Study on power envelope subtraction based on modulation transfer function

By Liu Yang

A thesis submitted to
School of Information Science,
Japan Advanced Institute of Science and Technology,
in partial fulfillment of the requirements
for the degree of
Master of Information Science
Graduate Program in Information Science

Written under the direction of
Associate Professor Masashi Unoki

September, 2012

Study on power envelope subtraction based on modulation transfer function

By Liu Yang (1010233)

A thesis submitted to
School of Information Science,
Japan Advanced Institute of Science and Technology,
in partial fulfillment of the requirements
for the degree of
Master of Information Science
Graduate Program in Information Science

Written under the direction of
Associate Professor Masashi Unoki

and approved by
Associate Professor Masashi Unoki
Professor Masato Akagi
Professor Jianwu Dang

August, 2012 (Submitted)

Contents

List of Figures

Chapter 1

Introduction

In real environments, significant features of speech are drastically smeared due to noise and reverberation. The quality of sound and intelligibility of speech are significantly reduced. Thus the effects of noise and reverberation must be removed for various speech signal processing, such as speech emphasis for transmission systems, hearing aid systems and the preprocesing for automatic speech recognition (ASR) systems.

There have already been some famous methods that can remove the effects of noise or reverberation in the real environment. There are, for example, the spectral subtraction method[1], the Kalman filtering method[2], the MMSE-STSA method[3], minimum-phase inverse filtering method[4] and the MINT method[5]. The first three methods can reduce the noise well in only noisy environments, while the last two methods can effectively reduce the effect of reverberation in only reverberant environments. All methods, however, cannot work well in both noisy and reverberant environment simultaneously because of different forms in noisy and/or reverberant environments.

Recently, Kinoshita *et al* proposed a method to restore the speech recorded in the noisy reverberant environment and this method can be divided into two processes: reducing the noise with spectral subtraction method for the noisy reverberant speech and use linear prediction method for dereverberation. Although this is a simple model for reducing the noise and reverberation, it is believed that the combination of different systems cannot deal with noise and reverberation simultaneously.

A novel concept of MTF had been proposed by Houtgast and Steeneken[6], to account for reducing speech intelligibility due to noise and reverberation in the room. A scheme of dereverberation and denoising based on the MTF-concept has been studied deeply by Unoki *et al.*[7] and the method based on MTF concept consists of two parts: power envelope subtraction and power envelope dereverberation. Although the MTF-based dereverberation[8] was reasonably designed, the MTF-based noise subtraction[9] is quite immature. The previous proposed noise subtraction method based on MTF concept equals to the method of subtracting the average value of noise temporal power envelope, but the fluctuationss of the noise power envelope still remain. These fluctuations therefore will be emphasized during dereverberation (inversed MTF).

The goal of our research is to propose a method to remove the fluctuations of the noise in the power envelope subtraction process and evaluate this method in noisy reverberant

environment.

1.1 Motivation

The Kalman filter method was proposed by Basu and Paliwal for the enhancement of speech. This method uses the linear predictive coefficients obtained from the clean speech signal and the other speech parameters obtained from the nonspeech sections. The advantages of Kalman filter are listed as follows:

- Kalman filter uses two moments of the noise. The first moment of noise is the mean value of the noise, the second moment of noise is the variance of the noise.
- Kalman filter can be used for stationary and non-stationary signals.
- Kalman filter can be used in time-variant and time-invariant systems.
- Kalman filter only exploit the current observed value and the most recent estimation value to estimate the current value.
- Kalman filter can overcome the musical tone problem and obtain the good speech quality of reducing the processing distortion of speech signal.
- Kalman filter not only takes advantage of the characteristics of the signal and noise, but also uses the production model of speech, ie, autoregressive model which is an effective model for human speech production system.

We can get a minimum mean-squared error (MMSE) estimation for the clean signal if the noise is Gaussian white noise and we can also get the linear minimum mean-square error estimation when the noise is non-Gaussian white noise. The Kalman filter can reduce the both the Gaussian white noise and non-Gaussian white noise. Based on these advantages of Kalman filter, we can consider to use Kalman filter to remove the fluctuations of the noise for stationary noise and non-stationary noise.

1.2 Thesis goal and outline

Our purpose is to remove the remaining fluctuations of the noise power envelope by Kalman filter method, after removing these noise fluctuations, MTF-based speech enhancement method can be improved greatly to restore the observed speech in both noisy and reverberant environment.

The rest of the thesis is organised as follows:

- **Chapter 2** introduces some classical methods for noise reduction and dereverberation. Since these method cannot deal with noise and reverberation simultaneously, we bring forward the concept of MTF which can solve this problem.

- **Chapter 3** presents the previous MTF based method for speech enhancement which can be divided into two parts: Noise reduction process and dereverberaiton process. The remaining problem of the noise reduction process is that this method can only reduce the mean value of the noise.

- **Chapter 4** proposes a method based on MTF concept by incorporating the Kalman filter to improve the noise reduction process of the previous method. In this chapter, we also discuss about the Linear Prediction Method which is used to get the parameters for the state equation of the Kalman filter and the derivations of the other parameters in the observation equation.

- **Chapter 5** describes the noisy conditions and the noisy reverberant condition for the experiments. I also talks about the two kinds of evaluation methods: Correlation and SER. The analysis of the results is also presented.

- Finally, **Chapter 6** summarize the contribution and achievements the thesis.

Chapter 2

Background

2.1 Classical noise reduction method

2.1.1 Spectral subtraction method

The method of acoustic noise suppression using spectral method was proposed by Boll. This noise suppression method is used for reducing the spectral effects of the additive noise of the speech. As we know, most of the digital speech processors in real environments require the subtraction of the noise from the digital waveform, the advantage of the spectral subtraction method is that it computes more efficiently and is independent on processor for effective speech analysis. This method suppress the stationary noise from speech by subtracting the spectral noise bias calculated from the nonspeech frames, then it subtract the residual noise left after the previous subtraction.

This approach is used to estimate the magnitude frequency spectrum of the clean speech by subtracting the noise magnitude spectrum from the noisy speech spectrum. This estimator needs the estimation of the current noise spectrum. It utilize the approximated average value of the noise magnitude estimated during the nonspeech section rather than the microphone source[18, 19].

This method can be used as a processor for speech recognition systems, speaker authentication systems and voice communication systems.

2.1.2 Kalman filtering method

The Kalman filtering method was first proposed to be applied to speech enhancement by Paliwal and Basu, this method uses the speech parameters and linear predictive coefficients which are obtained from clean speech and the noise characteristics obtained form the nonspeech sections. All these methods have to detect the nonspeech sections for the estimation of noise variance. The Kalman filter can provide the minimum mean-squared error estimation for the clean signal when the noise is Gaussian white noise and it can also provide the linear minimum mean-squared error estimation when the noise is non-Gaussian white noise.

Since the Klaman filter exploits the model of speech production, it has a better performance compared to the Wiener filter. The Kalman filtering method can improve both speech quality and speech intelligibility however the Wiener filtering method can only improve the speech quality.

2.2 Classical dereverberation method

2.2.1 Harmonicity-based blind dereverberation method

The Harmonicity-based dEReverBeration (HERB) method is used to reduce the amount of reverberation in the signal picked up by a single microphone. The method makes extensive use of harmonicity, a unique characteristic of speech, in the design of a dereverberation filter. In particular, harmonicity enhancement is proposed and demonstrated as an effective way of estimating a filter that approximates an inverse filter corresponding to the room impulse response. Two specific harmonicity enhancement techniques are presented and compared; one based on an average transfer function and the other on the minimization of a mean squared error function. Prototype HERB systems are implemented by introducing several techniques to improve the accuracy of dereverberation filter estimation, including time warping analysis. Experimental results show that the proposed methods can achieve high-quality speech dereverberation, when the reverberation time is between 0.1 and 1.0 s, in terms of reverberation energy decay curves and automatic speech recognition accuracy.

2.2.2 Multiple input/output inverse theorem method

The novel method of inverse filtering of room acoustic was proposed by Miyoshi and Kaneda to realize the exact inverse filtering of the room impulse response which is based on the principle of multiple-input/output iverse theorem (MINT). As we know, it is difficult to get the exact inverse filtering of the room acoustics with the previous methods because the impulse response has nonminimum phases. However, this propose method can achieve this goal for the inverse is constructed from the multiple FIR filters by adding extra acoustic signal-transmission channels produced by the multiple loudspeaker or microphones and the coefficients of these FIR filters can be obtained from the famous rules of matrix algebra.

This method only use one acoustic signal-transmission channel and is better than the previous methods. This method can reproduce and receive the sound without any distortion cause by reflected sounds in a room.

2.3 MTF concept

The MTF concept was proposed by Houtgast and Steeneken[6] to account for the relation between the degree of modulation of the envelopes of input and output signals and the characteristics of the enclosure and provide a way to predict the speech transfer index,

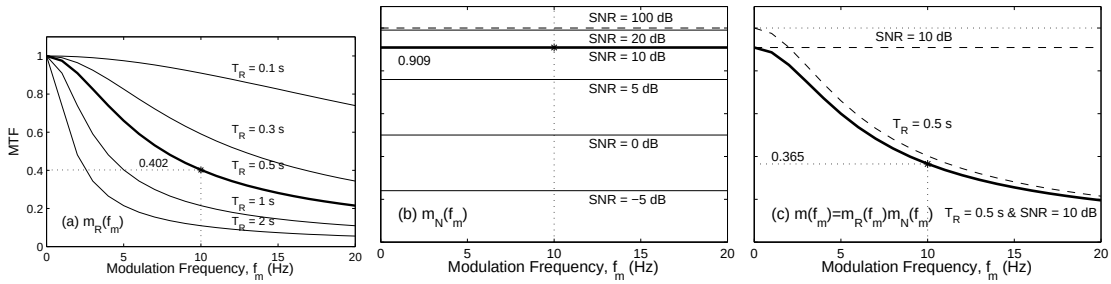

Figure 2.1: Theoretical representations of the MTFs, $m(f_m)$, in (a) reverberant environment, (b) noisy environment, and (c) both noisy and reverberant environments. The bold solid lines indicate the MTF with $T_R = 0.5$ s and SNR = 10 dB.

which is strongly related to intelligibility [6]. This concept was introduced as a measure in room acoustics for assessing the effect of the enclosure on intelligibility. They defined input and output temporal power envelopes as

$$\begin{aligned} \textbf{Input} &= \overline{I_i^2}(1+\cos(2\pi f_m t)), && (2.1)\\ \textbf{Output} &= \overline{I_o^2}\left\{1+m(f_m)\cos(2\pi f_m(t-\tau))\right\}, && (2.2) \end{aligned}$$

where $\overline{I_i^2}$ and $\overline{I_o^2}$ are the input and output intensities, f_m is the modulation frequency, and τ is the phase information. The modulation index of the power envelope is $m(f_m)$ and referred to as MTF. We will now explain the MTF in noisy and/or reverberant environments.

2.3.1 Model concept based on the MTF

We assume the output, the input, the impulse response, and the noise signals to be $y(t)$, $x(t)$, $h(t)$, and $n(t)$. These are modeled based on the MTF [8, 11, 12] as follows:

$$\begin{aligned} y(t) &= h(t) * x(t) + n(t), && (2.3)\\ x(t) &= e_x(t)c_x(t), && (2.4)\\ h(t) &= e_h(t)c_h(t) = a\exp(-6.9t/T_R)c_h(t), && (2.5)\\ n(t) &= e_n(t)c_n(t), && (2.6) \end{aligned}$$

where $e_x(t)$, $e_h(t)$, and $e_n(t)$ are the temporal envelopes of $x(t)$, $h(t)$, and $n(t)$. $c_x(t)$, $c_h(t)$, and $c_n(t)$ are carriers such as random variable. Here, $\langle c_l(t), c_l(t-\tau)\rangle = \delta(\tau)$ and $\langle\cdot\rangle$ is an ensemble average operation. T_R is the reverberation time. In this model, $e_y^2(t)$ can be derived as

$$\begin{aligned} \left\langle y^2(t)\right\rangle &= \left\langle h^2(t) * x^2(t)\right\rangle + \left\langle n^2(t)\right\rangle, && (2.7)\\ e_y^2(t) &= e_h^2(t) * e_x^2(t) + e_n^2(t). && (2.8) \end{aligned}$$

(see [8, 11] for a detailed derivation of Eq. (2.8)). We used the relationship between temporal power envelopes to restore $e_x^2(t)$ from the observed $e_y^2(t)$.

2.3.2 MTF in reverberant environments

In the reverberant condition, the input and output temporal power envelopes, $e_x^2(t)$ and $e_y^2(t)$, are represented as

$$e_x^2(t) = \overline{e_x^2}(1 + \cos(2\pi f_m t)), \tag{2.9}$$

$$e_y^2(t) = e_x^2(t) * e_h^2(t) = \frac{\overline{e_x^2}}{\alpha}\left\{1 + m_R(f_m)\cos(2\pi f_m t)\right\}, \tag{2.10}$$

where $\alpha = \int_0^\infty h^2(t)dt$ and $\beta = \int_0^\infty h^2(t)\exp(-j\omega_m t)dt$. The complex MTF in reverberant environments is defined as

$$m_R(f_m) = \left|\frac{\beta}{\alpha}\right| = \sqrt{1 + \left(2\pi f_m \frac{T_R}{13.8}\right)^2} \tag{2.11}$$

The MTF in reverberant environments depends on f_m. This means the low-pass characteristics as a function of T_R (as shown in Fig. 2.1(a)). In the case of a T_R of 0.5 s, $m(f_m)$ at $f_m = 10$ Hz is 0.402.

2.3.3 MTF in noisy environments

Where there is additive noise, $e_y^2(t)$ is represented as

$$e_y^2(t) = e_x^2(t) + e_n^2(t) = \left(\overline{e_x^2} + \overline{e_n^2}\right)\left\{1 + m_N(f_m)\cos(2\pi f_m t)\right\}, \tag{2.12}$$

where $\overline{e_n^2} = \frac{1}{T}\int_0^T e_n^2(t)dt$. Here, $e_n^2(t)$ is assumed to be constant in the time domain and T is the signal duration. The complex MTF in noisy environments, is defined as

$$m_N(f_m) = \frac{\overline{e_x^2}}{\overline{e_x^2} + \overline{e_n^2}} = \frac{1}{1 + 10^{-(\mathrm{SNR})/10}}, \tag{2.13}$$

where $\mathrm{SNR} = 10\log_{10}(\overline{e_x^2}/\overline{e_n^2})$ in dB. This MTF is independent of f_m and reduced as a function of SNR (Fig. 2.1(b)). In the case of SNR of 10 dB, $m(f_m)$ is 0.909.

