A NOISE ROBUST HEARABLE DEVICE WITH AN ADAPTIVE NOISE CANCELLER AND ITS DSP IMPLEMENTATION

Akihiko Sugiyama[1,4], *Ryoji Miyahara*[2], *Kouji Oosugi*[3]

[1]Data Science Research Laboratories, NEC Corporation,
[2]Embedded System Division, NEC Platforms, Ltd.
[3]Business Development Division, NEC Corporation,
1753 Shimonumabe, Nakahara-ku, Kawasaki 211–8666, Japan
[4] Faculty of System Design, Tokyo Metropolitan University
6-6 Asahigaoka, Hino, 191–0065, Japan

ABSTRACT

This paper presents a noise-robust hearable device with an adaptive noise canceller. The hearable device has an inner-ear and an outer-ear microphone to collect ear-canal speech signal and the outer-ear environmental signal, respectively. The environmental signal drives an adaptive filter to generate a noise replica which is subtracted from the inner-ear microphone signal to cancel the noise. Coefficients of the adaptive filter are updated by an NLMS (normalized least mean square) algorithm with an SNR-controlled stepsize for small speech distortion. The adaptive noise canceller is realized on a DSP (digital signal processor) chip with 39 McPS for 64 taps. Evaluation results in speech recognition demonstrates that a word recognition rate of 89% is achieved when the noise level is 80 dBA with 38% improvement over the inner-ear signal without adaptive noise cancellation.

Index Terms— Hearable device, Signal enhancement, Speech recognition, Noise canceller, SNR estimate, Stepsize

1. INTRODUCTION

Voice control capabilities of consumer products have been drawing attention through smart speakers. An important technology in such systems is automatic speech recognition that is sensitive to the speaker-microphone distance and surrounding noise. A close-talking or a bone-conduction microphone is effective, but bothersome.

Microphone earphones can be a solution, which are often Bluetooth-connected with a consumer device and not bothersome. Those with additional functions such as personal authentication for security [1] are called hearable device. Their built-in microphone in the ear canal captures user voice. Noise from outside the ear is attenuated by the ear bud, resulting in a higher signal-to-noise ratio (SNR) than that of the speech outside the ear. However, the noise leaking into the ear canal is not always sufficiently small. As a result, in adverse noise environment, speech recognition performance is not as good as required. Some of microphone earphones have active noise control function, as shown in Fig. 1, that cancels the incoming noise using an acoustic signal with the same magnitude but the opposite phase. One may think that it is useful, however, it is not the case.

A reference signal $r(k)$ obtained by an outer-ear microphone is modified by an adaptive filter to generate a noise replica $-\hat{n}_A(k)$. The adaptive filter models the acoustic characteristic from the outer-ear microphone to the speaker in the ear canal. The noise replica is mixed with a signal $y(k)$ to be heard to generate a signal $z(k) = y(k) - \hat{n}_A(k)$ which is radiated in the ear canal. When $z(k)$ is mixed

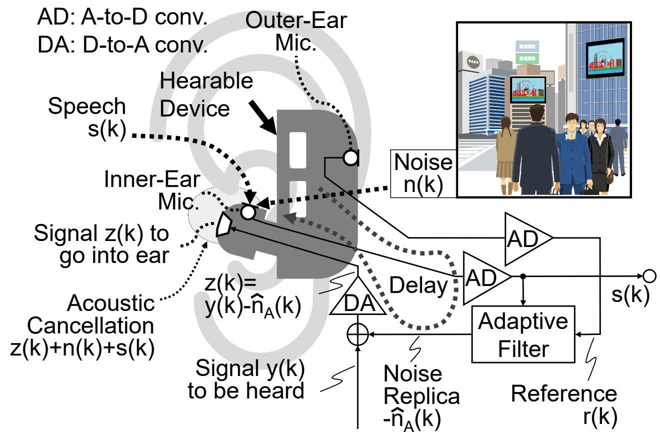

Fig. 1. Microphone earphone with active noise cancellation (Conventional microphone earphone). Noise is cancelled acoustically by radiating a noise replica in the ear canal.

