

A LOW DISTORTION NOISE CANCELLER WITH A NOVEL STEPSIZE CONTROL AND CONDITIONAL CANCELLATION

Akihiko Sugiyama and Ryoji Miyahara[†]

Information and Media Processing Labs., NEC Corporation

[†]Internet Terminal Division, NEC Engineering

1753, Shimonumabe, Nakahara-ku, Kawasaki-shi, Kanagawa 211-8666, JAPAN

ABSTRACT

This paper proposes a low-distortion noise canceller with a novel stepsize control and conditional cancellation. The coefficient adaptation stepsize is controlled by 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 output, and converted to a stepsize by an exponential function. This stepsize provides robustness to interference by the desired speech. Conditional cancellation guarantees that the noisy signal power is reduced by noise-replica subtraction. Comparison of the proposed noise canceller with five popular state-of-the-art commercial smartphones demonstrates good enhanced-signal quality with as much as 0.6 PESQ improvement.

Index Terms— Two microphone, Dual microphone, Low distortion, Noise canceller, Stepsize control

1. INTRODUCTION

Speech enhancement is an indispensable technology for communications and human-computer interaction in noisy environments. One of the most benefitting applications are mobilephone handsets. Most of today's handsets are equipped with two microphones for speech enhancement. There are three typical technologies for two-microphone, mobilephone speech enhancement; namely, two-channel noise suppression [1]–[10], acoustic beamforming [11]–[13], and noise cancellation [14]–[16].

Two-channel noise suppression uses the signal from the secondary microphone as additional information so that a more accurate noise estimate can be obtained. This accurate noise estimate is incorporated in the traditional single-channel noise suppression framework for better subtraction or better suppression with a more accurate spectral gain. However, auxiliary information obtained from the secondary microphone is not fully utilized because phase is still untouched in the process of suppression. Phase mismatch becomes a more serious problem in low signal-to-noise ratio (SNR) environments [17].

Acoustic beamforming, also known as microphone arrays (MAS), steers a beam and a null to enhance the target speech and attenuates undesirable interference. Although it manipulates magnitude and phase, it is useful only for point signal sources because it is based on directivity. Diffuse noise, which is often encountered in practical environments, cannot be attenuated with a limited number of microphones. Moreover, directivity in low frequencies is insufficient with a small microphone spacing [18] allowed for mobilephone handsets when they are placed side by side.

Noise cancellers (NCs) [15] do not have those limitations and have demonstrated potential in some applications [16]. A secondary microphone captures a signal which is correlated with the noise components in the primary-microphone signal. This signal drives an adaptive filter to generate a noise replica, which is subtracted from the primary-microphone signal to cancel noise. Adaptive filter coefficients are updated with the subtraction result, which consists of the speech to be enhanced 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 residual noise and enhanced speech [16].

As a solution to the interference problem, an adaptive noise canceller with a paired filter (ANC-PF) structure [19] 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 speech presence. A partitioned power-normalized proportionate normalized least-mean-square (PP-PNLMS) algorithm successfully scrapped the sub filter by calculating an SNR based on the main filter output [20]. However, it turned out by extensive evaluations that the algorithm is sometimes not sufficiently stable in extremely adverse environments.

This paper proposes a low-distortion noise canceller with a novel stepsize control and conditional cancellation. In the next section, SNR-based recursive stepsize control is reviewed in details to highlight an error propagation problem. Section 3 presents a new stepsize control and conditional cancellation. Finally, in Section 4, evaluation results of the

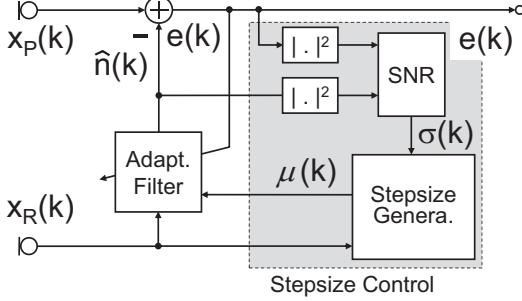


Fig. 1. Noise canceller with SNR-based recursive stepsize control.

new noise canceller are presented in comparison with the state-of-the-art commercial smartphones.

