A ROBUST NLMS ALGORITHM WITH A NOVEL NOISE MODELING BASED ON STATIONARY/NONSTATIONARY NOISE DECOMPOSITION

Akihiko Sugiyama

NEC Common Platform Software Research Laboratories Kawasaki 211-8666, JAPAN

ABSTRACT

This paper proposes a robust NLMS algorithm with a novel noise modeling based on stationary/nonstationary noise decomposition. The ambient noise including the near-end signal is modeled as a weighted sum of the stationary and the nonstationary components. These components are independently estimated with an appropriate time constant for better accuracy. The estimates are weighted by the stationary/nonstationary likelihood before summation. The integrated noise estimate controls the coefficient adaptation stepsize such that it is an upward convex function of the reference input with a noise offset for robustness. Evaluations in a fullband and a subband echo cancellation scenarios show that ERLE has been improved by as much as 40 dB over the algorithm with a conventional noise model in both single- and double-talk sections with no double-talk detection.

Index Terms— Adaptive filter, Robust algorithm, Noise model, Echo cancellation, Double-talk

1. INTRODUCTION

Adaptive filters have widespread applications in communications. These applications include echo cancellers, noise cancellers, and microphone arrays to name a few. Coefficients of an adaptive filter are updated based on the misadjustment defined as the difference between the desirable signal and the filter output. The misadjustment is often disturbed by additive noise and interference. Because the misadjustment is not separately available from the disturbance, coefficients are adapted using the error that consists of the misadjustment and the disturbance. Coefficient adaptation is seriously interfered unless the disturbance is sufficiently small compared to the misadjustment. Double-talk control in echo cancellation is an example of this problem. This problem has also been studied from a viewpoint of robust algorithms [1]–[4].

Hirano et al. proposed a noise-robust NLMS (normalized least mean squared) algorithm, which makes its stepsize smaller with a larger disturbing term [1, 2]. The disturbance is estimated as a longterm average of the instantaneous error. This is because averaging operation with a large time constant guarantees sufficient accuracy in the estimation. Although it is a good choice for stationary signals, it may not provide sufficient tracking capability for nonstationary signals including an impulse-like signal. If the time constant is set smaller, the accuracy for stationary signals is traded off for the tracking speed. Its descendant [2] has improved estimation of the disturbing term, however, it is effective only during the convergence process of the adaptive filter.

Valin et al. proposed robust algorithms with an adaptive stepsize to control the learning rate based on the power of the disturbance [3, 4]. They calculate an estimated ratio of the misadjustment to the error. To estimate the misadjustment power, it is expressed



Fig. 1. System Identification.

by the product of the echo power and a normalized filter misadjustment that is the inverse of the echo-return-loss enhancement (ERLE). Then, the echo power is approximated by the adaptive filter output. It is naturally assumed in [3, 4] that the normalized filter misadjustment, or equivalently, the ERLE, is slowly time-varying. However, it is true only from an analytical point of view. Reflecting coefficient adaptations, the ERLE is rapidly changing. Therefore, the same problem resides in the ERLE estimation as in the disturbingterm estimation of [1, 2].

This paper proposes a robust NLMS algorithm with a novel noise modeling based on stationary/nonstationary noise decomposition. An integrated noise estimate, which consists of a stationary and a nonstationary noise estimates with appropriate time constants, is used to control the stepsize for coefficient adaptation. In the following section, the conventional algorithm [1] is reviewed to highlight the trade-off in noise estimation. Section 3 presents a new noise model in combination with a robust NLMS algorithm. Finally in Section 4, evaluation results in the context of echo cancellation are demonstrated to show the robustness to the disturbing term including both the ambient noise and the near-end speech.

