Improvement in Variance Estimation in Variable-Step-Size Shared-Error NLMS Algorithm for Acoustic Echo and Noise Canceller

Kenta Iwai*
* Osaka Sangyo University, Japan
E-mail: iwai@ise.osaka-sandai.ac.jp

***Abstract*—In this paper, we investigate the variance estimation method used in the variable-step-size shared-error normalized least mean square algorithm for the acoustic echo and noise canceller. The variance of the shared-error signal is typically estimated using a recursive equation, as this signal is readily available. In contrast, the disturbances of the acoustic echo canceller and adaptive noise canceller cannot be directly extracted from the microphone input, which contains both the desired signal and disturbances. Consequently, their variances are approximated by subtracting the correlated components from the variance of the shared-error signal. However, owing to misadjusted cross-correlation vectors, the variable-step-size parameters often become inappropriate. To address this, the scaling method for the cross-correlation vectors within the variance estimation is modified in this paper. Simulation results demonstrate that the proposed method improves stability during and after the double-talk period. Nevertheless, the update of the adaptive digital filters is continued under the double-talk period, degrading the overall echo and noise reduction performance.**

I. Introduction

In this paper, we address the variance estimator problem in the variable-step-size shared-error normalized least mean square (VSS-SENLMS) algorithm used in the acoustic echo and noise canceller (AENC) [1], [2], [3], [4], [5]. The shared-error least normalized mean square (SENLMS) algorithm for the AENC was first proposed in [6]. This AENC consists of the acoustic echo canceller (AEC) and Widrow's adaptive noise canceller (ANC) [7], both implemented with time-domain adaptive digital filters. These filters are jointly optimized using the shared-error signal. The SENLMS-based AENC offers the advantage of lower computational complexity compared to deep neural network-based systems [8], [9], [10], [11].

To enhance robustness under nonstationary noise and double-talk conditions, a variable-step-size version of the SENLMS algorithm has been developed [12]. It is based on Benesty's non-parametric variable-step-size NLMS (NP-VSS-NLMS) algorithm [13], [14], derived via the minimization of the mean square of the *a posteriori* error.

In the VSS-SENLMS algorithm, step-size parameters are computed from the variances of several signals. The variance of the error signal can be easily calculated because the error signal can be directly obtained. In contrast, the variances of disturbance and near-end signals cannot be calculated because these signals cannot be obtained directly. Therefore, these variances are estimated using cross-correlation vectors. However, the misadjusted cross-correlation vectors lead to negative variance estimates, which cause instability in the step-size control.

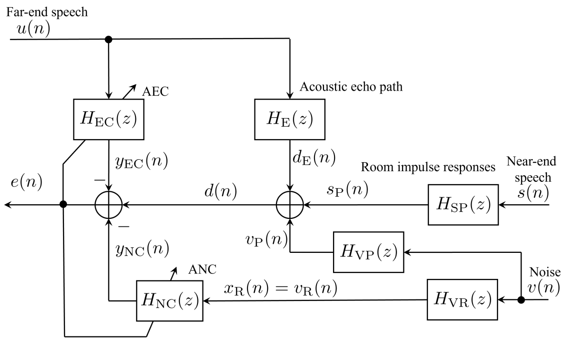


Fig. 1: Block diagram of AENC based on SENLMS algorithm.

In this paper, we propose a modified scaling method for the cross-correlation vectors in the variance estimation to address this issue. Mathematical analysis is performed to explain the limitation of the conventional method, followed by simulations to evaluate the effectiveness of the proposed modification.

II. AENC based on shared-error NLMS algorithm with variable step size

A. Shared-error NLMS algorithm [6]

The AENC based on the SENLMS algorithm employs two adaptive digital filters. The block diagram of this system is shown in Fig. 1. As illustrated in Fig. 1, it includes the time-domain AEC and Widrow's ANC [7]. These adaptive digital filters are updated using the shared-error signal, enabling joint optimization.

