

# DESIGN AND EVALUATION OF ENHANCEMENT TECHNIQUES FOR SINGLE-CHANNEL SPEECH

## A THESIS

*Submitted in partial fulfilment of the  
requirements for the award of the degree*

*of*

DOCTOR OF PHILOSOPHY

*in*

ELECTRICAL ENGINEERING

by  
**SACHIN SINGH**



DEPARTMENT OF ELECTRICAL ENGINEERING  
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE  
ROORKEE-247 667 (INDIA)

APRIL, 2015

©INDIAN INSTITUTE OF TECHNOLOGY ROORKEE, ROORKEE-2015  
ALL RIGHTS RESERVED





# INDIAN INSTITUTE OF TECHNOLOGY ROORKEE ROORKEE

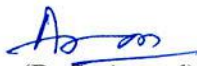
## CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in this thesis entitled "DESIGN AND EVALUATION OF ENHANCEMENT TECHNIQUES FOR SINGLE-CHANNEL SPEECH" in partial fulfilment of the requirements for the award of the Degree of Doctor of Philosophy and submitted in the Department of Electrical Engineering of the Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out during a period from July, 2011 to April, 2015 under the supervision of Dr. Manoj Tripathy, Assistant Professor and Dr. R. S. Anand, Professor, Department of Electrical Engineering, Indian Institute of Technology Roorkee, Roorkee.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other Institute.

  
(SACHIN SINGH)

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.

  
(R. S. Anand)  
Supervisor

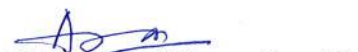
  
(Manoj Tripathy)  
Supervisor

The Ph.D. Viva-Voce Examination of **Mr. SACHIN SINGH**, Research Scholar, has been held on..26.10.2015

  
Chairman, SRC 26.10.15

 26/10/15  
Signature of External Examiner

This is to certify that the student has been made all the corrections in the thesis.

  
Signature of Supervisor (s)

 26/10  
Head of the Department

Date: 26/10/2015

## ABSTRACT

---

In real-world applications, a speech signal from the uncontrolled environment is often accompanied by various degradation components along with the actual speech components. The degradation components include background noise, reverberation and multi-talker speech. These unwanted interferences not only degrade perceptual speech quality and intelligibility which creates listening problem for human, but also give poor performance in automatic speech processing tasks like speech recognition, speaker recognition and hearing aid systems. Therefore, de-noising of corrupted single-channel speech has become a very necessary and important aspect for research in academia and industry.

The presently available single channel noise reduction methods include spectral subtraction, Wiener filter, minimum mean square error estimation (MMSE) and p-MMSE, log-MMSE, KLT, PKLT etc. These methods are applicable for specific environment of speech signal. Some of these perform better for one particular types of noise whereas others are suitable for other types of noise. Considering the limitations of these methods, different categories of speech signals have been treated separately. Based on this, the objectives of the present research work have been formulated as: (1) design of a suitable method for enhancement of mixed noisy speech of very low (Negative) input SNR conditions; (2) design and development of a suitable method for suppression of non-stationary noise in single-channel speech signal; (3) analysis and development of a suitable method for suppression of combined effect of background noise and reverberation; and (4) design and implementation of phase based single-channel speech enhancement technique. The mentioned objectives have been accomplished as follows:

In the first objective, single-channel speech enhancement based on modified Wiener gain function using Wavelet Packet Transform (WPT) is proposed for suppression of noise from multiple sources in both the low (negative) and high SNR speech signal ranging from -15 dB to +15 dB. The method includes steps as (1) decomposition of speech signal upto 3<sup>rd</sup> level to get speech signal in eight different bands; (2) the FFT of these bands is computed to get the wavelet packet soft threshold which is applied on the above FFT output; (3) the WP soft threshold is also used to determine the modified gain function; (4) finally to get the processed output speech, the IFFT of the product of the modified Wiener gain function and WP thresholded FFT output is computed. The overlap-add method is used to get the end reconstructed speech signal.

The performance of this proposed method is compared with other existing speech enhancement methods evaluating their performance parameters such as MSE, SNR, MOS, PESQ and SII. The dataset of low SNR ranging from -15 dB to +15 dB having mixed noise is used for performance evaluation of the implemented methods. The results show the improvement in terms of speech quality and intelligibility parameters. Proposed method gives highest improvement in comparison to other single-channel speech enhancement methods for all input SNR levels with various noise types.

To overcome the problem of using true speech or true noise in binary mask based methods of speech enhancement, a fuzzy mask is proposed here under second objective. It is based on soft and hard wavelet packet threshold. The method includes steps as (1) decomposition of speech signal upto 3<sup>rd</sup> level to get speech signal in eight different bands; (2) the FFT of these bands is computed to get the wavelet packet soft and hard threshold which is applied on the FFT output; (3) in this procedure, the modified Wiener gain function determined similarly as above is applied to get the denoised speech signal in frequency domain at first stage; (4) in second stage, fuzzy mask is applied on the output of first stage for further enhancement; (5) finally to get the processed output speech, the IFFT of the product of the fuzzy mask and WP soft and hard thresholded FFT output is computed. Again, the overlap-add method is used to get the end reconstructed speech signal.

Here again, the performance of this proposed method is compared with other existing speech enhancement methods comparing their performance parameters such as SNR, MOS, PESQ and STOI. The dataset of low SNR ranging from -15 dB to +15 dB having non-stationary noise is used for performance evaluation of the implemented methods. The results obtained from proposed method are much better than other existing single-channel speech enhancement methods.

Most of the above implemented algorithms are used for speech enhancement of noise and reverberation separately and they do not work effectively in case of their combination (i.e. reverberation with noise). To suppress the combined effect of early and late reverberations with various types of noise, a binary reverberation mask is implemented here for the fulfillment of the third objective.

In this proposed method signal-to-reverberant ratio (SRR) is calculated as a limit for ideal reverberant mask (IRM). The amplitudes with SRR greater than a preset threshold (i.e. -5dB) are used for reconstruction of dereverberated speech, while amplitudes with SRR values smaller than the threshold are eliminated. The construction of the SRR criterion assumes *a priori* knowledge of the input reverberant and target signal. Threshold values varying from

0dB to -90dB are analyzed for selection of IRM limit  $T$ . Finally, the dereverberated speech signal is constructed by multiplying noisy speech with reverberant mask.

The proposed reverberant mask based speech enhancement method is compared with other existing speech enhancement methods in terms of speech quality and intelligibility measure parameters such as PESQ, CD, SNR and MSE. The maximum improvement in reverberated noisy speech is obtained by proposed method in terms of speech quality and intelligibility at all input SNR levels ranging from -25 dB to -5 dB.

Most of the noise reduction algorithms perform the modification in amplitude only, while phase remains unchanged or discarded in the process of speech enhancement. Recently, it has been found that quality and intelligibility both can be improved upto a significant level by using either phase of speech signal only or phase with amplitude. Hence, signal phase ratio based single-channel speech enhancement method is implemented in fourth objective for further improvement in noisy speech signal which considers the phase of the noisy speech signal in processing. The phase ratio of noisy speech to noise signal is used in the phase based method. In this method two gain functions  $G_1$  and  $G_2$  are developed for correction in noisy phase by suppressing the noise coming from angles between  $0$  to  $\pm\pi/2$  and  $\pm\pi/2$  to  $\pm\pi$ , respectively. For the reconstruction of speech spectrum, both gains are multiplied together and lower values of the phases are neglected for getting desired speech spectrum. Results are compared with other phase based methods (such as phase spectrum compensation (PSC), exploiting conjugate symmetry of the short-time Fourier spectrum and STFT-phase for the MMSE-optimal) and are analyzed in terms of speech quality, intelligibility measures (like SNR, SSSNR, SIG, SII, BAK, OVL, and PESQ, etc.), informal subjective listening tests and spectrogram analysis. The performance measure parameters show that the proposed phase ratio based implemented method provides more effective improvement in noisy speech in comparison to other phase based speech enhancement methods.

Implemented algorithms are evaluated for various languages i.e. Hindi, Kannada, Bengali, Malayalam, Tamil, Telgu, and English. Indian language database used for evaluation are taken from IIIT-H Indic Speech Databases which was developed at Speech and Vision Lab, IIIT-Hyderabad for the purpose of building speech synthesis system among Indian languages. The speech data were recorded by native speakers of each language. The recording was done in a studio environment using a standard headset microphone connected to a Zoom handy recorder. A set of 1000 sentences were selected for each language. These sentences were selected to cover 5000 most frequent words in text corpus of the corresponding language. The NOIZEUS database of clean and noisy speech was used for English language sentences. This

database basically contains 30 IEEE sentences which were produced by three male and three female speakers in groups. The real-world sources of background noise at different SNRs were taken from AURORA and NOISEX-92 databases, respectively which include suburban train noise, babble, car, exhibition hall, restaurant, street, airport and train-station as noise sources.

In the nut shell, it can be said that the present work is an effort to determine suitability of various single-channel speech enhancement techniques to get the maximum speech quality and intelligibility.



## ACKNOWLEDGEMENTS

---

At the first place, I wish to bow before the almighty, which created this beautiful world and blessed us with wisdom to be able to appreciate his creation. I am highly grateful to the almighty for giving me courage and strength to help the suffering.

I wish to express my sincere gratitude to my guide Dr. Manoj Tripathy and Dr. R. S. Anand, Department of Electrical Engineering, Indian Institute of Technology Roorkee. They have been a constant source of inspiration and encouragement throughout this work. I am indeed fortunate to have them as guide and advisor. But for their interest and motivation, it is unlikely that this thesis could have taken shape. I have benefited immensely through their association. I sincerely thank them for all the help they have rendered.

I express my gratitude toward Dr. G. N. Pillai, Chairman, Departmental Research Committee and Dr. Vinod Kumar, Chairman, Student Research Committee, Department of Electrical Engineering, Dr. Manoj Mishra, External Expert, Department of Electronics and Computer Engineering, Dr. R. P. Maheshwari, Internal Expert, Department of Electrical Engineering, Indian Institute of Technology, Roorkee for being members of my research committee and sparing their valuable time in reviewing and critically examining the work.

My sincere thanks goes to my fellow researchers Mr. Pratul Arvind, Mr. Bhavik Patel, Mr. Nagashetty P. Biradar, Mr. Arvind Yadav, Mr. Arun Balodi, Mr. Jayendra Kumar, Mr. Roshan Kumar and Lab instructor Mr. Jogeshwar Prasad, who have spared their valuable time and valuable support.

My special thanks goes to my all lab mates and all other fellow researchers. A special thanks goes to Dr. Subrahmanyam Murala, Assistant Professor, Department of Electrical Engineering, IIT Ropar, for their friendly company and direct and indirect help provided by them throughout my research span.

I wish to thank my office superintendent and other office staff members who always been helpful during my work. I also wish to thank Instrumentation and Signal Processing lab supporting staff, Mr. Jogeshwar Prasad, Mr. Rajiv Gupta and Mr. Veer Chand Ji for all the necessary help provided by them.

I extend special thanks to my inmates in G. P. hostel and other hostels including Mr. Sanjiv Saxena, Mr. Yateendra Kumar, Mr. Manmohan, Mr. Bhavik Patel and their families for the nicest company provided by them during my family stay in the campus.

I am thankful to my parents, brothers and in-laws for their consistent support, patience and encouragement throughout my stay in Roorkee. Last but not the least I am thankful to my loving son OM (Kshitij Singh) and daughter SHIVI (Anvi Singh) who always been



supportive for carrying on my work. Words can hardly explain the strength and support provided by my wife Dr. Shobha Rani Katiyar, who always motivated and encouraged me in every hard walk of my work.

Last but not the least; I am thankful to all the people whose names are not blazing through my limited memory at this moment but, who directly or indirectly extended their whole-hearted help, encouragement and support in this research work.

(Sachin Singh)



## CONTENTS

<b>ABSTRACT.....</b>	<b>I</b>
<b>ACKNOWLEDGEMENTS.....</b>	<b>V</b>
<b>CONTENTS.....</b>	<b>VII</b>
<b>LIST OF FIGURES.....</b>	<b>X</b>
<b>LIST OF TABLES.....</b>	<b>XII</b>
<b>LIST OF ACRONYMS .....</b>	<b>XIII</b>
<b>LIST OF SYMBOLS .....</b>	<b>XIV</b>
<b>CHAPTER 1:INTRODUCTION .....</b>	<b>1</b>
1.1 Overview .....	1
1.2 Single-Channel Speech Enhancement: State of the Art.....	4
1.3 Performance Evaluation Parameters for Speech Enhancement Methods .....	9
1.3.1 Perceptive evaluation of speech quality (PESQ) .....	9
1.3.2 Mean opinion score (MOS) .....	10
1.3.3 Short-time objective intelligibility (STOI).....	11
1.3.4 Speech intelligibility index (SII).....	11
1.3.5 Distortion rating measures .....	12
1.3.6 Mean square error (MSE) .....	13
1.3.7 Signal -to- noise ratio (SNR) .....	13
1.3.8 Segmental SNR (SNRseg).....	13
1.3.9 Frequency weighted segmental SNR (fw-SNRseg).....	14
1.3.10 Cepstrum distance (CD) .....	14
1.3.11 Weighted spectral slop (WSS) metric .....	15
1.3.12 Itakura saito (IS).....	15
1.4 Objectives of the Present Work .....	16
1.5 Structure of the Thesis.....	17
<b>CHAPTER 2:SUPPRESSION OF MIXED NOISE .....</b>	<b>19</b>
2.1 Overview .....	19
2.2 Single-Channel Speech Enhancement Methods.....	20
2.2.1 Spectral subtraction method .....	20
2.2.2 Amplitude spectral subtraction method.....	20
2.2.3 Power spectral subtraction method .....	21

2.2.4	Multi-band spectral subtraction method.....	21
2.2.5	Scalart power spectral subtraction method.....	22
2.2.6	Parametric spectral subtraction method.....	22
2.2.7	Reduced delay convolution - spectral subtraction method.....	22
2.2.8	Wiener filtering method .....	22
2.2.9	MMSE-speech presence uncertainty (MMSE-SPU) estimation method .....	23
2.2.10	log-MMSE method.....	23
2.2.11	p-MMSE method.....	24
2.2.12	Cohen-MMSE method .....	24
2.2.13	KLT subspace method.....	24
2.2.14	PKLT subspace method.....	24
2.3	Proposed Wavelet Packet Transform Based Modified Wiener Gain Method.....	25
2.3.1	Mother wavelets and decomposition levels.....	27
2.3.2	Determination of soft WP threshold function.....	42
2.3.3	Gain functions and effects .....	42
2.3.4	Results and discussion.....	44
2.4	Summary .....	60
<b>CHAPTER 3: SUPPRESSION OF HIGHLY NON-STATIONARY NOISE .....</b>		<b>61</b>
3.1	Overview .....	61
3.2	Enhancement Techniques for Highly Non-Stationary Noise Environments.....	61
3.3	WPT Fuzzy Mask Based Speech Enhancement Method .....	64
3.3.1	Modified Wiener gain function for noise suppression.....	64
3.3.2	Fuzzy mask function .....	65
3.3.3	Results and discussion.....	68
3.4	Summary .....	81
<b>CHAPTER 4: COMBINED SUPPRESSION OF NOISE AND REVERBERATION .....</b>		<b>83</b>
4.1	Overview .....	83
4.2	Reverberation Suppression.....	84
4.2.1	Short-term dereverberation.....	84
4.2.2	Long-term dereverberation .....	85
4.3	Reverberation Modelling.....	86
4.3.1	Time domain .....	86
4.3.2	Statistical domain.....	87

4.4	Reverberant Mask Based Method .....	88
4.4.1	Database.....	90
4.4.2	Signal model .....	90
4.4.3	Reverberant mask calculation.....	91
4.4.4	Results and discussion.....	92
4.5	Summary .....	101
<b>CHAPTER 5:PHASE BASED SPEECH ENHANCEMENT .....</b>		<b>103</b>
5.1	Overview .....	103
5.2	Speech Enhancement Methods Using Phase .....	104
5.2.1	Phase spectrum compensation (PSC).....	104
5.2.2	Exploiting conjugate symmetry of short-time Fourier spectrum.....	105
5.2.3	MMSE-optimal spectral amplitude estimation based on STFT-phase.....	106
5.3	Proposed Phase Ratio Based Single-Channel Speech Enhancement Method.....	107
5.3.1	Signal model and notation .....	107
5.3.2	Signal phase ratio based approach .....	108
5.3.3	Results and discussion.....	110
5.4	Summary .....	125
<b>CHAPTER 6:CONCLUSIONS AND SCOPE FOR FUTURE WORK .....</b>		<b>127</b>
6.1	Conclusions .....	127
6.1.1	WPT based modified Wiener gain method.....	127
6.1.2	WPT based Fuzzy mask method.....	128
6.1.3	Reverberant mask based method.....	129
6.1.4	Signal phase ratio based method.....	130
6.2	Scope for Future Work.....	131
<b>PUBLICATIONS FROM THE WORK .....</b>		<b>133</b>
<b>REFERENCES.. .....</b>		<b>135</b>
<b>APPENDIX-A.....</b>		<b>155</b>
<b>APPENDIX-B.....</b>		<b>160</b>

## LIST OF FIGURES

Figure No.	Caption	Page No.
1.1	Basic overview for addition of noise with speech signal	2
1.2	Overview of any common speech enhancement system	3
2.1	Block diagram of proposed WPT modified Wiener gain based speech enhancement method	26
2.2	Waveforms of various wavelet function of wavelet family	29
2.3	Wavelet decomposition at 3 <sup>rd</sup> level for the input speech signals	30
2.4	Wavelet packet decomposition at 3 <sup>rd</sup> level for the input speech signals	30
2.5	Block diagram of the WPT based speech enhancement method	31
2.6	Variation of parameters with Input SNR in babble noise	32
2.7	Comparison of mother wavelets in babble noise	33
2.8	Variation of parameters with Input SNR in pink noise	34
2.9	Comparison of mother wavelets in pink noise	35
2.10	Variation of parameters with Input SNR in Volvo car noise	37
2.11	Comparison of mother wavelets in Volvo car noise	38
2.12	Variation of parameters with Input SNR in white noise	39
2.13	Comparison of mother wavelets in white noise	40
2.14	Variation of output SNR with Input SNR in mixed noise	47
2.15	Comparison of speech enhancement methods in terms of Output SNR	48
2.16	Variation of fw-SSNR with Input SNR in mixed noise	50
2.17	Comparison of speech enhancement methods in terms of fw-SSNR	51
2.18	Variation of PESQ with Input SNR in mixed noise	53
2.19	Comparison of speech enhancement methods in terms of PESQ values	54
2.20	Variation of Cepstrum Distance with Input SNR in mixed noise	56
2.21	Comparison of speech enhancement methods in terms of Cepstrum Distance	57
2.22	Variation of SII values with different speech enhancement methods and mixed noise	58
2.23	Shows comparative time-frequency spectrograms of (a) clean (b) noisy speech at 5dB (c) proposed method (d) log-MMSE (e) Wiener (f) SS (g) p-MMSE (h) IDBM (i) ICS (j) hard and (k) Soft wavelet thresholding method for Hindi speech pattern “ <i>apke hindi pasand karne par khushi hui</i> ”	60

3.1	Block diagram of signal-processing stages used in single-channel speech enhancement	63
3.2	Block diagram of proposed method for speech enhancement	66
3.3	Output SNR Scores in presence (a) babble (b) pink (c) f-16 (d) white noise	70
3.4	Representation of results in bar chart for output SNR	71
3.5	PESQ Scores in presence (a) babble (b) pink (c) f-16 (d) white noise case	73
3.6	Representation of results in bar chart for PESQ scores	74
3.7	MOS Scores in presence (a) babble (b) pink (c) f-16 (d) white noise case	76
3.8	Representation of results in bar chart for MOS scores	77
3.9	Output STOI Scores in presence (a) babble (b) pink (c) f-16 (d) white noise	79
3.10	Representation of results in bar chart for STOI scores	80
4.1	Waveform of a representative room impulse response	87
4.2	Block diagram of proposed method for speech enhancement	89
4.3	Variation of SNR for various speech enhancement algorithms	93
4.4	Variation of MSE for various speech enhancement algorithms	93
4.5	Variation of results between PESQ scores and input SNR	96
4.6	Variation of results between PESQ scores	97
4.7	Variation of results between CD scores and input SNR at reverberation+ noise	99
4.8	Variation of results between CD scores	100
5.1	Block diagram of phase ratio based proposed speech enhancement method	109
5.2	Variation of PESQ scores with input SNR at different noise types	112
5.3	Bar chart representation of PESQ scores at various input SNR levels	113
5.4	Variation of fw-SSNR scores with input SNR at different noise types	116
5.5	Bar chart representation of fw-SSNR scores at various input SNR levels	117
5.6	Variation of WSS scores with input SNR at different noise	119
5.7	Bar chart representation of WSS scores at various input SNR levels	120
5.8	Variation of OVL scores with input SNR at different noise	122
5.9	Bar chart representation of OVL scores at various input SNR levels	123
5.10	Spectrograms of (a) clean speech, (b) noisy speech (babble noise at -5 dB input SNR), and enhanced output signal given by speech enhancement methods such as (c) MMSE-Phase, (d) KLT, (e) PKLT, (f) spectral subtraction, (g) PSC, (h) MMSE, (i) Conjugate Symmetry, (j) proposed method.	125

## LIST OF TABLES

Table No.	Caption	Page No.
1.1	Rating scale for MOS	10
1.2	Rating scale for BAK, SIG and OVL measures	12
2.1	Analysis for babble noise	36
2.2	Analysis for pink noise	36
2.3	Analysis for Volvo car noise	41
2.4	Analysis for white noise	41
2.5	Results for gain functions	44
2.6	Output SNR values	46
2.7	fw-SSNR values	49
2.8	PESQ values	52
2.9	Cepstrum Distance measure values	55
2.10	Speech Intelligibility Index (SII) values	58
3.1	Output SNR score in presence of various noise types	69
3.2	Output PESQ score in presence of various noise types	72
3.3	MOS score in presence of various noise types	75
3.4	Output STOI score in presence of various noise types	78
4.1	Output SNR values with variation of input noise SNR and threshold values	94
4.2	Output MSE values with variation of input noise SNR and threshold values	94
4.3	PESQ measures values in reverberation condition	95
4.4	Cepstrum Distance values in reverberation condition	98
5.2	Performance Measure in terms of PESQ	111
5.3	Performance Measure in terms of fw-SSNR	115
5.4	Performance Measure in terms of WSS	118
5.5	Performance Measure in terms of OVL	121

## LIST OF TABLES

Table No.	Caption	Page No.
1.1	Rating scale for MOS	10
1.2	Rating scale for BAK, SIG and OVL measures	12
2.1	Analysis for babble noise	36
2.2	Analysis for pink noise	36
2.3	Analysis for Volvo car noise	41
2.4	Analysis for white noise	41
2.5	Results for gain functions	44
2.6	Output SNR values	46
2.7	fw-SSNR values	49
2.8	PESQ values	52
2.9	Cepstrum Distance measure values	55
2.10	Speech Intelligibility Index (SII) values	58
3.1	Output SNR score in presence of various noise types	69
3.2	Output PESQ score in presence of various noise types	72
3.3	MOS score in presence of various noise types	75
3.4	Output STOI score in presence of various noise types	78
4.1	Output SNR values with variation of input noise SNR and threshold values	94
4.2	Output MSE values with variation of input noise SNR and threshold values	94
4.3	PESQ measures values in reverberation condition	95
4.4	Cepstrum Distance values in reverberation condition	98
5.1	Performance Measure in terms of PESQ	111
5.2	Performance Measure in terms of fw-SSNR	115
5.3	Performance Measure in terms of WSS	118
5.4	Performance Measure in terms of OVL	121



## LIST OF SYMBOLS

---

$d_{sym}$	Average symmetrical disturbance
$d_{asym}$	Average asymmetrical disturbance
$D_{EST}$	Estimated noise spectrum
$X(n)$	Clean speech signal
$D(n)$	Noise signal
$Y(n)$	Noisy speech signal
$\hat{X}(n)$	Denoised or enhanced speech signal
$Y(n, k)$	Noisy speech signal in frequency domain
$D(n, k)$	Noise signal in frequency domain
$X(n, k)$	Clean speech signal in frequency domain
$\hat{X}(n, k)$	Denoised or enhanced speech signal in frequency domain
$n$	Number of frame
$k$	Number of frequency band
$G$	Gain function
$e$	Error signal
$Z(n)$	Reverberated speech signal
$T_{60}=1.0s$	Reverberation time
$X_1(n, k)$	Output speech signal in first stage
$P$	Quality of the reverberated speech signal
$P_{max}$	Maximum possible preference
$h_k$	Habets reverberation generation model
$\Delta$	Exponentially decaying parameter

*This chapter gives an overview of single-channel speech enhancement techniques and performance evaluation parameters. The sources of background noise that degrade the speech quality and intelligibility of speech signal are also introduced. State of the art for the single-channel speech enhancement is then discussed. The outline of thesis is also described at the end of this chapter.*

## **1.1 Overview**

Speech is the most desirable form of communication among human beings. The communication between speaker and listener is very easy and accurate if they communicate in a quiet environment. However, the information may be lost or degraded if the speaker and listener are at a distance and in noisy environment. In a highly non-stationary noise environment, a speech signal may be degraded upto that level from where the recognition of desired speech is very difficult. Hence, an effective algorithm is required to reduce the background noise and distortions for getting the desired speech signal. The process of background noise suppression from noisy speech signal is called the “*speech enhancement*” [1]. Research on speech enhancement was started with two patents given by Schroeder which can be traced back to 40-45 years ago [2]. Schroeder gave the analog implementation of spectral magnitude subtraction. Since then, the noise reduction or suppression of uncorrelated additive background noise has become an area of active research [3-4]. Over the years, researchers have presented a number of algorithms for the enhancement of noisy speech signal. Yet, due to complexities in speech signal and presence of highly non-stationary background noise poses a considerable challenge in speech signal enhancement for communication systems.

Speech enhancement methods can be classified into two main groups i.e. single and multi channel. For single-channel speech enhancement methods, only one microphone is used for signal input and it is multi-channel, if input is taken through more than one microphone. In real-time applications like automatic speech and speaker recognition, teleconferencing systems, mobile communication and hearing aids etc, normally single-channel signal is used. Since, single-channel speech enhancement methods have only one sensor for the observation of degraded speech and do not exploit any spatial information for the received noisy speech signal. It is difficult to design an efficient single-channel speech enhancement method which improves both quality and intelligibility of noisy speech signal [3]. The main advantage of designing a single-channel method is that they are simple and less expensive than multi-

channel speech enhancement systems. The different properties of speech and uncorrelated background noise are used in single-channel speech enhancement methods. The performance of these methods usually depends on level of background noise distortion in noisy speech. Since these methods assume the noise to be stationary signal during speech intervals, their performance is not satisfactory when applied in highly non-stationary noise environments.

A general schematic illustrating addition of background noise is shown in Fig. 1.1, where noisy speech signals are generated by combination of surrounding noise and clean speech signals. The speech enhancement algorithms are used for suppression/reduction of the additive surrounding noise.

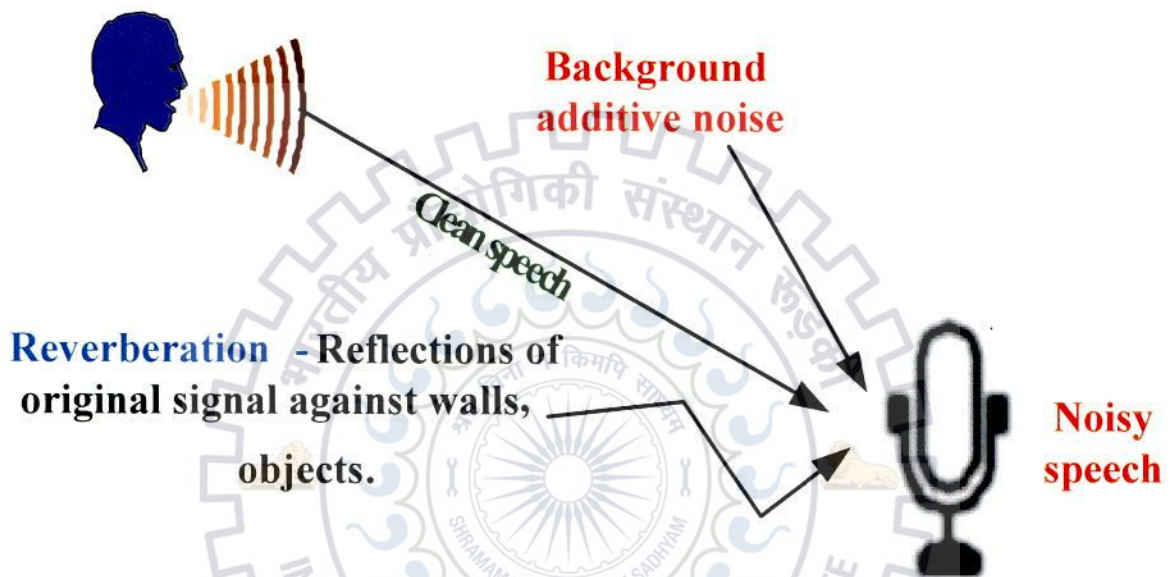


Fig. 1.1: Basic overview for addition of noise with speech signal.

Noise can be defined as an unwanted signal and there are many forms of noise. One of the most common sources of noise is the background noise, which is always present in different degrees in any location apart from a soundproof room. Operating a hands-free mobile phone in a car can be affected by at least three types of background noise, namely wind, road as well as engine noise. Other examples of noisy speech inputs are found in pay phones in noisy environments such as food courts and bus terminals, voice communication systems in cockpits, cellular phones in machine rooms, etc. A second source of noise is channel noise which affects both digital and analogue transmissions and therefore degrades the resulting speech at the receiver end. A third type of noise is quantization noise which results from an over compression of speech signals. Other noise types include, but are not limited to, competing speakers as well as echoes and reverberations which are delayed versions of a speech signal. Different types of noise require different noise models and their own unique set of solutions.

The general schematic of a speech enhancement system is shown in Fig. 1.2. Most of the speech enhancement algorithms use noisy speech phase straightway with modified speech magnitude. Here, noisy speech is first divided into short speech frames by using windowing and then its Discrete Fourier Transform (DFT) is computed. The noise is estimated next and then it is subtracted from the noisy speech. Now, inverse-DFT is computed using earlier phase. Finally, denoised (desired) speech is reconstructed using overlap-add method.

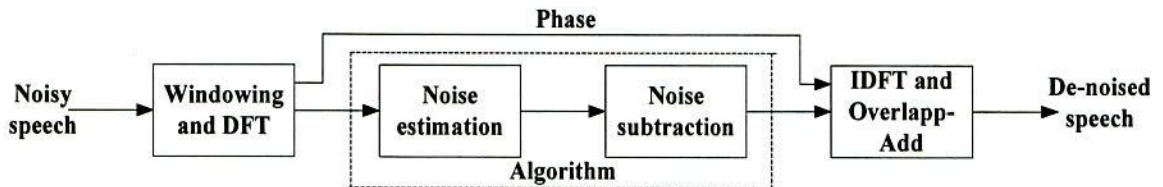


Fig. 1.2: Overview of any common speech enhancement system.

The speech production process starts by a series of muscular movements of the vocal tract where vocal tract is excited by the puffs of air released from the lungs. The excitation given to vocal tract can be classified into two groups i.e. impulse train generator and random noise generator [12]. These two generators are responsible for giving voiced and unvoiced speech, respectively. The voiced region has higher signal energy in comparison to unvoiced region i.e. high and low signal-to-noise ratio (SNR) regions, respectively. This SNR difference between regions plays a crucial role in perception [13]. The most significant regions are around the instants of glottal closure which are perceptually much significant [14-15]. There is big difference in speech characteristics at high and low signal-to-noise ratio (SNR) levels. The variations of speech characteristics with respect to time i.e. non-stationary and highly non-stationary, is also important for speech enhancement. Since speech is a more complicated and highly non-stationary signal, the perceptual aspects are usually not considered in presence of noise at low input SNR level [16].

With the ever increasing power and falling cost of digital signal processors and the availability of cheap memory chips, the use of speech processing systems for voice communication and recognition tasks is becoming more and more common. One outstanding example of a voice communication product is the cellular radio telephony system. Numerous examples of voice recognition products include hands-free input systems for voice dialling, voice activated security systems, automatic speech and speaker recognition, hearing aid devices for impaired people, etc. speaking is arguably a more natural way of communicating with a machine than typing. It is more efficient and faster if the recognition is accurate. As the presence of noise significantly degrades the performance of speech coders and voice

recognition systems, it is therefore imperative to incorporate single-channel speech enhancement as a pre-processing step in these systems. Since most of the single-channel speech enhancement methods are based on modification in speech amplitude only and unaltered or noisy speech phase is used at time of reconstruction of processed speech.

## 1.2 Single-Channel Speech Enhancement: State of the Art

The active research in speech enhancement can be traced back to 1950s. Initial traces of the work appeared in the year 1949 when N. Wiener proposed study of extrapolation, interpolation and smoothing of stationary time series signals in the Cambridge, MIT Press. It is known as Wiener filtering or linear MMSE [236].

The theme of single-channel speech enhancement came into existence, way back in late 70s. Pioneering efforts in this regard were made by J. S. Lim [32, 36] in 1978 and M. R. Sambur [56] introduced the adaptive noise cancelling for speech signals in the IEEE Transaction of acoustic speech and signal processing. In 1979, J. S. Lim proposed a method again for speech enhancement and bandwidth compression of noisy speech signals [16]. The spectral subtraction method of speech enhancement was introduced in 1979 by S. F. Boll [17] and remains one of the most widely used ways of reducing additive noise. A gain function was introduced by M. Berouti et al. in 1979 [18] to overcome the problem of residual broadband noise after processing in spectral subtraction. McAulay and Malpass in 1980 [239] observed that spectral subtraction performs poorly when there is no speech present and they introduced a two-state model for speech presence. Using a Gaussian model for the noise, they derived an expression for the probability of speech presence based on the true (or “a priori”) SNR and the ratio of noisy-signal to noise power (the “a posteriori” SNR). The efforts in development of enhancement system for speech processing made the pace in the era of 80s when Ephraim and Malah proposed an optimal MMSE estimation of the short-time spectral amplitude (STSA). Its structure is the same as that of spectral subtraction but, in contrast to the Wiener filtering motivation of spectral subtraction, it optimizes the estimate of the real amplitude rather than complex spectral amplitudes. Central to their procedures is the estimate of SNR in each frequency bin for which they proposed two algorithms: a maximum likelihood approach and a “decision directed” approach which they found to perform better. The maximum likelihood (ML) approach estimates the SNR (or “a priori” SNR) by subtracting unity from the low-pass filtered ratio of noisy -signal to noise power (the “a posteriori” or “instantaneous” SNR) and half-wave rectifying the result so that it is non-negative. The decision-directed approach forms the SNR estimate by taking a weighted average of this ML

estimate and an estimate of the previous frame's SNR determined from the enhanced speech. The weights used were 0.02 and 0.98 respectively. Both algorithms assume that the mean noise power spectrum is known in advance [22]. Subsequently in 1985, Ephraim and Malah [23] introduced an improved version of their procedure which minimized the mean square error of the log-spectrum, rather than that of the power spectrum itself. They reported that this gave noticeably lower background noise levels without introducing additional distortion.

Later in 1989 Ephraim et al. [240], put forward an HMM to model the speech but represented each state's output distribution using a mixture of LPC spectra rather than the conventional MFCC coefficients. It presented an approximate method of training the states which is supplemented by an exact method that gives an improvement in performance. G. S. Mallat discussed wavelet representation based theory for multi-resolution of speech signal decomposition [142]. In 1992, P. Lockwood et al. proposed a nonlinear spectral subtraction method which was basically a modification of the method proposed by author [18] by making the oversubtraction factor frequency dependent and the subtraction process nonlinear. Larger values are subtracted at frequencies with low SNR speech, and smaller values are subtracted at frequencies with high SNR speech. Based on Ephraim and Malah noise suppressor [22], O. Cappe proposed a method for elimination of the musical noise phenomenon [154] in 1994. A method utilizing time-frequency filtering for suppression of low residual noise was introduced by G. Whipple [156]. P. Scalart and J. Vieira-Filho [3] in the year 1996 suggested a modified speech enhancement method based on a priori signal-to-noise estimation technique to remove the musical and residual noise. Z. Goh, et al. [157] has proposed post-processing method for suppressing musical noise generated by spectral subtraction in the year 1998. J. Hansen and B. Pellom introduced an effective quality evaluation protocol for speech enhancement algorithms for further noise reduction in the same year. B. Sim and Y. Tong gave a parametric formulation of the generalized spectral subtraction method for low residual noise reduction [234].

In 1999 P. Satyanarayana [14] studied short segment analysis for single-channel speech enhancement and suppression of low residual noise. [241]. U. Mittal and N. Phamdo [241] in 2000, proposed an approach to deal with colored noise for speech enhancement using the Rayleigh Quotient method. H. Gustafsson et al. [19] in 2001, reported a modified spectral subtraction using reduced delay convolution and adaptive averaging. In the same year, J. Jensen and J. H. Hansen [215] presented a constrained iterative sinusoidal model for speech enhancement. In 2002, S. Kamath and P. C. Loizou [5] proposed a multi-band spectral subtraction method for enhancing speech corrupted by colored noise while R. Martin [210]

suggested a speech enhancement method using MMSE short time spectral estimation with Gamma distributed speech priors. Y. Hu and P. C. Loizou [20] in 2003 introduced a generalized subspace approach for enhancing speech corrupted by colored noise while F. Jabloun and B. Champagne [21] incorporated the human hearing properties in the signal subspace approach for speech enhancement in the same year, In 2004, a speech enhancement method [33] based on wavelet thresholding using multitaper spectrum was proposed by Y. Hu and P. C. Loizou. A speech Enhancement method using non-causal a-Priori SNR Estimator was introduced by I. Cohen [235]. In 2005, P. C. Loizou [24] suggested speech enhancement based on perceptually motivated Bayesian estimators of the magnitude spectrum for maximum noise suppression. R. Martin proposed statistical method for the enhancement of noisy speech [170]. Relaxed statistical model for speech enhancement and a priori SNR estimation was given by I. Cohen [176]. C. H. You et al. proposed  $\beta$ -order MMSE spectral amplitude estimation for speech enhancement [209] while T. Lotter and P. Vary illustrated speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model [211]. O. D. Deshmukh and C. Espy-Wilson gave speech enhancement using auditory phase opponency model [27]. P. C. Loizou et al. in the year 2005, implemented subspace based algorithm for noise reduction in cochlear implants [37]. Modified spectral subtraction method for enhancement of noisy speech was suggested by P. Krishnamurthy and S R M Prasanna [130]. Ghanbari Yasser and Mohammad Reza Karami [136] in 2006 illustrated a new approach for speech enhancement based on the adaptive thresholding of the wavelet packets. S. Rangachari and P. C. Loizou proposed a noise-estimation algorithm for highly non-stationary environments [150]. In the same year, C. Plapous et al. suggested an improved signal-to-noise ratio estimation for speech enhancement [174]. In 2007, O. D. Deshmukh et al. [25] discussed speech enhancement using the modified phase opponency model. Yu. Guoshen et al. illustrated audio signal denoising with complex wavelets and adaptive block attenuation [129] for highly noisy speech while J. S. Erkelens et al. proposed minimum mean-square error estimation of discrete Fourier coefficients with generalized Gamma priors for speech enhancement [213] in 2007.

Anthony P. Stark et al. [52] in 2008 studied about doise driven short-time phase spectrum compensation procedure for speech enhancement while K. Wojcicki and K. K. Paliwal suggested a speech enhancement method by exploiting conjugate symmetry of the short-time Fourier spectrum for speech enhancement [53]. In the same year, C. Breithaupt and R. Martin proposed parameterized MMSE spectral magnitude estimation for the enhancement of noisy speech [214]. In 2009, G. Kim [38] discussed an algorithm that improves speech

intelligibility in noise for normal-hearing listeners while J. Ma et al. suggested an objective measure for predicting speech intelligibility in noisy conditions based on new band-importance functions [88]. In the same year, S. So et al. proposed a method based on Kalman filter with phase spectrum compensation algorithm for speech enhancement [223]. In 2010, C. H. Taal et al. proposed a novel approach on predicting the difference in intelligibility before and after single-channel noise reduction [78] while C. Christiansen, et al. predicted speech intelligibility based on an auditory preprocessing model [79] and Bin Zhou proposed an improved wavelet-based speech enhancement method using adaptive block thresholding [131]. Gibak Kim and P. C. Loizou studied for improving speech intelligibility in noise using environment-optimized algorithms [181]. In the same year 2010, Gibak Kim and P. C. Loizou illustrated a new binary mask based on noise constraints for improved speech intelligibility [182]. In the year 2011, S. Jørgensen and T. Dau predicted speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing [76] and C. H. Taal et al. proposed an algorithm for intelligibility prediction of time-frequency weighted noisy speech [80]. K. Paliwal et al. studied about the role of modulation magnitude and phase spectrum towards speech intelligibility [127] while C. Breithaupt and R. Martin analyzed for the decision-directed SNR estimator in speech enhancement with respect to low-SNR and transient conditions [166]. In the same year 2011, G. Kim and P. C. Loizou introduced a gain-induced speech distortions and the absence of intelligibility benefit with existing noise-reduction algorithms in journal of the Acoustical Society of America [177]. In 2012, K. Wojcicki and P. C. Loizou introduced a channel selection method in the modulation domain for improved speech intelligibility in noise [128]. Tahsina Farah Sanam and Celia Shahnaz illustrated for the enhancement of noisy speech based on a custom thresholding function with a statistically determined threshold [133]. In the same year 2012, F. Chen and P. Loizou [178] studied about the impact of SNR and gain-function over- and under-estimation on speech intelligibility. In the year 2014, J. Jensen and C. H. Taal suggested the speech intelligibility prediction based on mutual information [86] and Yu Chengzhu et al. evaluated the importance of time-frequency contributions to speech intelligibility in noise [185]. In the same year 2014, Yanna Ma and Akinori Nishihara proposed a modified Wiener filtering method combined with wavelet thresholding multitaper spectrum for speech enhancement [179].

Study about reverberation suppression in noisy speech started with effort made by Oppenheim et al. in 1968 [242] who observed that the residual signal following linear prediction analysis contains peaks corresponding to the excitation events in voiced speech



together with additional peaks due to the reverberant channel and dereverberated speech were synthesised using the processed residual and the all-pole filter. In 1979, Neely and Allen [243] investigated about the homomorphic inverse filtering for dereverberation where the impulse response was decomposed into a minimum phase component and an all-pass component. Mourjopoulos et al. [244] in year 1994 proposed a single-channel least squares inverse filters for speech dereverberation. This requires extremely long inverse filter and results in large processing delay. In 2000, B. Yegnanarayana and P. Satyanarayana Murthy [45] studied about the enhancement of reverberant speech using LP residual signal. In 2001, K. Lebart and J. Boucher [44] suggested a new method based on spectral subtraction for speech dereverberation. B. W. Gillespie et al. [192] proposed a method for speech dereverberation via maximum-kurtosis subband adaptive filtering in 2001. E. A. P. Habets analyzed single-channel speech dereverberation based on spectral subtraction in 2004 [194]. M. Wu and D. Wang [43] in 2006 built a two-stage algorithm for one-microphone reverberant speech enhancement. In the year 2007, N. D. Gaubitch and P. A. Naylor illustrated spatiotemporal averaging method for enhancement of reverberant speech [193]. In 2008, E. A. P. Habets et al. [190] suggested for temporal selective dereverberation of noisy speech using one microphone. H. W. Lollmann and P. Vary [195] in the year 2009 proposed a blind speech enhancement algorithm for the suppression of late reverberation and noise. E. A. P. Habets et al. illustrated about late reverberant spectral variance estimation based on a statistical model in 2009 [196]. In the same year, J. S. Erkelens and R. Heusdens studied about single-microphone late-reverberation suppression in noisy speech by exploiting long-term correlation in the DFT domain [197]. K. Kokkinakis and O. Hazrati in the year 2011 has been proposed a channel-selection criterion for suppressing reverberation in cochlear implants [206].

Exhaustive study of available literature provides a deep insight into the work carrying out in the field of single-channel speech enhancement. Most of the previous researchers have studied for speech quality improvement with the use of speech amplitude in speech enhancement techniques. Very few algorithms are reported for intelligibility improvement. In this thesis, single-channel speech enhancement techniques are proposed for highly non-stationary noise with low and high input SNR of noisy speech signals. Furthermore, in this thesis, the designed speech enhancement methods are also tested for mixed noise types and reverberation conditions and phase is considered for speech enhancement in place of amplitude based methods.

### 1.3 Performance Evaluation Parameters for Speech Enhancement Methods

The rapid increase in usage of speech processing algorithms/techniques in multi-media and telecommunications applications raises the need for evaluation of speech for its quality and intelligibility. Accurate and reliable assessment of speech quality is thus becoming vital for the satisfaction of the end-user or customer of the deployed speech processing systems (e.g., cell phone, speech synthesis system, etc.). Assessment of speech quality can be done using subjective listening tests or using objective quality measures. Subjective evaluation involves comparisons of original and processed speech signals by a group of listeners who are asked to rate the quality of speech along a pre-determined scale. Objective evaluation involves a mathematical comparison of the original and processed speech signals. Objective measures quantify quality by measuring the numerical “distance” between the original and processed signals [89, 100-109].

The various parameters used for performance evaluation of single-channel speech enhancement methods are presented below:

#### 1.3.1 Perceptive evaluation of speech quality (PESQ)

The perceptual evaluation of speech quality (PESQ) measure, described in [89], was selected as the ITU-T recommendation P.862 [90] replacing the old P.861 recommendation [91]. The latter recommendation proposed a quality assessment algorithm called perceptual speech quality measure (PSQM). The scope of PSQM is limited to assess distortions introduced by higher-bit speech codecs operating over error-free channels. The final PESQ score is computed as a linear combination of the average disturbance value  $d_{sym}$  and the average asymmetrical disturbance value  $d_{asym}$  as follows:

$$PESQ = a_0 + a_1 \cdot d_{sym} + a_2 \cdot d_{asym} \quad (1.1)$$

Where,  $a_0 = 4.5$ , and  $a_1 = -0.1$  and  $a_2 = -0.0309$ .

The weighting of the additive frequencies is called average symmetrical  $d_{sym}$  and asymmetrical  $d_{asym}$  disturbance which are generally introduced by the codec. The range of the PESQ score is 0 to 4.5. However, for most of the cases output range will be a MOS-like score, i.e., a score between 1.0 and 4.5. High correlations ( $\rho > 0.92$ ) with subjective listening tests were reported in [35] using the above PESQ measure for a large number of testing conditions taken from mobile and voice over Internet Protocol (VoIP) applications. The PESQ can be used reliably to predict the subjective speech quality of codecs (waveform and CELP-type coders) in situations where transmission channel errors, packet loss or varying delays are

present in the signal. It should be noted that the PESQ measure does not provide a comprehensive evaluation of telephone transmission quality or other communication systems, as it only reflects the effects of one-way speech or noise distortion perceived by the end-user. Effects such as loudness loss, side tone and talker echo are not reflected in the PESQ scores. Higher value (approx 4) of PESQ shows for maximum speech enhancement.

### 1.3.2 Mean opinion score (MOS)

The most widely used direct method of subjective quality evaluation is the category judgment method in which listeners rate the quality of the test signal using a five-point numerical scale, with 5 indicating “excellent” quality and 1 indicating “unsatisfactory” or “bad” quality. This method has been recommended by the IEEE subcommittee on subjective methods [92] as well as by ITU [93-94]. The measured quality of the test signal is obtained by averaging the scores obtained from all listeners. This average score is commonly referred to as the Mean Opinion Score (MOS).

Table 1.1: Rating scale for MOS.

Rating	Speech quality	Level of distortion
5	Excellent	Imperceptible
4	Good	Just perceptible, but not annoying
3	Fair	Perceptible and slightly annoying
2	Poor	Annoying, but not objectionable
1	Bad	Very annoying and objectionable

The MOS test is administered in two phases: training and evaluation. In the training phase, listeners hear a set of reference signal that exemplify the high (excellent), the low (bad) and the middle judgment categories. This phase, also known as “anchoring phase”, is very important as it is needed to equalize the subjective range of quality ratings of all listeners; that is, the training phase should in principle equalize the “goodness” scales of all listeners to ensure, to the extent possible, that what is perceived “good” by one listener is perceived “good” by the other listeners, too. A standard set of reference signals need to be used and described when reporting the MOS scores [92]. In the evaluation phase, subjects listen to the test signal and rate the quality of the signal in terms of the five quality categories (1-5) shown in Table 1.1.

### 1.3.3 Short-time objective intelligibility (STOI)

The STOI is based on short-time segments i.e. 386 ms. This short segment is taken in order to get maximum correlation with subjective speech intelligibility. The STOI varies in between 0 and 1, where, 0 and 1 show the lowest and highest improvement in speech intelligibility, respectively. The intelligibility measure is defined as linear correlation coefficients between clean and enhanced time-frequency (TF) unit which is given by eq. (1.2) [80].

$$d_{n,k} = \frac{\sum_n (X_k - \frac{1}{N} \sum X_k) (\hat{X}_k - \frac{1}{N} \sum \hat{X}_k)}{\sqrt{\sum_n (X_k - \frac{1}{N} \sum X_k)^2 \sum_n (\hat{X}_k - \frac{1}{N} \sum \hat{X}_k)^2}} \quad (1.2)$$

In the equation (1.2),  $X_k(n)$  and  $\hat{X}_k(n)$  are the clean and enhanced signal respectively. The overall average of intelligibility measure from all bands and frames are given from equation (1.3). Where,  $N$  is the total number of frames and  $K$  is the number of one-third octave bands. The STOI is computed as [80]:

$$STOI = \frac{1}{N K} \sum_{N,K} d_{n,k} \quad (1.3)$$

### 1.3.4 Speech intelligibility index (SII)

This type of intelligibility measure depends on SNR values. The fraction of input SNR is calculated using eq. (1.4) [95].

$$fSNR_k = \begin{cases} \frac{\min(\overline{SNR}_k, SNR_k)}{SNR_k} & \text{if } SNR_k \geq SNR_L \\ 0 & \text{else} \end{cases} \quad (1.4)$$

Where,  $\overline{SNR}_k$  is the ratio of output SNR in band  $k$  to noise spectrum. The lowest SNR is  $SNR_L$  and  $fSNR_k$  is bounded within 0 to 1. Here, 0 and 1 indicate the lower and higher improvement in speech intelligibility, respectively. Weighted average is calculated across all bands for getting SII in eq. (1.5).

$$SII = \frac{1}{\sum_{k=1}^K W_k} \sum_{k=1}^K W_k \times fSNR_k \quad (1.5)$$

### 1.3.5 Distortion rating measures

Basically two types of distortions are considered in speech enhancement such as speech and noise distortion. If distortions are created by speech signal alone called speech distortions and distortions created by noise only are called noise distortions in the processed speech. Generally, the speech enhancement techniques/algorithms degrade the speech signal while suppressing the background noise in low SNR conditions. This situation of speech enhancement technique/algorithm complicates the intelligibility evaluation since it is not clear whether listeners rate their overall speech intelligibility on the speech distortion component, noise distortion component or both. The uncertainty that the individual listeners place different weight on the signal and noise distortion components introduces additional error because of variation in the listeners' ratings, which consequently decreases the reliability of the rating. These concerns were addressed by the ITU-T standard (P. 835) [96] that was designed to lead the listeners to integrate the effects of both signal and background noise distortion in making their ratings of overall quality [97]. The methodology proposed in [96] reduces the listener's uncertainty by requiring him/her to successively attend to and rate the waveform on the speech signal alone, the background noise alone, and the overall effect of speech and noise on quality. More precisely, the ITU-T P.835 method instructs the listener to successively attend to and rate the enhanced speech signal in terms of the five quality categories shown in Table 1.2.

Table 1.2: Rating scale for BAK, SIG and OVL measures.

Rating	BAK	SIG	OVL
1	Very conspicuous, very intrusive	Very unnatural, very degraded	Bad
2	Fairly conspicuous, somewhat intrusive	Fairly unnatural, fairly degraded	Poor
3	Noticeable but not intrusive	Somewhat natural, somewhat degraded	Fair
4	Somewhat noticeable	Fairly natural, little degradation	Good
5	Not noticeable	Very natural, no degradation	Excellent

1. The background noise distortions alone using a five-point scale of background intrusiveness **(BAK)**
2. The speech signal alone using a five-point scale of signal distortion **(SIG)**
3. The overall **(OVL)** effect using the scale of the Mean Opinion Score – [1=bad, 2=poor, 3=fair, 4=good, 5=excellent].

### 1.3.6 Mean square error (MSE)

The magnitude spectrum of clean speech is estimated from noisy speech spectrum by minimizing mean-square error between the magnitude spectra of clean and estimated speech. The lower value i.e. approx to 0, means maximum improvement in speech quality and quality decreases as MSE increases. The eq. (1.6) for measuring MSE is given as:

$$\epsilon = E \left[ \left( |X(n, k)| - |\hat{X}(n, k)| \right)^2 \right] \quad (1.6)$$

Where,  $\epsilon$  (error) is difference between clean and estimated spectrum of speech,  $X(n, k)$  is clean speech spectrum and  $\hat{X}(n, k)$  is estimated speech spectrum.

### 1.3.7 Signal -to- noise ratio (SNR)

The common definition of SNR is the ratio of power of the processed output signal to the noise signal power. For calculating the SNR of the processed speech signal output to the noise is given in eq. (1.7) as:

$$SNR = 10 \log_{10} \left[ \left( \frac{\hat{X}}{D} \right)^2 \right] \quad (1.7)$$

Where,  $\hat{X}$  and  $D$  indicates the processed output signal and noise signal, respectively. The calculated SNR is measured in decibel (dB). The higher value of SNR i.e. about 10dB shows good quality of speech signal.

### 1.3.8 Segmental SNR (SNRseg)

The segmental signal-to-noise ratio can be evaluated either in the time domain or frequency domain. The time-domain measure is perhaps one of the simplest objective measures used to evaluate speech enhancement or speech coding technique/algorithm. For this measure to be meaningful, it is important that the original and processed signal be aligned in time. The time domain segmental signal-to-noise ratio (SNRseg), given in eq. (1.8) is defined as [98]:

$$SNR_{seg} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \frac{\sum_{n=Nm}^{Nm+N-1} (X(n))^2}{\sum_{n=Nm}^{Nm+N-1} (X(n) - \hat{X}(n))^2} \quad (1.8)$$

Where,  $X(n)$  is the original (clean) signal,  $\hat{X}(n)$  is the corresponding enhanced speech,  $N$  is the frame length and  $M$  is the number of frames in the signal.

### 1.3.9 Frequency weighted segmental SNR (fw-SNRseg)

The frequency – weighted segmental SNR is calculated using the eq. (1.9) [99]

$$\frac{10}{n} \sum_{n=0}^{n-1} \frac{\sum_{k=1}^k W(n,k) \log_{10} \frac{(X(n,k))^2}{(X(n,k) - \hat{X}(n,k))^2}}{\sum_{k=1}^k W(n,k)} \quad (1.9)$$

Where,  $k$  is the number of band,  $n$  is the total number of frame and the critical-band magnitude of the clean signal is  $X(n,k)$  at  $n^{th}$  frequency band at the  $k^{th}$  frame. The  $\hat{X}(n,k)$  is the corresponding enhanced speech signal. In the given eq. (1.10)  $W(n,k)$  is the weight function and  $p$  is the power exponent which varies according to speech. The weighting function is given as:

$$W(n,k) = X(n,k)^p \quad (1.10)$$

The lower value i.e. approx to 0, means minimum improvement in speech quality. The quality of speech increases as value of fw-SNRseg increases.

### 1.3.10 Cepstrum distance (CD)

A linear prediction coefficients (LPC) based cepstrum distance measure is introduced as an effective evaluation measure, not only for coding distortion but also for other nonlinear distortion such as quadratic and logarithmic distortion. The LPC cepstrum distance CD is as in eq. (1.11) [100]:

$$CD = 10 / \log_{10} \sqrt{2 \sum_{n=1}^p (x(n) - \hat{x}(n))^2} \quad (1.11)$$

Where,  $x(n)$  and  $\hat{x}(n)$  are LPC cepstrum coefficients of input and output signal, respectively and  $p$  is the maximum order of the coefficients. The lower value (approx 0) of CD means maximum improvement in speech quality. The speech quality decreases as value of CD increases.

### 1.3.11 Weighted spectral slop (WSS) metric

Weighted differences between the spectral slops in each band, was proposed by Denis H. Klatt [108]. This measure was designed to penalize heavy differences in spectral peak (formants) locations while ignoring other differences between the spectral such as spectral tilt, overall level, etc. Those differences were found to have little effect on ratings of phonetic distance between pairs of synthetic vowels. This measure is computed by the spectral slop of each band. A first-order differencing operation is used to compute the spectral slops as given in eq. (1.12) and (1.13), respectively:

$$X(n) = C_x(n+1) - C_x(n) \quad (1.12)$$

$$\hat{X}(n) = \bar{C}_{\hat{x}}(n+1) - \bar{C}_{\hat{x}}(n) \quad (1.13)$$

Where,  $C_x(n)$  and  $\bar{C}_{\hat{x}}(n)$  denote the original (clean) and enhanced critical-band spectra, respectively, expressed in dB. The  $X(n)$  and  $\hat{X}(n)$  are the original (clean) and enhanced signals, respectively, of the  $k$ th band. The differences in spectral slops are then weighted according to, first, whether the band is near a spectral peak or valley and, second, according to whether the peak is the largest peak in the spectrum. The weight for bank  $k$ , denoted as  $W(k)$  in eq. (1.14), is computed as follows:

$$W(n) = \frac{K_{\max} K_{loc\max}}{[K_{\max} + C_{\max} - C_x(n)][K_{loc\max} + C_{loc\max} - C_x(n)]} \quad (1.14)$$

Where,  $C_{\max}$  is the largest log-spectral magnitude among all bands,  $C_{loc\max}$  is the value of the peak nearest to band  $n$ , and  $K_{\max}, K_{loc\max}$  are constants which can be adjusted using regression analysis to maximize the correlation between the subjective listening tests and values of the objective measure. The WSS measure is finally computed by eq. (1.15) [108]:

$$d_{WSS}(C_x, \bar{C}_{\hat{x}}) = \sum_{k=1}^L W(n) (X(n) - \hat{X}(n))^2 \quad (1.15)$$

Where,  $d_{WSS}$  is shown for calculated WSS values. The lower value (approx 0) of WSS means maximum improvement in speech quality. The increasing trend of WSS indicates degradation in speech quality.

### 1.3.12 Itakura saito (IS)

Initially, this IS measure was used successfully in speech recognition for comparing a reference power spectrum  $R(w)$  against a test spectrum  $X(w)$  according to eq. (1.16) [109]:

$$d_{IS}(X(n), R(n)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ \frac{R(n)}{X(n)} - \log \left( \frac{R(n)}{X(n)} \right) - 1 \right] dn \quad (1.16)$$



Owing to its asymmetric nature, the IS measure is known to provide more emphasis on spectral peaks than spectral valleys. A Bayesian estimator based on the Itakura-Saito (IS) measure was considered in [24, 110] between the estimated  $X$  and true  $\hat{X}$  short-time power spectra at the  $n$ th frequency bin and given by eq. (1.17) as:

$$d_{IS}(X^2, \hat{X}^2) = \frac{X^2}{\hat{X}^2} - \log\left(\frac{X^2}{\hat{X}^2}\right) - 1 \quad (1.17)$$

Where,  $d_{IS}$  is shown for calculated IS values. The lower value (approx 0) of IS means maximum improvement in speech quality. The speech quality decreases as value of IS increases.

#### 1.4 Objectives of the Present Work

As mentioned in the preceding sections, most of the single-channel enhancement techniques/algorithms process degraded speech for achieving improvement either in quality or intelligibility but do not improve both. Also, most of the existing speech enhancement techniques/algorithms do not perform equally well at both low and high SNR levels of noisy speech signals of any types of noise conditions.

In practical environment, there are various sources of noise which are responsible for degradation of speech signal. The additive background noise may be of any SNR level i.e. higher, medium and lower. Due to wide range of requirements of speech enhancement applications and limited performance of the available speech enhancement algorithms, there is a need to develop other methods/techniques which may prove to be robust in many noisy environment conditions. If generated background noise is highly non-stationary and of very low SNR level then the speech enhancement method must be very effective for noise reduction or suppression and give the processed speech of good quality and intelligibility. Considering all the aspects above, the objectives of the present work has been formulated as below:

- Design and development of a suitable algorithm for enhancement of single-channel mixed noisy speech of very low (Negative) SNR.
- Design and development of algorithm for suppression of non-stationary noise for single-channel speech signal.
- Design and development of algorithm for suppression of combined effect of background noise and reverberation.
- Design and implementation of phase based single-channel speech enhancement technique.

- Evaluation of the performance of proposed algorithms in comparison to existing speech enhancement algorithms.

## 1.5 Structure of the Thesis

The work presented in this thesis is organized in different chapters. The brief descriptions about these chapters are presented as following:

Chapter 2 discusses the importance of gain functions in single-channel speech enhancement methods. The description of the proposed modified Wiener gain function is given next. Further, the computed performance parameters are presented in tabulation form and their graphical and bar chart representation is given next.

Chapter 3 presents the description of wavelet packet transform (WPT) based fuzzy mask method for the enhancement of non-stationary and highly non-stationary noisy speech. Then the results and their comparative analysis are presented.

Chapter 4 describes the details about the reverberant mask criterion selection based method for the enhancement of noisy speech degraded by reverberation and sources of background noise both. A novel algorithm which was developed here for suppression of both reverberation and noise was discussed next. Thereafter, the comparison of the performance of the proposed algorithm with other speech enhancement algorithm discussed.

Chapter 5 introduces the concept of phase in speech enhancement process. In this chapter, a method based on phase of speech and noise is proposed and its details are discussed. The performance of proposed method is compared with other phase based method in terms of their quality and intelligibility measures.

In Chapter 6, conclusions and future scope of the work is reported. This covers the findings of this research work with an emphasis to extend research work in future.

*This chapter presents the research work done for improving quality and intelligibility both by suppressing multiple noise sources under very low input SNR conditions. It starts with the discussion of database generation of mixed noise sources and thereafter the proposed Wavelet Packet Transform based modified Wiener Gain method is introduced for suppression of mixed background noise sources. The results of this method for low SNR input Hindi and English speech signals are discussed at the end.*

### 2.1 Overview

The enhancement of multiple noise mixed speech signal can be done effectively and relatively easily, if the speech signals are collected simultaneously over two or more spatially distributed microphones. In such a case one could exploit the delay in speech signals produced by an individual at different microphone locations. The delays obtained for speakers become all the speakers cannot be placed at same location, simultaneously. The problem of enhancing speech degraded by multiple additive background noise and speech environment is a challenging task when only a single-channel is available for recording. Algorithms that use the spectral characteristics rely on the estimation of pitch of individual speaker and using this information, the speech of the desired speaker is enhanced by retaining only pitch and harmonic components while ignoring the remaining spectral components [114, 115]. Since speech energy of a particular speaker is concentrated at the pitch and harmonics, speech signal corresponding to the speaker is synthesized using amplitudes of short time spectrum at different harmonics that is at different frequencies [116-118]. But, it is generally difficult to obtain the pitch of an individual speaker from the multi-speaker signal. Alternatively, the algorithm that uses excitation information of speech relies on time-delay between the microphone signals and excitation characteristics of individual speaker for speech enhancement. The basis for this method is that, the relative positions of significant excitation in the direct component of the speech signal remain unchanged at each of the microphone for a given speaker. By estimating time delays and using the knowledge of excitation source characteristics, a weight function is derived for each speaker to identify the speech component of desired speaker relative to other speaker [119-123]. The high values of weight function indicate the temporal regions where the corresponding speaker's speech is predominant.

## 2.2 Single-Channel Speech Enhancement Methods

Most widely used speech enhancement methods are spectral subtraction based methods [16-17, 32, 34], Wiener method [3], and gain based method like MMSE STSA [22], log-MMSE [23], p-MMSE [24] etc. In addition the modulation channel selection based methods [124-129] are also used for improving both intelligibility and quality. These methods are explained in details as below:

### 2.2.1 Spectral subtraction method

The spectral subtraction method has been proposed by Berouti et al. in 1979 [17, 32, 34]. It is very popular method for reducing the effect of background (additive) noise. It is based on a simple principle that the spectrum of clean signal can be obtained by subtracting estimated noise spectrum from the noisy speech spectrum. In this method noise is considered as additive type. If a clean speech signal  $X(n)$  corrupted by background additive noise  $D(n)$  then the equation of noisy speech signal  $Y(n)$  is expressed as:

$$Y(n) = X(n) + D(n) \quad (2.1)$$

The Discrete Short-Time Fourier Transform (DSTFT) of the corrupted speech signal  $Y(n, k)$  is given as:

$$Y(n, k) = X(n, k) + D(n, k) \quad (2.2)$$

Where,  $Y(n, k)$ ,  $X(n, k)$  and  $D(n, k)$  are the Fourier transform of windowed noisy speech, clean speech and noise signal, respectively. Now, aim is to calculate the clean spectrum of speech signal. The estimated speech signal obtained from calculated clean spectrum has estimation error that produces musical noise in the estimated speech signal.

