Development of Novel Adaptive Beamforming Approaches for Multi-Channel Speech Enhancement

Submitted in partial fulfillment of the requirements
for the award of the degree of

DOCTOR OF PHILOSOPHY

by

S. SIVA PRIYANKA

(Roll No. 717030)

Under the Supervision of
Prof. T. KISHORE KUMAR

Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY
WARANGAL – 506004, T.S, INDIA

December – 2021

Dedicated to my beloved

Teachers, my mother, and my family

APPROVAL SHEET

This thesis entitled "**Development of Novel Adaptive Beamforming Approaches for Multi-Channel Speech Enhancement**" by **Smt.** S. Siva Priyanka is approved for the degree of **Doctor of Philosophy**.

Examiners

____

____

Supervisor

____

Prof. T. Kishore Kumar

Professor, Electronics, and Communication Engineering Department,

NIT WARANGAL

Chairman

____

Dr. P. Sreehari Rao

Head, Electronics, and Communication Engineering Department,

NIT WARANGAL

Date:

Place:

DECLARATION

I, hereby, declare that the matter embodied in this thesis entitled "**Development of Novel Adaptive Beamforming Approaches for Multi-Channel Speech Enhancement"** is based entirely on the results of the investigations and research work carried out by me under the supervision of **Prof. T. Kishore Kumar**, Department of Electronics and Communication Engineering, National Institute of Technology Warangal. I declare that this work is original and has not been submitted in part or full for any degree or diploma to this or any other University.

I declare that this written submission represents my ideas in my own words and where other ideas or words have been included. I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated, or falsified any idea/date/fact/source in my submission. I understand that any violation of the above will cause disciplinary action by the institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

S. Siva Priyanka

Roll No: 717030

Date:

Place: Warangal

Department of Electronics and Communication Engineering
National Institute of Technology
Warangal – 506 004, Telangana, India

CERTIFICATE

This is to certify that the dissertation work entitled **"Development of Novel Adaptive Beamforming Approaches for Multi-Channel Speech Enhancement,"** which is being submitted by Smt. S. Siva Priyanka (Roll No. 717030), is a bonafide work submitted to the National Institute of Technology Warangal in partial fulfillment of the requirement for the award of the degree of Doctor of Philosophy to the Department of Electronics and Communication Engineering of National Institute of Technology Warangal, is a record of bonafide research work carried out by her under my supervision and has not been submitted elsewhere for any degree.

Prof. T. Kishore Kumar
(Supervisor)
Professor, Department of ECE
National Institute of Technology
Warangal, India – 506004

ACKNOWLEDGEMENTS

I would like to thank a number of people who have contributed to my Ph.D. directly or indirectly and in different ways through their help, support, and encouragement.

Firstly, I would like to express my sincere gratitude to my supervisor **Prof. T. Kishore Kumar,** for the continuous support of my Ph.D. study and related research, for his patience, motivation, and guidance. My sincere thanks to him for providing me an opportunity to join the institute as a Ph.D. research scholar and giving me access to the research facilities. Without the precious support, it would not have been possible to conduct this research.

I thank the current Head of The Department, **Dr. P. Sreehari Rao**, and the former Heads, **Prof. L. Anjaneyulu, Prof. T. Kishore Kumar** and **Prof. N. Bheema Rao**, for providing all the necessary support from the department.

Besides my supervisor, I would like to thank the rest of my Doctoral Scrutiny Committee members**: Dr. P. Sreehari Rao, Dr. S. Anuradha, and Prof. Y.N. Reddy,** for their insightful comments and encouragement, but also for the hard questions which incented me to widen my research from various perspectives.

I thank **Prof. N.V. Ramana Rao**, Director, NIT Warangal, for providing official support and financial assistance to carry out the research.

I thank my teachers **Dr. P. Ravi Kumar, Dr. Grace Shanti, and Dr. Anuradha,** for providing moral support throughout my research work.

I thank my past and present fellow speech processing lab-mates at NIT Warangal, **Dr. P. Rakesh, Dr. Sudeep Surendran, Dr. B. Surekha Reddy, Dr. N. Prasad, Dr. N. Siva Prasad, Dr. V. Sunnydayal, Mr. B. Ravi and Mr. K Sunil Kumar,** with whom I have had the pleasure to work with.

I thank my friends, **Mr. Muhibullah Shah Pahelwan, and Mr. N. Gowtham Reddy,** for being with me throughout the journey and for their moral support.

Last but not least, I would like to thank my mothers, **Mrs. S. Veeramma and Mrs. A. Tulsiamma,** my brothers **S. Manohar Reddy and S. Madhusudhan Reddy and their**

families, my in-laws**, L. Tulsi Reddy, L. Gangi Reddy, R. Sankar Reddy, and R. Lakshmi Narasamma,** for supporting me throughout the Ph.D. period and in my life in general.

Particularly my mother, Mrs. S. Veeramma, my husband R. Amarnath Reddy and my daughter Ruchiraamruta Reddy Rangareddy for supporting me and staying with me throughout the Ph.D. period, without you nothing is possible.

S. SIVA PRIYANKA

ABSTRACT

Adaptive beamforming plays a crucial role in Multi-Channel Speech Enhancement (MCSE), especially for applications like teleconferences, mobile phones, hearing aids, etc., where real-time situations create various noisy environments while communicating. Multi-channel speech enhancement (MCSE) is prominent for noise-free communication in noisy real-time environments. This thesis considers the significance of adaptive beamforming approaches for multi-channel speech enhancement.

Generalized Sidelobe Canceller (GSC) beamformer is one of the adaptive beamforming used for multi-channel speech enhancement. GSC structure comprises a Fixed Beamformer (FBF), Blocking Matrix (BM), and an adaptive filtering block. Adaptive filtering plays a vital role in noise cancellation in the GSC structure. Existing GSC beamforming with existing adaptive algorithms shows substandard noise cancellation in the sidelobe canceling path under real-time noisy environmental conditions. They are ineffective at low SNR, i.e., at -10 dB, and high SNR, i.e., at 15 dB. Existing GSC adaptive beamformers also suffer from directional and diffuse noise. In the case of directional and diffuse noise in low frequencies, most adaptive beamformers reduce less noise. Existing multi-channel speech enhancement (MCSE) also suffers from residual noise in the output, which diminishes the quality and intelligibility of the desired speech.

Novel adaptive beamforming approaches should be developed for multi-channel speech enhancement to address the existing issues. In this thesis, novel adaptive filters and postfilter are implemented to the GSC adaptive beamforming. The proposed GSC structure comprises a fixed beamformer (e.g., delay-and-sum), Modified Blocking Matrix (MBM), and an adaptive filter. Delay and Sum Beamformer (DSB) is used as a fixed beamformer (FBF). It calculates the directional of arrival based on the delay from each microphone and which it sums up to give a reference speech signal. MBM blocks the speech signal and gives noise reference as input to the adaptive filter. An adaptive filter is updated till the noise gets reduced at the output of the GSC beamformer. All traditional algorithms are applied like LMS, NLMS, and RLS algorithms in the adaptive filter block of GSC. The performance and computational complexity are analyzed where noise reduction for only a few noise types and high computational complexity is addressed at the output of GSC. To address real-time noise at -10 dB SNR level with high convergence and low computational time. In this thesis novel, Fast Convergence NLMS (FCNLMS) is proposed to sidelobe canceling path of GSC and compared with GSC with existing adaptive algorithms. The proposed GSC-FCNLMS algorithm has achieved robust noise reduction at low SNRs.

This thesis proposes a convex combination of two Fast Convergence Normalized Least Mean Square (FCNLMS) filters to utilize the benefits of combining two adaptive filters through a mixing parameter. Further, it also offers a signed algorithm to a convex variety of FCNLMS. The combination approach provides a robust solution to alleviate the convergence speed vs. steady-state error tradeoff and efficiently increase the speech enhancement performance under various noisy environments for all real-time noises. This thesis addresses the directional and diffuse noise suppression in the adverse environment. It has been investigated on the performance of the GSC beamformer under directional, diffuse noisy conditions. The novel Improved Zelinski-TSNR multi-channel postfilter is implemented, i.e., GSC beamforming using improved Zelinski-TSNR multi-channel postfilter is proposed to suppress the directional and diffuse noise. Based on the speech presence probability using subband adaptive interference canceller, the degraded speech is enhanced with good quality and intelligibility. The last phase discusses the residual noise which has attained at the GSC beamformer using Zelinski multi-channel postfilter. The proposed GSC beamformer using combined postfilter and Sparse NMF algorithm addresses the residual noise suppression and source separation and achieves high quality and intelligibility for four and eight microphones. Using Sparse NMF at the output of the postfilter reduces the system-generated noise, i.e., residual noise. It also separates the interferences.

This work performs extensive computer simulations on all the proposed algorithms. The results demonstrate a significant performance in improving the terms of Perceptual Evaluation of Speech Quality (PESQ), output SNR, Segmental SNR (SSNR), Log Spectral Distance (LSD), Log-Likelihood Ratio (LLR), Short Time Objective Intelligibility (STOI), Signal to Distortion Ratio (SDR).

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANC – Adaptive Noise Cancellation

APA – Affine Projection Algorithm

ASR – Automatic Speech Recognition

BFE – Bayesian Feature Enhancement

BNMF – Basic Non-negative Matrix Factorization

CNMF – Constrained Non- negative Matrix Factorization

CVNMF – Convolutive Non- negative Matrix Factorization

DFT – Discrete Fourier Transform

D-GSC – Delay Generalized Sidelobe Canceller

DNMF – Discriminant Non-negative Matrix Factorization

DNN – Deep Neural Network

DOA – direction of arrival

DSB – Delay and Sum Beamformer

EM – Expectation Maximization

EMSE – Excess Mean Square Error

EVD – Eigenvalue Decomposition

FBF – Fixed Beamformer

FSB- Filter and Sum Beamformer

FCNLMS – Fast Convergence Normalized Least Mean Square

FIR – Finite Impulse Response

GA – Genetic Algorithm

GMM – Gaussian Mixture Model

GNMF – Generalized Non-negative Matrix Factorization

GSC – Generalized Sidelobe Canceller Beamformer

GSVD- Generalized Singular-Value Decomposition

HD-VoIP – High Definition Voice over IP

HMM – Hidden-Markov-Model

ISE – Iterative Signal Enhancement

KNMF – Kernel Non-negative Matrix Factorization

LCMV – Linear Constraint Minimum Variance

LDS – Linear Dynamical Systems

LLF – Log-Likelihood Function

LLR – Log-Likelihood Ratio

LMS – Least Mean Square

LSD – Log Spectral Distance

MAP – Maximum A Posteriori

MBM – Modified Blocking Matrix

MBM – Modified Blocking Matrix

MCSE – Multi-Channel Speech Enhancement

MMSE – Minimum Mean-Square-Error

MMSE-STSA – MMSE estimation of the Short-Time Spectral Amplitude

MNMF – NMF on Manifold

MOS - Mean Opinion Score

MSNR – Maximum Signal-to-Noise Ratio

MVDR – Minimum Variance Distortion Less Response

NMF – Non-negative Matrix Factorization

NMTF – Nonnegative Matrix Tri Factorization

NSS – Non-linear Spectral Subtraction

OM-LSA – Optimally Modified Log-Spectral Amplitude

ONMF – Orthogonal Non-negative Matrix Factorization

PDF – Probability Density Function

PESQ – Perceptual Evaluation of Speech Quality

PR – Posterior Regularization

PSD – Power Spectral Density

QRD- QR Decomposition

RGSVD – Recursive Generalized Singular-Value Decomposition

RPR – Reverberation Parameter Re-estimation

SCCFC – Signed Convex Combination of Fast Convergence

SDIRC – Speech Distortion and Interference Rejection Constraint Beamformer

SDR – Signal to Distortion Ratio

SDR-GSC – Speech Distortion Regularized Generalized Sidelobe Canceller

SDW-MWF – Speech Distortion Weighted Multi-channel Wiener filter

SNMF – Sparse Non-negative Matrix Factorization

STNMF – Structured Non- negative Matrix Factorization

SNR – Signal to Noise Ratio

SPP – Speech Presence Probability

SRR – Signal-to-Reverberation Ratio

SSNR – Segmental SNR

SS – Spectral Subtraction

STOI – Short Time Objective Intelligibility

SVD – Singular Value Decomposition

TF-GSC – Transfer-function Generalized Sidelobe Canceller

TSNR – Two-Step Noise Reduction

UFNLMS – Unconstrained Frequency domain Normalized Least Mean Square

VAD – Voice activity detectors

WERs – Word Error Rates

WNMF – Weighed Non- negative Matrix Factorization

CONTENTS

Chapter 1

Introduction

The speech signal is mostly corrupted with noise in real-world environments limiting its applicability in a wide range of applications like speech recognition in mobile phones, teleconferences, hearing aids, etc. In many common applications, it is desirable to suppress background noise and also to improve speech quality. That process of removing background noise in a degraded speech signal is referred to as speech enhancement or, more generally, as noise reduction. Speech enhancement has been a challenging problem for the past several decades [1–4] due to the complex nature of the speech signal. Speech enhancement techniques are classified as single-channel and multi-channel. Single-channel speech enhancement techniques [5] fail in finding the direction of unknown noise, suppressing multiple interferences, and in diffuse noise fields. So, multi-channel speech enhancement (MCSE) is the process involved in the removal of noise coming from various directions, and it separates the multiple inferences without any loss of information.

The scope of the thesis is the development and analysis of new adaptive beamforming approaches for multi-channel speech enhancement addressing various issues with the currently used methods. This chapter initially provides a brief introduction to multi-channel speech enhancement and the basics of beamforming techniques, and their classifications for noise reduction and interference suppression. The motivation for enhanced adaptive beamforming approaches for multi-channel speech enhancement is presented, followed by the problem statement, objectives of the thesis, and finally, the organization of the thesis.

1.1 Introduction to Multi-Channel Speech Enhancement

Speech enhancement refers to the improvement in quality and/or intelligibility of noise corrupted speech signals by using supervised or unsupervised speech enhancement methods. Speech enhancement deals with the processing of noisy speech signals, aiming at improving the perception of the human or decoding ability of machines [6]. It is used as a pre-processing unit for many speech communication applications. Speech enhancement is classified as single-channel and multi-channel speech enhancement. The classification is mainly based on the noisy background environment, multiple interference or speakers, and the number of microphones. Basically, single-channel speech enhancement can be performed in acoustic, stationary, and non-stationary noisy conditions, whereas in the case of reverberant, diffuse noise, and multi-speaker or interference from coming from various directions, single-channel speech enhancement fails to find the direction of arrival of the unknown signal. So, multi-channel speech enhancement techniques [7] achieved robustness in suppressing directional and diffuse noise in real-life environments. A simplified diagram of multi-channel speech enhancement system is shown Fig.1, where noisy input speech, i.e., a male speaker, radio sound, and female speaker from the crowd is taken whose direction is unknown are received at multi-microphone array which is given as input to multi-channel speech enhancement where background noise is reduced, and multiple speakers are separated, finally desired male speaker speech is obtained at the output.

In many speech communication systems, the presence of background interference degrades the quality or intelligibility of speech signals. There is a need to differentiate between the quality and intelligibility of speech, which in most cases are interchangeably but are quite different from each other. The quality of speech refers to how a speaker conveys an utterance and includes the attributes such as naturalness and speaker recognizing [8]. In very simple terms, quality is a measure of how well the examination resembles the original speech and how nice the speech sounds. Intelligibility is concerned with what the speaker has said, i.e., the meaning or information content behind the words [8]. It is a measure of how understandable the speech is and concentrates on the information-carrying content of speech.

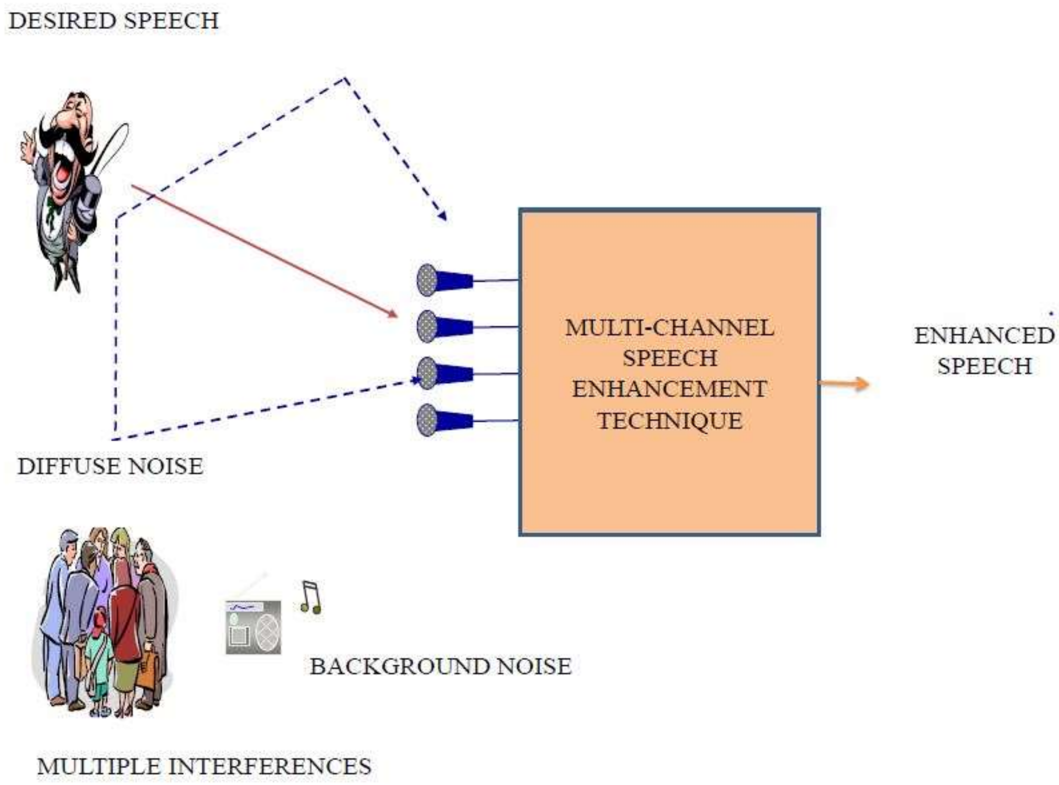

Figure 1.1: Multi-Channel Speech Enhancement System

The performance of a multi-channel speech enhancement system degrades rapidly in adverse environments. The presence of background noise causes the quality and intelligibility of speech to degrade. The performance of speech communication devices such as mobile phones, teleconferencing, automatic speech recognition, and electronic hearing aid, etc. which utilize speech processing systems to communicate and store speech signals, degrade significantly in the presence of background noise resulting in inaccurate information exchange and listener fatigue between the speaker and the listener. Thus, noisy environments reduce the speaker and the listeners' ability to communicate. Voice communication, for instance, over cellular telephone systems typically suffers from background noise present in the car, street, station, airport, restaurant, etc., at the transmitting end, which makes it difficult for the listener at the receiving end to understand the speaker. Thus, there are a wide variety of scenarios in which it is desired to enhance speech. Improving the quality and/or intelligibility of noisy speech effectively improves the performance of speech processing applications such as communication systems, speech recognition, speech coding, hearing aids, etc. The goal of a multi-channel speech enhancement system varies with respect to the application at hand. It could be to reduce the listener fatigue, to enhance the overall speech quality, to increase the intelligibility, etc., or a combination of these, depending on the application. In a speech recognition system of mobile phones, the recognition accuracy will suffer in the presence of noise, and hence the noisy speech signal can be pre-processed by a multi-channel speech enhancement algorithm before being fed to the system. In the teleconference system used by the military, the intelligibility has to be enhanced rather than the quality. For hearing-impaired listeners using hearing aids, it is always desired to enhance the noisy speech by removing the directional noise before amplifying the signal. Also, the characteristics of the noise and its relationship to the clean signal, like additive, convolutive, correlated, uncorrelated, etc., and the number of microphones available affect the design and development of the multi-channel speech enhancement system.

The performance of multi-channel speech enhancement systems is limited by the trade-off between interference cancellation, noise reduction, and multiple source separation [9]. Hence, the main challenge is to develop multi-channel speech enhancement algorithms, reducing the background noise from a particular direction, diffuse noise, and residual noise in an enhanced speech to improve the quality of the speech signal without reducing its intelligibility. Several multi-channel speech enhancement schemes have attempted to address the problem using various approaches.

1.2 Applications of Multi-Channel Speech Enhancement

Multi-channel speech enhancement has several practical application areas, which include telecommunication systems like mobile phones, teleconferences, speech/speaker recognition, hearing aids, etc. The multi-channel speech enhancement block can be placed as a front end to reduce the noise energy and improve the quality and intelligibility in telecommunications. Some of the most important applications are discussed in this section to show how important a role multi-channel speech enhancement has in our day-to-day lives.

Telephone communication has gone from home or office to a wide range of settings, including congested streets, vehicles, public transit, restaurants, and so on. Noise pollution can sometimes significantly reduce the quality and intelligibility of speech. As a result, it is necessary to avoid such deterioration, due to this noise reduction in mobile phones has been the subject of much study [10]-[11].

Teleconferencing permits a large number of people in a room to engage with one or more people in a hands-free experience. Due to its hands-free feature, listeners in a teleconference would be defenseless in case of adverse environment noise and directional noise. As a result, efforts have been undertaken to reduce the noise [12].

Various automatic speech recognition systems have been implemented into applications such as hands-free telephones, mobile phones, etc. When a noisy speech signal is utilised as an input to the system instead of a clean speech signal, the performance of system gets degrades. As a result, noise reduction in such systems has sparked a lot of study [13]-[14].

Noise has a greater impact on those with hearing loss than it does on people who are normally hearing. The people have a harder time distinguishing between noise and speech. As a result, research has been conducted to add speech enhancement modules into hearing aids to reduce the effect of noise contamination [15]–[17].

1.3 Beamforming for Speech Enhancement

Beamforming or spatial filtering is one of the multi-channel speech enhancement algorithms used in teleconferences, mobile phones, hearing aid applications. Beamforming methods [18] are useful to enhance the degraded speech from noisy real-time environments coming from unknown directions. Beamforming techniques are used to recover the desired clean speech signal from reverberation and noisy settings such as directional noise and diffuse noisy conditions. Spatial filtering [19] is used to reduce interference signals from undesirable directions. The signals from the microphone array are shaped into a beam pattern known as beamforming or spatial filtering. This is a time-honored technique for suppressing interference signals originating from various directions. Spatial filtering is the process of canceling out interference signals utilizing angles and frequencies from different directions. It is utilized to boost the quality of speech signals coming from the direction of view. Fixed and adaptive beamformers are the two types of beamformers covered below.

1.3.1 Fixed Beamforming

Fixed beamforming is a traditional multi-channel speech enhancement technique. Fixed beamformers [18] get their name from the fact that their weights co-efficient are fixed during the process. They are also known as data-independent filters because the filter is not reliant on the data from the microphone and follows fixed weights. Fixed beamformer approaches such as delay and sum beamformer improve speech signals by calculating and collecting the delay. FIR filters are employed in the filter and sum beamformer to improve the quality before summing up, which is suitable for narrowband speech improvement. A Delay and Sum Beamformer (DSB) as fixed beamformer is shown in Figure 1.2; here, noisy speech input is given to a multi-microphone array. Based on the distance and angle of arrival, delay from each microphone is calculated and added to have enhanced speech at the output. The delay and sum beamformer, on the other hand, will not work in a reverberant environment.

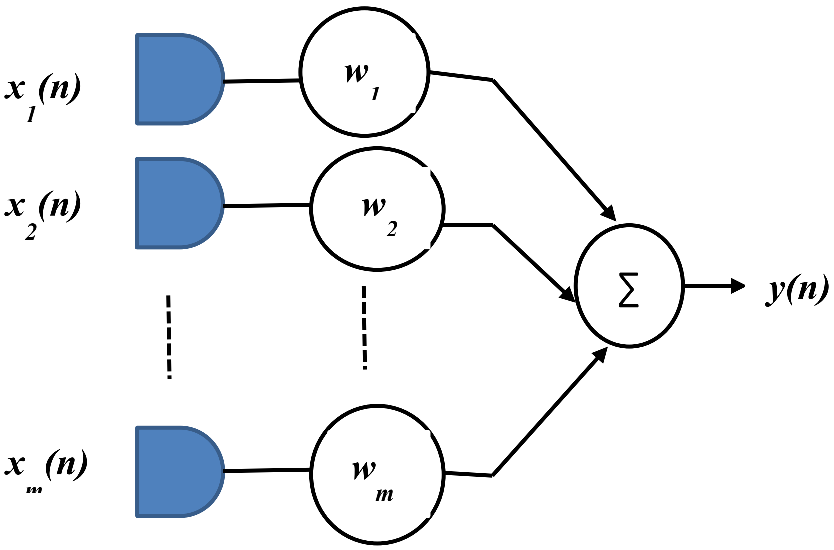

Figure 1.2: Fixed Beamformer

1.3.2 Adaptive Beamformer

Adaptive beamforming for speech enhancement necessitates thorough consideration of issues unique to degraded signal in adverse environment.

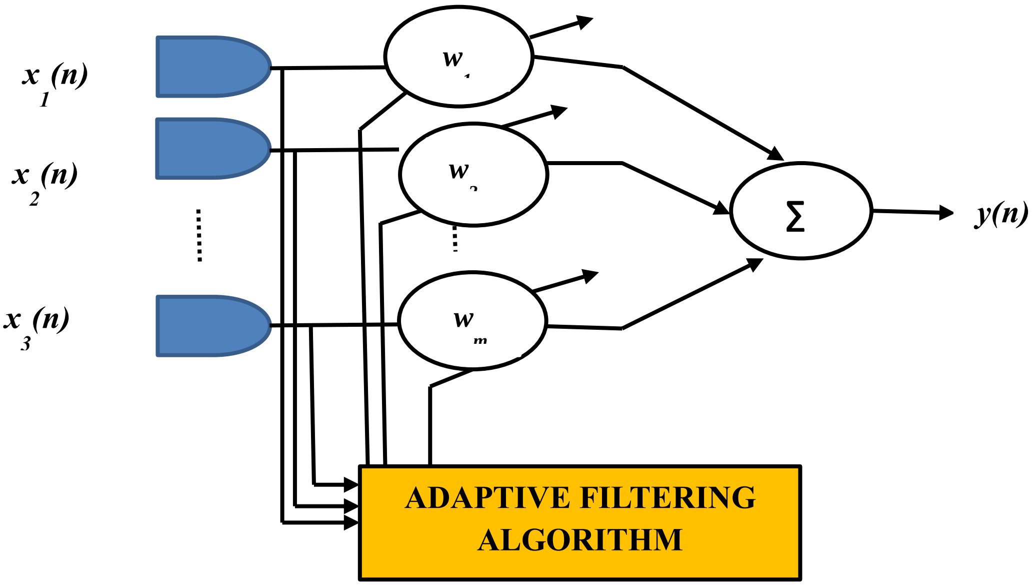

Figure 1.3: Schematic View of Adaptive Beamformer

Due to numerous reflections from the room walls, an acoustic field impulse response. The length of the filters in a typical workplace might approach thousands of taps. Furthermore, due to the speaker and objects, the impulse response is frequently time variable.

Adaptive beamformers [20] update themselves during the process. They are also known as data-dependent beamformers because they rely on the static features of desired, noisy speech signals entering the microphone. Adaptive beamforming is shown in Figure 1.3, where the weight is updated using the adaptive filtering algorithms, desired speech is obtained at the output until the error in the adaptive algorithm is minimized. The multi-channel speech enhancement general schematic flow is shown in Figure 1.4. When the unwanted signals are not pointing to sources, or there are too many interfering sources, the performance of some beamformers is restricted. Furthermore, due to the longer observation time necessary to determine signal statistics, several beamformers suffer from nonstationary interference.

Single-channel enhancement techniques can accomplish nonlinear spatial and/or spectral filtering and respond to changes in interference characteristics considerably more quickly. In this part, we'll look at how to employ algorithms like postfilter at the beamformer output. For the past few decades, various beamforming [21]-[22] methods have been introduced to remove directional noise. Existing adaptive beamforming approaches include the Minimum Variance Distortion Less Response (MVDR) [23] beamformer, Linear Constraint Minimum Variance (LCMV) [24] beamformer, and Speech Distortion and Interference Rejection Constraint beamformer (SDIRC) [25].

But, these existing adaptive beamforming techniques with postfilter fail in real-time environmental noises, directional, diffuse, and residual noise conditions. So, a novel adaptive beamforming approach should be developed to address directional, diffuse, and residual noise in real-time environmental noisy conditions like car, station, babble, street, restaurant, airport, etc.

1.4 Adaptive Beamforming with Multi-Channel Postfiltering

MCSE algorithms have lately gained popularity. It is generally known that beamforming techniques increase speech quality significantly [7]. As the diffuse noise is incoherent, here noise reduction becomes inadequate [26]. To process further post-processing using postfilter [27] is necessary. Furthermore, because non-stationary noise cannot be differentiated from speech signals in general, considerable performance deterioration is to be predicted in a non-stationary noise environment.

Most MCSE techniques include a DSB and GSC [21] followed by a Wiener filtering-based postfiltering mostly in grouping with SS). On the issue, many articles are stated including [28]–[36]. In general, postfilter are classified into two types. One is single-channel postfilter that operates on the beamformer output as a single-microphone speech enhancement technique. Next, is multi-channel postfilters, use the directional information collected by the GSC structure directly to improve speech signal separation from transient noise.

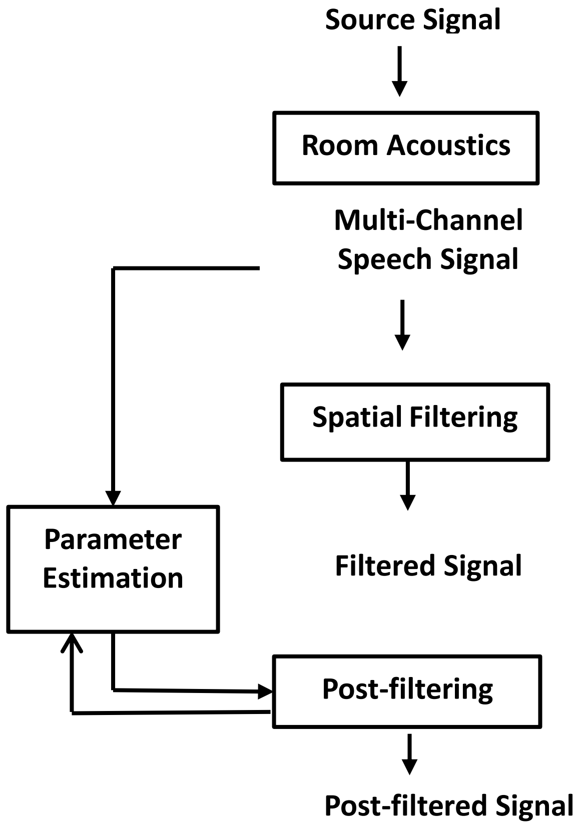

Figure 1.4: Generalized form of Adaptive Beamforming with Postfilter

1.5 Non-Negative Matrix Factorization (NMF)

One of the basic concepts deeply rooted in science and engineering is that there must be something simple, compact, and elegant playing the fundamental roles under the apparent chaos and complexity. This is also the case in signal processing, data analysis, data mining, pattern recognition, and machine learning. With the increasing quantities of available raw data due to the development in sensor and computer technology, how to obtain such an effective way of representation by appropriate dimensionality reduction technique has become important, necessary, and challenging in multivariate data analysis. Generally speaking, two basic properties are supposed to be satisfied: first, the dimension of the original data should be reduced; second, the principal components, hidden concepts, prominent features, or latent variables of the data, depending on the application context, should be identified efficaciously. Non-negative Matrix Factorization (NMF), which incorporates the non-negativity constraint and thus obtains the parts-based representation as well as enhancing the interpretability of the issue correspondingly, was initiated by Paatero and Tapper [37], together with Lee and Seung [38], [39].

NMF has become an imperative tool in multivariate data analysis and has been widely used in the fields of mathematics, optimization, neural computing, pattern recognition and machine learning [40], data mining [41], signal processing [42], image engineering, and computer vision [42], spectral data analysis [43], bioinformatics [44], finance and economics [45]. More specifically, such applications include text data mining [46], digital watermark, image restoration, image segmentation [47], facial expression recognition [48], audio pattern separation [49], music genre classification [50], speech recognition, microarray analysis, blind source separation [51], EEG signal processing [52], email surveillance [53], online discussion participation prediction, network security, automatic personalized summarization, identification of compounds in atmosphere analysis [17], earthquake prediction, stock market pricing [54] and so on.

Non-negative matrix factorization (NMF) and related probabilistic latent variable models (PLVMs) are data-driven machine learning techniques are used for the purpose of source separation. At a high level, when NMF/PLVMs is used for source separation, we decompose the audio spectrogram data, or equivalently the magnitude of the short-time Fourier transform (STFT) of an audio recording, is decomposed as a linear combination of the outer product of prototypical spectral components times vectors of amplitude over time. The spectral components for each sound source and their gains are learned from data, and the result is used to estimate the contribution of each source within an unknown mixture over time and eventually perform the separation.

NMF/PLVM methods can also be thought of as basis decomposition or dictionary-based methods and are closely related to sparse coding [55], principal component analysis [56], singular value decomposition [57], independent subspace analysis methods [58], and related matrix factorization methods. In addition to their audio applications, both NMF and PLVMs are also commonly used for processing images, text, and other data types and collectively have gained a significant research interest over the past decades.

1.6 Motivation

The speech processing systems used by people in everyday lives include digital mobile radio-communication systems, speech recognition systems, hearing aids, etc. These systems are prone to noise from various environments like background noise (airport noise, station noise, street noise, etc.), directional noise, diffuse noise residual noise, etc. This degrades the quality or intelligibility of these systems, which will affect people's lives as it makes the usage of these systems difficult. Speech enhancement or noise reduction algorithms restore or enhance the speech signals.

Single-channel speech enhancement techniques like spectral subtraction (SS), subspace algorithms, wiener filter, etc., fail to improve the quality of degraded speech signal coming from a particular direction and cannot find the directional of arrival under various noisy environments. Multi-channel speech enhancement algorithms provide better solutions to address these problems. A novel multi-channel speech enhancement algorithm has to be developed to suppress background, directional, diffuse, residual noises and separate the interference under various noisy environments. These are essential in applications like mobile phones, teleconferencing, hearing aids, etc. We need noise-free information for effective communication.

Multi-channel speech enhancement (MCSE) techniques like adaptive beamforming enable high-quality, hands-free communication in noisy environments. In the adaptive beamformer like Generalized Sidelobe Canceller (GSC) beamformer, the noise cancellation relies on the sidelobe canceling path. To improve speech in a noisy environment, a robust adaptive filter in the sidelobe canceling path must be constructed. GSC adaptive beamformer with effective noise cancellation makes the systems more reliable for noisy environments. A convex combination adaptive filter is used to overcome the challenges in noise reduction.

In multiple source environments, there is a need to suppress the directional noise. Similarly, in diffuse noise fields, as the noise from all direction looks similar, a novel adaptive beamforming technique need to be addressed. To further suppress the residual noise, a novel multi-channel speech enhancement system should be developed.

1.7 Problem Statement

Multi-Channel Speech Enhancement (MCSE) system has to be capable of suppressing the noise from the noisy speech signal. Most of the existing MCSE techniques use slow convergence with high computational complexity adaptive filters. Also, the MCSE gets affected due to noisy environments, and there is a need to develop an MCSE system that is robust to various noisy conditions. By using a combination of adaptive filters, the computational overhead of the MCSE system increases. MCSE systems also suffer from directional, diffuse, and residual noise. Therefore, there is a need to develop an MCSE system that gives better quality and intelligibility with directional, diffuse, residual noise suppression, and also it should separate multiple interferences.

1.8 Objectives

1. Implementation of novel Generalized Sidelobe Canceller (GSC) beamforming using different adaptive filtering algorithms like LMS, NLMS, RLS, and proposed FCNLMS for background (airport noise, station noise, street noise, etc.) noise reduction.

2. To develop adaptive beamforming using novel signed convex combination of adaptive filtering algorithm for speech enhancement with less computational complexity.

3. To implement adaptive beamforming using novel multi-channel postfilter for directional and diffuse noise suppression for speech enhancement.

4. To develop adaptive beamforming using combined postfilter and sparse NMF for residual noise suppression in an enhanced speech and multi-source separation.

1.9 Organization of Thesis

Chapter 1 gives the concept of a multi-channel speech enhancement system, and its applications are introduced. The motivation towards MCSE, objectives, and contributions towards the thesis are discussed in brief.

C**hapter 2** explains the state-of-the-art of problem. History of multi-channel speech enhancement techniques, adaptive beamformers, adaptive filters, postfilters, and non-negative matrix factorization techniques, and also about multi-channel speech enhancement simulation environment and the database used.

Chapter 3 proposes adaptive beamforming using different adaptive filters for speech enhancement. The chapter discusses different adaptive filtering algorithms like Least Mean Square (LMS), Normalized LMS (NLMS), and Recursive Least Square (RLS) algorithms to Generalized Sidelobe Canceller (GSC) beamformer and proposes Fast convergence NLMS algorithm to GSC beamformer under various noisy environments.

C**hapter 4** proposes a novel signed convex combination of fast convergence algorithm to GSC beamformer. A novel signed convex combination of fast convergence adaptive filters is proposed in the sidelobe canceling path of the GSC beamformer to provide a tradeoff for many noisy environments, and it is verified in various noisy situations. The analysis is carried out using different noises with SNRs ranging from -10 dB to 15 dB for a multi-channel speech enhancement system.

C**hapter 5** GSC beamforming using novel Zelinski-TSNR multi-channel postfilter for speech enhancement is proposed. The chapter explores directional noise and diffuses noise suppression. Directional noise is suppressed by the GSC beamformer. A novel Zelinski – Two-Step Noise Reduction (TSNR) multi-channel postfilter is implemented to the GSC beamformer to suppress diffuse noise.

C**hapter 6** proposes novel adaptive beamforming using combined postfilter and Sparse NMF for speech enhancement. The chapter describes the residual noise suppression at the output of the GSC adaptive beamformer with a combined postfilter and Sparse Non-negative Matrix Factorization (SNMF) algorithm. The simulation environment and analysis are explained under various SNR levels with a standard speech enhancement database.

C**hapter 7** gives the conclusions of the contributions of the thesis, and the future scope of this work is discussed in brief.

CHAPTER 2

Literature Survey

This chapter provides the literature on speech enhancement and adaptive beamforming approaches for multi-channel speech enhancement. The recent related techniques employed for post-filtering, directional, diffuse noise estimation and handling of residual noise are also discussed.

