Speech Enhancement Based on Linear Prediction and Correlation-Inputting Bias Free Equation Error ADF

Naoto Sasaoka[1], **Shinichi Wada**[2], **James Okello**[3], **Yoshio Itoh**[4], and **Masaki Kobayashi**[5], Non-members

ABSTRACT

In this paper, a speech enhancement technique to reduce background noise in noisy speech is proposed. We investigated the noise reconstruction system (NRS) based on linear prediction and system identification as a speech enhancement. Assuming that the background noise is generated from white noise by exciting a linear filter, the system identification estimates the background noise from estimated white noise. However, the white noise estimated by a linear prediction error filter (LPEF) includes residual speech, then the estimation accuracy of background noise is degraded at the system identification and the quality of enhanced speech is deteriorated. In order to reduce the influence of the residual speech, a lattice filter and a bias free equation error adaptive digital filter (ADF) are respectively introduced to the LPEF and system identification. The residual speech is reduced by the lattice filter which approximates a vocal-tract filter well. On the other hand, the bias free equation error ADF uses the cross-correlation between the whitened noise and a desired signal as a tap input. Since the speech does not have the correlation from the desired signal, the tap coefficients converge without the influence of speech.

Keywords: Speech Enhancement, Adaptive Filter

1. INTRODUCTION

Speech enhancement has been investigated to reduce the background noise in noisy speech. Speech enhancement is applied to a mobile phone, a hearing aid and a speech recognition system for improving the intelligibility and the recognition performance of speech. Spectral subtraction (SS) and MMSE (Minimum Mean Square Error) - STSA (Short Time Spectral Amplitude) [1], [2] are known as a speech enhancement based on the estimation of short time spectral amplitude with only one microphone [3]. However, the musical noise rises due to residual error at a SS system. MMSE-STSA is based on minimum mean square error between clean speech and estimated speech and can avoid the musical noise. Unfortunately, the calculation of a special function, for example a Bessel function, is required therefore it is difficult to achieve DSP implementation for a hearing aid. Furthermore, the SS and MMSE-STSA require prior estimation of a noise spectrum. This implies that voice activity detector (VAD) is required in noisy environments. Although the noise spectral estimation without a VAD is proposed [4], it is assumed that the input signals in first some frames do not include the speech components. Therefore, it is difficult to reduce the background noise which is suddenly generated in a speech section.

On the other hand, the speech enhancement with one microphone based on an adaptive filter is proposed [5], however, the accuracy estimation of a pitch period of speech is needed and an un-voiced sound cannot be enhanced. We have proposed the speech enhancement based on noise reconstruction system [6], [7] to solve the problems. The noise reconstruction system uses linear prediction error filter (LPEF) and system identification [6]. NRS assumes that the background noise is generated from white noise by a noise generating system, which is a linear system. At the system identification, the background noise is reconstructed from the noise whitened by LPEF through estimating the noise generating system. NRS does not require the prior estimation of a noise spectrum, a VAD and a pitch period detection. The enhanced speech does not include musical noise. However, the input signal of the adaptive filter for estimating background noise includes residual speech. Therefore, the estimation accuracy of background noise is deteriorated in a speech section, and the quality of enhanced speech is decreased.

In order to solve the problem, a bias free equation error adaptive digital filter (ADF) [8] and a lattice filter are respectively adopted to the system identification and the LPEF. At the bias free equation error ADF, the residual speech included in the input sig-

Manuscript received on July 31, 2011 ; revised on November 1, 2011.

[1,2,4] The authors are with with Graduate School of Engineering, Tottori University, 4-101 Koyamacho-minami, Tottori 680-8552, Japan , E-mail: sasaoka@ele.tottori-u.ac.jp, wada@dacom2.ele.tottori-u.ac.jp and itoh@ele.tottori-u.ac.jp

[3] The author is with the System IP Core Research Laboratories, NEC, Kawasaki-shi, Kanagawa 211-8666, Japan , E-mail: j-okello@bq.jp.nec.com

[5] The author is with Faculty of Engineering, Chubu University, Kasugai-shi, Aichi 487-8501, Japan , E-mail: mkoba@isc.chubu.ac.jp

nal is reduced by using the cross-correlation between the input signal and a desired signal as a tap input because the residual speech does not have the correlation with the desired background noise. Thus, the quality of enhanced speech is improved. Since the estimation accuracy of background noise is slightly degraded due to the bias free ADF, the sub adaptive filter is also used to improve the estimation accuracy in this paper. In addition, a lattice filter is used as the LPEF to decrease the residual speech in whitened noise because the vocal-tract filter for the speech production process is approximated by the lattice filter [9]. Therefore the lattice filter can improve the quality of enhanced speech.

This paper is organized as follows. In Section 2, the conventional NRS is explained. In Section 3, we propose NRS with a lattice filter and bias free equation error ADF. The experimental results of the proposed method are shown in Section 4. In Section 5, we conclude our paper.