The SENLMS algorithm is based on the normalized least mean square (NLMS) algorithm [15], [16], [17], where the error signal is replaced with the shared-error signal $e(n)$. The update equation of the filter coefficient vectors is derived by minimizing the mean square value of the shared-error signal defined as

$$e(n) = d_{\mathrm{E}}(n) - y_{\mathrm{EC}}(n) + v_{\mathrm{P}}(n) - y_{\mathrm{NC}}(n) + s_{\mathrm{P}}(n), \quad (1)$$

where $d_{\mathrm{E}}(n)$, $v_{\mathrm{P}}(n)$, and $s_{\mathrm{P}}(n)$ are the acoustic echo, background noise, and near-end speech signal, and $y_{\mathrm{EC}}(n) = \mathbf{h}_{\mathrm{EC}}^{\mathrm{T}}(n)\mathbf{u}(n)$, and $y_{\mathrm{NC}}(n) = \mathbf{h}_{\mathrm{NC}}^{\mathrm{T}}(n)\mathbf{x}_{\mathrm{R}}(n)$ are the outputs of the AEC and ANC, respectively. Moreover, $\mathbf{h}_{\mathrm{EC}}(n)$ and

$\mathbf{h}_{\rm NC}(n)$ are the filter coefficient vectors of the AEC and ANC, and $\mathbf{u}(n)$ and $\mathbf{x}_{\rm R}(n)$ are the reference signal vectors of the AEC and ANC, $\|\cdot\|$ is the l_2 norm of the vector, respectively. Taking the gradient of the cost functions with respect to the filter coefficient vectors yields the following update equations:

$$\mathbf{h}_{\rm EC}(n+1) = \mathbf{h}_{\rm EC}(n) + \alpha_{\rm EC}\frac{e(n)\mathbf{u}(n)}{\|\mathbf{u}(n)\|^2+\beta_{\rm EC}}, \quad (2)$$

$$\mathbf{h}_{\rm NC}(n+1) = \mathbf{h}_{\rm NC}(n) + \alpha_{\rm NC}\frac{e(n)\mathbf{x}_{\rm R}(n)}{\|\mathbf{x}_{\rm R}(n)\|^2+\beta_{\rm NC}}, \quad (3)$$

where $\alpha_{\rm EC}$ and $\alpha_{\rm NC}$ are the step size parameters of the AEC and ANC, and $\beta_{\rm EC}$ and $\beta_{\rm NC}$ are the regularization parameters of the AEC and ANC, respectively.

One challenge with the SENLMS algorithm is the tuning of the two step-size parameters to prevent divergence. Because the shared-error signal includes contributions from both filters, divergence in one component (AEC or ANC) can destabilize the entire system. This makes the parameter tuning particularly sensitive, especially under nonstationary noise or double-talk conditions.

B. Variable-step-size SENLMS algorithm [12]

To address the instability caused by fixed-step-size parameters in the SENLMS algorithm, a variable-step-size version (VSS-SENLMS) was proposed in [12]. This version builds upon the non-parametric variable-step-size NLMS (NP-VSS-NLMS) algorithm introduced by Benesty et al. [13], [14], and achieves the step-size control by minimizing the mean square of the *a posteriori* error. The *a posteriori* error is defined as

$$\varepsilon(n) = d_{\rm E}(n) - \mathbf{h}_{\rm EC}^{\rm T}(n+1)\mathbf{u}(n) + v_{\rm P}(n) - \mathbf{h}_{\rm NC}^{\rm T}(n+1)\mathbf{x}_{\rm R}(n) + s_{\rm P}(n). \quad (4)$$

The adaptive filters are updated with time-varying step-size parameters as follows:

$$\mathbf{h}_{\rm EC}(n+1) = \mathbf{h}_{\rm EC}(n) + \alpha_{\rm v,EC}(n)e(n)\mathbf{u}(n), \quad (5)$$

$$\mathbf{h}_{\rm NC}(n+1) = \mathbf{h}_{\rm NC}(n) + \alpha_{\rm v,NC}(n)e(n)\mathbf{x}_{\rm R}(n). \quad (6)$$