2. SNR-BASED RECURSIVE STEPSIZE CONTROL

Figure 1 depicts a blockdiagram of an NC with an N -tap adaptive filter based on SNR-based recursive stepsize control. The noise cancelled signal $e(k)$ is expressed by

$$e(k) = x_P(k) - \hat{n}(k) = s(k) + \Delta n(k) \quad (1)$$

$$\Delta n(k) = n(k) - \hat{n}(k) = n(k) - \sum_{l=k-N+1}^k x_R(l)w(k, k-l), \quad (2)$$

where $x_P(k)$, $x_R(k)$, $s(k)$, $n(k)$, and $\hat{n}(k)$ are the primary- and the reference-microphone signals, the desired speech, the noise to be cancelled, and a noise replica (adaptive filter output). $w(k, i)$ is the i -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 desired speech. With these replicas, $\hat{n}(k)$ and $e(k)$, of the noise and the desired speech, an estimated SNR, $\sigma(k)$, is calculated by

$$\sigma(k) = \text{ave}\{e^2(k)\}/\text{ave}\{\hat{n}^2(k)\}, \quad (3)$$

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

The SNR estimate, $\sigma(k)$ is then processed by an appropriate function $f\{\cdot\}$ to be converted to a stepsize $\tilde{\mu}(k)$ as in

$$\tilde{\mu}(k) = f\{\sigma(k)\} \cdot \mu_0. \quad (4)$$

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 $\sigma(k)$ such that a high SNR with a strong desired speech returns a small value for stable adaptation.

Because of time-averaging, the SNR estimate is somehow delayed. This delayed SNR estimate does not reflect rapid changes of the desired signal power, leading to an inappropriate stepsize and erroneous filter coefficients. Wrong coefficients violate the current assumption of good noise cancellation, and the SNR estimate becomes even worse. As a result, coefficients would not recover correct values because of

negative feedback or error propagation. Therefore, the coefficient adaptation algorithm has to be designed carefully paying more attention to the interfering desired speech.

3. PROPOSED NOISE CANCELLER

For sufficient stability in adverse conditions, the proposed NC incorporates three new functions; namely, individual stepsize control based on the reference power, global stepsize control based on an estimated SNR, and conditional cancellation of the noise. They cooperate for unrivalled robustness in the real environment.

3.1. Reference-Power Dependent Individual Stepsize Control

A coefficient $w(k, i)$ is updated by the NLMS algorithm as

$$w(k+1, i) = w(k, i) + \mu(k, i) \cdot \frac{e(k)x_R(k-i)}{\|x_R(k)\|^2}, \quad (5)$$

where $x_R(k)$ is a reference signal vector of the same size as the filter coefficient vector $w(k)$. From (1) and (5), the following equation is obtained.

$$\begin{aligned} w(k+1, i) = & w(k, i) + \mu(k, i) \cdot \frac{s(k)x_R(k-i)}{\|x_R(k)\|^2} \\ & + \mu(k, i) \cdot \frac{\Delta n(k)x_R(k-i)}{\|x_R(k)\|^2}. \end{aligned} \quad (6)$$

It was found through detailed investigations that coefficients with smaller magnitude tend to have larger variations, which sometimes lead to filter instability in adverse conditions. This fact can be explained by (6). The second term on the right-hand side in (6) is the interfering term in coefficient adaptation. A coefficient adaptation ratio $R(k, i)$ defined by

$$R(k, i) = \frac{\mu(k, i)|s(k)x_R(k-i)|}{\|x_R(k)\|^2 |w(k, i)|} = \frac{\mu(k, i)|s(k)|}{\|x_R(k)\|^2} \frac{|x_R(k-i)|}{|w(k, i)|} \quad (7)$$

represents a relative amount of coefficient adaptation to the coefficient value. Coefficients with too large a value of $R(k, i)$ may cause instability because they change drastically. This measure is relative because only small coefficients become unstable. It means that the first term on the right-hand side of (7) is common and only $|x_R(k-i)|/|w(k, i)|$ is of interest for us. Let us further define a normalized coefficient adaptation ratio $\bar{R}(k, i)$ as

$$\bar{R}(k, i) = R(k, i) \cdot \frac{\|x_R(k)\|^2}{\mu(k, i)|s(k)|} = \frac{|x_R(k-i)|}{|w(k, i)|}. \quad (8)$$

