ADAPTIVE NOISE CANCELLER WITH SNR ESTIMATE SWITCHOVER FOR STEPSIZE CONTROL

Akihiko Sugiyama[1,2]

[1]Data Science Research Laboratories, NEC Corporation
1753 Shimonumabe, Nakahara-ku, Kawasaki 211–8666, Japan
[2] Faculty of System Design, Tokyo Metropolitan University
6-6 Asahigaoka, Hino, 191–0065, Japan

ABSTRACT

This paper proposes an adaptive noise canceller with SNR estimate switchover for stepsize control. The SNR estimate is switched from a first estimate to a second one when the latter takes a smaller value than the former. The first estimate is a power ratio of the primary and the reference input signals, which is newly introduced to enable initial coefficient adaptation. The second estimate is a power ratio of the error signal and the adaptive filter output, which is used in a conventional stepsize control. Switchover from the first estimate to the second guarantees initial coefficient adaptation by the former and provides accurate SNR estimation by the latter. Evaluations with speech-speech inputs and speech-noise inputs demonstrate that use of the second SNR followed by the first SNR successfully makes the coefficient vector norm approach the true value which represents convergence.

Index Terms— Signal Enhancement, Noise canceller, SNR estimate, Switchover, Stepsize

1. INTRODUCTION

Noise cancellers (NCs) [1] are useful for signal enhancement when SNR (signal-to-noise ratio) is relatively low and single-channel speech enhancement [2, 3] cannot provide sufficient improvement. They have demonstrated high potentials in applications such as robots [4]-[6] and mobile phone handsets [7]. The signal from a reference microphone, which is correlated with the noise components in the primary-microphone signal, drives an adaptive filter to generate a noise replica. Coefficients of the adaptive filter are updated by using an error signal or the difference between the signal from the primary microphone and the noise replica. The error signal is obtained as a sum of the target signal to be enhanced and the pure misadjustment. The target signal has nothing to do with the misadjustment and simply interferes coefficient adaptation. As a result, distortions arise in the enhanced signal and the residual noise does not decrease sufficiently [5]. In the worst case, the adaptive filter coefficients blow up.

An adaptive noise canceller with a paired filter (ANC-PF) structure [8] introduced an auxiliary (or sub) adaptive filter for estimating an SNR that is used to slow down coefficient-adaptation in the main adaptive filter in the presence of the target signal. ANC-PF uses a power ratio of the error signal and the adaptive filter output as an SNR estimate. The auxiliary adaptive filter was eliminated by guaranteed stability with conditional cancellation [7]. However, in initial convergence, an estimated SNR equals infinity because the initial coefficients are set to zero and the adaptive filter output as the SNR denominator is zero. This SNR value results in a near-zero stepsize

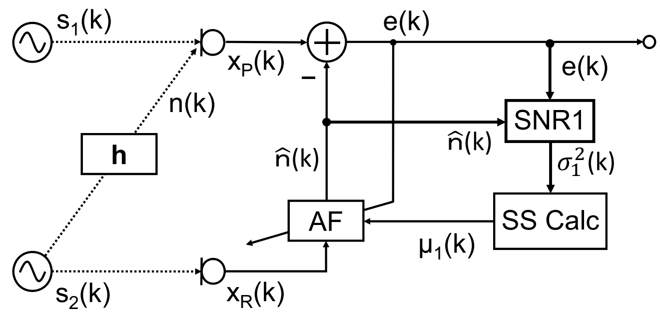

Fig. 1. Noise canceller with an SNR estimate.

and coefficients do not grow for a long time. Practically, this problem was alleviated by setting a relatively large fixed stepsize during initial convergence to guarantee coefficient growth. Nevertheless, selection of the initial fixed stepsize does not have a clear design rule. A too large stepsize is weak in the target-signal sections and may lead to divergence. An opposite setting may need long convergence time. Definition of initial convergence period is based on some inference. It is desirable to design a coefficient adaptation algorithm that automatically tracks initial convergence status and provide an appropriate stepsize.

This paper proposes a noise canceller with two SNR estimates for stepsize control. Use of a new SNR estimate in initial convergence and automatic switching to the conventional SNR estimate avoids SNR divergence.