2.3.4 MTF in noisy and reverberant environments

The MTF in noisy reverberation environments calculated from Eqs. (2.11) and (2.13), can be represented as

$$\begin{aligned} m(f_m) &= m_R(f_m)\cdot m_N(f_m) \\ &= \sqrt{1 + \left(2\pi f_m \frac{T_R}{13.8}\right)^2} \Big/ \left(1 + 10^{-\frac{\mathrm{SNR}}{10}}\right). \end{aligned} \tag{2.14}$$

The MTF in noisy reverberant environments depends on f_m. This means the low-pass characteristics resulting from reverberation as a function of T_R and the constant attenuation resulting from noise as a function of SNR (Fig. 2.1(c)). In the case of a T_R of 0.5

s and SNR = 10 dB, $m(f_m)$ at $f_m = 10$ Hz is 0.365 ($= 0.402 \times 0.909$). Hence, the effect of noise and reverberation can be suppressed by using the inverse filtering of MTF in Eq. (2.14).

Chapter 3

Previous proposed scheme

The previous method uses the power envelope restoration based the MTF concept. The block-diagram of the method is shown in Fig. 3.1. This method consists of (i) power envelope extraction, (ii) power envelope subtraction, and (iii) power envelope inverse filtering with parameter estimation. Here, the constant bandwidth filterbank was used to analyze the signal.

3.1 Power envelope subtraction process

3.1.1 Power envelope extraction

There are many well known methods for signal demodulation in AM transmission. For example, the low-pass half-wave rectification method[20] and the synchronous demodulation method. Both methods assume that the carrier signal is sinusoidal with a single frequency. So it cannot precisely extract the $e_y^2(t)$ if either of the methods is used in extracting the power envelope from an observed reverberant signal based on the MTF concept because the carrier is a white-noise signal.

We can use two methods to extract the power envelope in this research. The first

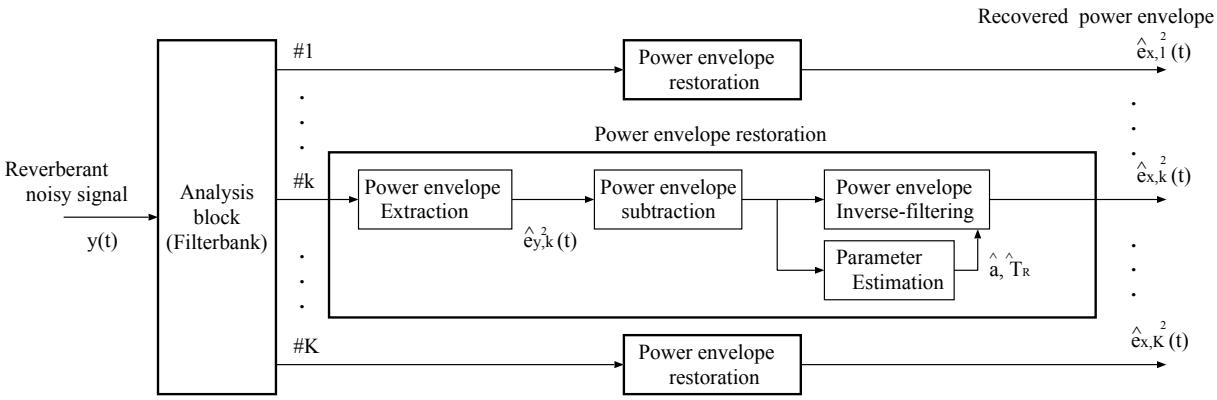

Figure 3.1: The power envelope restoration method.

method is called ensemble average method which is a straightforward method base on Eq.2.3, we assume that the product of each white noise signal becomes the other white noise signal. Let $\hat{n}(t)$ to be a set of white noise signals composed of a number of white noise, we assume that $\hat{y}(t) = y(t)\hat{n}(t)$ as a quasi-set of $y(t)$, we can use Eq. 2.3 to extract the power envelope from reverberant signal:

$$\hat{e}_y(t)^2 := \mathrm{LPF}[\langle \hat{y}(t)^2 \rangle] = \mathrm{LPF}[\langle (y(t)\hat{n}(t))^2 \rangle \tag{3.1}$$

In this equation, we use the low-pass filter to remove the higher frequency components in the power envelope caused by the estimation of $\hat{n}(t)$.

The second method is composed of the low-pass filtering and the Hilbert transform[21]. The Hilbert transform is widely used in calculating the instantaneous amplitude for signals. In this method, the carrier should be even or odd functions rather than single frequency sinusoidal signals. The extraction method of temporal power envelopes is as follows:

$$\hat{e}_y^2(t) = \mathrm{LPF}\left[\left|y(t) + \jmath\mathrm{Hilbert}\{y(t)\}\right|^2\right], \tag{3.2}$$

where LPF$[\cdot]$ is a low-pass filtering. Hilbert$[\cdot]$ is the hilbert transform. This method is based on calculation of the instantaneous amplitude of the signal, and used low-pass filtering as post-processing to remove the higher frequency components in the power envelopes. We use LPF with the cut-off frequency of 20 Hz because the most important modulaiton region for speech perception[22] and speech recognition is from 1 Hz to 16 Hz[23, 24].

3.1.2 Implementation

This section explains the previous noise suppression method based on the MTF concept. The modulation index and the averaged power in Eq. (2.13) are affected by noise. We restore the averaged power levels to suppress the noise effects. Eq. (3.3) is the offset value of he averaged power, given by

$$\mathbf{OV} = \frac{\overline{e_x^2}}{\overline{e_x^2} + \overline{e_n^2}}. \tag{3.3}$$

Substituting Eq. (3.3) into Eq. (2.13), the following equation can be obtained

$$\overline{e_x^2} + \overline{e_x^2} \cdot m(f_m) \cdot \cos(2\pi f_m t). \tag{3.4}$$

By multiplying a second term of Eq. (3.4) by $1/m(f_m)$ for restoring the modulation index. We obtain the following equation as

$$\begin{aligned} \hat{e}_x^2(t) &= \overline{e_x^2} + \left(\overline{e_x^2} \cdot m(f_m) \cdot \cos 2\pi f_m t\right) \times \frac{1}{m(f_m)}, & (3.5) \\ &= \overline{e_x^2}(1 + \cos(2\pi f_m t)). & (3.6) \end{aligned}$$

Here, the robust VAD method is used to calculate $\overline{e_n^2}$ (N) and $(\overline{e_x^2} + \overline{e_n^2})$ (SN) in Eq. (2.13) from the observed $e_y^2(t)$ in noise duration and signal+noise duration respectively. From this equation, we can see that this method equals to the method of subtracting the average

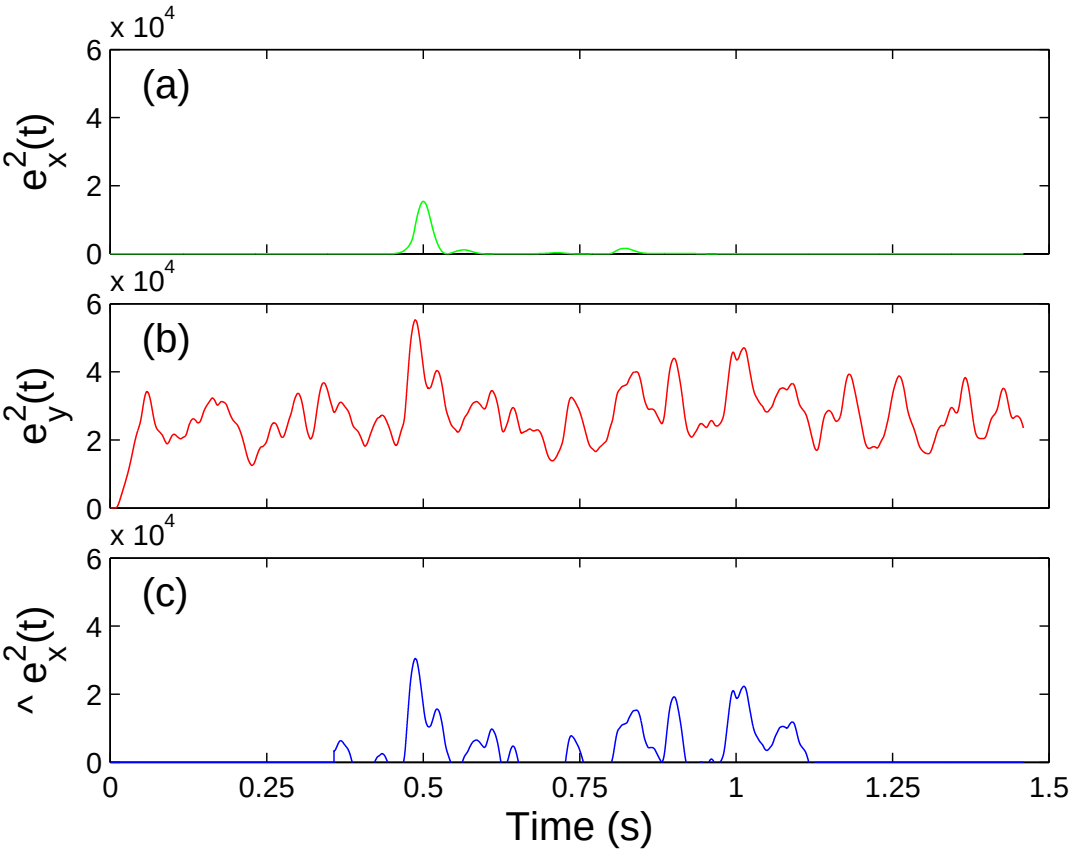

Figure 3.2: Example of power envelopes for one channel: (a) clean power envelope, (b) noisy power envelope, (c) restored power envelope of previous MTF based method

value of noise temporal power envelope form the output temporal power envelope. Figure 3.2 shows the comparison among clean power envelope, noisy power envelope and restored power envelope of previous method in one channel, we can see the clean power envelope has one dominant peak and two small peaks, but the restored power envelope has more than ten peaks after using the previous method. We can conclude that this method can only reduce the mean value of noise. Therefore, it is necessary to proposed an improved method to reduce the remaining fluctuations of the noise power envelope.

3.2 Dereverberation process

On the basis of this result, $e_x^2(t)$ can be recovered by deconvoluting $e_y^2(t) = e_x^2(t) * e_h^2(t)$ in Eq. (2.8) with $e_h^2(t)$. Here, the transmission functions of power envelopes $E_x(z)$, $E_h(z)$, and $E_y(z)$ are assumed to be the z-transforms of $e_x^2(t)$, $e_h^2(t)$, and $e_y^2(t)$. Thus, the $E_x(z)$

can be determined from

$$E_x(z) = \frac{E_y(z)}{a^2}\left\{1 - \exp\left(-\frac{13.8}{T_R \cdot f_s}\right) z^{-1}\right\}, \tag{3.7}$$

where f_s is the sampling frequency. The power envelope $e_x^2(t)$ can then be obtained from the inverse z-transform of $E_x(z)$. Here, two parameters (T_R and a) are obtained as [12]

$$\hat{T}_R = \underset{0 \le T_R \le T_{R,\max}}{\arg\min} \left\{\frac{dT_P(T_R)}{dT_R}\right\}, \tag{3.8}$$

$$T_P(T_R) = \min\left(\underset{t_{\min} \le t \le t_{\max}}{\arg\min} \left|\hat{e}_{x,n,T_R}(t)^2 - \theta\right|\right), \tag{3.9}$$

$$\hat{a} = \sqrt{1/\int_0^T \exp(-13.8t/\hat{T}_R)dt}. \tag{3.10}$$

3.3 Example

The example of how the power envelope restoration is related to the MTF concept is shown in Fig. 3.3. A sinusoidal power envelope as the original $e_x^2(t)$ ($= 0.5(1 + \sin(2\pi f_m t))$) and $x(t)$ calculated from $e_x^2(t)$ and a white noise carrier $c_x(t)$ using Eq. (2.4) are shown in Figs. 3.3(a) and (b); f_m was 10 Hz and $m(f_m)$ was 1. Figures 3.3(c) and (d) show $e_h^2(t)$ with $T_R = 0.5$ s and $h(t)$ of Eq. (2.5). An $e_n^2(t)$ and an $n(t)$ of Eq. (2.6) with an SNR of 3 dB are shown in Figs. 3.3(e) and (f), and we show $e_y^2(t)$ ($= e_x^2(t) * e_h^2(t) + e_n^2(t)$ and the observed noisy reverberant signal $y(t)$ ($=x(t)*h(t)+n(t)$) in Figs. 3.3(g) and (h). The left panels ((a), (c), (e), and (g)) show the power envelopes and the right panels ((b), (d), (f), and (h)) show the corresponding signals. As shown in this figure, $m(f_m)$ decreased from 1.0 (in Fig. 3.3(a)) to 0.404×0.5 (the maximum deviation of the envelope between the dotted lines in Fig. 3.3(e) relative to that in Fig. 3.3(a) and the reduction in Fig. 3.3(g)). The solid line in Fig. 3.3(g) shows the restored power envelope $\hat{e}_x^2(t)$ obtained from the noisy reverberant power envelope $e_y^2(t)$ (Fig. 3.3(g)) using Eqs.(3.7) with $T_R = 0.5$ s and SNR = 3 dB. It is shown that using power envelope restoration can precisely restore the power envelope from a noisy reverberant signal in terms of its shape and magnitude.