Coefficients with large $\bar{R}(k, i)$ values should not be adapted by limiting the coefficient change with scaling. Our task is to identify coefficients with such a large $\bar{R}(k, i)$ with a threshold

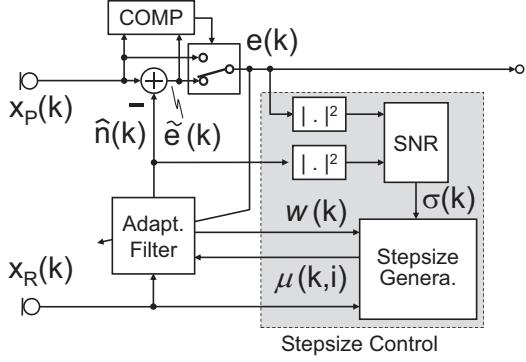


Fig. 2. Blockdiagram of the new noise canceller.

R_{th} . For computational savings, $\bar{R}_{max}(k, i)$ is used instead of $\bar{R}(k, i)$ as

$$\bar{R}_{max}(k, i) = \frac{|x_R(k - i)|}{\max\{|w(k, i)|\}} \leq \frac{|x_R(k - i)|}{|w(k, i)|} = \bar{R}(k, i). \quad (9)$$

As is clear from (9), use of $\bar{R}_{max}(k, i)$ makes the algorithm more stable. The final stepsize is given by

$$\mu(k, i) = \begin{cases} \frac{R_{th}}{\bar{R}_{max}(k, i)} \cdot \tilde{\mu}(k) & \bar{R}_{max}(k, i) > R_{th} \\ \tilde{\mu}(k) & \text{otherwise} \end{cases}, \quad (10)$$

where $\tilde{\mu}(k)$ is an SNR-dependent global stepsize.

3.2. SNR-Dependent Global Step size

Extensive evaluations revealed that approximation of the SNR-step size conversion function $f\{\cdot\}$ by a linear function does not always provide sufficiently small stepsize for some interference. Therefore, the proposed NC incorporates a decreasing exponential function as

$$\tilde{\mu}(k) = \max\{\min\{\alpha \exp\beta(\sigma(k) + \delta), \alpha\}, \epsilon\}. \quad (11)$$

Function $\tilde{\mu}(k)$ is illustrated in Fig. 3. Equation (11) indicates that the SNR-dependent global stepsize is a decreasing exponential function with a ceiling at α and a floor ϵ . It is shifted by δ toward left and scaled by α . Compared to an approximating linear function that crosses $\tilde{\mu}(k)$ at (δ, α) and (ρ, ϵ) , this function takes a small global stepsize more often in the transition range between δ and ρ . This fact guarantees higher stability for coefficient adaptation. Parameters in (10) and (11) were optimized with a wide range of realistic signals (SNRs, noise, crosstalk levels) and has proven insensitive.

3.3. Conditional Cancellation

Conditional cancellation subtracts the noise replica only when power reduction is guaranteed between the primary microphone signal and the error signal. It is implemented by the following equation:

$$e(k) = \begin{cases} x_p(k) - \hat{n}(k) & e^2(k) < x_p^2(k) \\ x_p(k) & \text{otherwise} \end{cases}. \quad (12)$$

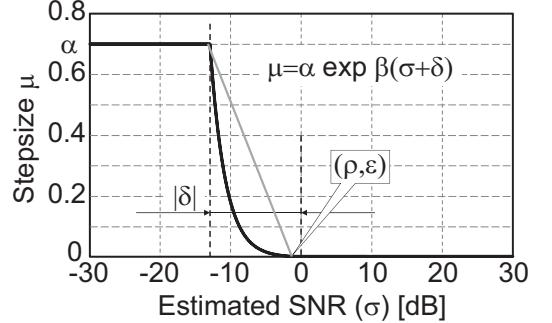


Fig. 3. SNR-dependent global stepsize.