2. NOISE RECONSTRUCTION SYSTEM

The speech enhancement using the noise reconstruction system is shown in Fig. 1. Noisy speech $x(n)$ is represented as a following equation.

$$x(n) = s(n) + \xi(n) \tag{1}$$

where $s(n)$ and $\xi(n)$ are respectively clean speech and background noise at time n. $w(n)$, $\hat{\xi}$(n) and $\hat{s}(n)$ are an output of a LPEF, reconstructed noise and enhanced speech respectively. The LPEF and the noise reconstruction filter (NRF) are transversal type filters. The transfer function of the LPEF and the NRF are respectively represented as $H_{LPEF}(z)$ and $H_{NRF}(z)$. These transfer functions are defined by

$$H_{LPEF}(z) = 1 - \sum_{k-1}^{M} h_k(n) z^{-k} \tag{2}$$

$$H_{NRF}(z) = \sum_{k=0}^{L} h'_k(n) z^{-k} \tag{3}$$

where $h_k(n)$ and $h'_k(n)$ are respectively the k-th tap coefficients of the LPEF and the NRF.

The whitened noise $w(n)$ is obtained by the LPEF. Tap coefficients of a LPEF converge such as a prediction error signal whitens [10]. Since a speech signal can be represented as a stationary and periodic signal in a short time interval, a speech signal $s(n)$ is estimated by a linear predictor. On the other hand, assuming that a background noise is generated by exciting a linear system $H_N(z)$ from white noise, the background noise whitens by a LPEF.

We then consider the background noise is reconstructed from a whitened signal by the NRF. Assuming that white noise generates background noise by exciting a linear system $H_N(z)$, the background noise can be reconstructed from a whitened noise $w(n)$ by estimating the transfer function of the linear system $H_N(z)$ [6]. This estimation is performed by a system identification model, where $\xi(n)$, $s(n)$ and $\hat{s}(n)$ are a desired signal, disturbance and an estimation error signal respectively. Finally, an enhanced speech signal is obtained by subtracting the reconstructed noise $\hat{\xi}(n)$ from $x(n)$.

However, the estimation accuracy of background noise is deteriorated due to disturbance and the residual speech component included in a tap input of the NRF.

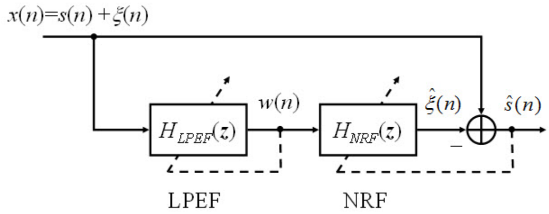

Fig.1: *Structure of Noise Reconstruction System.*

3. PROPOSED SPEECH ENHANCEMENT

In order to solve the above-mentioned problem, we need to reduce the influence of the residual speech in whitened noise. Thus, two methods are introduced to the conventional NRS. One is to reduce the residual speech by improving the estimation accuracy of speech at the LPEF. The other is to adopt a bias free ADF, which is robust to disturbance, as the NRF. At next sections, we explain these methods in detail.

3.1 Introducing Lattice Filter to LPEF

In order to reduce the residual speech included in an input signal of the NEF, a lattice filter is introduced to the LPEF. Since the LPEF estimates the vocal-tract filter of speech at NRS, the introduction of the lattice filter which approximates the vocal-tract filter of speech well improves the estimation accuracy of speech at the LPEF. Consequently, the estimation accuracy of background noise at NRF is also improved.

The structure of a lattice filter is shown in Fig. 2, where $f_m(n)$ and $b_m(n)$ are respectively the forward prediction error and the backward prediction error. The forward and backward prediction error $f_m(n)$ and $b_m(n)$ are defined by

$$f_m = f_{m-1}(n) + \alpha_m(n) b_{m-1}(n-1) \tag{4}$$

$$b_m = b_{m-1}(n-1) + \beta_m(n) f_{m-1}(n) \tag{5}$$

where $\alpha_m(n)$ and $\beta_m(n)$ represent forward and backward reflection coefficients respectively. In this paper, $f_M(n)$ which is M-th forward prediction error is used as the output signal of the LPEF.

As the algorithm for updating reflection coefficient, least square lattice (LSL) algorithm [10] is used. The algorithm is given as follows.

$$\begin{aligned}
\Delta_m(n) &= \lambda\Delta_m(n-1) + f_m(n-1)b_m(n-1)/\theta_m(n-1)\\
F_m(n) &= F_{m-1}(n) - \{\Delta_{m+1}(n)\}^2/B_m(n-1)\\
B_m(n) &= B_{m-1}(n-1) - \{\Delta_{m+1}(n)\}^2/B_m(n)\\
\alpha_m(n) &= -\Delta_m(n)/B_{m-1}(-1)\\
\beta_m(n) &= -\Delta_m(n)/F_{m-1}(n) \qquad (6)
\end{aligned}$$

where Δ_m indicates the covariance between $f_m(n)$ and $b_m(n)$. F_m and B_m are the variance of $f_m(n)$ and $b_m(n)$ respectively. λ is a forgetting factor. $\theta_m(n-1)$ is represented as

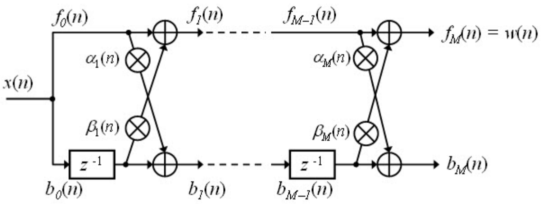

Fig.2: *Lattice type LPEF*

$$\theta_m(n-1) = \theta_{m-1}(n-1) - \frac{b^2_{m-1}(n-1)}{B_{m-1}(n-1)}. \qquad (7)$$

About divisions in Eq.(6), it is possible to decrease the computational load by a look-up table.

