Single Channel Noise Suppression Based on Speech and Noise Spectral Models

WEERAWUT THANHIKAM

SEPTEMBER 2012

Single Channel Noise Suppression Based on Speech and Noise Spectral Models

A dissertation submitted to
THE GRADUATE SCHOOL OF ENGINEERING SCIENCE
OSAKA UNIVERSITY
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY IN ENGINEERING

BY

WEERAWUT THANHIKAM

SEPTEMBER 2012

Abstract

The purpose of this research is to achieve single channel noise suppression based on speech and noise spectral models. This thesis consists of two main parts. The first part describes stationary noise suppression and the second part describes impulsive noise suppression.

First, a stationary noise suppression algorithm using Maximum a Posteriori (MAP) estimation with a speech spectral amplitude probability density function (speech PDF) is investigated. An estimated speech spectrum is given as a MAP solution which is obtained from the speech PDF. The speech PDF is hence the most important factor in this research. A useful speech PDF has been established and is entirely characterized by two shape parameters. As optimal shape parameters, certain fixed values have also been derived. Speech can be efficiently extracted when these parameters are properly applied so that the speech PDF fits to the real-speech PDF. However, the speech property should be considered as a time-variant function. In this case, the fixed speech PDF can not track the property change. In this research, under the assumption that the speech PDF changes according to signal to noise ratio (SNR), the author proposes adaptive shape parameters which are derived from real-speech PDFs in various narrow SNR intervals. The proposed adaptive shape parameters can track the change of the speech property, and give an appropriate MAP solution which is identical to the estimated speech spectrum. The effectiveness of the proposed method was examined and compared to conventional algorithms. The simulation results showed that the proposed method improved segmental SNR around 6 and 9 dB when the input speech signal was corrupted by white and tunnel noise signals at input SNR of 0 dB, respectively.

Second, an impulsive noise suppression method is investigated. This method utilizes a zero phase (ZP) signal which is defined as the IDFT of a spectral amplitude. In the impulsive noise suppression research, we assume that a speech signal has periodicity in a short observation, i.e., its spectral amplitude has values at equally spaced frequencies. In this case, the corresponding ZP signal becomes also periodic. This assumption is especially appropriate for a voiced speech which is mainly arisen in speech signals. On the other hand, we assume that a noise spectral amplitude is approximately flat. In this case, its ZP signal takes nonzero values only around the origin. Actually, many impulsive noise signals have such property. Under these assumptions, the ZP signal of a speech signal embedded in impulsive noise in an analysis frame becomes a periodic signal except around the origin. Hence, replacing the ZP signal around the origin with the ZP signal in the second or latter period, we get an estimated speech ZP signal. Taking DFT of it gives the estimated speech spectral amplitude. The IDFT of the estimated speech spectral amplitude with the observed spectral phase provides the estimated speech signal in time domain. The major advantage of this method is that it can suppress impulsive noise without a prior estimation of the noise spectral amplitude, while the a prior estimation of the noise is indispensable in most stationary noise suppression methods. Moreover, it is shown that the proposed impulsive noise suppressor can also be available to suppress stationary wide-band noise. Simulation results showed that the proposed noise suppressor improved the SNR more than 5dB for stationary tunnel noise and 13dB for impulsive clap noise in a low SNR environment.

Acknowledgements

First of all, I would like to express my deepest gratitude to Professor Youji Iiguni, Graduate School of Engineering Science, Osaka University for his numerous supports on this thesis. Without his invaluable suggestions and supports, it would be impossible to complete this thesis in time. Moreover, I am greatly indebted to him for not only the supervision of this thesis but also accepting me in this laboratory as an international student.

I also would like to express my sincere gratitude to Professor Masahiro Inuiguchi, and Professor Toshiyuki Ohtsuka, who are members of my dissertation committee, for their many helpful suggestions and comments for this thesis.

I would like to express my sincere gratitude to Assistant Professor Arata Kawamura, Graduate School of Engineering Science, Osaka University for his invaluable advice not only in academic but also in real life. I am so glad that I have spent five worthwhile years in researching digital speech processing with him.

I also would like to express my sincere gratitude to Professor Tadashi Itoh, and Professor Masahito Taya of Osaka University, who were examination committees for entrance to the Osaka University. Without their generous suggestions, contributions from this thesis would have never been happened.

I also would like to express my sincere thanks to Associate Professor Chalie Charoenlarpnopparut of Sirindhorn International Institute of Technology (SIIT) who is my nominal supervisor (in double supervisor system) for giving basic knowledge in the area of digital signal processing.

I also would like to thank all the members of Iiguni Laboratory, Osaka University for their support. They have provided me good experience,

knowledge, and friendship. I also would like to thank all Thai students in Toyonaka campus. They support me and make me feel like not far away from home. I would like to thank Mr. Kitiphong Tankavatanapibul, who passed away on Dec 2009. He visited me when I were in Osaka first time and gave me encouragements to further study in Osaka University as Ph.D student. Although I could not meet him in his last moment and show this thesis to him, his sincere friendship for me is still living in thesis.

Finally, I want to thank my family and my wife for all their love and support during five years I have been in Japan. I would like to dedicate this thesis to them.

Contents

List of Figures

List of Tables

1

Introduction

The continuous improvement of multimedia and communication systems has led to the widespread use of speech recording and processing devices, e.g., mobile phones, emergency telephones, and speech recognition tools. In practical situations, these devices are being used in environments where undesirable background noise exists. For example, mobile phone users have to communicate in the presence of undesirable background noise. As noise often degrades the quality of recorded speech, it is beneficial to carry out noise suppression. Also speech with background noise can cause problems for both mobile communication and speech recognition systems. For example, important conversation must be delivered correctly in emergency case. Fig.1.1 shows the example of situation that the speech signal is contaminated by background noise in mobile phone communication environments. Since general mobile phones employ a single microphone, a single channel noise suppressor is an important tool to improve the quality of speech communication. Hence, we will focus on single microphone noise suppression systems, while powerful dual or multi channel noise suppression algorithms exist [1]. Single channel noise suppression algorithms assume the existence of a single sensor (e.g., microphone) that captures the noisy speech. This type of algorithm has to estimate the background noise and enhance the speech from a single recording.

In this thesis, the author presents two efficient single channel noise suppression algorithms, individually. They effectively suppress stationary noise and impulsive noise, respectively.

First, single channel stationary noise suppression is investigated. A variety of stationary noise suppression methods have been proposed and extensively studied for

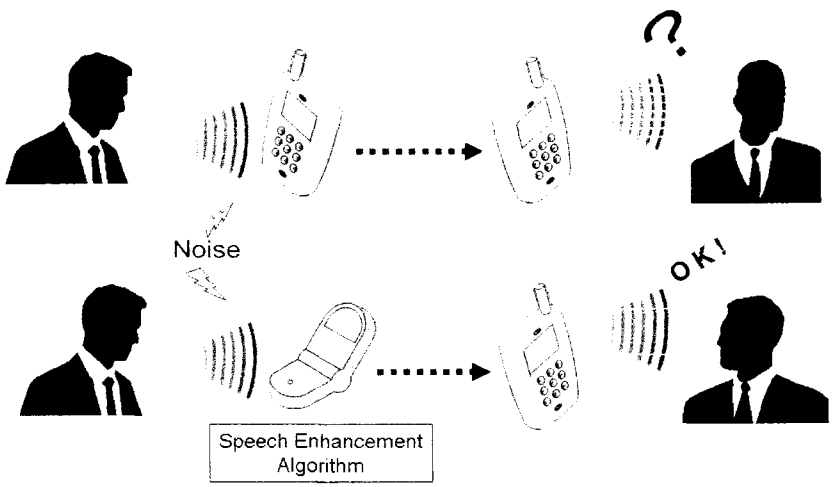


Figure 1.1: Noise Suppression in mobile phone communication.

decades [2] - [11]. In the same manner as most stationary noise suppression methods, an observed stationary noise is assumed to be Gaussian in this research. One of the well-known stationary noise suppression methods is the spectral subtraction algorithm proposed by Boll [2]. In this method, noise suppression is performed by simply subtracting an estimated noise spectral amplitude from an observed noisy speech spectral amplitude. The spectral subtraction method does not require speech spectral information. Although this method can be easily implemented, it is well known that it induces an artificial noise, called musical noise, in the enhanced speech. As another method that can enhance noisy speech with less residual musical noise, Ephraim and Malah have proposed the minimum mean square error short time spectral amplitude (MMSE-STSA) method [11], which utilizes a speech spectral amplitude probability density function (speech PDF). In the literature [11], the speech PDF was modeled by a Rayleigh density function. However, some researchers pointed out that the Rayleigh density function does not fit to a real speech PDF. A more efficient method that employs a maximum *a posteriori* (MAP) estimator has been proposed by Lotter and Vary [3]. In [3], the speech PDF has been modeled by a parametric super Gaussian function, controlled by two parameters. The parametric super Gaussian function has been developed from a histogram made from a large amount of real speech data in a single narrow signal to noise ratio (SNR) interval. However, the residual noise is still persistently perceived. Andrianakis and White [4] were aware that the speech PDF may

change in some SNR intervals. They utilized three histograms made from speech signals in three different narrow SNR intervals and approximate them with Gamma density function. As reported in [4], changing these three speech PDFs according to the SNR can improve the noise suppression capability. As a similar scheme to [4], an adaptive PDF method has been proposed by Tsukamoto et al. [5]. It is based on the assumption that the speech PDF continuously changes its shape according to the SNR. They employed the parametric super Gaussian function used in [3] and adaptively changed its shape parameters according to the SNR. Two histograms were used to make the adaptive parameter function and implement it in a noise suppressor. It resulted in a better noise suppression especially during non-speech segments. However, the shape of the speech PDF in a speech segment may be incorrectly estimated, because the shape parameters are determined from only two histograms which were made from speech signals in high and low SNRs, respectively. Specifically, the adaptive shape parameters simply connect such extreme speech PDFs without proper verification.

In this research, under the assumption that two speech histograms are not enough for estimating the shape parameters of the speech PDF, the conventional approach in [5] is sophisticated by evaluating many speech histograms, and a more efficient stationary noise suppression algorithm is derived. Firstly, histograms are made from the real-speech data in various narrow SNR intervals, and the fittings of the histograms are performed with the parametric speech PDF used in [3], [5]. Secondly, shape parameter functions are derived to mitigate fluctuations of experimental results. Finally, a noise suppression algorithm with the shape parameter functions are derived. Simulation results show that the proposed noise suppression algorithm can improve the enhanced speech quality, in both speech and non-speech segments.

As the second part of this thesis, a single channel impulsive noise suppressor is investigated. Examples of stationary and impulsive noise signals depicted in Fig. 1.2, where Fig. 1.2(a) shows a female speech signal corrupted with stationary white noise, and Fig. 1.2(b) shows one corrupted with non-stationary impulsive noise. They were sampled at 8kHz. All the above mentioned noise suppressors can effectively suppress the white noise, but cannot suppress the impulsive noise, because *a prior* information of impulsive noise can not be utilized. Hence, to suppress the impulsive noise, it must be established a noise suppressor which does not require *a prior* information of noise. Kamamori et al. proposed an impulsive noise suppressor based on a zero

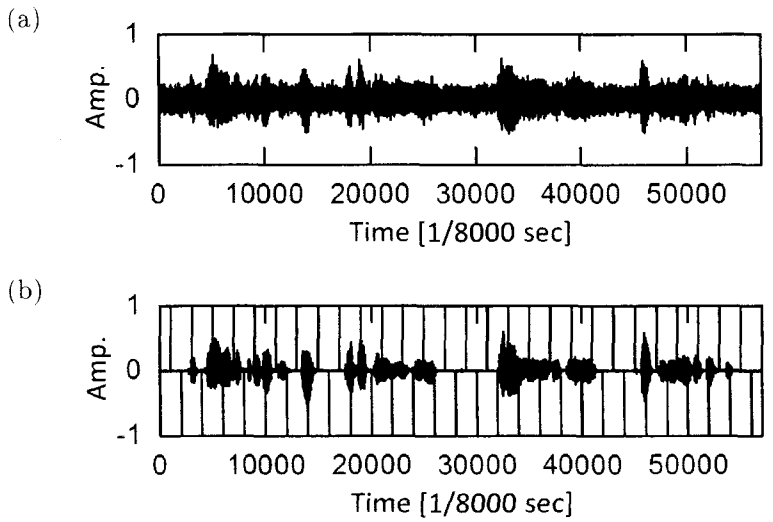


Figure 1.2: Examples of speech signal mixed with wide-band noise: (a) speech corrupted with white noise, (b) speech corrupted with impulsive noise.

phase (ZP) signal which is defined as the IDFT of a spectral amplitude [18]. The ZP signal becomes an impulse signal when the spectral amplitude is flat, and the ZP signal becomes a periodic signal when the spectral amplitude has values only at equally spaced frequencies. They assumed that a speech signal is periodic, i.e., its spectral amplitude has values only at equally spaced frequencies. As shown in [18], white noise and impulsive noise can be reduced by processing the ZP signal only at the origin. However, this method is not applicable for other impulsive noises.

To suppress real impulsive-type noise which has a duration that is normally more than one sample long, the author extends the concept of method in [18]. Assuming that a noise spectral amplitude is approximately flat, and a speech signal is periodic in a short observation, we can suppress the noise by replacing the noisy ZP signal around the origin with the ZP signal in the second or latter period. After this replacement, taking the DFT of the ZP signal gives the estimated speech spectral amplitude. The IDFT of the estimated speech spectral amplitude with the observed spectral phase provides the estimated speech signal in time domain. Unlike the method in [18], in the replacement technique, it has to be investigated about appropriate samples of the ZP signal used for replacement. In addition, a scaling function is introduced in this technique for

compensating a decay of ZP signal, where the decay is caused by segmenting and windowing an observed signal. Simulation results show that the proposed method is effective to suppress such impulsive noise signals.

The outline of this thesis is as follows: in Chapter 2, a common single channel noise suppression system is described and conventional stationary and impulsive noise suppression methods are reviewed. In Chapter 3, the proposed stationary noise suppressor is explained in detail. Additionally, some simulation results are carried out to confirm the effectiveness of the proposed method. In Chapter 4, the proposed impulsive noise suppressor is described and its noise suppression capability is evaluated. Chapter 5 concludes this thesis.

2

Conventional Single Channel Noise Suppressors

2.1 General Noise Suppression System

In this section, we present an overview of a general noise suppression system. As we will see in Sections 3 and 4, this system is a foundation of the proposed noise suppression methods.

The general noise suppression system is shown in Fig. 2.1, where $x(n)$ is an observed noisy signal at time n, and it consists of a clean speech $s(n)$ and an additional noise $d(n)$ given as

$$x(n) = s(n) + d(n). \tag{2.1}$$

The noisy signal $x(n)$ is transformed into frequency domain by segmentation and windowing with a window function $h(n)$, e.g., Hanning window. The DFT coefficient of the noisy signal at frame l and frequency bin k is calculated with

$$X_l(k) = \sum_{n=0}^{N-1} x(lQ+n)h(n)e^{-j2\pi nk/N}, \tag{2.2}$$

where N denotes the DFT frame size. The window is shifted by Q samples for the computation of the next DFT. The DFT coefficient $X_l(k)$ also consists of speech and noise parts, as given by

$$X_l(k) = S_l(k) + D_l(k), \tag{2.3}$$

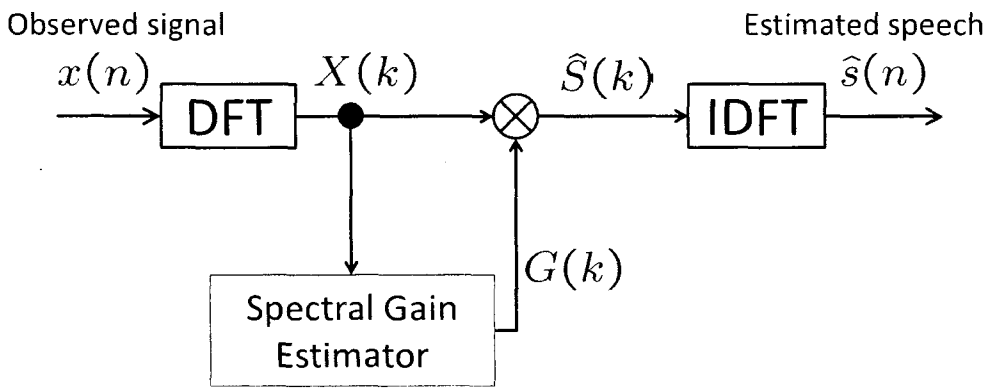


Figure 2.1: General spectral noise suppression system.

where $S_l(k)$ and $D_l(k)$ represent the DFT coefficients obtained from $s(n)$ and $d(n)$, respectively. As shown in Fig. 2.1, the noise suppressor calculates a speech spectral gain $G_l(k)$. Various definitions of $G_l(k)$ have been proposed for suppressing stationary noise [11]–[17], e.g., the spectral subtraction's spectral gain is $G_l(k) = 1 - |\hat{D}(k)|/|X_l(k)|$, where $|\hat{D}(k)|$ and $|X_l(k)|$ denotes *a priori* estimated noise spectral amplitude and noisy speech spectral amplitude, respectively. On the other hand, to achieve impulsive noise suppression, we have to estimate the speech spectral gain $G_l(k)$ without *a priori* estimation of noise spectral amplitude. One of the solutions is obtained by mapping the observed signal into the ZP domain as we will see in the latter section. After calculating $G_l(k)$, the enhanced speech spectrum $\hat{S}_l(k)$ is given by

$$\hat{S}_l(k) = G_l(k)X_l(k). \tag{2.4}$$

Finally, we obtain the enhanced speech $\hat{s}(n)$ in time domain by taking the IDFT of $\hat{S}_l(k)$ and overlap-add.