2. CONVENTIONAL ALGORITHM

Let us consider system identification of an unknown system with an impulse response **h** as shown in Fig. 1. $\mathbf{w}(k)$ is the coefficient vector of the adaptive filter with a time index k. $\mathbf{x}(k)$ is the reference input to the adaptive filter. Vectors **h**, $\mathbf{w}(k)$, $\mathbf{x}(k)$ are defined by

$$\mathbf{h} = [h_0 \ h_1 \ \cdots \ h_{N-1}]^T, \tag{1}$$

$$\mathbf{w}(k) = [w_0(k) \ w_1(k) \ \cdots \ w_{N-1}(k)]^T,$$
(2)

$$\mathbf{x}(k) = [x(k) \ x(k-1) \ \cdots \ x(k-N+1)]^T, \qquad (3)$$

where N is the number of adaptive-filter coefficients and $[\cdot]^T$ denotes a vector transpose. The unknown-system output m(k) and the adaptive-filter output $\hat{y}(k)$ are expressed by

$$m(k) = y(k) + n(k),$$

T (1)1 .

$$= \mathbf{x} (k)\mathbf{h} + n(k), \qquad (4)$$

$$\hat{y}(k) = \mathbf{x}^{*}(k)\mathbf{w}(k), \tag{5}$$

with the genuine output of the unknown system y(k) and a background noise n(k). From (4) and (5), the error e(k) becomes

$$e(k) = m(k) - \hat{y}(k),$$

$$= \mathbf{x}^{T}(k) \{\mathbf{h} - \mathbf{w}(k)\} + n(k).$$
(6)

The noise-robust NLMS algorithm [1] performs coefficient adaptation with a time-varying stepsize $\mu(k)$ by

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu(k)e(k)\mathbf{x}(k), \tag{7}$$

$$\mu(k) = \frac{\mu_0 \sigma_n^2(k)}{\sigma_x^4(k) + \alpha^2 \hat{\sigma}_n^4(k)}.$$
 (8)

 $\sigma_x^2(k)$ and $\hat{\sigma}_n^2(k)$ are the input-signal power and a noise power estimate, respectively. α is a positive constant. $\sigma_x^2(k)$ and $\hat{\sigma}_n^2(k)$ are given by

$$\sigma_x^2(k) = \mathbf{x}(k)^T \mathbf{x}(k) \tag{9}$$

$$\hat{\sigma}_{n}^{2}(k) = \begin{cases} \beta \hat{\sigma}_{n}^{2}(k-1) + (1-\beta)e^{2}(k) & \sigma_{x}^{2}(k) < \sigma_{0}^{2} \\ \hat{\sigma}_{n}^{2}(k-1) & otherwise \end{cases}, (10)$$

with a threshold σ_0^2 . The stepsize in (8) is an upward convex function of $\sigma_x^2(k)$ with a noise offset $\alpha^2 \hat{\sigma}_n^4(k)$ for robustness. For a large power of the noise, the stepsize naturally decreases to a small value for stability. Therefore, the robustness largely depends on the accuracy of the noise estimate $\hat{\sigma}_n^2(k)$.

The noise estimate is calculated as an average of the error e(k) that may be contaminated by noise. An averaging parameter β in (10) is set to a value close to 1 so that sufficient accuracy is guaranteed. However, it does not provide good tracking capability for quick changes widely encountered in nonstationary signal such as speech. If a smaller value of β is used for better tracking capability, sufficient accuracy for nonstationary noise may not be guaranteed. There is a trade-off in the selection of β .

3. PROPOSED ALGORITHM WITH A NEW NOISE MODEL

The proposed algorithm performs coefficient adaptation with the following time-varying stepsize.