3.2 Noise Estimation Based on Equation Error ADF

In order to make NRF robust to disturbance, the bias free equation error ADF which uses a correlation as a tap input [8] is introduced to the NRF. The equation error ADF takes advantage of independence between a desired signal and disturbance. Since the disturbance does not have the correlation from the desired background noise, the disturbance in a tap input of the NRF is reduced by using the correlation as a tap input signal. Thus the tap coefficients of the NRF converge without the influence of speech.

The transfer function of the NRF based on an equation error ADF is expressed as

$$H_{NRF}(z) = \hat{B}(z)/\hat{A}(z). \qquad (8)$$

where $\hat{A}(z)$ and $\hat{B}(z)$ are defined as follows.

$$\hat{A}(z) = \hat{a}_0(n) + \hat{a}_1(n)z^{-1} + \cdots \hat{a}_N(n)z^{-N} \qquad (9)$$

$$\hat{B}(z) = \hat{b}_0(n) + \hat{b}_1(n)z^{-1} + \cdots \hat{b}_N(n)z^{-N} \qquad (10)$$

and without loss of generality, $a_0 = 1$. Thus, the tap coefficient vectors are given by

$$\mathbf{a}(n) = [1, \hat{a}_1(n), \ldots, \hat{a}_N(n)]^T \qquad (11)$$

$$\mathbf{b}(n) = \left[\hat{b}_0(n), \hat{b}_1(n), \ldots, \hat{b}_N(n)\right]^T. \qquad (12)$$

Fig. 3 shows the structure of bias free equation error algorithm which uses the correlation as an input signal for updating the tap coefficient of the NRF. An error signal $e'(n)$ is used to derive an adaptive algorithm. $w(n-\tau_{n'})$ represents the input signal delayed by $\tau_{n'}$ samples. The iteration interval n' is far much less than the sampling rate of the input signal due to the relatively large value of the block length L_f+. n' is related to n by the following equation.

$$n' = div(n, L_f + 1) \qquad (13)$$

where $div(v, \omega)$ refers to the biggest integer less or equal to v, ω.

Let $R_{wx}(n, l-\tau_{n'})$ be the cross correlation between $w(n-\tau_{n'})$ and $x(n-l)$. $R_{ww}(n, l-\tau_{n'})$ is defined as the auto correlation between $w(n-\tau_{n'})$ and $w(n-l)$. $R_{wx}(n, l-\tau_{n'})$ and $R_{ww}(n, l-\tau_{n'})$ are given by

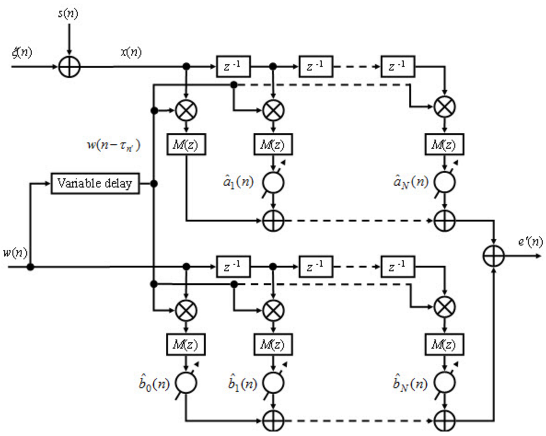

Fig.3: *Structure of bias free adaptive algorithm.*

$$R_{wx}(n, l-\tau_{n'}) = E\left[w(n-\tau_{n'})x(n-l)\right] \qquad (14)$$

$$R_{ww}(n, l-\tau_{n'}) = E\left[w(n-\tau_{n'})w(n-l)\right]. \qquad (15)$$

In these equations, an input signal of the NRF $w(n)$ is independent from disturbance $s(n)$. Therefore the Eq. (14) can be rewritten as follows.

$$\begin{aligned}
R_{wx}(n, l-\tau_{n'}) &= E\left[w(n-\tau_{n'})\{s(n-l) + \xi(n-l)\}\right]\\
&= E\left[w(n-\tau_{n'})\xi(n-l)\right]. \qquad (16)
\end{aligned}$$

Eq. (16) indicates that the speech which is disturbance is reduced. $\mathbf{x}'(n)$ and $\mathbf{w}'(n)$ are defined by

$$\mathbf{x}'(n) \approx \begin{bmatrix} R_{wx}(n,-\tau_{n'}) \\ R_{wx}(n,1-\tau_{n'}) \\ \vdots \\ R_{wx}(n,N-\tau_{n'}) \end{bmatrix} \approx \left[R_{wx}(n,-\tau_{n'})\mathbf{R}'_{wx}(n,\tau_{n'})^T\right]^T \quad (17)$$

$$\mathbf{w}'(n) \approx \begin{bmatrix} R_{ww}(n,-\tau_{n'}) \\ R_{ww}(n,1-\tau_{n'}) \\ \vdots \\ R_{ww}(n,N-\tau_{n'}) \end{bmatrix} \quad (18)$$

where $\mathbf{R}'_{wx}(n,\tau_{n'})$ is defined as

$$\mathbf{R}'_{wx}(n,\tau_{n'}) = \begin{bmatrix} R_{wx}(n,1-\tau_{n'}) \\ \vdots \\ R_{wx}(n,N-\tau_{n'}) \end{bmatrix}. \quad (19)$$