2. CONVENTIONAL NOISE CANCELLER WITH SNR ESTIMATION FOR STEPSIZE CONTROL

Figure 1 depicts a simplified blockdiagram of a conventional NC with SNR estimation for stepsize control [7]. The noise cancelled signal $e(k)$ is expressed by

$$\begin{aligned} e(k) &= x_P(k) - \hat{n}(k) \\ &= s(k) + \Delta n(k), \end{aligned} \tag{1}$$

$$\begin{aligned} \Delta n(k) &= n(k) - \hat{n}(k) \\ &= n(k) - \sum_{l=k-N+1}^{k} x_R(k-l)w(k,l), \end{aligned} \tag{2}$$

where $x_P(k)$, $x_R(k)$, $s(k)$, $n(k)$, and $\hat{n}(k)$ are the primary- and the reference-microphone signals, the target signal, noise to be cancelled in $x_P(k)$, and a noise replica (adaptive filter output). $w(k, l)$ is the l-th filter coefficient at time k. Assuming good noise cancellation by the adaptive filter, represented by $\Delta n(k) = 0$, $e(k)$ can be regarded as a replica of the target signal.

With these replicas, $\hat{n}(k)$ and $e(k)$, of the noise and the target signal, an estimated SNR, $\sigma_1^2(k)$, is calculated by (3) as the output of SNR1 in Fig. 1 and converted to a stepsize $\mu_1(k)$ by an appropriate function $f\{\cdot\}$ as in (4).

$$\sigma_1^2(k) = e^2(k)/\hat{n}^2(k), \tag{3}$$
$$\mu_1(k) = f\{\sigma_1^2(k)\} \cdot \mu_0. \tag{4}$$

μ_0 is the NLMS (normalized least mean-square) stepsize that satisfies $0 < \mu_0 < 2$. A function $f\{\cdot\}$ is designed as a decreasing function of $\sigma_1^2(k)$ such that a high SNR with a strong target signal returns a small value for stable adaptation. Adaptive filter coefficients $w(0, i)$ are initialized to zero for $i = 0, 1, \cdots, N-1$. Therefore, from (1) and (2), $e(k) = x_P(k)$ continues and no coefficient grows because $\Delta n(k) = n(k)$ does not change. $\sigma_1^2(k)$ initially does not work at all as an SNR estimate.

3. NEW NOISE CANCELLER WITH SNR ESTIMATE SWITCHOVER

3.1. New SNR estimate $\sigma_2^2(k)$ by SNR2

As a new SNR estimate useful in the initial period, a power ratio $\sigma_2^2(k)$ of the primary- and the reference-microphone signals is calculated by (5) in SNR2 of Fig. 2.

$$\sigma_2^2(k) = x_P^2(k)/x_R^2(k), \tag{5}$$

Assuming that the target signal $s(k)$ and noise $n(k)$ are not correlated,

$$E[\sigma_1^2(k)] = E[s^2(k)]/E[\hat{n}^2(k)], \tag{6}$$
$$E[\sigma_2^2(k)] = \{E[s^2(k)] + E[n^2(k)]\}/E[n^2(k)]. \tag{7}$$

Three relations between $E[\sigma_1^2(k)]$ and $E[\sigma_2^2(k)]$ follow from (6) and (7).

a. Initially, $E[\sigma_1^2(k)] \gg E[\sigma_2^2(k)]$ because $\hat{n}(0) = 0$.

b. $E[\sigma_1^2(k)]$ decreases as the adaptive filter approaches convergence ($\hat{n}(k)$ grows).

c. $E[\sigma_1^2(k)] < E[\sigma_2^2(k)]$ holds after convergence.

"Relation c" comes from the fact that $\hat{n}^2(k) \approx n^2(k)$ after convergence and (6) becomes

$$E[\sigma_1^2(k)] \approx E[s^2(k)]/E[n^2(k)]. \tag{8}$$

It follows from the three relations that there is a crossing between $E[\sigma_1^2(k)]$ and $E[\sigma_2^2(k)]$. Thus, $\sigma_2^2(k)$ can be initially used and switched to $\sigma_1^2(k)$ when their averages approximating $E[\sigma_2^2(k)]$ and $E[\sigma_1^2(k)]$ coincide. Please note that $E[\sigma_1^2(k)]$ never decreases unless $E[\sigma_2^2(k)]$ is initially used.

