เทคนิคการลดสัญญาณรบกวนในโดเมนความถี่สำหรับเครื่องช่วยฟังแบบดิจิทัล

นางสาว ธิริ ธันดา ออง

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2554

THE FREQUENCY DOMAIN NOISE REDUCTION TECHNIQUES FOR DIGITAL HEARING AIDS

Ms. Thiri Thandar Aung

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Electrical Engineering
Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2011

Thesis Title THE FREQUENCY DOMAIN NOISE REDUCTION TECHNIQUES FOR DIGITAL HEARING AIDS
By Ms. Thiri Thandar Aung
Field of Study Electrical Engineering
Thesis Advisor Nisachon Tangsangiumvisai, Ph.D.

Accepted by the Faculty of Engineering, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

……………………………………………Dean of the Faculty of Engineering
(Associate Professor Boonsom Lerdhirunwong, Dr.Ing.)

THESIS COMMITTEE

…………………………………………… Chairman
(Associate Professor Chedsada Chinrungrueng)

…………………………………………… Thesis Advisor
(Associate Professor Nisachon Tangsangiumvisai)

…………………………………………… Thesis Co-advisor
(Professor Akinori Nishihara)

…………………………………………… Examiner
(Assistant Professor Supavadee Aramvith)

…………………………………………… External Examiner

(Phaophak Sirisuk, Ph.D.)

ธิริ ธันดาออง : เทคนิคการลดสัญญาณรบกวนในโดเมนความถี่สำหรับเครื่องช่วยฟังแบบดิจิทัล. (THE FREQUENCY DOMAIN NOISE REDUCTION TECHNIQUES FOR DIGITAL HEARING AIDS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ.ดร.นิศาชล ตั้งเสงี่ยมวิสัย, ที่ปรึกษาวิทยานิพนธ์ร่วม: ศ.ดร. อากิโนริ นิชิฮาร่า 83 หน้า

ประสิทธิภาพของเครื่องช่วยฟังแบบดิจิทัลอาจถูกทำให้บิดเพี้ยน จากสัญญาณรบกวนบางประเภท ได้แก่ สัญญาณเสียงป้อนกลับแบบอะคูสติก สัญญาณแทรกสอด และสัญญาณรบกวนพื้นหลัง เป็นต้น วิทยานิพนธ์ฉบับนี้พิจารณาเทคนิคการลดเสียงรบกวนพื้นหลังแบบบวกสำหรับการใช้งานในเครื่องช่วยฟังแบบดิจิทัล ทั้งนี้ เทคนิคการลดเสียงรบกวนที่นำเสนอในวิทยานิพนธ์ฉบับนี้แสดงให้เห็นถึงแนวคิดที่จะเพิ่มประสิทธิภาพการทำงานของเครื่องช่วยฟังแบบดิจิทัลด้วยการลดเสียงรบกวนพื้นหลัง โดยใช้โครงสร้างไฮบริดระหว่างเทคนิคการลดเสียงรบกวนสองวิธีที่มีอยู่แล้ว คือ เทคนิคการลบสเปกตรัม (SS) ที่มีการใช้งานของการตัดออกความถี่ข้างที่มีลักษณะทั่วไป (GSC) แบบสองช่องสัญญาณ โดยจะเรียกว่า 2chGSC+SS ซึ่งถูกเสนอให้นำมาใช้กับโครงสร้างไฮบริดสำหรับการประมวลผลสัญญาณในช่วงความถี่ต่ำที่มีองค์ประกอบทางความถี่โดยส่วนมากของสัญญาณเสียงพูดอยู่ สำหรับการประมวลผลสัญญาณในช่วงความถี่สูงนั้น เทคนิคการลดเสียงรบกวนแบบสองช่องสัญญาณที่ใช้วิธีการลบสเปกตรัมข้ามแบบปรับปรุง(CSS)ถูกเสนอให้นำมาใช้งาน โดยจะเรียกว่า 2chCSS ผลการจำลองระบบได้ถูกศึกษาทั้งในสภาวะแวดล้อมแบบสัญญาณรบกวนแหล่งเดียว และแบบสัญญาณรบกวนหลายแหล่ง โดยที่ประสิทธิภาพของเทคนิคการลดเสียงรบกวนที่นำเสนอทั้งเชิงวัตถุวิสัยและเชิงจิตวิสัย แสดงให้เห็นถึงศักยภาพของวิธีที่นำเสนอ

นอกจากนี้ วิธีการปรับพารามิเตอร์ที่ใช้ในการลบสเปกตรัมของเสียงรบกวนให้แปรผันได้และขึ้นอยู่กับความถี่ถูกนำเสนอสองวิธี เพื่อใช้ในการควบคุมประสิทธิภาพการลดสัญญาณรบกวนและการก่อให้เกิดความบิดเพี้ยนของสัญญาณเสียงพูด สำหรับเทคนิคการลดสัญญาณรบกวนแบบ SS ผลการจำลองระบบแสดงให้เห็นว่า วิธีการปรับพารามิเตอร์ที่เสนอสามารถทำให้เทคนิคการลดสัญญาณรบกวนแบบ SS ลดผลกระทบของสัญญาณรบกวนพื้นหลังลงได้ และยังคงสามารถรักษาคุณภาพของสัญญาณเสียงพูด ได้ดีกว่าวิธีการลดสัญญาณรบกวนอื่นๆ ที่ถูกนำมาพิจารณาในวิทยานิพนธ์นี้

ภาควิชา วิศวกรรมไฟฟ้า ลายมือชื่อนิสิต

สาขาวิชา วิศวกรรมไฟฟ้า ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก

ปีการศึกษา 2554

5171864021: MAJOR ELECTRICAL ENGINEERING
KEYWORDS :DIGITAL HEARING AIDS/ NOISE REDUCTION/ SPEECH ENHANCEMENT

THIRI THANDAR AUNG: THE FREQUENCY DOMAIN NOISE REDUCTION TECHNIQUES FOR DIGITAL HEARING AIDS. ADVISOR: ASSOC.PROF.NISACHON TANGSANGIUMVISAI, Ph.D, CO-ADVISOR: PROF.AKINORI NISHIHARA, 83PP.

The overall performance of digital hearing aids can be degraded by some disturbing sources such as acoustic feedback signal, interfering signal, and background noise, etc. This dissertation focuses on the Noise Reduction (NR) techniques, especially for additive background noise in digital hearing aids. A proposed NR technique demonstrates an improved scheme based on the hybrid structure between two existing NR techniques, to reduce the noise efficiently, while preserving the speech spectral components. The Spectral Subtraction (SS) method that employs the two-channel Generalized Sidelobe Canceller (2chGSC+SS) is suggested to be employed by the proposed technique for processing the signal in the low frequency region where the dominant cues of the speech signals locate. For the high frequency region, the two-sensor NR technique based on the modified Cross-Spectral Subtraction (2chCSS) is employed. Simulation results in the one-noise-source and multiple-noise-source environments are undertaken, while objective and subjective performance measures are used to confirm the potential of the proposed method.

In addition, two approaches are proposed in order to control the noise subtraction parameter to be variable and frequency-dependent so that the amount of noise reduction and speech distortion can be managed for the SS-based NR methods. Simulation results are demonstrated where the SS method that employs these proposed variable parameters can sufficiently reduce the additive background noise while the ability to preserve speech quality is superior to the other investigated NR techniques.

Department : Electrical Engineering	Student's Signature ……………………
Field of Study: Electrical Engineering	Advisor's Signature ……………………
Academic Year: 2011	

Acknowledgements

The work contained in this thesis represents the accumulation of three years of my work in Thailand. Firstly, I would like to express my deep gratitude to my advisor Dr. Nisachon Tangsangiumvisai. It was a great privilege and honor to work and study under her guidance. I would like to thank her for her supports, friendship, empathy and great vision throughout the course and during the thesis period. And I would also like to thank my co-advisor Professor Akinori Nishihara for his valuable guidance.

In addition, I would like to thank the thesis committee members for their comments and recommendations.

My true gratitude is given to my scholarship donor, AUN/SEED-NET for their financial support. This research and studying at Chulalongkorn University would not be completed without this fellowship.

I am extremely grateful to my parents, U Than Htay Aung and Daw Tin Tin Aye, for their simultaneous supports, encouragements and educating and preparing me for my future life since I was a kid. This work is dedicated to them.

CONTENTS

Chapter Page

LIST OF TABLES

`

Page

Page

Page

LIST OF FIGURES

Page

Page

Page

CHAPTER I

INTRODUCTION

Hearing aids are small electronic devices that are used to make sounds louder so that hearing-impaired people can listen, communicate and participate more fully in daily activities. For example, hearing-impaired people may loss high-frequency information because the high-frequency components of the speech are weaker than the low-frequency ones. Therefore, hearing aids are used to amplify the frequencies where the speech has the weakest component and the hearing loss is the greatest.

There are different types of hearing aids depending on where they are worn: behind-the-ear, in-the-ear, in-the-canal and completely-in-the-canal as shown in Fig.1-1. Usually they have three basic parts: microphone, amplifier and speaker as shown in Fig.1-2 [1]

(a) (b) (c) (d)

Fig.1-1. Styles of hearing aids (a) behind-the-ear with earmold, (b) behind-the-ear with small tubing, (c) in-the-canal and (d) completely-in-the-canal

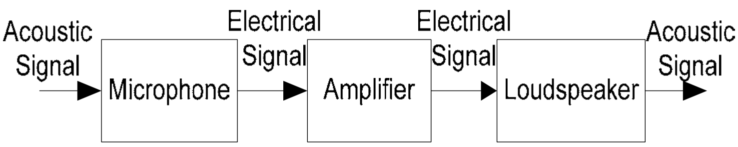


Fig.1-2. The process flow of analog hearing aids

1.1. The problem of background noise in Hearing Aids

There are two main types of hearing aids: analog and digital hearing aids [2]. Analog hearing aids convert sound waves into electrical signals, which are then amplified. Analog hearing aids have more than one program or setting for different listening environments; for example, for quiet rooms or outdoor, etc. Digital hearing aids, on the other hand, encode the sound waves before amplifying them. It is more flexible to adjust or program the digital hearing aids for different listening environments than the analog circuitry [2] However, digital hearing aids are normally more expensive than analog ones.

The overall performance of hearing aids is, however, degraded by some disturbing sources such as acoustic feedback signal, interfering signal and background noise, etc. Their effect is perceived as unpleasant, disturbing and interfering. Noise Reduction (NR) is, therefore, one of the most crucial functions in hearing aids, as hearing impaired people have great difficulty in understanding speech in noisy environments. NR is one means for Speech Enhancement (SE) i.e. several SE techniques aim to improve the quality of degraded speech signals by using signal processing tools [3] There are several noise reduction techniques which are normally employed to eliminate the background noise signal and to improve speech intelligibility [4] When the speech signal is very noisy, the hearing aids users tend to concentrate less after a long period of time. The speech intelligibility is thus reduced. The noise reduction algorithms can be categorized into two groups depending on the number of microphone channels; i.e. single-channel and multi-channel techniques.

The single-channel NR techniques have many interesting properties. For example, they can be integrated into most existing communication devices without requiring architectural changes, and their implementation cost is low. One example of the single-channel NR techniques that operates in the time-domain is the so-called Adaptive Noise Cancellation (ANC) [5]. On the other hand, various methods of the single-channel NR techniques have been proposed in the frequency-domain, such as the spectral subtraction (SS) technique [6]-[8] and various gain functions has been proposed to enhance the single-channel NR techniques [9],[10]. However, speech distortion is unavoidable when using the single-channel NR techniques that operate in the frequency domain, especially the SS technique. The amount of speech distortion is proportional to the amount of noise reduction; i.e. the more the background noise signal is reduced, the larger amount of speech distortion is obtained [10] and [11]. This is due to the fixed choice of the noise subtraction parameter of the SS method. When the noise subtraction parameter is chosen to be large in order to obtain a high amount of noise reduction, it will, however, introduce a large amount of speech distortion. On the other hand, by reducing the noise subtraction parameter to preserve the speech quality, the additive noise signal cannot be sufficiently eliminated. In order to control the amount of speech distortion while achieving sufficient level of noise reduction, the noise subtraction parameter of the SS method is proposed to be variable and frequency-dependent, instead of being fixed [11]. Alternatively, another approach is suggested in [48] to adjust the value of noise subtraction parameter as a function of the SNR at each particular frequency. When the speech spectral components are more dominant than the noise spectral ones, i.e. high SNR level, a large value of noise subtraction parameter should be chosen so that high noise subtraction is obtained. On the other hand, the value of noise subtraction parameter should be very small when the spectral level of noise is high, i.e. low SNR level, in order to preserve the speech spectral components.

The effect of speech distortion can be perceived by the hearing aid users. These randomly isolated spectral components of noise are referred to in the literature as ‘musical noise’, which is acoustically or electronically generated and randomly produced electronic signals [12].

One of the key factors for the noise reduction performance of the SS technique is the accuracy in estimating the noise spectrum. A better estimate of the noise spectral components can be obtained by the use of multiple microphones. Compared with single-channel NR techniques, multi-channel NR techniques have demonstrated great potential in suppressing the background noise signals and enhancing speech quality [13]-[15]. For this case, each microphone output can be modeled as the source speech signal convolved with the corresponding acoustic channel impulse response and then corrupted by background noise [13]. The aim of the multi-channel NR techniques is to estimate the desired speech signal from the multiple microphone observations [13]. In fact, the gain performance of multi-channel NR techniques is shown to be better than the single-channel NR ones due to the spatial filtering to suppress the interfering signals [16].

Generalized sidelobe canceller (GSC) is one of the various approaches, that is employed together with the NR techniques in order to eliminate or remove the interfering noise. Adaptive beamforming techniques are used to suppress the noise whose angle of arrival is different from the desired speech signal [17]. The example applications that apply GSC for the purpose of noise reduction are Acoustic Feedback Cancellation in hearing aids etc. [18]. There are two main beamforming processing; time-domain processing and frequency-domain processing. The wider the beamwidth, the poorer resolution can be obtained [19]. Therefore, the subband GSC for broadband beamforming has been proposed in [4]. In this method, the received broadband signals can be divided into many subbands and the beamformer is applied to each subband to increase the frequency resolution. In addition, the computational complexity can be reduced due to subband processing because the number of filter coefficients in each subband is less than those in the fullband case.

Usually, the GSC method contains three parts: fixed beamforming which is used for summing up all the multipath signals to get maximum signal-to-noise ratio (SNR); cancellation of speech signal, which is used to cancel the speech signal to produce the noise-only reference signal; and multichannel adaptive filter, which attempts to further remove the noise signal in tche fixed beamformer output [20]. Recently, two-channel noise reduction methods to produce the binaural outputs have been proposed to obtain better hearing performance by means of binaural benefits [4],[21],[39]. It is investigated in this dissertation a number of two-channel NR techniques; the SS method that employs the two-channel Generalized Sidelobe Canceller (2chGSC+SS)[4],[39], the two-sensor NR technique based on the modified Cross-Spectral Subtraction (2chCSS) [21]. Then, a hybrid structure between the 2chGSC+SS and the 2chCSS techniques is proposed for noise reduction in digital hearing-aid applications. This is due to the fact that the dominant cues of the speech signals are usually located in the frequency range below 1 kHz [22]. It is therefore suggested that there should be minimum distortion in this frequency region below 1 kHz so as not to degrade the quality of the enhanced speech signal. From the preliminary tests, we found that the 2chGSC+SS technique outperforms the 2chCSS method in terms of the ability to preserve speech quality, whereas the 2chCSS technique can remove the musical noise adequately. Hence, it is proposed that the 2chGSC+SS technique should be employed for processing the signal in the low-frequency region. As for the frequency region above 1 kHz, the 2chCSS method is employed. Comparison among these two-channel NR techniques and the proposed hybrid structure demonstrates improved performance of the two-channel NR techniques, as compared to the single-channel NR one.

1.2. Objectives

To propose a noise reduction technique for digital hearing aids in order to obtain efficient Noise Attenuation (NA) performance while preserving the speech quality of the original speech signal, as compared to the conventional NR techniques.

1.3. Scope of the Research

This dissertation focuses on the NR techniques for digital hearing aids that operate in the frequency domain.

1.4. Expected Outcome

- An improved NR technique for speech enhancement in digital hearing aids, as compared to some existing NR techniques.

1.5. Research Procedure

- Review relevant literature about NR techniques, especially in digital hearing aid applications
- MATLAB programming for simulation
- Comparison of the investigated algorithms based on objective measurement
- Propose a NR technique
- Simulation to confirm the proposed idea

1.6. Outline of Thesis

This thesis is organized as follows: The digital hearing aids are reviewed in the Chapter 2. In Chapter 3, the various noise reduction (NR) algorithms are described. The proposed NR technique, employing the frequency-selective noise subtraction parameter is presented in Chapter 4. The experiments and simulation results are given in Chapter 5, followed by the conclusions in Chapter 6.

CHAPTER 2
DIGITAL HEARING AIDS

2.1. Why Hearing Aids are needed?

Hearing-impaired people cannot hear some sounds at all. It is because the essential parts of some phonemes are inaudible [27]. The speech components in the high frequency region are normally weaker than that in the low frequency region. Therefore, hearing-impaired people mostly loss the high frequency information [27].

2.1.1. Causes of hearing loss

The human ears contain many parts, all of which need to work properly to hear sounds. A hearing loss occur if there is a problem at any point in the hearing pathway – in the outer, middle or inner ears, or in the complex auditory nerve pathway up to the brain. The structure of the human ear is shown in Fig.2-1. [26]

There are two main types of hearing loss depending on what part of the loss is not working properly [25]. If there is something not working in the ear canal, eardrum, middle ear bones or middle ear space, it is called the conductive hearing loss.

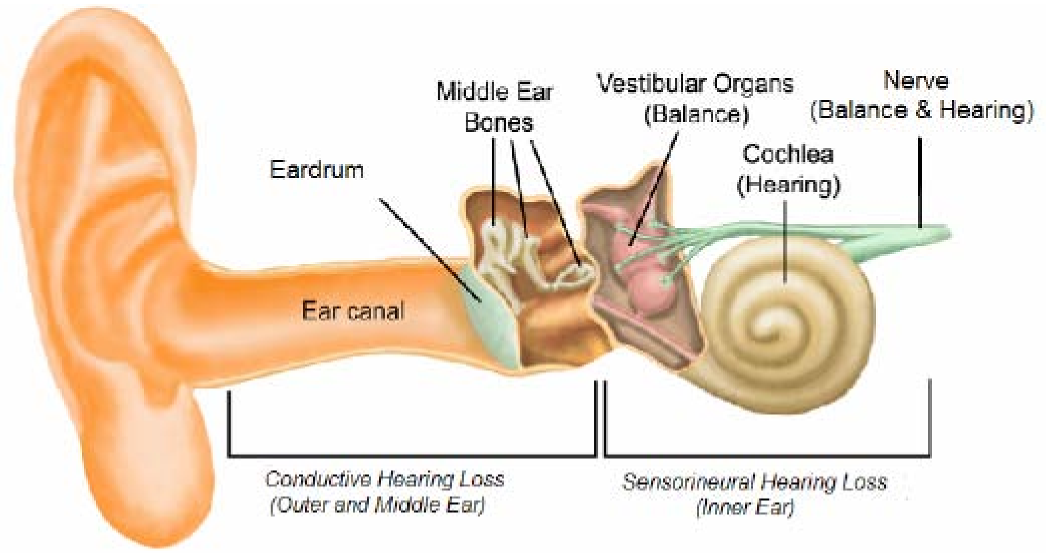


Fig.2-1. The structure of the human ear [26]

On the other hand, if there is something not working in the cochlea, auditory nerve or brain, it is called the sensorineural hearing loss.

2.1.1.1. Conductive hearing loss

This type of hearing loss is the loss of loudness. This is caused by the blockage or damage in the outer or middle ear. Depending on the cause of the hearing loss, hearing loss may be temporary or permanent. Conductive hearing loss [25] is almost the temporary hearing loss. People with a conductive hearing loss have a normal inner ear and they can hear their own voice louder than the voices of others. This kind of hearing loss often be helped by medical or surgical treatment.

2.1.1.2. Sensorineural hearing loss

Sensorineural hearing loss [25] is not only a loss of the loudness but also a loss of the clarity as well. It is usually the permanent hearing loss. People with a sensorineural hearing loss speak loudly, generally have more trouble understanding speech, and are especially bothered trying to understand speech in the presence of the background noise. The lack of clarity that may be associated with a sensorineural hearing loss is not completely possible by amplifying sounds.

To overcome this difficulty, hearing aids are needed to use in order to amplify for frequencies where speech has the weakest components and where hearing loss is the greatest. Hearing aids can provide different amount of gains in different frequency regions.

2.2. Digital Hearing Aids

As mentioned in chapter 1, there are two types of hearing aids; analog and digital ones. The analog hearing aids are typically designed with filters having fixed characteristics whereas the digital hearing aids are designed with the filters in which the given characteristics can be varied upon the incoming signals [24].

In digital hearing aids, the microphone picks up sounds and converts it to electrical signal [27]. An analog-to-digital converter (ADC) converts these electrical signals into binary numbers. The digital signal processor performs arithmetic operations on these numbers to get the original signal with minimal distortion resulting in excellent sound quality as shown in Fig.2-2.

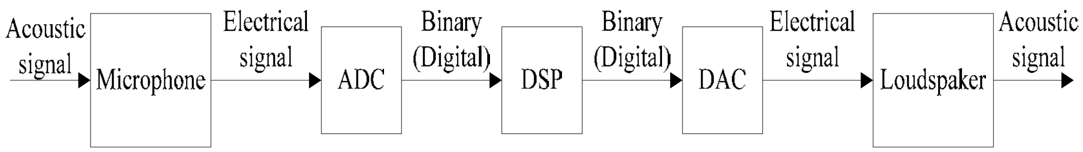


Fig.2-2. Digital Hearing Aids

Digital hearing aids amplify sounds depending on the frequency. This is due to the hearing-loss characteristics. In this dissertation, we focus on the noise reduction techniques in order to improve the digital hearing-aid performance. The input signal can be divided into several frequency bands where each band can be differently amplified and compressed. The first stage of digital signal processor is frame processing or windowing the input signal. The frame size or frame length is often chosen to be 64, 128, 256, 512 or 1024. The greater the number of input samples, the more individual frequencies can be classified. The second stage is transforming the signal into the frequency domain by operating fast Fourier transform (FFT) algorithm. Then noise reduction can be operated by employing the frequency-domain NR techniques, before converting back to the time-domain [27]. A block diagram of noise reduction in the frequency domain can be shown in Fig.2-3.

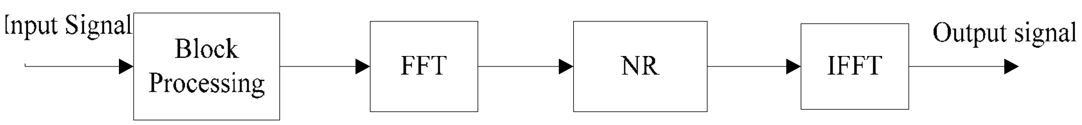


Fig.2-3. The process flow of noise reduction in the frequency domain

2.3. Summary

Hearing-impaired people use digital hearing aids so that they can hear the sound better. However, the performance of digital hearing aids is degraded by some interfering noise or background noise such as the sound from air-con, the car noise etc. Therefore, Noise Reduction (NR) is essential in digital hearing aids. The various types of NR algorithms will be described in the next Chapter.

CHAPTER 3

LITERATURE REVIEW AND RELATED WORKS IN NOISE REDUCTION TECHNIQUES

The presence of background noise such as breezing, alarms, other people talking, etc, can decrease speech intelligibility and interfere with the conversation [23]. NR is therefore required to reduce the background noise for the hearing aids applications so that the hearing-impaired people can hear sounds better. NR algorithms depend on the number of microphone channels, and the use of directional microphone can help not only to get a better range of hearing but also to improve noise reduction performance [24]. Therefore, we can divide the NR algorithms into single-channel and multi-channel NR techniques, as shown in Fig.3-1.