2.2 Stationary Noise Suppression Based on MAP Estimation

In stationary noise suppression, most conventional systems require to estimate a noise variance $\lambda_l(k) = E[|D_l(k)|^2]$, where $|\cdot|$ denotes the spectral amplitude, and $E[\cdot]$ is an expectation operator. In addition to $\lambda_l(k)$, *a priori* SNR $\xi_l(k)$ and *a posteriori* SNR $\gamma_l(k)$ for each DFT bin k are also required, where they are defined as

$$\xi_l(k) = \frac{E[|S_l(k)|^2]}{\lambda_l(k)} \ , \gamma_l(k) = \frac{|X_l(k)|^2}{\lambda_l(k)}. \tag{2.5}$$

By using these two SNRs, most speech spectral gains can be represented, e.g., [11]-[17]. For example, the MMSE-STSA method, its solution is completely characterized by $\lambda_l(k)$, $\xi_l(k)$ and $\gamma_l(k)$ (Appendix A.1). Whereas the *a posteriori* SNR $\gamma_l(k)$ defined in Eq.(2.5) can directly be computed, the *a priori* SNR $\xi_l(k)$ have to be estimated, because $\xi_l(k)$ is given as an expected value. The *a priori* SNR estimator $\hat{\xi}_l(k)$ of $\xi_l(k)$ has been proposed by Ephraim and Malah [11]. This estimation method is called as "decision-directed method", and it is represented as

$$\hat{\xi}_l(k) \ = \ \alpha_{snr}\hat{\xi}_{l-1}(k) + (1 - \alpha_{snr})F[\gamma_l(k) - 1], \tag{2.6}$$

where α_{snr} is a forgetting factor and

$$F[y] = \begin{cases} y; & y > 0 \\ 0; & \text{else.} \end{cases}$$

The *a priori* SNR has a high impact on the amount of noise suppression. It is useful to adjust a lower limit ξ_{thr} the *a priori* SNR according to

$$\hat{\xi}_l(k) \ = \ \begin{cases} \hat{\xi}_l(k), & \hat{\xi}_l(k) > \xi_{thr} \\ \xi_{thr}, & \text{otherwise.} \end{cases} \tag{2.7}$$

A general single channel stationary noise suppressor is shown in Fig. 2.2, where it includes some detail parts to be required for calculating the speech spectral gain. Here, we explain how to obtain the speech spectral gain $G_l(k)$ by using the MAP estimation [15]. We here omit the subscripts, the frame index l and the frequency index k, for simplicity. Let $p(|S|)$ and $p(\angle S)$ denote the probability density functions (PDFs) of the speech spectral amplitude and the phase, respectively. Here, $\angle\{\cdot\}$ denotes the spectral phase, $p(X)$ denotes the PDF of the input DFT coefficient, and $p(|S|, \angle S|X)$

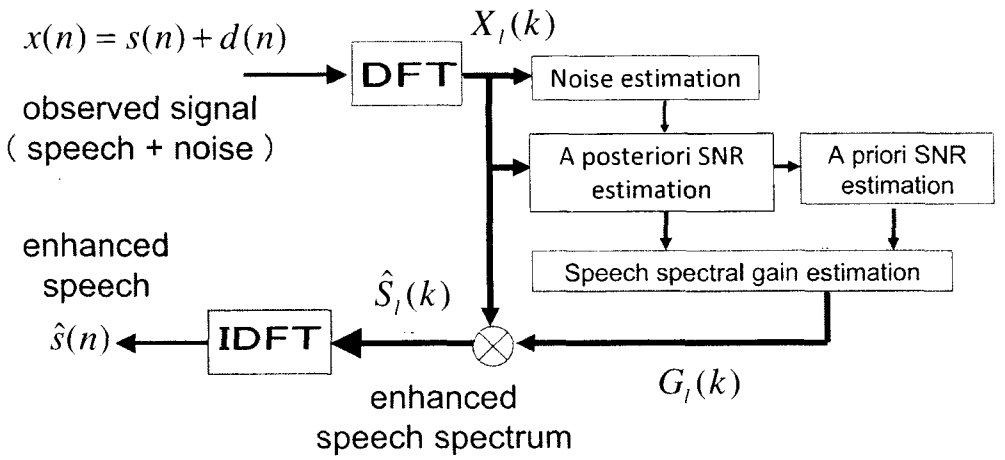


Figure 2.2: Overview of single channel stationary noise suppression system.

is the conditional speech PDF. To obtain a MAP estimate, we maximize the conditional speech PDF given by

$$p(|S|, \angle S|X) \propto p(X|\,|S|, \angle S)p(|S|, \angle S). \tag{2.8}$$

The MAP estimator gives the speech spectral amplitude $|\hat{S}|$ that maximizes $p(|S|, \angle S|X)$ represented as

$$|\hat{S}| = \arg\max_{|S|} p(X|\,|S|, \angle S)p(|S|, \angle S). \tag{2.9}$$

Note that we need to maximize only $p(X|\,|S|, \angle S)p(|S|, \angle S)$, since $p(X)$ is independent of $|S|$. Here, it is assumed that $p(X|\,|S|, \angle S)$ is Gaussian given as [3]

$$p(X|\,|S|, \angle S) = \frac{1}{\pi\lambda}\exp\left\{-\frac{|X-S|^2}{\lambda}\right\}, \tag{2.10}$$

and that $p(|S|)$ and $p(\angle S)$ are statistically independent. Moreover, $p(|S|)$ and $p(\angle S)$ are assumed to be

$$p(|S|) = \frac{\mu^{\nu+1}}{\Gamma(\nu+1)}\frac{|S|^{\nu}}{\sigma_S^{\nu+1}}\exp\left(-\mu\frac{|S|}{\sigma_s}\right), \tag{2.11}$$

$$p(\angle S) = \frac{1}{2\pi}, \tag{2.12}$$

where $\Gamma(\cdot)$ denotes Gamma function and, σ_S^2 is the variance of the speech spectrum. The PDF $p(|S|)$ shown in Eq. (2.11) has been proposed by Lotter and Vary (Appendix A.2), and is completely characterized by positive parameters μ and ν [3]. Substituting Eqs.(2.11) and (2.12) into Eq.(2.9), and solving it for $|S|$, we have (see Appendix A.3)

$$|\hat{S}| = \tilde{G} \cdot |X| \tag{2.13}$$

with

$$\tilde{G} = u + \sqrt{u^2 + \frac{\nu}{2\gamma}}, \tag{2.14}$$

$$u = \frac{1}{2} - \frac{\mu}{4\sqrt{\gamma\xi}}. \tag{2.15}$$

Note that another MAP solution has also been derived under the same speech and noise models in [3] (Appendix A.4). The parameter values are recommended by Lotter and Vary as fixed at $\mu = 1.74$ and $\nu = 0.126$ [3]. When using the MAP method, we have to properly determine the parameters μ and ν in Eq.(2.11). The recommended parameters are derived by using a large amount of signals in speech segments as reported in [3].

However, the derived speech PDF is also used for noise suppression in non-speech segments. When the estimated speech PDF does not agree with the actual one, it results in low quality of the enhanced speech. It can be expected that the speech PDF in non-speech segments is different from the one in speech segments.

Tsukamoto *et al.* [5] have developed a method of adjusting the parameters μ and ν in the parametric speech PDF according to whether the input signal is in a speech segment or in a non-speech segment. Fig.2.3 shows the parametric speech PDF with $\mu = 3.2$ and $\sigma_s = 1$ for different values of $\nu = 0$, 1, 2. This figure shows that the peak of $p(S)$ gets close to 0 as ν approaches to 0. When $\nu = 0$, $p(S)$ is equal to an exponential distribution defined as

$$p(S) = \frac{\mu}{\sigma_S} \exp\left(-\mu \frac{S}{\sigma_S}\right). \tag{2.16}$$

On the other hand, as ν gets larger, the peak goes apart from 0. When $\nu = 2$, it is very close to Rayleigh distribution which is defined as

$$p(S) = \frac{2S}{\sigma_S^2} \exp\left(-\frac{S^2}{\sigma_S^2}\right). \tag{2.17}$$

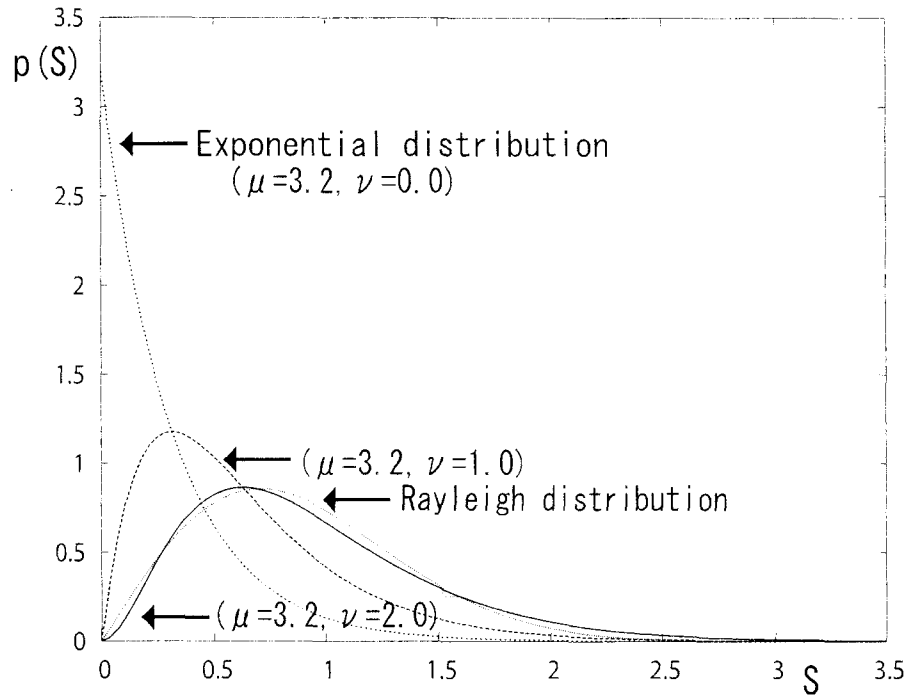


Figure 2.3: Curves of the parametric PDF.

As reported in [5], the actual speech PDF in speech segments can be approximated by Rayleigh distribution as shown in Fig. 2.4, i.e., the parametric PDF with $\mu = 3.2$ and $\nu = 2$. While in non-speech segments, the actual PDF is explicitly expressed as a Delta function, because it does not include speech component. Tsukamoto approximated the Delta function with the exponential distribution, i.e., the parametric speech PDF with $\mu = 3.2$ and $\nu = 0$. A simple adaptive method to change ν has been derived in [5]. It just smoothly changes ν value from 0 to 2 according to SNR. For adaptively changing ν, Tsukamoto utilized the input power to the noise power ratio given as

$$R(l) = \frac{\sum_{k=0}^{N-1} |X_l(k)|^2}{\sum_{k=0}^{N-1} \lambda_l(k)}. \tag{2.18}$$

where it becomes large in a speech segment and small in a non-speech segment. The

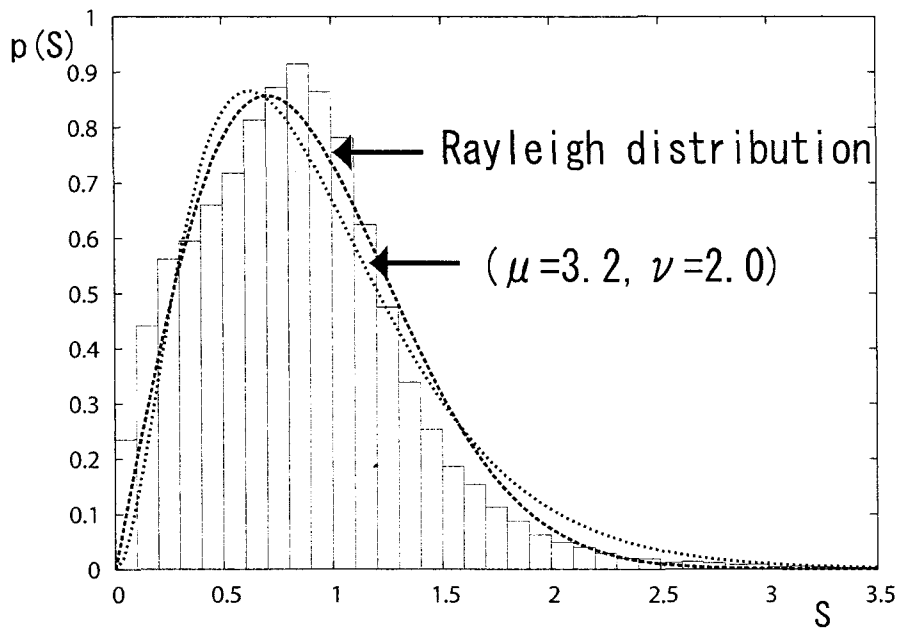


Figure 2.4: Actual speech PDF.

adaptive parameter ν_l is given as

$$\tilde{\nu}_l = 0.1 \cdot 10 \log_{10} R(l), \tag{2.19}$$

$$\nu_l = \begin{cases} 2, & \tilde{\nu}_l > 2, \\ \tilde{\nu}_l, & 0 \leq \tilde{\nu}_l \leq 2, \\ 0, & \tilde{\nu}_l < 0, \end{cases} \tag{2.20}$$

where Eq. (2.19) has been derived in an empirical manner and the parameter ν_l is restricted to the range $[0, 2]$. In non-speech segments, ν_l approaches to 0, i.e., the speech PDF approaches to the Delta function. In speech segments, ν_l gets close to 2, i.e., the speech PDF approaches to Rayleigh distribution. It results in a better noise suppression especially during non-speech segments as reported in [5]. However, the shape of the speech PDF in a speech segment may be incorrectly estimated, because the shape parameters are determined from only two histograms which are made from speech signals in high and low SNRs, respectively. Specifically, the adaptive shape parameters simply connect such extreme speech PDFs without proper verification. In Chapter 3, the approach in [5] will be sophisticated by evaluating many speech histograms, under

the assumption that two speech histograms are not enough for estimating the shape parameters of the speech PDF.

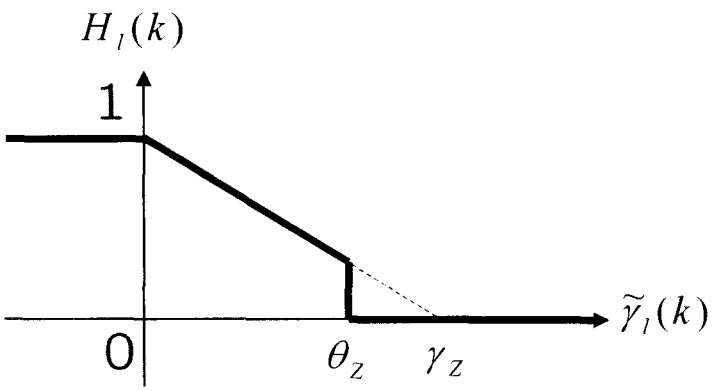


Figure 2.5: Weighting function

We should note that noise estimation also plays an important role in stationary noise suppression systems. As shown in Fig. 2.2, the SNR estimation blocks calculate *a priori* SNR $\xi_l(k)$ and *a posteriori* SNR $\gamma_l(k)$ for each DFT bin k. The SNR calculation needs an estimation of the noise variance $\lambda_l(k)$. A useful noise estimator can significantly improve the noise suppression capability. One of the most useful noise estimators is the weighted noise estimator [6] which exhibits better performance than the methods based on minimum statistics [8], [7]. Since the stationary noise suppressor proposed in Chapter 3 also requires a beneficial noise estimator, the weighted noise estimator is employed for obtaining the noise variance $\lambda_l(k)$. A brief procedure of it is as follows. The noise variance is recursively updated by

$$\lambda_l(k) = \begin{cases} \beta\lambda_{l-1}(k) + (1-\beta)H_l(k)|X_l(k)|^2, & H_l(k) > 0 \\ \lambda_{l-1}(k), & H_l(k) = 0 \end{cases}, \tag{2.21}$$

where $H_l(k)$ is the weight function as shown in Fig. 2.5, and a forgetting factor β is restricted to $0 < \beta < 1$. The weight coefficient is assigned so that it is almost inversely proportional to the estimated SNR as follows:

$$H_l(k) = \begin{cases} 1, & \tilde{\gamma}_l(k) \leq 0 \\ -\frac{1}{\gamma_z}\tilde{\gamma}_l(k) + 1, & 0 < \tilde{\gamma}_l(k) \leq \theta_z, \\ 0, & \theta_z \leq \tilde{\gamma}_l(k) \end{cases}$$

$$\tilde{\gamma}_l(k) = 10\log_{10}\left(\frac{|X_l(k)|^2}{\lambda_{l-1}(k)}\right), \tag{2.22}$$

where γ_z is a constant to decide a slope of graph and θ_z is a threshold to eliminate an unreliable $\tilde{\gamma}_l(k)$. We adjust $\theta_z = 7$ and $\gamma_z = 10$ as shown in [5], [6].

2.3 Impulsive Noise Suppression Based on Zero Phase Signal

In a practical environment, there is impulsive noise which is generated from thunder, clap, other bangs, and so on. Here, we will briefly discuss about an impulsive noise suppression algorithm proposed by Kamamori *et al.* [18]. For simplicity, we omit the frame index l. The DFT coefficient of an observed signal $x(n)$ can be expressed as

$$X(k) = |X(k)|e^{j\angle X(k)}. \tag{2.23}$$

The ZP signal of $x(n)$, $x_0(n)$, is defined as

$$x_0(n) = \frac{1}{N}\sum_{k=0}^{N-1} |X(k)|^\rho e^{j\frac{2\pi n}{N}k}, \tag{2.24}$$

where ρ is a certain constant. Obviously, $|X(k)|^\rho$ can be reproduced from the DFT of the ZP signal $x_0(n)$ as

$$|X(k)|^\rho = \sum_{n=0}^{N-1} x_0(n)e^{-j\frac{2\pi k}{N}n}. \tag{2.25}$$

Since $\rho = 1$ is appropriated for noise suppression as shown in [18], the same value is also applied through this thesis. In addition, we assume that $x(n)$ is a real valued signal. In this case, the ZP signal $x_0(n)$ come to real even signals.

Here, we show a few examples of the ZP signal. Let the spectral amplitude $|X(k)|$ be a constant α_0 (≥ 0). Substituting $|X(k)| = \alpha_0$ into Eq. (2.24) with $\rho = 1$, we have

$$x_0(n) = \alpha_0 \delta(n), \tag{2.26}$$

where $\delta(n)$ denotes the Kronecker's delta function. Eq.(2.26) shows that the ZP signal of any flat spectral amplitude is expressed as the delta function. Next, let $|X(k)|$ be equally-spaced line-spectral pairs (i.e., $x(n)$ is periodic), where each frequency interval is k_c ($0 < k_c < N/2$). That is

$$|X(k)| = \sum_{m=1}^{\left\lfloor \frac{N}{2k_c} \right\rfloor} \frac{\alpha_m}{2} \{\delta(k - mk_c) + \delta(k + mk_c - N)\}, \tag{2.27}$$

where $\lfloor \cdot \rfloor$ denotes a floor function, and α_m is an amplitude of the m^{th} frequency. Substituting Eq. (2.27) into Eq. (2.24) with $\rho = 1$, we have

$$x_0(n) = \sum_{m=1}^{\left\lfloor \frac{N}{2k_c} \right\rfloor} \frac{\alpha_m}{N} \cos \frac{2\pi m k_c}{N} n. \tag{2.28}$$

Hence, the ZP signal of a periodic signal becomes also a periodic signal whose period is N/k_c. The ZP signal becomes an impulse signal when the spectral amplitude is flat, and the ZP signal becomes a periodic signal when the spectral amplitude has values only at equally spaced frequencies.