with the incoming noise $n(k)$ and the target speech $s(k)$, active noise cancellation is achieved. Assuming that the user is not listening when speaking, $y(k) = 0$ and the inner ear microphone captures a signal $s(k)+n(k)-\hat{n}_A(k)$. Coefficients of the adaptive filter are updated to minimize the error-signal power $\{n(k) - \hat{n}_A(k)\}^2$. Nevertheless, acoustic cancellation necessitates analog-to-digital (AD) and digital-to-analog (DA) conversions in the noise-replica generation path that introduce delay. This delay degrades noise cancellation in high frequencies where a certain delay corresponds to more samples than in low frequencies. Moreover, active noise control equalizes a peaky frequency characteristic of the speaker in the signal path, leading to a long FIR (finite impulse response) filter. Active noise control does not provide sufficient cancellation for the inner ear microphone. In order to provide wideband noise cancellation with a reasonable size of the filter, electrical noise control is desirable.

This paper presents a noise robust hearable device with an adaptive noise canceller and its implementation on a digital signal processor (DSP). The noise leaking into the ear canal is cancelled with the help of an adaptive filter controlled by a novel adaptation algorithm, resulting in low distortion speech suitable for speech recognition as well as communication. In the following section, the noise robust hearable device is presented followed by details of DSP implementation. Section 5 presents evaluation results by computer simulation and a hardware device operating in realtime.

2. ELECTRICAL NOISE CONTROL WITH TWO MICROPHONES

There are three typical approaches to two-microphone electrical noise control; namely, two-channel noise suppression [2]–[12], acoustic beamforming [13]–[19], and noise cancellation [20]–[28]. Two-channel noise suppression uses the signal from the secondary microphone as additional information to obtain a more accurate noise estimate. 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 ignored in the process of suppression. Phase mismatch becomes a more serious problem in low signal-to-noise ratio (SNR) environments [29].

Acoustic beamforming, also known as microphone arrays, steers a beam and a null to enhance the target speech and attenuates 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, in hearable devices, one microphone is placed inside and the other outside the ear canal. There is an acoustic barrier in between that prevents the two microphone signals from forming directivity.

Adaptive noise cancellers do not have those limitations and have demonstrated potential in some applications [28, 30, 31]. 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 [24]. This problem can be offset by appropriately controlling a stepsize for coefficient adaptation based on an estimated SNR. Ikeda et al. [21] proposed use of a sub adaptive filter for SNR estimation and Sugiyama et al. [28] later showed that an SNR can be estimated without the sub adaptive filter.

3. HEARABLE DEVICE WITH AN ADAPTIVE NOISE CANCELLER

Figure 2 depicts a blockdiagram of the proposed noise-robust hearable device with an adaptive noise canceller. Noise replica $\hat{n}(k)$ is generated in time because its path has a single ADC, which is equal to the signal path for $x_p(k)$. An N-tap adaptive filter based on SNR-based stepsize control generates a noise replica to be "electrically" subtracted from the inner-ear microphone signal. The noise cancelled signal $e(k)$ is expressed by

$$e(k) = \begin{cases} x_P(k) - \hat{n}(k) \\ = s(k) + \Delta n(k) & e^2(k) < x_P^2(k) \\ x_P(k) & otherwise \end{cases} \quad (1)$$

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

where $x_P(k)$, $x_R(k)$, $s(k)$, $n(k)$, and $\hat{n}(k)$ are the inner-ear and the outer-ear 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. Equation (1) represents conditional subtraction [28] which performs subtraction only when a condition is satisfied. Conditional subtraction is marked as a boxed "+" in Fig. 2.

