Wiener solution considering cross-spectral term between echo and near-end speech for acoustic echo reduction

Masahiro Fukui¹,*; Suehiro Shimauchi¹,†; Yusuke Hioka²,‡; Akira Nakagawa¹,§; Yoichi Haneda³,§; Akitoshi Kataoka⁴,‖ and Hitoshi Ohmuro¹,∗∗

¹NTT Media Intelligence Laboratories, NTT Corporation, 3–9–11 Midori-cho, Musashino, 180–8585 Japan
²Department of Electrical and Computer Engineering, University of Canterbury, Private Bag 4800, Christchurch, 8140 New Zealand
³Faculty of Informatics and Engineering, The University of Electro-Communications, 1–5–1 Chofugaoka, Chofu, 182–8585 Japan
⁴Faculty of Science and Technology, Ryukoku University, 1–5 Yokotani, Seta Oe-cho, Otsu, 520–2914 Japan

(Received 30 July 2013, Accepted for publication 8 January 2014)

Abstract: This paper introduces a frequency-domain acoustic echo reduction process based on a new Wiener-filtering method taking into account the cross-spectral term between the acoustic echo and the near-end speech. The conventional echo reduction method based on Wiener filtering estimates the gain based on the assumption that the cross-spectral term of the echo and the near-end speech is zero because the acoustic echo and the near-end speech are statistically uncorrelated. However, this assumption does not always hold true in practice because the gain is estimated in a very short period where the amount of statistical data, which is used to calculate the ensemble averages of the observed signals, is insufficient. As a result, the conventional method occasionally causes the perceptual degradation of sound quality during a double-talk situation; therefore, the performance is still not sufficient. Our goal was to accurately calculate the echo-reduction gain to decrease the speech distortions produced by the echo-reduction process. The proposed method solves a least mean square error of the Wiener-filtering method by taking into account the cross-spectral term between the echo and the near-end speech to obtain a better echo-reduction gain. The performance of this method was demonstrated by objective and subjective results in which speech distortions were decreased.

Keywords: Acoustic echo, Echo reduction, Wiener filtering, Acoustic-coupling level

PACS number: 43.60.Dh [doi:10.1250/ast.35.150]

1. INTRODUCTION

An acoustic echo canceller (AEC) is a key technology in handsfree telecommunication systems (i.e., video conferencing, and videophones). It is necessary to prevent annoying howling and to eliminate an undesired acoustic echo component from the microphone input signal that includes the echo signal, near-end speech signal, and background noise signal. In most AECs, adaptive digital filter (ADF) [1,2] and echo reduction (ER) [3–14] processes are used to achieve such purposes to remove the echo signal. The ADF is a linear-processing method; therefore, it can cancel out only the echo signal from a microphone input signal by adaptively modeling an unknown acoustic echo path. However, the ADF cannot sufficiently eliminate the echo right after an echo-path change because of the slow convergence speed of the ADF. And so in practice the ER processes are also used in series after the ADF as a post filter to suppress the residual echo signal.

The ER suppresses the echo signal by multiplying the microphone input signal by an echo-reduction gain, which is called ER gain, in the frequency domain. In general,
Wiener-filtering (WF) method [15,16] has been widely used in ER-gain estimation [3–10]. This method calculates the ER gain based on the assumption that echo and near-end speech signals are statistically uncorrelated and so a cross-spectral term of their signals is zero. However, as a practical issue in the WF, the ER gain needs to be calculated from a very short period because most echo and near-end speech signals are usually nonstationary; therefore, the cross-spectral term between echo and near-end speech signals will not always become zero because the number of samples used in calculating the cross-spectral term is small. As a result, the ER process based on the conventional WF method causes speech distortions during double-talk situations.

In this study, for improving the performance of ER, we developed a modified ER-gain estimation method to decrease speech distortions. While conventional WF assumes that the cross-spectral term of the echo and near-end signals is zero, the proposed method estimates the ER gain based on the assumption that the cross-spectral term of their signals is not zero because the time-sequence period is short. The proposed ER gain is obtained by solving a least mean square error of the WF method by taking into account the cross-spectral term of the signals. An advantage of this strategy is that it is able to accurately calculate the ER gain even in a short period.