Initially, the applications and classifications of speech enhancement methods are discussed. Then, a detailed description of adaptive beamforming approaches is provided, which will form the underlying theory of the algorithms developed in the later chapters. The noise estimation techniques for speech enhancement are briefly described, and the currently used methods in adaptive beamforming algorithms to handle the case of directional, diffuse, and noise in real-time environments are also mentioned. Then the adaptive filters and combined adaptive filtering algorithms of speech signals for enhancement. Finally, NMF for speech enhancement and source separation is shown in the last section. The issues with the existing methods of speech enhancement obtained from the literature survey are provided from which the framework of the research work is decided.

2.1 Introduction

Speech enhancement is a challenging task in real-world environments like automatic speech recognizers and other communication systems. It aims at improving the quality and intelligibility of speech signals corrupted with a variety of noise conditions like airport, car, restaurant, train, street, diffuse-field effects, speech signals from other speakers, etc., to name a few [5]. A robust speech enhancement system should be able to perform well in any of these noisy situations.

2.2 Classification of Speech Enhancement

Typically, the speech enhancement methods can be broadly divided into single-channel and multi-channel enhancement techniques [59] depending on the number of microphones used to collect the acoustic signal and noise. The performance of a speech enhancement algorithm is limited by the number of noise sources available [6], [60]–[63]. In most of the widely used applications like hearing aids and mobile phones, where mostly only a single channel is available, single-channel enhancement is used. Single-channel enhancement techniques are very easy to build and are less expensive when compared to their multi-channel counterpart.

2.2.1 Single Channel Speech Enhancement

In single-channel or single-microphone enhancement, it is assumed that only the noisy signal containing both the clean speech and the additive noise is available from a single microphone for speech enhancement. There is no second signal which could provide information regarding the reference noise or speech. In most real-time applications, such as speaker and speech recognition, mobile communication, and hearing aids, usually, a second channel is not available. Hence this is one of the most challenging problems in speech enhancement. This is widely studied because of its simplicity and universal applicability since in most real-life situations, and only single microphone systems are available such as in speech communication, speech coding, and speech recognition in noisy environments. These systems are easy to build and comparatively less expensive than multiple-input systems. Single-channel speech enhancement methods have only a single input having the noisy speech from which enhanced speech has to be extracted [64]–[74]. Single-channel systems constitute one of the most difficult situations of speech enhancement since no reference signal to the noise is available, and the speech cannot be pre-processed prior to being affected by the noise. Usually, they make use of different statistics of speech and noise. Traditional single-channel speech enhancement methods are Spectral Subtractive (SS) algorithms, wiener filtering, statistical-model-based algorithms, subspace algorithms are explained as follows.

2.3 Multi-Channel Speech Enhancement

The number of microphones available can influence the performance of speech enhancement algorithms [75]. Typically, the larger the number of microphones, the easier is the speech enhancement task. Adaptive cancellation techniques can be used when at least one microphone is placed near the noise source. The multi-channel system uses the noise reference obtained in an adaptive noise cancellation (ANC) device. It uses phase alignment to reject undesired noise components. The system even uses both the noise reference and the phase alignment [76]. These systems tend to be more complex. The multi-channel speech enhancement method gives a better performance in non-stationary noise conditions due to the presence of a reference channel [77] – [85]. Phase alignment can be performed in one of the channels to reject the undesired noise components. The main drawbacks of multi-channel speech enhancement techniques are fabrication cost and complexity.

For economic reasons, most systems are single-microphone-based solutions where the speech enhancement is done on the output of a single microphone, although better speech enhancement results can be achieved by using a microphone array system with more than one microphone, but with increased complexity and expenses. The speech enhancement techniques can also be classified as supervised or unsupervised speech enhancement methods.

Supervised methods achieve noise reduction by considering a model for both the speech and noise signals, which require a training phase to estimate the parameters. Some of the supervised techniques include HMM-based methods [86] – [90], Gaussian Mixture

Models (GMM) [91] - [92], codebook based algorithms [93] – [94]. DNN based approaches [95] - [97], and Nonnegative Matrix Factorization (NMF) based methods [98]–[101].

A Speech enhancement method that reconstructs clean speech signal from a sinusoidal model and a set of acoustic speech features like a voicing classification, fundamental frequency, and spectral envelope, estimated from noisy speech using a single statistical model, is proposed by Philip Harding and Ben Milner [101]. By constraining the enhanced signal to be produced by a model of speech production, the output is free from noise. Tian Gao et al. [102] proposed a unified DNN approach to reduce both background noise and speech interference in a speaker-dependent scenario.

The DNN system is trained to unify speech enhancement and speech separation. The signals of speech interference are considered as one noise type. The unified system achieves good results compared with specific systems where only noise or speech interference is present and better performance for noise and speech interference mixed conditions. The results demonstrate the effectiveness of the ensemble method in low SNR environments. The performance of the supervised approaches depends on the prior information fed to the system, which limits its performance in non-stationary noise environments. There are a number of unsupervised speech enhancement methods that are not provided any data. Clean speech is estimated from noisy observations without any prior information on the noise type or speaker identity.

These multi-channel interfaces often have higher improvement possibilities than single-channel interfaces. They enable the creation of multi-channel spatial filters that selectively amplify or suppress sounds in certain directions (or volumes) by leveraging spatial variety, such as phase and level discrepancies, or, more broadly, the variations in acoustic characteristics between channels. Single-channel spectrum filters, on the other hand, need a considerably more thorough understanding of the target and the noise, and thus often result in a lesser quality increase. Indeed, it can be demonstrated that the greatest potential quality improvement attainable with only two microphones is already significantly larger than with a single microphone and that it continues to increase with more microphones [103].

Over the previous four decades, hundreds of MCSE methods have been suggested in the literature along two historical research lines. Microphone array processing arose from the theory of sensor array processing for telecommunications and focused primarily on the localization and enhancement of speech in noisy or reverberant environments [8], [60], [104]-[106], whereas Blind Source Separation (BSS) was later popularized by the machine learning community and addressed “cocktail party” scenarios involving multiple sound sources mixed together [107]-[112]

2.3.1 Beamforming

Beamforming is the way of forming a spatial-temporal filter. Broadband arrays are made up of a series of filters applied to each incoming microphone signal, then summing. The fundamental purpose of the beamformer is to extract the desired signal having an adverse effect on the array at a specific location from noisy array data. Interference signals frequently share the same frequency spectrum as the intended signal. The delay-and-sum beamformer is the most basic construction, compensating for the relative delay between different microphone inputs before summarizing the steered signal to produce a single output. If the number of microphones is reasonably high, this beamformer, which is still commonly employed, can be highly efficient at reducing non-coherent, i.e., spatially white, noise sources. However, if the noise source is coherent, noise reduction (NR) is highly dependent on the direction in which the noise signal arrives. As a result, the DSB performance is ineffective in reverberant environments. The delay and sum idea was expanded by Jan and Flanagan [123]-[124] and Rabinkin et al. [125] by adding the FSB. This structure, which is meant for multipath conditions such as reverberant enclosures, replaces the better delay compensator with a corresponding filter to achieve better performance.

In general, the beam pattern form microphone array can be tailored to have a certain reaction. This may be accomplished by correctly adjusting the weights of the MCF. Whereas, in dynamic acoustical settings, the use of data-independent design approaches is severely constrained.

The statistical characteristics of the intended and interference signals are used to build statistically optimum beamformers. They generally seek to boost the intended signal while rejecting the interfering signal. Several criteria, such as maximum signal-to-noise ratio (MSNR), minimal mean-squared error (MMSE), and linearly restricted minimum variance, can be used in beamformer design (LCMV). [19] - [117] provides an overview of many design criteria.

Beamforming techniques rely on signal statistics (at least second-order statistics), which are typically unavailable and must be inferred from data. Furthermore, the acoustical environment varies over time as a result of talker and object movement, as well as sudden changes in noise characteristics (e.g., passing cars). As a result, adaptive mechanisms are necessary. Each of the predefined design criteria can have an adaptable equivalent. Sondhi and Elko [126], Kaneda and Ohga [127], and Van Compernolle [128] made early contributions to the topic of adaptive beamformer design. By including echo cancellers into the beamformer architecture, Kellermann [129] solved the problem of integrated echo cancellation and noise reduction. In a vehicle scenario, Nordholm et al. [130]-[131] used microphone arrays and constructed a beamformer that used calibration signals to improve the results. Martin [132] looked at beamforming techniques for microscopic microphone arrays.

The well-known MCWF [133] is the result of minimizing the MSE in the context of array processing. Doclo and Moonen [134]–[136] suggested an efficient Wiener filter implementation based on the microphone data matrix's GSVD. This approach produces an optimum (in the MMSE sense) estimation of the required signal component of one of the microphone signals. The authors also developed efficient techniques for RGSVD updates. A post-filtering stage for adaptive noise cancellation is also offered as an alternative. In such a method, an optimum noise channel is estimated, moreover to attain desired signal, an optimal estimation is designed. To improve the speech signal even more, this evaluated noise signal coefficients are treated as reference noise signal [133]. Spriet [137] introduced a subband based GSVD [137] technique, but Rombouts [138]-[139] recommended solving the problem with the efficient QRD.

Acoustic arrays are used in various adaptive beamforming techniques. ATF combines the speech and the noise estimations. The multi-channel Wiener filter, on the other hand, relies solely on estimations of the recorded noisy signal's second-order statistics and the noise signal and makes no a priori assumptions about the signal model. However, as Chen et al. [140] point out, even if the Wiener filter is the best in terms of MMSE, but unable to produce desired speech at output. However, this issues are be addressed by modified MMSE which enables undistorted signal at output. This modification is used in a strategy described in [141]-[142].

Frost's [20] LCMV beamformer aims to minimize output power under linear limitations on the array's responsiveness to the intended speech signal. He presented an adaptive LMS algorithm [133]. To circumvent this restricted adaptation, author in [21] introduced the GSC, model and later modified by Affes [143] and Gannot [144]. Improved GSC is extended Transfer-Function Generalized Sidelobe Canceller (TF-GSC) comprises: a FBF, BM and ANC for NR in sidelobe cancelling path.

Nordholm [145] investigate the upper limits of the GSC's achievable NR in an isotropic noise field. Bitzer et al. address their issue in [146]-[148]. The authors of [146] construct a formula for the NR as a function of the noise field and assess the deterioration as a function of the reverberation duration (T60). [147] discusses the unique two-microphone case. GSC with wiener and LMS filter are shown in [148]. Marro et al. [149] and Nordholm et al. [150] introduce a frequency-band GSC structure.

Huarng and Yeh [151] solved the distortion problem by calculating the desired signal leakage into the GSC structure's reference noise branch. However, the delay-only ATFs assumption is applied, and the anticipated deterioration due to pointing mistakes alone is assessed. Nordholm et al. [152] illustrate the performance decrease caused by limiting Wiener filters to a finite impulse response (FIR) construction. The resultant performance limits of the GSC structure are significantly dependent on the cross-correlation between the sensors' signals caused by the noise field, as demonstrated in the references above and by Cox [153].

Beamformers are frequently sensitive to signal mismatch. The GSC, in particular, is plagued by two fundamental issues. For starters, non-ideal FBF might result in non-coherent filter-and-sum operations. To increase the resilience of beamformers, Doclo [135] and Nordholm et al. [158] utilize spatial and frequency-domain restrictions. The second issue addressed by this survey is the leaking phenomena induced by imperfect BM. If the intended speech seeps into the noise reference signals U(k,l), the noise canceller filters will remove speech components from the FBF output, resulting in self-cancellation and hence severe distortion. It should be noted that self-cancellation is unavoidable even when the ANC filters are modified during noise-only times. The purpose of this section is to provide some ideas for improving the resilience of the GSC structure and lowering its susceptibility to signal mismatch. Cox et al. [117] conducted an in-depth examination of array sensitivity.

Hoshuyama et al. [159]-[160] suggested many approaches for dealing with the robustness problem, focusing on the self-cancellation phenomena produced by the leakage of the intended speech signal to the BM outputs. This effect is accentuated in reverberant settings, as the BM merely adjusts for the relative delay [as in [142]]. In general, there are two approaches to address the leaking issue. First, better spatial filtering may be integrated into the BM design. Claesson and Nordholm [130] recommended that a spatial high-pass filter be used to cancel out all signals within a given frequency and angular range. Huarng and Yeh [151] investigated the leaking problem and imposed a derivative restriction to the array response, resulting in greater tolerance to pointing mistakes.

A second solution to leakage concerns is to put restrictions on the ANC filters. Hoshuyama et al. [159] suggested a number of configurations that combined changes for both the BM and ANC blocks. An adaptive BM based on signal cancellers replaces the traditional delay-compensation BM. Two limiting techniques can be used for the filters in question.

The first method makes use of norm-constraint, whereas the second makes use of the leaky LMS adaption scheme. Haykin [168] demonstrated that the two techniques are equal. The leaky LMS algorithm or Cox's norm-constrained adaptation mechanism is used to adjust the ANC filter (see [30]). As a final note summarising Hoshuyama's techniques, we draw the reader's attention to the similarity between the suggested modification of the BM filters and the subspace tracking procedure given by Affes and Grenier shown in [143].

Doclo [161] demonstrate that the output SNR after NR using the aforementioned speech distortion weighted multi-channel Wiener filter (SDW-MWF) is always greater than or equal to the input SNR, regardless of filter length or value of the trade-off parameter between NR and speech distortion. This ANC filter solution is known as the speech distortion regularised generalized sidelobe canceller (SDR-GSC) structure. Spriet et al. [141] also recommended incorporating a single-channel postfilter to correct for the structure's distortion in the event of voice leakage into the reference signals. The scope of this study does not allow for further examination of this structure. The authors offer a stochastic gradient-based implementation of their criterion in [162].

Spriet et al. [16] investigate the resilience of both the multi-channel Wiener filter and the GSC structures in the context of hearing-aid applications. Improvements in the GSC structure are an ongoing topic, particularly in the sidelobe canceling path. Interfering signals can significantly degrade the quality of the desired signal received by a sensor array. This issue arises in a variety of array processing applications and is exacerbated when the interfering signals are nonstationary [163]-[166]. Moving interfering sources or sudden changes in the propagation channel might produce nonstationary interfering signals.

Furthermore, when interfering sources are placed in dense multipath settings, such as acoustic environments, it causes interference. It is quite difficult to make noise reduction without degrading the quality of the target signal. In these instances, minimizing interfering signals sometimes involves the employment of FIR filters with a large number of taps, which have a high computational cost and a slow convergence rate [167]-[168]. Broadband adaptive beamforming systems are extensively employed to handle this problem (see, for example, [167]–[171]) since they are very successful at receiving the desired source signal while simultaneously decreasing interfering components, especially in dense multipath settings.

One of the most often used adaptive beamforming methods for broadband signals is the generalized sidelobe canceller (GSC) [21]. It consists of a fixed beamformer (e.g., delay-and-sum or filter-and-sum [7]) and an adaptive route that reduces noisy components created by unwanted interfering sources, resulting in decreased noise power at the system output. Many adaptive beamforming methods for nonstationary noisy settings were based on the GSC in the literature [144], [160], [172]-[173]. The adaptive filtering algorithm employed in the sidelobe canceling path is entirely responsible for the success of a GSC system.

In general, gradient-based adaptive algorithms, such as the least mean squares (LMS) algorithm, can be used to adjust filters in the time domain (see, for example, [174]). Although this family of methods has a cheap computing cost, when the filter length is fairly high, the convergence is rather sluggish [174], making the adaption of the filter weights impractical in real-time applications. Hessian-based adaptive filtering, which is common in algorithms such as the recursive least squares (RLS) filter, is another time-domain standard method. When compared to gradient-based algorithms, the latter method achieves quicker convergence [174]. However, due to the high computing cost of RLS adaptive filtering, adaptation may become prohibitively expensive. Furthermore, depending on the characteristics of the required source signals, the RLS method may perform worse than the LMS algorithm in a nonstationary environment [175].

The affine projection algorithm (APA) family [176], which is widely used in adaptive beamforming [19]–[20], provides a good compromise between performance and computational load because it has faster convergence rates and manageable computational complexity when compared to other time-domain algorithms. Furthermore, when compared to other traditional time-domain adaptive algorithms, APA is the greatest fit for processing colored signals. Despite its strong capabilities [176], APA is hampered by unfavorable environmental conditions, particularly in the presence of many nonstationary sources, which render the adaptation process unstable and impair performance.

To overcome this issue, we offer resilient array beamforming techniques based on the adaptive combination of MISO filtering systems, which are simply filtered banks in this case. The adaptive combination of adaptive filters is a particularly effective and versatile method for balancing the tradeoffs inherent in adaptive filter settings [177]-[178]. Combined adaptive schemes are often implemented using filters from the same family and complementing characteristics, such as distinct step sizes or filter lengths. They are also employed with filters from various families that utilize different update algorithms or cost functions [179]–[183]. The combined scheme is capable of switching between filters adaptively based on the highest performing filter, ensuring that the best possible filtering is always provided [177].

In this article, beamforming designs that use an adaptive combination of filters to enhance system tracking in the face of broadband nonstationary interfering signals. One approach is to merge two MISO systems that use the same updating method but have different step size values. In fact, it has been demonstrated that combining a fast filter (with a big step size value) and a slow filter (with a small step size value) leads to faster convergence, reduced residual misalignment, and enhanced tracking capabilities when compared to separate filters [184], [185] , [177]. Another way for improving tracking capabilities in nonstationary situations is to combine two filters with distinct updating methods, namely one gradient-based and one Hessian-based [182]-[186]. This filter combination makes use of the Hessian-based filter's rapid convergence and the gradient-based filter's performance capabilities, which may beat the Hessian-based filter in nonstationary circumstances [182]-[186]. In comparison to the performance of a combination of filters with different step sizes, which is never better than the performance of individual filters in terms of excess mean square error (EMSE), the performance of a combination of filters with different updating approaches may outperform the performance of individual correspondent filters in terms of EMSE [186]. In terms of adaptive combinations, we focus on the convex constrained combination with sigmoid nonlinearity on the output stage in our study since it introduces less gradient noise than unconstrained and affine constrained combinations [181]-[183].

We present two distinct beamforming architectures based on the integration of adaptive MISO systems with various update methods. The first approach employs a system-by-system (SS) combination in which the overall output of the first MISO system is convexly combined with the overall output of the second MISO system. The second architecture is a filter-by-filter (FF) combination scheme in which each adaptive filter from the first MISO system is convexly coupled with the comparable filter from the second MISO system. All adaptive filters in both systems are updated using an APA.

We utilize various projection orders for each MISO system to differentiate them. Furthermore, in order to employ the optimum parameter setting for each filter and enhance tracking performance even further, we suggest a multistage combination method in which the filtering procedure is performed in two phases.

2.4 Multi-Channel Postfiltering

The use of postfilter approaches in MCSE has recently gained popularity. It is well known that beamforming approaches increase speech quality significantly [7]. The noise reduction is insufficient when the noise field is spatially incoherent or diffuse [26], and an extra postfilter is usually necessary [27]. Furthermore, because nonstationary noise cannot be differentiated from speech signals in general, a large performance deterioration in a nonstationary noise or noisy real-time environment is expected.

A multi-microphone component (either delay and sum beamformer or GSC [21]) is usually followed by a postfilter based on Wiener filtering (occasionally in conjunction with spectral subtraction) in most MCSE. Several articles have been published on the subject, including [28]–[35], to name a few. The postfilters can be classified into two kinds in general. On the beamformer output, the first is a single-channel postfilter that functions as a single-microphone speech enhancement method. Multi-channel postfilters, on the other hand, make explicit use of the spatial information recovered by the GSC structure to distinguish the speech signal from the transient noise or diffuse noise.

While the theory suggests that using a Wiener post-filter improves performance, obtaining good estimates of the signal and noise spectral densities required to calculate the post-filter transfer function can be problematic in practice. The most frequent way for estimating these spectral densities is to use the multi-channel input signals' auto- and cross-spectral densities. Marro et al. [149] investigate this type of post-filter estimation in-depth, and it is largely based on Zelinski's work [28]. While the Zelinski post-filter produces acceptable results, it is reliant on a number of assumptions. The assumption of zero correlation between the noise on distinct channels, which corresponds to a fully incoherent noise field, is made in particular. In actuality, such an incoherent noise field is unusual, and the noise correlation between channels can be strong, especially at low frequencies. This is especially true for sensors that are near together, like in speech enhancement applications.

This work shows how the Zelinski post-filter estimator's assumption of incoherent noise can be replaced with the more generic assumption of a known noise field coherence function. Theoretical noise fields, such as spherically isotropic (diffuse) or cylindrically isotropic noise fields, can be used to represent a variety of realistic noise fields, such as those found in workplaces or cars. The coherence functions for these theoretical noise fields are already used in a number of well-known beamforming techniques, including super directive beamformers ([117], [187]-[188]).

The use of theoretical noise coherence models is extended to postfilter estimates enabling the development of a more appropriate post-filter for various noise conditions. The

Zelinski post-filter, which corresponds to a unity coherence matrix, is included as a special instance.

According to Simmer et al. [8] the multi-channel Wiener filter provides the ideal solution to the problem of multi-channel noise reduction for broadband inputs in the minimum mean square error (MMSE) sense and may be further decomposed into an MVDR beamformer followed by a Wiener post-filter. As a result, in order to increase the performance of microphone arrays in noisy practical settings, a post-filter based on Wiener theory is usually required [8].

In the literature, a variety of post-filtering strategies have been published [28], [189]-[199]). Zelinski [28] was the first to introduce a widely used multi-channel post-filter based on the Wiener filter. This post-filter is based on the premise that noise from separate microphones is uncorrelated, resulting in a perfectly incoherent noise field. This assumption, however, is rarely met in real-world situations, particularly in the case of closely spaced microphones and low frequencies, which are characterized by high-correlated noise.

Fischer et al. proposed combining the generalized sidelobe canceller (GSC) with the Zelinski post-filter to suppress the spatially correlated and uncorrelated noise to suppress the high-correlated noise [192]. However, neither the GSC nor the Zelinski post-filter operates well at low frequencies, according to Bitzer et al. [147]. Meyer et al. provide an alternative technique that uses spectral subtraction to suppress the high-correlated noise components [33].

Due to the voice activity detector (VAD) based noise estimate technique, this method creates fake "musical noise" and fails to deal with non-stationary noise. McCowan and Bourlard have devised a universal expression for the Zelinski post-filter based on the noise field's a priori coherence function [189].

Although employing office room recordings, this post-filter was demonstrated to increase voice quality and speech recognition accuracy when compared to the Zelinski postfilter. Its performance is likely to be severely impaired when the "real" and presumed coherence function differs [189].

The optimally modified log-spectral amplitude (OM-LSA) estimator, a single-channel noise suppression technique, was recently described for minimizing log-spectral amplitude distortion in non-stationary noise settings [193]. When multi-channel inputs are available, the OM-LSA estimator was also extended to a multi-channel post-filtering approach, which was shown to be effective in reducing highly non-stationary noise components from the desired source components based on the energy-based speech presence probability estimator ([191], [194]. A speech presence probability estimator based on these spatial characteristics was provided to improve the performance of the OM-LSA post-filter [195]-[196] by taking into account the geographically stable characteristics of noise fields. The inherent sensitive implementation parameters involved in the variations of the OM-LSA post-filter [191], [193]-[194], on the other hand, severely decrease their performance in actual contexts.

A diffuse noise field has been proved to be a viable model for a wide range of practical noise situations, including reverberant rooms and automobile environments [7], [189], [33]. Traditional post-filters, such as the Zelinski and Mc Cowan post-filters, fail to minimize diffuse noise despite being based on Wiener theory [3]-[4]. OM-LSA post-filters, on the other hand, may be able to deal with diffuse noise with appropriate implementation parameters, but they are not based on Wiener theory, therefore breaching the framework of the multi-channel Wiener filter [191], [194]. Novel postfilter has to be explored in order to concentrate on both low and high frequencies of signal to enable speech presence samples.

2.5 Non-negative Matrix Factorization Algorithms

Single-channel sound source separation or enhancement methods are motivated by many outstanding issues in signal processing and machine learning, such as speech denoising, speech enhancement, audio-based forensics, music transcription, and music remixing. One of the most effective approaches for these purposes is based on NMF [39], [198], and [199] and its probabilistic latent variable model counterparts [200] and [201]. These methods model spectrogram data or equivalently the magnitude of the short-time

Fourier transform (STFT) of an audio recording as a linear combination of prototypical spectral components over time. The resulting spectral components and their resulted gains are then used to separate each source within the mixture. These methods can achieve good separation or enhancement results using supervised or semi-supervised techniques. In these techniques, isolated training data is used to learn individual models of distinct sound sources also separates an unknown mixture of similar-sounding sources [202].

Most of the variants that have been proposed to improve the performance of NMF consist in adding a regularization term to the log-likelihood function (LLF) of the observed data. Defining the right penalty (or regularizer) is one of the most important steps for incorporating the user-annotation constraint into the given latent model [203]. In [258], the user annotations are used to obtain the posterior regularization (PR) terms. There are several ways to incorporate these annotations into latent variable models, for instance, by using the suitable regularization functions and expectation maximization (EM) algorithms. In this method, annotations control the regularization parameters.

In [204], Chung et al. have proposed a speech enhancement approach in which clean speech spectral components and spectral noise components were modeled by Gaussian Mixture Model (GMM). The corresponding Log-Likelihood Function (LLF) was used as regularization to the cost function of conventional NMF to extract the statistical characteristics of the signals. Non-negative dynamical system (NDS) was introduced to model the speech and audio power spectra [205]. The NDS model can be interpreted both as an adaptation of linear dynamical systems (LDS) to non-negative data and as an extension of non-negative matrix factorization (NMF) to support Markovian dynamics. The performance of the proposed NDS algorithm was significantly better than the state-of-the-art algorithm in terms of SDR in real environmental sounds like babble noise, helicopter, bees, fire, and shaking chapter. Recently proposed online semi-supervised NMF algorithms have only been evaluated using noisy mixtures shorter than 30 seconds. The performance degrades when the speech signal starts to appear after 2 minutes. To solve this problem, [206] proposed a rotational reset strategy. In the proposed method, instead of updating of entire speech dictionary continuously, periodically and rotationally reset speech dictionary elements. The proposed algorithm performs better than existing algorithms in non-stationary noisy environments (: birds, casinos, cicadas, computer keyboard, eating chips, frogs, jungle, machine guns, motorcycles, and ocean) in various SNR conditions.

In [207], proposed a sparse and low-rank NMF with Kullback-Leibler divergence to estimate the noise spectrum from the input noisy speech spectrogram without any prior knowledge of speech and noise. In the proposed method, noise and speech were estimated by decomposing the input noisy magnitude spectrogram into a sparse speech-like part and low-rank noise part.

2.6 Issues with the Existing Methods of Multi-Channel Speech Enhancement

The major issues with the existing multi-channel speech enhancement (MCSE) techniques identified from the literature survey are the following:

1. Existing adaptive beamforming algorithms show substandard noise cancellation in sidelobe canceling path under noisy environmental conditions and also ineffective at -10 dB SNR. Existing beamforming algorithms are unable to find the direction of the unknown signal and fail in suppressing the real-time environmental noise.

2. Combined adaptive filters in existing adaptive beamforming are ineffective to achieve better noise reduction at -10 dB to 15 dB SNR levels and also gain more computational burden. Adaptive beamforming with a single adaptive filter is limited to particular real-time noise reductions.

3. In the case of directional and diffuse noise in low frequencies, most of the multi-channel speech enhancement methods give a less noise reduction. To remove diffuse noise from noisy speech, postfiltering of the noise signal is a concern. But existing postfiltering shows less noise reduction in the low-frequency region, where an exact speech signal exists.

4. Residual noise is generated after speech enhancement. Most of the speech enhancement algorithms assume the noise to be additive and are ineffective while dealing with the case adverse environment with multiple interferences. The performance of the algorithms designed for residual noise is also inefficient while handling real-time noise.

2.7 Framework of Research Work

From the issues, it is identified that to solve the problems of the existing methods of multi-channel speech enhancement; new algorithms have to be developed based on different adaptive beamforming approaches. A generalized sidelobe canceller (GSC) would be appropriate to provide a balance between finding the direction of arrival of unknown signal and noise reduction in a noisy real-time environment. The innovative integrated adaptive algorithms should be introduced to adaptive beamforming systems to maintain a fair tradeoff between noise reduction and computational complexity. In order to deal with adverse environments to reduce diffuse noise and smoothen the output, efficient postfiltering techniques could be employed to adaptive beamforming. Residual noise could be addressed by employing an NMF algorithm that could adaptively enhance speech with respect to the varying noise levels. A novel technique integrating beamforming, postfiltering, and NMF must be devised to address the case of speech enhancement and interference separation. Waveform and spectrogram plots, as well as objective measures, are used to evaluate the developed speech enhancement algorithms.

Objective measures should be able to assess the performance of the developed methods with respect to the specific issues addressed. The developed algorithms could be used in a variety of applications like teleconferencing, mobile phones, speech/speaker recognition, hearing aids, communication systems, etc.

2.8 Summary

In this chapter, the applications, the previous works, and the current advancements in the area of speech enhancement have been discussed. It provides the issues identified in the existing speech enhancement techniques and the framework for the research work in the thesis. The classification of speech enhancement techniques into single and multi-channel enhancement based on the availability of a number of microphones/channels is discussed and various techniques employed are studied. Due to low complexity, better quality, and intelligibility, multi-channel techniques are more popular than single-channel techniques though the performance increases with the number of channels. There has always been an effort to develop speech enhancement algorithms that provide balance between real-time environment noise, diffuse and residual noise. Existing multi-channel speech enhancement methods like Adaptive beamforming shows ineffective performance in reducing the directional noises in noisy real-time environmental conditions. So, a novel Adaptive filtering algorithm has to be implemented. Conventional algorithms produce more residual noise in an attempt to reduce the diffuse noise in noisy environmental conditions like offices or cars. It has been shown in the works done that multi-channel speech enhancement techniques like adaptive beamforming methods give a better balance between the two than any other existing adaptive beamforming technique. Improvements in signal distortion have been reported when postfiltering stages are used after enhancement. Also, diffuse noise reduction could be achieved by employing signal estimators designed for it. Diffuse noise estimation and separation is an important stage in any practical speech enhancement algorithm. Employing the VAD-based noise estimation algorithms would degrade the performance of speech enhancement techniques since most of them update noise only during the speech absent frames and assume noise to be stationary during active speech regions. The most efficient noise estimator would be that of a continuous estimator based on speech presence probability. Speech enhancement in residual noise scenarios is a challenge, and a new technique has to be devised to handle it. Different techniques used so far in the beamforming approach to handling the case of diffuse and residual have been thoroughly studied. Most of the algorithms are inefficient in handling the case since they assume the nature of the noise to be additive white. For those designed to address the issue of real-time environmental and diffuse noise specifically, the sidelobe canceling path in beamformers should be carefully set to be efficient. NMF has been shown to deal with residual noise better in certain signal processing applications and could be employed to handle the case of residual for speech enhancement and separation. The development of speech enhancement algorithms addressing the issues identified from the literature would add to the performance of a wide range of speech processing applications like teleconferencing, mobile phones in speech/speaker recognition in pre-processing, and also in hearing aids communication systems, etc.

Chapter 3

Adaptive Beamforming Using Different Adaptive Filters for Speech Enhancement

This chapter proposes the use of various adaptive filtering algorithms like LMS, NLMS, and RLS to GSC beamformers under noisy real-time environments. And also proposes FCNLMS adaptive filtering algorithms to the sidelobe canceling path of the GSC beamformer for speech enhancement under different SNR levels.

3.1 Motivation

In the GSC beamformer, the noise cancellation relies on the sidelobe canceling path. Existing adaptive beamforming algorithms show substandard noise cancellation in sidelobe canceling path under noisy environmental conditions and also ineffective at -10 dB SNR. In order to improve speech in a noisy environment, a robust adaptive filter in the sidelobe canceling path must be developed. Novel adaptive filtering for multi-channel speech enhancement is proposed to address different noisy types with reduced computational time.

3.2 Introduction

In multi-microphone array, environmental noise degrades the desired speech quality and intelligibility. This is a major issue in speech communication applications like teleconferences, mobile phones, etc. when the desired speaker is non-stationary [7], i.e., in a noisy real-time environment, reducing the noise becomes quite difficult. In these cases, for noise reduction and interference suppression [106], in the place of conventional Finite Impulse Response (FIR) filters, which result in high computational complexities, the adaptive filters like Least Mean Square (LMS), Normalized LMS (NLMS) are widely used. However, in the case of single-channel speech enhancement, noise from a specific direction cannot be found using these basic adaptive filters. Single-channel speech enhancement algorithms fail in reverberant noise and also in finding the direction of the arrival of the input source. So, in multi-channel speech enhancement, Griffiths and Jim [21] introduced a GSC beamforming structure. It comprises three major blocks: fixed beamformer, blocking matrix, and an adaptive filtering block. In the fixed beamformer such as Delay and Sum Beamformer (DSB) [144], the microphone array receives the desired speech along with the noise. Delay from each microphone is calculated and then summed together to obtain the partially enhanced output [9], [214].

The performance of a multi-channel speech enhancement system depends completely on the blocking matrix and the adaptive filtering [227] block, which eliminates the unwanted noise and increases the quality of the desired speech. The adaptive filter block in the GSC beamforming plays a crucial role in noise reduction performance [213]. In the time domain, the gradient descent adaptive algorithms are used to update the weight of the filter. One such algorithm is the LMS algorithm which has low computational complexity but is not stable in noisy real-time conditions when the filter tap gets increased [113], [216]-[217]. Another popular adaptive algorithm is Recursive Least Squares (RLS) filter, which is based on Hessian adaptive filtering. It gives faster convergence when compared to LMS, whereas computational cost is high and is too expensive for noisy real-time environments [218], [172]. In order to have a better noise reduction in a noisy real-time environment, a GSC beamformer with different adaptive filters is implemented in this chapter.

3.3 The Proposed GSC Adaptive Beamforming for Speech Enhancement

The Generalized Sidelobe Canceller (GSC) beamformer is one of the most popular adaptive beamforming techniques in the multi-channel speech enhancement domain. A GSC structure is composed of a fixed beamformer (e.g., delay-and-sum), Modified Blocking Matrix (MBM), and with different adaptive filters is proposed as shown in Figures 3.1. The input to the proposed system is considered using a microphone array setup with noisy real-time conditions in a virtual conference room. The virtual conference room is designed based on the Image method [225], which takes the Room Impulse Response in the form of a Mex function using RIR generator [226] in MATLAB. As a fixed beamformer, delay and sum beamformer (DSB) is used. It calculates the direction of arrival (DOA) based on the delay and distance from each microphone. An unknown noisy input signal with partial enhancement is found at the output of DSB.

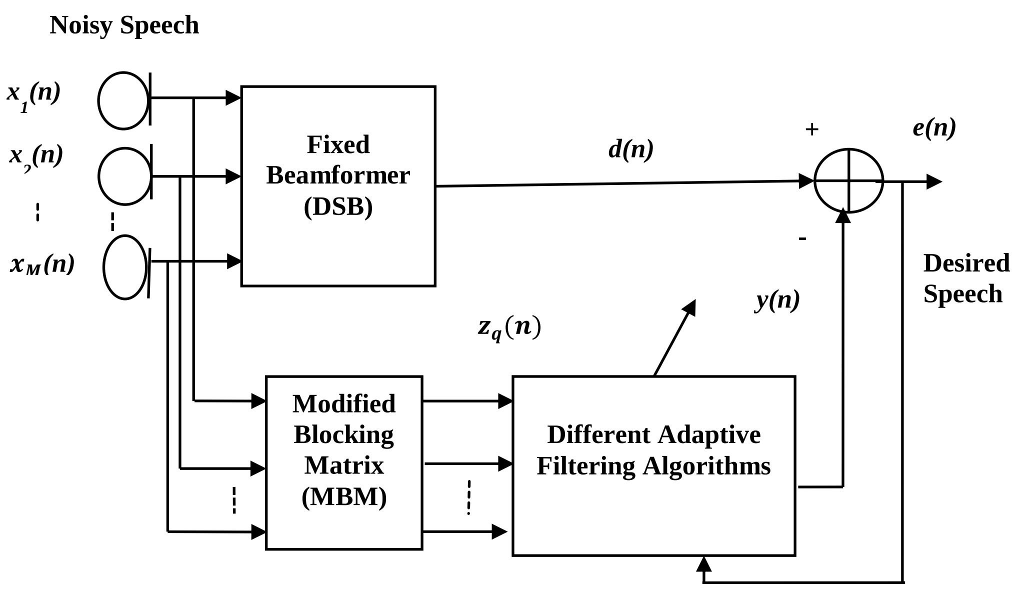

Figure 3.1: The Proposed GSC Beamformer for Speech Enhancement

MBM, on the other hand, blocks the speech signal and feeds a noise reference to the adaptive filter. To further reduce the noise in the signal, proposed a novel Fast Convergence NLMS adaptive algorithm in the sidelobe canceling path of the GSC beamformer and also verified with existing adaptive algorithms. An adaptive filter is updated till the noise gets reduced at the output of the GSC beamformer. In the adaptive filter block, different adaptive filtering algorithms like LMS, NLMS, and RLS algorithms are implemented, and also novel FCNLM adaptive filter is implemented for the noisy real-time environment in the coming sections to achieve noise reduction and low complexity in a non-stationary environment.

3.3.1 Fixed Beamformer (FBF)

FBF is used to find the direction of arrivals for unknown signals. To make the proposed method robust, considered the microphone array setup in the adverse environment using the room impulse response-based Mex function (in MATLAB). The DSB is one of the fixed beamforming techniques. It is used as FBF in the proposed method, which calculates the direction of arrival (DOA) based on the delay between each microphone. The structure of the DSB beamformer is shown in Figure 3. 2

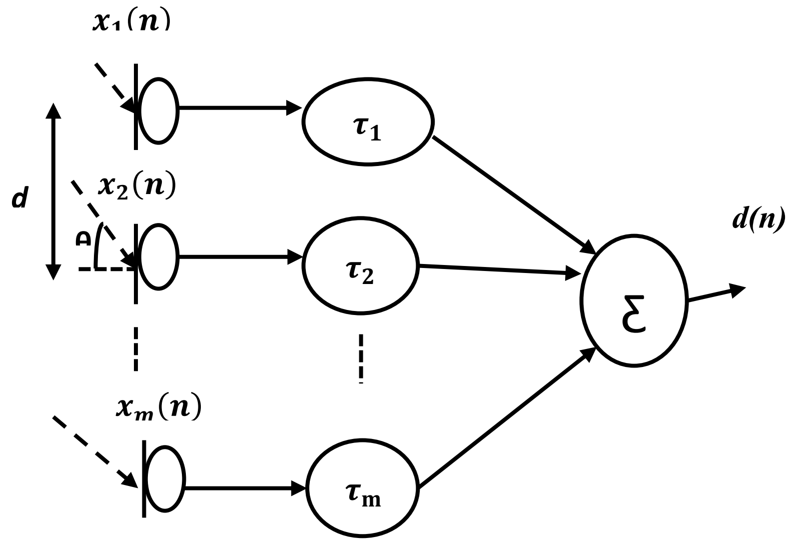

Figure 3.2: Fixed Beamformer (FBF)

Consider *d(n)* as the desired signal and $I_m(n)$ as the total noise and interferences observed at the output of m^{th} microphones. The input noisy speech signal $s_m(n)$ at the output of m^{th} the microphone is given by

$$s_m(n) = d(n) + I_m(n) \tag{3.1}$$

In the DSB structure, the microphones are placed linearly by giving d as the spacing between each microphone and angle θ for receiving the input signal from a particular direction. Here $s_1(n), s_2(n), ...s_m(n)$ is the combination of desired speech (unknown source), directional interference, and diffuse noise, which are input signals to the microphone. The input of each microphone is delayed with an angle θ and then summed to have a partially enhanced speech with the diffuse noise at the output of DSB, which is shown in Figure.3.2.