In the proposed system, the cross-correlation and auto correlation are estimated by

$$\begin{aligned} \mathbf{x}'_n &= \alpha\mathbf{x}'(n-1) + \{w(n-\tau_{n'})\mathbf{x}(n)\} \\ &\quad - \alpha^L\{w(n-\tau_{n'}-L-1)\mathbf{x}(n-L-1)\} \quad (20) \\ \mathbf{w}'_n &= \alpha\mathbf{w}'(n-1) + \{w(n-\tau_{n'})\mathbf{w}(n)\} \\ &\quad - \alpha^L\{w(n-\tau_{n'}-L-1)\mathbf{w}(n-L-1)\} \quad (21) \end{aligned}$$

where

$$\begin{aligned} \mathbf{x}(n) &= [x(n), x(n-1), \ldots, x(n-N)]^T \\ \mathbf{w}(n) &= [w(n), w(n-1), \ldots, w(n-N)]^T \quad (22) \end{aligned}$$

and α is a forgetting factor and α^L is nearly equal to one but less than one. In Fig. 3, M(z) is a linear filter that calculates Eqs. (20) and (21).

The error signal $e'(n)$ to update the tap coefficient is given by

$$\begin{aligned} e'(n) &= \mathbf{x}'(n)^T\hat{\mathbf{a}}(n) - \mathbf{w}'(n)^T\hat{\mathbf{b}}(n) \\ &\approx R_{wx}(-\tau_{n'}) + \mathbf{R}'_{wx}(n,\tau_{n'})^T\hat{\mathbf{a}}(n) - \mathbf{w}'(n)^T\hat{\mathbf{b}}(n) \end{aligned} \quad (23)$$

where $\hat{a}_0(n) = 1$. $\hat{\mathbf{a}}'(n)$ is part of vector $\hat{\mathbf{a}}'(n)$ such that $\hat{\mathbf{a}}'(n) = [1\hat{\mathbf{a}}'^T(n)]^T$. Then, the error $e'(n)$ is represented as

$$e'(n) \approx R_{wx}(-\tau_{n'}) - \mathbf{u}(n)^T\hat{\mathbf{h}}(n) \quad (24)$$

where

$$\begin{aligned} \hat{\mathbf{h}}(n) &= \left[\hat{\mathbf{a}}'(n)^T \quad \hat{\mathbf{b}}(n)^T\right]^T \\ \mathbf{u}(n) &= \left[-\mathbf{R}'_{wx}(n,\tau_{n'})^T \quad \mathbf{w}'(n)^T\right]^T \quad (25) \end{aligned}$$

The proposed algorithm which minimizes the mean square of $e'(n)$ is based on NLMS (Normalized Least Mean Square) algorithm [10] shown as follows:

$$\hat{\mathbf{h}}(n'+1) = \hat{\mathbf{h}}(n') + \mu'\frac{\mathbf{u}(n')e'(n')}{\mathbf{u}^T(n')\mathbf{u}(n')} \quad (26)$$

where μ' is the step size of adaptation. The step size μ' must be such that

$$0 < \mu' < 2 \quad (27)$$

Assuming that the input signal is stationary, the correlations given by Eqs. (20) and (21) are constant. Therefore, the tap input signal of the NRF is constant. Under such a condition, $\hat{\mathbf{h}}(n)$ will never be equal to an ideal tap coefficient vector. In order to avoid the situation where the correlation vector $\mathbf{u}(n)$ is entirely constant over time, we make the value of the delay $\tau_{n'}$ to change after certain time interval. The variable delay is defined by

$$\tau_{n'} = mod(div(n, L_f+1), \tau_{max}). \quad (28)$$

where $mod(v,w)$ is a function which returns the remainder when dividing v by w, and τ_{max} is the maximum delay.

3.3 Sub-Noise Reconstruction Filter

Although the quality of enhanced speech is improved by the lattice filter and the bias free equation error ADF, the noise reduction performance is degraded because the auto correlation of background noise is faded out due to the delay $\tau_{n'}$ and NRF cannot estimate the noise components whose auto correlation is weak. For improving the noise estimation accuracy, we introduce the sub-noise reconstruction filter after the bias free equation error ADF.

Fig. 4 shows the proposed noise reconstruction system with a sub-NRF. The LPEF is the lattice filter explained in Section 3.1. The NRF is the equation error ADF explained in Section 3.2. The sub-NRF is a transversal type adaptive filter whose transfer function is represented as

$$H_{SNRF}(z) = \sum_{k=0}^{M_s} h_{S,k}(n)z^{-k} \quad (29)$$

where $h_{S,k}(n)$ is the k-th tap coefficient of the sub-NRF. $\hat{\xi}'(n)$ represents the background noise estimated by the bias free equation error ADF. The output signal of sub-NRF is $\hat{\xi}(n)$. The background noise is estimated by the Sub-NRF where $\xi(n)$, $s(n)$ and $\hat{s}(n)$ are a desired signal, disturbance and estimation error respectively. Though the transversal type ADF is not robust to speech, the speech component in $\hat{\xi}'(n)$ is reduced enough and then the sub-NRF estimates the background noise accurately while maintaining the high quality of enhanced speech.