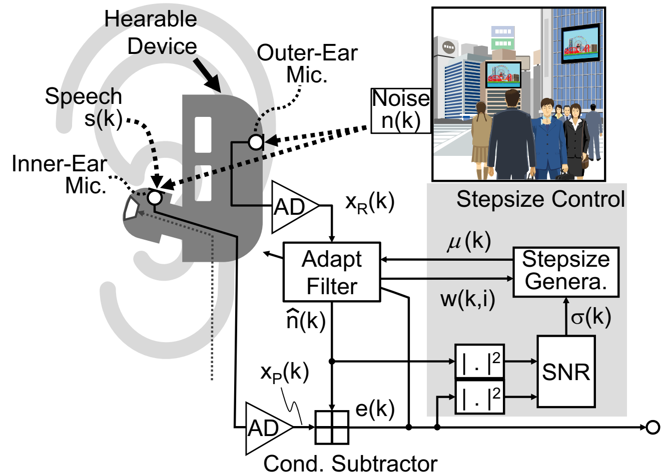

Fig. 2. Proposed hearable device with an adaptive noise canceller. Noise is cancelled electrically with a conditional subtractor in contrast to Fig. 1.

A coefficient $w(k,i)$ is updated by an NLMS (normalized least mean-square) algorithm as

$$w(k+1,i) = w(k,i) + \mu(k)\frac{e(k)x_R(k-i)}{||x_R(k)||^2}, \quad (3)$$

$$\mu(k) = \begin{cases} \frac{R_{th}}{\bar{R}_{max}(k,i)} \cdot \tilde{\mu}(k) & \bar{R}_{max}(k,i) > R_{th} \\ \tilde{\mu}(k) & otherwise \end{cases}, \quad (4)$$

$$\bar{R}_{max}(k,i) = \frac{|x(k-i)|}{\max\{|w(k,i)|\}}, \quad (5)$$

where $\boldsymbol{x}_R(k)$ is a reference signal vector of size N and $\tilde{\mu}(k)$ is an SNR-dependent stepsize [21, 28] given by

$$\tilde{\mu}(k) = \max\{\min\{\alpha \exp\beta(\sigma(k)+\delta), \alpha\}, \epsilon\}, \quad (6)$$

$$\sigma(k) = e^2(k)/\hat{n}^2(k), \quad (7)$$

where $\sigma(k)$ is an SNR estimate. $\tilde{\mu}(k)$ is a decreasing function of $\sigma(k)$ [21, 28] such that a high SNR with a strong desired speech returns a small value for stable adaptation.

4. DSP IMPLEMENTATION

The adaptive noise canceller was implemented on *Kalimba* DSP core with 24-bit arithmetic operating at 120 MHz which is integrated into Qualcomm CSR8675 [32]. Specifications of the DSP core are summarized in Tab. 1.

The total number of computations was originally 52 million cycles per second (McPS) for the adaptive noise canceller (ANC) alone and 162 McPS for ANC+cVc[1]. It should be noted that the number of taps for adaptive noise canceller was reduced to $N = 64$ for easy implementation. Code optimizations was performed in the following four steps:

[1] cVc is a noise cancellation technology by Qualcomm [33].

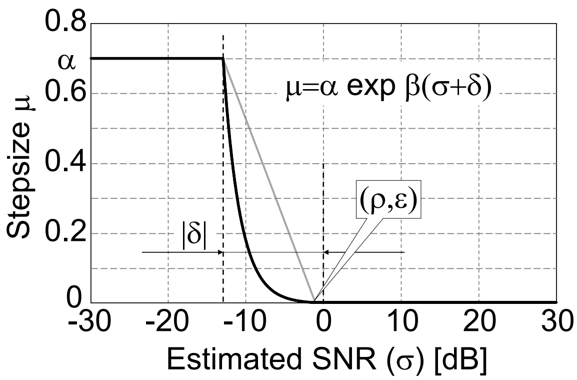

Fig. 3. Estimated SNR $\sigma(k)$ vs. stepsize $\tilde{\mu}(k)$.

Table 1. Specifications of Kalimba

Data accuracy	24-bit fixed point arithmetic
Speed	120 MIPS
Instruction	32-bit×12K word
Data memory	24-bit×64K word
MAC	24×24-bit MPY & 56-bit ACC
Barrel shifter	56-bit in/24-bit out
Division	12-cycle

1. Reduction in the number of adaptive filter taps (N=256 → N=64).
2. Parallelization of filtering, power calculation, and coefficient adaptation.
3. Algorithm simplification by disabling branches.
4. Limiting the maximum number of cycles in loops.