DSB output is defined as

$$x(n) = \frac{1}{M} \sum_{m=1}^{M} s_m(n - \tau_m) \tag{3.2}$$

Where *x(n)* is the output of DSB, M is the number of microphones, $s_m(n)$ is the incoming source at the m^{th} microphone and the delay from source to each microphone is τ_m. The main lobe position in the directivity pattern is changed by modifying the phase weight $\phi_m(f)$ i.e,

$$\varphi_m(f) = 2\pi\alpha(m-1)d \tag{3.3}$$

Where

$$\alpha = \frac{\sin\theta}{\lambda} \tag{3.4}$$

θ is the direction of arrival of an incoming signal and λ is used to determine the wavelength of frequency. The phase shift in the frequency domain relates to a time delay in the time domain. The time delay τ_m can be analyzed as given below

$$\tau_m = \frac{\varphi_m(f)}{2\pi f} \tag{3.5}$$

$$\tau_m = \frac{2\pi\alpha(m-1)d}{2\pi f} \tag{3.6}$$

$$\tau_m = \frac{2\pi\alpha(m-1)d\sin\theta}{2\pi f} \tag{3.7}$$

$$c = f\lambda \tag{3.8}$$

$$\tau_m = \frac{2\pi\alpha(m-1)d\sin\theta}{c} \tag{3.9}$$

Likewise, the delays (τ_m) from the microphones are calculated and summed at the output of DSB represented as *x(n)* in the proposed method.

3.3.2 Modified Blocking Matrix (MBM)

In GSC beamforming, the blocking matrix plays a crucial role. It is used to block the desired speech signal and provides only the noise reference as input to the adaptive interference canceller, which is described as follows.

The lower path of the GSC beamforming is the blocking matrix [21], which is used to block the desired signal $d(n)$. As the desired signal is common to all the microphones from Equation (3.1), blocking is confirmed if the rows of the blocking matrix sum up to zero.

If b_m^T is the mth row of blocking matrix

$$b_m^T 1 = 0 \quad \text{for all values of m} \tag{3.10}$$

and b_m is linearly independent so that $l_m(n)$ it will have M-1 linearly independent components, which makes the row dimensions of the blocking matrix to M-1. Griffiths [21] considered the number of microphones as four, i.e., M=4, and gave two matrices. The first blocking matrix is defined as

$$BM_1 = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 \end{bmatrix}$$

Similarly, the second blocking matrix is defined as

$$BM_2 = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

The rows are mutually orthogonal and are the elements of the binary-valued Walsh function. Here BM_2 represents the difference between the adjacent microphone outputs. Each row BM_1 represents different amplitude responses, whereas each row BM_2 has identical patterns. But by using these matrices, the spatial information is not completely utilized. So, MBM is designed to subtract the desired speech from the noisy input signal using adjacent microphones. In the proposed GSC beamforming, MBM is used to utilize the complete spatial information on adjacent microphones and also on other microphones by using the identical pattern in the matrix. MBM is designed as

$$MBM = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}$$

The number of columns in the matrix indicates the number of microphones which is considered to be four, and the efficiency of MBM is 3. MBM gives the details of the complete noise present in the target signal and blocks the desired speech, and thus acts as noise reference for adaptive filter. The number of columns in the matrix indicates the number of microphones which is considered to be four, and the efficiency of MBM is 3. Finally, the DSB output is a speech reference signal, and the MBM output is a noise reference that is given to the adaptive filter block. To further reduce the noise in the signal, proposed a different adaptive filter in the sidelobe canceling path of the GSC beamformer. Next, the adaptive filtering block is explained as follows:

3.3.3 Adaptive filtering Algorithms

An adaptive filter with a robust convergence rate is essential in speech enhancement. In the lower part of GSC, the second block is an adaptive filter. In this chapter, adaptive filter plays a prominent role in reducing the error between the desired and noisy reference of a GSC. This can be achieved by using different adaptive algorithms [113] in the adaptive filter block. The blocking matrix noisy reference is given as input to the adaptive filter, where the weights are updated to enhance the corrupted speech at the GSC output. In this chapter GSC with different adaptive filters is proposed in order to improve the performance of GSC in terms of speed and complexity. Here, introduced three adaptive filters like LMS, NLMS, RLS is analyzed. The traditional LMS and NLMS adaptive filters give limited noise reduction, low convergence, and high computational complexity [215]. So, novel FCNLMS is implemented in the adaptive filter block of GSC. Different adaptive algorithms are explained below.

3.3.3.1 LMS Algorithm

In adaptive signal processing, the least-mean-squares (LMS) algorithm [215] is extensively used due to its stable nature and simplicity during implementation. In stationary conditions, LMS shows the best steady-state performance [216]. The standard LMS algorithm is explained in step by step manner below.

1. In the first step, the filter weight coefficients are initialized

$$\bar{w}\ (n) = \left[\ w_1(n)w_2(n)w_3(n) \dots w_p(n) = \ 0\right] \tag{3.11}$$

Where P is the order of the filter

2. The adaptive filter output is calculated as

$$y(n) = \overline{w}^T(n)z_q(n) \tag{3.12}$$

3. The error signal e(n) is calculated as

$$e(n) = d(n) - y(n) \tag{3.13}$$

LMS update equation is given by

$$\overline{\mathrm{w}}(\mathrm{n}+1) = \overline{\mathrm{w}}(\mathrm{n}) + \mu\mathrm{e}(\mathrm{n})z_q(\mathrm{n}) \tag{3.14}$$

Where μ is the step-size, the convergence rate of the filter weights is purely based on μ value. Equation (3.14) LMS weights update equation, which is employed in the GSC structure's adaptive filter block to update degraded speech and decrease error. The computational complexity of LMS is given by the $2N$ number of additions/Subtractions and $2N+1$Multiplications/ Divisions with N=256.

3.3.3.2 NLMS Algorithm

The normalized LMS (NLMS) algorithm is in addition to the standard LMS algorithm [216]. In NLMS, the weight vector will be changed minimum times from one iteration to other. The step size μ in the NLMS algorithm is in a time-varying parameter which is used to calculate convergence of the adaptive filter. Step size μ is given as

$$\mu(n) = \frac{\alpha}{c + \left\| z_q(n) \right\|^2} \tag{3.15}$$

The convergence rate of NLMS is optimized by adaption constant α , which ranges from 0<α<2, 'c' in the Equation (3.15) is a constant term used for normalization of the filter, which is limited to c< 1.

Finally, the NLMS algorithm updates the filter coefficients by using the following equation.

$$\bar{w}(n+1) = \bar{w}(n) + \frac{\alpha}{c+||\ \bar{x}'(n)||2} e(n) z_q(n) \tag{3.16}$$

NLMS algorithms converge faster compared to LMS because of the normalization factor μ . The error $e(n)$ of NLMS is less compared to the LMS algorithm. Computational Complexity [217 of NLMS is given by the $2N^2 + 2N$ number of additions/Subtractions and $2N^2 + 3N$ Multiplications/ Divisions.

3.3.3.3 RLS Algorithm

Recursive Least Squares (RLS) algorithm is robust adaptive algorithms to fasten the convergence rate compared to LMS and NLMS [218]. By using the RLS algorithm, the adaptive filter coefficients are found recursively to minimize the weighted least square of cost function corresponding with the input. RLS algorithm is strong in spontaneously adjusting the filter coefficients without knowing the input signal statistical information. At each instant, the RLS algorithms minimize the sum of squares of the desired speech signal estimated errors [219]. Noise cancellation capacity is high compared to LMS and NLMS but requires complicated mathematical operations. Because of this, RLS requires more computational resources [220]. The RLS algorithm is explained below in a step-by-step manner.

1. In the first step, the RLS algorithm filter coefficients are initialized

$$\bar{w}(0) = [0\ 0\ 0 \dots \dots .\ 0\ 0\ ]^T \tag{3.17}$$

2. In the second step, the inverse matrix $P(0)$ is initialized with the diagonal matrix maintaining the main diagonal with $\delta - 1$ a value

$$z_q(n) = [z_q(n) z_q(n-1) .... z_q(n-M+1)]^T \tag{3.18}$$

where the $z_q(n)$ is the adaptive filter input vector

3. In the final step, the RLS updated by calculating the following equations at each segment of the input signal.

$$\bar{w}(n) = \bar{w}(n-1) + R(n)e(n), \tag{3.19}$$

$$R(n) = \lambda^{-1}\Pi(n) / \left(1 + \lambda^{-1} z_q{}^H(n)\Pi(n)\right) \tag{3.20}$$

$$P(n) = \lambda^{-1}P(n-1)\, \lambda^{-1}R(n) z_q{}^H(n) P\,(n-1) \tag{3.21}$$

The error is estimated as follows

$$e\,(n) = d(k) - \bar{w}^H(n-1) z_q(n) \tag{3.22}$$

The computational complexity of RLS is given by $3N^2 + 4N$ a number of additions/Subtractions and $3N^2 + 6N$ Multiplications/ Divisions.

3.3.3.4 Proposed Fast Convergence NLMS Algorithm

A fast convergence and low complexity adaptive algorithm named Fast Convergence NLMS (FCNLMS) [232] is presented in this section, where updating the filter coefficients depends on adaption gain and likelihood variable of the fast transversal filter. In FCNLMS, the forward prediction error *e(n)* of the fast transversal filter [231] is calculated by applying a de-correlated technique to the input signal. This is used in analyzing the dual Kalman gain. The step by step procedure of FCNLMS is as follows:

1. Initialization: Initialize

Initialize $C_N(0)$ adaptation gain vector, $h_N(0)$ estimated filter coefficient vector, and also $\gamma_N(0)$ and $\gamma_1(0)$ likelihood variables for N samples.

$$C_N(0) = h_N(0) = 0 \tag{3.23}$$

$$\gamma_N(0) = 0 ,\gamma_1(0) = 0$$

$$\alpha(0) = \gamma_1(0) = E_0 \tag{3.24}$$

Where E_0 is an initialization constant with $\alpha(0)$ forward prediction error's variance.

2. Prediction error: $e(n)$

The prediction coefficient estimation can be calculated as

$$a(n) = \frac{r_1(n)}{r_0(n)+c_a} \tag{3.25}$$

Where $r_1(n)$ and $r_0(n)$ can be estimated recursively according to

$$r_1(n) = \lambda_a r_1(n-1) + z_q(n) z_q(n-1) \tag{3.26}$$

$$r_0(n) = \lambda_a r_0(n-1) + {z_q}^2(n) \tag{3.27}$$

Where $z_q(n-1)$ is input vector at the time 'n,' 'λ_a' is exponential forgetting factor and is. 'c_a' a small positive constant.

To compute the prediction error using a first-order prediction model:

$$e(n) = \bar{x}'(n) - a(n)z_q(n-1) \tag{3.28}$$

The forward prediction error variance is defined as

$$\alpha(n) = \lambda\alpha(n-1) + e^2(n) \tag{3.29}$$

Adaption Gain is given by

$$\begin{bmatrix} \widetilde{C_N}(n) \\ c(n) \end{bmatrix} = \begin{bmatrix} -\frac{e(n)}{\lambda\alpha_N(n-1)+c_0} \\ \widetilde{C_N}(n-1) \end{bmatrix} \tag{3.30}$$

$\widetilde{C_N}(n)$ dual Kalman gain,

$$\delta(n) = c(n)z_q(n-N) + \frac{(n)e(n)}{\lambda\alpha_N(n-1)+c_0} \tag{3.31}$$

$$\gamma_N(n) = \frac{\gamma_N(n-1)}{1+\gamma_N(n-1)\delta(n)} \tag{3.32}$$

Error for an adaptive filter is given as

$$\varepsilon_N(n) = d(n) - h_N^T(n-1)\bar{x}'_N(n) \tag{3.33}$$

Finally, the FCNLMS updating equation is defined as

$$h_N(n) = h_N(n-1) - \mu\varepsilon_N(n)\,\gamma_N(n)\widetilde{C_N}(n) \tag{3.34}$$

FCNLMS algorithm converges faster compared to RLS, NLMS, and LMS because of adaption gain. FCNLMS exhibits faster convergence with low complexity compared to LMS NLMS and RLS. The computational complexity of FCNLMS is 3N. FCNLMS algorithm converges faster compared to NLMS because of adaption gain. FCNLMS exhibits faster convergence with low complexity compared to LMS NLMS and RLS. The computational complexity of FCNLMS is 3N multiplications whereas LMS with 2N+1, NLMS with 2N^2+3N, and RLS with. The computational complexity of FCNLMS is low compared to LMS NLMS and RLS algorithms. All the four adaptive filtering algorithms LMS, NLMS, RLS, and the proposed FCNLMS, are applied to the adaptive filter block of GSC, and the error is minimized under various real-time noisy environments. The existing GSC-LMS and GSC-NLMS algorithm performance are less under noisy real-time conditions. The proposed GSC-FCNLMS algorithm gives enhanced speech with a minimal error when compared to GSC-LMS and NLMS. GSC-FCNLMS achieves faster convergence when compared with existing GSC-NLMS. The performance evaluation of the proposed algorithms is shown in the next section.

3.4 Results and Discussion

In this section, the simulation of the proposed GSC-FCNLMS, GSC-RLS, GSC-NLMS, GSC-LMS methods in noisy real-time conditions is evaluated and explained. The proposed GSC beamformer with different adaptive algorithms considers the following simulation parameter as shown in Table 3.1. A Multi-channel room impulse response is generated using a Mex function with a reverberation time of 300 ms following a Mex setup using Mex function, i.e., rir-generator.cpp [226] in MATLAB. The Mex function was taken from International Audio Laboratories Erlangen at Friedrich Alexander University Erlangen-Nuremberg. (https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator).

The noisy real-time condition is created by adding desired speech and real-time noises from unknown directions. The desired speech is taken using the DARPA TIMIT [227]-[228] database. The database is maintained with a sampling frequency of 8 kHz, which consists of 6300 male and female sentences where each of the 630 speakers speaks 10 sentences each. The real-time noises (Car, Restaurant, Babble, Airport, Station, Street noises) are taken from the NOIZEUS database [229]-[231]. These input signals are provided to the Mex setup, which gives a combination of the desired speech with real-time noise for different SNRs (-10 dB to 15 dB).

The degraded speech is an input to the DSB to evaluate the delay from each microphone and obtain a reference enhanced signal. After that, the input degraded speech is given to MBM. Using the MBM matrix, the subtraction of the delays caused on the adjacent microphones is calculated. Further, at the MBM output, a noise reference is generated. Finally, the same reference noise is applied to the adaptive filtering block (where different adaptive algorithms are analyzed) as input, where the weights of the filter are updated for each algorithm. Due to the proposed FCNLMS adaptive algorithm in sidelobe canceller, the error is minimized better compared to traditional algorithms like LMS, NLMS, RLS, and enhanced speech is attained at the output of the GSC beamformer. GSC-FCNLMS gives a high-quality enhanced speech at the GSC output, as shown in Figure 3.1.

Table 3.1: Simulation Parameters Considered for Proposed GSC Beamforming

Parameters	**Specifications**
Number of microphones(m)	m=4
Spacing to each microphone	5cm
Real-time noisy environment	Car, Restaurant, Babble, Airport, Station, and Street
Input SNR Levels	-10 dB, -5 dB, 0 dB, 5 dB, 10 dB
Room dimensions	6 m X 5 m X 3 m (Image Method), RIR generator [225]-[226]
Database	DARPA TIMIT [227] and Noizeus [229]-[230]
Tools	MATLAB and Python
Processor	Intel Core I7 Processor, Clock Speed-2.20 GHz, 8 GB RAM

3.4.1 Performance Analysis of the Proposed Method with Existing Methods

The performance of the proposed GSC-FCNLMS, and other GSC-LMS, GSC-NLMS, and GSC-RLS algorithms are analyzed in terms of objective parameters, namely Perceptual Evaluation of Speech Quality (PESQ) [231], Signal to Noise Ratio (SNR), and Log-Likelihood Ratio (LLR) [231].

3.4.1.1 Perceptual Evaluation of Speech Quality (PESQ)

PESQ [231] is an objective comprehensible measure. The range of PESQ as per the Standards International Telecommunication Union Telecommunication (ITU-T) lies between "0.5 to 4.5". The more the PESQ, the better is the intelligibility. Intelligibility measure PESQ at 10 dB for GSC- FCNLMS is 4.080 dB whereas for GSC-LMS GSC-NLMS and GSC-RLS is 3.407 dB, 3.960 dB and 4.010 under station noise, similarly for same station noise at -10 dB the PESQ for GSC- FCNLMS is 2.769 dB whereas, for GSC-LMS, GSC-NLMS and GSC-FCNLMS are 2.494 dB, 2.753 dB, 2.75 dB respectively. These measures show that GSC-FCNLMS beamformer gives an improved performance with less computation complexity, i.e., better PESQ compared to GSC-LMS, GSC-NLMS and GSC-RLS.

3.4.1.2 Signal to Noise Ratio (SNR)

The SNR (signal to noise ratio) is the ratio of signal to noise power. The higher the SNR value, the better the quality of the received output will be.

$$SNR(dB) = 10\, log_{10} \frac{\sum_{K=0}^{N-1} y^2(k)}{\sum_{K=0}^{N-1}[\hat{x}(k)-y(k)]^2} \tag{3.35}$$

Quality measure output SNR at 10 dB for GSC-FCNLMS is 31.4 dB whereas for GSC-LMS, GSC-NLMS, and GSC-RLS is 25 dB, 26 dB and 33.8 under station noise, similarly for same station noise at -10 dB the output SNR for GSC- FCNLMS is 6.9 dB.

Table 3.2 Performance Comparison of PESQ and SNR for GSC-FCNLMS With Existing Algorithms

SNR in dB	Noise Types	GSC-LMS		GSC-NLMS		GSC-RLS		GSC-FCNLMS	
		PESQ	SNR	PESQ	SNR	PESQ	SNR	PESQ	SNR
-10	Car	2.198	7.0	2.369	7.8	2.461	10.1	2.444	8.4
	Restaurant	2.338	9.3	2.516	11.5	2.713	14.2	2.687	12.7
	Babble	2.239	8.5	2.412	9.3	2.671	12.1	2.574	10.1
	Station	2.494	6.3	2.53	6.8	2.75	7.5	2.769	6.9
	Airport	2.273	6.3	2.465	7.5	2.771	10.1	2.612	8.8
	Street	2.378	10.3	2.601	11.0	2.879	12.9	2.768	12.0
	White	2.037	7.8	2.135	9.0	2.46	12.1	2.35	10.8
-5	Car	2.474	11.0	2.651	12.3	2.741	14.0	2.698	12.9
	Restaurant	2.58	13.3	2.82	17.3	2.93	20.6	2.88	18.8
	Babble	2.49	12.3	2.705	13.5	3.0	17.3	2.961	14.9
	Station	2.764	11.0	3.039	11.8	3.056	12.7	3.05	12.0
	Airport	2.576	9.8	2.818	11.8	2.942	14.3	2.894	12.8
	Street	2.62	14.5	2.886	17.0	2.99	18.2	2.91	17.6
0	Car	2.707	18.3	2.935	18.5	2.99	20.4	2.941	18.9
	Restaurant	2.785	20.0	3.089	21.5	3.192	23.0	3.099	22.2
	Babble	2.729	18.3	3.0	19.8	3.16	24.2	3.100	22.8
	Station	3.012	17.0	3.344	17.9	3.472	19.3	3.410	18.7
	Airport	2.872	14.0	3.132	16.5	3.539	20.5	3.436	18.6
	Street	2.822	20.8	3.16	21.5	3.44	26.6	3.360	23.2
5	Car	2.972	22.5	3.255	24.0	3.428	29.9	3.350	26.6
	Restaurant	3.039	24.0	3.451	26.8	3.66	28.0	3.578	27.5
	Babble	2.972	22.8	3.301	24.5	3.505	25.3	3.421	25.0
	Station	3.221	21.0	3.691	21.8	3.737	23.9	3.700	22.4
	Airport	3.094	21.8	3.487	22.0	3.548	25.1	3.513	23.7
	Street	3.043	25.5	3.464	27.0	3.876	30.0	3.686	28.1
10	Car	3.158	27.5	3.50	28.8	3.641	34.0	3.740	31.3
	Restaurant	3.26	28.8	3.711	30.8	3.77	32.5	3.353	31.7
	Babble	3.174	27.3	3.576	29.5	3.634	30.8	3.598	29.9
	Station	3.407	25.0	3.960	26.0	4.010	33.8	4.080	31.4
	Airport	3.315	26.0	3.774	26.8	3.86	28.5	3.824	27.0
	Street	3.256	29.8	3.74	31.0	3.805	33.5	3.785	32.5

Whereas for GSC-LMS, GSC-NLMS, and GSC-FCNLMS are 6.3 dB, 6.8 dB, 7.5 dB respectively. These measures show that GSC-FCNLMS beamformer gives improved performance with less computation complexity, i.e., better SNR compared to GSC-LMS, GSC-NLMS. Whereas GSC-RLS gives better performance than GSC-FCNLMS, but computation complexity is high compared to the proposed GSC-FCNLMS.

3.4.1.3 Log-Likelihood Ratio (LLR)

LLR [231] is an objective measure defined based on the LPC co-efficient, where α_p is the clean speech LPC vector and processed speech LPC vector α_p. R_c Clean speech auto-correlation matrix.

$$d_{LLR}(\alpha_p, \alpha_c) = \log\left(\frac{\alpha_p R_{c\alpha_p}{}^T}{\alpha_c R_{c\alpha_c}{}^T}\right) \tag{3.36}$$

Lowering the LLR more will be speech performance quality.

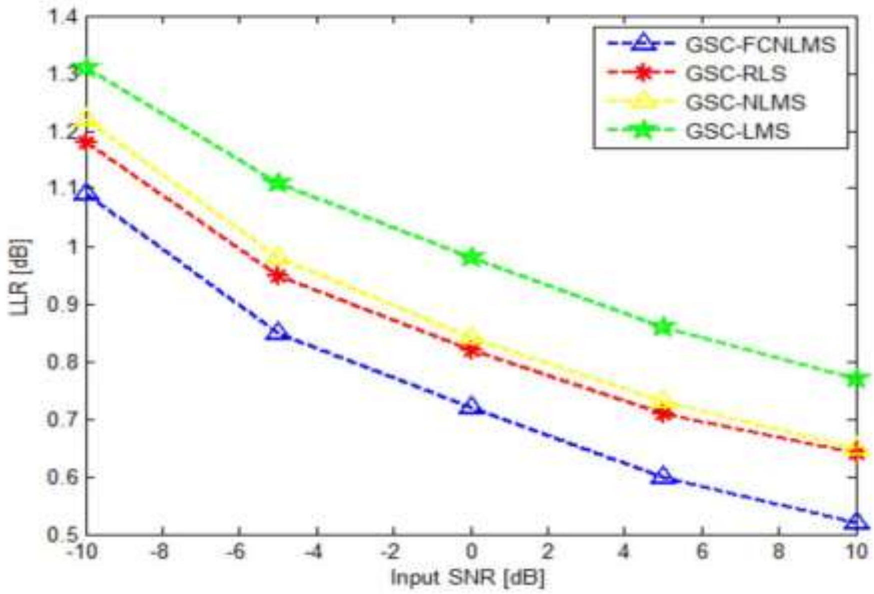

Figure: 3.3 Performance Comparison of Log-Likelihood Ratio (LLR)

LLR measure for the GSC-FCNLMS gives a lower LLR of 0.532 dB at 10 dB and 1.09 dB at -10 dB under station noise, which means the proposed GSC-FCNLMS. The LLR performance of GSC- FCNLMS, GSC-LMS, and GSC-NLMS is shown in Figure 3.3

3.4.1.4 Waveforms

In Figure 3.4 and Figure 3.5, the time domain plots and spectrograms of the proposed multi-channel speech enhancement system are illustrated, which shows the proposed GSC-FCNLMS with existing GSC-RLS GSC-NLMS and GSC-LMS noise reduction performance for 5 dB street noise. The enhanced speech signal of the proposed GSC-FCNLMS algorithm shown in Figures 3.4 and 3.5 gives better noise reduction compared to other algorithms.

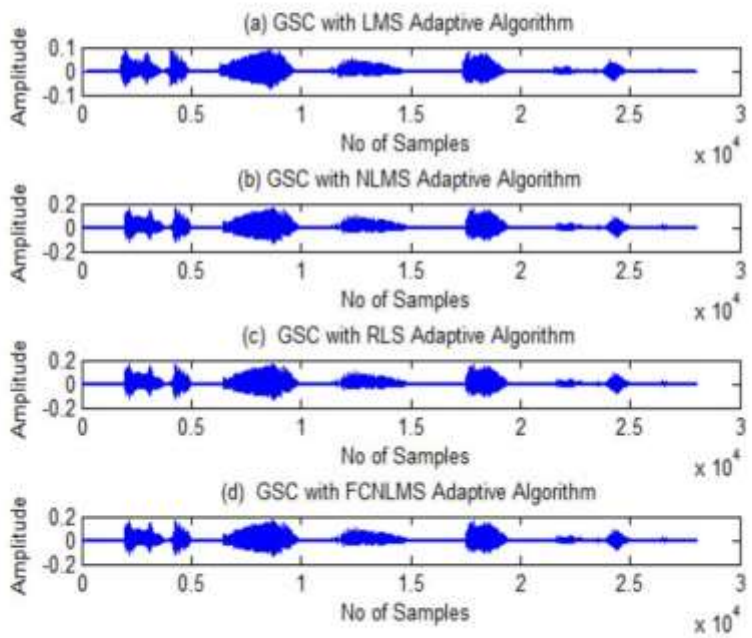

Figure 3.4: Time Domain Plot of Proposed GSC-FCNLMS with Existing Algorithms

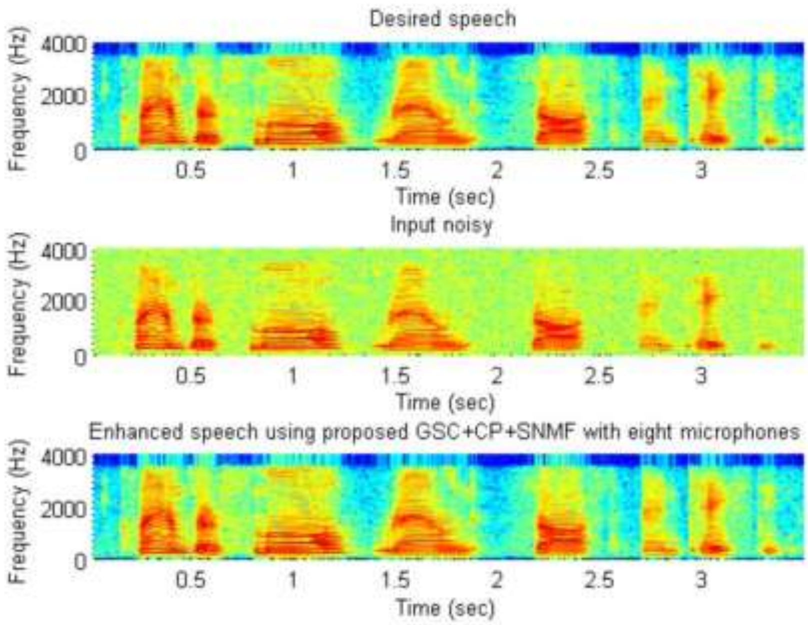

Figure 3.5: Spectrogram of Proposed GSC-FCNLMS with Existing Algorithms

3.5 Summary

Adaptive beamforming using FCNLMS adaptive filters for speech enhancement is proposed in this chapter. GSC-FCNLMS beamformer gives fast convergence and low complexity when compared with existing GSC-LMS, GSC-NLMS, and GSC-RLS algorithms under various noisy conditions. The quality of the speech signal for the proposed GSC-FCNLMS gives superior performance compared to existing algorithms. At -10 dB, the PESQ for proposed GSC-FCNLMS is 2.769 dB, whereas GSC-LMS, GSC-NLMS and GSC-RLS it is 2.494 dB, 2.53 dB and 2.75 dB under station noise conditions. Similarly for at -10 dB input SNR, GSC-FCNLMS output SNR is 6.9 dB, whereas GSC-LMS and GSC-NLMS is 6.3 dB, and 6.8 dB, respectively. Both quality and intelligibility of speech are improved for GSC with FCNLMS compared to LMS, NLMS, under various noise types even at lower SNRs. But only a few noisy types are giving better noise reduction using the proposed GSC-FCNLMS algorithm in this chapter. In order to address all the real-time environmental noises with the high objective quality measures under higher SNRs, i.e., above 10 dB, the novel adaptive filtering algorithm to sidelobe canceling path of GSC is implemented in chapter 4.

Chapter 4

Signed Convex Combination of Fast Convergence Algorithm to GSC Beamformer

This chapter proposes a convex combination of two FCNLMS adaptive filtering algorithms to the sidelobe canceling path of the GSC beamformer for speech enhancement. And also, proposed the signed algorithm to a convex combination of fast convergence filters to reduce real-time environmental noise and computational burden in the sidelobe canceling path.

4.1 Motivation

Combined adaptive filters in existing adaptive beamforming are ineffective to achieve better noise reduction at higher SNR levels and also gain more computation burden. Compared to a single adaptive filter, combination adaptive filters provide better noise reduction for all types of noises. Existing adaptive beamforming algorithms with convex combination adaptive filter gives noise cancellation for only particular real-time noises. In order to improve speech in a noisy environment, a robust adaptive filter in the sidelobe canceling path must be developed. A novel convex combination adaptive filtering method with a signed scheme is proposed for multi-channel speech enhancement to address various noisy types with less computational time.

4.2 Introduction

Fast convergence [224] algorithm has less computational complexity but gives less performance under various noisy conditions at high SNRs. And also, when the positions of the source signal change, the weight coefficient information used to update the adaptive filter will be lost; due to this, poor performance in the non-stationary environment combined adaptive filter [233]-[235] are designed, which give good convergence transition compared to the single adaptive filter.

Adaptive beamforming with an Affine Projection Algorithm [236] (APA) is presented to increase adaption performance, which provides higher noise reduction than existing time-domain techniques but fails in a noisy real-time environment. In the combined adaptive beamforming method [236], a combination of LMS-RLS adaptive filters in sidelobe canceller fails in noisy real-time conditions, and the computational burden is raised due to the mixing parameter. Another existing algorithm for noise reduction in recent times is, GSC beamformer with linear prediction filter [237], which is used in multi-channel speech enhancement systems addresses dereverberation and noise reduction but has high computation complexity. Barnov., in 2019, introduced GSC beamforming using controlled white Gaussian gain [238], where non-stationary environments are only limited to a single speaker. A modified change prediction [239] to GSC beamforming is applied, which holds good for echo cancellation but fails in interference suppression.

The above-mentioned algorithms give the motivation for the further improvement of the sidelobe canceller path of the GSC beamformer to achieve both noise reduction and less computational burden. A robust beamforming method should be designed to overcome these disadvantages.

In this chapter, a GSC beamformer with SCCFC adaptive filters is proposed to address the above-mentioned issues. The novelty of the GSC structure lies in the sidelobe canceling path. In this chapter, novelty is achieved in two steps. The first step is to consider FCNLMS as an adaptive filter in the convex combination algorithm to give a better noise reduction and a low computational complexity.

The second step is to employ a signed algorithm to further reduce the computational complexity in the mixing parameter design. In this way, using the signed algorithm with a convex combination of FCNLMS adaptive filters, both noise reduction and low computational complexity are achieved under various real-time noisy conditions. The proposed GSC beamforming using the SCCFC algorithm shows better noise reduction and lower computation complexity when compared to the existing algorithms.

The main contributions of the chapter4 are as follows:

(1) To improve the sidelobe canceling path of a robust GSC beamformer, a novel convex combination of fast convergence filters is proposed.

(2) To maintain a trade-off between computational complexity and noise reduction, a signed algorithm is introduced to the proposed filtering method.

(3) Tested the proposed multi-channel speech enhancement system under various real-time noisy conditions.

(4) The performance is shown in terms of computational complexity and noise reduction.

4.3 The Proposed GSC Beamforming with SCCFC Adaptive Filtering Algorithm

This section describes the multi-channel speech enhancement system in a real-time environment, as shown in Figure 4.1. GSC beamformer comprises of three major blocks: a fixed beamformer and modified blocking matrix (MBM) and whereas in the sidelobe canceling path where novel Signed Convex Combination of Fast Convergence (SCCFC) adaptive algorithm is proposed. The input to the proposed system is considered using a microphone array setup with noisy real-time conditions in a virtual conference room. The virtual conference room is designed based on the Image method [225], which takes the Room

Impulse Response in the form of a Mex function in MATLAB. The fixed beamformer block and MBM are explained in section 3.3.1 and section 3.3.2, respectively. Whereas in this chapter, the novel MBM is designed in order to utilize complete spatial information and is discussed below.

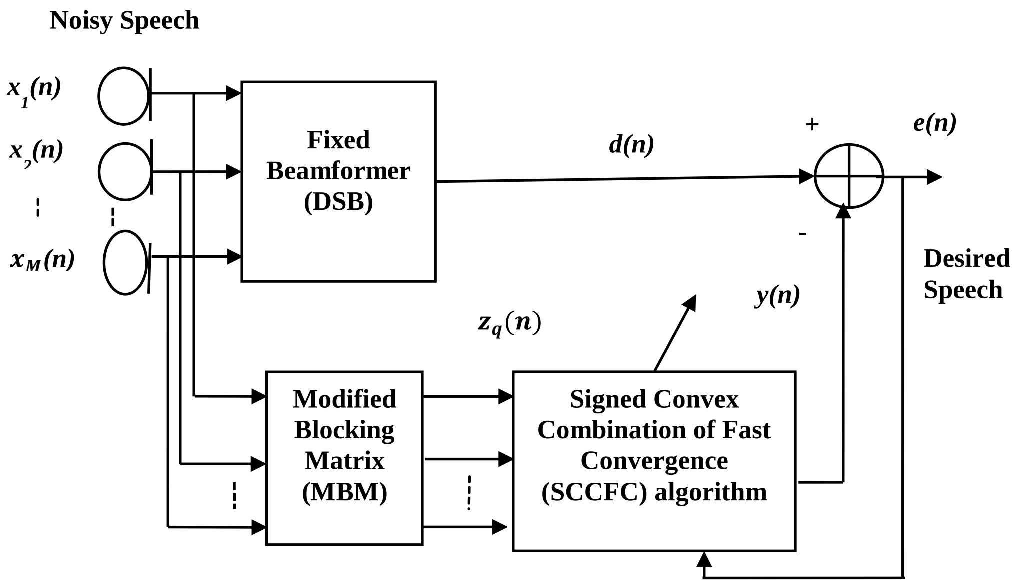

Figure 4.1: GSC Beamforming with Proposed SCCFC

$$\mathrm{MBM} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & & 0 & 0 \\ 0 & 0 & 0 & & 0 & 0 \\ & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & & -1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -1 \end{pmatrix}$$

The number of columns in the matrix indicates the microphone here with q = 1, . . . , Q where Q = M - 1, where M is the number of microphones. MBM gives the details of the complete noise present in the target signal and blocks the desired speech, and thus acts as noise reference for SCCFC. These constraints are considered to show the effectiveness of the proposed SCCFC in the GSC structure. The noise reference signals are adapted using the proposed SCCFC algorithm. The error at the output of the GSC beamformer is the difference between SCCFC output and speech reference. Then, the GSC-SCCFC beamformer output is given by $e(n) = d(n) - y(n)$. The error is updated using the proposed SCCFC algorithm until it is minimized. The derivation of the SCCFC algorithm is shown in the below section. Firstly, a convex combination of the FCNLMS adaptive filter is drawn, and then the signed algorithm is applied using the transfer approach in the next section.

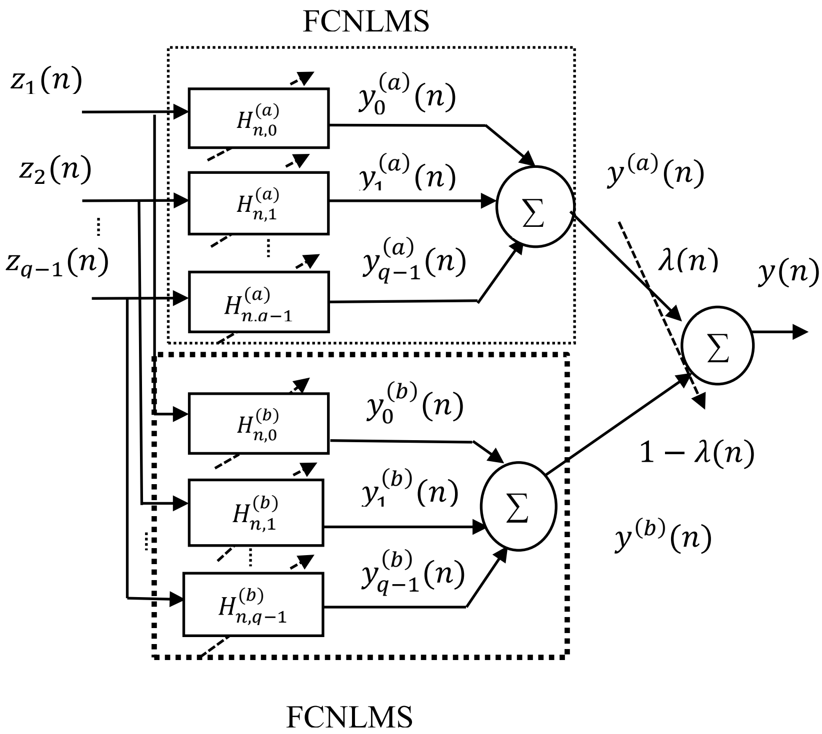

Figure 4.2: The Proposed SCCFC Adaptive Filter

4.4 Signed Convex Combination of Fast Convergence (SCCFC) Adaptive Filtering Algorithm

The proposed SCCFC block is a signed convex combination of two same fast convergence adaptive filters, i.e., FCNLMS, as shown in Figure 2 with updating rule which is given by:

$$H_{n,q}^{(l)} \in \mathbb{R}^N = \left(h_{0,q}^{(l)}(n), h_{1,q}^{(l)}(n), ...h_{N-1,q}^{(l)}(n)\right)^T \tag{4.1}$$

Where, $H_{n,q}^{(l)}$ is the vector with q^{th} filter coefficients of l^{th} system, with $l = a, b$ a n^{th} time instant, $l = a$ implies first FCNLMS filter and $l = b$ implies second FCNLMS filter. The q^{th} noise reference vector is expressed similarly.

$$Z_{n,q} \in R^N = \left(z_q\left(n\right)z_q\left(n-1\right)....z_q\left(n-N+1\right)\right)^T \tag{4.2}$$

The combined adaptive filter is obtained by combining the two adaptive filter outputs using the mixing parameter. $y^{(l)}(n)$ is the output of the combined adaptive filter, which is defined as

$$y^{(l)}(n) = \sum_{q=1}^{Q} y_q^{(l)}(n) \tag{4.3}$$

The convex combination of $y^{(a)}(n)$ and $y^{(b)}(n)$ is given by

$$y(n) = \lambda(n)y^{(a)}(n) + (1-\lambda(n))y^{(b)}(n) \tag{4.4}$$

Where, $\lambda(n)$ is a mixing parameter, and ranges from [0,1] [234]. When $\lambda(n)=0$, the small step size filter (slow filter) works effectively by maintaining low steady-state error. When $\lambda(n)=1$, the large step size filter (fast filter) works better with high convergence to limit the $\lambda(n)$ range between [0,1], the mixing parameter is expressed by a sigmoid function and an auxiliary parameter $I(n)$.

$$\lambda(n) = \frac{1}{1+e^{-I(n)}} \tag{4.5}$$

The convex combined filter error is minimized by adapting $I(n)$ and is defined as:

$$I(n+1) = I(n) + \mu_I e(n)\left[y^{(a)}(n) - y^{(b)}(n)\right]\lambda(n)[1-\lambda(n)] \tag{4.6}$$

When $\lambda(n)$ it is equal to 0 or 1, to reduce idleness, the auxiliary parameter is limited to $[-I^+, I^+]$, such that the mixing parameter is made to move in $[1-\lambda^+, \lambda^+]$. Here, I^+ and λ^+

are small positive constants. The update rule of the weight vector $H_{n,q}^{l}(n+1)$ for l^{th} an adaptive filter $(l=a,b)$ is written as where $\tilde{C}_N(n)$ is dual Kalman gain [224], $\gamma_N(n)$ is the Likelihood variable [224].