3.2. Switching from $\sigma_2^2(k)$ to $\sigma_1^2(k)$

$E[\sigma_1^2(k)]$ and $E[\sigma_2^2(k)]$ cannot be obtained in real products but can be approximated by their time averages, assuming their ergodicity,

$$E[\sigma_1^2(k)] \approx \overline{\sigma_1^2(k)} \tag{9}$$
$$E[\sigma_2^2(k)] \approx \overline{\sigma_2^2(k)}, \tag{10}$$

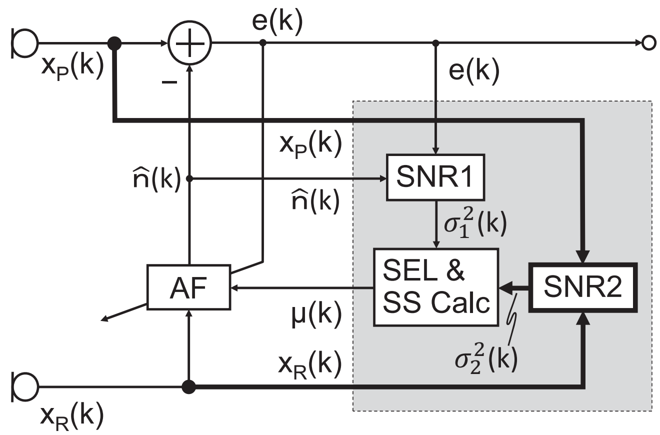

Fig. 2. New noise canceller with SNR estimate switchover.

where $\overline{\cdot}$ is an averaging operator. For switching of SNR estimates, from $\sigma_2^2(k)$ to $\sigma_1^2(k)$, a flag $\xi(k)$ is introduced with an initial value of zero as

$$\xi(k) = \begin{cases} 1 & \overline{\sigma_1^2(k)} < \overline{\sigma_2^2(k)} \ and \ \xi(k-1) = 0 \\ \xi(k-1) & otherwise \end{cases}. \tag{11}$$

The final SNR estimate $\sigma^2(k)$ is obtained by

$$\sigma^2(k) = \{1 - \xi(k)\}\sigma_1^2(k) + \xi(k)\sigma_2^2(k). \tag{12}$$

The proposed SNR estimation can also be applied to CTRANC [9]-[11] and NCs with a cross-coupled structure [12, 13].

3.3. Coefficient update with a new switchover estimate $\sigma^2(k)$

$\sigma^2(k)$ is converted to a stepsize $\mu(k)$ [7] by

$$\mu(k) = \max\{\min\{\alpha \exp\beta(\sigma^2(k) + \delta), \alpha\}, \epsilon\}. \tag{13}$$

The function in (13) is illustrated in Fig. 3. Equation (13) indicates that the stepsize $\mu(k)$ is a decreasing exponential function with a ceiling α and a floor ϵ. It is shifted by δ toward left and scaled by α. Compared to an approximating linear function that crosses $\mu(k)$ at (δ, α) and (ρ, ϵ), this function takes a small stepsize more often in the transition range between δ and ρ. This fact guarantees higher stability for coefficient adaptation. Parameters in (13) were optimized with a wide range of realistic signals (SNRs, noise, crosstalk levels) and has proven insensitive.

A coefficient vector $\boldsymbol{w}(k)$ is updated by the NLMS algorithm as

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu(k) \cdot \frac{e(k)\boldsymbol{x}_R(k)}{||\boldsymbol{x}_R(k)||^2}, \tag{14}$$

where $\boldsymbol{x}_R(k)$ is a reference signal vector of the same size as the filter coefficient vector $\boldsymbol{w}(i)$.