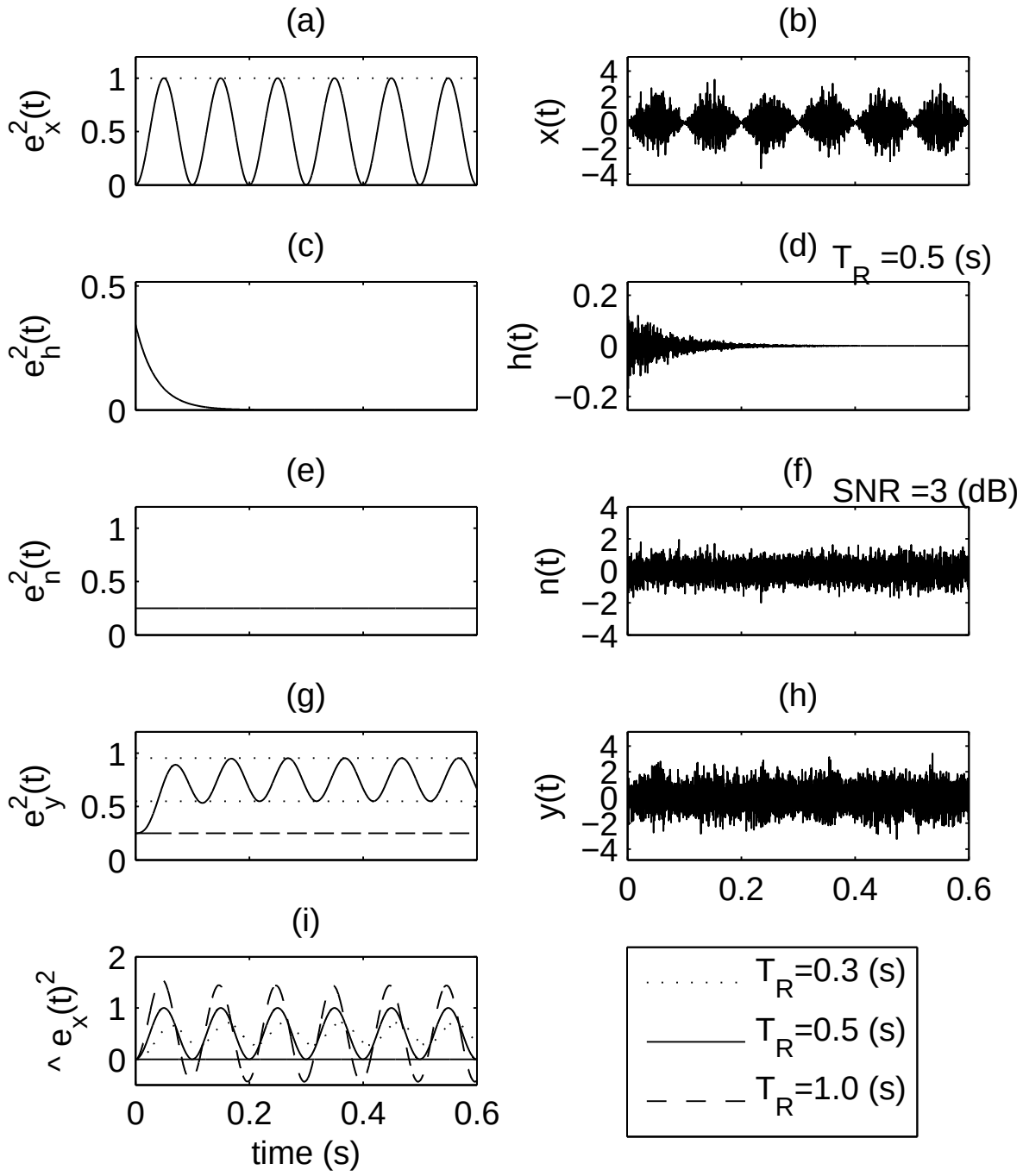

Figure 3.3: Example of relationship between power envelopes of system based on the MTF concept: (a) power envelope $e_x^2(t)$ of (b) original signal $x(t)$, (c) power envelope $e_h^2(t)$ of (d) simulated room impulse response $h(t)$ ($T_R = 0.5$ s), (e) power envelope $e_n^2(t)$ of (f) noise signal $n(t)$, (g) power envelope $e_y^2(t)$ derived from $e_x^2(t) * e_h^2(t) + e_n^2(t)$, (h) noisy reverberant signal $y(t)$ derived from $x(t) * h(t) + n(t)$, and (i) restored power envelope $\hat{e}_x^2(t)$.

Chapter 4

Improved proposed scheme

The block-diagram of the improved method is shown in Fig. 4.1. The method consists of: (i)power envelope extraction, (ii)power envelope subtraction of previous proposed method and (iii) power envelope subtraction using Kalman filter with linear prediction method.

4.1 Kalman filter based on MTF concept

4.1.1 Kalman filter definition

The Kalman filter, together with its basic variants, are some of the most widely applied tools in fields related to statistical processing, especially in casual, real-time applications. The Kalman filter can be used to resolve the problems of residual and musical noise, and to achieve quite good quality by reducing the distortion. It not only exploits the statistical characteristics of signal and noise but also utilizes the speech production model. Therefore, Kalman filter can be used to reduce the fluctuations of the noise. The state and observation equation are the main equations in Kalman filter and they are represented as:

$$\boldsymbol{X}(k) = F\boldsymbol{X}(k-1) + \boldsymbol{U}(k), \tag{4.1}$$

$$\boldsymbol{Y}(k) = H\boldsymbol{X}(k) + \boldsymbol{V}(k), \tag{4.2}$$

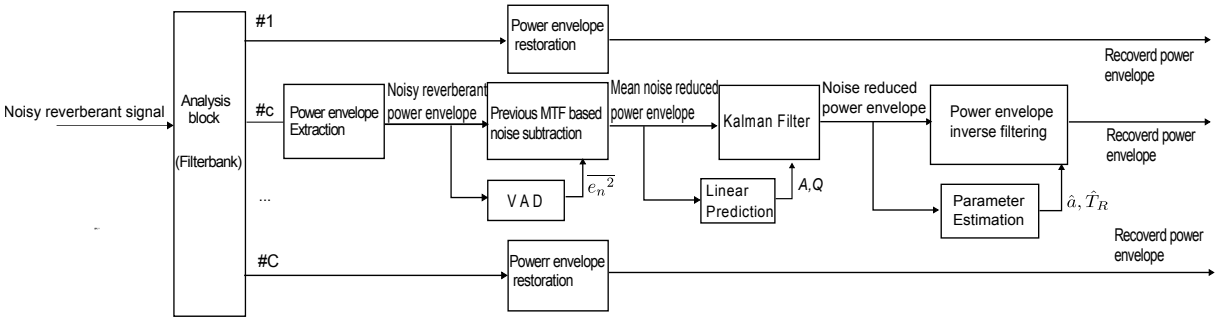

Figure 4.1: Proposed model.

where $\boldsymbol{X}(k)$ is the system state of time k, $\boldsymbol{Y}(k)$ is the observation value of time k. $\boldsymbol{U}(k)$ and $\boldsymbol{V}(k)$ are driving noise and observation noise. They are assumed to be white noise.

4.1.2 Kalman filter with MTF concept

We combine the Kalman filter with MTF concept. The state equation of power envelope based on Kalman filter is defined as:

$$\boldsymbol{e}_x^2(k) = A\hat{\boldsymbol{e}}_x^2(k-1) + \epsilon(k), \tag{4.3}$$

where $\boldsymbol{e}_x^2(k)$ is the state vector of time k, in this research, the sampling frequency is 20 kHz and the k-th sampling point corresponds to the time k. The speech signal can be modelled with an AR process of order p, the state vector then can be represented as:

$$\begin{aligned} \boldsymbol{e}_x^2(k) &= \big[\hat{e}_x^2(k-p+1), \hat{e}_x^2(k-p+2), \\ &\qquad \hat{e}_x^2(k-p+3), \cdots, \dot{e}_x^2(k)\big]^{\mathrm{T}}, \end{aligned} \tag{4.4}$$

where $\hat{e}_x^2(k)$ is the power envelope of the optimal estimation of time k and $\dot{e}_x^2(k)$ is the estimation from the state equation. In this research, we choose $p = 6$. $\hat{\boldsymbol{e}}_x^2(k-1)$ is the state vector of time $k-1$ as

$$\begin{aligned} \hat{\boldsymbol{e}}_x^2(k-1) &= \big[\hat{e}_x^2(k-p), \hat{e}_x^2(k-p+1), \\ &\qquad \hat{e}_x^2(k-p+2), \cdots, \hat{e}_x^2(k-1)\big]^{\mathrm{T}}. \end{aligned} \tag{4.5}$$

F is transition matrix which can be obtained by linear prediction method. $\epsilon(k)$ is assumed to be white noise and the variance of $\epsilon(k)$ is Q.

The observation equation of power envelope based on Kalman filter is defined as:

$$e_y^2(k) = H\,\boldsymbol{e}_x^2(k) + (e_n^2(k) - \overline{e_n^2}(k)), \tag{4.6}$$

where $e_y^2(k)$ is noisy power envelope of time k, this power envelope is the mean value of the noise reduced power envelope derived from the previous MTF method. H is the observation matrix, in this research, $H = [0, 0 \cdots 1]$ and $(e_n^2(k) - \overline{e_n^2}(k))$ is the mean value of the noise reduced noise power envelope whose mean value is zero and the variance of the noise power envelope is R which is calculated by the robust VAD method. We need five steps to calculate the optimal estimations:

Step 1: We set the initial state vector $\boldsymbol{e}_x^2(1|1) = (10^{-12} \cdots 10^{-12})$. Then we can estimate the power envelope of time 2 from initial state vector. Repeat this step, we can estimate the power envelope of time k from the optimal estimation of time $k-1$.

$$\boldsymbol{e}_x^2(k|k-1) = A\hat{\boldsymbol{e}}_x^2(k-1|k-1). \tag{4.7}$$

Step 2: Update the covariance of $\boldsymbol{e}_x^2(k|k-1)$ form the covariance of $\hat{\boldsymbol{e}}_x^2(k-1|k-1)$. we set the initial $P(1|1) = diag(R \cdots R)$.

$$P(k|k-1) = AP(k-1|k-1)A^T + Q. \tag{4.8}$$

Step 3: Estimate the current value and smooth the previous value.

$$\hat{\boldsymbol{e}}_x^2(k|k) = \boldsymbol{e}_x^2(k|k-1) + G(k)(e_y^2(k) - H\boldsymbol{e}_x^2(k|k-1)), \tag{4.9}$$

where $G(k)$ is the Kalman gain and $G(k)(e_y^2(k) - H\boldsymbol{e}_x^2(k|k-1))$ is called new information. In this step, it also innovate the value of the previous estimation.

Step 4: Update the Kalman gain.

$$G(k) = P(k|k-1)H^T/(HP(k|k-1)H^T + R). \tag{4.10}$$

Step 5: Update the covariance of $\hat{\boldsymbol{e}}_x^2(k|k)$.

$$P(k|k) = (I - G(k)H)P(k|k-1), \tag{4.11}$$

where I is unit matrix.

4.2 Parameter estimation of linear prediction method

Linear predictive coding or LPC analysis is one of the most famous speech analysis methods. The basic idea of LPC analysis is that each speech sample can be represented by a linear combination of previous samples. In this section, the linear prediction model and the most famous two methods for calculating the linear prediction parameters are introduced.

4.2.1 linear prediction model

We use the linear prediction method to estimate the parameter A in Kalman filter. Assume that the sampling sequence of clean power envelope is $e_x^2(k)$, $k = 1, 2, \cdots, K$. We can use the previous p samples to estimate the current value. The model of linear prediction is represented as:

$$\hat{e}_x^2(k) = \sum_{i=1}^{p} \alpha_i e_x^2(k-i), \tag{4.12}$$

where $\hat{e}_x^2(k)$ is the optimal estimation of $e_x^2(k)$ under the principle of MMSE, $\alpha_1, \alpha_2, \cdots, \alpha_p$ are linear prediction coefficients. p is the order. In this research, we choose $p = 6$.

4.2.2 Estimation of LPC coefficients

There are two famous methods for calculating coefficients of LPC:

- Autocorrelation
- Covariance

Both methods use short-term filter coefficients and the energy of residual signal is minimized. They both use the principle of minimum squared-errors.

Windowing

Signal analysis assumes that the property of signal relatively slow with time. We can use short-term analysis of a signal. The signal is divided into continuous segments. The signal $s(n)$ is multiplied by a fixed length analysis window $w(n)$ to extract a particular segment of one time. A typical window function called Hamming window is widely used which has the form of:

$$w(n) = \begin{cases} 0.54 - 0.46\cos(2\pi\frac{n}{N_w-1}), \text{if} \quad 0 \leq n \leq N_w - 1 \\ 0, \qquad\qquad\qquad\qquad\quad \text{otherwise} \end{cases} \tag{4.13}$$

Although the Hamming method provides improved side-lobe behavior, but it will broaden the main-lobe of the spectral estimator. In order to maintain the resolution properties that are needed to justify representing the speech spectral properties, the window width must be more than 2.5 times of the average pitch period.

Covariance method

The covariance method is similar to autocorrelation method. The covariance method windows the error signal while the autocorrelation method windows the original speech signal. The energy of windowed error signal is

$$E = \sum_{n=-\infty}^{\infty} e_w^2(n) = \sum_{-\infty}^{\infty} e^2(n)w(n) \tag{4.14}$$

for $1 \leq k \leq p$, we have the following p linear equations:

$$\sum_{k=1}^{p} \varphi(i,k)\alpha_k = \varphi(i,0), \tag{4.15}$$

where $\varphi(i,k)$ is the covariance function of $s(n)$ which is defined as:

$$\varphi(i,k) = \sum_{-\infty}^{\infty} w(n)s(n-i)s(n-k) \tag{4.16}$$

The covariance matrix is symmetric and it may be not invertible, if the LPC filter is unstable the equations may have no solution.

Autocorrelation method

The speech signal $s(n)$ is windowed by $w(n)$ to get the windowed speech segment $s_w(n)$:

$$s_w(n) = w(n)s(n) \tag{4.17}$$

The residual is defined as:

$$e = \sum_{-\infty}^{\infty} e^2(n) = \sum_{-\infty}^{\infty} \left(s_w(n) - \sum_{k=1}^{p} \alpha_k s_w(n-k) \right)^2 \tag{4.18}$$

If we assume that $\frac{\partial e}{\partial \alpha_k} = 0$, we can get:

$$\sum_{k=1}^{p} \alpha_k \sum_{n=-\infty}^{\infty} s_w(n-i)s_w(n-k) = \sum_{n=-\infty}^{\infty} s_w(n-i)s_w(n), \tag{4.19}$$

The autocorrelation function of the windowed segment $s_w(n)$ is defined as:

$$R(i) = \sum_{n=i}^{N_w-1} s_w(n)s_w(n-i), \qquad 1 \le i \le p, \tag{4.20}$$

where N_w is the length of the window. After substituting the values from equation 4.20 to equation 4.19, we can get:

$$\sum_{k=1}^{p} R(|i-k|)\alpha_k = R(i), \tag{4.21}$$

This allow the linear equations to be calculated by some efficient recursive procedures and the most famous method is Durbin's algorithm.