### 2.2.2 Amplitude spectral subtraction method

It is the basic spectral subtraction method for noise reduction from noisy speech signal. The noisy speech signal can be given as [18]:

$$Y(n) = X(n) + D(n) \quad (2.3)$$

After taking Fourier transform of both sides of eq. (2.3), get eq. (2.4)

$$Y_w(e^{j\omega}) = X_w(e^{j\omega}) + D_w(e^{j\omega}) \quad (2.4)$$

Where,  $Y_w(e^{j\omega})$ ,  $X_w(e^{j\omega})$  and  $D_w(e^{j\omega})$  are noisy speech, clean speech and noise signal respectively and  $w$  is given for windowing. Multiplying both sides by their complex conjugates as:

$$|Y(e^{j\omega})|^2 = |X(e^{j\omega})|^2 + |D(e^{j\omega})|^2 + 2|X(e^{j\omega})||D(e^{j\omega})|\cos(Dq) \quad (2.5)$$

Where,  $Dq$  is the phase difference between speech and noise. Now, taking expected value of both sides and get:

$$E\{|Y(e^{j\omega})|^2\} = E\{|X(e^{j\omega})|^2\} + E\{|D(e^{j\omega})|^2\} + 2E\{|X(e^{j\omega})|\}E\{|D(e^{j\omega})|\}E\{\cos(Dq)\} \quad (2.6)$$

In the eq. (2.6) reasonable assumptions are made: Noise and speech magnitude spectrum values are independent of each other. The phase of noise and speech are independent of each other and of their magnitude and it is assumed that  $E\{\cos(Dq)\}=0$ , now eq. (2.6) is given as:

$$X_w(e^{j\omega}) = Y_w(e^{j\omega}) - D_w(e^{j\omega}) \quad (2.7)$$

The magnitude spectrum of the noise is averaged during speech inactive periods and clean speech spectrum is estimated by subtracting the average spectrum of noise from each segment of noisy speech spectrum.

### 2.2.3 Power spectral subtraction method

In power spectral subtraction it is assumed that  $E\{\cos(Dq)\}=0$ . Put this value in eq. (2.4) and get eq. (2.8) and (2.9) as [17]:

$$E\{|Y(e^{j\omega})|^2\} = E\{|X(e^{j\omega})|^2\} + E\{|D(e^{j\omega})|^2\} \quad (2.8)$$

$$X_w(e^{j\omega}) = Y_w(e^{j\omega}) - E\{D_w(e^{j\omega})\} \quad (2.9)$$

The power spectrum of clean speech is obtained by subtracting power spectrum of noise from the power spectrum of the noisy speech in the current frame. The power spectrum of noise is estimated during speech inactive periods.

### 2.2.4 Multi-band spectral subtraction method

A multi-band spectral subtraction approach takes into account the fact that colored noise affects the speech spectrum differently at various frequencies. This method outperforms the

standard power spectral subtraction method resulting in superior speech quality and largely reduced musical noise [5]. The clean speech spectrum is given in eq. (2.10) as:

$$|X(e^{j\omega})|^2 = |Y(e^{j\omega})|^2 - \alpha |D(e^{j\omega})|^2 \quad (2.10)$$

Where,  $\alpha$  is over-subtraction factor which is a function of segmental SNR.

### 2.2.5 Scalart power spectral subtraction method

In this type of spectral subtraction Scalart used a priori SNR criteria to estimate the noise spectrum. This *a priori* SNR is given in eq. (2.11) as [3]:

$$SNR_{prio} = \frac{E\{|X(e^{j\omega})|^2\}}{E\{|D(e^{j\omega})|^2\}} \quad (2.11)$$

Now, the noise suppression function is given as:

$$G = \sqrt{\frac{SNR_{prio}}{1 + SNR_{prio}}} \quad (2.12)$$

The noise spectrum is estimated by using eq. (2.12), and this noise spectrum is subtracted from noisy speech spectrum to obtain clean speech spectrum.

### 2.2.6 Parametric spectral subtraction method

A parametric formulation of the basic spectral subtraction is given as [234]:

$$X_w(e^{j\omega}) = A(Y_w(e^{j\omega})) - B(E\{D_w(e^{j\omega})\}) \quad (2.13)$$

Where, A and B are parameters in the parametric formulation. If A=B=1, then the eq. (2.13) becomes as eq. (2.9). The clean speech spectrum is obtained by using eq. (2.13) and another set of values are taken as A=1, B = noise subtraction factor [234].

### 2.2.7 Reduced delay convolution - spectral subtraction method

In this type of spectral subtraction, reduced delay convolution and adaptive averaging is used for noise-reduction. In this method a filter gain is used for calculating clean speech [19].

$$X_w(e^{j\omega}) = G(Y_w(e^{j\omega})) \quad (2.14)$$

In eq. (2.14)  $G$  is a non-causal filter, it is phase shifted to obtain causality and to avoid circular effect.

### 2.2.8 Wiener filtering method

The aim of Wiener filter is to minimize the mean square error between the desired signal

(clean signal) and the estimated output. In the Wiener filtering, it is assumed that the spectral property of speech is uncorrelated with the spectral property of noise. The Wiener filter is obtained by minimizing the eq. (2.15) [3].

$$e = E \left[ \left| X - \hat{X} \right|^2 \right] \quad (2.15)$$

Where,  $X$  is desired speech and  $\hat{X}$  is estimated output. The distortions introduced by this method were suppressed by using priori information of the signal and based on the priori information MMSE based methods were proposed for speech enhancement.

### 2.2.9 MMSE-speech presence uncertainty (MMSE-SPU) estimation method

Minimum Mean Square Error under Speech Presence Uncertainty estimator is an efficient algorithm that is named as MMSE-SPU which has been proposed by Ephraim and Malah in 1984 [22]. It is an optimal magnitude spectrum estimator like Wiener filtering. An efficient spectral gain is calculated by using *a priori* SNR. This MMSE-SPU algorithm is motivated by the fact that speech might not be present at all time and frequencies. Therefore, a two state model was introduced which was based on speech presence and speech absence frames in noisy speech signal. This is given as:

$$\hat{X}_N = E(X_N | Y_N, H_1^N) P(H_1^N | Y_N) \quad (2.16)$$

Where,  $X_N$  is desired speech,  $\hat{X}_N$  is estimated output and  $Y_N$  is noisy speech.  $P(H_1^N | Y_N)$  denotes the conditional probability that speech is present in frequency bin  $N$  and hypothesis is given as  $H_1^N$ . *A priori* SNR can be estimated recursively (frame-wise) using the “decision-directed” approach. It is noted that when *a priori* SNR is estimated using the “decision-directed” approach, the enhanced speech has no “musical noise”.

### 2.2.10 log-MMSE method

Log-spectrum based minimum mean square error has been derived by Ephraim and Malah from the basic concept of MMSE [13]. The log-MMSE algorithm assumes a Gaussian model for the complex spectral amplitudes of both speech and noise. It gives the optimum results of the log-spectrum of the clean speech signal. *A priori* SNR is estimated from the “Decision-directed” approach. In this approach distortions are minimized by using log estimator as mentioned in eq. (2.17):

$$e = E \left[ \left( \log|X| - \log|\hat{X}| \right)^2 \right] \quad (2.17)$$

Where,  $X$  is desired speech and  $\hat{X}$  is estimated output.

### 2.2.11 p-MMSE method

The auditory masking effects are generated due to squared-error cost function used in traditional MMSE method. To overcome the problem of spectral peaks which takes account of auditory masking effects, p-MMSE estimator is used. To estimate clean signal  $x$ , the postulate-MMSE estimator is defined as in eq. (2.18) [24]:

$$\hat{X}^{MMSE}(Y) = E(X|Y; P_{post}, \sigma_{post}^2) \quad (2.18)$$

Where,  $p_{x|y}(X|Y; P_{post}, \sigma_{post}^2)$  is the conditional distribution of  $X$  given  $Y$  under the  $x$  distribution and noise variance  $\sigma^2$  specified as parameter.

### 2.2.12 Cohen-MMSE method

The non-causal estimator is used for the *a priori* SNR. This estimator effectively discriminates between speech onsets and noise irregularities and minimizes the conditional expected value of the distortion measure. The clean speech signal is estimated by applying a spectral gain function to the noisy speech signal which is given in eq. (2.19) [235].

$$X_w(e^{j\omega}) = G(\xi, \gamma)(Y_w(e^{j\omega})) \quad (2.19)$$

Where,  $\xi$  and  $\gamma$  are *a priori* and *a posteriori* SNRs, respectively.

### 2.2.13 KLT subspace method

It is a generalized subspace approach in which pre-whitening is used for enhancing speech signal corrupted by coloured noise. In this approach, clean speech is estimated by cancelling the noise subspace signal from the noisy subspace signal. A non-unitary transform is used to project the noisy signal on to a signal-plus-noise subspace and a noise subspace [20].

### 2.2.14 PKLT subspace method

It is a perceptually motivated subspace approach which incorporates human hearing properties for enhancing speech signal. The enhancement of corrupted speech is performed by assuming that the clean component is concentrated in signal subspace, while the noise occupies the noise subspace. The noise reduction is obtained by removing the noise subspace and by removing the noise contribution in the signal subspace. Thus, a key issue in developing a subspace-based model is to select the optimal rank that will reconstruct the



enhanced signal in an optimal way. The Eigen value Decomposition (EVD) of the noisy speech is given as [21]:

$$R_{YY} = R_{XX} + R_{DD} \quad (2.20)$$

Where,  $R_{YY}$ ,  $R_{XX}$ , and  $R_{DD}$  denote noisy speech matrix, clean speech covariance metric and noise autocorrelation matrix.

### 2.3 Proposed Wavelet Packet Transform Based Modified Wiener Gain Method

The WPT based modified Wiener Gain method is proposed here for suppression of mixed noise of low SNR. The process of enhancement starts with the decomposition of received noisy speech input and given in Figure 2.1. Procedure used for the proposed speech enhancement method is described in following steps:

Step1: Generation of mixed noisy single-channel speech of low input SNR.

Step2: Input speech decomposition by using Db10 mother wavelet packet upto 3<sup>rd</sup> level.

Step3: The decomposed noisy speech signal is reconstructed into eight energy bands.

Step4: After reconstruction, FFT is applied to all eight reconstructed signals.

Step5: The soft WP threshold function is applied in first stage of the noise reduction process.

Step6: After applying soft WP threshold function in first stage, the processed speech signal is given to second stage of modified Wiener Gain function to calculate the gain function.

Step7: Noisy speech signal is multiplied with Modified Wiener gain function for second step noise reduction.

Step8: An enhanced speech spectrum is recovered by applying Inverse - FFT and overlap-add method.

Now, all processed signals of given eight energy bands of input speech are added to get the desired speech signal.

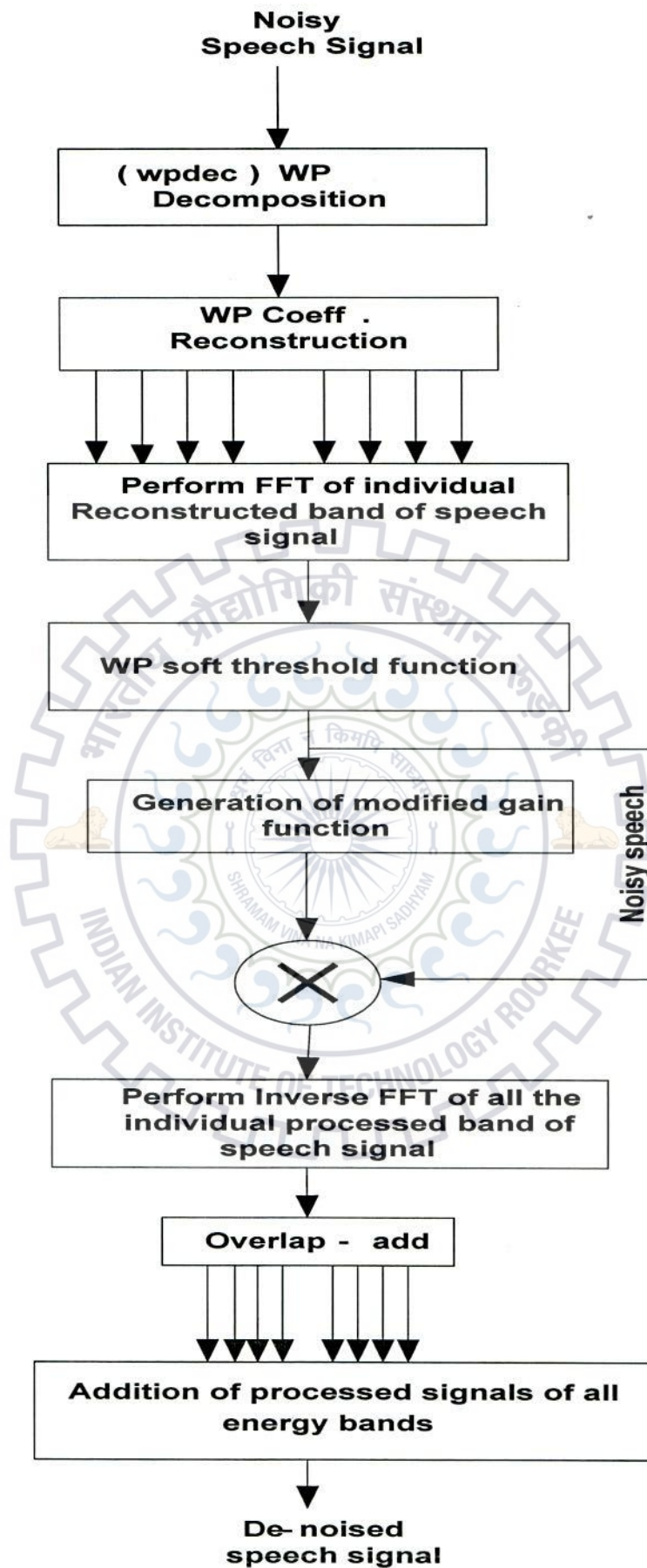


Fig. 2.1: Block diagram of proposed WPT modified Wiener gain based speech enhancement method.

### 2.3.1 Mother wavelets and decomposition levels

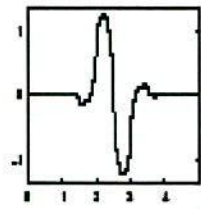
The transform of a signal is just another form of representing the signal. It does not change the information content present in the original signal. The Wavelet Packet Transform (WPT) provides a time-frequency representation of signal [141]. It was developed to overcome the shortcoming of Short-Time Fourier transform (STFT), which can also be used to analyze non-stationary signals. While STFT gives a constant resolution at all frequencies, the WPT uses multi-resolution technique by which different frequencies are analyzed with different resolutions [142].

There are number of basic functions that can be used as the mother wavelet for WPT. Since the mother wavelet produces all WP functions used in the transformation through translation and scaling, it determines the energy bands of decomposed speech signal into approximate and details coefficients. Therefore, the details of the particular application are taken into account for getting maximum energy of the decomposed signal. Further, the appropriate mother wavelet must be chosen in order to use the WPT effectively. The commonly used wavelet functions are: Haar, Daubechies, Symlets, Coiflets, Meyer, Morlet and Mexican Hat etc.

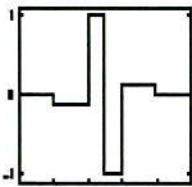
The Haar wavelet is one of the oldest and simplest wavelet functions. The Daubechies wavelets are the most popular wavelets. They represent the foundation of wavelet signal processing and are used in numerous applications. These are also called Maxflat wavelets as their frequency response has maximum flatness at frequencies 0 and  $\pi$ . This is a very desirable property in some applications.

The Haar, Daubechies, Symlets, and Coiflets are orthogonal wavelets. These wavelets along with Meyer wavelets are capable of perfect reconstruction. The Meyer, Morlet and Mexican Hat wavelets are symmetric in shape. The wavelets are chosen based on their shape and suitability for analyzing the signal in a particular application [143-147]. The Daubechies family wavelets are written as Db 'N', where 'N' is the order and Db the "surname" of the wavelet. Figure 2.2 illustrates the various wavelet functions of the members of the wavelet family [146-149].

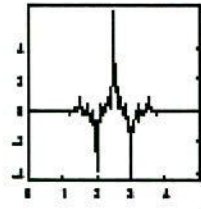
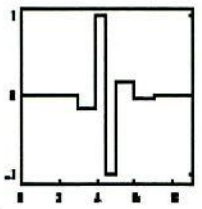
By hit and trial method, the optimum level of maximum wavelet level decomposition (i.e. third) is decided which improves the speech quality and intelligibility to a better level. Decomposed levels of given input speech are obtained with application of various Wavelet families.



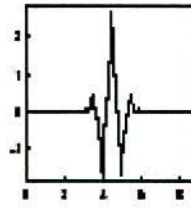
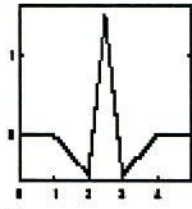
bior1.3



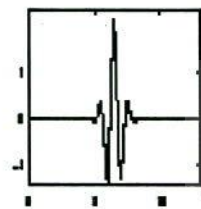
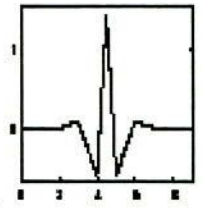
bior1.5



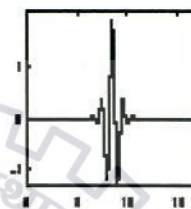
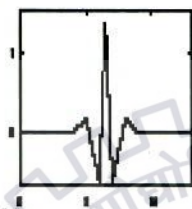
bior2.2



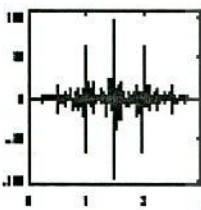
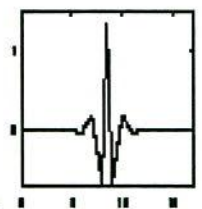
bior2.4



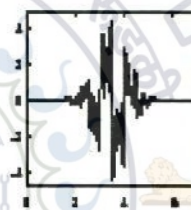
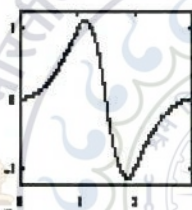
bior2.6



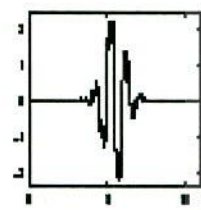
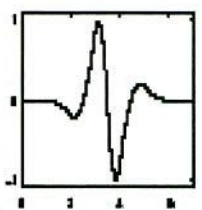
bior2.8



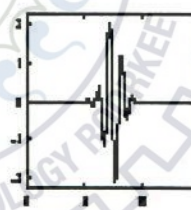
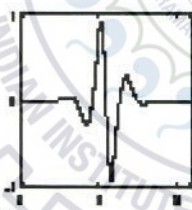
bior3.1



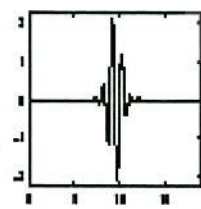
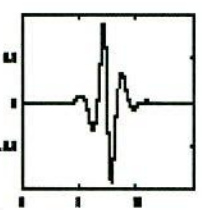
bior3.3



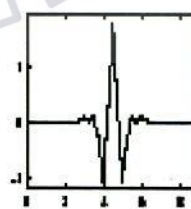
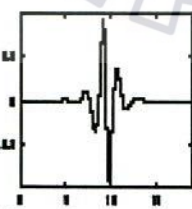
bior3.5



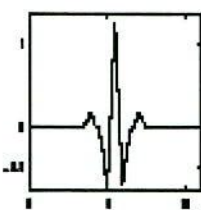
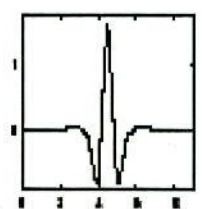
bior3.7



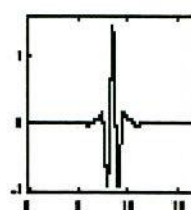
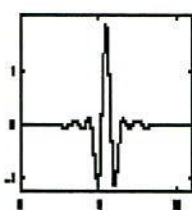
bior3.9



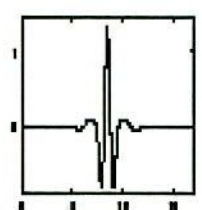
bior4.4



bior5.5



bior6.8



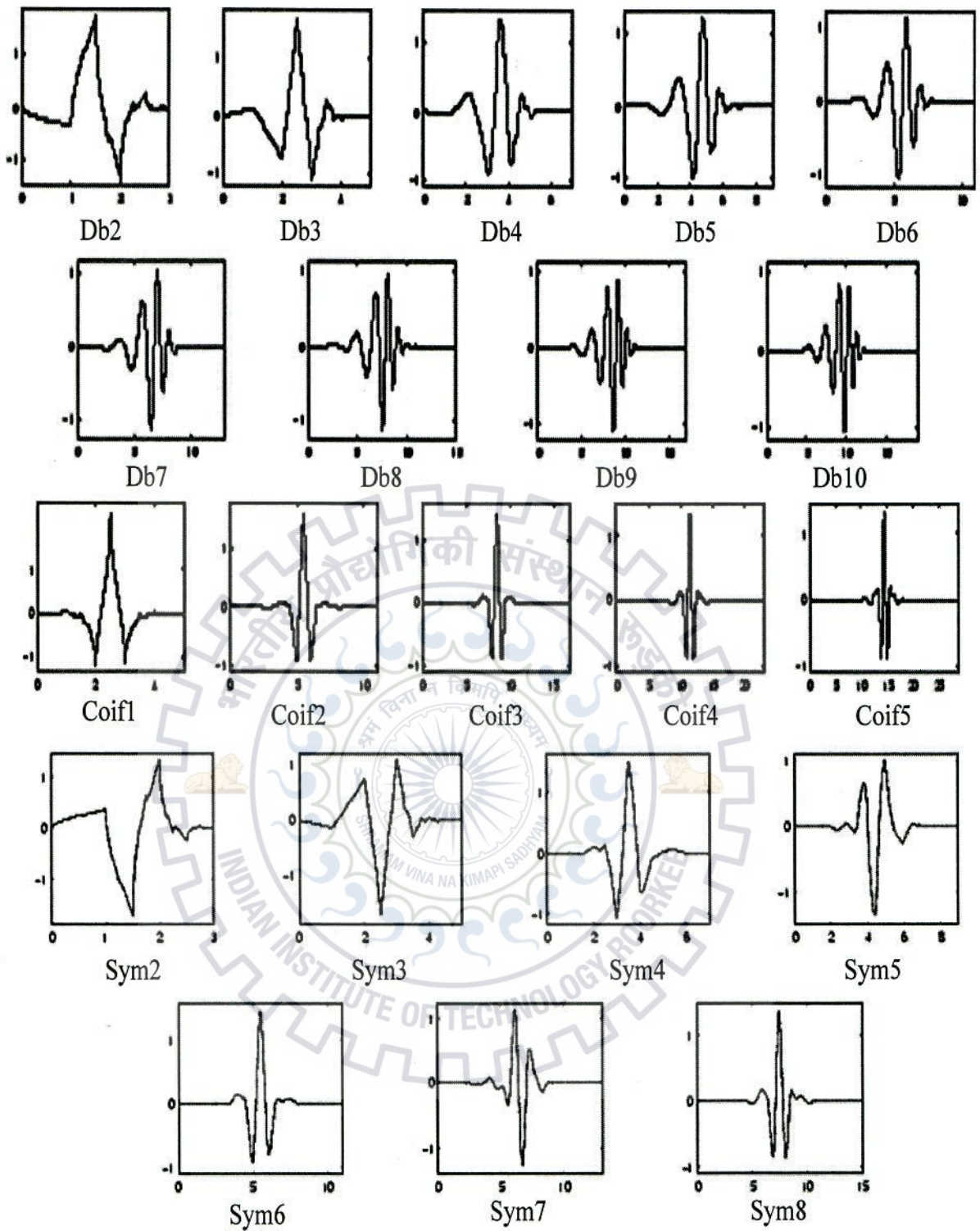


Fig. 2.2: Waveforms of various wavelet function of wavelet family [146-149].

The features of decomposed levels are available in the form of approximated and detailed coefficients. Figure 2.3 and 2.4 illustrates the 3<sup>rd</sup> level decomposition of the WT and WPT, respectively. A third level of decomposition is used for getting eight energy band of the input speech.

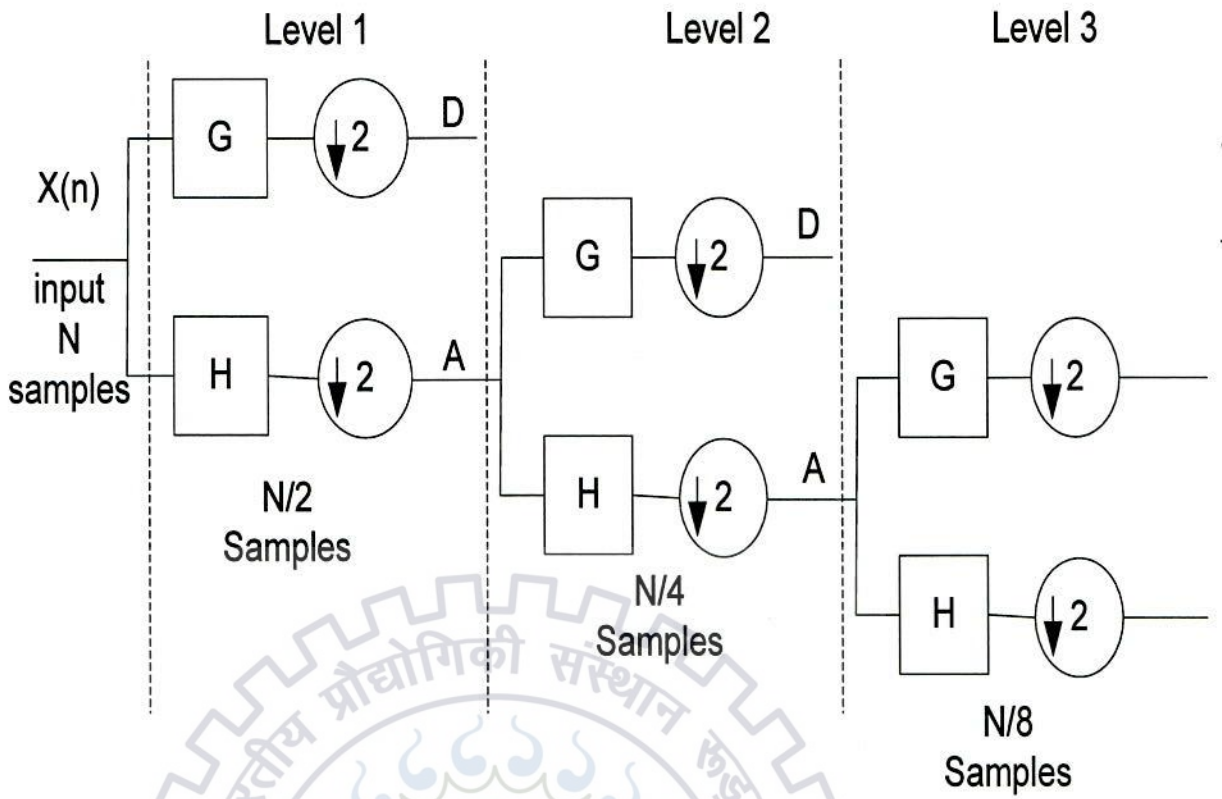


Fig. 2.3: Wavelet decomposition at 3<sup>rd</sup> level.

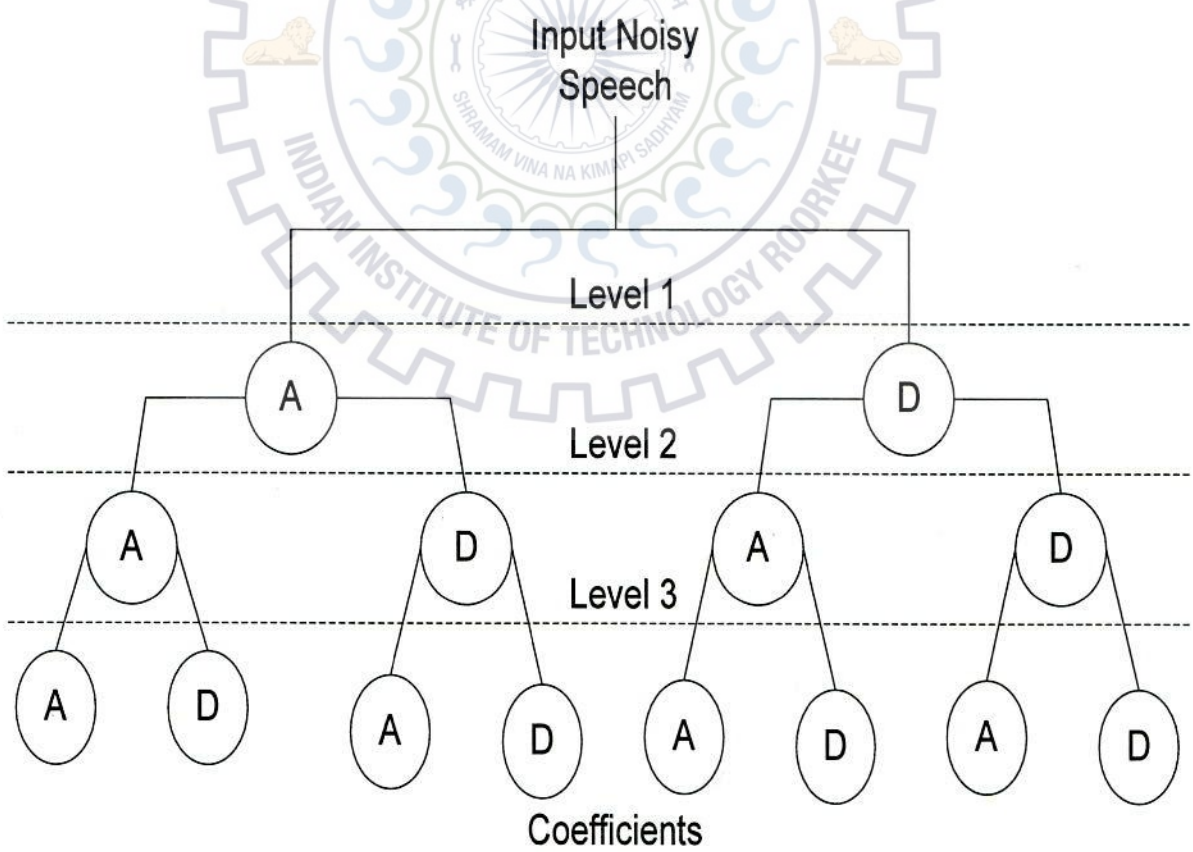


Fig. 2.4: Wavelet Packet decomposition at 3<sup>rd</sup> level.

Various mother wavelets are analyzed and compared with each other to obtain the best suitable mother wavelet for decomposition of speech signal. The block diagram illustrated in Figure 2.5 is used for the evaluation of the different mother WP. Various mother wavelets are considered at WP decomposition step and coefficients of WP decomposed levels of noisy speech signal corrupted by different types of noise like babble, pink, Volvo car and white (at different levels of SNR) are given to threshold function (described in section 2.3.2) for getting the denoised speech spectrum and enhanced speech signal is recovered by applying inverse WPT to this spectrum.

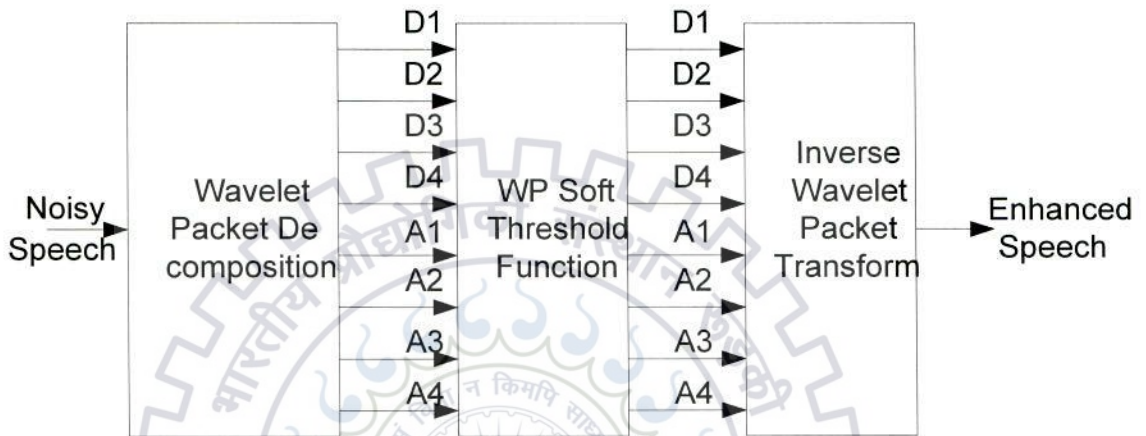


Fig. 2.5: Block diagram of the WPT based speech enhancement method.

The effectiveness of different mother wavelets used for reduction of noise (babble, pink Volvo car and white noise) is measured in terms of Perceptive Evaluation of Speech Quality (PESQ), Cepstrum Distance (CD) and output SNR. The computed values of these quantities for different types of noise with varying SNR levels of noisy speech signals and different types of mother wavelets are presented in Tables 2.1 to 2.4. The observation on these values indicates that the Daubechies family is found to provide promising results. Among the Daubechies family the Db10 wavelet performs well for all input SNR levels of noisy speech signals. Hence, Daubechies wavelet Db10 at 3<sup>rd</sup>, level is selected for signal decomposition in different applications here.

Table 2.1 shows the results for babble noise. The maximum PESQ and output SNR values are obtained for Db10 mother wavelet in comparison with other mother wavelets; whereas, minimum values of CD is obtained for Db10 at all input SNR levels. This illustrates that better quality improvement in noisy speech for Db10 mother wavelet in case of babble noise. For direct illustration of the performance, the graphical and bar graph representation of these values for different types of mother wavelets for babble noise are presented in Figures 2.6 and 2.7, respectively.

Table 2.2 shows the results for pink noise. These results illustrate the better quality improvement in noisy speech for Db10 mother wavelet in case of pink noise also. The graphical and bar graph representation of these values for different types of mother wavelets for pink noise are presented in Figures 2.8 and 2.9, respectively. These again illustrate the best performance of Db10 mother wavelet.

Table 2.3 and 2.4 show the results for Volvo car and white noise. The graphical and bar chart representation for direct illustration of their comparative performance is shown in Figures 2.10 to 2.12. Here again Db10 mother wavelet produces the best results.

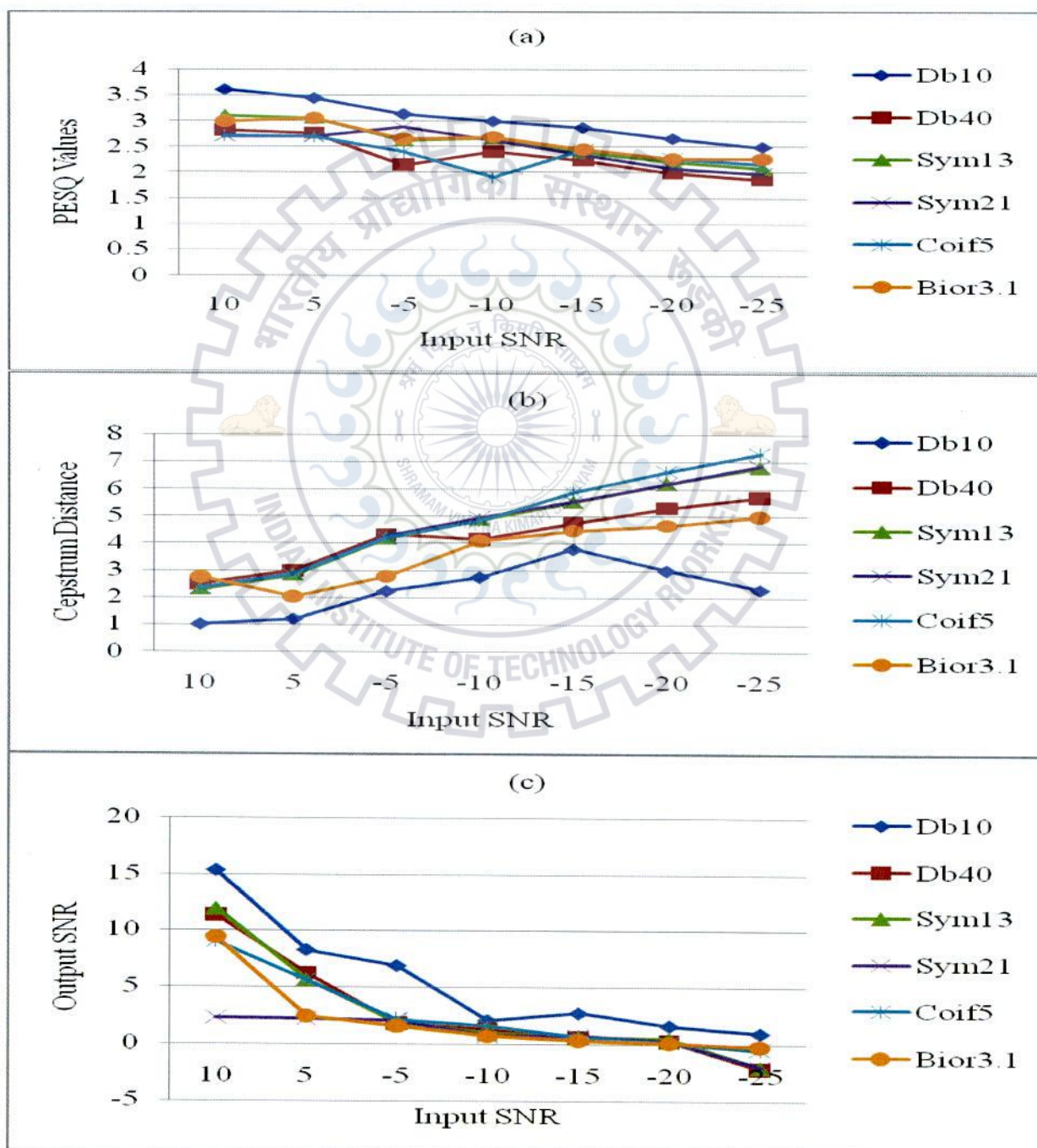


Fig. 2.6: Variation of parameters with Input SNR in babble noise.



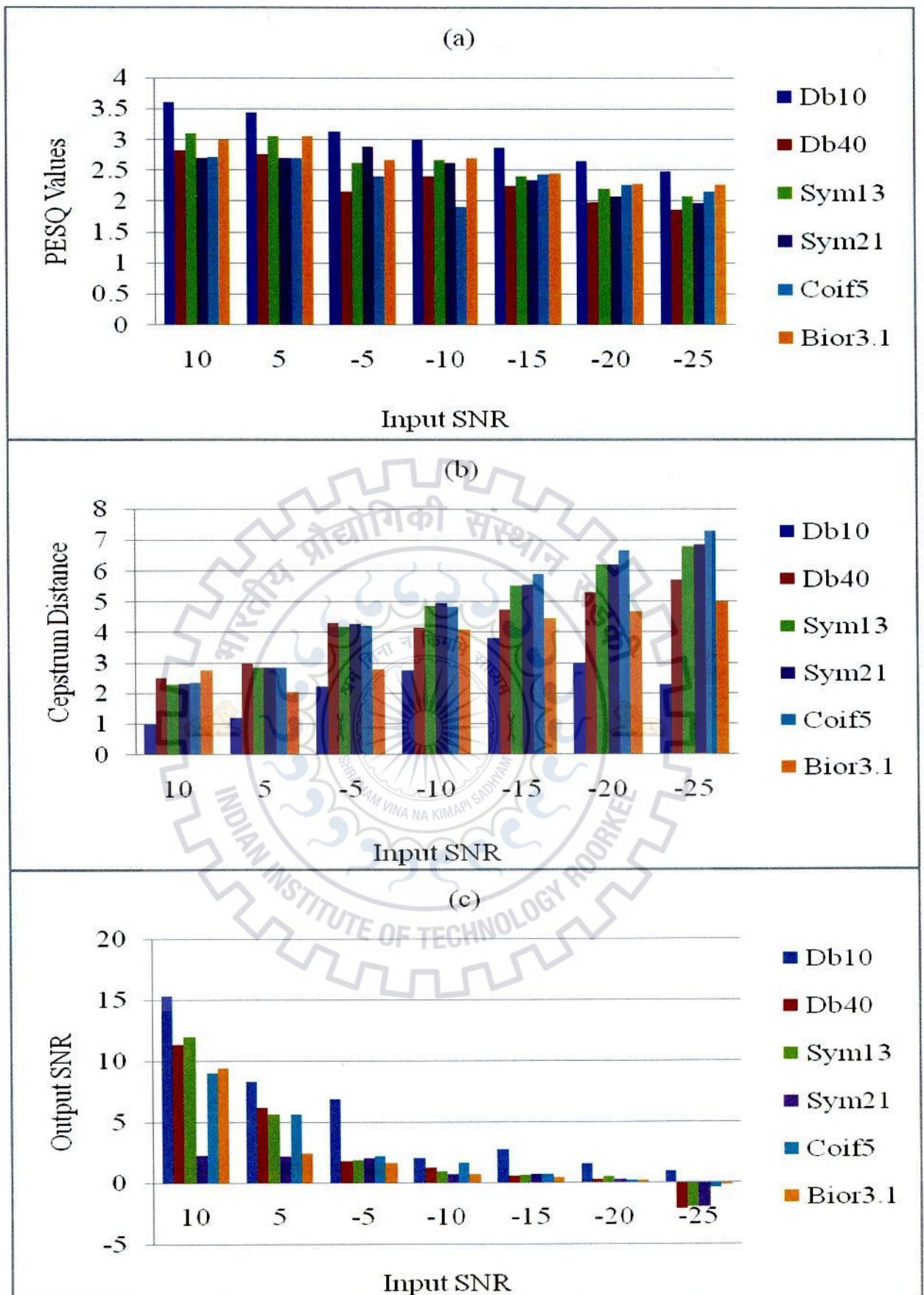


Fig. 2.7: Comparison of mother wavelets in babble noise.

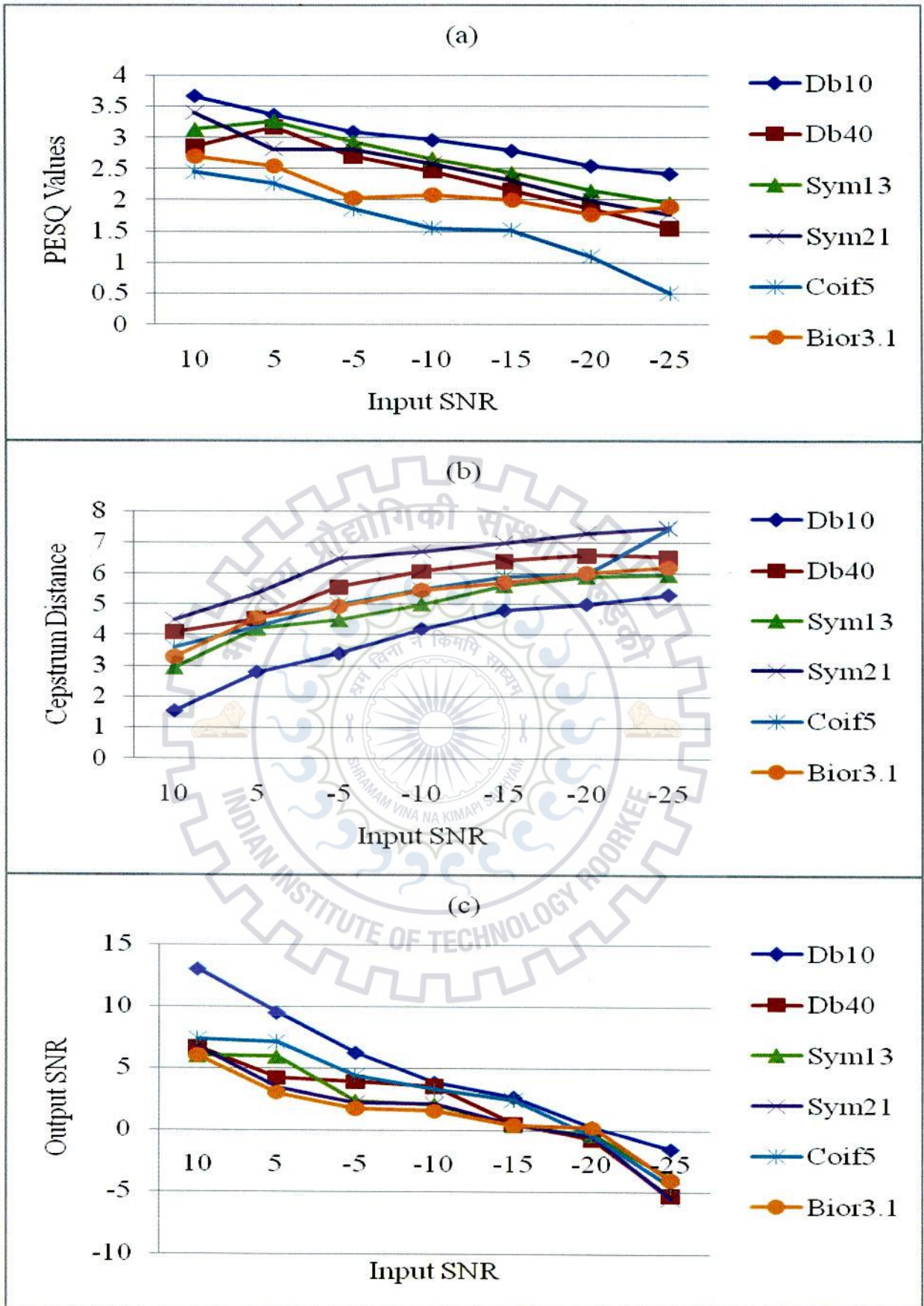


Fig. 2.8: Variation of parameters with Input SNR in pink noise.

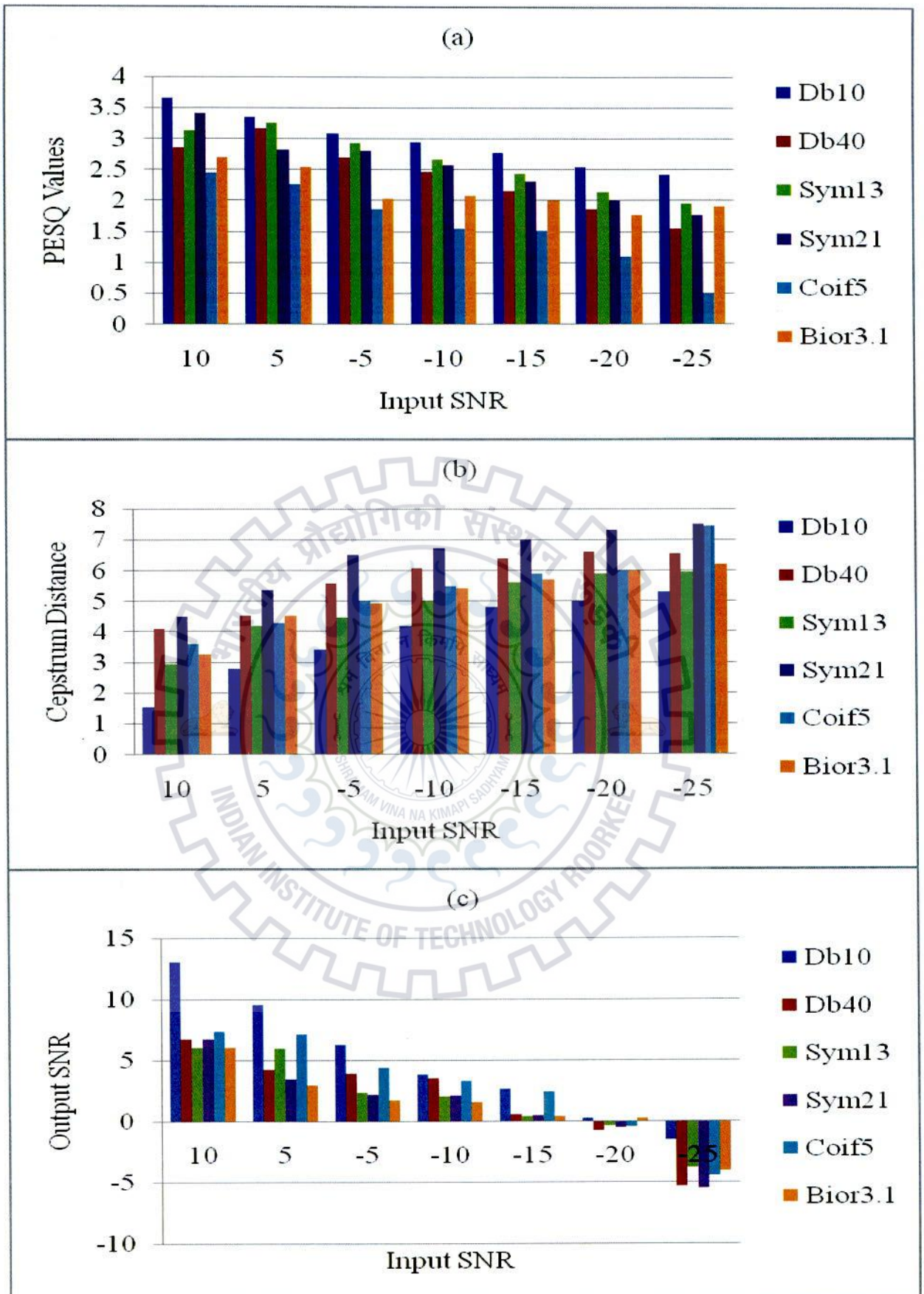


Fig. 2.9: Comparison of mother wavelets in pink noise.

Table 2.1: Analysis for babble noise.

<b>Input SNR</b>	<b>10</b>	<b>5</b>	<b>-5</b>	<b>-10</b>	<b>-15</b>	<b>-20</b>	<b>-25</b>
PESQ values							
<b>Db10</b>	<b>3.6105</b>	<b>3.4464</b>	<b>3.1367</b>	<b>2.9939</b>	<b>2.8698</b>	<b>2.6521</b>	<b>2.4869</b>
Db40	2.82	2.76	2.15	2.4	2.2407	1.9848	1.8659
Sym13	3.1	3.06	2.62	2.6638	2.3972	2.2011	2.0809
Sym21	2.7	2.7	2.8841	2.622	2.339	2.0793	1.9746
Coif5	2.72	2.7	2.4	1.9	2.4393	2.2643	2.15
Bior3.1	3	3.06	2.66	2.6926	2.4549	2.274	2.2669
Cepstrum Distance values							
<b>Db10</b>	<b>1</b>	<b>1.2</b>	<b>2.24</b>	<b>2.75</b>	<b>3.8</b>	<b>3</b>	<b>2.3</b>
Db40	2.5065	2.9905	4.3204	4.16	4.74	5.31	5.7
Sym13	2.303	2.8386	4.1903	4.8784	5.5348	6.1966	6.7894
Sym21	2.3343	2.863	4.276	4.9566	5.5526	6.2128	6.8659
Coif5	2.3423	2.8584	4.2079	4.8318	5.89	6.66	7.3
Bior3.1	2.75	2.0329	2.7848	4.1	4.48	4.67	5
Output SNR values							
<b>Db10</b>	<b>15.3455</b>	<b>8.2506</b>	<b>6.8818</b>	<b>2.04</b>	<b>2.6981</b>	<b>1.5922</b>	<b>0.9351</b>
Db40	11.3537	6.2048	1.7779	1.2518	0.5266	0.2361	-2.2088
Sym13	11.9388	5.596	1.8805	0.9121	0.6194	0.4907	-2.0434
Sym21	2.2647	2.1548	2.0424	0.6981	0.6718	0.2056	-2.0128
Coif5	9	5.6	2.1561	1.6369	0.6767	0.1271	-0.5041
Bior3.1	9.4	2.4559	1.61	0.7096	0.3663	0.1197	-0.2138

Table 2.2: Analysis for Pink noise.

<b>Input SNR</b>	<b>10</b>	<b>5</b>	<b>-5</b>	<b>-10</b>	<b>-15</b>	<b>-20</b>	<b>-25</b>
PESQ values							
<b>Db10</b>	<b>3.66</b>	<b>3.356</b>	<b>3.0823</b>	<b>2.951</b>	<b>2.78</b>	<b>2.5384</b>	<b>2.4128</b>
Db40	2.85	3.1653	2.6999	2.4576	2.1588	1.8621	1.5501
Sym13	3.13	3.261	2.9264	2.6594	2.4295	2.1462	1.9496
Sym21	3.4056	2.82	2.8139	2.5739	2.3092	1.997	1.7723
Coif5	2.45	2.26	1.86	1.55	1.52	1.1	0.5
Bior3.1	2.7	2.54	2.0292	2.08	2	1.77	1.9
Cepstrum Distance values							
<b>Db10</b>	<b>1.54</b>	<b>2.8</b>	<b>3.4</b>	<b>4.2</b>	<b>4.8</b>	<b>5</b>	<b>5.31</b>
Db40	4.1	4.54	5.57	6.08	6.4	6.6	6.53
Sym13	2.94	4.2	4.48	5	5.6	5.89	5.95
Sym21	4.4976	5.3662	6.5035	6.7259	7.0012	7.3175	7.4943
Coif5	3.6	4.29	5	5.5	5.89	6	7.447
Bior3.1	3.2694	4.54	4.93	5.44	5.7	6	6.2
Output SNR values							
<b>Db10</b>	<b>13.0353</b>	<b>9.4925</b>	<b>6.244</b>	<b>3.826</b>	<b>2.6414</b>	<b>0.2506</b>	<b>-1.5848</b>
Db40	6.7003	4.2073	3.904	3.5401	0.5299	-0.7995	-5.354
Sym13	6.0036	5.934	2.3727	2.051	0.3936	-0.3993	-3.7899
Sym21	6.724	3.4303	2.1878	2.0789	0.4451	-0.5524	-5.5344
Coif5	7.3699	7.1293	4.3775	3.3048	2.4323	-0.4987	-4.54
Bior3.1	6.0148	3.0023	1.728	1.5354	0.3795	0.2043	-4.0623

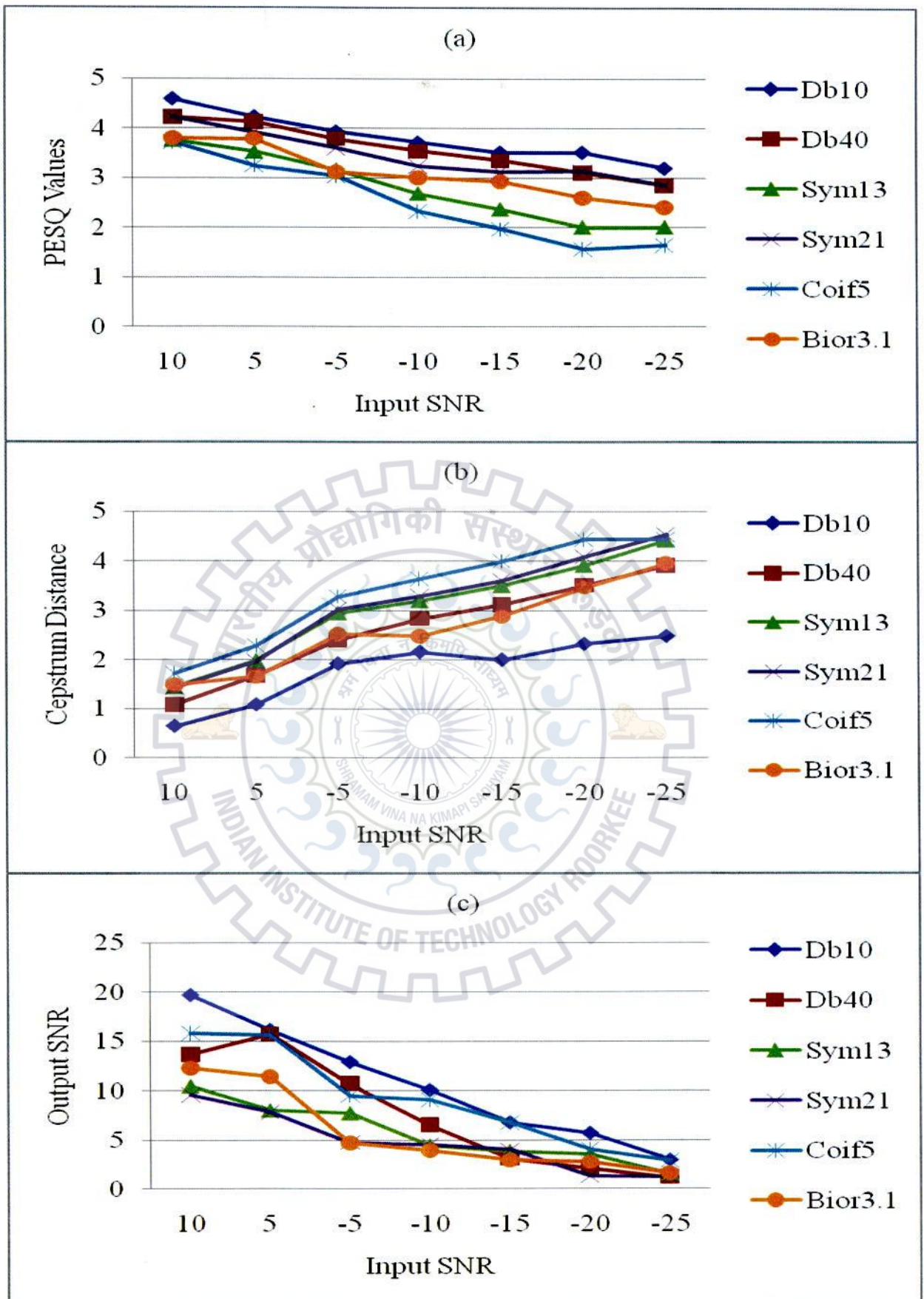


Fig. 2.10: Variation of parameters with Input SNR in Volvo car noise.

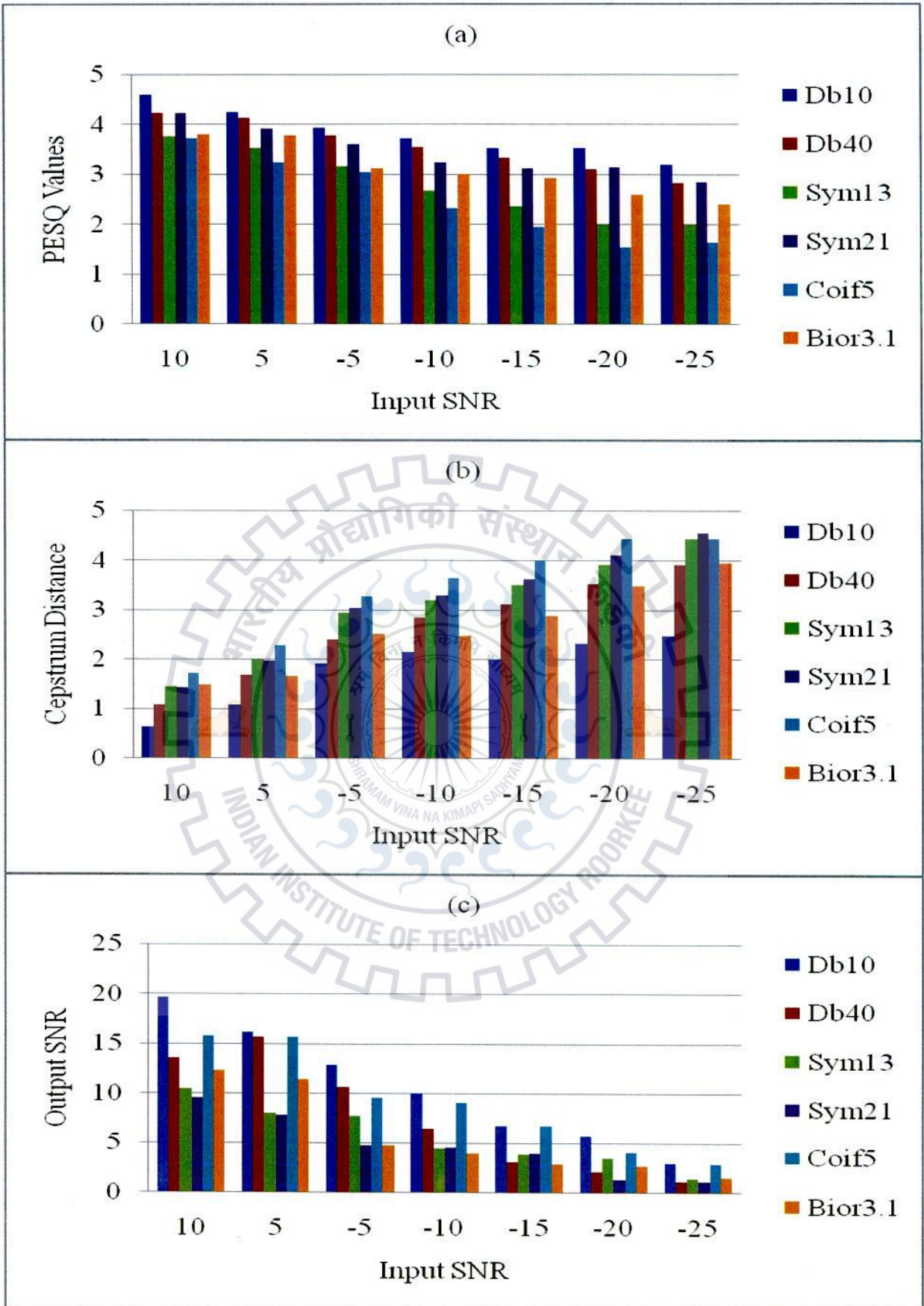


Fig. 2.11: Comparison of mother wavelets in Volvo car noise.

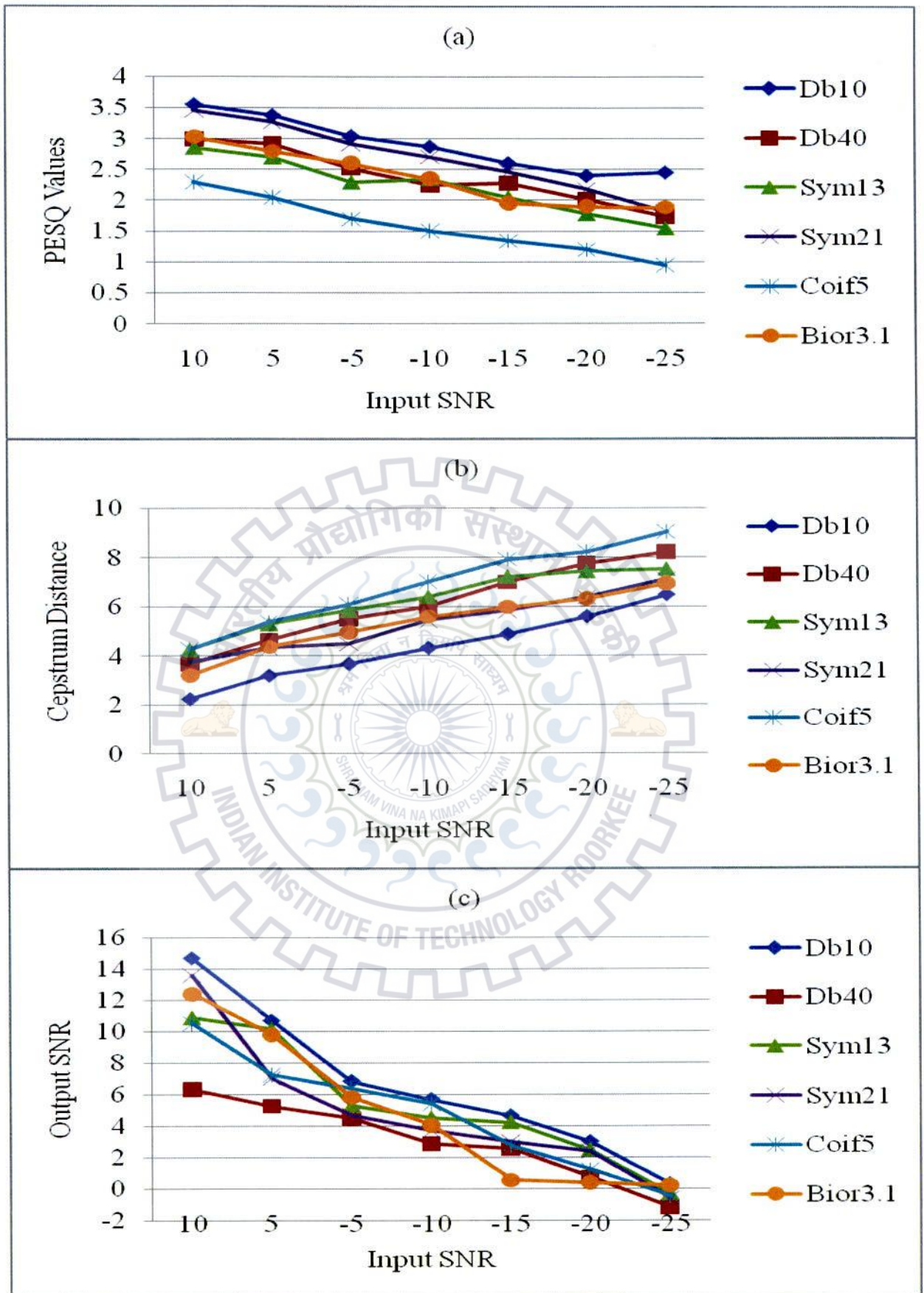


Fig. 2.12: Variation of parameters with Input SNR in white noise.