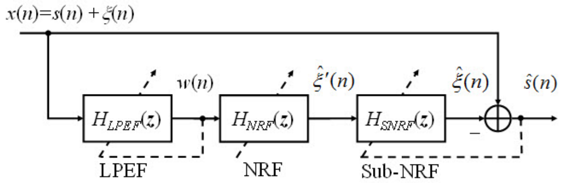

Fig.4: *Structure of Proposed NRS*

4. SIMULATION RESULTS

4.1 Simulation Conditions

The performance of the proposed speech enhancement was evaluated. All sound data prepared in simulations were sampled by 8 kHz in 16 bit resolution. The input signals were generated by artificially adding background noise to clean speech. As a speech signal, 2 male speech data and 2 female speech data recorded in Acoustic Society of Japan-Japanese Newspaper Article Sentences (ASJ-JNAS) were used. The tunnel noise and factory noise were used as the actual noise. The tunnel noise is recorded inside the tunnel of an expressway. The tunnel noise includes wideband noise, noise arising from the ventilation fans and the noise arising from automobiles' engines. The factory noise is in the Noisex-92 database. Table 1 shows each parameter in this simulation.

The signal to noise ratio (SNR) and quality of speech (QS) were used to evaluate the noise reduction ability and quality of enhanced speech, respectively. These indices are defined by

$$SNR_{in} = 10\log_{10}\sum_{j=1}^{N'} s^2(j) \Bigg/ \sum_{j=1}^{N'} \xi^2(j) \qquad (30)$$

$$SNR_{out} = 10\log_{10}\sum_{j=1}^{N'} \hat{s}_s^2(j) \Bigg/ \sum_{j=1}^{N'} \hat{s}_\xi^2(j) \qquad (31)$$

$$QS = 10\log_{10}\sum_{j=1}^{N'} s^2(j) \Bigg/ \sum_{j=1}^{N'} \{s(j) - \hat{s}_s(j)\}^2 \qquad (32)$$

where SNR_{in} and SNR_{out} represent the input and output SNR, respectively. N' is the number of samples. $\hat{s}_s(j)$ and $\hat{s}_\xi(j)$ are the speech and the noise included in $\hat{s}(j)$, respectively. $s(j) - \hat{s}_s(j)$ represents the distortion of speech components due to a filter, thus QS increases as the quality of enhanced speech is increased.

4.2 Decreasing residual speech by lattice filter

In order to evaluate the performance of lattice type LPEF, the residual speech in whitened noise $w(n)$ at lattice type LPEF is compared with that at the

Table 1: *Parameters of Filters*

Conventional NRS

LPEF (LMS)	Number of tap coefficients M	128
	Step Size μ	0.1
NRF (NLMS)	Number of tap coefficients L	128
	Step Size μ'	0.02

Proposed NRS

LPEF (LSL)	Number of tap coefficients M	128
	Forgetting Factor β	0.999
	Step Size μ	0.00095
NRF	Order of Filter N	7
	Block length L_f	800
	Forgetting factor α	0.999
	Maximum delay τ_{max}	10
	Step size μ'	0.05
Sub-NRF (NLMS)	Number of tap coefficients M_s	128
	Step Size μ'	0.002

LPEF using a finite impulse response (FIR) filter. The tunnel noise, which has been explained in Sect. 4.1, was used as the background noise. As speech, the female speech was used. SNR_{in} was set to 0 dB. $E_{RS}(n)$ and $\overline{E_{RS}(n)}$ were used to evaluate the residual speech in whitened noise w(n). $E_{RS}(n)$ and $\overline{E_{RS}(n)}$ are defined by

$$E_{RS}(n) = 10\log_{10}\sum_{j=0}^{127} w_s^2(n-j) \Bigg/ \sum_{j=0}^{127} s^2(n-j) \quad (33)$$

$$\overline{E_{RS}(n)} = 10\log_{10}\sum_{n=1}^{N'} w_s^2(n-j) \Bigg/ \sum_{n=1}^{N'} s^2(n) \quad (34)$$

where $w_s(n)$ is speech component in whiten noise $w(n)$. Fig. 5 shows the performance of LPEF. Figs. 5(a) and 5(b) illustrate clean speech and $E_{RS}(n)$. From the simulation results, E_{RS} is improved in a speech section and $\overline{E_{RS}(n)}$ is improved by about 1.6 dB. The figure indicates that the lattice type LPEF has a potential of decreasing the residual speech in $w(n)$.