In this research, we use the autocorrelation method to calculate the parameters, the sampling frequency is 20 kHz, we divide the noisy power envelope into 15 ms frames and the frame size is 300 samples. The order p is set to be 6, the AR coefficients were updated for every frame. The data for computing the AR parameters were 600 samples, i.e. the current noisy frame and the previous enhanced frame. In this process, we use the Levinson-Durbin method. Then we can get the transition matrix for Kalman filter:

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \alpha_p & \alpha_{p-1} & \alpha_{p-2} & \cdots & \alpha_1 \end{bmatrix} \tag{4.22}$$

In our research, we use the delayed Kalman filter because the use of estimation of the clean power envelope of time k-p+1 calculated at time k would result in better performance relative to the value that was filtered for the first time (since more information is incorporated for in calculating this value).

Figure 4.2 shows the comparison among clean power envelope, noisy power envelope and restored power envelope using proposed method. We use the same clean power envelope as Fig. 3.2, the restored power envelope has one dominant peak and two small peaks and they are corresponded to the same time with clean power envelope. We can conclude that the proposed method can remove the fluctuations of the noise effectively.

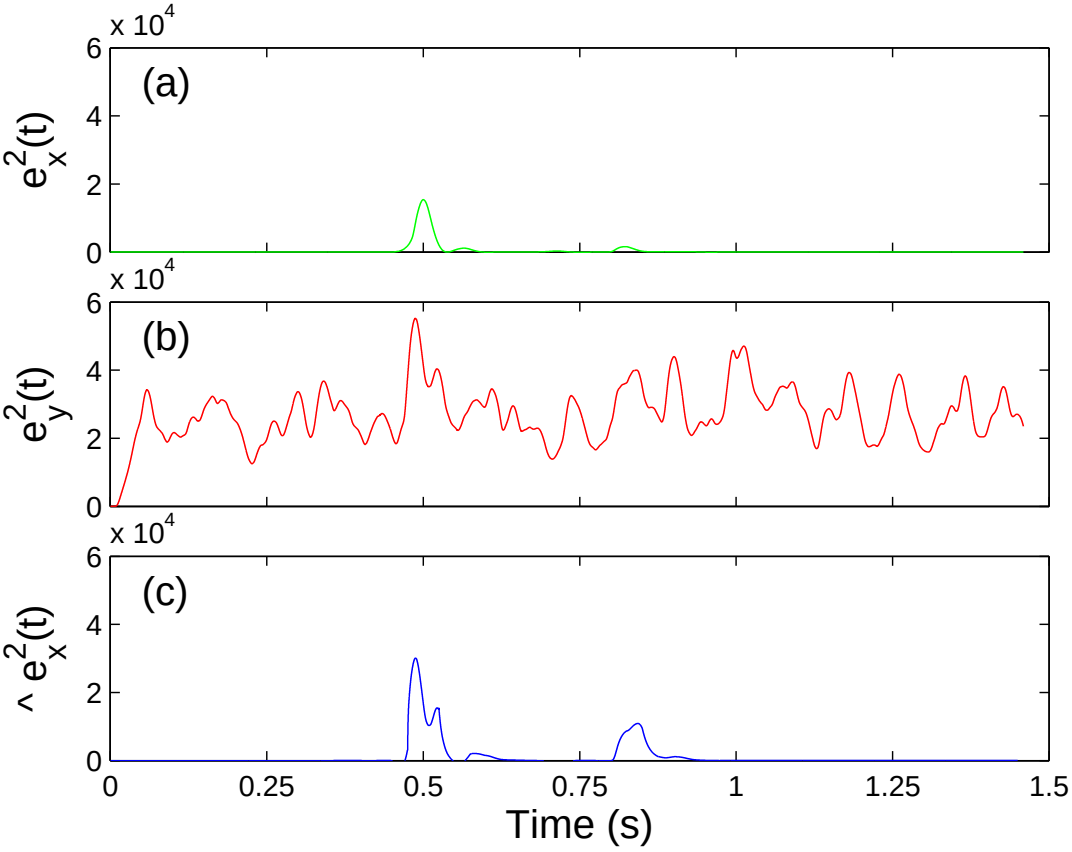

Figure 4.2: Example of power envelopes of improved method: (a) clean power envelope, (b) noisy power envelope, (c) restored power envelope of the improved MTF based method.

Chapter 5

Experiments and evaluation

5.1 Database and conditions

5.1.1 Noisy condition

We carried out the following simulations to evaluate the proposed model in noisy environment. The speech signals were three Japanese sentences (/aikawarazu/,/shinbun/,/joudan/) uttered by ten speakers (five males and five famales) from ATR database [14]. We used 3 kinds of noise signals $n(t)$: white noise, pink noise and factory noise. Signal to noise ratios (SNRs) were fixed at 20, 10, 5, 0, and -5 dB. Sampling frequency of signal is 20 kHz. We used filterbank for speech restoration, and divided signal into 100 bands. The bandwidth of each channel was set to be 100 Hz.

5.1.2 Noisy reverberant condition

We carried out the following simulations to evaluate the proposed method in noisy reverberant environment. The speech signals were three Japanese sentences (/aikawarazu/, /shinbun/, /joudan/) uttered by ten speakers (five males and five females) from the ATR database [14]. We used 100 artificial impulse responses $h(t)$ and 100 white noise signals $n(t)$. Five reverberation times (T_R = 0.1s, 0.3s, 0.5s, 1s, 2s) were used. Signal to noise ratios (SNRs) between $x(t)$ and $n(t)$ were fixed at 20, 10, 5, 0 and -5 dB. All reverberant signals and all noisy signals were generated by convolving $x(t)$ with $h(t)$ and by adding $n(t)$ to $x(t)$. All noisy reverberant signals $y(t)$ were also used. The sampling frequency of signal f_s is 20 kHz.

5.2 Measurements

SER (where S is the original power envelope and E is the difference between the original power envelope and the restored power envelope). This is one of the best evaluation measures for measuring the restoration error between temporal envelopes (with magnitude)

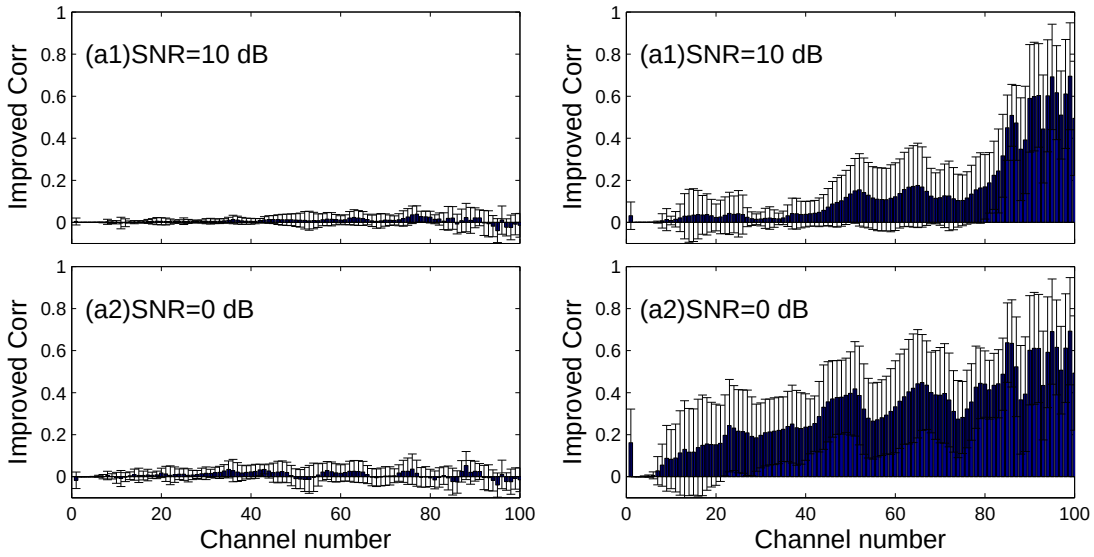

Figure 5.1: Improvement correlation for white noise in noisy environment: (left) Improvement of previous method, (right) Improvement of improved method .

but cannot be used to judge the similarity between the envelopes (with shape). In order to evaluate both error and similarity of the power envelopes, we use correlation and SER.The correlation (Corr) and SER can be calculated by the following equations:

$$\mathrm{Corr}(e_x^2, \hat{e}_x^2) = \frac{\int_0^T \left(e_x^2(t) - \overline{e_x^2}\right)\left(\hat{e}_x^2(t) - \overline{\hat{e}_x^2}\right) dt}{\sqrt{\left\{\int_0^T (e_x^2(t) - \overline{e_x^2})dt\right\}\left\{\int_0^T (\hat{e}_x^2(t) - \overline{\hat{e}_x^2})dt\right\}}}, \tag{5.1}$$

$$\mathrm{SNR}(e_x^2, \hat{e}_x^2) = 10\log_{10}\frac{\int_0^T (e_x^2(t))^2 dt}{\int_0^T (e_x^2(t) - \hat{e}_x^2(t))^2 dt}, \tag{5.2}$$

where $\overline{e_x^2}$ is the average value of $e_x^2(t)$, $\hat{e}_x^2(t)$ is the restored temporal power envelope. The improvements in Corr and SNR are calculated from $\mathrm{Corr}(e_x^2, \hat{e}_x^2) - \mathrm{Corr}(e_x^2, e_y^2)$, and $\mathrm{SNR}(e_x^2, \hat{e}_x^2) - \mathrm{SNR}(e_x^2, e_y^2)$. Note that the positive values indicate the temporal power envelope and waveform of speech were restored from the noisy signal to a certain degree.

5.3 Improvement of Corr and SER in noisy environment

Figure. 5.1 to Figure. 5.9 show the results of the improvements of previous method, the improved method using Kalman filter and the improvement between these two methods for white noise, pink noise and factory noise.

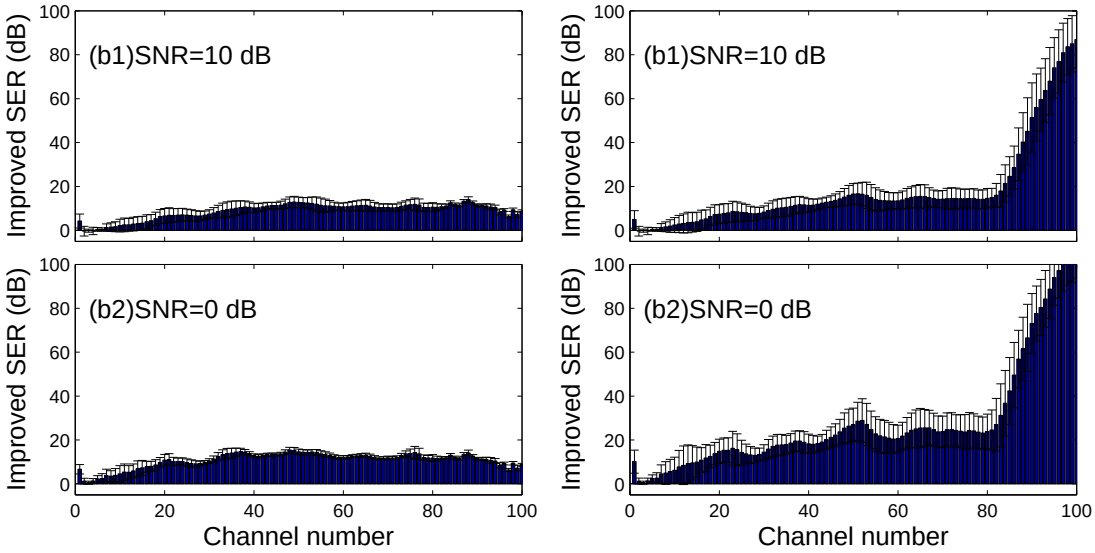

Figure 5.2: Improvement SER for white noise in noisy environment: (left) Improvement of previous method, (right) Improvement of improved method.

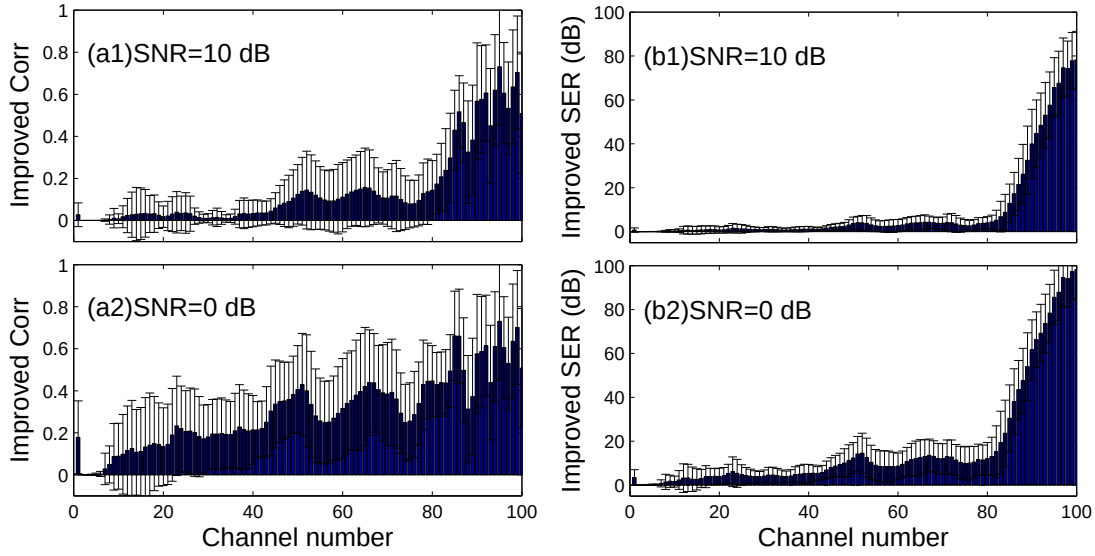

Figure 5.3: Improvement of between both methods for white noise in noisy environment: (left) Correlation improvement, (right) SER improvement.