$$\mu(k) = a(k)\mu_{\min}(k), \tag{11}$$

$$u_{\min}(k) = \min\{\bar{\mu}(k) \; \bar{\mu}(k-1) \; \cdots \; \bar{\mu}(k-M+1)\}, \, (12)$$

$$\bar{\mu}(k) = \frac{\mu_0 \sigma_n(k)}{\sigma_x^4(k) + \sigma_A^4(k)},$$
(13)

$$a(k) = \max\{sgn(\sigma_x^2(k) - \sigma_0^2), 0\}$$
(14)

a(k) in (14) is introduced to (11) such that coefficient adaptation is skipped for low input power. min $\{\cdot\}$ is an operator to take the minimum value of the arguments. (12) takes the minimum stepsize of the past M samples for robustness. $\alpha^2 \hat{\sigma}_n^4(k)$ in (8) is replaced with a better noise estimate $\sigma_A^4(k)$ that is integrated from the stationary



Fig. 2. The proposed algorithm in an acoustic echo canceller.

and the nonstationary noise estimates. Figure 2 illustrates a blockdiagram of the proposed algorithm in an acoustic echo canceller.

Separate estimations for the stationary and the nonstationary noise are introduced for eliminating the trade-off in the β selection. They are estimated with different values of β and mixed depending on the stationarity likelihood $c_1(k)$ as **NEW NOISE MODEL**

$$\sigma_A^2(k) = c_1(k)\sigma_{sn}^2(k) + \{1 - c_1(k)\}\sigma_{nn}^2(k), \quad (15)$$

where $\sigma_A^2(k)$, $\sigma_{sn}^2(k)$, and $\sigma_{nn}^2(k)$ are the integrated noise estimate, the stationary noise estimate, and the nonstationary noise estimate, respectively. The simplest way is to select either the stationary or the nonstationary noise as follows:

$$c_1(k) = \max\{sgn(\epsilon_0 \sigma_{sn}^2(k) - e^2(k)), 0\}$$
(16)

 ϵ_0 is a postive constant. (15) and (16) mean that, when the error $e^2(k)$ is within a certain range of the estimated stationary noise, it is considered as stationary noise, otherwise, nonstationary one.

This principle is reasonable when there is a negligible residual echo because the variance from the estimate is relatively small for stationary noise and large for nonstationary one. Nonetheless, it does not hold when the residual echo is not negligible such as in the convergence process of the adaptive filter. In that case, (16) is replaced by (17), where the second condition is newly introduced.

$$c_{1}(k) = \begin{cases} 1 & e^{2}(k) < \epsilon_{1}\sigma_{sn}^{2}(k) \\ 1 & e^{2}(k) > \epsilon_{1}\sigma_{sn}^{2}(k) \\ & and & \sigma_{x}^{2}(k) > \sigma_{0}^{2} \\ & and & e^{2}(k) < \sigma_{x}^{2}(k) \\ 0 & otherwise \end{cases}$$
(17)

When there is possible residual of y(k), $e^2(k)$ smaller than the reference signal $\sigma_x^2(k)$ is additionally considered as nonstationary noise. Possibility of the residual is evaluated by the reference signal power.

Convergence is declared once an averaged gradient $\overline{\Delta}_w(k)$ of the coefficient-vector norm has become smaller than a threshold $\overline{\Delta}_0$. $\overline{\Delta}_w(k)$ is given by

$$\overline{\Delta}_w(k) = \frac{1}{L} \sum_{j=k-L+1}^k \Delta_w(j), \qquad (18)$$

$$\Delta_w(k) = \frac{\mathbf{w}^T(k)\mathbf{w}(k) - \mathbf{w}^T(k-1)\mathbf{w}(k-1)}{\mathbf{w}^T(k)\mathbf{w}(k)}.$$
 (19)

 $\overline{\Delta}_0$ and L should be optimized for given values of μ_0 and N.



Fig. 3. Separate estimation of stationary/nonstationary noise and their integration.