This paper is based on a workshop paper which has been published in the International Workshop on Acoustic Signal Enhancement (IWAENC 2012) [17]. In this paper, we further added comparison between the conventional and proposed methods in terms of the estimation accuracy of the ER gains. In addition, this paper newly performs the subjective experimental test complying ITU-R BS.1534-1 [18], which is called “MULTi Stimulus test with Hidden Reference and Anchor (MUSHRA),” in order to evaluate the speech quality during the double-talk period.

The remainder of this paper is organized as follows. Section 2 presents the principle of the ER processes based on the WF method and its problem. Section 3 provides details of the proposed method and the simulation results of the gain estimation accuracy. Section 4 describes the experimental results, and the paper is concluded in Sect. 5.

2. ECHO REDUCTION AND ITS PROBLEM

2.1. STSA-based Echo Reduction

This section derives the echo-reduction (ER) process on the basis of a short-time spectral amplitude (STSA) estimation [19,20]. The ER process included in an AEC is illustrated in Fig. 1. The microphone input signal $y(n)$ can be written a summation of the acoustic echo signal $d(n)$ and the desired (target) near-end speech signal $s(n)$:

$$y(n) = d(n) + s(n), \quad (1)$$

where $d(n)$ is modeled by the convolution of the received speech signal $x(n)$ and the echo path $h(n)$:

$$d(n) = x(n) * h(n), \quad (2)$$

where $*$ denotes convolution. The short-time spectrum of $y(n)$ is represented as

$$Y_i(\omega) = D_i(\omega) + S_i(\omega), \quad (3)$$

where $\omega$ is a discrete frequency index, $i$ is a discrete time-frame index, and $D_i(\omega)$ and $S_i(\omega)$ are the short-time spectra of $d(n)$ and $s(n)$, respectively. The ER can be expressed as

$$\hat{S}_i(\omega) = G_i(\omega)Y_i(\omega), \quad (4)$$

where $\hat{S}_i(\omega)$ is the estimate of $S_i(\omega)$, and $G_i(\omega)$ is the ER gain. The obtained estimate $\hat{S}_i(\omega)$ is transformed into the time domain signal $\hat{s}(n)$, which is the send signal, by an inverse fast Fourier transform (IFFT).

The ER gain is for example calculated by Wiener filtering (WF) method [15,16] obtained by the following equation:

$$G_i^W(\omega) = \frac{E[|S_i(\omega)|^2]}{E[|S_i(\omega)|^2] + E[|D_i(\omega)|^2]}, \quad (5)$$

**Fig. 1** Block diagram of ER process.
\[
\hat{H}_i(\omega) = \left[ \sum_{r=-R_\theta}^{R_\theta} \sum_{k=0}^{N-1} X_{i-k} \omega (r) Y_{i-k} (\omega + r) \right]^2
\]

where * is a complex conjugate, and \( N \) and \( R_\theta \) indicate the number of frames and number of frequency bins required for the average calculations, respectively. In this calculation method, the ACL can be rapidly estimated by focusing on time and frequency spectral domains for an averaged cross-spectral calculation. This method also has the advantage of being able to estimate the ACL in a short period, even during double-talk situations [21].

### 2.2. Wiener Filtering
The WF method estimates the ER gain \( G_i(\omega) \) by minimizing a squared error \( \epsilon \), which is given as follows:

\[
\epsilon = \| S_i(\omega) - G_i(\omega) Y_i(\omega) \|^2,
\]

where \( \cdot \) represents a norm of a vector and boldface denotes a time-sequence vector of a short-time spectral amplitude: \( P_i = [P_i, \ldots, P_{i-L+1}]^T \). The number of frames is \( L \), where \( L < N \). By solving the differential equation