Dual Kalman gain is defined as:

$$H_{n,q}^{(l)}(n+1) = H_{n-1,q}^{(l)} - \mu e_q^{(l)}(n)\gamma_N(n)\tilde{C}_N(n) \tag{4.7}$$

$$\widetilde{C_N}(n) = -\frac{Z_{n,p}(n)}{\frac{\lambda}{1-\lambda}\sigma_z^2 + c_o} \tag{4.8}$$

Where c_o and λ is a small positive constant. Likelihood variable is defined as:

$$\gamma_N(n) = \frac{1}{1-\sum_{k=1}^{N} v(n-k+1)} \tag{4.9}$$

Where $v(n) = C_N^{(-1)}x(n)$ is the shifting component, $e_q^{(l)}(n)$ in Equation (4.7). is the error estimator of FCNLMS filter with q^{th} error signal, expressed as

$$e_q^{(l)} = d(n) - y_q^{(l)}(n) \tag{4.10}$$

Where $y_q^{(l)}(n)$ is the FCNLMS filter output of q^{th} filter and is expressed as

$$y_q^{(l)}(n) = Z_{n,q}^{T} H_{n-1,q}^{(l)} \tag{4.11}$$

Where μ_l is the step size of l^{th} adaptive filter. The overall weight coefficient of the convex combination of the adaptive filters is expressed as

$$H_{n,q} = \lambda(n)H_{n,q}^{(a)}(n) + [1-\lambda(n)]H_{n,q}^{(b)} \tag{4.12}$$

By updating the filter with the help of the mixing parameter, there is a decent trade-off between the convergence speed and steady-state error. However, such algorithms require the fixing of mixing parameters while updating the weights resulting in the loss of information. Complexity burden increases due to $I(n)$ the update rule and also fails to work for real-time noises. A GSC beamformer should be constructed with fewer operations in the $I(n)$ update rule for various real-time noise reductions to avoid computation complexity.

To overcome the computational burden on mixing parameters and overall real-time noise reduction. In this chapter, a signed algorithm is proposed for the convex combination of fast convergence adaptive filters, which is described in the next section.

4.5 Signed Algorithm to a Convex Combination of Fast Convergence Adaptive Filter

We propose the SCCFC algorithm in this section. By opting for this signed algorithm, the mixing parameter update rule is changed to limit the squared estimation error.

$$J(n) = \frac{1}{2}e^2(n) \tag{4.13}$$

The gradient $\nabla_I J(n)$, is normalized and $I(n)$, is updated recursively, and is expressed as:

$$I(n+1) = I(n) - \mu_I \frac{\nabla_I J(n)}{\|\nabla_I J(n)\|} \tag{4.14}$$

Here μ_I is a step-size and is a small positive constant, $\nabla_I J(n),$ is defined as

$$\nabla_I J(n) = -e(n)(y_a(n) - y_b(n))\lambda(n)(1-\lambda(n)) \tag{4.15}$$

The normalized gradient $\frac{\nabla_I J(n)}{\|\nabla_I J(n)\|}$ in Equation (4.14), can be expressed as

$$\frac{\nabla_I J(n)}{\|\nabla_I J(n)\|} = \operatorname{sgn}(\nabla_I J(n)) \tag{4.16}$$

where $\operatorname{sgn}(.)$ is a sign function [234] and is defined as

$$\operatorname{sgn}(.) = \frac{u}{\|z\|} = \begin{cases} 1 & if \quad z > 0 \\ 0 & if \quad z = 0 \cdot \rightarrow \\ -1 & if \quad z < 0 \end{cases} \tag{4.17}$$

Therefore, Equation (4.15), can be written as

$$I(n+1) = I(n) + \mu_I \operatorname{sgn}\left(e(n)y^{(a)}(n) - y^{(b)}(n)\lambda(n)(1-\lambda(n))\right) \tag{4.18}$$

As $\lambda(n) > 0$ & $1-\lambda(n) > 0$ the parameter $I(n)$ in Equation (4.18), can also be represented as

$$I(n+1) = I(n) + \mu_I \operatorname{sgn}\left(e(n)(y^{(a)}(n) - y^{(b)}(n))\right) \tag{4.19}$$

$$I(n+1) = I(n) + \mu_I \operatorname{sgn}\left(e(n)(e^{(a)}(n) - e^{(b)}(n))\right) \tag{4.20}$$

The proposed SCCFC algorithm can reduce computational complexity and attain robustness by replacing $e(n)[y^{(a)}(n)-y^{(b)}(n)]\lambda[1-\lambda(n)]$ it with a normalized gradient $\frac{\nabla_I J(n)}{\left\|\nabla_I J(n)\right\|}$.

An instant transfer approach [234] can be utilized to improve speech further with less computation while keeping high convergence.

if nmod $D_o = 0$ and $I(n+1) = I^+$ then

$$H_{n,q}^{(b)}(n+1) = H_{n,q}^{(a)}(n+1)$$

endif

Where D_o is the length of the Window. During convergence transition, an instant transfer algorithm is applied when the first FCNLMS is effective than the second FCNLMS filter. The computation cost of this algorithm is smaller compared to the traditional combination filters. Due to the predefined window length, the computation burden is still reduced so that the proposed SCCFC works effectively for various real-time noises with low complexity in updating the adaptive filter.

As discussed to the sidelobe canceller path SCCFC algorithm is designed and computation issue is solved using signed scheme to the convex combination algorithm which is explained here.

Overall steps involved in the proposed SCCFC algorithm is summarized below

Algorithm : Summary of Proposed SCCFC Algorithm

Initialize

$D_o, \mu_l, (l = a, b), \mu_{I,} I^+,$ $\widetilde{C_N}(0) = 0$ $\gamma_N(0) = 0, I(0) = 0, \lambda(0) = 0.4,$

$H_{n,q}^{(a)}(0) = 0, H_{n,q}^{(b)}(0) = 0$

Loop $n = 1 \rightarrow$

$$e_q^{(a)}(n) = d(n) - y_q^{(a)}(n)$$

$$y(n) = \lambda(n) y^{(a)}(n) + (1 - \lambda(n)) y^{(b)}(n)$$

$$H_{n,q}^{(a)} = H_{n-1,q}^{(a)} + \mu \frac{Z_{n,q} e_q^{(a)}(n)}{Z_{n,q}^T Z_{n,q} + \delta}$$

$$I(n+1) = I(n) + \mu_I e(n) \left[y^{(a)}(n) - y^{(b)}(n) \right] \lambda(n) [1 - \lambda(n)]$$

$$\lambda(n+1) = \frac{1}{1 + e^{-I(n+1)}}$$

Signed Algorithm

If

$I(n+1) < -I^{+}$

$I(n+1) = -I^{+}$

$\lambda(n+1) = 0$

endif

if

$\lambda(n+1) = 1$

if $(\mathrm{mod}(n-1), D_o = 0)$

$H_{n,q}^{(b)}(n+1) = H_{n,q}^{(a)}(n)(n+1)$

Endif

Else

$H_{n,q} = \lambda(n) H_{n,q}^{(a)}(n) + [1-\lambda(n)] H_{n,q}^{(b)}$

let n=n+1

end

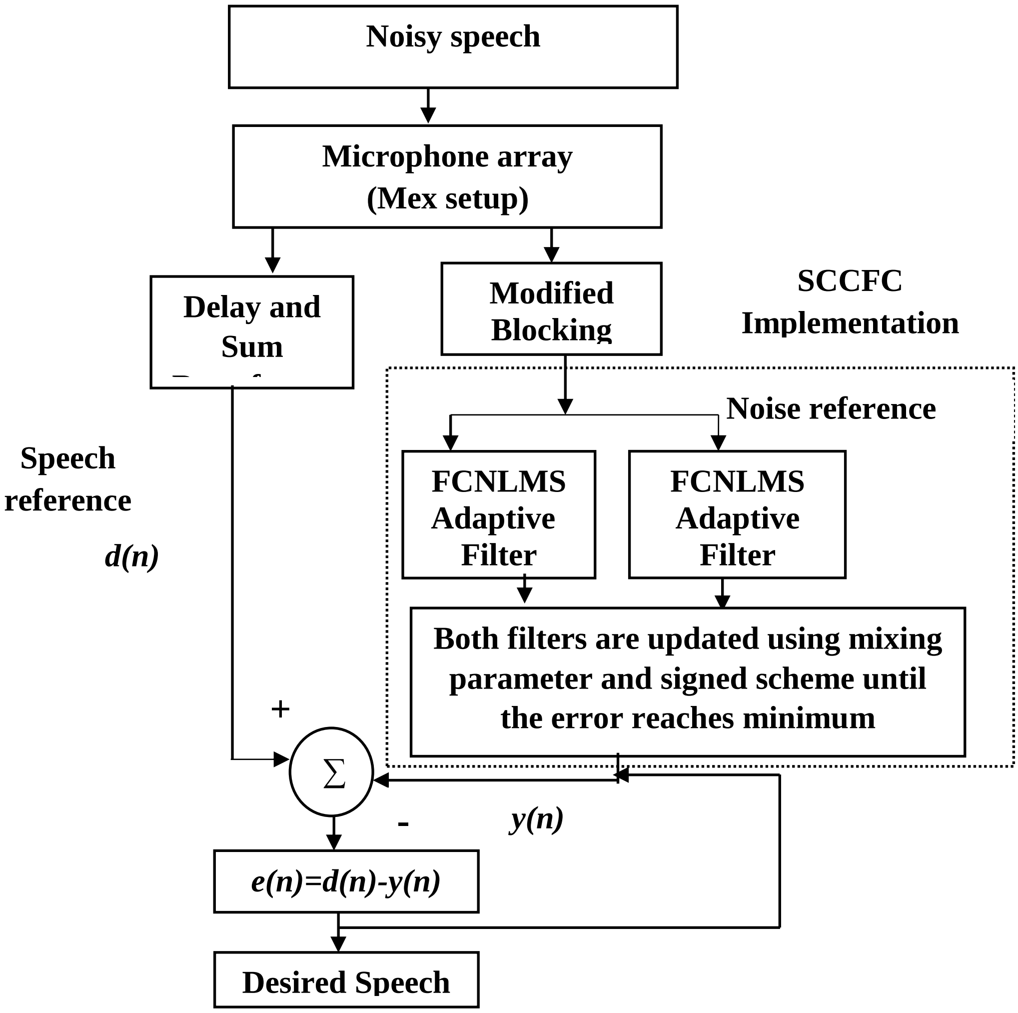

Figure 4.3: Workflow of the Proposed GSC-SCCFC

The workflow of the proposed multi-channel speech enhancement system (GSC-SCCFC) is as shown in Figure 4.3.

4.6 Computational Complexity

The computational complexity of the LMS [216], NLMS [217], FCNLMS [224], CLMS [234], and the proposed SCCFC algorithms are compared in this section. Here the length of the adaptive algorithm is given by N. For a regular LMS algorithm takes 2N+1 multiplications to update the filter. The basic NLMS and FCNLMS algorithms require a 2N number of multiplications. The proposed SCCFC algorithm, which is a combination of the two same filters FCNLMS, requires 4N multiplications to update the filter components. According to Equation (4.14), updating $I(n)$, the proposed SCCFC requires only three multiplications, whereas the existing CLMS algorithm requires six multiplications to update the same $I(n)$ parameter. Due to the usage of the signed algorithm with known window length, the proposed SCCFC algorithm reduces the computational operations compared to the conventional algorithms. When it comes to stability, the relative variations in $e(n)$ is maintained by taking μ as a small positive constant. Also, the mixing parameter $I(n)$ is independent on $J(n)$, $I(n)$ becomes more stable when it $\nabla_I J(n)$ is small. Finally, the proposed GSC-SCCFC gives less computation complexity with 4N multiplications, where N=256 is the length of the filter and requires three primary combinations in the update rule, which is very less compared to existing algorithms. The computational complexity of the proposed multi-channel speech enhancement system is compared with the existing algorithms, as shown in Table 4.1. The proposed algorithm also gives good trade-off stability compared to the other algorithms.

Table 4.1: Comparison of Computation Complexity

Algorithms	**Multiplications**	**Primary Combinations**	**Precise Weight Calculations**	**Weight Transfer**
LMS [216]	2N+1	-	-	-
NLMS [217]	2N	-	-	-
FCNLMS [224]	2N	-	-	-
CLMS [234]	4N+2	6	2N	2N
SCCFC (proposed)	**4N**	**3**	**2N**	**-**

4.7 Results and discussions

In this section, the simulation of the proposed GSC-SCCFC in noisy real-time conditions is evaluated and explained. In order to show the performance of the sidelobe canceller of GSC, the same noisy input in a virtual room is considered. The proposed GSC-SCCFC method considers the simulation parameter as explained in chapter 3, section 3.4, and description as shown in Table 3.1 as well.

The input signals are same as in chapter 3 provided to the Mex setup, which gives a combination of the desired speech with real-time noise. Whereas in order to evaluate the proposed GSC-SCCFC at higher SNRs, in this chapter, 15 dB input SNR is also considered.

The degraded speech is an input to the DSB to evaluate the delay from each microphone and obtain a reference enhanced signal. After that, the input degraded speech is given to MBM. Using the MBM matrix, the subtraction of the delays caused on the adjacent microphones is calculated. Further, at the MBM output, a noise reference is generated as discussed in chapter 3, section 3.5; here, a novel MBM is implemented to completely utilize spatial information. Finally, the same reference noise is applied to the SCCFC block as input, where the weights of the individual filters are updated and combined using a mixing parameter. Due to the proposed SCCFC algorithm, the error is minimized, and enhanced speech is attained at the output of the GSC beamformer.

The proposed GSC-SCCFC algorithm is compared with different existing algorithms like Combined adaptive beamforming [236], GSC with improved linear prediction [237], GSC with controlled white Gaussian [238], combined beamforming and echo cancellation [239], which are represented as GSC-CC [236], GSC-LP [237], GSC-CWGN [238], and GSC-CBE [239] respectively. GSC-CC algorithm uses a combination of adaptive filters [LMS-RLS] for noise reduction. GSC-LP multi-channel improves linear predictors to improve the spatial filter. Both GSC-CBE and GSC-CWGN are used for noise reduction under white noise.

4.7.1 Performance Analysis of Proposed GSC-SCCFC Algorithm

The performance of the proposed GSC-SCCFC algorithm is evaluated using standard speech processing performance metrics, namely Perceptual Evaluation of Speech Quality

[231] (PESQ), Segmental SNR (SSNR) [240], Log-Spectral Distance (LSD) [232], and Log-Likelihood Ratio [231].

4.7.1.1 Comparison of PESQ Score for the Proposed Algorithms

PESQ [231] standards are discussed in the section. 3.4.1.1. Table.4.2 shows the PESQ score comparison of GSC-SCCFC over existing methods. Under station noise, for -10 dB, the proposed GSC- SCCFC PESQ score is 3.302, but for GSC-CC, it is 2.411. Similarly, at 15 dB input SNR for street noise, PESQ for the proposed GSC-SCCFC is 4.393, but for GSC-CWGN and GSC-CBE, it is 3.401 and 3.567, respectively.

At -10 dB car noise, the proposed GSC-SCCFC method gives a PESQ of 2.632, but for GSC with CWGN and CBE, it is 2.305 and 2.401, respectively. Similarly, at 15 dB PESQ for GSC-CWGN, GSC-CBE, and the proposed GSC-SCCFC are 3.232, 3.451, and 4.365, respectively. Similarly, for the remaining noises too, the perception is improved for the enhanced speech using the proposed GSC-SCCFC algorithm when compared with conventional algorithms, as shown in Table 4.2. For the proposed method, an improvement in PESQ of 4.393 is achieved, which is very much closer to the maximum PESQ that can be achieved. Due to SCCFC, at the output, the desired speech perception is attained.

4.7.1.2 Segmental SNR (SSNR)

SSNR [240] SSNR is the renowned objective measure for speech enhancement. In SNR, the complete signal is taken into consideration, whereas, for SSNR, the segments with 256 samples per frame are considered. (k=256, with 50 percent overlap). The higher the Segmental SNR, the more will be the speech quality.

SSNR is defined as

$$SSNR = \frac{10}{N} \frac{\sum_{q=0}^{N-1} 10 \log \sum_{q=0}^{M-1} z^2 (q + \frac{nM}{2})}{\sum_{q=0}^{M-1} [(q + \frac{nM}{2}) - e(q + \frac{nM}{2})]^2} \tag{4.21}$$

Table 4.2: Performance Comparison of PESQ and SSNR for Proposed GSC-SCCFC with Existing Algorithms

SNR in dB	Noise Type	GSC-CC [236]		GSC-LP [237]		GSC-CWGN [238]		GSC-CBE [239]		GSC-SCCFC (Proposed)	
		PESQ	**SSNR**	**PESQ**	**SSNR**	**PESQ**	**SSNR**	**PESQ**	**SSNR**	**PESQ**	**SSNR**
-10	Car	2.401	2.9	2.482	3.7	2.305	4.2	2.401	5.6	**2.632**	**11.2**
-10	Restaurant	2.325	4.6	2.062	4.9	2.232	5.8	2.591	5.9	**3.013**	**15.3**
-10	Babble	2.303	2.8	2.123	4.2	2.200	5.2	2.501	6.1	**3.022**	**16.1**
-10	Station	2.411	5.2	2.102	3.2	2.428	4.5	2.656	5.3	**3.302**	**12.1**
-10	Airport	2.510	3.7	2.323	4.2	2.398	5.7	2.618	6.3	**2.801**	**13.3**
-10	Street	2.241	4.4	2.208	5.5	2.511	6.2	2.674	7.7	**3.011**	**16.7**
-5	Car	2.008	3.6	2.569	4.2	2.507	5.1	2.604	7.2	**2.804**	**17.2**
-5	Restaurant	2.211	4.7	2.381	3.6	2.316	4.8	2.623	7.8	**3.093**	**21.5**
-5	Babble	2.007	5.1	2.312	4.5	2.421	5.9	2.729	8.1	**3.201**	**18.1**
-5	Station	2.118	3.5	2.421	3.8	2.551	6.7	2.634	8.4	**3.104**	**16.8**
-5	Airport	2.092	2.8	2.383	4.1	2.483	6.2	2.715	9.4	**3.302**	**15.2**
-5	Street	2.183	5.7	2.572	5.9	2.501	7.5	2.749	9.5	**3.259**	**20.3**
0	Car	2.010	7.2	2.454	6.9	2.611	5.9	2.734	9.2	**3.405**	**21.4**
0	Restaurant	2.486	3.1	2.687	7.3	2.643	6.4	2.787	8.5	**3.401**	**25.3**
0	Babble	2.201	5.4	2.532	7.9	2.571	6.3	2.663	10.5	**3.569**	**24.2**
0	Station	2.229	7.5	2.556	6.8	2.691	6.9	2.719	9.7	**3.582**	**22.8**
0	Airport	2.237	4.6	2.399	8.1	2.582	7.1	2.697	10.6	**3.691**	**21.5**
0	Street	2.597	6.9	2.573	7.9	2.660	8.2	2.793	11.5	**3.710**	**25.2**
5	Car	2.602	8.8	2.735	9.3	2.812	9.5	2.867	10.7	**3.408**	**21.7**
5	Restaurant	2.676	7.2	2.812	8.9	2.752	9.8	2.702	11.3	**3.421**	**25.1**
5	Babble	2.698	5.8	2.790	9.5	2.862	10.2	2.923	11.7	**3.543**	**24.2**
5	Station	2.702	4.7	2.809	10.2	2.951	10.7	2.921	12.1	**3.521**	**22.6**
5	Airport	2.818	4.9	2.901	9.9	3.028	10.5	3.052	11.9	**3.671**	**21.9**
5	Street	2.992	8.9	3.095	12.8	3.191	11.7	3.179	13.8	**3.722**	**25.2**

10	Car	2.901	10.6	3.011	11.7	3.221	12.7	3.328	13.3	**3.992**	**31.9**
10	Restaurant	2.822	12.4	3.039	11.6	3.219	13.7	3.222	13.5	**4.072**	**32.4**
10	Babble	2.899	15.1	3.156	11.9	3.312	13.2	3.356	14.2	**4.287**	**34.1**
10	Station	2.907	12.2	3.121	12.7	3.224	12.9	3.401	14.8	**4.356**	**32.8**
10	Airport	2.974	13.2	3.111	11.6	3.212	13.2	3.456	14.1	**4.456**	**34.1**
10	Street	3.012	14.7	3.223	15.6	3.431	16.3	3.582	17.7	**4.311**	**31.6**
15	Car	3.061	15.3	3.151	16.2	3.232	16.3	3.451	19.8	**4.365**	**32.5**
15	Restaurant	2.921	15.9	3.164	16.2	3.379	1.8	3.511	20.9	**4.346**	**34.3**
15	Babble	3.056	15.2	3.178	15.4	3.245	16.9	3.489	20.3	**4.310**	**34.1**
15	Station	3.110	15.8	3.208	15.9	3.212	16.2	3.501	22.6	**4.355**	**33.8**
15	Airport	3.089	15.5	3.219	16.9	3.302	17.8	3.451	21.7	**4.387**	**34.8**
15	Street	3.121	16.2	3.410	17.4	3.401	18.4	3.567	22.1	**4.393**	**34.6**

From Table 4.2 at -10 dB with car noise, SSNR for GSC-SCCFC algorithm is 11.2, but for GSC-CC, GSC-LP, GSC-CWGN, and GSC-CBE, it is 2.9, 3.7, 4.2, and 5.6, respectively. Similarly, SSNR for 15 dB GSC-SCCFC is 32.5 dB while that for GSC-CC, GSC-LP, GSC-CWGN, and GSC-CBE are 15.3 dB, 16.2 dB, 17.6 dB, and 19.8 dB, respectively. SSNR for the proposed GSC-SCCFC shows improved performance as noise present in each frame is reduced. Also, for 15 dB station noise, SSNR for GSC-SCCFC is 33.8, but for GSC-CC, GSC-LP, GSC-CWGN, and GSC-CBE, it is 15.8 dB, 15.9 dB, 16.2 dB, and 22.6 dB, respectively. Likewise, for 15 dB street noise, GSC-SCCFC, GSC-CC, GSC-LP, GSC-CWGN, and GSC-CBE results in SSNRs of 34.6 dB, 16.2 dB, 17.4 dB, 18.4 dB, and 22.1 dB, respectively. Likewise, the performance of SSNR is improved gradually for different real-time noises, which are represented in Table 4.2. SSNR for the proposed GSC-SCCFC with four microphones gives better noise reduction in the segmental analysis.

4.7.1.3 Log Spectral Distance (LSD)

Log spectral distance (LSD) [232] is an advanced metric; the reduction in the spectral distance is calculated using LSD. The expression LSD is provided in Eq. (4.22),

$$LSD = \frac{10}{N}\sum_{n=0}^{(N-1)}\frac{1}{(M+1)}\sum_{n=0}^{(\frac{M}{2})}[\log z_q(n) - \log e(n)]^2 \tag{4.42}$$

LSD for the proposed GSC-SCCFC algorithm is compared with existing algorithms for various real-time noises, as shown in Figures 4.4 (A) to 4.4 (F). The proposed algorithm showing lower values of LSD implies better performance. The reduction of the spectral distance is achieved using MBM by utilizing the complete spatial information. As the distance between the frames decreases, the distortion gets reduced. At 10 dB for car noise, LSD for GSC-SCCFC is 0.91 dB, but for GSC-CC, GSC-LP, GSC-CWGN, and GSC-CBE, it is 2.04 dB, 2.22 dB, 2.39 dB, 2.21 dB. For 15 dB input SNR under station noise, LSD for GSC-SCCFC, GSC-CC, GSC-LP, GSC-CWGN, and GSC-CBE is 0.51 dB, 1.54 dB, 2.16 dB, 2.03 dB, and 1.73 dB, respectively. The proposed GSC-SCCFC achieves better performance when compared to the existing algorithms. LSD gradually decreases for the remaining noises, which are shown in Figure 4.4 (A) to (F).

A smaller spectral distance for the proposed GSC-SCCFC for 15 dB at 0.41 dB is observed under street noise. Using the proposed SCCFC algorithm in the adaptive filtering block of GSC beamforming, better quality is achieved for the output speech, which is represented in terms of LSD as shown in Figures 4.4 (A) to (F). 10 dB for car noise, LSD for GSC-SCCFC is 0.91 dB, but for GSC-CC, GSC-LP, GSC-CWGN, and GSC-CBE, it is 2.04 dB, 2.22 dB, 2.39 dB, 2.21 dB. For 15 dB input SNR under station noise, LSD for GSC-SCCFC, GSC-CC, GSC-LP, GSC-CWGN, and GSC-CBE is 0.51 dB, 1.54 dB, 2.16 dB, 2.03 dB, and 1.73 dB, respectively. The proposed GSC-SCCFC achieves better performance when compared to the existing algorithms. LSD gradually decreases for the remaining noises, which are shown in Figure 4. A smaller spectral distance for the proposed GSC-SCCFC for 15 dB at 0.41 is observed under street noise. Using the proposed SCCFC algorithm in the adaptive filtering block of GSC beamforming, better quality is achieved for the output speech, which is represented in terms of LSD as shown in Figures 4.4 (A) to (F).

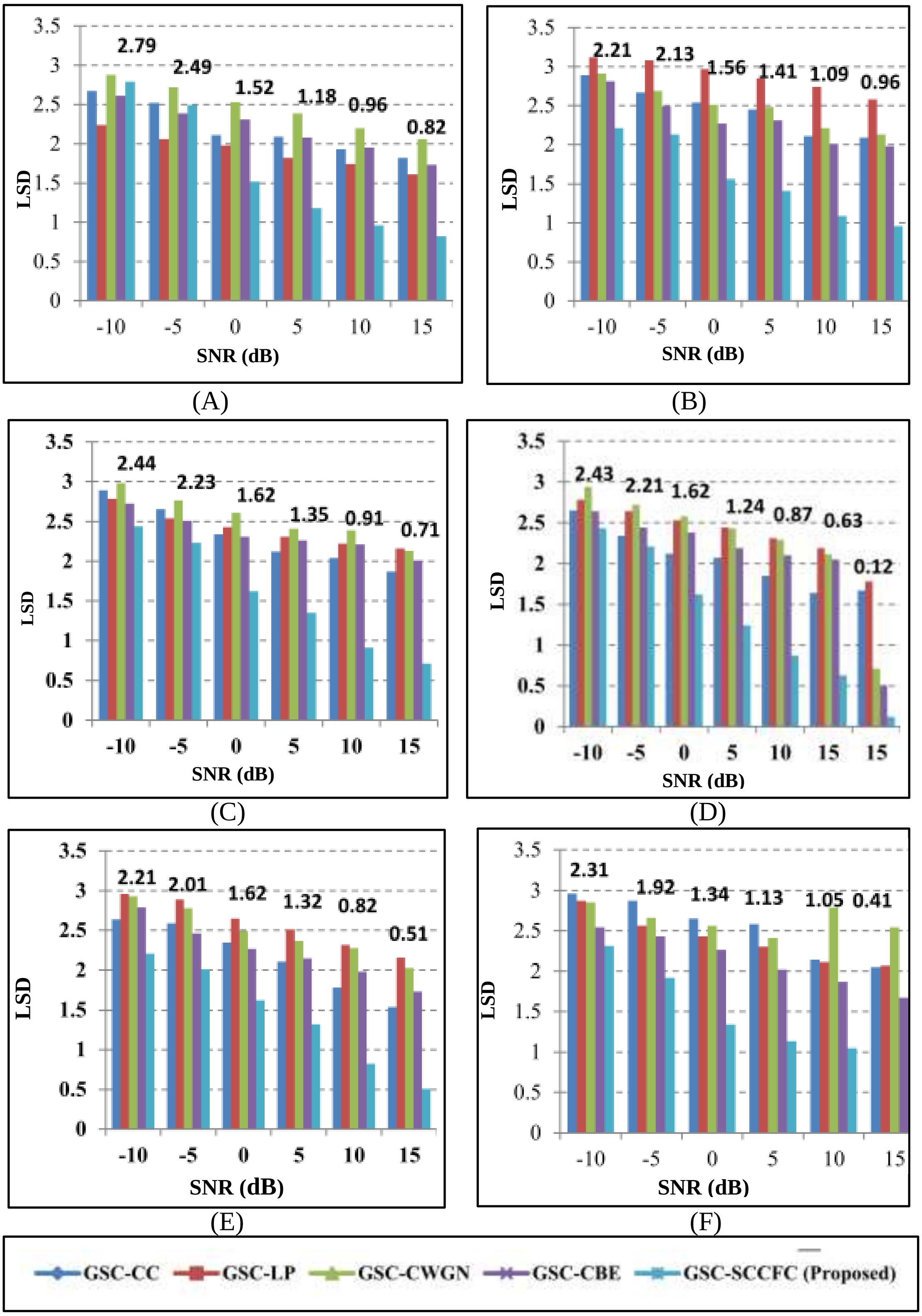

Figure 4.4: Performance Comparison of LSD Under Different Noises (A) Airport (B) Babble (C) Car (D) Restaurant (E) Station (F) Street

4.7.1.4 Comparison of LLR Score for the Proposed Algorithm

LLR formulation is explained in section 3.4.1.3. Lowering the LLR will be more speech performance quality. For car noise at 15 dB input SNR, LLR is 0.36 for the proposed GSC-SCCFC, but for GSC-CC, GSC-LP, GSC-CWGN, and GSC-CBE, it is 0.89, 0.87, 0.83, and 0.72, respectively. For station noise with 15 dB input SNR, GSC-SCCFC results in an LLR of 0.07, whereas GSC-CC, GSC-LP, GSC-CWGN, and GSC-CBE, it is 1.52, 1.41, 0.42, and 0.73, respectively. At 15 dB input SNR, LLR of 0.04 under airport noise is achieved by the proposed GSC-SCCFC, which is very low when compared to the other conventional algorithms as shown in Figures 4.5 (A) to 4.5 (F).

4.7.1.5 Computational Time

The computational time is calculated in this section. An input degraded speech signal from the real-time environment with a duration of 2.814 seconds is considered. The simulations are executed on an intel i7 core processor with a 2.20 GHz clock speed with 8 GB RAM. The operating system used is Windows 10. The GSC-SCCFC is compared with the conventional algorithm in Table 4.4. GSC-SCCFC shows less computation of 0.93 s is shown in Table 4.4. The conventional algorithm shows low performance in noise reduction and gives high computation time is shown in Table 4.4. The proposed GSC-SCCFC method gives better performance with lower computational time.

Table 4.3 Computation Time

Methods	**Computation time (s)**
GSC-CC [236]	2.38
GSC-LP [237]	1.98
GSC-CWGN [238]	2.71
GSC-CBE [239]	2.29
GSC-SCCFC (proposed)	0.93

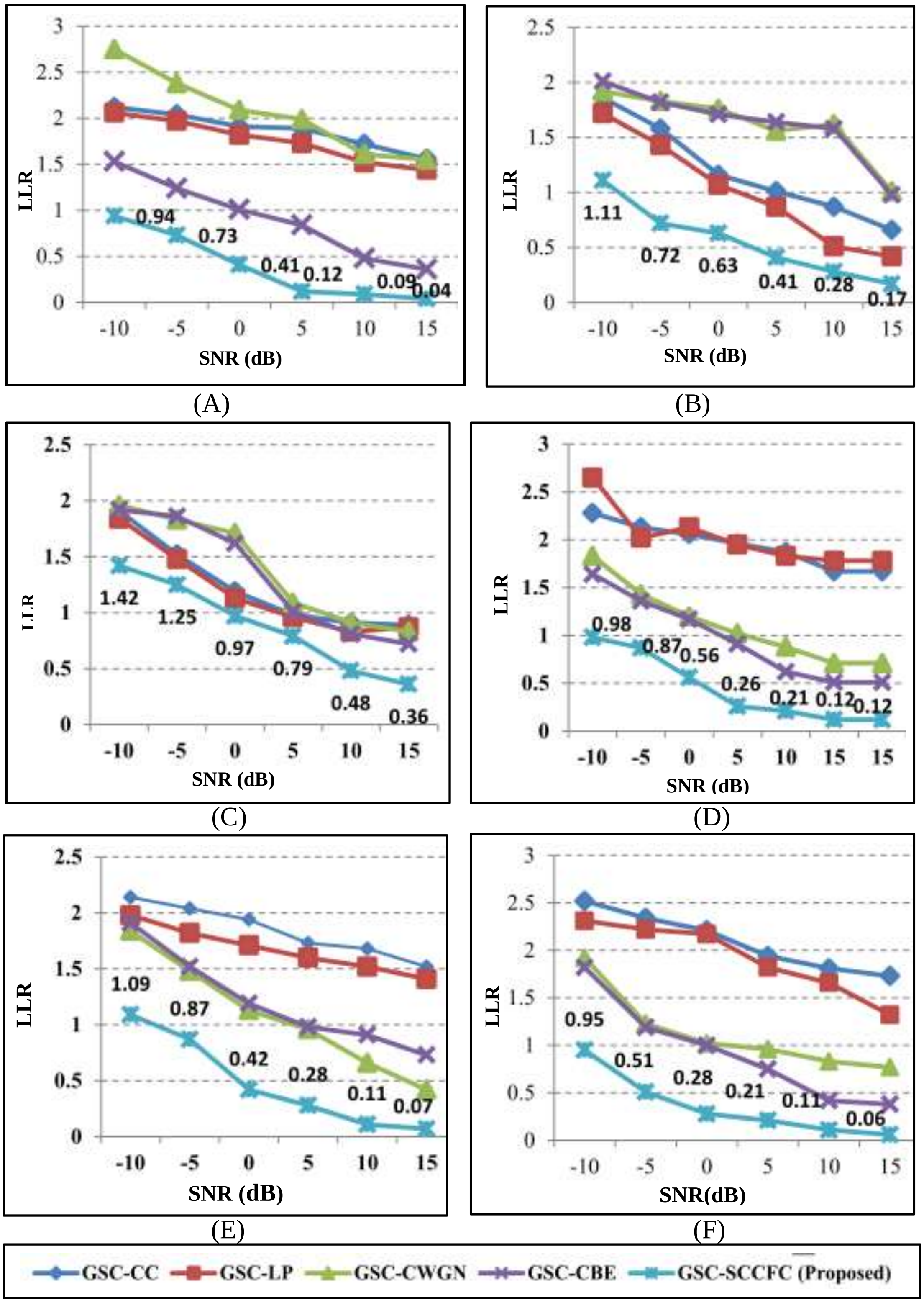

Figure 4.5: Performance Comparison of LLR Under Different Noises (A) Airport (B) Babble (C) Car (D) Restaurant (E) Station (F) Street

4.7.1.6 Waveforms

In Figure 4.6 and Figure 4.7, the time domain plots and spectrograms of the proposed multi-channel speech enhancement system are illustrated, which shows the proposed GSC-SCCFC noise reduction performance for 5 dB car noise. The enhanced speech signal of the proposed GSC-SCCFC algorithm shown in Figure 6 looks similar to the clean speech signal. The enhanced speech signal is also attained at low SNRs. PESQ of 4.393 is obtained using the proposed GSC-SCCFC method, which is the highest when compared to GSC-CC [236], GSC-LP [237], GSC-CWGN [238], and GSC-CBE [239], which have scores of 3.121, 3.410, 3.401, and 3.567 for street noise at 15 dB input SNR, respectively.