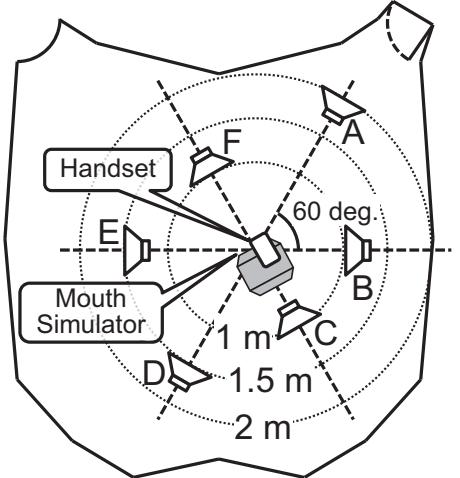


Fig. 4. Experimental setup.

This is a conservative operation, however, for sufficient stability, it plays an important role.

4. EVALUATIONS

4.1. Evaluation Conditions

Evaluations were performed with an $N = 512$ tap adaptive filter and compared with the state-of-the-art, popular smartphones; namely, iPhone4S, 5, 5C, 5S and Galaxy S4. A noise suppressor [21] is used for the proposed NC as post-processing. For fair comparison, the enhanced speech was encoded and decoded for the proposed NC by an AMR codec [22] at a bitrate of 12.2 kbit/s. The experimental setup is depicted in Fig. 4. Six loudspeakers were driven by the same signal that mostly consists of street noise, babble noise, or their mix. The primary and the reference microphones were mounted on an iPhone 5S at exactly the same microphone positions. The smartphone was placed 0.3 m above a table whose height was 1.0 m. The primary microphone was facing the table. The sound pressure level of the speech and the noise was approximately 80dBA at the primary microphone. SNRs for street, babble, and mixed noise conditions

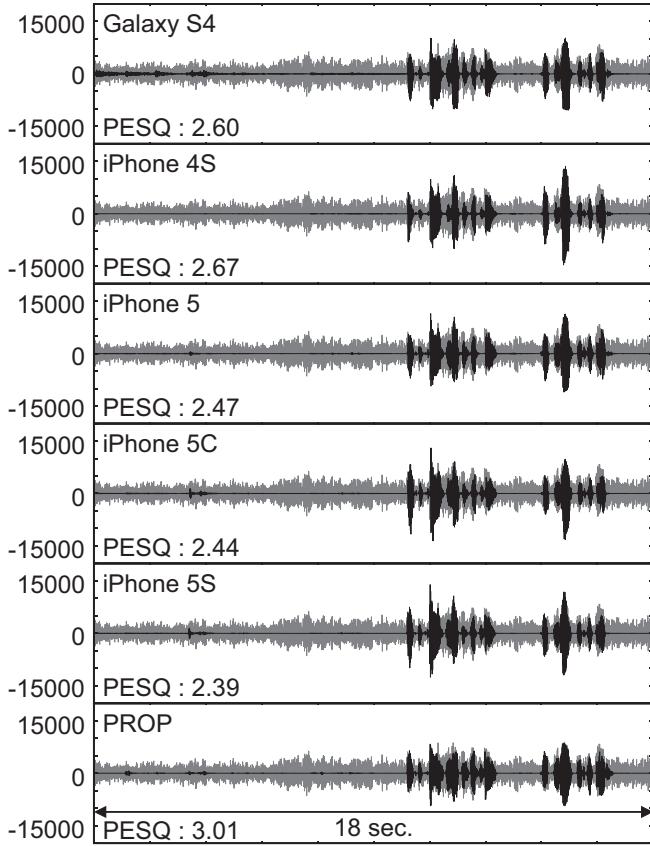


Fig. 5. Output (Street noise).

were -2 , $+8$, and $+1$ dB, respectively.

4.2. Evaluation Results

Figure 5 depicts the noisy speech (gray) and the enhanced speech (black) with the street noise. Although the signals look similar to each other, the PESQ scores [23] are considerably different. The proposed NC achieved as much as 0.6 better score over other five commercial smartphones. This is a sign of low distortion of the proposed NC.

Figure 6 shows PESQ scores for three different noise conditions. PESQ differences defined as the PESQ difference from the one by the proposed NC are also included. It is confirmed that the proposed NC always provides the best PESQ among the six smartphones.