Substituting (5) and (6) into (4), we obtain the mean square value of the *a posteriori* error as

$$\mathbb{E}\left[\varepsilon^2(n)\right] = \sigma_e^2(n)\left[1-\|\mathbf{u}(n)\|^2\alpha_{\rm v,EC}(n) - \|\mathbf{x}_{\rm R}(n)\|^2\alpha_{\rm v,NC}(n)\right]^2, \quad (7)$$

where $\sigma_e^2(n) = \mathbb{E}\left[e^2(n)\right]$, and $\|\mathbf{u}(n)\|^2$ and $\|\mathbf{x}_{\rm R}(n)\|^2$ are treated as constant at time n. If we set $\mathbb{E}\left[\varepsilon^2(n)\right] = \mathbb{E}\left[s_{\rm P}^2(n)\right] = \sigma_s^2(n)$, we obtain the constraint:

$$\|\mathbf{u}(n)\|^2\alpha_{\rm v,EC}(n) + \|\mathbf{x}_{\rm R}(n)\|^2\alpha_{\rm v,NC}(n) = 1-\sqrt{\frac{\sigma_s^2(n)}{\sigma_e^2(n)}}. \quad (8)$$

This equation involves two unknowns, $\alpha_{\rm v,EC}(n)$ and $\alpha_{\rm v,NC}(n)$, and thus is underdetermined.

To solve this, in [12], Benesty's NP-VSS-NLMS algorithm directly adopted for estimating $\alpha_{\rm v,EC}(n)$: [1]

$$\alpha_{\rm v,EC}(n) = \frac{\alpha_{\rm VSS,EC}(n)}{\|\mathbf{u}(n)\|^2} = \frac{1}{\|\mathbf{u}(n)\|^2}\left[1-\sqrt{\frac{\sigma_{v,\rm EC}^2(n)}{\sigma_e^2(n)}}\right], \quad (9)$$

where $\sigma_{v,\rm EC}^2(n) = \mathbb{E}\Big[\{v_{\rm P}(n) - y_{\rm NC}(n) + s_{\rm P}(n)\}^2\Big]$. Substituting (9) into (8) yields

$$\alpha_{\rm v,NC}(n) = \frac{\alpha_{\rm VSS,NC}(n)}{\|\mathbf{x}_{\rm R}(n)\|^2} = \frac{1}{\|\mathbf{x}_{\rm R}(n)\|^2}\left[1-\sqrt{\frac{\sigma_s^2(n)}{\sigma_e^2(n)}} - \alpha_{\rm VSS,EC}(n)\right]. \quad (10)$$

To compute (9) and (10), the variances $\sigma_e^2(n)$, $\sigma_s^2(n)$ and $\sigma_{v,\rm EC}^2(n)$, must be estimated. The shared-error signal $e(n)$ can be directly obtained, and its variance is recursively estimated as [13]

$$\sigma_e^2(n) = \lambda\sigma_e^2(n-1) + (1-\lambda)\,e^2(n), \quad (11)$$

where $\lambda(0 << \lambda < 1)$ is the smoothing factor. In contrast, $s_{\rm P}(n)$ and $v_{\rm EC}(n) = v_{\rm P}(n) - y_{\rm NC}(n) + s_{\rm P}(n)$ cannot be directly observed, and their variances $\sigma_s^2(n)$ and $\sigma_{v,\rm EC}^2(n)$ are estimated using the following [18]:

$$\sigma_s^2(n) = \sigma_e^2(n) - \frac{\|\mathbf{r}_{eu}(n)\|^2}{\sigma_u^2(n)} - \frac{\|\mathbf{r}_{ex}(n)\|^2}{\sigma_x^2(n)}, \quad (12)$$

$$\sigma_{v,\rm EC}^2(n) = \sigma_e^2(n) - \frac{\|\mathbf{r}_{eu}(n)\|^2}{\sigma_u^2(n)}, \quad (13)$$