These properties of the ZP signal are shown in Fig. 2.6 that a speech signal $s(n)$ is periodic and an additional impulsive noise $d(n)$ has a flat spectral amplitude. The ZP signal of $x(n) = s(n) + d(n)$ is approximately represented as

$$x_0(n) \approx \begin{cases} s_0(n) + d_0(n), & n = 0 \\ s_0(n), & \text{otherwise} \end{cases}, \tag{2.29}$$

where $s_0(n)$ and $d_0(n)$ are the ZP signal of $s(n)$ and $d(n)$, respectively. Since $s_0(n)$ is periodic signal, we have $s_0(0) = \max\{s_0(n)\} \approx \max_{n \neq 0}\{x_0(n)\}$, where $\max\{\cdot\}$ denotes the operator to extract the maximum value. Kamamori *et al.* [18] have proposed the following impulsive noise suppression role in ZP domain as

$$\hat{s}_0(n) = \begin{cases} \max_{m \neq 0}\{x_0(m)\}, & n = 0 \\ x_0(n), & \text{otherwise} \end{cases}. \tag{2.30}$$

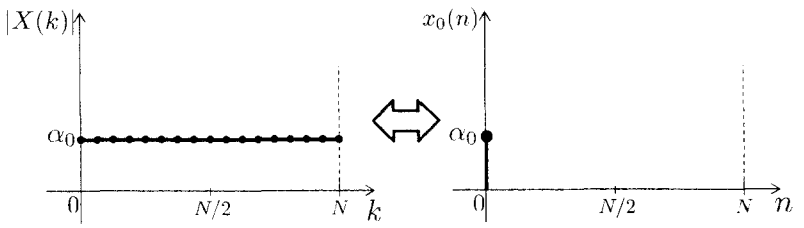
where $m = 0, 1, \ldots, N-1$. The signal $\hat{s}_0(n)$ denotes the estimated speech ZP signal. Then, the estimated speech spectral amplitude is obtained as

$$|\hat{S}(k)| = \sum_{n=0}^{N-1} \hat{s}_0(n) e^{-j\frac{2\pi k}{N}n} \tag{2.31}$$

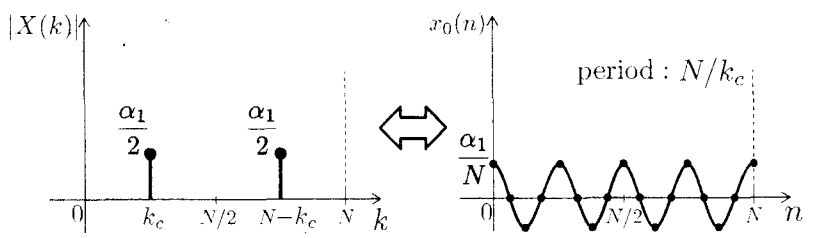
Taking the IDFT of $|\hat{S}(k)|$ with the observed spectral phase gives the estimated speech signal in time domain as

$$\hat{s}(n) = \frac{1}{N} \sum_{k=0}^{N-1} |\hat{S}(k)| e^{j\angle X(k)} e^{j\frac{2\pi n}{N}k}. \tag{2.32}$$

As shown in [18], a white noise and an impulsive noise can be reduced by processing the ZP signal only at the origin. However, this method is not applicable for other impulsive noises. To suppress many kinds of impulsive noise, the concept of [18] is extended and a new technique is proposed in Chapter 4.



(a) constant spectral amplitude



(b) equally spaced line spectra

Figure 2.6: Examples of zero phase signal: (a) constant, (b) equally spaced line spectra.

3

Stationary Noise Suppression Using Real-Speech PDF in Various Narrow SNR Intervals

As discussed in Section 2.2, the shape parameters of the speech spectral amplitude PDF, μ and ν, had been derived from a large amount of speech data in a single narrow SNR interval. However, in a practical situation, a speech signal includes both of activity segments and pause segments. Since the value of the speech spectral amplitude is always zero in the pause segments, its PDF can be modeled as an expected delta function. On the other hand, in the activity speech segments, the PDF of the speech spectral amplitude obeys other functions. As shown in Section 2.2, Tsukamoto *et al.* [5] have noticed the fact and investigated an adaptive method to change the PDF of the speech spectral amplitude, according to the SNR. They have chosen Lotter's PDF defined in Eq. (2.11) as the adaptive PDF, because its shape is easily controlled by ν and μ. Here, the examples of Lotter's PDF with different shape parameters are shown in Fig. 3.1. It is noticed from this figure that the PDF can fit the exponential distribution and the Rayleigh distribution by adjusting the shape parameters. Utilizing real speech histograms, Tsukamoto *et al.* derived adaptive shape parameters and showed its effectiveness through the computer simulations. This basic idea is useful for speech enhancement in a practical situation. Unfortunately, a reliability of the derived adaptive shape parameter is comparatively low, because it derived from only two speech histograms. To sophisticate Tsukamoto's adaptive shape parameter, this research has

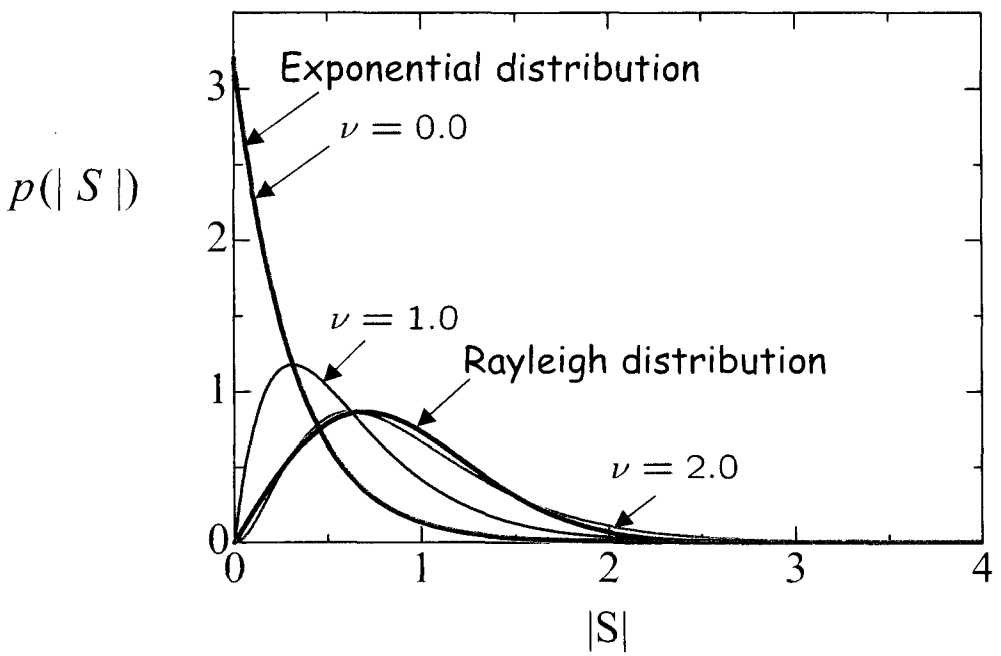


Figure 3.1: Speech PDF function with different shape parameters.

made and evaluated many real speech histograms in various narrow SNR intervals. One of the objective of this research is to fit the speech histograms with Eq. (2.11), and revealed an interesting curve of the shape parameters for narrow SNR intervals.

3.1 Derivation of Shape Parameter Function

The speech PDF with shape parameters μ and ν has been introduced in [3] which is given as

$$p(|S_l(k)|) = \frac{\mu^{\nu+1}}{\Gamma(\nu+1)} \frac{|S_l(k)|^{\nu}}{\sigma_S^{\nu+1}(k)} \exp\left(-\mu \frac{|S_l(k)|}{\sigma_S(k)}\right), \tag{3.1}$$

where $\Gamma(\cdot)$, and $\sigma_S^2(k)$ denote the Gamma function, and the variance of the speech spectrum, respectively. The speech PDF shown in Eq. (3.1) can represent many shapes of PDF, e.g., Super Gaussian which is employed in [3], Gamma in [4], also Rayleigh in [5] by changing its shape parameters. The objective is to find the optimal values for

both parameters that give the best fit of the speech PDF to the speech histogram in each SNR interval. The fitting can be performed by minimizing the distance between the histograms and the speech PDF. To find optimal fitting, the Kullback-Leibler (KL) divergence [31] is employed which is theoretically considered the optimal method for distance measurement. The KL divergence is defined as:

$$KL = \sum_{i=1}^{N_{bin}} (p_h(i) - p_s(i)) \ln \left(\frac{p_h(i)}{p_s(i)} \right), \tag{3.2}$$

where $p_h(i)$ denotes the value of the speech histogram at interval i, and $p_s(i)$ is one of the speech PDFs. N_{bin} is the number of histogram bins. The clean speech signals sampled at 8 kHz from the LDC database [32] are used to make histograms. The speech signals are spoken by 10 male speakers and 10 female speakers with total length around 11 minutes. Firstly, the spectral amplitude data of speech signals are normalized over frequency bins. Then, the *a priori* SNR as a ratio of the speech signal to a stationary microphone noise is calculated which is generally occurred from the microphone and recorded with the speech signals. The normalized spectral amplitude data will be categorized into the *a priori* SNR. After categorizing the spectral amplitude data, SNR-specified histograms are created. Lastly, the author find the optimal shape parameters that minimize the KL divergence between the histogram and the speech PDF in Eq. (3.1). The summary of the procedure of getting the optimal shape parameters is as follows:

1. Obtaining normalized speech spectral amplitude $S_l(k)$

$$\begin{aligned} \tilde{S}_l(k) &= |S_l(k)|/\sigma_S(k), \\ \sigma_S^2(k) &= \frac{1}{M} \sum_{l=0}^{M-1} \{|S_l(k)| - \bar{S}(k)\}^2, \\ \bar{S}(k) &= \frac{1}{M} \sum_{l=0}^{M-1} |S_l(k)|, \end{aligned} \tag{3.3}$$

   where M is the number of frames.

2. Calculating *a priori* SNR $P_l(k)$ [dB].

$$P_l(k) = 10 \log_{10} \hat{\xi}_l(k), \tag{3.4}$$

$$\hat{\xi}_l(k) = \alpha_{\mathrm{snr}} \frac{|\hat{S}_{l-1}(k)|^2}{\lambda_l(k)} + (1-\alpha_{\mathrm{snr}}) F[\gamma_l(k)-1], \tag{3.5}$$

where $\hat{\xi}_l(k)$ is an estimation of $\xi_l(k)$, α_{snr} is a forgetting factor, and $F[\cdot]$ is the half-wave rectifier given as

$$F[y] = \begin{cases} y, & y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Eq. (3.5) is "decision-directed method" which is also explained in Section 2.2. Here, we set $\alpha_{\mathrm{snr}} = 0.98$ according to [11]. Since the observed signal is a speech signal from the corpus, it used $G_l(k) = 1$ in Eq. (2.4) to obtain $\hat{S}_l(k)$. The noise variance $\lambda_l(k)$ was estimated as the averaged value of $|Y_l(k)|$ in the first 6 frames.

3. Categorizing $\tilde{S}_l(k)$ into each SNR interval.

   The author defined a narrow interval of $P_l(k)$ as any interval from 0 to 80 dB, having a 1 dB gap which is sufficiently narrow interval [3], [4], i.e., 0-1, 1-2, ..., 79-80 dB.

4. Making the histograms.

   The number of bins as is adjusted to 30 in an empirical manner to remove jitters from the histograms (30 bins were also selected in [3] and [5] for making speech histograms).

5. Finding the optimal shape parameters of Eq. (A.13) based on KL divergence measurement.

   To obtain the optimal shape parameters, the full search method is applied. The full search method scans μ and ν parameters that give minimum KL divergence in each SNR interval, in the range $0.0 \leq \mu \leq 20$ and $0.0 \leq \nu \leq 3$ with 0.1 gap. Evaluation of the KL divergence $0.0 \leq \tilde{S}_l(k) \leq 3.0$ that covers the main part of the histogram [3].

Fig. 3.2 (a) and (b) show the obtained optimal value of shape parameters. The fitting results may include fluctuations due to the limited amount of the speech data. To reduce the fluctuation, the author used the averaged values of fitting results. Since a higher linearity could be found by dividing the region into several parts, e.g., 19-33 dB

Table 3.1: Shape parameter functions $R_l^{\mu}(k)$ and $R_l^{\nu}(k)$.

SNR range [dB]	$R_l^{\mu}(k) = F[a_0 P_l(k) + b_0]$		$R_l^{\nu}(k) = F[c_0 P_l(k) + d_0]$	
	a_0	b_0	c_0	d_0
$P_l(k) \leq 20$	-0.087	3.50	0.060	-1.04
$20 < P_l(k) \leq 33$	0.045	0.84	0.060	-1.04
$33 < P_l(k) \leq 49$	-0.079	4.90	-0.035	2.11
$49 < P_l(k) \leq 65$	-0.011	1.60	0.039	-1.56
$65 < P_l(k)$	-0.074	5.60	0	1.00

and 33-50 dB, the linear least squares method [33] is used to obtain a linear curve in each range. The results are also shown in Fig. 3.2 (a) and (b), where the solid lines show the results of the linear least square fitting. Table 3.1 shows $R_l^{\mu}(k)$ and $R_l^{\nu}(k)$ that represent the derived linear curves. They are so called the shape parameter functions.

Here, some examples of the speech histogram and the speech PDFs are shown. Fig. 3.3 depicts the histogram of speech amplitude, which is obtained from the 19-20 dB SNR intervals. Fig. 3.3 also shows the conventional speech PDFs [3]–[5] and the proposed speech PDF with the derived shape parameter functions, respectively. The conventional [3], [4], and the proposed speech PDFs give good fitting results, while the speech PDF from [5] is different from other methods in this SNR interval. To observe fitting result in another range, the speech histogram and the speech PDFs in 49-50 dB SNR interval in Fig. 3.4 are shown. Here, it appears that the proposed parameters set provides the best fit for speech histogram. These results support the assumption that the speech histogram has various shapes, and the fixed values of shape parameters from the other conventional methods are no longer appropriate.

3.2 Stationary Noise Suppression Algorithm

In this section, an adaptive stationary noise suppression algorithm is proposed. The proposed algorithm is based on the derived shape parameter functions shown in Table 3.1. Usually, the speech PDF in the present frame cannot not be independent from one in the previous frame. Although the instantaneous variables $R_l^{\mu}(k)$ and $R_l^{\nu}(k)$ exist, using these variable directly might not agree with real estimation, which are dependent

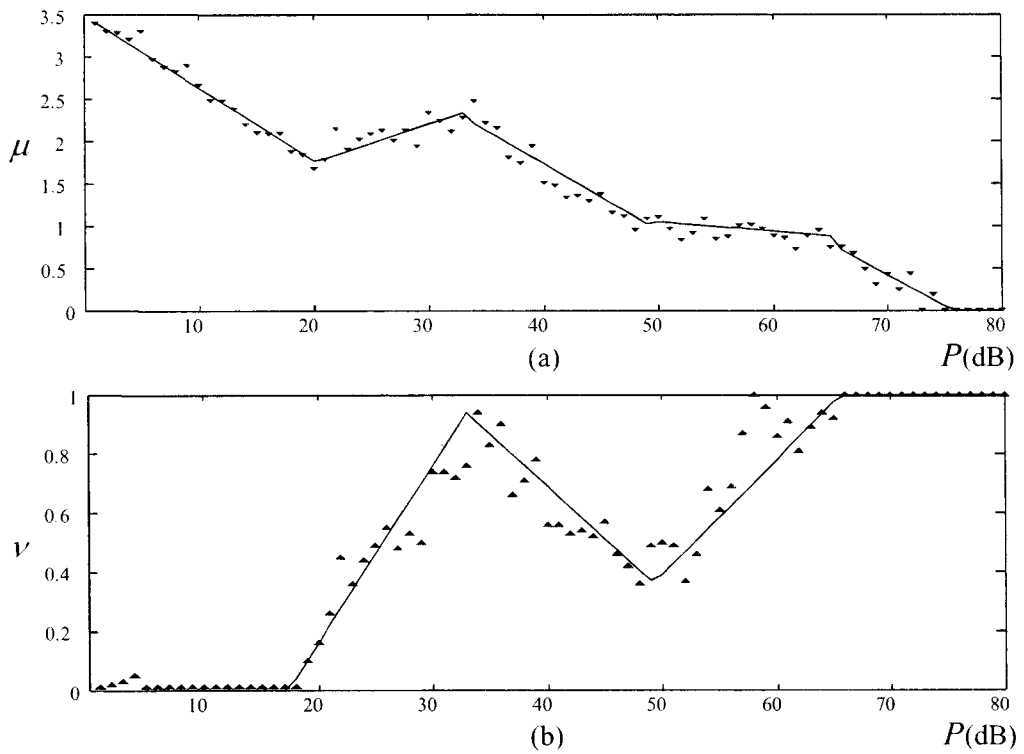


Figure 3.2: Relation between shape parameters and SNR intervals (a) μ for SNR (b) ν for SNR.

upon the present SNR only. For this reason, a forgetting factor to use an averaged value of $R_l^{\mu}(k)$ and $R_l^{\nu}(k)$ is introduced. The proposed speech spectral gain is as follows:

$$G_l(k) = u_l(k) + \sqrt{u_l^2(k) + \frac{\nu_l(k)}{2\gamma_l(k)}}, \tag{3.6}$$

$$u_l(k) = \frac{1}{2} \cdot \frac{\mu_l(k)}{4\sqrt{\gamma_l(k)\hat{\xi}_l(k)}}, \tag{3.7}$$

$$\mu_l(k) = \alpha\mu_{l-1}(k) + (1-\alpha)R_l^{\mu}(k), \tag{3.8}$$

$$\nu_l(k) = \alpha\nu_{l-1}(k) + (1-\alpha)R_l^{\nu}(k), \tag{3.9}$$

where α is the forgetting factor, and $\mu_l(k)$ and $\nu_l(k)$ are the adaptive shape parameters. In the proposed method, the additional computations are 5 multiplications and

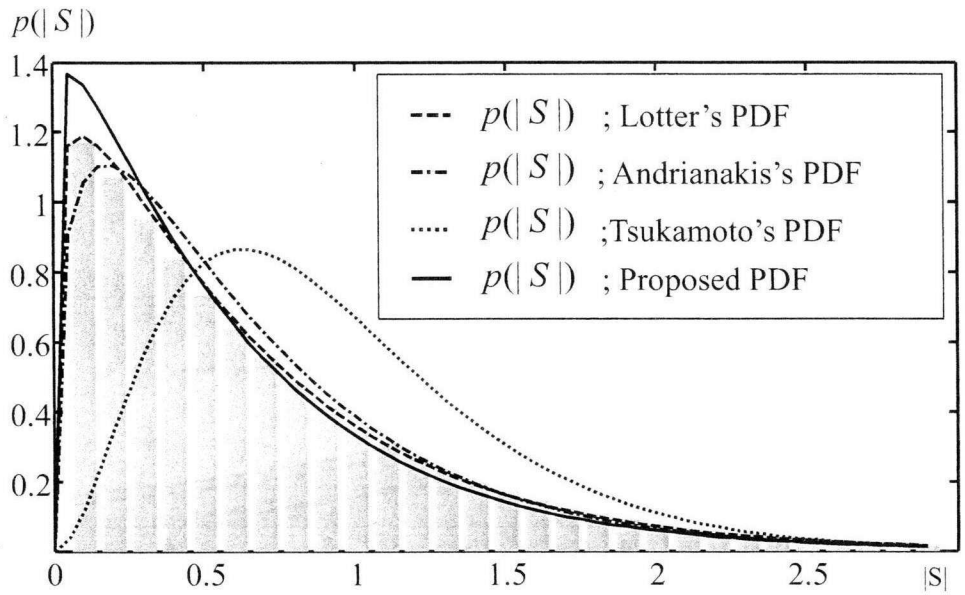


Figure 3.3: Speech histogram in 19-20 dB SNR interval and speech PDFs which are Lotter's PDF [3] (dashed line), Andrianakis's PDF [4] (dotted-dash line), Tsukamoto's PDF [5] (dotted line), and proposed PDF (solid line).