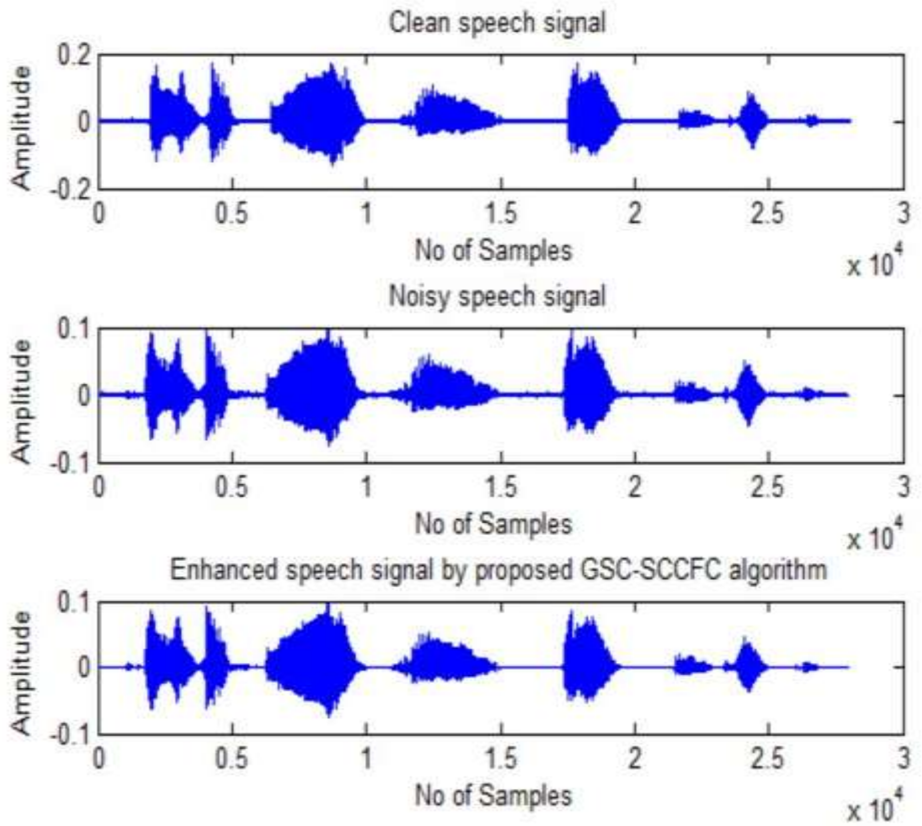

Figure 4.6 Time Domain Plot of Proposed GSC-SCCFC at 5 dB Car Noise

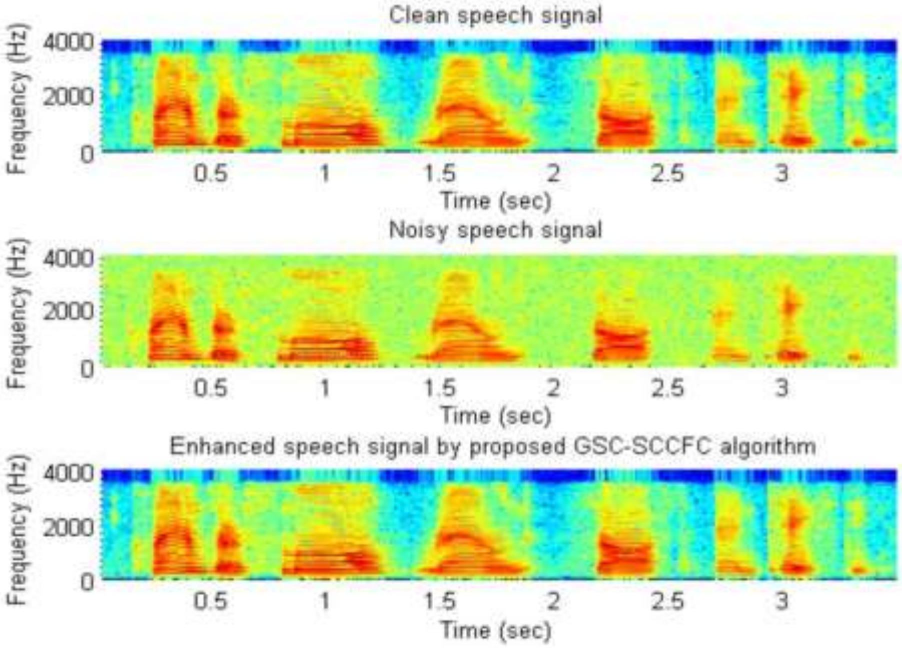

Figure 4.7 Spectrogram of Proposed GSC-SCCFC at 5 dB Car Noise

The PESQ score of the proposed method almost reaches the maximum achievable PESQ score of 4.5. In the same way, the proposed method has significantly higher SSNR, and lower LSD, LLR, and also lower computational complexity values, clearly showing its superiority in performance and its ability to provide a better trade-off between noise reduction and computational complexity compared to other methods.

4.8 Summary

A multi-channel speech enhancement system using the GSC-SCCFC algorithm is proposed in this chapter. Both noise reduction and low computational complexity is achieved using GSC-SCCFC. GSC beamforming using the proposed SCCFC algorithm is compared with the existing algorithms under various real-time noisy conditions. In the proposed multi-channel speech enhancement system, a signed algorithm is adapted into the convex combination of two same adaptive filters (FCNLMS) with different step sizes, which effectively reduces the computational burden in updating the weight coefficient and also reduces the real-time noises present in the input signal. The proposed system gave better speech intelligibility scores of 4.393 of PESQ and SSNR of 34.8 dB for 15 dB airport noise, respectively. Other measures like LSD and LLR gave values of 0.41 for 15 dB street noise and 0.04 for 15 dB airport noise respectively. For the proposed GSC-SCCFC algorithm, LSD and LLR are smaller values compared to the conventional algorithms. Lower LLR and LSD values, showing the lower distance between the frames, which resembles improved speech quality. The proposed algorithm is essential for smooth communication through speech in noisy real-time conditions. In the same way, if we consider a diffuse noise environment, the input signal will not be analyzed only with an adaptive filter in order to evaluate noisy input speech in an adverse environment. Novel postfilters have to be implemented in the frequency domain, which verifies the low frequencies and high frequencies of the input noisy signal frame by frame to eliminate speech absence frames. So, the GSC beamforming in the frequency domain is designed for an adverse environment in chapter 5 and evaluated for different SNR levels at the output of the postfilter.

Chapter 5

Adaptive Beamforming using Zelinski-TSNR Multi-Channel Postfilter for Speech Enhancement

This chapter proposes the use of a postfilter to GSC beamformer to suppress direction and diffuse noise. A novel multi-channel postfilter is proposed at the output of the GSC beamformer for multi-channel speech enhancement.

5.1 Motivation

In the case of directional and diffuse noise in low frequencies, most of the multi-channel speech enhancement methods give a lacking noise reduction. Adaptive beamforming with postfilters provide better solutions to address these problems. A robust multi-channel postfilter should be developed to suppress directional and diffuse noise, under various noisy environments, which are very much essential in applications like mobile phones, teleconferencing, and hearing aids, etc. We need noise-free information for effective communication.

5.2 Introduction

In the case of adverse environments like diffuse noise fields, a particular interference speakers or noise comes from an unknown direction. Aiming the speech enhancement for degrade signal becomes quite difficult. In order to estimate the diffuse noise such as car and office noise, it is required to approach postfilters for adaptive beamformers. From the past few decades, various beamforming [7], [20]-[22] methods have been introduced to remove directional noise. O. L. Frost [20] had introduced a beamformer with an array structure for adaptive broadband processing. Similarly, L. J. Griffiths [21] proposed an alternative structure to Frost's [20], named GSC beamformer, that suppresses interferences from different directions and also provides low computational complexity. In a reverberant environment, the Widrow [22] adaptive noise canceller may face signal cancellation due to improper microphones and steering vector errors. Also, the adaptive filter block of the GSC beamformer [241] produces transient noise due to fixed step size. In the previous chapter, we had seen GSC beamformers using different adaptive filters to suppress various real-time noisy environments. Owing to these shortcomings, which are discussed in chapters 3 and 4, there is a significant need to combine some filtering methods to have better noise reduction in a multi-source environment.

In the case of diffuse noise, i.e., car noise, office noise, etc., where the noise spectrum power is uniform in all directions, Zelinski postfilter [28] is applied, which estimates cross and auto-correlation to obtain an enhanced speech. The generalized expression of the Zelinski postfilter can be analyzed based on prior knowledge of the noise field. Mc Cowan [189] has given a generalized expression for Zelinski postfilter for office room recordings, but it fails while considering highly correlated noise. J. Li [244] introduced Improved Zelinski (IZ) postfilter to enhance desired speech from the diffuse noise fields by applying Wiener postfilter for low frequencies. Apart from the Wiener postfilter, a Two-Step Noise Reduction (TSNR) method was given by C. Plapous [245] to have a better noise reduction in adverse environments. When a person is in motion, the Decision Directed (DD) approach is applied for estimating *a priori SNR* of the current frame. I. Cohen [232] introduced a multi-channel speech presence probability-based postfilter for non-stationary environments.

S Gannot [194] introduced transfer function GSC with a multi-channel postfilter and compared it with a single channel postfilter. This method fails in diffuse noise fields. An improved GSC with multi-channel postfiltering is presented by K. Li [248], which eliminates the directional noise but is unable to suppress diffuse noise and has caused more speech distortion in low frequencies. A GSC beamforming is designed to suppress directional noise in an adverse environment to address the above limitations. Whereas to reduce the diffuse noise in each subband, Zelinski-TSNR multi-channel postfilter is proposed and applied to the GSC beamforming.

5.3 Proposed GSC Beamforming with Multi-Channel Postfilter

In the proposed multi-microphone or multi-channel speech enhancement method, we consider an adverse environment with directional and diffuse noise, and then it is applied to the linear array of four microphones. Whereas adverse environment is created using Mex function as mentioned in fixed beamformer at section 3.3. 1, in chapter 3. The signal received at each microphone contains directional noise from a particular direction, a diffuse noise that propagates uniformly in all directions, and the desired speech simultaneously. The proposed multi-microphone array speech enhancement is shown in Figure 5.1.

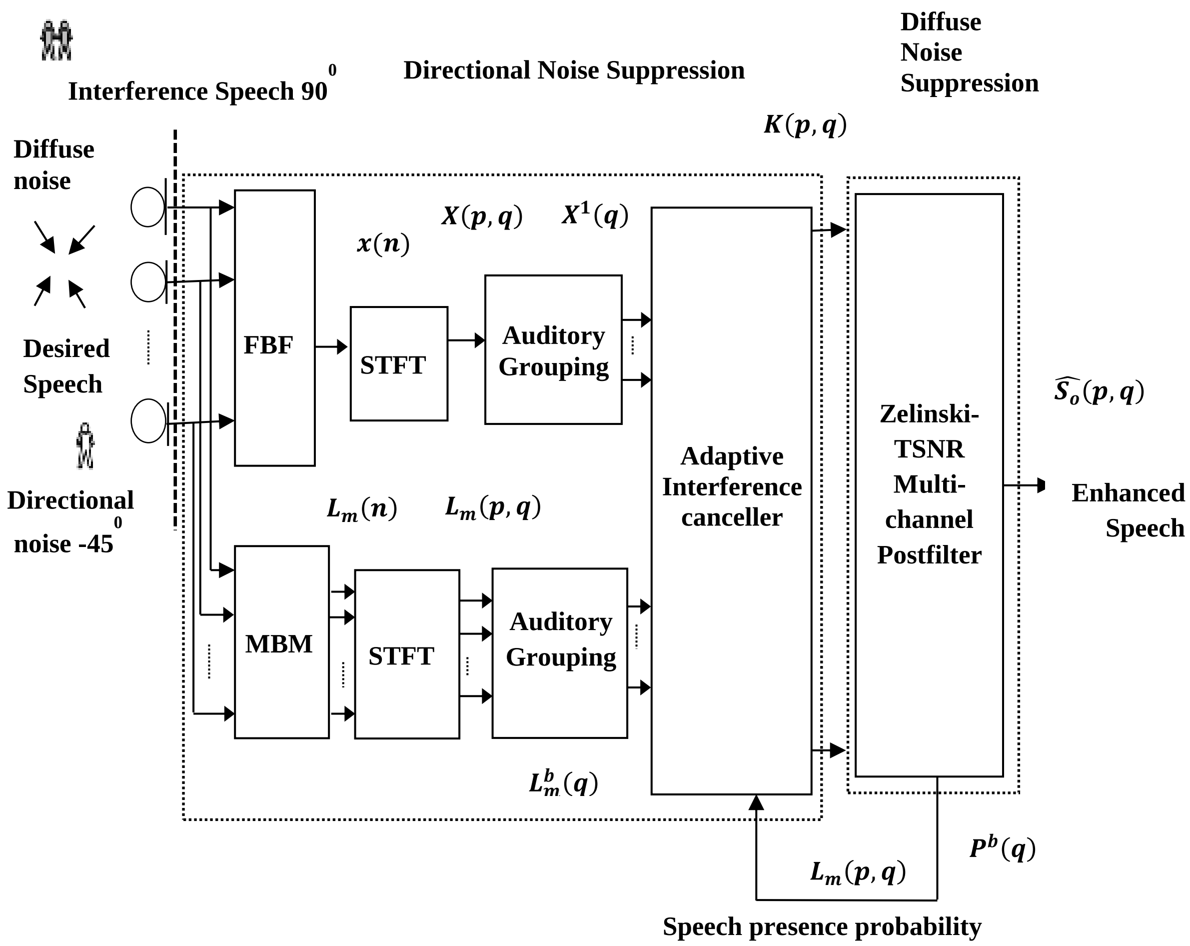

Figure 5.1: Proposed GSC Beamformer Using Zelinski-TSNR Multi-Channel Postfilter

In this chapter, a novel GSC beamforming with Zelinski-TSNR multi-channel postfilter is proposed for speech enhancement. It is a combination of two main blocks, the GSC beamforming with UFNLMS, which reduces the directional noise, and the second part is a Zelinski-TSNR for diffuse noise reduction. The workflow of the proposed method is shown in Figure 5.2. At first, the Fixed Beamforming (FBF) and the Modified Blocking Matrix (MBM) are analyzed in the frequency domain using Short Time Fourier Transform (STFT). A CP is a combination of Improved Zelinski (IZ) and Two-Step Noise Reduction (TSNR) postfilters, where the IZ evaluates the gain of high frequencies and the TSNR evaluates the gain of low frequencies to reduce the diffuse noise. The SPP in each subband is derived using Cohen's multi-channel postfilter and is explained in directional noise and diffuse noise suppression in the coming sections clearly.

5.4 Directional Noise Suppression

In this section, the GSC beamforming using the UFNLMS algorithm is proposed for directional noise (sources coming from known and unknown directions) reduction. As discussed in section 3.3 in chapter 3, GSC structure, there are three main parts: an FBF, an MBM, and an adaptive interference canceller as Unconstrained Frequency domain Normalized Least Mean Square (UFNLMS) algorithm, which is essentially designed to have a better noise reduction from the interfering speech coming from different directions. FBF and MBM are analyzed in sections 3.3.1 and 3.3.2 in chapter 3. Unknown signals coming from various directions are analyzed using a fixed beamformer (DSB). The number of columns in the matrix indicates the number of microphones which is considered to be four, and the efficiency of MBM is 3.

Consider FBM and MBM outputs as $x(n)$ and $l_m(n)$ respectively. Applying STFT to segments, the time domain signal is converted into the frequency domain to obtain $X(p,q)$ and $L_m(p,q)$ in Equation (5.1) and Equation (5.2). When the signal is segmented into frames, tracking the signal becomes easy.

$$x(n) \xrightarrow{STFT} X(p,q) \tag{5.1}$$

$$l(n) \xrightarrow{STFT} L_m(p,q) \tag{5.2}$$

Where p is the temporal frame index, q is the frequency bin, and m = 1, 2, · · ·M – 1 are the number of microphones. Auditory grouping is then applied to $X(p,q)$ and $L_m(p,q)$; it regroups all the frequencies into bark frequency components based on the bark scale, which works on human auditory frequencies, and from this, we separate the total signal into low and high-frequency components. In b^{th} the group, the vector of bins is represented as $X^b(q)$ and $L_m^b(q)$. According to Widrow, classical adaptive noise cancellation [22] and unconstrained minimization is considered and is expressed as

$$\epsilon_m^b(q) = E[X^b(q) - w_m^b(q)L_m^b(q)^2] \tag{5.3}$$

Where $\epsilon_m^b(q)$ is the b^{th} band energy. $E[.]$ and $(.)^H$ are the expectation operator and Hermitian transpose operator, respectively, which can be minimized as follows and $x(n)$

$$w_m^b(q)_{opt} = \frac{\phi_{L_m}^b X(q)}{\phi_{L_m}^b L_m(q)}, \qquad if \ \frac{e_m^b(q)}{dw_m^b(q)} = 0 \tag{5.4}$$

Where

$$\phi_{l_m}^b X(q) = E(L_m^b(q)(X^b(q))^H) \tag{5.5}$$

$$\phi_{l_m}^b l_m(q) = E(L_m^b(q)(L_m^b(q))^H) \tag{5.6}$$

In the GSC beamforming structure, the third part is the adaptive interference canceller: In this chapter, the UFNLMS algorithm is used with different norm constraints to update the weight coefficient based on each subband in SPP which is explained below.

UFNLMS weight update equation is given as

$$w_m^b(q+1) = w_m^b(q) + \mu \frac{L_m^b(q)(X^b(q))^H}{P_{est}^b} \tag{5.8}$$

$$P_{est}^b(q) = \alpha P_{est}^b(q-1) + (1-\alpha \textstyle\sum_m^{M-1} \left|Y_m^b\right|^2 \tag{5.9}$$

Adaptive interference cancellers, i.e., UFNLMS algorithm weight coefficients, are updated based on SPP $P_{est}^b(q)$, which uses the power of the noise reference signal. When we update weight $w_m^b(q)$, there is a chance of signal cancellation in the speech presence region. So we use frequency domain representation of input sensor signal $Y_m^b(q)$. Due to this, the weight update becomes small, which improves the quality during speech presence. As illustrated in the next section, the speech presence probability is calculated using a postfilter, and SPP is passed back to the adaptive interference canceller to update the UFNLMS algorithm.

In general, for any speech signal, a huge amount of speech samples exists in the low-frequency region. It's critical to use non-uniform filters to make the low-frequency bands narrower and the high-frequency bands wider in order to improve degraded speech. This enables the adaptive interference canceller to converge smoothly. In this chapter, high convergence is achieved because of the effective auditory subband method, which utilizes the speech presence frames in each subband and leaves speech absence frames.

In a practical scenario, a lot of speech leakage issues are observed. When the speaker is in motion, speech information is lost in the reference channels in reverberant and echo environments. Desired speech information may be lost if the frequency response of the microphone position is not clear. While updating the adaptive interference canceller, some speech information is canceled, and such errors are minimized by $\epsilon_m^b(\mathrm{q})$ in Equation (5.3). Speech presence in each subband is considered by omitting speech absence frames while updating UFNLMS weights to solve speech leaking difficulties. In subband adaptive interference canceller, updating rate in Equation (5.8) is slow for speech presence frames. The steady-state error and convergence are decided by the step size of an adaptive filter. The time-varying step-size for the UFNLMS algorithm, which is feedback by SSP in each subband 12, is given below:

$$P^b(q) = (1 - \frac{1}{M_b} \Sigma_{i=b_1 b_2} P^i(q))\mu \tag{5.10}$$

Where M_b is the number of frequency bins within the b^{th} subband,

$$0 < p^b(q) < 1$$

$$P^b(q) = \frac{1}{M_b} \sum_{i=b_1 b_2} P^i(q) \tag{5.11}$$

$$\mu^b(q) = (1 - p^b(q))\mu \tag{5.12}$$

Where $p^b(q)$ is SPP of combined postfilter and the range of SPP is $0 < p^b(q) < 1$. $p^b(q)$ is the presence of desired speech in b^{th} a subband of b^{th} the frame. A large $\mu^b(q)$ Equation (5.13) gives a slow update rate of the UFNLMS, which protects the speech components. In fast-changing environments, the update rate should be faster to update the UFNLMS algorithm.

$$= (1 - \frac{1}{M_b} \sum_{i=b_1 b_2} P^i(q))\mu \tag{5.13}$$

5.5 Diffuse Noise Reduction Using Zelinski-TSNR Multi-Channel Postfilter

Zelinski and Mc Cowan [28], [189] introduced diffuse noise reduction postfilters where noise is partially reduced considering spectral constraints. Here, a novel postfilter is implemented, which is a combination of Improved Zelinski- Two-Step Noise Reduction (IZ–TSNR) [244]-[245] postfilter and is named as combined postfilter for diffuse noise reduction. According to human auditory frequencies, speech samples remain more at low frequencies. Here the CP is implemented in two steps: first, low frequencies are analyzed by TSNR postfilter [245], and then high frequencies are analyzed by IZ postfilter. IZ postfilter is implemented by calculating the cross-spectral density of GSC output which serves as input to the CP. For high frequencies, to minimize the mean square error between speech and its estimate, the IZ postfilter is designed. In the IZ postfilter, transient frequencies are analyzed by following the microphone array geometry. The auto and cross-spectral densities of desired and noisy speech for high frequencies are defined as

$$\bar{\phi}_{x_i x_i}(p,q) = K(p,q) + L_m(p,q) \tag{5.14}$$

$$\bar{\phi}_{x_i x_j}(p,q) = K(p,q) \tag{5.15}$$

In IZ postfilter [245], the gain function can be analyzed as

$$G_{IZ}(p,q) = \frac{\frac{1}{\left|\Omega_{m(p)}\right|}\sum_{\{i,j\}\in\Omega_{m(p)}} \Re\{\bar{\phi}_{x_i x_{j(p,q)}}(p,q)\}}{\frac{1}{\left|\Omega_{m(p)}\right|}\sum_{\{i,j\}\in\Omega_{m(p)}} [\frac{1}{2}\bar{\phi}_{x_i x_{j(p,q)}}(p,q) + \bar{\phi}_{x_i x_{j(p,q)}}(p,q)]} \tag{5.16}$$

Two-Step Noise Reduction (TSNR) filter is applied for low frequencies to reduce the noise and improve the intelligibility of the desired speech signal. This filter is implemented in two steps; in the first step using the DD algorithm, the spectral gain $G_{DD}(p,q)$ *i*s analyzed. In the second step, the spectral gain of the next frame is calculated and applied to the current frame.

Here *a priori* and *a posteriori* SNR are evaluated to determine the spectral gain of the DD approach using equations given below,

$$\widehat{snr}_{post}(p,q)=\frac{|K(p,q)|^2}{(\widehat{L}m(p,q))} \tag{5.17}$$

$$\widehat{snr}_{prio}^{DD}(p,q)=\frac{\beta(\hat{s}(p-1),q)^2}{(\widehat{L}m(p,q))}+(1-\beta)P(\widehat{snr}_{post}(p,q)-1) \tag{5.18}$$

$$G_{DD}(p,q)=\frac{\widehat{snr}_{prio}^{DD}(p,q)}{1+\widehat{snr}_{prio}^{DD}(p,q)} \tag{5.19}$$

In the TSNR filter, the second step is to calculate the *a priori SNR* based on the DD approach of the first step

$$\widehat{snr}_{prio}^{TSNR}(p,q)=\widehat{snr}_{prio}^{DD}(p+1,q) \tag{5.20}$$

$$=\frac{\beta'|G_{DD}(p,q)k(p,q)|^2}{(\hat{L}_m(p,q))+(1-\beta^1)P[\widehat{snr}_{post}(P+1,q)-1]} \tag{5.21}$$

If $\beta'=1$, the diffuse noise is degraded by the DD approach, and Equation (5.16) is modified as

$$\widehat{snr}_{prio}^{TSNR}(p,q)=\frac{\beta'|G_{DD}K(p,q)|^2}{(\hat{L}_m(p,q))} \tag{5.22}$$

The spectral gain for low frequencies is defined as

$$G_{TSNR}(p,q)=h(\widehat{snr}_{prio}^{TSNR}(p,q),\widehat{snr}_{post}^{prio}(p,q)) \tag{5.23}$$

In low frequencies, the noisy speech is enhanced, based on TSNR gain and reference estimate

$$\hat{S}(p,q) = G_{TSNR}(p,q)K(p,q) \tag{5.24}$$

Finally, TSNR spectral gain is determined as

$$G_{TSNR}(p,q) = \frac{\widehat{snr}_{prio}^{TSNR}(p,q)}{1+\widehat{snr}_{prio}^{TSNR}(p,q)} \tag{5.25}$$

The output of the combined postfilter can be expressed as

$$G(p,q) = G_{TSNR}(p,q) + G_{IZ}(p,q) \tag{5.26}$$

Finally, to estimate the diffuse noise in the multi-microphone array, the CP and reference signal estimate is analyzed as follows

$$\hat{S}_o = G(p,q)K(p,q) \tag{5.27}$$

Using SPP, the filter coefficients are updated in each subband as mentioned in the above equations. The time-frequency units of each subband are averaged to estimate constrained filter updates.

The workflow of proposed GSC-Zelinski-TSNR multi-channel speech enhancement (MCSE) is shown in, where the directional noise coming from

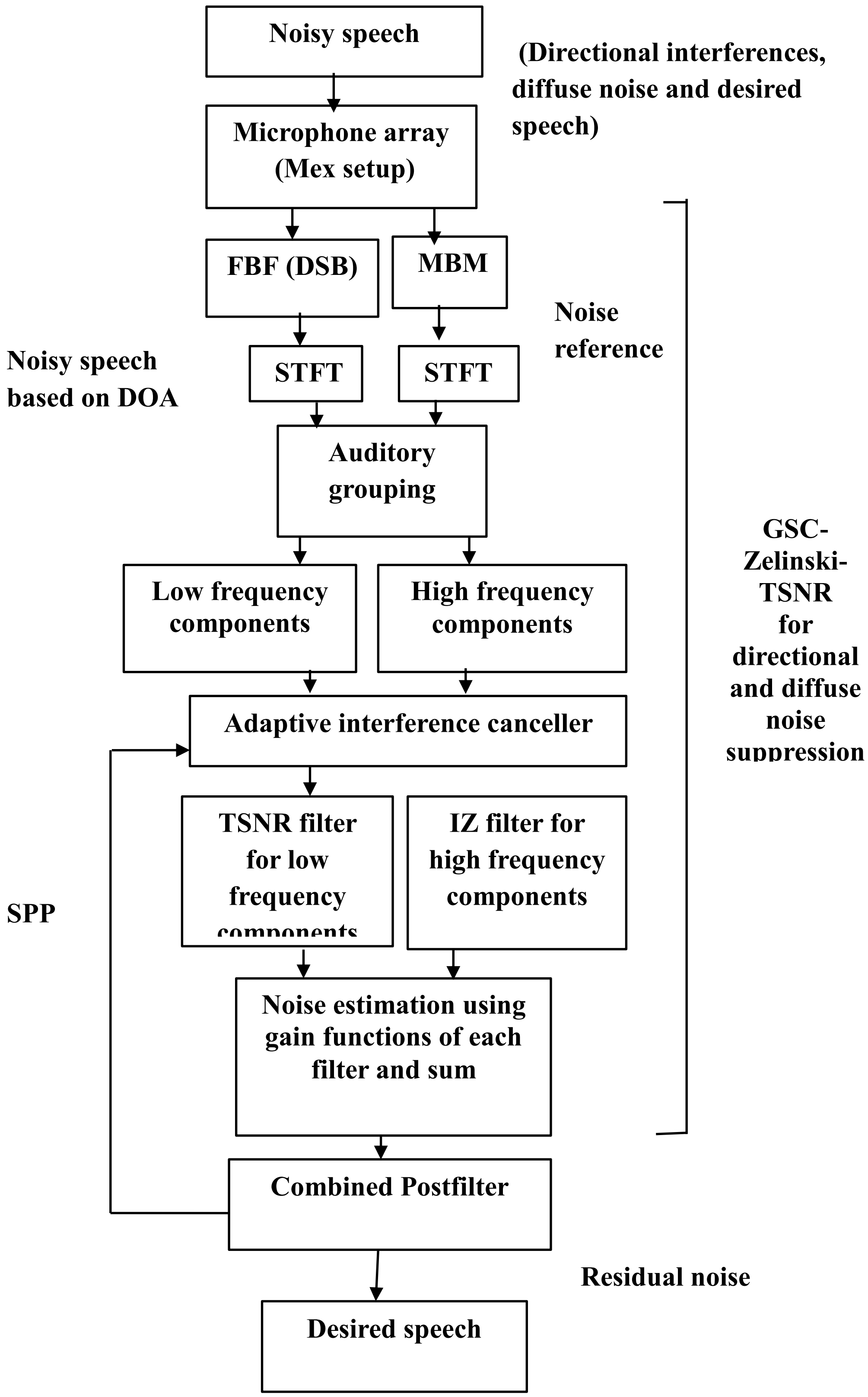

Figure 5.2: Workflow of Proposed GSC Beamforming with Zelinski-TSNR Multi-channel Postfilter

5.6 Results and Discussions

In this section, simulation of the proposed GSC beamforming with Zelinski-TSNR multi-channel postfilter in an adverse environment is evaluated and discussed. The simulation parameter with specifications considered for the proposed GSC-CP-SNMF method is shown in Table 5.2. Image method [225] is applied to generate multi-channel room impulse response. Where, a linear array of four microphones with a distance of 5 cm between each microphone, and distance of 1 m between the source and microphone array, in a conference room with 6 m x 3 m x 5 m and reverberation time of 300 ms following a Mex setup using Mex function, i.e., rir-generator.cpp [226] in MATLAB. The Mex function was taken from International Audio Laboratories Erlangen at Friedrich Alexander University Erlangen-Nuremberg. (https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator).

An adverse environment is considered taking the desired speech from an unknown direction, directional interferences like white noise at -45^o and female speech at 90^o using the DARPA TIMIT [227]-[228], i.e., database of 6300 male and female sentences, 10 sentences spoken by every 630 speakers with a sampling frequency of 8 kHz and also a diffuse noise [190] from NOIZEUS database [229]-[230], i.e., a car noise are given to Mex setup, finally forms a noisy input signal (i.e., a combination of desired speech, directional interferences, and diffuse noise) with different SNR levels from -10 dB to 10 dB.

This noisy input signal is applied to FBF, which analyzes the DOA of (known and also unknown) input signal, and all the delays from the microphones are added based on the DSB principle. Finally, a partially enhanced signal is obtained at the output of FBF. In the next stage, from noisy input speech, the delays on the adjacent microphones are subtracted using MBM so that a noise reference is generated. Both the partially enhanced signal and the noise reference are parallelly applied to the auditory grouping.

Later, directional noise is suppressed by GSC using UFNLMS, and diffuse noise is suppressed by Improved Zelinski-TSNR multi-channel postfilter in each subband. Noisy multi-channel speech is processed with the existing methods to show the performance comparison of the proposed GSC with Improved Zelinski-TSNR multi-channel postfilter.

1. GSC algorithm in the time domain (GSC-TD) [21].
2. GSC algorithm in the frequency domain (GSC-FD) [241].
3. GSC-FD with modified blocking matrix (GSC-FD*) [246].
4. GSC-FD* with Subband-Feedback-Controlled Adaptive Filter (GSC-FD*-SFC) [247].
5. Cohen's algorithm [232].
6. Kai Li Method [248].
7. Proposed GSC-Zelinski-TSNR

Table 5.1 Simulation Parameters Considered for the Proposed GSC-Zelinski-TSNR

Parameters	**Specifications**
Conference Room Dimensions	6 m X 5 m X 3 m (Using Image Method [225] with Mex setup using Mex function i.e., rir-generator.cpp [226] in MATLAB)
Microphones (m)	m=4
Distance between each microphone	5 cm
Distance from source to microphone	1 m
Diffuse noise	Car noise from Noizeus [230]
Input SNR Levels	-10 dB, -5 dB, 0 dB, 5 dB and 10 dB
Database	Darpa Timit [227]-[228] and Noizeus [229]-[230]
Directional Interferences	white noise at -45 and female speech at 90
Desired speech and diffuse noise	unknown direction

5.6.1 Performance Evaluation of the Proposed Zelinski-TSNR

The performance of the above-mentioned algorithms is shown in terms of three objective parameters, namely, perceptual evaluation of speech quality [231] (PESQ), second segmental SNR (SSNR) [240], and log spectral distance (LSD) [232].

5.6.1.1 Comparison of PESQ for Proposed GSC- Zelinski-TSNR

PESQ [231] is an objective intelligibility measure, the standard range of which lies between 0.5 to 4.5 dB, as explained in chapter 3, section 3.4.1.1. The higher the PESQ Score better will be the perception. Table 5.2 compares the PESQ score for the proposed method with existing algorithms, respectively. In Kai Li Method [248], the PESQ at -10 dB attains 2.38 dB whereas for the proposed GCS-Zelinski-TSNR is 2.45. The proposed method shows superior performance, as SPP is used for minimizing the noise power instead of step size in diffuse noise reduction. For the proposed method, a PESQ of 3.42 using four microphones is attained at the Zelinski-TSNR output at 10 dB input SNR using four microphones.

5.6.1.2 Comparison of SSNR for Proposed GSC- Zelinski-TSNR

Segmental SNR [240] is one of the most popular objective measures for speech enhancement methods. In normal SNR calculation, the whole signal is considered, whereas while calculating SSNR, segments are taken with 256 samples per frame (k=256, with 50 percent overlap). Higher the SSNR will be better the quality of speech.

Segmental SNR is calculated as

$$SSNR = \frac{10}{L} \frac{\sum_{q=0}^{L-1} 10log \sum_{m=0}^{M-1} X^2(m+q\frac{m}{2})}{\sum_{m=o}^{M-1}\left[X\left(m+q\frac{m}{2}\right)-\hat{S}_o(m+q\frac{m}{2}\right]^2} \tag{5.26}$$

In the Zelinski-TSNR multi-channel postfilter, the low-frequency region, which is below 4kHz, is processed with a TSNR filter, and the high-frequency region, which is in the range 4kHz - 8kHz, is processed with an IZ filter. The diffuse noise in each subband is eliminated by considering speech presence segments that result in improved performance of SSNR.

Table 5.2: PESQ Comparison for Proposed GSC-Zelinski-TSNR

Input SNR(dB)	**-10**	**-5**	**0**	**5**	**10**
GSC-TD[21]	2.01	2.06	2.10	2.18	2.21
GSC-FD[241]	2.03	2.21	2.40	2.58	2.62
GSC-FD*[246]	2.04	2.24	2.52	2.71	2.91
GSC-FD*-SFC[247]	2.09	2.26	2.44	2.62	3.06
Cohen Method[232]	2.13	2.32	2.51	2.83	3.15
Kai Li Method[248]	2.38	2.62	2.83	3.03	3.28
GSC-Zelinski-TSNR (Proposed)	**2.45**	**2.68**	**2.94**	**3.20**	**3.42**

Table 5.3: SSNR Comparison for Proposed GSC-Zelinski-TSNR

Input SNR(dB)	**-10**	**-5**	**0**	**5**	**10**
GSC-TD[21]	2.1	3.9	5.8	7.2	13.2
GSC-FD[241]	3.2	4.3	6.6	8.5	14.1
GSC-FD*[246]	4.1	5.4	7.3	9.1	16.6
GSC-FD*-SFC[247]	6.2	8.4	9.8	11.4	18.9
Cohen Method[232]	8.8	12.6	16.0	18.7	19.1
Kai Li Method[248]	9.2	14.1	16.5	19.8	22.4
GSC-Zelinski-TSNR (Proposed)	**9.8**	**15.3**	**18.7**	**20.2**	**26.8**

At 10 dB input SNR under adverse environment, SSNR for the proposed GSC-Zelinski-TSNR is 26.8 whereas Kai Li method [248] it is 22.4 dB. Similarly at -10 dB input SNR, SSNR for proposed method is 9.8 dB whereas for existing Kai Li method [248] and Cohen Method [232] it is 9.2 and 8.8, the proposed method outperforms the existing methods in suppressing diffuse noise in each segment.

5.6.1.3 Comparison of LSD for Proposed GSC-Zelinski-TSNR

Log Spectral Distance measure [232] is an objective measure for the calculation of the spectral distance between the frames. Better intelligibility can be achieved with a reduction in spectral distance.

LSD is calculated as

$$LSD = \frac{10}{L}\sum_{(q=0)}^{L-1}\left\{1+\left(\frac{M}{2}+1\right)\sum_{q=0}^{\frac{M}{2}}\left[logX(p,q) - log\hat{S}_o(p,q)\right]^2\right\} \tag{5.27}$$

In Table 5.4, the LSD measure for the proposed method is compared with the competing methods.

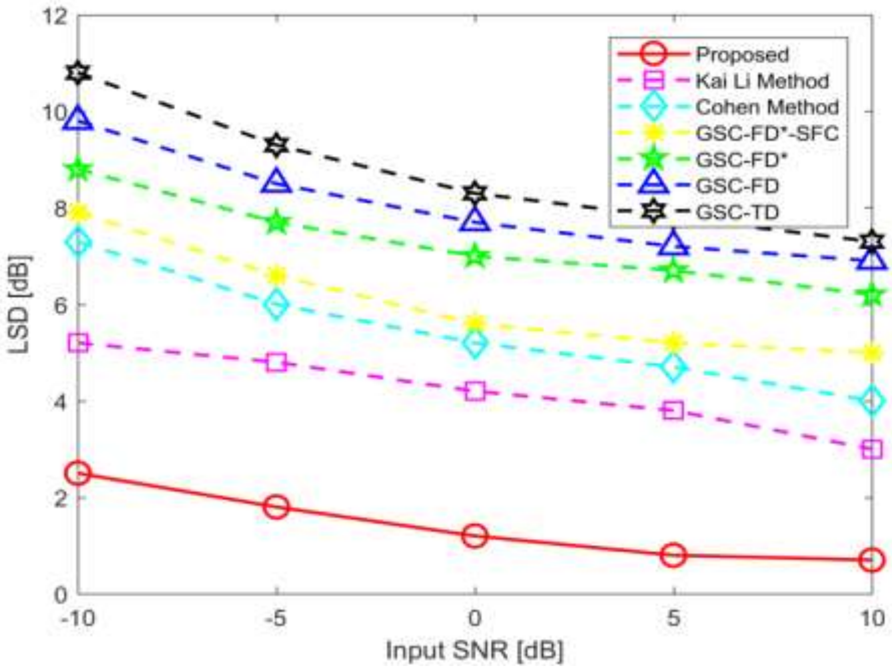

Figure 5.3: LSD Comparison of Proposed GSC-Zelinski TSNR with Existing Methods

The proposed method shows the lower LSD, resulting in better noise reduction. The spatial information is completely utilized by taking MBM into consideration. The spectral distance between two frames is reduced, and the proposed GSC-Zelinski-TSNR achieves better performance compared to the other classical methods. As the distance between the frames decreases, the distortion gets reduced. LSD for the proposed GSC- Zelinski-TSNR at -10 dB is 2.6 dB, whereas, for the K. Li [248] method, it is 11.2 dB for four microphones, which shows that the proposed method GSC- Zelinski-TSNR has the lower LSD compared to an existing method and so on.

5.6.1.4 Spectrograms

In Figure 5.4, the spectrograms of the proposed GSC-Zelinski TSNR multi-channel postfilter at 10 dB input SNR, where the noise reduction of noisy input speech using proposed GSC-Zelinski-TSNR for four microphones is shown.

