

A Low-Distortion Noise Canceller with an SNR-Modified Partitioned Power-Normalized PNLMS Algorithm

Akihiko Sugiyama, Masanori Kato, and Masahiro Serizawa
 NEC Common Platform Software Research Laboratories
 Kawasaki 211-8666, JAPAN

Abstract— This paper proposes a low-distortion noise canceller with an SNR-modified partitioned power-normalized PNLMS algorithm. The coefficient adaptation stepsize is controlled by two factors; an estimated signal-to-noise ratio (SNR) at the primary input and a relative coefficient magnitude normalized by the reference power. The SNR is estimated based on the noise replica and the noise-canceller output. The SNR-controlled stepsize provides robustness to interference of the desired speech to the error. A newly developed partitioned power-normalized proportionate normalized least-mean-squares (PP-PNLMS) algorithm helps stabilize coefficient fluctuations caused by a delay in recursive SNR estimation. Subjective evaluation results have demonstrated statistically significant differences in the low-SNR scores over a 3GPP noise suppressor.

I. INTRODUCTION

Speech enhancement is an indispensable technology for communications and human-computer interaction in noisy environments. Interference and noise are sometimes point-sources of sound and sometimes diffused ones. As multiple-microphone solutions, microphone arrays have been extensively studied. However, they are not suitable for a diffused interference because of its dependency on directivity. For diffused noise, adaptive noise cancellers have demonstrated their potential in some applications [1].

A noise canceller was originally proposed by Widrow et al. in mid 70's [2]. An auxiliary (or reference) microphone captures a signal which is correlated with the noise components in the primary-microphone signal. This correlated signal drives an adaptive filter to generate a noise replica, which is then subtracted from the primary-microphone signal for noise cancellation. Coefficients of the adaptive filter are updated with the subtraction result, which is a mixture of the speech to be enhanced, *i.e.* the desired speech, and the misadjustment. It is clear that the desired speech has nothing to do with the misadjustment and plays a role of an interference. As a result, coefficient adaptation is disturbed, resulting in distortions in the enhanced speech and residual noise.

As a solution to the interference problem, an adaptive noise canceller with a paired filter (ANC-PF) structure [3] introduces an auxiliary (or sub) adaptive filter for estimating a signal-to-noise ratio (SNR) that is used to control the coefficient-adaptation stepsize in the main adaptive filter. A small stepsize for high SNRs, when the desired speech seriously interferes the misadjustment, provides steady and accurate adaptation

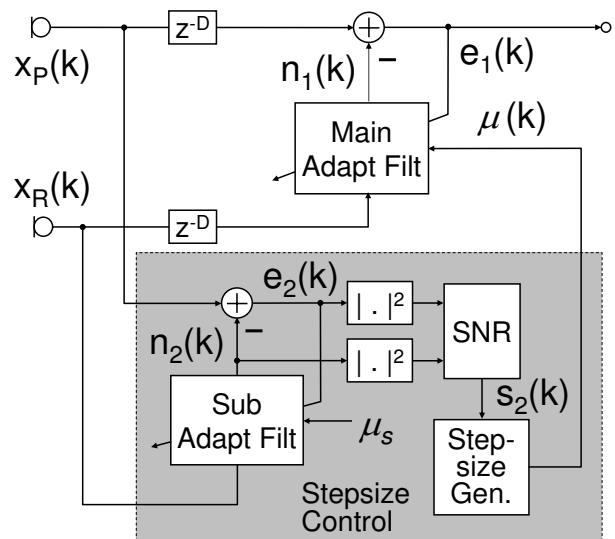


Fig. 1. Blockdiagram of ANC-PF.

of coefficients, leading to low-distortion and small misadjustment. However, the necessity for an auxiliary filter for SNR estimation doubles the total number of computations. In addition, a delay is introduced in the enhanced speech to compensate for the SNR-estimation delay with the auxiliary filter. This delay reduces the margin for the total delay imposed by international standards for communications terminals.