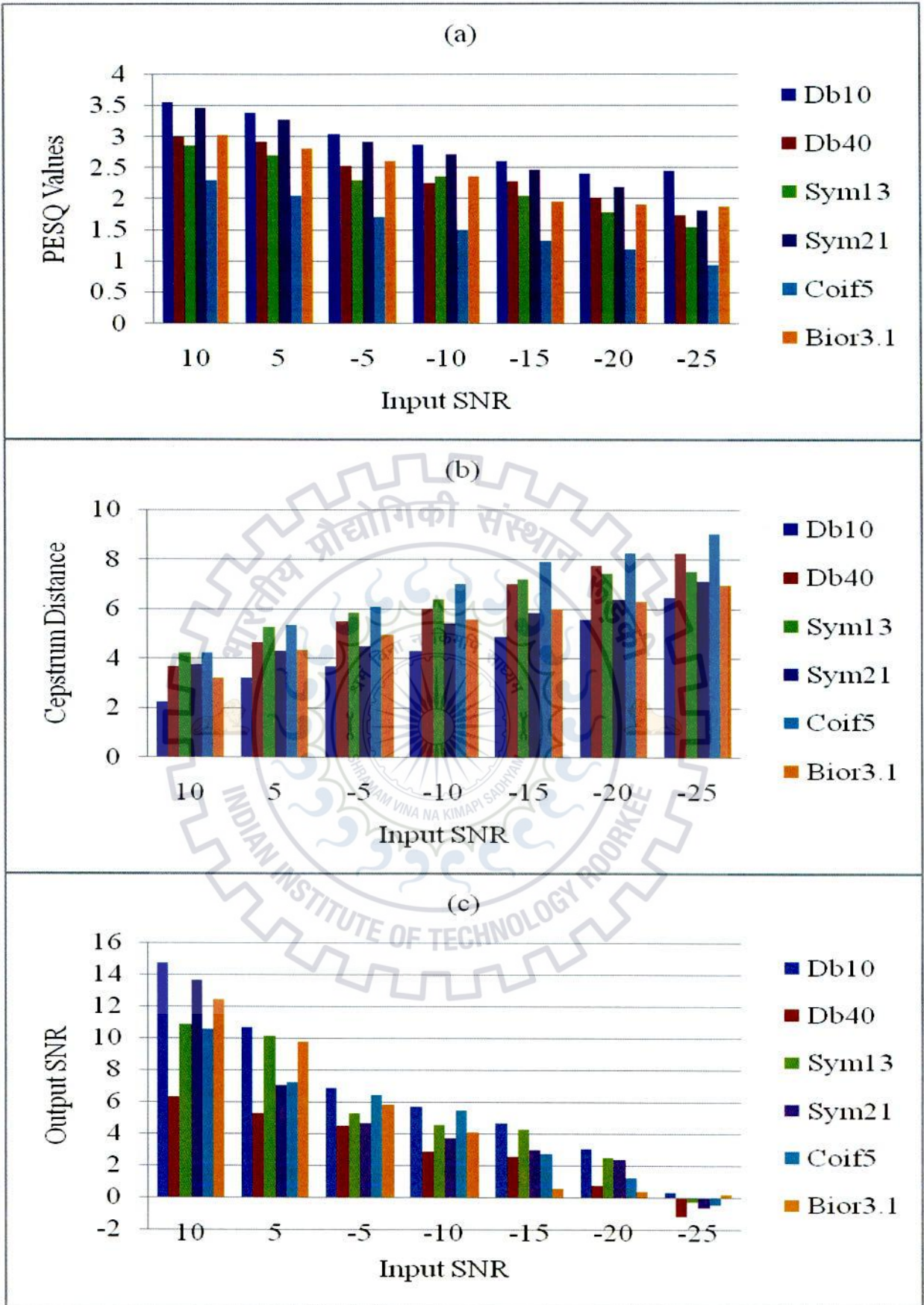


Fig. 2.13: Comparison of mother wavelets in white noise.



Table 2.3: Analysis for Volvo Car noise.

Input SNR	10	5	-5	-10	-15	-20	-25
PESQ values							
Db10	<b>4.6</b>	<b>4.2368</b>	<b>3.9382</b>	<b>3.72</b>	<b>3.52</b>	<b>3.52</b>	<b>3.2</b>
Db40	4.2323	4.1281	3.7872	3.5369	3.3426	3.098	2.8389
Sym13	3.76	3.52	3.16	2.68	2.36	2	2
Sym21	4.2314	3.92	3.6	3.24	3.12	3.1321	2.8427
Coif5	3.72	3.24	3.04	2.32	1.96	1.56	1.64
Bior3.1	3.7977	3.7836	3.12	3	2.92	2.6	2.4
Cepstrum Distance							
Db10	<b>0.64</b>	<b>1.08</b>	<b>1.92</b>	<b>2.16</b>	<b>2</b>	<b>2.32</b>	<b>2.48</b>
Db40	1.08	1.68	2.4	2.84	3.12	3.52	3.92
Sym13	1.4494	1.9874	2.9536	3.2043	3.5113	3.9219	4.4289
Sym21	1.4332	1.9853	3.0368	3.2988	3.6207	4.099	4.5478
Coif5	1.72	2.28	3.28	3.64	4	4.44	4.4375
Bior3.1	1.4834	1.6577	2.52	2.48	2.88	3.48	3.96
Output SNR							
Db10	<b>19.7162</b>	<b>16.1441</b>	<b>12.9289</b>	<b>10.1066</b>	<b>6.821</b>	<b>5.7453</b>	<b>2.9897</b>
Db40	13.6	15.6805	10.689	6.5237	3.1052	2.1434	1.155
Sym13	10.4681	7.997	7.7373	4.4315	3.8627	3.4895	1.4803
Sym21	9.608	7.8315	4.7685	4.547	4.018	1.3146	1.2076
Coif5	15.8298	15.6771	9.5449	9.1344	6.7668	4.0843	2.888
Bior3.1	12.3183	11.4222	4.7113	3.9343	2.906	2.7062	1.5782

Table 2.4: Analysis for White noise.

Input SNR	10	5	-5	-10	-15	-20	-25
PESQ values							
Db10	<b>3.5534</b>	<b>3.3751</b>	<b>3.0405</b>	<b>2.8707</b>	<b>2.6006</b>	<b>2.3985</b>	<b>2.45</b>
Db40	3	2.92	2.52	2.25	2.2837	2.0146	1.7323
Sym13	2.86	2.7	2.3	2.35	2.05	1.78	1.55
Sym21	3.4627	3.2785	2.9217	2.7068	2.4676	2.1827	1.8135
Coif5	2.3	2.05	1.7	1.5	1.34	1.2	0.94
Bior3.1	3.03	2.8	2.6	2.35	1.95	1.9	1.88
Cepstrum Distance							
Db10	<b>2.24</b>	<b>3.2</b>	<b>3.68</b>	<b>4.32</b>	<b>4.88</b>	<b>5.6</b>	<b>6.48</b>
Db40	3.68	4.64	5.52	6	7.015	7.76	8.24
Sym13	4.24	5.28	5.84	6.4	7.2204	7.4307	7.5266
Sym21	3.76	4.32	4.48	5.44	5.84	6.4	7.12
Coif5	4.24	5.36	6.08	7.0175	7.92	8.24	9.04
Bior3.1	3.2	4.3567	4.96	5.6	6	6.32	6.96
Output SNR							
Db10	<b>14.7145</b>	<b>10.6956</b>	<b>6.8557</b>	<b>5.7119</b>	<b>4.6537</b>	<b>3.0067</b>	<b>0.2942</b>
Db40	6.3147	5.2437	4.4711	2.8684	2.5447	0.7121	-1.1963
Sym13	10.8549	10.1112	5.2745	4.529	4.2274	2.4819	-0.3069
Sym21	13.6167	7.0146	4.684	3.7217	2.9885	2.3818	-0.6571
Coif5	10.5273	7.2272	6.4173	5.444	2.7268	1.2108	-0.4603
Bior3.1	12.4235	9.7846	5.8369	4.0406	0.5442	0.3871	0.1947

### 2.3.2 Determination of soft WP threshold function

A noisy single channel Hindi speech signal is modelled as the sum of the clean Hindi speech and two real non-stationary additive background noise for low SNR. This noisy speech is given in eq. (2.21) as:

$$Y(n) = X(n) + D_1(n) + D_2(n) \quad (2.21)$$

Where,  $Y(n)$ ,  $X(n)$  and  $D_i(n)$  denote frames of noisy speech, clean speech and additive background noise, respectively. The Discrete Short-Time Fourier Transform (DSTFT) of the corrupted speech signal is given in eq. (2.22) as:

$$Y(n, k) = X(n, k) + D_1(n, k) + D_2(n, k) \quad (2.22)$$

For decomposition, Db10 mother wavelet packet transform is used upto 3<sup>rd</sup> levels. After getting eight band wavelet packet coefficients, soft WP threshold function is applied in first stage of noise reduction. The soft threshold is defined as:

$$X_1 = \begin{cases} Y - \text{sgn}(Y) & \text{if } |Y| \geq T \\ 0 & \text{if } |Y| < T \end{cases} \quad (2.23)$$

Where,  $Y$  represents the WP coefficients of decomposed noisy speech and  $t$ ,  $T$  are threshold limits which are given as:

$$t = \frac{\text{med}(|Y|)}{0.6745} \quad \text{for hard threshold} \quad (2.24)$$

$$T = t \sqrt{2 \log(Y)} \quad \text{for soft threshold} \quad (2.25)$$

The desired speech signal obtained by using this function is utilized for further noise reduction in proposed WPT Modified Wiener gain function based speech enhancement method.

### 2.3.3 Gain functions and effects

Most of the single-channel speech enhancement algorithms such as Wiener, SS, MMSE-SPU, log-MMSE, p-MMSE etc., use gain functions for noise reduction. For the noise reduction or suppression, a gain function is multiplied with noisy speech spectrum. The gain reduction value depends on estimated ratio from noisy speech and calculated noise in frame. The range of gain function varies in between 0 to 1. The applied gain value will be low i.e.

near to 0 if estimated ratio is high and vice-versa. If the ratio is very high then the effect of gain reduction will be very low.

In prior studies of speech enhancement, Wiener gain function has been used for the noise reduction due to its easy implementation and requirement of less computation [150]. It is necessary to know the effectiveness of gain function for the improvement in quality and intelligibility of speech. A gain function may create distortion in enhanced speech if it is not well designed. Since speech intelligibility is reduced if a gain function generates the distortion in the processed speech.

From the literature, it is confirmed that maximum improvement in quality and intelligibility is obtained by Wiener gain function but it also generates some distortion in the processed speech [150]. Hence, a modified Wiener gain function is proposed in this work. The performance of this proposed method is measured in terms of fw-SSNR and speech intelligibility index (SII).

The process starts with the noisy speech input dataset and each noisy speech sentence is segmented into 20ms frames and 50% overlap between adjacent frames [150]. Hanning window is applied on each speech frame and a modified gain function is obtained by using eqs. (2.26) to (2.30).

The background noise spectrum  $D_{EST}$  is estimated by using Rangachari and Loizou method [150]. The noise spectrum  $D_{EST}$  is estimated as:

$$D_{EST}(n, k) = \alpha_s(n, k) D(n-1, k) + (1 - \alpha_s(n, k)) |Y(n, k)|^2 \quad (2.26)$$

Where,  $\alpha_s$  is a frequency-dependent smoothing factor. This smoothing factor is calculated as per following eq. (2.27):

$$\alpha_s^\Delta = \alpha_d + (1 - \alpha_d) p(n, k) \quad (2.27)$$

where,  $\alpha_d$  is a constant and  $\alpha_s$  takes values in the range of  $\alpha_d \leq \alpha_s \leq 1$ . The  $p(n, k)$  gives the speech - presence probability value as suggested by Rangachari and Loizou [150]:

$$p(n, k) = \alpha_p p(n-1, k) + (1 - \alpha_p) I(n, k) \quad (2.28)$$

Where,  $\alpha_p$  is smoothing constant. The  $I(n, k)$  is either taken as 1 or 0 for speech presence or speech absence, respectively.

$$\gamma = \min((Y / D_{EST}, 20) \quad (2.29)$$

The gain function G is estimated as

$$G = \max(0, (\gamma - 1.3) / \gamma) \quad (2.30)$$

Where,  $Y$  is the noisy speech. The  $\gamma$  is calculated minimum value from noisy speech and estimated noise spectrum.

A modified Wiener gain function is applied at the second stage of the proposed speech enhancement method. For the calculation of denoised speech spectrum, a gain factor  $G$  is required to be multiplied with the speech signal  $X_1$  given in eq. (2.23).

The enhanced speech signal thus can be obtained as mentioned by eq. (2.31):

$$\hat{X} = G * X_1 \quad (2.31)$$

Where,  $\hat{X}$  and  $X_1$  are denoted as enhanced speech and first stage speech spectrum, respectively.

The effect of modified Wiener gain function  $G$  (defined in terms of fw-SSNR and SII parameters) on speech enhancement is presented in Table 2.5 along with the effect of other gain functions for the babble noise for English language.

Table 2.5: Results for gain functions.

Gain function Parameters	Gain function					Modified Wiener gain
	Wiener	Spectral Sub.	MMSE-SPU	log-MMSE	p-MMSE	
fw-SSNR	7.54893	7.44161	6.96737	7.27897	7.03386	<b>9.63616</b>
SII	0.07621	0.08837	0.04723	0.07354	0.09404	<b>0.47239</b>

The gain functions give different values of fw-SSNR and SII, which show that all gains are not equally useful for speech enhancement. It is observed from the table that MMSE-SPU and p-MMSE gain functions give less improvement in processed speech signal whereas the modified Wiener gain function gives maximum improvement.

### 2.3.4 Results and discussion

Hindi language speech patterns have been taken from International Institute of Information Technology Hyderabad (IIIT-H) Indic speech database [138] as clean speech dataset. The IIIT-H indic speech database contains common Indian languages like Hindi, Kannada, Bengali, Malayalam, Tamil, Telgu and Marathi and each language has 1000 clean speech sentences. These 1000 sentences per language are selected to cover 7000 most frequent words in text corpus of the corresponding language. These texts are taken from Wikipedia articles on Indian languages. The text data is made available in IT3 (a

transliteration scheme) as well as in Unicode (UTF-8 format). The noise dataset are taken from NOIZEX-92 database [139] for generating mixed noise signals for analysis. The clean speech patterns of Hindi language have been added with noise sources and speech sentences of NOIZEUS database to generate the dataset for analysis purpose here [140]. The sampling frequency of original speech sentences is 16 kHz. The NOIZEUS clean speech database contains 30 IEEE sentences which have been produced by three male and three female speakers [140]. The input dataset are generated as making combinations like f16 + babble + NOIZEUS speech, machinegun + pink + NOIZEUS speech, and factory floor + white+ NOIZEUS speech mixed with Hindi speech patterns. The generated dataset have been used for evaluation of speech enhancement methods.

Distortions introduced by noise-suppressive gain functions during the process of processing the noisy speech, degrade its quality and intelligibility at low SNR noise conditions. In this section, the proposed speech enhancement algorithms is analyzed and compared with previously discussed methods such as Wiener, spectral subtraction, MMSE-SPU, p-MMSE, log-MMSE, ICS and IdBM. The performance measures such as output SNR, fw-SSNR, PESQ, CD and SII have been used for comparison. Comparative evaluation of the results indicate that the effectiveness of the proposed WPT modified Wiener gain method for improvement in both quality and intelligibility over unprocessed speech stimuli at all input noisy speech of low SNR levels. The mixed noise patterns (i.e. f16 + babble + NOIZEUS clean speech, machinegun + pink + NOIZEUS clean speech and factory floor + white + NOIZEUS clean speech) are used for comparative evaluation of speech enhancement methods. The obtained results for these different groups of signals using different speech enhancement methods are given in Tables from 2.6 to 2.10.

The output SNR values under mixed noise environment for Hindi speech are presented in Table 2.6. The tabulation values are represented in graphical form in Figures 2.14 and 2.15. An observation on Figure 2.14 and 2.15 illustrates that for low SNR speech (-15dB to -5dB). The performance of Wiener method (in terms of SNR improvement) is better among all the existing speech enhancement methods whereas for higher SNR speech signals (0dB to 5dB), spectral subtraction method gives better results. The MMSE based methods produce the poor output SNR comparatively. The Wavelet based methods (i.e. soft, hard, ICS and IDBM) perform better than the above methods. Among these, the hard wavelet threshold function based methods provides maximum SNR value (6.2080 dB) at 5dB input SNR. It is, however, found that although the hard wavelet threshold function based method gives the maximum output SNR but it is relatively poor to the proposed method.

The output SNR given by proposed method is maximum (13.641 dB) at +5 dB input SNR which is shown in case of factory floor and white noise mixed environment. The maximum output SNR in Machinegun + Pink+ NOIZEUS speech+ Hindi patterns and F16 + Babble + NOIZEUS speech+ Hindi patterns are 9.4900 and 7.4173 dB, respectively. The maximum improvement obtained by proposed WPT modified Wiener Gain method is obtained for all input SNR levels. Results show that the maximum speech quality and intelligibility can be obtained by proposed method in mixed noise environment.

Table 2.6: Output SNR values.

Noise Sources	SNR Input	Wiener	Spectral Sub.	p-MMSE	log-MMSE	SOFT	HARD	ICS	IDBM	Proposed Method
<b>F16 + Babble+ NOIZEUS speech+ Hindi patterns</b>	-15	-2.3280	-6.3402	-2.7278	-1.4599	0.1074	0.1227	-0.341	1.7234	<b>1.9144</b>
	-10	-1.7747	-3.9115	-3.2184	-1.6779	0.3564	0.3975	0.4341	1.9081	<b>4.9554</b>
	-5	-0.4984	-1.2080	-1.2184	-1.0523	1.1027	1.1781	1.6116	2.1099	<b>5.0651</b>
	0	0.5938	0.9557	0.9455	0.3557	2.9073	3.0064	2.3550	2.2943	<b>8.7918</b>
	5	0.9135	3.6325	2.9951	1.2099	6.0987	6.2080	2.7208	2.3954	<b>7.4173</b>
<b>Machinegun + Pink+ NOIZEUS speech+ Hindi patterns</b>	-15	-0.7753	-5.0933	-1.7526	-1.6779	0.0951	0.1188	-0.340	1.9557	<b>5.1966</b>
	-10	-0.1275	-4.3398	-1.8198	-1.4599	0.3175	0.3849	0.4341	2.1763	<b>6.5033</b>
	-5	1.3932	-0.2685	-1.0912	-1.0523	1.0053	1.1570	1.6116	2.3215	<b>6.5157</b>
	0	3.2380	3.2606	0.7479	0.3557	2.7176	2.9725	2.3550	2.3828	<b>6.9647</b>
	5	3.2257	6.0620	4.4728	1.2099	5.8121	6.1535	2.7208	2.4372	<b>9.4900</b>
<b>Factory floor +White+ NOIZEUS speech+ Hindi patterns</b>	-15	-0.3544	-0.8270	-1.8356	-0.8670	0.0914	0.1186	0.2282	1.8533	<b>2.5929</b>
	-10	-0.2839	-0.6202	-1.6846	-0.7145	0.3095	0.3840	1.2293	2.0681	<b>3.7844</b>
	-5	0.3937	-0.0985	0.3967	0.2028	0.9900	1.1596	2.5925	2.2488	<b>9.4924</b>
	0	0.4275	2.1069	1.8154	0.8680	2.6975	2.9735	3.9886	2.3678	<b>9.6020</b>
	5	0.4498	4.5250	3.7515	0.9420	5.7831	6.1485	4.8562	2.4343	<b>13.641</b>

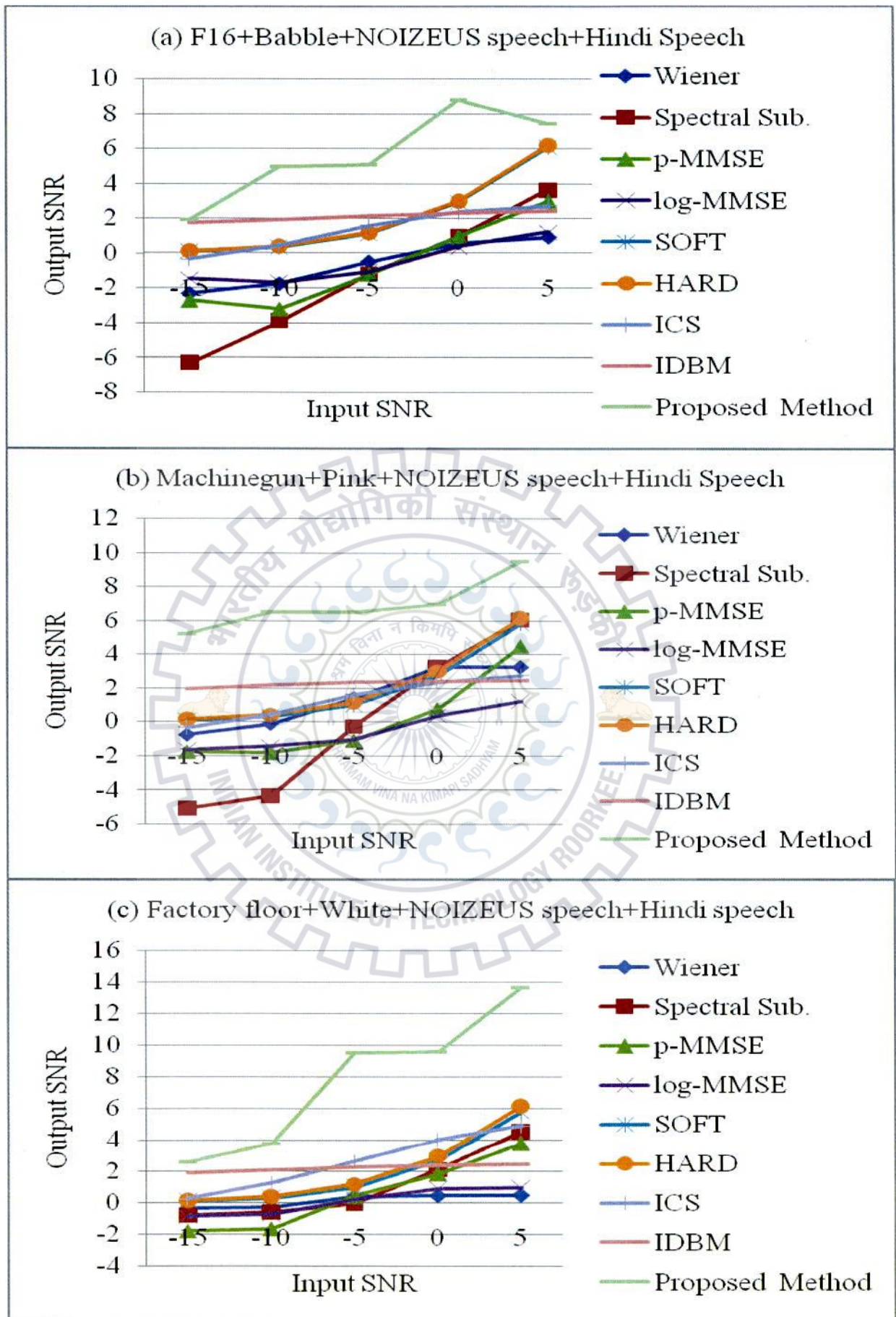


Fig. 2.14: Variation of output SNR with Input SNR in mixed noise.

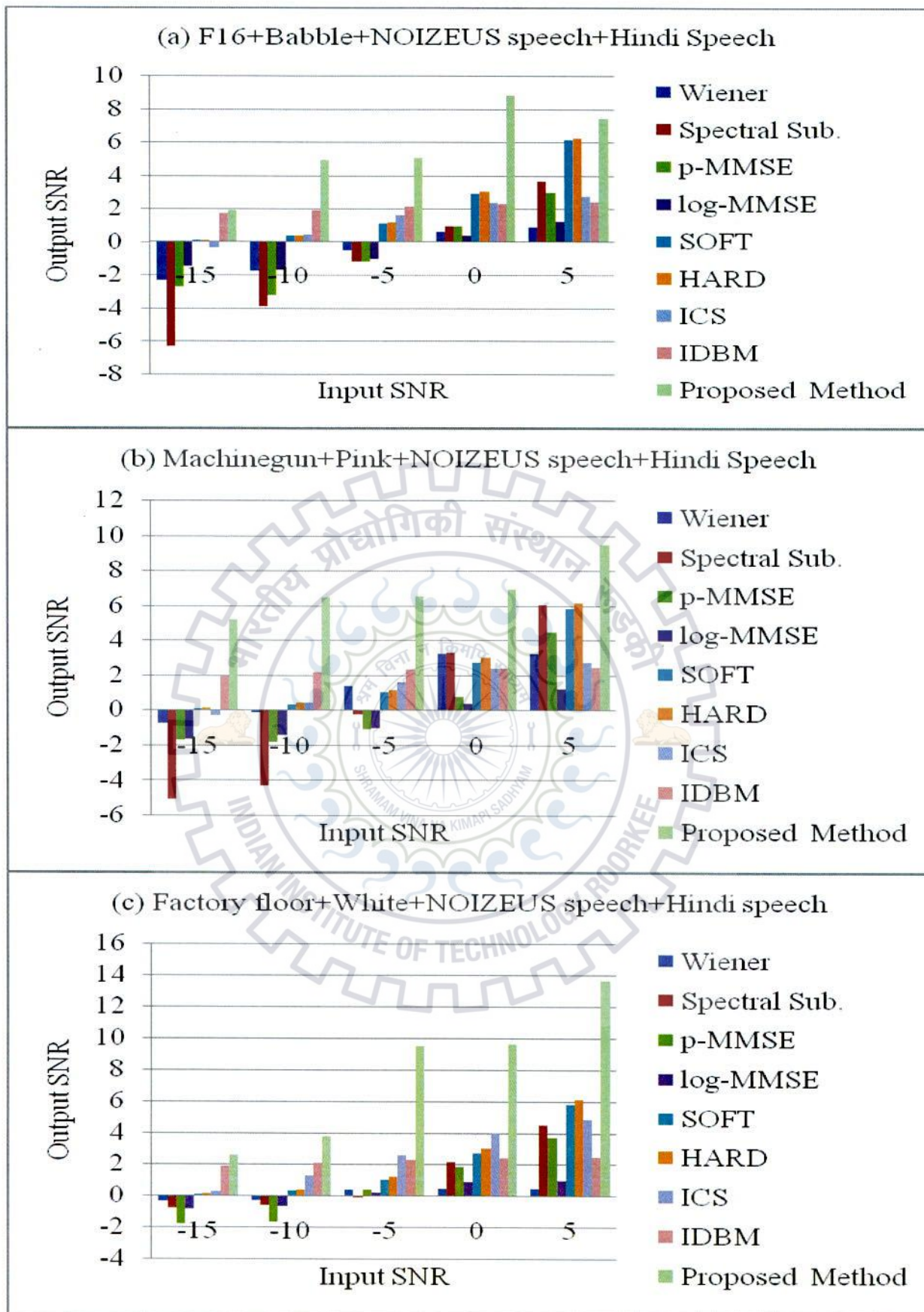


Fig. 2.15: Comparison of speech enhancement methods in terms of Output SNR.



Table 2.7: fw-SSNR values.

↓ Noise Sources	SNR Input	Wiener	Spectral Sub.	p-MMSE	log-MMSE	SOFT	HARD	ICS	IDBM	Proposed Method
<b>F16+ Babble+ NOIZEUS speech+Hindi patterns</b>	-15	5.0571	3.1652	4.0320	3.6735	-0.748	-0.147	5.5686	5.3070	<b>10.5472</b>
	-10	6.0642	3.5508	4.9666	4.7690	-0.446	0.1070	6.3331	6.3098	<b>11.5256</b>
	-5	7.0349	4.3527	5.7113	5.7564	0.1922	0.7147	7.1150	7.9000	<b>12.6038</b>
	0	7.8101	5.3783	6.8051	6.5908	1.3700	1.8111	7.8795	9.6244	<b>13.7704</b>
	5	8.5420	6.2126	7.6541	7.4505	3.1060	3.4668	8.5125	11.0828	<b>15.2217</b>
<b>Machinegun+ Pink+ NOIZEUS speech+Hindi patterns</b>	-15	5.8261	3.6454	4.5390	4.3748	0.0777	0.8930	6.1386	5.6329	<b>11.1857</b>
	-10	7.0447	3.8034	5.5766	5.4821	0.6654	1.4324	6.9527	7.0784	<b>12.1470</b>
	-5	8.1942	4.2869	6.5798	6.5276	1.6390	2.3279	8.0529	8.6422	<b>13.1690</b>
	0	8.9267	5.3417	7.5296	7.6897	3.0707	3.7346	9.2448	10.6607	<b>14.5452</b>
	5	9.7453	6.5323	8.1045	8.4328	5.0220	5.7362	10.407	12.7112	<b>16.1519</b>
<b>Factory floor+ White+ NOIZEUS speech+Hindi patterns</b>	-15	4.4029	3.4376	4.8502	4.5135	0.1634	1.0884	4.9541	5.5395	<b>11.0368</b>
	-10	5.5310	4.4487	5.6798	5.5474	0.6235	1.5122	5.9086	6.7019	<b>11.9830</b>
	-5	6.6120	5.2682	6.8295	6.7192	1.3315	2.2791	6.9053	8.6436	<b>13.0345</b>
	0	7.4657	6.2192	7.5922	7.7587	2.7032	3.6208	8.1404	10.5455	<b>14.2575</b>
	5	8.4698	7.3728	8.6906	8.3784	4.5114	5.5029	9.3946	12.3724	<b>15.7827</b>

In Table 2.7, fw-SSNR values are given at various level of input SNR for different speech enhancement methods. The Figures 2.16 and 2.17 present the graphical and bar chart representation of the values given in Table 2.7. These results shows that the Wiener method gives better results in comparison to spectral subtraction, p-MMSE, log-MMSE, SOFT, HARD but results given by above methods are not better than the ICS, IDBM and proposed methods. The maximum value of fw-SSNR for ICS, IDBM and proposed methods are 10.407, 12.71 and 16.1519, respectively. Among these three, the performance of proposed method is superior to other speech enhancement methods. The higher value of fw-SSNR (about 16) shows good improvement in quality of enhanced speech while lower values (less than 10) show less improvement in quality of speech.

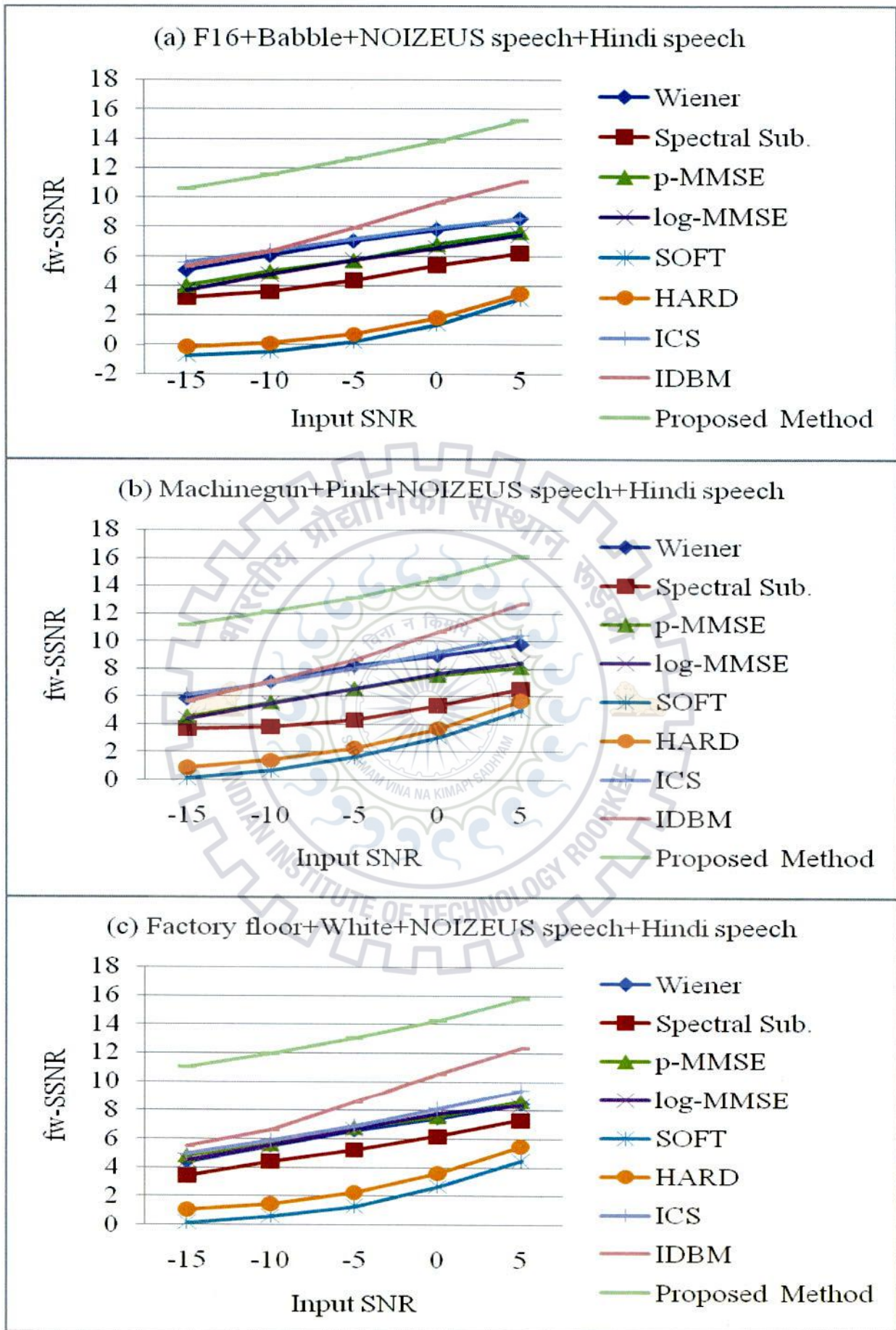


Fig. 2.16: Variation of fw-SSNR with Input SNR in mixed noise.

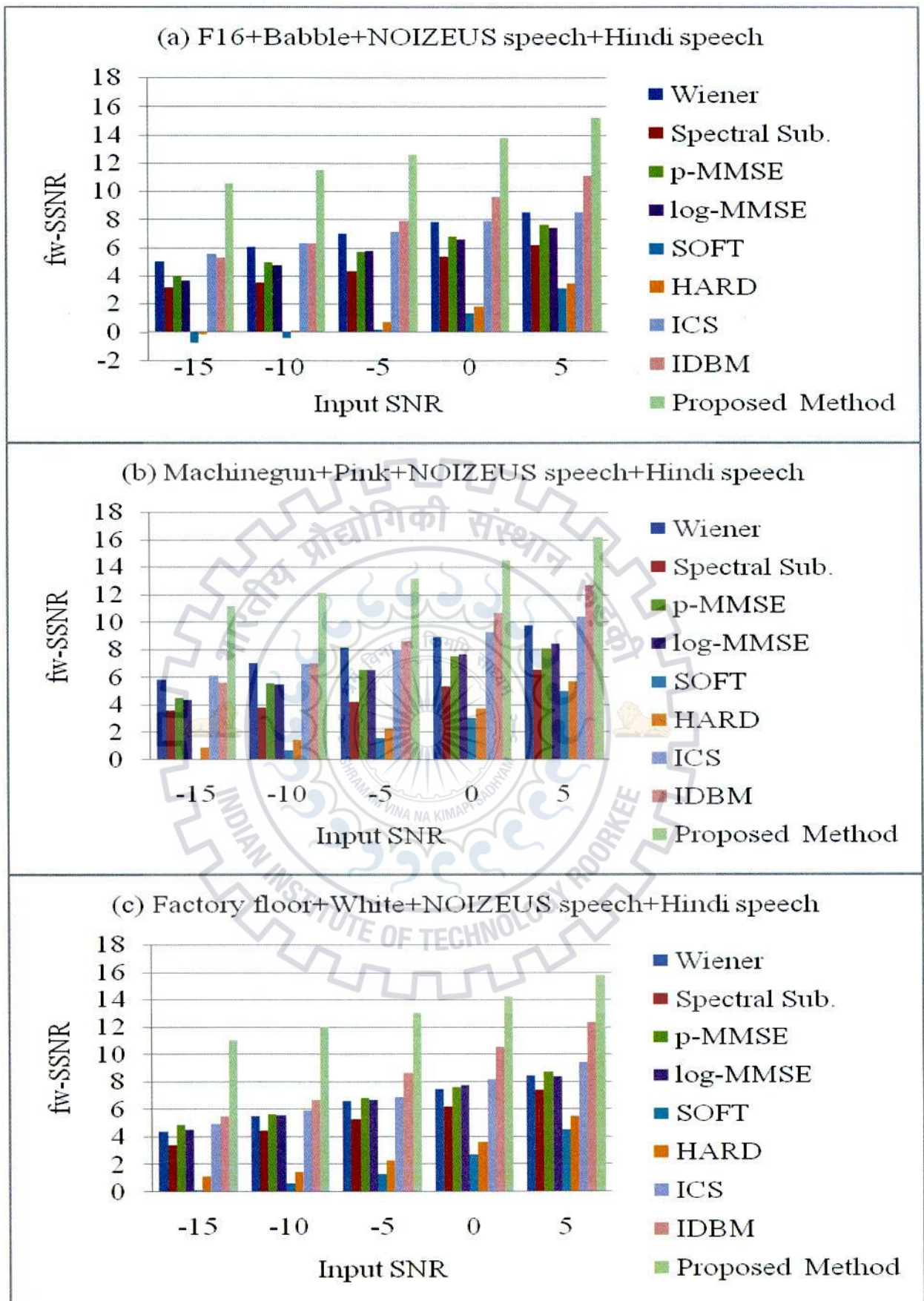


Fig. 2.17: Comparison of speech enhancement methods in terms of fw-SSNR.



In Table 2.8, PESQ values are illustrated for Hindi speech signal in mixed noise environment. In comparison, Wiener method is best in terms of PESQ parameter than existing speech enhancement methods (such as spectral subtraction, p-MMSE, log-MMSE, SOFT, HARD) but the results given by Wiener method are poor to ICS, IDBM and proposed methods. The performance given by proposed method is much better than the existing speech enhancement methods. The PESQ parameter value (3.7283) obtained by proposed WPT modified Wiener based speech enhancement method is highest in mixed noise. Here, figures 2.18 and 2.19 present the graphical and bar chart representation of the values given in Table 2.8. The comparison given in these figures also illustrates that the proposed method gives better performance in comparison with existing speech enhancement methods at all input SNR levels and for all types of noise. The maximum PESQ values obtained by proposed method indicate for the maximum speech quality and intelligibility improvement.

Table 2.8: PESQ values.

↓ Noise Sources	SNR Input	Wiener	Spectral Sub.	p-MMSE	log-MMSE	SOFT	HARD	ICS	IDBM	Proposed Method
<b>F16 + Babble+ NOIZEUS speech+ Hindi patterns</b>	-15	1.4113	0.7360	1.1122	0.8631	0.6088	0.5178	1.6374	2.1603	<b>2.9461</b>
	-10	1.7369	1.2246	1.6300	1.5443	0.6570	0.6227	1.9248	2.4629	<b>3.1161</b>
	-5	1.9502	1.5909	1.9117	1.9385	0.9095	0.8751	2.0734	2.7196	<b>3.3128</b>
	0	2.0538	1.9101	2.1188	2.0642	1.1928	1.1543	2.2435	2.9735	<b>3.4904</b>
	5	2.1666	2.1459	2.2772	2.1359	1.5488	1.5081	2.3775	3.2065	<b>3.6692</b>
<b>Machinegun+ Pink+ NOIZEUS speech+ Hindi patterns</b>	-15	1.7724	0.7781	1.0428	0.9297	0.3475	0.3362	2.0334	2.4390	<b>3.0693</b>
	-10	2.1900	1.2577	1.7752	1.7312	0.6439	0.6150	2.3538	2.6606	<b>3.2843</b>
	-5	2.3156	1.5517	2.1287	2.0665	1.0599	1.0054	2.6068	2.9464	<b>3.4962</b>
	0	2.4644	1.9179	2.3978	2.2364	1.5037	1.4469	2.8100	3.1747	<b>3.6290</b>
	5	2.3163	2.2206	2.5526	2.4324	1.9344	1.8887	3.0038	3.4010	<b>3.7283</b>
<b>Factory floor+ White+ NOIZEUS speech+ Hindi patterns</b>	-15	1.2016	0.5885	1.3526	1.2638	0.5855	0.4971	1.6822	2.3729	<b>2.9461</b>
	-10	1.5249	1.0285	1.6659	1.5962	0.8421	0.7599	1.9231	2.6667	<b>3.1161</b>
	-5	1.8298	1.4618	1.9585	1.9130	1.1432	1.0697	2.1347	2.9649	<b>3.3128</b>
	0	1.9031	1.8881	2.1148	2.0634	1.5136	1.4554	2.3160	3.1514	<b>3.4904</b>
	5	2.0063	2.1964	2.3665	2.0749	1.9444	1.8910	2.4765	3.4042	<b>3.6692</b>

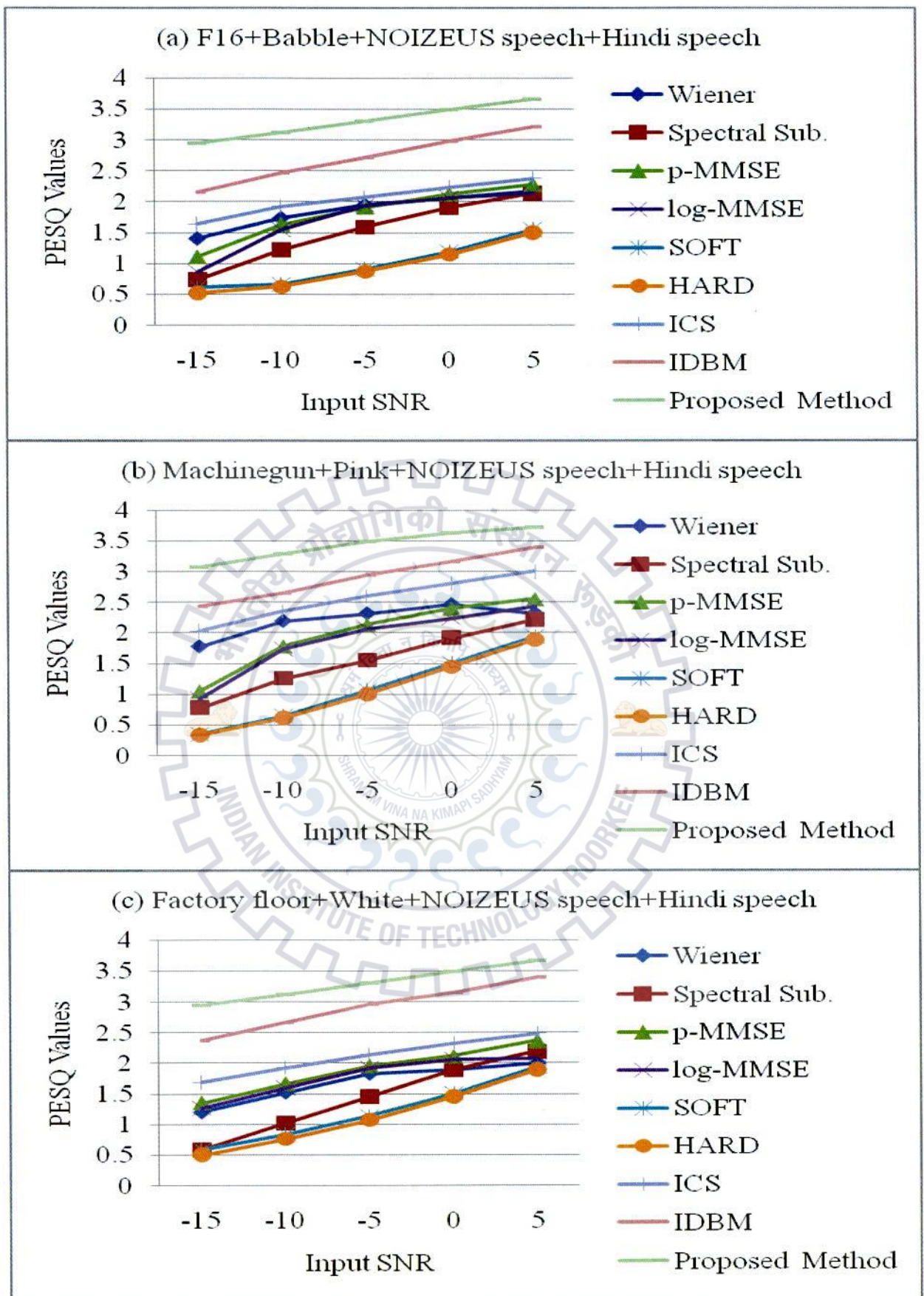


Fig. 2.18: Variation of PESQ with Input SNR in mixed noise.

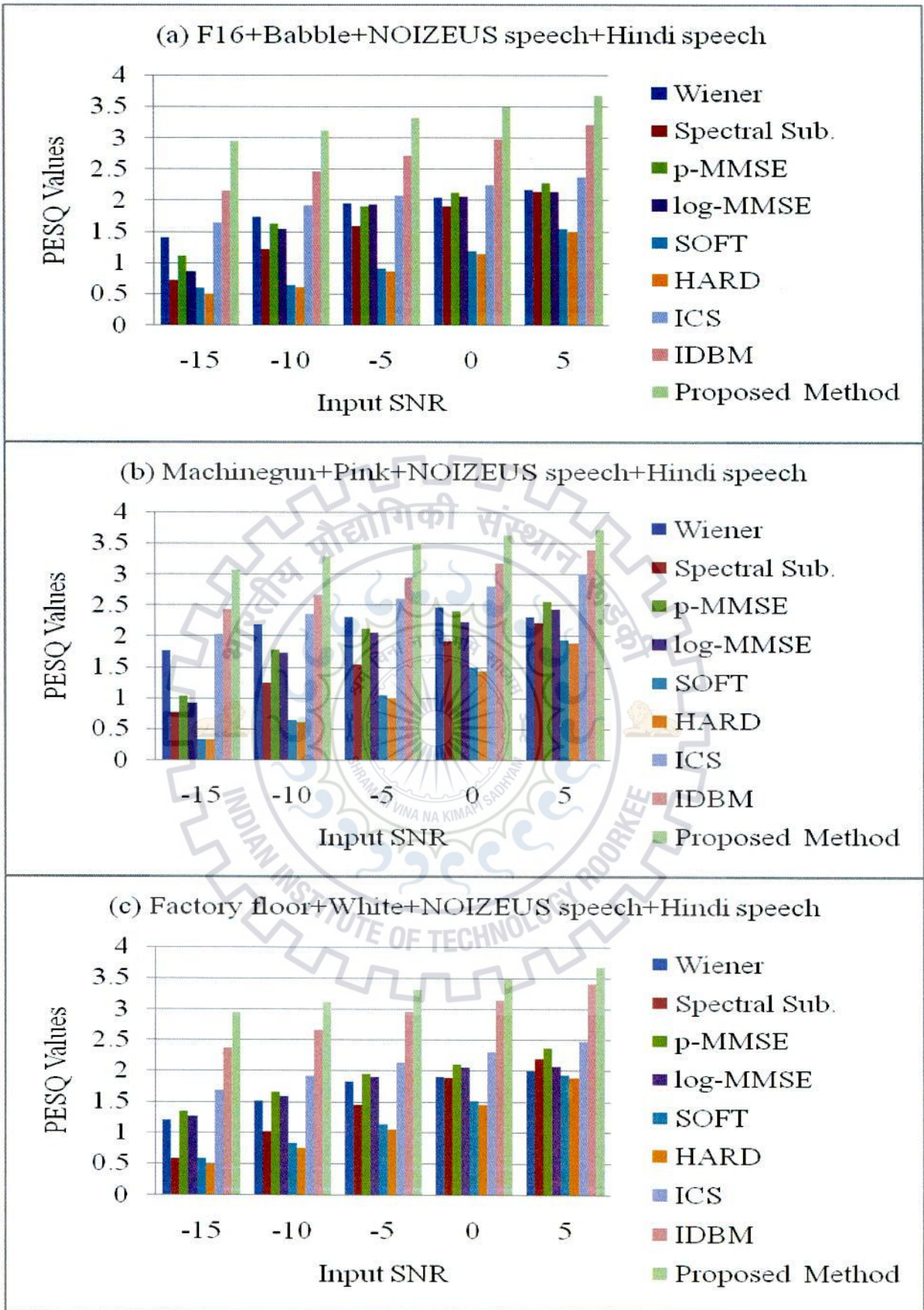


Fig. 2.19: Comparison of speech enhancement methods in terms of PESQ values.

Table 2.9, illustrates the output Cepstral Distance (CD) values for all levels of input SNR in mixed noise environment. The lower value of CD shows better speech enhancement. Wiener method is much better than other existing speech enhancement methods since the minimum value given by Wiener method is 4.1148. The minimum value given by Wiener method shows maximum improvement in speech quality but this is again poor to ICS, IDBM and WPT modified Wiener gain based proposed method. Among these methods, the proposed method gives minimum value of CD (2.4343) and hence maximum improvement in speech quality and intelligibility. The value of CD parameter obtained by proposed method is lower in comparison to other speech enhancement methods for all types of mixed noise. This shows that the output speech signal has maximum improvement in quality and intelligibility. The graphical and bar chart representation for direct illustration of their comparative performance is shown in Figures 2.20 and 2.21.

Table 2.9: Cepstral Distance measure values.

↓ Noise Sources	SNR Input	Wiener	Spect. Sub.	p-MMSE	log-MMSE	SOFT	HARD	ICS	IDBM	Proposed Method
<b>F16 + Babble+ NOIZEUS speech+ Hindi patterns</b>	-15	6.3712	8.9649	6.4998	6.8541	8.9740	8.1510	5.7062	6.6023	<b>3.8327</b>
	-10	5.7469	8.3147	5.7121	6.0816	8.8841	8.0589	5.1584	6.0197	<b>3.5062</b>
	-5	5.0810	7.7327	5.2491	5.3721	8.6720	7.8635	4.5837	5.1234	<b>3.1802</b>
	0	4.5786	6.9332	4.6580	4.7810	8.3275	7.6145	4.1056	4.0709	<b>2.8688</b>
	5	4.1148	6.2579	4.2262	4.3377	7.8272	7.2667	3.7389	3.4918	<b>2.4457</b>
<b>Machinegun + Pink+ NOIZEUS speech +Hindi patterns</b>	-15	6.9649	9.2130	6.9907	7.2290	8.7038	7.7673	6.2689	6.7802	<b>3.7350</b>
	-10	6.3223	8.6247	6.1562	6.4585	8.6388	7.6427	5.6466	6.0449	<b>3.4162</b>
	-5	5.6094	8.1748	5.4169	5.7169	8.4468	7.4515	4.9943	5.2407	<b>3.0929</b>
	0	4.9475	7.2596	4.8335	5.0060	8.1577	7.2002	4.3914	4.3674	<b>2.8114</b>
	5	4.3986	6.3997	4.3537	4.5193	7.6599	7.0104	3.8266	3.5898	<b>2.4343</b>
<b>Factory floor + White+ NOIZEUS speech</b>	-15	6.9740	7.4833	6.8650	7.0829	8.9829	7.6844	6.4056	6.9002	<b>3.8550</b>
	-10	6.3688	7.0698	6.2460	6.5174	8.8797	7.6085	5.8588	6.3690	<b>3.5749</b>
	-5	5.7123	6.5540	5.5772	5.8053	8.6688	7.4638	5.1973	5.4502	<b>3.2958</b>
	0	5.0298	5.9057	4.9941	5.1695	8.2869	7.2162	4.5903	4.6898	<b>3.0416</b>
	5	4.4418	5.1847	4.4824	4.6432	7.8035	7.0270	3.9377	4.1103	<b>2.7106</b>

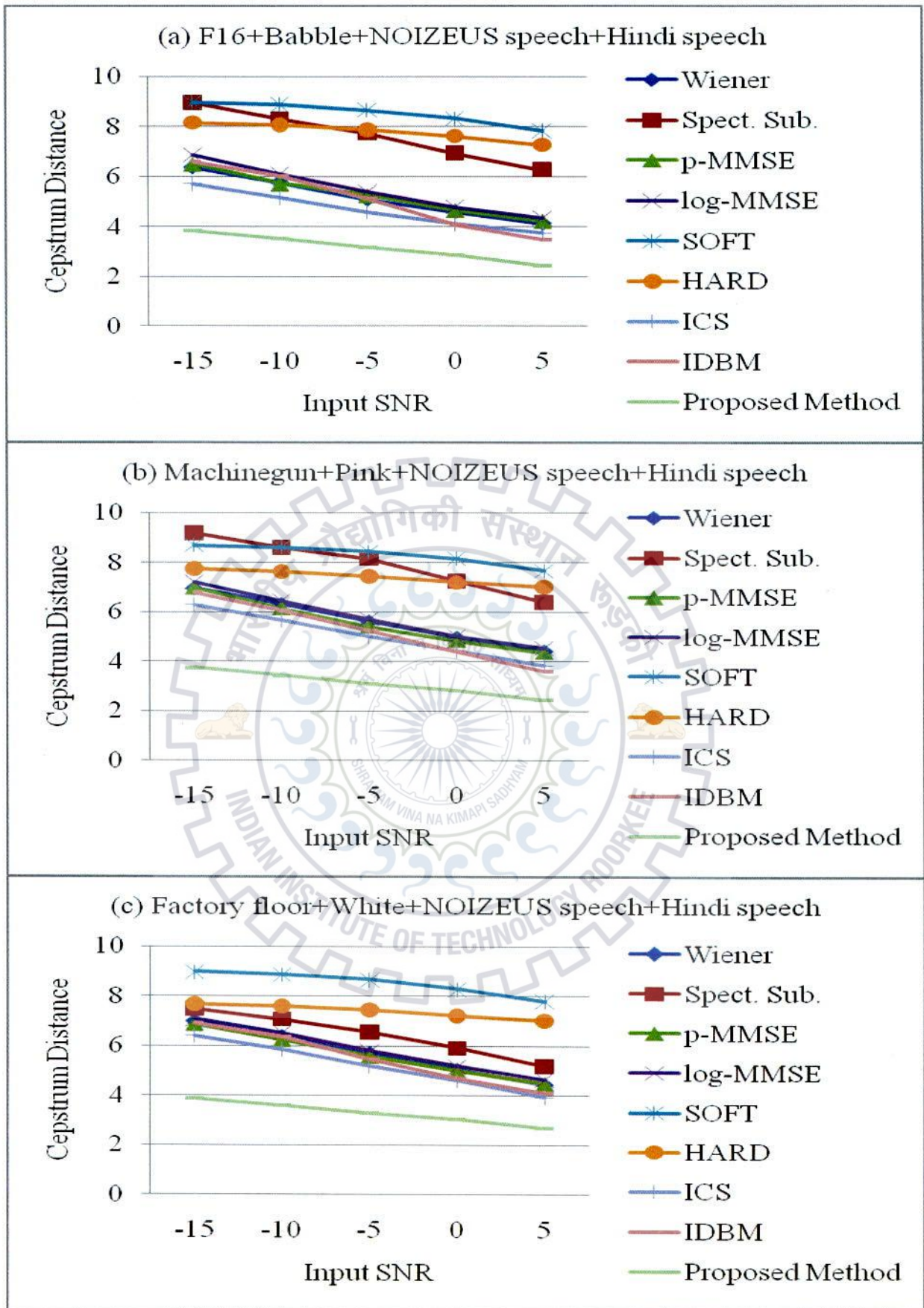


Fig. 2.20: Variation of Cepstrum Distance with Input SNR in mixed noise.



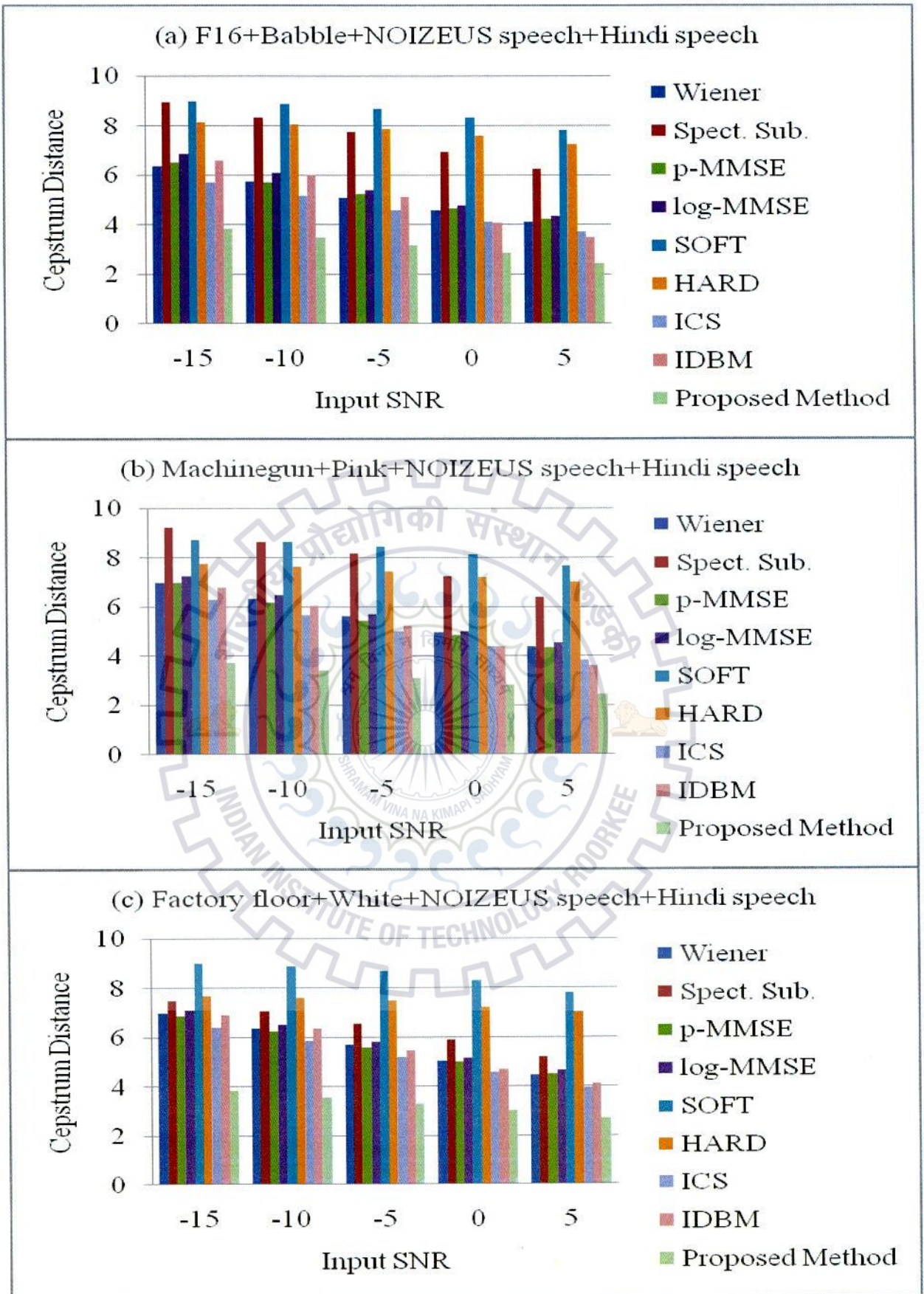


Fig. 2.21: Comparison of speech enhancement methods in terms of Cepstrum Distance.

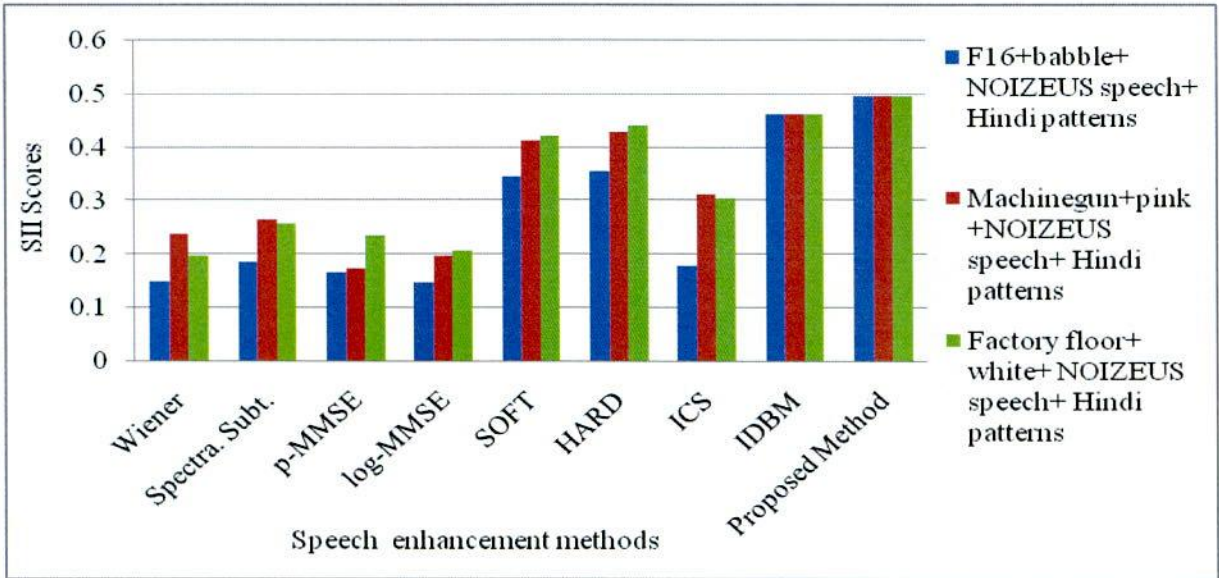


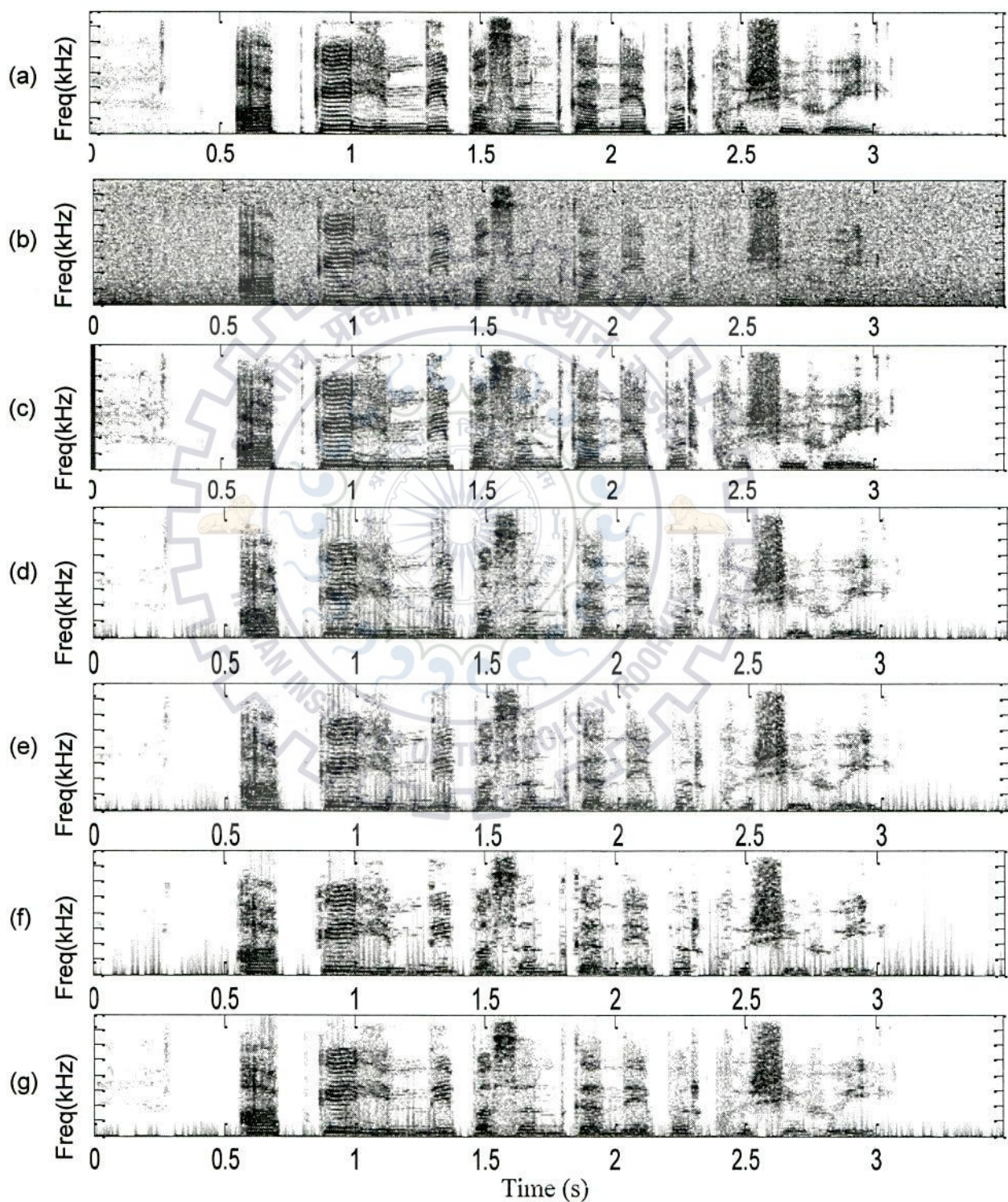
Fig. 2.22: Variation of SII values with different speech enhancement methods and mixed noise.

SII parameter is used to measure the intelligibility of speech signal. The range of SII is from 0 to 1 scale. Where, 0 shows the lower intelligibility and near to 1 for maximum intelligible speech. The output SII values for mixed noise case are illustrated in Table 2.10 for Hindi speech. The maximum SII value of 0.4442 is obtained in machinegun + pink noise mixed environment at +5 dB input SNR. This maximum improvement in SII parameter is given by proposed WPT modified Wiener Gain method in comparison to commonly used speech enhancement methods and hence maximum speech intelligibility is achieved by proposed method in low SNR noise conditions.

Table 2.10: Speech Intelligibility Index (SII) values.

↓ Noise Types	SNR Input	Wiener	Spect. Subt.	p-MMSE	log-MMSE	SOFT	HARD	ICS	IDBM	Proposed Method
<b>F16 + Babble + NOIZEUS speech + Hindi patterns</b>	5dB	0.1499	0.1854	0.1676	0.1485	0.3444	0.3549	0.1779	0.4618	0.4942
<b>Machinegun + Pink+ NOIZEUS speech+ Hindi patterns</b>	5dB	0.2391	0.2633	0.1731	0.1972	0.4118	0.4292	0.3129	0.4618	0.4942
<b>Factory floor + White + NOIZEUS speech + Hindi patterns</b>	5dB	0.1979	0.2568	0.2356	0.2069	0.4220	0.4407	0.3044	0.4618	0.4942

Figure 2.22 shows the comparative analysis of various speech enhancement methods in terms of SII at 5dB SNR of mixed noise. In this comparison, all speech enhancement methods show less improvement in speech intelligibility index than proposed WPT modified Wiener Gain method. Hence, it is concluded that the proposed method is very much effective for improving quality and intelligibility in low SNR mixed noise environment.