4.3 Decreasing residual speech by bias free equation error ADF

In order to evaluate the performance of bias free equation error ADF, the residual speech in the noise estimated by the bias free equation error ADF is compared with that in the noise estimated by the conventional NRS. The tunnel noise was used as the background noise. As speech, the female speech was used. SNR_{in} was set to 0 dB. $E'_{RS}(n)$ and $\overline{E'_{RS}(n)}$

were used to evaluate the residual speech in estimated background noise, which are represented as

$$E'_{RS}(n) = 10\log_{10}\sum_{j=0}^{127}\hat{\xi}_s^2(n-j)\bigg/\sum_{j=0}^{127}s^2(n-j) \quad (35)$$

$$\overline{E'_{RS}(n)} = 10\log_{10}\sum_{n=1}^{N'}\hat{\xi}_s^2(n)\bigg/\sum_{n=1}^{N'}s^2(n) \quad (36)$$

where $\xi_s(n)$ is speech component in estimated noise. Fig. 6 shows the performance of bias free equation error ADF. Figs. 6(a) and 6(b) represent clean speech and $E'_{RS}(n)$. Comparing the bias free equation error ADF with NRF which is composed by a FIR filter, $E'_{RS}(n)$ are improved in a speech section. $\overline{E'_{RS}(n)}$ is improved by about 2.6 dB. Therefore we verify that the proposed equation error ADF can decrease the residual speech in the estimated background noise.

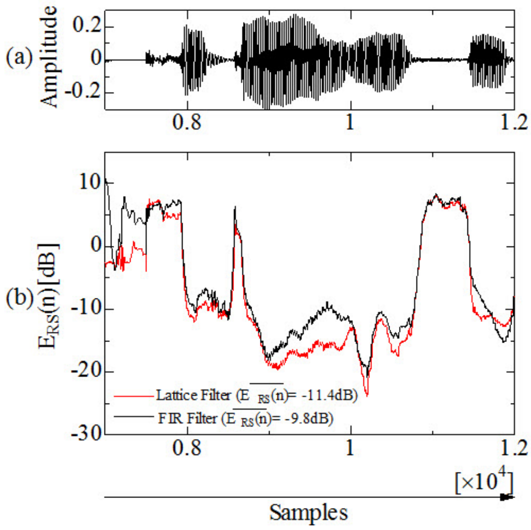

Fig.5: *Performance of LPEF (a)Clean speech (b)$E_{RS}(n)$*

4.4 Improving estimation accuracy of background noise by sub-NRF

In this section, the effectiveness of the bias free equation error ADF and sub-NRF is evaluated. The tunnel noise was used as the background noise. As speech, the female speech was used. SNR_{in} was set to 0 dB. $E_N(n)$ and $\overline{E_N(n)}$ were used to evaluate the estimation accuracy of background noise, which are defined by

$$E_N(n) = 10\log_{10}\frac{\sum_{j=0}^{127}\xi^2(n-j)}{\sum_{j=0}^{127}\{\xi(n-j)-\hat{\xi}(n-j)\}^2} \quad (37)$$

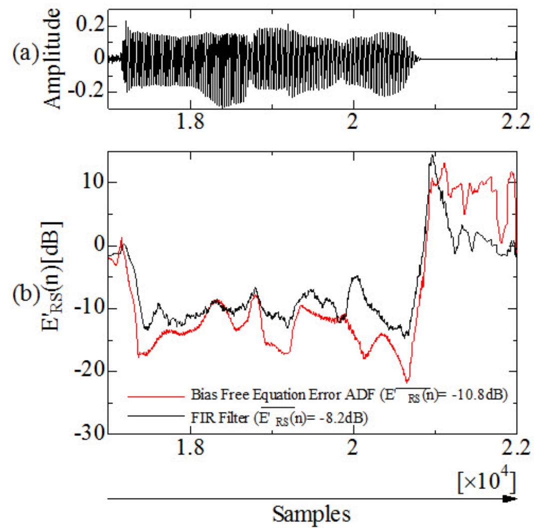

Fig.6: *Performance of NRF (a)Clean speech (b) $E'_{RS}(n)$*

$$\overline{E_N(n)} = 10\log_{10}\frac{\sum_{j=1}^{N'}\xi^2(n)}{\sum_{j=1}^{N'}\{\xi(n)-\hat{\xi}(n)\}^2}. \quad (38)$$

Fig. 7(a) represents clean speech. Fig. 7(b) shows the estimation accuracy of the NRF using a FIR filter [6], the NRF using the bias free equation error ADF without the sub-NRF and the NRF using the bias free equation error ADF with the sub-NRF. Comparing the bias free equation error ADF without the sub-NRF with the FIR filter, the estimation accuracy is degraded because the auto-correlation of background noise is faded out due to variable delay. Comparing the bias free equation error ADF with the sub-NRF with the FIR filter, we observe that $E_N(n)$ is improved. In addition, $\overline{E_N(n)}$ is improved by about 0.5 dB. It can be seen that the sub-NRF has potential to improve the estimation accuracy of background noise.

4.5 Waveforms of tunnel noise reduction

Fig. 8 shows the speech enhancement results when the tunnel noise was used in the simulation. As the speech, the female speech was used. SNR_{in} was set to 0 dB. Fig. 8(a) and 8(b) represent clean speech and noisy speech, respectively, in a 0dB environment. The speech enhancement results by conventional NRS [6], the proposed NRS without sub-NRF and the proposed NRS are illustrated in Fig. 8(c), 8(d) and 8(e), respectively at SNR_{out} = 7.3 dB. Comparing Fig. 8(d) with Fig. 8(c), QS is increased by about 5.3 dB at the same SNR_{out}. It is verified that the Lattice type LPEF and the bias free ADF improve the quality of enhanced speech. On the other hand, comparing Fig. 8(e) with Fig. 8(c), QS is increased by about 7.2 dB. We observe that the proposed NRS has potential to decrease the residual speech in whiten noise effectively and reduce the background noise while maintaining the high quality of enhanced speech.