This paper proposes a low-distortion noise canceller with stepsize control that requires no additional filter nor any delay in the enhanced speech. In the next section, SNR estimation in ANC-PF is reviewed in details to highlight a delay problem. Section III presents a new low-distortion noise canceller. Finally, in Section IV, evaluation results of the new noise canceller are presented with objective and subjective measures.

II. SNR ESTIMATION IN ANC-PF

Figure 1 depicts a blockdiagram of ANC-PF. It has a sub adaptive filter (SAF) connected in parallel to the main adaptive filter (MAF). The SAF is introduced for providing an appropriate stepsize for the MAF through an estimated SNR. Its stepsize is fixed to μ_s that is relatively large compared to the MF stepsize $\mu(k)$. μ_s provides good tracking capability to the changes of the SNR as well as its estimate.

Assuming good noise cancellation by the SAF, its output $n_2(k)$ is a faithful noise replica. It is subtracted from the primary-microphone signal $X_P(k)$ to obtain the desired speech as

$$e_2(k) = x_P(k) - n_2(k) \quad (1)$$

Therefore, the subtraction result $e_2(k)$ can be regarded as a replica of the desired speech. Based on these replicas, $n_2(k)$ and $e_2(k)$, of the noise and the desired speech, an estimated SNR, $s_2(k)$, is calculated by

$$s_2(k) = \text{ave}\{e_2^2(k)\} / \text{ave}\{n_2^2(k)\}, \quad (2)$$

where $\text{ave}\{\cdot\}$ is a time-averaging operator to absorb imperfections in the SAF behavior for better accuracy.

The SNR estimate, $s_2(k)$ is then processed by an appropriate function $f\{\cdot\}$ to convert it to a stepsize $\mu(k)$ as in

$$\mu(k) = f\{s_2(k)\} \cdot \mu_0. \quad (3)$$

μ_0 is the NLMS stepsize that satisfies $0 < \mu_0 < 2$. The function $f\{\cdot\}$ is designed as a decreasing function of $s_2(k)$ such that a high SNR with a strong desired speech returns a small value for stable adaptation.

The most significant drawback of this structure is additional computations necessary for the SAF. The SAF has the same number of taps, thus, the total number of computations is doubled. Moreover, there is a delay problem.

Because of the time-averaging operator, the SNR estimate is somehow delayed. This delayed estimate of the SNR in the primary signal does not provide an appropriate stepsize for the MAF. The primary signal also has to be delayed so that it becomes in-phase with the SNR estimate. For this purpose, D -sample delay is introduced in both the primary and the reference signals as in Fig. 1. This delay accounts for the total delay of a communication system which must satisfy the maximum delay requirement imposed by international standards. The delay may be eliminated by using the main-filter output as the noise replica for SNR estimation. However, in this case, the main filter may become unstable because an appropriate but delayed stepsize based on the SNR estimate is not available.

III. PROPOSED NOISE CANCELLER

The SNR is estimated based on the MAF output for eliminating the SAF. For the instability problem caused by the delay, the coefficient adaptation algorithm is made more robust to the interference. A blockdiagram of the proposed noise canceller is depicted in Fig. 2.

A newly developed partitioned power-normalized proportionate normalized least-mean-squares (PP-PNLMS) algorithm solves this instability. PP-PNLMS calculates a stepsize for a block of coefficients based on average powers of the coefficients and the reference signal in the block. Assuming L equisize blocks, the PP-PNLMS stepsize $\tilde{\mu}_j(k)$ for the j -th