5 additions in comparison to Lotter's method (2.14). In this research, the author puts the initial values as $\mu_0(k) = 20$ and $\nu_0(k) = 0$ that implies $P_0(k) = -190$ dB, where it is assumed that the *a priori* SNR in the first frame is extremely low. Here, sensitivity of the forgetting factor is evaluated. For evaluation, the author uses SegSNR of the enhanced speech. The SegSNR is defined as:

$$\text{SegSNR} = \frac{1}{M}\sum_{l=0}^{M-1} 10\log_{10} f(l), \tag{3.10}$$

$$f(l) = \frac{\sum_{n=0}^{L-1} s^2(lQ+n)}{\sum_{l=0}^{L-1}[s(lQ+n) - \hat{s}(lQ+n)]^2}, \tag{3.11}$$

where M is the total frame number of the input signal. The function $f(l)$ is limited by -10 dB for lower bound and 35 dB for upper bound. Input signals are noisy speech

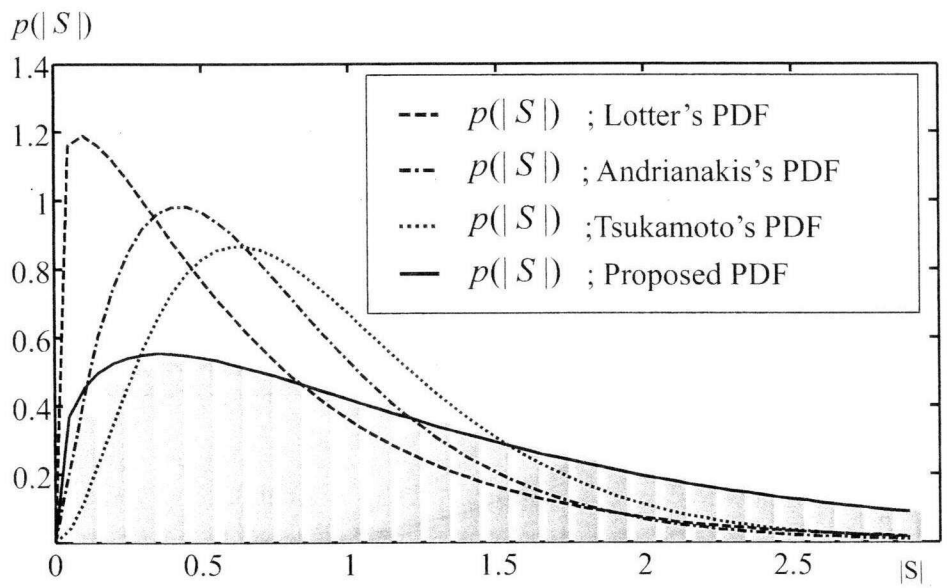


Figure 3.4: Speech histogram in 49-50 dB SNR interval and speech PDFs which are Lotter's PDF [3] (dashed line), Andrianakis's PDF [4] (dotted-dash line), Tsukamoto's PDF [5] (dotted line), and proposed PDF (solid line).

spoken by 2 male speakers and 2 female speakers which are corrupted by tunnel noise with SegSNR = 0. The enhanced speech output SegSNR is shown in Fig. 3.5. From this figure, it shows that the output SegSNR slowly increases as value of α gets higher, and it rapidly increases once value of α is higher than 0.9. It can be obtained from the result that α=0.98 gave the highest SegSNR. In summary, the adaptive shape parameters with proper forgetting factor contribute to improve estimation accuracy of the speech PDF in the current frame.

The characteristic of the proposed stationary noise suppressor is further examined by focusing the quantity of the spectral gain with respect to the *a posteriori* SNR $\gamma_l(k)$. Firstly, roughly analyzing of the proposed spectral gain is proposed as follows:

1. When the *a posteriori* SNR $\gamma_n(k)$ becomes very large, the following approximation from Eq. (3.6) can be made,

$$G_n(k) \approx 2u_l(k). \tag{3.12}$$

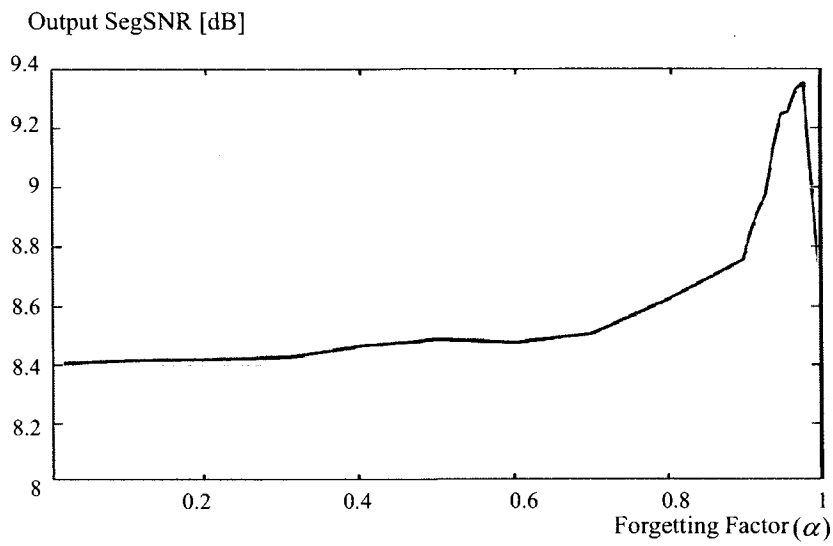


Figure 3.5: Evaluation of sensitivity for forgetting factor α.

As we can see from Fig. 3.2, the value of $\mu_l(k)$ approaches 0 when the SNR gets higher. Then, $u_l(k)$ in Eq. (3.7) can be approximated by

$$u_l(k) \approx \frac{1}{2}. \tag{3.13}$$

By considering Eq. (3.12) and Eq. (3.13), we can conclude that $G_n(k) \approx 1$. This result shows that the output signal is similar to the observed signal, and hence the proposed algorithm can preserve the speech components in the high SNR environment.

2. When $\gamma_l(k)$ is lower than or close to 0 dB, $\nu_l(k)$ is further reduced as shown in Fig. 3.2. By applying $\nu_l(k) \approx 0$, the approximation of Eq. (3.12) is obtained again. From Eq. (3.7), we can say that, when the value of $\mu_l(k)$ increases, $u_l(k)$ is decreased. This leads to a small $G_l(k)$. In this case, the noise components in the low SNR are strongly reduced.

From above analysis, it can be noticed that the effectiveness of the proposed spectral gain in both high and low SNR, in other words, the proposed method is reasonable in both speech and non-speech segments.

Next, the property of the proposed spectral gain is analyzed by observing the theoretical gain curve. Fig. 3.6 (a), (b), (c) and (d) show the gain curves of Lotter's method [3], Andrianakis's method [4], Tsukamoto's method [5] and the proposed method, respectively. Here, the spectral gains are depicted for *a priori* SNR ξ and the instantaneous value of ξ, $\gamma - 1$. We first focus on the effect of the spectral gains in high *a posteriori* SNR ξ situation. As we can see from Fig. 3.6 (a) and (b) when SNR ξ is higher than zero and $\gamma - 1$ is lower than zero, the value of gain reaches to 1 steadily. It means that the gain curves of methods [3] and [4] have less capability to remove existed background noise in high SNR ξ. While in case of method [5] and the proposed gain curves, in high *a priori* SNR ξ situation they show a good capability of noise removal when low value of $\gamma - 1$ persists. Then, we move on to the next observation. When the *a posteriori* SNR ξ is low and $\gamma - 1$ is high, Fig. 3.6 (c) and (d) becomes relatively small. It means that Tsukamoto's method [5] and the proposed spectral gain perform better at suppressing the noise in low SNR situation and non-speech segments.

To confirm that the proposed method reduces background noise effectively, especially in a non-speech segment, we perform noise suppression simulations in next section. In the simulation, the proposed method is compared with the conventional methods [3], [4], and [5].

3.3 Simulation

We carried out noise suppression simulations to confirm the effectiveness of the proposed stationary noise suppressor. All speech signals used in this section were taken from ATR-Promotion database [21] and sampled at 8 kHz. In the noise suppression system, we used a 50% overlapping frame with 256 samples at 8 kHz sampling frequency (i.e., L=256, Q=128). We set the forgetting factor α=0.98.

Firstly, we evaluate the efficiency of the proposed algorithm when the signal contains mainly noise (i.e., non-speech segments). We performed noise suppression for the speech signal corrupted by a tunnel noise with 0 dB of SNR, where the noise is recorded in a tunnel on an expressway in Japan. Fig. 3.7 shows the averaged amplitude frequencies of the enhanced speech in the non-speech segments, i.e., it shows the residual noise level. The results obtained from Lotter's method [3], Andrianakis's method [4], Tsukamoto's method [5] and the proposed method are represented by the dash line, bold dotted

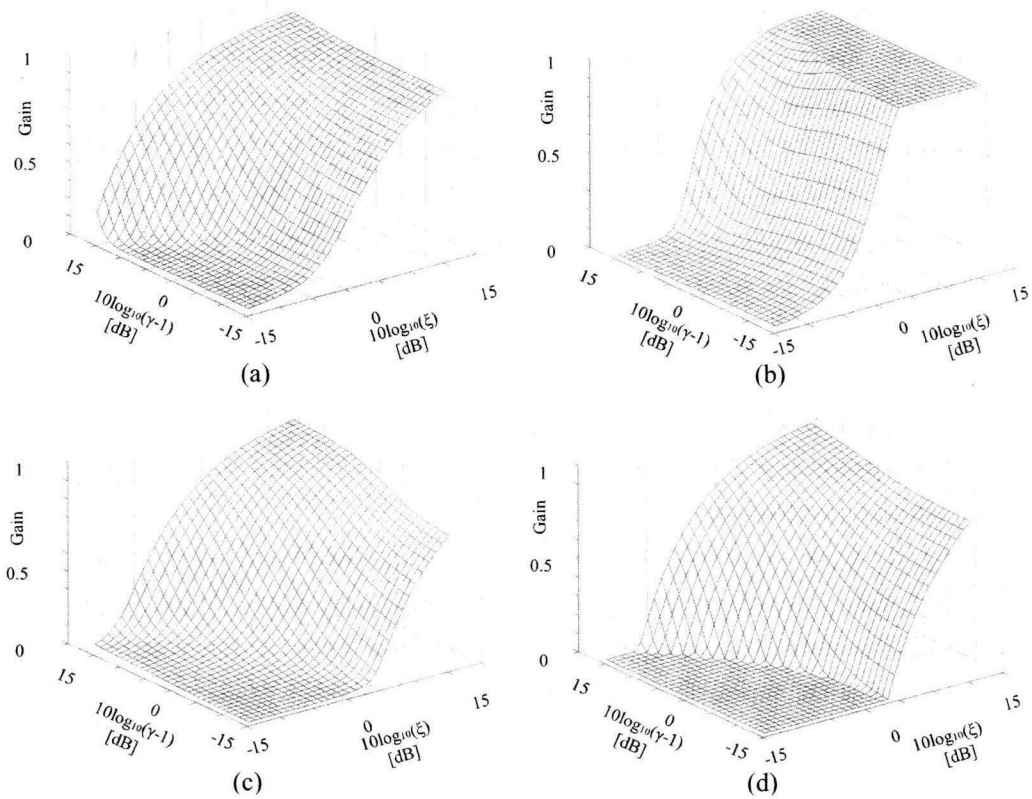


Figure 3.6: Gain curves as a function of the *a priori* SNR ξ and instantaneous SNR γ -1. (a) Lotter's method [3], (b) Andrianakis's method [4], (c) Tsukamoto's method [5] and (d) proposed method.

line, dotted line, and bold line, respectively. The thin solid line is the amplitude of the observed signal. The result of the proposed algorithm exhibits further noise suppression, as compared to the conventional methods. Result from Fig. 3.7 shows that the proposed method can suppress background noise more than the conventional methods, especially, 20 dB further reduced from the conventional method [4] in the frequency range of 1.5 – 2.5 kHz.

Next, the output waveforms and spectrograms of the tunnel noise suppression results are compared. Fig. 3.8 shows the results, where the left hand side shows the waveforms and the right hand side shows the corresponding spectrograms. In Fig. 3.8, (a) shows

the observed signal, (b)–(e) show the noise suppression results by Lotter's methods [3], Andrianakis's method [4], Tsukamoto's method [5], and the proposed method, respectively. From the spectrogram of Fig. 3.8(b), we can observe many spurious spectral peaks which are perceived as a musical noise (e.g. around 3.5 sec). Moreover, we can confirm from the waveforms that the residual noise level of (b) is higher than other results. This result agrees with the fact that Lotter's method gives the smallest zero gain area of the gain curve in comparison to the other method as shown in Fig 3.6. On the other hand, we see from Fig. 3.8(c) that Andrianakis's method tends to remove speech spectral components especially in a low SNR, although the noise suppression capability is superior to (b). This result also agrees with the fact that its spectral gain cannot increase in a low ξ when $\gamma - 1$ becomes large (see Fig. 3.6(b)). Hence, we can expect that the proposed method or Tsukamoto's method is appropriate for noise suppression in comparison to the other methods.

Then, the noise suppression capability of the the proposed method by the SegSNR is evaluated. Since the difference between the proposed method and [5] (or [3]) is just the speech PDF used for calculating the spectral gain, the change of the SegSNR is caused by the speech PDF sophistication of the proposed method. Twenty sentences of clean speech spoken by 5 male speakers and 5 female speakers were corrupted by white, babble and tunnel noise. To avoid the misuse of experimental data, those twenty sentences from all speakers are totally different from the speech used for making histogram in Section 3.1. The white and babble noise were taken from the NOISEX-92 database [35] and added to the clean speech with different input SegSNRs, i.e., 0, 5, and 10 dB. Table 3.2 shows the results of the SegSNR for the proposed and conventional algorithms, where the results by the traditional spectral subtraction method [2] is also shown. We see from Table 3.2 that, for each condition, the proposed method gave the best results in comparison to the other methods. For the white noise suppression with Input SegSNR = 0 dB, noise suppression capabilities of the conventional methods [2] and [3] were comparatively low, where they did not employ an adaptive speech PDF. On the other hand, the adaptive speech PDF methods [4], [5], and the proposed method are superior to [2] and [3], especially, the proposed method attained 6 dB of the SegSNR. For the tunnel noise suppression with Input SegSNR = 0 dB, the proposed method has attained 9.2 dB of the SegSNR, while Tsukamoto's method [5] has improved 7.4 dB. It implies that the proposed sophistication method improved the noise suppression

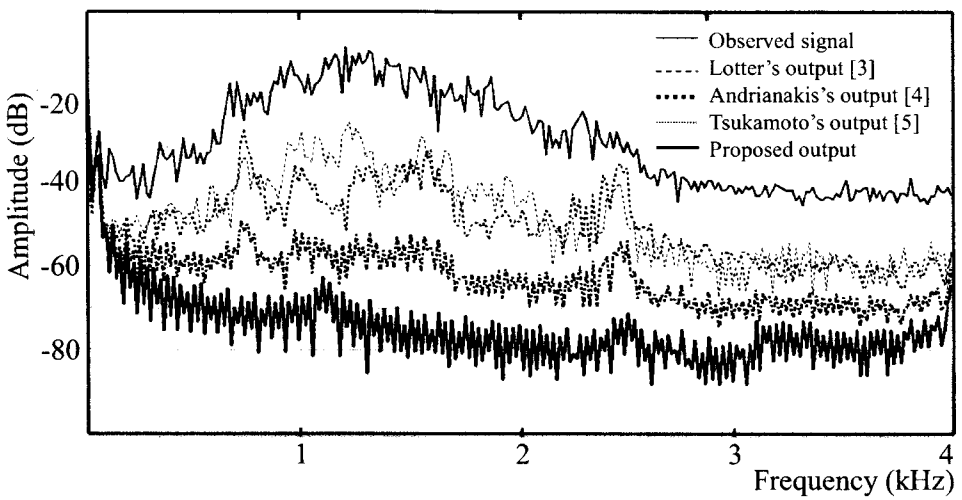


Figure 3.7: Averaged amplitude frequencies in the non-speech segments. Thin solid line represents noise signal, dash line represents output of Lotter's method [3], bold dotted line represents output of Andrianakis's method [4], dotted line represents output of Tsukamoto's method [5], bold line represents output of the proposed algorithm.

capability of the approach in [5]. Although the babble noise suppression capabilities of the adaptive speech PDF methods were comparatively low, the proposed method was slightly superior to [4] and [5]. In this chapter, the variable speech PDF has been derived from real-speech histograms in various narrow SNR intervals, and utilized it in a MAP noise suppressor. The variable speech PDF adaptively changes according to the *a priori* SNR. From spectrograms of the simulation results, we were able to confirm that the proposed method reduces noise effectively, especially in the non-speech segments. Other evaluation results have shown that the proposed method improved SegSNR around 6 and 9 dB when the input speech signal was corrupted by white and tunnel noises at 0 dB, respectively.

