	2914.5	3.4602×10^{13}	3768.4

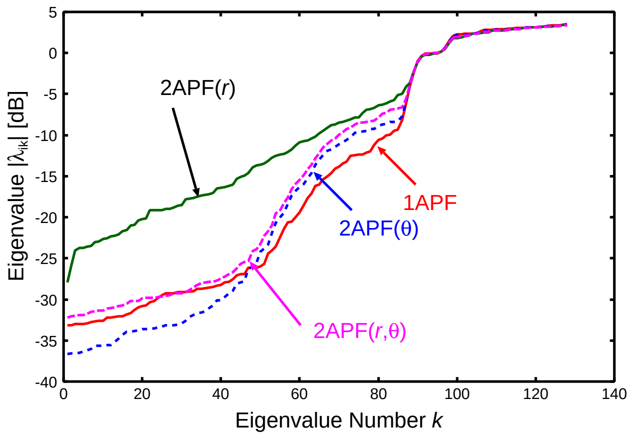

Fig. 6. Eigenvalue.

fined by

$$NCE(n) = 10 * log_{10} \frac{\sum_{j=1}^{2} \|\mathbf{h}_{j,i} - \mathbf{w}_{j,i}(n)\|^2}{\sum_{j=1}^{2} \|\mathbf{h}_{j,i}\|^2}. \quad (32)$$

The second-order pre-processors ("2AFP(r)", "2APF(r,θ)") converges more than twice as fast as the first-order one ("1APF"). In an early stage, "2APF(θ)" converges faster than "1APF". However, it becomes slower in a final stage. Its residual error is slightly larger than that of the others.

Analyses using estimated correlation matrices $\mathbf{R}$ have also be performed. Averaging for $\mathbf{R}$ was carried out by the following manners: the time average "TA," the ensemble average with multiple $\mathbf{x}(n)$ "EA(1)," the ensemble average with multiple $\mathbf{x}(n)$ and $random(n)$ "EA(2)." The time averaging was carried out for $M = 10000$ samples. The ensemble averaging was for 3000 independent runs. Results for the ensemble average with multiple $random(n)$ are not shown because these results heavily depend on the time index n and are not suitable for the analyses.

Table 2 compares the eigenvalue spread for pre-processed signals. Results by "TA" agree with "EA(2)." The order of the eigenvalue spread by these two methods; "2APF(r)", "2APF(r, θ)", "1APF", "2APF(θ)", agrees with the convergence speed shown in Fig. 5. Results from "EA(1)" do not agree with the simulations. Because of its simplicity, time averaging is good choice for these analyses.

The eigenvalue spread, however, explain nothing for the behavior in the early stage until $NCE(n)$ reaches $-50dB$. Examining the all eigenvalues also turned out not to be enough as shown in Fig. 6; it simply suggests that "2APF(r)" is fastest.

Figure 7 demonstrates $\|\mathbf{v}_i(n)\|^2$ estimated by using the time-averaged $\mathbf{R}$. It agrees with the NCE in Fig. 5 in an early stage. The convergence of "2APF(θ)" is faster than "1APF" untile 2×10^5 sample, and then becomes slower.

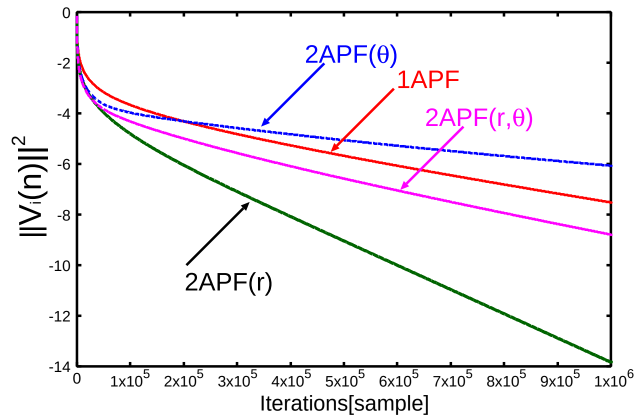

Fig. 7. Analytical results.

6. CONCLUSION

This paper has proposed a stereophonic acoustic echo canceller with randomly time-varying second-order pre-processing filters. A quantitative approach for the convergence speed assessment is also introduced. Analyses based on time-averaged correlation matrix provide up a good overview on the convergence of SAEC's. Simulation results show almost twice faster convergence speed than that of the conventional SAEC. Analyses also agrees with the simulations.

7. REFERENCES

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

[2] A. Sugiyama, Y. Joncour, and A. Hirano, "A stereo echo canceler with correct echo-path identification based on an input-sliding technique," *IEEE Trans. SP*, vol. 49, no. 11, pp. 2577–2587, Nov. 2001.

[3] M. Ali, "Stereophonic acoustic echo cancellation system using time-varying all-path filtering for signal decorrelation," *Proc. of ICASSP'98*, pp. 3689–3692, May 1998.

[4] A. Hirano, K. Nakayama, and K. Watanabe, "Convergence analysis of stereophonic echo canceller with pre-processing — relation between pre-processing and convergence —," *Proc. of ICASSP '99*, pp. 861–864, Mar. 1999.

[5] B. Widrow and S. D. Stearns, "Adaptive noise canceling: Principles and applications," *Proc. of IEEE*, vol. 63, no. 12, pp. 1692–1716, Dec 1975.

[6] J. Nagumo and A. Noda, "A learning method for system identification," *IEEE Trans. AC*, vol. 12, no. 3, pp. 282–287, Mar 1967.