\[
\frac{\partial \epsilon}{\partial G_i(\omega)} = 2[G_i(\omega)\|Y_i(\omega)\|^2 - \langle S_i(\omega), Y_i(\omega) \rangle] \rightarrow 0,
\]

the ER gain is then obtained as follows:

\[
G_i(\omega) = \frac{\langle S_i(\omega), Y_i(\omega) \rangle}{\|Y_i(\omega)\|^2},
\]

\[
= \frac{\langle Y_i(\omega) - D_i(\omega), Y_i(\omega) \rangle + \delta_1}{\|Y_i(\omega)\|^2},
\]

\[
= \frac{\|Y_i(\omega)\|^2 - \langle D_i(\omega), S_i(\omega) \rangle + \delta_1 + \delta_2}{\|Y_i(\omega)\|^2},
\]

\[
\|Y_i(\omega)\|^2 - \|D_i(\omega)\|^2 - \langle D_i(\omega), S_i(\omega) \rangle + \delta_1 + \delta_2,
\]

where

\[
\delta_w = \frac{\delta_1 + \delta_2 - \langle D_i(\omega), S_i(\omega) \rangle}{\|Y_i(\omega)\|^2},
\]

\[
\|Y_i(\omega)\|^2 - \|D_i(\omega)\|^2 - \langle D_i(\omega), S_i(\omega) \rangle + \delta_1 + \delta_2.
\]

The WF-based gain calculated from Eq. (7) is finally obtained by approximating Eq. (17) for a very short period.

### 2.3. Problem of Wiener Filtering
It is assumed that the inner product between the echo and near-end vectors is \( \langle D_i(\omega), S_i(\omega) \rangle = 0 \) with the WF method. This assumption holds only if a very long period of data is available because the statistical properties of data are used. However, the ER gain needs to be calculated from a very short period in practice because most echo and near-end speech signals are usually nonstationary. In that case, the number of samples used in calculating the inner product becomes insufficient and so the inner product \( \langle D_i(\omega), S_i(\omega) \rangle \) is not always zero. Therefore, \( G_i(\omega) \) is sometimes estimated to be smaller than the actual gain because the inner product between the echo and near-end vectors is ignored in Eq. (17). As a result, the ER based on the WF method suffers from speech distortions and quite often causes perceptual degradation of sound quality. In Sect. 3, we propose the new method to solve the problem with the conventional method and to reduce speech-quality degradation in double-talk situations.

### 3. PROPOSED METHOD

#### 3.1. Strategy for High-quality Echo Reduction
In this section, we explain the concept of the proposed gain estimation method used in the ER process. As described in Sect. 2.3, the inner product between the echo and near-end vectors are not always zero in practice and so its inner product cannot be ignored with respect to the ER-gain calculation. The proposed method derives a better ER gain \( G_i(\omega) \) based on the inner product from Eq. (13) as follows:

\[
G_i(\omega) = \frac{\|Y_i(\omega)\|^2 - \langle D_i(\omega), Y_i(\omega) \rangle + \delta_1}{\|Y_i(\omega)\|^2},
\]
where

\[ \gamma_i(\omega) = \frac{\langle D_i(\omega), Y_i(\omega) \rangle}{\|D_i(\omega)\|^2} \]

and

\[ \delta_p = \frac{\delta_1}{\|Y_i(\omega)\|^2}. \]

The parameter \( \gamma_i(\omega) \) denotes the ratio of the inner product between echo and input vectors divided by the square norm of the echo vector. The parameter \( \delta_p \) is the approximation error of the proposed method. In the conventional WF method, \( \gamma_i(\omega) \) and \( \delta_p \) are set to constant values \( \gamma_i(\omega) = 1 \) and the zero parameter \( \delta_p = 0 \), respectively, and the gain obtained from these approximations are equivalent to that of \( \delta_w = 0 \) in Eq. (17). However, the assumption \( \gamma_i(\omega) = 1 \) cannot hold true during double-talk situations in practice due to \( \langle D_i(\omega), Y_i(\omega) \rangle \neq \|D_i(\omega)\|^2 \) in a very short period.