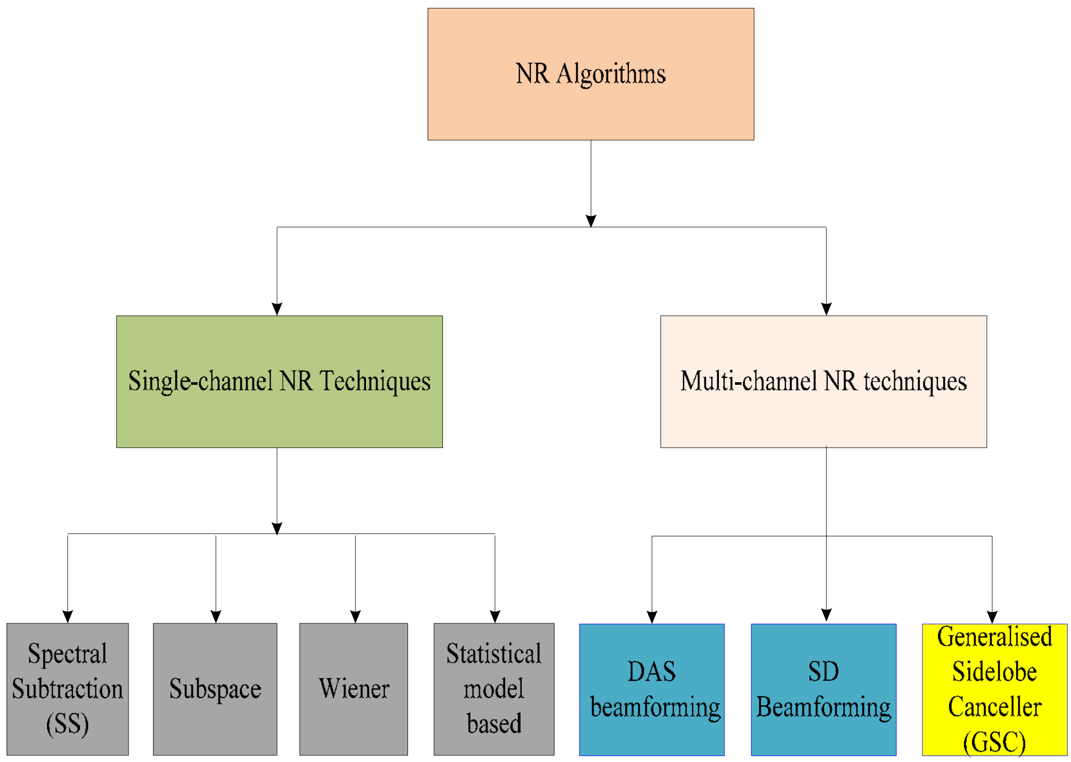


Fig.3-1. Categories of NR techniques

In the following section, we will explain in details those single-channel and multi-channel NR techniques.

3.1. Single-channel Noise Reduction Techniques

In this section, the SS method, the subspace method, the Wiener method will be described, followed by the statistical-model-based MMSE algorithm.

3.1.1. Spectral Subtraction (SS) method

For a single-channel NR technique, the observed microphone signal is modeled as a sum of the clean speech , $s(n)$, and the additive background noise, $b(n)$, as expressed by

$$x(n) = s(n) + b(n) \tag{3.1}$$

where n is the discrete-time index. By assuming that $s(n)$ and $b(n)$ are uncorrelated to each other, the noisy signal can be analyzed in the frequency domain as

$$X(\omega) = S(\omega) + B(\omega) \tag{3.2}$$

Due to non-stationary characteristics of speech signal, the speech analysis is normally applied to a short frame of speech signal [6]. Short-time Fourier Transform (STFT) is applied to the analysis spectrum so that noise reduction can be performed in each small subband as given by

$$X(k,l) = S(k,l) + B(k,l) \tag{3.3}$$

where $l = 1, 2, \ldots$ is the analysis block and the index $k = 1, 2, \ldots, N$ represents frequency bin or frequency components of the analyzed spectrum while employing N-point Short-Time Fourier Transform (STFT).

The estimated enhanced speech spectrum, $\hat{S}(k,l)$, is obtained as shown in Fig.3-2, and is given by

$$\left|\hat{S}(k,l)\right| = \begin{cases} \left|X(k,l)\right|^p - \alpha \left|\hat{B}(k,l)\right|^p, & \left|X(k,l)\right|^p \geq \left|\hat{B}(k,l)\right|^p \\ \beta \left|X(k,l)\right|^p, & \text{otherwise} \end{cases} \tag{3.4}$$

where $E\{\cdot\}$ is the expectation operator, p is the exponent and α controls the amount of subtraction β is the spectral flooring factor. When $p = 1$, it is known as the Magnitude SS (MSS) while $p = 2$, it will be referred to as the Power SS (PSS).

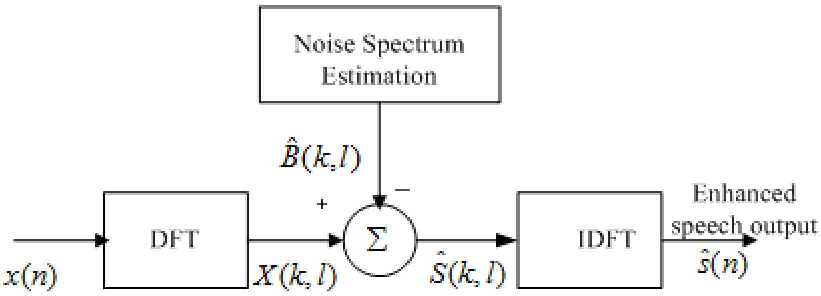


Fig.3-2. A block diagram of the spectral subtraction algorithm

The noise spectrum estimate can be obtained during the non-speech activity frame as given by [6]-[8],

$$\left|\hat{B}(k,l)\right|^p = \lambda\left|\hat{B}(k,l-1)\right|^p + (1-\lambda)\left|X(k,l)\right|^p \tag{3.5}$$

where λ is a forgetting factor. This noise spectrum estimation technique can be seen as 'low-pass filtering' the noise-only spectral components [7]. The spectrum of the enhanced speech signal is therefore given by

$$\hat{S}(k,l) = \left|\hat{S}(k,l)\right| e^{j\theta_x(k,l)} \tag{3.6}$$

where $\theta_X(k,l)$ is the phase of the noisy spectrum. By using the Inverse STFT, the enhanced speech signal $\hat{s}(n)$ can be obtained.

The spectral subtraction algorithm is one of the most popular speech enhancement algorithms because of its low computational complexity and high efficiency. Conventionally, the SS algorithm [29] estimates the magnitude or power noise spectrum during non-speech-activity. The use of Voice Activity Detector (VAD) is therefore necessary [7]. If the noise spectrum can be estimated very accurately, superior performance of the SS algorithm can therefore be obtained. On the other hand, residual noise and musical noise will be introduced in the enhanced speech spectrum if the noise spectrum estimate is insufficiently correct.

Since the single-channel SS method can be alternatively considered as Spectral Suppression, i.e.by calculating a proper spectral gain function, $G(k,l)$, the enhanced speech spectrum can be obtained from the multiplication of the gain function to the noisy spectrum, as given by

$$\hat{S}(k,\ell) \quad = \quad G(k,\ell)\,X(k,\ell) \tag{3.7}$$

Several methods have been introduced to derive efficient the spectral gains [43]-[44], e.g. Wiener gain, MMSE-based gain, etc.

3.1.1.1. Wiener algorithm

Wiener filter is one means of NR based on linear processing models and minimum-mean-square-error (MMSE) estimation [35]. The estimated clean speech spectrum depends on the Fourier Transform of noisy signal, $X(\omega)$, and a spectral gain function, $G(\omega)$.

This NR algorithm also brings about the musical noise which reduces the speech quality. Several methods have been proposed to obtain efficient spectral gain function, including a priori SDR estimation, Adaptive short-time analysis-synthesis [36], [37].

3.1.1.2. Statistical-model-based MMSE algorithm

For this type of NR technique, MMSE short-time spectral amplitude (STSA) estimation is discussed. The Discrete Fourier Transform (DFT) is applied to noisy signal to obtain its spectral magnitude and phase. The enhanced speech spectrum, which can be modeled as the multiplication of the magnitude of the noisy signal $|X(k)|$ by the spectral gain function, $G(k)$, and then by the phase of the noisy signal $\angle X(k)$, as represented previously in eq.(3.7). However, if the noise is non-stationary, the speech intelligibility will be poor [35].

3.1.2. Subspace algorithm

To reduce speech distortion while eliminating the background noise, subspace-based methods have been introduced in [29]–[32]. The noisy signal is decomposed into two subspaces; the signal-plus-noise subspace and the noise subspace. Speech enhancement is performed by removing the noise subspace and by retaining only the clean signal in the signal subspace. The decomposition of the space into two subspaces can be done using either the Singular Value Decomposition (SVD) [33], [34] or the Eigen Value Decomposition (EVD) [29]–[32]. Compared with the spectral subtraction algorithm, the subspace algorithm gives less amount of residual noise. However, its computational cost is high because of the eigenvalue decomposition.

It can be seen that these single-channel NR algorithms still encounter with the problem of residual noise and the musical noise, especially when the noise is non-stationary.

3.1.3.Voice Activity Detector (VAD)

VAD is essential in various speech communication applications such as speech recognition, speech coding, hands-free telephony and echo cancellation [51]. In Noise Reduction (NR) algorithms, it is difficult to accurately estimate the statistical properties of the desired signal and noise from the noisy signal. In traditional NR algorithm, the noise spectrum is estimated and subtracted from the noisy spectrum without knowing whether the incoming signal is a mixture of the speech and noise signals or noise only signal [52]. If the speech is present in the noise spectrum, it will also be suppressed at the output. On the other hand, if there is no speech, no complicated algorithm is needed since the noise spectrum can be immediately suppressed. Therefore, Voice Activity Detector (VAD) is necessary to determine whether the speech signal is present or not [52]. In particular, most of the single-channel NR algorithms are needed the voice activity detector (VAD) in order to estimate the noise spectrum.

Various VAD algorithms assume that the background noise statistics are stationary over a longer period of time than those of speech. The decision rule for determining the presence or absence of speech is based on the comparison of the observed signal statistics and the estimated noise statistics in the current frame [50].

In [49]-[51], the VAD is based on the likelihood ratio test. The a priori and a posteriori SNRs are required in order to estimate VAD. These algorithms have higher computational complexity than that in [52] and [53]. In [53], VAD is applied to microphone array processing and decision rule is based on the VAD filter.

The choice of the VAD filter does not require any knowledge of the speech signal [53]. The VAD filter which is a whitening filter [53] is given by

$$\left|H_{VAD}(\omega)\right|^2 = \frac{1}{\phi_n(\omega)} \tag{3.8}$$

where $\phi_n(\omega)$ is the noise spectral density of the noisy signal. If the VAD filter is not accurately designed, there will be speech distortion at the output. In [53], the decision rule is formulated in terms of the average subband SNR formulated by

$$SNR(l) = \frac{1}{D}\sum_{d=0}^{D-1} Q_{SNR}(l,d) \tag{3.9}$$

where $Q_{SNR}(l,d)$ is the signal energy in each l^{th} frame, d is the subband index for $d = 0,1,...,D-1$. The signal energy is calculated by

$$Q_{SNR}(l,d) = Q_p(l,d) - E_N(d) \tag{3.10}$$

where $Q_p(l,d)$ is the p sampling quantile formulated by

$$Q_p(l,d) = (1-f)E_{(l)}(l,d) + fE_{(l+1)}(l,d), \qquad l = \lfloor 2pN \rfloor,\ f = 2pN - l \tag{3.11}$$

and $E_N(d)$ is estimated as the median filter of the set $\{E(0,d), E(1,d),..., E(N-1,d)\}$. The log-energies in D subbands is computed by means of

$$E(l,d) = \log(\frac{D}{L}\sum_{a=a_d}^{a_{d+1}-1} X_l(a)), \quad a_d = \frac{A}{2D}d, \quad \text{d=0,1,...,D-1} \tag{3.12}$$

If the average SNR is greater than the threshold in the current frame, it is classified as speech. Otherwise, it is classified as non-speech.

$$X(k,l) = \begin{cases} S(k,l) & \text{if } SNR(l) > threshold \\ B(k,l) & otherwise \end{cases} \tag{3.13}$$

By using the VAD, the noise can be accurately estimated. However, the NR algorithms with VAD have higher computational complexity than the other NR algorithms which do not need any VAD.

3.2. Multi-channel Noise Reduction techniques

The beamforming techniques can create a pattern of constructive and destructive interference and exploit them. The simplest beamforming technique is delay-and-sum beam forming (DAS). In this technique, time-domain sensor signals are delayed and then summed to give a single channel output, as shown in Fig.3-3 [1],[38]. Therefore, the beamforming technique can be considered as the multi-channel NR technique. However, the low directivity of DAS beamformer results in low noise reduction performance indicating that the directivity is effectively zero and forming the beam pattern nearly flat [4].

In order to overcome such low-frequency directivity problems of the DAS technique, Superdirective (SD) beamformer [31] is designed. In this technique, the channel filters was calculated by maximizing the array factor of directivity. However, the noise reduction performance will be degraded in non-diffuse or time-varying conditions since it is designed for diffuse noise condition and time-invariant.

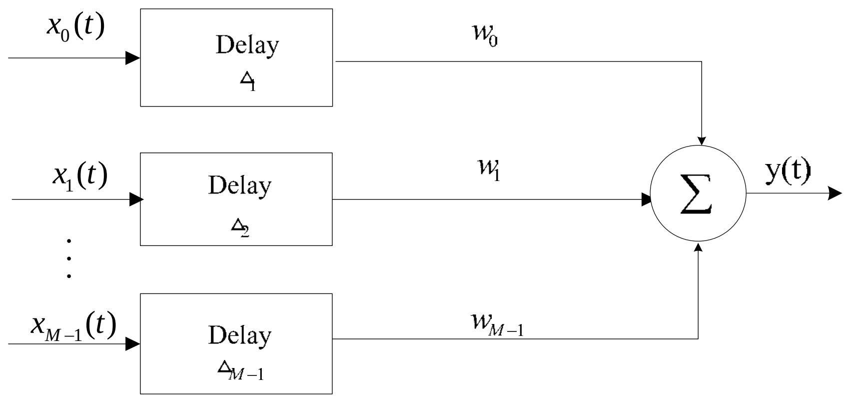


Fig.3-3. A block diagram of DAS beamformer

3.2.1. Generalized sidelobe canceller

To deal with the problem of DAS and SD beamformers, Generalized Sidelobe Canceller (GSC) is proposed to reduce the interfering noise [39]. Its block diagram is shown in Fig.3-4. This method can suppress coherent noise only when the number of noise source is less than that of microphones when the number of noise source is greater than or equal to that of microphone, low noise attenuation performance is obtained [40].

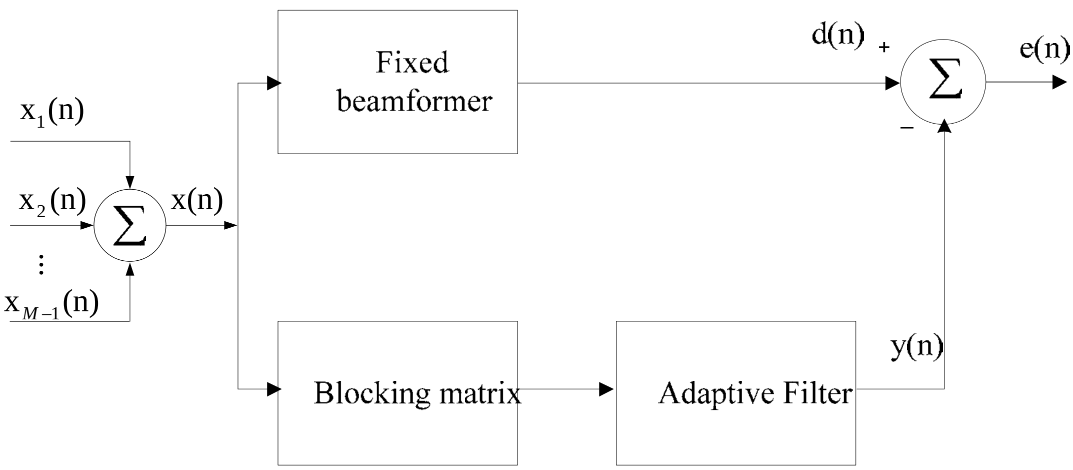


Fig.3-4. A block diagram of the Generalized Sidelobe Canceller (GSC)

3.2.2. Extension to two-microphone NR in highly non-stationary multiple-noise-source

Two-channel beamformer based on short-time spectral amplitude (STSA) estimation has been proposed in [41]. However, the performance of this algorithm deteriorates in the case of multiple highly non-stationary interfering signals.

To improve the noise attenuation performance of the two-channel GSC technique in [41], a two-microphone noise reduction technique is proposed to be employed in highly non-stationary multiple-noise-source environments [4]. Its block diagram is as shown in Fig.3-5.

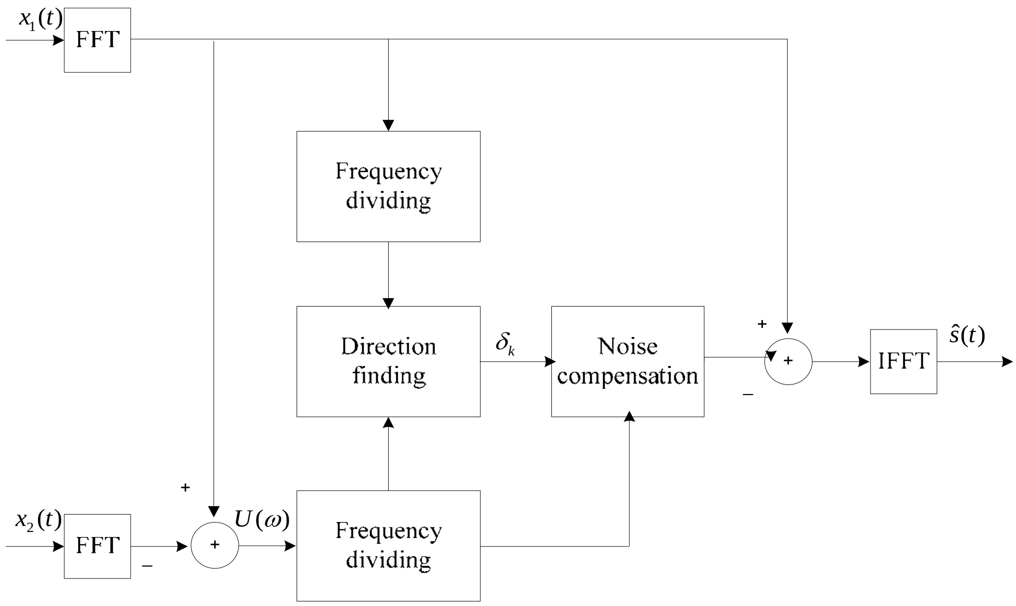


Fig.3-5. A block diagram of the two-channel GSC in [4]

The desired speech spectrum is first carried out to obtain the noise-only spectrum. The direction of the virtual integrated noise signal is estimated by using DOA technique. Then the estimated spectrum of the integrated noise can be obtained. Finally, the enhanced speech spectrum is obtained by subtracting the estimated noise spectrum from the noisy speech spectrum on the first microphone as shown in Fig.3-5.

3.2.3. Two-channel GSC with binaural outputs

The two-channel GSC in [4] is then extended to obtain binaural benefits e.g. in hearing-aid applications. A two-microphone NR technique has been proposed in [42] to be employed in highly non-stationary multiple-noise-source environments with binaural outputs. We will further investigate this technique in more details and will refer to this approach as '2chGSC'.

The two noisy speech signals obtained from the first microphone and second microphone can be modeled as a sum of clean speech signal $s(n)$ and the m^{th} interfering noise signals, $b(n)$, as represented by

$$x_i(n) = s(n) + \sum_{m=1}^{M} b_m(n-(i-1)\delta_m), \quad i = 1,2 \tag{3.14}$$

where i^{th} is the number of microphone, M is the number of noise sources and δ_m is the time delay of the m-th interfering noise source.

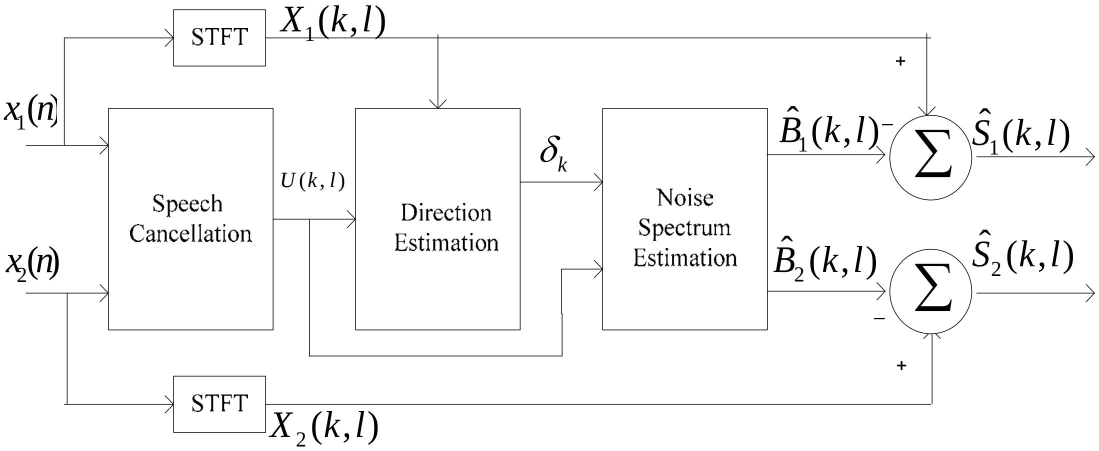


Fig.3-6. A block diagram of two-channel GSC for noise reduction

The two noisy speech signals are then transformed into the frequency domain using STFT. The speech spectrum is cancelled by subtracting the noisy spectrum obtained in the second microphone from that obtained in the first microphone to get the noise-only spectrum as shown in Fig.3-6.

The analyzed spectrum is divided into several small subbands to reduce the computational complexity. The speech cancelled spectrum in the k^{th} subband is given in terms of the integrated noise spectrum in the k^{th} subband, $N_k(\tilde{\omega})$, as

$$U(\tilde{\omega}) = 2jN_k(\tilde{\omega})e^{-j\tilde{\omega}\frac{\delta_k}{2}}\sin(\tilde{\omega}\frac{\delta_k}{2}), \quad \omega_{k-1} \le |\tilde{\omega}| < \omega_k, \quad k = 1,2,..,K \tag{3.15}$$

where $\tilde{\omega}$ is the angular frequency between the adjacent k^{th} subband when K is the number of subband.

Assuming that the speech signal and the interfering noise signals are uncorrelated, the Direction of Arrival (DoA) of the virtual noise source, $\hat{\delta}_k$, of the k^{th} subband in the first microphone can be calculated as

$$\hat{\delta}_k = 2\hat{\delta}'_k = 2\arg\max_t \left[IFFT\left[\frac{U(\tilde{\omega})X_1^*(\tilde{\omega})}{|U(\tilde{\omega})||X_1^*(\tilde{\omega})|} \right] \right] \tag{3.16}$$

where $\hat{\delta}'_k$ is the half of the estimate of the virtual noise direction $\hat{\delta}_k$.

Similarly, the same procedure is taken to find $\hat{\delta}_k$ in the second microphone [9]. The noise compensator is used to obtain the accurate noise estimation.

With the noise compensator, which is given by

$$H_{k,i}(\tilde{\omega}) = \frac{e^{j\tilde{\omega}\frac{\hat{\delta}_k}{2}}}{2j\sin(\tilde{\omega}\frac{\hat{\delta}_k}{2})}, \tag{3.17}$$

the integrated noise spectrum in the k^{th} subband is then obtained to be

$$\left|\hat{N}_{k,i}(\tilde{\omega})\right| = \begin{cases} U(\tilde{\omega})H_{k,i}(\tilde{\omega}), & \left|\sin(\tilde{\omega}\hat{\delta}_k/2)\right| \geq \varepsilon \\ \gamma\left|\hat{N}_{k,i}^{pre}(\tilde{\omega})\right|^2 + (1-\gamma)E[\left|\hat{N}_{k,i}(\tilde{\omega})\right|^2 \| X_i(\tilde{\omega})|^2], & \text{otherwise} \end{cases} \tag{3.18}$$

where ε is a small positive value and $0 < \gamma < 1$ is the forgetting factor controlling the rate of noise reduction, $\hat{N}_{k,i}^{pre}(\tilde{\omega})$ is the estimate of the noise spectrum in the previous frame. The noise spectrum in each subband is estimated by the multi-channel NR approach if $\left|\sin(\tilde{\omega}\hat{\delta}_k/2)\right| \geq \varepsilon$, otherwise it is estimated by the single-channel NR approach. The noise spectrum in each microphone can be calculated over the entire frequency region as

$$\hat{B}_i(\omega) = \sum_{k=1}^{K} \left|\hat{N}_{k,l}(\tilde{\omega})\right|, \quad \omega_{k-1} \leq \tilde{\omega} < \omega_k, \; i = 1, 2 \tag{3.19}$$

Finally, the enhanced speech spectrum at each microphone can be obtained by subtracting the estimated noise spectrum from the noisy speech spectrum in each channel as shown in eq.(3.4). The flow diagram of this technique can be shown in Fig.3-7.