$$\sigma_x^2(n) = \lambda\sigma_x^2(n-1) + (1-\lambda)\,x_{\rm R}^2(n), \quad (14)$$

$$\sigma_u^2(n) = \lambda\sigma_u^2(n-1) + (1-\lambda)\,u^2(n), \quad (15)$$

$$\mathbf{r}_{eu}(n) = \lambda\mathbf{r}_{eu}(n-1) + (1-\lambda)\,e(n)\mathbf{u}(n), \quad (16)$$

$$\mathbf{r}_{ex}(n) = \lambda\mathbf{r}_{ex}(n-1) + (1-\lambda)\,e(n)\mathbf{x}_{\rm R}(n). \quad (17)$$

By adapting $\alpha_{\rm v,EC}(n)$ and $\alpha_{\rm v,NC}(n)$ using (9) and (10), the VSS-SENLMS algorithm can suppress divergence while maintaining effective echo and noise reduction performance.

However, it has been reported in [19] that under double-talk conditions, the inaccurate estimation of $\sigma_s^2(n)$ and $\sigma_{v,\rm EC}^2(n)$ can still cause instability. This underscores the need for further investigation into improved variance estimation methods.

III. Investigation of Variance Estimation

As discussed in Sect. II, the estimation of $\sigma_s^2(n)$ and $\sigma_{v,\rm EC}^2(n)$ using (12) and (13) is prone to low accuracy. This is primarily because these equations rely on scaled cross-correlation terms, which are sensitive to the wrong estimation of the variances and can lead to the instability of the variable-step-size parameters. In particular, the numerator of each term involves cross-correlation vectors scaled by the squared l_2 norms of the reference signals, potentially introducing mismatches in time.

[1] $v_{\rm P}(n)$, $y_{\rm NC}(n)$, and $s_{\rm P}(n)$ are the disturbances for the AEC. Hence, $v_{\rm EC}(n) = v_{\rm P}(n) - y_{\rm NC}(n) + s_{\rm P}(n)$.

To illustrate this issue, consider expanding the term $\|\mathbf{r}_{eu}(n)\|^2$ from (13) as follows:

$$\|\mathbf{r}_{eu}(n)\|^2 = \lambda^2 \|\mathbf{r}_{eu}(n-1)\|^2 + \lambda(1-\lambda)\, e(n)\mathbf{r}_{eu}^{\mathrm{T}}(n-1)\mathbf{u}(n) + \lambda(1-\lambda)\, e(n)\mathbf{u}^{\mathrm{T}}(n)\mathbf{r}_{eu}(n-1) + (1-\lambda)^2\, e^2(n)\|\mathbf{u}(n)\|^2. \quad (18)$$

Dividing $\|\mathbf{r}_{eu}(n)\|^2$ by $\sigma_u^2(n)$ yields

$$\frac{\|\mathbf{r}_{eu}(n)\|^2}{\sigma_u^2(n)} = \lambda^2 \frac{\|\mathbf{r}_{eu}(n-1)\|^2}{\sigma_u^2(n)} + \lambda(1-\lambda)\, \frac{e(n)\mathbf{r}_{eu}^{\mathrm{T}}(n-1)\mathbf{u}(n)}{\sigma_u^2(n)} + \lambda(1-\lambda)\, \frac{e(n)\mathbf{u}^{\mathrm{T}}(n)\mathbf{r}_{eu}(n-1)}{\sigma_u^2(n)} + (1-\lambda)^2 \frac{e^2(n)\|\mathbf{u}(n)\|^2}{\sigma_u^2(n)}. \quad (19)$$

Here, it is evident that $\mathbf{r}_{eu}(n-1) \propto \mathbf{u}(n-1)$, whereas the current scaling uses $\sigma_u^2(n)$, introducing mismatches in time. In other words, $\sigma_u^2(n)$ is not synchronized with $\mathbf{r}_{eu}(n-1) \propto \mathbf{u}(n-1)$. This time lag can degrade the estimation accuracy of the cross-correlation vectors.