### 3.2. Comparison of Approximation Accuracy

The approximation errors of the conventional and proposed methods \( \delta_w \) and \( \delta_p \) during the double-talk situation are shown in Fig. 2. The approximation error is defined as the difference between the target and estimated ER gains, and this is the value that occurs by neglecting the unknown terms included in the estimation of ER gains. The vertical and horizontal axes are the approximation errors and the number of frame \( L \), respectively. The simulation conditions are listed in Table 1. These approximation errors are calculated from average for all frames of two male and two female signals. As seen in Fig. 2, with both cases the approximation errors decreases in proportion as increasing the amount of statistics which is determined by \( L \). However, with the conventional method, the error \( \delta_w \) significantly increases when \( L \) is small which is an amount of error that should not be ignored. On the other hand, in the proposed method, the error is smaller even if \( L \) is small. This shows that the proposed method works effectively for calculating the gain in a short period accurately.

We also evaluated the time transitions of approximation errors \( \delta_w \) and \( \delta_p \) of when \( L = 2 \) and \( L = 4 \) during the double-talk situation, which are plotted in Figs. 3 and 4, respectively. The proposed method showed the smaller approximation error over the entire period than that of the conventional method.

### 3.3. Calculation Method of Echo Reduction Gain

Accurately estimating the parameter \( \gamma_i(\omega) \) is a key in solving the problem in which speech distortions are caused...
during double-talk situations. To calculate $\gamma_l(\omega)$ in practice, the proposed method substitutes the unknown vector $D_i(\omega)$ with $|\hat{H}_i(\omega)|^2|X_i(\omega)|^2$ in Eq. (8), i.e.,

$$\hat{\gamma}_l(\omega) = \frac{\langle \hat{D}_i(\omega), Y_l(\omega) \rangle}{\|\hat{D}_i(\omega)\|^2},$$

(24)

where

$$\hat{D}_i(\omega) = [\hat{D}_i(\omega), \cdots, \hat{D}_{i-L+1}(\omega)]^T.$$  

(25)

The ER gain is therefore represented using the estimated correlation $\hat{\gamma}_l(\omega)$ as

$$G_{Q_i}^Q(\omega) = \frac{\|Y_i(\omega)\|^2 - \hat{\gamma}_l(\omega)\|\hat{D}_i(\omega)\|^2}{\|Y_i(\omega)\|^2}.$$  

(26)

The parameter $\hat{\gamma}_l(\omega)$ varies corresponding to the rate of the near-end speech component included in the microphone input signal. Equation (24) therefore takes a value closer to one during single-talk situations whereas it takes a value larger than one during double-talk situations.

The ER gain is finally represented by using the estimated $\hat{\gamma}_l(\omega)$ and by replacing the norms $\|\hat{D}_i(\omega)\|^2$ and $\|Y_i(\omega)\|^2$ into instantaneous values $|\hat{D}_i(\omega)|^2$ and $|Y_i(\omega)|^2$ as follows:

$$G_{Q_i}^Q(\omega) = \frac{|Y_i(\omega)|^2 - \hat{\gamma}_l(\omega)|\hat{D}_i(\omega)|^2}{|Y_i(\omega)|^2}.$$  

(27)

4. EVALUATION

The performance of our new method was evaluated using both simulation and subjective listening tests. The proposed and conventional methods were used to calculate the ER gain using Eqs. (27) and (7), respectively. Table 2 lists the experimental conditions. Figure 5 shows the frequency characteristics of impulse response used in the computer simulation. The numbers of frames, $N$ and $L$, are set to 100 and 4, respectively.

4.1. Simulation Experiments

We conducted simulations to compare the proposed ER-gain estimation method to the conventional method (i.e. WF). The received signal $x(n)$ and near-end signal $s_l(n)$ are shown in Figs. 6 and 7, respectively. Periods A and B are received and send single-talk situations, respectively.
Double-talk situation occurs during period C. The received single talk is defined as the far-end speaker talking to near-end speaker. On the contrary, the send single talk means that the near-end speaker is talking to far-end speaker. The double talk occurs when both the near-end and far-end speakers are talking concurrently.