4. EVALUATION

Evaluations were performed using male and female speech as well as one station noise sampled at 8 kHz for $s(k)$ and $n(k)$ with an impulse response of length $N = 1024$ identified in a room with a smartphone handset. Figure 4 illustrates the impulse response from the noise source to the primary microphone. Convolution of the impulse response and the male speech or the station noise was added

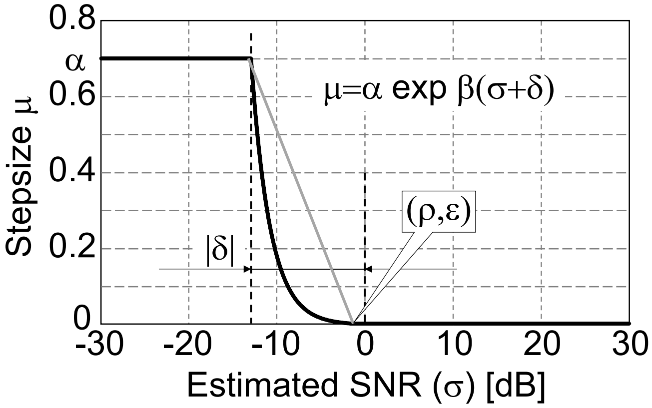

Fig. 3. SNR-dependent stepsize.

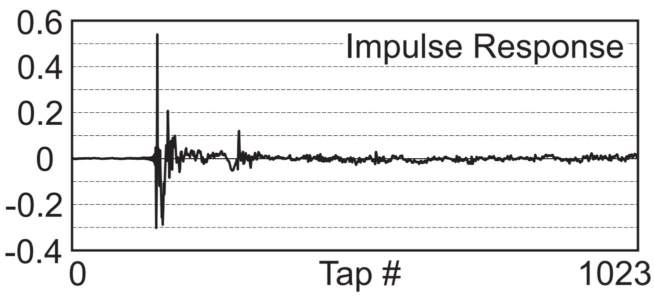

Fig. 4. Impulse response from the noise source to the primary microphone ($\boldsymbol{h}$ in Fig. 1).

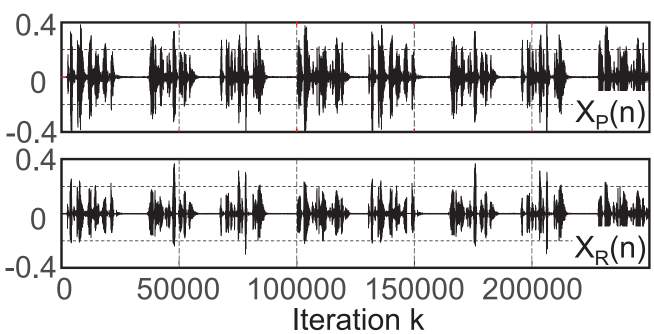

Fig. 5. Input signals, $x_P(k)$ and $x_R(k)$, both of which are speech.

to the female speech as the primary signal. The male speech or the station noise was used as the reference signal. Parameters for coefficient adaptation were set as in [7]. Estimated SNR values were averaged over 256 samples for (9) and (10) for SNR-estimate switchover decision.

Figure 5 illustrates the primary signal $x_P(k)$ and the reference signal $x_R(k)$ for speech-speech inputs. They have significant overlaps which is hard for the adaptive filter to converge without a good stepsize control. The conventional SNR estimate $\sigma_1^2(k)$ in black bold and the new SNR estimate $\sigma_2^2(k)$ in gray dots are compared in Fig. 6. As explained earlier, $E[\sigma_1^2(k)]$ decreases as adaptation and $E[\sigma_1^2(k)] < E[\sigma_2^2(k)]$ in the right half of the figure. Shown in Fig.