Frequency distribution of frame PESQ are compared for iPhone5S and the proposed NC in Fig. 7. The proposed NC exhibits smoother and more natural distribution with a single peak around 3.8. iPhone5S, on the contrary, shows an irregular curve with many ups and downs with multiple peaks. The proposed NC achieves reasonable PESQ score depending on the frame SNR.

Shown in Fig. 8 are SNR improvement (SNRI) for the six smartphones. All smartphones except Galaxy show similar SNRI values. The proposed NC provides the best or

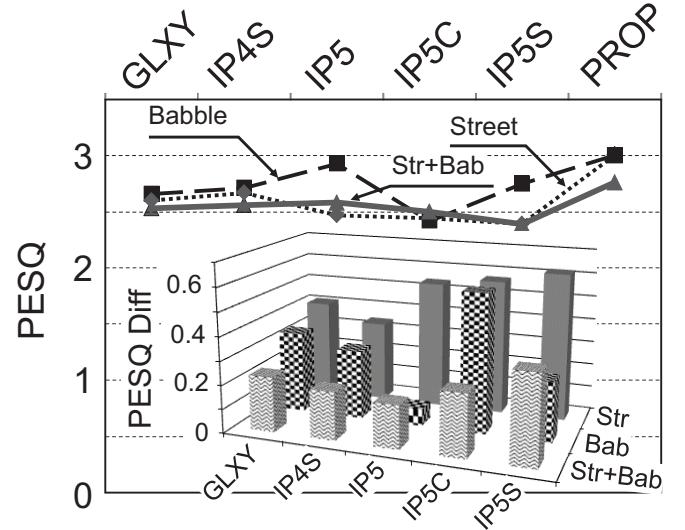


Fig. 6. PESQ and PESQ Diff. Diff is measured from the score of PROP.

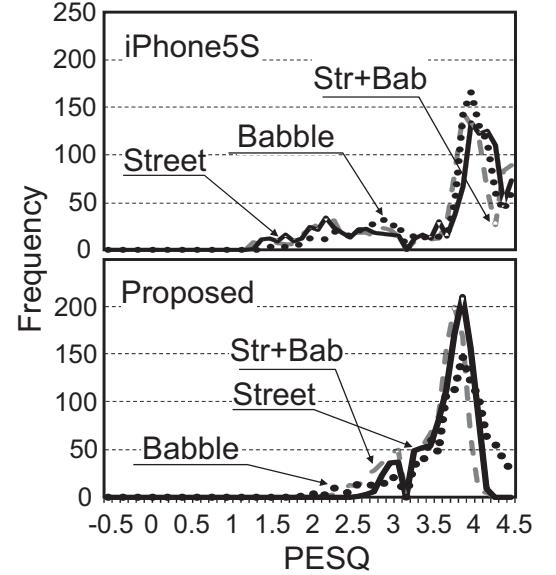


Fig. 7. Frequency distribution of frame PESQ.

comparable-to-the-best scores. For the babble noise, the SNRI by the proposed NC is lower than some of the others. However, such a difference is hardly audible in such an SNRI range higher than 25 dB.

5. CONCLUSION

A low-distortion noise canceller with a novel stepsize control and conditional cancellation has been proposed. The stepsize is controlled by adaptation amount of each coefficient and significance of interference. Conditional cancellation guarantees reduction of noise power for additional stability. Comparison of PESQ scores and SNRI values has demonstrated superior

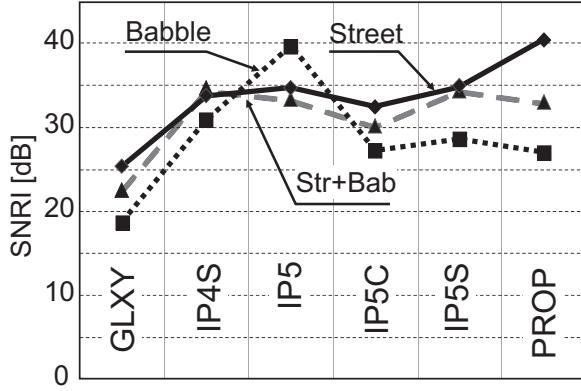


Fig. 8. Signal-to-noise ratio improvement (SNRI).

performance of the proposed noise canceller.