In the next chapter, an impulsive noise suppressor will be described, while the proposed and the conventional stationary noise suppressors are impractical for the impulsive noise.

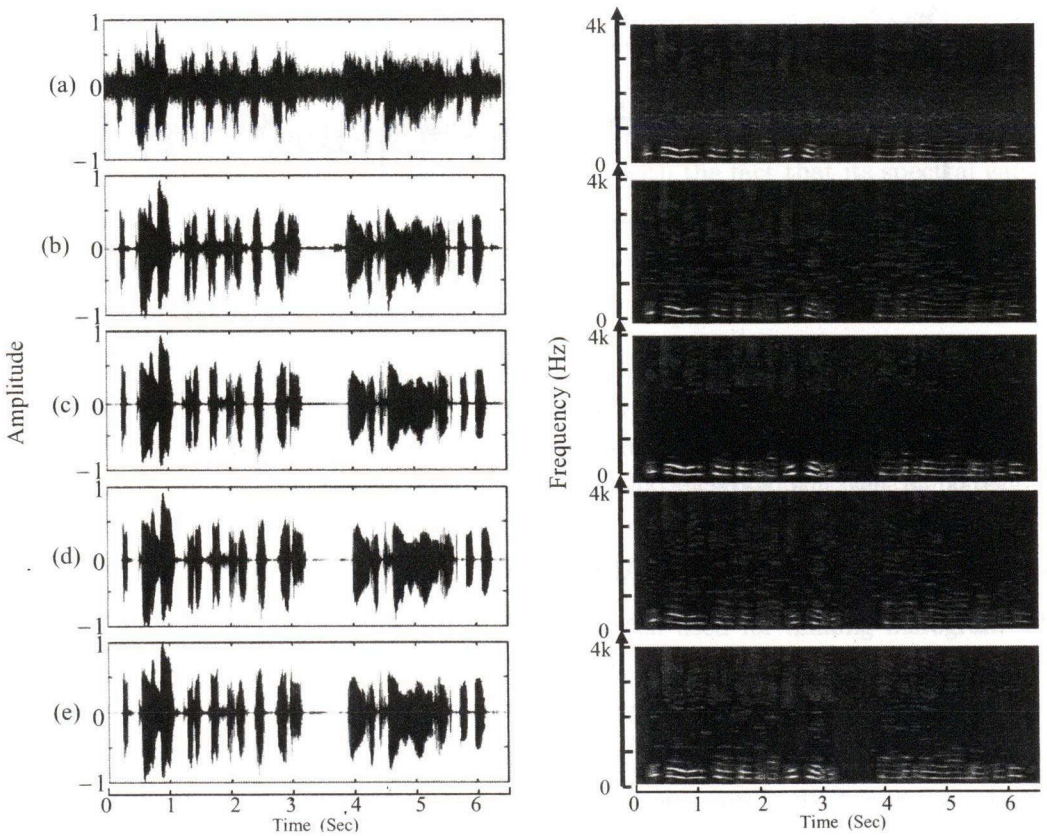


Figure 3.8: Waveforms and spectrograms of tunnel noise suppression: (a) Observed signal, (b) Output by Lotter's method [3], (c) Output by Adrinakis's method [4], (d) Output by Tsukamoto's method [5], (e) Output by proposed method.

Table 3.2: Evaluation results of SegSNR.

Noise	Input SegSNR	Output SegSNR [dB]				
		Spectral subtraction [2]	Lotter's method [3]	Andrianakis's method [4]	Tsukamoto's method [5]	Proposed Method
White	0	1.4	2.5	4.3	4.9	6.0
	5	5.5	6.2	6.5	8.0	8.6
	10	9.1	9.9	9.0	11.0	11.1
Tunnel	0	0.8	3.4	9.1	7.4	9.2
	5	4.7	7.2	10.5	10.8	12.0
	10	8.4	10.6	12.5	13.3	14.0
Babble	0	0.6	1.4	3.0	2.5	3.3
	5	4.3	5.5	6.7	6.4	7.1
	10	8.0	9.5	10.2	10.2	10.7

3. STATIONARY NOISE SUPPRESSION USING REAL-SPEECH PDF IN VARIOUS NARROW SNR INTERVALS

4

Impulsive Noise Suppression Using Zero Phase Signal Replacement Technique

In this chapter, an impulsive noise suppression scheme is derived in a frame work of spectral gain approaches. To obtain an appropriate spectral gain, the proposed method utilizes a zero phase (ZP) signal which is defined as the IDFT of a spectral amplitude [18]. The ZP signal becomes an impulse signal when the spectral amplitude is flat, and the ZP signal becomes a periodic signal when the spectral amplitude has values only at equally spaced frequencies. As shown in [18], a white noise and an impulsive noise can be reduced by processing the ZP signal only at the origin. However, this method is not applicable for other noises. To suppress real impulsive-type noise which has a duration that is normally more than one sample long, we extend the concept of [18]. This research assume that a noise spectral amplitude is approximately flat, and a speech signal is periodic in a short observation. Then, we can suppress the noise by replacing the noisy ZP signal around the origin with the ZP signal in the second or latter period. Unlike [18], in the replacement technique, it has to be investigated about an appropriate ZP samples used for replacement. In addition, a scaling function is introduced in this technique for compensating a decay of ZP signal, where the decay is caused by segmenting and windowing an observed signal.

4.1 Zero Phase Signal of Noise Signals

As the same manner of conventional noise suppression methods [11]–[17], it is also assumed that the spectral phase of the estimated speech signal is equal to that of the observed signal, i.e., $\angle\hat{S}(k) = \angle X(k)$. It means that

$$x_0(n) = s_0(n) + d_0(n), \tag{4.1}$$

where $s_0(n)$ and $d_0(n)$ are the ZP signals of $s(n)$ and $d(n)$, respectively. Under the assumption, we derive a wide-band noise suppression system which can suppress both of stationary and non-stationary wide-band noises, without a priori estimation of noise spectral amplitudes.

Firstly, a speech signal $s(n)$ in a short observation is modeled as a HNM (Harmonic plus Noise Model) [19], [20] given by

$$s(n) = \sum_{m=1}^{\left\lfloor \frac{N}{2k_c} \right\rfloor} \alpha_m \cos(2\pi \frac{k_c}{N} mn + \theta_m) + \varepsilon(n), \tag{4.2}$$

where k_c/N is the normalized fundamental frequency, and α_m and θ_m are the amplitude and the phase of the m^{th} harmonic frequency, respectively. The signal $\varepsilon(n)$ is a noise signal generated by passing a white noise through an all-pole filter [20]. Here, we assume that the energy of $\varepsilon(n)$ in an observation frame is sufficiently small in comparison to one of the harmonic part. This assumption is appropriate for a voiced speech, but it is not appropriate for an unvoiced speech. Although this assumption may give a degradation to an enhanced speech, the degradation is not fatal. Because, voiced speech energy is usually much greater than unvoiced one.

Next, the properties of practical noise and speech signals in the ZP domain are shown. The ZP signals of some practical wide-band noises and a female speech signal are plotted in Fig. 4.1, where (a) shows a tunnel noise, (b) shows a motor noise, (c) shows a babble noise, (d) shows a clap noise, (e) and (f) show voiced and unvoiced speech signals, respectively. Here, all the signals were sampled at 8kHz and $N = 256$.

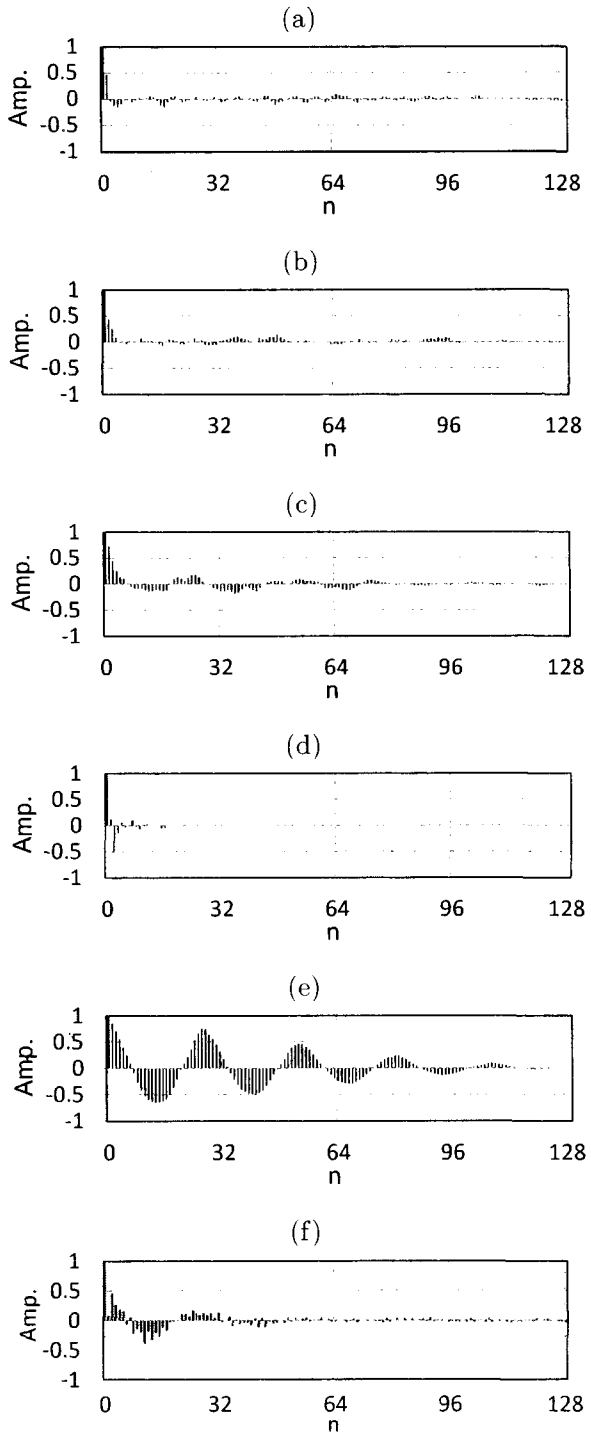


Figure 4.1: Zero phase signals. (a) tunnel noise, (b) motor noise, (c) babble noise, (d) clap noise, (e) voiced speech signal, (f) unvoiced speech signal.

We see from Figs. 4.1(a)–(d) that the energy of all wide-band noises is concentrated around time 0 in the ZP domain. Hence, if we remove the ZP signal around the origin, then the noise is greatly reduced. On the other hand, from Fig. 4.1(e), we see that the voiced speech becomes a periodic signal with amplitude attenuation in the ZP domain. This attenuation arises due to the window function. Since the window function is known, we can compensate the attenuation. We also see from Fig. 4.1(e) that the effect of $\varepsilon(n)$ is extremely low for the voiced speech. On the other hand, the ZP signal of the unvoiced speech shown in Fig. 4.1(f) is similar to that of the noises. As shown in Figs. 4.1(e) and (f), the energy of the unvoiced speech is less than the voiced one. In this research, we concentrate on extracting the voiced speech rather than the unvoiced one.

4.2 Impulsive Noise Suppression Algorithm

The noise ZP signal has nonzero values mainly around origin. Hence, we assume that the noise ZP signal $d_0(n)$ at $(n > L)$ is sufficiently small for $x_0(n)$. Then we have

$$x_0(n) \approx \begin{cases} s_0(n) + d_0(n), & 0 \le n \le L \\ s_0(n), & L < n \le \frac{N}{2}, \end{cases} \tag{4.3}$$

$$x_0(n) = x_0(N-n), \quad \frac{N}{2} < n < N. \tag{4.4}$$

When the pitch period of the speech ZP signal, $T = N/k_c$, is greater than L, we can estimate T as the time index of the second peak of $x_0(n)$ as shown in Fig. 4.2. Since the observed ZP signal $x_0(n)$ in $T \le n < N + L$ does not include the noise components, we obtain the estimated speech ZP signal $\hat{s}_0(n)$ by the following replacement.

$$\hat{s}_0(n) = \begin{cases} \mathrm{sc}(n) \cdot x_0(T+n), & 0 \le n \le L \\ x_0(n), & L < n \le \frac{N}{2} \end{cases}, \tag{4.5}$$

where $\mathrm{sc}(n)$ is a scaling function to compensate the envelope attenuation of the speech ZP signal. It is obtained as the reciprocal function of the window for signal segmentation. When we use the hanning window, the scaling function $\mathrm{sc}(n)$ is given as (see Appendix A.5)

$$\mathrm{sc}(n) = \frac{1 + \cos\frac{2\pi}{N}n}{1 + \cos\frac{2\pi}{N}(n+T)}. \tag{4.6}$$

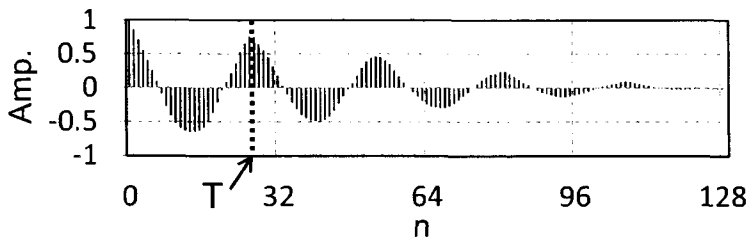


Figure 4.2: T obtained from second peak of ZP signal.

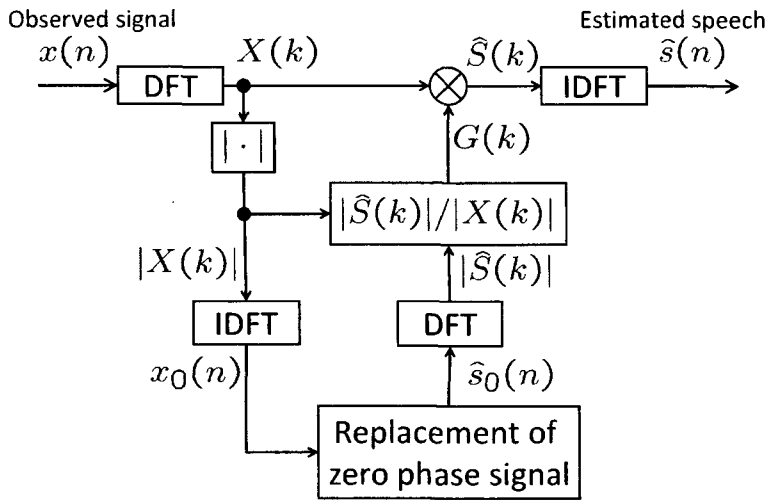


Figure 4.3: Proposed wide-band noise suppression system using zero phase signal.

Comparing Eq. (2.30) and (4.5), we see that the novelty of the proposed method is to extend the replacement samples from one to L, and to introduce the scaling function $\mathrm{sc}(n)$. After the replacement (4.5), the DFT of $\hat{s}_0(n)$ gives the estimated speech spectral amplitude $|\hat{S}(k)|$. Finally, taking the IDFT of $|\hat{S}(k)|e^{j\angle X(k)}$, we have the estimated speech signal $\hat{s}(n)$ in time domain.

Fig. 4.3 shows the block diagram of the proposed wide-band noise suppression system, where the spectral gain is given as $G(k) = |\hat{S}(k)|/|X(k)|$. Here, this system requires the additional DFT and IDFT to achieve stationary and non-stationary wide-band noise suppression without a priori estimation of noise spectral amplitudes. The most important parameters in the proposed method are the pitch period T and the re-

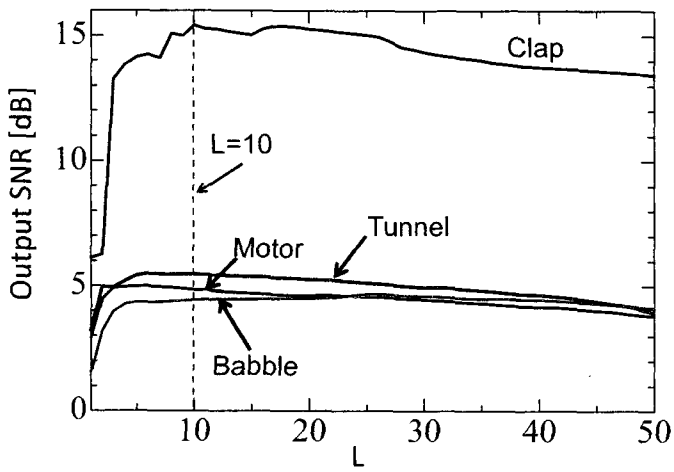


Figure 4.4: Practical wide-band noise suppression results for various L with Input SNR= 0dB.

placement size L shown in (4.5). In the next section, we investigate about an estimation method of them.

We first describe about an estimation method of the pitch period T, and then derive an appropriate replacement size L in an empirical manner.

From the definition (2.24), we see that any ZP signal takes the maximum value at the origin. On the other hand, as shown in Fig. 4.2, a voiced speech provides a periodic ZP signal with amplitude attenuation. Hence, as we stated in the previous section, the index of the second peak in the speech ZP signal gives T. As reported in [22], an averaged pitch period of male speakers is about 8ms, and that of female speakers is about 4ms. Hence, an computationally efficient peak search method can be established by restricting the search range. The pitch period T is given as

$$T = \arg \max_{t_L \leq n \leq t_H} \{x_0(n)\}, \tag{4.7}$$

where, t_L is the lowest index number of the search range, and t_H is the highest one.