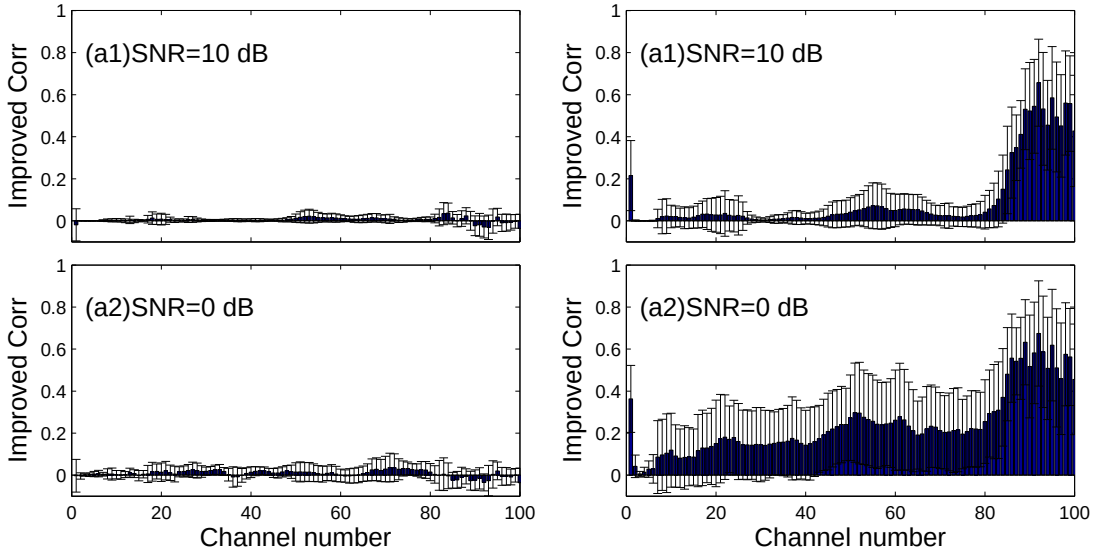

Figure 5.4: Improvement correlation for pink noise in noisy environment: (left) Improvement of previous method, (right) Improvement of improved method.

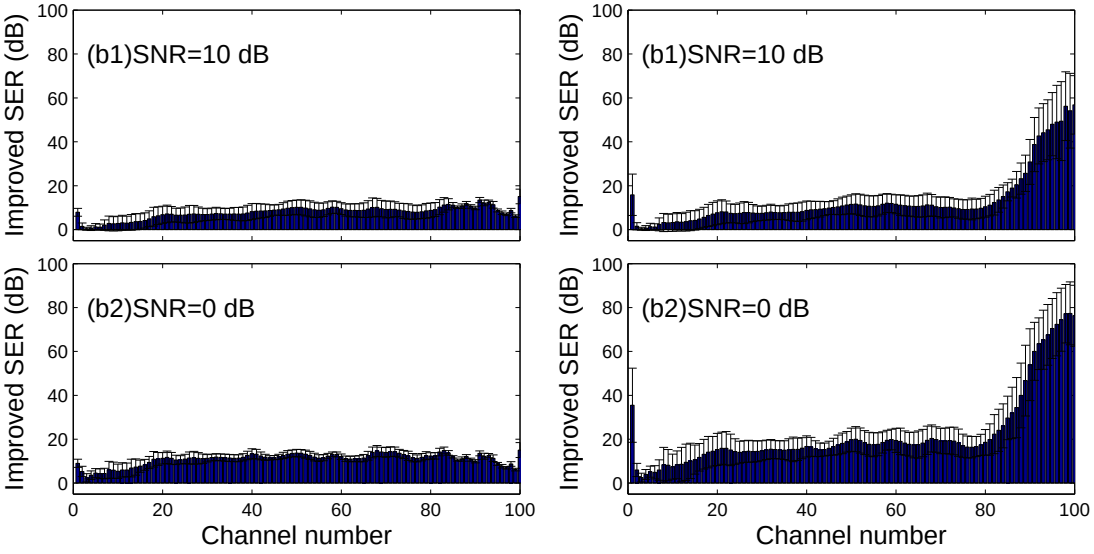

Figure 5.5: Improvement SER for pink noise in noisy environment: (left) Improvement of previous method, (right) Improvement of improved method.

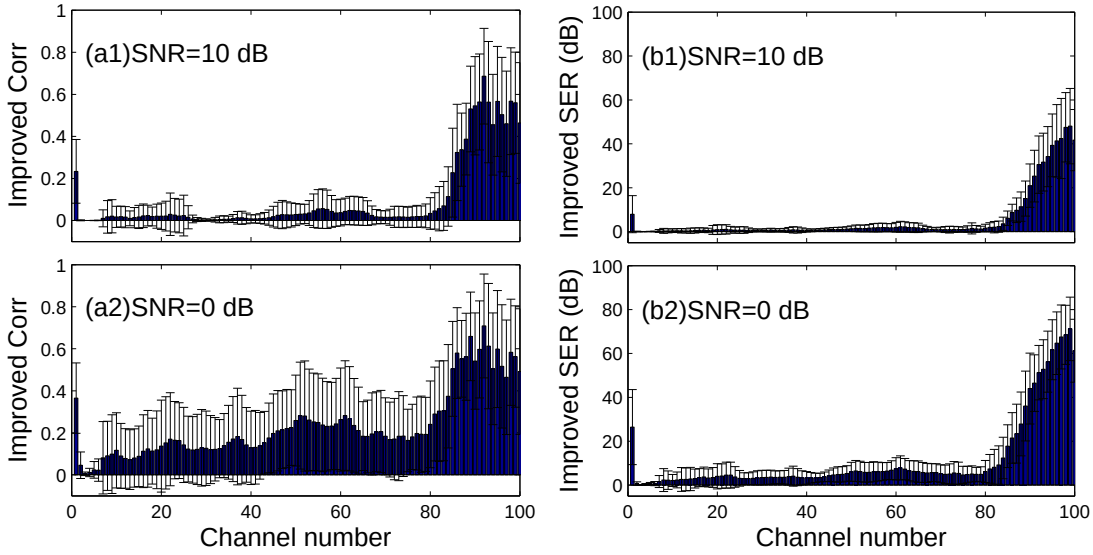

Figure 5.6: Improvement between both methods for pink noise in noisy environment: (left) Correlation improvement, (right) SER improvement.

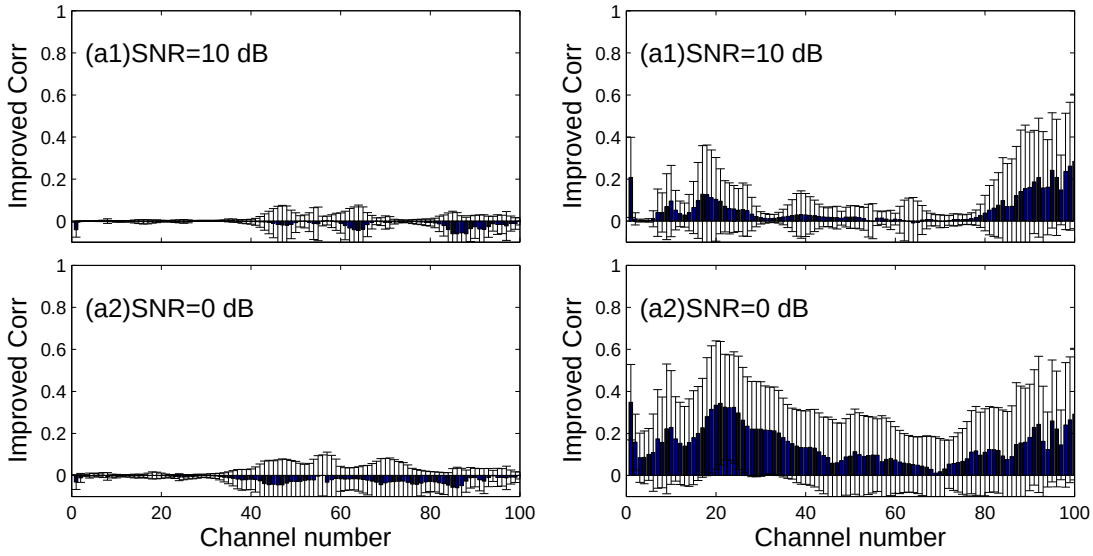

Figure 5.7: Improvement correlation for factory noise in noisy environment: (left) Improvement of previous method, (right) Improvement of improved method.

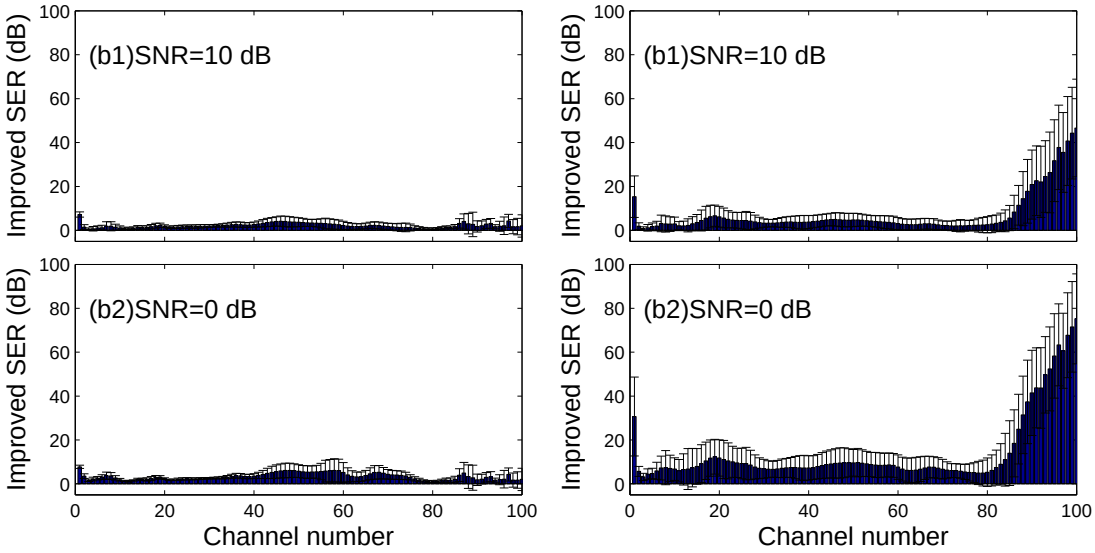

Figure 5.8: Improvement SER for factory noise in noisy environment: (left) Improvement of previous method, (right) Improvement of improved mehtod.

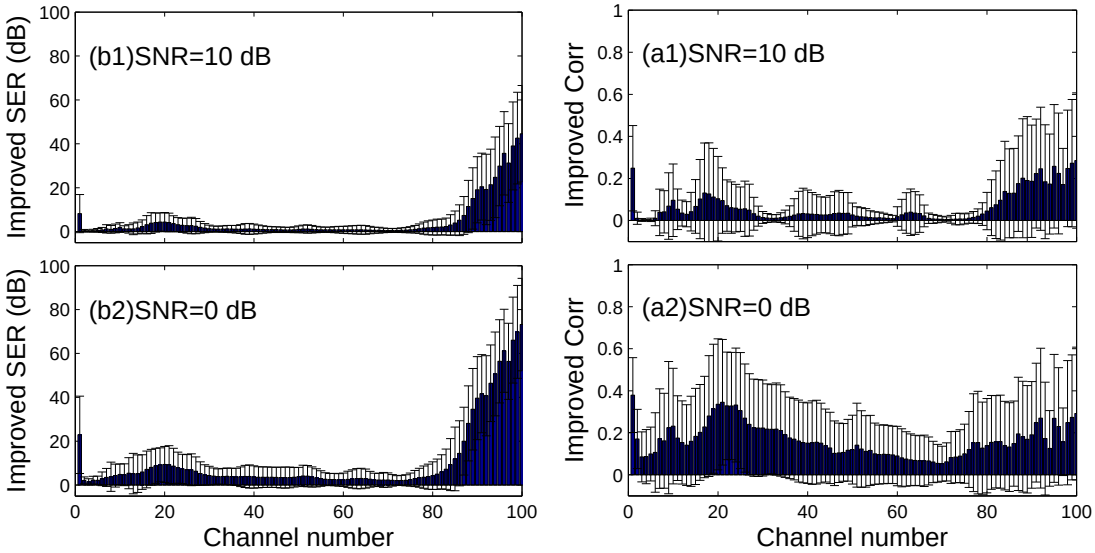

Figure 5.9: Improvement between both methods for factory noise: (left) Correlation improvement, (right) SER improvement.

In these results, the height of bar shows the mean value and the error bar shows the standard deviation. The improvement of correlation was almost zero by the previous method but the proposed method had much higher improvement. The proposed method also had a small improvement of SER in low frequency bands and a large improvement in high frequency bands. It is because when we add the noise to the clean signal, it has a characteristic that the higher frequency bands have a lower correlation and SER and the noisy power envelope have this characteristic for all the stimuli from the experiment results. The previous method can only reduce the mean value of the noise power envelope, so it almost cannot improve any correlation. The proposed method can remove the fluctuations of the noise power envelope, so the proposed method can improve much correlation, we know that the high frequency bands have low correlation and SER, so there are much higher improvements for the high frequency bands. These results showed that the proposed method could improve much correlation and SER compared to the previous method in noisy conditions.

5.4 Improvement of Corr and SER in noisy reverberant environment

Figure. 5.10 to Figure. 5.18 show the improvement of correlation and SER for the previous method and the improved method of the white noise, pink noise and factory noise in noisy reverberant environment. We can see the improved method can improve much correlation and a little SER.

In these results, the height of bar shows the mean value. These improvements showed that the proposed method can be used to get better improvement for the temporal power envelope from the noisy reverberant signals compared to the previous method. In the proposed method, we use the same dereverberation method with the previous method, so the trends of the improvement for the noisy reverberant power envelope is similar to noisy power envelope.