6. REFERENCES

- [1] M.-S. Choi and H.-G. Kang, “A two-channel minimum mean-square error log-spectral amplitude estimator for speech enhancement,” Proc. HSCMA2008, pp.152–155, May 2008.
- [2] J. Freudenberger, S. Stenzel and B. Venditti, “A noise PSD and cross-PSD estimation for two-microphone speech enhancement systems,” Proc. SSP2009, pp.709–712, Aug. 2009.
- [3] S. -Y. Jeong, K. Kim, J. -H. Jeong, K. -C. Oh, and J. Kim, “Adaptive noise power spectrum estimation for compact dual channel speech enhancement,” Proc. ICASSP2010, pp.1630–1633, Apr. 2010.
- [4] K. Kim, S. -Y. Jeong, J. -H. Jeong, K. -C. Oh, and J. Kim, “Dual channel noise reduction method using phase difference-based spectral amplitude estimation,” Proc. ICASSP2010, pp.217–220, Apr. 2010.
- [5] L. Watts, “Real-time, high-resolution simulation of auditory pathway, with application to cell-phone noise reduction,” Proc. ISCAS2010, pp.3821–3824, May 2010.
- [6] N. Yousefian and P. C. Loizou, “A dual-microphone speech enhancement algorithm based on the coherence function,” IEEE Trans. ASLP, Vol. 20, No. 2, pp.599–609, Feb. 2012.
- [7] M. Jeub, C. Herglotz, C. Nelke, C. Beaugeant, and P. Vary, “Noise reduction for dual-microphone mobile phones exploiting power level differences,” Proc. ICASSP2012, pp.217–220, Mar. 2012.
- [8] J. Zhang, R. Xia, Z. Fu, J. Li, and Y. Yan, “A fast two-microphone noise reduction algorithm based on power level ratio for mobile phone,” Proc. ICSLP2012, pp.206–209, Dec. 2012.
- [9] Z.-H. Fu, F. Fan and J. -D. Huang, “Dual-microphone noise reduction for mobile phone application,” Proc. ICASSP2013, pp.7239–7243, May 2013.
- [10] J. Taghia, R. Martin, J. Taghia and A. Leijon, “Dual-channel noise reduction based on amixture opf circular-symmetric complex gaussians on unit hypersphere,” Proc. ICASSP2013, pp.7289–7293, May 2013.
- [11] J. Chen, L. Shue, K. Phua and H. Sun, “Theoretical comparisons of dual microphone systems,” Proc. ICASSP2004, pp.73–76, May 2004.
- [12] J. Chen, L. Shue, K. Phua and H. Sun, “Experimental study of dual microphone systems,” Proc. ICME2004, pp.1519–1522, Jun. 2004.
- [13] Z. Koldovský, P. Tichavský, and D. Botka “Noise reduction in dual-microphone mobile phones using a bank of pre-measured target cancellation filters,” Proc. ICASSP2013, pp.679–683, May 2013.
- [14] X. Zhang, H. Zeng, and A. Lunardhi, “Noise estimation based on an adaptive smoothing factor for improving speech quality in a dual-microphone noise suppression system,” Proc. ICSPCS2011, pp.1–5, Dec. 2011.
- [15] 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), pp.1692–1716, 1975.
- [16] 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.
- [17] A. Sugiyama and R. Miyahara, “Phase randomization, - A new paradigm for single-channel signal enhancement,” Proc. ICASSP2013, pp.7487–7491, May 2013.
- [18] A. Sugiyama and R. Miyahara, “A new generalized side-lobe canceller with a compact array of microphones suitable for mobile terminals,” Proc. ICASSP2014, pp.820–824, May 2014.
- [19] 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.
- [20] A. Sugiyama, M. Kato, and M. Serizawa, “A low-distortion noise canceller with an SNR-modified partitioned power-normalized PNLMs algorithm,” Proc. APSIPA ASC 2009, pp.222–225, Oct. 2009.
- [21] 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.
- [22] “Digital cellular telecommunications system (Phase 2+); Adaptive Multi-Rate (AMR); speech processing functions; General description,” 3GPP TS 06.71 Release 98.
- [23] “Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” ITU-T P.862, Feb. 2002.