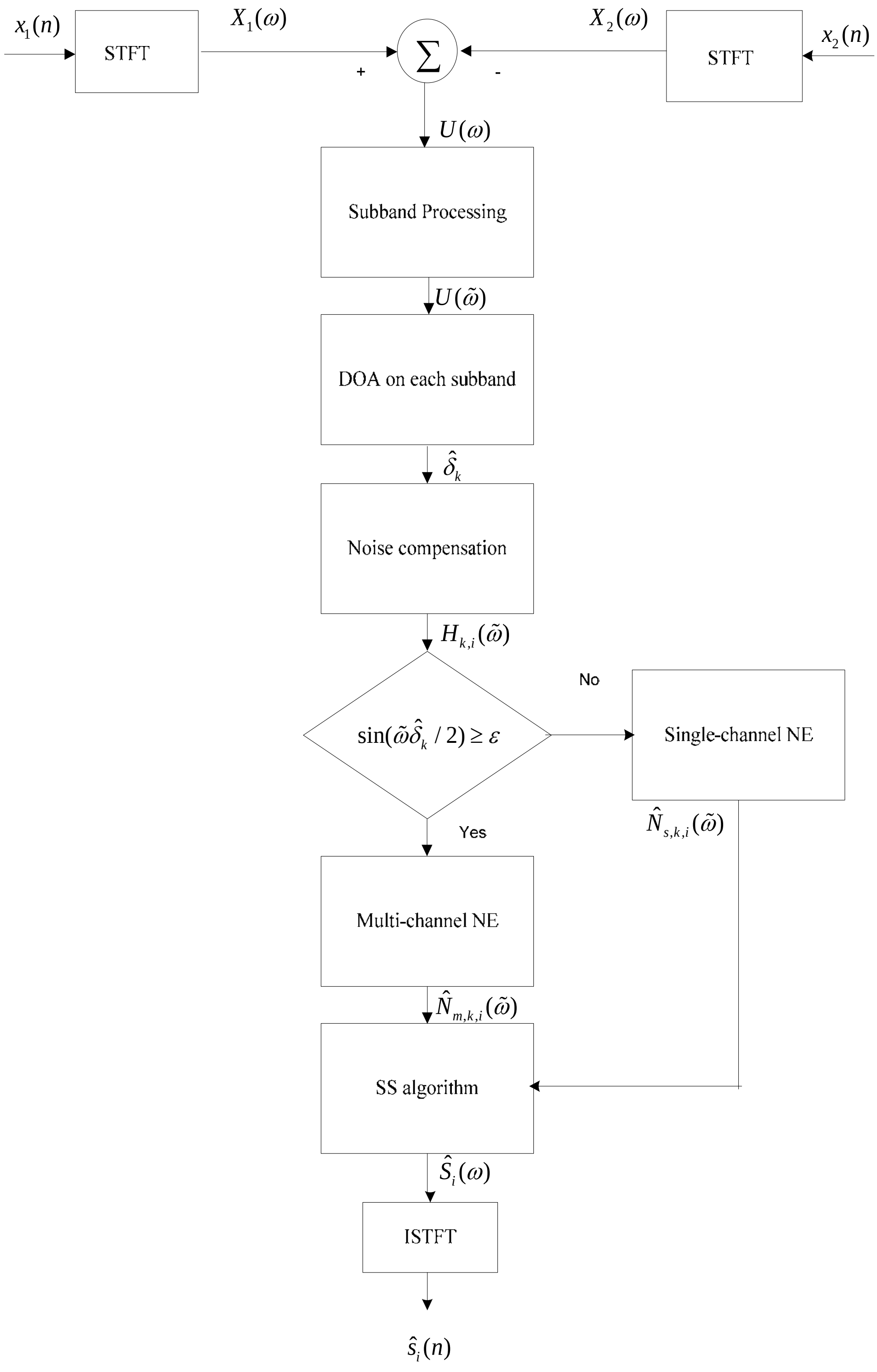


Fig.3-7. The flow diagram of a two microphone noise reduction method in [40]

3.2.4. The NR technique based on the modified Cross-Spectral Subtraction (2chCSS)

This two-channel NR technique was introduced for the applications of hands-free car kit [22]. Since the single-channel SS method can be alternatively considered as Spectral Suppression, i.e.by calculating a proper spectral gain function, $G(k,l)$, the enhanced speech spectrum can be obtained from the multiplication of the gain function to the noisy spectrum, as given by

$$\hat{S}(k,\ell) \quad = \quad G(k,\ell)\,X(k,\ell) \tag{3.20}$$

Several methods have been introduced to derive efficient the spectral gains [43]-[44].

The noise-only reference signal is suggested in [22] to be obtained by the use of the decorrelation of the noises when two microphones are sufficiently spaced. Based on the Cross-Spectral Subtraction (CSS) [45], the coherence between the two noisy spectrums X_1 and X_2 can be realized as a spectral gain function, or a filter, as given by

$$G_{\text{css}}(k,\ell) \quad = \quad \frac{\left|\gamma_{x_1x_2}(k,l)\right| - \left|\gamma_{b_1b_2}(k,l)\right|}{\sqrt{\gamma_{x1}(k,l)\gamma_{x2}(k,l)}} \tag{3.21}$$

where $\gamma_{x_i}(k,l)$ and $\gamma_{x_1x_2}(k,l)$ are the noisy signal power spectral densities (psd) for $i=1,2$ and cross-psd of the speech signals, respectively. $\gamma_{b_1b_2}(k,l)$ is the noise cross-psd. As a result, by applying the coherence filter to the noisy spectrum, the noise spectral components can be removed. The noisy signal psd and cross psd of the speech signals are obtained recursively as

$$\gamma_{x_i}(k,l) \quad = \quad \lambda\gamma_{x_1}(k,l-1) + \left(1-\lambda\right)X_1(k,l)X_1^*(k,l) \tag{3.22}$$

and

$$\gamma_{x_1x_2}(k,l) \quad = \quad \lambda\gamma_{x_1x_2}(k,l-1) + \left(1-\lambda\right)X_1(k,l)X_2^*(k,l) \tag{3.23}$$

for $i=1,2$. The forgetting factor $0<\lambda\le 1$ is suggested to take small values for speech-activity frames and high values during non-speech-activity frames in order to control the residual noise during non-speech-activity frames. The adaptive expression for the forgetting factor was proposed in [22] to be

$$\lambda(k,l) \quad = \quad 0.98-0.3G_{\mathrm{css}}(k,l-1) \tag{3.24}$$

However, the authors in [4] found that the musical noise still occurred during speech-activity frames. Hence, it was proposed in [22] to control the musical noise effect by the overestimation of the noise cross-psd, $\gamma_{b_1b_2}(k,l)$, as follows.

$$\sqrt{\gamma_{b_1}(k,l)\gamma_{b_2}(k,l)} \quad = \quad d\left(\widetilde{SNR}_{\mathrm{post}}(k,l)\right)\sqrt{\gamma_{b_1}(k,l-1)\gamma_{b_2}(k,l-1)} \tag{3.25}$$

where γ_{b_i} is the noise psd in each channel, $i=1,2$. The function $d\left(\widetilde{SNR}_{\mathrm{post}}\right)$ depends on the posteriori SNR and real positive constants h,g and L.

$$d(\widetilde{SNR}_{\mathrm{post}}) \quad = \quad L+(1-L).\frac{1}{1+1/(g.\widetilde{SNR}_{\mathrm{post}})}.\left(1+\frac{1}{1+g.h.\widetilde{SNR}_{\mathrm{post}}}\right) \tag{3.26}$$

The a posteriori modified SNR is given by

$$\widetilde{SNR}_{post}(k,l) = \frac{\left|X_1(k,l)X_2(k,l)\right|}{\sqrt{\lambda_{b_1}(k,l-1)\lambda_{b_2}(k,l-1)}} \tag{3.27}$$

where h is the maximum boundary value that the function $d\left(\widetilde{SNR}_{\mathrm{post}}\right)$ can take and $g=1/(1-h)$ is normally chosen.

Therefore, the two-channel NR technique in [22] modified the CSS method and will be referred to in this paper as the 2chCSS technique. The enhanced speech spectrum for each channel is thus obtained as

$$\hat{S}_i(k,\ell) \quad = \quad G_{\mathrm{CSS}}(k,\ell)\,X_i(k,\ell) \tag{3.28}$$

and its block diagram can be demonstrated as given in Fig. 3-8.

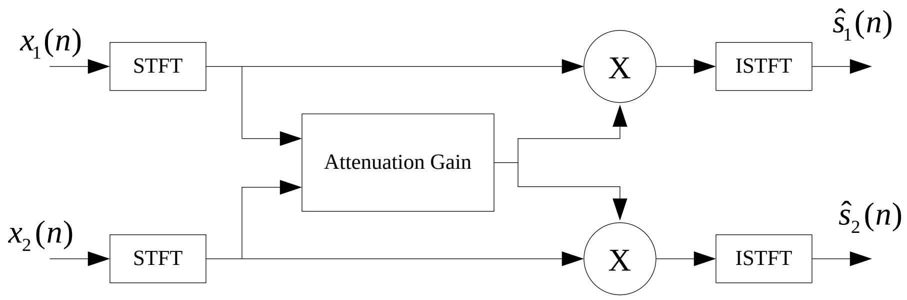


Fig.3-8. A block diagram of the 2chCSS technique.

3.3. Summary

In section 3.1, the single-channel noise reduction techniques are described. The use of VAD is necessary in order to clarify that the current frame is speech-activity or non-speech-activity. In section 3.2, multi-channel noise reduction algorithms are described. In these NR algorithms, the noisy signals are considered from the two microphone signals. The noise is estimated without any need for a VAD. Therefore, multi-channel NR techniques will be considered in this dissertation to improve their noise reduction performance while preserving the enhanced speech quality, which will be described in the next Chapter.

CHAPTER 4

THE PROPOSED NOISE REDUCTION TECHNIQUES FOR DIGITAL HRAING AIDS

In this chapter, the two new approaches for controlling the noise subtraction parameter of the SS-based noise reduction techniques will be presented. In the two proposed method, the parameter α will be frequency-dependent. So as to mitigate the tradeoff of the conventional SS technique, i.e. in the conventional SS technique, α is fixed for every frequency components i.e. for the whole frequency spectrum. If we would like to increase α for further noise subtraction, large α will, however, introduce speech distortion and musical noise. By reducing the α parameter to preserve speech quality, the additive noise source **cannot** be sufficiently eliminated. Finally, a proposed NR technique, which is a hybrid structure between the 2chGSC+SS technique and 2chCSS method is given.

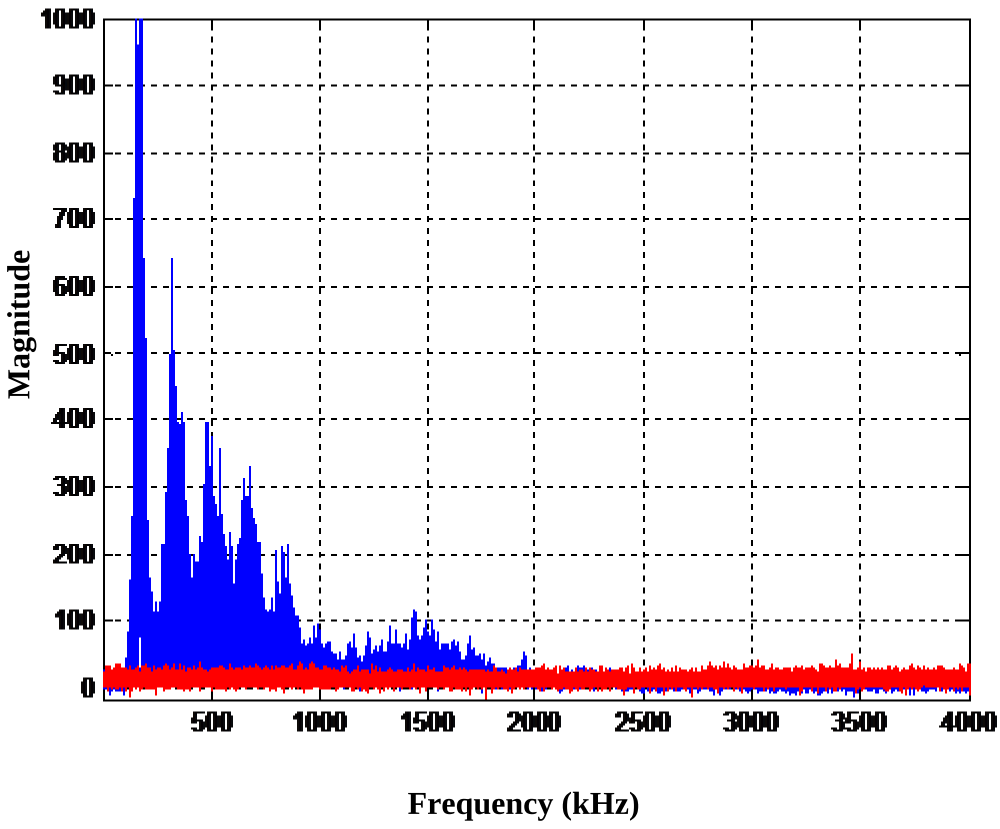


Fig.4-1. Power Spectrum of clean speech and noise

4.1. Variable Noise Subtraction Parameters for SS-based Noise Reduction method

When the noise subtraction parameter is chosen to be large in order to obtain a high amount of noise reduction, it will, however, introduce a large amount of speech distortion. On the other hand, by reducing the noise subtraction parameter to preserve the speech quality, the additive noise signal cannot be sufficiently eliminated. In order to control the amount of speech distortion while achieving sufficient level of noise reduction, the noise subtraction parameter of the SS method is proposed to be variable and frequency-dependent, instead of being fixed, as will be described in the following subsections.

4.1.1. Method I

From Fig.4-1, it can be seen that the speech components are more dominant than noise components in low frequency band. By considering at the spectral properties of the clean speech signal, $s(n)$, and the additive background noise, $b(n)$, the subtraction factor, α, should be frequency dependent. When the speech spectral components are more dominant than the noise spectral ones, i.e. high SNR level, a large value of noise subtraction parameter, α, should be chosen so that high noise subtraction is implemented, i.e. $1 \leq \alpha < \alpha_{\max}$, where $\alpha_{\max}$ is the maximum value ($\alpha_{\max} = 1.1$). On the other hand, in order to preserve the speech spectral components, the value of noise subtraction parameter should be very small $0 < \alpha < 1$ when the spectral level of noise is high, i.e. low SNR level, in order to preserve the speech spectral components. Based on the simulation results, the significant amount of noise was introduced when the SNR is less than 20dB. It is proposed in this dissertation that the variable and frequency-selective α should be a function of SNR as given by

$$\left|\hat{S}(k,l)\right|^p = \left|X(k,l)\right|^p - \alpha(k,l)\left|B(k,l)\right|^p \tag{4.1}$$

where $\alpha(k,l)$ is given as a function of $SNR(k,l)$ as follows:

$$\alpha\left(k,l\right) \;=\; \begin{cases} \alpha_0 + \mu \mathrm{SNR}(k,l), & \mathrm{SNR}(k,l) \le 20 \\ \alpha_{\max}, & \alpha\left(k,l\right) \ge \alpha_{\max} \\ \alpha_0, & \text{otherwise} \end{cases} \tag{4.2}$$

where $0 < \alpha_0 \le 1$ is the initialized value of the noise subtraction parameter and it is set as 0.8, $\alpha_{\max} = 1.1$ in order to obtain less speech distortion by avoiding noise overestimtion , μ is a small positive value ($\mu = 0.01$) and the noisy signal-to-noise ratio for each frequency component within the l^{th} frame can be calculated as

$$SNR(k,l) \;=\; 10 \times \log_{10} \left\{ \frac{\sum_{l=1}^{J} \left| X(k,l) \right|^2}{\sum_{l=1}^{J} \left| \hat{B}(k,l) \right|^2} \right\} \tag{4.3}$$

where J denotes the number of speech-activity frames. Hence, the enhanced speech spectrum can be obtained as given by

$$\left| \hat{S}(k,l) \right|^p \;=\; \left| X(k,l) \right|^p - \alpha(k,l) \left| \hat{B}(k,l) \right|^p \tag{4.4}$$

4.1.2. Method II

Another approach for frequency selective noise spectral subtraction is proposed based on 2ChGSC+MSS. As mentioned earlier, if the parameter α is to be increased for further noise subtraction, large α , however, introduces speech distortion. On the other hand, by reducing the α factor to preserve speech quality, the additive noise source is not sufficiently eliminated.

Therefore the noise subtraction parameter, α , is suggested to be dependent on the frequency components of the difference between the noisy speech spectrum and the enhanced speech spectrum. For each l^{th} analysis frame, the variable noise subtraction parameter is obtained and updated by

$$\alpha(k,l+1) \;=\; \mu\alpha(k,l) \;+\; \gamma E(k,l) \tag{4.5}$$

where $0<\mu<1$, $\gamma > 0$, k represents the frequency components. The difference between the noisy spectrum and the enhanced speech spectrum is given by

$$E(k,l) = X(k,l) - \hat{S}(k,l) \tag{4.6}$$

In this method, $\alpha(k,l+1)$ is updated based on the $\alpha(k,l)$ in the previous analysis frame and the difference between the noisy spectrum and enhanced speech spectrum $E(k,l)$. When $E(k,l)$ is low, it means there is little amount of noise in the enhanced signal. Therefore $\alpha(k,l+1)$ should be low. On the other hand, when $E(k,l)$ is large, it means that the amount of noise in the enhanced signal is large. So $\alpha(k,l+1)$ should be large.

4.2. A Hybrid Structure

The proposed technique is a hybrid structure between the 2chGSC+SS and the 2chCSS techniques. Since the dominant cues of the speech signals are usually located in the frequency range below 1 kHz [22], minimum speech distortion is allowed in the low-frequency region (0-1 kHz). From the preliminary tests, we found that the 2chGSC+SS technique outperforms the 2chCSS method in terms of the ability to preserve speech quality, whereas the 2chCSS technique can remove the musical noise adequately. Hence, it is proposed that the 2chGSC+SS technique should be employed for processing the signal in the low-frequency region. As for the frequency region above 1 kHz up to the Nyquist rate, the 2chCSS method is employed. The block diagram of the proposed technique is shown in Fig.4-2. The two sensor NR technique can be extended more sensors by putting them like a necklace. However, it will be cost when the more sensors are used. Therefore, two-microphone NR technique is focused in this dissertation.

4.3. Summary

In this chapter, the two new approaches for controlling the noise subtraction parameter based on 2chGSC+MSS was presented. In addition, the hybrid structure was proposed in order to obtain the efficient noise attenuation (NA) performance while preserving the speech quality of the original speech signal.

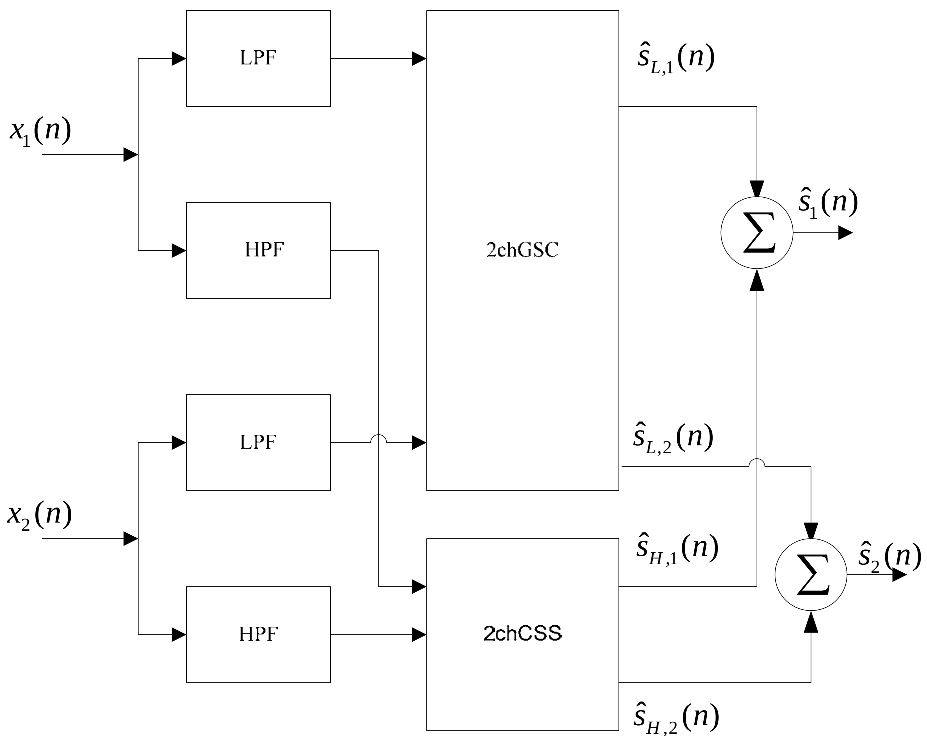


Fig.4-2. A block diagram of the proposed two-channel NR technique.

4.4. Statement of Originality

As far as the author is aware, Chapter 4 of this thesis contains substantial parts which are original contributions to the area of Noise Reduction. The following aspects of the thesis are believed to be original, with the most significant contributions considered to be:

1. The proposal of two approaches in order to control the noise subtraction parameter of the SS-based NR methods to be variable and frequency-dependent, as presented in Section 4.1.1 and Section 4.1.2.

2. The development of the hybrid structure between the 2chGSC+SS and 2chCSS methods, as described in Section 4.2.

CHAPTER 5
EXPERIMENTS AND RESULTS

In this chapter, the performance of the conventional SS method, the magnitude SS method that employs the two-channel Generalized Sidelobe Canceller (2chGSC+MSS), the two-sensor NR technique based on modified Cross-Spectral Subtraction (2chCSS) and the proposed method is evaluated via the noise attenuation performance and the ability to preserve speech quality. Seven experiments were carried out. In the first experiment, the 2chGSC+MSS method was employed in the one-noise-source and three-noise-source environment to observe their superior performance to the single-channel NR technique; MSS. The second experiment was set up to demonstrate the effect of fixed and variable noise subtraction parameter of the 2chGSC+MSS method under three-noise-source environment. This is because the 2chCSS method does not employ the noise subtraction parameter and utilizes the gain function instead. From preliminary results and it will be shown later in this chapter that the 2chGSC+MSS technique that employs the proposed variable $\alpha(k,l)$ as a function of $SNR(k,l)$ yields better NA performance than the one with fixed α and the one with variable $\alpha(k,l)$ in Method II. Therefore, 2chGSC+MSS method that employs the proposed variable $\alpha(k,l)$ as a function of $SNR(k,l)$ was considered for the proposed hybrid method. In the third experiment, the performance of various two-channel NR techniques was compared and discussed, especially for the multiple-noise-source case. Then, another speech signal was used to demonstrate the potential of the proposed methods in experiment 4. The effect of other types of additive noise was shown through experiment 5 to 7. The performance of these NR techniques was observed through the two measurements: output SNR and Log Spectral Distance (LSD) [36].

5.1. Experimental Configuration

Throughout the first three experiments, a clean speech signal of a male speaker at the sampling frequency of 8 kHz was degraded by the additive white Gaussian noise at various SNR levels, i.e. from 5 to 25 dB. Then another clean speech signal of a female speaker with the same sampling rate was employed. For the final three experiments, other types of additive noises; pink noise, factory noise, and car noise were used instead of white Gaussian noise. The speech and noise signals were assumed to be uncorrelated to each other. The noisy signal model as in eq. (3.14) was used. The analysis frame for STFT was 25 ms with the 12.5 ms frame shift and the Hamming window was employed.

The noise attenuation performance is measured via the Output Signal-to-Noise Ratio (SNR), which is given by

$$\mathrm{OutputSNR(dB)} = 10 \times \log_{10} \left\{ \frac{\sum_{n=0}^{Z-1} s^2(n)}{\sum_{n=0}^{Z-1} (s(n) - \hat{s}_1(n))^2} \right\} \tag{5.1}$$

where Z is the sample number and $\hat{s}_1(n)$ is used for the first microphone. On the other hand, $\hat{s}_2(n)$ is used for the second microphone. A higher level output SNR than the input SNR indicates improved noise attenuation performance. On the other hand, a lower level output SNR than the input SNR indicates that the enhanced speech signal still contains some noise or interfering signals.

The measurement for speech distortion is via the Log Spectral Distance (LSD), which is defined to be the difference between the log spectrum of clean speech and that of the enhanced signal, given by

$$\text{LSD(dB)} = \frac{10}{J}\sum_{l=0}^{J-1}\left\{\log_{10}(\left|\hat{S}(k,l)\right|^2) - \log_{10}(\left|S(k,l)\right|^2)\right\} \tag{5.2}$$

where J is the number of speech activity frames and $S(k,l)$ and $\hat{S}(k,l)$ are the clean speech spectrum and the enhanced speech spectrum in the k^{th} frequency bin and l^{th} block. A low LSD level indicates low level of speech distortion. The performance of the conventional SS, 2chGSC+MSS, 2chmCSS and the proposed method is evaluated via the noise attenuation performance and the ability to preserve speech quality.