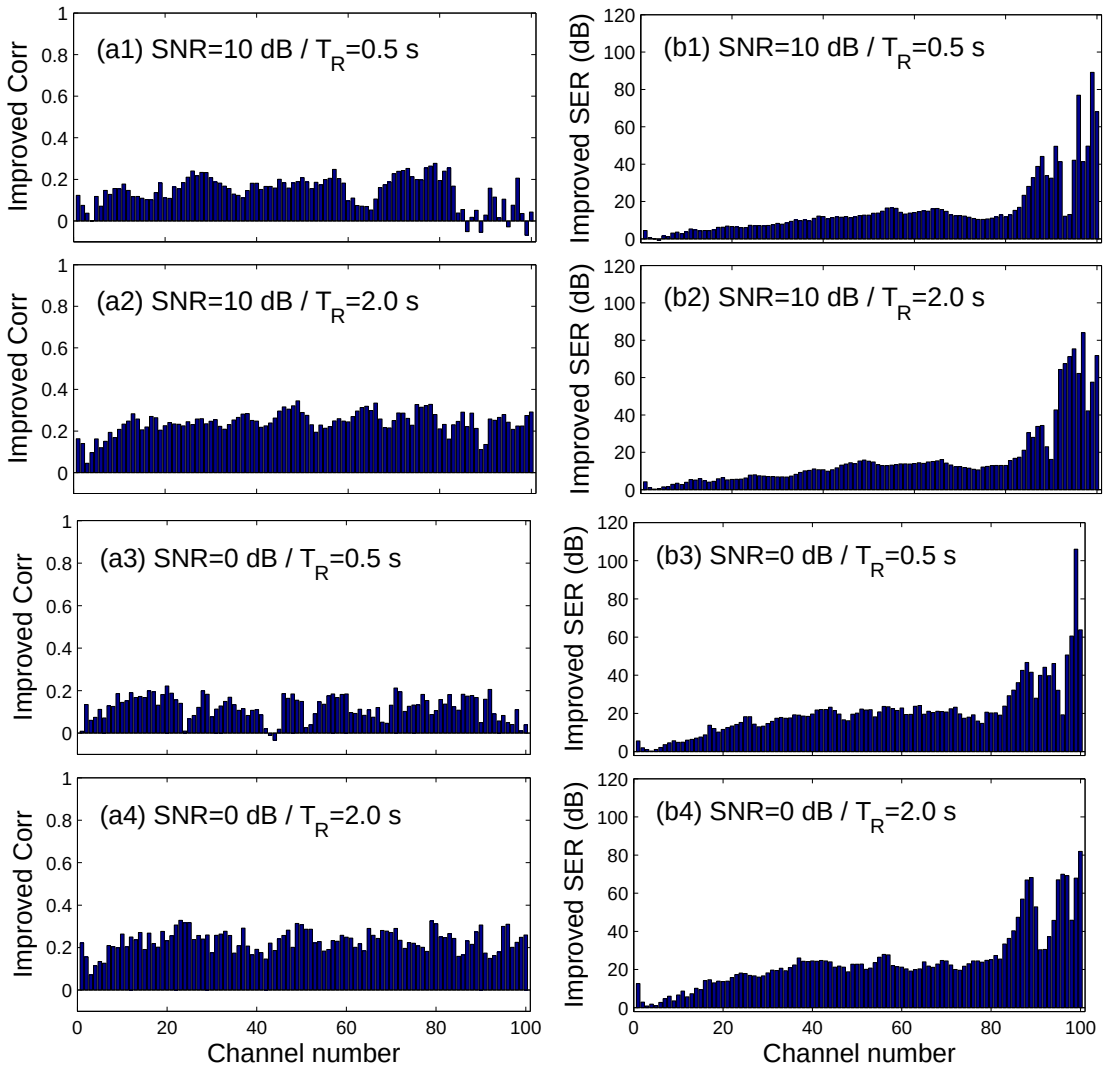

Figure 5.10: Improvement in restoration accuracy of the previous method for white noise: (a) improved Corrs and (b) improved SERs. $T_R = 0.5$ and 2.0 s. SNR = 10 and 0 dB.

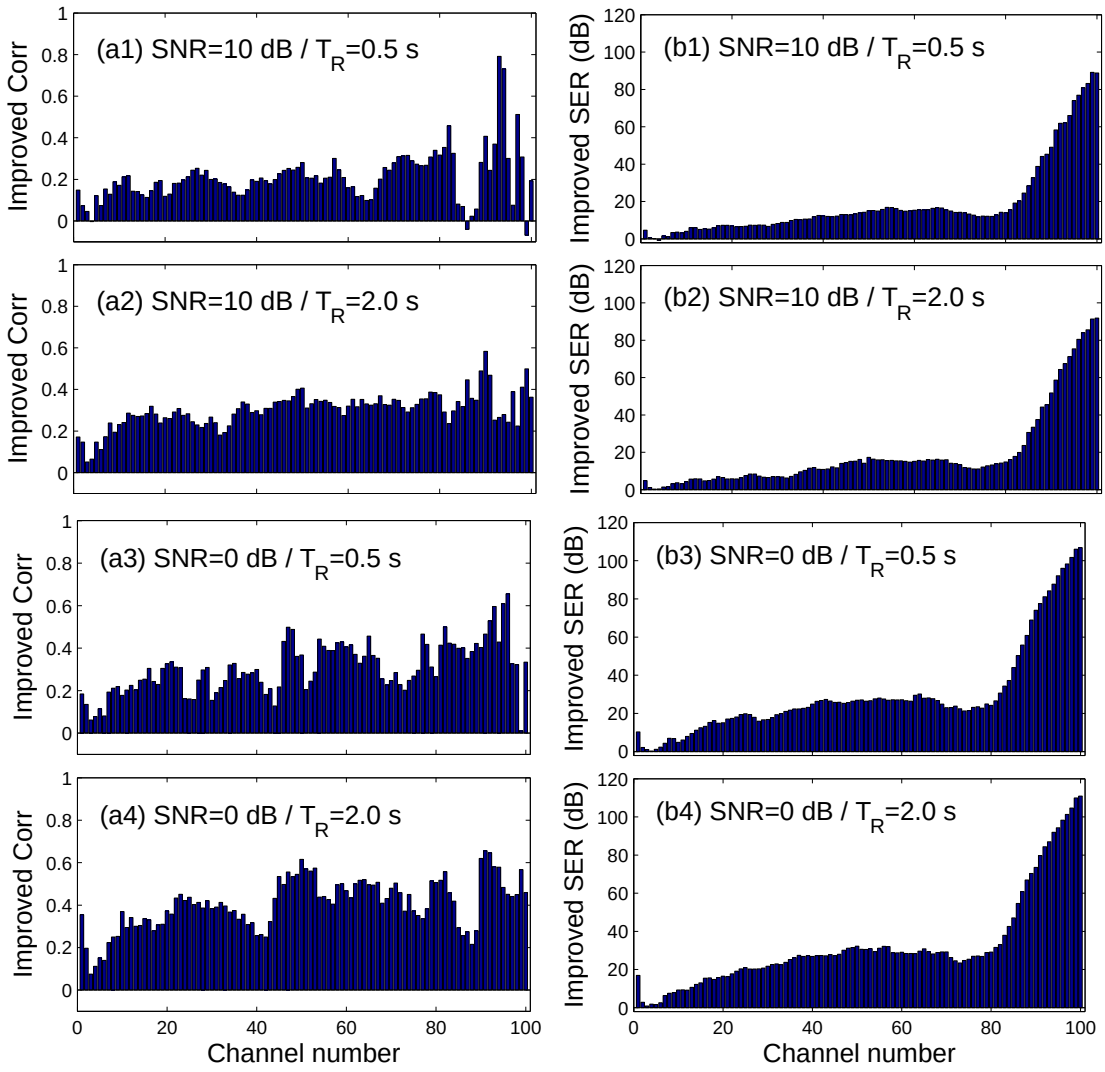

Figure 5.11: Improvement in restoration accuracy of the improved method for white noise: (a) improved Corrs and (b) improved SERs. $T_R = 0.5$ and 2.0 s. SNR = 10 and 0 dB.

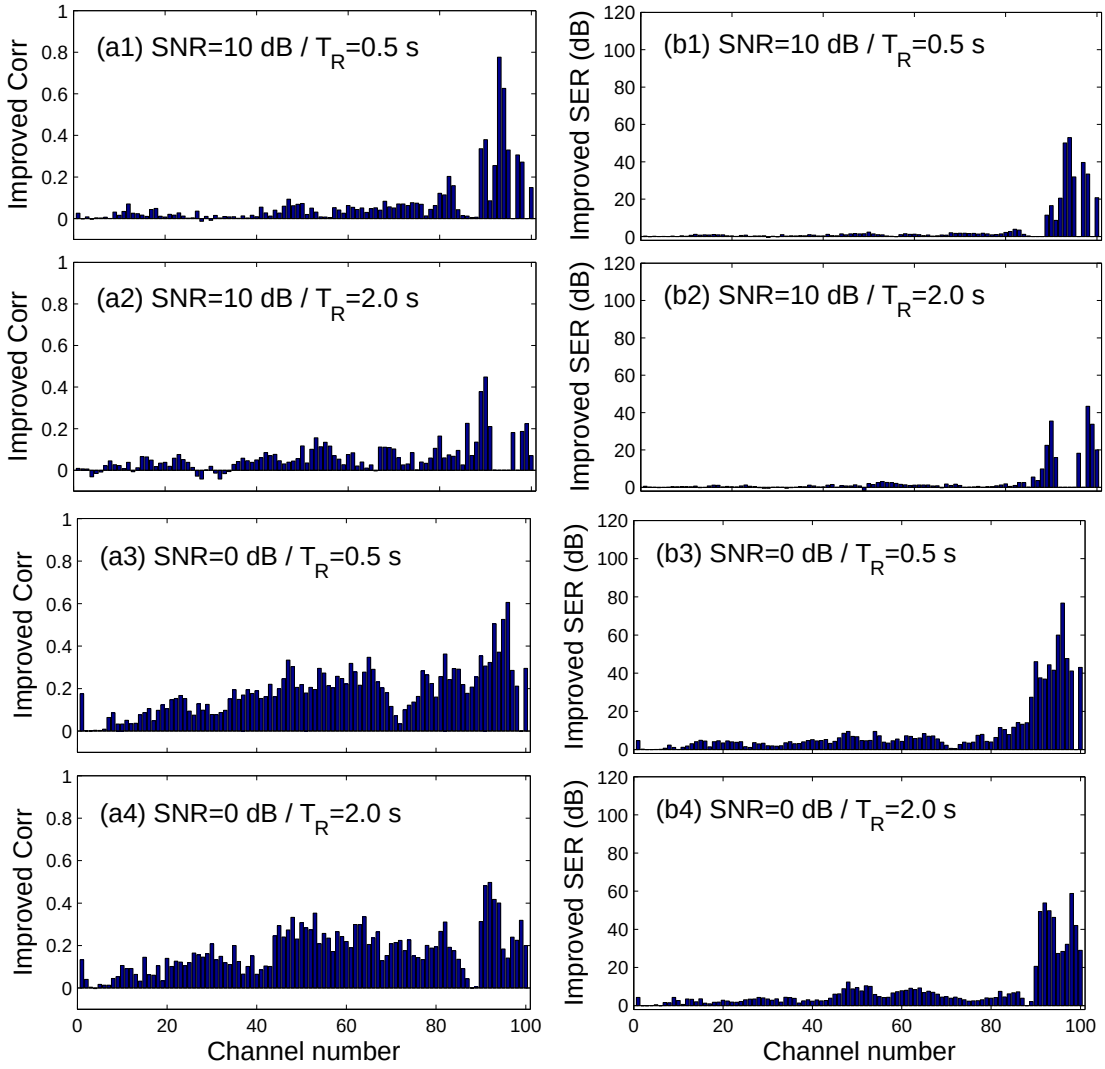

Figure 5.12: Improvement in restoration accuracy between the improved method and the previous method for white noise: (a) improved Corrs and (b) improved SERs. $T_R = 0.5$ and 2.0 s. SNR = 10 and 0 dB.

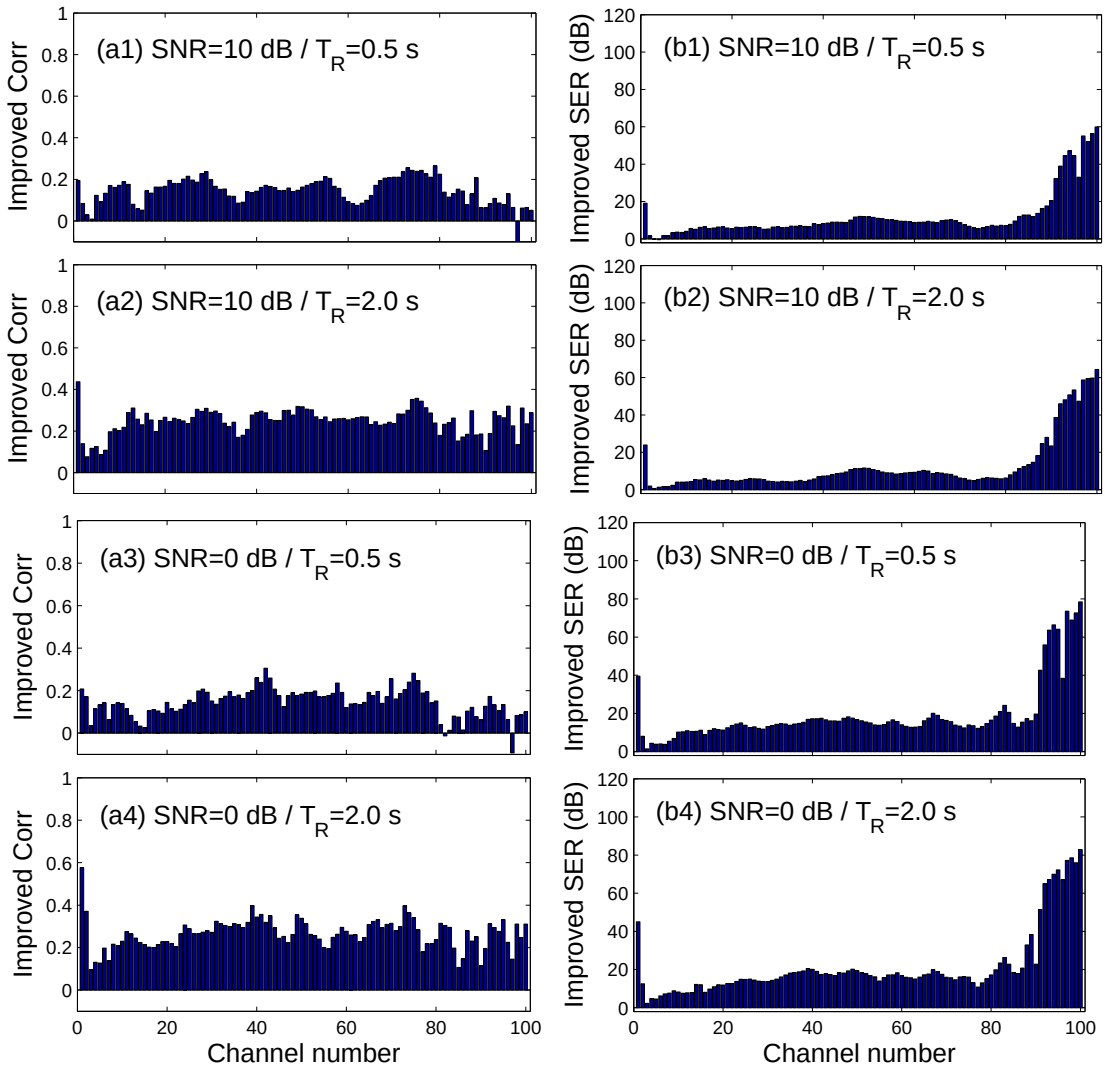

Figure 5.13: Improvement in restoration accuracy for the previous method for pink noise: (a) improved Corrs and (b) improved SERs. $T_R = 0.5$ and 2.0 s. SNR = 10 and 0 dB.