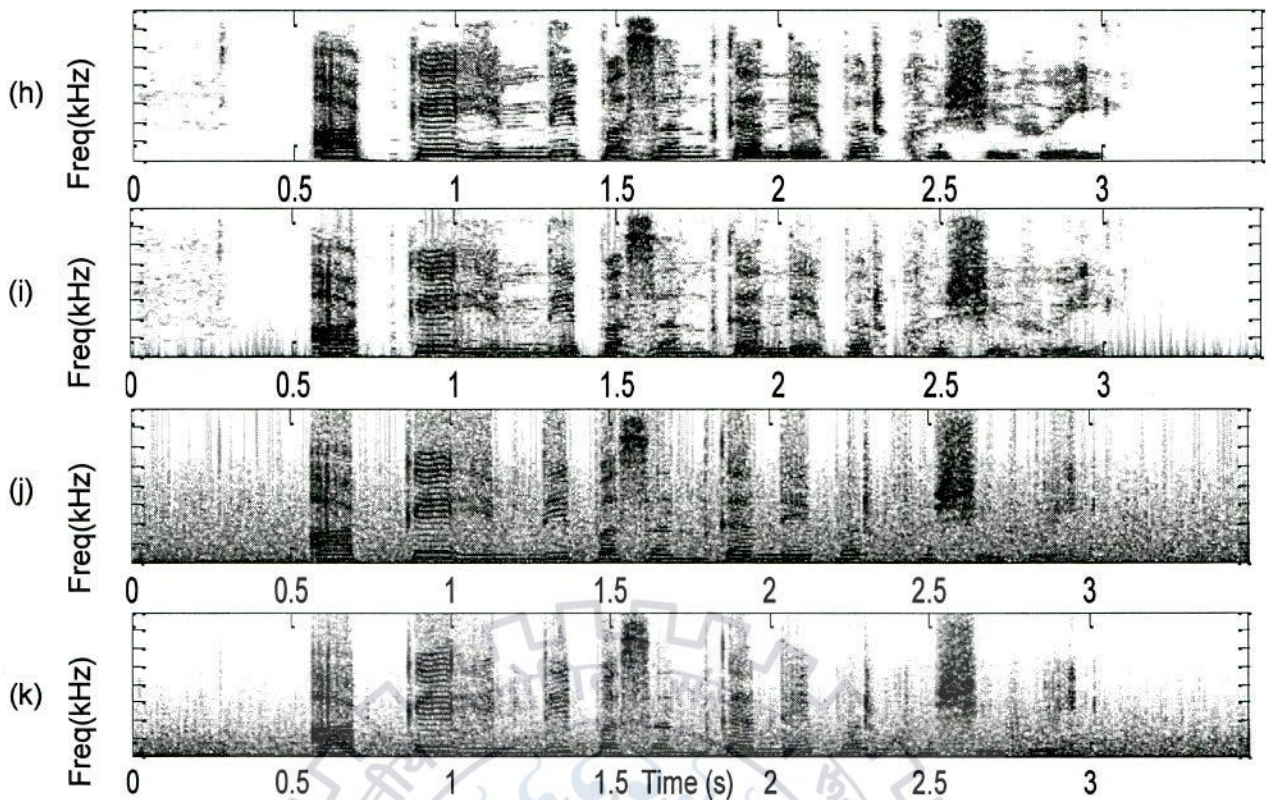


Fig. 2.23: Shows comparative time-frequency spectrograms of (a) clean (b) noisy speech at 5dB (c) proposed method (d) log-MMSE (e) Wiener (f) Spectral Sub. (g) p-MMSE (h) IDBM (i) ICS (j) hard and (k) Soft wavelet thresholding method for Hindi speech pattern “*apke hindi pasand karne par khushi hui*”.

Figure 2.23 shows the comparative analysis of spectrograms for Hindi speech at 5dB SNR of mixed noise sources i.e. white + factory floor + NOIZEUS speech + Hindi speech. In spectrogram comparison, all speech enhancement methods representing musical noise in the spectrogram except proposed WPT modified Wiener Gain method. Hence, it is clear that the proposed method is very much effective for improving quality and intelligibility in low SNR mixed noise environment and removes musical noise effectively.

#### 2.4 Summary

In this chapter, WPT modified Wiener Gain method is proposed for suppression of mixed noise of low SNR. The noisy dataset of mixed noise of low SNR range from -15 dB to 5 dB are generated for evaluation of the proposed speech enhancement method. The performance of proposed method is compared with other commonly used speech enhancement methods such as Wiener, SS, p-MMSE, log-MMSE, soft and hard WP threshold function ICS, IDBM and results show the improvement in terms of speech quality and intelligibility parameters i.e. SII, CD, PESQ, fw-SSNR and SNR. The WPT modified Wiener Gain method shows maximum improvement in comparison to other speech enhancement methods at all levels of input SNR.

*This chapter describes the work related to suppression of residual noise and speech distortion in highly non-stationary low SNR noise environments. It starts with some background for requirement of efficient speech enhancement algorithms for suppression of highly non-stationary noise. Then, the solution to overcome the limitations of traditional speech enhancement methods in the presence of highly non-stationary noise is suggested. Next, author's contribution is explained with supporting results which are given by WPT fuzzy mask based proposed speech enhancement method that efficiently work in the environments of non-stationary noisy speech of low and high input SNR.*

### 3.1 Overview

The highly non-stationary noise is encountered in many environments, such as stations, theatres, airports, vehicles, factories, cafeterias, bars, streets, etc. The denoised speech is obtained with the application of enhancement methods to noisy observations. The spectral subtractive, mask based methods and statistical model-based conventional approaches for noise suppression are based on assumption of stationary noise condition and they do not consider the environment of non-stationary and highly non-stationary noise. These conventional approaches have limited application in the area of noise suppression. High residual noise or high speech distortion are introduced in the processed speech spectrum if the signal power spectrum and the noise power spectrum are estimated under the assumption of stationary noise condition. Furthermore, in the mask based approaches clean speech signal information is used for noise suppression which is not possible for real-world applications because of non-availability of clean speech signal. Hence, the problem of enhancing speech degraded by highly non-stationary noise is a challenging task in real world applications.

### 3.2 Enhancement Techniques for Highly Non-Stationary Noise Environments

Single-channel speech enhancement algorithms are generally based on short-time spectral attenuation (STSA). The example of STSA based speech enhancement methods are spectral subtraction method (proposed by Berouti et al. [18]) and MMSE based short-time spectral amplitude estimator (proposed by the Ephraim–Malah [22]). But these conventional approaches work on the assumption of stationary noise condition. Some modifications in the basic suppression rules have been proposed by few researchers in order to overcome the problem of assumption of stationary noise signal [18, 151-152, 158], but these techniques only reduce the “musical” noise (i.e. pure tones or isolated peaks in the residual noise) but do

not eliminate it completely. The complete elimination of “musical” noise phenomenon is only obtained by a crude overestimation of the average noise spectrum and as a consequence, the short-time spectrum is attenuated more than it would be necessary: Due to this fact, audible distortions in the audio signal can be generated [153]. Cappe analyzed the behaviour of Ephraim and Malah estimator [154] and demonstrated that *a priori* SNR follows the shape of *a posteriori* SNR with a delay of one frame. This bias is due to the use of the speech spectrum estimated at the previous frame to compute the current *a priori* SNR. Since the gain depends on *a priori* SNR, it does not match with the current frame and thus it degrades the performance of the noise suppression system. Two-Step Noise Reduction (TSNR) technique has been presented by Cyril Plapous et al. [155] to refine the estimation of *a priori* SNR which suppresses these drawbacks while maintaining advantages of the decision-directed approach, like the highly reduced musical noise effect. This technique suffers from a delay of one frame in first step which is removed by the second step of TSNR algorithm. However, it has not been much effective in real world noise environments.

The algorithm described by Whipple utilizes a simple 2D analysis of magnitude spectrograms to find energy bursts that are localized in time and frequency [156]. Such bursts are replaced with zero energy to suppress the musical noise. Another similar approach for processing of a 2D spectrogram has been described by Goh et al. [157]. The local variance of coefficients is used for detection of musical noise, and a median filter is used to repair regions detected as musical noise. In the work of Linand Gourban et al. [159], a non-adaptive 2D smoothing of a magnitude spectrogram is used to detect speech/noise regions by applying a magnitude threshold. The spectrogram for regions that are classified as noise is time-smoothed with a box filter. The speech regions are processed by the Ephraim-Malah method [22]. The algorithm by Soon et al. [160] uses a 2D Fourier transform applied to a matrix of time-domain STFT (short-time Fourier transform) windows, which is equivalent to applying 1-dimensional DFT to every row of a complex spectrogram. This allows to effectively analyze the time correlations of STFT coefficients, but it doesn't exploit the frequency correlation of spectrograms effectively. To alleviate this problem, the application of above explained algorithms [156] is suggested as a post processing step and adaptive 2D spectrogram smoothing based algorithms achieves effective reduction of musical noise artifacts with minimal damage to the target signal [161-162]. Thomas Esch and Peter Vary described a post processing method for the spectral weighting gains to suppress musical noise [163]. This post filter adaptively smoothens the weighting gains over frequency based on soft-decisions of a low SNR detector and gives minimum residual noise in the processed speech.

From the literature, it is clear that most of the speech enhancement algorithms [151-178] estimate processed speech signal by multiplying magnitude of noisy speech signal with calculated gain function. However, a number of studies have analyzed that gain function based approaches generate residual noise in the processed speech and distort the speech [6, 165-178]. A detailed study done by P. C. Loizou and G. Kim illustrates the reasons why current speech enhancement algorithms do not improve speech intelligibility [35]. Based on these reasons, a modified Wiener filtering method combined with wavelet thresholding multitaper spectrum has been developed by Yanna Ma and Akinori Nishihara for suppression of speech distortions [179]. These types of algorithms can be described in two stages as shown in Figure 3.1. The application of gain function at the first stage is followed by binary mask decision [128] at the second stage. The first stage of gain reduction depends on calculated SNR from estimated noise spectrum and noisy speech spectrum in each band. The value of gain reduction is a ratio calculated from each band. The gain reduction has a range from 0 to 12 dB in applications like hearing aids. The speech frames with high *a-priori* SNR give high gain (about to 1) where as a low value (about to 0) of *a-priori* SNR in speech frames has low gain. Since noise and speech signals spectrally overlap, the concept of gain-reduction is not much beneficial for speech intelligibility improvement. Speech distortion is introduced by gain-reduction process. To overcome this distortion, second stage is introduced as a binary mask for further smoothing of the distorted speech signal.



Fig. 3.1: Block diagram of signal-processing stages used in single-channel speech enhancement.

Most of the two stage type speech enhancement methods use clean stimuli database for calculating the threshold level in binary mask or take a fixed value of SNR as threshold to overcome speech distortions and residual noise [180-185]. The clean stimuli and noise database used in gain functions and calculation of the mask limits are not practically available. None of the techniques were evaluated for both quality and intelligibility improvement at lower and higher levels of input SNR (such as -15 to 15 dB). Typical approaches to binary mask estimation use low-level features and bottom-up techniques. One reason that estimation techniques focus on the low-level features is that the Ideal Binary Mask (IdBM) itself is defined based on the local SNR at the Time-Frequency (T-F) unit level.

In this Chapter, a description for mask estimation is presented which is based on fuzzy mask function and WPT. The goal of the present study is to improve the quality and intelligibility both at lower and higher levels of input SNR varying from -15 dB to +15 dB and to reduce the distortions introduced by gain function. Hard and soft wavelet packet threshold values are used as lower and higher limits, respectively in the fuzzy mask for getting maximum resolution. In this proposed approach, a modified Wiener gain function is used at first stage and at second stage wavelet packet threshold values are used in the fuzzy mask for further improvement in speech.

### **3.3 WPT Fuzzy Mask Based Speech Enhancement Method**

The hard and soft WP threshold based fuzzy mask method is proposed for suppression of highly non-stationary noise in speech signal. The noisy input speech signal is applied to WPT for decomposing the input noisy speech signal into various energy bands and after decomposition into energy bands and reconstruction into signals a modified Wiener gain function is applied for noise suppression. The modified Wiener gain function has already been explained in Chapter 2. Further, the processed speech is given to fuzzy mask function for suppression of gain induced speech distortions and reconstruction of denoised speech signal is performed at last. Details of all these processes as applied in the present approach are described in two steps. Gain function is applied in its first stage and in second stage a fuzzy mask is used for further improvement in speech. These steps are explained in detailed in following sections.

#### **3.3.1 Modified Wiener gain function for noise suppression**

In practical environment, the background noise level and characteristics are constantly changing. Good estimation of the speech signal is required to alleviate the distortion caused by speech enhancement algorithms. Among the numerous techniques that were developed, the Wiener filter can be considered as one of the most fundamental speech enhancement approaches, which has been delineated in different forms and adopted in various applications [22-24]. Where, it is very necessary to use an effective gain function to overcome the noise and distortions. With this aim, a modified gain function is used for noise reduction:

Let us consider an additive noise model

$$Y(n) = X(n) + D(n) \tag{3.1}$$



Where,  $Y(n)$ ,  $X(n)$  and  $D(n)$  denote discrete-time signals of noisy speech, clean speech and noise, respectively. The discrete short-time Fourier transform (DSTFT) of the corrupted speech signal  $x(n)$ , is given as:

$$Y(n, k) = X(n, k) + D(n, k) \quad (3.2)$$

After experimentation in Chapter 2, it was found that the Db10 mother WPT is much suitable for speech signal decomposition in comparison to other mother wavelets. Hence, in this Chapter Db10 is applied for decomposition of input noisy speech signal. Decomposition levels provide significant information to avoid unreasonable maximum level values. Eq. 2.7 is used for selection of WP decomposition level which is mentioned in Chapter 2. According to this equation, 3<sup>rd</sup> level is more suitable for decomposition of input speech signal. The fuzzy mask based proposed method use Db10 mother WP with 3<sup>rd</sup> level decomposition of input speech signal to reduce the highly non-stationary noise and gain-induced speech distortions.

A modified gain function  $G(n, k)$  is multiplied with the noisy speech for getting the first stage denoised speech. Now, denoised speech signal can be expressed by eq. (3.3):

$$\hat{X}_1(n, k) = G(n, k) * Y(n, k) \quad (3.3)$$

The processed speech  $\hat{X}_1(n, k)$  is used in fuzzy mask for reduction of gain-induced speech distortions for further enhancement.

### 3.3.2 Fuzzy mask function

The prior research have suggested that the quality and intelligibility are improved by using true speech  $x(n)$  spectrum or true noise spectrum in binary mask for estimating enhanced speech spectrum [181-185]. Here, true speech means the clean speech that is used for generating noisy speech by mixing noise and true noise means the noise which is used for constructing the noisy speech signal. In practical environment only a noisy speech signal is available i.e. clean speech signal and true noise signal are not available. This makes estimation of mask a different task. To overcome the problem of using true speech or true noise in a binary mask, a fuzzy mask is proposed here which is based on soft and hard wavelet packet threshold. Figure 3.2 shows the block diagram of proposed method used in de-noising of noisy speech spectrum.

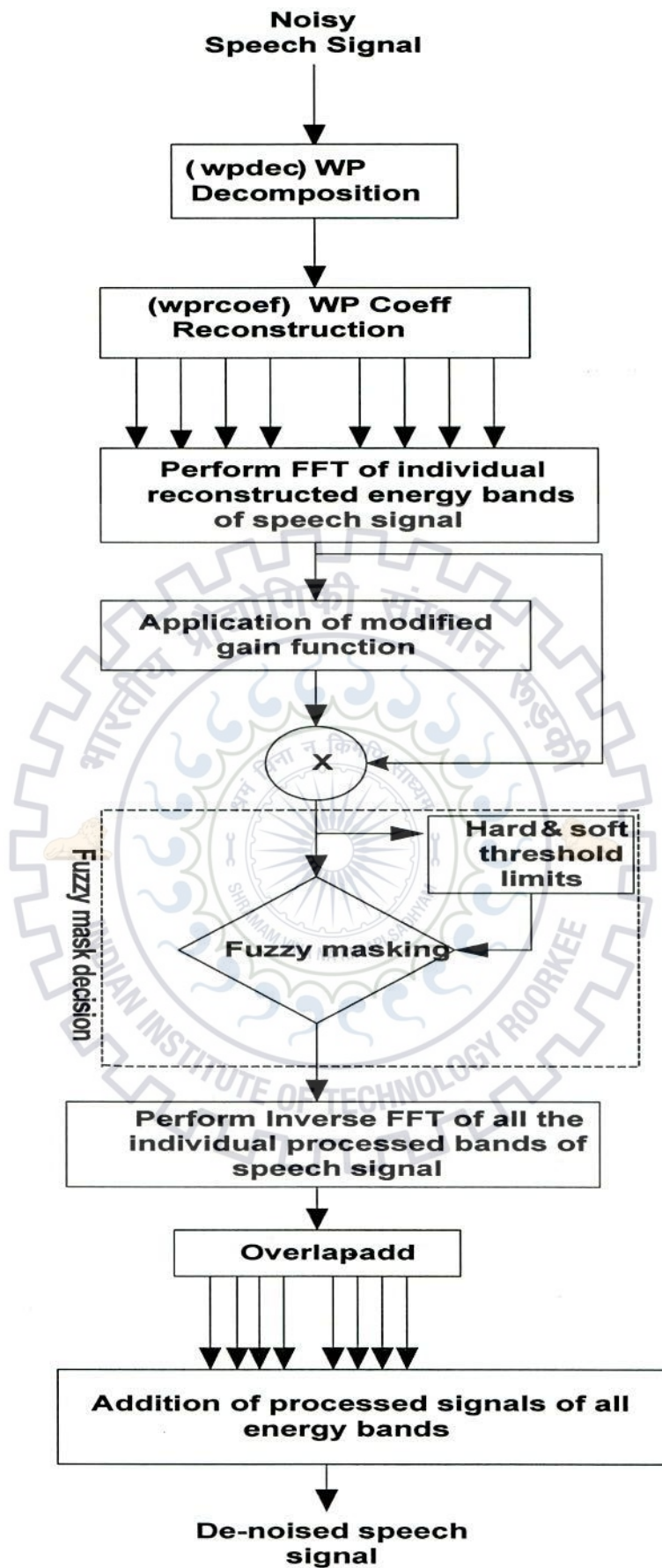


Fig. 3.2: Block diagram of proposed method for speech enhancement.

Following steps are used for estimating hard and soft WP threshold values in the proposed fuzzy masking approach:

Step1. Decomposition of input noisy speech signal using Db10 mother WP transform upto 3<sup>rd</sup> levels.

Step2. Reconstruct wavelet packet coefficients of the decomposed signal into eight noisy speech energy bands.

Step3. Perform FFT of noisy speech bands.

Step4. Estimate the processed speech spectrum  $\hat{X}_1(n, k)$  by using a modified gain function  $G(n, k)$

Step5. Limits  $a$  and  $b$  are computed by using eq. (3.4) to (3.6) where  $a$ , and  $b$  are hard and soft WP threshold limits [179].

$$a = \frac{\text{median}(|\hat{X}_1(n, k)|)}{0.6745} \quad (3.4)$$

$$T = a\sqrt{2 \cdot \log(\hat{X}_1(n, k))} \quad (3.5)$$

$$b = a + T \quad (3.6)$$

Step6. These parameters are used in fuzzy mask function in eq. (3.7).

$$f(\hat{X}_1(n, k); a, b) = \begin{cases} 0, & \hat{X}_1(n, k) \leq a \\ 2 \left( \frac{\hat{X}_1(n, k) - a}{b - a} \right)^2, & a \leq \hat{X}_1(n, k) \leq \frac{a+b}{2} \\ 1 - 2 \left( \frac{\hat{X}_1(n, k) - b}{b - a} \right)^2, & \frac{a+b}{2} \leq \hat{X}_1(n, k) \leq b \\ 1, & \hat{X}_1(n, k) \geq b \end{cases} \quad (3.7)$$

Step7. The desired speech spectrum coefficients are obtained by using fuzzy mask function.

Step8. Inverse-FFT and over-lap add method is applied to reconstruct the processed speech signal of each energy level.

Now, all processed signals of all eight energy bands of input speech signal are added to get the desired speech signal.

### 3.3.3 Results and discussion

The clean English speech input dataset are taken from NOIZEUS [140] corpus. The real-world sources of background noise are taken from AURORA [204] and NOISEX-92 [139] databases. These datasets have suburban train noise, babble, car, exhibition hall, restaurant, street, airport and train-station noise. These various types of noise are used for generation of noisy input speech.

The distortions introduced by noise-suppressive gain functions degrade the quality and intelligibility of the processed speech. In this section, performance of the proposed speech enhancement method is analyzed and compared with given speech enhancement methods such as Wiener, spectral subtraction, MMSE-SPU, p-MMSE, and log-MMSE in terms of output SNR, PESQ, MOS, and STOI parameters. Comparative evaluation of the results indicate the effectiveness of proposed method for improvements in both quality and intelligibility over unprocessed speech stimuli at all input noisy speech SNR levels from -15dB to 15dB. The results obtained are illustrated in Table 3.1 to 3.4.

The Figures 3.3 and 3.4 represent the graphical and bar chart representation of values given in Table 3.1. An observation on Table 3.1 and Figures 3.3, 3.4 indicates that for low input SNR (-15dB to -5dB) the performance of Wiener method in terms of SNR improvement is better among all the existing speech enhancement methods whereas for higher input SNR (0dB to 5dB), spectral subtraction method gives better result.

The MMSE based methods produce the poor output SNR comparatively. The minimum improvement is obtained by MMSE-SPU algorithm since the output SNR decreases rapidly with increasing input SNR (output SNR from -0.9161 dB to -7.1942 dB with input SNR from -15 dB to 15 dB, respectively) which shows that the gain function used in the algorithm is not much effective in reducing noise and hence much noise is left in processed speech. All these methods (such as Wiener, spectral subtraction, MMSE-SPU, p-MMSE, and log-MMSE) perform poorer than the proposed method. The output SNR obtained by aforementioned methods are decreasing with increasing input SNR but for the proposed method the output SNR values are increasing with increasing input SNR. The maximum output SNR values given by proposed method are 21.1661, 19.9634, 20.4319 and 20.3651dB SNR at 15 dB input SNR in babble, pink, f-16 and white noise. From this comparison, it is illustrated that proposed method outperforms the other existing methods and hence maximum improvement in speech quality and intelligibility is achieved.

Table 3.1: Output SNR Score in presence of various noise types.

Noise Type	Input SNR(dB)	Wiener	Spec. Sub.	MMSE-SPU	p-MMSE	log-MMSE	Fuzzy Mask	Proposed Method
<b>Babble</b>	<b>-15</b>	-1.6248	-3.0628	-0.9161	-2.6807	-1.4685	1.2639	<b>1.2876</b>
	<b>-10</b>	-0.8606	-1.6308	-0.3830	-1.3025	-0.6867	3.0527	<b>3.8137</b>
	<b>-5</b>	0.1228	0.4497	-0.2623	0.3705	0.6201	5.1278	<b>7.3879</b>
	<b>0</b>	0.3655	2.7938	-1.1846	2.3972	0.6598	6.5880	<b>11.3732</b>
	<b>5</b>	0.5560	4.7733	-1.9024	4.0540	1.1075	7.2565	<b>15.4439</b>
	<b>10</b>	0.5152	7.7444	-3.9847	6.1688	1.0174	7.5013	<b>18.8483</b>
	<b>15</b>	0.5334	10.5792	-7.1942	8.5556	0.8708	7.5798	<b>21.1661</b>
<b>Pink</b>	<b>-15</b>	-0.2943	-0.3913	-0.4945	-1.3664	-0.7095	0.7116	<b>0.3442</b>
	<b>-10</b>	-0.5045	-0.7419	-0.4192	-1.6068	-0.7159	2.4417	<b>1.7663</b>
	<b>-5</b>	-0.4067	-0.9633	-0.9239	-0.9312	-0.8058	4.7528	<b>4.6866</b>
	<b>0</b>	-0.1184	-0.2859	-0.7880	0.4606	0.0055	6.4175	<b>8.5616</b>
	<b>5</b>	0.2040	2.3865	-0.7660	2.1988	0.6066	7.1917	<b>12.6826</b>
	<b>10</b>	0.2726	4.3463	-2.7544	4.1307	0.6461	7.4785	<b>16.6544</b>
	<b>15</b>	-0.0466	7.8992	-4.6816	5.1177	0.5309	7.5798	<b>19.9634</b>
<b>f-16</b>	<b>-15</b>	-0.9183	-1.3101	-1.1139	-3.1839	-1.8632	0.6527	<b>0.1501</b>
	<b>-10</b>	-0.6528	-1.8024	-1.5140	-3.7675	-2.2559	2.3379	<b>2.1643</b>
	<b>-5</b>	0.1031	-0.4786	-0.7188	-1.1627	0.2874	4.6648	<b>5.5853</b>
	<b>0</b>	0.7172	0.8981	0.8761	1.5892	1.1103	6.3936	<b>9.5105</b>
	<b>5</b>	1.4096	2.7942	0.4943	3.1153	1.6383	7.2097	<b>13.5659</b>
	<b>10</b>	1.4223	5.1909	-0.7133	4.0056	1.7299	7.5037	<b>17.3403</b>
	<b>15</b>	1.3144	8.1706	-3.8082	5.9823	1.6504	7.5905	<b>20.4319</b>
<b>White</b>	<b>-15</b>	-3.5186	-6.4321	-1.6131	-3.7496	-2.3578	0.8345	<b>1.0934</b>
	<b>-10</b>	-2.6637	-4.4210	-1.5610	-2.5652	-1.7333	2.8686	<b>3.4797</b>
	<b>-5</b>	-1.3864	-2.4705	-1.1013	-0.7980	-1.0125	5.1418	<b>6.8096</b>
	<b>0</b>	-0.7655	0.8868	-1.1623	0.9690	-0.1488	6.6421	<b>10.4854</b>
	<b>5</b>	-0.3111	2.9205	-1.6204	2.4689	0.4242	7.2645	<b>14.1393</b>
	<b>10</b>	-0.2424	5.8020	-2.7799	4.3175	0.4384	7.4796	<b>17.5540</b>
	<b>15</b>	-0.1393	9.3381	-4.6100	6.8198	0.0872	7.5686	<b>20.3651</b>

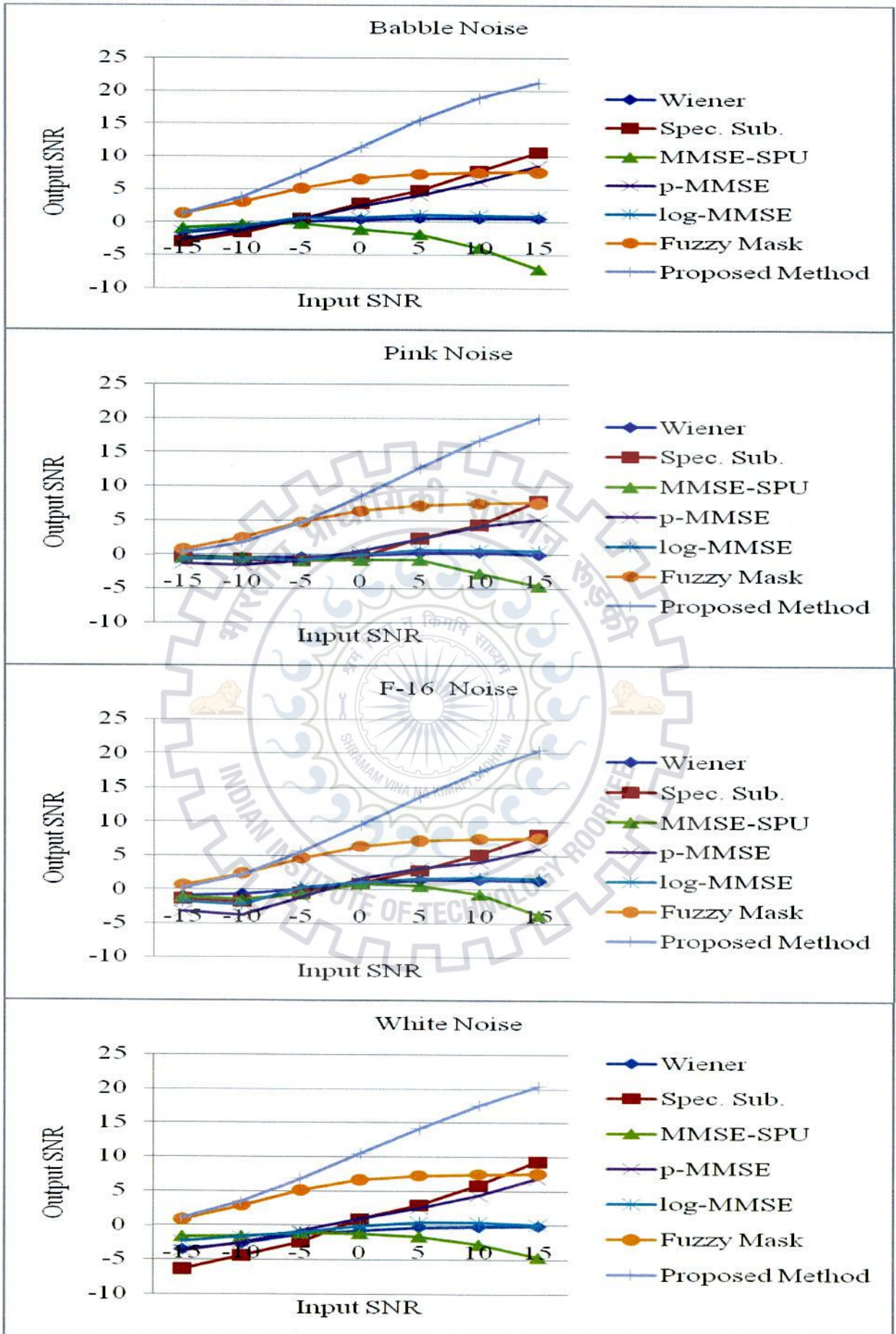


Fig. 3.3: Output SNR Scores in presence (a) babble (b) pink (c) f-16 (d) white noise.

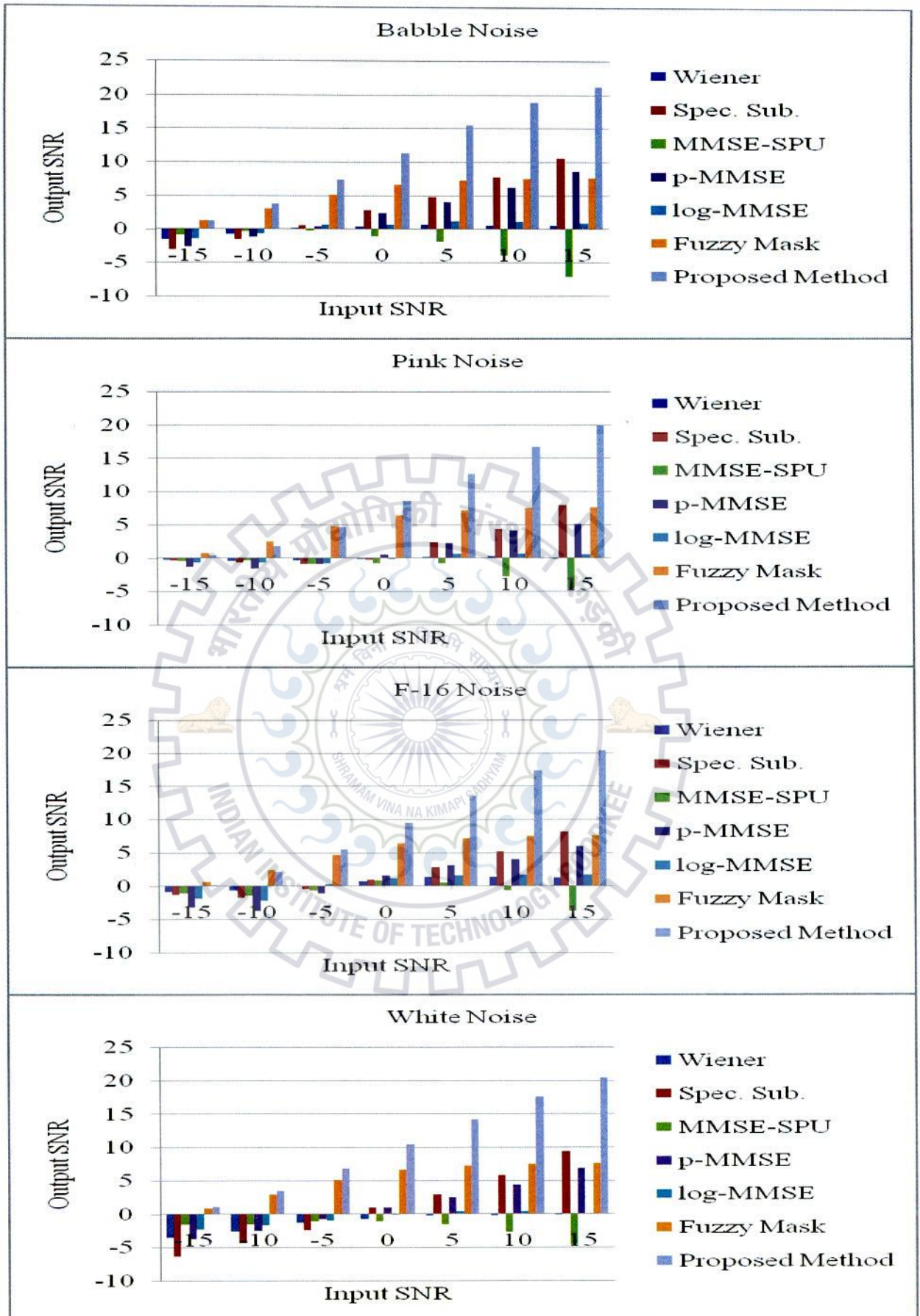


Fig. 3.4: Representation of results in bar chart for output SNR.

The PESQ parameter values computed for given speech enhancement methods is presented in Table 3.2. The graphical and bar chart representation for direct illustration of their comparative performance is shown in Figures 3.5 and 3.6. From these Figures and Table 3.2, it is observed that the maximum distortion is created by MMSE-SPU algorithm since this speech enhancement method shows the lowest PESQ values (1.6866, 2.0838, 2.1753 and 2.0375 in babble, pink, f-16 and white noise) at 15 dB input SNR. From the comparison, it is also indicated that fuzzy mask method obtains better results for all input SNR levels. The other speech enhancement methods (such as Wiener, spectral subtraction, MMSE-SPU, p-MMSE, log-MMSE) give poor results and hence lower improvement in processed speech signal. These speech enhancement methods are also compared with WPT fuzzy mask based proposed method in terms of PESQ parameter. The proposed speech enhancement method shows maximum improvement as compared to all other speech enhancement methods in terms of speech quality and intelligibility parameters at all levels of input noise SNR. There is a continuous improvement in quality with increasing input SNR and hence lowest distortion is observed in the processed speech.

Table 3.2: Output PESQ Scores in presence of various noise types.

Noise Type	Input SNR(dB)	Wiener	Spec. Sub.	MMSE - SPU	p-MMSE	log-MMSE	Fuzzy Mask	Proposed Method
Babble	-15	1.6813	1.5545	1.6855	1.5092	1.6339	1.8998	<b>1.6827</b>
	-10	1.2301	1.8331	1.1190	1.1982	1.0858	1.1307	<b>1.9930</b>
	-5	2.3253	2.2233	2.2648	2.2230	2.3977	2.3704	<b>2.3979</b>
	0	2.4117	2.3960	2.2959	2.5898	2.3786	2.6224	<b>2.6431</b>
	5	2.4556	2.5635	2.2339	2.7121	2.4321	2.8979	<b>2.9202</b>
	10	2.4778	2.9118	1.9946	2.9200	2.4919	3.2468	<b>3.1957</b>
	15	2.4817	3.0776	1.6866	3.1323	2.5006	3.5100	<b>3.5268</b>
Pink	-15	1.5915	0.9638	1.6875	1.7086	1.6284	1.3499	<b>1.8913</b>
	-10	1.8897	1.4961	1.9001	1.0717	1.9326	1.6963	<b>1.9710</b>
	-5	2.1771	1.7567	2.1956	2.1434	2.2461	2.0692	<b>2.4822</b>
	0	2.3073	2.0649	2.4186	2.4377	2.4747	2.4059	<b>2.5647</b>
	5	2.3950	2.3414	2.4062	2.5792	2.5010	2.7077	<b>2.7096</b>
	10	2.4417	2.5315	2.2094	2.7298	2.5007	3.0604	<b>3.0641</b>
	15	2.4504	2.9180	2.0838	2.8261	2.5126	3.4330	<b>3.4792</b>
f-16	-15	1.7929	0.1580	1.6276	1.6818	1.5438	1.5779	<b>1.8377</b>
	-10	2.0629	0.8926	2.0570	2.1260	2.0221	1.8788	<b>2.1461</b>
	-5	2.2578	1.5944	2.3126	2.3953	2.4124	2.1102	<b>2.5267</b>
	0	2.3951	1.9427	2.4149	2.5471	2.5103	2.3990	<b>2.6166</b>
	5	2.5349	2.3752	2.4099	2.7067	2.5639	2.6909	<b>2.7586</b>
	10	2.5443	2.6817	2.4094	2.8097	2.6527	3.0820	<b>3.1069</b>
	15	2.5516	2.8651	2.1753	3.0056	2.6527	3.4409	<b>3.4295</b>
White	-15	1.9357	1.1429	1.7455	1.8185	1.7558	1.3015	<b>1.9772</b>
	-10	2.1643	1.6273	2.1424	2.1893	2.1331	1.4821	<b>2.4001</b>
	-5	2.2787	1.8394	2.2899	2.3608	2.3651	1.7697	<b>2.4608</b>
	0	2.3462	1.9940	2.3712	2.4956	2.4590	2.0927	<b>2.5662</b>
	5	2.3701	2.2662	2.3036	2.4989	2.4056	2.4269	<b>2.5078</b>
	10	2.3897	2.5622	2.2053	2.6343	2.4185	2.7546	<b>2.8388</b>
	15	2.3922	2.8059	2.0375	2.9020	2.3681	3.1114	<b>3.2036</b>



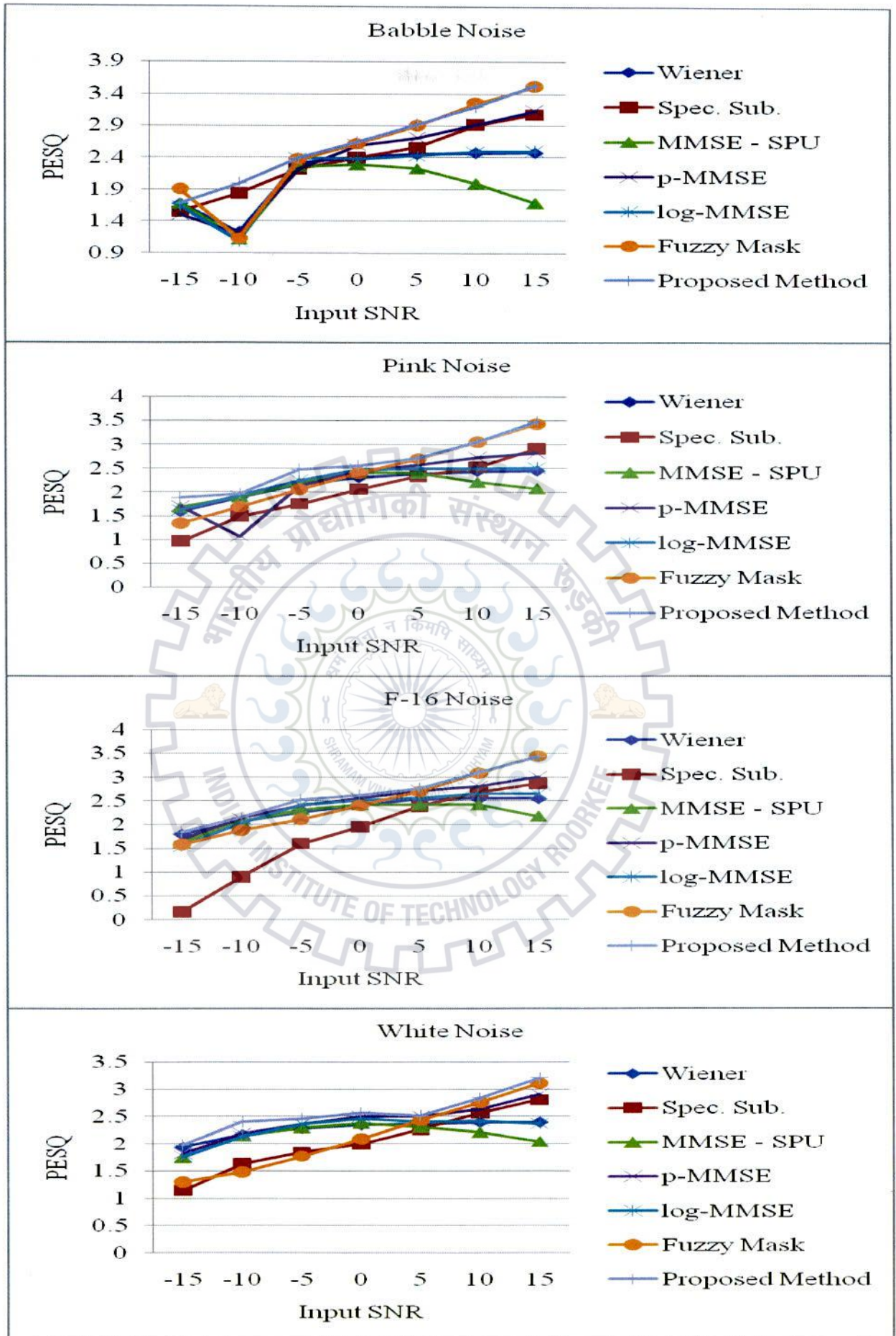


Fig. 3.5: PESQ Scores in presence (a) babble (b) pink (c) f-16 (d) white noise case.

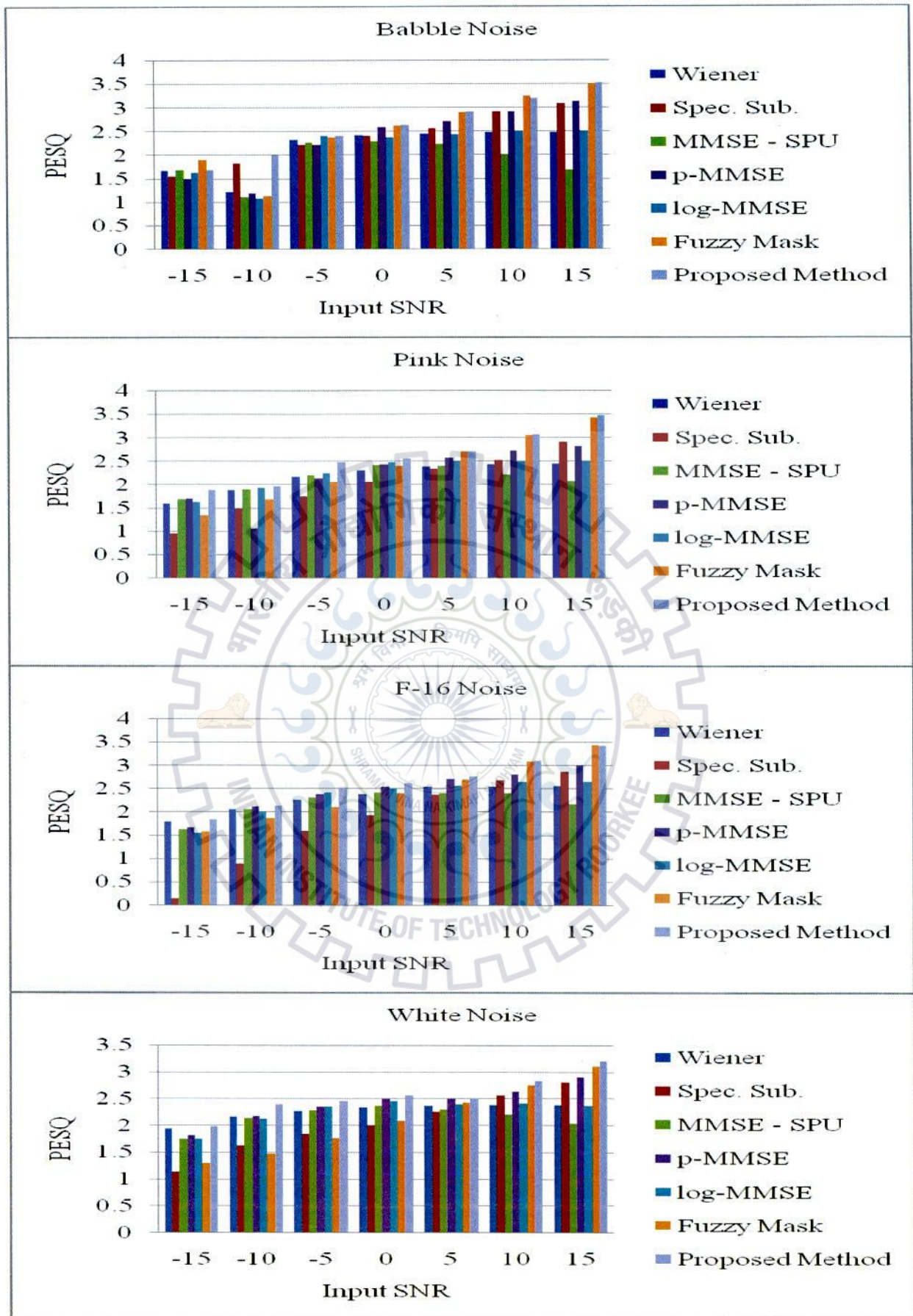


Fig. 3.6: Representation of results in bar chart for PESQ scores.

The MOS scores are presented in Table 3.3 and Figures 3.7 and 3.8 present the graphical and bar chart representation of the values given in Table 3.3 at different input SNR levels and various types of noise. The MOS rating lies in between 0 and 1 where 0 stands for lowest and 1 for highest improvement in speech intelligibility. By comparing the MOS scores of existing speech enhancement methods, it is observed that Wiener method shows better results for low input SNR values only (i.e. -5 to -15 dB) while spectral subtraction shows better results for higher input SNR (i.e. 0 to 5 dB). The MMSE-SPU method shows distortions in the processed speech which results in lowest MOS values at all levels of input SNR. Since the proposed method gives highest MOS scores 0.7182, 0.7050, 0.6911, 0.6252 in babble, pink, f-16 and white noise, respectively; hence, the maximum improvement and lowest distortion is observed in processed speech. The performance of proposed method is increasing with increasing input SNR levels (from -15 to +15 dB).

Table 3.3: MOS Scores in presence of various noise types.

Noise Type	Input SNR (dB)	Wiener	Spec. Sub.	MMSE - SPU	p-MMSE	log-MMSE	Fuzzy Mask	Proposed Method
<b>Babble</b>	-15	0.3398	0.2703	0.2839	0.2866	0.2783	0.3112	<b>0.3436</b>
	-10	0.3674	0.3020	0.3465	0.3612	0.3406	0.3486	<b>0.3752</b>
	-5	0.3871	0.3661	0.3744	0.4088	0.4030	0.3969	<b>0.4175</b>
	0	0.4062	0.4026	0.3808	0.4495	0.3987	0.4579	<b>0.4634</b>
	5	0.4164	0.4428	0.3682	0.4819	0.4109	0.5344	<b>0.5409</b>
	10	0.4217	0.5385	0.3255	0.5409	0.4251	0.6380	<b>0.6228</b>
	15	0.4226	0.5876	0.2840	0.6039	0.4272	0.7190	<b>0.7182</b>
<b>Pink</b>	-15	0.2739	0.2305	0.2841	0.2865	0.2777	0.2529	<b>0.2923</b>
	-10	0.3247	0.2648	0.3112	0.3382	0.3160	0.2851	<b>0.3418</b>
	-5	0.3572	0.2922	0.3607	0.3509	0.3706	0.3378	<b>0.3835</b>
	0	0.3832	0.3370	0.4078	0.4122	0.4209	0.4049	<b>0.4356</b>
	5	0.4024	0.3905	0.4049	0.4468	0.4273	0.4807	<b>0.4812</b>
	10	0.4131	0.4348	0.3634	0.4868	0.4272	0.5839	<b>0.5835</b>
	15	0.4152	0.5403	0.3403	0.5138	0.4301	0.6921	<b>0.7050</b>
<b>f-16</b>	-15	0.2968	0.2093	0.2776	0.2835	0.2692	0.2725	<b>0.3052</b>
	-10	0.3367	0.2275	0.3357	0.3477	0.3299	0.3082	<b>0.3596</b>
	-5	0.3730	0.2742	0.3843	0.4025	0.4064	0.3449	<b>0.4107</b>
	0	0.4024	0.3175	0.4069	0.4387	0.4295	0.4033	<b>0.4473</b>
	5	0.4356	0.3980	0.4058	0.4804	0.4429	0.4761	<b>0.4947</b>
	10	0.4380	0.4737	0.4057	0.5091	0.4659	0.5889	<b>0.5963</b>
	15	0.4398	0.5249	0.3568	0.5661	0.4659	0.6943	<b>0.6911</b>
<b>White</b>	-15	0.3164	0.2395	0.2909	0.3001	0.2921	0.2494	<b>0.3275</b>
	-10	0.3548	0.2776	0.3507	0.3595	0.3490	0.2636	<b>0.3668</b>
	-5	0.3772	0.3028	0.3796	0.3948	0.3957	0.2938	<b>0.4027</b>
	0	0.3916	0.3254	0.3971	0.4260	0.4172	0.3418	<b>0.4281</b>
	5	0.3968	0.3747	0.3824	0.4268	0.4048	0.4097	<b>0.4289</b>
	10	0.4012	0.4425	0.3626	0.4611	0.4078	0.4936	<b>0.5174</b>
	15	0.4018	0.5080	0.3324	0.5356	0.3964	0.5977	<b>0.6252</b>

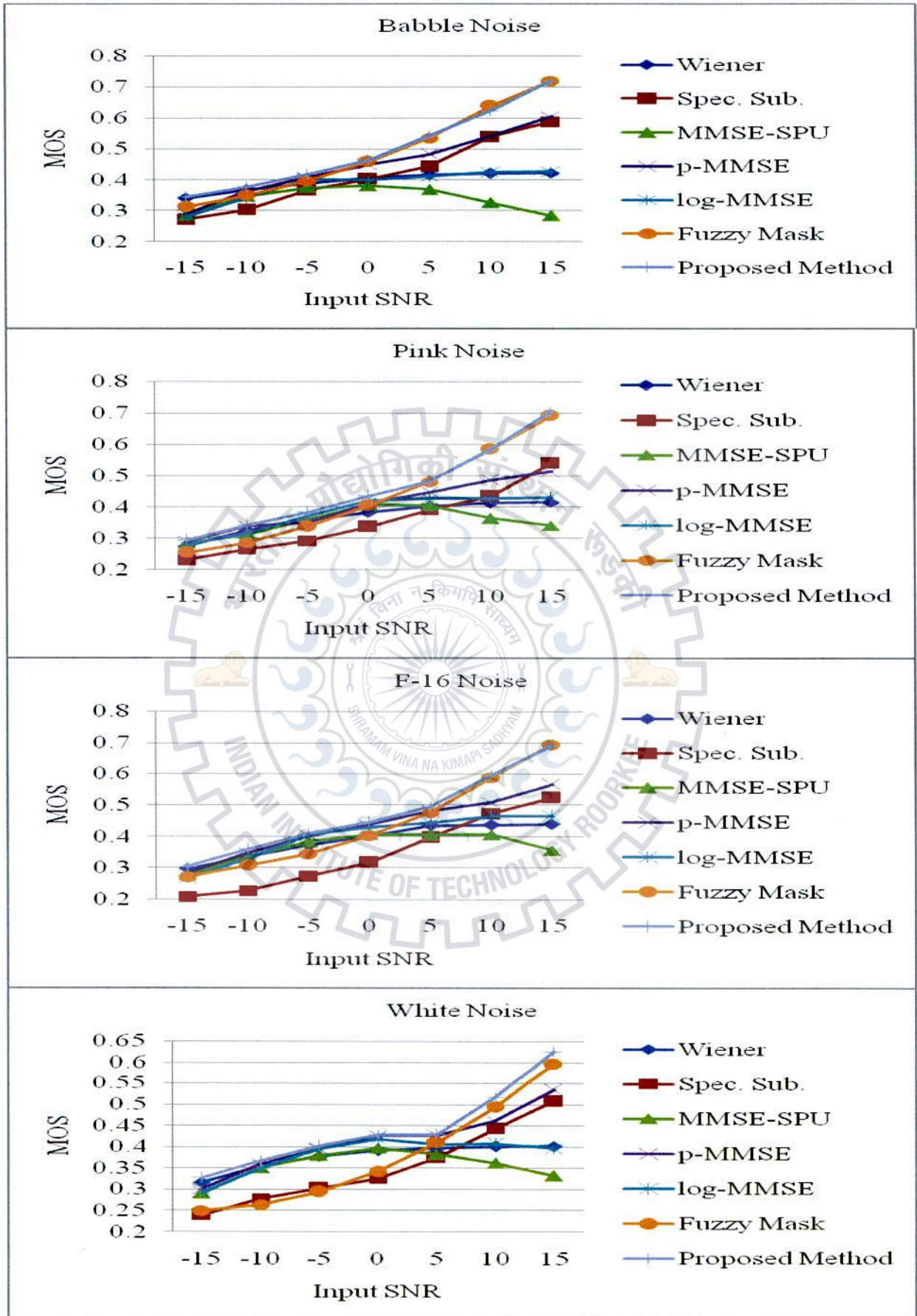


Fig. 3.7: MOS Scores in presence (a) babble (b) pink (c) f-16 (d) white noise case.

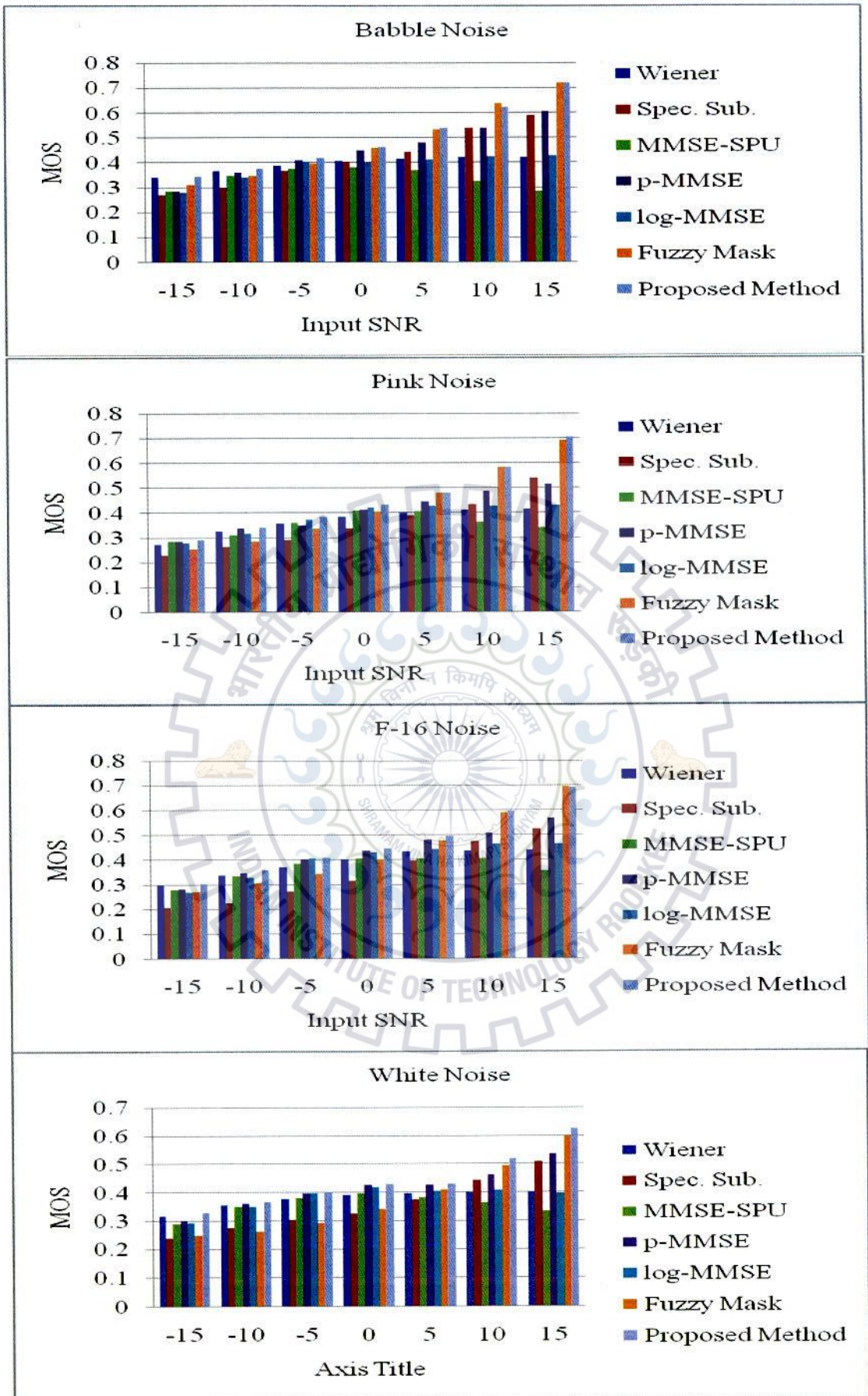


Fig. 3.8: Representation of results in bar chart for MOS scores.

Short-time objective intelligibility (STOI) is a parameter for speech intelligibility measure and the range of STOI is in between 0 to 1. The increment in STOI values indicates improvement in quality and intelligibility of the noisy speech signal. By comparing the output SNR values of previously given speech enhancement methods in Table 3.4, it has been observed that Wiener method shows better results for low input SNR values (i.e. 5 to -15 dB) only and spectral subtraction for higher input SNR (i.e. 10 to 15 dB). The MMSE-SPU method does not give better STOI scores and hence results are not considered satisfactory.

Table 3.4: Output STOI Scores in presence of various noise types.

Noise Type	Input SNR (dB)	Wiener	Spec. Sub.	MMSE - SPU	p-MMSE	log-MMSE	Fuzzy Mask	Proposed Method
<b>Babble</b>	<b>-15</b>	0.7402	0.6216	0.7296	0.7372	0.7099	0.5803	<b>0.7546</b>
	<b>-10</b>	0.7651	0.6940	0.7845	0.7815	0.7747	0.6933	<b>0.7990</b>
	<b>-5</b>	0.7990	0.7590	0.8021	0.8063	0.8109	0.7873	<b>0.8195</b>
	<b>0</b>	0.8147	0.7921	0.7950	0.8164	0.8201	0.8487	<b>0.8476</b>
	<b>5</b>	0.8235	0.8181	0.7966	0.8325	0.8303	0.8861	<b>0.9026</b>
	<b>10</b>	0.8123	0.8499	0.7779	0.8532	0.8281	0.9120	<b>0.9300</b>
	<b>15</b>	0.8119	0.8820	0.7218	0.8759	0.8166	0.9288	<b>0.9472</b>
<b>Pink</b>	<b>-15</b>	0.5776	0.4200	0.6284	0.6347	0.5756	0.4957	<b>0.6410</b>
	<b>-10</b>	0.6521	0.5092	0.7108	0.7201	0.6772	0.5834	<b>0.7211</b>
	<b>-5</b>	0.7411	0.6052	0.7757	0.7630	0.7595	0.6820	<b>0.7851</b>
	<b>0</b>	0.7848	0.6945	0.8163	0.8078	0.8151	0.7634	<b>0.8227</b>
	<b>5</b>	0.8006	0.7601	0.8187	0.8159	0.8193	0.8315	<b>0.8292</b>
	<b>10</b>	0.8181	0.7980	0.8093	0.8369	0.8154	0.8796	<b>0.8890</b>
	<b>15</b>	0.8214	0.8442	0.7891	0.8375	0.8221	0.9183	<b>0.9359</b>
<b>f-16</b>	<b>-15</b>	0.6237	0.3430	0.6827	0.6900	0.6243	0.4650	<b>0.6980</b>
	<b>-10</b>	0.7072	0.5169	0.7612	0.7691	0.7393	0.5665	<b>0.7694</b>
	<b>-5</b>	0.7701	0.6624	0.8097	0.7976	0.7971	0.6691	<b>0.8111</b>
	<b>0</b>	0.8027	0.7255	0.8257	0.8067	0.8185	0.7685	<b>0.8390</b>
	<b>5</b>	0.8282	0.7815	0.8197	0.8033	0.8224	0.8455	<b>0.8625</b>
	<b>10</b>	0.8295	0.8336	0.8242	0.8313	0.8312	0.8926	<b>0.9172</b>
	<b>15</b>	0.8393	0.8591	0.7987	0.8542	0.8371	0.9274	<b>0.9564</b>
<b>White</b>	<b>-15</b>	0.6624	0.5091	0.6425	0.6490	0.5920	0.4851	<b>0.6669</b>
	<b>-10</b>	0.7290	0.6048	0.7216	0.7124	0.6966	0.5720	<b>0.7302</b>
	<b>-5</b>	0.7440	0.6503	0.7566	0.7527	0.7465	0.6483	<b>0.7611</b>
	<b>0</b>	0.7701	0.6819	0.7765	0.7742	0.7679	0.7243	<b>0.7861</b>
	<b>5</b>	0.7873	0.7273	0.7980	0.7991	0.7921	0.8007	<b>0.8040</b>
	<b>10</b>	0.8122	0.7919	0.7761	0.8037	0.7936	0.8631	<b>0.8783</b>
	<b>15</b>	0.8055	0.8400	0.7671	0.8441	0.8024	0.9066	<b>0.9291</b>

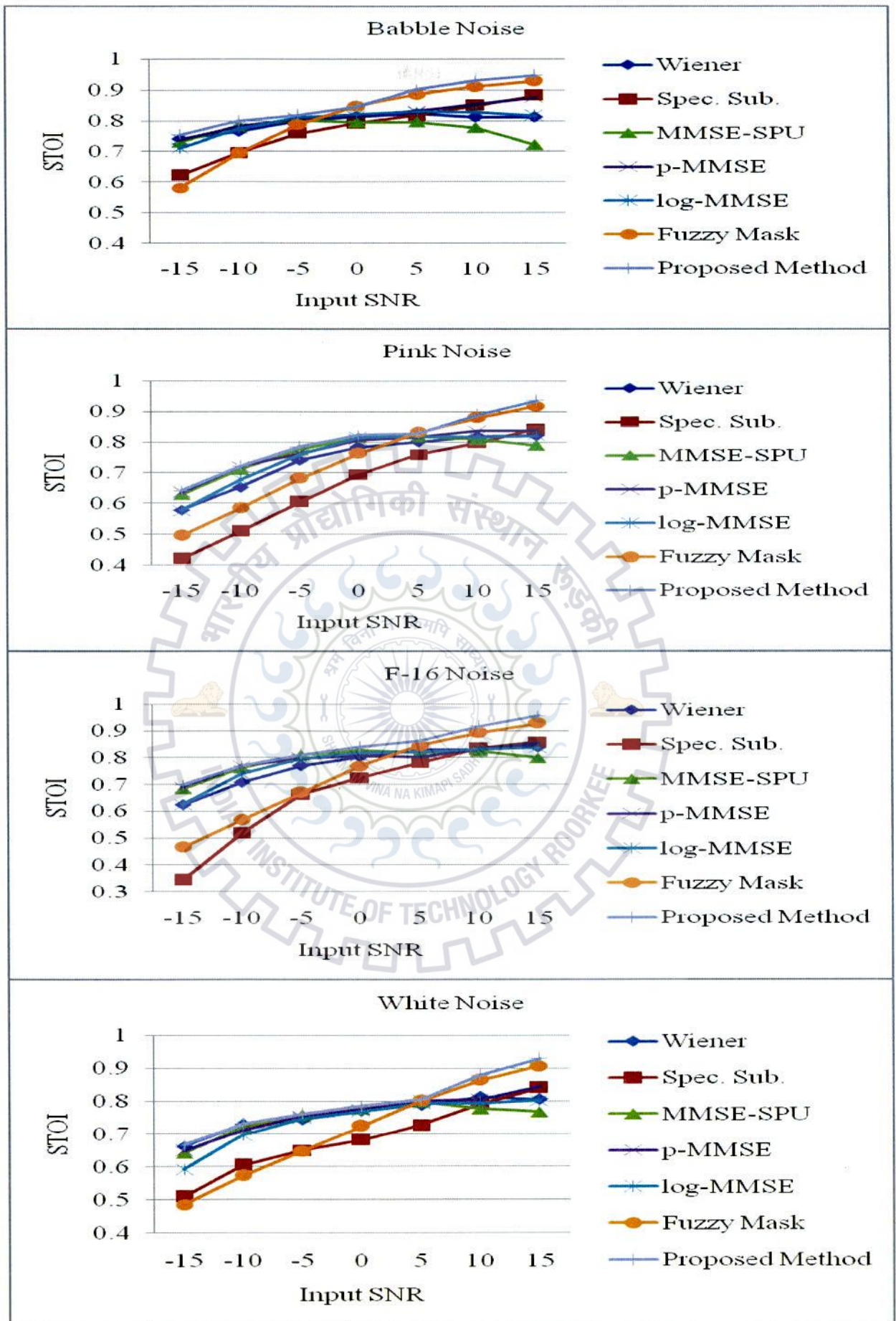


Fig. 3.9: Output STOI Scores in presence (a) babble (b) pink (c) f-16 (d) white noise.