5.2. Simulation Results

Experiment I: One-noise-source v.s. three noise-source environments

In this section, the performance of MSS and 2ChGSC+MSS is evaluated via the noise attenuation performance and the ability to preserve speech quality. Simulation results for one-noise-source and three-noise-source are demonstrated and discussed in below.

It can be seen from Table 5-1 that both the MSS and 2chGSC+MSS methods exhibit similar NA performance in the one-noise-source environment. As for the speech distortion, from Table 5-2, the 2chGSC+MSS method, however, introduces higher level of speech distortion than the original MSS technique. This is evidently shown via the spectrogram plots in Fig. 5-1. For the three-noise-source environment, the 2chGSC+MSS method gives higher output SNR than the MSS method, as shown in Table 5-3. In addition, the 2chGSC+MSS yields less amount of speech distortion than the MSS one, as can be seen in Table 5-4. Their spectrogram plots in Fig. 5-2 confirm these results. The NA performance and LSD improvement for MSS and 2chGSC+MSS techniques in one-noise-source environment can be seen in Fig 5-2 and Fig 5-3. The NA performance and LSD improvement for MSS and 2chGSC+MSS techniques in three-noise-source environment is shown in Fig 5-5 and Fig 5-6.

Table.5-1. NA performance of MSS and 2chGSC+MSS techniques in one-noise-source environment

Input SNR (dB)	Output SNR (dB)	
	MSS	2chGSC+MSS
5	7.42	7.44
10	11.80	11.90
15	16.55	16.57
20	21.51	21.53
25	26.44	26.54

Table.5-2. LSD measurement of MSS and 2chGSC+MSS techniques in one-noise-source environment

Input SNR (dB)	LSD (dB)	
	MSS	2chGSC+MSS
5	6.84	8.76
10	6.23	7.42
15	4.84	6.49
20	3.81	5.11
25	3.02	4.05

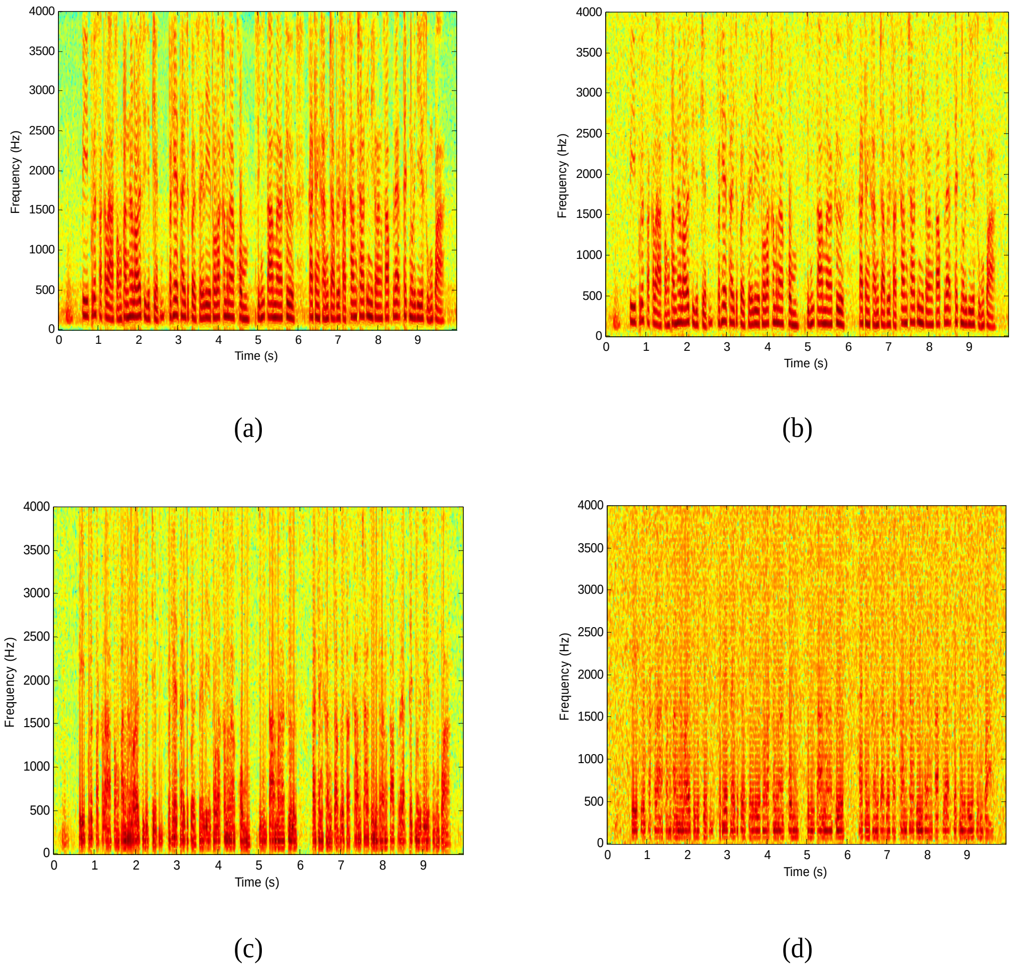


Fig. 5-1. Spectrogram plots for (a) the clean speech signal, (b) the noisy speech signal, the enhanced speech signals using (c) theMSS method, (d) the 2chGSC+MSS method, when the input SNR = 25 dB under one-noise-source environment.

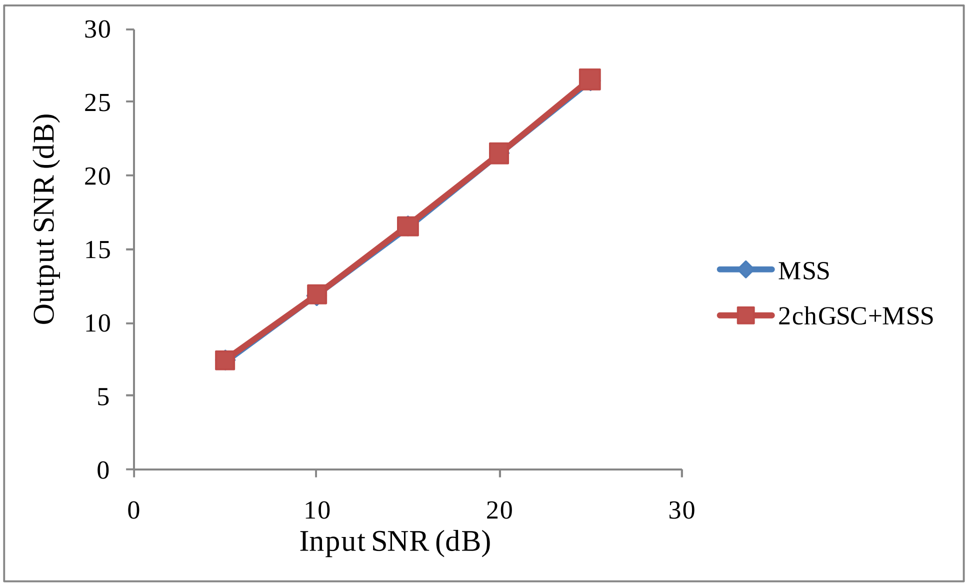


Fig.5-2.Output SNR of MSS and 2chGSC+MSS techniques in one-noise-source environemt

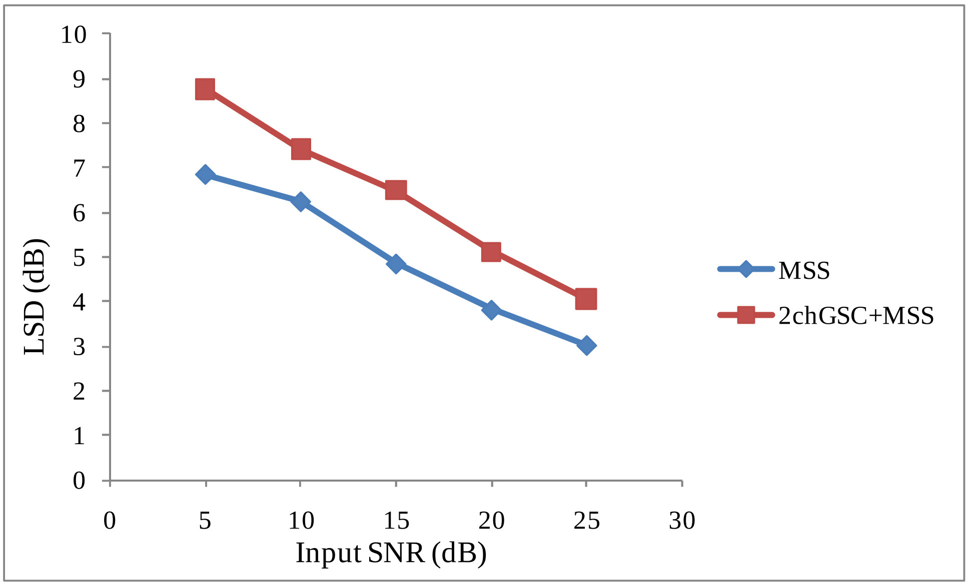
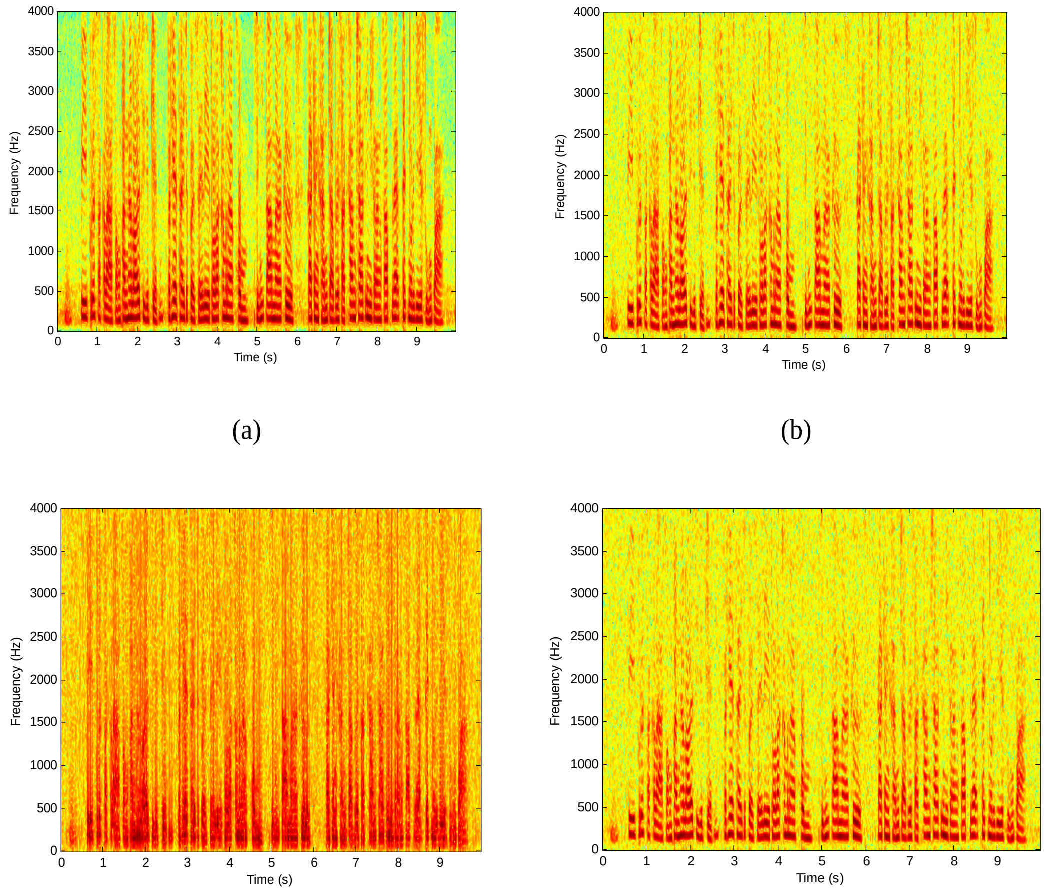


Fig.5-3.LSD measurement of MSS and 2chGSC+MSS techniques in one-noise-source environemt



(a) (b)

(c) (d)

Fig. 5-4. Spectrogram plots for (a) the clean speech signal, (b) the noisy speech signal, the enhanced speech signals using (c) theMSS method, (d) the 2chGSC+MSS method, when the input SNR = 25 dB under three-noise-source environment.

Table.5-3. NA performance of the MSS and 2chGSC+MSS in 3-noise-source environment

Input SNR (dB)	Output SNR (dB)	
	MSS	2chGSC+MSS
5	7.87	8.77
10	11.08	12.64
15	15.31	17.26
20	20.66	22.22
25	25.58	27.24

Table.5-4. LSD measurement of the MSS and 2chGSC+MSS in 3-noise-source environment

Input SNR (dB)	LSD (dB)	
	MSS	2chGSC+MSS
5	8.75	8.45
10	7.45	6.97
15	6.14	5.78
20	5.20	4.59
25	4.31	4.03

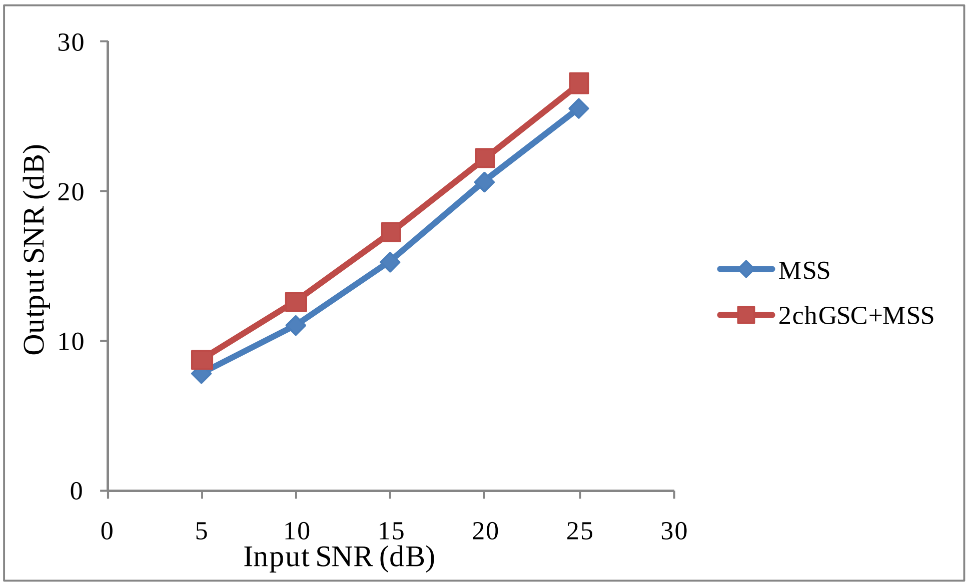


Fig.5-5.Output SNR of MSS and 2chGSC+MSS techniques in three-noise-source environemt

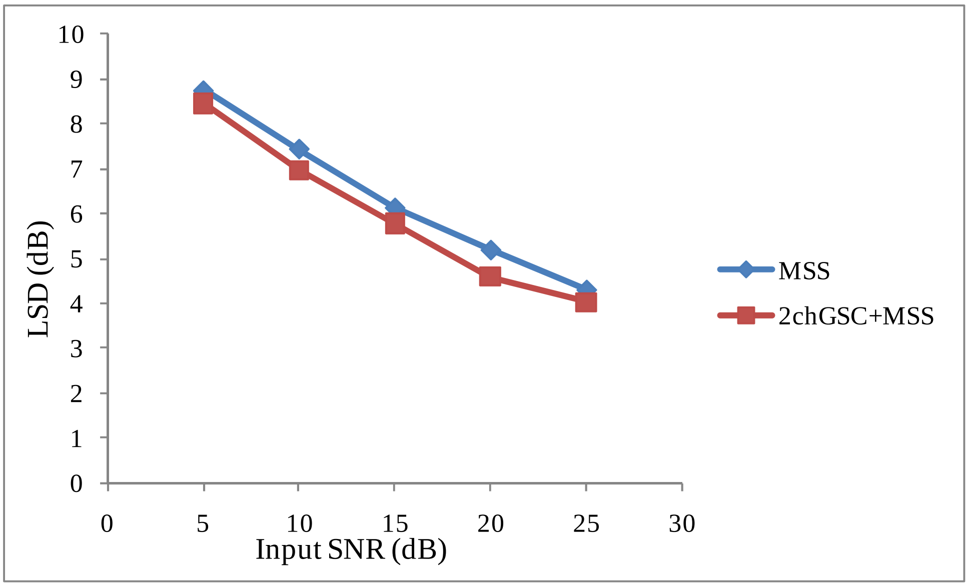


Fig.5-6.LSD measurement of MSS and 2chGSC+MSS techniques in three-noise-source environemt

Experiment2 Fixed noise subtraction parameter v.s. variable noise subtraction parameter

In this experiment, the effect of fixed and variable noise subtraction parameters of the 2chGSC+MSS method is focused under three-noise-source environment. This is because the 2chCSS method does not employ the noise subtraction parameter but utilizes the gain function instead. Since the proposed hybrid structure employs the 2chCSS method, it will not be investigated either. From Table 5-5, it is clear that the 2chGSC+MSS technique that employs the proposed variable $\alpha(k,l)$ (method I) as a function of SNR(k,l) yields the best NA performance than the other two techniques. Furthermore, the least amount of speech distortion can be obtained by employing the proposed variable $\alpha(k,l)$ in method I as compared to the other two techniques. Speech spectral components can be preserved much more than the fixed one, as demonstrated in Table 5-6. For method I and method II, the method I gives better NA performance than method II while these two methods have nearly the same amount of speech distortion. Therefore, method I is focused on for the proposed hybrid structure. From Fig.5-7, it is clear that the 2chGSC+MSS technique that employs the proposed variable $\alpha(k,l)$ as a function of $SNR(k,l)$ yields better NA performance than the one with fixed α and the one with variable $\alpha(k,l)$ in Method II.

Table.5-5. NA performance of the 2chGSC technique with fix α and variable $\alpha(k,l)$

Input SNR (dB)	Output SNR (dB)		
	2chGSC(fix subtraction)	methodI	methodII
5	8.77	12.47	11.45
10	12.64	15.79	14.78
15	17.26	20.05	18.97
20	22.22	24.76	23.77
25	27.24	29.68	28.67

Table.5-6. LSD measurement of the 2chGSC technique with fix α and variable $\alpha(k,l)$

Input SNR (dB)	LSD (dB)		
	2chGSC(fix subtraction)	methodI	methodII
5	8.45	7.36	7.75
10	6.97	6.62	6.43
15	5.78	5.68	5.56
20	4.59	4.33	4.68
25	4.03	3.37	3.54

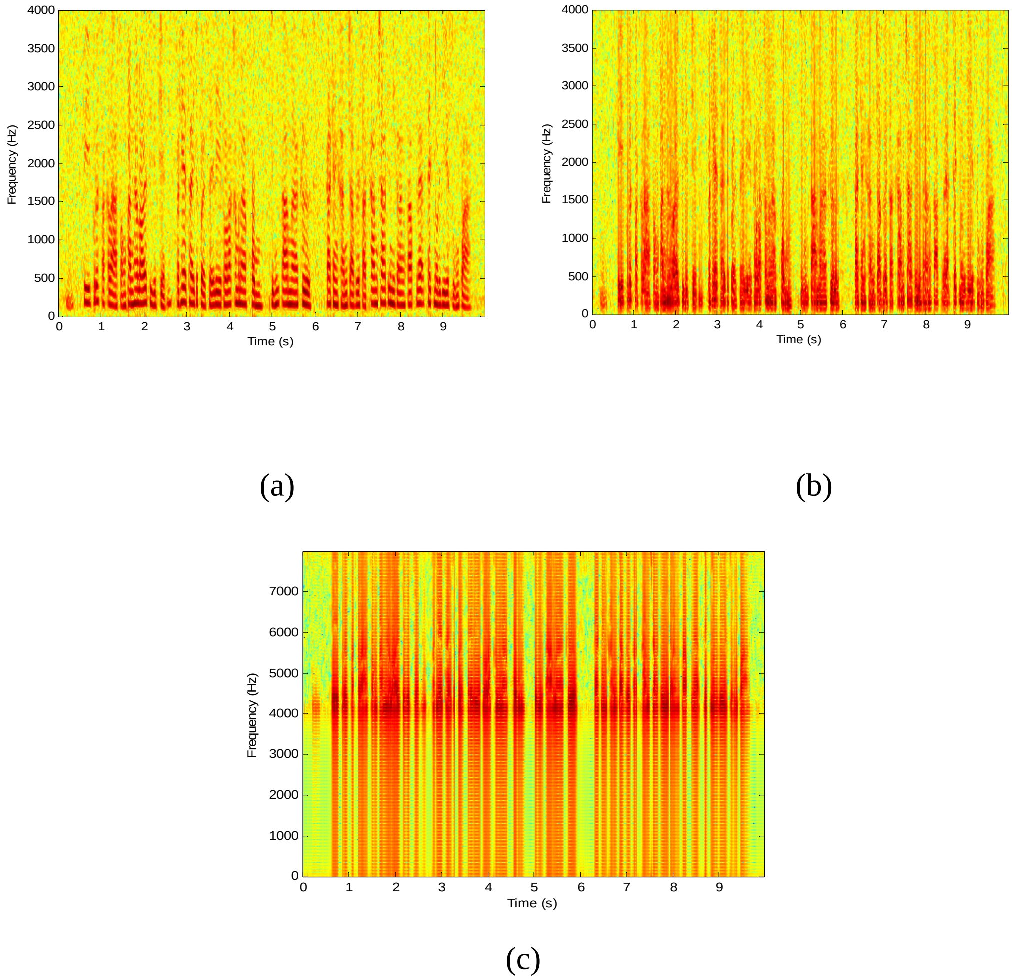


Fig. 5-7. Spectrogram plots for the enhanced speech signals using the 2chGSC+MSS method with (a) fixed noise subtraction parameter, the proposed variable noise subtraction parameter (b) with method I and (c) with method II when the input SNR = 25 dB, under three-noise-source environment.

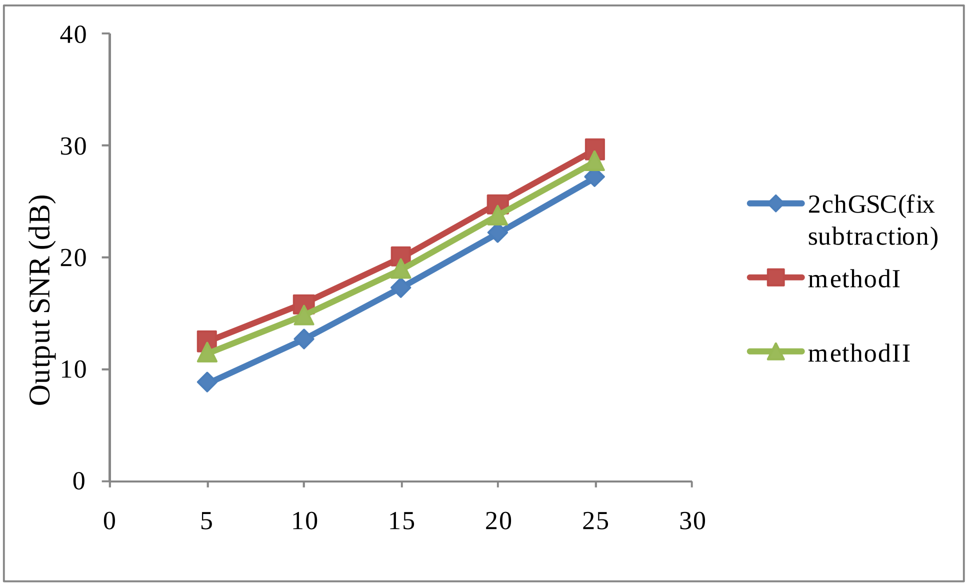


Fig.5-8.Output SNR of 2chGSC+MSS with fixed α and variable $\alpha(k,l)$

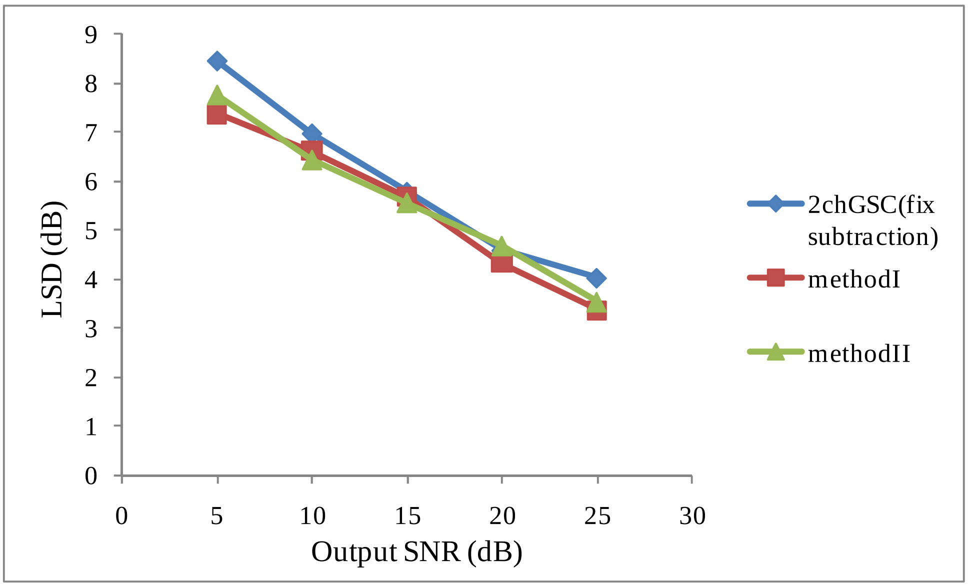


Fig.5-9.LSD measurement of 2chGSC+MSS with fixed α and variable $\alpha(k,l)$

Experiment3 Comparison among various two-channel NR techniques in the three-noise-source environment

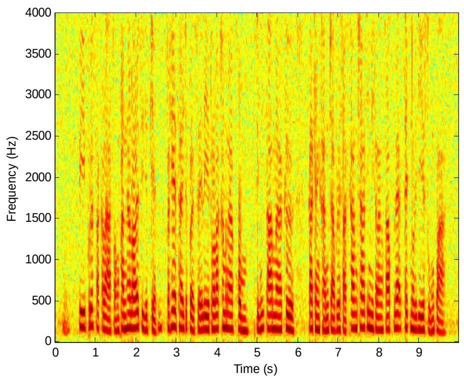
For all the investigated two-channel NR techniques, 2chGSC+MSS employed fixed noise subtraction parameters and the proposed hybrid structure employed variable noise subtraction parameters while the 2chCSS method employed the gain function instead. From Table 5-7, the NA performance of the proposed method was better than that of the 2chGSC+MSS one. However, the 2chCSS yields the best NA performance among these techniques for all input SNR levels. As for the speech distortion via the LSD measurement in Table 5-8, the proposed method shows the least amount of speech distortion, whereas the 2chCSS was the worst. Spectrogram plots in Fig. 5-10 support these objective results.