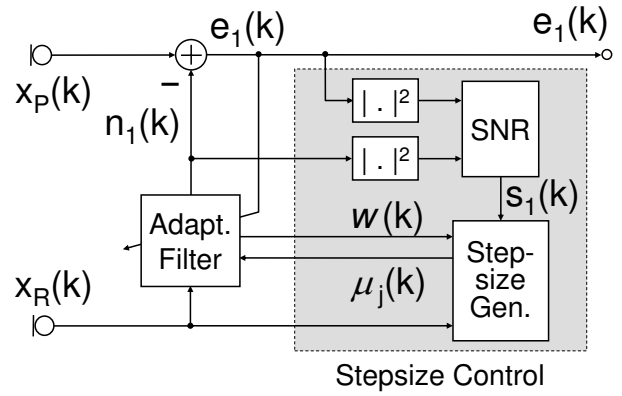


Fig. 2. Blockdiagram of the new noise canceller.

block ($j = 0, 1, 2, \dots, L - 1$) is given by

$$\tilde{\mu}_j(k) = \mu_{mod}(j, k) \cdot \frac{w_b(j, k)}{w_b(j_{max}, k)} \cdot \mu_0 \quad (4)$$

$$\mu_{mod}(j, k) = \frac{x_b(j_{max}, k)}{x_b(j, k)}, \quad (5)$$

where $w_b(j, k)$ and $x_b(j, k)$ are the sum of squared coefficients and reference signal powers in Block j . j_{max} is the index j to the block with the maximum value of $w_b(j, k)$ for all values of j . The right-hand side of (4) without $\mu_{mod}(j, k) \cdot \mu_0$ is a partitioned PNLMS algorithm¹. The rest of the right-hand side is the power-normalization factor for each block. It is a blockwise reference signal power normalized by that for the block with the maximum sum of squared coefficients.

The possible instability becomes more evident when the SNR goes up. In this case, the stepsize should decrease. However, a delayed SNR estimate causes a delay in the stepsize decrease. As a result, the stepsize stays large and disturbs coefficient adaptation.

This problem is solved in two ways in (4). μ_{mod} , which is associated with the block reference power, is effective when the SNR is increased by the reference signal. Some points should be noted before explaining this point in more details. j_{max} is generally the first block, because the room impulse response is an exponentially dying off function. When the reference signal power increases, the block power $w_b(j, k)$ for $j = 0, 1, \dots, L - 1$ is roughly an increasing series. It means that $\mu_{mod}(j, k)$ is a decreasing series. Therefore, multiplication of $\mu_{mod}(j, k)$ offsets the increase of the reference power.

An SNR increase also happens when the desired speech increases in power. In this case, the other factor associated with coefficients takes care of the delay in SNR estimation. The scaling factor for μ_0 with respect to the sum of squared coefficients makes the stepsize smaller for all coefficients except those in Block j_{max} . Large coefficients that have

¹The original PNLMS algorithm [4] has a ratio of each absolute coefficient to an average absolute coefficient as a scaling factor for the NLMS stepsize. The algorithm in this paper takes the squared coefficient instead of the absolute coefficient and the maximum instead of the average. It is named "partitioned" because it is implemented in a blockwise manner.

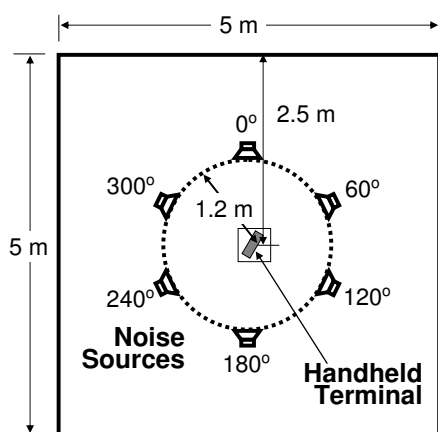


Fig. 3. Experimental setup.

significant impact on the noise replica are updated as in the conventional algorithm. Other small coefficients are adapted with a small stepsize, leading to robustness to the desired speech even when only a delayed estimate of the SNR is available.