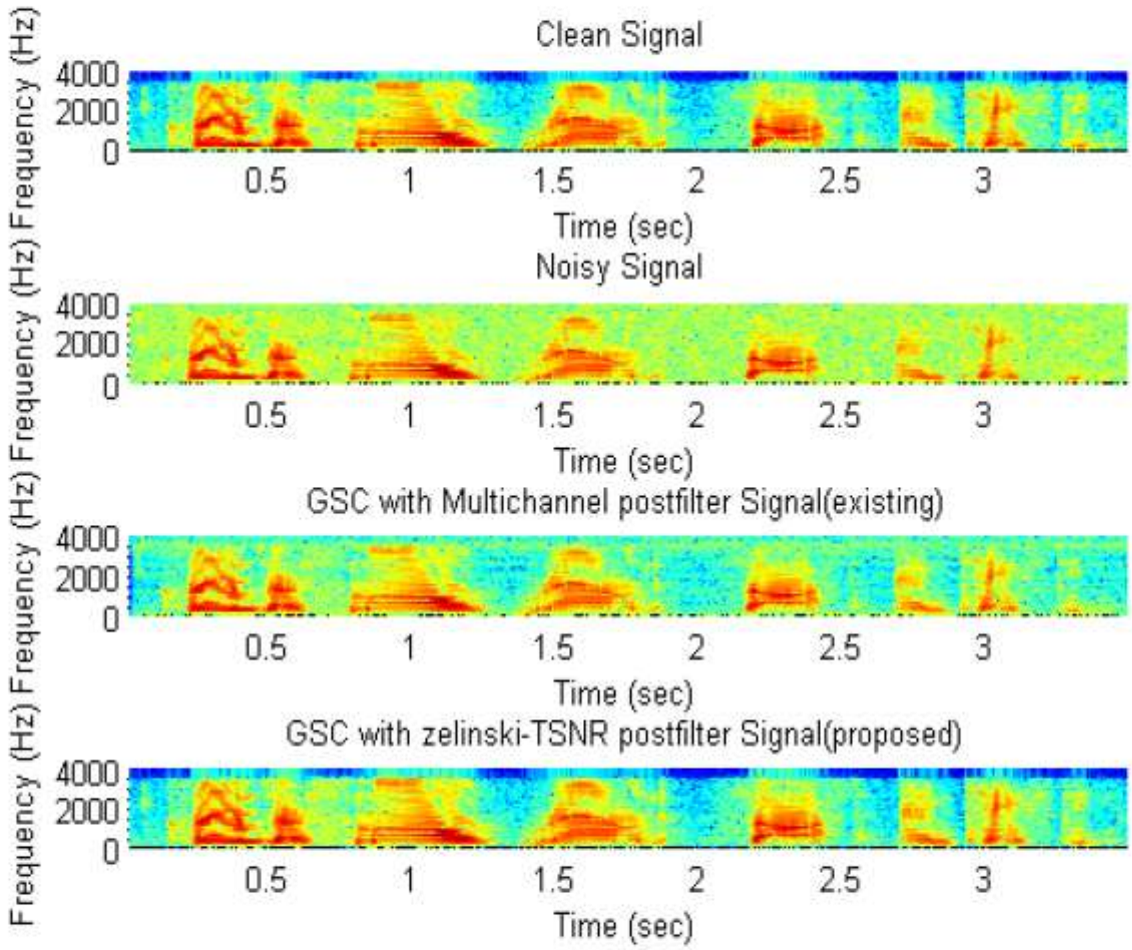

Figure 5.4: Spectrogram of Proposed GSC- Zelinski TSNR Multi-Channel Postfilter

The proposed GSC Zelinski-TSNR method shows better noise reduction with improved quality and intelligibility in each subband. SPP eliminates speech absence frames using adaptive interference canceller.

5.7 Summary

Multi-microphone adaptive beamforming, i.e., GSC beamforming with Zelinski-TSNR multi-channel postfilter, is proposed to enhance the degraded speech in directional and diffuse noise conditions. In this chapter, directional interference is eliminated by using GSC with auditory interference canceller with subband feedback control using speech presence probability. By using UFCNLMS, the iteration speed gets increases compared to time-domain methods. Diffuse noise is suppressed using Zelinski-TSNR multi-channel postfilter in which speech signals at low-frequency regions are enhanced by using TSNR, and high frequencies are enhanced by improved Zelinski postfilter. At 10 dB input SNR, PESQ and SSNR for the proposed GSC-Zelinski-TSNR is 3.42 and 26.8 dB, shows that the diffuse noise suppression in each subband. Degraded speech in the low-frequency region is enhanced completely, which made the proposed algorithm show better performance compared to the existing algorithms in terms of PESQ, Segmental SNR, and LSD. After processing of noisy speech using proposed GSC beamformer using Zelinski- TSNR multi-channel postfilter which is later called as Combined Postfilter (CP) produces a system-generated noise in the desired speech, which can be called residual noise. To eliminate residual noise and also to separate interference speakers from unknown directions in real-time environments, a novel adaptive beamforming with multi-channel postfilter is implemented and discussed in chapter 6.

Chapter 6

Adaptive Beamforming Using Combined Postfilter and Sparse NMF for Speech Enhancement

This chapter proposes the sparse NMF to GSC beamformer with combined postfilter to suppress residual noise. Sparse NMF is proposed to reduce residual noise generated at the output of GSC with a combined postfilter for multi-channel speech enhancement.

6.1 Motivation

Residual noise is a major problem in multi-channel speech enhancement. In addition to the requirement for minimal distortion of the original speech, which was discussed in Chapter 3, it is important that the residual noise, i.e., the noise remaining after the enhancement process, does not sound annoying. Therefore, there is a great need to reduce residual noise, to reduce listener fatigue, and improve intelligibility. Existing multi-channel speech enhancement (MCSE) also suffers from residual noise in the output, and that reduces the quality and intelligibility of the desired signal. Multi-channel speech enhancement techniques (MCSE) such as adaptive beamforming with postfiltering enables high-quality, hands-free communication in noisy conditions. But there exists residual noise in the desired speech. So, a robust multi-channel speech enhancement algorithm should be developed to suppress residual noise and also to separate interference speakers coming from an unknown direction.

6.2 Introduction

Residual noise occurs due to the existence of randomly spaced peaks in the spectrum of the reconstructed signal because of the overestimates and underestimates of the clean signal in adjacent spectral groupings. Sometimes resulting from the crude estimation of the noisy signal power spectrum. These peaks sound similar to tones with frequencies that change randomly at the analysis frame rate. Residual noise is more prominent in the unvoiced segments of speech where the noise power is comparable to the speech power and is sometimes more disturbing than the original distortions caused by the interfering noise, which is challenging for multi-channel speech enhancement (MSCE).

Interference noise or speaker separation refers to the problem of separating one or more desired signals from mixtures of multiple signals. This problem can be encountered in many different applications, such as medical [249] [250], military [251], and multimedia [252]. This challenge is commonly approached by using numerous sensors, each of which monitors a different mixture of a source signal to acquire enough information about the incoming source signals to perform the effective separation. In most cases, the source signals are assumed to be statistically independent, and no extra prior information about the source signals is assumed available.

The more complicated problem is that of separating multiple source signals from an unknown direction. This problem is usually defined as the multi-channel speech enhancement and interference separation problem. The goal of multi-channel speech enhancement and interference separation is to recover the original source signals from a multi-microphone recording of their linear mixture, as shown in Figure 6.1. Since the problem is underspecified, prior knowledge or training data for the source signals are assumed to be available. In this thesis, the multi-channel source separation and enhancement problems are considered for the adverse environment. The adverse environment is considered directional noise, diffuse noise, real-time noises, and interference speakers from different directions. The multi-channel speech enhancement and interference separation problem is encountered in many applications such as: separating instruments in music recordings [253], separating speech signals from multiple simultaneous speakers recording [200], [254], separating speech signals from background music signals [255], speech denoising [97], and improving automatic speech recognition systems by removing the background signals [256].

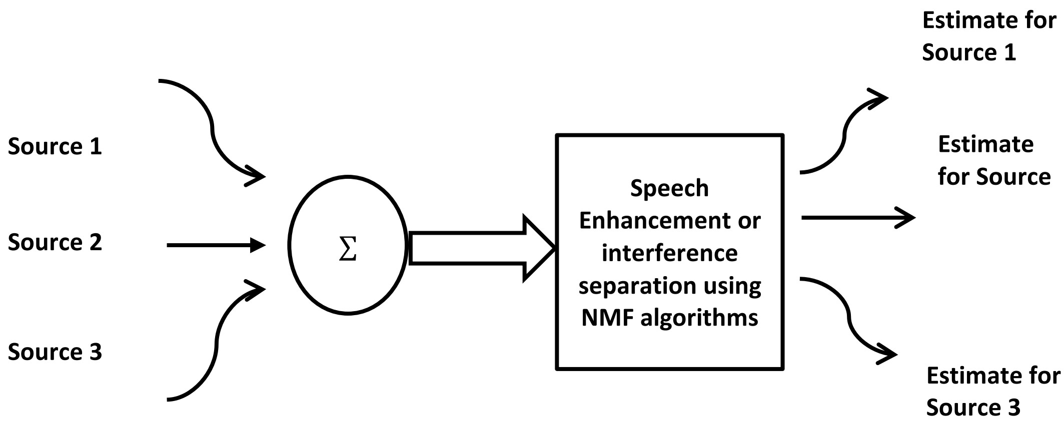

Figure 6.1 Multi-Channel Speech Enhancement Using NMF.

6.2.1 Multi-Channel Speech Enhancement and Interference Separation Using NMF Algorithms

There are many proposed approaches to estimate the desired speech from the observed noisy speech signal from an adverse environment. Most of these approaches rely on training data about the input signals that are in the mixture. In many approaches, the training and the mixed signals are usually processed in magnitude or power spectral domain [257]-[258]. In other approaches, the signals are processed in the log-spectral domain [259].

Another approach for multi-channel speech enhancement or interference separation is to decompose the mixed-signal spectral frames as a weighted linear combination of the training data spectral frames. In [260]-[261], the mixed signal is decomposed as a linear combination of a number of exemplars from a large exemplar dictionary of training data for each source signal.

The most used approach for solving the MCSE and interference separation problem is nonnegative matrix factorization (NMF) [38] to train a set of nonnegative basis vectors (dictionary) for the training data of each source. In the separation stage, NMF is used to decompose the mixed signal as a weighted linear combination of the trained basis vectors. The estimate of each source is found by summing its corresponding trained basis terms from the NMF decomposition during the separation stage [262]. The NMF is used in this framework in the magnitude spectral or power spectral domain where the non-negativity constraint is necessary. The number of the trained basis vectors is usually less than the dimension of the spectral frames of the training data. Due to the efficient update rule solutions of NMF [38], and since every source is represented by a few numbers of basis vectors, this approach is considered to be fast and very simple, which makes it the most used approach in multi-channel speech enhancement and interference separation. Another advantage of using NMF in multi-channel speech enhancement and interference separation is that there is no limitation on the energy level for the training and mixed signals.

Much research has been done to improve the performance of NMF by encouraging the NMF decomposition matrices to satisfy specific features of the source signals to be evaluated. In [263], harmonicity and smoothness were enforced in Bayesian NMF and applied to music transcription. In [264], spatial decorrelation and other priors were incorporated with NMF for different applications. In [198], regularized NMF with Itakura-Saito (IS-NMF) divergence was introduced with Markov chain prior models for smoothness within a Bayesian framework. The conjugate prior distributions on the NMF weights and basis matrices with the Poisson observation model within the Bayesian framework were introduced in [265]. In [266], the discriminative constraint was applied to the NMF solution. When NMF algorithms are used for source separation, a good separation can be expected only when speaker-dependent basis matrices are learned. In contrast, for noise reduction, even if a general speaker-independent basis matrix of speech is learned, a good enhancement can be achieved [98]. Since the basic NMF allows a large degree of freedom, the performance of the source separation algorithms can be improved by imposing extra constraints and regularizations, motivated by the sparsity of the basis vectors and NMF coefficients or smoothness of the NMF coefficients. In probabilistic NMFs, these constraints can be applied in the form of prior distributions. Among different priors, significant attention has been paid to model the temporal dependencies in the signals because this important aspect of audio signals is ignored in a basic NMF approach [205], [267].

Schmidt et al. [269] presented an NMF-based unsupervised batch algorithm for noise reduction. In this approach, it is assumed that the entire noisy signal is observed, then the noise basis vectors are learned during the speech pauses. In the intervals of speech activity, the noise basis matrix is kept fixed, and the rest of the parameters (including speech basis and speech and noise NMF coefficients) are learned by minimizing the Euclidean distance with an additional regularization term to impose sparsity on the NMF coefficients. The reported results show that this method outperforms a spectral subtraction algorithm, especially for highly non-stationary noises. The schematic view of enhancing degraded speech using NMF is shown in Figure 6.2.

In [270], a supervised NMF-based denoising scheme is proposed in which a heuristic regularization term is added to the cost function. By doing so, the factorization is enforced to follow the pre-obtained statistics. In this method, the basis matrices of speech and noise are learned from training data offline. Also, as part of the training, the mean and covariance of the log of the NMF coefficients are computed. The negative probability of a Gaussian distribution is used to regularize the cost function during the enhancement using these statistics (with the computed mean and covariance).

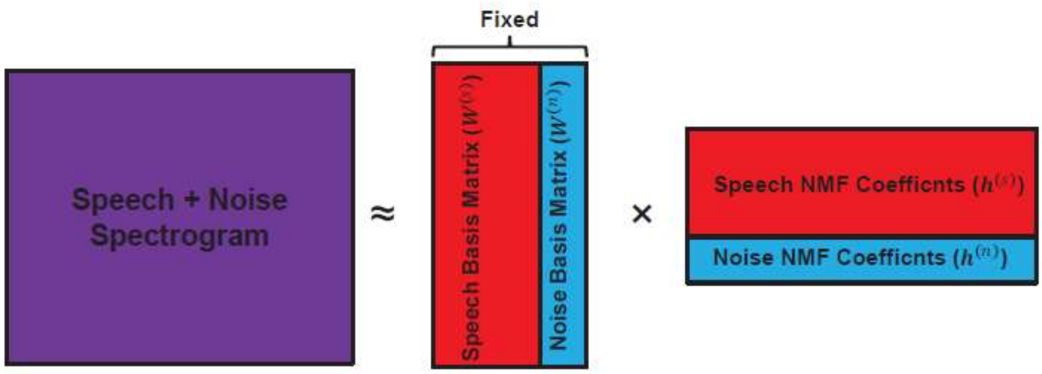

Figure 6.2: Schematic View of NMF

To the multi-channel speech enhancement and multiple interference separation environments, S Gannot [194] introduced the transfer function GSC with a multi-channel postfilter and compared it with a single channel postfilter. This method fails in diffuse noise fields. An improved GSC with multi-channel postfiltering is presented by K. Li [248], which eliminates the directional noise but is unable to suppress diffuse noise and has caused more speech distortion in low frequencies. Li. Pfeifenberger [271] introduced GSC adaptive beamforming with directional to diffuse noise postfilter, it separates the directional and diffuse noise components, but this method fails in case of interferences. X. Wang [272] proposed a modified SPP-based multi-channel postfilter for reverberant noise; The intended speech is obtained using this procedure, although it is accompanied by residual noise. J. Park [273] proposed a GSC beamforming using Wiener postfilter for composite noise suppression. But, when diffuse noise is considered, it does not give the desired result due to the employment of a single-channel Wiener postfilter at the output of GSC. In the method proposed by L. Zhang [274], post-secondary filtering is introduced to a time-domain GSC beamforming to reduce the diffuse noise, point noise, and speech interferences.

In enhancing or separating the speech signal from interfering noisy source, machine learning methods like NMF plays a significant role. G. Rithwik [275] introduced a speaker-independent speech enhancement in which NMF based postfilter is used to reduce the noise. S. Priyanka [276] presented GSC adaptive beamforming using Zelinski TSNR postfilter, but when the number of interferences increased, it was unable to separate the desired speech. S Gannot [19] developed a consolidated perspective on multi-microphone speech enhancement and source separation methods which are interrelated to each other. And also address that the performance of the system depends on the number of microphones, which improves speech communication in noisy and reverberant environments. P.D.O. Grady [277] presents a convolutive NMF with a sparse constraint to represent speech phones in auditory data. C. Fevotte [278] introduced NMF with sparse constraints for single-channel audio source separation. But these algorithms individually help in either separating interference or reducing the noise.

During the simulation of the GSC-CP algorithm for the adverse environment, residual noise is generated by the system. In the real-time environment, the number of interferences will be more it will lead to the overlapping of the sources. This makes it very critical while communicating in teleconference applications. An innovative strategy should be implemented to overcome these hurdles in a challenging environment, such as minimizing residual noise and separating unknown interferences.

A GSC beamforming approaches to suppress directional and diffuse noise is implemented both in a time domain and frequency domain under noisy real-time environments in the previous chapter 3, chapter 4, and chapter 5. In Chapter 5, a combination postfilter (CP) is proposed and applied to the GSC beamforming to reduce diffuse noise in each subband. In this chapter, to reduce residual noise generated at the output of GSC-CP and also to separate unknown interferences, Sparse NMF (SNMF) is proposed. The SNMF is introduced to the proposed GSC-CP from chapter 5 to suppress residual noise and also to separate the interference speakers in the real-time environment.

6.3 Proposed GSC Adaptive Beamforming using CP and Sparse NMF

In the proposed multi-microphone speech enhancement method, we consider an adverse environment with directional and diffuse noise, and then it is applied to the linear array of four and eight microphones. As considered in chapter 5, the same generated noisy input is considered here, where the signal received at each microphone contains directional noise from a particular direction, a diffuse noise that propagates uniformly in all directions, and the desired speech simultaneously. The proposed multi-microphone array speech enhancement is shown in Figure 6.3.

In this chapter, a novel GSC beamforming with CP and SNMF is proposed for speech enhancement. It is a combination of three main blocks as the GSC beamforming with UFNLMS reduces the directional noise, the second part is a combined postfilter (CP) for diffuse noise reduction, which is already implemented in chapter 5, section 5.4, and finally, the third is SNMF which suppress the residual noise generated at the output of CP which is explained in the next section

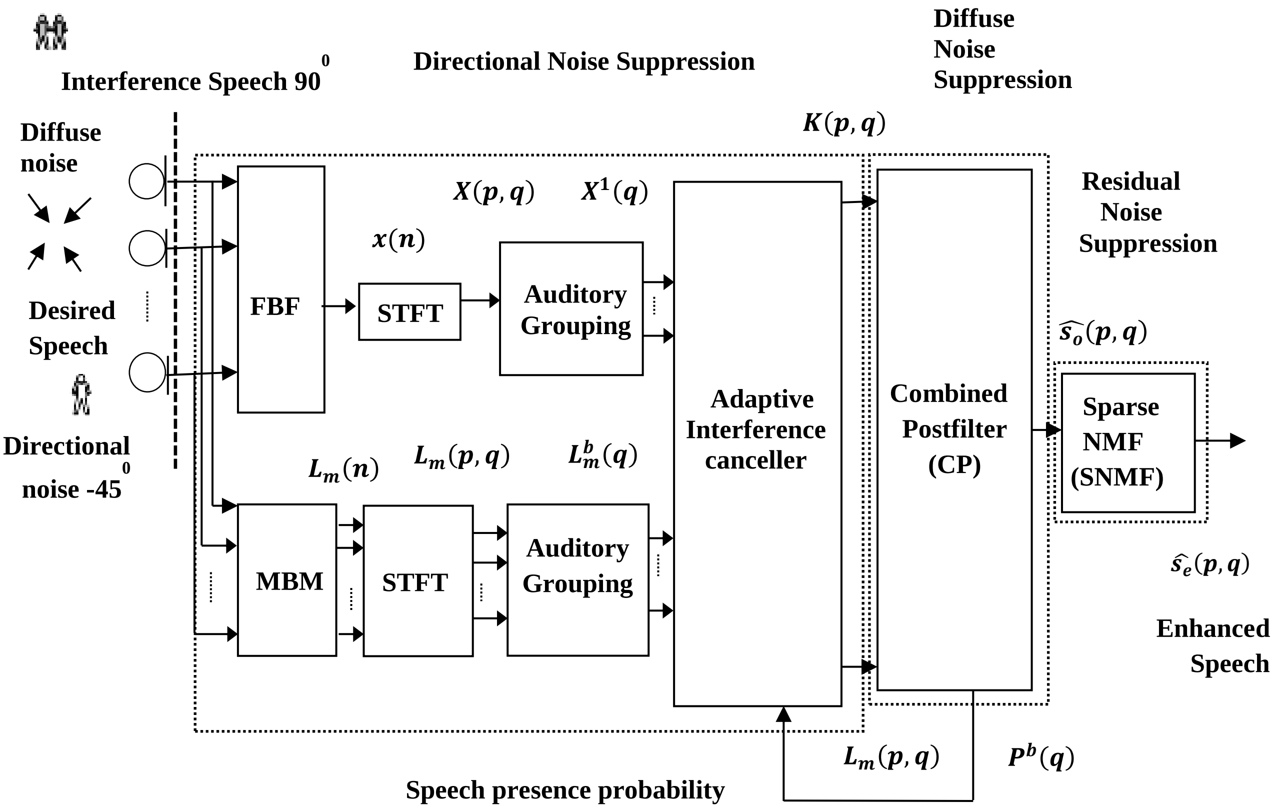

Figure 6.3: The Proposed GSC Beamforming with CP and SNMF for a Multi-Channel Speech Enhancement

6.3.1 Sparse NMF for Residual Noise Suppression

The purpose of sparse linear coding [279]-[280] is to identify a decomposition in which the hidden components are sparse; that is, their probability densities are significantly peaked at zero, and their tails are long. This essentially indicates that every given input vector may be properly represented with only a few non-zero hidden coefficients. NMF's ability to provide a sparse representation of data is one of its most valuable features. This type of representation encodes a large amount of data with a small number of 'active' components, making the encoding simple to understand.

Sparse coding is a representational system in which only a few units (from a huge population) are employed to adequately represent typical data vectors [281]. In practice, this means that most units take values near zero, with just a few taking values that are notably non-zero. The sparsest feasible vector (with just one non-zero component) should have a sparseness of one on a normalized scale, whereas a vector with all components equal should have a sparseness of zero. We employ a sparseness metric based on the connection between the L1 and L2 norms in this chapter which helps in updating the activation function

For desired speech from a known residual noise and interference separation from a noisy speech, we employ the following procedure:

1. Get training data for the residual noise $S_o(t)$ and $S_e(t)$ desired speech, make a magnitude spectrogram for both, then use SNMF to extract associated frames w_t^o and w_t^e.

2. Create a combined basis set $w_t^{oe} = [w_t^o \mid w_t^e]$, which yields a basis twice the size of R.

3. Make a magnitude spectrogram of a mixture made up of two unknown sources. SNMF with fixed to w_t is used to fit the mixture to w_t^{oe} , and only the related activations *h* are learned.

4. Split h into noisy residual speech h^o and desired speech, h^e components that correspond to their corresponding bases, $h = \left[h^o \mid h^e \right]$.

5. Create a magnitude spectrogram for both sources using their respective bases and activations: $S^o = \sum_{t=0}^{T_o-1} w^o h^o$ and $S^e = \sum_{t=0}^{T_o-1} w^e h^e$.

6. Create an audible reconstruction for both sources using the phase information from the mixture.

In this procedure, the residual noise is separated from the GSC-CP output. A related formulation for updating the basis vector and activation function in the SNMF model is as follows.

In SNMF, the noisy speech signal *v* is the linear multiplication of the basis vector and the activation coefficients *w* and *h.* SNMF calculates *w* and *h* by reducing the sparseness of activation coefficient *h* using L_2 normalization is defined as

$$w, h = \min_{w,h} D\left(v \,\|\, wh\right) + \mu h_1 \tag{6.1}$$

The spacing between v, w, h can be determined by Euclidian space. Here iterative multiplicative update is used to determine w and h then,

$$h \leftarrow \frac{h \cdot * \left(\overline{w}^{T} v + \overline{w} h \right)}{\left(\overline{w}^{T} + \mu \right)} \tag{6.2}$$

$$w \leftarrow w \cdot * \frac{\frac{v}{wh} h^{T} + 1\left(1h^{T} .* \overline{w}\right) \cdot * \overline{w}}{1\overline{h}^{T} + 1\left(\frac{v}{wh} \cdot * \overline{w} \right) \cdot * \overline{w}} \tag{6.3}$$

In the above Equation (6.1) and Equation (6.2), the column-wise L_2 normalization of w is $\overline{w}$. The Hadamard product $\cdot *$ and division / are used to determine w and h in Equation (6.1) and Equation (6.2)

Using SNMF at the output of GSC-CP, the major advantages are residual noise reduction and multiple interference separation based on w and h matrix multiplication by taking corresponding frequencies of interferences at each time instant. The same procedure is verified for noisy input with four microphones and eight microphones, respectively. Therefore, the proposed method removes noise coming from different directions and separates interferences in an adverse environment which is shown clearly in simulation results.

The workflow of the proposed method is shown in Figure 6.4; at first, the Fixed Beamforming (FBF) and the Modified Blocking Matrix (MBM) are analyzed in the frequency domain using Short Time Fourier Transform (STFT). Next, the auditory grouping is performed based on the bark scale, and the frequencies are converted to bark frequency components. Using auditory subband adaptive interference canceller (UFNLMS), the noise is suppressed in each subband based on speech enhancement. Directional noise is reduced using a GSC beamformer in each subband. The diffuse noise is reduced using a combined postfilter (CP) based on SPP using UFNLMS algorithms. At last residual noise is suppressed, introducing SNMF to the GSC-CP method. Finally, desired speech is obtained at the output.

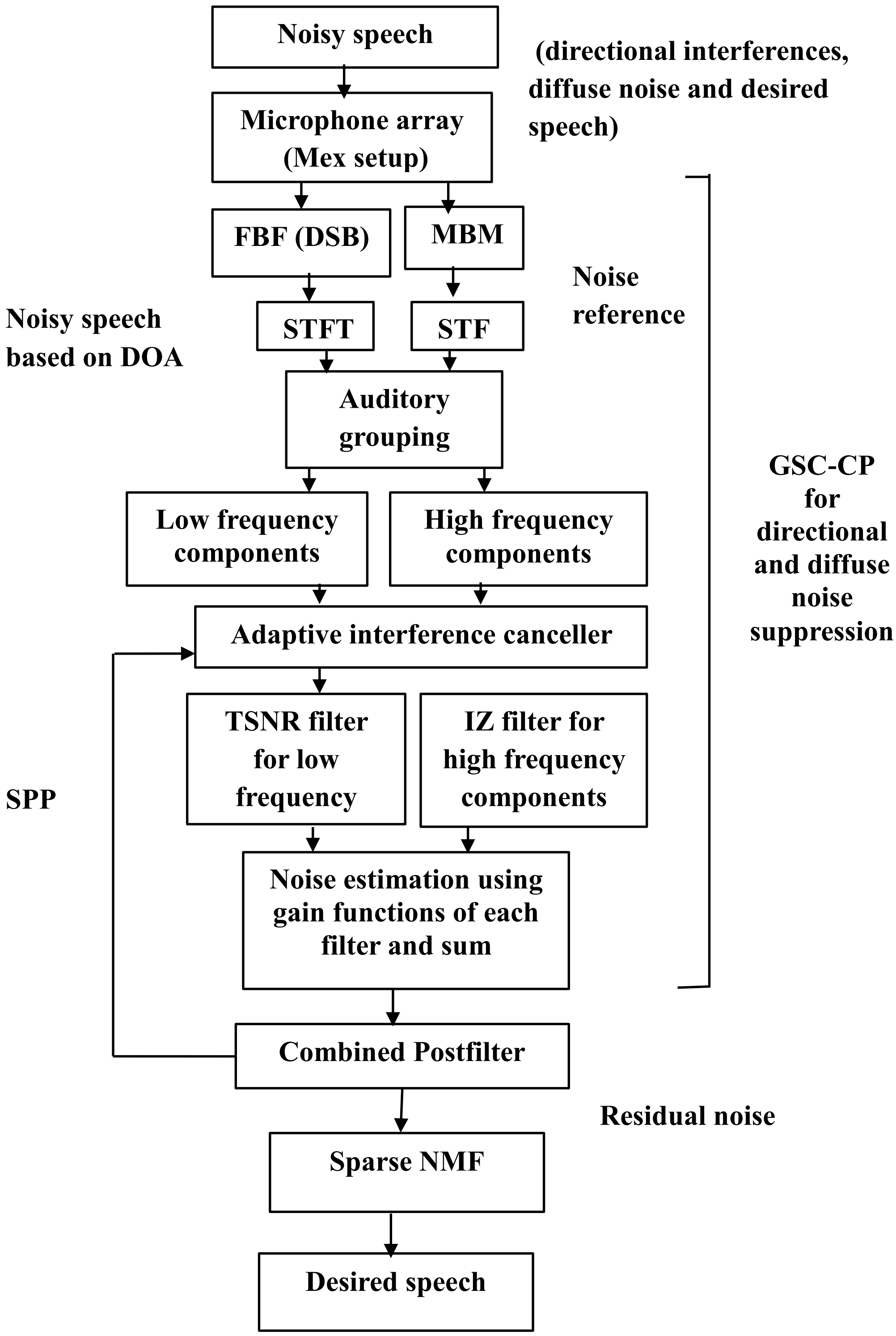

Figure 6.4: Workflow of the Proposed GSC-CP-SNMF

6.4 Simulation Results

In this section, simulation of the proposed GSC beamforming with CP and SNMF in an adverse environment is evaluated and discussed. The simulation parameter with specifications considered for the proposed GSC-CP-SNMF method is shown in Table 6.1 Using image method [225], a multi-channel room impulse response is generated, considering a linear array of four microphones with a distance of 5 cm between each microphone, and distance of 1 m between the source and microphone array, in a conference room with 6 m x 3 m x 5 m and reverberation time of 300 ms following a Mex setup using Mex function, i.e., rir-generator.cpp in MATLAB. The Mex function was taken from International Audio Laboratories Erlangen at Friedrich Alexander University Erlangen-Nuremberg. (https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator) same as chapter 5, but the number of microphones considered here are four and eight to show the efficiency of the system.

An adverse environment is considered taking the desired speech from an unknown direction, directional interferences like white noise at -45^{o}C and female speech at 90^{o}C using the DARPA TIMIT [227]-[228], i.e., database of 6300 male and female sentences, 10 sentences spoken by every 630 speakers with a sampling frequency of 8 kHz and also a diffuse noise [229] from NOIZEUS database [230], i.e., a car noise are given to Mex setup, finally form a noisy input signal (i.e., a combination of desired speech, directional interferences, and diffuse noise) with different SNR levels from -10 dB to 10 dB.

This noisy input signal is applied to FBF, which analyzes the DOA of (known and also unknown) input signal, and all the delays from the microphones are added based on the DSB principle. Finally, a partially enhanced signal is obtained at the output of FBF. In the next stage, from noisy input speech, the delays on the adjacent microphones are subtracted using MBM so that a noise reference is generated. Both the partially enhanced signal and the noise reference are parallelly applied to the auditory grouping.

Later, directional noise is suppressed by GSC using UFNLMS, and diffuse noise is suppressed by CP in each subband. However, in these processes, residual noise is generated, which is removed by proposing an SNMF block to GSC-CP. The same procedure is repeated for eight microphones to increase the performance of the proposed multi-microphone array speech enhancement. The simulation result is compared to the existing methods to demonstrate the performance of the proposed GSC-CP-SNMF approach in terms of intelligibility and quality.

6.4.1 Performance Analysis of the Proposed GSC-CP-SNMF

The performance of the proposed GSC with CP and SNMF algorithm is analyzed in terms of five objective parameters, namely Perceptual Evaluation of Speech Quality (PESQ) [231], Segmental SNR (SSNR) [240], Short Time Object Intelligibility (STOI) [282], Signal to Distortion Ratio (SDR) [283] and Log-Spectral Distance (LSD) [232].

Table 6.1 Simulation Parameters Considered for the Proposed GSC-CP-SNMF

Parameters	**Specifications**
Conference Room Dimensions	6 m X 5 m X 3 m (Using Image Method [225] with Mex setup using Mex function i.e., rir-generator.cpp [226] in MATLAB)
Microphones (m)	m=4, 8
Distance between each microphone	5 cm
Distance from source to microphone	1 m
Diffuse noise	Car noise from Noizeus [230]
Input SNR Levels	-10 dB, -5 dB, 0 dB, 5 dB and 10 dB
Database	Darpa Timit [227]-[228] and Noizeus [229]-[230]
Directional Interferences	white noise at -45 and female speech at 90
Desired speech and diffuse noise	unknown direction

6.4.1.1 Comparison of PESQ for Four and Eight Microphones

PESQ [231] is an objective intelligibility measure, the standard range of which lies between 0.5 to 4.5 dB. The higher the PESQ score better will be the perception. Table 6.2 and Table 6.3 compares the PESQ score for the proposed method with four and eight microphones, respectively. In S. Priyanka [276] method, the PESQ at -10 dB attains 2.45 dB and 2.82 dB with 4 and 8 microphones, respectively.

Table 6.2: PESQ Comparisons for Four Microphones.

Input SNR (dB)	**-10**	**-5**	**0**	**5**	**10**
Gannot[9]	2.02	2.21	2.49	2.66	2.81
K. Li [248]	2.38	2.62	2.83	3.03	3.38
Pfeifenberger [271]	2.06	2.13	2.39	2.53	2.74
X. Wang [272]	2.10	2.22	2.28	2.37	2.54
J. Park [273]	2.56	2.62	2.74	3.11	3.34
L. Zhang [274]	2.17	2.31	2.43	2.58	2.77
G. Rithwik [275]	2.43	2.52	2.65	3.11	3.62
S. Priyanka [276]	2.45	2.68	2.94	3.20	3.42
GSC-CP-SNMF (Proposed)	**3.32**	**3.68**	**3.83**	**4.11**	**4.29**

The proposed GSC-CP-SNMF at -10 dB is 3.32 dB and 3.49 dB with four and eight microphones, respectively. The proposed method shows superior performance, as SPP is used for minimizing the noise power instead of step size in diffuse noise reduction. At the output, the desired speech perception is attained, and interferences are separated due to SNMF. For the proposed method, PESQ of 4.29 dB and 4.34 dB are obtained at 10 dB with 4 and 8 microphones, respectively.

Table 6.3: PESQ Comparisons for Eight Microphones.

Input SNR (dB)	**-10**	**-5**	**0**	**5**	**10**
Gannot[9]	2.11	2.27	2.56	2.83	2.95
K. Li [248]	2.77	2.84	2.89	2.98	3.46
Pfeifenberger [271]	2.18	2.25	2.47	2.68	2.94
X. Wang [272]	2.18	2.32	2.48	2.63	2.67
J. Park [273]	2.71	2.83	2.91	3.27	3.62
L. Zhang [274]	2.25	2.41	2.55	2.69	2.84
G. Rithwik [275]	2.46	2.67	2.83	3.25	3.82
S. Priyanka [276]	2.82	3.03	3.25	3.39	3.81
GSC-CP-SNMF (Proposed)	**3.49**	**3.78**	**3.96**	**4.28**	**4.34**

6.4.1.2 Comparison of SSNR for Four and Eight Microphone

Segmental SNR [240] is one of the most popular objective measures for speech enhancement methods. In normal SNR calculation, the whole signal is considered, whereas while calculating SSNR, segments are taken with 256 samples per frame (k=256, with 50 percent overlap). Higher the SSNR will be better the quality of speech.

Segmental SNR is calculated as

$$SSNR = \frac{10}{L}\frac{\sum_{q=0}^{L-1} 10log \sum_{m=0}^{M-1} X^2(m+q\frac{m}{2})}{\sum_{m=o}^{M-1}\left[X\left(m+q\frac{m}{2}\right)-\hat{s}_e(m+q\frac{m}{2}\right]^2}$$

Table 6.4 SSNR Comparisons for Four Microphones.

Input SNR (dB)	**-10**	**-5**	**0**	**5**	**10**
Gannot[9]	5.2	7.8	9.3	12.7	15.6
K. Li [248]	9.2	14.1	16.5	19.8	22.4
Pfeifenberger [271]	6.4	7.8	9.2	14.3	17.6
X. Wang [272]	7.2	8.5	9.8	15.8	18.3
J. Park [273]	8.2	12.6	13.8	14.3	16.1
L. Zhang [274]	8.7	10.2	11.3	16.2	19.5
G. Rithwik [275]	9.8	11.6	12.5	18.2	20.8
S. Priyanka [276]	9.8	15.3	18.7	20.2	26.8
GSC-CP-SNMF (Proposed)	**11.2**	**16.4**	**19.6**	**22.6**	**28.3**

Table 6.5 SSNR Comparisons for Eight Microphones

Input SNR (dB)	**-10**	**-5**	**0**	**5**	**10**
Gannot[9]	6.5	8.5	9.8	14.3	16.5
K. Li [248]	9.7	16.8	18.2	21.6	23.4
Pfeifenberger [271]	7.3	8.9	10.2	15.3	18.9
X. Wang [272]	8.4	9.9	12.7	18.2	20.4
J. Park [273]	9.6	13.2	14.8	15.7	18.6
L. Zhang [274]	9.4	12.3	16.8	19.9	21.7
G. Rithwik [275]	10.2	14.3	18.3	21.6	22.8
S. Priyanka [276]	11.6	17.4	20.1	22.9	27.7
GSC-CP-SNMF (Proposed)	**13.4**	**18.5**	**21.3**	**25.9**	**29.9**

In the combined postfilter, the low-frequency region, which is below 4kHz, is processed with a TSNR filter, and the high-frequency region, which is in the range 4kHz to 8kHz, is processed with an IZ filter. The diffuse noise in each subband is eliminated by considering speech presence segments that result in improved performance of SSNR. The SSNR for four microphones is 22.6 dB, and 28.3 dB at 5 dB and 10 dB input SNR, respectively, which is better than K. Li [248], and other methods are shown in Table 6.4. Similarly, the SSNR for eight microphones is 25.9 dB, and 29.9 dB at 5 dB and 10 dB input SNR, respectively, which is better than S Gannot [9], K. Li [248], G. Rithwik [275], and other methods are shown in Table 6.5. The proposed GSC-CP-SNMF method shows the highest SSNR values compared to existing methods.

6.4.1.3 Short Time Object Intelligibility (STOI) Comparison for Four and Eight Microphones

STOI [282] is based on the processed and reference signal correlation coefficient of each frame. The standard range of STOI is 0 to 1. The higher the STOI value, the better is the intelligibility of speech. Table 6.6 and Table 6.7 show the STOI comparison for the proposed method using four and eight microphones. STOI performance with four and eight microphones for the proposed GSC-CP-SNMF gives the highest values of 0.802 and 0.892, respectively, at 10 dB input SNR, which is close to 1. It is because UFNLMS is in a subband adaptive feedback controller, which reduces the noise in each subband of the frame by generating a variable step size. The highest STOI is achieved for the proposed GSC-CP-SNMF method.