To mitigate this issue, in this paper, we here propose a modified scaling approach where the cross-correlation vector is normalized using the standard deviation $\sigma_u(n)$ in the same time frame. Specifically, the estimation is redefined as

$$\sigma_{v,\mathrm{EC}}^2(n) = \sigma_e^2(n) - \|\mathbf{r}_{eu}(n)\|^2, \quad (20)$$

$$\mathbf{r}_{eu}(n) = \lambda \mathbf{r}_{eu}(n-1) + (1-\lambda)\, e(n)\frac{\mathbf{u}(n)}{\sigma_u(n)}. \quad (21)$$

Expanding $\|\mathbf{r}_{eu}(n)\|^2$ under this modification, we obtain

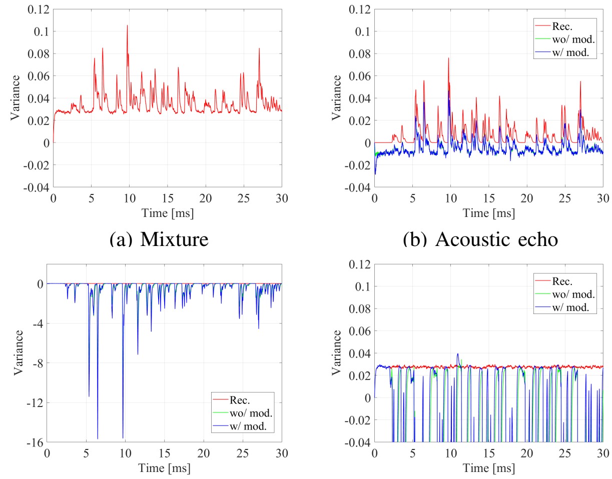
$$\|\mathbf{r}_{eu}(n)\|^2 = \lambda^2 \|\mathbf{r}_{eu}(n-1)\|^2 + \lambda(1-\lambda)\, \frac{e(n)\mathbf{r}_{eu}^{\mathrm{T}}(n-1)\mathbf{u}(n)}{\sigma_u(n)} + \lambda(1-\lambda)\, \frac{e(n)\mathbf{u}^{\mathrm{T}}(n)\mathbf{r}_{eu}(n-1)}{\sigma_u(n)} + (1-\lambda)^2 \frac{e^2(n)\|\mathbf{u}(n)\|^2}{\sigma_u^2(n)}. \quad (22)$$

As can be seen from (22), $\sigma_u^2(n)$ is synchronized with $\mathbf{u}(n)$. Using (21), we can expand the second term of (22) as

$$\frac{e(n)\mathbf{r}_{eu}^{\mathrm{T}}(n-1)\mathbf{u}(n)}{\sigma_u(n)} = \lambda e(n)\mathbf{r}_{eu}^{\mathrm{T}}(n-2)\frac{\mathbf{u}(n)}{\sigma_u(n)} + (1-\lambda)\, e(n)e(n-1)\frac{\mathbf{u}^{\mathrm{T}}(n-1)\mathbf{u}(n)}{\sigma_u(n)\sigma_u(n-1)}. \quad (23)$$

In (23), $\sigma_u(n)$ is synchronized with $\mathbf{u}(n)$. For the third term of (22), $\sigma_u(n)$ is synchronized with $\mathbf{u}(n)$.

Fig. 2 shows the dynamics of the variances of the mixed signal, far-end speech signal, and background noise at the reference microphone estimated using (12)–(17). In Fig. 2, "Rec." depicts the variance estimated using the recursive equation (11). In the case of using (11), it is assumed that the acoustic echo and background noise are separately obtained. Hence, the estimate obtained using (11) is the ideal value. From Fig. 2, the variance estimated using the cross-correlation is negative during both single- and double-talk periods. On the other hand, the variance estimated using the recursive equation is larger than zero. Moreover, the variance of the acoustic echo is negative when the far-end speech is absent. Here, from Fig. 2(d), the variance of the background noise estimated using the modified equation converges to the ideal value faster than that estimated using the conventional equation, although the variance estimated using the modified equation is frequently negative. Therefore, the modified equation (22) is appropriate for estimating the variance.