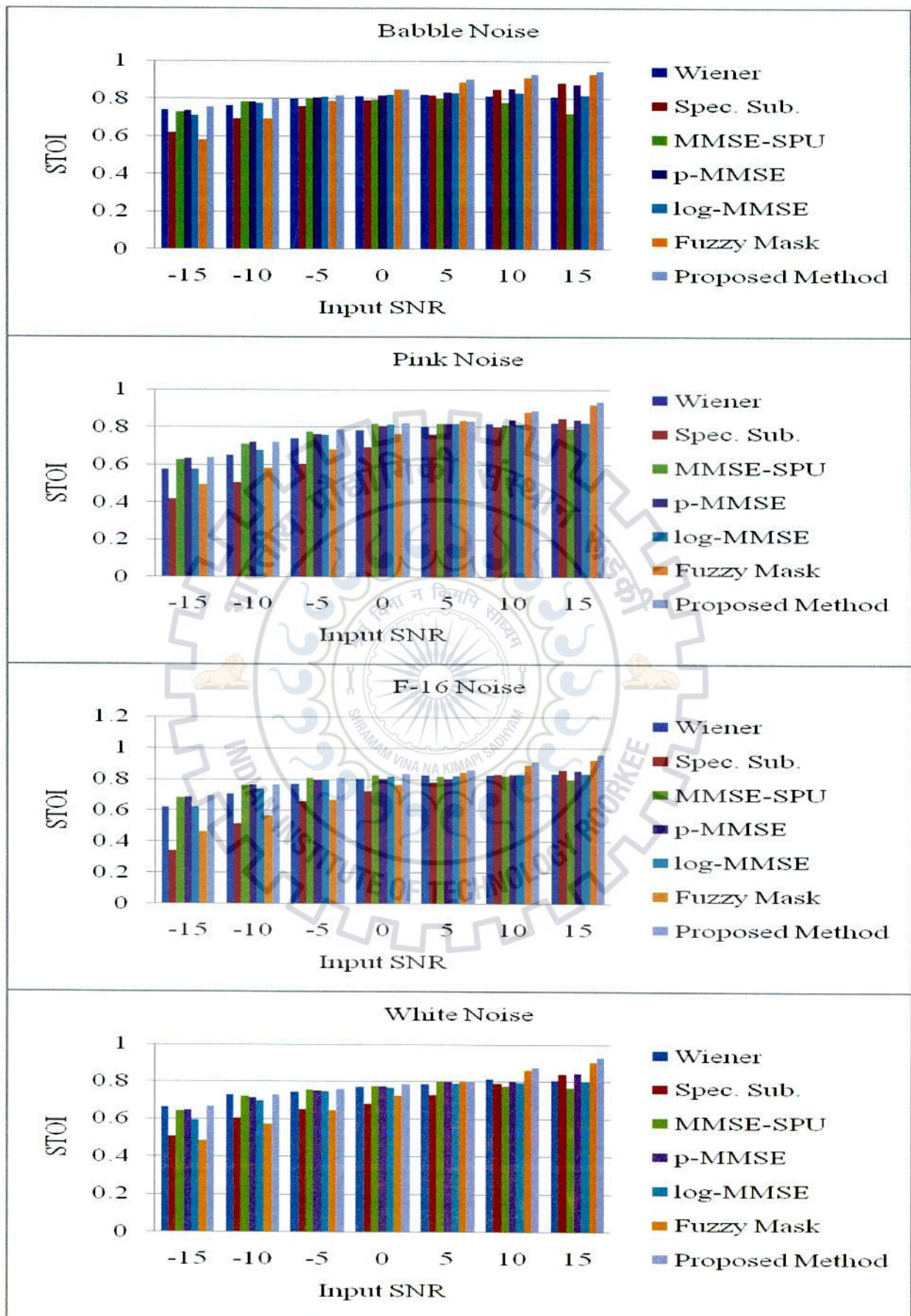


Fig. 3.10: Representation of results in bar chart for STOI scores.



The existing speech enhancement methods obtained improvement in STOI with increasing input SNR from -15dB to 15dB but not more than the fuzzy mask method and proposed method. The maximum STOI value (0.9472, 0.9359, 0.9564 and 0.9291) for babble, pink, f-16 and white noise, respectively) is obtained by WPT fuzzy mask based proposed method at 15 dB input. The STOI values obtained for proposed method are consistently increasing with increasing input SNR levels and for all noise types. For direct illustration of the results, the graphical and bar chart representation of the values given in Table 3.4 are presented in Figures 3.9 and 3.10.

Both, the clean (true) and desired speech signal (processed) are played back through simulation in Matlab and it is observed that they are generating almost the same voice. All performance measure parameters illustrate that the performance of proposed WPT Fuzzy mask method is comparatively better for all noise conditions and input SNR levels.

Experiments have also been performed for Indian languages such as Bengali, Kannada, Malayalam and Hindi. These results are given in appendix A and B. The results given in these appendixes are in favour of the proposed method. The maximum improvement in performance parameters are shown by WPT Fuzzy mask based proposed method.

### 3.4 Summary

A WPT Fuzzy mask based method has been proposed here for suppression of highly non-stationary noise sources of low SNR input noisy speech. This fuzzy mask replaces the need of true speech or noise signal with the use of WP soft and hard threshold. The noisy dataset of SNR range from -15 dB to 15 dB are generated for comparative analysis. The performance of proposed method is compared with other available speech enhancement methods and results show the improvement in terms of speech quality and intelligibility parameters. The WPT Fuzzy mask method gives maximum quality and intelligibility improvement when compared to other speech enhancement methods at all levels of input SNR.

*This chapter presents the research work for single-channel speech enhancement by combined suppression of reverberation and noise. It starts with overview of reverberation suppression. Thereafter discussion about the methodology of the algorithms for their suppression is presented. Thereafter, approach to the problem is discussed. The performance of the proposed reverberant mask based speech enhancement method is compared with other existing methods in terms of performance measure parameters.*

### 4.1 Overview

Reverberation is one of the most common phenomenon which affects the quality and intelligibility of speech communication systems. In reverberation, delayed copies of the speech acoustic waveform, called echoes, are added to the direct speech. The received signal over a distant microphone or uncontrolled environment generally consists of direct sound, reflections that arrive shortly after the direct sound (known as early reverberation), and reflections that arrive after the early reverberation (known as late reverberation) etc. The combination of the direct sound and early reverberation is sometimes referred as the early sound component [41]. The early reverberation components enhance both audibility and intelligibility of direct speech but they also produce spectral distortions called '*coloration*'. In contrast to early reverberation, late reverberation impairs speech intelligibility [43].

Enhancement of reverberated speech signals has gained considerable research interest recently, as they can be used in many emerging communication applications such as hands-free communication, voice control, and hearing aids. Reverberated speech signals collected by microphones degrade the performance of communication systems: for example, the recognition accuracy of voice control system is decreased with decreasing the speech quality/intelligibility.

Berkley was the first researcher to propose the perception of reverberated speech in two parts: coloration and echo [186]. He further mentioned that the coloration correlates with room spectral deviation  $\sigma$  and echo correlates with  $T_{60}$ . Based on these findings, Allen [187] proposed that the quality of reverberated speech can be estimated by

$$P = P_{\max} - \sigma T_{60} \quad (4.1)$$

Here,  $P$  is the subjective preference;  $P_{\max}$  and  $\sigma$  is defined for the maximum possible preference and the room spectral variance, respectively. The  $T_{60}$  is defined as the

Reverberation Time (RT) which is required for reflections of a direct sound to decay by 60 dB to the level of the direct sound.

Room spectral variance  $\sigma$  is determined by Signal-to-Reverberant Ratio (SRR). The  $\sigma$  increases monotonically as SRR decreases and saturates to a fixed value when SRR drops below 0 dB [188]. When SRR is larger than 0 dB, reverberation effect is mainly due to the early reflections of room impulse response (RIR) and thus coloration is more pronounced. Because the total energy of all the reflections is similar everywhere in a room and the direct path energy depends mainly on the distance  $d_{sm}$  between the sound source and microphone, spectral variance  $\sigma$  is also determined by the distance  $d_{sm}$ . In order to preserve speech quality, it is required to reduce both spectral deviation  $\sigma$  and reverberation time T60, corresponding to reduced coloration (also termed as short term reverberation) and echo (also termed as long term reverberation).

Several enhancement methods have been developed for enhancement of single-channel reverberant speech. These methods are categorized into short and long-term dereverberation methods. This is due to the fact that the two types of reverberations demonstrate different characteristics and is difficult to remove both at the same time. A two stage method eliminating coloration and echo separately is usually used [43, 189-199]. Few algorithms are reported in the literature for suppression of short and long term reverberations. Some of these are described as below: [189-199].

## 4.2 Reverberation Suppression

### 4.2.1 Short-term dereverberation

Most of the short term dereverberation methods rely on Linear Prediction (LP) analysis [192]. A speech signal  $x(n)$  can be predicted by its past components  $p$  such as:

$$X(n)=[X(n-1),X(n-2),\dots\dots\dots X(n-p)] \text{ as: } X(n)=-a^T X(n-1)+e(n) \quad (4.2)$$

Here,  $a=[a_1, a_2, \dots\dots\dots a_p]^T$  comprises the Linear Prediction Coefficients (LPC),  $e(n)$  is the prediction error, and  $p$  the prediction order.

Let, the prediction error vector of reverberated speech be denoted as  $e_l=[e(n_l), e(n_l+1), \dots\dots\dots, e(n_l+L-1)]^T$  and the prediction error vector of enhanced speech as  $\hat{e}_l=[\hat{e}(n_l), \hat{e}(n_l+1), \dots\dots\dots, \hat{e}(n_l+L-1)]^T$ . Early dereverberation methods correspond to different methods of attaining  $e_l$  from  $\hat{e}_l$ . For example, Gillespie et al. [192] and Gaubitch et

al. [193] proposed kurtosis maximization and LP residual averaging methods for short-term dereverberation, respectively. Wu et al. [43] and Habets et al. [190] applied these schemes as the first part of their two stage dereverberation methods.

Gaubitch et al. [193] proposed to average the LP residual between neighboring larynx cycles in voiced speech based on the following phenomena: some random peaks appear in LP residual of reverberated speech, while the main features of clean speech LP residual between consecutive larynx-cycles change slowly. The smoothed LP residual  $\hat{e}_l$  for larynx cycle  $l$  can be represented as:

$$\hat{e}_l = (I - W)e_l + \frac{1}{2\Gamma + 1} \sum_{i=-\Gamma}^{\Gamma} W e_{l+i} \quad (4.3)$$

Here,  $e_{l+i}$  is the LP residual of larynx cycle  $l+i$  of the reverberated speech,  $I$  the identity matrix and  $W$  a diagonal weighting function. The  $\Gamma$  is the number of frames used for smoothing. To further eliminate reverberation in unvoiced speech segments and utilize past correct larynx-cycle frames, a  $l_i$  tap FIR inter larynx filter  $\hat{g}_l$  is trained by minimizing  $\|g_l^T e_l - \hat{e}(n_l)\|^2$ . The minimizing filter  $\hat{g}_l$  is used to update a smoothed inter larynx filter only in voiced speech segments with smoothing factor  $\gamma$

$$\hat{g}(n_l) = \gamma \hat{g}(n_{l-1}) + (1 - \gamma) \hat{g}_l \quad (4.4)$$

The smoothed filter is applied to both voiced and unvoiced segments of speech. The drawback of these methods is that the previous described methods were not effectively working for long term reverberation. To overcome the long-term reverberation, some methods were introduced as explained below.

#### 4.2.2 Long-term dereverberation

Long-term reverberation demonstrates similar characteristics as noise. The late reflection part of RIR is often modeled as an exponentially damped Gaussian noise process. Late reverberation is often treated as additive noise, so that denoising methods such as spectral subtraction [18] and MMSE estimation [22] can be used for dereverberation. When spectral coefficients of clean speech and noise are modeled as independent complex Gaussian random variables, spectral subtraction is a maximum-likelihood (ML) estimation of the spectral variance of clean speech. On the other hand, the MMSE method is a Bayesian estimation of the magnitude and phase spectrum of clean speech. Estimation of late

reverberation spectral variance (LRSV) is a key problem. Several LRSV estimation methods have been developed recently. They can be classified into two categories: estimation based on a statistical model [194-196] and estimation based on a weighted sum of past DFT components [43, 189], [197].

Recently, developed spectral subtraction based reverberant speech enhancement methods plays an important role in the enhancement of reverberant speech. The spectral subtraction based enhancement methods aims at the suppression of late reverberation to improve speech intelligibility [41, 44]. There is another class of excitation source information based reverberant speech enhancement algorithms which primarily aim to emphasize the high Signal to Reverberant Ratio (SRR) regions relative to the low SRR regions of the reverberant speech signal in the temporal domain [45-46]. The basis for the temporal processing technique is that in case of reverberant environments, the excitation source signal of voiced speech segments contains the original impulses followed by several other peaks due to multi-path reflections. Consequently, dereverberation is achieved by attenuating the peaks in the excitation sequence due to multi-path reflections, and synthesizing the enhanced speech waveform using the modified excitation source signal and the time-varying all-pole filter with coefficients derived from the reverberant speech. The high SRR regions are emphasized by deriving the weight function to modify the excitation source characteristics at fine and gross levels [45]. However, until now there are no practical and robust dereverberation techniques available mainly because, the degradation is non-stationary, correlated with the signal and cannot easily be modeled for combined suppression of reverberation and noise sources. High performance and innovative algorithms are needed for joint noise and reverberation suppression to restore high quality speech inputs for communication systems.

### 4.3 Reverberation Modelling

Basically, two types of modelling are used for generation of reverberant speech signal. These are classified as time domain methods and statistical modelling methods. The description of these methods is given as follows:

#### 4.3.1 Time domain

In a reverberant room, the reverberated speech  $Z(n)$  results from the convolution of the clean speech signal  $X(n)$  and the room impulse response (RIR)  $h(n)$  as

$$Z(n) = \sum_{i=0}^{Q-1} h(i)X(n-i) \quad (4.5)$$

Where,  $Q$  is the length of  $h(n)$ . Figure 4.1 depicts a representative RIR generated by using eq. (4.5), which is called as image method [199]. The RIR can be partitioned into three components: the direct signal, early reflections, and late reflections. The direct signal is the strongest impulse corresponding to the direct path from the speech source to the listener. Early reflections are the impulses that arrive within 50 ms after the direct signal. Early reflections are known to cause short-term reverberation or “coloration” effects. Early reflections boost the energy of the direct signal as well as emphasize modulation frequency content around 4 Hz [200], and they have minimal effects on intelligibility. Late reflections, in turn, which arrive at time intervals greater than 50ms post the direct impulse, smear the speech signal and can severely reduce signal quality and intelligibility. Late reflections cause long-term reverberations or echoes.

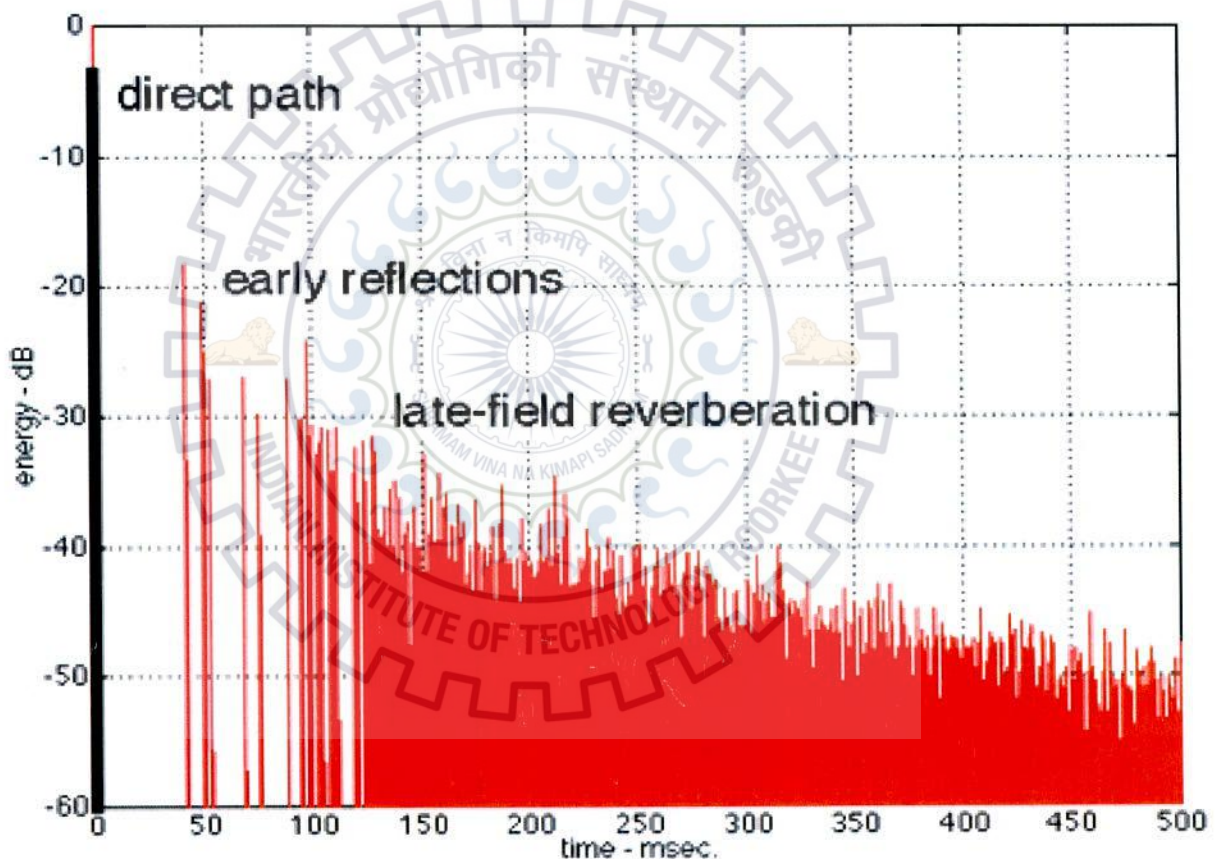


Fig. 4.1: Waveform of a representative room impulse response.

### 4.3.2 Statistical domain

Moorer [201] reported that the convolution of a clean speech and a Gaussian noise modulated by exponentially decaying envelope will generate natural reverberation effect. Based on this phenomenon, Polack [202] proposed a model for modelling the RIR as the product of a stationary noise process and an exponentially decaying envelope as:

$$h(t) = b(t)e^{-\Delta t}, \quad t \geq 0 \quad (4.6)$$

Here,  $b(t)$  is a zero-mean Gaussian stationary noise and the exponentially decaying parameter  $\Delta$  are linked to reverberation time (RT) through eq. (4.7):

$$\Delta = \frac{3 \ln(10)}{T_{60}} \quad (4.7)$$

Where,  $T_{60}$  is defined as the RT required for reflections of a direct sound to decay by 60 dB below the level of the direct sound. Because RT is frequency dependent, the statistical model in eq. (4.8) can be implemented in each acoustic frequency bin  $k$  as:

$$h_k(t) = b_k(t)e^{-\Delta_k t}, \quad t \geq 0, \quad k, k = 1, \dots, \dots, k \quad (4.8)$$

Where  $b_k(t)$  is band pass Gaussian noise in the  $k^{\text{th}}$  bin.

The above mentioned model works well when the distance between source and microphone is larger than the critical distance. Critical distance is defined as the source-microphone distance at which the energy of direct path and the energy of all reflections are equal. When source-microphone distance is smaller than the critical distance, Habets [196] proposed a more accurate model as:

$$h_k(t) = \begin{cases} b_d(t), & t < T_r \\ b_r(t)e^{-\Delta t}, & t \geq T_r \end{cases} \quad (4.9)$$

Where,  $b_d(t)$  and  $b_r(t)$  are two separate zero mean Gaussian noise processes;  $T_r$  is a time constant chosen so that  $b_d(t)e^{-\Delta t}$  only contain the direct path component and  $b_r(t)e^{-\Delta t}$  contains all reflections of RIR.

#### 4.4 Reverberant Mask Based Method

Reverberant mask based method is proposed for combined suppression of reverberation and noise. A channel selection criterion based on SRR of the individual channel is used for calculation of reverberant mask. The involved steps are illustrated in Figure 4.2.

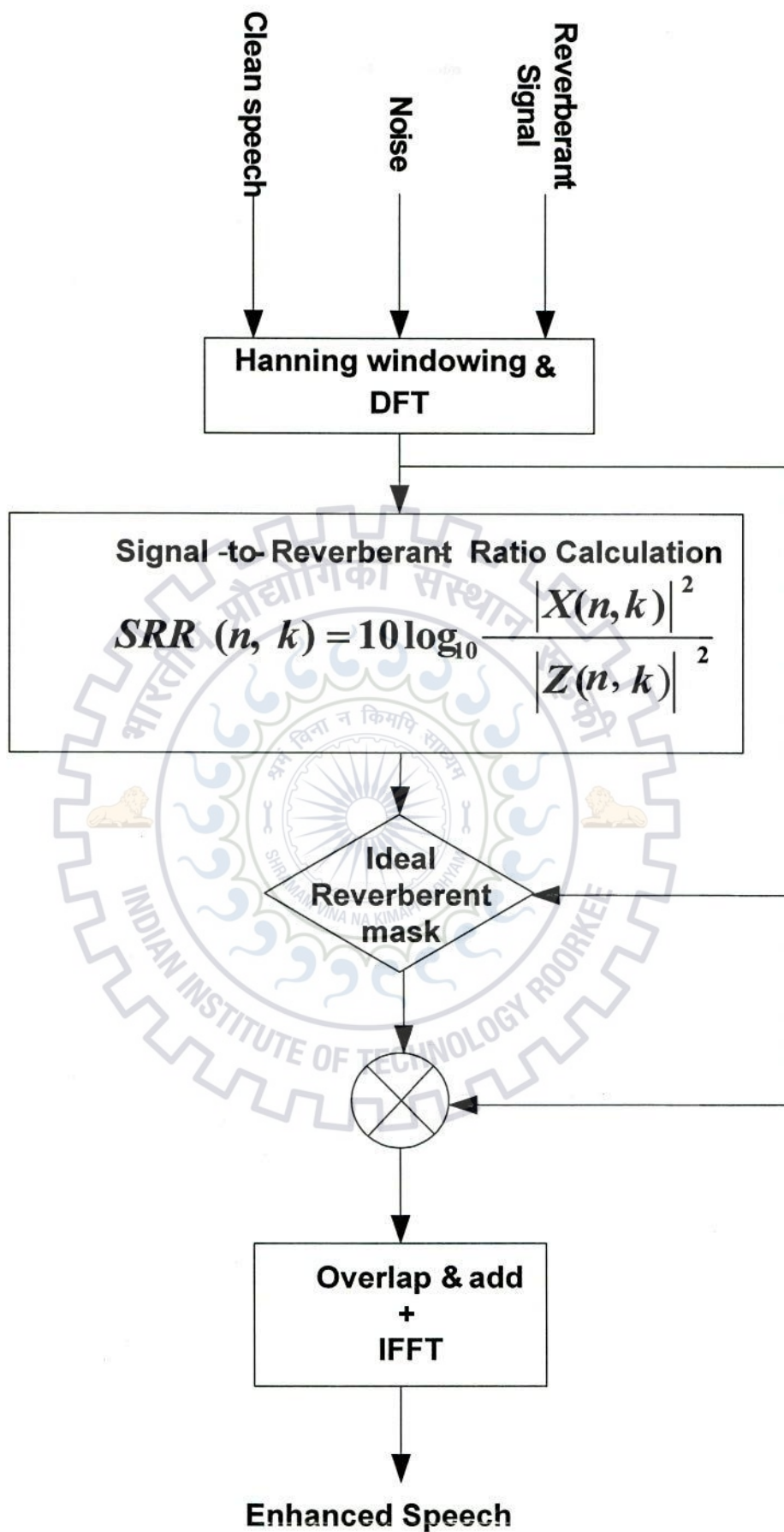


Fig. 4.2: Block diagram of proposed method for speech enhancement.



The amplitudes with SRR greater than a preset threshold (i.e. -5dB) are used for reconstruction of denoised speech, while amplitudes with SRR values smaller than the threshold are eliminated. The SRR reflects implicitly the ratio of the energies of the signal originating from the early (and direct) reflections and the signal originating from the late reflections. The construction of the SRR criterion assumes *a priori* knowledge of the input reverberant and target signal. Threshold values varying from 0dB to -90dB are analyzed for selection of ideal reverberant mask (IRM) limit  $T$ . Enhanced speech signal is constructed by multiplying noisy speech with reverberant mask. The suppression of combined effect of reverberation and noise (i.e. -25 dB to -5 dB) is carried out for single-channel speech enhancement. The block diagram of the proposed method is given in Figure 4.2 which illustrates all steps involved for the suppression of combined effect of all types of noise and reverberation. The processing of speech enhancement starts with the generation of noisy input speech database.

#### 4.4.1 Database

The test samples of English language are taken from the Institute of Electrical and Electronics Engineers (IEEE) database of NOIZEUS [140, 203] and the AURORA noise data base [204]. Each sentence of NOIZEUS database is a combination of 7-12 alphabets and there are 72 lists of 10 sentences each, which have been produced by a single talker. The sampling rate was 16 kHz at time of recording. The MATLAB (2014a) software is used for program implementation of proposed algorithm. The effect of varying multitalker babble noise with different SNR threshold has been taken in this study. The reverberated speech signal with additive noise is divided into frames for performance evaluation of the proposed method. One frame is 20ms of the speech and 50% overlap is used in between successive frames.

#### 4.4.2 Signal model

In an acoustic environment reverberation and noise sources are present which degrade the actual speech signal. A noisy speech signal which is corrupted from reverberation and stationary, non-stationary noise is defined by eq. (4.10) such as:

$$Y(n) = Z(n) + D(n), \quad n \in [0, N-1], \quad (4.10)$$

Where,  $N$  is frame index number. The noisy speech signal  $Y(n)$  is comprised of reverberated speech signal  $Z(n)$ , stationary and non-stationary noise  $D(n)$ . Reverberated speech  $Z(n)$  results from the convolution of the clean speech signal  $X(n)$  and the room

impulse response (RIR)  $h(n)$ . To generate reverberant signals, clean signals are convolved with real room impulse response which are recorded by Vanden Bogaert et al. with reverberation time  $T_{60}=1.0s$  and direct reverberant ratio is  $-0.49$  dB [205]. In Vanden Bogaert setup 1m distance has been used between the single source signal and the microphone. The multi-talker babble noise is added to the reverberant signals at 0dB, 5dB, 10dB and 15dB SNR with varying threshold from  $-23$  dB to  $0$  dB. This reverberant speech signal with noise signal served as the target signal for Mean Square Error (MSE) measurement.

#### 4.4.3 Reverberant mask calculation

The speech enhancement techniques which are based on maximum selection criterion take erroneously peaks of amplitudes during the gaps of unvoiced segments of the utterance. Due to this reason, the vowel and consonant are smeared and make it difficult for the listener to identify the words. In the proposed method, the amplitudes corresponding to the direct sound and early reflections have been taken and the amplitudes corresponding to the late reflections and various additive noise have been discarded. The SRR is used for amplitude selection, which selects the frame that has amplitude value more than the preset threshold. Here, a range of different threshold values are considered for testing that varies from  $-23$  dB to  $0$  dB so that maximum speech quality and intelligibility can be achieved. In this method, noisy speech signal is divided in frames of 20 ms noisy speech signal with 50% overlapping between frames by using Hanning window. In the frequency domain, DFT decomposes the signal into  $n$  frequency channels, where,  $n$  is the frame index which is selected based on the SRR. This SRR is computed as [206]:

$$SRR(n, k) = 10 \log_{10} \frac{|X(n, k)|^2}{|Z(n, k)|^2}, \quad (4.11)$$

Where,  $X(n, k)$  and  $z(n, k)$  denote the clean and reverberant signals, respectively,  $n$  corresponds to the time-frame index, and  $k$  defines the frequency or channel index. Higher value of SRR will give the direct signal amplitude (and early reflections) for generation of de-reverberated signal. In contrast, a small SRR value will give the reverberant energy. The sum of the energies from early reflections will give denoised speech. Now, the overlap-masking effect is minimized by removing the reverberant energy residing in the gaps of different speech frames. The denoised speech signal is calculated by multiplying the reverberant signal with ideal reverberant mask (IRM) function. The IRM function is based on SRR and preset threshold value. It takes 1 when SRR is greater than prescribed threshold (T) and shows zero

if SRR is less than threshold (T). The de-reverberated speech signal is obtained as shown in eq. (4.12):

$$\hat{X}(n, k) = Y(n, k) * IRM(n, k), \quad (4.12)$$

The IRM is given in eq. (4.13)

$$IRM(n, k) = \begin{cases} 1, & \dots, SRR(n, k) > Threshold(T) \\ 0, & \dots, otherwise \end{cases} \quad (4.13)$$

Overlap-add method and Inverse FFT is applied on processed speech spectrum to find the denoised speech signal in time domain.

#### 4.4.4 Results and discussion

The AURORA noise database [204] and reverberation of speech signal are used for the performance evaluation of the implemented methods. The multi-talker babble noise and reverberation signal are mixed with clean speech signal for generating noisy input speech.

Table 4.1 illustrate the SNR improvement with variation of threshold values from 0 dB to -90 dB. For the analysis four levels (0, 5, 10 and 15dB) of input noise SNR are used and mixed with reverberated speech. From the Table, it is found that the output SNR becomes approximately constant when threshold value is set below -70 dB. The same conclusion is drawn from the Table 4.2 in variation of MSE values. The MSE values become approximately constant when threshold value is set below -70 dB. The cut-off threshold value is obtained from these results and it is found that the maximum improvement in reverberated noisy speech environment is obtained at -5dB threshold (T).

The output SNR given by different speech enhancement methods are compared with proposed method in Figure 4.3. The proposed method gives maximum SNR (5.9 dB) improvement in comparison to other speech enhancement methods such as MMSE, logMMSE, pMMSE, Spectral Subtraction and Wiener method. The pMMSE method shows minimum improvement in output SNR (3.1dB). Figure 4.4 illustrates the MSE values obtained from processed output speech signal. The minimum MSE value is obtained for proposed method in comparison to other speech enhancement methods. Figures 4.3 and 4.4 illustrate that the proposed method gives maximum improvement in speech quality in comparison to existing speech enhancement methods.

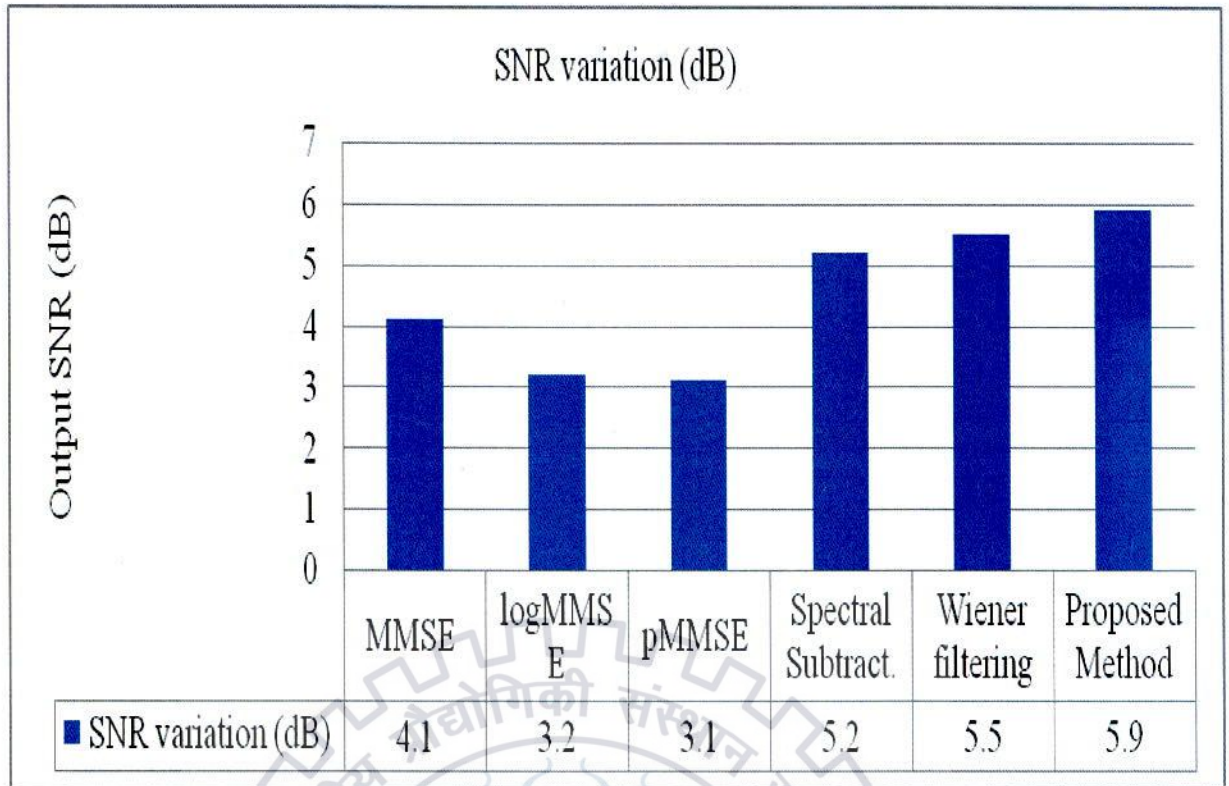


Fig. 4.3: Variation of SNR for various speech enhancement algorithms.

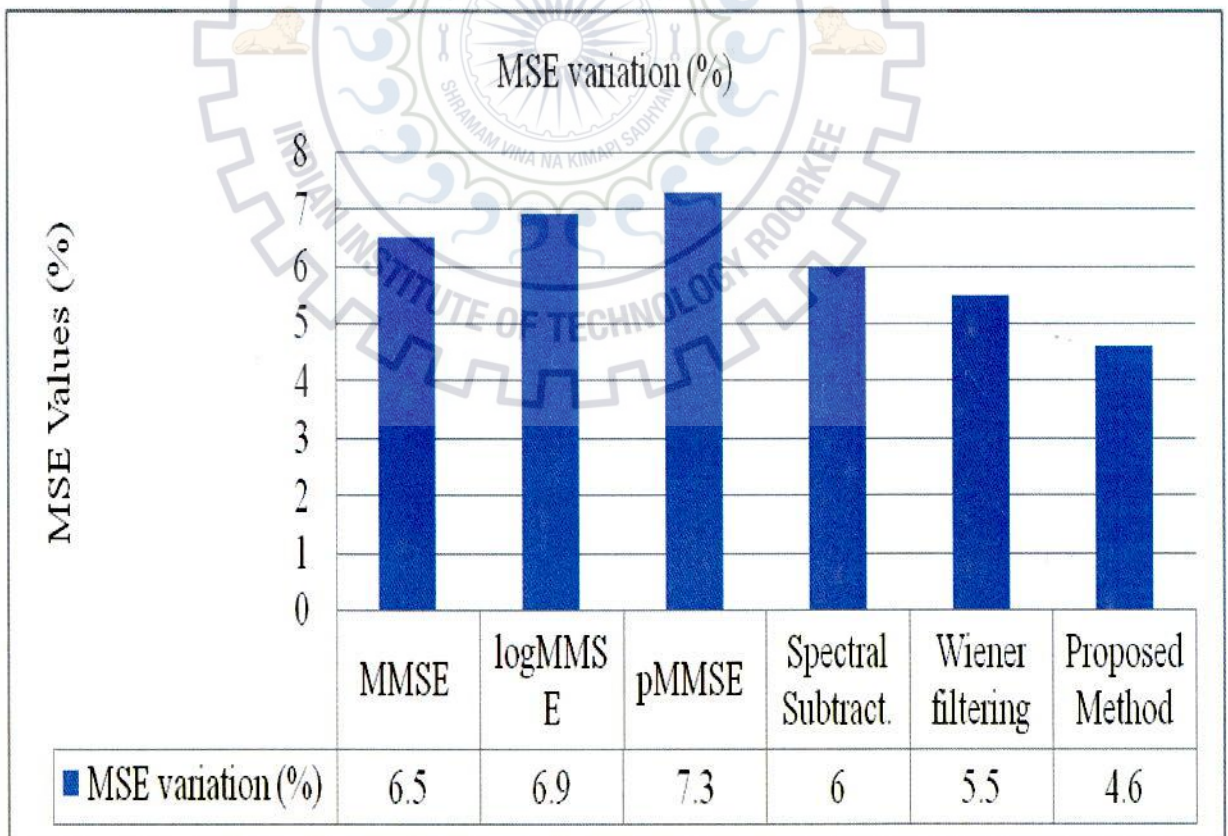


Fig. 4.4: Variation of MSE for various speech enhancement algorithms.

Table 4.1: Output SNR values with variation of input noise SNR and threshold values.

Threshold	S+N(0dB)	S+N(5dB)	S+N(10dB)	S+N(15dB)
0	-20.3	-27.9	-33.94	-37.6
-5	-11.7	-18.8	-24.1	-26.44
-8	-0.82	-7	-15.4	-19
-10	1.54	4.63	6.83	6.4
-15	3.17	7.1	9.6	11.32
-20	3.82	7.83	10.0	11.57
-30	5.13	9.0	10.63	11.79
-40	6.6	9.94	11.19	11.94
-50	8.0	10.8	11.67	12.08
-60	9.3	11.45	11.98	12.17
-70	10.7	11.93	12.1	12.19
-80	11.5	11.98	12.13	12.19
-90	11.6	11.99	12.13	12.19

Table 4.2: Output MSE values with variation of input noise SNR and threshold values.

Threshold	S+N(0dB)	S+N(5dB)	S+N(10dB)	S+N(15dB)
0	0.0228	0.0184	0.0151	0.0222
-5	0.0218	0.0159	0.0121	0.0209
-8	0.0215	0.0147	0.0106	0.0206
-10	0.0218	0.0135	0.0096	0.0205
-15	0.0206	0.0117	0.0077	0.0201
-20	0.0200	0.0101	0.0069	0.0196
-30	0.0162	0.0078	0.0056	0.0155
-40	0.0114	0.0061	0.0041	0.0108
-50	0.0084	0.005	0.0032	0.0084
-60	0.0051	0.0045	0.0030	0.0052
-70	0.0053	0.0045	0.0029	0.0048
-80	0.0043	0.0045	0.0029	0.0037
-90	0.0042	0.0045	0.0029	0.0036

The PESQ values computed from speech signal given by various speech enhancement methods is given in Table 4.3. For direct interpretation of the results given in Table 4.3, the graphical and bar graph representation of the speech enhancement methods (log-MMSE, p-MMSE, MMSE, spectral subtraction and Wiener) for different types of mixed noise are presented in Figures 4.5 and 4.6.

Wiener and spectral subtraction methods obtained the maximum PESQ values than other existing speech enhancement methods at all noise sources and input SNR levels but these speech enhancement methods are inferior to proposed method. The maximum PESQ values (3.0659, 2.8641, 2.6112, 2.3759, and 2.1906) for -5, -10, -15, -20 and -25 dB input SNR levels, respectively are provided by proposed method. Since, the maximum PESQ values are given by proposed method hence it results in maximum quality and intelligibility improvement in processed speech.

Table 4.3: PESQ scores in reverberation condition.

Noise Types	Enhancement Techniques	PESQ				
		-5	-10	-15	-20	-25
<b>Reverberation + Babble</b>	log-MMSE	2.1676	1.8356	1.5905	1.3440	1.0751
	p-MMSE	0.9854	0.3448	0.6554	0.3286	0.3126
	Wiener	2.3672	1.9566	1.8550	1.6701	1.5180
	MMSE	2.2137	1.9104	1.7599	1.3934	1.0462
	Spectral sub	2.1904	1.7972	1.4384	0.9406	0.6061
	Proposed method	<b>2.3538</b>	<b>2.1889</b>	<b>1.9834</b>	<b>1.8244</b>	<b>1.6987</b>
<b>Reverberation + Pop Music</b>	log-MMSE	2.0154	1.6723	1.5449	1.4590	1.4132
	p-MMSE	0.9996	0.6905	0.4257	0.4623	0.4630
	Wiener	2.4156	2.1180	1.9778	1.7468	1.6612
	MMSE	2.3403	1.9484	1.7709	1.5985	1.2230
	Spectral sub	2.0594	1.8465	1.4637	0.9515	0.3823
	Proposed method	<b>3.0659</b>	<b>2.8641</b>	<b>2.6112</b>	<b>2.3759</b>	<b>2.1906</b>
<b>Reverberation + Restaurant</b>	log-MMSE	2.1506	1.9256	1.7537	1.6989	1.5302
	p-MMSE	0.9895	0.7448	1.0114	0.9777	0.8514
	Wiener	2.4108	2.2613	1.9726	1.7775	1.6469
	MMSE	2.3388	2.2600	1.8659	1.5832	1.2471
	Spectral sub	2.2379	1.8774	1.4880	1.2344	0.8345
	Proposed method	<b>2.5046</b>	<b>2.2852</b>	<b>2.1121</b>	<b>1.9329</b>	<b>1.6865</b>
<b>Reverberation + Exhibition</b>	log-MMSE	2.3856	2.1236	1.8751	1.6452	1.5514
	p-MMSE	0.9279	0.6747	0.8491	1.0593	0.8192
	Wiener	2.6743	2.4085	2.1916	1.9271	1.7671
	MMSE	2.5058	2.1417	1.8827	1.6182	1.4338
	Spectral sub	2.2129	1.9737	1.7409	1.4215	1.1908
	Proposed method	<b>2.7390</b>	<b>2.4769</b>	<b>2.2463</b>	<b>2.0383</b>	<b>1.7918</b>

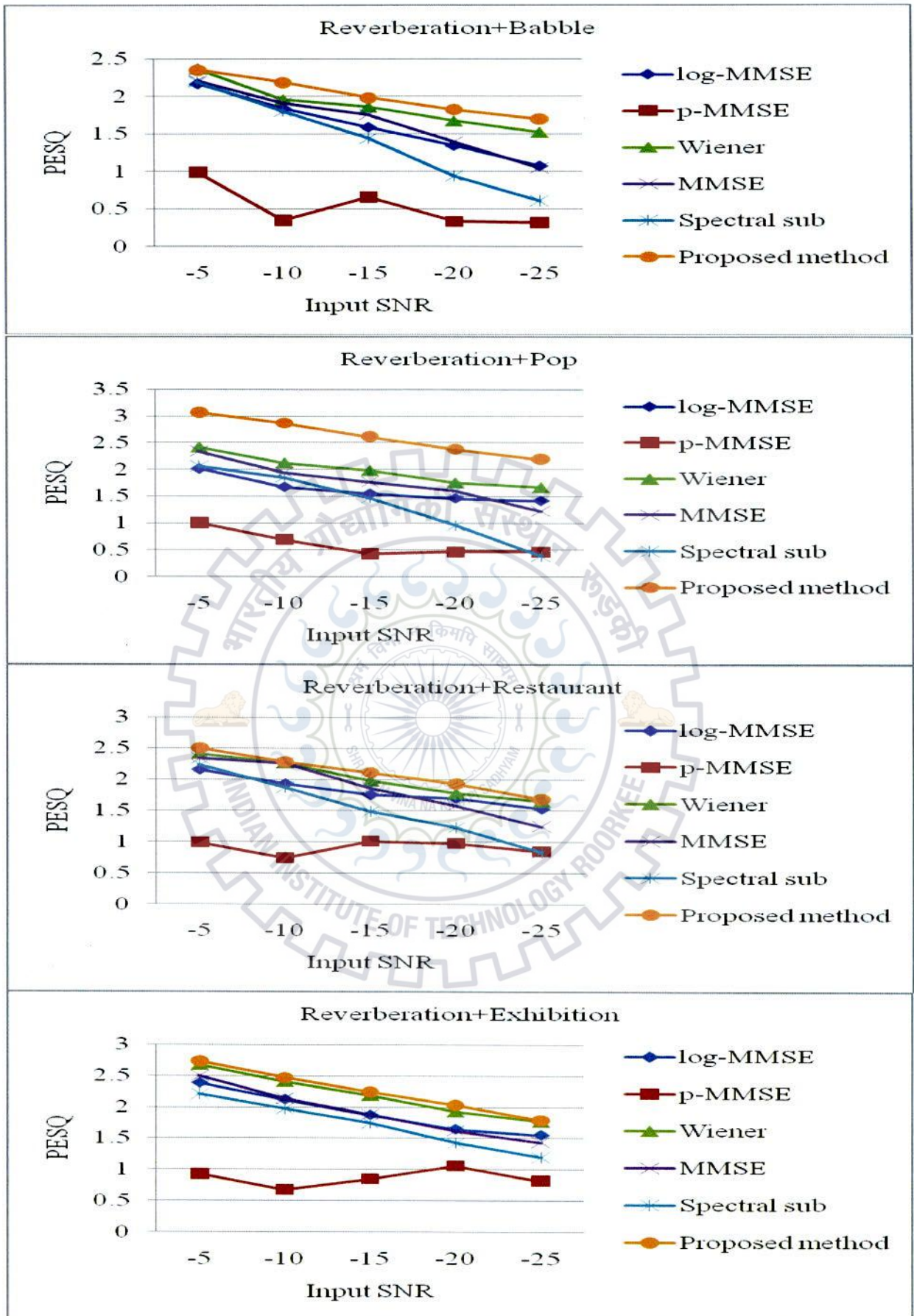


Fig. 4.5: Variation of results between PESQ scores and input SNR.

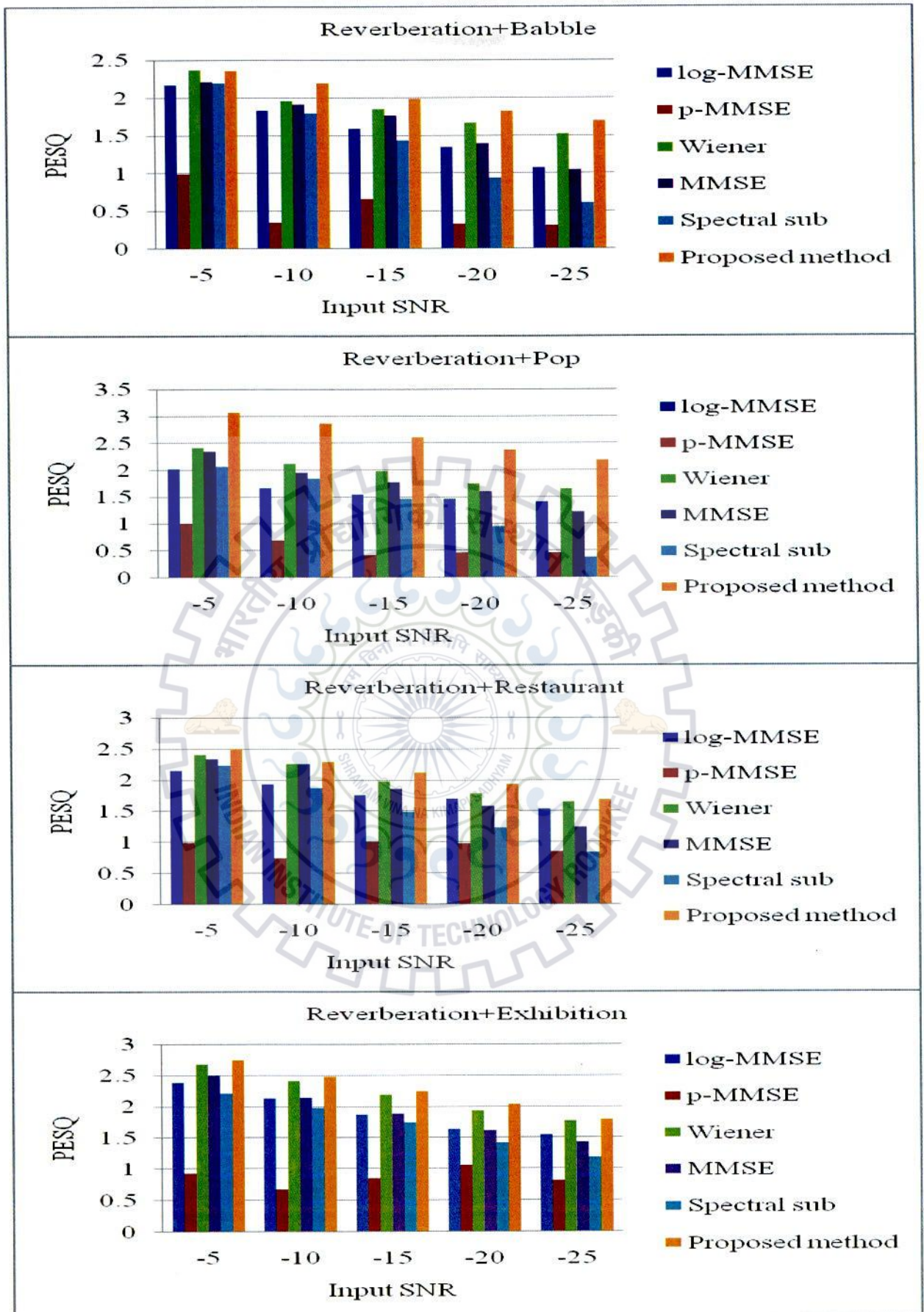


Fig. 4.6: Variation of results between PESQ scores.



Table 4.4 presents the values of CD at various levels of input SNR in presence of reverberation and different types of noise (i.e. babble, pop music, restaurant and exhibition). Results obtained by Wiener method are better to other existing speech enhancement methods. The CD values given by Wiener method are 4.4733, 4.9254, 5.3910, 5.5810 and 6.2071 at -5, -10, -15, -20 and -25 dB input SNR, respectively but Wiener method performs inferior to the proposed speech enhancement method. The CD values (4.4359, 4.8991, 5.2019, 5.4335 and 5.4520 at -5, -10, -15, -20 and -25 input SNR, respectively) given by proposed method are lower than the Wiener method hence the proposed method is superior than other existing speech enhancement methods including Wiener method. The Figures 4.7 and 4.8 present the graphical and bar chart representation of the values given in Table 4.4 to clearly indicate the above observations.

Table 4.4: Cepstrum Distance values in reverberation condition.

Noise Types	Enhancement Techniques	Cepstrum Distance				
		-5	-10	-15	-20	-25
<b>Reverberation+ Babble</b>	log-MMSE	4.5921	4.9528	5.4306	6.1142	6.7387
	p-MMSE	8.2647	8.6265	8.8290	8.8998	8.9039
	Wiener	4.4733	4.9254	5.3910	5.5810	6.2071
	MMSE	5.4830	6.2020	7.1333	7.7539	8.6447
	Spectral sub.	5.6965	6.3803	7.0233	7.6207	7.9614
	Proposed method	<b>4.4359</b>	<b>4.8991</b>	<b>5.2019</b>	<b>5.4335</b>	<b>5.4520</b>
<b>Reverberation+ Pop Music</b>	log-MMSE	5.2935	5.7240	6.2458	6.7787	7.3119
	p-MMSE	7.5246	7.8952	8.0725	8.1274	8.1438
	Wiener	5.1802	5.6672	6.0720	6.1722	6.5200
	MMSE	6.4523	7.4001	8.0207	8.4666	8.8016
	Spectral sub.	7.0658	7.6249	8.2241	8.5619	8.6890
	Proposed method	<b>5.0162</b>	<b>5.5138</b>	<b>5.9538</b>	<b>5.9862</b>	<b>5.9638</b>
<b>Reverberation+ Restaurant</b>	log-MMSE	4.8125	4.7132	<b>4.8997</b>	<b>5.1571</b>	5.6619
	p-MMSE	8.1935	8.5898	8.8087	8.89100	8.8966
	Wiener	<b>3.9227</b>	<b>4.3357</b>	4.9270	5.4738	5.8093
	MMSE	5.5893	6.4502	7.4589	8.0319	8.5047
	Spectral sub.	5.0776	5.8999	6.3150	6.7133	7.2347
	Proposed method	4.4982	4.9467	5.2900	5.3374	<b>5.3237</b>
<b>Reverberation+ Exhibition</b>	log-MMSE	5.7104	6.2065	6.6398	7.2652	7.8753
	p-MMSE	8.0783	8.4198	8.5976	8.6747	8.7115
	Wiener	5.5191	6.1476	6.3742	7.1502	6.5564
	MMSE	6.6778	7.3998	7.8321	8.1800	8.3435
	Spectral sub.	7.7872	8.3901	8.9271	9.2187	9.3881
	Proposed method	<b>5.2649</b>	<b>6.0708</b>	<b>6.2453</b>	<b>6.3502</b>	<b>6.2115</b>

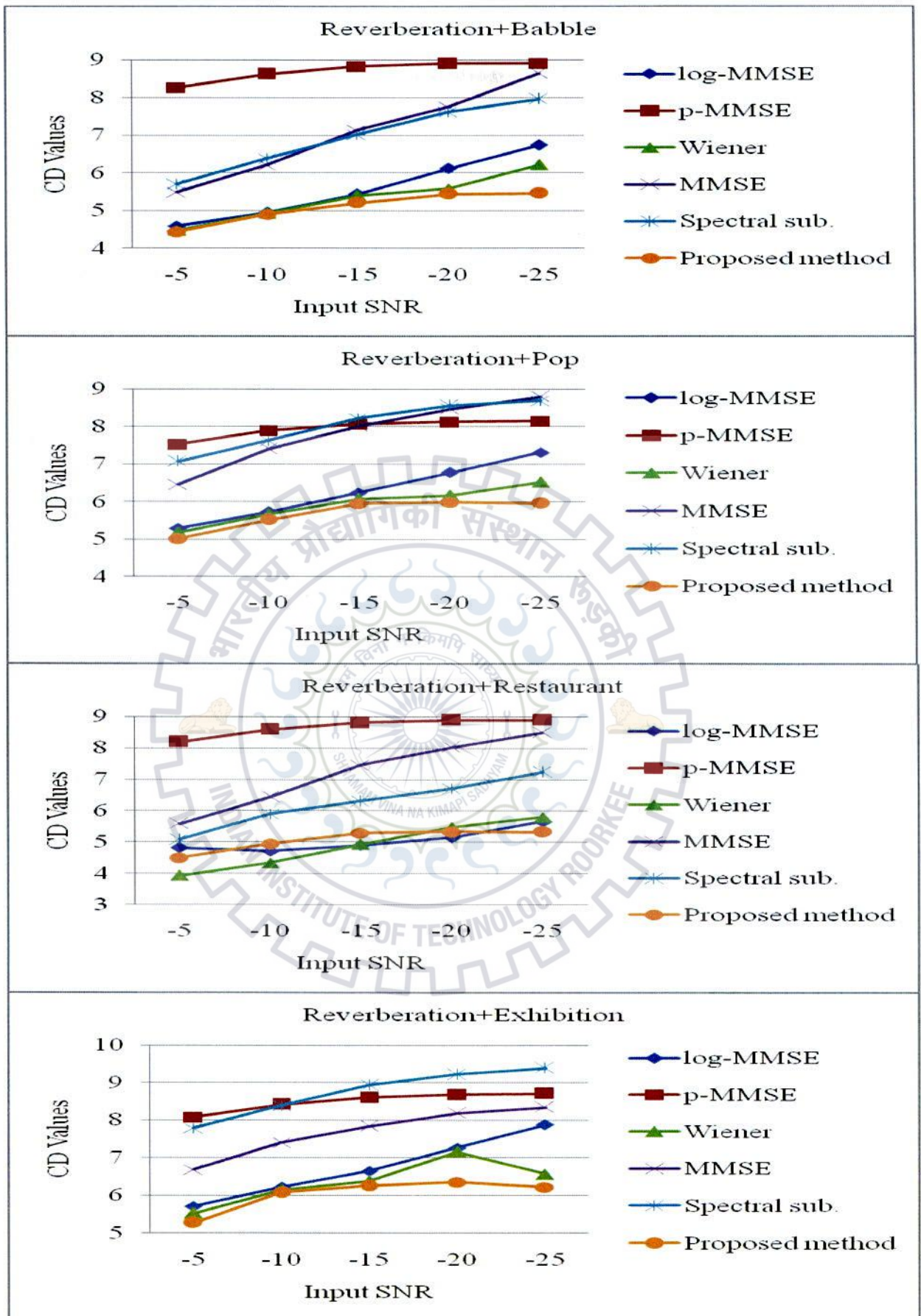


Fig. 4.7: Variation of results between CD scores and input SNR at reverberation+ noise.

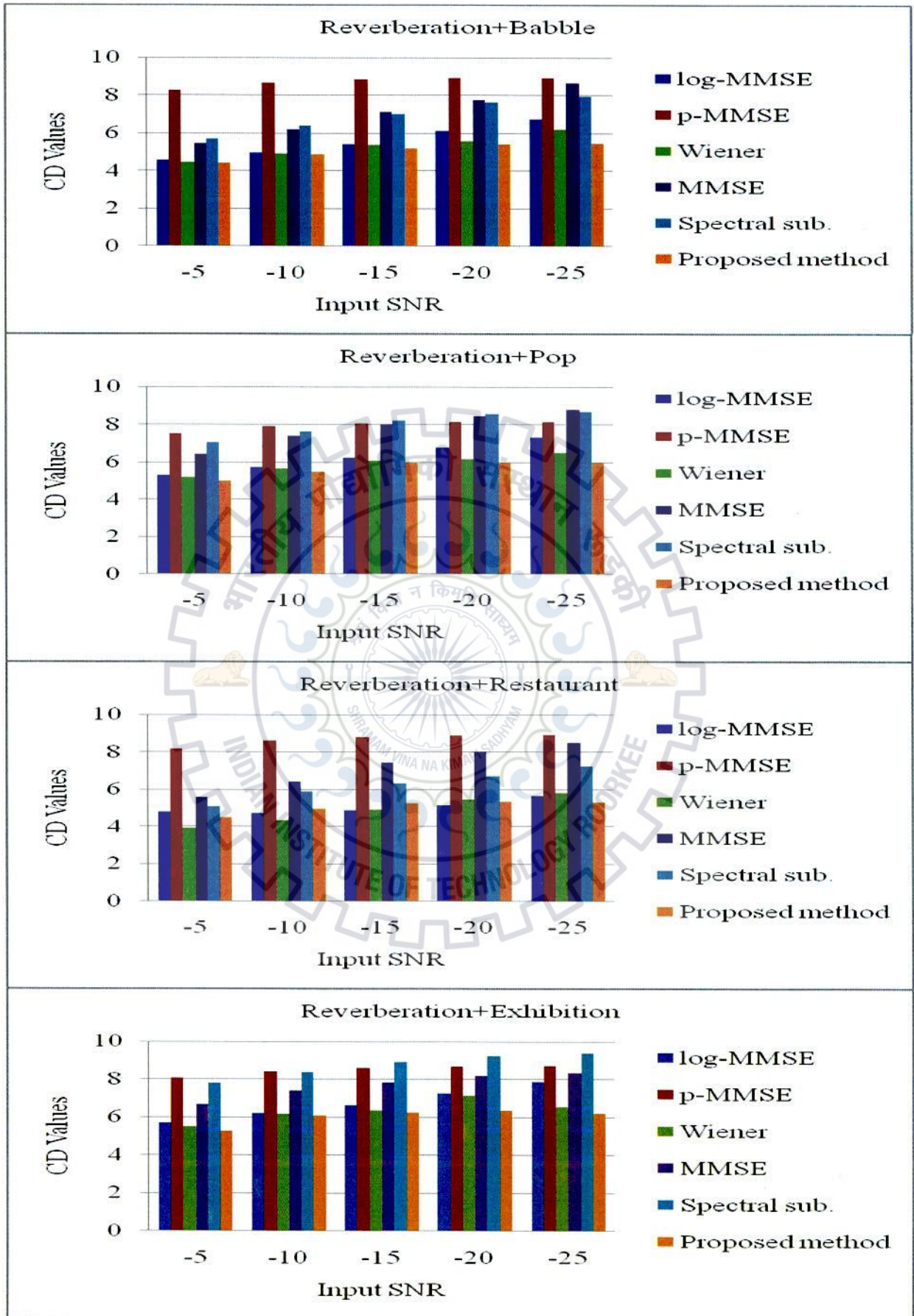


Fig. 4.8: Variation of results between CD scores.

Thus it is summarized that, all the results given in Tables (from 4.1 to 4.4) and Figures (from 4.3 to 4.8) show that, the proposed method gives maximum speech enhancement than those obtained by log-MMSE, p-MMSE, MMSE, Wiener and spectral-subtraction methods.

#### 4.5 Summary

Reverberant mask criterion based method is proposed for combined suppression of reverberation and highly non-stationary noise of low SNR. Threshold values from 0 to -90 dB are used for analysis and -5 dB is selected as trade off in ideal reverberant mask selection. The performance of proposed method is compared with other existing speech enhancement methods. Reverberant mask based proposed method gives maximum improvement at all levels of input SNR.



*This chapter describes the importance of phase in single-channel speech enhancement methods. It starts with some background of phase estimation and its use in speech enhancement by considering altered speech phase and without changing in speech amplitude. Then, the description of some important phase based speech enhancement method is presented. In addition to this the proposed phase ratio based speech enhancement method is introduced and compared with other phase based speech enhancement methods.*

## 5.1 Overview

Nowadays, speech enhancement algorithms are implemented for many communicating electronic gadget such as smart phones and hearing aids etc. As these devices are often used in noisy environments and therefore, recent research addresses robustness in non-stationary noise, e.g. babble of low signal-to-noise ratios etc. Though phase is usually considered to be insignificant for human perception as compared to amplitude, this is true only for high SNR ( $>5$  dB) while for lower SNRs phase leads to the speech distortion. However, the phase enhancement is much more difficult and complex than amplitude based speech enhancement.

Typical speech enhancement methods modify only the magnitude spectrum and keep the phase spectrum unchanged. The ignorance of phase spectrum is direct result of studies undertaken in the 1980s which showed that the phase spectrum provide no perceptual difference in the enhanced signal [47, 207-208]. Wang and Lim [47] have done experiments in which they investigated the importance of phase estimation in speech enhancement. They synthesized noisy speech by taking the amplitude and phase from signals with different SNR and observed that improving the noisy spectral amplitude is more important for the signal quality than improving the noisy spectral phase and on the basis of this fact it was concluded that clean speech phase estimation is unimportant in speech enhancement, and until today most of the researcher estimate the clean speech amplitude only, while keeping the noisy phase unaltered.

In amplitude estimators, the speech spectral amplitude is altered while the noisy phase is left unchanged [322-323, 209-214]. Moreover, in common speech enhancement techniques, like sinusoidal modelling etc., it has been proposed to combine improved spectral amplitudes with the noisy STFT phase [215]. Despite the general trend of neglecting STFT phase estimation, Paliwal et al. argue that, potentially, the role of the phase in speech enhancement has been underestimated in the past [51] and showed that if the segment overlap and the length of the Fourier transform are increased, the impact of the clean speech phase is larger

than observed by Wang and Lim [47]. Thus, the speech phase can indeed be beneficial for design of speech enhancement method. While Paliwal et al. proposed methods for speech enhancement that involve modification of complex spectral coefficients [51], the direct estimation of the clean spectral phase is considered a difficult task and only few proposals exist [51, 216-224]. For instance, Griffin and Lim proposed to estimate the spectral phase by iteratively analyzing and synthesizing the signal starting from the spectral amplitudes [216]. However, approach was computationally complex and requires knowledge of the clean speech spectral amplitudes.

Now, it has been proved that speech phase has much importance in speech enhancement process [51]. It could not be neglected where speech quality and intelligibility is very necessary. Some effective phase based speech enhancement approaches are given in following section with detailed description.

## 5.2 Speech Enhancement Methods Using Phase

Despite the use of phase for other speech processing applications (such as features for automatic speech recognition (ASR) [218-220] or speaker identification [222]), phase information has only been used in few methods for single-channel speech enhancement. The phase based speech enhancement techniques (i.e. phase spectrum compensation (PSC) [52, 223], exploiting conjugate symmetry of the short-time Fourier spectrum [53] and STFT-phase for the MMSE-optimal spectral amplitude estimation [50]) are described below.

### 5.2.1 Phase spectrum compensation (PSC)

. Phase Spectrum Compensation (PSC) utilises the synthesis procedure (i.e. IFFT and overlap-add reconstruction) which is commonly used in speech enhancement methods where an enhanced waveform is required for playback. Since the incoming speech signal is real-valued, the DFT coefficients are conjugated symmetric. The PSC controls the amount of reinforcement or cancellation that occurs during synthesis by adding a noise-weighted anti-symmetry function  $\Lambda^i(k)$  to the noisy speech signal in the complex frequency domain [52]:

$$Y_{\Lambda}^i = Y^i + \Lambda^i \quad (5.1)$$

For the frequencies having low noise magnitudes, the anti-symmetry function causes little change to the original signal. For high noise components however, the anti-symmetry function causes the conjugate pairs to cancel during the synthesis stage. Whilst PSC has

shown promising improvements in human intelligibility, it requires reconstruction in the time domain which is unnecessary and sometimes undesirable for speech applications.

### 5.2.2 Exploiting conjugate symmetry of short-time Fourier spectrum

This method is based on analysis modification synthesis (AMS) framework where a real-valued noisy speech signal is used at analysis stage, and therefore, its discrete short-time Fourier transform (DSTFT) is a conjugate symmetric, i.e.,  $Y(n, k) = Y^*(n, N - k)$ . In this approach, emphasis is given to the degree to which the conjugates reinforce or cancel by altering their angles. Thus, the changed phase spectrum is computed as following.

The noisy complex spectrum is offset by an additive real-valued frequency-dependent  $\Lambda(k)$  function [53]:

$$Y_{\Lambda}(n, k) = Y(n, k) + \Lambda(n, k) \quad (5.2)$$

Where,  $\Lambda(k)$  should be made anti-symmetric about  $F_s/2$  (half the sampling rate) to achieve the cancellation effect. Here, anti-symmetric  $\Lambda(n, k)$  function is given as:

$$\Lambda(n, k) = \begin{cases} +\lambda & 0 \leq k < N/2 \\ -\lambda & N/2 \leq k \leq N-1 \end{cases} \quad (5.3)$$

Where  $\lambda$  is a real-valued constant and  $N$  is the length of frequency analysis assumed to be even.  $Y_{\Lambda}(n, k)$  is used to compute the changed phase spectrum through a four-quadrant version of the arctangent function as:

$$\angle Y_{\Lambda}(n, k) = \arctan \left( \frac{\text{Im} \{ Y_{\Lambda}(n, k) \}}{\text{Re} \{ Y_{\Lambda}(n, k) \}} \right) \quad (5.4)$$

Where,  $\text{Im}\{\bullet\}$  and  $\text{Re}\{\bullet\}$  denote imaginary and real operators, respectively. This changed phase spectrum is a pseudo-phase spectrum. The pseudo-phase spectrum is recombined with the noisy magnitude spectrum to produce a modified complex spectrum

$$\hat{X}_{\Lambda}(n, k) = |Y(n, k)| e^{j\angle Y_{\Lambda}(n, k)} \quad (5.5)$$

In the synthesis stage, Inverse-DSTFT is used to convert frequency-domain frames,  $\hat{X}_{\Lambda}(n, k)$  to time-domain representation and time-domain enhanced signal is reconstructed by employing overlap-add procedure.