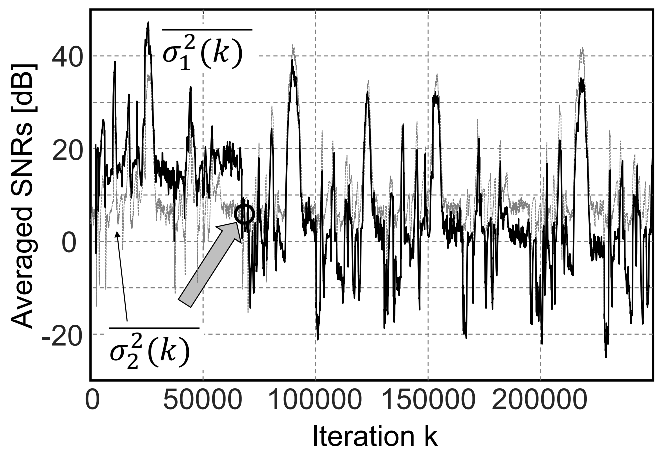

Fig. 6. Comparison of two estimated SNRs, $\overline{\sigma_1^2(k)}$ and $\overline{\sigma_2^2(k)}$ for speech-speech inputs. SNR at primary microphone is 0 dB.

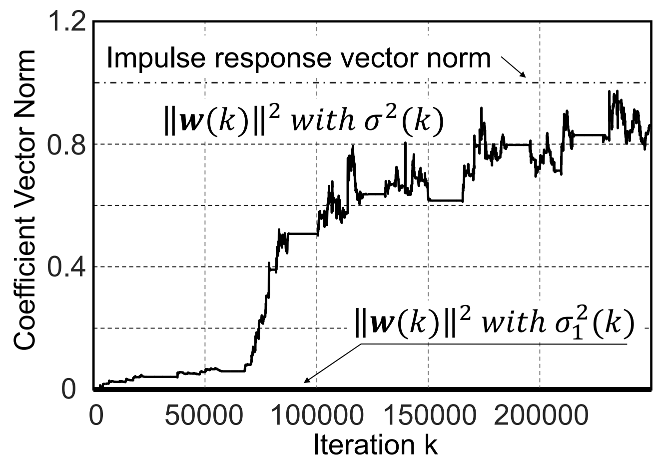

Fig. 7. Convergence characteristics by coefficient vector norm for speech-speech inputs. SNR at primary microphone was 0 dB

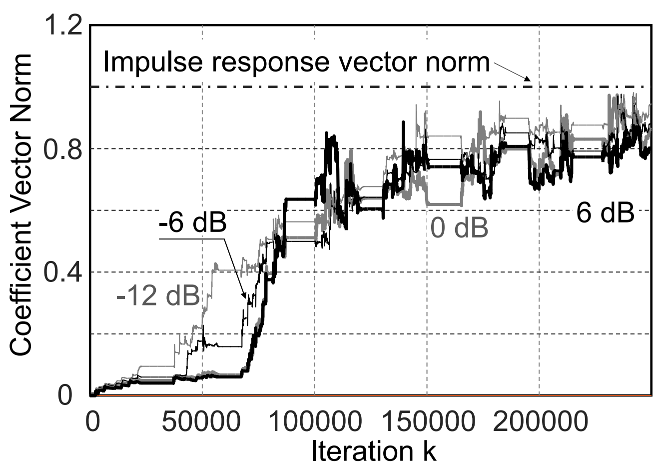

Fig. 8. Convergence characteristics by coefficient vector norm for speech-speech input at SNRs of -12, -6, 0, and $+6$ dB.

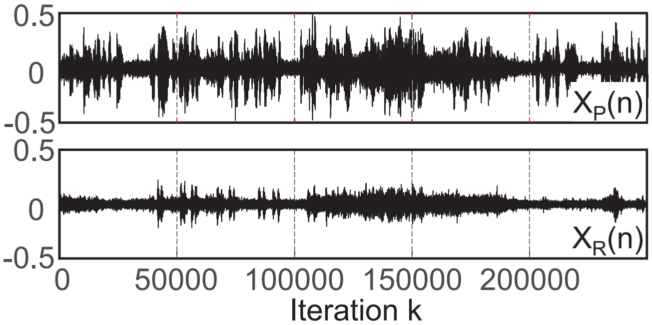

Fig. 9. Input signals, $x_P(k)$ and $x_R(k)$, which are speech and station noise, respectively.