Finally, $\tilde{\mu}_j(k)$ in (4) is modified by a function of the estimated SNR $s_1(k)$ to obtain the final stepsize $\mu_j(k)$ as

$$\mu_j(k) = f\{s_1(k)\} \cdot \tilde{\mu}_j(k), \quad (6)$$

$$s_1(k) = \text{ave}\{e_1^2(k)\} / \text{ave}\{n_1^2(k)\}. \quad (7)$$

The SNR factor, $f(s_1(k))$, changes inversely proportional to the SNR estimate, $s_1(k)$, as in [3]. A design example of $f\{\cdot\}$ is also disclosed in the same reference. The stepsize of the proposed noise canceller is obtained by (4)–(7). $\tilde{\mu}_j(k)$ is calculated for $j = 0, 1, \dots, L-1$.

With these $\mu_j(k)$, adaptation for a new coefficient vector \mathbf{w}_{k+1} is performed based on

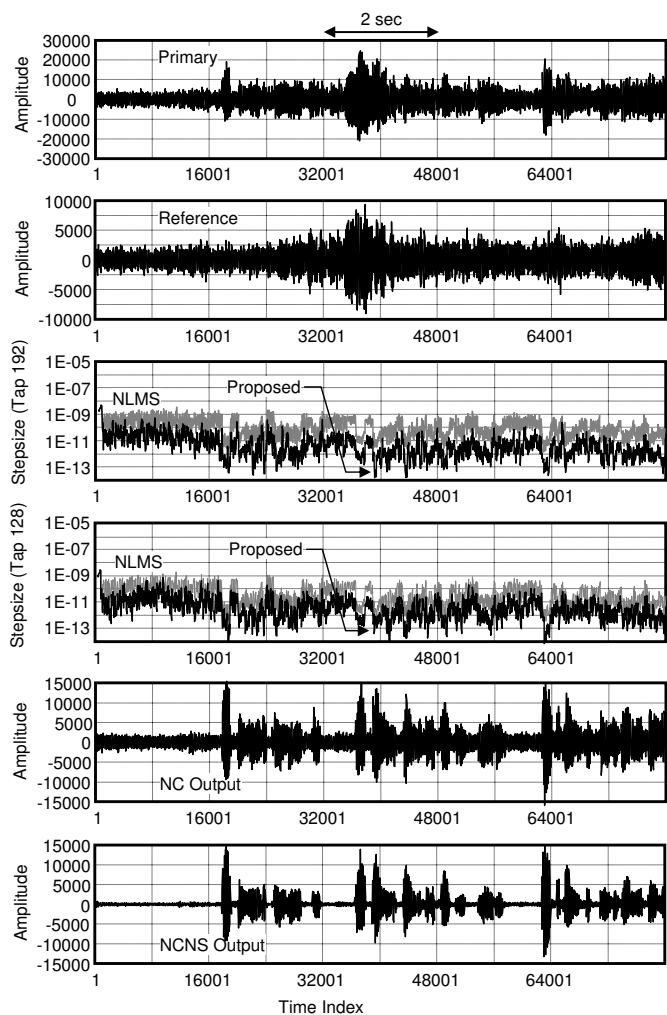
$$\mathbf{w}_{k+1} = \mathbf{w}_k + M(k) \cdot \frac{e(k)\mathbf{x}(k)}{|\mathbf{x}(k)|^2}, \quad (8)$$

$$M(k) = \begin{bmatrix} \mu_0(k) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mu_0(k) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu_{L-1}(k) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mu_{L-1}(k) \end{bmatrix}. \quad (9)$$

IV. EVALUATIONS

A. Objective Evaluation

Objective evaluations were performed with $N = 256$ and $L = 32$, where N denotes the number of taps of the adaptive filter. The experimental setup is depicted in Fig. 3. Six loudspeakers were driven by the same signals that mostly consists of station noise. The primary and the reference microphones were mounded on the opposite sides of a handheld terminal. This terminal was placed 0.3 m above a table whose height


 Fig. 4. Stepsize and output ($\text{SNR}_{PP}=5\text{dB}$).

was 1.0 m. The primary microphone was facing the table. The sound pressure level of the desired speech was 95dB at the primary microphone.