### 5.2.3 MMSE-optimal spectral amplitude estimation based on STFT-phase

In this method clean speech amplitude was estimated with the use of STFT phase. This was based on minimization of the mean squared error between the compressed clean speech amplitudes  $A^\beta$  and the estimator for compressed amplitudes  $\hat{A}^\beta$  [20]. A compression factor  $\beta < 1$  allows emphasizing estimation errors of low amplitudes, and for  $\beta \rightarrow 0$  a logarithmic spectral amplitude estimator was approximated [224, 214]. The  $\hat{A}^\beta$  is given as [50]:

$$\begin{aligned}\hat{A}^\beta &= E\left(A^\beta \mid r, \phi_Y, \phi_S\right) \\ &= \int_{-\infty}^{\infty} a^\beta p_{A|R, \phi_Y, \phi_S}(a \mid r, \phi_Y, \phi_S) da\end{aligned}\quad (5.6)$$

Using Bayes' theorem it is given as

$$\hat{A}^\beta = \frac{\int_{-\infty}^{\infty} a^\beta p_{R, \phi_Y} \mid A, \phi_S(r, \phi_Y \mid a, \phi_S) p_{A, \phi_S}(a, \phi_S) da}{\int_{-\infty}^{\infty} p_{R, \phi_Y \mid A, \phi_S}(r, \phi_Y \mid a, \phi_S) p_{A, \phi_S}(a, \phi_S) ds}\quad (5.7)$$

As in [211-214], the real and imaginary parts of the complex noise spectral coefficients are independent and Gaussian distributed. Thus, the conditional probability density function (PDF) of the noisy coefficients was obtained if the speech coefficients are given. This PDF is obtained as:

$$p_{R, \phi_Y \mid A, \phi_S}(r, \phi_Y \mid a, \phi_S) = \frac{r}{\pi \sigma_N^2} \exp\left(-\frac{r^2 + a^2 - 2ar \cos(\phi_Y - \phi_S)}{\sigma_N^2}\right)\quad (5.8)$$

To model the PDF of the speech spectral amplitudes by using X-distribution with shape parameter  $\mu$ , the X-distribution is a special case of the generalized Gamma-distribution [213]. The X-distribution is defined as

$$p_A(a) = \frac{2}{\Gamma(\mu)} \left(\frac{\mu}{\sigma_S^2}\right)^\mu a^{2\mu-1} \exp\left(-\frac{\mu}{\sigma_S^2} a^2\right)\quad (5.9)$$

Now, inserting eq. (5.9) and (5.8) into eq. (5.7) [26] to get eq. (5.10) and the actual estimator as:



$$\hat{A} = \left( E \left( A^\beta \middle| r, \phi_Y, \phi_S \right) \right)^{\frac{1}{\beta}}$$

$$= \sqrt{\frac{1}{2} \frac{\xi}{\mu + \xi} \sigma_N^2 \left( \frac{\Gamma(2\mu + \beta)}{\Gamma(2\mu)} \frac{D_{-(2\mu+\beta)}(v)}{D_{-(2\mu)}(v)} \right)^{\frac{1}{\beta}}}$$
(5.10)

Where,  $D$  is the parabolic cylinder function and  $\xi$  is the *a-priori* SNR. This method was not investigated for low input SNR and highly non-stationary noise.

### 5.3 Proposed Phase Ratio Based Single-Channel Speech Enhancement Method

Despite the original statement made by Wang and Lim in 1982 [47], and the recent findings in [211-214], researchers have so far been unable to develop an appropriate method for estimating either the noise or speech phase spectrum. Motivated by the lack of suitable solution, the contributions contained in this chapter include a signal phase ratio based method for single-channel speech enhancement. The proposed approach is described in subsequent sections.

#### 5.3.1 Signal model and notation

Let us consider noisy speech is an additive superposition of noise over clean speech as expressed in eq. (5.11).

$$Y(n) = X(n) + D(n)$$
(5.11)

Where,  $Y(n)$ ,  $X(n)$  and  $D(n)$  denote for the noisy, clean and noise signals in discrete-time domain, respectively. Now, discrete short-time Fourier transform (STFT) of the corrupted speech signal  $Y(n)$  is given by eq. (5.12) as:

$$Y(n, k) = \sum_{p=-\infty}^{\infty} y(p) w(n-p) e^{-j2\pi k p / N}$$
(5.12)

Where,  $k$  denotes the  $k^{\text{th}}$  discrete-frequency of  $N$  uniformly spaced frequencies and  $w(n)$  is function of analysis window. Hanning window with frame length of 240 samples were used in speech processing and 8 kHz sampling frequency with 50 percentage overlapping.

Now eq. (5.11) can be represented as:

$$Y(n, k) = X(n, k) + D(n, k)$$
(5.13)

Where,  $Y(n, k)$ ,  $X(n, k)$  and  $D(n, k)$  are the STFT of the noisy, clean and noise signals, respectively. The representation of eq. (5.13) in terms of STFT magnitude and STFT phase spectrum of noisy speech signal is given as in eq. (5.14).

$$Y(n, k) = |Y(n, k)|e^{j\angle Y(n, k)} \quad (5.14)$$

Where,  $|Y(n, k)|$  and  $\angle Y(n, k)$  are magnitude and phase spectrum, respectively.

With the aim of speech enhancement, a signal phase ratio based approach is implemented for speech enhancement where noisy speech phase is altered. In this method two gain functions are used for speech phase enhancement. The procedure of the phase ratio based proposed approach is shown in Figure 5.1 and it is explained in section 5.3.2.

### 5.3.2 Signal phase ratio based approach

For the speech phase enhancement, a phase ratio based algorithm is implemented and evaluated. In this method, phase ratio is calculated from noise and noisy speech. The values of all constants are determined in such a way as to maximize speech intelligibility. The two gain functions  $G_1$  and  $G_2$  are calculated for suppressing noise coming from angles  $0$  to  $\pm\pi/2$  and  $\pm\pi/2$  to  $\pm\pi$ , respectively. Phase ratio based single-channel speech enhancement method has mainly three steps:

Step1: Calculation of phase ratio

Step2: By using phase ratio find out  $G_1$  and  $G_2$  for correcting phase to suppress noise coming from angles  $0$  to  $\pm\pi/2$  and  $\pm\pi/2$  to  $\pm\pi$ , respectively.

Step3: extracting correct phase by using  $G = G_1 \cdot G_2$  by using eq. (5.15) to (5.18).

For calculating phase ratio, the angle of STFT noise and noisy spectrum is calculated as given in eq. (5.15) and (5.16):

$$P_D(n, k) = \angle D(n, k) \quad (5.15)$$

$$P_Y(n, k) = \angle Y(n, k) \quad (5.16)$$

$$P_{YD}(n, k) = \lambda P_Y(n, k) + (1 - \lambda) P_D(n, k) \quad (5.17)$$

$$\text{Phase Ratio}(n, k) = P_{YD}(n, k) / (P_D(n, k) * P_Y(n, k) + \xi) \quad (5.18)$$

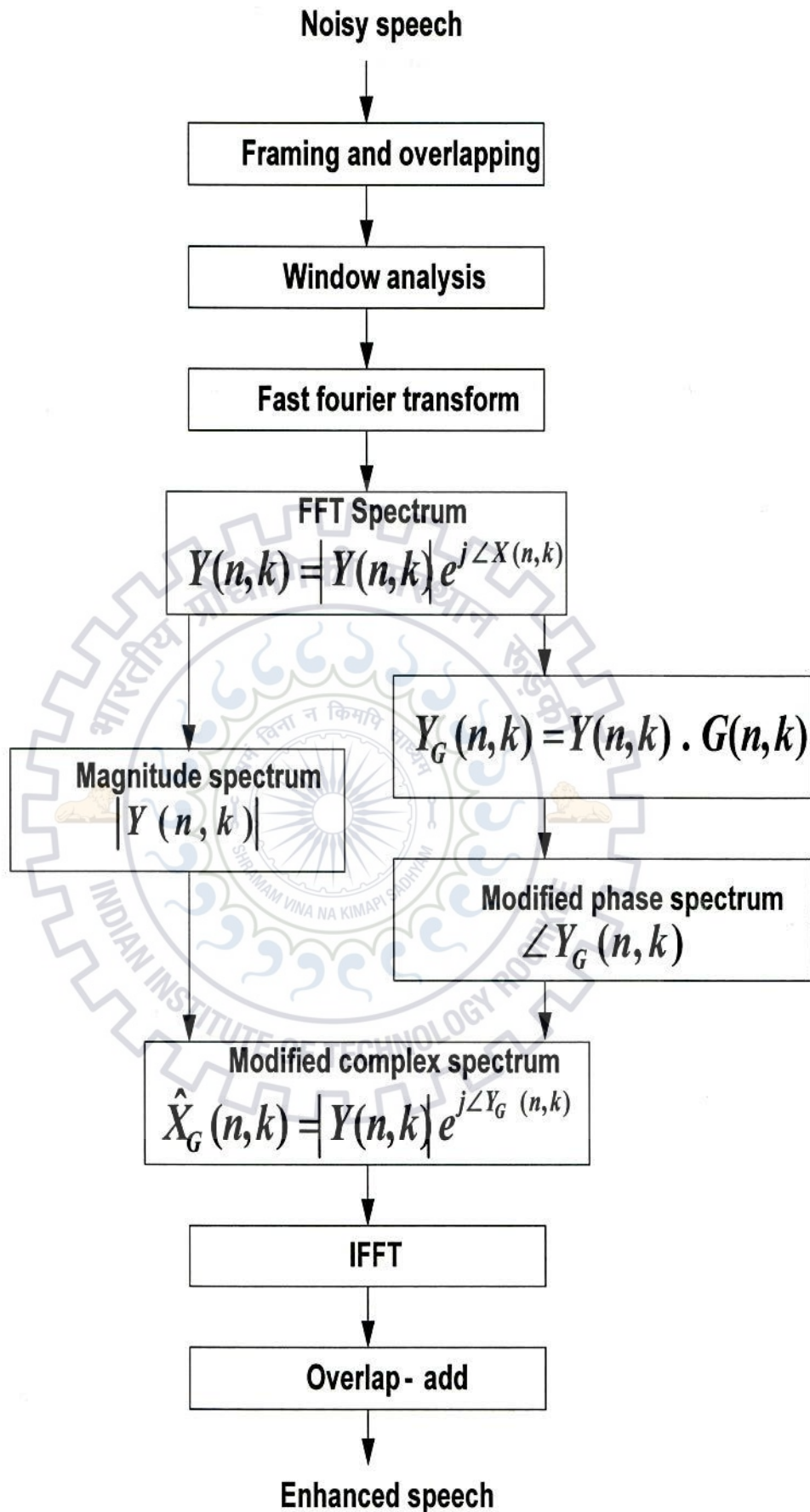


Fig. 5.1: Block diagram of the phase ratio based single-channel speech enhancement method.

Where,  $P_D(n, k)$ , and  $P_Y(n, k)$  give angle of noise and noisy spectrum, respectively. The combined angle  $P_{YD}(n, k)$  of noise and noisy signals are measured by using eq. (5.17) and eq. (5.18), respectively. The forgetting factors  $\lambda$ , and  $\xi$  are held to a fixed value 0.65 and  $10^{-12}$ , respectively. These factors are used to control the noise part so that distortions are reduced in processed speech [11].

Now, to cancel the noise signals coming from angles 0 to  $\pm\pi/2$ , gain functions  $G_1$  and  $G_2$ , for correcting phase to suppress noise coming from angles 0 to  $\pm\pi/2$  and  $\pm\pi/2$  to  $\pm\pi$ , respectively are calculated by using eq. (5.19), and (5.20), respectively.

$$G_1(n, k) = 1 - \text{Phase Ratio} \quad (5.19)$$

$$G_2(n, k) = \begin{cases} 0.05, & \text{if } G_1 < \mu \\ 1, & \text{Otherwise} \end{cases} \quad (5.20)$$

Where,  $\mu$  is constant value of -0.3 for suppressing noise coming from angle more than  $\pm\pi/2$  [11]. Final filter  $G$  is calculated by using eq. (5.19) and (5.20).

$$G(n, k) = G_1(n, k) * G_2(n, k) \quad (5.21)$$

The corrected phase spectrum is calculated by using eq. (5.21) and (5.14).

$$Y_G(n, k) = Y(n, k) * G(n, k) \quad (5.22)$$

The enhanced single-channel speech  $\hat{S}_G(n, k)$  is calculated by using eq. (5.22) and (5.14) as given in eq. (5.23).

$$\hat{X}_G(n, k) = |Y(n, k)| e^{j\angle Y_G(n, k)} \quad (5.23)$$

The inverse STFT is used to get the processed speech signal in time-domain. Now, overlap-add method is employed for the reconstruction of enhanced single-channel speech signal.

### 5.3.3 Results and discussion

For evaluating speech enhancement algorithms, NOIZEUS clean speech database is used which have been recorded from six speakers (three males and three females) [140]. This database comes with various additive non-stationary noise types and noisy speech at different

SNR levels (i.e. -25, -20, -15, -10, -5, 0, and 5 dB). The noise sources which are used for evaluation are such as M109 or Tank, Buccaneer or Jet cockpit, Leopard or Military vehicle and babble.

Table 5.1: Performance Measure in terms of PESQ.

Noise Type	Enhancement Techniques	PESQ						
		Noise Levels	-25	-20	-15	-10	-5	0
Babble	MMSE-Phase	2.3824	2.5737	2.7683	2.9503	3.1763	3.4931	3.7218
	KLT	0.9386	0.8479	0.8283	1.6447	1.8966	2.2526	2.5489
	PKLT	0.9204	0.7014	0.8279	0.5738	1.7319	2.0773	2.3854
	SPEC SUB	1.1121	1.0317	1.0722	1.7488	2.0126	2.2108	2.5457
	PSC	1.0320	1.1335	1.0831	1.8500	2.1949	2.4392	2.6529
	MMSE	1.0404	0.9857	0.9372	0.9501	2.0885	2.3563	2.5993
	Conjugate Sy.	1.0689	1.0585	1.0650	1.0176	2.0122	2.2667	2.5169
	Proposed	<b>2.9880</b>	<b>2.9445</b>	<b>3.0260</b>	<b>3.2159</b>	<b>3.4261</b>	<b>3.6514</b>	<b>3.8392</b>
M109 or Tank	MMSE-Phase	2.3935	2.6221	2.7992	3.0517	3.3396	3.5614	3.7149
	KLT	1.1191	1.2645	1.5544	1.7355	2.2567	2.6292	2.9324
	PKLT	0.2427	0.2639	1.2399	1.6989	1.9391	2.2960	2.6749
	SPEC SUB	1.3193	1.5293	1.6361	2.0088	2.3070	2.5421	2.9918
	PSC	1.0432	1.5671	1.9427	2.2405	2.4796	2.7189	3.0327
	MMSE	1.5515	2.0061	2.0957	2.3288	2.5331	2.8158	3.1586
	Conjugate Sy.	0.9598	1.2590	1.9038	2.2003	2.4561	2.6840	2.9996
	Proposed	<b>2.9256</b>	<b>2.9556</b>	<b>3.0970</b>	<b>3.3529</b>	<b>3.5553</b>	<b>3.7170</b>	<b>3.8485</b>
Buccaneer or Jet cockpit	MMSE-Phase	1.3427	1.5198	1.5137	1.7438	2.0286	2.2436	2.3748
	KLT	1.5804	1.4979	1.3545	1.4836	1.8854	2.2775	2.5758
	PKLT	0.4302	0.7221	0.5649	1.0303	1.5401	1.7814	2.1788
	SPEC SUB	1.2275	1.2522	1.2282	1.5148	1.7529	2.0793	2.3909
	PSC	1.3383	1.4264	1.3301	1.3739	1.9478	2.2565	2.5172
	MMSE	1.5610	2.5710	2.6100	2.0100	2.0173	2.3047	2.5723
	Conjugate Sy.	1.4720	1.5519	1.4404	1.4623	1.9905	2.1931	2.3999
	Proposed	<b>2.9023</b>	<b>2.9103</b>	<b>2.9755</b>	<b>3.1527</b>	<b>3.2694</b>	<b>3.4286</b>	<b>3.5949</b>
Leopard or Military vehicle	MMSE-Phase	1.9425	1.9606	1.9030	2.0829	2.1940	2.3014	2.3840
	KLT	1.1809	1.3340	1.4920	1.9840	2.4359	2.8298	3.1533
	PKLT	0.2414	1.3075	1.0639	1.4822	2.0397	2.4686	2.8744
	SPEC SUB	1.4400	1.5734	1.8982	2.1804	2.4441	2.7325	3.2315
	PSC	1.7358	2.0645	2.3577	2.6116	2.8852	3.1841	3.3863
	MMSE	1.8631	2.1654	2.4530	2.7656	3.0043	3.2989	3.5926
	Conjugate Sy.	1.6592	1.9390	2.2476	2.5270	2.7846	3.0481	3.2257
	Proposed	<b>3.0517</b>	<b>3.1183</b>	<b>3.3168</b>	<b>3.6045</b>	<b>3.7758</b>	<b>3.9399</b>	<b>4.0637</b>

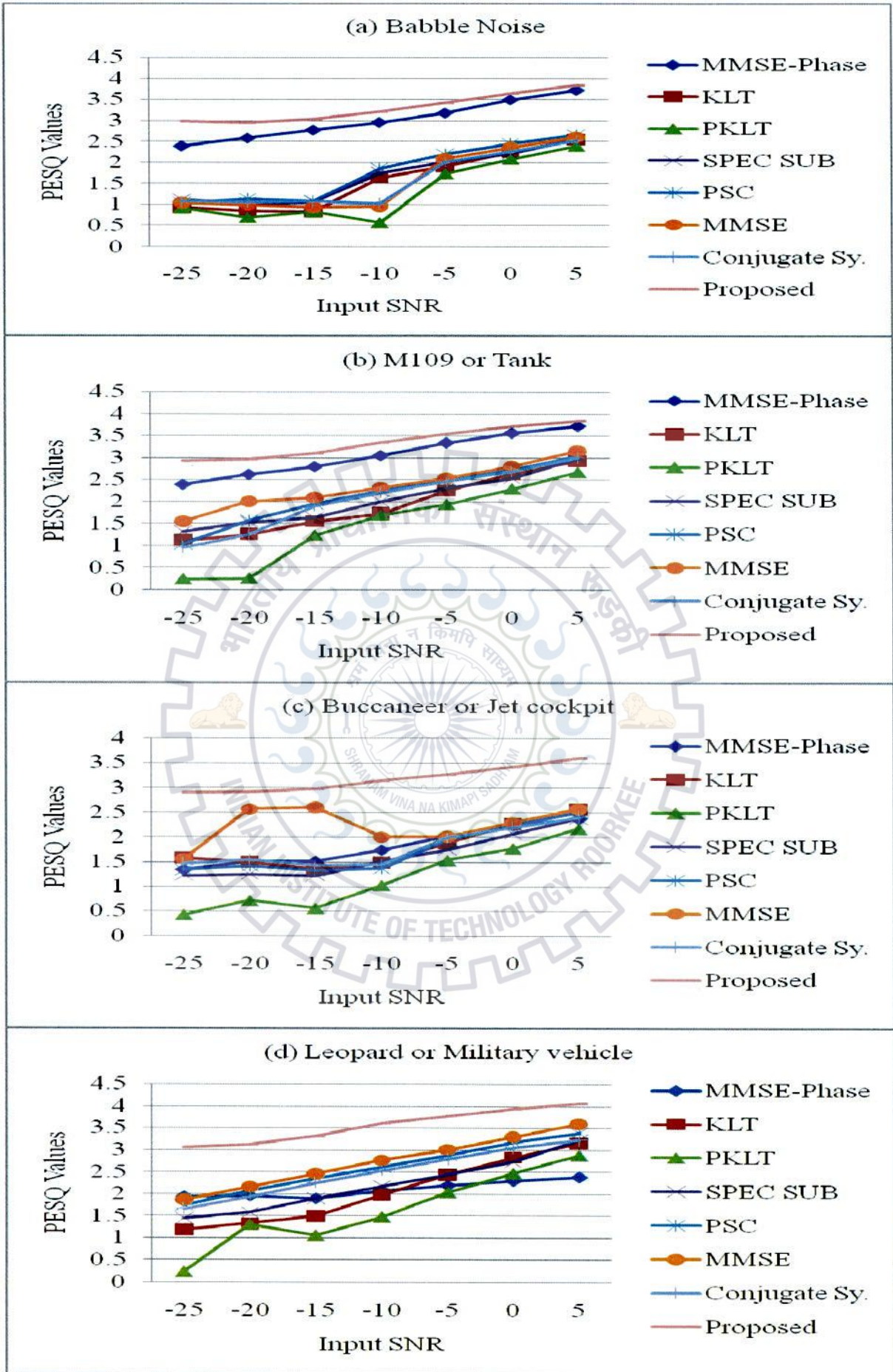


Fig. 5.2: Variation of PESQ scores with input SNR at different noise types.

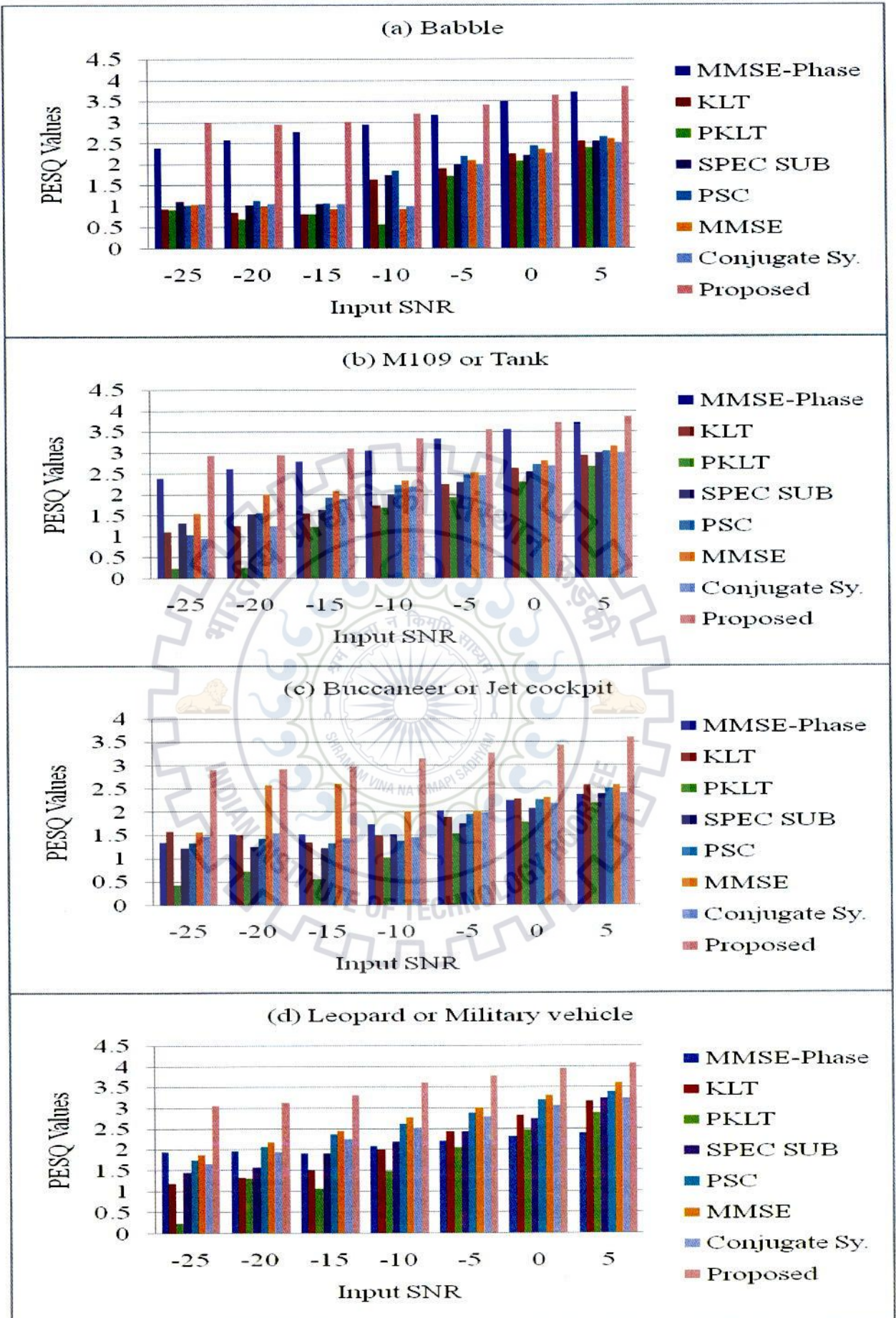


Fig. 5.3: Bar chart representation of PESQ scores at various input SNR levels.

Tables 5.2 to 5.5 present the results of performance parameters PESQ, fw-SSNR, WSS and OVL, respectively. Figures 5.2 to 5.9 present the graphical and bar chart representation of the values given in Tables 5.2 to 5.5, respectively.

The PESQ values are given in Tables 5.2. The MMSE-Phase based method outperforms the existing speech enhancement methods such as spectral subtraction, Conjugate Symmetry, MMSE, PSC, KLT and PKLT. The maximum PESQ values obtained by MMSE-Phase based method are 2.3824, 2.5737, 2.7683, 2.9503, 3.1763, 3.4931 and 3.7218 at -25, -20, -15, -10, -5, 0 and 5 dB, respectively. But here the results given by this method are not better than the proposed phase based speech enhancement method.

The maximum values of PESQ is obtained by the proposed method which are 3.0517, 3.1183, 3.3168, 3.6045, 3.7758, 3.9399, and 4.0637 at -25, -20, -15, -10, -5, 0 and 5 dB input SNR levels, respectively. Hence, the maximum improvement in speech is obtained by signal phase ratio based proposed method. The comparative graphical and bar chart representation for direct illustration of their performance is shown in Figures 5.2 and 5.3.

Table 5.3 illustrates the fw-SSNR values at various input SNR levels for different noise types. In case of babble noise, MMSE-Phase method outperforms spectral subtraction, Conjugate Symmetry, MMSE, PSC, KLT and PKLT. The KLT method gives the lowest performance at all input SNR levels and noise types. The maximum value by MMSE-Phase and KLT is 15.884 and 10.044 at 5dB input SNR, respectively for military vehicle noise. In comparison, the signal phase ratio based proposed method ( fw-SSNR 18.582) outperforms MMSE-Phase based method, spectral subtraction, Conjugate Symmetry, MMSE, PSC, KLT and PKLT methods (giving fw-SSNR as 15.884, 10.397, 10.505, 11.664, 9.8543, 10.044 and 8.7544, respectively).

Thus, the signal phase ratio based proposed method produces maximum speech enhancement. The graphical and bar chart representation of these values are illustrated in Figures 5.4 and 5.5 for direct interpretation of the above observation.



Table 5.2: Performance Measure in terms of fw-SSNR.

Noise Type	Enhancement Techniques	Frequency Weighted SSNR						
		Noise Levels	-25	-20	-15	-10	-5	0
Babble	MMSE-Phase	3.5273	4.7565	6.1965	8.1271	9.6345	11.305	13.189
	KLT	0.3222	0.5855	0.8925	1.6416	2.7685	4.8097	7.1844
	PKLT	0.5954	0.7380	0.9482	1.7446	2.8360	4.9762	6.9008
	SPEC SUB	1.4591	1.7560	2.1691	2.6552	3.6928	5.0067	8.2673
	PSC	1.6212	1.9657	2.5086	3.2495	4.0552	5.2823	7.1401
	MMSE	1.1695	1.5165	1.9593	2.6409	3.9313	5.6793	7.5121
	Conjugate Sy.	0.9833	1.2214	1.7036	2.4445	3.5514	5.1996	7.2665
	Proposed	<b>9.9469</b>	<b>10.846</b>	<b>11.762</b>	<b>12.903</b>	<b>14.104</b>	<b>15.333</b>	<b>16.748</b>
M109 or Tank	MMSE-Phase	3.7851	4.9114	6.2514	7.9152	9.4426	11.334	13.512
	KLT	1.9202	2.1493	2.5165	3.0599	4.7705	6.8869	9.0349
	PKLT	1.5156	1.7187	2.0165	2.5880	4.0885	5.7733	8.0204
	SPEC SUB	2.8886	3.1926	3.5460	4.1088	5.5284	6.7455	9.9780
	PSC	3.1657	3.4685	4.0035	4.7791	6.0122	7.7986	9.8633
	MMSE	3.4339	3.7248	4.1888	5.0660	6.4864	8.4310	10.513
	Conjugate Sy.	3.4323	3.6751	4.1784	4.7766	5.9948	7.4724	9.5593
	Proposed	<b>9.5644</b>	<b>10.552</b>	<b>11.610</b>	<b>12.646</b>	<b>13.604</b>	<b>14.919</b>	<b>16.318</b>
Buccaneer or Jet cockpit	MMSE-Phase	3.1091	4.1551	5.0649	6.3712	7.6279	9.1008	10.829
	KLT	0.9752	1.1059	1.4889	2.0083	3.2265	4.7707	6.5053
	PKLT	1.5671	1.7012	1.9382	2.3724	2.9859	4.2075	5.3372
	SPEC SUB	1.3405	1.4366	1.6210	2.1611	2.9158	4.2932	5.9777
	PSC	2.0987	2.1626	2.3398	2.8479	3.8510	5.1390	6.8470
	MMSE	1.4906	1.5767	1.8769	2.4747	3.5319	5.3996	6.9037
	Conjugate Sy.	1.3322	1.3658	1.4674	1.8102	2.5568	3.7894	5.2167
	Proposed	<b>9.9394</b>	<b>10.702</b>	<b>11.651</b>	<b>12.391</b>	<b>13.102</b>	<b>13.918</b>	<b>15.086</b>
Leopard or Military vehicle	MMSE-Phase	5.0510	6.6879	8.0063	9.9031	11.754	13.678	15.884
	KLT	2.0546	2.3855	2.7340	3.9979	5.3810	7.8376	10.044
	PKLT	0.0945	1.2357	1.6055	2.5475	4.1049	6.4193	8.7544
	SPEC SUB	3.2312	3.6135	3.9685	4.4602	5.4305	7.2698	10.397
	PSC	3.1980	3.9762	4.8725	5.8203	7.0390	8.4264	9.8543
	MMSE	3.8835	4.5870	5.2133	6.4568	7.9655	9.7882	11.664
	Conjugate Sy.	3.2498	3.8781	4.7533	5.9420	7.2585	8.9965	10.505
	Proposed	<b>10.128</b>	<b>11.197</b>	<b>12.469</b>	<b>13.944</b>	<b>15.368</b>	<b>16.751</b>	<b>18.582</b>

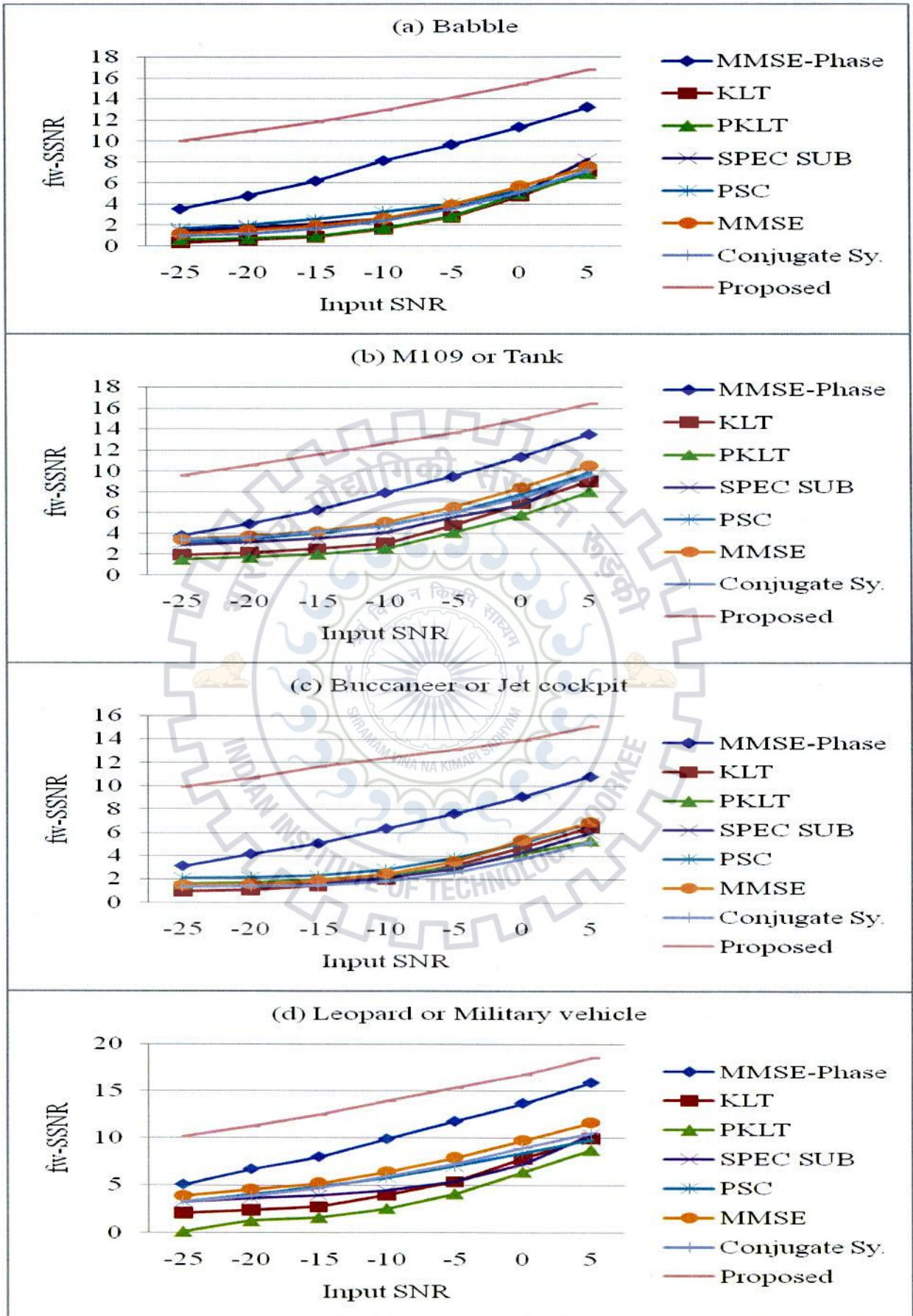


Fig. 5.4: Variation of fw-SSNR scores with input SNR at different noise types.

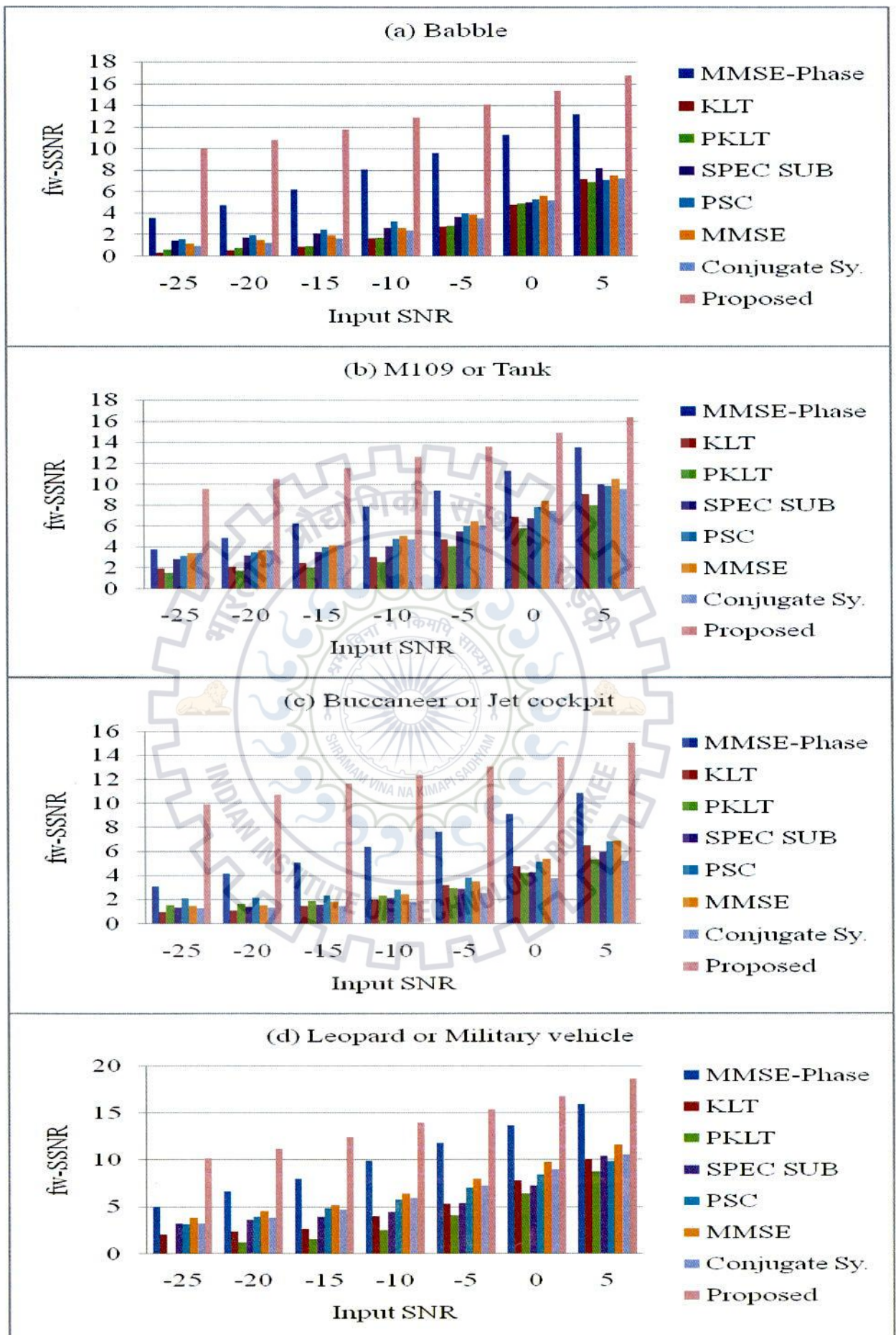


Fig. 5.5: Bar chart representation of fw-SSNR scores at various input SNR levels.

The variation of WSS with input SNR levels is illustrated in Table 5.4. The graphical and bar chart representation for direct illustration of their performance is shown in Figures 5.6 and 5.7. The MMSE-Phase based method outperforms to existing speech enhancement methods such as spectral subtraction, Conjugate Symmetry, MMSE, PSC, KLT and PKLT methods. The lowest performance is given by spectral subtraction method. The other existing methods also give poor performance in comparison to the proposed method. The minimum value of WSS is 14.4586 at 5 dB input SNR level for proposed method which results in maximum speech enhancement. At all input SNR levels, the minimum values of WSS are obtained by proposed method which indicates the maximum suppression of noise and improvement in the speech quality.

Table 5.3: Performance Measure in terms of WSS.

Noise Type	Enhancement Techniques	WSS						
		Noise Levels → -25	-20	-15	-10	-5	0	5
Babble	MMSE-Phase	67.9852	62.0664	54.7294	46.7552	40.0040	34.5434	29.5462
	KLT	121.3283	117.4654	114.4330	102.7649	89.0387	72.7817	62.6379
	PKLT	97.4068	96.2676	93.6595	84.8683	72.9789	59.8312	50.4380
	SPEC SUB	129.3368	125.6523	119.0640	110.2210	98.5172	87.2732	69.3789
	PSC	120.0904	115.4200	109.0679	100.2740	88.5606	75.3503	63.9619
	MMSE	127.0232	123.8617	116.4540	106.3762	92.6271	81.0146	71.4111
	Conjugate Sy.	93.3738	91.9290	89.1098	82.6612	72.3165	62.4125	52.5151
	Proposed	<b>54.4779</b>	<b>50.1016</b>	<b>44.8156</b>	<b>37.1932</b>	<b>31.8529</b>	<b>24.6646</b>	<b>19.6877</b>
M109 or Tank	MMSE-Phase	66.7720	61.9093	58.0951	47.2683	42.0446	37.5644	32.3064
	KLT	137.6189	137.3932	132.4593	119.3720	99.4317	79.1611	64.7961
	PKLT	100.2506	98.7130	92.7143	79.7230	65.1467	52.6999	41.6770
	SPEC SUB	133.2038	130.5794	121.0903	110.8230	93.5196	79.8651	63.4469
	PSC	121.4662	117.3131	109.6984	98.2452	83.6718	68.4479	55.6452
	MMSE	110.2545	105.4071	97.7061	86.6748	75.6126	62.7963	50.4232
	Conjugate Sy.	91.6034	89.2146	83.9034	75.4857	64.5922	53.8724	42.9048
	Proposed	<b>64.73335</b>	<b>57.3459</b>	<b>49.6683</b>	<b>41.2609</b>	<b>35.2530</b>	<b>28.1324</b>	<b>21.8501</b>
Buccaneer or Jet cockpit	MMSE-Phase	62.2532	58.0021	57.0966	51.6070	45.5851	40.1528	36.3499
	KLT	171.0636	167.4567	161.1343	150.6597	130.788	111.217	91.5862
	PKLT	106.6629	103.5043	100.1480	89.2949	76.5075	63.3130	54.6485
	SPEC SUB	140.8174	138.9545	135.5267	127.4261	116.384	102.857	89.5121
	PSC	137.9379	136.6822	133.2392	125.6793	114.115	101.478	89.1987
	MMSE	110.7734	110.4905	108.1185	100.3574	90.4451	79.2963	69.8801
	Conjugate Sy.	91.5691	90.8474	89.0236	84.6315	76.2368	67.4235	58.7509
	Proposed	<b>55.2276</b>	<b>52.1525</b>	<b>49.3513</b>	<b>43.9642</b>	<b>38.1284</b>	<b>32.2926</b>	<b>26.1241</b>
Leopard or Military vehicle	MMSE-Phase	61.3296	55.8793	45.3908	36.7490	28.8412	23.1471	19.2922
	KLT	95.7167	87.9608	83.7462	76.9802	68.2710	58.4352	45.0863
	PKLT	80.4575	79.2520	76.7726	70.3466	61.8777	47.5845	36.2147
	SPEC SUB	98.1188	93.9264	87.5414	82.4838	73.0006	60.3028	45.7017
	PSC	103.6240	97.2185	89.2340	79.3934	67.4311	55.1597	45.7826
	MMSE	82.4947	77.0706	69.4333	60.4101	51.1064	41.4909	34.9684
	Conjugate Sy.	74.9725	70.8043	66.3962	58.5295	49.2674	40.1795	32.6572
	Proposed	<b>54.8275</b>	<b>48.7972</b>	<b>40.5751</b>	<b>29.8282</b>	<b>22.5407</b>	<b>18.3344</b>	<b>14.4586</b>

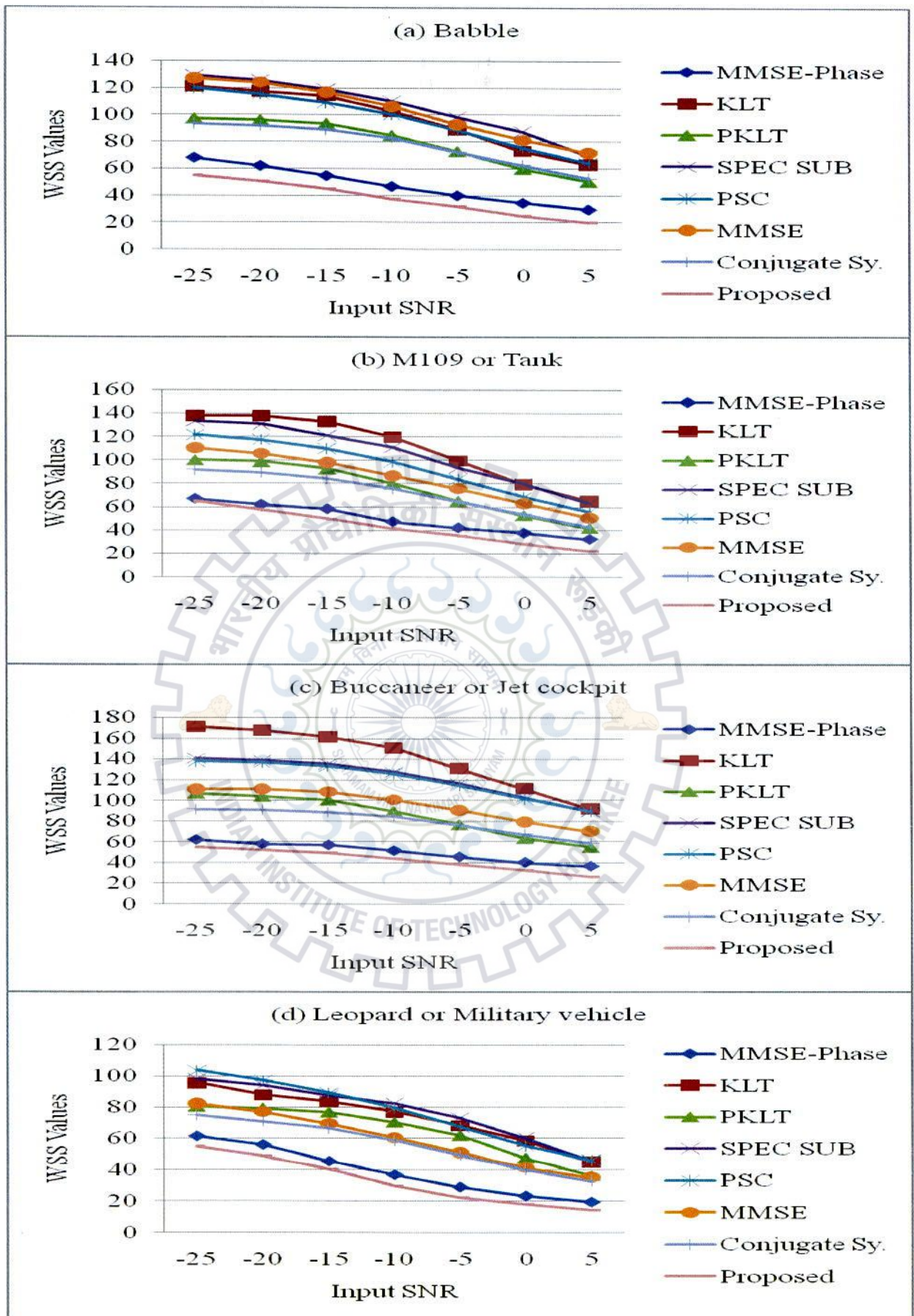


Fig. 5.6: Variation of WSS scores with input SNR at different noise.

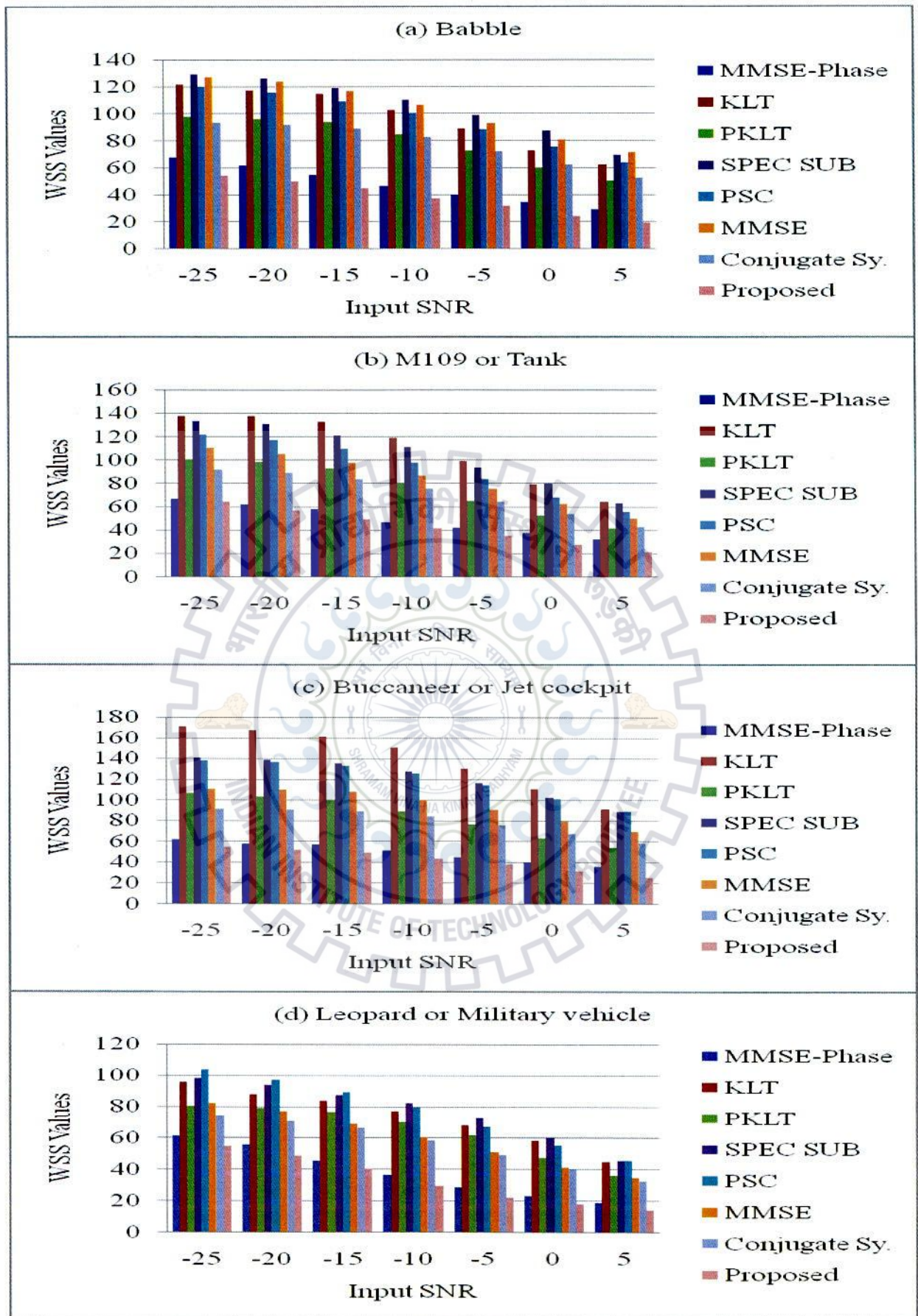


Fig. 5.7: Bar chart representation of WSS scores at various input SNR levels.

The overall performance of the existing speech enhancement methods is compared in Table 5.5 at different input SNR levels for various noise types. The comparative graphical and bar chart representation for direct illustration of their performance is shown in Figures 5.8 and 5.9, respectively. The MMSE-Phase based method shows maximum speech enhancement than spectral subtraction, Conjugate Symmetry, MMSE, PSC, KLT and PKLT methods but do not give better enhancement than the proposed signal phase ratio based method. The PKLT method obtains the lowest speech enhancement for babble, M109 and leopard noise types. The conjugate symmetry method gives better results in buccaneer noise. The proposed speech enhancement method gives maximum values of OVL score at all input SNR levels and for all noise types.

Table 5.4: Performance Measure in terms of OVL.

Noise Type	Enhancement Techniques	OVL						
		Noise Levels →	-25	-20	-15	-10	-5	0
Babble	MMSE-Phase	1.4432	1.7346	2.0731	2.5876	2.9959	3.4178	3.8830
	KLT	1.0000	1.0000	1.0000	1.0006	1.3420	1.8510	2.2864
	PKLT	1.0000	1.0000	1.0000	1.0000	1.2332	1.6715	2.0604
	SPEC SUB	1.0305	1.0664	1.1231	1.2775	1.5246	1.8327	2.5094
	PSC	1.0414	1.1016	1.2349	1.4287	1.6376	1.9658	2.2448
	MMSE	1.0000	1.0000	1.0062	1.1605	1.4205	1.7669	2.1371
	Conjugate Sy.	1.0000	1.0000	1.0032	1.1679	1.3722	1.7593	2.2368
	Proposed	<b>3.0773</b>	<b>3.2788</b>	<b>3.5288</b>	<b>3.8333</b>	<b>4.1006</b>	<b>4.3684</b>	<b>4.7272</b>
M109 or Tank	MMSE-Phase	1.4944	1.6896	2.0995	2.5508	2.9813	3.4018	3.8360
	KLT	1.2616	1.2436	1.3218	1.5047	1.9613	2.4210	3.0175
	PKLT	1.1842	1.1736	1.2442	1.3528	1.6724	2.0870	2.5293
	SPEC SUB	1.3106	1.4147	1.4598	1.5851	1.9213	2.1991	3.0296
	PSC	1.5934	1.6653	1.8026	1.9606	2.2870	2.7102	3.1137
	MMSE	1.5027	1.5776	1.6123	1.8130	2.1592	2.6188	3.0820
	Conjugate Sy.	1.7125	1.7720	1.8718	1.9958	2.2660	2.6071	3.0945
	Proposed	<b>2.9577</b>	<b>3.2717</b>	<b>3.5787</b>	<b>3.8685</b>	<b>4.0715</b>	<b>4.3219</b>	<b>4.7228</b>
Buccaneer or Jet cockpit	MMSE-Phase	1.1097	1.4125	1.6203	1.9925	2.3257	2.7026	3.0380
	KLT	1.0000	1.0000	1.0485	1.0798	1.3424	1.6314	1.9630
	PKLT	1.1546	1.1600	1.1879	1.2173	1.2784	1.4422	1.6738
	SPEC SUB	1.0000	1.0000	1.0000	1.0442	1.1480	1.4993	1.8742
	PSC	1.2689	1.2712	1.2970	1.4123	1.5930	1.7811	2.0683
	MMSE	1.0000	1.0182	1.0956	1.1978	1.3786	1.7037	2.0896
	Conjugate Sy.	1.0000	1.0000	1.0000	1.0000	1.1138	1.3759	1.6008
	Proposed	<b>3.1630</b>	<b>3.2842</b>	<b>3.5323</b>	<b>3.7790</b>	<b>3.8517</b>	<b>4.0150</b>	<b>4.2413</b>
Leopard or Military vehicle	MMSE-Phase	1.9481	2.3358	2.6000	3.0592	3.5233	4.0600	4.6420
	KLT	1.1370	1.1968	1.2617	1.6613	2.0477	2.6696	3.2904
	PKLT	1.0000	1.0000	1.0868	1.2742	1.6136	2.1312	2.6941
	SPEC SUB	1.5636	1.6655	1.7347	1.8468	2.0170	2.5574	3.2884
	PSC	1.6481	1.8266	2.0915	2.3646	2.6850	3.0254	3.1658
	MMSE	1.7530	1.8953	2.0483	2.3932	2.7493	3.0929	3.4642
	Conjugate Sy.	1.6736	1.7842	1.9876	2.2589	2.6146	2.9464	3.1730
	Proposed	<b>3.0102</b>	<b>3.3067</b>	<b>3.5807</b>	<b>4.0672</b>	<b>4.4429</b>	<b>4.7340</b>	<b>5.0000</b>

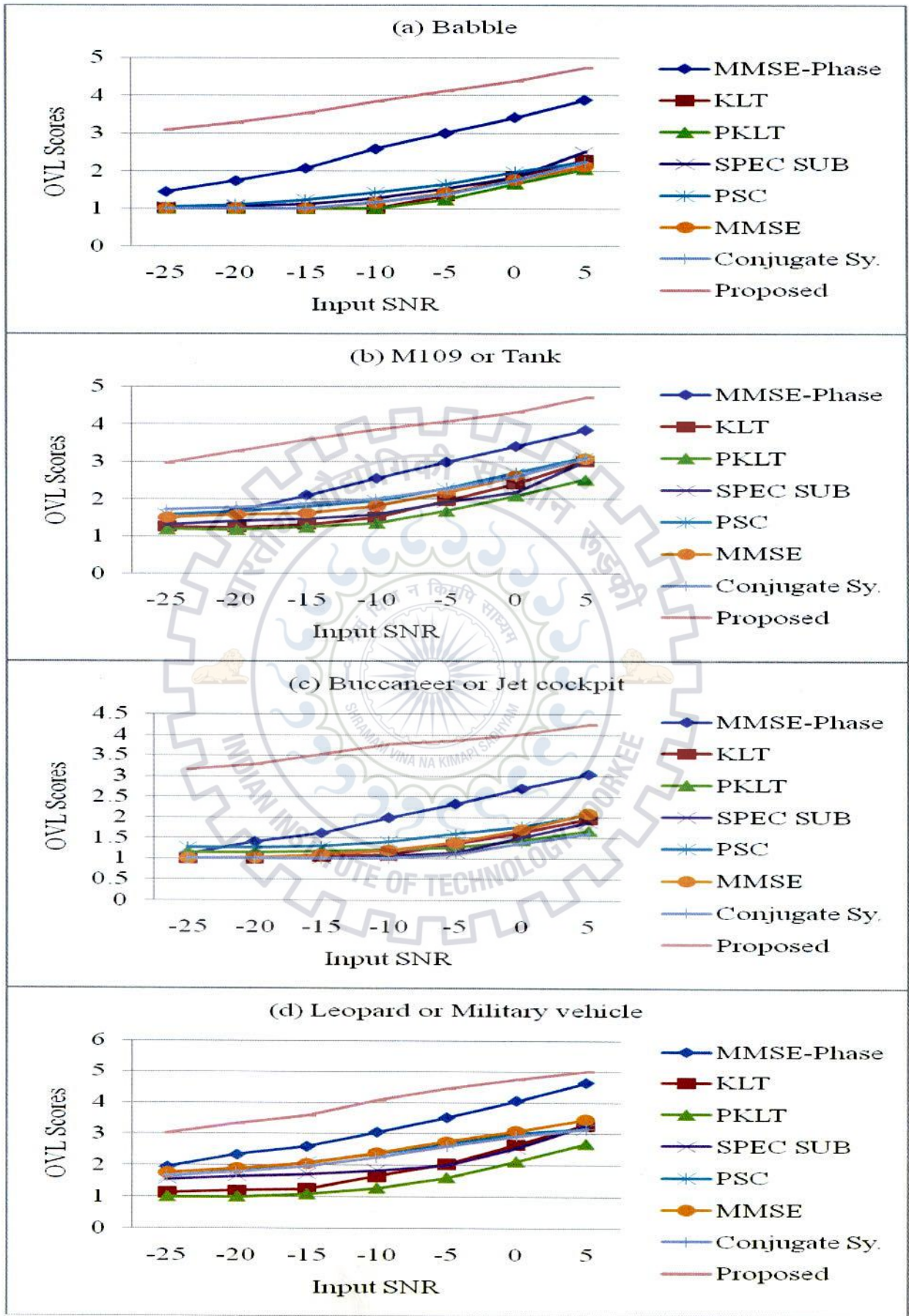


Fig. 5.8: Variation of OVL scores with input SNR at different noise.



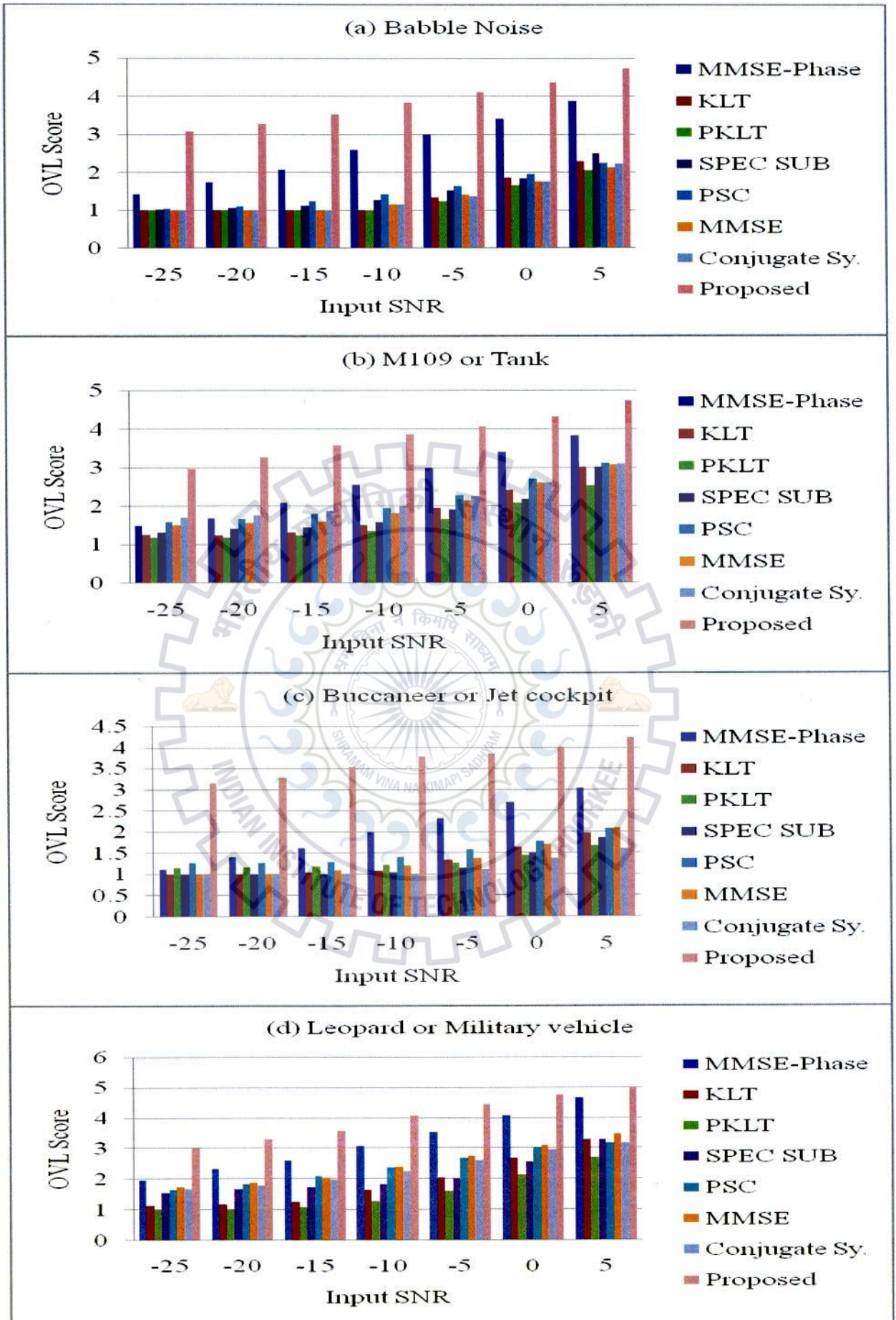
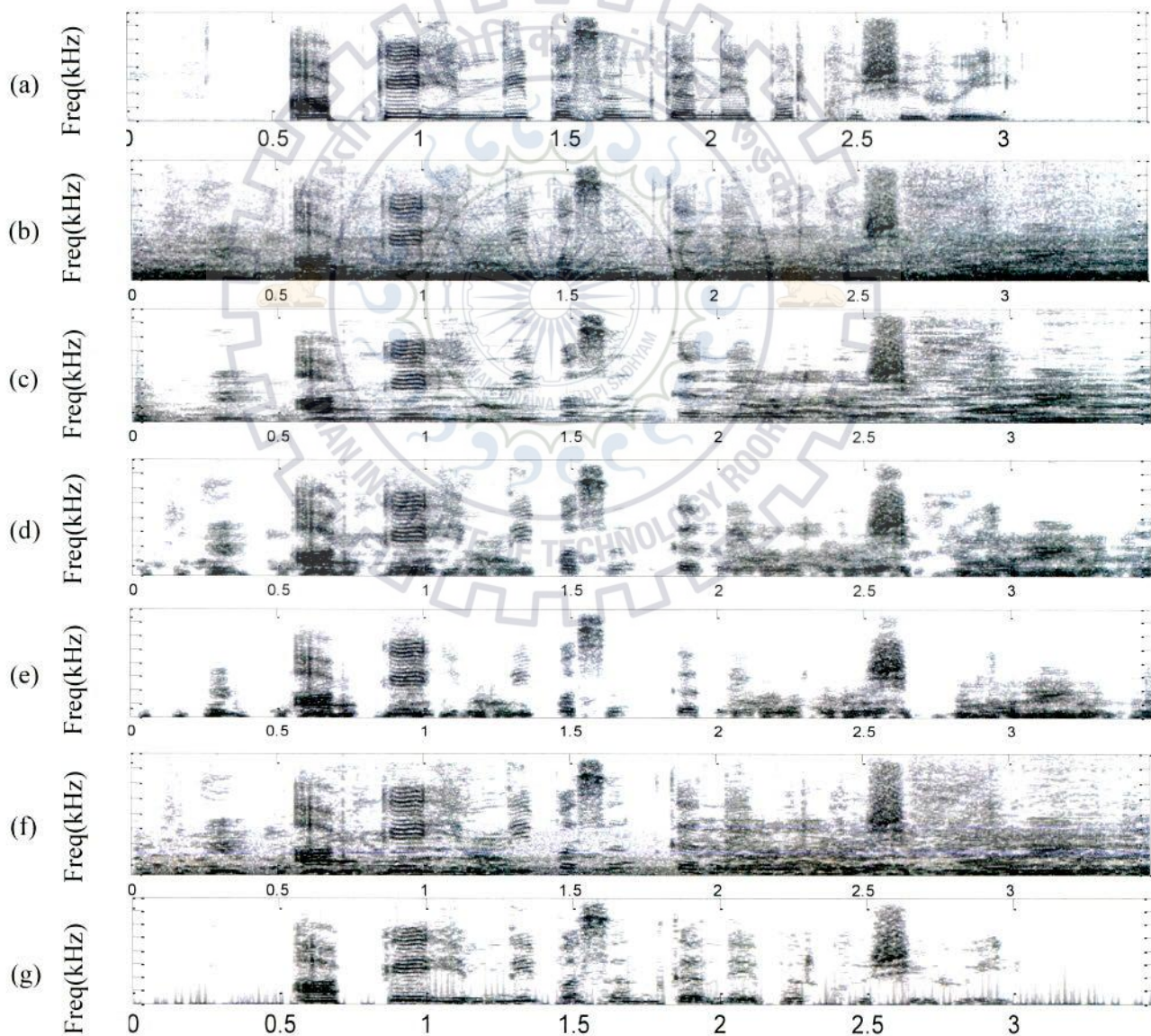


Fig. 5.9: Bar chart representation of OVL scores at various input SNR levels.