7 are sum of coefficients or the coefficient vector norm $||\boldsymbol{w}(k)||^2$ as a measure of adaptive filter convergence. With the use of the new SNR estimate $\sigma^2(k)$, the coefficient vector norm $||\boldsymbol{w}(k)||^2$ successfully grows and approaches the ground truth, *i.e.* the impulse response vector norm. The conventional SNR estimate $\sigma_1^2(k)$ does not grow coefficients and the coefficient vector norm stays at zero. The actual changeover happened at a gray circle around 70000. Figure 8 summarizes coefficient growth by coefficient vector norm at different SNRs, namely, $+6$, 0, -6, and -12 dB. The coefficients grow similarly to each other in a wide range of SNRs.

Figure 9 shows the primary signal $x_P(k)$ and the reference signal $x_R(k)$ for speech-noise inputs. Compared to speech-speech inputs case in Fig. 5, the station noise fills the gap of the speech so that significant pauses are not visible. Comparison between the conventional SNR estimate $\sigma_1^2(k)$ in black bold and the new SNR estimate $\sigma_2^2(k)$ in gray dots appears in Fig. 10. Similarly to the speech-speech inputs case, $E[\sigma_1^2(k)]$ decreases as coefficient adaptation and $E[\sigma_1^2(k)] < E[\sigma_2^2(k)]$ is observed in the right half of the figure. Convergence characteristics of filter coefficient are illustrated in Fig. 11 by sum of coefficients or the coefficient vector norm $||\boldsymbol{w}(k)||^2$. The new SNR estimate $\sigma^2(k)$ successfully grows the coefficient vector norm $||\boldsymbol{w}(k)||^2$ as for the speech-speech inputs and it approaches the ground truth, *i.e.* the impulse response vector norm. The conventional SNR estimate $\sigma_1^2(k)$ fails to grow coefficients and the coefficient vector norm keeps the zero value. A gray circle at around 70000 represent the actual switchover from $\sigma_2^2(k)$ $\sigma_1^2(k)$. This switchover apparently accelerates convergence. Figure 8 compares coefficient growth by coefficient vector norm at different SNRs, namely, $+6$, 0, -6, and -12 dB. The coefficients growth is similar to each other in a wide range of SNRs.

5. CONCLUSION

An adaptive noise canceller with SNR estimate switchover for step-size control has been proposed. An approximation to the conventional SNR estimate has been developed for the initial convergence period when the conventional SNR estimate does not work. Their switchover takes place when an average of the conventional SNR estimate becomes smaller than an average of the new SNR estimate. Evaluations with speech-speech inputs and speech-noise inputs have demonstrated that the coefficient vector norm approached its true value, representing convergence. For SNRs between -12 and $+6$ dB, coefficient grows similarly to each other, which is a sign of robustness to different SNRs.

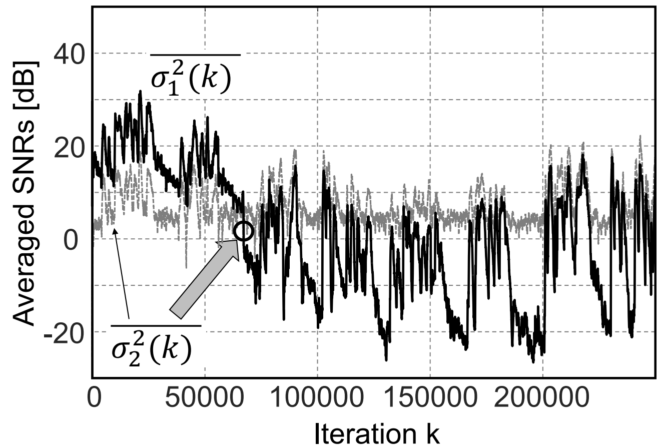

Fig. 10. Comparison of two estimated SNRs, $\sigma_1^2(k)$ and $\sigma_2^2(k)$ for speech-noise inputs.