Figures 4 and 5 illustrate the coefficient adaptation stepsize and the enhanced speech for a peak-to-peak SNR (SNR_{PP}) of 5dB and 24 dB. The stepsize of the proposed noise canceller is much smaller than that of the NLMS. The difference is more distinct at high-SNR positions in Fig. 5.

The enhanced speech is demonstrated at the outputs of the new noise canceller and a postprocessing low-distortion noise suppressor [6]. The former is labeled as NC Output and the latter, NCNS Output. From Fig. 4, it is observed that the noise canceller alone does not provide sufficiently small residual noise level, especially, in noise sections. This is because the number of taps for the adaptive filter was set to 256 that may be too small. However, the residual noise is successfully suppressed by the postprocessing noise suppressor. In the case of $\text{SNR}_{PP} = 24\text{dB}$, the NC Output and the NCNS Output look quite the same. This is because the noise is sufficiently small that there is little residual noise in the NC Output for the postprocessing NS to suppress.

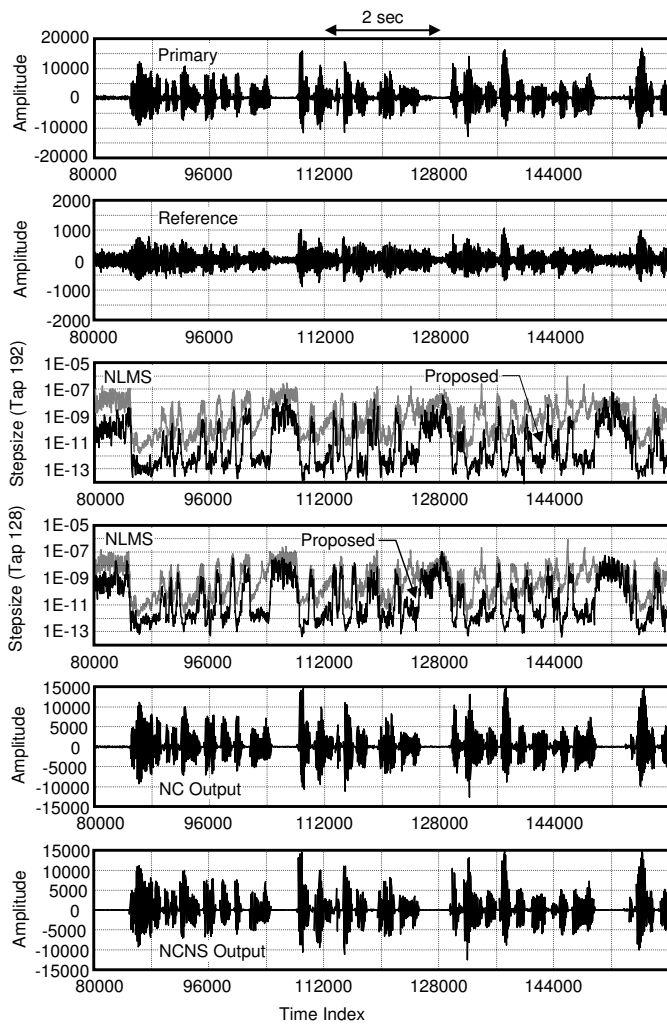


Fig. 5. Step size and output ($SNR_{PP}=24dB$).

B. Subjective Evaluation

In the subjective evaluation, car, street, and babble noise signals with SNRs of 0, 6, 12, and 18 dB in combination with four speech signals were used. All signals were sampled at 8 kHz. A total of 22 subjects participated in the subjective evaluation. Evaluation was performed based on the 5-grade ACR (absolute category rating) scale [5] in Tab. I. The proposed noise canceller followed by a low-distortion noise suppressor [6] for residual noise suppression (NC+LDNS) was compared with a 3GPP noise suppressor [7] (NS) and a noisy speech with a 3 dB higher SNR ($SNR+3$) than the original SNR.