Table.5-7. NA performance of the 2chGSC+MSS, 2chCSS and proposed technique

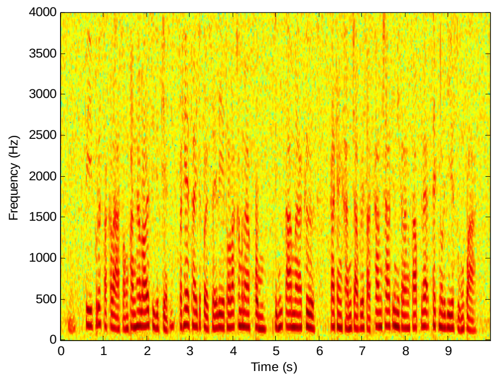
Input SNR (dB)	Output SNR (dB)		
	2chGSC+MSS	2chCSS	Proposed
5	8.77	31.08	18.38
10	12.64	35.13	23.22
15	17.26	39.90	28.09
20	22.22	44.88	33.16
25	27.24	49.87	38.07

Table.5-8. LSD measurement of the 2chGSC+MSS, 2chCSS and proposed technique

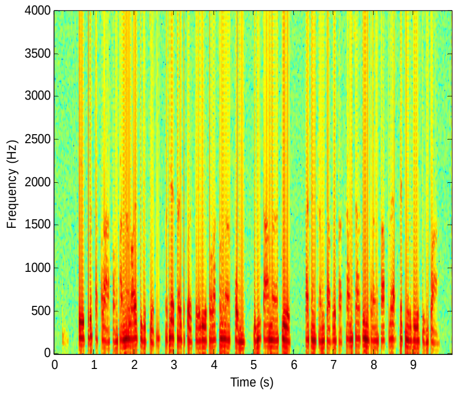
Input SNR (dB)	LSD (dB)		
	2chGSC+MSS	2chCSS	Proposed
5	8.45	8.45	3.37
10	6.97	5.97	3.26
15	5.78	5.51	3.25
20	4.59	5.42	3.25
25	4.03	5.14	3.25



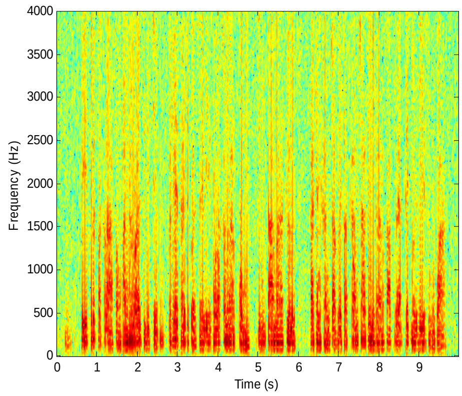
(a)



(b)



(c)



(d)

Fig. 5-10. Spectrogram plots for (a) the noisy speech signal, the enhanced speech signals using (b) the 2chGSC+MSS method, (c), the 2chCSS method, and (d) the proposed NR technique, when the input SNR = 25 dB under three-noise-source environment.

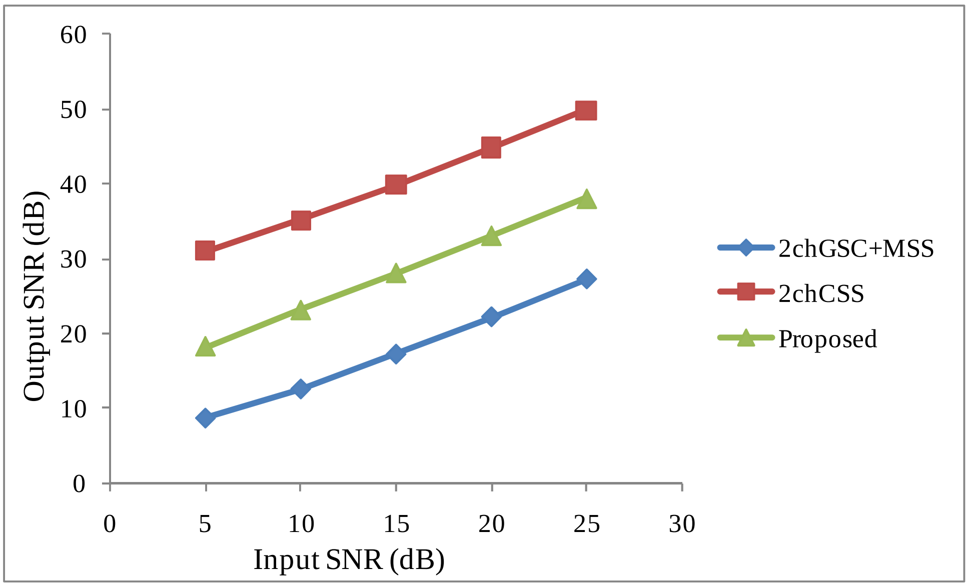


Fig.5-11.Output SNR of the 2chGSC+MSS, 2chCSS and proposed method

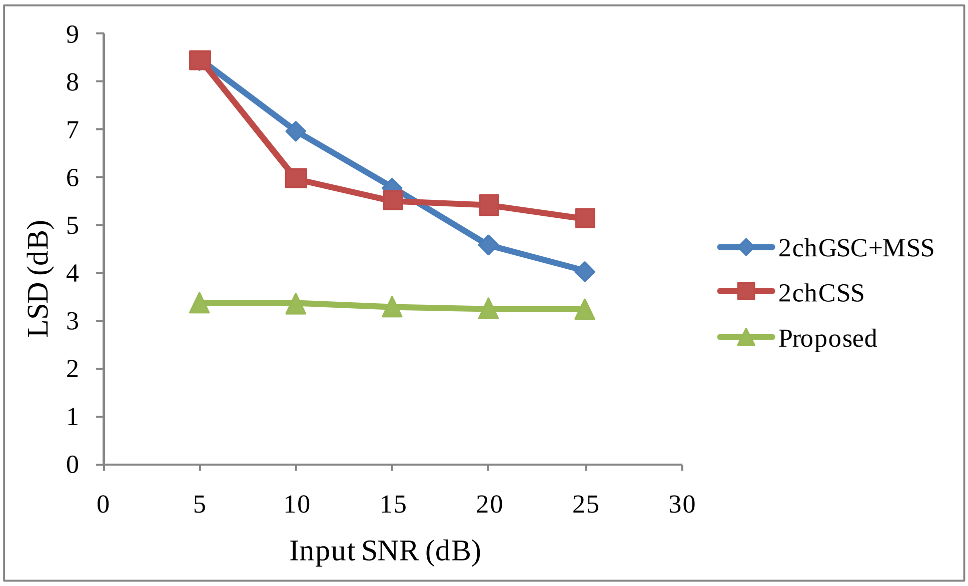


Fig.5-12.LSD measurement of 2chGSC+MSS, 2chCSS and proposed method

Experiment4. Comparison among the two-channel NR techniques and the Proposed technique

In this experiment, another speech signal was added with WGN. From Table 5-9, 2chGSC+MSS gives better noise attenuation performance than MSS one. However, from Table 5-10, 2chGSC+MSS introduces high level of speech distortion than the original MSS technique. From Table 5-11 and 5-12, it can be clearly seen that the 2chGSC+MSS technique that employs the proposed variable $\alpha(k,l)$ as a function of $SNR(k,l)$ yields better NA performance with less amount of speech distortion as compared to the 2chGSC+MSS that employs fixed α . From Table 5-13, the proposed hybrid method yields the best NA performance than the other two techniques. Furthermore, the least amount of speech distortion can be obtained by employing the proposed technique as compared to the other two techniques. In addition, the spectrogram plots of these investigated NR techniques can be demonstrated in Fig.5-11 and 5-12. It can be seen that the proposed hybrid method can eliminate the additive background noise better than the MSS method, whereas high frequency components of the enhanced speech signal can be preserved by employing the hybrid method better than the MSS method. However, some amount of speech distortion can be seen in Fig.5-13 (c), (d) and Fig.5-14 (a), (b).

Table.5-9. NA performance of MSS and 2chGSC+MSS techniques

Input SNR (dB)	Output SNR (dB)	
	MSS	2chGSC+MSS
5	7.50	7.58
10	11.80	12.58
15	16.45	17.60
20	21.40	22.70
25	26.37	27.72

Table.5-10. LSD measurement of MSS and 2chGSC+MSS techniques

Input SNR (dB)	LSD (dB)	
	MSS	2chGSC+MSS
5	9.60	11.95
10	8.71	10.9
15	7.61	9.61
20	6.58	8.65
25	5.93	7.46

Table.5-11. NA performance of the 2chGSC technique with fix α and variable $\alpha(k,l)$

Input SNR (dB)	Output SNR (dB)	
	2chGSC+MSS	Method I
5	7.58	8.82
10	12.58	13.12
15	17.60	17.85
20	22.70	22.79
25	27.72	27.77

Table.5-12. LSD measurement of the 2chGSC technique with fix α and variable $\alpha(k,l)$

Input SNR (dB)	LSD (dB)	
	2chGSC+MSS	Method I
5	11.95	11.28
10	10.90	9.84
15	9.61	8.56
20	8.65	7.79
25	7.46	6.73

Table.5-13. NA performance of the 2chGSC+MSS, 2chCSS and proposed technique

Input SNR (dB)	Output SNR (dB)		
	2chGSC+MSS	2chCSS	Proposed
5	7.58	10.57	26.59
10	12.58	15.40	31.41
15	17.60	20.21	36.44
20	22.70	25.14	41.36
25	27.72	30.06	46.35

Table.5-14. LSD measurement of the 2chGSC+MSS, 2chCSS and proposed technique

Input SNR (dB)	LSD (dB)		
	2chGSC+MSS	2chCSS	Proposed
5	11.95	12.03	6.41
10	10.90	11.60	6.35
15	9.61	11.13	6.33
20	8.65	10.50	6.32
25	7.46	9.94	6.30

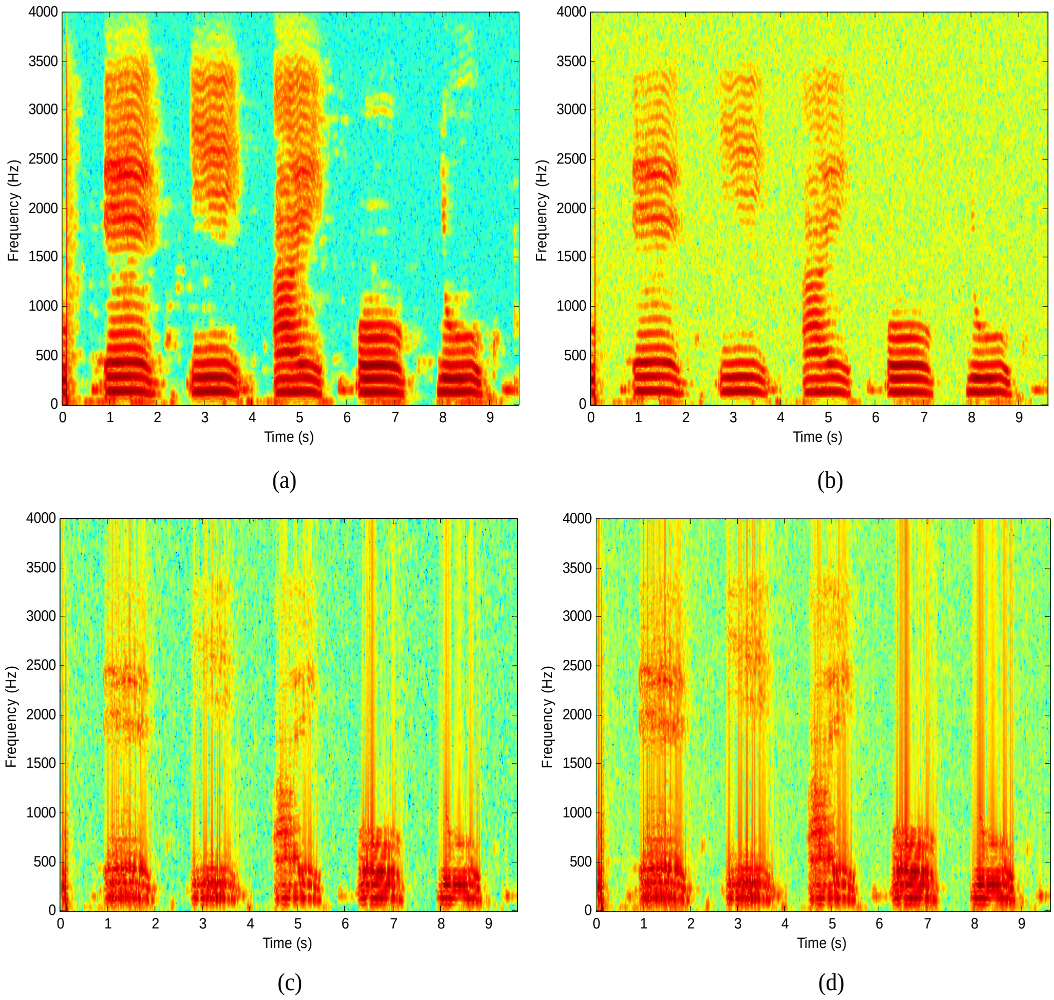


Fig. 5-13. Spectrogram plots for (a) the clean speech signal, (b) the noisy speech signal when the input SNR = 25 dB, the enhanced speech signals using (c) the MSS method, and (d) the 2chCSS+MSS method.

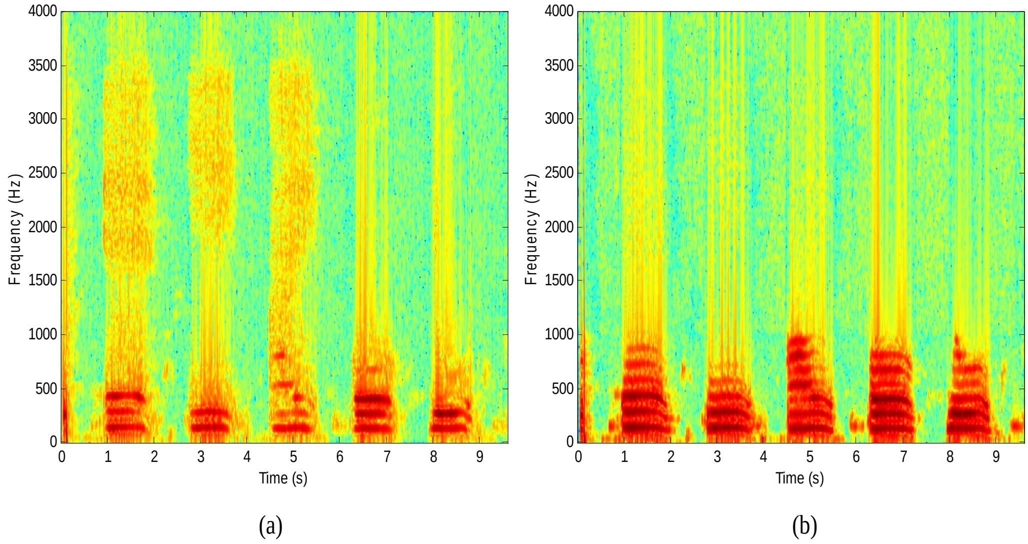


Fig. 5-14. Spectrogram plots for the enhanced speech signals using (a) the 2chCSS method, and (b) the proposed NR technique, when the input SNR = 25 dB