4.6 Output SNR and Quality of Speech

Figs. 9 and 10 respectively show the noise reduction ability and the quality of enhanced speech for an input SNR. As background noise, the tunnel and factory noise

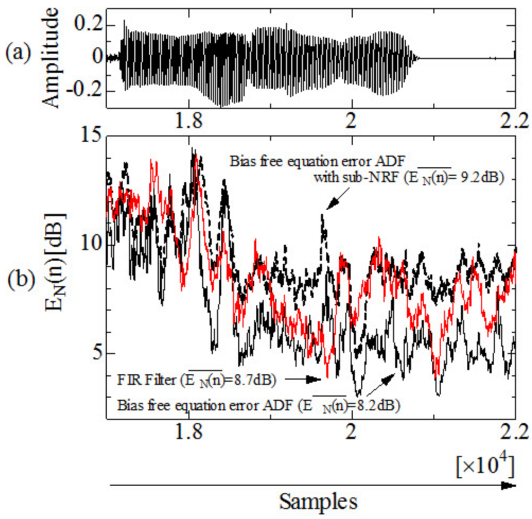

Fig.7: *Noise Estimation Accuracy of NRF and Sub-NRF (a)Clean speech (b)Estimation accuracy of tunnel noise*

were used. The 2 kinds of female speech and 2 kinds of male speech were used as speech signals. The average SNR_{out} and the average QS for -5 dB to 15 dB SNR_{in} are respectively shown in Fig. 9 and Fig. 10. The vertical axis in Fig. 9 and Fig. 10 respectively represents the average of SNR_{out} and QS for 4 kinds of speech. Comparing the proposed NRS with conventional NRS [6], QS are always higher when the SNR_{out} of the conventional and the proposed NRS are the same. Although the QS of the conventional NRS is decreased as SNR_{in} increases, the QS of the proposed NRS is increased. From the simulation results, it can be seen that the proposed NRS improves the quality of enhanced speech.

5. CONCLUSIONS

In this paper, the speech enhancement based on lattice type LPEF and bias free equation error ADF is proposed. At the conventional NRS the residual speech in whitened noise influences the estimation accuracy of background noise. In this paper, the lattice filter which approximates the vocal-tract filter is used as the LPEF for decreasing the residual speech. In

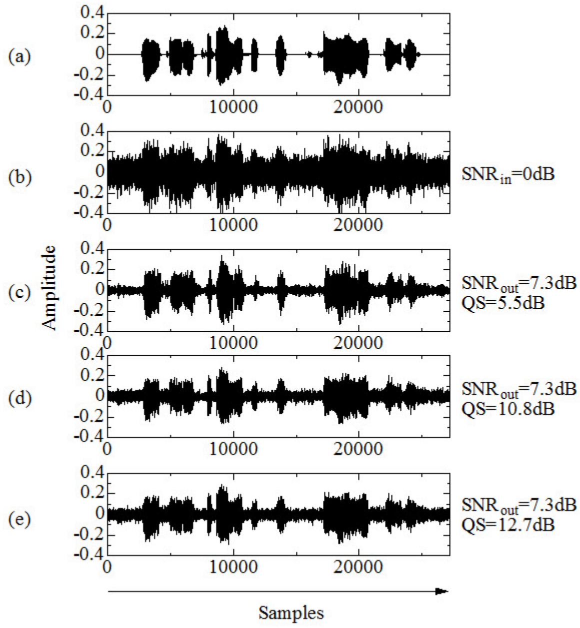

Fig.8: *Waveforms of Simulation results (a)Clean speech (b)Noisy speech (c) Enhanced speech by conventional NRS (d) Enhanced speech by proposed NRS without sub-NRF (e) Enhanced speech by proposed NRS*

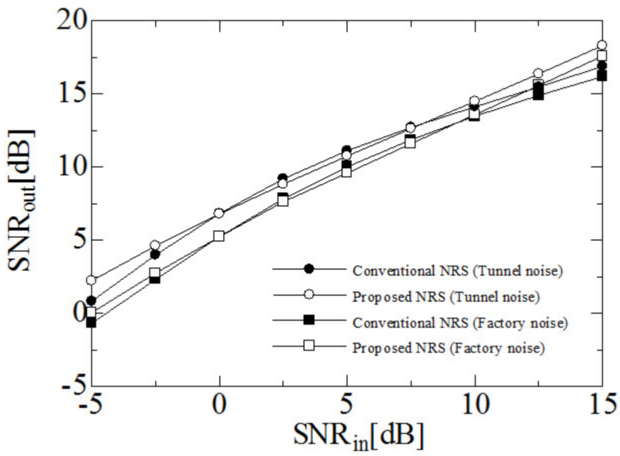

Fig.9: *SNR_{out} Performance*

addition, the bias free equation error ADF is adopted to the NRF for reducing the influence of the residual speech. The bias free equation error ADF takes advantage of identification between the desired signal and disturbance. The influence of the residual speech is reduced by using the cross-correlation between the desired signal and disturbance as tap inputs. However, the bias free equation error ADF causes the degradation of estimation accuracy of background noise. Then, the sub-NRF is used after the bias free equation error ADF. Since the residual speech is reduced enough by lattice type LPEF and the bias free equation error ADF, the sub-NRF can estimate the