Fig. 6 shows subjective evaluation results. NC+LDNS provides better scores than NS in the low SNR cases such as 0 and 6 dB. At a 0 dB SNR, there are statistically significant differences between NC+LDNS and NS in all the three noise signals. NC+LDNS obtained comparable scores to that of NS in the high SNR cases. Therefore, NC+LDNS provides subjective quality better than or comparable to the 3GPP endorsed noise suppressor.

TABLE I
5-POINT GRADE.

| Description | Bad | Poor | Fair | Good | Excellent |
|-------------|-----|------|------|------|-----------|
| Score | 1 | 2 | 3 | 4 | 5 |

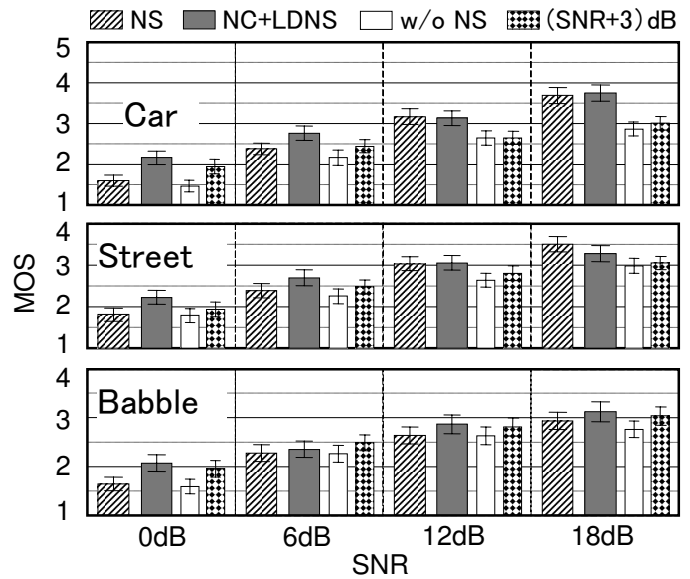


Fig. 6. Subjective evaluation result.

V. CONCLUSION

A low-distortion noise canceller with an SNR-modified partitioned power-normalized PNLMS algorithm has been proposed. Unlike the conventional noise canceller with an SNR-controlled stepsize, it has been designed to have no auxiliary adaptive filter for SNR estimation. An estimated SNR based on the noise replica and the ANC output combined with a newly developed PP-PNLMS algorithm has been shown to provide good stepsize control for a wide range of SNRs. Subjective evaluation results have demonstrated statistically significant differences in the low-SNR scores over a 3GPP noise suppressor.

REFERENCES

- [1] A. Sugiyama, "Low-distortion noise cancellers – Revival of a classical technique," *Speech and audio processing in adverse environment*, Chap. 7, Hänslér and Schmidt, ed. Springer, 2008.
- [2] B. Widrow, J. R. Glover, Jr., J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, Jr., R. C. Goodlin: Adaptive noise cancelling: principles and applications, *Proc. IEEE*, **63**, (12), 1692–1716, 1975.
- [3] S. Ikeda and A. Sugiyama, "An Adaptive Noise Canceller with Low Signal-Distortion for Speech Codecs," *IEEE Trans. Sig. Proc.*, pp.665–674, Mar. 1999.
- [4] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. SAP*, vol. 8, No. 5, pp.508–518, Sep. 2000.
- [5] "Minimum performance requirements for noise suppressor application to the AMR speech encoder," 3GPP TS 06.77 V8.1.1, Apr. 2001.
- [6] K. Yamato, A. Sugiyama, M. Kato, "Post-processing noise suppressor with adaptive gain-flooring for cell-phone handsets and IC recorders," *ICCE'07*, 6.1–4, Jan. 2007.
- [7] M. Kato A. Sugiyama, S. Serizawa, "A low-complexity noise suppressor with nonuniform subbands and a frequency-domain highpass filter," *Proc. ICASSP2006*, pp.473–476, May 2006.