Table 6.6 STOI Comparisons for Four Microphones

Input SNR (dB)	**-10**	**-5**	**0**	**5**	**10**
Gannot[9]	0.219	0.252	0.306	0.347	0.405
K. Li [248]	0.432	0.498	0.520	0.597	0.612
Pfeifenberger [271]	0.251	0.305	0.394	0.473	0.501
X. Wang [272]	0.253	0.329	0.426	0.519	0.563
J. Park [273]	0.264	0.342	0.459	0.537	0.575
L. Zhang [274]	0.331	0.381	0.505	0.552	0.614
G. Rithwik [275]	0.432	0.469	0.580	0.607	0.634
S. Priyanka [276]	0.512	0.532	0.551	0.639	0.706
GSC-CP-SNMF (Proposed)	**0.522**	**0.567**	**0.628**	**0.716**	**0.802**

Table 6.7 STOI Comparisons for Eight Microphones

Input SNR (dB)	**-10**	**-5**	**0**	**5**	**10**
Gannot[9]	0.268	0.295	0.358	0.384	0.429
K. Li [248]	0.474	0.513	0.579	0.608	0.629
Pfeifenberger [271]	0.272	0.332	0.453	0.489	0.532
X. Wang [272]	0.287	0.367	0.509	0.573	0.589
J. Park [273]	0.326	0.501	0.551	0.620	0.685
L. Zhang [274]	0.376	0.398	0.552	0.591	0.628
G. Rithwik [275]	0.491	0.526	0.598	0.611	0.689
S. Priyanka [276]	0.526	0.541	0.611	0.699	0.756
GSC-CP-SNMF (Proposed)	**0.538**	**0.578**	**0.647**	**0.768**	**0.892**

6.4.1.4 Signal to Distortion Ratio (SDR) Comparison for Four and Eight Microphones

SDR [283] is an objective quality measure to calculate the distortion in each subband. The higher the SDR value, the higher will be the quality of the desired speech signal.

Table 6.8 SDR Comparisons for Four Microphones

Input SNR (dB)	**-10**	**-5**	**0**	**5**	**10**
Gannot[9]	2.45	3.38	4.59	7.21	10.43
K. Li [248]	3.81	4.33	8.28	10.13	11.65
Pfeifenberger [271]	2.57	3.55	5.62	8.93	10.86
X. Wang [272]	3.07	3.76	6.13	9.59	11.02
J. Park [273]	4.92	6.89	8.56	10.98	11.19
L. Zhang [274]	3.87	4.07	7.34	10.25	11.47
G. Rithwik [275]	4.22	4.65	8.39	10.87	11.79
S. Priyanka [276]	6.211	9.11	10.98	11.69	12.61
GSC-CP-SNMF (Proposed)	**7.19**	**10.12**	**11.76**	**12.96**	**13.39**

Table 6.9 SDR Comparisons for Eight Microphones

Input SNR (dB)	**-10**	**-5**	**0**	**5**	**10**
Gannot[9]	2.71	4.31	7.02	8.78	10.91
K. Li [248]	3.96	5.79	9.32	10.92	12.15
Pfeifenberger [271]	2.89	3.80	6.29	9.22	11.05
X. Wang [272]	3.43	3.98	7.67	10.03	11.32
J. Park [273]	5.07	7.65	9.58	11.66	12.57
L. Zhang [274]	4.05	4.74	8.92	10.45	11.88
G. Rithwik [275]	4.97	5.07	10.11	11.58	12.14
S. Priyanka [276]	6.89	9.92	11.12	12.79	13.42
GSC-CP-SNMF (Proposed)	**7.78**	**10.78**	**12.57**	**13.24**	**13.95**

SDR for the GSC-CP-SNMF is higher compared to K. Li [248], X. Wang [272], and other existing methods. SDR at -10 dB for four microphones is 7.19 dB, and for eight microphones, it is 7.78 dB. Due to SNMF at the output of CP, the residual noise generated is reduced. SDR at 10 dB for K. Li [248] method is 12.15 dB, and for the proposed method, it is 13.95 dB for eight microphones which shows the better noise reduction over existing methods which is shown in Table 6.8 and Table 6.9.

6.4.1.5 Comparison of LSD for Four and Eight Microphones

Log Spectral Distance measure [232] is an objective measure for the calculation of the spectral distance between the frames. Better intelligibility can be achieved with a reduction in spectral distance.

LSD is calculated as

$$LSD = \frac{10}{L}\sum_{(q=0)}^{L-1}\left\{1 + \left(\frac{M}{2}+1\right)\sum_{q=0}^{\frac{M}{2}}[logX(p,q) - log\hat{s}_e(p,q)]^2\right\}$$

In Table 6.10 and Table 6.11 LSD measure for the proposed method is compared with the competing methods. The proposed method shows the lower LSD, resulting in better noise reduction. The spatial information is completely utilized by taking MBM into consideration. The spectral distance between two frames is reduced, and the proposed GSC-CP-SNMF achieves better performance compared to the other classical methods. As the distance between the frames decreases, the distortion gets reduced.

Table 6.10 LSD Comparisons for Four Microphones

Input SNR (dB)	**-10**	**-5**	**0**	**5**	**10**
Gannot[9]	9.1	7.9	7.4	7.1	6.8
K. Li [248]	7.9	6.6	5.6	5.2	5.0
Pfeifenberger [271]	7.3	6.0	5.2	4.7	4.0
X. Wang [272]	5.2	4.8	4.2	3.8	3.2
J. Park [273]	4.9	3.8	3.0	2.9	2.4
L. Zhang [274]	4.7	4.1	3.6	3.4	3.0
G. Rithwik [275]	4.0	3.6	3.2	3.0	2.8
S. Priyanka [276]	3.5	2.6	2.0	1.5	1.3
GSC-CP-SNMF (Proposed)	**2.6**	**1.8**	**1.1**	**0.8**	**0.6**

Table 6.11 LSD Comparisons for Eight Microphones

Input SNR (dB)	-10	-5	0	5	10
Gannot[9]	8.3	7.0	6.2	5.7	5.2
K. Li [248]	7.2	6.4	5.8	4.9	4.2
Pfeifenberger [271]	6.7	5.8	5.0	4.2	3.9
X. Wang [272]	4.6	4.3	3.8	3.3	3.0
J. Park [273]	3.6	3.1	2.8	2.2	1.9
L. Zhang [274]	3.8	3.6	3.2	2.8	2.3
G. Rithwik [275]	3.5	3.2	2.5	2.3	2.0
S. Priyanka [276]	2.8	2.2	1.8	1.2	1.0
GSC-CP-SNMF (Proposed)	2	1.5	1.1	0.7	0.4

. The LSD for eight microphones is lower when compared to that for four microphones. LSD for the proposed GSC-CP-SNMF at 10 dB is 0.4 dB, whereas, for the G. Rithwik [275] method, it is 2.0 dB, for K. Li [248] method, it is 4.2 dB for eight microphones, which shows that the proposed method GSC-CP-SNMF has the lower LSD compared to existing methods.

6.4.1.6 Spectrogram for Four and Eight Microphones

In Figure 6.5 and Figure 6.6, the spectrograms of the proposed speech enhancement method using four and eight microphones are illustrated at 10 dB input SNR, which shows the noise reduction of noisy input speech using the proposed GSC-CP-SNMF for four and eight microphone cases, respectively. The proposed GSC-CP-SNMF method shows better noise reduction with improved quality and intelligibility.

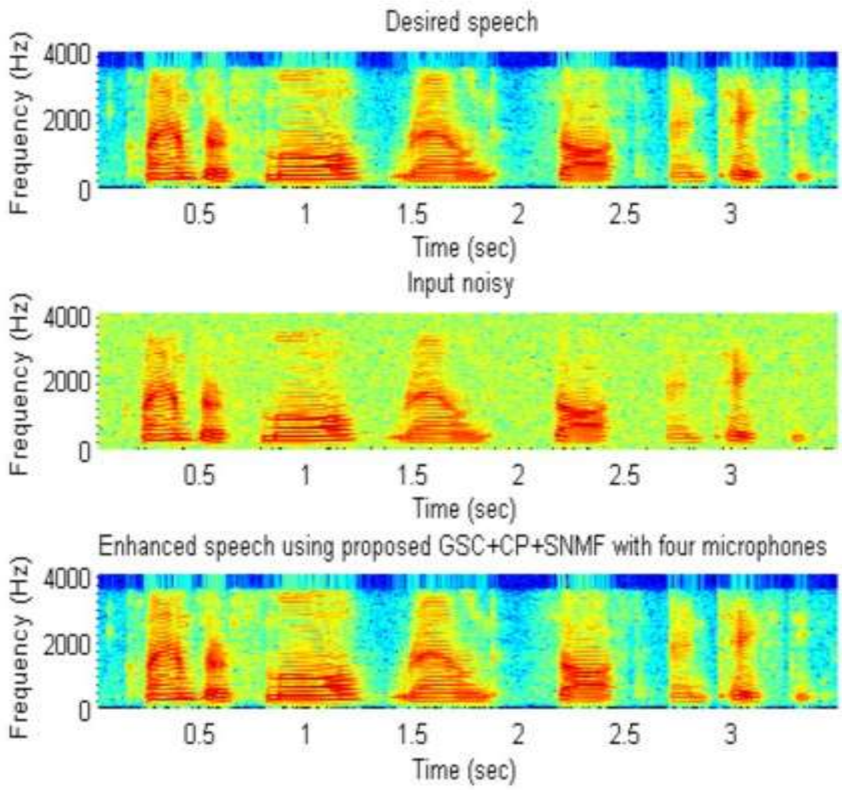

Figure 6.5: Spectrogram for the proposed GSC-CP-SNMF method using four microphones at 10 dB input SNR.

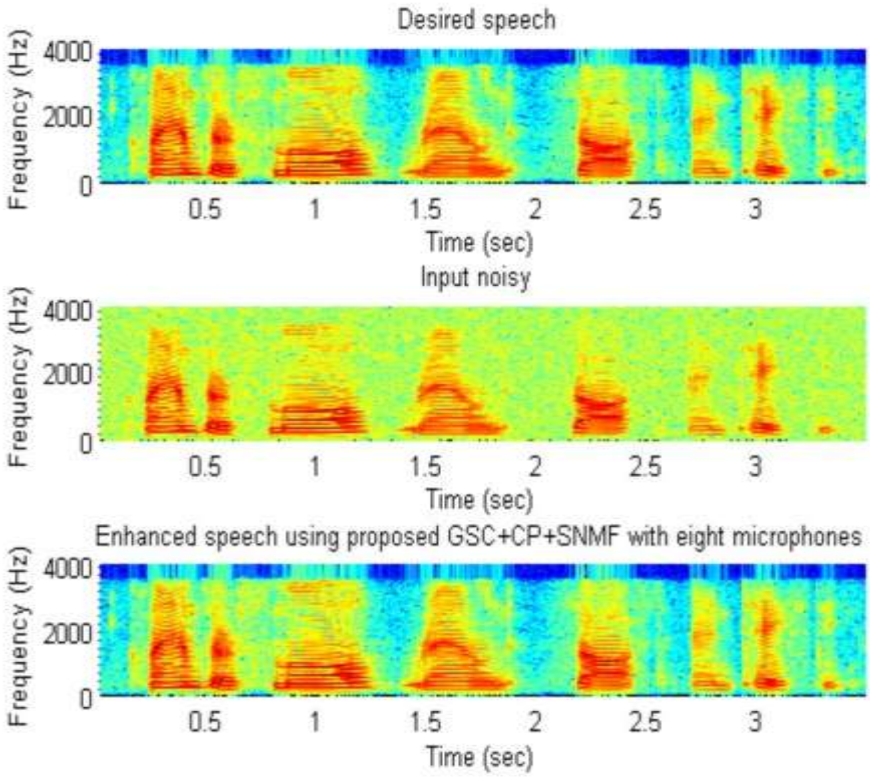

Figure 6.6 Spectrogram for the Proposed GSC-CP-SNMF Method Using Eight Microphones at 10 dB Input SNR

6.5 Summary

A multi-microphone speech enhancement method using GSC beamforming with combined postfilter and SNMF is proposed to enhance the desired speech from directional interferences, diffuse noise, and residual noise. Directional interferences are eliminated using GSC beamforming with UFCNLMS, and the diffuse noise is reduced using combined postfilter and the residual noise by SNMF. Four and eight microphones systems were considered to evaluate the performance, and as the number of microphones increased, the performance was also improved. Using SNMF, the GSC-CP becomes more robust to the real-time environment in the case of multiple speakers. The proposed GSC-CP-SNMF method outperforms existing methods in terms of PESQ, SSNR, STOI, SDR and LSD which is quantified from the results. The proposed method outperforms the current approach in better noise reduction.

The addition of the SNMF algorithm to GSC-CP proves that in teleconference applications, the multi-microphone speech enhancement and multi-microphone speech separation methods can be used in an inter-related manner to have noise-free communication. Better quality and intelligibility were achieved using the proposed method.

Chapter 7

Conclusions and Future Scope

This chapter provides the conclusions of the thesis. The future scope is also provided, which suggests some of the potential research areas in multi-channel speech enhancement.

7.1 Conclusions

This thesis focuses on developing novel adaptive beamforming approaches for Multi-Channel Speech Enhancement (MCSE) algorithms on real-time noisy conditions and adverse environment in four contributions which is as follows.

The first contribution generalized sidelobe canceler (GSC) beamformer using different adaptive filtering algorithms is proposed to address the different real-time noisy types in the existing multi-channel speech enhancement. For the desired speech from the noisy environment, GSC adaptive beamformer initially finds the direction of arrival of the noisy speech signal, based on calculating the delay from each microphone using delay and sum beamformer (DSB). Later, it cancels the received noise using the adaptive filtering algorithms. A virtual conference room setup is designed by following the image method with a Mex rir generator to generate real-time noise input signals in a multi-channel environment like a teleconference. Then, a noisy input speech using multiple microphones is simulated and is applied for the estimation. GSC beamforming with traditional adaptive filtering algorithms like LMS, NLMS reduces to low noise. Whereas, another conventional algorithm i.e., RLS adaptive algorithm produces high computational complexity in the sidelobe canceling path of the GSC beamformer. So, a novel fast convergence NLMS (FCNLMS) algorithm is implemented in the sidelobe canceling path of the GSC beamformer to various real-time noises under different SNR levels. At -10 dB, the PESQ for proposed GSC-FCNLMS is 2.769 dB, whereas GSC- LMS, GCS-NLMS, and GSC-RLS it is 2.494 dB, 2.53 dB and 2.75 dB under station noise conditions. Similarly, for at -10 dB input SNR, GSC-FCNLMS output SNR is 6.9 dB, whereas GSC-LMS and GSC-NLMS are 6.3 dB and 6.8 dB, respectively. The proposed GSC beamforming with the FCNLMS algorithm gives the best performance in terms of intelligibility and quality is compared to the existing GSC with LMS, NLMS, and RLS algorithms at low SNRs.

In the second contribution, a novel signed convex combination of two FCNLMS algorithms is implemented to address all real-time noises further and reduce the computation burden on the GSC beamforming in the high SNRs, i.e., above 10 dB. The adaptive filter error is minimized using a mixing parameter to update the filter weights in the convex combinational technique. Meanwhile, the GSC beamformer computational cost is lowered compared to GSC-LMS and GSC-NLMS by using the signed algorithm to the convex combinational approach. The proposed GSC beamformer with signed convex combinational of fast convergence approach improves the speech quality with temporal characteristics at high SNRs with PESQ of 4.393 dB at 15 dB input SNR for street noise and 4.355 for station noise with low computation time is achieved of 0.97 ms is achieved. Better quality and intelligibility are attained using the proposed GSC-SCCFC for all real-time noises given at microphone array input and achieved low computational complexity compared to the existing algorithm.

Most of the diffuse noise fields are addressed by postfilters, but in the adverse environment, along with diffuse noise, there exist interference speakers coming from unknown directions are discussed in contribution three. A GSC beamforming with improved Zelinski-TSNR postfiltering is proposed to address directional and diffuse noise. Directional noise is suppressed by the GSC beamformer, whereas diffuse noise is suppressed by Zelinski-TSNR multi-channel postfilter. The Zelinski filter is applied for high frequencies, and the TSNR filter is applied for low frequencies. The speech absence frame is eliminated using speech presence probability (SPP) and adaptive interference canceller. Desired speech is obtained as GSC beamformer using improved Zelinski-TSNR multi-channel postfilter. In each subband, the speech absence frames are eliminated. Both directional and diffuse noises are suppressed using GSC with Zelinski-TSNR postfilter. The proposed GSC with Zelinski-TSNR postfilter gives a PESQ of 2.45 dB at -10 dB input SNR and 3.42 dB at 10 dB input SNR in an adverse environment when compared with existing Cohen and Kai Li methods of 2.13 dB and 2.38 dB at -10 dB input SNR, 3.15 dB, and 3.28 dB at 10 dB. The proposed outperforms the existing techniques in suppressing directional and diffuse noise.

In contribution of four to reduce the production of residual noise, which is generated at the output GSC with combined postfilter (CP). A novel sparse NMF (SNMF) algorithm is proposed at the output of GSC-CP. The basis vector and activation functions are updated using the sparse constraint. The non-negative factorization (NMF) model separates the speech absence using the basis vector and activation function effectively. The data in the given spectrogram like speech to one matrix and noise to another matrix, i.e., W and H. The proposed GSC-CP-SNMF beamformer separates residual noise and produces the desired speech signal at the output. GSC-CP-SNMF separates noise from a noisy speech and separates interference speakers, if any, in the environment. To attain better quality and intelligibility in the adverse environment, eight microphones are also used to simulate GSC-CP-SNMF and attained. The proposed GSC-CP-SNMF gives a PESQ of 3.49 dB at -10 dB input SNR and 4.34 dB at 10 dB input SNR for eight microphone array input and 4.29 3.32 for four microphones at 10 dB and -10 dB input SNR. The proposed method outperforms the existing Kai li and Ritwik methods of 2.38 dB and 2.43 dB at -10 dB input SNR, 3.38 dB, and 3.62 dB at 10 dB input SNR.

The thesis shows the effectiveness and robustness of the developed adaptive beamforming approaches for multi-channel speech enhancement (MCSE). Various real-world non-stationary noisy environments with a wide range of SNRs were considered for the performance evaluation. The performance of the developed algorithms in terms of waveforms spectrograms and objective parameters is presented, which shows the superiority of the proposed algorithms when compared to the existing speech enhancement algorithms in dealing with issues like directional noise, diffuse noise, and handling of residual noise in a real-time environment. The proposed GSC-CP-SNMF also supports source separation in an adverse environment.

7.2 Future Scope

As future work, novel early and late fusion Convolutional Neural Networks (CNNs) are proposed for multi-channel speech enhancement. Two beamformers, namely Delay-and-Sum (DS) and Minimum Variance Distortion less Response (MVDR), are used as pre-filters to suppress the effect of noise in the input microphone array. Enhanced outputs of the two beamformers are to form two-channel input to the CNN, and it is known as the early fusion CNN model. On the other hand, outputs of the beamformers are considered as inputs to the two individual CNN's separately. Further, outputs of CNNs are concatenated to form an input to the fully connected layers, and it is known as the late fusion CNN model.

References

[1] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 32, no. 6, pp. 1109–1121, 1984.

[2] D. E. Tsoukalas, J. N. Mourjopoulos, and G. Kokkinakis, “Speech enhancement based on audible noise suppression,” IEEE Transactions on Speech Audio Processing, vol. 5, no. 6, pp. 497–514, 1997.

[3] S. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Transactions on Acoustic Speech Signal Processing, vol. 27, no. 2, pp. 113–120, 1979.

[4] P. C. Loizou and S. Member, “Speech enhancement based on perceptually motivated bayesian estimators of the magnitude spectrum,” IEEE Transactions on Speech Audio Processing, vol. 13, no. 5, pp. 857–869, 2005.

[5] J. Benesty, M. M. Sondhi, and Y. A. Huang, “Springer handbook of speech processing” Springer, vol. 8, 2008.

[6] M. Berouti, R. Schwartz, and J. Makhoul, “Enhancement of speech corrupted by acoustic noise,” in acoustics, speech, and signal processing," IEEE International Conference on ICASSP 79, vol. 4, pp. 208–211, 1979.

[7] M. Brandstein and D. Ward, Microphone Arrays: “Signal Processing and Applications,” Springer, 2001.

[8] Y. Hu and P. C. Loizou, “Subjective comparison and evaluation of speech enhancement algorithms,” Speech Communication, vol. 49, no. 7-8, pp. 588 – 601, 2007.

[9] S. Gannot, E. Vincent, S. Markovich-Golan and A. Ozerov, “A consolidated perspective on multi-microphone speech enhancement and source separation,” IEEE/ACM Transactions on Audio, Speech, Language Processing, vol. 25, pp. 692–730, 2017.

[10] M. M. Goulding and J. S. Bird, “Speech enhancement for mobile telephony,” IEEE Transactions on Vehicle Technology, vol. 39, no. 4, pp. 316–326, 1990.

[11] N. D. Degan and C. Parti, “Acoustic noise analysis and speech enhancement techniques for mobile radio applications,” Signal Processing, vol. 15, pp. 43–56, 1988.

[12] R. Le Bouquin Jeannes, P. Scalart, G. Faucon, and C. Beaugeant, “Combined noise and echo reduction in hands-free systems: A survey,” IEEE Transactions on Speech Audio Processing, vol. 9, no. 8, pp. 808–820, 2001.

[13] T. Yamada, M. Kumakura, and N. Kitawaki, “System under noise conditions using objective,” IEEE Transactions on Audio, Speech, Language Processing, vol. 14, no. 6, pp. 2006–2013, 2006.

[14] J. H. L. Hansen and M. A. Clements, “Constrained iterative speech enhancement with application to speech recognition,” IEEE Transactions on Signal Processing, vol. 39, no. 4, pp. 795–805, 1991.

[15] O. Roy and M. Vetterli, “Rate-constrained collaborative noise reduction for wireless hearing aids,” IEEE Transactions on Signal Processing, vol. 57, no. 2, pp. 645–657, 2009.

[16] Spriet, M. Moonen, and J. Wouters, “Robustness analysis of multi-channel Wiener filtering and generalized sidelobe cancellation for multi-microphone noise reduction in hearing aid applications,” IEEE Transactions on Speech Audio Processing, vol. 13, no. 4, pp. 487–503, 2005.

[17] M. Li, H. G. McAllister, N. D. Black, and T. A. De Perez, “Perceptual time-frequency subtraction algorithm for noise reduction in hearing aids,” IEEE Transactions on Biomedical Engineering, vol. 48, no. 9, pp. 979–988, 2001.

[18] B. D. Van Veen and K. M. Buckley, “Beamforming: A versatile approach to spatial filtering,” IEEE ASSP Magazine, pp. 4-24, 1988.

[19] H. Krim and M. viberg, “Two decades of array signal processing research,” IEEE Signal Processing Magazine, 1996.

[20] O. L. Frost, “An algorithm for linearly constrained adaptive array processing,” Proceeding IEEE, vol. 60, pp. 926–935, 1972.

[21] L. J. Griffths and C. W. Jim, “An alternative approach to linearly constrained adaptive beamforming,” IEEE Transactions on Antenna Propagation, vol. 30, pp. 27–34, 1982.

[22] B. Widrow, “Adaptive noise canceling: Principles and applications,” Proceeding of IEEE, vol. 63, pp. 1692–1716, 1975.

[23] E. A. P. Habets, J. Benesty, I. Cohen and S. Gannot, “New insights into the MVDR beamformer in room acoustics,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 1, 2010.

[24] Mehrez Souden, Jacob Benesty, and Sofiène Affes, “A study of the LCMV and MVDR noise reduction filters,” IEEE Transactions on Signal Processing, vol. 58, no. 9, 2010.

[25] Emanuël A. P. Habets, Jacob Benesty, and Patrick A. Naylor, “A speech distortion and interference rejection constraint beamformer,” IEEE Transactions On Audio, Speech, And Language Processing, vol. 20, no. 3, 2012.

[26] S. Gannot, D. Burshtein, and E. Weinstein, “Analysis of the power spectral deviation of the general transfer function GSC,” IEEE Transactions on Signal Processing, vol. 52, no. 4, pp. 1115-1121, 2004.

[27] K.U. Simmer, J. Bitzer and C. Marro, Post-filtering techniques. “Microphone Arrays: Signal Processing Techniques and Applications,” ed. by M.S. Brand stein, D.B. Ward Springer, Berlin, Heidelberg, pp. 39-60, 2001.

[28] R. Zelinski, “A microphone array with adaptive post-filtering for noise reduction in reverberant rooms,” IEEE ICASSP, pp. 2578-2581, New York, 1988.

[29] R. Zelinski, “Noise reduction based on microphone array with LMS adaptive post-filtering,” Electronics Letters, vol. 26, no. 24, pp. 2036-2581, 1990.

[30] S. Fischer and K.U. Simmer, “An adaptive microphone array for hands-free communication,” International Workshop Acoustics on Echo Noise Control (IWAENC), pp. 44-47, Røros, 1995.

[31] S. Fischer and K.U. Simmer, “Beamforming microphone arrays for speech acquisition in noisy environments,” Speech Communication, vol. 20, no. 3-4, pp. 215-227, 1996.

[32] S. Fischer and K.D. Kammeyer, “Broadband beamforming with adaptive postfiltering for speech acquisition in a noisy environment,” Proceedings of ICASSP, vol. 1, pp. 359-362, Munich, 1997.

[33] J. Meyer and K.U. Simmer, “Multi-channel speech enhancement in a car environment using Wiener filtering and spectral subtraction,” Proceedings of ICASSP, pp. 21-24, Munich, 1997.

[34] K.U. Simmer, S. Fischer and A. Wasiljeff, “Suppression of coherent and incoherent noise using a microphone array,” Annales des Télécommunications, Vol. 49, no. 7-8, pp. 439-446, 1994.

[35] J. Bitzer, K. Simmer and K.D. Kammeyer, “Multi microphone noise reduction by post-filter and superdirective beamformer,” International Workshop Acoustics on Echo Noise Control (IWAENC), pp. 100-103, Pocono Manor, 1999.

[36] J. Bitzer, K.U. Simmer and K.D. Kammeyer, "Multi microphone noise reduction techniques as front end devices for speech recognition," Speech Communication, vol. 34, pp. 3-12, 2001.

[37] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," Environmetrics-Wiley Online Library, vol. 5, no. 2, pp. 111– 126, 1994.

[38] D. Lee and H. Seung, "Algorithms for nonnegative matrix factorization," Advanced Neural Information Processing System, no. 1, pp. 556–562, 2001.

[39] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature, vol. 401, no. 6755, pp. 788–91, 1999.

[40] S. A. A. Cichocki, R. Zdunek and A. Phan, "Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation," John Wiley & Sons, 2009.

[41] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," Computer Statistics and Data Analysis, vol. 52, no. 1, pp. 155–173, 2007.

[42] Buciu, "Non-negative matrix factorization, a new tool for feature extraction : theory and applications non-negative matrix factorization approach," Journal of Computers, Communications, and Control, vol. 3, pp. 67–74, 2008.

[43] M. Chu and R. Plemmons, "Nonnegative matrix factorization and applications," Bulletin of the International Linear Algebra Society, vol. 34, no. 1, pp. 2–7, 2005.

[44] S. Sra and I. Dhillon, "Nonnegative matrix approximation: Algorithms and applications," Computer Science Department, University of Texas at Austin, 2006.

[45] K. Drakakis and S. Rickard, "Analysis of financial data using non-negative matrix factorization," International Mathematical Forum, vol. 3, no. 38, pp. 1853–1870, 2008.

[46] C. Ding, T. Li, W. Peng, and H. Park, "Orthogonal nonnegative matrix t-factorizations for clustering," Proceedings of the 12th ACM SIGKDD, pp. 126–135, 2006.

[47] R. Sandler and M. Lindenbaum, "Nonnegative matrix factorization with earth mover's distance metric for image analysis," IEEE Transactions on pattern analysis and machine intelligence, vol. 33, no. 8, pp. 1590–1602, 2011.

[48] Kotsia, S. Zafeiriou, and I. Pitas, "A novel discriminant non-negative matrix factorization algorithm with applications to facial image characterization problems,"

IEEE Transactions on Information Forensics and Security, vol. 2, no. 3, pp. 588–595, 2007.

[49] W. Wang, “Squared Euclidean distance-based convolutive non-negative matrix factorization with multiplicative learning rules for audio pattern separation,” ISSPIT, IEEE International Symposium on Signal Processing and Information Technology, pp. 347–352, 2007.

[50] E. Benetos and C. Kotropoulos, “Non-negative tensor factorization applied to music genre classification,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 8, pp. 1955–1967, 2010.

[51] R. Zdunek and A. Cichocki, “Non-negative matrix factorization with quasi-Newton optimization,” In International conference on artificial intelligence and soft computing, vol. 4029, pp. 870–879, 2006.

[52] P. C. Loizou, “Speech Enhancement: Theory and Practice,” 1st edition, Boca Raton, FL: CRC Press, 2007.

[53] R. Le Bouquin Jeannes, P. Scalart, G. Faucon, and C. Beaugeant, “Combined noise and echo reduction in hands-free systems: A survey,” IEEE Transaction on Speech Audio Processing, vol. 9, no. 8, pp. 808–820, 2001.

[54] M. Li, H. G. McAllister, N. D. Black, and T. A. De Perez, “Perceptual time-frequency subtraction algorithm for noise reduction in hearing aids,” IEEE Transaction on Biomedical Engineering, vol. 48, no. 9, pp. 979–988, 2001.

[55] S. S. Chen, D. L. Donoho, and M. a. Saunders, “Atomic decomposition by basis pursuit,” SIAM J. Sci. Comput., vol. 43, no. 1, pp. 129-159, 2001.

[56] Jolliffe, "Principal Component Analysis," Wiley Online Library, 2005.

[57] G. H. Golub and C. Reinsch, “Singular value decomposition and least squares solutions,” In Linear algebra, Springer, Berlin, Heidelberg, pp. 134-151, 1971.

[58] M. a Casey and A. Westner, “Separation of mixed audio sources by independent subspace analysis,” In Proceedings of the International Computer Music Conference, pp. 154–161, 2000.

[59] P. C. Loizou, “Speech Enhancement: Theory and Practice, CRC Press,” In Taylor Francis Group, Boca Raton, 2007.

[60] J. Benesty, J. Jensen, M. G. Christensen., and J. Chen, “Speech Enhancement,” Oxford: Academic Press, 2014. [Online]. Available: http://www.sciencedirect.com/science/book/9780128001394

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

[62] Y. Ephraim, "Statistical-model-based speech enhancement systems," Proceedings of the IEEE, vol. 80, no. 10, pp. 1526-1555, 1992.

[63] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," In 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, pp. 44164-44164, 2002.

[64] S. Lee, D. K. Han, and H. Ko, "Single-channel speech enhancement method using reconstructive NMF with spectrotemporal speech presence probabilities," Applied Acoustics, vol. 117, pp. 257-262, 2017. Available: http://www.sciencedirect.com/science/article/pii/S0003682X16301165

[65] P. Mowlace, J. Stahl, and J. Kulmer, "Iterative joint map single-channel speech enhancement given non-uniform phase prior," Speech Communication, vol. 86, pp. 85-96, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167639316300188

[66] K. Lee, "Application of non-negative spectrogram decomposition with sparsity constraints to single-channel speech enhancement," Speech Communication, vol. 58, pp. 69-80, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167639313001672

[67] S. So and K. K. Paliwal, "Modulation-domain kalman filtering for single-channel speech enhancement," Speech Communication, vol. 53, no. 6, pp. 818-829, 2011.

[68] K. Barros and N. Ohnishi, "Single channel speech enhancement by efficient coding," Signal Processing, vol. 85, no. 9, pp. 1805-1812, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168405001015

[69] K. Paliwal, K. WAşjcicki, and B. Schwerin, "Single-channel speech enhancement using spectral subtraction in the short-time modulation domain," Speech Communication, vol. 52, no. 5, pp. 450-475, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167639310000282

[70] P. Mowlace, M. Blass, and W. B. Kleijn, "New results in modulation-domain single-channel speech enhancement," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 11, pp. 2125-2137, 2017.

[71] M. Krawczyk and T. Gerkmann, "STFT phase reconstruction in voiced speech for an improved single-channel speech enhancement," IEEE/ACM Transactions Audio, Speech. and Language Processing, vol. 22, no. 12, pp. 1931-1940, 2014.

[72] C. S. J. Doire, M. Brookes, P. A. Naylor, C. M. Hicks, D. Betts, M. A. Dmour, and S. H. Jensen, "Single-channel online enhancement of speech corrupted by reverberation and noise," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25. no. 3, pp. 572-587, 2017.

[73] M. Pirolt, J. Stahl, P. Mowlaee, V. I. Vorobiov, S. Y. Barysenka, and A. G. Davydov, "Phase estimation in single-channel speech enhancement using phase invariance constraints," In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5585-5589, 2017.

[74] P. Mowlace and J. Kulmer, "Harmonic phase estimation in single-channel speech enhancement using phase decomposition and SNR information," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 9, pp. 1521-1532, 2015.

[75] J. Gonzalez-Rodriguez, J. L. Sanchez-Bote, and J. Ortega-Garcia, "Speech dereverberation and noise reduction with a combined microphone array approach," In 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings, vol. 2, pp. 1037-1040, 2000.

[76] K. Kokkinakis and P. C. Loizou, "Multi-microphone adaptive noise reduction strategies for coordinated stimulation in bilateral cochlear implant devices," The Journal of the Acoustical Society of America, vol. 127, no. 5, pp. 3136-3144, 2010. [Online]. Available: http://dx.doi.org/10.1121/1.3372727

[77] J. Meyer and K. U. Simmer, "Multi-channel speech enhancement in a car environment using Wiener filtering and spectral subtraction," IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, pp. 1167-1170, 1997.

[78] J. Tu and Y. Xia, "Effective kalman filtering algorithm for distributed multi-channel speech enhancement," Neurocomputing, vol. 275, pp. 144-154, 2017. [Online]. Available: http://www. sciencedirect.com/science/article/pii/S0925231217308986

[79] S. M. Kim, H. K. Kim, S. J. Lee, and Y. K. Lee, "Adaptation mode control with residual noise estimation for beamformer-based multi-channel speech enhancement," In 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 301-304, 2012.

[80] N. Moritz, K. Adiloğlu, J. Anemijller, S. Goetze, and B. Kollmeier, "Multi-channel speech enhancement and amplitude modulation analysis for noise-robust automatic speech recognition," Computer Speech & Language, vol. 46, pp. 558-573, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0885230816300729

[81] J. AnemÄijller and H. Kayser, "Multi-channel signal enhancement with speech and noise covariance estimates computed by a probabilistic localization model," In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 156-160, 2017.

[82] D. Mahmoudi and A. Drygajlo, "Wavelet transform based coherence function for multi-channel speech enhancement," In 9th European Signal Processing Conference EUSIPCO, pp. 1-4, 1998.

[83] W. Nabi, N. Aloui, and A. Cherif, "An improved speech enhancement algorithm for dual-channel mobile phones using wavelet and genetic algorithm," Computers & Electrical Engineering, vol. 62, pp. 692-705, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S004579017301684

[84] J. Heymann, L. Drude, and R. Haeb-Umbach, "A generic neural acoustic beamforming architecture for robust multi-channel speech processing," Computer Speech & Language, vol. 46, pp. 374-385, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0885230816300730

[85] R. Tong. G. Bao, and Z. Ye, "A higher-order subspace algorithm for multi-channel speech enhancement," IEEE Signal Processing Letters, vol. 22, no. 11, pp. 2004-2008, 2015.

[86] Y. Ephraim, "A Bayesian estimation approach for speech enhancement using hidden Markov models," IEEE Transactions on Signal Processing, vol. 40, no. 4, pp. 725-735, 1992.

[87] H, Sameti, H. Sheikhzadeh, L. Deng, and R. L. Brennan. "HMM-based strategies for enhancement of speech signals embedded in nonstationary noise," IEEE Transactions on Speech and Audio Processing, vol. 6, no. 5, pp. 445-455, 1998.

[88] D. Y. Zhao and W. B. Kleijn, "HMM-based gain modeling for enhancement of speech in noise," IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 3, pp. 882-892, 2007.

[89] N. Mohammadiha, R. Martin, and A. Leijon, "Spectral-domain speech enhancement using HMM state-dependent super-Gaussian priors," IEEE Signal Processing Letters. vol. 20, no. 3, pp. 253-256, 2013.

[90] H. Veisi and H. Sameti, "Speech enhancement using hidden Markov models in Mel frequency domain," Speech Communication, vol. 55, no. 2, pp. 205-220, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167639312000970

[91] S. Chehrehsa and T. J. Moir, "Speech enhancement using maximum a-posteriori and gaussian mixture models for speech and noise periodogram estimation," Computer Speech & Language, vol. 36, pp. 58-71, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0885230815000868

[92] S. Chehresa and M. H. Savoji, "MMSE speech enhancement using GMM," In Artificial Intelligence and Signal Processing (AISP), CSI International Symposium, pp. 266-271, 2012.

[93] T. V. Sreenivas and P. Kirnapure, "Codebook constrained Wiener filtering for speech enhancement," IEEE Transactions on Speech and Audio Processing, vol. 4, no. 5, pp. 383-389, 1996.

[94] S. Srinivasan, J. Samuelsson, and W. B. Kleijn, "Codebook driven short-term predictor parameter estimation for speech enhancement," IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, no. 1, pp. 163-176, 2006.

[95] Y. Xu, J. Du, L. R. Dai, and C. H. Lee, "An experimental study on speech enhancement based on deep neural networks," IEEE Signal Processing Letters, vol. 21, no. 1, pp. 65-68, 2014.

[96] Y. Xu, J. Du, L. R. Dai, and C. H. Lee, "A regression approach to speech enhancement based on deep neural networks," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 1, pp. 7-19, 2014.

[97] K. W. Wilson, B. Raj, P. Smaragdis, and A. Divakaran, "Speech denoising using nonnegative matrix factorization with priors," In 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4029-4032, 2008.

[98] N. Mohammadiha, J. Taghia, and A. Leijon, "Single channel speech enhancement using Bayesian NMF with recursive temporal updates of prior distributions," In 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4561-4564, 2012.

[99] H. T. Fan, J. w. Hung, X. Lu, S. S. Wang, and Y. Tsao. "Speech enhancement using segmental nonnegative matrix factorization," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 4483-4487, 2014.

[100] L. Qiao, X. Zhang, X. Chen, and J. Yang. "Speech enhancement using non-negative matrix factorization solved by improved alternating direction method of multipliers," In 2016 International Conference on Progress in Informatics and Computing (PIC), pp. 374-378, 2016.

[101] P. Harding and B. Milner, "Reconstruction-based speech enhancement from robust acoustic features," Speech Communication, vol. 75, pp. 62-75, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167639315001065

[102] T. Gao, J. Du, L.-R. Dai, and C.-H. Lee, "A unified DNN approach to speaker-dependent simultaneous speech enhancement and speech separation in low SNR environments," Speech Communication, vol. 95, pp. 28-39, 2017.

[103] E. Vincent, R. Gribonval, and M. D. Plumbley, "Oracle estimators for the benchmarking of source separation algorithms," Signal Processing, vol. 87, no. 8, pp. 1933–1950, 2007.

[104] S. L. Gay and J. Benesty, Eds., "Acoustic signal processing for telecommunication," Springer, vol. 551, 2012.

[105] J. Benesty, S. Makino, and J. Chen, Eds., "Speech Enhancement," Springer, 2005.

[106] Cohen, J. Benesty, and S. Gannot, Eds., "Speech processing in modern communication: Challenges and perspectives," Springer, 2010.