Equations (4) and (5) are calculated for each coefficient and conditional subtraction in Fig. 2 is heavy for DSP implementation. Thus, in Step 3, (4) and (5) were replaced with

$$\mu(k) \quad = \quad \begin{cases} 2/3 & \tilde{\mu}(k) > 2/3 \\ \tilde{\mu}(k) & otherwise \end{cases} , \qquad (8)$$

where the stepsize $\mu(k)$ is common to all coefficients. Conditional subtraction was replaced by subtraction and detailed stability check for each coefficient is removed. Step 4 was to offset a Kalimba specific constraint. The progress in code optimization is summarized in Fig. 4. The required computations for $N = 64$ is 29 and 39 McPS for adaptive noise canceller and adaptive noise canceller and cVc. It is sufficiently small for realtime operation on a 120MHz DSP core.

5. EVALUATIONS

5.1. Computer Simulation

Evaluations were performed with 50 phoneme-balanced speech sentences spoken by a male speaker in a soundproof room and recorded as $s(k)$. A babble noise from an ETSI (European Telecommunications Standards Institute) noise data base was reproduced from a loudspeaker at another time in the same room and recorded as $n(k)$. The speech and the noise were sampled at 16 kHz upon recording. The room was $1.8 \times 2.2 \times 2.2$ m in size and had a layout as illustrated in Fig. 5. The speech and the noise were mixed at different SNRs. The number of adaptive filter taps was set to $N = 256$. Other parameters for coefficient adaptation were set as in [28].

Figure 6 depicts the clean speech as well as the raw signal $x_P(k)$ direct from the inner-ear microphone (black) and $e(k)$ after noise cancellation (gray) at SNRs of 0, 6, and 12 dB. The figure demonstrates that good noise cancellation is attained.

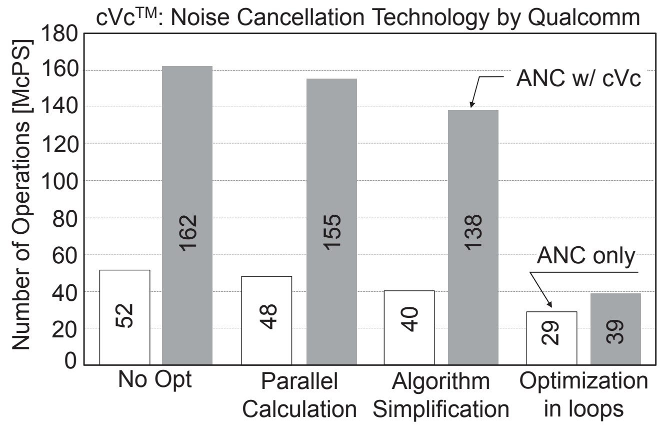

Fig. 4. Progress in code optimization.

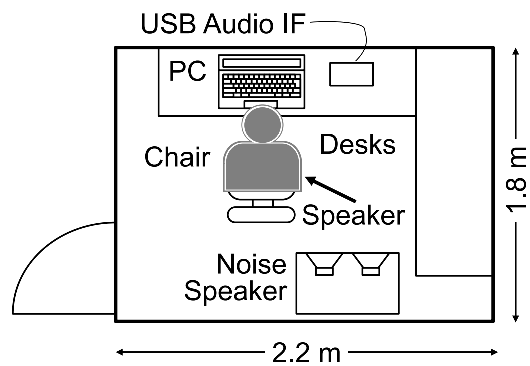

Fig. 5. Layout of the sound proof room. The speaker wears the proposed hearable device under evaluation.

Full-text matching rates, which were calculated character-by-character without delimiters, are shown in Fig. 7 at SNRs of 0, 6, and 12 dB in addition to the clean speech condition. Although it is a challenging task, the adaptive noise canceller improved the full-text matching rate as much as 38% at an SNR of 0 dB. The raw inner-microphone signal always failed in text matching whereas the noise-cancelled signal achieved 38%.