Stationary noise is estimated by leaky integration in (20) that is controlled by (21).

$$\sigma_{sn}^2(k) = \begin{cases} \beta_0 \sigma_{sn}^2(k-1) + (1-\beta_0)e^2(k) & c_2(k) = 1\\ \sigma_{sn}^2(k-1) & c_2(k) = 0 \end{cases}, (20)$$

$$c_2(k) = \begin{cases} 1 & \sigma_x^2(k) < \sigma_0^2 \\ & and \ \epsilon_2 \overline{\sigma}_{sn}^2(k) < e^2(k) < \epsilon_3 \overline{\sigma}_{sn}^2(k) , \\ 0 & otherwise \end{cases}$$
(21)

where ϵ_2 and ϵ_3 are positive constants. β_0 is a constant satisfying $0 \le \beta_0 \le 1$. $\overline{\sigma}_{sn}^2(k)$ is an average of $\sigma_{sn}^2(k)$ given by

$$\overline{\sigma}_{sn}^2(k) = \beta_1 \overline{\sigma}_{sn}^2(k-1) + (1-\beta_1) \sigma_{sn}^2(k), \qquad (22)$$

where β_1 is a nonnegative constant ($0 \le \beta_1 \le 1$). In another word, the update is performed only when the reference input has a small power (*i.e.* the echo is small) and the error is within a certain range of an averaged stationary noise estimate. The second condition detects signal components which have significantly larger power than the stationary noise and skips update of the stationary noise estimate.

This concept is illustrated in Fig. 3, where $\sigma_{nn}^2(k)$ is set equal to $e^2(k)$ for simplicity. $e^2(k)$, representing $\sigma_{nn}^2(k)$, and $\sigma_{sn}^2(k)$ are expressed by a dashed and a dotted lines, respectively. When $e^2(k)$ is smaller than a scaled-up version of the stationary noise estimate as is highlighted by a shaded area, $\sigma_{sn}^2(k)$ is selected as $\sigma_A^2(k)$ (circles in Fig. 3). Otherwise, $e^2(k)$ is used as represented by squares in the figure. The resulting integrated noise estimate is the line connecting the circles and squares from the left in Fig. 3.

The initial value $\sigma_{sn}^2(0)$ of the stationary noise estimate should be a nonzero positive number. Otherwise, $\sigma_{sn}^2(k)$ in (20) does not grow. One of the good initial values is an average of the error $e^2(k)$ for the initial *P* samples. It is based on a reasonable assumption that $e^2(k)$ contains only stationary ambient noise with no echo nor nearend signal right after the system is switched on. This is a widely used assumption in noise suppression for calculating an initial value for the estimated noise [5].

For the nonstationary noise, a different time constant from that for the stationary-noise estimation in (20) is used as follows

$$\hat{\sigma}_{nn}^2(k) = \gamma \hat{\sigma}_{nn}^2(k-1) + (1-\gamma)e^2(k),$$
 (23)

where γ is a nonnegative constant $(0 \le \gamma \le 1)$. A smaller value of γ provides faster tracking speed. $\hat{\sigma}_{nn}^2(k)$ finally results in

$$\sigma_{nn}^2(k) = \frac{\delta}{|\sigma_x(k)|} \hat{\sigma}_{nn}^2(k), \qquad (24)$$

where the scaling factor on the right-hand side plays a similar role to that of α in (8).

Although a new algorithm was developed based on [1] and assuming echo cancellers, it is possible to apply the same technique



Fig. 4. Estimated noise and the microphone signal m(k).

 Table 1. Parameters for evaluations.

N	1280	β_0	0.99
M	80	β_1	0.99
σ_0	$2^{14}N$	ϵ_0	1
L	8N	ϵ_1	2
$\overline{\Delta}_0$	$(16N)^{-1}$	ϵ_3	0.385
P	4N	ϵ_4	2
δ	7.6×10^{22}	γ	0

as in (15), (16), (20), and (23) to other algorithms such as [3] and other applications. The essence is to mix a stationary estimate and a nonstationary estimate after independent estimations, which are performed with different time constants.