[107] P. O'Grady, B. Pearlmutter, and S. T. Rickard, "Survey of sparse and non-sparse methods in source separation," International Journal of Imaging Systems and Technologies, vol. 15, pp. 18–33, 2005.

[108] S. Makino, T.-W. Lee, and H. Sawada, Eds., "Blind speech separation," Springer, 2007.

[109] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra, "Convolutive blind source separation methods," In Springer Handbook of Speech Processing, Springer, pp. 1065–1094, 2008.4

[110] P. Comon and C. Jutten, Eds., "Handbook of Blind Source Separation, Independent Component Analysis, and Applications," Academic Press, 2010.

[111] E. Vincent, M. Jafari, S. A. Abdallah, M. D. Plumbley, and M. E. Davies, "Probabilistic modeling paradigms for audio source separation," In Machine Audition: Principles, Algorithms, and Systems, pp. 162–185. IGI Global, 2010.

[112] E. Vincent, N. Bertin, R. Gribonval, and F. Bimbot, “From blind to guided audio source separation: How models and side information can improve the separation of sound,” IEEE Signal Processing Magazine, vol. 31, no. 3, pp. 107–115, 2014.

[113] S. Haykin, "Adaptive Filter Theory," Pearson Education India, 2008.

[114] B. Widrow and S.D. Stearns, "Adaptive Signal Processing," Prentice-Hall, Upper Saddle River,1985.

[115] D. Monzingo and T. Miller, "Introduction to Adaptive Arrays," Wiley, New York 1980.

[116] H.L.V. Trees, "Detection, Estimation, and Modulation Theory," Vol. 4, Wiley, New York, 2002.

[117] H. Cox, R. Zeskind and M. Owen, “Robust Adaptive Beamforming,” IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 35, no. 10, pp. 1365–1376, 1987.

[118] S.L. Gay and J. Benesty, "Acoustic Signal Processing for Telecommunication," Kluwer Academic, Dordrecht, 2001.

[119] J. Benesty and Y. Huang (Eds.), “Adaptive Signal Processing: applications to real-world problems,” Springer, Berlin, Heidelberg, 2003.

[120] E. Hänsler and G. Schmidt, "Acoustic echo and noise control: a practical approach,” Wiley, New York , 2004.

[121] W. Herbordt, “Sound capture for human/machine interfaces – Practical aspects of microphone array signal processing,” Springer, Berlin, Heidelberg, 2005.

[122] S. Gannot, J. Benesty, J. Bitzer, I. Cohen, S. Doclo, R. Martin and S. Nordholm, “Advances in multi-microphone speech processing,” EURASIP Journal of Applied Signal Processing, 2006.

[123] E. Jan and J. Flanagan, “Microphone arrays for speech processing,” International Symposium on Signals, Systems, and Electronics, ISSSE, URSI, San Francisco, pp. 373–376, 1995.

[124] E. Jan and J. Flanagan, “Sound capture from spatial volumes: Matched-filter processing of microphone arrays having randomly-distributed sensors,” Proc. ICASSP, pp. 917–920, Atlanta, 1996.

[125] D. Rabinkin, R. Renomeron, J. Flanagan and D. Macomber, “Optimal truncation time for matched filter array processing,” IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP, pp. 3629–3632, Seattle, 1998.

[126] M. Sondhi and G. Elko, "Adaptive optimization of microphone arrays under a nonlinear constraint," IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP, vol. 11, pp. 981–984, Tokyo, 1986.

[127] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," IEEE Transactions on Acoustics Speech and Signal Processing, vol. 34, no. 6, pp. 1391–1400, 1986.

[128] D.V. Compernolle, "Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings," In International Conference on Acoustics, Speech, and Signal Processing ICASSP, pp. 833–836, Albuquerque, 1990.

[129] W. Kellermann, "Acoustic echo cancellation for beamforming microphone arrays." In Microphone Arrays: Signal Processing Techniques and Applications, ed. by M.S. Brandstein, D.B. Ward, (Springer, Berlin, Heidelberg), pp. 281–306, 2001.

[130] Claesson and S. Nordholm, "A spatial filtering approach to robust adaptive beamforming," IEEE Transactions on Antennas Propagation Letters, vol. ~40, no. 1, pp. 1093–1096, 1992.

[131] S. Nordholm, I. Claesson and N. Grbi´c, "Optimal and adaptive microphone arrays for speech input in automobiles," In Microphone Arrays: Signal Processing Techniques and Applications, ed. by M.S. Brandstein, D.B. Ward (Springer, Berlin, Heidelberg 2001) pp. 307–330

[132] R. Martin, "Small microphone arrays with postfilters for noise and acoustic echo reduction" Microphone Arrays: Signal Processing Techniques and Applications, ed. by M.S. Brandstein, D.B. Ward (Springer, Berlin, Heidelberg 2001) pp. 255–280

[133] B. Widrow, J.R. Glover Jr., J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, E. Dong Jr, and R.C. Goodlin, "Adaptive noise canceling: Principals and applications," Proceedings of the IEEE, vol. 63, no. 12, pp. 1692–1716, 1975.

[134] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," IEEE Transactions on Signal Processing, vol. 50, no. 9, pp. 2230–2244, 2002.

[135] S. Doclo and M. Moonen, "Design of far-field and nearfield broadband beamformers using eigenfilters," Signal Processing, vol. ~83, no. 12, pp. 2641-2673, 2003.

[136] S. Doclo and M. Moonen, "Multi-microphone noise reduction using recursive GSVD-based optimal filtering with ANC postprocessing stage," IEEE Trans. Speech Audio Process, vol. ~13, no. 1, pp. 53–69, 2005.

[137] Spriet, M. Moonen and J. Wouters, "A multi-channel subband GSVD approach for speech enhancement," In Proc. Int. Workshop on Acoustic Echo and Noise Control, IWAENC, pp. 187-191, 2001.

[138] G. Rombouts and M. Moonen, "QRD-based unconstrained optimal filtering for acoustic noise reduction," Signal Processing, vol. 83, no. 9, pp. 1889–1904, 2003.

[139] G. Rombouts and M. Moonen, "Fast QRD lattice-based unconstrained optimal filtering for acoustic noise reduction," IEEE Transactions on Speech Audio Processing, vol. 13, no. 6, pp. 1130–1143, 2005.

[140] J. Chen, J. Benesty, Y. Huang and S. Doclo, "New insights into the noise reduction Wiener filter," IEEE Transactions on Speech Audio Processing, vol. 14, no. 4, pp. 1218–1234, 2006.

[141] Spriet, M. Moonen and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," Signal Processing, vol. 84, no. 12, pp. 2367–2387, 2004.

[142] S. Doclo, A. Spriet, J. Wouters and M. Moonen, "Speech distortion weighted multi-channel Wiener filtering techniques for noise reduction," In Speech Enhancement, Signals and Communication Technology, ed. by J. Benesty, S. Makino, J. Chen (Springer, Berlin, Heidelberg), pp. 199–228, 2005.

[143] S. Affes and Y. Grenier, "A source subspace tracking array of microphones for double talk situations," IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings, ICASSP, pp. 269–272, Munich, 1997.

[144] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," IEEE Transactions on Signal Processing, vol. 49, no. 8, pp. 1614–1626, 2001.

[145] S. Nordholm and Y.H. Leung, "Performance limits of the broadband generalized sidelobe canceling structure in an isotropic noise field," The Journal of the Acoustical Society of America, vol. 107, no. 2, pp. 1057–1060, 2000.

[146] J. Bitzer, K.U. Simmer and K.D. Kammeyer, "Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement," Proc. ICASSP, pp. 2965–2968, Phoenix, 1999.

[147] J. Bitzer, K. Simmer, and K. Kammeyer, "Multi-channel noise reduction - algorithms and theoretical limits," Proc. EUSIPCO, vol. 1, pp. 105–108, Rhodes, 1998.

[148] J. Bitzer, K.-D. Kammeyer and K. Simmer, “An alternative implementation of the Superdirective beamformer,” IEEE Workshop on Application of Signal Processing to Audio and Acoustics,” New Paltz, NY 1999.

[149] C. Marro, Y. Mahieux and K. Simmer, “Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering,” IEEE Transactions on Speech Audio Processing, vol. 6, no. 3, pp. 240–259, 1998.

[150] S. Nordholm, I. Claesson and M. Dahl, “Adaptive microphone array employing calibration signals: an analytical evaluation,” IEEE Transactions on speech and audio processing, vol. 7, no. 3, pp. 241–252, 1999.

[151] K.C. Huarng and C.C. Yeh, “Performance analysis of derivative constraint adaptive arrays with pointing errors,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 38, no. 2, pp. 209–219, 1990.

[152] S. Nordholm, I. Claesson and P. Eriksson, “The broadband Wiener solution for griffiths-jim beamformers,” IEEE Transactions on Signal Processing, vol. 40, no. 2, pp. 474–478, 1992.

[153] H. Cox, “Resolving power and sensitivity to mismatch of optimum array processors,” The Journal of the acoustical society of America, vol. 54, no. 3, pp. 771–785, 1973.

[154] D. Brandwood, “A complex gradient operator and its application in adaptive array theory,” In IEEE Proceedings H-Microwaves, Optics and Antennas, vol. 130, pp. 11–16, 1983.

[155] G. Strang, “Linear Algebra and its Application,” 2nd edn. (Academic, New York 1980)

[156] R. Crochiere, “A weighted overlap-add method for short-time Fourier analysis/synthesis,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 28, no. 1, pp. 99–102, 1980.

[157] J. Shynk, “Frequency-domain and multirate and adaptive filtering,” IEEE Signal Processing Magazine, vol. 9, no. 1, pp. 14–37, 1992.

[158] S. Nordholm, H.Q. Dam, N. Grbi´c and S.Y. Low, “Adaptive microphone array employing spatial quadratic soft constraints and spectral shaping,” In Speech Enhancement, Signals and Communication Technology, ed. by J. Benesty, S. Makino, J. Chen (Springer, Berlin, Heidelberg 2005), pp. 229–246.

[159] O. Hoshuyama and A. Sugiyama, “Robust adaptive beamforming,” In Microphone Arrays, ed. By M. Brandstein, D. Ward (Springer, Berlin, Heidelberg) pp. 87–109, 2001.

[160] O. Hoshuyama, A. Sugiyama and A. Hirano: "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," IEEE Transactions on Signal Processing, vol. 47, no. 10, pp. 2677–2684, 1999.

[161] S. Doclo and M. Moonen, "On the output SNR of the speech-distortion weighted multi-channel Wiener filter," IEEE Signal Processing Letters, vol. 11, no. 12, pp. 809–811, 2005.

[162] Spriet, M. Moonen and J. Wouters, "Stochastic gradient-based implementation of spatially preprocessed speech distortion weighted multi-channel Wiener filtering for noise reduction in hearing aids," IEEE Transactions on Signal Processing, vol. 53, no. 3, pp. 911–925, 2005.

[163] A.B.Gershman, U.Nickel, and F.Böhme, "Adaptive beamforming algorithms with robustness against jammer motion," IEEE Transactions on Signal Processing, vol. 45, pp. 1878–1885, 1997.

[164] A.B.Gershman, L.Stanković, and V.Katkovnik, "Sensor array signal tracking using a data-driven window approach," Signal Processing, vol. 80, no. 12, Pp. 2507–2515, 2000.

[165] S.A. Vorobyov, A.B.Gershman, Z.-Q.Luo and N.Ma, "Adaptive beamforming with joint robustness against mismatched signal steering vector and interference nonstationarity," IEEE Signal Processing Letters, vol. 2, pp. 108–111, 2004.

[166] S. Markovich, S.Gannot and I.Cohen, "Multi-channel eigenspace beamforming in a noisy reverberant environment with multiple interfering speech signals," IEEE Transactions on Audio Speech and Language Processing, vol. 17, no. 6, pp. 1071–1086, 2009.

[167] J. Li and P.Stoica (Eds.), "Robust Adaptive Beamforming," Telecommunications and Signal Processing, Wiley & Sons Inc., Hoboken, NJ, 2005.

[168] S. Haykin and K.J.R.Liu (Eds.), "Handbook on Array Processing and Sensor Networks," Wiley & Sons, Inc., Hoboken, NJ, 2010.

[169] M. Agrawal and S.Prasad, "Robust adaptive beamforming for wide-band, moving, and coherent jammers via uniform linear arrays," IEEE Transactions on Antennas and Propagation, vol. 47, no. 8, pp. 1267–1275, 1999.

[170] E.D. Di Claudio and R.Parisi, "Robust ML wideband beamforming in reverberant fields," IEEE Transactions on Signal Processing, vol. 51, no. 2, pp. 338–349, 2003.

[171] W.Liu, R.Wu, and R.J.Langley, “Design and analysis of broadband beamspace adaptive arrays,” IEEE Transactions on Antennas and Propagation, vol. ~55, no. 12, pp. 3413–3420, 2007.

[172] S. J. Yu and F. B. Ueng, “Blind adaptive beamforming based on generalized sidelobe canceller,” Signal Processing, vol. 80, no. 12, pp. 2497–2506, 2000.

[173] A. Krueger, E.Warsitz and R. Haeb Umbach, “Speech enhancement with a GSC-like structure employing eigenvector-based transfer function ratios estimation,” IEEE Transactions on Audio, Speech and Language Processing, vol. ~19, pp. 206–219, 2011.

[174] A.H.Sayed, Adaptive Filters, John Wiley&Sons, Hoboken, NJ, 2008.

[175] E. Eweda, “Comparison of RLS, LMS and signal algorithms for tracking randomly time-varying channels,” IEEE Transactions on Signal Processing, vol. 42, no. 11, pp. 2937–2944, 1994.

[176] K. Ozeki and T.Umeda, “An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties,” Electronics and Communications in Japan, vol. 67, pp. 19–27, 1984.

[177] J. Arenas-García, A. R. Figueiras Vidal, A. H. Sayed, “Mean-square performance of a convex combination of two adaptive filters,” IEEE Transactions on Signal Processing, vol. 54, no. 3, pp. 1078–1090, 2006.

[178] S. S. Kozat, A. C. Singer, “Multi-stage adaptive signal processing algorithms,” In Proceedings of the IEEE Sensor Array Multi-channel Signal Processing Workshop, pp. 380–384, Cambridge, MA, 2000.

[179] D.Mandic, P. Vayanos, C. Boukis, B. Jelfs, S. L. Goh, T. Gautama, T. Rutkowski, “Collaborative adaptive learning using hybrid filters,” In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '07), vol. 3, Honolulu, HI, pp. 921–924, 2007.

[180] Y. Zhang, J. A. Chambers, “Convex combination of adaptive filters for a variable tap-length LMS algorithm,” IEEE Signal Processing Letters, vol. 13, no. 10, pp. 628–631, 2006.

[181] N. J. Bershad, J. C. M. Bermudez, J. Tourneret, “An affine combination of two LMS adaptive filters transient mean square analysis,” IEEE Transactions on Signal Processing, vol. 56, no. 5, pp. 1853–1864, 2008.

[182] M. T. M. Silva, V. H. Nascimento, "Improving the tracking capability of Adaptive filters via convex combination," IEEE Transactions on Signal Processing, vol. 56, no. 7, pp. 3137–3149, 2008.

[183] S. S. Kozat, A. T. Erdogan, A. C. Singer, A. H. Sayed, "Steady-state MSE performance analysis of mixture approaches to adaptive filtering," IEEE Transactions on SignalProcessing, vol. 58, no. 8, pp. 4050–4063, 2010.

[184] J. Arenas-García, M. Martínez-Ramón, A. Navia-Vázquez and A. R. Figueiras Vidal, "Plant identification via adaptive combination of transversal filters," Signal Processing, vol. 86, no. 9, pp. 2430–2438, 2006.

[185] J. Arenas-García, V.Gomez-Verdejo, A. R. Figueiras-Vidal, "New algorithms for an improved adaptive convex combination of LMS transversal filters," IEEE Transactions on Instrumentation and Measurement, vol. 54, no. 6, pp. 2239–2249, 2005.

[186] V. H. Nascimento, M. T. M. Silva, L. A. Azpicueta-Ruiz and J. Arenas-García, "On the tracking performance of the combination of least mean squares and recursive least squares adaptive filters," In Proceedings of the IEEE International Conference on Acoustics, Speech, and SignalProcessing(ICASSP '10), Dallas, TX, pp. 3710–3713, 2010.

[187] J. Bitzer and K. U.K. Uwe Simmer, "Superdirective microphone arrays," Microphone Arrays, M. Brandstein and D.Ward, Eds: Springer, pp. 19–38, 2001.

[188] H. Cox, R. Zeskind, and T. Kooij, "Practical super gain," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 34, no. 3 pp. 393–397, 1986.

[189] Mc Cowan, H. Bourlard, "Microphone array post-filter based on noise field coherence," IEEE Transactions on Speech and Audio Processing, vol. 11, no. 6, pp. 709–716, 2003.

[190] I. Cohen, B. Berdugo, "Microphone array post-filtering for nonstationary noise suppression," IEEE international conference on acoustics, speech, and signal processing, vol. 1, pp. 901, 2002.

[191] I. Cohen, "Multi-channel post-filtering in non-stationary noise environments," IEEE Transactions on Signal Processing, vol. 52, no. 5, pp. 1149–1160, 2004.

[192] S. Fischer, K. D. Kammeyer, K. U. Simmer, "Adaptive microphone arrays for speech enhancement in coherent and incoherent noise fields," In Procedding of 3rd joint meeting of the Acoustical Society of America and the Acoustical Society of Japan, Honolulu, Hawaii, 1996.

[193] Cohen, B. Berdugo, “Speech enhancement for non-stationary noise environments,” IEEE Transactions on Signal Processing, vol. 81, no. 11, pp. 2403–2418, 2001.

[194] S. Gannot, I. Cohen, “Speech enhancement based on the general transfer function GSC and postfiltering,” IEEE Transactions on Speech and Audio Processing, vol. 12, no. 6, pp. 561–571, 2004.

[195] J. Li, X. Lu, M. Akagi, “A noise reduction system in arbitrary noise environments and its applications to speech enhancement and speech recognition,” In IEEE International Conference on Acoustics Speech and Signal Processing, pp. 277–280, 2005

[196] J. Li, M. Akagi, “A noise reduction system based on hybrid noise estimation technique and postfiltering in arbitrary noise environments,” Speech Communication, vol. 48, no. 2, pp. 111–126, 2006.

[197] Y. X. Wang and Y. J. Zhang, “Nonnegative Matrix Factorization: A comprehensive review,” IEEE Transactions on Knowledge in Data Engineering, vol. 25, no. 6, pp. 1336–1353, 2013.

[198] C. Févotte, N. Bertin, and J. L. Durrieu, “Nonnegative matrix factorization with the Itakura-Saito divergence: with application to music analysis.,” Neural Computation, vol. 21, no. 3, pp. 793–830, 2009.

[199] P. Smaragdis, “Non-negative matrix factorization for polyphonic music transcription,” IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, no. 1, pp. 177–180, 2003.

[200] B. Raj and P. Smaragdis, “Latent variable decomposition of spectrograms for single-channel speaker separation,” IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 17–20, 2005.

[201] P. Smaragdis, B. Raj, and M. Shashanka, “A probabilistic latent variable model for acoustic modeling,” Advances in Models for Acoustic Processing, NIPS, no. 1, pp. 148, 2006.

[202] P. Smaragdis, B. Raj, and M. Shashanka, “Supervised and semi-supervised separation of sounds from single-channel mixtures,” International Conference on Independent Component Analysis and Signal Separation, Springer, no. 1, pp. 414–421, 2007.

[203] N. J. . Bryan and G. J. Mysore, “An efficient posterior regularized latent variable model for interactive sound source separation,” In. Proceedings of the 30 th International Conference on Machine Learning, Atlanta, Georgia, USA, pp. 208-216, 2013.

[204] B. C. Chung, Hanwook, Eric Plourde, “Regularized NMF-based speech enhancement with spectral components modeled by gaussian mixtures,” IEEE International Workshop on Machine Learning for Signal Processing, 2014.

[205] J. R. H. Févotte Cédric, Jonathan Le Roux, “Non-negative dynamical system with application to speech and audio,” IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 3158–3162, 2013.

[206] Z. D. Zhou, Jun, Shuo Chen, “Rotational reset strategy for online semi-supervised NMF-based speech enhancement for long recordings,” In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2015.

[207] M. Sun, Y. Li, J. F. Gemmeke, and X. Zhang, “Speech enhancement under low SNR conditions via noise estimation using sparse and low-rank NMF with Kullback – Leibler divergence,” IEEE/ACM Transactions on Audio Speech Language Processing, vol. 23, no. 7, pp. 1233–1242, 2015.

[208] Michael Brandstein and Darren Ward, “Microphone arrays: signal processing techniques and applications,” Springer, 2001.

[209] Ricardo Kehrle costaa, João Paulo C.L. da Costaa, and Felix Antreich, “High accuracy and low complexity adaptive generalized sidelobe cancelers for colored noise scenarios,” Digital Signal Processing, Elsevier, vol. 34, pp. 48–55, 2014.

[210] Mamun Ahmed and Nasimul Hyder Maruf Bhuyan, “Comparison of LMS and NLMS algorithm with the using of 4 linear microphone array for speech enhancement,” European Journal of Engineering Research and Science(EJERS), vol. 2, no. 4, pp. 15-19, 2017.

[211] Anna Barnov, Alejandro Cohen, Morag Agmon, Vered Bar Bracha, Shmulik Markovich-Golan, and Sharon Gannot, “A dynamic TF-GSC beamformer for distributed arrays with dual-resolution speech-presence-probability estimators,” International Conference on the Science of Electrical Engineering (ICSEE), pp. 1-5, 2016.

[212] Xianxiang Wang, Yi Zhou, Yongbao Ma, and Qiang Li, “A robust generalized sidelobe canceller controlled by a priori sir estimate,” 3rd International Conference on Multimedia Systems and Signal Processing, (ICMSSP), pp. 84-89, 2018.

[213] Pogula Rakesh, S Siva Priyanka, and Dr. T Kishore Kumar, “Performance evaluation of beamforming techniques for speech enhancement,” Fourth International Conference on Signal Processing, Communication and Networking (ICSCN), Chennai, India, pp. 1-5, 2017.

[214] S. Siva Priyanka. "Review on adaptive beamforming for speech enhancement," Innovations in Power and Advanced Computing Technologies (i-PACT), pp. 1-6, 2017.

[215] Priyanka Gupta, Mukesh Patidar, and Pragya Nema, "Performance analysis of speech enhancement using LMS, NLMS and UNANR algorithms," International Conference on Computer, Communication and Control (IC4), Indore, India, pp. 1-5, 2015.

[216] M. Turki-Hadj Alouane and M. Jaídane-Saídane, "A new nonstationary LMS algorithm for tracking Markovian time-varying systems," Signal Processing, vol. 86, No. 1, pp. 50-70, 2006.

[217] A.H. Sayed, "Fundamentals of Adaptive Filtering," Wiley, 2003.

[218] J. H. Husoy and M. S. E. Abadi, "A comparative study of some simplified RLS type algorithm," Proceedings of International Symposium on Control, Communications, and Signal Processing, pp. 705-708, 2004.

[219] Komal R. Borisagar and Dr. G.R.Kulkarni, "Simulation and comparative analysis of LMS and RLS algorithms using real-time speech input signal," Global Journal of Research in Engineering, Vol. 10, No. 5, pp. 44-47, 2010.

[220] Pogula Rakesh, T. Kishore Kumar, "A novel RLS based adaptive filtering method for speech enhancement," World Academy of Science, Engineering and Technology International Journal of Electronics and Communication Engineering, vol. 9, no. 2, pp. 624-628, 2015.

[221] John H.L, Hansen and Xianxian Zhang "Analysis of CFA-BF: Novel combined fixed/adaptive beamforming for robust speech recognition in real car environments," ELSEVIER Speech Communication, vol. 52, no. 2, pp. 134-149, 2010.

[222] Akila Sayoud, Mohamed Djendi, Soumia Medahi, Abderrezak Guessoum, "A dual fast NLMS adaptive filtering algorithm for blind speech quality enhancement," Elsevier Applied Acoustics, vol. 135, pp. 101–110, 2018.

[223] Cioffi J, Kailath T, "Fast recursive least squares transversal filters for adaptive filtering," IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP, vol. 32, no. 2, pp. 304–337, 1984.

[224] Benallal A and Aezki M, "A fast convergence normalized least mean-square type algorithm for adaptive filtering," International Journal of Adaptive Control Signal Processing, vol. 28, no. 10, pp. 1073-1080, 2013.

[225] Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," The Journal of the Acoustical Society of America, vol. 65, pp. 943–950, 1979.

[226] RIR Generator. Multi-channel Simulation Software https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator, accessed on Jan. 09, 2021.

[227] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, and N. Dahlgren, "The DARPA TIMIT acoustic-phonetic continuous speech corpus," NTIS speech disc, NTIS order number PB91-100354, 1990.

[228] E. H. Rothauser, "IEEE recommended practice for speech quality measurements." IEEE Transactions on Audio and Electroacoustics, vol. 17, no. 3, pp. 225-246, 1969.

[229] http://ecs.utdallas.edu/loizou/speech/noizeus/.

[230] Hu Y. and Loizou P, "Subjective evaluation and comparison of speech enhancement algorithms," Speech Communication, vol. 49, pp. 588-601, 2007.

[231] Y. Hu and P. Loizou, "Evaluation of objective quality measures for speech enhancement," IEEE Transactions on audio, speech, and language processing, vol. 16, no. 1, pp. 229–238, 2008.

[232] I. Cohen, "Multi-channel postfiltering in nonstationary noise environments," IEEE Transactions on Signal Processing, vol. 52, no. 5, pp. 1149–1160, 2004.

[233] Nascimento V. H., Silva M. T. M., Azpicueta-Ruiz L. A., and Arenas-Garcia J, "On the tracking performance of the combination of least mean squares and recursive least squares adaptive filters," IEEE International Conference on Acoustics, Speech, Signal Processing, Dallas, TX, pp. 3710-3013, 2010.

[234] Lu L. and Zhao H., "Novel sign adaption scheme for a convex combination of two adaptive filters." Elsevier International Journal of Electronics and Communication, vol. 69, no. 11, pp. 1590-1598, 2015.

[235] Nascimento, Victor h and Lamare, "A low complexity strategy for speeding up the convergence of convex combinational of adaptive filters." IEEE International Conference on Acoustics, Speech And Signal Processing, pp. 3553-3556, 2012.

[236] Comminiello D, Scarpiniti M, Parisi R, Uncini A, "Combined adaptive beamforming techniques for noise reduction in changing environments," In 36th IEEE International Conference on Telecommunications and Signal Processing, Italy, pp. 690-694, 2013.

[237] Dietzen T, Doclo S and Moonen M, "Joint multi-microphone speech dereverberation and noise reduction using integrated sidelobe cancellation and linear prediction," In 16th IEEE International Workshop on Acoustic Signal Enhancement (IWAENC), Japan, pp.221-225, 2018.

[238] Barnov, Anna, Bracha V. B., Markovich-Golan S, and Gannot S, “Spatially robust GSC beamforming with controlled white noise gain,” In 16th IEEE International Workshop on Acoustic Signal Enhancement (IWAENC), Japan, pp.231-235, (2018).

[239] Kuhl, Stefan, Bohlender A, Schrammen M, and Jax P. “Improved change prediction for combined beamforming and echo cancellation with application to a generalized sidelobe canceller,” Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pp. 363-367, 2019.

[240] Cohen I, Gannot S and Berdugo B, “An integrated real-time beamforming and postfiltering system for non-stationary noise environments,” EURASIP Journal on Applied Signal Processing, vol. 11, pp. 1-10, 2003.

[241] Y.H.Chen and H.D.Fang, “Frequency-domain implementation of Griffiths-Jim adaptive beamformer,” Journal of Acoustic Society of America, vol. 91, no. 6, pp. 3354–3366, 1992.

[242] Tobias Wolff, Markus Buck, “A Practical beamformer-postfilter system for adaptive speech enhancement in non-stationary noise environments,” IEEE Conference on Joint Workshop on Hands-free Speech Communication and Microphone Arrays, pp. 159-160, 2011.

[243] Parinaz Khayeri, Hamid Reza Abutalebi, and Vahid Abootalebi, “A hybrid near field superdirective GSC and post-filter for speech enhancement,” IEEE Conference on 20th European Signal Processing Conference (EUSIPCO), Bucharest, Romania, pp. 2065-2069, 2012.

[244] Junfeng Li and Masato Akagi, “A hybrid microphone array post-filter in a diffuse noise field,” Applied Acoustics ELSEVIER, vol. 69, no. 6, pp. 546–557, 2008.

[245] Cyril Plapous, Claude Marro, and Pascal Scalart, “Improved signal-to-noise ratio estimation for speech enhancement,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, no. 6, pp. 2098 – 2108, 2006.

[246] G.F. Franklin, J. David Powel, and A. Emami-Naeini, “Feedback control of dynamic systems.” Addison-Wesley, Reading, MA, 1994.

[247] H. Zhang, “Speech enhancement using improved adaptive null-forming in the frequency domain with postfilter,” IEICE TRANSACTIONS on Fundamentals of Electronics, Communications, and Computer Sciences, vol. 91, no. 12, pp. 3812–3816, 2008.

[248] K. Li, Kai, Qiang Fu, and Yonghong Yan “Speech enhancement using improved generalized sidelobe canceller in the frequency domain with multi-channel

postfiltering," In 11th Annual Conference of the International Speech Communication and Association, ISCA, 2010, pp. 973–976.

[249] N. Correa, T. Adali, Y. O. Li, and V. D. Calhoun, "Comparison of blind source separation algorithms for FMRI using a new Matlab toolbox: Gift," ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 5, pp. 401–404, 2005.

[250] T. K. A. Cichocki, S. L. Shishkina, T. Musha, Z. Leonowicz and T. Asada, "EEG filtering based on blind source separation (BSS) for early detection of Alzheimer's disease," Clinical Neurophysiology, vol. 116, no. 3, pp. 729–737, 2005.

[251] N. Doukas and N. V. Karadimas, "A blind source separation based cryptography scheme for mobile military communication applications," WSEAS Trans. Commun., vol. 7, no. 12, pp. 1235–1245, 2008.

[252] C. H. Choi, W. Chang, and S. Y. Lee, "Blind source separation of speech and music signals using harmonic frequency-dependent independent vector analysis," Electronics Letters, vol. 48, no. 2, pp. 124-125, 2012.

[253] B. Wang and M. Plumbley, "Investigating single-channel audio source separation methods based on non-negative matrix factorization," In Proc. ICA Research Network International Workshop, pp. 17-20, 2006

[254] A. M. Reddy and B. Raj, "Soft mask methods for single-channel speaker separation," IEEE Transactions on Audio, Speech and Language Processing, vol. 15, no. 6, pp. 1766–1776, 2007.

[255] E. Grais and H. Erdogan, "Single-channel speech music separation using nonnegative matrix factorization and spectral masks," Digital Signal Processing (DSP), pp. 1-6, 2011.

[256] C. Demir, M. Saraçlar, and A. T. Cemgil, "Single-channel speech-music separation for robust ASR with mixture models," IEEE Transactions on Audio, Speech and Language Processing, vol. 21, no. 4, pp. 725–736, 2013.

[257] B. Raj, M. V. S. Shashanka, and P. Smaragdis, "latent dirichlet decomposition for single-channel speaker separation," IEEE International Conference on Acoustics, Speech and Signal Processing Proceedings, vol. 5, pp. 5-5, 2006.

[258] G. J. Mysore and P. Smaragdis, "A non-negative approach to semi-supervised separation of speech from noise with the use of temporal dynamics" In 2011 IEEE

International Conference on Acoustics, Speech, and Signal Processing Proceedings (ICASSP), pp. 17–20, 2011.

[259] M. H. Radfar and R. M. Dansereau, "Single-channel speech separation using soft mask filtering," IEEE Transactions on Audio, Speech and Language Processing, vol. 15, no. 8, pp. 2299–2310, 2007.

[260] Antti Hurmalainen and Tuomas Virtanen, "Modelling spectro-temporal dynamics in factorization-based noise-robust automatic speech recognition," IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings, pp. 4113–4116, 2012.

[261] J. F. Gemmeke, T. Virtanen, and A. Hurmalainen, "Exemplar-based sparse representations for noise-robust automatic speech recognition," IEEE Transactions on Audio, Speech and Language Processing, vol. 19, no. 7, pp. 2067–2080, 2011.

[262] M. N. Schmidt and R. K. Olsson, "Single-channel speech separation using sparse non-negative matrix factorization," Spok. Lang. Process. ISCA International Conference INTERSPEECH, vol. 2, pp. 2–5, 2006.

[263] N. Bertin, R. Badeau, and E. Vincent, "Enforcing harmonicity and smoothness in Bayesian non-negative matrix factorization applied to polyphonic music transcription," IEEE Transactions on Audio, Speech and Language Processing, vol. 18, no. 3, pp. 538–549, 2010.

[264] A. Cichocki, R. Zdunek, and S. Amari, "New algorithms for non-negative matrix factorization in applications to blind source separation," IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings, vol. 5, pp. 621–624, 2006.

[265] T. Virtanen, A. T. Cemgil, and S. Godsill, "Bayesian extensions to non-negative matrix factorization for audio signal modeling," ICASSP, IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings, no. 2, pp. 1825–1828, 2008.

[266] N. Y. Guan, D. C. Tao, Z. G. Luo, and B. Yuan, "Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent," IEEE Transactions on Image Processing, vol. 20, no. 7, pp. 2030–2048, 2011.

[267] A. Lef and F. Bach, "Itakura-Saito nonnegative matrix factorization with group sparsity CNRS / LTCI Statistics team," In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), no. 1, pp. 21-24, 2011.

[268] P. Smaragdis, “Convolutive speech bases and their application to supervised speech separation,” IEEE Transactions on Audio, Speech and Language Processing, vol. 15, no. 1, pp. 1–12, 2007.

[269] M. N. Schmidt and J. L. Larsen, “Reduction of non-stationary noise using a non-negative latent variable decomposition,” In IEEE Workshop on Machine Learning for Signal Processing (MLSP), pp. 486-491, 2008.

[270] K. W. Wilson, B. Raj, and P. Smaragdis, “Regularized non-negative matrix factorization with temporal dependencies for speech denoising,” Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH, pp. 411–414, 2008.

[271] L. Pfeifenberger and F. Pernkopf, “A multi-channel postfilter based on the diffuse noise sound field,” IEEE 22nd European Signal Processing Conference (EUSIPCO), Lisbon, Portugal, pp. 686-690, 2014.

[272] X. Wang, Y. Guo, Q. Fu, and Y. Yan, "Speech enhancement using multi-channel postfiltering with modified signal presence probability in a reverberant environment," Chinese Journal of Electronics, vol. 25, no. 3, pp. 512-519, 2016.

[273] J. Park, W. Kim, D. K. Han, and H. Ko, “Two-microphone generalized sidelobe canceller with postfilter based speech enhancement in composite Noise,” ETRI Journal, vol. 38, pp. 366-375, 2016.

[274] L. Zhang, M. Wang, Q. Zhang, and H. Chen, “Post-Secondary Filtering Improvement of GSC Beamforming Algorithm,” IEEE International Conference on Audio Language and Image Processing (ICALIP), Shanghai, China, pp.338-343, 2018.

[275] G. Ritwik, H.Karim, and Z.Tao, “A novel target speaker-dependent postfiltering approach for multi-channel speech enhancement,” IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, pp. 46-50, 2017.

[276] S. Priyanka and K. Kishore Kumar, “Adaptive beamforming using Zelinski-TSNR multi-channel postfilter for speech enhancement,” IEEE 9th Int. Conference on Computing Communication and Networking Technologies, (ICCCNT), pp. 1-6, 2018.

[277] P. D. O Grady and B. A. Pearlmutter, “Discovering speech phones using convolutive non-negative matrix factorization with a sparseness constraint,” Neurocomputing, vol. 72, no. 1-3, pp. 88–101, 2008.

[278] Fevotte, Cedric, E. Vincent, and A. Ozerov, “Single-channel audio source separation with NMF: divergences, constraints and algorithms,” Audio Source Separation, Springer, pp. 1–24, 2018.

[279] G. F. Harpur and R. W. Prager, “Development of low entropy coding in a recurrent network,” Network: Computation in Neural Systems, vol. 7, pp. 277–284, 1996.

[280] P. O. Hoyer, “Modeling receptive fields with non-negative sparse coding,” in E. D. Schutter (ed.), Computational Neuroscience: Trends in Research 2003, Elsevier, Amsterdam, In press.)

[281] D. J. Field. What is the goal of sensory coding? Neural Computation, 6:559–601, 1994.

[282] Taal, Cees H., Richard C. Hendriks, Richard Heusdens, and Jesper Jensen. "An algorithm for intelligibility prediction of time-frequency weighted noisy speech." IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 7, pp. 2125-2136, 2011.

[283] Vincent, Emmanuel, Rémi Gribonval, and Cédric Févotte. "Performance measurement in blind audio source separation." IEEE transactions On Audio, Speech, And Language Processing, vol. 14, no. 4, pp. 1462-1469, 2006.

List of Publications

International Journals:

[1] S. Siva Priyanka and Prof. T. Kishore Kumar, "**Signed Convex Combination of Fast Convergence Algorithm to Generalized Sidelobe Canceller Beamformer for Multi-Channel Speech Enhancement," Traitement du Signal**, vol. 38, no. 3, pp. 785-795, June 2021, DOI:https://doi.org/10.18280/ts.380325. **(SCI)** [Online].Available: https://www.iieta.org/journals/ts/chapter/10.18280/ts.380325.

[2] S. Siva Priyanka and Prof. T. Kishore Kumar, "**Generalized Sidelobe Canceller Beamforming Using Combined Postfilter and Sparse NMF for Speech Enhancement," Fluctuation and Noise Letters (FNL),** vol. 20, no. 2, 2150014 (2021), DOI: 10.1142/S0219477521500140. **(SCI)** [Online]Available:https://www.worldscientific.com/doi/10.1142/S0219477521500140

International Conferences:

[1] S. Siva Priyanka and Prof. T. Kishore Kumar, **"GSC Beamforming Using Different Adaptive Algorithms for Speech Enhancement,"** in 10th International Conference on Computing, Communication and Networking Technologies [ICCCNT 2019], IIT Kanpur, July 6-8, 2019. (**IEEE**)

[2] S. Siva Priyanka and Prof. T. Kishore Kumar, **"GSC Adaptive Beamforming Using Fast NLMS Algorithm for Speech Enhancement,"** in 3rd International Conference on Imaging, Signal Processing, and Communication (ICISPC 2019), NTU Singapore, July 27-29, 2019. (**IEEE**)

[3] S. Siva Priyanka and Prof. T. Kishore Kumar, "**Adaptive Beamforming using Zelinski-TSNR Multi-channel Postfilter for Speech Enhancement**," 9th International Conference on Computing, Communication and Networking Technologies [ICCCNT 2018], IISC Bangalore, July 10-12, 2018. (**IEEE**)