Next, we choose the replacement size L in an empirical manner. For various L, we performed wide-band noise suppression simulations, and evaluated its capability by using

$$\text{InputSNR} = 10\log_{10}\frac{\sum_{n=0}^{M-1} s^2(n)}{\sum_{n=0}^{M-1} d^2(n)}, \tag{4.8}$$

$$\text{OutputSNR} = 10\log_{10}\frac{\sum_{n=0}^{M-1} s^2(n)}{\sum_{n=0}^{M-1} \{\hat{s}(n) - s(n)\}^2}, \tag{4.9}$$

where M is the number of samples. The results for the four practical noises with Input SNR of 0dB are shown in Fig. 4.4. We see from this figure that the proposed method is effective for wide-band noise suppression, especially suppressing the non-stationary clap noise. Although the respective maximum Output SNRs gave different values of L, all they were less than 10. Hence, we employ $L = 10$ as an appropriate value.

In the next section, we perform other noise suppression simulations to confirm the effectiveness of the proposed method with $L = 10$.

4.3 Simulation

We evaluate the capability of the proposed method in further detail. The speech signals used in the simulations were taken from ATR-promotion database [21]. All signals used in simulations were sampled at 8kHz. We put $N = 256$ and $L = 10$, and used the Hanning window for signal segmentation. We put $t_L = 16$ and $t_H = 64$ that implies the pitch search range from 2ms to 8ms. The proposed method was compared with some conventional methods.

Firstly, we performed the noise suppression simulations for input SNR of 0dB, where we used artificially generated white and impulsive noises. In the simulations, we compared the proposed method with the most traditional spectral subtraction method [2].

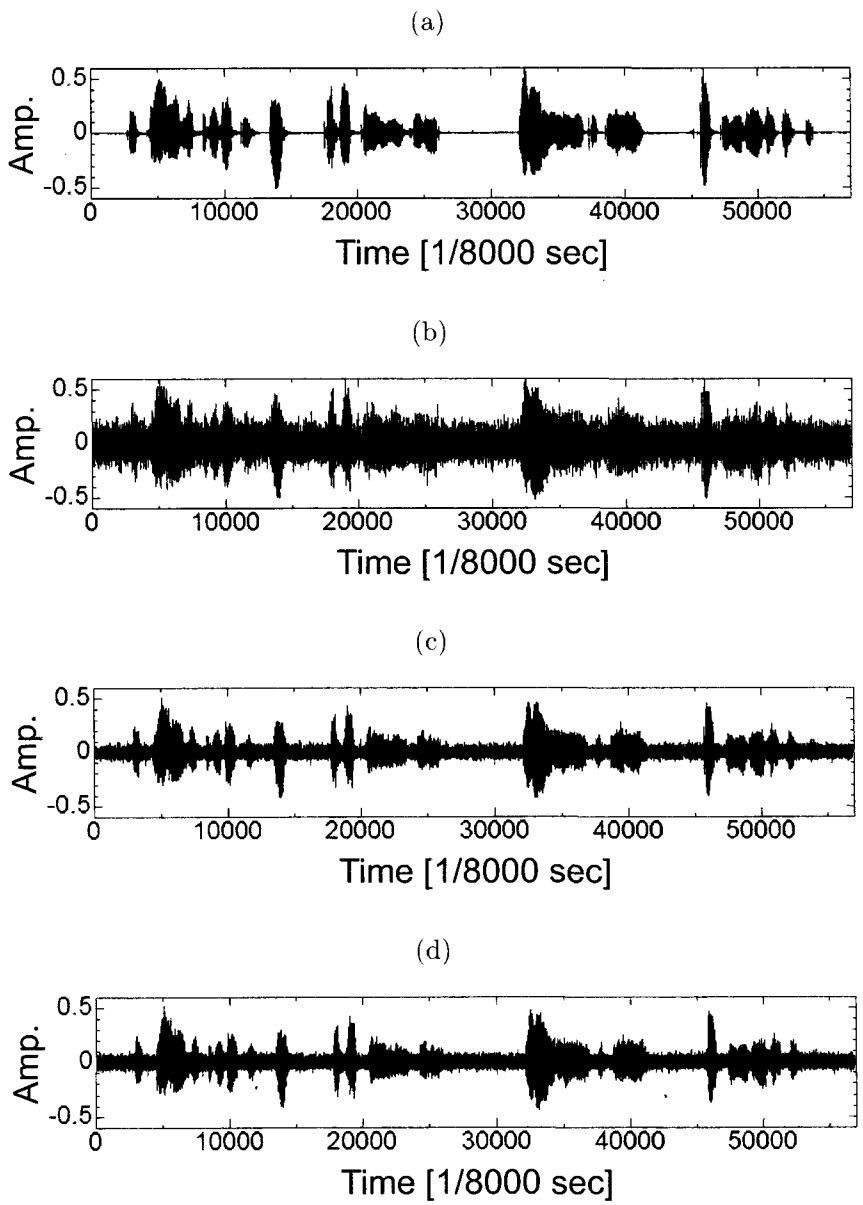


Figure 4.5: Results of white noise suppression: (a) clean speech signal, (b) speech signal corrupted by white noise (SNR=0.0dB), (c) the estimated speech by the spectral subtraction method (SNR=7.0dB), (d) the estimated speech by the proposed method (SNR=6.8dB).

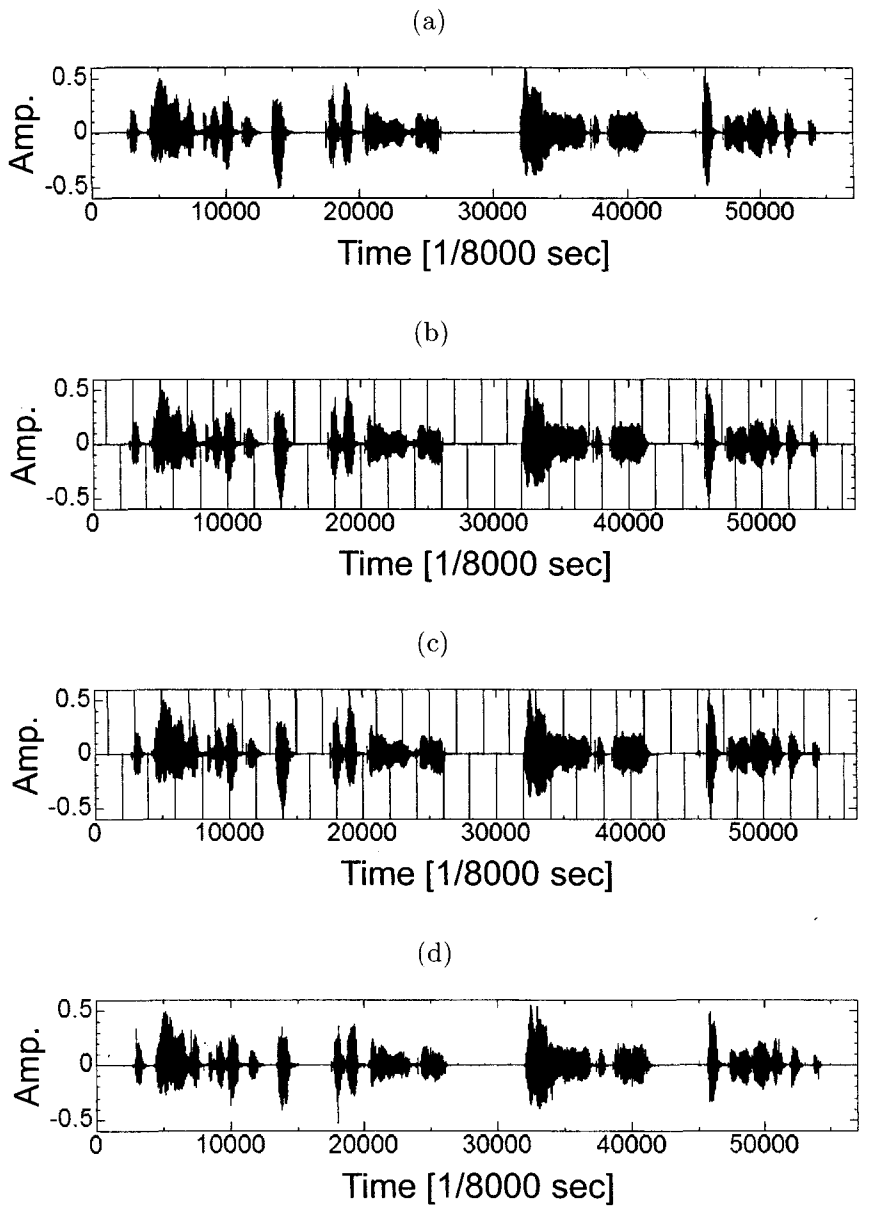


Figure 4.6: Results of impulsive noise suppression: (a) clean speech signal, (b) speech signal corrupted by impulsive noise (SNR=0.0dB), (c) the estimated speech by the spectral subtraction method (SNR=−0.1dB), (d) the estimated speech by the proposed method (SNR=13.3dB).

The waveforms of the white noise suppression results are shown in Figs. 4.5(a)–(d), where (a) shows the clean speech signal, (b) shows the speech signal corrupted by the white noise with the input SNR of 0.0dB, (c) shows the estimated speech signal obtained by the spectral subtraction method, where the output SNR was 7.0dB, and (d) shows the estimated speech signal obtained by the proposed method, where the output SNR was 6.8dB. From these results, we see that the proposed method can suppress the stationary wide-band noise, without a prior estimation of noise spectral amplitudes. For the impulsive noise suppression simulation, we imposed the condition that amount of impulse per block-segment is one or zero on this simulation.

The results are summarized in Figs. 4.6(a)–(d), where (a) shows the clean speech signal, (b) shows the speech signal corrupted by the impulsive noise with the input SNR of 0.0dB, (c) shows the estimated speech signal by the spectral subtraction method, where the output SNR was −0.1dB, and (d) shows the estimated speech signal by the proposed method, where the output SNR was 13.3dB. From these results, we see that the proposed method can suppress the non-stationary impulsive noise with the same procedure of stationary wide-band noise suppression. The spectral subtraction method cannot suppress such non-stationary wide-band noise.

Next, we carried out noise suppression simulations for 8 kinds of wide-band noises with different input SNRs. For the stationary wide-band noises, we used a white noise, tunnel noise, motor noise, and babble noise. On the other hand, impulsive noise, clap noise, white mixed with impulsive noise, and train noise were used as non-stationary wide-band noises. Here, the motor and babble noises were obtained from a SPIB database [23], train noise was obtained from a noise database distributed from Sunrise Music inc. [24], and clap and tunnel noises were practically recorded by the authors. The speech signals are spoken by 10 male and 10 female from ATR-promotion database [21]. For evaluating noise suppression capability, we used the Output SNR as a time domain criterion and Itakura-Saito Distance (ISD) [22] as a frequency domain criterion. The ISD is defined as

$$\text{ISD} = \frac{1}{J}\sum_{j=0}^{J-1}\frac{1}{N}\sum_{k=0}^{N-1}(\log\frac{f(k,j)}{g(k,j)}+\frac{f(k,j)}{g(k,j)}-1), \tag{4.10}$$

where J is the number of frames, and $f(k,j)$ and $g(k,j)$ are k^{th} bin of spectral envelopes in the j^{th} frame obtained by the maximum likelihood estimation. The spectral envelope

$f(k,j)$ is given as [22]

$$f(k,j) = \frac{1}{N}\frac{\sigma_f^2}{1+2\sum_{i=1}^{P} A_i \cos(2\pi ki/N)}, \tag{4.11}$$

$$A_i = \sum_{m=0}^{P-|i|} a_m a_{m+|i|}, \tag{4.12}$$

where a_m ($m = 1, 2, \ldots, P$) is the m^{th} linear predictor coefficient for the speech signal $s(n)$ in the j^{th} frame. P denotes the order of the linear predictor, and σ_f^2 is the variance of the residual error. The same procedure for the estimated speech $\hat{s}(n)$ gives the other spectral envelope $g(k,j)$. For all of the following simulation results, we compared the proposed method with the spectral subtraction (SS) [2], a variable Maximum a Posteriori estimation method (VMAP) [5], and the conventional ZP signal method (CZPS) [18]. Here, VMAP is a recently proposed spectral gain method and CZPS is a noise suppression method utilizing the ZP signal only at the origin.

Table 4.1 and Figs. 4.7(a)–(h) show the output SNR of the wide-band noise suppression results. We see from the results for the stationary wide-band noise shown in Figs. 4.7(a)–(d) that the noise suppression capability of the proposed method is almost the same or slightly low in comparison to ones of SS and VMAP methods which require the prior estimation of the noise spectral amplitude. On the other hand, CZPS and the proposed method do not require any prior estimation of the noise spectral amplitude. In the stationary practical noise cases (Figs. 4.7(b)–(d)), the proposed method is superior to the CZPS. This improvement is caused by removing the noise ZP signal $d_0(n)(1 \leq n \leq L)$ which cannot be reduced by the CZPS method. Note that the capability of the proposed method exactly reaches to ones of SS or VMAP when we utilize the prior estimation of the noise spectral amplitude. On the other hand, for non-stationary wide-band noises (Figs. 4.7(e)–(h)), the noise suppression capability of the proposed method and CZPS are superior to SS and VMAP. When the input SNR was 0dB in clap noise situation, the proposed method improved the SNR to 13.5dB which is 7dB higher than the result of the CZPS. Table 4.2 and Figs. 4.8(a)–(h) show the ISD of the simulation results, where it expresses speech spectral envelope distortion. Note that the lower value of ISD is better than the higher one. We see from Fig. 4.8(a) that SS, CZPS and the proposed method gave almost the same results for the white

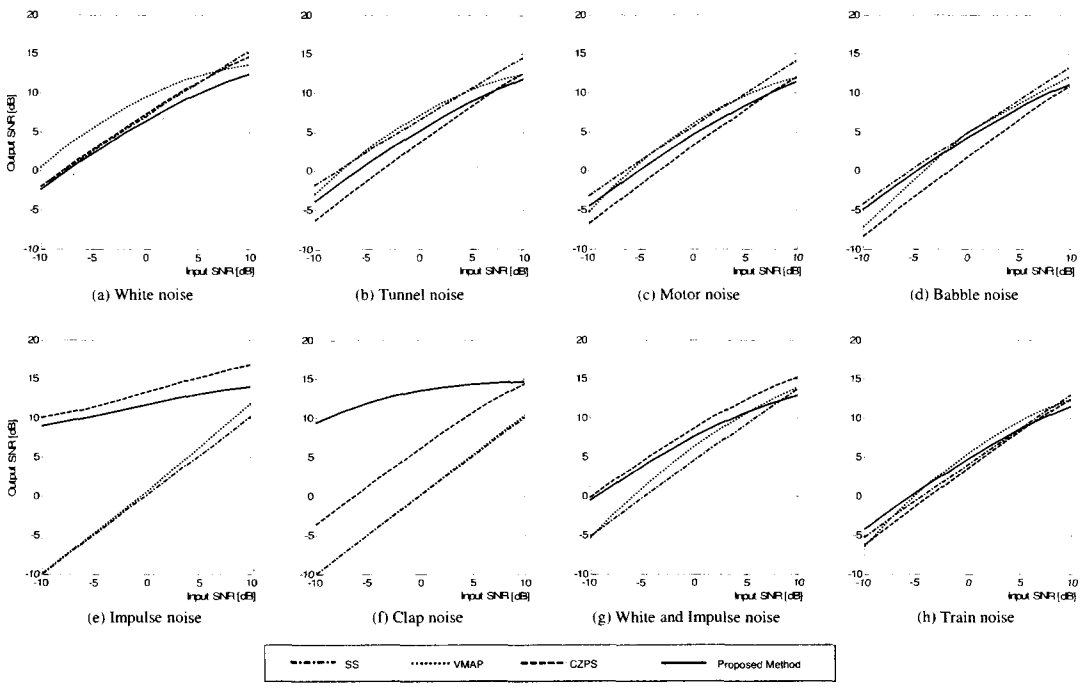


Figure 4.7: Output SNR of noise suppression results

noise. We also notice from Figs. 4.8(b)-(d) that the proposed method can improve noise suppression capability in comparison to the CZPS. On the other hand, the proposed method significantly improve noise suppression capability for the non-stationary wide-band noise. As shown in Figs. 4.8(e)-(f), ISD results from the proposed method gave the lowest value among other ones. The proposed method also gave comparatively low ISD results for the white and impulsive noise and the train noise as shown in Figs. 4.8(g)-(h).

After that, we evaluated speech quality by the formal listening test. The speech quality was rated by a scale of 1 (bad) to 5 (excellent). We average those scores obtained from 15 listeners as the mean opinion score (MOS). Table 4.3 and Fig. 4.9 show MOS results for the four methods under the tunnel and clap noise conditions. In the tunnel noise case, the proposed method gave a better result than ones from SS and CZPS. While VMAP gave a high speech quality result in the tunnel noise case, it gave the

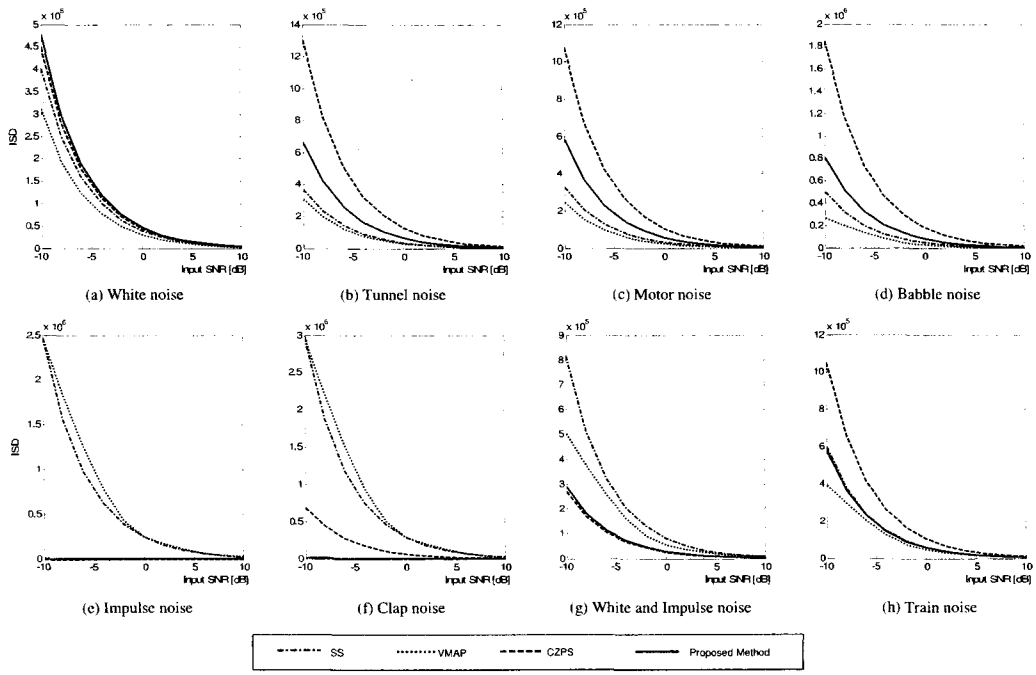


Figure 4.8: ISD of noise suppression results

lowest one in the clap noise case. We can see from Fig. 4.9 that the proposed method gave the best speech quality result in the clap noise case. In Table 4.3, it also shows standard deviation of the listening test results where the proposed method gave the smallest deviation among the others.