Spectrogram of the output speech signals are presented in Figure 5.10. These are obtained at -5 dB SNR level in presence of babble noise. In Figure 5.10, Spectrograms are (a) clean speech, (b) noisy speech, and from (c) – (j) denoised speech signals obtained by speech enhancement methods such as (c) for MMSE-Phase, (d) for KLT, (e) for PKLT, (f) for spectral subtraction, (g) for PSC, (h) for MMSE, (i) for Conjugate Symmetry, (j) for proposed method, respectively. An observation in Figure 5.10 illustrates that, the proposed method does not give more distortions and impulses in output spectrogram, whereas all other methods do so. This indicates that the proposed method effectively reduces the background noise from the noisy input speech signal. Hence, the maximum speech quality and intelligibility is achieved by proposed method than those obtained by existing phase based speech enhancement methods.



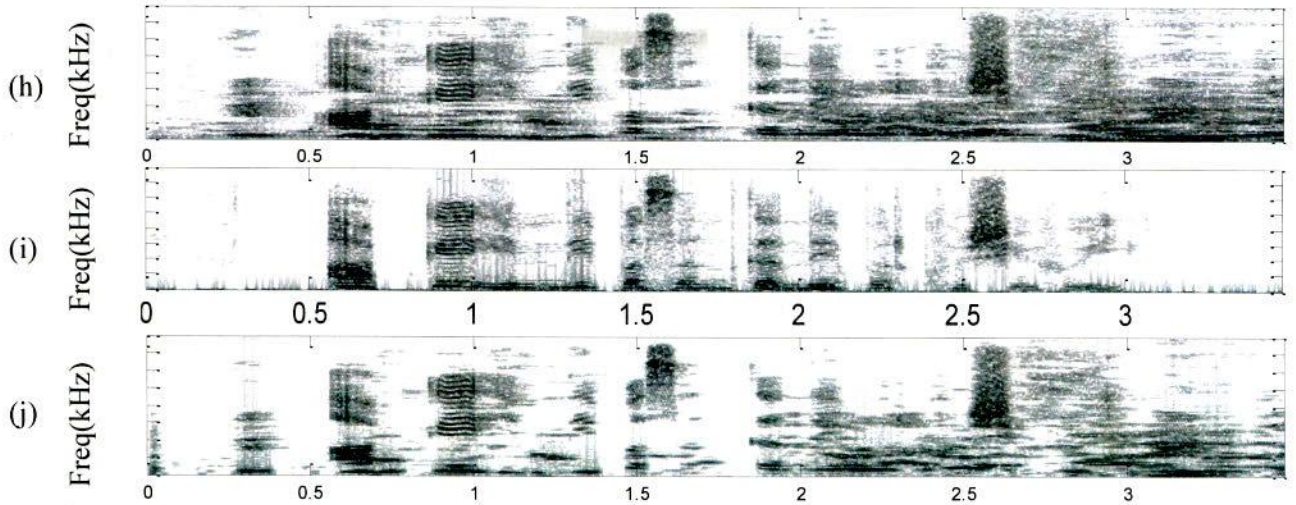


Fig. 5.10: Spectrograms of (a) clean speech, (b) noisy speech (babble noise at -5 dB input SNR), and enhanced output signal given by speech enhancement methods such as (c) MMSE-Phase, (d) KLT, (e) PKLT, (f) spectral subtraction, (g) PSC, (h) MMSE, (i) Conjugate Symmetry, (j) proposed method.

#### 5.4 Summary

Signal phase ratio based speech enhancement method is proposed for removing noise from noisy speech signal and improving in speech intelligibility. In this phase based single-channel speech enhancement method, two gain functions  $G_1$  and  $G_2$  are calculated for noise suppression. The performance of proposed method is compared with existing phase based speech enhancement methods. The comparative results indicate the better performance of proposed phase based speech enhancement method.

*This chapter describes conclusions and future scope of the research work. The major contributions of present thesis are summarized in this chapter while few contributions for future research are listed in section of scope for future work.*

### 6.1 Conclusions

In this thesis, an attempt has been made to design and evaluate the new speech enhancement algorithms applicable for single-channel speech. These algorithms have been focused for getting speech of good quality and intelligibility. With above consideration, the following issues of single-channel speech enhancement have been addressed in this thesis.

- Enhancement of mixed noisy speech of very low (Negative) SNR.
- Enhancement of speech in non-stationary noise case.
- Suppression of combined effect of background noise and reverberations.
- Enhancement of speech considering the phase effect.

In this thesis, different methods have been designed and implemented to solve the speech enhancement problem for English and major Indian languages considering many noise types. The performance of these developed methods is compared with existing speech enhancement methods. Various quality and intelligibility measure parameters are considered for the evaluation of these methods.

In this research work, four different methods have been proposed which are WPT based modified Wiener gain method, WPT based Fuzzy mask method, reverberant mask based method and signal phase ratio based speech enhancement method. On the basis of the obtained results conclusions drawn are summarized in following subsections.

#### 6.1.1 WPT based modified Wiener gain method

Various methods like Wiener, spectral subtraction, KLT, PKLT and MMSE based algorithms have been reported for enhancement of noisy speech having single noise in the case of single-channel speech. But no work has been reported in speech enhancement for mixed noise case. To consider the mixed noise case for single-channel speech enhancement, WPT based modified Wiener gain based method has been proposed here. The performance of modified gain function based method is compared with other existing gain function based methods. The performance is evaluated in terms of quality and intelligibility parameters such as SII, CD,

PESQ, fw-SSNR and SNR. The results obtained show that the proposed method is effective for all noise types and for varied SNR levels for English and Indian languages like Hindi, Kannada, Bengali and Malayalam.

The proposed WPT based modified Wiener gain based method give 7.4173 dB output SNR at 5 dB input SNR in (f16+babble+NOIZEUS speech+ Hindi speech) mixed noise signals while other existing speech enhancement methods gave output SNR in the range from 0.9135 dB to 6.0987 dB. The maximum improvement in PESQ parameter is 3.7283 at 5 dB SNR for the proposed method while other methods give PESQ values in the range of 1.8887 to 3.4010. The proposed method shows the best performance in other speech patterns having mixed noise cases (such as machinegun+ pink+ NOIZEUS speech+ Hindi and factory floor+ white+ NOIZEUS+ Hindi speech) also. The other existing speech enhancement methods show poor performance and which result in lower speech quality and intelligibility.

### 6.1.2 WPT based Fuzzy mask method

A WPT Fuzzy mask based method has been proposed here for suppression of highly non-stationary noise in noisy speech of low SNR. This fuzzy mask replaces the need of true speech and/or information about noise signal for determination of WP soft and hard threshold. The noisy dataset of SNR ranging from -15 dB to 15 dB are generated for comparative analysis. In this approach, the noisy input speech dataset are applied to WPT for decomposing the input noisy speech signal into eight energy bands which are later converted into signals of different energy bands. A modified gain function is applied for denoising of these separate bands of speech signals. The output speech is then given to fuzzy mask function for suppression of gain induced speech distortions. The performance parameters such as output SNR, PESQ, MOS, and STOI for the noise suppressed speech signal are evaluated to compare the performance of this method with the other speech enhancement methods.

The above given parameters are evaluated for babble, pink, f-16 and white noise case. The performance of the existing speech enhancement methods increases with increasing input SNR levels except for MMSE-SPU speech enhancement method, where the performance decreases with increasing input SNR levels from -15 dB to 15 dB (i.e. -0.9161 dB and -7.1942 dB output SNR at -15 dB and 15 dB, respectively). The fuzzy mask and spectral subtraction methods give better performance among other existing speech enhancement methods such as p-MMSE, log-MMSE and Wiener but the performance of proposed WPT Fuzzy mask method is much better than other existing speech enhancement methods and is increasing with increasing input SNR levels. The output SNR values given by proposed methods are 1.2876

dB and 21.1661 dB at -15 dB and 15 dB input SNR, respectively. In case of proposed method, the maximum values of STOI, PESQ and MOS parameters at 15 dB input are given as 0.9472, 3.5268 and 0.7182, respectively. The all evaluated performance measure parameters of the processed speech signals (also in other noise cases such as pink, f-16 and white) indicate an almost noise free speech signal having been generated by proposed method. The WPT Fuzzy mask based proposed speech enhancement method gives maximum quality and intelligibility improvement when compared to other speech enhancement methods at all levels of input SNRs.

### 6.1.3 Reverberant mask based method

Here, a reverberant mask based speech enhancement method is proposed for combined suppression of reverberation and noise. In this method signal-to-reverberant ratio (SRR) is calculated as a limit. The amplitudes with SRR greater than a preset threshold (i.e. -5dB) are used for reconstruction of denoised speech, while amplitudes with SRR values smaller than the threshold are eliminated. The SRR reflects implicitly the ratio of the energies of the signal originating from the early (and direct) reflections and the signal originating from the late reflections. The construction of the SRR criterion assumes *a priori* knowledge of the input reverberant and target signal. Threshold values varying from 0dB to -90dB are analyzed for selection of ideal reverberant mask (IRM) limit  $T$ . Enhanced speech signal is constructed by multiplying noisy speech with reverberant mask.

For the performance analysis, the babble, pop music, restaurant and exhibition noise are added individually with the reverberated speech. The combined reverberated speech with noise is used for testing the effectiveness of the speech enhancement methods. The proposed reverberant mask based speech enhancement method is compared with other existing speech enhancement methods in terms of speech quality and intelligibility measure parameters such as PESQ, CD, SNR and MSE.

Among the existing speech enhancement methods, Wiener gives the maximum improvement and p-MMSE gives minimum improvement in all reverberated noisy speech conditions. The proposed speech enhancement method performs even better than Wiener method for combined suppression of reverberation and noise. The proposed method gives PESQ score of 3.0659 in comparison to 2.4156 by Wiener method at -5 dB input SNR in reverberation speech with pop music signal. The minimum values of CD and MSE parameters are obtained by proposed method and hence maximum improvement is obtained in speech quality and intelligibility at all noisy speech conditions.

#### 6.1.4 Signal phase ratio based method

In the last part of the work, another speech enhancement method, which uses signal phase ratio in place of the amplitude of input speech signal, has been proposed for suppression of noise from low and high SNR noisy speech signals. In this method, phase ratio is calculated from noise and noisy speech. The values of all constants are determined in such a way that it maximizes the speech intelligibility. The two gain functions  $G_1$  and  $G_2$  are calculated for suppressing noise derived from angles  $0$  to  $\pm\pi/2$  and  $\pm\pi/2$  to  $\pm\pi$ , respectively. The calculation of  $G_1$  and  $G_2$  parameters used in this method is based on phase information of speech signal which does not require signal amplitude which is used in other existing speech enhancement method.

The performance of this method is compared with other existing speech enhancement methods such as phase spectrum compensation (PSC), exploiting conjugate symmetry of the short-time Fourier spectrum and STFT-phase for the MMSE-optimal spectral amplitude estimation KLT, PKLT etc. The performance is measured in terms of quality and intelligibility measure parameters namely, the PESQ, fw-SSNR, WSS and OVL. The babble, tank, buccaneer and leopard noise are added to speech signals to generate the noisy speech patterns having SNR levels between -25dB to 5dB.

Among the existing speech enhancement methods MMSE phase based method gives the maximum improvement while PKLT method gives minimum improvement for all input SNR levels. The MMSE phase based method gives maximum PESQ score of 2.3824 in comparison to 0.9204 by PKLT method at -25 dB input SNR. The maximum value of PESQ given by proposed method is 2.9880 at -25 dB input SNR which is better than Wiener method. The all other evaluated performance measure parameters (such as fw-SSNR, WSS OVL) of the processed speech signals (in other noise cases such as M109, Buccaneer and leopard) also indicate an almost noise free speech signal having been generated by proposed method. It is analyzed here that the phase is also as important as is amplitude for speech enhancement. The proposed method shows better performance than existing speech enhancement methods for all noise types and input noisy speech SNR levels.

By using the Wiener gain method to enhance the speech quality and intelligibility parameters for a mixed noise of low SNR speech signals selection of threshold and gain is very crucial. It required prior knowledge of SNR range and type of noises present in speech signal. Therefore the heuristic knowledge is one of the key factor in designing such algorithm. The same challenge was attempt by an intelligent method called WPT Fuzzy which is better as compared to modified Wiener gain method and used for suppression of highly non-

stationary noise sources of low SNR input noisy speech. Another challenge for improvement of speech quality and intelligibility is reverberation as a noise which was suppressed by ideal reverberant mask based method. The main hurdle to handle this challenge is threshold selection which is taken as a trade of value for this method. Signal phase ratio based method is also used for noisy speech enhancement which provides two gain functions for noise suppression. This method use phases of noisy speech and noise for gain calculation. It shows better results as compare to the other reported methods which are using amplitude. All these proposed algorithms are effective for any type of noise source and language.

In nutshell, on the basis of experimental results, it is concluded that better performance is exhibited by above proposed methods and they significantly contribute towards the state of the art in single-channel speech enhancement.

## 6.2 Scope for Future Work

The current research work has also an open space to be carried out by the future researchers. Few of them are suggested as below:

This research work has been focused on low level input SNR, reverberations, highly non-stationary noise and their various combinations in case of single-channel speech applications. Based on the achievements here, the suggestions for future work are given as below:

- For WPT-fuzzy mask based single-channel speech enhancement method, the fuzzy mask parameters used in this method can be calculated using intelligente approaches (such as ANN, DNN etc.) in place of experimental approach.
- In case of combined suppression of noise and reverberation from reverberated noisy speech the intelligent approaches may also be used for selection of adaptive threshold.
- Further improvement in quality with intelligibility is possible by adding additional features such as magnitude and phase components of LP residual signal separately for capturing the language-specific phonotactic information present in excitation source.



- In phase based speech enhancement method, only phase of the noise and noisy speech are used for gain calculation. Results can further be improved by considering the magnitude also for gain calculations in addition to phase.



## PUBLICATIONS FROM THE WORK

---

### Journals

- [1] Sachin Singh, Manoj Tripathy, R. S. Anand, "Subjective and objective analysis of speech enhancement algorithms for single channel speech patterns of indian & english languages". *IETE Journal of Technical Review*, **Taylor & Francis**, vol. 31, issue 1, pp. 34-46, Jan. 2014.
- [2] Sachin Singh, Manoj Tripathy, R. S. Anand, "Suppression of combined effect of late reverberation and masking noise for speech enhancement using channel selection method", *International Journal of Signal and Imaging Systems Engineering*, **Inderscience**, 2014. (Published online)
- [3] Sachin Singh, Manoj Tripathy, R. S. Anand, "A wavelet based method for removal of highly non-stationary noise from single-channel hindi speech patterns of low input SNR", *International Journal of Speech Technology*, **Springer**, , vol. 18, Issue 2, pp. 157-166, 2015.
- [4] Sachin Singh, Manoj Tripathy, R. S. Anand, "Binary mask based method for enhancement of mixed noise speech of low snr input", *International Journal of Speech Technology*, **Springer**, August, 2015.  
DOI 10.1007/s10772-015-9305-5 (Published online)
- [5] Sachin Singh, Manoj Tripathy, R. S. Anand, "Investigation of speech enhancement algorithms for intelligibility improvement", *IETE Journal of Technical Review*, **Taylor & Francis**, 2015. (First Revision Submitted)
- [6] Sachin Singh, Manoj Tripathy, R. S. Anand, "Speech intelligibility improvement using single-channel noise-reduction algorithms for Indian languages", *International Journal of Sadhana Acedemic*, **Springer**, 2015. (Under Review)
- [7] Sachin Singh, Manoj Tripathy, R. S. Anand, "STFT phase based single-channel speech enhancement using phase ratio", *Journal of acoustical society of America*, 2014. (Under Review)
- [8] Sachin Singh, Manoj Tripathy, R. S. Anand, "A fuzzy mask and modified gain function based single-channel speech enhancement using hard and soft wavelet packet threshold", *Journal of Speech Communication*, **Elsevier**, 2015. (Under Review)

### Book Chapter

- [9] Sachin Singh, Manoj Tripathy, R. S. Anand, "Single channel speech enhancement for mixed non-stationary noise environments", *Advances in Signal Processing and Intelligent Recognition Systems*, (LNCS Series) **Springer**, vol. 264, pp 545-555, 2014.

### **International conferences**

- [10] Sachin Singh, Manoj Tripathy and R. S. Anand, "Performance analysis of speech enhancement techniques on single channel speech patterns of Hindi language", in International Conference **SAP-BEATS**, Rajasthan India, January, 2013.
- [11] Sachin Singh, Manoj Tripathy and R. S. Anand, "Noise removal in single channel Hindi speech patterns by using binary mask thresholding function in various mother wavelets," in IEEE International Conference on Signal Processing, Computing and Control (**ISPCC**), Shimla, India, 26-28 Sept. 2013.
- [12] Sachin Singh, Manoj Tripathy and R. S. Anand, "A fuzzy mask based on wavelet packet for improving speech quality and intelligibility", in IEEE International Conference on Signal Processing & Integrated Networks (**SPIN**), Noida, India, 20-21, Feb. 2014.
- [13] Sachin Singh, Manoj Tripathy and R. S. Anand, "Evaluation of noise calculation techniques in low SNR environment for speech enhancement", in IEEE International Conference on Recent Advances and Innovations in Engineering (**ICRAIE**), Jaipur India, 09-11 May 2014.
- [14] Sachin Singh, Manoj Tripathy and R. S. Anand, "Wavelet packet based multiple noise suppression in single channel speech using binary mask threshold", in IEEE International Conference on Signal Propagation and Computer Technology (**ICSPCT**), Ajmer, India, 12-13 July 2014.

## REFERENCES

---

- [1]. Y. Hu and P. C. Loizou, "A comparative intelligibility study of single-microphone noise reduction algorithms," *J. Acoust. Soc. Am.*, vol. 122, no. 3, pp. 1777–1786, 2007.
- [2]. M. R. Schroeder, *Apparatus for suppressing noise and distortion in communication signals*. U.S. Patent no. 3180936, April 27, 1965.
- [3]. P. Scalart and J. Vieira-Filho, "Speech enhancement based on a priori signal to noise estimation," in *Proc. of the 21<sup>st</sup> IEEE Int. Conf. on Acoust., Speech, and Signal Process. (ICASSP-96)*, pp. 629–632, Atlanta, Georgia, 1996.
- [4]. J. Chen, J. Benesty and Y. Huang, "New insights into the noise reduction wiener filter," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 4, pp. 1218–1234, pp. 2006.
- [5]. S. Kamath and P. C. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *Student Research Abstracts of Proc. IEEE Int. Conf. on Acoust., Speech, and Signal Process. (ICASSP)*, vol. 4, pp. IV-4164, Orlando, FL, USA, 2002.
- [6]. Y. Hu and P. C. Loizou, "Subjective evaluation and comparison of speech enhancement algorithms," *Speech Commun.*, vol. 49, no. 7, pp. 588–601, 2007.
- [7]. Y. Hu and P. C. Loizou, "Subjective comparison of speech enhancement algorithms," in *Proceedings of the IEEE Int. Conf. on Acoust., Speech, and Signal Process. (ICASSP)*, vol. 1, pp. 153–156, 2006.
- [8]. L. Rabiner and B.-H. Juang, *Fundamentals of speech recognition*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.
- [9]. J. John R. Deller, J. G. Proakis and J. H. Hansen, *Discrete Time Processing of Speech Signals*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1993.
- [10]. Y. Ephraim, H. L. Ari and W. Roberts, "A brief survey of speech enhancement," in *The Electronic Handbook*, 2nd ed. CRC Press, Apr. 2005.
- [11]. Y. Ephraim and I. Cohen, "Recent advancements in speech enhancement," in *The Electrical Engineering Handbook*. CRC Press, ch. 15, pp. 12–26, 2006.
- [12]. L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals: 1<sup>st</sup> Edition* Pearson education, 1978.
- [13]. C. Cherry and R. Wiley, "Speech communication in very noisy environments," *Nature*, vol. 214, pp.1184, Jun. 1967.
- [14]. P. Satyanarayana, "*Short segment analysis of speech for enhancement*," Ph.D. dissertation, Indian Institute of Technology Madras, Dept. of Computer Science and Engg., Chennai, India, Feb. 1999.

- [15]. S. R. M. Prasanna, "Event based analysis of speech," Ph.D. dissertation, Indian Institute of Technology Madras, Dept. of Computer Science and Engg., Chennai, India, Mar. 2004.
- [16]. J. Lim and A. Oppenheim, "Enhancement and bandwidth compression of noisy speech," in Proc. IEEE, vol. 67, no. 12, pp. 1586–1604, Dec. 1979.
- [17]. S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust. Speech Signal Proces, vol. 27, no. 2, pp. 113–120, 1979.
- [18]. M. Berouti, R. Schwartz and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in IEEE Int. Conf. on Acoust., Speech, and Signal Process. (ICASSP), vol. 4, pp. 208–211, 1979.
- [19]. H. Gustafsson, S. E. Nordholm and I. Claesson, "Spectral subtraction using reduced delay convolution and adaptive averaging," IEEE Trans. Speech Audio Proces, vol. 9, no. 8, pp. 799–807, 2001.
- [20]. Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," IEEE Trans. Speech Audio Proces., vol. 11, no. 4, pp. 334–341, 2003.
- [21]. F. Jabloun and B. Champagne, "Incorporating the human hearing properties in the signal subspace approach for speech enhancement," IEEE Trans. Speech Audio Proces., vol. 11, no. 6, pp. 700–708, 2003.
- [22]. Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," IEEE Trans. Acoust. Speech Signal Proces., vol. 32, no. 6, pp. 1109–1121, 1984.
- [23]. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Trans. Acoust. Speech Signal Process., vol. 33, no. 2, pp. 443–445, 1985.
- [24]. P. C. Loizou, "Speech enhancement based on perceptually motivated Bayesian estimators of the magnitude spectrum," IEEE Trans. Speech Audio Proces., vol. 13, no. 5, pp. 857–869, 2005.
- [25]. O. D. Deshmukh, C. Espy-Wilson and L.H. Carney, "Speech Enhancement Using The Modified Phase Opponency Model", Journal of Acoustic Society of America, vol. 121, no. 6, pp 3886-3898, 2007.
- [26]. Jani Nurminen, Hanna Silen, Victor Popa, Elina Helander and Moncef Gabbouj, "Voice Conversion," in Speech Enhancement, Modeling and Recognition - Algorithms and Applications, Intech 2012.

- [27]. O. D. Deshmukh and C. Espy-Wilson, "Speech enhancement using auditory phase opponency model," in Proceedings of the Eurospeech, pp. 2117–2120, Lisbon, Portugal, 2005.
- [28]. Jihwan Park, Jong-Woong Kim and Joon-Hyuk Chang, Yu Gwang Jin, Nam Soo Kim, "Estimation of speech absence uncertainty based on multiple linear regression analysis for speech enhancement," Applied Acoustics, vol. 87, pp. 205-211, 2015.
- [29]. Jani Nurminen, Hanna Silén, Elina Helander and Moncef Gabbouj, "Evaluation of Detailed Modeling of the LP Residual in Statistical Speech Synthesis," in Proc. of IEEE Int. Symposium on Circuits and Systems, ISCAS, Beijing, China, 19-23 May 2013.
- [30]. W.J. Hardcastle and A. Marchal, Editors, Speech Production and Speech Modeling, NATO ASI Series, Series D: Behavioral and Social Sciences, vol. 55, The Netherlands: Kluwer Academic Publishers, 1990.
- [31]. O. D. Deshmukh, C. Espy-Wilson, M. Azalone and L. H. Carney, "A noise reduction strategy for speech based on phase-opponency detectors," in 149<sup>th</sup> Meeting of the Acoustical Society of America, Vancouver, Canada, 2005.
- [32]. J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," IEEE Trans. Acoust. Speech Signal Process., vol. 26, no. 3, pp. 197–210, 1978.
- [33]. Y. Hu and P. C. Loizou, "Speech enhancement based on wavelet thresholding the multitaper spectrum," IEEE Trans. Speech Audio Proces., vol. 12, no. 1, pp. 59–67, 2004.
- [34]. P. C. Loizou, *Speech Enhancement: Theory and Practice*. CRC, Boca Raton, 2007.
- [35]. P. C. Loizou and G Kim, "Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions," IEEE Trans. Audio Speech Lang. Process., vol.19, no.1, pp. 47–56, 2011.
- [36]. J. S. Lim, "Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise," IEEE Trans. Acoust. Speech Signal Proces., vol. 26, no. 5, pp. 471–472, 1978.
- [37]. P. C. Loizou, A. Lobo and Y. Hu, "Subspace algorithms for noise reduction in cochlear implants," J. Acoust. Soc. Am., vol. 118, no. 5, pp. 2791–2793, 2005.
- [38]. G. Kim, Y. Lu, Y. Hu and P. C. Loizou, "An algorithm that improves speech intelligibility in noise for normal-hearing listeners," J. Acoust. Soc. Am., vol. 126, no. 3, pp. 1486–1494, 2009.

- [39]. E. W. Healy, S. E. Yoho, Y. Wang and D. L. Wang, "An algorithm to improve speech recognition in noise for hearing-impaired listeners," *J. Acoust. Soc. Am.*, vol. 134, no. 4, pp. 3029–3038, 2013.
- [40]. K. Hu, P. Divenyi, D. Ellis, Z. Jin, B. G. Shinn-Cunningham and D. L. Wang, "Preliminary intelligibility tests of a monaural speech segregation system," in *ISCA Tutorial and Research Workshop on Statistical and Perceptual Audition (SAPA)*, pp. 11–16, 2008.
- [41]. E. Habets, "*Single-and multi-microphone speech dereverberation using spectral enhancement*," Ph.D. dissertation, Technische Universiteit Eindhoven, The Netherlands, Jun. 2007.
- [42]. Sandipan Dandapat, Sudeshna Sarkar and Anupam Basu, "Automatic Part-of-Speech Tagging for Bengali: An approach for morphologically rich languages in poor resource scenario," *ACL June 2007*.
- [43]. M. Wu and D. Wang, "A two-stage algorithm for one-microphone reverberant speech enhancement," *IEEE Trans. Audio, Speech, Language process.*, vol. 14, pp. 774–784, May 2006.
- [44]. K. Lebart and J. Boucher, "A new method based on spectral subtraction for speech dereverberation," *Acta Acoustica*, vol. 87, pp. 359–366, 2001.
- [45]. B. Yegnanarayana and P. Satyanarayana Murthy, "Enhancement of reverberant speech using LP residual signal," *IEEE Trans. Speech Audio process.*, vol. 8, pp. 267–281, May 2000.
- [46]. B. Yegnanarayana, S. R. M. Prasanna, R. Duraiswami and D. Zotkin, "Processing of reverberant speech for time-delay estimation," *IEEE Trans. Speech Audio process.*, vol. 13, pp. 1110–1118, Nov. 2005.
- [47]. D. L. Wang and J. S. Lim, "The unimportance of phase in speech enhancement," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 30, no.4, pp. 679-681, 1982.
- [48]. B. J. Shannon and K. K. Paliwal, "Role of Phase Estimation in Speech Enhancement", *INTERSPEECH-ICSLP*, Pittsburgh, Pennsylvania, 2006.
- [49]. K. K. Paliwal and L. D. Alsteris, "On the usefulness of STFT phase spectrum in human listening tests," *ELSEVIER, Speech communication*, vol. 45, pp 153-170, 2005.
- [50]. Timo Gerkmann and Martin Krawczyk, "MMSE-optimal spectral amplitude estimation given the STFT-phase," *IEEE Signal Processing Letters*, vol. 20, no.2, pp. 1-4, Feb. 2013.

- [51]. K. K. Paliwal, K. Wojcicki and B. Shannon, "The importance of phase in speech enhancement," *ELSEVIER Speech Commun.*, vol. 53, no. 4, pp. 465-494, Apr. 2011.
- [52]. Anthony P. Stark, K. Wojcicki, James G. Lyons and K. K. Paliwal, "Noise driven short-time phase spectrum compensation procedure for speech enhancement," *Inter Speech*, Brisbane Australia, Sept. 22-26, 2008.
- [53]. K. Wojcicki, Mitar Milacic, Anthony P. Stark, James G. Lyons and K. K. Paliwal, "Exploiting conjugate symmetry of the short-time Fourier spectrum for speech enhancement," *IEEE Signal Processing Letters*, vol. 15, pp. 461-464, 2008.
- [54]. I. B. Thomas, and A. Ravindran, "Intelligibility enhancement of already noisy speech signals," *J. Audio Eng. Soc.*, vol. 22, no. 3, pp. 234-236, 1974.
- [55]. R. A. Curtis and V. Neiderjohn, "An investigation of several frequency domain processing methods for enhancing the intelligibility of speech in wideband random noise," in *Proc. IEEE Int. Conf. Acoust. Speech and Signal Processing*. pp. 606-609, 1978.
- [56]. M. R. Sambur, "Adaptive noise cancelling for speech signals. *IEEE Trans. Acoust., Speech and Signal Processing*, vol. 26, no.6, pp. 419-423, 1978.
- [57]. M. H. Hecker and N. Guttman, "A survey of methods for measuring speech quality," *J. Audio Eng. Soc.*, vol. 15, no. 5, pp. 400-413, 1967.
- [58]. ANSI S3.5, *American National Standard Methods for the Calculation of the Articulation Index*, American National Standards Institute, New York, NY, USA, 1969.
- [59]. Rajesh Kumar Dubey and Arun Kumar, "Non-intrusive speech quality assessment using several combinations of auditory features," *Intl. Journal of Speech Technology*, Springer, vol. 16, no. 1, pp. 89-101, March 2013.
- [60]. Kamlesh Dutta, Nupur Prakash and Saroj Kaushik, "Resolving Pronominal Anaphora in Hindi," *Web Journal of Formal Computation and Cognitive Linguistics*, vol. 10, January 2008.
- [61]. K. Audhkhasi and Arun Kumar, "Two scale auditory features based non-intrusive speech quality evaluation," *IETE Journal of Research*, vol. 56, no. 2, pp.111-118, 2010.
- [62]. Kamlesh Dutta, Nupur Prakash and Saroj Kaushik, "Probabilistic Neural Network Approach to the Classification of demonstrative Pronouns for Indirect Anaphora in Hindi," *International research journal Expert Systems with Applications*, Elsevier, vol 37, no. 8, pp. 5607-5613, Aug. 2010,



- [63]. Abhijit Karmakar, Arun Kumar and R. K. Patney, "A multiresolution model of auditory excitation pattern and its application to objective evaluation of perceived speech quality", *IEEE Trans. on Audio, Speech and Language Process.*, vol. 4, no. 6, pp. 1912-1923, Nov. 2006.
- [64]. N. R. French and J. C. Steinberg, "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Amer.*, vol. 19, no.1, pp. 90–119, 1947.
- [65]. K. D. Kryter, "Methods for the calculation and use of the articulation index," *J. Acoust. Soc. Amer.*, pp. 1689–1697, 1962.
- [66]. ANSI S3.5, *Methods for the Calculation of the Speech Intelligibility Index*, American National Standards Institute, New York, NY, USA, 1995.
- [67]. K. S. Rhebergen and N. J. Versfeld, "A speech intelligibility index based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," *J. Acoust. Soc. Amer.*, vol. 117, no. 4, pp. 2181–2192, 2005.
- [68]. J. M. Kates and K. H. Arehart, "Coherence and the speech intelligibility index," *J. Acoust. Soc. Amer.*, vol. 117, no. 4, pp. 2224–2237, 2005.
- [69]. Sound system equipment–part 16: *Objective rating of speech intelligibility by speech transmission index*, IEC 60268-16, Int. Electrotechnical Commission, Geneva, Switzerland, 2003.
- [70]. H. J. M. Steeneken and T. Houtgast, "A physical method for measuring speech-transmission quality," *J. Acoust. Soc. Amer.*, vol. 67, pp. 318–326, 1980.
- [71]. G. Gweon, M. Jain, J. Mc Donough, B. Raj and C. P. Rosé, "Measuring Prevalence of Other-Oriented Transactive Contributions Using an Automated Measure of Speech Style Accommodation," *International Journal of Computer Supported Collaborative Learning*, vol. 8, no. 2, pp 245-265, June 2013.
- [72]. T. Houtgast and H. J. M. Steeneken, "Evaluation of speech transmission channels by using artificial signals," *Acustica*, vol. 25, pp. 355–367, 1971.
- [73]. V. Hohmann and B. Kollmeier, "The effect of multichannel dynamic compression on speech intelligibility," *J. Acoust. Soc. Amer.*, vol. 97, pp. 1191–1195, 1995.
- [74]. R. Drullmann, "Temporal envelope and fine structure cues for speech intelligibility," *J. Acoust. Soc. Amer.*, vol. 97, no. 1, pp. 585–592, Jan. 1995.
- [75]. R. L. Goldsworthy and J. E. Greenberg, "Analysis of speech-based speech transmission index methods with implications for nonlinear operations," *J. Acoust. Soc. Amer.*, vol. 116, no. 6, pp. 3679–3689, Dec. 2004.

- [76]. S. Jørgensen and T. Dau, "Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing," *J. Acoust. Soc. Amer.*, vol. 130, no. 3, pp. 1475–1487, Sep. 2011.
- [77]. C. Ludvigsen, C. Elberling and G. Keidser, "Evaluation of a noise reduction method—comparison between observed scores and scores predicted from STI," *Scand. Audio. Suppl.*, vol. 38, pp. 50–55, 1993.
- [78]. C. H. Taal, R. C. Hendriks, R. Heusdens and J. Jensen, "On predicting the difference in intelligibility before and after single-channel noise reduction," in *Proc. Int. Workshop, Acoust. Echo Noise Control*, 2010.
- [79]. C. Christiansen, M. S. Pedersen and T. Dau, "Prediction of speech intelligibility based on an auditory preprocessing model," *Speech Commun.*, vol. 52, pp. 678–692, 2010.
- [80]. C. H. Taal, R. C. Hendriks, R. Heusdens and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 2125–2136, Sep. 2011.
- [81]. T. M. Cover and J. A. Thomas, *Elements of information theory*. New York, NY, USA: Wiley, 1991.
- [82]. J. B. Allen, "The articulation index is a Shannon channel capacity," in *Auditory Signal Processing*, D. Pressnitzer, A. Cheveigné, S. McAdams, and L. Collet, Eds. New York, NY, USA: Springer, pp. 313–319, 2005.
- [83]. Kamlesh Dutta, Nupur Prakash and Saroj Kaushik, "Hybrid Framework for Information Extraction for Geographical Terms in Hindi Language Texts," in *Proc. of IEEE Int. Conf. on Natural Language Processing and Knowledge Engineering*, pp. 577–581, Nov. 2005.
- [84]. A. Leijon, "Articulation index and shannon mutual information," in *Hearing - From Sensory Processing to Perception*, B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey, Eds. Berlin/Heidelberg, Germany: Springer, pp. 525–532, 2007.
- [85]. J. Taghia, R. Martin and R. C. Hendriks, "On mutual information as a measure of speech intelligibility," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 65–68, 2012.
- [86]. J. Jensen and C. H. Taal, "Speech intelligibility prediction based on mutual information," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 2, pp. 430–440, Feb. 2014.

- [87]. Priti Jagwani, Shivendra Tiwari and Saroj Kaushik, "Defending Location Privacy using Zero Knowledge Proof Concept in Location Based Services," in IEEE 13<sup>th</sup> International Conference on Mobile Data Management, Bengaluru, India, July 23-26, 2012.
- [88]. J. Ma, Y. Hu and P. C. Loizou, "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions," The Journal of the Acoustical Society of America, vol. 125, pp. 3387-3405, 2009.
- [89]. A. Rix, J. Beerends, M. Hollier and Hekstra, "A Perceptual evaluation of speech quality (PESQ) - A new method for speech quality assessment of telephone networks and codecs," in. Proc. IEEE Int. Conf. Acoust, Speech, Signal Processing, vol. 2, pp. 749-752, 2001.
- [90]. ITU, Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs. ITU-T Recommendation p. 862, 2000.
- [91]. J. Beerends and J. Stemerdink, "A perceptual speech-quality measure based on a psychoacoustic sound representation," J. Audio Eng. Soc., vol.42, no.3, pp.115-123, 1994.
- [92]. IEEE Subcommittee, "IEEE Recommended Practice for Speech Quality Measurements," IEEE Trans. Audio and Electroacoustics, vol. 17, no. 3, pp. 225-246, 1969.
- [93]. International Telecommunication Union - Radio communication Sector, Recommendation BS. 562-3, Subjective assessment of sound quality, 1990.
- [94]. International Telecommunication Union - Telecommunication Sector, Recommendation, Subjective performance assessment of telephone band and wideband digital codecs, 1998.
- [95]. P. C. Loizou and J. Ma, "Extending the articulation index to account for non-linear distortions introduced by noise-suppression algorithms," The Journal of the Acoustical Society of America, vol. 130, no. 2, pp. 986-995, 2011.
- [96]. ITU-T, Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm, ITU-T Recommendation, 2003.
- [97]. Debi Prasad Das and Ganapati Panda, "Active Mitigation of Nonlinear Noise Processes Using a Novel Filtered-s LMS Algorithm," IEEE Transactions on Speech and Audio Processing, vol. 12, no. 3, MAY 2004.

- [98]. J. Hansen and B. Pellom, "An effective quality evaluation protocol for speech enhancement algorithms," in Proc. Inter. Conf. on Spoken Lang. Proc., vol. 7, pp. 2819–2822, 1998.
- [99]. J. Tribolet, P. Noll, B. McDermott and R. E. Crochiere, "A study of complexity and quality of speech waveform coders," in Proc. IEEE Int. Conf. on Acoust., Speech, and Signal Process., pp. 586–590, 1978.
- [100]. N. Kitawaki, H. Nagabuchi and K. Itoh, "Objective quality evaluation for low bit-rate speech coding systems," IEEE J. Select. Areas in Comm., vol. 6, no. 2, pp. 262-273, 1988.
- [101]. Arun Kumar and S. K. Mullick, "Attractor dimension, entropy and modeling of speech time series," IEE Electronics Letters, vol. 26, no. 21, pp. 1790 - 1792, 1990.
- [102]. W.J. Phillips, W. Robertson and S. Sarkar; "Systolic designs for speech processing," in Proc. Canadian Conf. on Electrical and Computer Engineering, pp.274-276, Vancouver, BC, Sept.14-17 1993.
- [103]. B. Suhm, L. Levin, N. Coccaro, J. Carbonell, K. Horiguchi, R. Isotani, A. Lavie, L. Mayfield, C. P. Rosé, C. Van Ess-Dykema and A. Waibel, "Speech-Language Integration in a Multi-Lingual Speech Translation System," in Proc. of the American Association of Artificial Intelligence Workshop on Integration of Natural Lang. and Speech Process., 1994.
- [104]. V. Gupta, J. Ajmeria, A. Kumar and A. Verma, "A language independent approach to audio search," in Proc. Interspeech, pp. 1125-1128, Florence, Italy, Aug. 2011.
- [105]. Nithin V. George and Ganapati Panda, "Advances in active noise control: A survey, with emphasis on recent nonlinear techniques," Elsevier, Signal Processing, vol. 93, pp. 363-377, 2013.
- [106]. S. Sen, W. Robertson and W. J. Phillips, "The effects of reduced precision bit lengths on feed forward neural networks for speech recognition," IEEE Int. Conf. on Neural Networks, ICNN'96, June 1996.
- [107]. M. Woszczyna, N. Coccaro, A. Eisele, A. Lavie, A. McNair, T. Polzin, I. Rogina, C. P. Rosé, T. Sloboda, J. Tsutsumi, N. Aoki-Waibel, A. Waibel and W. Ward, "Recent Advances in JANUS: A Speech Translation System," in ARPA Proceedings of the Human Language Technologies Workshop, 1993.
- [108]. D. Klatt, "Prediction of perceived phonetic distance from critical band spectra," in Proc. IEEE Int. Conf. Acoust. , Speech, Signal Processing, vol. 7, pp. 1278-1281, 1982.

- [109]. F. Itakura and S. Saito, "An analysis-synthesis telephony based on maximum likelihood method," in Proc. 6<sup>th</sup> Int. Conf. Acoust., pp. 17-20, 1968.
- [110]. S. Quackenbush, T. Barnwell and M. Clements, *Objective measures of speech quality*. NJ: Prentice-Hall, Eaglewood Cliffs, 1988.
- [111]. V. Gupta, J. Ajmeria, A. Kumar and A. Verma, "A language independent approach to audio search," in Proc. Interspeech, pp. 1125-1128, Florence, Italy, Aug. 2011.
- [112]. M. Paralič, M. Kuba and R. Jarina, "Two Dimensional Cepstrum and Mel Filter Cepstral Coefficients analysis in Speech Recognition," konferencia Transcom, Žilina, 2005.
- [113]. R. Jarina, Ing. M. Kuba and M. Paralič, "Compact Representation of Speech Using 2-D Cepstrum - an Application to Slovak Digits Recognition," in 8<sup>th</sup> Int. Conf. on Text, Speech and Dialogue, Karlovy Vary, 2005.
- [114]. T. Parsons, "Separation of speech from interfering speech by means of harmonic selection," J. Acoust. Soc. Am., vol. 60, pp. 911-918, Oct. 1976.
- [115]. D. Morgan, E. George, L. Lee and S. Kay, "Cochannel speaker separation by harmonic enhancement and suppression," IEEE Trans. Speech Audio process. vol. 5, pp. 407-424, Sep. 1997.
- [116]. O. D Deshmukh and C. Espy-Wilson, "Modified Phase Opponency Based Solution to the Speech Separation Challenge," in Proc. of Interspeech, pp. 101-104, Pittsburgh, 2006.
- [117]. Jani Nurminen, Hanna Silén and Moncef Gabbouj, "Speaker-specific retraining for enhanced compression of unit selection text-to-speech databases," in Proc. Interspeech, Lyon, France, Aug. 2013.
- [118]. M. Portnoff, "Short-time fourier analysis of sampled speech," IEEE Trans. Acoust., Speech, Signal process., vol. 29, no. 3, pp. 364-373, Jun. 1981.
- [119]. B. Yegnanarayana, S.R.M. Prasanna and M. Mathew, "Enhancement of speech in multispeaker environment," in Proc. European Conf. Speech process., Technology, Geneva, Switzerland, pp. 581-584, 2003.
- [120]. V. Mitra, H. Nam, C. Espy-Wilson, E. Saltzman and L. Goldstein, "Tract variables for noise robust speech recognition," IEEE Trans. on Audio, Speech & Language Processing, vol. 19, no. 7, pp. 1913-1924, 2011.
- [121]. V. Mitra, H. Nam, C. Espy-Wilson, E. Saltzman and L. Goldstein, "Noise Robustness of Tract Variables and their Application to Speech Recognition," in Proc. of Interspeech, pp. 2759-2762, 2009.

- [122]. S. R. M. Prasanna, "Event based analysis of speech," Ph.D. dissertation, Indian Institute of Technology Madras, Dept. of Computer Science and Engg., Chennai, India, Mar. 2004.
- [123]. S. Shukla, S. R. M. Prasanna and S. Dandapat, "Stressed speech processing: Human vs automatic in non-professional speakers scenario," in National Conference on Communications (NCC), IEEE, pp. 1 –5, January 2011.
- [124]. F. Dubbelboer and T. Houtgast, "The concept of signal-to-noise ratio in the modulation domain and speech intelligibility," J. Acoust. Society America, vol. 124, no. 6, pp. 3937-3946, 2008.
- [125]. S. Jorgensen and T. Dau, "Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing," J. Acoust. Society America. 1475-1487, 2011.
- [126]. Kisoo Kwon, Jong Won Shin and Nam Soo Kim, "NMF-Based Speech Enhancement Using Bases Update," IEEE Signal Process. Letters, vol. 22, no. 4, Apr. 2015.
- [127]. K. Paliwal, B. Schwerin and K. Wojcicki, "Role of modulation magnitude and phase spectrum towards speech intelligibility," Speech Communication, pp. 327-339, 2011.
- [128]. K. Wojcicki and P. C. Loizou, "Channel selection in the modulation domain for improved speech intelligibility in noise," J. Acoust. Society America, pp. 2904-2913, 2012.
- [129]. Yu. Guoshen, E. Bacry and S. Mallat, "Audio signal denoising with complex wavelets and adaptive block attenuation," in Proc. IEEE Int. Conf. Acoustic, speech signal processing (ICASSP), vol. 3, pp. 869-872, 2007.
- [130]. P. Krishnamurthy and S R M Prasanna, "Modified spectral subtraction method for enhancement of noisy speech," in Proc 3<sup>rd</sup> Int. Conf. on intelligent sensing and information processing, pp 146-150, Bangalore, India, 2005.
- [131]. Bin Zhou, "An improved wavelet-based speech enhancement method using adaptive block thresholding," in Proc. IEEE Conference, 2010.
- [132]. Nithin V. George and Ganapati Panda, "Active control of nonlinear noise processes using cascaded adaptive nonlinear filter," Elsevier, Applied Acoustics, vol. 74, pp. 217-222, 2013.
- [133]. Tahsina Farah Sanam and Celia Shahnaz, "Enhancement of noisy speech based on a custom thresholding function with a statistically determined threshold," Int. J. Speech Technology, April 2012.
- [134]. D. L. Donoho, "De-noising by soft thresholding," IEEE Trans. Inform. Theory, vol. 41,

- pp. 613-627, 1995.
- [135]. D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425-455, 1994.
- [136]. Ghanbari Yasser and Mohammad Reza Karami, "A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets. *Speech Communication*, vol. 48, pp. 927 – 940, 2006.
- [137]. R. Kumar, R. Gangadhara, Sharath Rao, K. Prahallad, C. P. Rosé and A. W. Black, "Building a Better Indian English Voice Using More Data," in 6<sup>th</sup> ISCA Workshop on Speech Synthesis, Bonn, Germany, 2007.
- [138]. Kishore Prahallad, E. Naresh and Venkatesh Keri, *The IIIT-H Indic Speech Databases*. In Proceedings of Interspeech Portland, Oregon, USA. , 2012.  
(<http://speech.iiit.ac.in/index.php/research-svl/69.html>)
- [139]. P. Varga and H. J. M. Steeneken: Technical report, DRA Speech Research Unit, 1992. (<http://www.speech.cs.cmu.edu/comp.speech/Sect-ion1/Data/noisex.html>)
- [140]. NOIZEUS: A noisy speech corpus for evaluation of speech enhancement algorithms. <http://www.utdallas.edu/~loizou/speech/noizeus>
- [141]. H. L. Shashidhara, T. S. Suneel, V. M. Gadre and S.S. Pande, "Accuracy improvement for CNC system using wavelet-neural networks," in Proc. of the IEEE Int. Conf. on Industrial Technology, vol. 2, pp. 341-346, 2000.
- [142]. G. S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 2, no. 7, pp. 674-693, 1989.
- [143]. Kamlesh Dutta, Nupur Prakash and Saroj Kaushik, "Probabilistic Neural Network Approach to the Classification of demonstrative Pronouns for Indirect Anaphora in Hindi," *Int. research journal Expert Systems with Applications*, Elsevier, vol 37, no. 8, pp. 5607-5613, Aug. 2010,
- [144]. W. J. Phillips, C. Tosuner and W. Robertson; "Speech recognition techniques using RBF networks," WESCANEX 1995.
- [145]. Kamlesh Dutta, Nupur Prakash and Saroj Kaushik, "Resolving Pronominal Anaphora in Hindi," *Web Journal of Formal Computation and Cognitive Linguistics*, vol. 10, January 2008.
- [146]. M. Misiti and Y. Misiti, "*Wavelet Toolbox, for use with Matlab*", Georges Oppenheim, Jean-Michel Poggi, User guide version 1-624, 1996.

- [147]. O. Farooq and S. Datta, "Wavelet based robust sub-band features for phoneme recognition, *IEE Proc. Visual Image Signal Process.*, vol. 151, no. 3, pp. 187-193, June 2004.
- [148]. R. Polikar, "The story of wavelets", *IMACS/IEEE CSCC'99 Proceedings*, pp. 5481-5486, 1999.
- [149]. S.R. Pillai, W. Robertson and W. Phillips; "Subband filters using allpass structures", in *Proceedings of the IEEE, Int. Conf. on Acoust., Speech, and Signal Process.*, Toronto, May 1991.
- [150]. S. Rangachari and P. C. Loizou, "A noise-estimation algorithm for highly non-stationary environments," *Speech Communication*, vol. 48, pp. 220-231, 2006.
- [151]. S. Vaseghi and R. Frayling-Cork, "Restoration of old gramophone recordings," *J. Audio Eng.*, vol. 40, no. 10, pp. 791-801, 1992.
- [152]. Ganesh Sivaraman, Vikramjit Mitra and Carol Y. Espy-Wilson, "Fusion of acoustic, perceptual and production features for robust speech recognition in highly non-stationary noise," *The 2<sup>nd</sup> CHiME Workshop on Machine Listening in Multisource Environments*, Vancouver, Canada, June 1<sup>st</sup>, 2013.
- [153]. O. Cappe and J. Laroche, "Evaluation of short-time spectral attenuation techniques for the restoration of musical recordings," *IEEE Trans. Speech Audio Processing*, 1994.
- [154]. O. Cappe, "Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor," *IEEE Trans. Speech Audio Processing*, vol. 2, no. 2, pp. 345-349, April 1994.
- [155]. Cyril Plapous, Claude Marro, Laurent Mauuary and Pascal Scalart, "A two-step noise reduction technique", *ICASSP*, 2004.
- [156]. G. Whipple "Low residual noise speech enhancement utilizing time-frequency filtering," in *Proceedings of IEEE Int. Conf. on Acoust., Speech, and Signal Process.*, pp. I/5-I/8, Apr.1994.
- [157]. Z. Goh, K.C. Tan and B.T.G. Tan, "Post-processing Method for Suppressing Musical Noise Generated by Spectral Subtraction," *IEEE Trans. on Speech and Audio Process.*, vol.6,no. 3, pp. 287-292, May 1998.
- [158]. Moncef Gabbouj and Edward J. Coyle, "Minimum Mean Absolute Error Stack Filtering With Structural Constraints," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 955-968, June 1990.



- [159]. Z. Lin and R. Gourban, "Musical noise reduction in speech using two-dimensional spectrogram enhancement," in Proceedings of the 2<sup>nd</sup> IEEE Int. Workshop on Haptic, Audio and Visual Environments and Their Applications, pp. 61-64, Sept. 2003.
- [160]. I. Y. Soon and S.N. Koh, "Speech Enhancement Using 2-D Fourier Transform," IEEE Trans. on Speech and Audio Process., vol. 11, no.6, pp. 717-724, Nov. 2003.
- [161]. Alexey Lukin and Jeremy Todd, "Suppression of Musical Noise Artifacts in Audio Noise Reduction by Adaptive 2D Filtering," in Audio Engineering Society, Presented at the 123<sup>rd</sup> Convention, New York, NY, USA, Oct. 5–8, 2007.
- [162]. Ganapati Panda, Bernard Mulgrew, Colin F. N. Cowan and Peter M. Grant, "A self-orthogonalizing efficient block adaptive filter," IEEE Trans. on Acoust., Speech, and Signal Process., vol. 34, no. 6, Dec. 1986.
- [163]. Thomas Esch and Peter Vary, "Efficient musical noise suppression for speech enhancement systems," ICASSP, 2009.
- [164]. Minje Kim and Paris Smaragdis, "Mixtures of Local Dictionaries for Unsupervised Speech Enhancement," IEEE Signal Process. Letters, vol. 22, no. 3, March 2015.
- [165]. Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," IEEE Trans. Audio Speech Lang. Process., vol. 16, pp. 229–238, 2008.
- [166]. C. Breithaupt and R. Martin, "Analysis of the decision-directed SNR estimator for speech enhancement with respect to low-SNR and transient conditions," IEEE Trans. Audio Speech Lang. Process., vol.19, pp.277–289, 2011.
- [167]. Samit Bhattacharya, Sudeshna Sarkar and Anupam Bas, "Sanyog: A Speech Enabled Communication System for the Speech Impaired and People with Multiple Disorders," Journal of Technology in Human Services, vol. 25, no. 2, pp 177-180, 2007.
- [168]. Samit Bhattacharya, Sudeshna Sarkar and Anupam Basu. "Speech Enabled Communication Tool for the Speech Impaired and People with Multiple Disorders," in Proceedings of the seventh international conference of Human Services Information Technology Applications, Hong Kong, China, 2004.
- [169]. J. Erkelens, J. Jensen and R. Heusdens, "A data-driven approach to optimizing spectral speech enhancement methods for various error criteria," Speech Commun., Vol. 49, pp.530–541, 2007.
- [170]. R. Martin, "Statistical methods for the enhancement of noisy speech," Speech Enhancement. Springer book ch., Berlin, pp. 43–64, 2005.
- [171]. R. Martin, "Bias compensation methods for minimum statistics noise power spectral density estimation," Signal Process., vol. 86, pp. 1215–1229, 2006.

- [172]. A. Papoulis, S. Pillai, *Probability Random Variables and Stochastic Processes*, 4th ed. McGraw Hill, Inc., New York, 2002.
- [173]. S. Shukla, S. Dandapat and S. R. M. Prasanna, "Spectral slope based analysis and classification of stressed speech," *Int J Speech Tech*, vol. 14, pp. 245–258, 2011.
- [174]. C. Plapous, C. Marro and P. Scalart, "Improved signal-to-noise ratio estimation for speech enhancement," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, pp. 2098–2108, 2006.
- [175]. Roman Jarina, Martin Paralič, Michal Kuba, Ján Olajec, Andrej Lukáč and Miroslav Dzurek, "Development of a Reference Platform for Generic Audio Classification," in *Proc. of 9<sup>th</sup> Int. Workshop on Image Analysis for Multimedia Interactive Services*, pp. 239-242, Klagenfurt, Austria, May 2008.
- [176]. I. Cohen, "Relaxed statistical model for speech enhancement and a priori SNR estimation," *IEEE Trans. Speech Audio Process.*, vol. 13, pp. 870–881, 2005.
- [177]. G. Kim and P. C. Loizou, "Gain-induced speech distortions and the absence of intelligibility benefit with existing noise-reduction algorithms," *Journal of the Acoustical Society of America*, vol. 130, no. 3, pp. 1581-1596, 2011.
- [178]. F. Chen and P. Loizou, "Impact of SNR and gain-function over- and under-estimation on speech intelligibility," *Speech Communication*, vol. 54, pp.272-281, 2012.
- [179]. Yanna Ma and Akinori Nishihara, "A modified wiener filtering method combined with wavelet thresholding multitaper spectrum for speech enhancement," *Eurasip: Journal on audio, speech, and music process*, vol. 32, no. 7, 2014.
- [180]. Kamil K. Wojcicki and P. C. Loizou, "Channel selection in the modulation domain for improved speech intelligibility in noise," *J. Acoust. Soc. Am.*, vol. 131, no. 4, 2904-2913, 2012.
- [181]. Gibak Kim and P. C. Loizou, "Improving speech intelligibility in noise using environment-optimized algorithms," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 18, no. 8, pp. 2080-2090, 2010.
- [182]. Gibak Kim and P. C. Loizou, "A new binary mask based on noise constraints for improved speech intelligibility," *Interspeech-Japan*, sept. 2010.
- [183]. D. Brungart, P. Chang, B. Simpson and D. Wang, "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," *J. Acoust. Soc. Am.*, vol. 120, pp. 4007-4018, 2006.

- [184]. N. Li and P. C. Loizou, "Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction," *J. Acoust. Soc. Am.*, vol.123, pp. 1673-1682, 2008.
- [185]. Yu Chengzhu, Kamil K. Wojcicki, P. C. Loizou and J. H. L. Hansen, "Evaluation of the importance of time-frequency contributions to speech intelligibility in noise," *J. Acoust. Soc. Am.*, vol. 135, no. 5, pp. 3007-3016, 2014.
- [186]. D. A. Berkley. *Acoustical factors affecting hearing aid performance*. University Park Press, 1982.
- [187]. J. B. Allen, "Effects of small room reverberation on subjective preference," *J. Acoust. Soc. Amer.*, page S5, 1982.
- [188]. J. J. Jetzt, "Critical distance measurement of rooms from the sound energy spectral response," *J. Acoust. Soc. Amer.*, pp. 1204–1211, 1979.
- [189]. A. Sugiyama and L. Gatttonit, "Two-stage dereverberation with integrated reverberation and noise estimation," in 12<sup>th</sup> Digital Signal Processing Workshop, pp.322–327, 2006.
- [190]. E. A. P. Habets, N. D. Gaubitch and P. A. Naylor, "Temporal selective dereverberation of noisy speech using one microphone," in International Conference on Acoustics, Speech and Signal Processing, pp.4577–4580, 2008.
- [191]. Nima Youse fian, John H. L. Hansen and Philipos C. Loizou, "A Hybrid Coherence Model for Noise Reduction in Reverberant Environments," *IEEE Signal Process. Letters*, vol. 22, no. 3, March 2015.
- [192]. B. W. Gillespie, H. S. Malvar and D. A. F. Florencio, "Speech dereverberation via maximum-kurtosis subband adaptive filtering," in Int. Conf. on Acoust., Speech, and Signal Process., pp. 3701–3704, 2001.
- [193]. N. D. Gaubitch and P. A. Naylor, "Spatiotemporal averaging method for enhancement of reverberant speech," in Int. Conf. on Digital Signal Processing, pp. 607–610, 2007.
- [194]. E. A. P. Habets, "Single-channel speech dereverberation based on spectral subtraction," in 15<sup>th</sup> Annual Workshop on Circuits, Systems and Signal Processing, pp. 250–254, 2004.
- [195]. H. W. Lollmann and P. Vary, "A blind speech enhancement algorithm for the suppression of late reverberation and noise," in Int. Conf. on Acoust., Speech and Signal Process., pp. 3989–3992, 2009.
- [196]. E. A. P. Habets, S. Gannot and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Processing Letters*, vol. 9, pp.770–773, 2009.

- [197]. J. S. Erkelens and R. Heusdens, "Single-microphone late-reverberation suppression in noisy speech by exploiting long-term correlation in the DFT domain," in Int. Conf. on Acoustics, Speech and Signal Process., pp. 3997–4000, 2009.
- [198]. Sandipan Dandapat, Sudeshna Sarkar and Anupam Basu, "A Hybrid Model for Part-of-Speech Tagging and its Application to Bengali," Int. Conf. on Computational Intelligence, pp. 169-172, 2004.
- [199]. J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Amer., pp. 943–950, 1978.
- [200]. T. H. Falk, C. Zheng and W.-Y. Chan, "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech," IEEE Trans. on Audio, Speech and Lang. Process., pp. 1766–1774, 2010.
- [201]. J. A. Moorer, "About this reverberation business," Computer Music Journal, vol. 2, pp. 13–28, 1979.
- [202]. J. D. Polack. La transmission de l'énergie sonore dans les salles. PhD thesis, Université du Maine, 1998.
- [203]. AU17, 225-246, "IEEE recommended practice speech quality measurements," IEEE Transaction Audio Electroacoust, 1969.  
<http://www.utdallas.edu/~loizou/speech/noizeus/>.
- [204]. <http://www.ee.columbia.edu/~dpwe/sounds/noise/>
- [205]. T. Van den Bogaert, S. Doclo, J. Wouters and M. Moonen, "Speech enhancement with multichannel Wiener filter techniques in multi-microphone binaural hearing aids," The Journal of the Acoustical Society of America, Vol. 124, pp.360–371, 2009.
- [206]. K. Kokkinakis and O. Hazrati, "A channel-selection criterion for suppressing reverberation in cochlear implants," Journal of Acoustical Society of America, vol. 129, pp. 3221–3232, 2011.
- [207]. A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," in Proceedings of the IEEE, vol. 69, no. 5, pp. 529-541, 1981.
- [208]. Josef Kulmer and Pejman Mowlae, "Phase Estimation in Single Channel Speech Enhancement Using Phase Decomposition," IEEE Signal Process. Letters, vol. 22, no. 5, May 2015.
- [209]. C. H. You, S. N. Koh and S. Rahardja, "β-order MMSE spectral amplitude estimation for speech enhancement," IEEE Speech Audio Process., vol. 13, no. 4, pp. 475–486, Jul. 2005.

- [210]. R. Martin, "Speech enhancement using MMSE short time spectral estimation with Gamma distributed speech priors," in IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), pp. 253–256, May 2002.
- [211]. T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model," EURASIP J. Applied Signal Process., vol. 2005, no. 7, pp. 1110–1126, Jan. 2005.
- [212]. I. Andrianakis and P. R. White, "MMSE speech spectral amplitude estimators with Chi and Gamma speech priors," in IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Toulouse, France, pp. 1068–1071, May 2006.
- [213]. J. S. Erkelens, R. C. Hendriks, R. Heusdens and J. Jensen, "Minimum mean-square error estimation of discrete Fourier coefficients with generalized Gamma priors," IEEE Audio, Speech, Lang. Process., vol. 15, no. 6, pp. 1741–1752, Aug. 2007.
- [214]. C. Breithaupt, M. Krawczyk and R. Martin, "Parameterized MMSE spectral magnitude estimation for the enhancement of noisy speech," in IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), pp. 4037–4040, Apr. 2008.
- [215]. J. Jensen and J. H. Hansen, "Speech enhancement using a constrained iterative sinusoidal model," IEEE Speech Audio Process., vol. 9, no. 7, pp. 731–740, Oct. 2001.
- [216]. D. Griffin and J. Lim, "Signal estimation from modified short-time Fourier transform," IEEE Trans. Acoust., Speech, Signal Process., vol. 32, no. 2, pp. 236–243, Apr. 1984.
- [217]. S. Ramamohan and S. Dandapat, "Sinusoidal Model Based Analysis and Classification of Stressed Speech," IEEE Trans. of Speech and Audio Processing, May, 2006.
- [218]. R. Schluter and H. Ney, "Using phase spectrum information for improved speech recognition performance," in Proceeding of the IEEE Int. Conf. on Acoust., Speech, and Signal Process., vol. 1, pp. 133-136, May 2001.
- [219]. M. Paralič and R. Jarina, "Variable component approach in GMM - based speaker modelling," 33<sup>rd</sup> Int. Acoustical Conf. - EAA Symposium, Acoustics High Tatras, 2006.
- [220]. D. Zhu and K. K. Paliwal, "Product of power spectrum and group delay function for speech recognition," in Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Process., vol. 1, pp. 125-128, 2004.
- [221]. Sumitra Shukla, S.R.M. Prasanna and S. Dandapat, "Speech recognition under stressed condition", in Proc. NCC, pp. 294-298, Jan 2009.

- [222]. L. Wang, S. Ohtsuka and Nakagawa, "High improvement of speaker identification and verification by combining MFCC and phase information," in Proc. of the IEEE Int. Conf. on Acoust., Speech, and Signal Process., Taipei, pp. 4529-4532, 2009.
- [223]. S. So, K. K. Wojcicki, J. G. Lyons, A. G. Stark and K. K. Paliwal, "Kalman filter with phase spectrum compensation algorithm for speech enhancement," in Proc. of the IEEE Int. Conf. on Acoust., Speech, and Signal Process., Taipei, pp. 4405-4408, 2009.
- [224]. C. H. You, S. N. Koh and S. Rahardja, " $\beta$ -order MMSE spectral amplitude estimation for speech enhancement," IEEE Speech Audio Process., vol. 13, no. 4, pp. 475-486, Jul. 2005.
- [225]. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals Series and Products*, 7th ed. San Diego, CA, USA, Academic Press, 2000.
- [226]. ITU-T, Perceptual evaluation of speech quality (PESQ) and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs. ITU-T Recommendation P. 862, 2001.
- [227]. Y. Hu and P. C. Loizou, "Evaluation of objective measures for speech enhancement," in Proc. Interspeech, 2006.
- [228]. C. H. Taal, R. C. Hendriks, R. Heusdens, J. Jensen and U. Kjems, "An evaluation of objective quality measures for speech intelligibility prediction," in Proc. Interspeech, pp. 1947-1950, 2009.
- [229]. J. Ma, Y. Hu and P. Loizou, "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions," J. Acoust. Soc. Amer., vol. 125, no. 5, pp. 3387-3405, 2009.
- [230]. W. Bastiaan Kleijn and R. C. Hendriks, "A Simple Model of Speech Communication and its Application to Intelligibility Enhancement," IEEE Signal Process. Letters, vol. 22, no. 3, March 2015.
- [231]. Xinhui Zhou, Daniel Garcia-Romero, Nima Mesgarani, Maureen Stone, Carol Espy-Wilson and Shihab Shamma, "Automatic intelligibility assessment of pathologic speech in head and neck cancer based on auditory-inspired spectro-temporal modulations," Interspeech, 2012.
- [232]. J. B. Boldt and D. P. W. Ellis, "A simple correlation-based model of intelligibility for nonlinear speech enhancement and separation," in Proc. EUSIPCO, 2009, pp. 1849-1853.
- [233]. ITU-T, Mean opinion score (MOS) ITU-T Recommendation P.800.1, July 2006.