Fig. 8 shows word recognition rates which correspond, with respect to the SNRs, to Fig. 7. The word recognition rate was degraded as the SNR decreased for the raw inner-microphone signal. They were 92%, 76%, and 22% for 12, 6, and 0 dB SNR. On the contrary, the noise-cancelled signal kept good word recognition rates at low SNRs. 93%, 91%, and 87% were obtained at SNRs of 12, 6, and 0 dB. The improvements were 0%, 14%, and 65%, respectively.

5.2. Realtime Evaluation by Hardware

In this evaluation, the male speaker wore the hearable device under evaluation and read the same 50 phoneme-balanced sentences in the soundproof room. The ETSI babble noise was reproduced simultaneously from the loudspeaker. A natural mixing of the speech and the noise took place in the sound proof room. The sampling frequency of the hearable device was 16 kHz and the number of taps with the adaptive filter was reduced to $N = 64$ for a constraint on the computational load. Other parameters were equal to the evaluation by computer simulation.

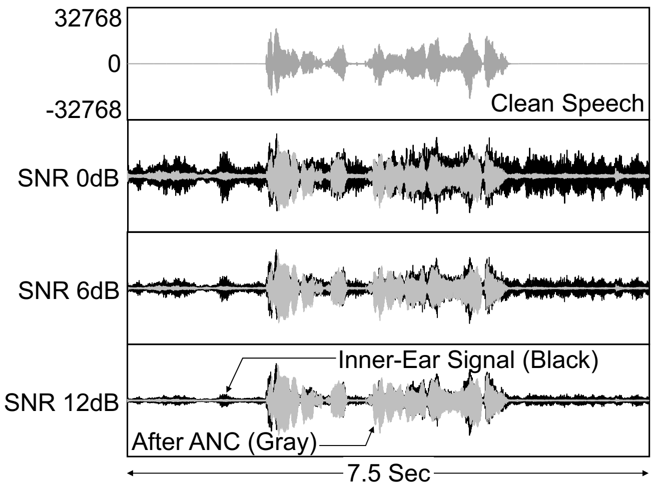

Fig. 6. Signal before and after noise cancellation in the same scale.

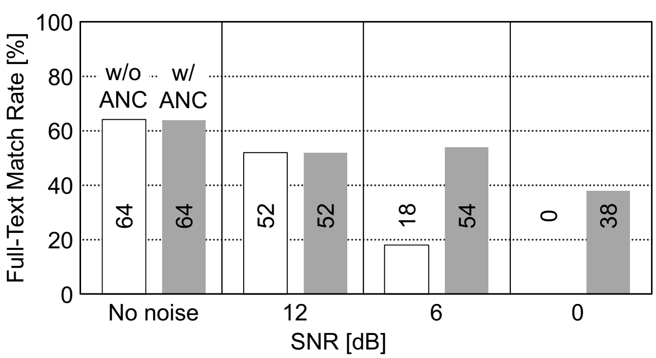

Fig. 7. Full-text matching rate (Computer simulation).

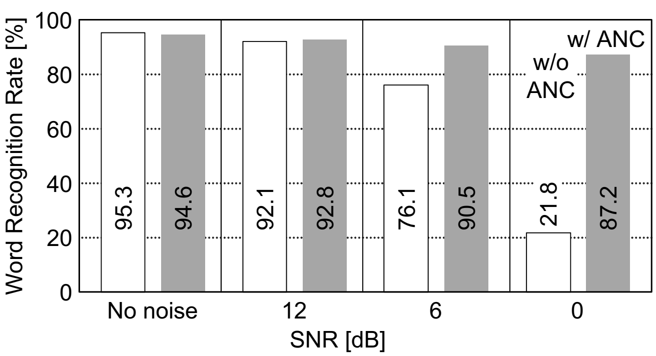

Fig. 8. Word recognition rate (Computer simulation).

Figure 9 illustrates the raw signal $x_P(k)$ direct from the inner-ear microphone (black) and $e(k)$ after noise cancellation (gray). Good noise cancellation is demonstrated.

Full-text matching rates are shown in Fig. 10 with outer-ear noise levels of 35, 70, 80, and 90 dBA. The 35 dBA noise level represents almost noise-free environment. The adaptive noise canceller improved the full-text matching rate as much as 48% at a noise level of 80 dBA. The raw inner-microphone signal attained 4% whereas the noise-robust hearable device achieved 52%.