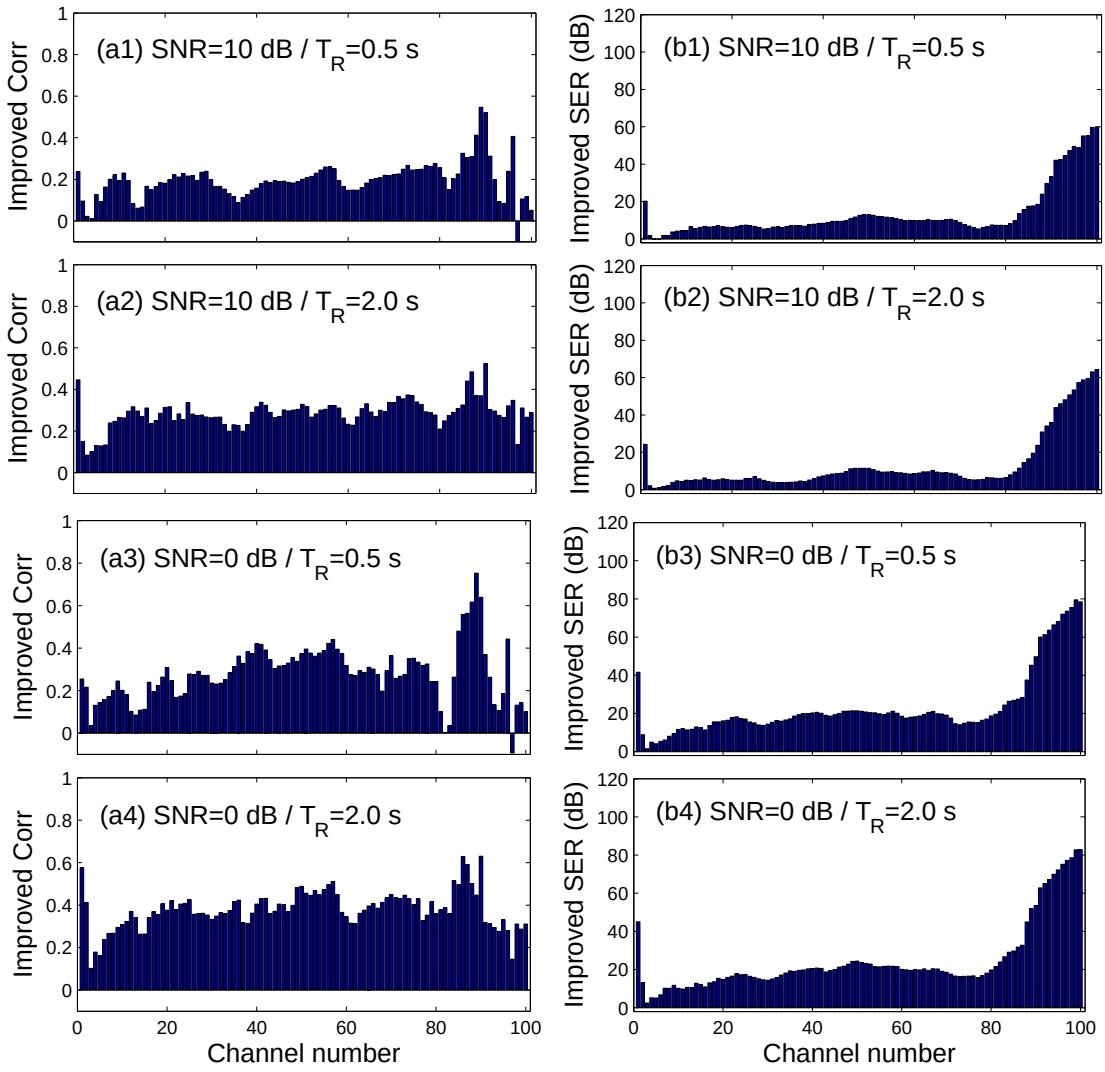

Figure 5.14: Improvement in restoration accuracy for the improved method for pink noise: (a) improved Corrs and (b) improved SERs. $T_R = 0.5$ and 2.0 s. SNR = 10 and 0 dB.

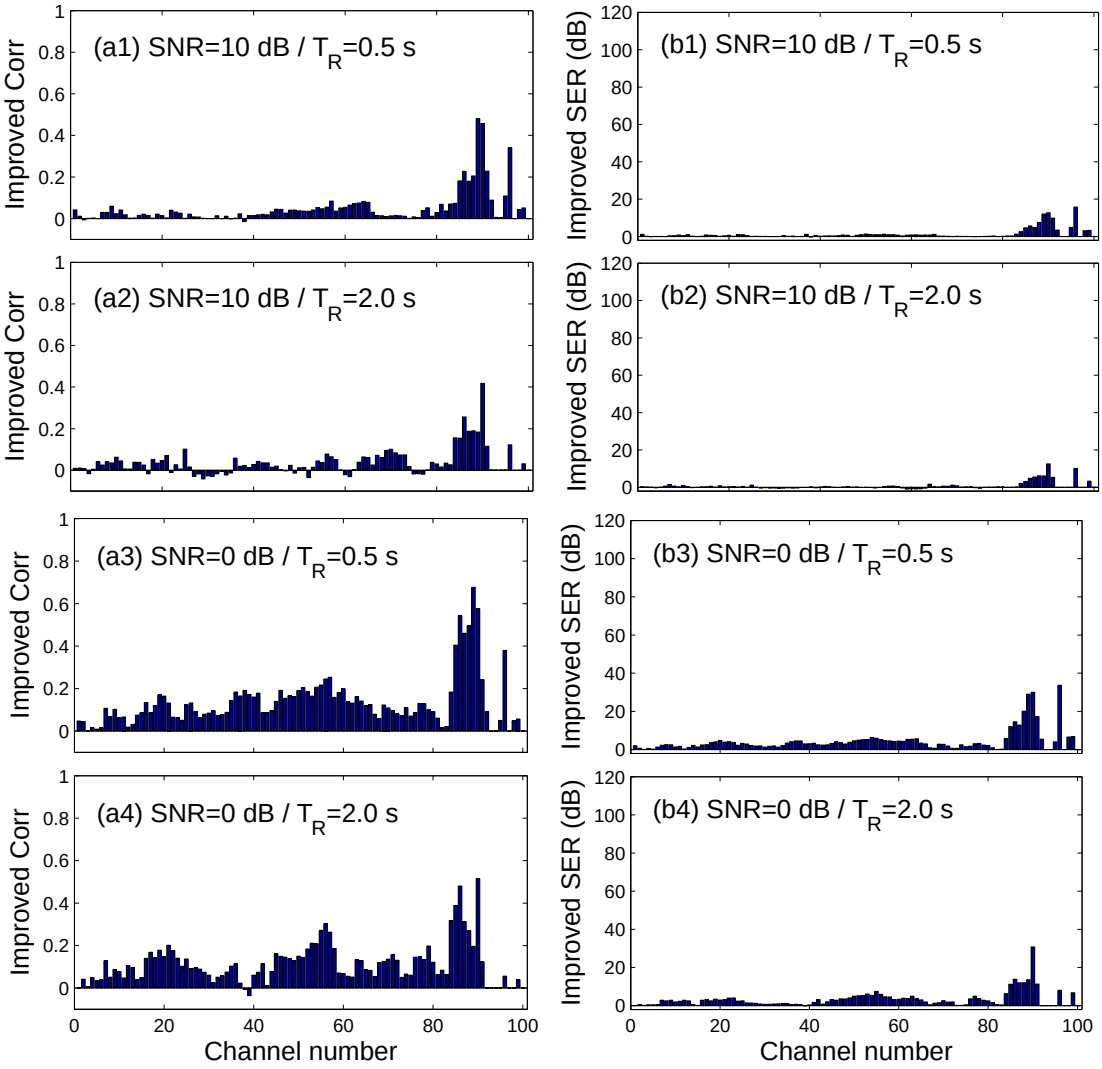

Figure 5.15: Improvement in restoration accuracy between the improved method and previous method for pink noise: (a) improved Corrs and (b) improved SERs. $T_R = 0.5$ and 2.0 s. SNR = 10 and 0 dB.

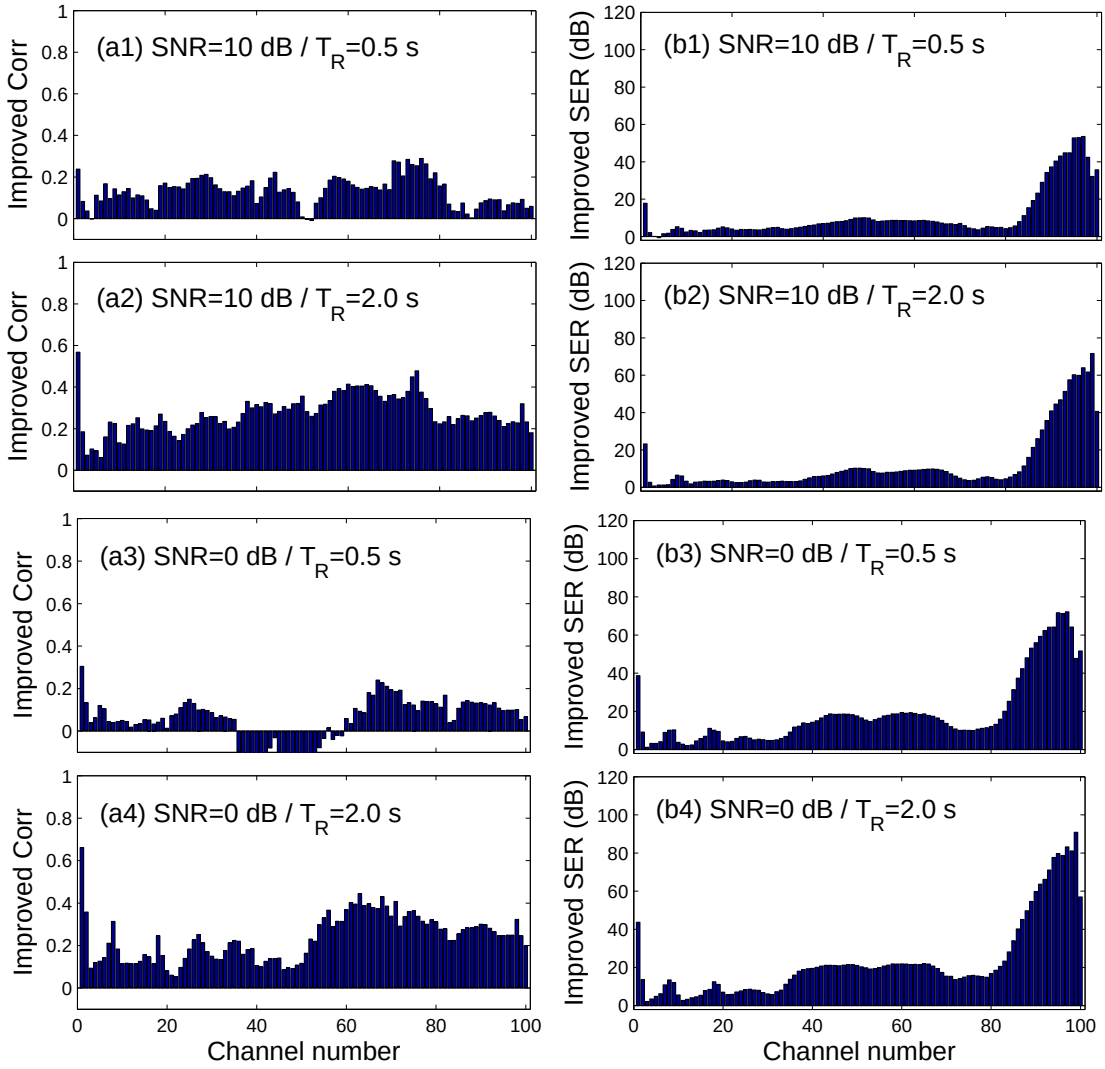

Figure 5.16: Improvement in restoration accuracy for the previous method for factory noise: (a) improved Corrs and (b) improved SERs. $T_R = 0.5$ and 2.0 s. SNR = 10 and 0 dB.