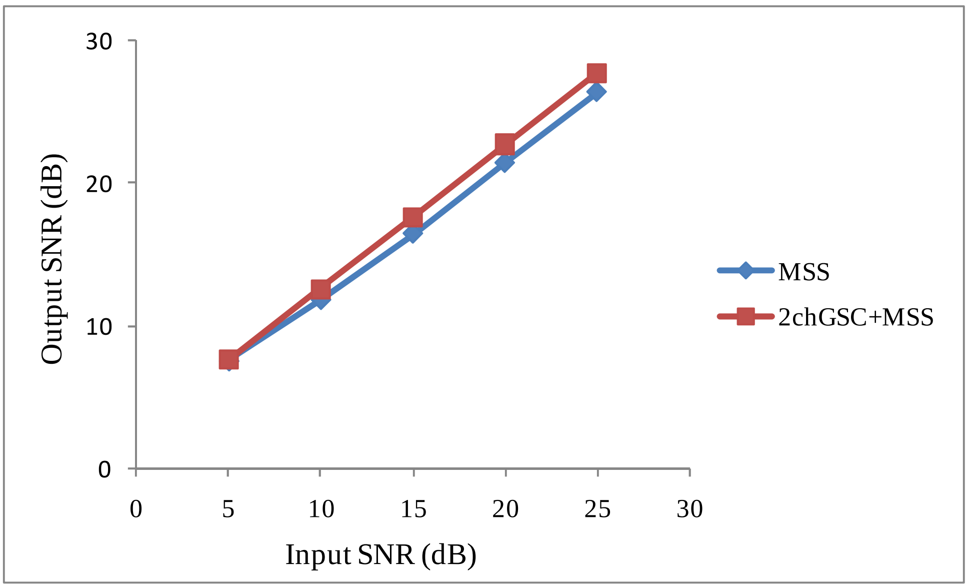


Fig.5-15.Output SNR of MSS and 2chGSC+MSS techniques

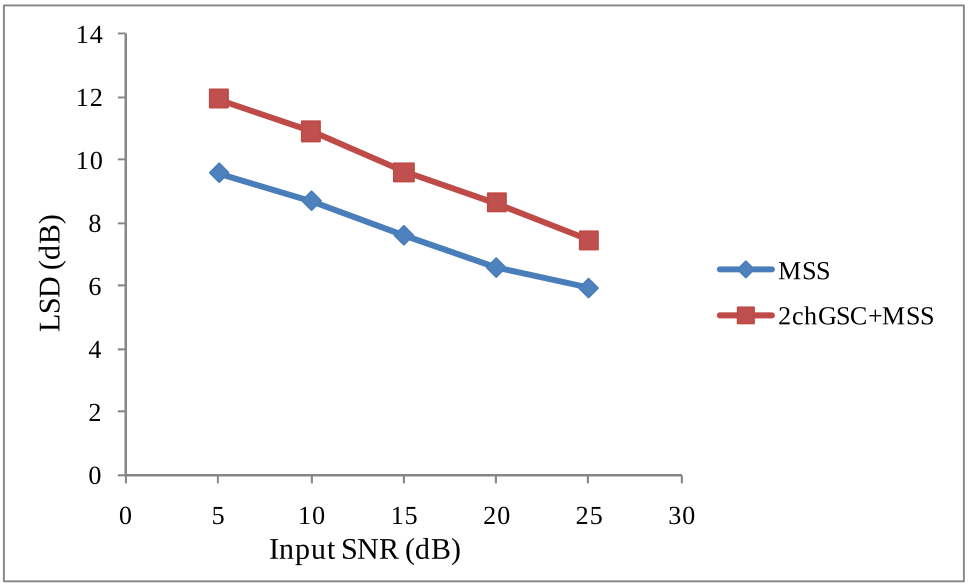


Fig.5-16.LSD measurement of MSS and 2chGSC+MSS techniques

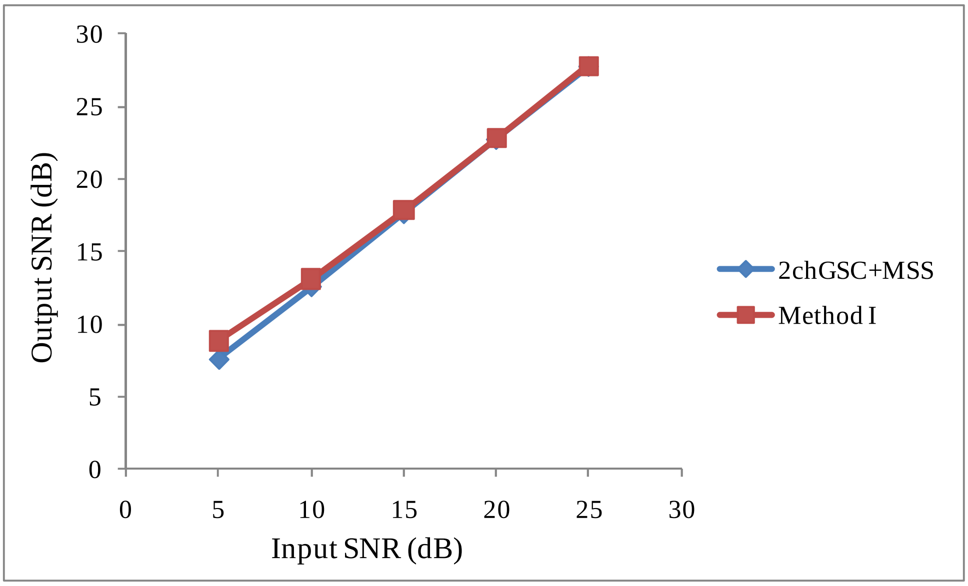


Fig.5-17.Output SNR of 2chGSC+MSS with fixed α and variable $\alpha(k,l)$

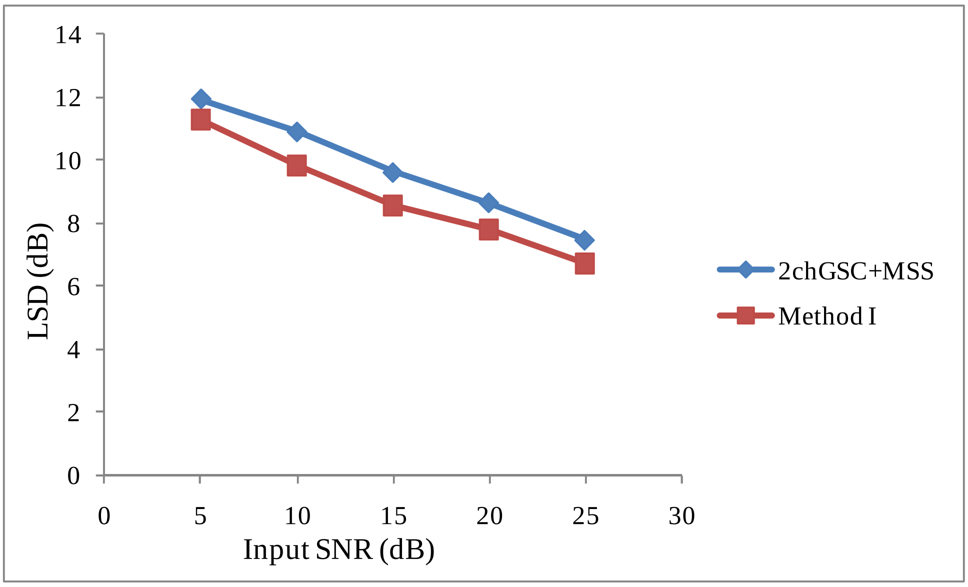


Fig.5-18.LSD measurement of 2chGSC+MSS with fixed α and variable $\alpha(k,l)$

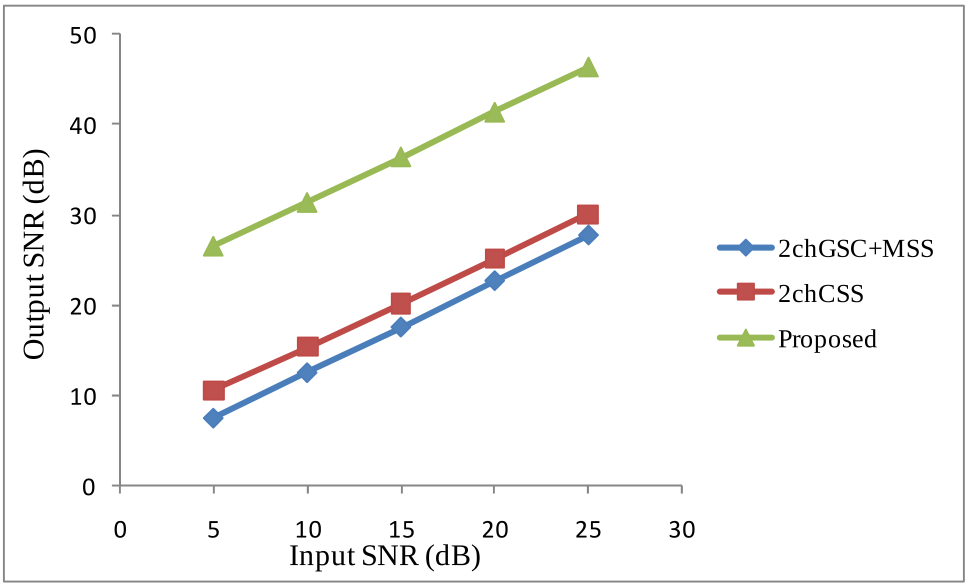


Fig.5-19.Output SNR of the 2chGSC+MSS, 2chCSS and proposed method

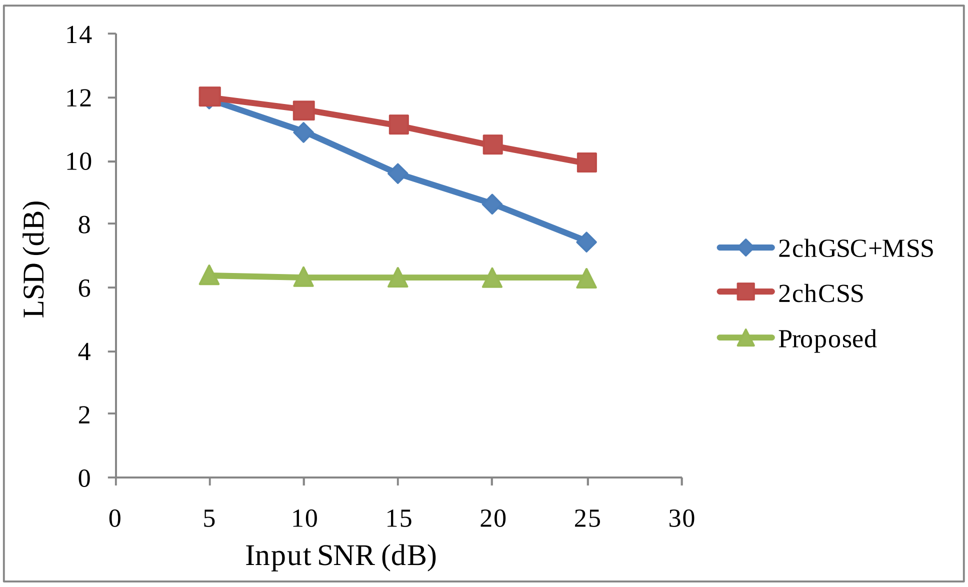


Fig.5-20.LSD improvement of the 2chGSC+MSS, 2chCSS and proposed method

Experiment5. Investigation on two-channel NR techniques with Pink noise

In this experiment, the pink noise was added to the speech signal. For the subtraction factor, $\alpha_0 = 1$ and $\alpha_{\max} = 1.5$ were chosen. From Table 5-15, it can be seen that the 2chGSC+MSS gives better NA performance than MSS one. However, the 2chGSC+MSS gives high level of speech distortion, as shown from the LSD measurement in Table 5-16. In Table5-17, the 2chGSC+MSS technique with variable $\alpha(k,l)$ gives better NA performance than the fixed one for high values of SNR. Moreover, low level of speech distortion is obtained by using the 2chGSC+MSS technique with variable $\alpha(k,l)$. From Table 5-19 and 5-20, the proposed hybrid technique gives better NA performance and less speech distortion than the other two investigated techniques. The spectrogram plots of these investigated NR techniques can be seen in Fig.5-21. It can be seen that the effect of additive pink noise is alleviated for both the 2chCSS and hybrid methods. The hybrid method demonstrates higher level of speech spectral components both in low and high frequency regions. This confirms with the NA performance and LSD measurement in Table 5-19 and 5-20.

Table.5-15. NA performance of MSS and 2chGSC+MSS techniques

Input SNR (dB)	Output SNR (dB)	
	MSS	2chGSC+MSS
5	5.57	12.20
10	10.55	16.92
15	15.54	20.87
20	20.54	24.40
25	25.54	26.77

Table.5-16. LSD measurement of MSS and 2chGSC+MSS techniques

Input SNR (dB)	LSD (dB)	
	MSS	2chGSC+MSS
5	2.80	5.10
10	2.07	4.31
15	1.64	3.72
20	1.54	3.45
25	1.47	3.32

Table.5-17. NA performance of the 2chGSC technique with fix α and variable $\alpha(k,l)$

Input SNR (dB)	Output SNR (dB)	
	2chGSC+MSS	Method I
5	12.20	10.63
10	16.92	15.61
15	20.87	20.60
20	24.40	25.60
25	26.77	30.61

Table.5-18. LSD measurement of the 2chGSC technique with fix α and variable $\alpha(k,l)$

Input SNR (dB)	LSD (dB)	
	2chGSC+MSS	Method I
5	5.10	4.35
10	4.31	3.40
15	3.72	2.83
20	3.45	2.63
25	3.32	2.43

Table.5-19. NA performance of the 2chGSC+MSS, 2chCSS and proposed technique

Input SNR (dB)	Output SNR (dB)		
	2chGSC+MSS	2chCSS	Proposed
5	12.20	12.45	20.71
10	16.92	17.45	25.71
15	20.87	22.45	30.70
20	24.40	27.45	35.71
25	26.77	32.45	38.35

Table.5-20. LSD measurement of the 2chGSC+MSS, 2chCSS and proposed technique

Input SNR (dB)	LSD (dB)		
	2chGSC+MSS	2chCSS	Proposed
5	5.10	4.39	3.60
10	4.31	4.03	3.55
15	3.72	3.77	3.53
20	3.45	3.64	3.52
25	3.32	3.58	3.51

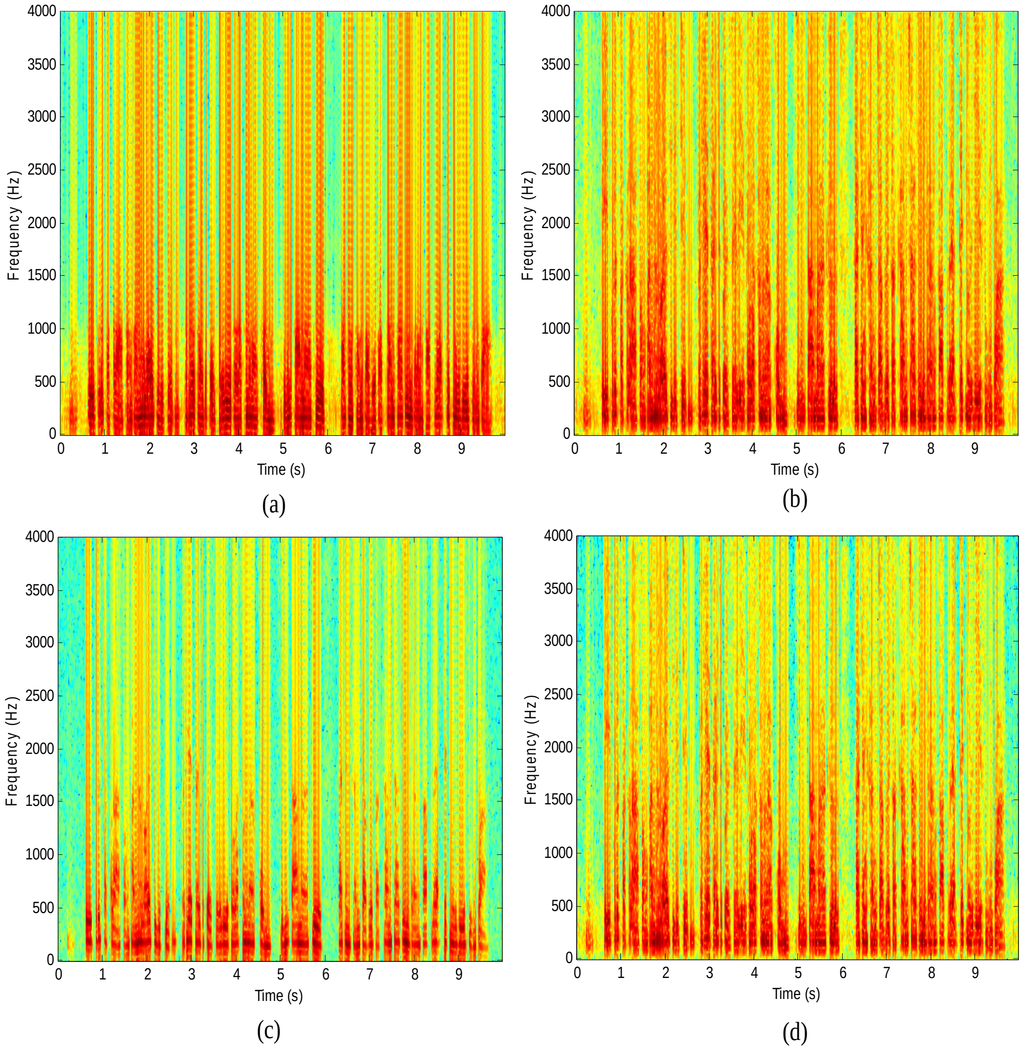


Fig. 5-21. Spectrogram plots for the enhanced speech signals using (a) the MSS technique, (b) the 2chGSC+MSS technique, (c) the 2chCSS method, and (d) the proposed NR technique, when the input SNR = 25 dB with pink noise

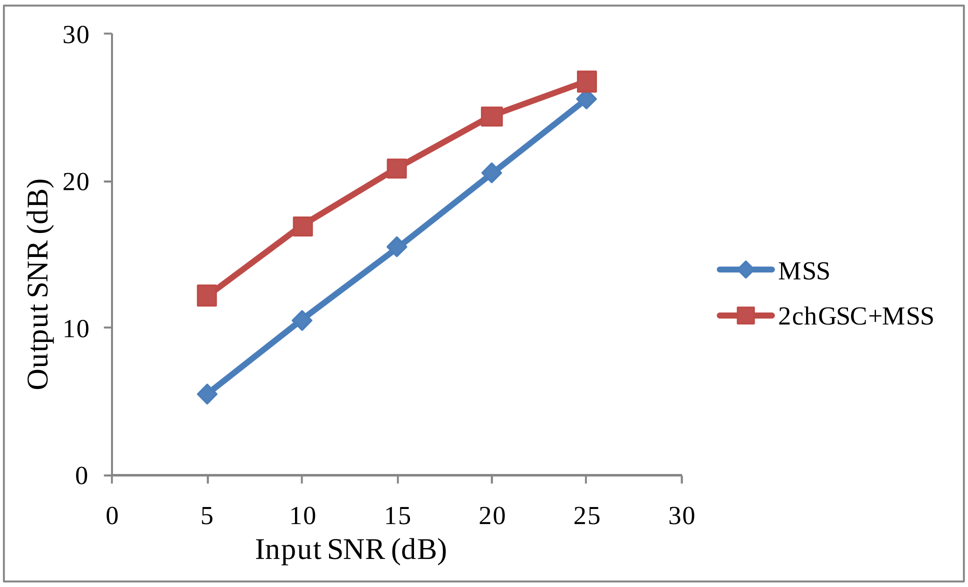


Fig.5-22.Output SNR of MSS and 2chGSC+MSS techniques with pink noise

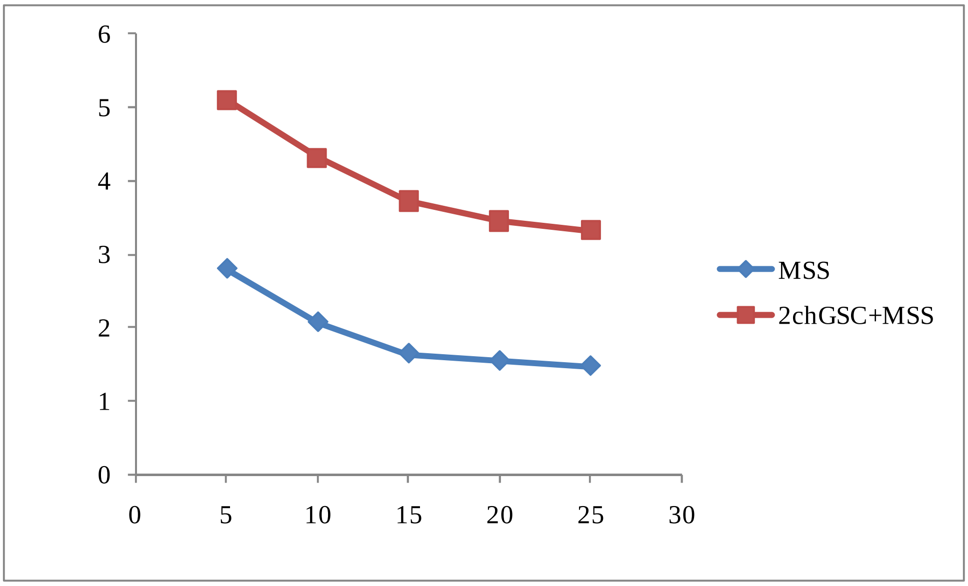


Fig.5-23.LSD measurement of MSS and 2chGSC+MSS techniques with pink noise

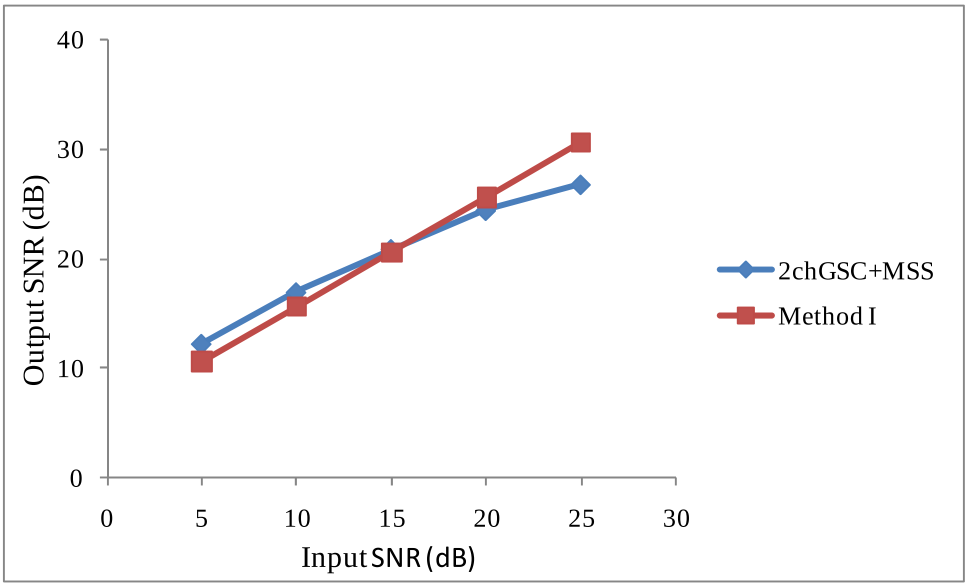


Fig.5-24.Output SNR of 2chGSC+MSS with fixed α and variable $\alpha(k,l)$ with pink noise

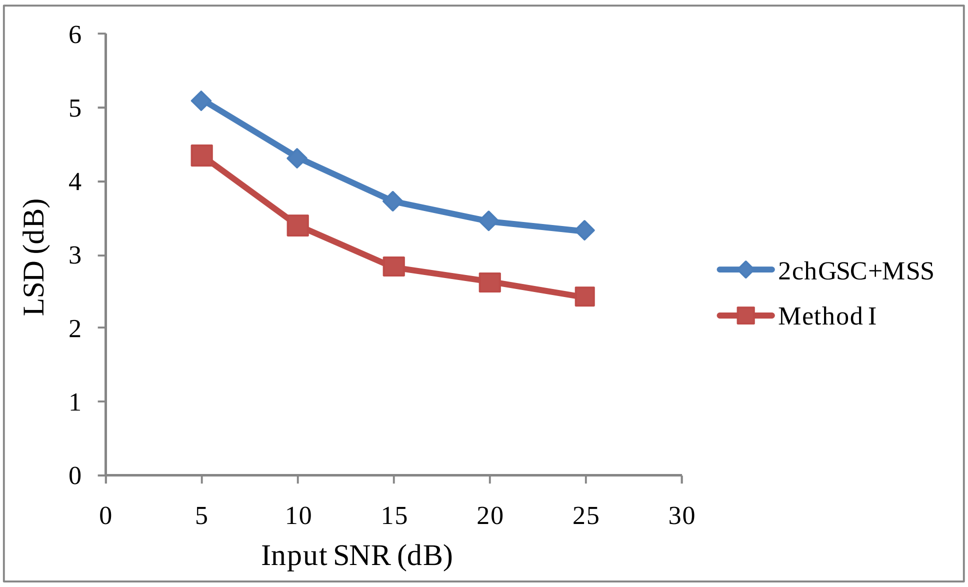


Fig.5-25.LSD measurement of 2chGSC+MSS with fixed α and variable $\alpha(k,l)$ with pink noise

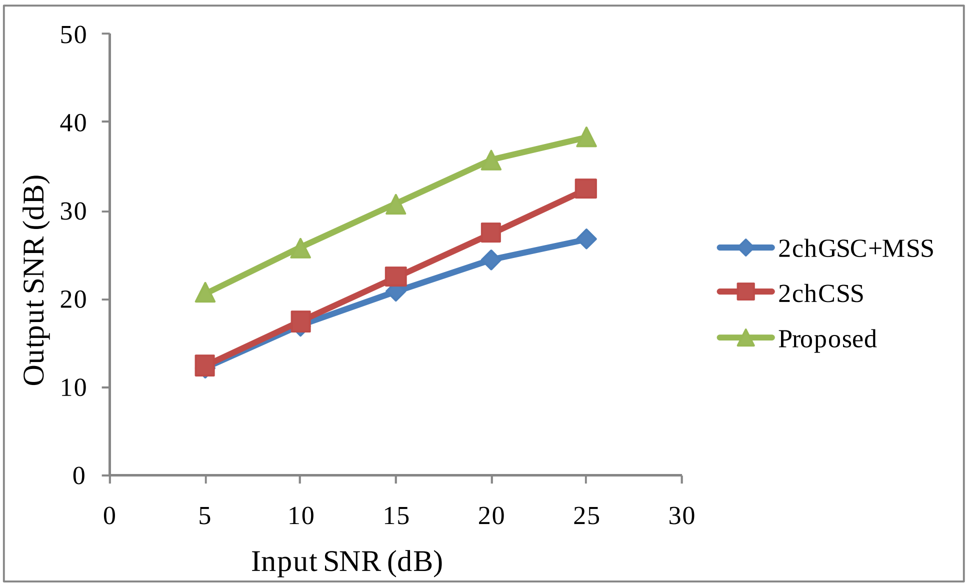


Fig.5-26.Output SNR of the 2chGSC+MSS, 2chCSS and proposed method with pink noise

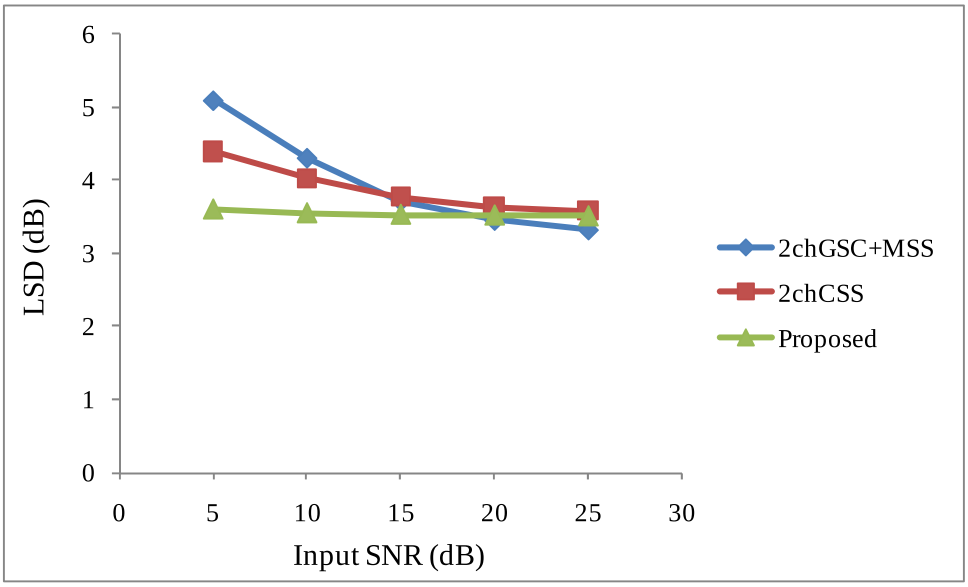


Fig.5-27.LSD measurement of the 2chGSC+MSS, 2chCSS and proposed method with pink noise

Experiment6. Investigation on two-channel NR techniques with Factory noise

In this experiment, factory noise was added to the speech signal. For the subtraction factor, $\alpha_0 = 1$ and $\alpha_{\max} = 1.5$ were chosen. As in experiment5, 2chGSC+MSS gives better NA performance than MSS one with the high LSD level. And the 2chGSC+MSS with variable $\alpha(k,l)$ gives better NA performance than the fixed one for high values of SNR. However, for speech distortion, 2chGSC+MSS with variable $\alpha(k,l)$ gives less amount of speech distortion for every SNR level. Moreover, the proposed hybrid technique yields the best NA performance and less amount of speech distortion among these NR techniques. It can be clearly seen in Table 5-21 – 5-26. The spectrogram plots of these investigated NR techniques can be seen in Fig.5-28.