Shown in Fig. 11 are word recognition rates which correspond to Fig. 10 with respect to the noise levels. The word recognition rate was degraded as noise level increased for the raw inner-microphone signal. They were 51% and 23% for 80 and 90 dBA noise. On the contrary, the noise robust hearable device kept good word recognition rates at high noise levels. 89% and 72% were obtained with 80 and 90 dBA noise. The improvements were 38% and 44.8%, respectively.

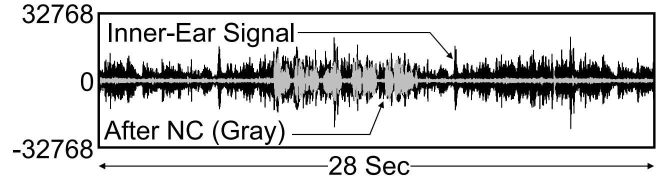

Fig. 9. Signal before and after noise cancellation (w/ 80 dBA noise).

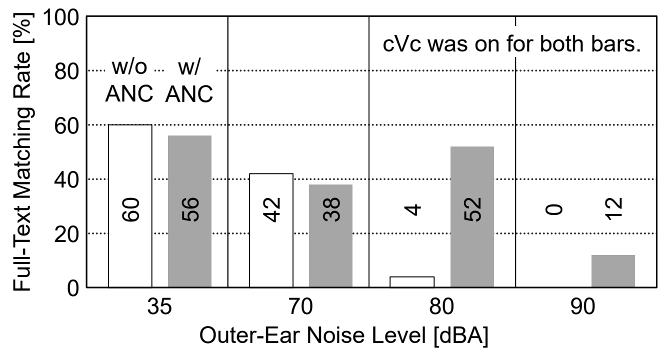

Fig. 10. Full-text matching rate (DSP implementation).

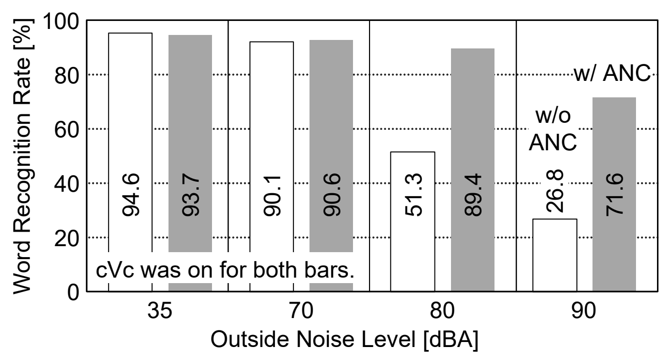

Fig. 11. Word recognition rate (DSP implementation).

Figures 10 and 7 as well as Figs. 11 and 8 exhibit similar results. The results without ANC demonstrates that cVc was not sufficiently effective because cVc cannot be disabled. The hearable device with an adaptive noise canceller was successfully implemented on a DSP.

6. CONCLUSION

A noise-robust hearable device with an adaptive noise canceller has been presented with DSP implementation. An adaptive FIR filter with an NLMS algorithm incorporating an SNR controlled stepsize has been employed for better noise-cancelled speech quality. Noise leaking into the ear canal and contaminates the desired speech captured in the ear canal has successfully been cancelled with good speech quality. The DSP code has been optimized to 39 McPS for a 64-tap adaptive FIR filter in the adaptive noise canceller. Evaluation results in speech recognition have demonstrated that a word recognition rate of 89% is achieved when the noise level is 80 dBA with 38% improvement over the raw inner-ear signal.

7. REFERENCES

[1] T. Arakawa, T. Koshinaka, S. Yano, H. Irisawa, R. Miyahara and H. Imaoka, "Fast and accurate personal authentication using ear acoustics," Proc. APSIPA2016 Dec. 2016.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] L. Watts, "Real-time, high-resolution simulation of auditory pathway, with application to cell-phone noise reduction," Proc. ISCAS2010, pp.3821–3824, May 2010.

[7] 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.