The microphone input signals $y(n)$ are plotted in Fig. 8. Figures 9 and 10 plot the send signals after processing by conventional and proposed methods, respectively. The power envelopes of send signals of conventional and proposed methods in period C are plotted in Fig. 11. As seen in Figs. 8, 9, and 10, the proposed and conventional methods sufficiently suppressed echo signals over the entire period. Table 3 shows echo-suppression levels of the conventional and proposed methods during the single-talk situation of the period A. The echo-suppression levels were 33.79 dB with the conventional method and 33.72 dB with the proposed method.

However, as seen in Fig. 11, the conventional method seems to result in speech distortions during the double-talk situation compared with the proposed method. The amount of speech distortions during the double-talk situation were evaluated using a linear predictive coding (LPC) cepstral distance [22], which is computed by

\[
CD(n) = \frac{10}{\log 10} \sqrt{2 \sum_{k=1}^{16} [c(k,n) - \hat{c}(k,n)]^2},
\]

where $c(k,n)$ and $\hat{c}(k,n)$ are the $k$-th cepstral coefficients of near-end speech and send signals, respectively, and $CD(n)$ is the LPC cepstral distance. The results from comparing the conventional and proposed methods using the LPC cepstral distance are shown in Fig. 12. As these results indicate, the better scores in the LPC cepstral distance were observed than the conventional method over the entire period and significant improvement was confirmed.
4.2. Subjective Experiments

A multi-stimulus test with hidden reference and anchor (MUSHRA) using a 100-point scale, compliant with ITU-R BS.1534-1 [18], was used to test the quality of speech. All reference and evaluation signals are played to both ears with headphones (Sennheiser HD 280 Pro). Eight experienced listeners evaluated speech under six conditions: near-end speech signal (c00: hidden reference), near-end speech signal filtered by 3.5 kHz low-pass filter (c01: anchor A), microphone input signal (c02: anchor B), target signal of ER (c03: anchor C), and send signals of conventional and proposed methods (c04: conventional method and c05: proposed method). The target signal $\tilde{s}_T(n)$ expresses the limiting value of ER, which is simulated using the following equation:

$$\tilde{s}_T(n) = \text{IFFT}[|S_i(\omega)|e^{i\theta_Y}],$$

(29)

where IFFT[·] and $\theta_Y$ denotes the IFFT operation and the phase component of $Y_i(\omega)$, respectively.

The MUSHRA test results comparing the conventional and proposed methods during the double-talk situation of the period C are shown in Fig. 13. The vertical lines in the figure denote a 95% confidence interval. For the double-talk period (periods C), mean scores were awarded for four sound signals by eight listeners. As these results indicate, a better score was observed in the double-talk period by using the proposed method that calculates the ER gain considering the cross-spectral term between echo and near-end speech signals. The proposed method improved the sound quality by about six points on a 100-point scale compared with the conventional method, and a significant improvement was confirmed.

5. CONCLUSION

This paper proposed a new modified gain-estimation method for the echo-reduction process. To reduce the speech distortion produced by echo reduction, the proposed gain was calculated based on the assumption that the echo and near-end signals is uncorrelated but the cross-spectral term of their signals obtained in the short-time period is not zero. The experimental results showed that the proposed gain-estimation method performed better than the conventional method by using the echo-reduction process, and significant improvement was confirmed.

---

Table 3 Comparison of echo-suppression levels during single-talk situation of period A.

| Method  | Echo-suppression level [dB] |
|---------|-----------------------------|
| Conventional | 33.79                      |
| Proposed  | 33.72                      |

Fig. 11 Each power envelope in each signal during double-talk situation of period C.

Fig. 12 Comparison of LPC cepstrum distances during double-talk situation in period C.