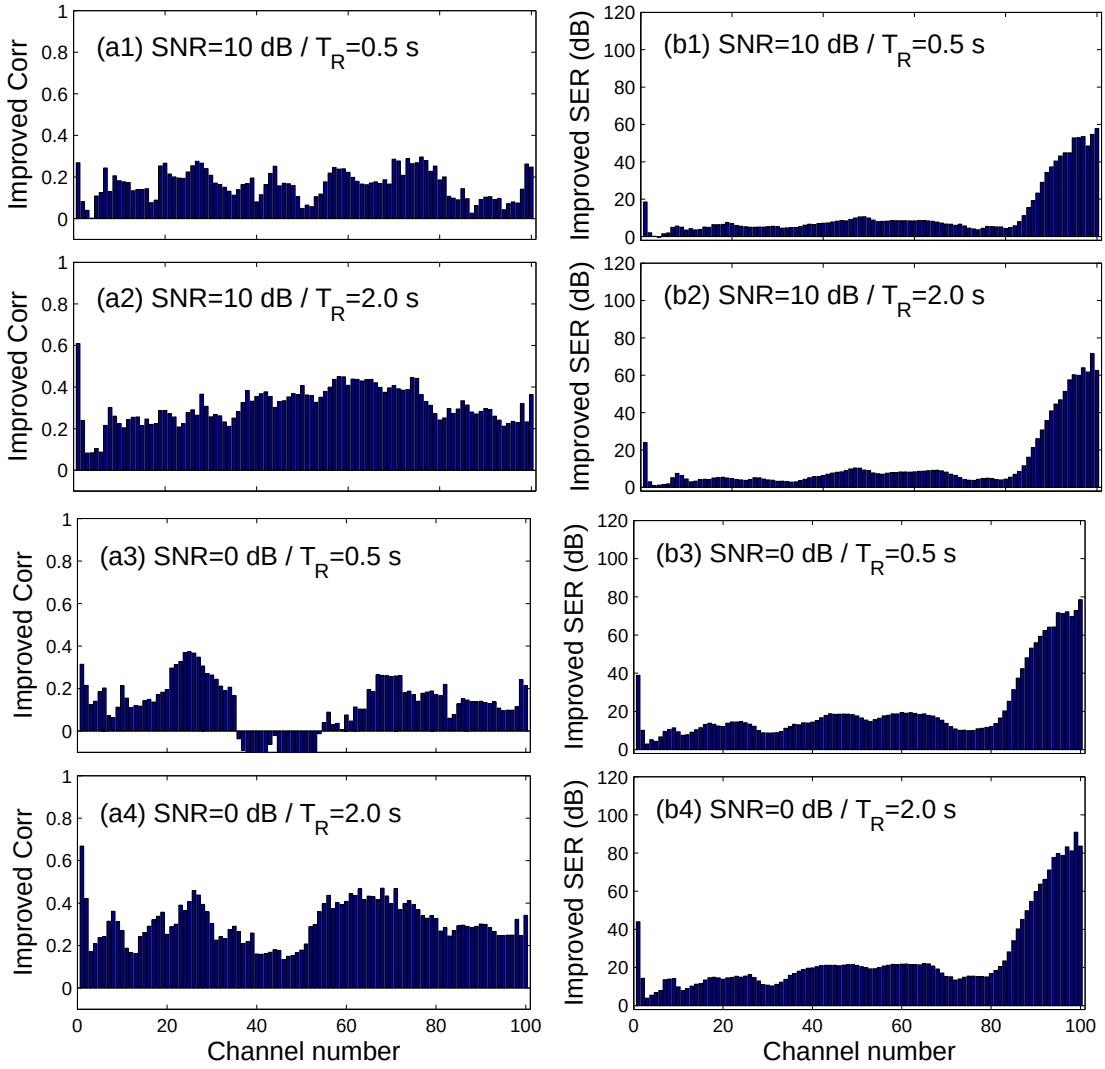

Figure 5.17: Improvement in restoration accuracy for the improved method for factory noise: (a) improved Corrs and (b) improved SERs. $T_R = 0.5$ and 2.0 s. SNR = 10 and 0 dB.

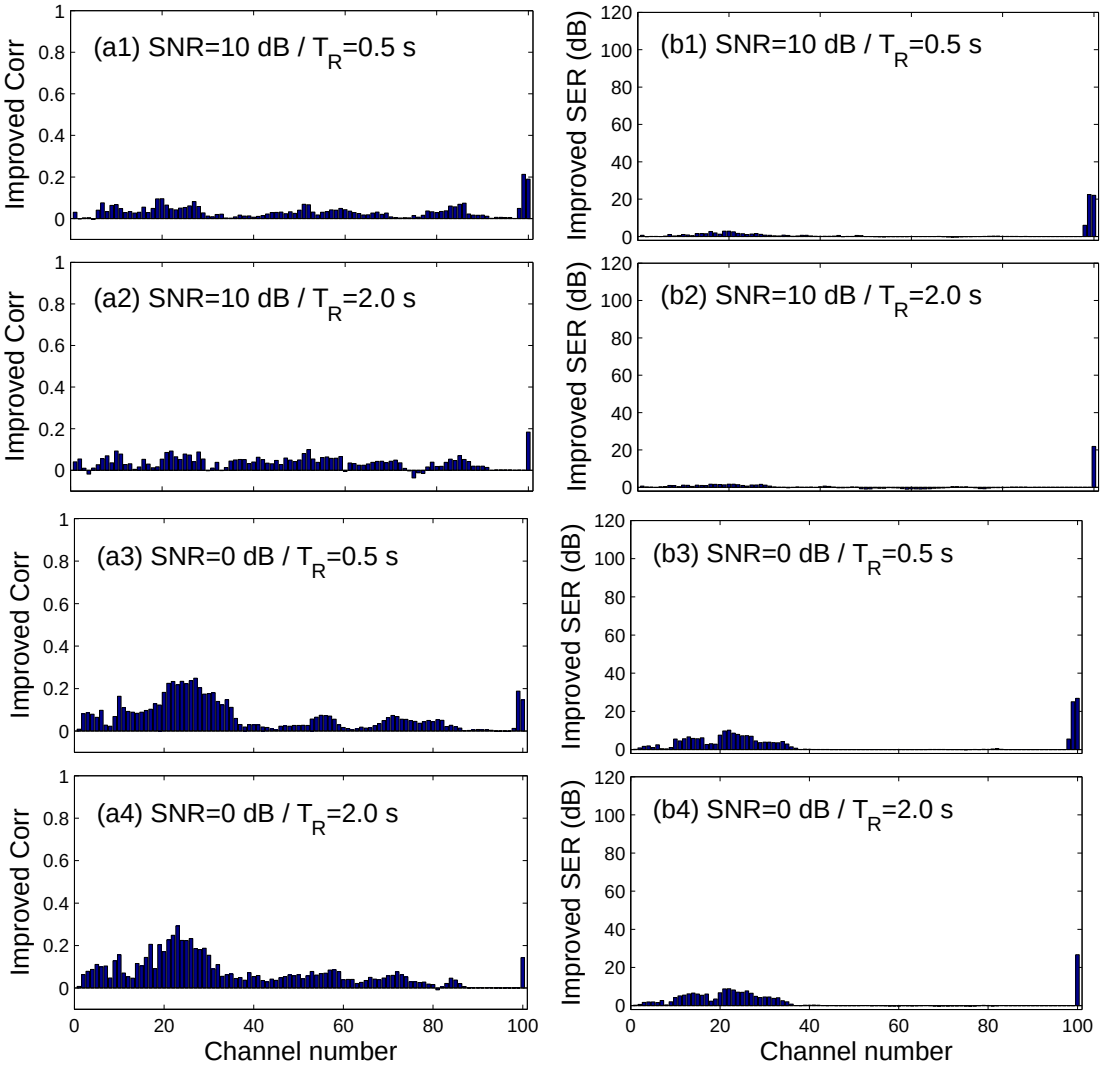

Figure 5.18: Improvement in restoration accuracy between the improved method and previous method for factory noise: (a) improved Corrs and (b) improved SERs. $T_R = 0.5$ and 2.0 s. SNR = 10 and 0 dB.

Chapter 6

Conclusion

6.1 Summary

In this thesis, we use the method based on the MTF concept to suppress the noise and reverberation simultaneously to improve the intelligibility of the speech. The method based on MTF concept has been proposed consists of two parts: power envelope subtraction process and power envelope dereverberation process. The power envelope subtraction method can only reduce the mean value of noise power envelope,in order to removing the fluctuations of the noise, we proposed a improved power envelope subtraction method using the Kalman filter based on the MTF concept and we also apply the linear prediction method to estimate the parameters for the state equation for the Kalman filter. We have carried out simulations to evaluate the improved proposed method using white noise, pink noise and factory noise in noisy environment and noisy reverberant environment. The results showed that the proposed method can achieve the goal of removing the fluctuations of the noise power envelope and then improve much more correlation and SER for stationary noise and non-stationary noise in both noisy environment and noisy reverberant environment compared to the previous MTF based method.

6.2 Future work

In our research, we use the AR model for the state equation of the Kalman filter. We calculate the parameters of AR model using clean power envelope (non-blind method) and noisy power envelope by iteration method (blind method) respectively. The parameters calculated from clean power envelope is the ideal case and the improvement of Corr and SER are relatively high, although the parameters calculated from noisy power envelope by iteration method have some improvement, they are much smaller than the ideal case. The remaining work is to consider a better blind estimation method which has the similar improvement with the non-blind method.

6.3 Contribution

The method based on MTF concept does not require the impulse response and noise conditions in the room acoustics to be measured and it can enhance the noisy reverberant speech simultaneously. We improved the power envelope subtraction process in this research and it can provide a better input for various kinds of speech processing applications, such as:speech-emphasis for transmission systems and hearing aid systems, as well as the preprocessing for speech recognition systems. All of these applications are closely related to our daily life.

Bibliography

[1] S. F. Boll, "Uppression of acoustic noise in speech using spectral subtraction," *IEEE Trans.ASSP.*, **27**(2), 113-120, 1979.

[2] K. K. Paliwal and A, Basu, "A speech enhancement method based on Kalman filtering," Proc. *ICASSP'87*, **1**, 117-180, 1987.

[3] Y. Ephraim and D. Mlah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator, "*IEEE Trans. Acoust. Speech, Signal Process., ASSP-32(6)*, 1109-1211, Dec. 1984.

[4] S. T. Neely and J. B. Allen, "Invertilibility of a room impulse response," *J. Acoust. Soc. Am*, **66**(1), 166-169, July 1979.

[5] M. Miyoshi, Y.Kaneda, "Inverse filtering of room acoustics," *IEEE Trans.ASSP.*,Vol. **36**(2), 145-152, Feb. 1988.

[6] T. Houtgast and H.J. Steeneken,"The Modulation Transfer Function in Room Acoustics as a Predictor of Speech Inteligibility", *Acustica*, **28**, 66-73, 1973.

[7] M. Unoki, Y. Yamasaki, and M. Akagi, "MTF-based power envelope restoration in noisy reverberant environments,"Proc. *EUSIPC2009, CDROM*, 2009.

[8] M. Unoki, M. Furukawa, K. Sakata and M. Akagi, "An improved method based on the MTF concept for restoring the power envelope from a reverberant signal," *Acoust. Sci. & Tech.* **25**(4), 232-242, 2004.

[9] Y. Yamasaki and M. Unoki, "Study on a Method of Suppressing Noise Based on the MTF Concept," *Journal of Signal Processing.* **13**(4), 335-338. July.

[10] K. Kinoshita, M. Delcroix, T. Nakatani, and M. Miyoshi,"Muti-step linear prediction based speech enhancement in noisy reverberant environment," *Proc,Interspeech-2007.*, 854-857,Aug.2007.

[11] M. Unoki, K. Sakata, M. Furkawa, and M. Akagi,"A speech dereverberation method based on the MTF concept in power envelope restoration," *Acoustic. Sci. & Tech.*, **25**(4), 243-254. 2004.

[12] M. Unoki,M. Toi, and M. Akagi, "Development of the MTF-based speech dereverberation method using adaptive time-frequency division," *Proc. Forum Acusiticum2005*, 51-56.Budapest,2005.

[13] X. Lu,M. Unoki, and M. Akagi,"Comparative evaluation of modulation-transfer-function-based blind restoration of sub-band power envelopes of speech as a front-end processor for automatic speech recognition systems,'*Acoust.Sci. & Tech* ., **29**(6), 351-361,2008.

[14] T. Takeda *et al.*,Speech Database User's Manual,ATR Technical Report, TR-I-0028,1988.

[15] D.Ying,Y. Shi,X. Lu,J. Dang,and F.Soong,"Robust voice activity detection based on noise eigenspace," *Acout. Sci. & Tech.*, **28**(6), 413-423. 2007.

[16] Sadaoki Furui and M. Mohan Sondi,*Advances in Speech Signal Procesing*, Marcel Dekker,Inc.,NY,1992.

[17] N. Ma,M. Bouchard and R. A. Goubran, "A Perceptual Klaman Filtering-Based Approach for Speech Enhancement," *International Symposium for Signal Processing and Applications.*, ISSPA, Jul.2003.

[18] B. Winrow *et al*,"Adaptive noise cancelling: Principles and applications," *Proc.IEEE*,vol.63,pp.1692-1716,Dec.1975.

[19] S. F. Boll and D. Pulsipher,"Noise suppression methods for robust speech processing,"Dep. Comput. Sci.,Univ.Utah,Salt Lake City, Semi-Annu. Tech. Rep., Utec-CSc-77-202, pp. 50-54. Oct. 1977.

[20] F. J Taylor,*Principles of Signals and Systems* (MacGraw-Hill,Inc., New York, 1994).

[21] A. V. Oppenheim and R .W. Schafer, *Digital Signal Processing* (Prentice-Hall, Inc., London, 1975).

[22] T .Arai, M. Pavel, H. Hermansky and C. Avendano, "Syllable intelligibility for temporally filtered LPC cepstral trajectories," *J. Acoust. Soc. Am.*, **105**, 2783-2791 (1999).

[23] N. Kanedera, T. Arai, H .Hermansky and M. Pavel,"On the importance of various modulation frequencies for speech recognition," *Proc. EuroSpeech 97*, 1079-1082 (1997).

[24] N. Kanedera, T. Arai and T. Funada, "Robust automatic speech recognition emphasizing important modulation spectrum," *IEICE Trans. D-II*, **J84-D-II**, 1261-1269 (2001).

Acknowledgements

My master course at School of Information Science in JAIST will soon come to an end, at the completion of my graduation thesis; I wish to express my sincere appreciation to all those who have offered me invaluable help during the whole period of my study and research. First and foremost, I would like to express my deepest gratitude to my supervisor, Professor Unoki, who led me into the world of Speech enhancement and gave me a chance to explore my potential in this research topic at the full capacity. He generously spent much time reading through each draft. He provided me a lot of instructive advices, useful suggestions, insightful criticism and professional guidance, without his consistent and illuminating instruction, this thesis could not be presented in its current form. Secondly, I should give my hearty thanks to Professor Akagi and Professor Dang, for their valuable comments and suggestions. They have put considerable time and effort into their comments on my research. They gave me much help and helped me work out my problems during the difficult course of my research, from whom I benefited a lot. Also, I would like to thank my friends, Haitao Zhang, Ngo Nhut Minh. They kindly gave me a hand when I was in frustration or depression. Their encouragement and unwavering support has sustained me through the hard times. Last but not the least, my thanks would go to my family for their love and great confidence for me all through these years. They have supported me continuously, which is the biggest motivity for my study. Once again, I would like to thank my supervisor Professor Unoki, his rigorous, conscientious and earnest attitude to scholarly studies impressed me most, this would be an invaluable wealth for my future.

Publications

[1] Yang Liu, Masashi Unoki, "Study on power envelope subtraction based on Modulation Transfer Function," *IEICE Technical Report*, EA2012-58, Sendai,2012.

[2] Yang Liu, Masashi Unoki, "Study on noise subtraciton method based on Modulation Transfer Function," Acoustical Society of Japan, Nagano, September, 2012.