(a) Mixture (b) Acoustic echo (c) Background noise (d) Background noise (expanded view)

Fig. 2: Variances of each signal obtained by primary microphone.

TABLE I: Simulation conditions.

Far-end speech	Female speech (4507)[22]
Near-end speech	Male speech (1188)[22]
Background noise	White noise
Signal-to-noise ratio (SNR)	-5 dB
Sampling frequency	16 kHz
Tap length of each path	1024
Tap length of ADF of AEC/ANC	1024 / 1024
Regularization parameter	1.0×10^{-6}

IV. Computer Simulation

To evaluate the effectiveness of the proposed variance estimation method described in Sect. III, we conducted a computer simulation to compare the conventional and modified approaches. In this simulation, the impulse responses of each path were generated by the RIR simulator [20]. The arrangement of the sound source and microphone is shown in Fig. 3. Here, the direct-to-reverberant ratio (DRR) [21] of the acoustic echo path was set to about 26 dB. Simulation conditions are shown in Table I. The smoothing factor λ was set to $\lambda = 1 - 3/2N$ with the tap length of the adaptive digital filter (ADF) N. The time waveforms of the acoustic echo, near-end speech, and background noise are shown in Fig. 4. It can be seen from Fig. 4 that the double-talk period was approximately between 10 and 12 s.

The time waveform of the error signal is shown in Fig. 5, which indicates that the system with the modified estimation method achieves the decrease of the error signal and a faster

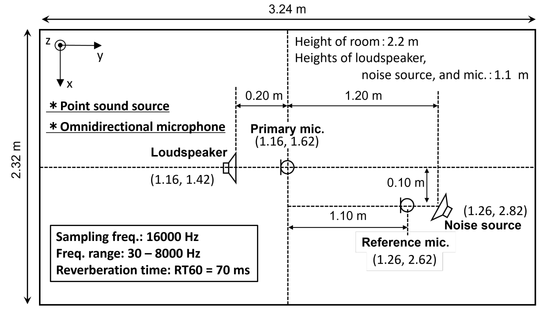
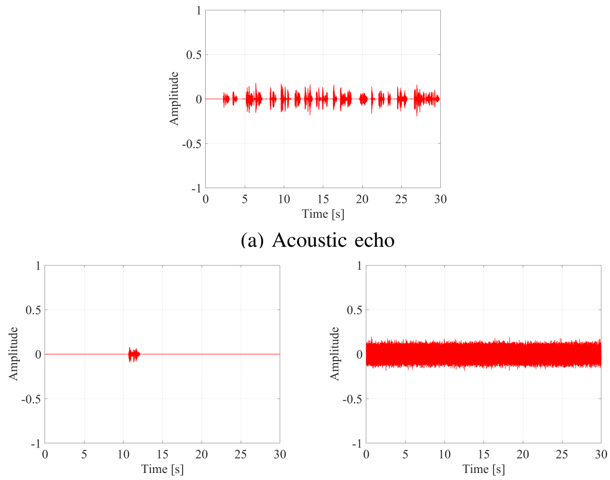


Fig. 3: Arrangement of sound source and microphone in generation of room impulse response.



(a) Acoustic echo

(b) Near-end speech signal

(c) Background noise

Fig. 4: Observed signals of primary microphone.

convergence. On the other hand, the update of the adaptive filters is not stopped under double-talk condition, and the adaptive filters are misadjusted. Then, the acoustic echo superimposes on the near-end speech. This is because the accuracy of the variance estimated using the modified equation is not sufficiently high, leading to the misadjustment of the variable-step-size parameters.