Finally, we evaluated the pitch estimation accuracy since it is one of the important factors for the proposed method. The pitch estimation performance is evaluated by Gross Pitch Error (GPE) which is given as [25]

$$\text{GPE} = \frac{N_{F0E}}{N_V}, \tag{4.13}$$

where N_V is the number of total frames considered as voiced speech segment. The value of N_{F0E} is the number of frames that satisfies

$$\left| \frac{F0_{\text{estimated}}}{F0_{\text{reference}}} - 1 \right| > \gamma\%, \tag{4.14}$$

where γ is a threshold which is set to 10 in this research. $F0_{\text{estimated}}$ denotes the estimated pitch frequency obtained by the proposed method, and $F0_{\text{reference}}$ denotes the true pitch frequency, where we obtained $F0_{\text{reference}}$ by the discrete-time Fourier transform (DTFT) with 1 Hz gap. In this GPE evaluation, we also used speech data spoken by 10 male and 10 female from [21]. To evaluate GPE, we added 8 kinds of noises at different input SNRs. The results are plotted in Fig. 4.10(a) and (b). We can see from Fig. 4.10(a) and (b) that GPE decreases when SNR gets higher in all simulated situations. As shown in Fig. 4.10(a), when speech is corrupted by the stationary noises at 0dB, GPE varies between 0.15 to 0.35. On the other hand, when the speech is corrupted by non-stationary clap noise, we can get GPE values less than 0.15 even if SNR is extremely low. It means that the proposed method is effective especially for stationary noises in high SNR situations, and for impulsive noises in any SNR situation.

In this chapter, we have proposed a wide-band noise suppression method based on the ZP signal replacement. The noise suppression is achieved by replacing the observed ZP signal around the origin with the ZP signal in the second period. The proposed method does not require a prior estimation of noise spectral amplitudes, and can suppress not only stationary wide-band noises but also non-stationary wide-band noises. Many simulation results have shown the effectiveness of the proposed noise suppression method. The stationary wide-band noise suppression capability of the proposed method is almost the same or slightly low in comparison to the spectral subtraction method and the variable MAP method which require a priori estimation of noise spectral amplitudes. The most advantage of the proposed method is that it can provide a high noise suppression performance for non-stationary wide-band noises. In the clap noise suppression, the proposed method attained the output SNR of 13.5dB when the input SNR was 0dB. The effectiveness of the proposed method for the other wide-band noises was also confirmed. Future works include to derive an extraction method of unvoiced speech in the ZP domain.

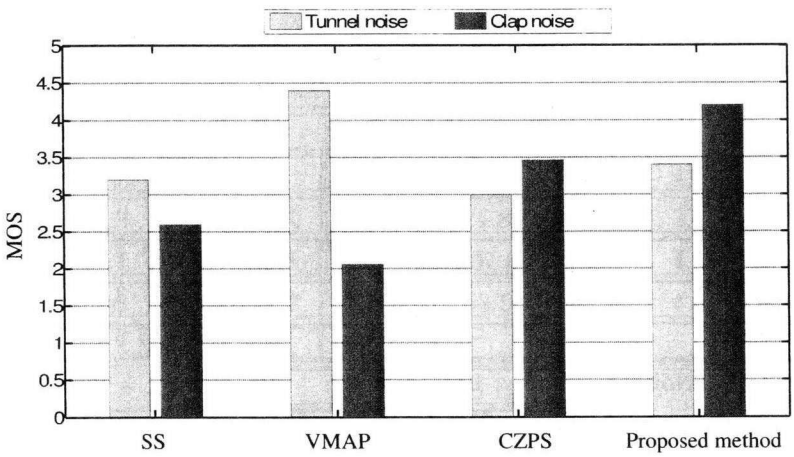


Figure 4.9: Formal listening test results for tunnel noise and clap noise.

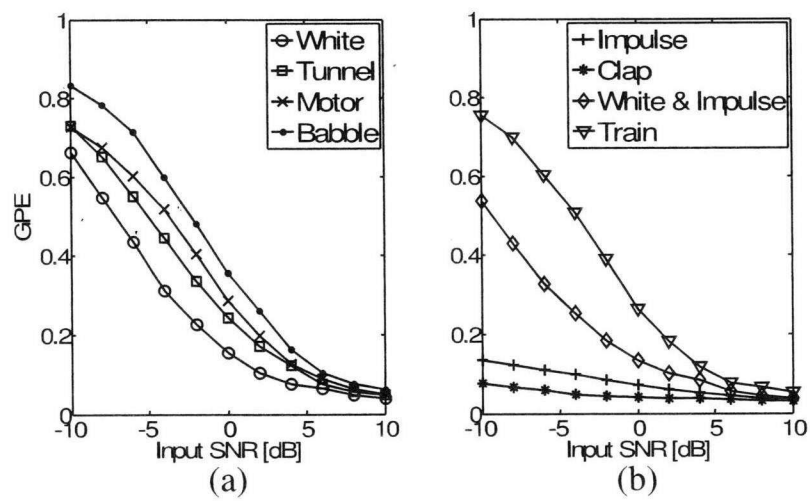


Figure 4.10: GPE results for various kinds of noise.

4. IMPULSIVE NOISE SUPPRESSION USING ZERO PHASE SIGNAL REPLACEMENT TECHNIQUE

Table 4.1: Output SNR of wide band noise suppression results [dB]

System \ Noise	-10.0 [dB]				0.0 [dB]			
	W	Tn	M	B	W	Tn	M	B
SS [2]	−2.0	−1.8	−3.2	−4.2	6.9	6.5	5.6	4.7
VMAP [5]	0.4	−3.0	−5.2	−7.1	9.4	7.1	6.0	4.8
CZPS [18]	−2.0	−6.3	−6.7	−8.3	7.2	3.5	3.1	1.6
Proposed method	−2.4	−3.9	−4.5	−4.9	6.3	5.1	4.5	4.1

System \ Noise	10.0 [dB]			
	W	Tn	M	B
SS [2]	15.3	14.5	14.1	13.3
VMAP [5]	13.6	12.4	12.2	12.1
CZPS [18]	14.6	12.5	12.0	10.9
Proposed method	12.4	11.8	11.4	11.1

W : white noise Tn : tunnel noise M : motor noise B : babble noise

System \ Noise	-10.0 [dB]				0.0 [dB]			
	I	C	WI	Tr	I	C	WI	Tr
SS [2]	−9.9	−9.9	−5.1	−5.2	0.1	0.1	4.5	4.0
VMAP [5]	−9.8	−9.9	−5.3	−6.4	0.4	0.1	6.3	5.4
CZPS [18]	10.1	−3.6	−0.1	−6.2	13.3	6.1	8.6	3.5
Proposed method	9.0	9.4	−0.5	−4.2	11.7	13.5	7.6	4.7

System \ Noise	10.0 [dB]			
	I	C	WI	Tr
SS [2]	10.2	10.1	13.7	13.0
VMAP [5]	11.9	10.4	13.9	12.4
CZPS [18]	16.8	14.4	15.2	12.4
Proposed method	14.0	14.7	12.9	11.5

I : impulsive noise C : clap noise WI : white and impulsive noise Tr : train noise

Table 4.2: ISD of wide band noise suppression results ($\times 10^4$)

System \ Noise	-10.0 [dB]				0.0 [dB]			
	W	Tn	M	B	W	Tn	M	B
SS [2]	40.1	36.8	32.8	50.2	4.0	3.7	3.3	5.0
VMAP [5]	31.0	30.7	24.8	26.6	3.1	3.0	2.5	2.7
CZPS [18]	44.8	130.5	107.6	185.1	4.5	13.1	10.7	18.5
Proposed method	47.7	66.6	58.5	81.0	4.8	6.6	5.8	8.1

System \ Noise	10.0 [dB]			
	W	Tn	M	B
SS [2]	0.4	0.4	0.3	0.5
VMAP [5]	0.3	0.3	0.2	0.3
CZPS [18]	0.4	1.3	1.1	1.8
Proposed method	0.5	0.7	0.6	0.8

W : white noise Tn : tunnel noise M : motor noise B : babble noise

System \ Noise	-10.0 [dB]				0.0 [dB]			
	I	C	WI	Tr	I	C	WI	Tr
SS [2]	247.2	297.2	81.7	59.4	24.7	29.7	8.2	5.9
VMAP [5]	247.1	297.0	50.5	39.2	24.8	29.7	5.4	4.4
CZPS [18]	0.0	68.7	26.8	104.7	0.0	6.9	2.7	10.5
Proposed method	0.0	1.7	29.1	57.6	0.0	0.2	2.9	5.7

System \ Noise	10.0 [dB]			
	I	C	WI	Tr
SS [2]	2.5	2.9	0.8	0.6
VMAP [5]	2.6	3.0	0.7	0.6
CZPS [18]	0.0	0.7	0.3	1.1
Proposed method	0.0	0.02	0.3	0.6

I : impulsive noise C : clap noise WI : white and impulsive noise Tr : train noise

Table 4.3: Formal listening results of tunnel and clap noise suppression at 0dB.

Noise suppression system	Tunnel noise		Clap noise	
	MOS	Standard deviation	MOS	Standard deviation
SS	3.20	0.86	2.60	0.91
VMAP	4.40	0.83	2.06	0.70
CZPS	3.00	1.13	3.47	0.83
Proposed method	3.46	0.74	4.20	0.56

5

Conclusion

This thesis has described about single channel noise suppression based on speech and noise spectral models. This thesis consisted of two main parts. The first part has described about stationary noise suppression and the second part has described about impulsive noise suppression.

In Chapter 3, the author has proposed a stationary noise suppression algorithm using Maximum a Posteriori (MAP) estimation with a speech spectral amplitude probability density function (speech PDF). The estimated speech spectrum is given as a MAP solution which is obtained from the speech PDF. The speech PDF is hence the most important factor in this research. Since, the speech property can be considered as a time-variant function, the author assumed that the speech PDF changes according to SNR. Under this assumption, the author proposed adaptive shape parameters which were derived from real-speech PDFs in various narrow SNR intervals. The proposed adaptive shape parameters can pursue the change of the speech property, and give an appropriate MAP solution which is identical to the estimated speech spectrum. The effectiveness of the proposed method was examined and compared to the conventional algorithms. The simulation results have shown that the proposed method improved segmental SNR around 6 and 9 dB when the input speech signal was corrupted by white and tunnel noise signals at input SNR of 0 dB, respectively.

In Chapter 4, an impulsive noise suppression method has been investigated. This method utilizes a zero phase (ZP) signal which is defined as the IDFT of a spectral amplitude. In the impulsive noise suppression research, it was assumed that a speech signal has periodicity in a short observation, i.e., its spectral amplitude has values at

equally spaced frequencies. In this case, the corresponding ZP signal becomes also periodic. On the other hand, it was assumed that a noise spectral amplitude is approximately flat. In this case, its ZP signal takes nonzero values only around the origin. Actually, real impulsive-type noise has such property as shown in Section 4.1. Under these assumptions, the ZP signal of a speech signal embedded in impulsive noise in an analysis frame becomes a periodic signal except around the origin. The author has proposed the ZP signal replacement method which replaces the ZP signal around the origin with the ZP signal in the second or latter period. Then, speech ZP signal can be estimated. Taking DFT of it gives the estimated speech spectral amplitude. The IDFT of the estimated speech spectral amplitude with the observed spectral phase provides the estimated speech signal in time domain. The major advantage of this method is that it can suppress impulsive noise without a prior estimation of the noise spectral amplitude, while it is indispensable in most stationary noise suppression methods. Moreover, it has been shown that the proposed impulsive noise suppressor can also be available to suppress stationary wide-band noise. Simulation results showed that the proposed noise suppressor improved the SNR more than 5dB for stationary tunnel noise and 13dB for impulsive clap noise in a low SNR environment.

Reverberation of signals often exists in practical situations, and it degrades speech quality. However, suppression techniques of the reverberation have not been discussed in this thesis. Hence, suppressing the reverberation should be the future work in this study. In addition, speech spectral models discussed in this thesis are comparatively simple yet. In the future work, the author would like to more sophisticate these speech spectral models to improve speech quality of extracted signals.

Appendix A

Derivations

A.1 Derivation of MMSE-STSA

The MMSE-STSA method is derived by minimizing a conditional mean square value of the short time spectral amplitude. The cost function to be minimized is given by

$$\begin{aligned} J_{\mathrm{MMSE}} &= E\left[|S-\hat{S}|^2 \middle| X\right] \\ &= \int_{-\infty}^{\infty} |S|^2 p(S|X)ds + |\hat{S}|^2 - \hat{S}\int_{-\infty}^{\infty} S^* p(S|X)ds \\ &\quad -\hat{S}^*\int_{-\infty}^{\infty} Sp(S|X)ds, \end{aligned} \tag{A.1}$$

where $p(S|X)$ denotes the conditional PDF of S. The estimated speech spectrum which minimizes J_{MMSE} is given as

$$\hat{S}_{\mathrm{MMSE}} = \int_{-\infty}^{\infty} Sp(S|X)ds = E[S|X]. \tag{A.2}$$

As shown in [15], when we assume $p(S)$ and $p(D)$ as Gauss functions, (A.2) produces the Wiener filter again. On the other hand, Ephraim and Malah considered the PDFs of the speech spectral amplitude and phase, i.e., $p(|S|)$ and $p(\angle S)$. They assumed that $p(|S|)$ and $p(\angle S)$ as the Rayleigh distribution and the uniform distribution, respectively [36]. They assumed $p(D)$ as the Gauss function, where the noise variance σ_d^2 is assumed

to split equally into real and imaginary parts. These PDFs are expressed as

$$p(|S|) = \frac{2|S|}{\sigma_s^2} \exp\left\{-\frac{|S|^2}{\sigma_s^2}\right\}, \tag{A.3}$$

$$p(\angle S) = \frac{1}{2\pi}, \tag{A.4}$$

$$p(X|S) = \frac{1}{\pi\sigma_d^2} \exp\left\{-\frac{|X-S|^2}{\sigma_d^2}\right\}, \tag{A.5}$$

where $P(X|S)$ is corresponding to $p(D)$. Assuming $p(S) = p(|S|)p(\angle S)$, we can calculate (A.2) by using the relation $p(S|X) = p(X|S)p(S)/p(X)$. After tedious and complex computations, the spectral gain is given as [11]

$$G_{\text{MMSE}} = \frac{(\pi v)^{1/2}}{2\gamma} \exp\left(\frac{-v}{2}\right) \times \left[(1+v)I_0\left(\frac{v}{2}\right) + vI_1\left(\frac{v}{2}\right)\right], \tag{A.6}$$

where $I_i(\cdot)$ is the modified Bessel function of order i, and

$$v = \frac{\xi}{1+\xi}\gamma, \ \gamma = \frac{|X|^2}{\sigma_d^2}. \tag{A.7}$$

Here, γ is called as the *a posteriori* SNR. As shown in [11], the optimal spectral phase in the sense of MMSE-STSA is identical to the observed one. Hence, G_{MMSE} is also a real value.

A.2 Derivation of Spectral Amplitude Model

In the following, a simple statistical model for the speech spectral amplitude will be presented, which is closer to the real distribution than the commonly applied Gaussian model. Considering noise, the Gaussian assumption holds due to comparably low correlation in the analysis frame. Assuming statistical independence of real and imaginary parts, the PDF of the noise amplitude $|D_n(k)|$ can easily be found as Rayleigh distributed by polar integration [30],

$$p(|D|) = \frac{2|D|}{\lambda} \exp\left\{\frac{|D|^2}{\lambda}\right\}, \tag{A.8}$$

The real and imaginary part of the Fourier coefficients can be considered statistically independent with high accuracy. Then, $p(|S|)$ can in general be calculated by

$$p(|S|) = \int_0^{2\pi} |S| \cdot p(|S|\cos\phi) \cdot p(|S|\sin\phi) d\phi, \tag{A.9}$$

Considering Gaussian components, the rotational invariance greatly facilitates the polar integration. Similar to Eq.(A.8) the amplitude is Rayleigh distributed:

$$p(|S|) = \frac{2|S|}{\lambda} \exp\left\{-\frac{|S|^2}{\lambda}\right\} \tag{A.10}$$

Apparently, the slope of the Gamma amplitude PDF differs from that of the Laplace amplitude PDF. Hence, a parameter μ is introduced, which enables to approximate both. After normalizing $|S|$ by the standard deviation σ_S, we thus assume

$$p(|S|) \sim \exp\left\{-\mu\frac{|S|}{\sigma_S}\right\} \tag{A.11}$$

At low values of $|S|$, the PDF of the Laplace and Gamma amplitude is much higher than Rayleigh PDF. Considering the Rayleigh PDF according to Eq.(A.10), the behavior at low values is mainly due to the linear term of $|S|$, whereas the exponential term plays a minor role at small values.

Both the PDF of the Laplace amplitude and the PDF of the Gamma amplitude can be approximated by abandoning a linear term in $|S|$. Instead, $|S|$ is taken to the power of a parameter ν after normalization to the standard deviation of speech, i.e., $p(|S|) \sim (\frac{|S|}{\sigma_S})^\nu$ in order to be able to approximate a large variety of PDFs. The smaller the parameter ν, the more amplitude PDF distributed at low values. The term hardly influences the behavior of the function at high value due to the dominance of the exponential decay

$$p(|S|) \sim \frac{|S|^\nu}{\sigma_S^\nu} \exp\left\{-\mu\frac{|S|}{\sigma_S}\right\} \tag{A.12}$$

After taking $\int_0^\infty d|S| = 1$ into account, the approximating function with parameters ν, μ is finally obtained

$$p(|S|) = \frac{\mu^{\nu+1}}{\Gamma(\nu+1)} \frac{|S|^\nu}{\sigma_S^{\nu+1}} \exp\left\{-\mu\frac{|S|^2}{\sigma_S}\right\}. \tag{A.13}$$

Here, Γ denotes the Gamma function.