Fig. 13 Double-talk quality assessments for period C.
REFERENCES

[1] J. Nagumo and A. Noda, “A learning method for system identification,” IEEE Trans. Autom. Control, 12, 282–297 (1967).
[2] S. Haykin, Adaptive Filter Theory, 3rd ed. (Prentice-Hall, Englewood Cliffs, N.J., 1996).
[3] C. Avendano, “Acoustic echo suppression in the STFT domain,” IEEE Workshop Appl. Signal Process. Audio Acoust., pp. 175–178 (2001).
[4] C. Faller and J. Chen, “Suppressing acoustic echo in a spectral envelope space,” IEEE Trans. Speech Audio Process., 13, 1048–1062 (2005).
[5] C. Faller and C. Tourneri, “Estimating the delay and coloration effect of the acoustic echo path for low complexity echo suppression,” Proc. IWAENC 2005, pp. 53–56 (2005).
[6] C. Faller and C. Tourneri, “Robust acoustic echo control using a simple echo path model,” Proc. ICASSP 2006, Vol. 5, pp. 281–284 (2006).
[7] S. Sakauchi, A. Nakagawa, Y. Haneda and A. Kataoka, “Implementing and evaluating of an audio teleconferencing terminal with noise and echo reduction,” Proc. IWAENC 2003, pp. 191–194 (2003).
[8] Y. S. Park and J. H. Chang, “Frequency domain acoustic echo suppression based on soft decision,” IEEE Signal Process. Lett., 16, 53–56 (2009).
[9] M. Fukui, A. Nakagawa, S. Suehiro, Y. Haneda and A. Kataoka, “Accurate echo power estimation for echo reduction,” Proc. 2009 IEICE Gen. Conf., A-4-18 (2009).
[10] A. Favrot, C. Faller and F. Kuech, “Modeling late reverberation in acoustic echo suppression,” Proc. IWAENC 2012, pp. 1–4 (2012).
[11] C. Beaugeant, V. Turbin, P. Scalart and A. Gilloire, “New optimal filtering approaches for hands-free telecommunication terminals,” Signal Process., 64, 33–47 (1998).
[12] S. Gustafsson, R. Martin and P. Vary, “Combined acoustic echo control and noise reduction for hands-free telephony,” Signal Process., 64, 21–32 (1998).
[13] V. Turbin, A. Gilloire and P. Scalart, “Comparison of three post-filtering algorithms for residual acoustic echo reduction,” Proc. ICASSP 97, Vol. 1, pp. 307–310 (1997).
[14] E. Hänsler and G. U. Schmidt, “Hands-free telephones: Joint control of echo cancellation and postfiltering,” Signal Process., 80, 2295–2305 (2000).
[15] J. S. Lim and A. V. Oppenheim, “Enhancement and bandwidth compression of noisy speech,” Proc. IEEE, 67, 1586–1604 (1979).
[16] R. Le Bouquin Jeannes, P. Scalart, G. Fauco’n and C. Beaugeant, “Combined noise and echo reduction in hands-free systems: A survey,” IEEE Trans. Speech Audio Process., 9, 808–820 (2001).
[17] M. Fukui, A. Nakagawa, S. Shimauchi, A. Nakagawa, Y. Haneda and A. Kataoka, “Echo reduction using Wiener gains considering short-time correlation between echo and near-end speech,” Proc. IWAENC 2012, pp. 1–4 (2012).
[18] ITU-R Recommendation BS.1534-1, “Method for the subjective assessment of intermediate quality level of coding systems” (2003).
[19] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. Acoust. Speech Signal Process., 27, 113–120 (1979).
[20] Y. Ephraim and D. Malah, “Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator,” IEEE Trans. Acoust. Speech Signal Process., 32, 1109–1121 (1984).

[21] M. Fukui, S. Shimauchi, A. Nakagawa, Y. Haneda and A. Kataoka, “Acoustic-coupling level estimation for performance improvement of echo reduction,” Proc. IWAENC 2008, pp. 1–4 (2008).
[22] A. H. Gray Jr. and J. D. Markel, “Distance measures for speech processing,” IEEE Trans. Acoust. Speech Signal Process., 24, 380–391 (1976).