The dynamics of each variance is shown in Fig. 6. From Fig. 6, it can be seen that $\sigma^2_{v,\mathrm{EC}}(n)$ estimated using the modified equation becomes quickly asymptotic reaching to $\sigma^2_e(n)$ during the non-speech period in which the background noise only exists. During the non-speech period, $v_{\mathrm{EC}}(n) = v_{\mathrm{P}}(n) - y_{\mathrm{NC}}(n)$ and $e(n) = v_{\mathrm{P}}(n) - y_{\mathrm{NC}}(n)$ are satisfied, then $\sigma^2_{v,\mathrm{EC}}(n)$ ideally becomes $\sigma^2_e(n)$. On the other hand, $\sigma^2_{v,\mathrm{EC}}(n)$ and $\sigma^2_{\mathrm{S}}(n)$ sometimes show extremely small values. It is expected that the second and third terms in (12) and (13), which are related to the cross-correlation vector, will be larger than the variance $\sigma^2_e(n)$. Then, the variance will become negative, and will be outside of the range. In this simulation, the variance is forced to be zero if the estimated variance becomes negative.

From the above results, the effectiveness of the modified variance estimation was demonstrated.

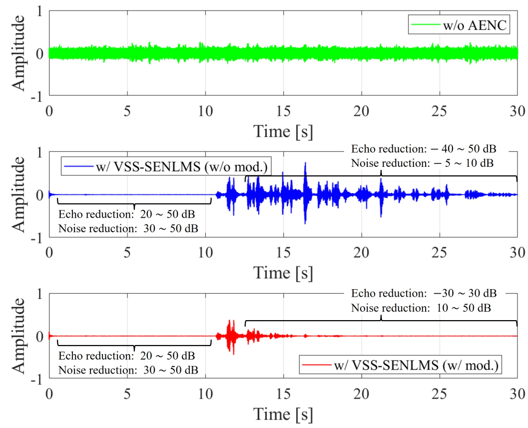
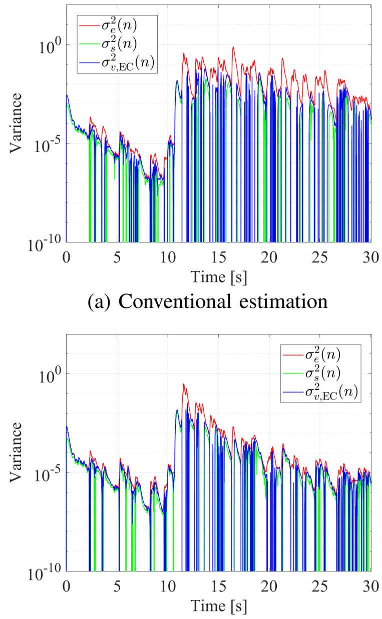


Fig. 5: Time waveform of error signal.



(a) Conventional estimation

(b) Proposed estimation

Fig. 6: Dynamics of variance.

V. Conclusion

In this study, we investigated the variance estimation method used in the VSS-SENLMS algorithm for AENC. From the mathematical analysis, the proposed variance estimation method modifies the scaling of the cross-correlation vectors. Consequently, the variance is synchronized with the cross-correlation vectors. Simulation results demonstrate that the proposed method improves stability during and after the double-talk period. Nevertheless, the update of the adaptive digital filters is continued under the double-talk period, degrading the echo and noise reduction performance. As our future work, we plan to investigate a variance estimation technique that does not rely on cross-correlation vectors, aiming to further improve the robustness of the VSS-SENLMS algorithm.

Acknowledgment

This work was supported by JSPS KAKENHI (21K17773).

References

[1] S. Gustafsson, R. Martin, and P. Vary, “Combined acoustic echo control and noise reduction for hands-free telephony,” *Signal Process.*, vol. 64, no. 1, pp. 21–32, 1998.

[2] S. Park, C. Cho, C. Lee, and D. Youn, “On integrating acoustic echo and noise cancellation systems for hands-free telephony,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2001, pp. 961–964 vol.2.

[3] M. Fukui, S. Shimauchi, Y. Hioka, A. Nakagawa, and Y. Haneda, “Acoustic echo and noise canceller for personal hands-free video IP phone,” *IEEE Trans. Consum. Electron.*, vol. 62, no. 4, pp. 454–462, 2016.