Table.5-21. NA performance of MSS and 2chGSC+MSS techniques

Input SNR (dB)	Output SNR (dB)	
	MSS	2chGSC+MSS
5	5.43	12.06
10	10.38	16.63
15	15.36	20.47
20	20.36	23.78
25	25.35	26.43

Table.5-22. LSD measurement of MSS and 2chGSC+MSS techniques

Input SNR (dB)	LSD (dB)	
	MSS	2chGSC+MSS
5	3.35	6.01
10	2.69	5.03
15	2.07	4.25
20	1.88	3.68
25	1.60	3.40

Table.5-23. NA performance of the 2chGSC technique with fix α and variable $\alpha(k,l)$

Input SNR (dB)	Output SNR (dB)	
	2chGSC+MSS	Method I
5	12.06	10.57
10	16.63	15.52
15	20.47	20.50
20	23.78	25.50
25	26.43	30.5

Table.5-24. LSD measurement of the 2chGSC technique with fix α and variable $\alpha(k,l)$

Input SNR (dB)	LSD (dB)	
	2chGSC+MSS	Method I
5	6.01	5.06
10	5.03	4.30
15	4.25	3.50
20	3.68	2.89
25	3.40	2.64

Table.5-25. NA performance of the 2chGSC+MSS, 2chCSS and proposed technique

Input SNR (dB)	Output SNR (dB)		
	2chGSC+MSS	2chCSS	Proposed
5	12.06	11.46	16.92
10	16.63	16.45	21.88
15	20.47	21.44	26.87
20	23.78	26.44	31.86
25	26.43	31.44	36.86

Table.5-26. LSD measurement of the 2chGSC+MSS, 2chCSS and proposed technique

Input SNR (dB)	LSD (dB)		
	2chGSC+MSS	2chCSS	Proposed
5	6.01	4.78	3.43
10	5.03	4.37	3.36
15	4.25	4.01	3.33
20	3.68	3.75	3.31
25	3.40	3.62	3.30

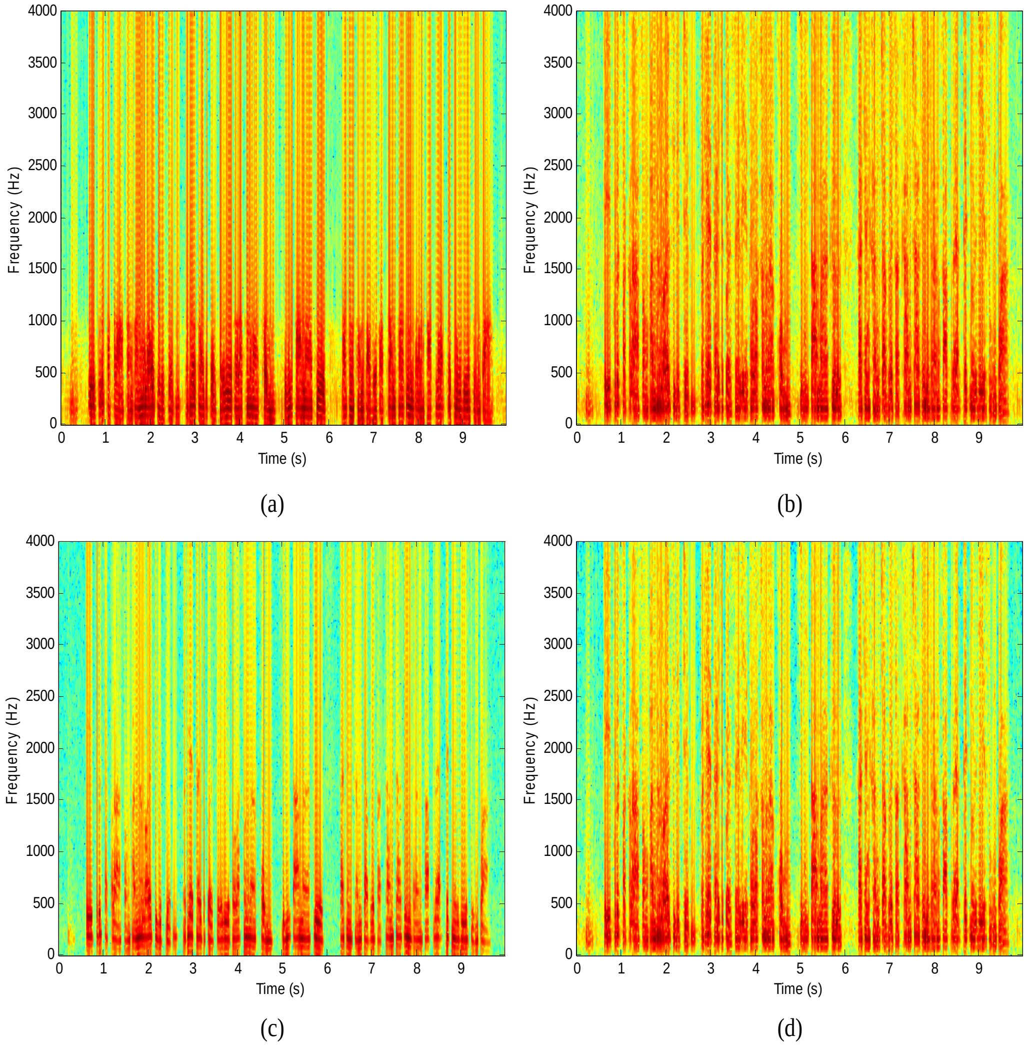


Fig. 5-28. Spectrogram plots for the enhanced speech signals using (a) the MSS technique, (b) the 2chGSC+MSS technique, (c) the 2chCSS method, and (d) the proposed NR technique, when the input SNR = 25 dB with factory noise

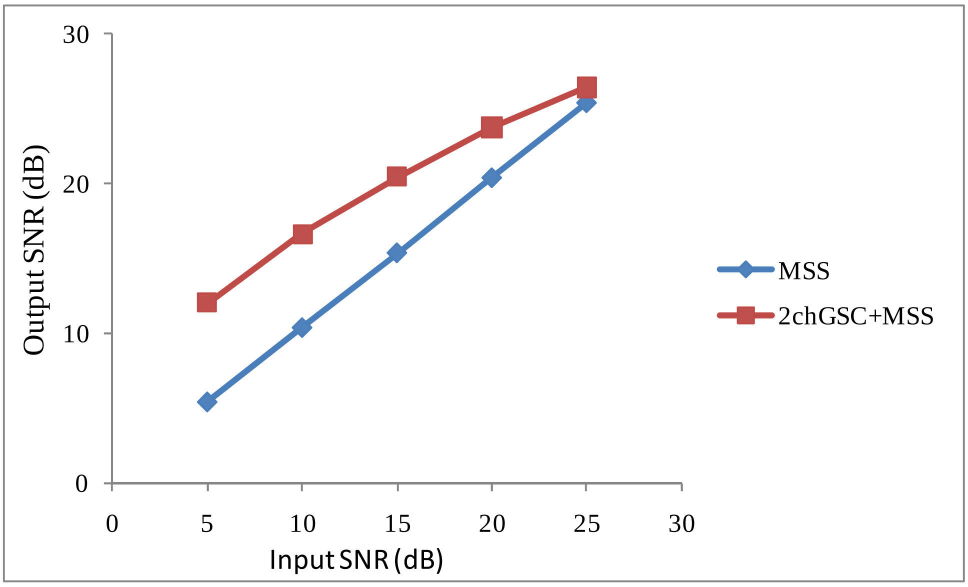


Fig.5-29.Output SNR of MSS and 2chGSC+MSS techniques with factory noise

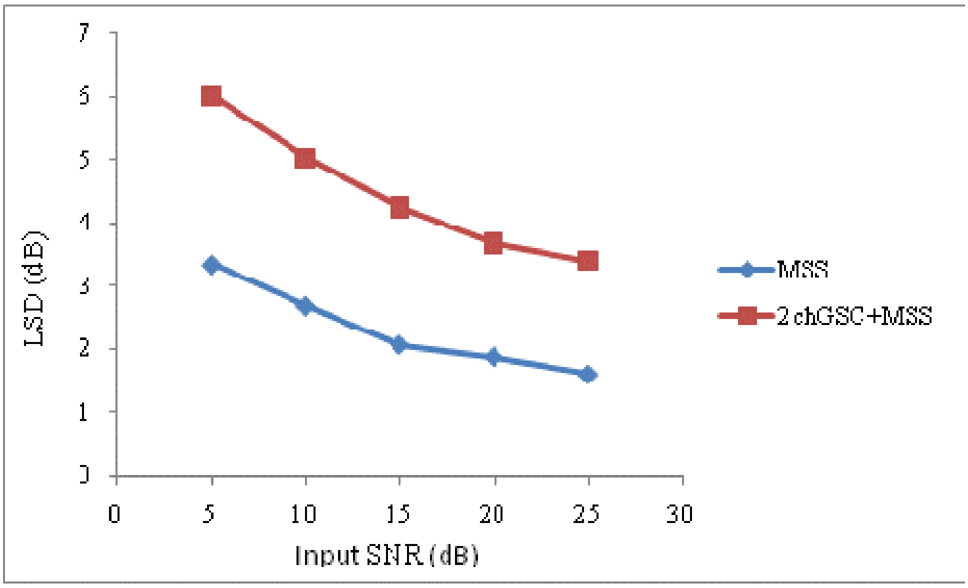


Fig.5-30.LSD measurement of MSS and 2chGSC+MSS techniques with factory noise

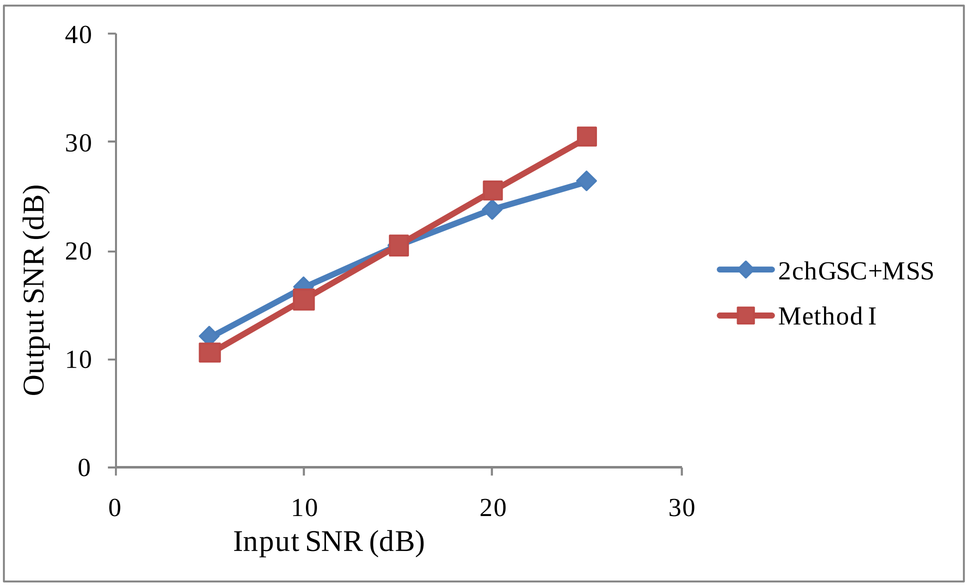


Fig.5-31.Output SNR of 2chGSC+MSS with fixed α and variable $\alpha(k,l)$ with factory noise

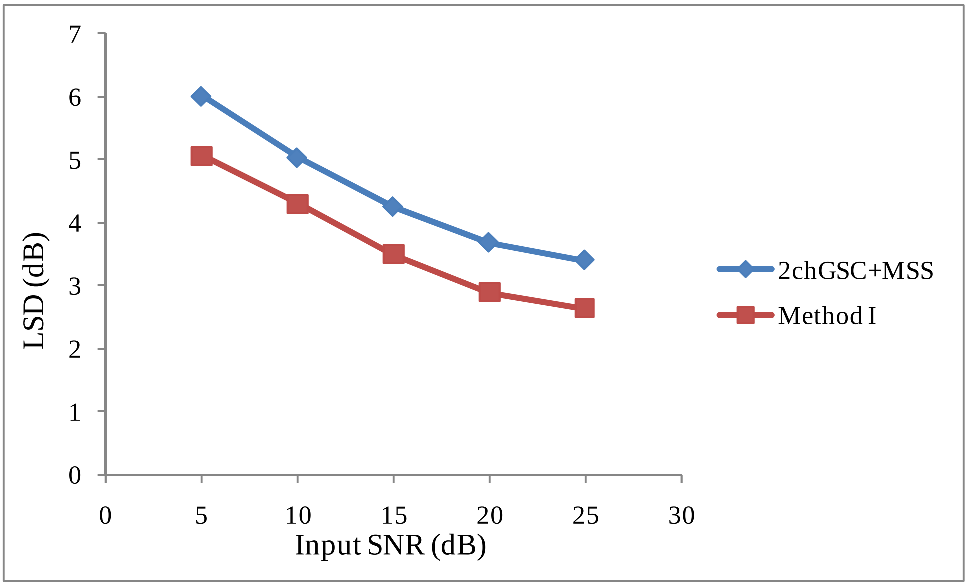


Fig.5-32.LSD measurement of 2chGSC+MSS with fixed α and variable $\alpha(k,l)$ with factory noise

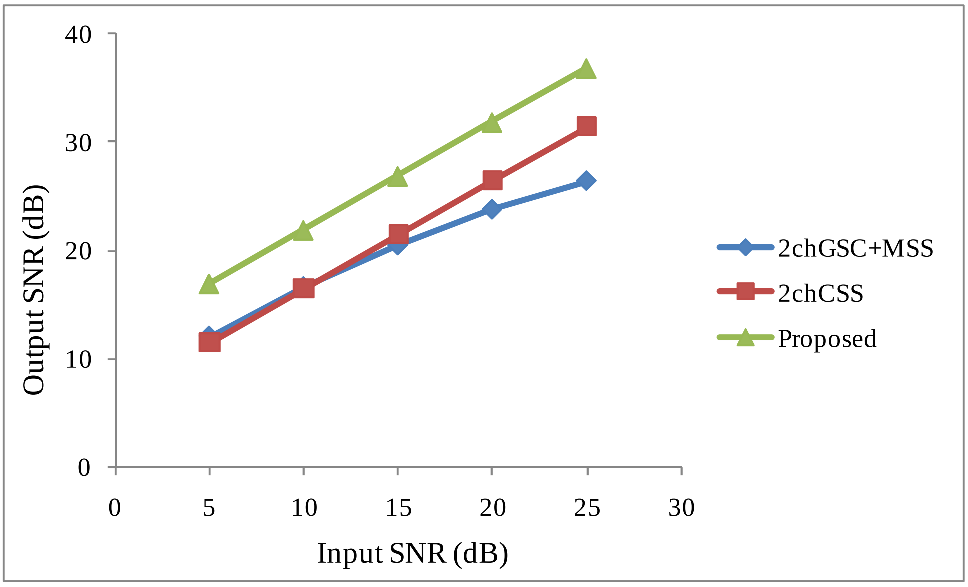


Fig.5-33.Output SNR of the 2chGSC+MSS, 2chCSS and proposed method with factory noise

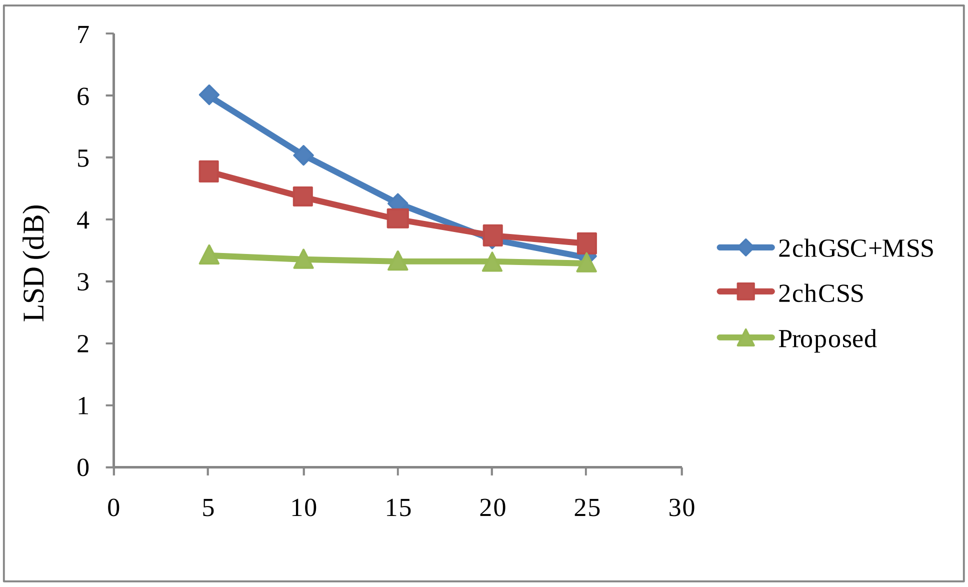


Fig.5-34.LSD measurement of the 2chGSC+MSS, 2chCSS and proposed method with factory noise

Experiment7. Investigation on two-channel NR techniques with Car noise

In this experiment, the car noise was added to the speech signal. For the subtraction factor, $\alpha_0 = 1$ and $\alpha_{\max} = 1.5$ were chosen. As in experiment 5 and 6, 2chGSC+MSS gives better NA performance than MSS one while the 2chGSC+MSS with variable $\alpha(k,l)$ gives better NA performance than the fixed one. Among these NR techniques, the proposed hybrid technique yields the best NA performance with less amount of speech distortion than the other investigated techniques. The spectrogram plots of these investigated NR techniques can be seen in Fig.5-35.

Table.5-27. NA performance of MSS and 2chGSC+MSS techniques

Input SNR (dB)	Output SNR (dB)	
	MSS	2chGSC+MSS
5	6.78	16.10
10	11.61	19.13
15	16.56	21.38
20	21.54	24.43
25	26.54	26.61

Table.5-28. LSD measurement of MSS and 2chGSC+MSS techniques

Input SNR (dB)	LSD (dB)	
	MSS	2chGSC+MSS
5	1.44	4.74
10	1.42	4.71
15	1.41	4.68
20	1.37	4.66
25	1.38	4.64

Table.5-29. NA performance of the 2chGSC technique with fix α and variable $\alpha(k,l)$

Input SNR (dB)	Output SNR (dB)	
	2chGSC+MSS	Method I
5	16.10	10.51
10	19.13	15.34
15	21.38	20.29
20	24.43	25.27
25	26.61	30.26

Table.5-30. LSD measurement of the 2chGSC technique with fix α and variable $\alpha(k,l)$

Input SNR (dB)	LSD (dB)	
	2chGSC+MSS	Method I
5	4.74	2.38
10	4.71	2.37
15	4.68	2.37
20	4.66	2.35
25	4.64	2.34

Table.5-31. NA performance of the 2chGSC+MSS, 2chCSS and proposed technique

Input SNR (dB)	Output SNR (dB)		
	2chGSC+MSS	2chCSS	Proposed
5	16.10	32.68	34.41
10	19.13	37.67	39.31
15	21.38	42.67	44.29
20	24.43	47.67	49.28
25	26.61	52.67	54.28

Table.5-32. LSD measurement of the 2chGSC+MSS, 2chCSS and proposed technique

Input SNR (dB)	LSD (dB)		
	2chGSC+MSS	2chCSS	Proposed
5	4.74	4.26	4.87
10	4.71	4.23	4.7
15	4.68	4.19	4.61
20	4.66	4.18	4.57
25	4.64	4.17	4.56

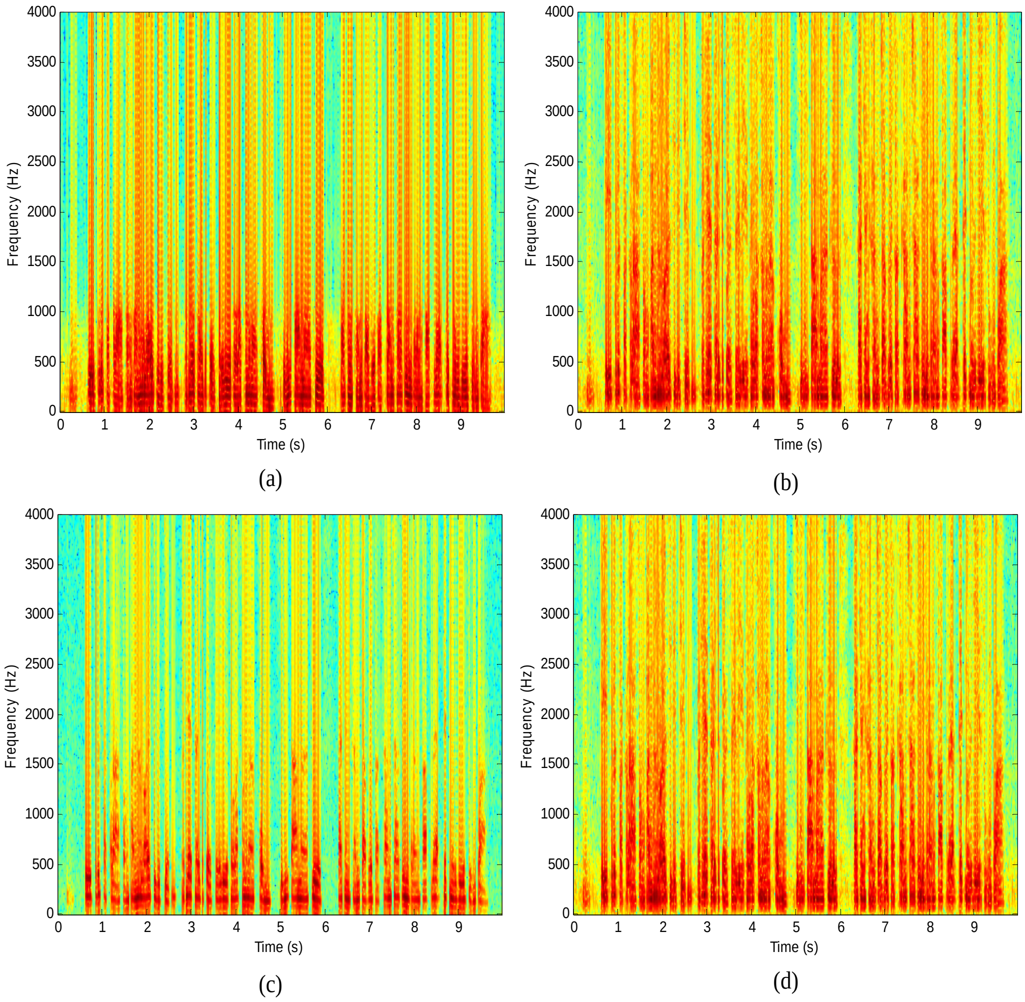


Fig. 5-35. Spectrogram plots for the enhanced speech signals using (a) the MSS technique, (b) the 2chGSC+MSS technique, (c) the 2chCSS method, and (d) the proposed NR technique, when the input SNR = 25 dB with car noise

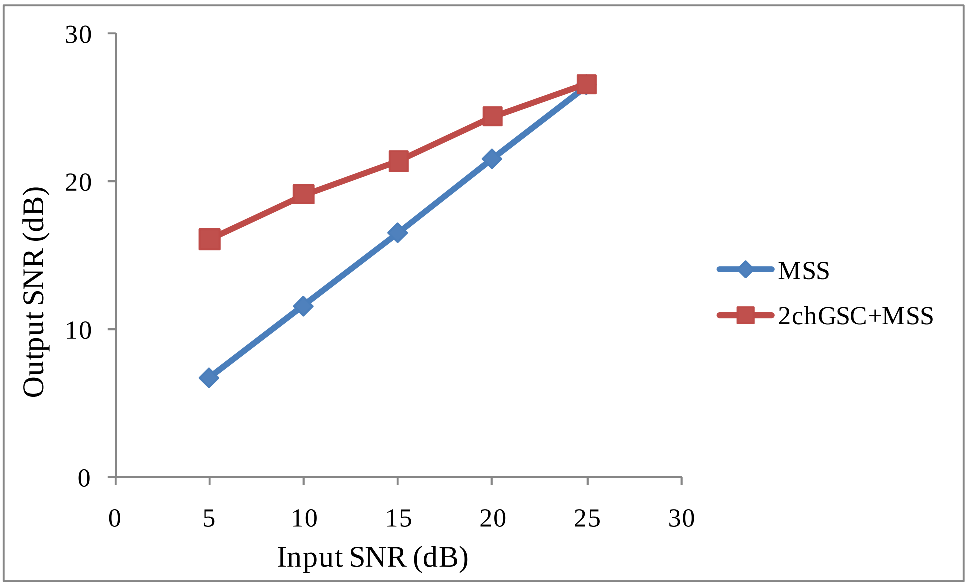


Fig.5-36.Output SNR of MSS and 2chGSC+MSS techniques with car noise

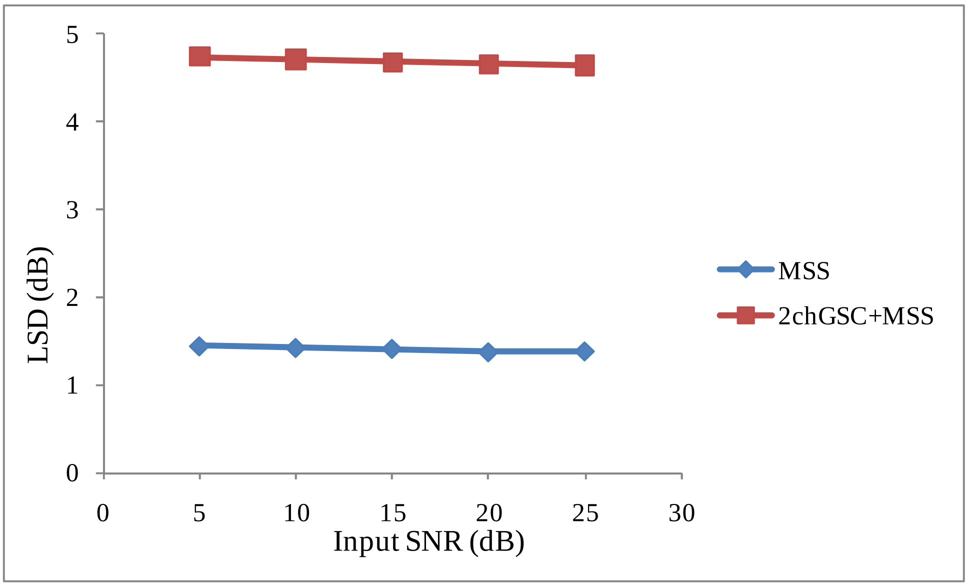


Fig.5-37.LSD measurement of MSS and 2chGSC+MSS techniques with car noise

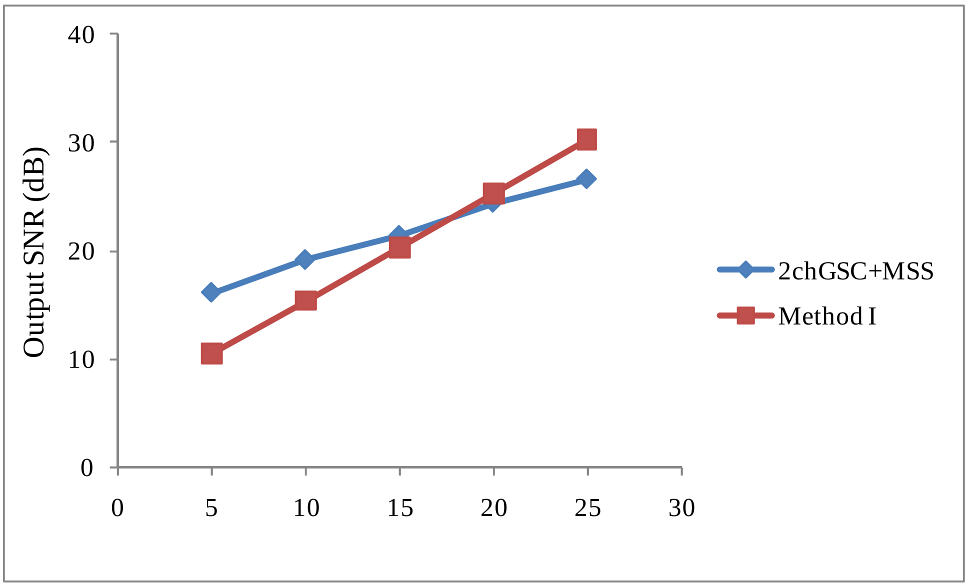


Fig.5-38.Output SNR of 2chGSC+MSS with fixed α and variable $\alpha(k,l)$ with car noise