A.3 Derivation of Speech Spectral Gain (2.14)

For simplicity, the frame index n and frequency index k are omitted. Let $p(\cdot)$ denote the PDF (Probably Density Function). A joint MAP solution is given as

$$|\hat{S}| = \arg\max_{|S|} p(X||S|, \angle S)p(|S|, \angle S), \tag{A.14}$$

$$\angle\hat{S} = \arg\max_{\angle S} p(X||S|,\angle S)p(|S|,\angle S). \tag{A.15}$$

As proposed in [3], we put $p(X||S|,\angle S)$ and $p(\angle S)$ as

$$p(X||S|,\angle S) = \frac{1}{\pi\lambda}\exp\left(-\frac{|X-S|^2}{\lambda}\right), \tag{A.16}$$

$$p(\angle S) = \frac{1}{2\pi}. \tag{A.17}$$

Under the assumption that $p(|S|)$ is statistically independent with $p(\angle S)$, i.e., $p(|S|,\angle S) = p(|S|)p(\angle S)$, we have

$$p(X||S|,\angle S)p(|S|,\angle S)$$

$$= \frac{1}{2\pi^2\lambda}\exp\left(\frac{|X-S|^2}{\lambda}\right)\frac{\mu^{\nu+1}}{\Gamma(\nu+1)}\frac{|S|^\nu}{\sigma_S^{\nu+1}}\exp\left(-\mu\frac{|S|}{\sigma_S}\right). \tag{A.18}$$

Since the natural logarithm greatly facilitates the optimization of (A.14) or (A.15), we take the logarithm of (A.18) as

$$\ln p(X||S|,\angle S)p(|S|,\angle S)$$

$$= -\frac{|X|^2+|S|^2-X^*|S|e^{j\angle S}-X|S|e^{-j\angle S}}{\lambda}+\nu\ln|S|-\mu\frac{|S|}{\sigma_S}$$

$$+\ln\left(\frac{\mu^{\nu+1}}{2\pi^2\lambda\sigma_S^{\nu+1}\Gamma(\nu+1)}\right). \tag{A.19}$$

Differentiating (A.19) with respect to $\angle S$ and setting it to zero yield

$$e^{j(\angle S-\angle X)} - e^{j(\angle X-\angle S)} = 0. \tag{A.20}$$

Therefore, the estimation of $\angle S$ which maximizes $(A.19)$ is given by

$$\angle \hat{S} = \angle X. \tag{A.21}$$

This is the solution of (A.15). Then, differentiating (A.19) with respect to $|S|$, setting it to zero and replacing $\angle S$ with $\angle X$, we get

$$\begin{aligned} -\frac{2|S|}{\lambda} + \frac{\nu}{|S|} + \left(\frac{2|X|}{\lambda} - \frac{\mu}{\sigma_S}\right) &= 0, \\ |S|^2 - 2\left(\frac{1}{2} - \frac{\mu}{4\sqrt{\gamma\xi}}\right)|S||X| - \frac{\nu}{2\gamma}|X|^2 &= 0, \\ |S|^2 - 2u|S||X| - \frac{\nu}{2\gamma}|X|^2 &= 0, \end{aligned} \tag{A.22}$$

where u is defined in (2.14). Since the solution of $|S|$ is positive, the estimation of $|S|$ which maximize (A.19) is given by

$$|\hat{S}| = \left(u + \sqrt{u^2 + \frac{\nu}{2\gamma}}\right)|X|. \tag{A.23}$$

This is the solution of (A.14). Since $G = |\hat{S}|e^{j\angle\hat{S}}/X$, we have the spectral gain given by (2.14).

A.4 Derivation of Another Speech Spectral Gain Based on MAP Estimation

A computationally efficient MAP solution is given as

$$|\hat{S}| = \arg\max_{|S|} p(|S|||X|) = \arg\max_{|S|} \frac{p(|X||S|)p(|S|)}{p(|X|)}. \tag{A.24}$$

Now, the super-Gaussian function is used to model the PDF of the speech spectral amplitude $p(|S|)$. Then Gaussian assumption of noise allows to apply for $p(|X||S|)$. We need to maximize only $p(|X||S|)p(|S|)$, since $p(|X|)$ is independent of $|S|$. A closed form solution can be found if the modified Bessel function I_0 is considered asymptotically, with

$$I_0(s) \approx \frac{1}{\sqrt{2\pi x}} e^s. \tag{A.25}$$

After insertion of Eq.(A.25) into Eq.(2.10), we get for $p(|X|||S|)p(|S|)$:

$$p(|X|||S|)p(|S|) \sim |S|^{\nu-\frac{1}{2}} \exp\left\{-\frac{|S|^2}{\lambda} - A(\frac{\mu}{\sigma_S} - \frac{2|X|}{\lambda})\right\}. \tag{A.26}$$

Instead of differentiating $p(|X|||S|)p(|S|)$, the maximization can be performed better after applying the natural logarithm, because the product of the polynomial and exponential converts into a sum:

$$\frac{d\log\left[p(|X|||S|)p(|S|)\right]}{d|S|} = \left(\mu - \frac{1}{2}\right)\frac{1}{|S|} - \frac{2|S|}{\lambda} - \frac{\mu}{\sigma_S} + \frac{2|X|}{\lambda} \doteq 0. \tag{A.27}$$

After multiplication with $|S|$, one reasonable solution $|\hat{S}|=G \cdot |X|$ to the quadratic equation is found, because the second solution delivers spectral amplitudes $|S| < 0$ at least for $\nu > 0.5$. The second derivative at $|\hat{S}|$ is negative, thus a local maximum is guaranteed. The speech enhancement algorithm based on MAP estimation is as follows:

$$\hat{S} = G \cdot X, \tag{A.28}$$

$$G = \mu + \sqrt{\mu^2 + \frac{\nu - \frac{1}{2}}{2\gamma}}, \tag{A.29}$$

$$\mu = \frac{1}{2} - \frac{\mu}{4\sqrt{\gamma\xi}}. \tag{A.30}$$

A.5 Derivation of scaling function (4.6)

The segmented speech signal is given by

$$\tilde{s}(n) = s(n) \cdot h(n). \tag{A.31}$$

Under the assumption that the power of $|\varepsilon(n)|$ is small enough to be neglected in comparison to one of harmonic part in (4.2). Then, we can approximate a speech signal $s(n)$ as

$$s(n) \approx \sum_{m=1}^{\left\lfloor \frac{N}{2k_c} \right\rfloor} \alpha_m \cos(2\pi \frac{k_c}{N} mn + \theta_m). \tag{A.32}$$

We utilize the Hanning window function given as

$$h(n) = \frac{1}{2}\left\{1 - \cos\left(\frac{2\pi n}{N}\right)\right\}. \tag{A.33}$$

Then, the spectral amplitude of $\tilde{s}(n)$ is given by

$$\begin{aligned} |S(k)| &= \sum_{m=1}^{\left\lfloor \frac{N}{2k_c} \right\rfloor} \frac{\alpha_m}{2} \left\{ \frac{1}{2}\delta(k - mk_c + 1) + \delta(k + mk_c) \right. \\ &\quad + \frac{1}{2}\delta(k - mk_c - 1) + \frac{1}{2}\delta(k + mk_c - N + 1) \\ &\quad \left. + \delta(k + mk_c - N) + \frac{1}{2}\delta(k + mk_c - N - 1) \right\}. \end{aligned} \tag{A.34}$$

By substituting (A.34) into (2.24) with $\beta = 1$, we get

$$\begin{aligned} s_0(n) &= \sum_{m=1}^{\left\lfloor \frac{N}{2k_c} \right\rfloor} \frac{\alpha_m}{2} \left\{ \frac{1}{2}\cos\frac{2\pi(mk_c - 1)}{N}n \right. \\ &\quad \left. + \cos\frac{2\pi(mk_c)}{N}n + \frac{1}{2}\cos\frac{2\pi(mk_c + 1)}{N}n \right\} \\ &= \left(1 + \cos\frac{2\pi}{N}n\right) \cdot \sum_{m=1}^{\left\lfloor \frac{N}{2k_c} \right\rfloor} \frac{\alpha_m}{N}\cos\frac{2\pi mk_c}{N}n. \end{aligned} \tag{A.35}$$

The scaling function for $s_0(n + T)$ is given as

$$\begin{aligned} sc(n) &= \frac{s_0(n)}{s_0(n + T)} \\ &= \frac{\left(1 + \cos\frac{2\pi}{N}n\right) \cdot \sum_{m=1}^{\left\lfloor \frac{N}{2k_c} \right\rfloor} \frac{\alpha_m}{N}\cos\frac{2\pi mk_c}{N}n}{\left\{1 + \cos\frac{2\pi}{N}(n+T)\right\} \cdot \sum_{m=1}^{\left\lfloor \frac{N}{2k_c} \right\rfloor} \frac{\alpha_m}{N}\cos\frac{2\pi mk_c}{N}(n+T)}. \end{aligned} \tag{A.36}$$

Using the following relation

$$\cos\frac{2\pi mk_c}{N}(n + T) = \cos\frac{2\pi mk_c}{N}n, \tag{A.37}$$

we have (4.6).

References

[1] T.H. Dat, K. Takeda, and F. Itakura, "Multichannel speech enhancement based on speech spectral magnitude estimation using generalized Gamma prior distribution," IEEE International Conference on Acoustics, Speech and Signal, ICASSP 2006 Processing, Vol.4, page IV, May 2006.

[2] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoustics, Speech, and Signal Processing, Vol. ASSP-27, No. 2, pp. 113–120, Apr. 1979.

[3] T. Lotter, and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model," EURASIP Journal on Applied Signal Processing, Vol. 7, pp. 1110–1126, Sept. 2005.

[4] I. Andrianakis, and P.R. White, "Speech spectral amplitude estimators using optimally shaped Gamma and Chi priors," Speech Communication, Vol.51, Issue 1, Jan. 2009.

[5] Y. Tsukamoto, A. Kawamura, and Y. Iiguni, "Speech enhancement based on MAP estimation using a variable speech distribution," IEICE Trans. Fundamentals, Vol.E90-A, No.8, pp.1587-1593, Aug. 2007.

[6] M. Kato, A. Sugiyama, and M. Serizawa, "Noise suppression with high speech quality based on weighted noise estimation and MMSE-STSA," IEICE Trans, Fundamentals, Vol.E85-A, No.7, pp.1710-1718, Jul. 2002.

[7] 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.

[8] P. Vary, and R. Martin, Digital Speech Transmission: Enhancement, Coding and Error Concealment, Wiley, 2005.

[9] M. Muneyasu and A. Taguchi, Nonlinear digital signal processing, Asakura Publishing Company, Tokyo, 1999.

[10] A. Kawamura, Y. Iiguni and Y. Itoh, "A noise reduction method based on linear prediction with variable step-size," IEICE Trans. Fundamentals, Vol.E88-A, No.4, pp.855–861, April 2005.

[11] 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.

[12] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," Proceedings of the IEEE, Vol.67, No. 12, pp. 1586-1604, 1979.

[13] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," Proc. 4th IEEE Int. Conf. Acoust. Speech Signal Peocess., ICASSP-79, Vol.4, pp. 208–211, 1979.

[14] P. Scalart and J. V. Filho, "Speech enhancement based on a priori signal to noise estimation," in Proc. 21st IEEE Int. Conf. Acoust. Speech Signal Process., ICASSP-96, Vol.2, pp. 629-632, 1996.

[15] P.J. Wolfe and S.J. Godsill, "Efficient alternatives to the Ephraim and Malah suppression rule for audio signal enhancement," EURASIP Journal on Applied Signal Processing, Vol.10, pp.1043–1051, Oct. 2003.

[16] A. Kawamura, W. Thanhikam, and Y. Iiguni, "A speech spectral estimator using adaptive speech probability density function," Proc. of EUSIPCO 2010, pp.1549–1552, Aug. 2010.

[17] W. Thanhikam, A. Kawamura and Y. Iiguni, "Speech enhancement using speech model parameters refined by two-step technique" Proc. of the Second APSIPA Annual Summit and Conference, p.11, Dec. 2010.

[18] Y. Kamamori, A. Kawamura, and Y. Iiguni, "Zero phase signal analysis and its application to noise reduction," IEICE Trans. Fundamentals, Vol.J93-A, No.10, pp.658–666, Oct. 2010.

[19] D.W. Griffin and J.S. Lim. "Multiband-excitation vocoder," IEEE Trans. Acoust., Speech, Signal Processing, ASSP-36, pp. 1223–1235, Aug. 1988.

[20] J. Laroche, Y. Stylianou, and E. Moulines, "HNS: Speech modification based on a harmonic + noise model" in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing '93, Minneapolis, MN, pp. 550–553, Apr. 1993.

[21] "SLDB, Natural Speech and Language Database," ATR-Promotions, http://www.atr-p.com

[22] S. Furui, Digital speech processing, Tokai University Press, Tokyo, 1985.

[23] "Signal processing information base," http://spib.rice.edu/

[24] "Noise database," Sunrise Music Inc., http://www.sunrisemusic.co.jp/database/fl/noisedata01_ fl. html

[25] L. Rabiner, M. Cheng, A. Rosenberg, and C. McGonegal, "A comparative performance study of several pitch detection algorithms," IEEE Trans. on Acoust., Speech, Signal Processing, Vol.24, No. 5, pp.399–418, 1976.

[26] B. Widrow, J.G.R. Glover, Jr., J.M. Mccool, J. Kaunitz, C.S. Williams, R. H. Hearin, J.R. Zeidler, E. Dong, Jr., and R.C. Goodlin, "Adaptive noise cancelling: Principles and applications," Proceedings of The IEEE, Vol. 63, No. 12, pp. 1692–1719, 1975.

[27] R. Martin, "Speech enhancement based on minimum mean-square error estimation and super-Gaussian priors," IEEE Trans. Speech and Audio Processing, Vol. 13, No. 5, pp. 845–856, 2005.

[28] R. Martin, "Spectral subtraction based on minimum statistics," EU-RIPCO'94, pp.1182-1185, Sept. 1994.

[29] P. Vary and R. Martin, "Digital Speech Transmission: Enhancement, Coding and Error Concealment, 2nd Ed." John Wiley & Sons, Ltd., 2007.

REFERENCES

[30] P. Vary, "Noise suppression by spectral magnitude estimation - Mechanisms and theoretical limits," Signal Processing, vol. 8, pp. 38-400, 1985.

[31] S. Kullback, Information Theory and Statistics, Dover Publication, 1968.

[32] David Graff, Kevin Walker, and David Miller, Switchboard Cellular Part 1 Audio, Linguistic Data Consortium, Philadelphia, 2001.

[33] N.R. Draper and H. Smith, Applied Regression Analysis, 3rd Ed., John Wiley & Son, New York, 1998.

[34] D.R. Brillinger, Time Series: Data Analysis and Theory, Holden-Day, 1981.

[35] A. Varga and H.J.M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," Speech Commun., Vol.12, No.3, pp.504-512, Jul. 2001.

[36] S.L. Miller and D.G. Childers, Probability and Random Processes, *Elsevier Academic Press*, 2004.

List of Publications

Journal Papers

1. Weerawut Thanhikam, Arata Kawamura, and Youji Iiguni, "Speech enhancement based on real-speech PDF in various narrow SNR intervals", IEICE Trans. Fundamentals, Vol.E95-A, No.3, pp.623-630, Mar. 2012.

2. Weerawut Thanhikam, Arata Kawamura, and Youji Iiguni, "Stationary and non-stationary wide-band noise reduction using zero phase signal", IEICE Trans. Fundamentals, Vol.E95-A, No.5, pp.843-852, May 2012.

3. Arata Kawamura, Weerawut Thanhikam, and Youji Iiguni, "Single Channel Speech Enhancement Techniques in Spectral Domain", ISRN Mechanical Engineering, Vol.2012, Article ID 919234, July 2012.

International Conferences

1. Weerawut Thanhikam, Arata Kawamura and Youji Iiguni, "Speech enhancement using speech model parameters refined by two-step technique", Proc. of the Second APSIPA Annual Summit and Conference (Student Symposium), p.11, Dec. 2010.

2. Arata Kawamura, Weerawut Thanhikam, and Youji Iiguni, "A speech spectral estimator using adaptive speech probability density function", Proc. of 18th European Signal Processing Conference (EUSIPCO-2010), pp.1549-1552, Aug. 2010.

3. Weerawut Thanhikam, Arata Kawamura, and Youji Iiguni, "Noise suppression based on replacement of zero phase signal", Proc. of ISPACS2011, PID242, Dec. 2011.

4. Weerawut Thanhikam, Arata Kawamura, and Youji Iiguni, "A speech enhancement method using adaptive speech PDF", Proc. of ISPACS2011, PID134, Dec. 2011.

5. Weerawut Thanhikam, Sayuri Kohmura, Arata Kawamura, and Youji Iiguni, "An impulsive noise suppressor based on zero phase signal," Proc. of ITC-CSCC2012, paper No. D-T2-02, July 2012.

Domestic Conferences

1. Ararta Kawamura, Weerawut Thanhikam, and Youji Iiguni, "A speech enhancement method using spectral probability density function of non-speech segments,", IEICE Tech. Report, SIP2008-101, pp.65-70, Sep. 2008.

2. Weerawut Thanhikam, Arata Kawamura, and Youji Iiguni, "Speech enhancement algorithm based on variable speech spectral amplitude distribution", IEICE Tech. Report, SIP2008-153, pp.173-176, Jan. 2009.

3. Ararta Kawamura, Weerawut Thanhikam, and Youji Iiguni, "Noise suppression using speech periodicity in zero phase domain", Acoustical Society of Japan (ASJ), 2010 Autumn Meeing, pp.709-710, Sep. 2010.

4. Weerawut Thanhikam, Arata Kawamura, and Youji Iiguni, "Noise suppression in zero phase domain", Proc. of the 2011 IEICE General Conference, p.79, March 2011.

5. Ararta Kawamura, Weerawut Thanhikam, and Youji Iiguni, "Replaced zero phase signal method for noise suppression", ASJ, 2011 Autumn Meeing, pp.777-778, Sep. 2011.