[4] S. Ruiz, T. van Waterschoot, and M. Moonen, “Distributed combined acoustic echo cancellation and noise reduction in wireless acoustic sensor and actuator networks,” *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 30, pp. 534–547, 2022.

[5] K. Nathwani, “Joint acoustic echo and noise cancellation using spectral domain Kalman filtering in double-talk scenario,” in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*, 2018, pp. 326–330.

[6] K. Iwai and T. Nishiura, “Acoustic echo and noise canceller using shared-error normalized least mean square algorithm,” in *Proc. 2022 Asia-Pacific Signal Inf. Process. Association Annual Summit Conf. (APSIPA ASC)*, 2022, pp. 281–285.

[7] B. Widrow et al., “Adaptive noise cancelling: Principles and applications,” *Proc. IEEE*, vol. 63, no. 12, pp. 1692–1716, 1975.

[8] C. Zhang, J. Liu, and X. Zhang, “A complex spectral mapping with inplace convolution recurrent neural networks for acoustic echo cancellation,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2022, pp. 751–755.

[9] K. N. Watcharasupat, T. N. T. Nguyen, W.-S. Gan, S. Zhao, and B. Ma, “End-to-end complex-valued multidilated convolutional neural network for joint acoustic echo cancellation and noise suppression,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2022, pp. 656–660.

[10] H. Zhao et al., “A low-latency deep hierarchical fusion network for fullband acoustic echo cancellation,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2023, pp. 1–2.

[11] Y. Liu, Y. Shi, Y. Li, K. Kalgaonkar, S. Srinivasan, and X. Lei, “SCA: Streaming cross-attention alignment for echo cancellation,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2023, pp. 1–5.

[12] K. Iwai and T. Nishiura, “Step size control of shared-error normalized least mean square algorithm for acoustic echo and noise canceller,” in *Proc. 2023 Asia-Pacific Signal Inf. Process. Association Annual Summit Conf. (APSIPA ASC)*, 2023, pp. 1150–1154.

[13] J. Benesty, H. Rey, L. Vega, and S. Tressens, “A nonparametric VSS NLMS algorithm,” *IEEE Signal Process. Lett.*, vol. 13, no. 10, pp. 581–584, 2006.

[14] C. Paleologu, S. Ciochină, J. Benesty, and S. Grant, “An overview on optimized NLMS algorithms for acoustic echo cancellation,” *EURASIP J. Adv. Signal Process.*, vol. 2015, no. 1, 2015. DOI: 10.1186/s13634-015-0283-1.

[15] S. Haykin, *Adaptive Filter Theory*, Fourth. Upper Saddle River, NJ: Prentice Hall, 2001.

[16] A. Zerguine, “Convergence behavior of the normalized least mean fourth algorithm,” in *Proc. Conf. Record Thirty-Fourth Asilomar Conf. Signals Syst. Comput.*, 2000, pp. 275–278 vol.1.

[17] E. Eweda and A. Zerguine, “A normalized least mean fourth algorithm with improved stability,” in *Proc. Conf. Record Forty Fourth Asilomar Conf. Signals Syst. Comput.*, 2010, pp. 1002–1005.

[18] M. Iqbal and S. L. Grant, “Novel variable step size NLMS algorithms for echo cancellation,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2008, pp. 241–244.

[19] K. Iwai and T. Nishiura, “Performance evaluation of acoustic echo and noise canceller with variable-step-size shared-error NLMS algorithm under double-talk conditions,” in *Proc. 2023 Asia-Pacific Signal Inf. Process. Association Annual Summit Conf. (APSIPA ASC)*, 2024, pp. 1–5.

[20] E. Habets, *Room impulse response generator*, the International Audio Laboratories Erlangen, 2010.

[21] J. Eaton, A. H. Moore, P. A. Naylor, and J. Skoglund, “Direct-to-Reverberant Ratio estimation using a null-steered beamformer,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2015, pp. 46–50.

[22] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, “Librispeech: An ASR corpus based on public domain audio books,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2015, pp. 5206–5210.