4. EVALUATIONS

Evaluations were performed in echo cancellation scenarios with two real speech signals sampled at 16 kHz for the far-end signal (FES) and the near-end signal (NES). Two echo canceller structures were used in the evaluation, namely, the fullband structure and a 64-band complex oversampled subband structure [6]. A real echo-path impulse response in an office was used to generate the echo. Parameters for the fullband structure are illustrated in Tab. 1. The subband structure basically used the same parameter except that some of them are scaled according to the decimation factor 80. μ_0 was set to 0.2.

Figure 4 depicts the estimated noise and the microphone signal m(k) for the fullband (FB) and the subband structures. In the latter case, results in subbands 9 (around 1kHz) and 17 (around 2kHz) are illustrated. FB, 1kHz, and 2kHz represent the result by the fullband structure and those in subbands 9 and 17. Single-talk with no NES is maintained until a double-talk section starts at around 5 sec.

Shown in these subfigures are the estimated noise by the pro-

posed noise model (A: dotted line), one by the conventional model (B: solid line), and the microphone signal m(k) in gray. It is clearly seen in Fig. 4 that the estimated noise by the proposed model (dotted line) takes a larger value than that by the conventional model (solid line) in double-talk periods. The proposed model often results in a larger noise estimate even in single-talk periods. This is because the parameters to discriminate the stationary/nonstationary noise, namely, ϵ_0 and ϵ_1 , were set to modest values. Setting these parameters to aggressive values may lead to smaller estimates in double-talk periods. The current values in Tab. 1 were optimized from a viewpoint of subjective output quality.

Figure 5 compares the output, the ERLE, and the stepsize for the proposed algorithm with a new noise model and the conventional algorithm. "A" and "B" in each figure represent (8) and (13) with the same stationary noise estimation in (20). The stepsize by (13) takes a small value when there is nonstationary noise such as the NES. A reasonable ERLE is maintained in both single- and double-talk sections The ERLE is improved by as much as 40 dB due to independent stationary/nonstationary noise estimation. It is clear that the output signal is close to the NES with no double-talk detection, which represents good performance.

Results in the subband structure are illustrated in Fig. 6. The third subfigure from the top is the fullband output after the synthesis filterbank. The ERLEs and the stepsizes for subbands 17 (around 2kHz) and 9 (around 1kHz) are shown in the fourth through the seventh subfigures. Similar results to those by the fullband structure were obtained with the same FES and the NES as for the fullband structure.

5. CONCLUSION

A robust NLMS algorithm with a novel noise modeling has been proposed. Independent stationary and nonstationary noise estimations with different time constants followed by integration has been introduced based on a new noise model to better control the stepsize for robustness. Evaluations in a fullband and a subband echo cancellation scenarios have demonstrated that as much as 40 dB higher ERLE were obtained compared to the results with the conventional noise model in double-talk sections with no double-talk detection.

6. REFERENCES

- A. Hirano and A. Sugiyama, "A noise-robust stochastic gradient algorithm with an adaptive step size suitable for mobile hands-free telephones," Proc. ICASSP'95, pp.1392–1395, Apr. 1995.
- [2] Y. Hiratani, A. Hirano, and M. Kanazawa, "A noise-robust echo canceller on V830 multimedia RISC processor integrated into a car navigation system," Proc. ICASSP'98, pp.1753– 1756, Apr. 1998.
- [3] J.-M. Valin, "On adjusting the learning rate in frequency domain echo cancellation with double-talk," IEEE Trans. on Audio, Speech and Language Processing, Vol.15, No.3, pp.1030– 1034, Mar. 2007.
- [4] J.-M. Valin and I. B. Collings, "A new robust frequency domain echo canceller with closed-loop learning rate adaptation," Proc. ICASSP2007, pp.I93–I96, Apr. 2007.
- [5] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-27, no. 2, pp.113-120, Apr. 1979.
- [6] B. F. Boroujeny, "Adaptive filters, Theory and applications," John Wiley & Sons, 1998.



Fig. 5. Evaluation results (Fullband implementation).



Fig. 6. Evaluation results (Subband implementation).