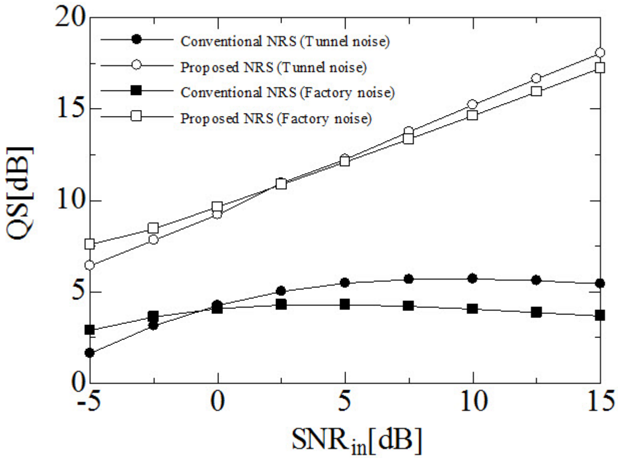

Fig.10: *QS Performance*

background noise accurately with the high quality of enhanced speech. The simulation results show that the proposed system can reduce the actual noise while maintaining the high quality of enhanced speech. In a future work, we will research the more improvement of noise reduction ability and the DSP implementation for a hearing aid.

References

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

[2] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust. Speech Signal Process.*, Vol.ASSP-32, No.6, pp.1109–1121, Dec. 1984.

[3] P.C. Loizou, *Speech Enhancement*, CRC Press, New York, 2007.

[4] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, Vol.9, No.5, pp.504–512, Jul. 2001.

[5] M.R. Sambur, "Adaptive noise canceling for speech signals," *IEEE Trans. Acoust. Speech Signal Process.*, Vol.ASSP-26, No.5, pp.419–423, Oct. 1978.

[6] A. Kawamura, K. Fujii, Y. Itoh and Y. Fukui, "A new noise reduction method using estimated noise spectrum," *IEICE Trans. Fundamentals*, Vol.E85-A, No.4, pp.784–789, Apr. 2002.

[7] N. Sasaoka, M. Watanabe, Y. Itoh and K. Fujii, "A variable step size algorithm for speech noise reduction method based on noise reconstruction system," *IEICE Trans. Fundamentals*, Vol.E92-A, No.1, pp.244–251, Jan. 2009.

[8] R. Kanagawa, E. Kuno, J. Okello, N. Sasaoka, Y. Itoh and M. Kobayashi, "An equation error algorithm based on correlation function with variable delay," *Proc. 16th European Signal Processing Conference*, Aug. 2008.

[9] S. Furui, *Digital Speech Processing*, Tokai University Press, Tokyo, 1985.

[10] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, New Jersey, 1996.

Naoto Sasaoka received B.E., M.E. and Dr.E degrees in electrical and electronic engineering from Tottori University, Tottori, Japan, in 2002, 2004 and 2006 respectively. He is currently an assistant professor in the Graduate School of Engineering, Tottori University. His research interests are in the area of digital signal processing. He is a member of IEEE, IEICE and ASJ.

Shinichi Wada received B.E degree in electrical and electronic engineering from Tottori University, Tottori, Japan, in 2011. Currently he is a master's degree candidate at the Graduate School of Engineering, Tottori University. His research interests are in the area of digital signal processing. He is a member of IEICE.

James Okello received a B. tech. degree in electrical and communication engineering from Moi University, Eldoret, Kenya, in 1992 and the M.E. and Dr.E. degrees from Tottori University, Tottori, Japan, in 1998 and 2001, respectively. Between 2001 and 2003, he was an Assistant Professor with Faculty of Information Engineering, Kyushu Institute of Technology, Iizuka, Japan. Currently he is an assistant manager, Research Engineer with the System IP Core Research Laboratories, NEC. His main areas of interest are in digital signal processing and communications systems. He has also received the Gallenkamp International Award of Moi University, the Telecom System Award 1997, and the Nikkei Large Scale Integration (LSI) Design Award, Japan. He is a member of IEEE and IEICE.

Yoshio Itoh received B.E. degree in electronics engineering from Osaka Institute of Technology, M.E. and Dr.E. degrees in electrical engineering from Osaka Prefecture University in 1979, 1981 and 1991, respectively. He is now a professor in the Graduate School of Engineering, Tottori University, Tottori, Japan. His research interests are in the area of digital signal processing and communication systems. From 2000 to 2004, he was as Associate Editor of Trans. on Fundamentals of IEICE. He is a member of IEEE, IEICE and ASJ.

Masaki Kobayashi received B.E. degree from Meijo University, M.E. degree in electronics engineering from Nagoya Institute of Technology and Dr.E. degree in electrical engineering from Osaka Prefecture University, Sakai, Japan in 1967, 1969 and 1987, respectively. He is now a professor in the faculty of engineering at Chubu University. His research interests are in the area of digital signal processing and digital communication systems. He is a member of IEEE, IEICE and the IEE of Japan.