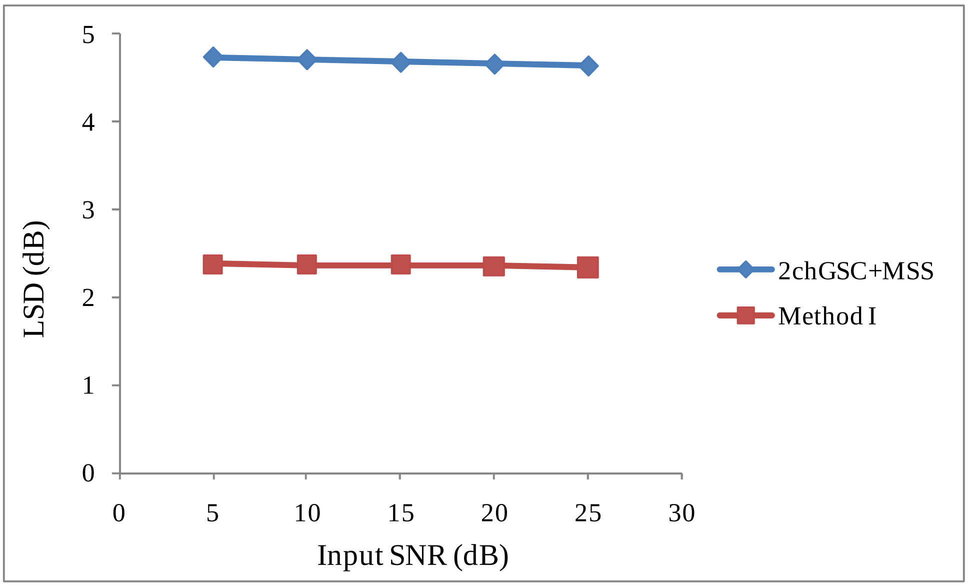


Fig.5-39.LSD measurement of 2chGSC+MSS with fixed α and variable $\alpha(k,l)$ with car noise

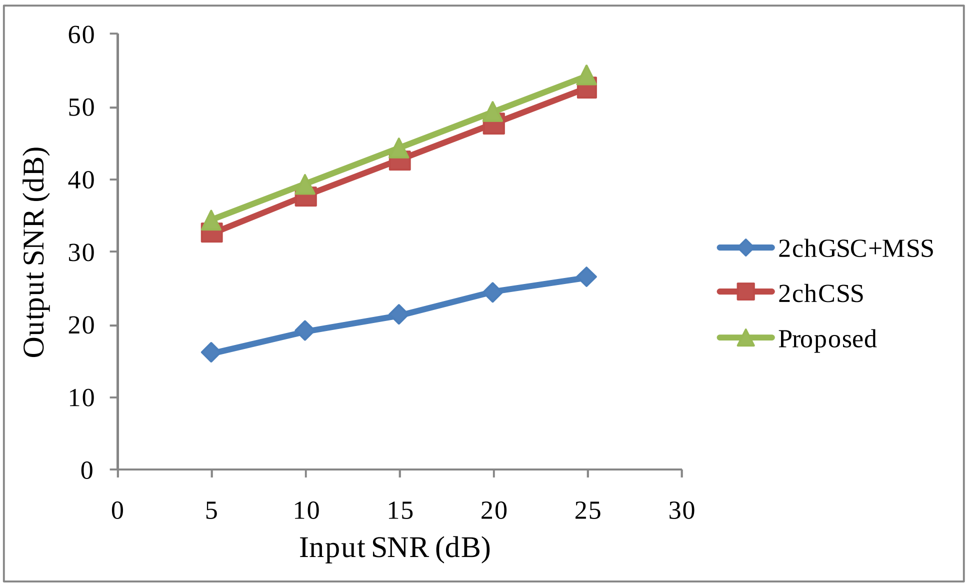


Fig.5-40.Output SNR of the 2chGSC+MSS, 2chCSS and proposed method with car noise

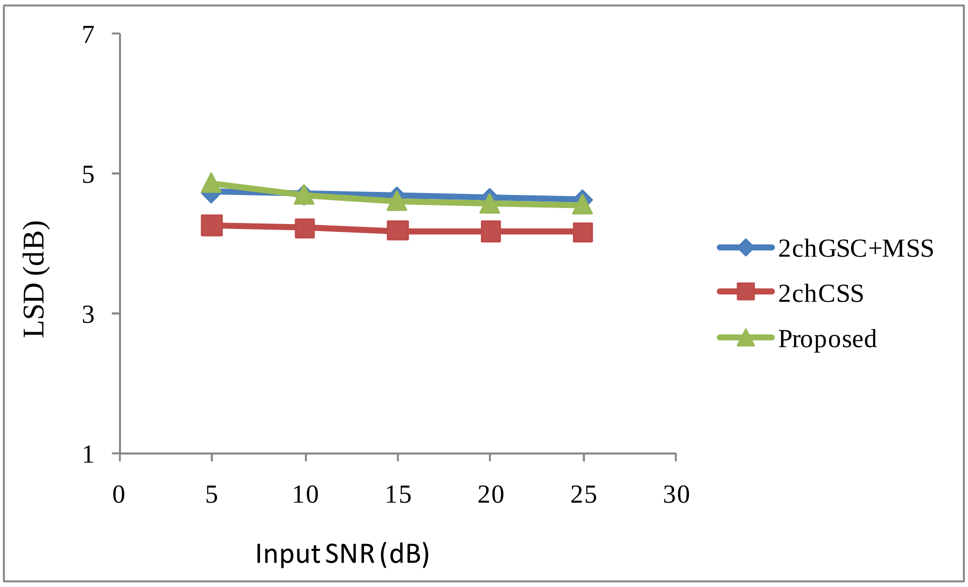


Fig.5-41.LSD measurement of the 2chGSC+MSS, 2chCSS and proposed method with car noise

5.3. Subjective Evaluations

All the noise reduction algorithm; MSS, 2chGSC+MSS (fix α) , 2chGSC+MSS (variable α), 2chCSS and the proposed hybrid structure were also evaluated with the subjective listening tests. The enhanced speeches of these NR algorithms were randomly presented to each listener through the microphone. The quality of the enhanced speech was demonstrated based on their preference in terms of mean opinion score (MOS): 1=bad, 2=poor, 3=fair, 4=good, 5=excellent [4] Among the tested NR algorithms, the proposed hybrid structure has the highest MOS rates. Therefore, the subjective listening tests can confirm that the proposed hybrid structure is successful in reducing the multiple background noises while preserving the quality of the speech signal.

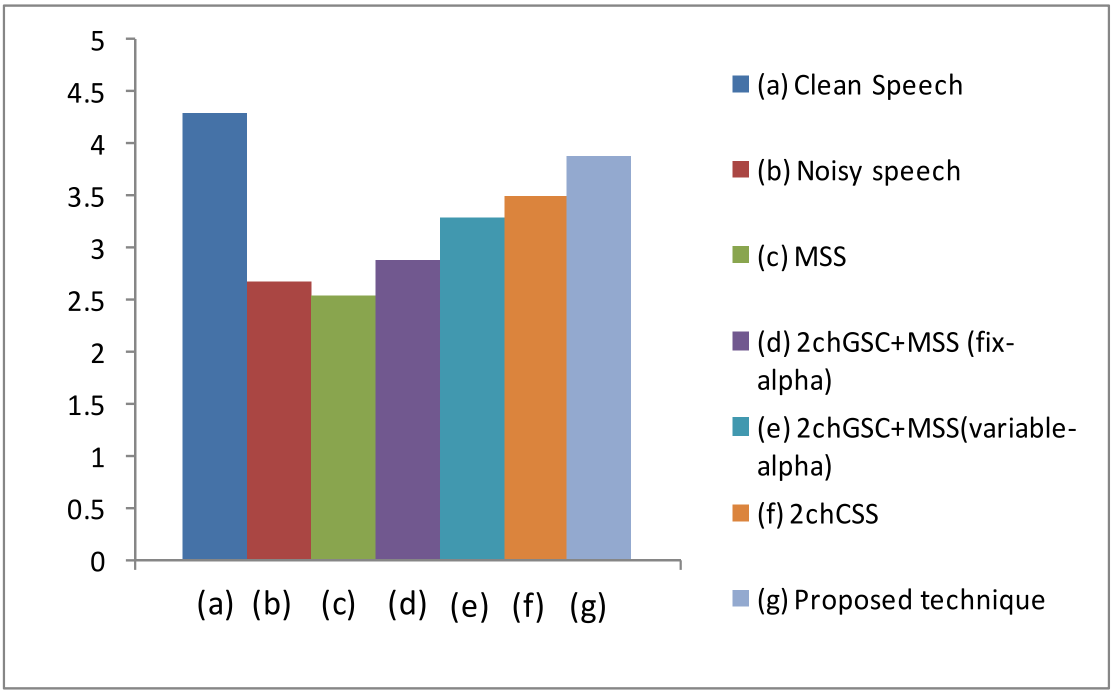


Fig.5-42. Mean opinion score (MOS) of the MSS, 2chGSC+MSS with fixed and variable noise subtraction parameter, 2chCSS and the proposed hybrid structure

5.4. Summary

From the simulation results based on the objective and subjective measurements, it was clearly seen that the proposed hybrid structure can compromise the background noise with the efficient NA performance while the other NR algorithms still introduce a large amount of speech distortion.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

This dissertation focuses on the improvement of the frequency-domain noise reduction techniques, especially in the applications of digital hearing aids. The needs of digital hearing aids and their characteristics are given in Chapter 2. The problem of additive background noise and a discussion of existing Noise Reduction (NR) techniques including single-channel and multi-channel NR methods are given in Chapter 3. As for the conventional single-channel Spectral Subtraction (SS) method, although it has been found to reduce the background noise efficiency and simple to be implemented, speech distortion is, however, unavoidable. The amount of speech distortion is proportional to the amount of noise reduction. Two approaches have been introduced in Chapter 4 in order to control the noise subtraction parameter of the SS method to be variable and frequency-dependent. By employing the two proposed methods, it has been shown via simulation results that it is possible to manage both the amount of noise reduction and speech distortion.

The Magnitude Spectral Subtraction (MSS), which is one version of the conventional frequency-domain NR technique, with the use of the two-channel Generalized Sidelobe Canceller (2chGSC), has been shown to eliminate the additive background noise sufficiently well. Despite the fact that the two-channel noise reduction technique that employs the modified Cross-Spectral Subtraction (2chCSS) can reduce the musical noise well, it, however, demonstrates a large amount of speech distortion. It has been proposed in Chapter 4 of this dissertation to obtain maximum advantages of both of these NR techniques via the hybrid structure; emphasizing on the achievement of noise reduction performance at minimal speech distortion level. For signal processing in the low-frequency region, the 2chGSC+MSS technique is implemented, while for the high-frequency region, the 2chCSS method is employed to reduce the musical noise. Comparative studies between the stand-alone 2chGSC+MSS and 2chCSS, and the proposed hybrid structure have been illustrated in Chapter 5, based on simulation results. The objective performance both in terms of noise attenuation and speech distortion support the proposed idea, especially for multiple-noise-source environments.

6.2 Future Work

The work within this dissertation could be further extended in a number of directions, as shown by the following issues:

- The utilizaton of Wavelet transform can be applied for signal analysis. [46] Various techniques based on wavelet have been proposed for denoising speech signal. The choice of wavelet function is to maximize the output Signal-to-Noise Ratio (SNR). However, the computational complexity of this wavelet-based NR technique is another main factor to be considered. Normally, the computational complexity of Discrete Wavelet Transform (DWT) increases with the number of vanishing moments. [47]

REFERENCES

[1] The Technology of Hearing Aids: Analog vs. Digital [Online]. 2010. Available from: http://www.hearingcentral.com/hearingaidtech.asp [2010,March]

[2] National Institute of health: Hearing Aids [Online]. 2010. Available from: http://www.nidcd.nih.gov/health/hearing/hearingaid.asp [2010,March]

[3] Y.Hu and P.C.Loizou. Subjective Comparison of Speech Enhancement Algorithms, Proc.ICASSP 2006, Vol. 1, pp. I-153 – I-156, May 2006.

[4] J. Li, M. Akagi and Y. Suzuki. A two-microphone noise reduction method in highly non-stationary multiple-noise-source environments, IEICE Trans. Fundamentals, Vol. E91-A, No. 6, pp. 1337-1346, June 2008.

[5] S.Haykin. Adaptive Filter Theory: Fourth edition.. Prentice Hall Information and System Sciences Series (2002): .

[6] McOlash, S.M.; Niederjohn, R.J. and Heinen, J.A. A Spectral Subtraction Method for the Enhancement of Speech Corrupted by Non-White, Non-Stationary Noise, Proc. IEEE, Industrial Electronics, Control, and Instrumentation, Vol. 2, No. 20, pp. 872 – 877, November 1995.

[7] 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, 1979.

[8] S.V. Vaseghi. Advanced Digital Signal Processing and Noise Reduction: 2nd edition.. John Wiley and Sons Ltd (2000): .

[9] H. Chen and C-H.Wu. Speech enhancement based on audible noise spectrum and short-time spectral amplitude estimator, Electronics Letters, Vol. 38, No. 10, pp. 485 – 487, May 2002.

[10] J. Chen, J. Benesty, Y. Huang, and S. Doclo. New insights into the noise reduction Wiener filter, IEEE Trans. Audio, Speech, Lang. Process., Vol. 14, No. 4, pp. 1218–1234, July 2006.

[11] T.T. Aung, N. Tangsangiumvisai, and A. Nishihara. A frequency-selective noise spectral subtraction approach, Proceedings of the International Conference on Science and Engineering (ICSE), Myanmar, December 2010.

[12] Noise (music) [Online] 2010. Available from: http://en.wikipedia.org/wiki/Noise_%28music%29 [2010 July]

[13] J.Chen, Benesty and Y.Huang. A Minimum Distortion Noise Reduction Algorithm With Multiple Microphones, IEEE Trans. Audio, Speech, Language Processing, Vol.16, No. 3, pp. 481 – 493, March 2008.

[14] C. Marro, Y. Mahieux, and K. U. Simmer. Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering, IEEE Trans. Speech and Audio Processing, Vol. 6, No. 3, pp. 240–259, May 1998.

[15] K. U. Simmer, J. Bitzer, and C. Marro. Post-Filtering Techniques: Berlin.. Springer-Verlag (2001), ch. 3, pp. 39–60.

[16] Cohen,I. Multichannel post-filtering in nonstationary noise environments, IEEE Trans. Signal Processing, Vol. 52, No. 5, pp. 1149 – 1160, May 2004.

[17] H. Cox, R.M. Zeskind, and M.M. Owen. Robust adaptive beamforming, IEEE Trans. Acoust. Speech Signal Process., Vol. 35, No. 10, pp. 1365–1375, October 1987.

[18] Reuven,G., Gannot,S. and Cohen,I. Multichannel Acoustic Echo Cancellation and Noise Reduction in Reverberant Environments using the transfer-function GSC, Proc.ICASSP 2007, Vol. 1, pp. I-81 – I-84, April 2007.

[19] Choo Leng Koh and Stephan Weiss. Constant Beamwidth Generalized Sidelobe Canceller, IEEE/SP 13th Workshop, Statistical Signal Processing, pp.283 – 288 , 2005.

[20] Z.L.Yu and M.H.Er. An extended generalized sidelobe canceller in time and frequency domain, Proc. ISCAS 2004, Vol. 3, pp. III 629-632, May 2004.

[21] Whitmal,N.A.; Rutledge,J.C. Noise reduction algorithms for digital hearing aids, IEEE, Engineering in Medicine and Biology Society, Vol. 2, No. 8, pp. 1294 – 1295, November 1994.

[22] A. Guerin, R.L. Bouquin-Jeannes, and G. Faucon. A two-sensor noise reduction system: Applications for hands-free car kit, EURASIP Journal on Applied Signal Processing, No. 11, pp. 1125-1134, 2003.

[23] F.L. Wightman and D.J. Kistler. The dominant role of low-frequency interaural

Hearing Loss [Online] 2010. Available from:

time differences in sound localization, Journal on Acoustical Society of America, 91, 3 (March 1992): 1648-1661.

[24] Spriet,A., Moonen,M. and Wouters,J. Robustness analysis of multichannel Wiener filtering and generalized sidelobe cancellation for multimicrophone noise reduction in hearing aid applications, IEEE Trans. Speech and Audio Processing, Vol. 13, No. 4, pp. 487 – 503, July 2005.

[25] Australian Hearing: Causes of Hearing loss [Online] 2010. Available from: http://www.hearingawarenessweek.org.au/wordfiles/Causes%20of%20Hearing%20Loss.pdf [2010 August]

[26] Harvard Medical School Center For Hereditary Deafness: Common Causes of Hearing Loss [Online] 2010. Available from: http://hearing.harvard.edu/info/common-causes-of-hearingloss.pdf [2010 August]

[27] Harvey Dillon. Hearing Aids: Boomerang Press, pp. 53 – 56, 2001.

[28] National Instruments: Short-Time Fourier Transform [Online] 2010. Available from; http://zone.ni.com/reference/en-XX/help/372656A-01/lvasptconcepts/aspt_stft/ [2010 September]

[29] Y. Ephraim and H. L. Van Trees. A signal subspace approach for speech enhancement, IEEE Trans. Speech Audio Processing, Vol. 3, No. 4, pp. 251–266, July 1995.

[30] U. Mittal and N. Phamdo. Signal/noise KLT based approach for enhancing speech degraded by colored noise, IEEE Trans. Speech Audio Processing, Vol. 8, No. 6, pp. 159–167, March 2000.

[31] A. Rezayee and S. Gazor. An adaptive KLT approach for speech enhancement, IEEE Trans. Speech Audio Processing, Vol. 9, No. 2, pp. 87–95, February 2001.

[32] Y. Hu and P. C. Loizou. A subspace approach for enhancing speech corrupted by colored noise, in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Vol. 1, No. 8, pp. 573–576 May 2002.

[33] M. Dendrinos, S. Bakamidis, and G. Carayannis. Speech enhancement from noise: A regenerative approach, Speech Commun., vol. 10, pp. 45–57, 1991.

[34] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sørensen. Reduction of broad-band noise in speech by truncated QSVD, IEEE Trans. Speech Audio Processing, Vol. 3, No. 6, pp. 439–448, November 1995.

[35] J.Benesty, S.Makino and J.Chen. Speech Enhancement: Signals and Communication Technology, Signal and Communication Technology, pp. 1 – 3, 2005.

[36] R. Thoonsaengngam and N. Tangsangiumvisai. The a priori SDR Estimation Techniques with Reduced Speech Distortion for Acoustic Echo and Noise Suppression, IEICE Trans. Commun, Vol. E92-B, No. 10, pp. 3022-3033, October 2009.

[37] D. Rudoy, P.Basu, T.F. Quatieri, B.Donn and P.J. Wolfe. Adaptive short-time analysis-synthesis for Speech Enhancement, ICASSP, Acoustics, Speech and Signal Processing, pp. 4905-4908, May 2008.

[38] Beamforming Basics [Online] 2010. Available from: http://cnx.org/content/m12563/latest/ [2010 October]

[39] L.J. Griffiths and C.W. Jim. An alternative approach to linearly constrained adaptive beamforming, IEEE Trans. Antennas Propag., Vol. AP-30, No. 1, pp. 27–34, January 1982.

[40] J. Bitzer, K.U. Simmer, and K.D. Kammeyer. Multichannel noise reduction—Algorithms and theoretical limits, Proc. EUSIPCO1998, pp.105–108, 1998.

[41] H.Y. Kim, F. Asano, Y. Suzuki, and T. Sone. Speech enhancement based on short-time spectral amplitude estimation with two channel beamformer, IEICE Trans.Fundamentals, Vol. E79-A, No. 12, pp. 2151–2158, December 1996.

[42] J. Li, M. Akagi and Y. Suzuki. Extension of the Two-Microphone Noise Reduction Method for Binaural Hearing Aids, ICALIP 2008, Audio, Language and Image Processing, pp.97-101, July 2008.

[43] Y. Ephraim and D. Malah. Speech enhancement using a minimum mean square error short-time spectral amplitude estimator, IEEE Trans. Acoustics, Speech, and Signal Processing, Vol. 32, No. 6, pp. 1109-1121, 1984.

[44] O. Cappe. Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor, IEEE Trans. Speech Audio Processing, Vol.2, No.2, pp. 345-349, April 1994.

[45] A. A. Azirani, R.L. Bouquin-Jeannes, and G. Faucon. Enhancement of speech degraded by coherent and inconherent noise using cross-spectral subtraction, IEEE Trans. Speech, and Audio Processing, Vol. 5, No. 5, pp. 484-487, 1997.

[46] K. Ramchandran, M. Vetterli, and C. Herley. Wavelets, Subband Coding, and Best Bases, Invited paper in the proceedings of the IEEE, vol. 84, no. 4, April 1996.

[47] M.S. Chavan, M.N.Chavan, and M.S. Gaikwad. Studies on Implementation of Wavelet for Denoising Speech Signal, International Journal of Computer Applications, vol. 3, no. 2, June 2010.

[48]T.T.Aung, N.Tangsangiumvisai, A.Nishihara, Submitted to International Journal on Signal Processing, Elsevier, September 2011.

[49] J.Sohn and W.Sung. A voice activity detector employing soft decision based noise spectrum adaptation, Proc. ICASSP 1998, Vol. 1, pp. 365-368, May 1998.

[50] J.Sohn, N.S.Kim and W.Sung. A statistical model-based voice activity detection, IEEE Trans. Signal Processing, Vol. 6, pp. 1-3, January 1999.

[51] D.Enqing, L.Guizhong, Z.Yatong and C.Yu. Voice activity detection based on short-time energy and noise spectrum adaptation, Proc. ICOSP 2002, Vol. 1, pp. 464-467, August 2002.

[52] L.Krasny and S. Oraintara. Voice activity detector for microphone array processing in hand-free systems, Sensor Array and Multichannel Signal Processing Workshop Processings, pp. 224-228, August 2002.

[53] J.Ramirez, J.C. Segura, C. Benitez, A.de la Torre and A.Rubio. A new voice activity detector using subband order-statistics filters for robust speech recognition, Proc. ICASSP 2004, Vol. 1, pp. I-849 - I-852, May 2004.

Appendix

<u>List of Publications</u>

The following publications were written during the course of this research work.

1. T. T. Aung, N. Tangsangiumvisai , and A. Nishihara, “A comparative study on noise reduction techniques in multiple-noise-source environments”, Proceedings of the 2nd International Worlshop on Multimedia and Communication Technology (ISMAC 2010, Manila, PHILIPPINES), September 2010.

2. T. T. Aung, N. Tangsangiumvisai , and A. Nishihara, “A Frequency-selective Noise Spectral Subtraction Approach”, Proceedings of the International Conference on Science and Engineering (ICSE 2010, Yangon, MYANMAR), December 2010.

3. T. T. Aung, N. Tangsangiumvisai , and A. Nishihara, “A hybrid structure for noise reduction technique in hearing aids”, submitted to an International Journal of Signal Processing, Elsevier, since 4 September 2011.

Vitae

Thiri Thandar Aung was born in 1982 in Yenanchaung, Myanmar. She was awarded a Bachelor and Master degree of Electronics Engineering at the Department of Electronics Engineering, Mandalay Technological University and Yangon Technological University in 2004 and 2006. She has worked as a Lecturer in Mawlamyine Technological University since 2006. She has been a graduate student in the Doctoral Degree Program in Electrical Engineering at the Department of Electrical Engineering, Chulalongkorn University, Thailand, since 2008.