Masahiro Fukui received the B.E. degrees in information science from Ritsumeikan University, Shiga, Japan, in 2002. He received the M.E. degree in information science from Nara Institute of Science and Technology, Nara, Japan, in 2004. Since joining Nippon Telegraph and Telephone Corporation (NTT) in 2004, he has been engaged in research on acoustic echo cancellers and speech coding. He is now a Research Engineer at NTT Media Intelligence Laboratories. He is a member of IEEE, IEICE, and ASJ.

Suehiro Shimauchi received the B.E., M.E., and Ph.D. degrees from Tokyo Institute of Technology in 1991, 1993, and 2007, respectively. Since joining Nippon Telegraph and Telephone Corporation (NTT) in 1993, he has been engaged in research on acoustic signal processing for acoustic echo cancellers. He is now a Senior Research Engineer at NTT Media Intelligence Laboratories. He is a member of IEEE, IEICE, and ASJ.

Yusuke Hioka received his B.E., M.E., and Ph.D. degrees in engineering in 2000, 2002, and 2005 from Keio University, Yokohama, Japan. From 2005 to 2012, he was with the NTT Cyber Space Laboratories (now NTT Media Intelligence Laboratories), Nippon Telegraph and Telephone Corporation (NTT). From 2010 to 2011, he was also a Visiting Research Fellow at Massey University and a visiting researcher at Victoria University of Wellington, both located in Wellington, New Zealand. In 2013 he joined the Department of Electrical and Computer Engineering at the University of Canterbury, Christchurch, New Zealand where he is currently a Lecturer. His research interests include microphone array signal processing and room acoustics. He is a senior member of the IEEE and a member of the ASJ and the IEICE.

Akira Nakagawa received the B.E. and M.E. degrees from Kyushu Institute of Technology in 1992 and 1994, respectively. Since joining Nippon Telegraph and Telephone Corporation (NTT) in 1994, he has been investigating acoustic signal processing and acoustic echo cancellers. He is now a Senior Research Engineer at NTT Media Intelligence Laboratories. He received a paper award from the Acoustical Society of Japan in 2001. He is a member of ASJ.
Yoichi Haneda received the B.S., M.S., and Ph.D. degrees from Tohoku University, Sendai, in 1987, 1989, and 1999. From 1989 to 2012, he was with Nippon Telegraph and Telephone Corporation (NTT), Japan. In 2012, he joined the University of Electro-Communications, where he is a Professor. His research interests include modeling of acoustic transfer functions, microphone arrays, loudspeaker arrays, and acoustic echo cancellers. He received paper awards from the ASJ and from the IEICE of Japan in 2002. Dr. Haneda is a senior member of the IEEE and is also a member of ASJ.

Akitoshi Kataoka received the B.E., M.E., and Ph.D. degrees in Electrical Engineering from Doshisha University of Kyoto in 1984, 1986, and 1999 respectively. Since joining NTT Laboratories in 1986, he has been engaged in research on noise and reverberation reduction, acoustic microphone arrays, and medium bit-rate speech and wideband coding algorithms for the ITU-T standard. He contributed to establishing ITU-T G.729 standards. He is currently Professor of the Faculty of Science and Technology at Ryukoku University. He received the Technology Development Award from ASJ, the Telecommunication Systems Technology Prize awarded by The Telecommunications Advancement Foundation in 1996, and the Prize of the Commissioner of the Japan Patent Office from the Japan Institute of Invention and Innovation in 2003. Dr. Kataoka is a member of the ASJ and the IEICE.

Hitoshi Ohmuro Senior Research Engineer, Supervisor, Speech, Acoustics and Language Laboratory, NTT Media Intelligence Laboratories. He received the B.E. and M.E. degrees in electrical engineering from Nagoya University, Aichi, in 1988 and 1990, respectively. He joined NTT in 1990. He has been engaged in research on highly efficient speech coding and the development of VoIP applications. He is a member of IEEE, IEICE and ASJ.