[8] K. Li, Y. Guo, Q. Fu, and Y. Yan, "A two microphone-based approach for speech enhancement in adverse environments," Proc. ICCE2012, pp.41–42, Jan. 2012.

[9] 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.

[10] 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.

[11] Z.-H. Fu, F. Fan and J. -D. Huang, "Dual-microphone noise reduction for mobile phone application," Proc. ICASSP2013, pp.7239–7243, May 2013.

[12] J. Taghia, R. Martin, J. Taghia and A. Leijon, "Dual-channel noise reduction based on a mixture of circular-symmetric complex gaussians on unit hypersphere," Proc. ICASSP2013, pp.7289–7293, May 2013.

[13] J. Chen, L. Shue, K. Phua and H. Sun, "Theoretical comparisons of dual microphone systems," Proc. ICASSP2004, pp.73–76, May 2004.

[14] J. Chen, L. Shue, K. Phua and H. Sun, "Experimental study of dual microphone systems," Proc. ICME2004, pp.1519–1522, Jun. 2004.

[15] J. Beh, R. H. Barran and H. Ko, "Dual-channel based speech enhancement using novelty filter for robust speech recognition in automobile environment," IEEE Trans. CE, Vol. 52, No. 2, pp.583–589, May. 2006.

[16] N. -V. Vu, H. Ye, J. Whittington, J. Devlin, and M. Mason, "Small footprint implementation of dual-microphone delay-and-sum beamforming for in-car speech enhancement," Proc. ICASSP2010, pp.1482–1485, Apr. 2010.

[17] 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.

[18] Z. Koldovský, J. Malek, P. Tichavský, and F. Nesta "Semi-blind noise extraction using partially known position of the target source," IEEE Trans. ASLP, Vol. 21, No. 10, pp.2029–2041, Oct. 2013.

[19] A. Sugiyama and R. Miyahara, "A new generalized sidelobe canceller with a compact array of microphones suitable for mobile terminals," Proc. ICASSP2014, pp.820-824, May 2014.

[20] 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.

[21] 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.

[22] S. Ikeda and A. Sugiyama, "An Adaptive Noise Canceller with Low Signal-Distortion in the Presence of Crosstalk," IEICE Trans. Fund., vol. E82-A, No.8, pp.1517–1525, Aug. 1999.

[23] M. Sato, A. Sugiyama, and S. Ohnaka, Miki Sato, Akihiko Sugiyama, and Shino?=fichi Ohnaka, "An Adaptive Noise Canceller with Low Signal-Distortion Based on Variable Stepsize Subfilters for Human-Robot Communication," IEICE Trans. Fund., vol. E88-A, No.8, pp.2055–2061, Aug. 2005.

[24] 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.

[25] 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.

[26] 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.

[27] S. Khorram, H. Sameti, and H. Veisi, "An optimum MMSE post-filter for adaptive noise cancellation in automobile environment," Proc. ISSPA2012, pp.431–435, Jul. 2012.

[28] 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.

[29] A. Sugiyama and R. Miyahara, "Phase randomization, - A new paradigm for single-channel signal enhancement," Proc. ICASSP2013, pp.7487–7491, May 2013.

[30] M. Sato, A. Sugiyama, O. Hoshuyama, N .Yamashita, S. Ohnaka, and Y. Fujita, "Audio Interface of a Personal Robot 'PaPeRo," Journal of Acoustical Society of Japan, vol. 62, No.3, pp.173–181, Mar. 2006 (in Japanese).

[31] M. Sato, T. Iwasawa, A. Sugiyama, T. Nishizawa, and Y. Takano, "A Single-Chip Speech Dialogue Module and its Evaluation on a Personal Robot, PaPeRo-mini," IEICE Trans. Fund., vol. E93-A, No.1, pp.261–271, Jan. 2010.

[32] "Qualcomm CSR8675 Bluetooth Audio Platform." [Online] Available: https://www.qualcomm.com/media/documents/files/csr8675-product-brief.pdf

[33] "cVc noise cancellation technologies." [Online] Available: https://www.qualcomm.com/solutions/voice-and-music/features/cvc-noise-cancellation-technology