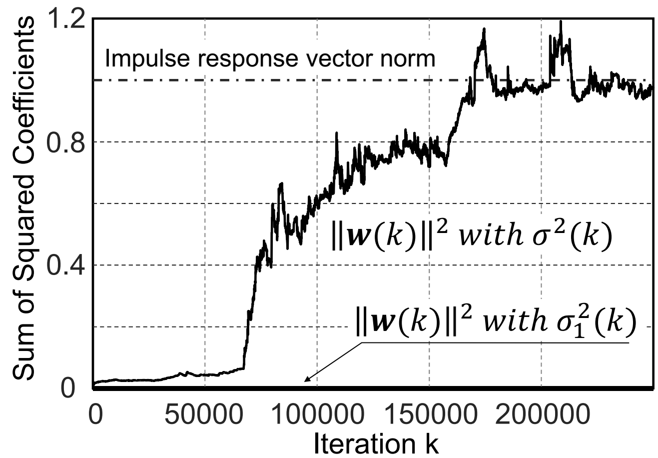

Fig. 11. Convergence characteristics by coefficient vector norm for speech-noise inputs.

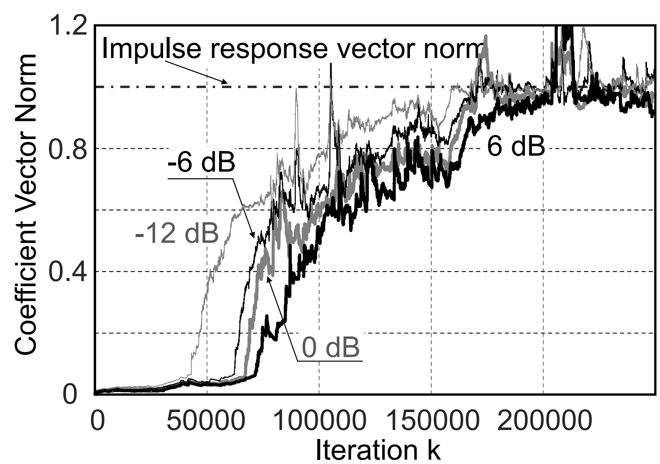

Fig. 12. Convergence characteristics by coefficient vector norm for speech-noise inputs at SNRs of -12, -6, 0, and $+6$ dB.

6. REFERENCES

[1] 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.

[2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. ASSP, vol. 27, no. 2, pp.113–120, Apr. 1979.

[3] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. ASSP, vol. 32, no. 6, pp.1109–1121, Dec. 1984.

[4] M. Sato, A. Sugiyama, and S. Ohnaka, " An adaptive noise canceller with low signal-distortion based on variable stepsize subfilters for human-robot communication,"IEICE Trans. Fundamentals, vol. E88-A, No.8, pp.2055–2061, Aug. 2005.

[5] A. Sugiyama, "Low-distortion noise cancellers – Revival of a classical technique," *Speech and audio processing in adverse environment*, Chap. 7, Hänsler and Schmidt, ed. Springer, 2008.

[6] M. Sato, T. Iwasawa, A. Sugiyama, T. Nishizawa, Y. Takano, " A single-chip speech dialogue module and its evaluation on a personal robot, PaPeRo-mini, " IEICE Trans. Fundamentals, vol. E93-A, No.1, pp.261–271, Jan. 2010.

[7] A. Sugiyama and R. Miyahara, "A Low Distortion Noise Canceller with a Novel Stepsize Control and Conditional Cancellation," Proc. EUSIPCO2014, TH-P2.8, Sep. 2014.

[8] 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.

[9] G. Mirchandani, R. L. Zinser, J. B. Evans: A new adaptive noise cancellation scheme in the presence of crosstalk, *IEEE Trans. CAS-II*, 681–694, Oct. 1992.

[10] V. Parsa, P. A. Parker, R. N. Scott, "Performance analysis of a crosstalk resistant adaptive noise canceller," IEEE Trans. Circuits and Systems, Vol. 43, 473–482, 1996.

[11] R. L. Zinser, G. Mirchandani, J. B. Evans, "Some experimental and theoretical results using a new adaptive filter structure for noise cancellation in the presence of crosstalk," Proc. ICASSP '85, 1253–1256, Mar. 1985.

[12] M. J. Al-Kindi, J. Dunlop, "A low distortion adaptive noise cancellation structure for real time applications," Proc. ICASSP '87, 2153–2156, Apr. 1987.

[13] S. Ikeda and A. Sugiyama, "An adaptive noise canceller with low signal-distortion in the presence of crosstalk," IEICE Trans. Fund, 1517–1525, Aug. 1999.