- [234]. B. Sim and Y. Tong, "A Parametric Formulation of the Generalized Spectral Subtraction Method", IEEE Trans. on Speech and Audio Process., vol. 6, no. 4, pp. 328-337, July 1998.
- [235]. I. Cohen, "Speech Enhancement Using a Non-causal *a-Priori* SNR Estimator," IEEE Signal Processing Letters, vol. 11, no. 9, pp. 725-728, Sep. 2004.
- [236]. N. Wiener, *Extrapolation, interpolation and smoothing of stationary time series: with engineering applications*. In the Cambridge, MIT Press, 1949.
- [237]. J. Capon, "High-resolution frequency-wave number spectrum analysis", in Proc. of IEEE, vol. 57, no. 8, pp. 1408-1418, 1969.
- [238]. O. L. Frost, "An algorithm for linearly constrained adaptive array processing", in Proc. of IEEE, vol. 60, no. 8, pp. 926-935, 1972.
- [239]. R. McAulay and M. Malpass, "Speech enhancement using a soft-decision noise suppression filter", IEEE Trans. Acoust., Speech, Signal Process., vol. 28, no. 2, pp. 137-145, April 1980.
- [240]. Y. Ephraim, D. Malah and B.-H. Juang, "On the application of hidden Markov models for enhancing noisy speech", IEEE Trans. Acoust., Speech, Signal Process., vol. 37, no. 12, pp. 1846-1856, 1989.
- [241]. U. Mittal and N. Phamdo, "Signal/noise KLT based approach for enhancing speech degraded by colored noise", IEEE Trans. Speech Audio Process., vol. 8, no. 2, pp. 159-167, March 2000.
- [242]. A. V. Oppenheim, R. W. Schafer and T. G. Stockham, "Nonlinear filtering of multiplied and convolved signals", IEEE Trans. Audio Electroacoust., vol. 16, no. 3, pp. 437-466, 1968.
- [243]. S. T. Neely and J. B. Allen, "Invertibility of a room impulse response", J. Acoust. Soc. Am., vol. 66, no. 1, pp. 165-169, 1979.
- [244]. J. N. Mourjopoulos, "Digital equalization of room acoustics", Journal Audio Eng. Soc., vol. 42, no. 11, pp. 884-900, 1994.

**APPENDIX - A**

Table A1: Analysis for English language

Methods	Measures	Non-stationary							Highly non-stationary					Mixed noise	Combined
		White	Volvo-car	Pink	Machine-gun	Tank	Leopard vehicle	Hf-channel	Factory Floor1	Buccaneer 2	Babble	F-16 plane	Destroyer operation	F16+Babble+NOIZEUS	Reverberation +Babble
Wiener filtering method	SNR	6.3346	7.6280	4.5242	7.2886	4.6569	8.6454	-2.0859	3.6562	1.4230	4.1732	1.9769	2.5944	8.5560	9.4964
	SSNR	-4.9067	-7.2589	-6.3218	-3.9448	-8.5721	-8.3849	-9.5711	-7.5951	-8.8284	-7.8298	-8.1383	-8.8677	-3.8526	-2.5453
	PSNR	85.5355	85.6810	84.2530	86.3820	84.2853	87.3358	80.5299	83.4229	82.2118	83.9322	82.6131	83.0013	86.1234	85.0718
	MSE	1.8177	1.7578	2.4422	1.4958	2.4241	1.2009	5.7557	2.9566	3.9076	2.6288	3.5626	3.2580	0.00015	0.00021
Spectral sub.	SNR	7.2809	1.9303	7.8899	5.8584	8.8554	5.2860	4.8226	7.8434	7.0166	8.2884	7.1828	7.5236	8.1731	9.2633
	SSNR	-2.3878	-1.5745	-2.9814	-0.629	-7.7386	-7.8766	-8.3859	-5.6054	-7.8384	-6.2833	-6.3310	-7.7812	-3.2253	-2.1634
	PSNR	90.9144	75.1588	86.1414	94.5427	86.1864	92.8582	82.6358	84.8365	84.6452	85.5890	84.6851	84.9359	85.5038	84.7171
	MSE	5.5268	.0020	1.5810	0.22846	1.5647	0.33671	3.5441	2.1352	2.2313	1.7955	2.2109	2.0868	0.00018	0.00022
MMSE-SPU	SNR	6.3306	7.4675	4.6899	7.4272	4.7653	8.5404	-1.8121	3.8078	1.6084	4.2507	2.1948	2.7773	1.1846	5.24899
	SSNR	-4.8362	-7.2791	-6.2302	-3.8620	-8.5228	-8.3564	-9.5363	-7.7189	-8.7605	-7.7829	-8.0596	-8.8239	-6.2149	-3.21494
	PSNR	85.4396	85.4635	84.2680	86.3934	84.2566	87.1513	80.5827	83.4226	82.1983	83.9025	82.6548	83.0213	81.2658	80.2684
	MSE	1.8583	1.8481	2.4338	1.4919	2.4402	1.2530	5.6861	2.9568	3.9197	2.6475	3.5286	3.2430	2.3194	3.20152
p-MMSE	SNR	5.3135	7.1101	3.5465	6.2619	3.7721	7.6330	-3.5685	2.5144	.3672	3.2808	.9718	1.5100	5.4752	8.4045
	SSNR	-5.3328	-7.5971	-6.6629	-4.4457	-8.6928	-8.4766	-9.6914	-8.0597	-8.9951	-7.9960	-8.3496	-9.0118	-6.0390	-4.8453
	PSNR	84.8328	86.6482	83.6967	85.6281	83.7582	86.5870	80.1041	82.8569	81.7011	83.4825	82.1669	82.4715	84.4488	84.8998
	MSE	2.1370	1.4069	2.7759	1.7794	2.7369	1.4268	6.3485	3.3681	4.3952	2.9163	3.9481	3.6807	0.00023	0.00021
log-MMSE	SNR	6.3346	7.6280	4.5242	7.2886	4.6569	8.6454	-2.0859	3.6562	1.4230	4.1732	1.9769	2.5944	6.6395	8.9506
	SSNR	-4.9067	-7.2589	-6.3218	-3.9448	-8.5721	-8.3849	-9.5711	-7.5951	-8.8284	-7.8298	-8.1383	-8.8677	-5.3974	-4.2037
	PSNR	85.5355	85.6810	84.2530	86.3820	84.2853	87.3358	80.5299	83.4229	82.2118	83.9332	82.6131	83.0013	85.1900	85.1736
	MSE	1.8177	1.7578	2.4422	1.4958	2.4241	1.0273	5.7557	2.9566	3.976	2.6288	3.5626	3.2580	0.000197	0.000198
Binary Mask (IDBM)	SNR	7.2591	4.8380	5.9335	11.288	4.5367	7.8703	0.2791	3.7566	4.1103	2.979	3.0038	3.2502	2.2341	1.2166
	SSNR	-4.3263	-6.1638	-5.5886	-1.8041	-8.7061	-8.3763	9.3547	-7.5746	-8.5180	-7.8050	-7.8102	-8.7689	-3.4562	-3.5891
	PSNR	86.1930	81.3173	84.9055	89.6213	84.0332	86.5294	81.3334	83.2894	83.5880	82.6432	82.9201	83.1332	67.5112	61.1176
	MSE	1.5624	4.8012	2.1015	0.70950	2.5690	1.4459	4.7834	3.0489	2.8463	3.5380	3.3195	3.1605	0.0115	0.0503
WPT-IDBM (Modified Wiener)	SNR	10.7124	11.1585	8.6652	7.4051	12.3968	14.5360	9.6353	9.8644	10.4070	11.0051	9.6133	13.0812	10.9032	11.3553
	SSNR	1.2310	0.3706	1.9203	5.3203	0.1349	1.8938	0.4202	1.0900	1.1626	0.2608	0.1617	2.3087	0.3242	1.0472
	PSNR	88.5757	88.9142	86.2090	84.6174	90.2407	92.4732	87.4883	87.5846	88.2567	88.8355	87.4499	91.0483	88.7522	86.9888
	MSE	0.00009	0.00084	0.00016	0.00023	0.00062	0.00037	0.00001	0.0001	0.00097	0.00085	0.00012	0.00005	0.000087	0.00013
WPT Fuzzy	SNR	12.0609	8.1131	10.5220	6.4400	13.8772	11.1565	11.5604	13.7225	12.3953	6.7466	9.0982	11.7353	6.5880	11.9868
	SSNR	7.1176	6.8408	7.0513	6.8156	6.9460	7.1497	6.9129	7.0892	7.2276	6.9441	7.0235	7.0733	4.5374	6.4384
	PSNR	93.8915	89.4624	92.1943	87.4041	95.7838	92.8939	93.3929	95.5252	94.2484	87.8312	90.6729	93.4915	96.6661	91.8013
	MSE	0.00027	0.00074	0.000039	0.00012	0.00001	0.000034	0.000029	0.000018	0.000024	0.00012	0.00056	0.00003	0.00014	0.000043
Phase ratio method	SNR	13.1873	30.0835	13.1392	22.3228	16.0288	23.4410	11.7532	13.9077	13.2971	14.5451	12.4817	15.6564	12.5489	11.92846
	SSNR	8.7125	9.492	9.5530	2.5802	7.252525	8.3728	7.6455	7.6326	8.6984	7.262321	8.534	9.4522	10.59874	7.2681
	PSNR	90.6236	107.526	90.5151	99.8057	93.4975	100.876	89.2951	91.2450	90.7079	92.0469	89.9802	93.0996	95.2614	96.2651
	MSE	0.00027	0.00005	0.00002	0.000010	0.00003	0.000004	0.00004	0.00008	0.00005	0.00008	0.00003	0.00001	0.00002	0.000001

Results for English language speech signals are analyzed in Table A1. In this table last three methods give better performance in comparison to existing speech enhancement methods. Since minimum error is obtained by WPT Fuzzy and phase ratio based methods hence these methods give good quality and intelligibility speech signal.



Table A2: Performance analysis for Hindi language

Methods	Measures	Non-stationary							Highly non-stationary					Mixed noise	Combined
		White	Volvo-car	Pink	Machine-gun	Tank	Leopard vehicle	Hf-channel	Factory Floor1	Buccaneer2	Babble	F-16 plane	Destroyer operations	F16+Babble+NOIZEUS	Reverberation+Babble
Wiener filtering method	SNR	7.5294	4.2057	7.8933	7.7158	4.4559	6.4924	3.9079	6.3352	4.4923	5.00	5.53	3.4461	9.3245	10.5585
	SSNR	-4.0826	-6.4439	-5.2249	-2.8421	-8.7064	-8.4229	-9.0971	-6.6962	-8.533	-6.8861	-7.4932	-8.6008	-4.7016	-3.7610
	PSNR	85.1100	81.5482	87.3334	87.6955	84.1123	85.8632	84.0945	85.6434	83.83	84.37	84.96	83.1399	88.4356	88.0334
	MSE	0.634	0.553	0.202	0.105	0.523	0.686	0.5	0.773	0.69	0.38	0.1	0.156	0.00009	0.0001
Spectral sub.	SNR	6.2021	1.6788	7.3765	8.7929	4.3004	6.8597	5.0438	6.0702	4.7718	5.2211	5.6177	4.1075	9.6695	10.5495
	SSNR	-3.6007	-2.7192	-4.1359	-1.2131	-8.5368	-8.3104	-8.8860	-6.0537	-8.2997	-6.4008	-7.0117	-8.3968	-4.1883	-3.4027
	PSNR	81.0824	76.0443	86.0940	89.1990	83.1124	85.9178	84.2356	84.5022	83.3487	83.8309	84.1609	82.8624	88.4399	87.8426
	MSE	0.068	0.0016	0.599	0.820	0.176	0.665	0.452	0.306	0.01	0.692	0.495	0.364	0.00009	0.00011
MMSE-SPU	SNR	8.1594	7.1712	5.2378	3.2576	0.6251	3.6146	-0.9314	3.1145	1.1015	2.6363	2.5941	0.0273	4.0124	4.3816
	SSNR	-5.1748	-7.7221	-6.5502	-5.6631	-9.3604	-8.9389	-9.5795	-8.0420	-9.2585	-8.2419	-8.5189	-9.3581	-5.2761	-4.2681
	PSNR	87.3995	86.0951	85.8273	84.549	83.0226	84.5038	82.2636	84.0923	82.7232	83.7877	83.7508	82.4353	79.2641	80.5492
	MSE	1.184	1.599	1.6996	2.281	0.242	0.305	0.861	0.534	0.474	0.719	0.742	0.7115	0.9125	0.2005
p-MMSE	SNR	7.9184	8.1144	4.1124	1.7639	-1.0591	1.9760	-2.7702	1.9048	-0.5084	1.1871	1.2994	-1.8551	5.9054	8.6376
	SSNR	-5.6390	-8.0672	-7.1104	-6.4912	-9.4791	-9.1076	-9.712	-8.3936	-9.4357	-8.7146	-8.8382	-9.5764	-6.4319	-5.4398
	PSNR	86.8119	87.6343	85.2644	83.7319	82.3786	83.7495	81.7202	83.5509	82.2223	83.3436	83.2422	82.4511	86.1896	87.0508
	MSE	1.3549	1.211	1.9348	0.7535	0.7603	0.743	0.3758	0.8707	0.8981	0.012	0.082	0.2143	0.00016	0.00013
log-MMSE	SNR	8.8741	7.1865	5.1508	3.0807	0.5283	3.5837	-1.0396	3.2116	1.0943	2.6625	2.5144	-0.0774	6.9233	9.2578
	SSNR	-5.2455	-7.7531	-6.6371	-5.6733	-9.3867	-8.9658	-9.6058	-8.0630	-9.2906	-8.2502	-8.5724	-9.386	-5.9083	-4.9558
	PSNR	87.4336	86.2445	85.8647	84.5037	83.0542	84.5699	82.3256	84.2611	82.8471	83.9132	83.7983	81.8835	86.8301	87.3594
	MSE	1.1741	1.5439	1.685	2.3052	0.2185	0.2703	0.8065	0.4376	0.3757	0.641	2.7117	0.6981	0.00013	0.00012
Binary Mask (IDBM)	SNR	5.4539	3.1652	2.5948	7.62	1.6363	2.3219	1.5672	1.7176	0.8920	0.3989	1.4361	0.4371	2.1412	1.3016
	SSNR	-5.151	-6.7116	-6.5898	-3.03	-9.1123	-8.94	-9.404	-7.832	-9.2056	-8.1902	-8.4015	-9.252	-4.0967	-3.4831
	PSNR	85.186	80.8695	83.749	87.7694	82.7294	83.56	83.1284	83.0586	82.4048	82.2113	82.8694	82.2378	68.5513	63.2892
	MSE	1.97	0.3227	0.743	1.0868	0.4685	0.8646	0.164	0.2153	0.7377	0.9079	0.3585	0.8842	0.0091	0.0305
WPT-IDBM (Modified Wiener)	SNR	9.2872	10.9254	9.5745	9.0629	8.3289	7.5698	6.7924	6.8978	7.5340	11.5764	7.3500	6.6364	9.7918	11.0772
	SSNR	4.7200	4.0496	4.4544	6.6610	4.1598	4.3175	4.7156	4.1846	4.4811	3.6021	4.5513	4.5943	13.7704	2.1089
	PSNR	87.985	85.1788	84.2235	83.0334	85.2038	85.9984	84.7579	87.7964	88.8260	90.9430	88.9001	89.4527	91.8140	80.6470
	MSE	0.0103	0.00006	0.0049	0.00032	0.0020	0.0016	0.0014	0.0011	0.0017	0.00052	0.0084	0.00037	0.00004	0.00056
WPT Fuzzy	SNR	10.6235	9.1792	9.9869	10.8274	9.8917	10.1734	8.7272	7.5090	8.1377	7.0628	9.1778	8.1635	9.9948	11.6105
	SSNR	2.4167	2.7660	2.6928	2.6618	2.8362	3.2879	1.2127	-0.9134	1.7756	0.3904	2.4875	-3.4373	11.5689	1.8780
	PSNR	89.9188	86.6272	86.8226	82.7472	87.5105	87.7404	86.0095	89.6008	88.9032	91.1215	88.3995	92.6774	90.1364	83.0954
	MSE	0.00013	0.00014	0.00001	0.00035	0.0001	0.00011	0.00017	0.00023	0.00019	0.00025	0.00015	0.00018	0.0002	0.00032
Phase ratio method	SNR	13.6734	11.9547	11.4491	14.2141	10.1347	11.9463	10.3263	10.2331	12.0136	12.7543	11.6926	10.3215	11.2613	11.9536
	SSNR	4.7806	3.73158	4.6125	3.1058	4.4725	4.0452	4.6125	4.4952	4.7125	4.7523	4.6422	4.7652	6.2354	7.5246
	PSNR	95.1163	90.5779	90.3838	93.1371	91.3464	86.3235	91.4206	97.5137	90.8526	91.9056	90.4513	95.8790	94.2165	90.5842
	MSE	0.0019	0.00001	0.00025	0.0007	0.0006	0.0001	0.0003	0.0012	0.0004	0.0002	0.00007	0.00201	0.00003	0.00001

Results for Hindi language speech signals are analyzed in Table A2. Tabulated last three methods give better performance in comparison to existing speech enhancement methods. The phase ratio, WPT fuzzy and WPT modified Wiener method give better performance in all noise types. The minimum error and maximum SNR is obtained by these methods.

Table A3: Analysis for Kannada language

		Non-stationary							Highly non-stationary					Mixed noise	Combined
Methods	Measures	White	Volvo-car	Pink	Machine-gun	Tank	Leopard vehicle	Hf-channel	Factory Floor1	Buccaneer2	Babble	F-16 plane	Destroyer operations	F16+Babble+NOIZEUS	Reverberation+Babble
Wiener filtering method	SNR	7.6017	3.5582	7.6576	7.6622	3.7207	5.9474	2.5359	5.7299	4.8220	3.7551	5.0498	3.4910	8.1956	9.5624
	SSNR	-3.5397	-5.5023	-4.5446	-2.2741	-8.7985	-8.7054	-9.3119	-6.3873	-8.3361	-7.5606	-7.2340	-8.4978	-3.2489	-2.3937
	PSNR	92.2779	80.7037	87.2625	87.8498	83.8220	85.9774	82.9528	85.3731	85.0380	83.2968	84.7012	83.2398	87.3231	87.0917
	MSE	0.38485	5.5298	1.2213	1.0669	2.6970	1.6419	3.2946	1.8870	2.0384	3.0437	2.2027	3.0839	0.00012	0.00013
Spectral sub.	SNR	6.2556	1.5567	6.8950	8.9412	3.7333	6.1608	3.3367	5.3217	5.3005	3.8123	4.9478	3.7601	8.4248	9.4716
	SSNR	-3.3423	-1.8671	-3.7156	-1.3608	-8.5950	-8.6361	-9.1357	-5.8031	-7.9776	-7.4597	-6.7371	-8.3921	-2.9517	-2.1880
	PSNR	92.7650	75.9668	85.7935	88.7581	82.9866	85.8404	82.9634	84.0790	84.5719	82.9892	83.7039	82.8544	87.1937	86.7558
	MSE	0.34401	.0016	1.7129	0.86552	3.2691	1.6945	3.2866	2.5420	2.2693	3.2671	2.7713	3.3701	0.00013	0.00014
MMSE-SPU	SNR	6.7167	7.2953	5.5244	4.5052	0.2638	3.0362	-2.8847	2.4548	0.6740	1.3149	1.9604	0.8017	1.0264	4.02647
	SSNR	-4.1486	-72391	-6.1169	-4.3747	-9.4416	-9.2076	-9.8242	-8.0730	-9.2351	-8.6270	-8.4698	-9.3221	-2.1284	-1.25471
	PSNR	89.7986	86.4134	86.4361	85.4911	83.0486	84.4787	81.6990	84.2971	83.4242	82.8166	83.8600	83.0283	81.5524	80.5894
	MSE	0.68112	1.4851	1.4773	1.8364	3.2227	2.3185	4.3973	2.4175	2.9557	3.3995	2.6735	3.2378	1.6548	0.02584
p-MMSE	SNR	3.8459	8.5848	4.2253	3.4900	-1.2864	1.8426	-5.5951	1.0315	-1.0225	0.1364	0.5335	-0.6523	4.6857	7.8178
	SSNR	-4.4832	-7.6229	-6.7727	-5.3279	-9.5828	-9.3802	-9.9002	-8.5046	-9.4467	-8.9986	-8.8132	-9.5003	-5.5187	-4.2061
	PSNR	89.2921	88.3171	85.8015	84.9571	82.5828	84.0807	81.2339	83.7667	82.7996	82.6714	83.4128	82.7112	85.4535	86.5357
	MSE	0.76536	0.95802	1.7098	2.0767	3.5876	2.5410	4.8943	2.7316	3.4129	3.5151	2.9635	3.4831	0.00019	0.00014
log-MMSE	SNR	4.6569	7.3857	5.4695	4.4795	0.2253	3.1194	-2.8426	2.5106	0.6179	1.4515	1.9876	0.7807	5.9256	8.4662
	SSNR	-4.2115	-7.2970	-6.1590	-4.2846	-9.4587	-9.2260	-9.8240	-8.0759	-9.2728	-8.6207	-8.4999	-9.3395	-4.6648	-3.5417
	PSNR	89.8575	86.6451	86.4910	85.5617	83.1091	84.6659	81.8068	84.4219	83.5054	82.9856	83.9707	83.0739	86.1397	86.7712
	MSE	6.7194	1.4079	1.4588	1.8068	3.1781	2.2207	4.2895	2.3491	2.9009	3.2698	2.6062	3.2040	0.00016	0.00014
Binary Mask (IDBM)	SNR	3.0841	1.5097	2.8283	3.3274	2.6367	2.7691	2.5399	2.6822	2.6616	2.4880	2.6119	2.4248	2.0154	1.3566
	SSNR	1.0522	0.4220	1.7874	1.8841	1.2618	1.9280	1.6606	3.1854	1.0906	1.6014	1.7804	1.2665	3.8815	2.6281
	PSNR	73.5976	69.6894	73.8160	73.9660	74.4778	73.7856	75.1243	73.7635	73.7463	75.5090	74.0131	74.8076	68.6602	64.3106
	MSE	0.0028	0.0070	0.0027	0.0026	0.0023	0.0027	0.0020	0.0020	0.0027	0.0018	0.0026	0.0021	0.0089	0.0241
WPTIDB M (Modified Wiener)	SNR	4.6687	1.7570	3.9168	4.6953	8.3292	3.1542	9.2608	5.8040	6.7281	4.3584	11.0907	9.5649	7.6888	3.5625
	SSNR	3.9564	3.2950	3.5566	4.5560	3.4717	3.4227	3.4667	3.2439	3.6438	2.9957	3.5192	3.2642	2.8015	4.4039
	PSNR	82.3113	76.3803	81.1375	82.3353	87.1359	79.7432	88.1600	83.8828	85.1592	81.8111	90.1764	88.5960	85.9215	79.3749
	MSE	0.00004	0.0015	0.00005	0.00038	0.000126	0.00069	0.00099	0.00027	0.0002	0.0002	0.00043	0.00063	0.00009	0.00017
WPT Fuzzy	SNR	9.0267	10.1691	8.3870	6.5743	9.0643	9.4675	7.6100	6.4066	7.0897	6.7260	11.3260	10.0058	7.3448	8.3019
	SSNR	3.2892	3.1145	3.5144	3.0471	3.7451	3.9438	2.1116	2.0467	2.5803	0.4068	3.0762	4.1573	3.2002	5.1651
	PSNR	87.6504	88.6492	87.1092	82.5158	88.0233	88.4981	86.0830	84.7418	85.4955	84.9510	86.8346	85.9967	85.5591	84.0702
	MSE	0.00012	0.000089	0.00013	0.00036	0.00011	0.000092	0.00016	0.00022	0.00018	0.00021	0.00013	0.00016	0.00018	0.00026
Phase ratio method	SNR	9.8742	12.1852	12.4792	12.1752	12.8676	13.5163	9.8346	12.0837	8.5279	11.8455	12.593	11.5371	9.87461	10.54926
	SSNR	4.3155	3.28721	4.1658	6.2414	4.0156	3.9454	4.2625	4.0824	4.3045	3.6847	4.2647	4.1924	5.21064	5.94152
	PSNR	88.0207	97.5283	91.9634	94.6687	92.1754	92.9957	89.0516	91.4565	78.6791	91.2506	91.881	91.1173	90.54123	85.04697
	MSE	0.0007	0.00008	0.00005	0.0003	0.00004	0.00002	0.0005	0.00007	0.0002	0.00005	0.0007	0.00004	0.00001	0.000214

Table A3 shows the results for Kannada language speech. All the given speech enhancement methods are analyzed in above given noise types. The phase ratio based method gives better performance in mixed and reverberated noise also. The maximum quality and intelligibility of speech signal is obtained by last three speech enhancement methods.

Table A4: Analysis for Bengali language

		Non-stationary							Highly non-stationary					Mixed noise	Combined
Methods	Measures	White	Volvo car	Pink	Machnegun	Tank	Leopard vehicle	Hf-channel	Factory Floor1	Buccaneer2	Babble	F16 plane	Destroyer operation	F16+Babble +NOIZEUS	Reverberation +Babble
Wiener filtering	SNR	9.7341	8.4188	7.8274	8.4748	4.6130	3.9684	4.0583	3.8061	4.7679	1.9819	3.6708	2.6353	2.0547	0.9762
	SSNR	1.9598	3.3679	3.5685	2.0218	6.6965	6.6676	7.1603	5.3175	6.4313	6.5931	5.7850	6.8042	5.2215	4.1215
	PSNR	80.0612	78.2964	78.4951	79.4160	75.9674	75.5827	75.7770	75.5336	76.2563	74.3354	75.4199	74.7088	59.0757	51.4213
	MSE	6.4116	9.6258	9.1954	8.1082	.0016	.0018	.0017	.0017	.0015	.0024	.0019	.0022	0.0804	0.4688
Spectral sub.	SNR	7.44551	8.3623	6.8580	2.3727	7.8273	6.1189	5.8123	5.0427	5.4394	4.6330	6.2720	5.0987	5.3044	6.7434
	SSNR	4.9629	4.9172	5.3151	4.7374	5.4912	5.8794	3.7527	3.3987	3.9547	1.7579	4.8912	5.8831	2.7263	1.8538
	PSNR	78.1326	78.9104	77.6812	73.1291	78.9769	77.5363	76.4736	75.1817	75.8308	74.5617	77.0122	76.4407	75.4190	72.5242
	MSE	0.00099	0.00084	0.0011	0.0032	0.00082	0.0011	0.0015	0.0020	0.0017	0.0023	0.0013	0.0015	0.0019	0.0036
MMSE-SPU	SNR	12.7022	12.0503	9.7935	8.0874	6.2262	6.5570	5.7363	8.6154	7.6350	7.1692	7.6502	4.9919	6.25849	5.29712
	SSNR	1.6627	4.3531	2.7753	3.0886	6.6188	6.2260	6.7537	4.3427	5.7593	4.6654	5.0310	6.6686	5.04691	4.68452
	PSNR	82.8180	82.0668	81.0626	78.6336	77.3567	77.4294	76.9380	79.1817	78.3469	77.6973	78.3967	76.3284	79.0541	80.2791
	MSE	3.3985	4.0402	5.0913	8.9068	0.0090	0.0012	0.0013	0.0013	9.5146	0.0011	9.4062	0.0015	0.00481	0.00561
p-MMSE	SNR	11.7313	11.7236	9.8611	7.1107	4.8546	5.6459	4.6536	7.4847	6.6592	6.2960	6.6497	3.8000	5.3869	8.3227
	SSNR	2.0948	4.8675	3.3619	3.9812	7.1398	6.6370	7.1403	5.0209	6.1725	5.3846	5.5763	7.1593	6.7063	4.5253
	PSNR	82.1312	82.0749	80.4619	77.9623	76.5364	76.9374	76.3735	78.4828	77.8284	77.3511	77.7946	75.6644	76.6621	75.2851
	MSE	3.987	4.0326	5.8464	.0010	.0014	.0013	.0015	.0015	.0011	.0012	.0011	.0018	0.0014	0.0019
log-MMSE	SNR	12.6737	12.1029	9.7501	8.0042	6.1772	6.5111	5.6043	8.5388	7.5599	7.1062	7.5834	4.8756	6.4236	8.3449
	SSNR	1.5991	4.4005	2.7020	2.8932	6.6465	6.2355	6.7993	4.3056	5.8063	4.6545	5.0315	6.6903	6.2155	4.1378
	PSNR	82.9003	82.2371	81.1354	78.6043	77.4170	77.4962	76.9242	79.2206	78.4008	77.7414	78.4347	76.3255	77.2707	75.0799
	MSE	3.3347	3.8848	5.0066	8.9671	.0012	.0012	.0013	.0013	9.3972	.0011	9.3243	.0015	0.0012	0.0020
Binary Mask (IDBM)	SNR	10.4650	13.6724	3.2584	7.5248	5.7311	1.4186	5.0586	7.6572	7.6724	6.3833	10.7094	7.1235	6.8069	4.2099
	SSNR	5.1410	4.3587	4.5534	4.4899	4.2819	4.3638	4.7878	4.3468	4.3587	3.8701	4.5521	4.2496	4.4816	0.3313
	PSNR	79.7225	83.1439	70.1273	76.4174	74.0503	65.7799	73.1795	79.9922	83.1439	74.9434	80.0734	78.3366	80.1240	77.7378
	MSE	0.00069	0.00032	0.0063	0.0015	0.0026	0.0172	0.0031	0.00065	0.00032	0.0021	0.000064	0.000954	0.00063	0.0109
WPT IDBM (Modified Wiener)	SNR	13.0217	10.8079	13.992	11.2802	9.7994	9.0419	9.1173	10.4611	8.4486	9.4090	10.7019	8.2558	8.5681	8.3790
	SSNR	0.3080	2.7085	1.0576	0.9435	5.3564	5.1690	5.8790	2.7444	4.7711	3.1505	3.3124	4.8539	5.0798	3.5439
	PSNR	86.9002	80.3661	83.8767	81.3579	79.8370	79.2191	79.3463	81.4422	80.4089	79.1612	80.6744	78.4115	78.4290	79.6531
	MSE	0.0084	0.0064	0.0084	0.0087	0.0097	0.0079	0.0079	0.0067	0.0067	0.0057	0.00584	0.0068	0.00094	0.0022
WPT Fuzzy	SNR	15.7162	15.3500	12.848	12.0569	11.951	10.8690	9.6597	11.2030	9.2335	11.2308	12.2646	11.3238	9.214854	9.15489
	SSNR	5.4588	4.4222	5.4325	4.48236	5.3915	5.05562	5.40125	5.25525	5.4445	5.20532	5.3955	5.38287	6.1542	4.58136
	PSNR	85.3509	84.8811	82.5017	77.1334	82.5979	80.5967	80.1769	81.8385	82.9560	80.7697	81.8966	81.1197	86.2843	85.2641
	MSE	0.00045	0.0045	0.0005	0.0013	0.0006	0.0008	0.0002	0.00006	0.00005	0.00009	0.00005	0.0004	0.0001	0.0021
Phase ratio method	SNR	19.5622	17.9352	14.327	12.4930	12.273	11.3285	10.879	11.8780	10.8993	12.5141	10.9248	13.7366	10.8601	10.3655
	SSNR	6.01847	5.08051	6.0506	.1476	5.0142	5.0486	5.6381	2.0327	4.491	2.6183	2.7650	4.5139	7.7347	3.4120
	PSNR	89.3933	75.8323	84.0339	82.3401	79.8844	79.2745	80.0348	81.5503	80.5324	79.0098	80.5671	78.4049	88.3859	89.5645
	MSE	0.00008	0.0001	0.0005	0.00004	0.0004	0.00001	0.0004	0.0001	0.00004	0.000561	0.00025	0.00001	0.00094	0.0023

Results for Bengali language speech signals are analyzed in Table A4. In this table last three methods give better performance in comparison to existing speech enhancement methods in all noise conditions. Since minimum error is obtained by WPT Fuzzy and phase ratio based methods hence these methods give good quality and intelligibility of speech signal

Table A5. Analysis for Malayalam language

Methods	Measures	Non-stationary							Highly non-stationary					Mixed noise	Combined
		White	Volvo car	Pink	Machinegun	Tank	Leopard vehicle	HF channel	Factory Floor1	Buccaneer2	Babble	F16 plane	Destroyer operation	F16+Babble+NOIZEUS	Reverberation +Babble
Wiener filtering	SNR	8.3386	3.2876	7.2644	7.8039	4.0973	6.7857	3.3499	5.0439	4.8080	4.2180	4.5010	3.8734	7.7053	9.0658
	SSNR	4.6769	6.2425	5.7720	3.9451	9.0807	8.9841	9.2874	7.3608	8.7840	8.0084	7.8877	8.9312	4.5498	3.8215
	PSNR	91.6207	81.0026	87.4165	88.3581	84.4301	87.1963	84.1213	85.3067	85.4129	84.5175	84.8072	84.1318	87.4803	87.3904
	MSE	0.44772	5.1625	1.1788	0.94901	2.3446	1.2401	2.5174	1.9161	1.8698	2.2979	2.1496	2.5113	0.00012	0.00012
Spectral sub.	SNR	7.7486	1.3680	6.2793	8.4605	3.9340	7.0537	3.7774	4.5131	4.8290	4.2112	4.2203	3.7824	7.8240	9.0700
	SSNR	4.3007	2.3181	4.7470	3.3257	9.0150	8.9744	9.2200	6.9154	8.6776	7.8795	7.6468	8.8755	4.2138	3.6692
	PSNR	91.8701	76.0964	85.8430	88.9363	83.7766	87.3698	84.2692	84.0852	85.0149	84.1180	83.8672	83.6190	87.4120	87.3162
	MSE	0.42274	.0016	1.6935	0.83071	2.7253	1.1915	2.5411	2.5384	2.0492	2.5193	2.6691	2.8260	0.00012	0.00012
MMSE-SPU	SNR	8.6398	6.6394	5.5160	6.3220	2.3786	5.1980	0.4021	3.4174	2.9535	2.4505	2.2716	2.4264	1.26714	3.28947
	SSNR	5.8290	8.1695	7.4255	5.1739	9.3648	9.1193	9.5583	8.4954	9.1926	8.7731	8.9426	9.2985	3.02385	3.4569
	PSNR	89.3776	86.5319	86.7033	87.1791	84.1693	86.0848	83.2458	85.0379	84.6837	83.9507	84.3670	84.1316	80.2.184	81.6412
	MSE	0.75045	1.4451	1.3892	1.2450	2.4897	1.6018	3.0797	2.0384	2.2116	2.6183	2.3789	2.5114	2.62554	1.02641
p-MMSE	SNR	7.5711	7.5424	4.4702	5.7538	1.3242	4.7252	0.8780	2.5042	2.0444	1.7397	1.0425	1.6149	5.0644	7.2603
	SSNR	6.3172	8.6295	7.8535	5.9717	9.4890	9.2440	9.6286	8.7644	9.3045	9.0084	9.1682	9.4003	6.8158	6.0793
	PSNR	88.6707	88.1090	86.1926	86.9980	83.9946	86.1716	82.9375	84.7784	84.4532	84.0065	83.9764	84.1129	86.0549	86.5449
	MSE	0.88311	1.0050	1.5625	1.2980	2.5919	1.5701	3.4796	2.1639	2.3322	2.5848	2.6028	2.5223	0.00016	0.00014
log-MMSE	SNR	8.5809	6.7332	5.4442	6.3351	2.2902	5.3373	0.2653	3.3448	2.8612	2.5132	2.1305	2.3371	6.0248	7.9459
	SSNR	5.8920	8.1852	7.4692	5.1548	9.3851	9.1443	9.5735	8.5205	9.2215	8.7968	8.9694	9.3205	6.2115	5.4154
	PSNR	89.4319	86.7463	86.7601	87.3240	84.2194	86.3925	83.3104	85.0757	84.7497	84.1466	84.3718	84.1911	86.6071	86.9083
	MSE	0.74112	1.3755	1.3711	1.2042	2.4619	1.4922	3.0342	2.0207	2.1783	2.5028	2.3763	2.4773	0.00014	0.00013
Binary Mask (IDBM)	SNR	9.1882	7.6922	7.1543	4.8812	0.3543	3.0906	4.2987	3.0804	5.0803	5.0683	7.5815	6.9937	2.012647	3.2167
	SSNR	8.4565	7.5755	8.8555	6.3189	4.67845	4.47423	4.94521	8.70548	9.99552	8.5845	8.8358	9.84545	6.5987	6.2159
	PSNR	88.7957	101.674	92.2607	83.3720	69.5661	80.6121	89.0594	92.1407	63.0437	82.5171	81.1379	82.0707	80.5791	80.5492
	MSE	0.85804	0.044228	0.38638	2.9915	0.0072	5.6477	0.80749	0.39720	0.0323	0.0364	0.2569	0.40365	0.138425	0.0541
WPT IDBM (Modified Wiener)	SNR	9.7046	7.5657	7.8276	6.0655	7.5683	7.1296	6.5329	4.9588	8.9836	5.8393	8.0033	7.4408	2.1555	6.4124
	SSNR	12.9368	10.0749	8.1453	6.4178	6.2014	7.0760	7.7107	8.8178	10.7923	9.8429	9.2661	9.1695	7.8682	6.3302
	PSNR	76.8685	71.010	76.3485	80.8055	80.6173	76.6941	75.9748	76.2272	76.0937	85.4066	87.0553	86.0568	89.7872	85.2939
	MSE	0.0013	0.0052	0.0015	0.0027	0.0018	0.0014	0.0016	0.0016	0.0016	0.0016	0.0013	0.0016	0.0068	0.0192
WPT Fuzzy	SNR	10.3962	8.6010	9.1027	7.9547	7.9385	8.5263	6.8505	5.0663	10.0125	8.2008	8.2886	7.7024	8.6353	7.2343
	SSNR	11.4036	11.5683	10.8277	8.7720	8.8114	9.8257	9.9731	9.6539	10.0624	10.2539	9.8434	90.0154	7.4011	7.7125
	PSNR	82.4479	85.5104	88.5480	81.0608	81.4980	81.2010	85.2336	61.9647	84.6414	86.9602	82.3431	76.9894	87.7181	88.3249
	MSE	0.0004	0.00018	0.00091	0.0509	0.0730	0.0049	0.000195	0.0414	0.2233	0.0131	0.00038	0.0013	0.00011	0.0191
Phase ratio method	SNR	11.9755	9.2960	10.3089	10.1921	8.8545	9.3128	7.8605	6.2820	10.4491	9.2709	10.0469	94.5501	10.7794	9.9710
	SSNR	14.8700	14.8870	13.5666	4.6198	12.8472	10.8129	10.5201	11.8945	11.4382	11.8345	9.0205	9.8514	8.6432	8.6603
	PSNR	88.4414	88.6441	90.8932	83.4884	88.8264	89.1323	87.1748	85.4759	86.019	90.4264	87.6660	87.2031	96.0606	94.8282
	MSE	0.00009	0.000089	0.00011	0.00029	0.00008	0.000079	0.000013	0.00018	0.00017	0.00019	0.00011	0.00013	0.00016	0.00022

Table A5 shows the results obtained by single-channel speech enhancement methods for Malayalam language. These results are obtained in various noise types and results are obtained in quality and intelligibility measure parameters. The last three speech enhancement methods give better quality and intelligibility of the speech signal in all noise conditions.

**APPENDIX - B**

**Table B1: Speech Intelligibility Evaluation for Hindi Language**

Methods	Measures	Babble	Airport	Car	Exhibition	Restaurant	White	Pink	Railway Platform	
SS Boll	MOS	0.3350	0.3260	0.3498	0.3964	0.3290	0.3444	0.3342	0.3901	
	SII	0.2248	0.1982	0.1870	0.4894	0.2286	0.3169	0.1926	0.3519	
	STOI	0.8234	0.7796	0.8024	0.8436	0.7922	0.8181	0.7948	0.7961	
	MULTI-BAND SS	MOS	0.2669	0.2883	0.3142	0.2674	0.2596	0.2943	0.2989	0.3670
		SII	0.4135	0.3529	0.3583	0.3524	0.3257	0.2087	0.2427	0.3581
		STOI	0.7474	0.7269	0.7475	0.6935	0.7141	0.7185	0.7457	0.8053
Berouti SS	MOS	0.2621	0.2809	0.2909	0.2606	0.2566	0.2897	0.2894	0.3614	
	SII	0.4184	0.3577	0.3644	0.3923	0.3523	0.2499	0.2687	0.3692	
	STOI	0.7571	0.7459	0.7668	0.7327	0.7392	0.7600	0.7716	0.8284	
Parametric SS	MOS	0.2688	0.2892	0.3260	0.2807	0.2594	0.3175	0.3201	0.3686	
	SII	0.4289	0.3733	0.3350	0.3881	0.3735	0.3173	0.2857	0.3568	
	STOI	0.7348	0.7035	0.7049	0.6897	0.6669	0.7096	0.7092	0.7668	
Scalart SS	MOS	0.2667	0.2893	0.3152	0.2845	0.2699	0.3444	0.3428	0.3780	
	SII	0.4381	0.3792	0.3636	0.4873	0.4214	0.4131	0.3311	0.3895	
	STOI	0.7630	0.7486	0.7629	0.7553	0.7400	0.7961	0.7884	0.8363	
WIENER	MOS	0.3784	0.3459	0.3477	0.3683	0.3338	0.3490	0.3663	0.3711	
	SII	0.1595	0.1620	0.1721	0.3487	0.1810	0.2483	0.1753	0.1793	
	STOI	0.8390	0.8393	0.8408	0.8496	0.8185	0.8484	0.8480	0.8548	
MMSE	MOS	0.3958	0.3716	0.3831	0.4198	0.3450	0.3541	0.3746	0.4270	
	SII	0.1945	0.1694	0.1603	0.3934	0.1960	0.2775	0.1858	0.1861	
	STOI	0.8515	0.8397	0.8540	0.8773	0.8232	0.8414	0.8500	0.8670	
ICS	MOS	0.2195	0.2191	0.2157	0.2232	0.2257	0.2159	0.2188	0.2153	
	SII	0.0251	0.0370	0.0276	0.0308	0.0399	0.0249	0.0264	0.0587	
	STOI	0.4639	0.4447	0.4926	0.4888	0.4633	0.4446	0.4844	0.4521	
IdBM	MOS	0.6082	0.6082	0.5875	0.6035	0.6123	0.6013	0.6046	0.6539	
	SII	0.4538	0.3867	0.3909	0.7030	0.4819	0.5244	0.4027	0.4034	
	STOI	0.9605	0.9456	0.9418	0.9641	0.9602	0.9655	0.9539	0.9545	
WPT Modified Wiener	MOS	0.5016	0.5165	0.4910	0.6161	0.5512	0.5966	0.5209	0.5330	
	SII	0.4385	0.3811	0.3761	0.6990	0.4808	0.5155	0.3937	0.3927	
	STOI	0.9600	0.9487	0.9345	0.9701	0.9590	0.9625	0.9491	0.9391	
WPT FUZZY	MOS	0.2901	0.3152	0.3152	0.3144	0.2900	0.3946	0.3891	0.3618	
	SII	0.3882	0.3593	0.3593	0.5705	0.4286	0.4844	0.3675	0.3566	
	STOI	0.7474	0.7482	0.7482	0.7952	0.7680	0.8331	0.8228	0.8152	
WPT FUZZY modified Wiener	MOS	0.2686	0.2877	0.2973	0.3012	0.2662	0.3842	0.3401	0.3588	
	SII	0.3191	0.2673	0.2667	0.5589	0.3809	0.4449	0.2804	0.2556	
	STOI	0.7750	0.7588	0.8156	0.8181	0.7831	0.8742	0.8372	0.8529	
Phase ratio based method	MOS	0.2594	0.2795	0.2916	0.3064	0.2646	0.3262	0.2915	0.2996	
	SII	0.3615	0.3080	0.3095	0.5731	0.4085	0.4876	0.3256	0.2925	
	STOI	0.7926	0.7731	0.7970	0.7994	0.7855	0.8263	0.8169	0.8122	

Table B2: Speech Intelligibility Evaluation for Kannada Language

Methods	Measures	Babble	Airport	Car	Exhibition	Restaurant	White	Pink	Railway Platform
SS Boll	MOS	0.2923	0.2509	0.2621	0.2431	0.2792	0.2627	0.2728	0.3668
	SII	0.2256	0.1323	0.0684	0.1641	0.3752	0.0888	0.0666	0.2399
	STOI	0.6589	0.6331	0.6350	0.6781	0.7129	0.6871	0.7291	0.7777
MULTI-BAND SS	MOS	0.3233	0.3256	0.3371	0.3013	0.3178	0.3448	0.3666	0.4000
	SII	0.2090	0.2104	0.1759	0.2866	0.2280	0.2017	0.1892	0.2066
	STOI	0.6579	0.6716	0.6540	0.6573	0.6355	0.6788	0.6880	0.7268
Berouti SS	MOS	0.3169	0.3320	0.3438	0.3038	0.3206	0.3422	0.3754	0.4214
	SII	0.2554	0.2559	0.2180	0.3530	0.2764	0.2626	0.2494	0.2505
	STOI	0.6838	0.7016	0.6944	0.6934	0.6746	0.7119	0.7138	0.7806
Parametric SS	MOS	0.3016	0.2917	0.3067	0.2784	0.2985	0.2993	0.3216	0.4005
	SII	0.1576	0.1672	0.1189	0.2071	0.4114	0.1254	0.1029	0.2965
	STOI	0.5988	0.6578	0.6562	0.7004	0.7031	0.7119	0.7412	0.7694
Scalart SS	MOS	0.3034	0.3260	0.3390	0.3118	0.3265	0.3306	0.3500	0.4419
	SII	0.2796	0.3327	0.2867	0.4449	0.4515	0.3685	0.3278	0.3744
	STOI	0.6585	0.7353	0.7369	0.7338	0.7354	0.7541	0.7665	0.8327
WIENER	MOS	0.4051	0.3863	0.3992	0.4092	0.3854	0.4111	0.4148	0.4485
	SII	0.2064	0.1709	0.1565	0.3544	0.1912	0.2264	0.2090	0.2054
	STOI	0.8705	0.8578	0.8452	0.8604	0.8429	0.8537	0.8606	0.8760
MMSE	MOS	0.4432	0.4027	0.4629	0.4591	0.4236	0.4715	0.4612	0.5052
	SII	0.2179	0.1952	0.1755	0.3618	0.2116	0.2986	0.2224	0.2144
	STOI	0.8679	0.8547	0.8597	0.8823	0.8612	0.8669	0.8705	0.8757
ICS	MOS	0.2288	0.2189	0.2241	0.2198	0.2185	0.2188	0.2175	0.2139
	SII	0.0189	0.0422	0.0407	0.0379	0.0315	0.0217	0.0416	0.0311
	STOI	0.5337	0.5300	0.5768	0.4677	0.5122	0.4835	0.5314	0.5291
IdBM	MOS	0.6450	0.6650	0.6438	0.6284	0.6507	0.6566	0.6448	0.7072
	SII	0.4448	0.4293	0.3803	0.8058	0.4806	0.5423	0.4626	0.4107
	STOI	0.9519	0.9490	0.9439	0.9569	0.9550	0.9471	0.9460	0.9554
WPT IdBM	MOS	0.5215	0.5625	0.5550	0.6410	0.5558	0.6702	0.5958	0.5848
	SII	0.4244	0.4097	0.3628	0.7766	0.4631	0.5266	0.4422	0.3934
	STOI	0.9568	0.9430	0.9398	0.9659	0.9546	0.9590	0.9486	0.9480
WPT FUZZY	MOS	0.3442	0.3577	0.3879	0.3335	0.3358	0.3670	0.4307	0.4449
	SII	0.3983	0.3912	0.3389	0.5590	0.4384	0.5124	0.4203	0.3654
	STOI	0.7444	0.7618	0.7665	0.7798	0.7456	0.8027	0.8061	0.8117
WPT FUZZY modified Wiener	MOS	0.3329	0.3561	0.3814	0.3539	0.2973	0.4567	0.4411	0.4344
	SII	0.2937	0.2996	0.2733	0.5133	0.3982	0.4742	0.3244	0.2743
	STOI	0.7685	0.7771	0.8012	0.8005	0.7509	0.8467	0.8377	0.8444
Phase ratio based method	MOS	0.3108	0.3469	0.3416	0.3609	0.2928	0.4115	0.3774	0.3506
	SII	0.3657	0.3662	0.3224	0.5271	0.4367	0.5027	0.3881	0.3215
	STOI	0.8149	0.8081	0.8165	0.8215	0.7803	0.8570	0.8424	0.8399

Table B3: Speech Intelligibility Evaluation for Bengali Language

Methods	Measures	Babble	Airport	Car	Exhibition	Restaurant	White	Pink	Railway Platform
SS Boll	MOS	0.3052	0.2710	0.2653	0.2607	0.2753	0.2741	0.2738	0.2981
	SII	0.3781	0.1562	0.1044	0.2339	0.2194	0.0495	0.0479	0.1396
	STOI	0.8024	0.7134	0.7996	0.6867	0.6663	0.8342	0.8328	0.7498
MULTI-BAND SS	MOS	0.3174	0.3383	0.3540	0.3316	0.3101	0.3440	0.3875	0.4254
	SII	0.2787	0.2040	0.1887	0.2994	0.2042	0.1945	0.1980	0.2057
	STOI	0.7558	0.7406	0.7929	0.7462	0.7016	0.7586	0.7922	0.8253
Berouti SS	MOS	0.3046	0.3223	0.3374	0.3104	0.3022	0.3427	0.3730	0.4171
	SII	0.2944	0.2194	0.2081	0.3321	0.2311	0.2250	0.2258	0.2196
	STOI	0.7790	0.7671	0.8210	0.7649	0.7386	0.7968	0.8301	0.8598
Parametric SS	MOS	0.3083	0.2998	0.3248	0.3033	0.2743	0.3311	0.3405	0.3761
	SII	0.3960	0.1689	0.1063	0.2976	0.2664	0.0596	0.0548	0.1477
	STOI	0.7644	0.7083	0.8126	0.7098	0.6303	0.8151	0.8232	0.7555
Scalart SS	MOS	0.3161	0.3008	0.3087	0.2966	0.2795	0.2970	0.3220	0.3894
	SII	0.4020	0.2465	0.2207	0.4040	0.3425	0.2748	0.2345	0.2384
	STOI	0.7951	0.7673	0.8246	0.7873	0.7481	0.8275	0.8541	0.8504
WIENER	MOS	0.3751	0.4006	0.4140	0.4277	0.4162	0.4106	0.3722	0.4260
	SII	0.1704	0.1224	0.1234	0.2509	0.1664	0.1747	0.1239	0.1277
	STOI	0.8793	0.8781	0.8791	0.8599	0.8718	0.8675	0.8759	0.8871
MMSE	MOS	0.4380	0.4493	0.4395	0.4937	0.4431	0.4550	0.4218	0.4840
	SII	0.1867	0.1371	0.1219	0.3290	0.1929	0.2335	0.1524	0.1369
	STOI	0.8838	0.8819	0.8706	0.9106	0.8733	0.8920	0.8878	0.8849
ICS	MOS	0.2214	0.2177	0.2197	0.2206	0.2201	0.2176	0.2241	0.2141
	SII	0.0200	0.0157	0.0132	0.0378	0.0254	0.0098	0.0118	0.0137
	STOI	0.4483	0.5075	0.5115	0.4360	0.4703	0.4426	0.5142	0.5187
IdBM	MOS	0.4613	0.4637	0.4577	0.4636	0.4699	0.4705	0.4637	0.4671
	SII	0.4112	0.3197	0.3046	0.6472	0.4244	0.4444	0.3511	0.3055
	STOI	0.9757	0.9686	0.9665	0.9705	0.9732	0.9714	0.9705	0.9660
WPT IdBM	MOS	0.6742	0.6994	0.6749	0.7818	0.7012	0.7758	0.7170	0.6859
	SII	0.3975	0.3024	0.2920	0.6262	0.4079	0.4309	0.3367	0.2930
	STOI	0.9745	0.9685	0.9648	0.9793	0.9740	0.9809	0.9733	0.9595
WPT FUZZY	MOS	0.3046	0.3170	0.3300	0.3034	0.2798	0.3184	0.3539	0.3791
	SII	0.3780	0.2819	0.2666	0.4660	0.3780	0.4084	0.3136	0.2655
	STOI	0.7935	0.7944	0.8440	0.8115	0.7893	0.8680	0.8818	0.8644
WPT FUZZY modified Wiener	MOS	0.3043	0.3238	0.3468	0.3414	0.2577	0.4367	0.4360	0.4451
	SII	0.2835	0.2370	0.2477	0.4392	0.3277	0.3795	0.2633	0.2351
	STOI	0.8171	0.8068	0.8663	0.8281	0.8114	0.8823	0.8880	0.8977
Phase ratio based method	MOS	0.3031	0.3160	0.3397	0.3403	0.2505	0.4316	0.3590	0.3371
	SII	0.3294	0.2688	0.2714	0.4575	0.3685	0.4205	0.3046	0.2589
	STOI	0.8282	0.8192	0.8568	0.8252	0.8156	0.8741	0.8678	0.8585

Table B4: Speech Intelligibility Evaluation for Malayalam Language

Methods	Measures	Babble	Airport	Car	Exhibition	Restaurant	White	Pink	Railway Platform
SS Boll	MOS	0.3020	0.2771	0.2656	0.2743	0.2903	0.2765	0.2693	0.3325
	SII	0.3626	0.2585	0.1167	0.2628	0.3029	0.0289	0.1119	0.2639
	STOI	0.7122	0.6964	0.7169	0.7409	0.6877	0.8042	0.7598	0.7263
MULTI-BAND SS	MOS	0.3137	0.3134	0.3401	0.2951	0.3129	0.3185	0.3383	0.3941
	SII	0.2441	0.1714	0.1681	0.2509	0.2144	0.2024	0.1608	0.2300
	STOI	0.6216	0.5941	0.6490	0.6366	0.6254	0.6755	0.6578	0.6964
Berouti SS	MOS	0.3127	0.3215	0.3364	0.3044	0.3106	0.3326	0.3596	0.3966
	SII	0.2852	0.2255	0.2066	0.3213	0.2689	0.2639	0.2092	0.2785
	STOI	0.6633	0.6486	0.6997	0.6840	0.6768	0.7320	0.7168	0.7503
Parametric SS	MOS	0.3035	0.2964	0.3042	0.2968	0.2871	0.3071	0.3133	0.3745
	SII	0.3548	0.2707	0.1176	0.2621	0.3297	0.0272	0.1223	0.2742
	STOI	0.6312	0.6884	0.7185	0.6897	0.6687	0.8117	0.7585	0.7389
Scalart SS	MOS	0.3178	0.3242	0.3335	0.2973	0.3106	0.3039	0.3566	0.4170
	SII	0.4189	0.3597	0.2723	0.4982	0.4280	0.3505	0.3086	0.3708
	STOI	0.7113	0.7234	0.7521	0.7647	0.7182	0.8088	0.7954	0.7809
WIENER	MOS	0.3834	0.4146	0.4149	0.4273	0.4471	0.4331	0.3832	0.4356
	SII	0.1919	0.1808	0.1676	0.3077	0.2163	0.2598	0.1519	0.2180
	STOI	0.8587	0.8591	0.8621	0.8519	0.8602	0.8611	0.8486	0.8647
MMSE	MOS	0.4292	0.4441	0.4846	0.4784	0.4514	0.4814	0.4490	0.4824
	SII	0.1829	0.1799	0.1821	0.3596	0.2158	0.2885	0.1878	0.2240
	STOI	0.8637	0.8545	0.8409	0.8817	0.8564	0.8682	0.8481	0.8533
ICS	MOS	0.2203	0.2157	0.2237	0.2216	0.2212	0.2183	0.2166	0.2161
	SII	0.0268	0.0300	0.0341	0.0232	0.0349	0.0388	0.0259	0.0228
	STOI	0.4665	0.4787	0.5322	0.4051	0.4620	0.4412	0.4903	0.4556
IdBM	MOS	0.6746	0.7095	0.6857	0.7416	0.7258	0.7454	0.7174	0.7434
	SII	0.4527	0.4274	0.3655	0.7189	0.5124	0.5665	0.4055	0.4386
	STOI	0.9461	0.9369	0.9215	0.9650	0.9456	0.9543	0.9375	0.9359
WPT IdBM	MOS	0.5878	0.6292	0.5972	0.7468	0.6472	0.7363	0.6402	0.6689
	SII	0.4280	0.4075	0.3502	0.6954	0.4876	0.5675	0.4103	0.4158
	STOI	0.9376	0.9227	0.9107	0.9696	0.9416	0.9546	0.9281	0.9233
WPT FUZZY	MOS	0.3217	0.3378	0.3480	0.3480	0.3077	0.3396	0.3896	0.4236
	SII	0.3828	0.3805	0.3229	0.3229	0.4385	0.5064	0.3647	0.3821
	STOI	0.7336	0.7430	0.7519	0.7519	0.7508	0.8386	0.8203	0.7923
WPT FUZZY modified Wiener	MOS	0.3085	0.3571	0.3537	0.3493	0.2896	0.4512	0.4530	0.4735
	SII	0.2433	0.2685	0.2625	0.4887	0.3049	0.4174	0.2761	0.2820
	STOI	0.7589	0.7464	0.7737	0.8333	0.7654	0.8746	0.8242	0.8117
Phase ratio based method	MOS	0.3190	0.3744	0.3599	0.3943	0.3017	0.4821	0.4135	0.3871
	SII	0.2700	0.3062	0.2910	0.5571	0.3602	0.4866	0.3103	0.2919
	STOI	0.7895	0.7726	0.7825	0.8444	0.7832	0.8732	0.8281	0.8016



Table B5: Speech Intelligibility Evaluation for Telgu Language

Methods	Measures	Babble	Airport	Car	Exhibition	Restaurant	White	Pink	Railway Platform
SS Boll	MOS	0.2635	0.2615	0.2531	0.2425	0.2270	0.2600	0.3293	0.2828
	SII	0.2936	0.1779	0.0358	0.1827	0.1170	0.0315	0.0007	0.0746
	STOI	0.5415	0.5299	0.4963	0.4870	0.4261	0.6129	0.6202	0.6234
MULTI-BAND SS	MOS	0.2949	0.2797	0.3096	0.2641	0.2854	0.3103	0.3308	0.3948
	SII	0.3351	0.0979	0.1015	0.2606	0.1584	0.1369	0.0988	0.1321
	STOI	0.5280	0.4539	0.4920	0.4800	0.4619	0.5353	0.5248	0.5738
Berouti SS	MOS	0.3005	0.2801	0.3112	0.2663	0.2854	0.2982	0.3256	0.3911
	SII	0.3726	0.1340	0.1385	0.3122	0.1984	0.1827	0.1410	0.1668
	STOI	0.5545	0.4658	0.5120	0.5015	0.4738	0.5427	0.5439	0.5963
Parametric SS	MOS	0.2661	0.2695	0.2676	0.2698	0.2458	0.2663	0.3461	0.3215
	SII	0.2200	0.1891	0.0264	0.2338	0.1112	0.0445	0.0025	0.0812
	STOI	0.4736	0.4966	0.5382	0.5105	0.4429	0.6146	0.6170	0.6034
Scalart SS	MOS	0.2904	0.2997	0.2877	0.2748	0.2704	0.2881	0.3030	0.3472
	SII	0.4166	0.2692	0.1911	0.3944	0.2939	0.2806	0.2119	0.2253
	STOI	0.5537	0.5377	0.5453	0.5789	0.5076	0.5900	0.5726	0.6069
WIENER	MOS	0.4352	0.4362	0.4234	0.4459	0.4510	0.4207	0.4186	0.4907
	SII	0.2176	0.1204	0.1056	0.3337	0.2291	0.2151	0.1786	0.1219
	STOI	0.7326	0.7487	0.7388	0.7327	0.7522	0.7151	0.7083	0.7779
MMSE	MOS	0.4287	0.4721	0.4608	0.4757	0.4581	0.4738	0.4556	0.4942
	SII	0.2526	0.1232	0.1199	0.3757	0.2340	0.2782	0.1546	0.1401
	STOI	0.7589	0.7589	0.7531	0.7687	0.7517	0.7325	0.7215	0.7749
ICS	MOS	0.2227	0.2243	0.2208	0.2236	0.2275	0.2187	0.2186	0.2189
	SII	0.0307	0.0244	0.0230	0.0316	0.0337	0.0111	0.0529	0.0219
	STOI	0.3743	0.4262	0.4280	0.3365	0.4071	0.3014	0.3860	0.4261
IdBM	MOS	0.6398	0.6686	0.6466	0.6747	0.6688	0.7245	0.6738	0.6997
	SII	0.5390	0.3364	0.3132	0.8754	0.5534	0.5758	0.4149	0.3137
	STOI	0.8893	0.8848	0.8691	0.8936	0.8962	0.8821	0.8650	0.8918
WPT IdBM	MOS	0.7128	0.7314	0.7228	0.7818	0.7208	0.7852	0.7474	0.7353
	SII	0.5189	0.3169	0.2991	0.8436	0.5281	0.5550	0.3877	0.3032
	STOI	0.9169	0.9097	0.8973	0.9299	0.9183	0.9232	0.9121	0.8981
WPT FUZZY	MOS	0.3029	0.3156	0.3370	0.2908	0.2965	0.3317	0.3823	0.3979
	SII	0.4759	0.2957	0.2766	0.5840	0.4829	0.5338	0.3473	0.2688
	STOI	0.5907	0.5732	0.5956	0.6091	0.5796	0.6515	0.6334	0.6338
WPT FUZZY modified Wiener	MOS	0.3120	0.3487	0.3605	0.3287	0.2923	0.4750	0.4983	0.5029
	SII	0.3820	0.2385	0.2532	0.5832	0.4511	0.5253	0.2682	0.2212
	STOI	0.6197	0.5924	0.6382	0.6404	0.5967	0.7062	0.7000	0.7137
Phase ratio based method	MOS	0.3286	0.3632	0.3816	0.3680	0.2950	0.4968	0.4532	0.4113
	SII	0.4667	0.2937	0.2913	0.6002	0.5153	0.5609	0.3437	0.2655
	STOI	0.6626	0.6542	0.6912	0.6576	0.6234	0.7330	0.7315	0.7209

Table B6: Speech Intelligibility Evaluation for Tamil Language

Methods	Measures	Babble	Airport	Car	Exhibition	Restaurant	White	Pink	Railway Platform
SS Boll	MOS	0.3175	0.3274	0.3275	0.2490	0.2883	0.2529	0.2713	0.4282
	SII	0.5834	0.3536	0.3221	0.2993	0.6181	0.1520	0.1797	0.3446
	STOI	0.7754	0.8135	0.8261	0.6585	0.7993	0.7096	0.7194	0.8704
MULTI-BAND SS	MOS	0.3085	0.3046	0.2964	0.2840	0.2742	0.3066	0.3316	0.4158
	SII	0.5094	0.2918	0.1695	0.2433	0.3930	0.1793	0.1652	0.3675
	STOI	0.7499	0.7241	0.6643	0.6593	0.7112	0.7143	0.7288	0.8665
Berouti SS	MOS	0.3065	0.3040	0.3034	0.2916	0.2778	0.3215	0.3362	0.3960
	SII	0.5236	0.3082	0.1996	0.2865	0.4217	0.2210	0.2014	0.3727
	STOI	0.7616	0.7623	0.7256	0.7045	0.7394	0.7582	0.7704	0.8667
Parametric SS	MOS	0.3052	0.2992	0.3155	0.2705	0.2707	0.2893	0.2949	0.3876
	SII	0.5882	0.3667	0.3263	0.2968	0.6373	0.1741	0.2083	0.3374
	STOI	0.7413	0.7563	0.7610	0.6471	0.7637	0.7178	0.7194	0.8163
Scalart SS	MOS	0.3188	0.3224	0.3362	0.3071	0.2950	0.3478	0.3599	0.4020
	SII	0.5950	0.3820	0.3580	0.4913	0.6438	0.3870	0.3210	0.3675
	STOI	0.7747	0.7901	0.8092	0.7537	0.7947	0.8085	0.8127	0.8585
WIENER	MOS	0.4003	0.4367	0.4380	0.4052	0.4344	0.4131	0.4481	0.4539
	SII	0.2357	0.1557	0.1531	0.2932	0.2508	0.2393	0.1911	0.1408
	STOI	0.8551	0.8588	0.8558	0.8249	0.8555	0.8463	0.8642	0.8752
MMSE	MOS	0.4104	0.4402	0.4587	0.4296	0.4083	0.4395	0.4550	0.4559
	SII	0.2837	0.1477	0.1496	0.4290	0.2729	0.3028	0.1780	0.1166
	STOI	0.8688	0.8643	0.8635	0.8708	0.8578	0.8591	0.8710	0.8468
ICS	MOS	0.2159	0.2126	0.2140	0.2220	0.2177	0.2183	0.2141	0.2096
	SII	0.0312	0.0357	0.0355	0.0175	0.0264	0.0222	0.0382	0.0424
	STOI	0.4310	0.4023	0.4520	0.3985	0.3783	0.3734	0.4227	0.3852
IdBM	MOS	0.6271	0.6280	0.6113	0.6173	0.6313	0.6446	0.6266	0.6645
	SII	0.6012	0.3925	0.3828	0.7673	0.6526	0.6845	0.4419	0.3811
	STOI	0.9590	0.9493	0.9404	0.9673	0.9597	0.9672	0.9530	0.9548
WPT IdBM	MOS	0.5871	0.5689	0.5313	0.6794	0.5976	0.7248	0.5782	0.5728
	SII	0.5796	0.3721	0.3602	0.7504	0.6278	0.6581	0.4157	0.3655
	STOI	0.9586	0.9487	0.9353	0.9737	0.9647	0.9685	0.9524	0.9397
WPT FUZZY	MOS	0.3069	0.3194	0.3245	0.3164	0.2883	0.3658	0.3883	0.3709
	SII	0.5478	0.3393	0.3182	0.6750	0.5875	0.6411	0.3670	0.2996
	STOI	0.7553	0.7837	0.7942	0.7946	0.7820	0.8511	0.8492	0.8253
WPT FUZZY modified Wiener	MOS	0.3355	0.3527	0.3576	0.3697	0.2990	0.5150	0.4730	0.4230
	SII	0.4137	0.2623	0.2648	0.6837	0.5732	0.6202	0.2751	0.2334
	STOI	0.8102	0.8301	0.8525	0.8445	0.8280	0.9156	0.8999	0.8832
Phase ratio based method	MOS	0.3457	0.3540	0.3549	0.3798	0.3040	0.4860	0.4276	0.3972
	SII	0.5121	0.3283	0.3282	0.6905	0.6134	0.6649	0.3628	0.2761
	STOI	0.8158	0.8229	0.8413	0.8149	0.8209	0.8863	0.8754	0.8701

Table B7: Speech Intelligibility Evaluation for Marathi Language

Methods	Measures	Babble	Airport	Car	Exhibition	Restaurant	White	Pink	Railway Platform
SS Boll	MOS	0.3435	0.3554	0.3552	0.3115	0.3439	0.3528	0.3842	0.5097
	SII	0.3918	0.2216	0.1653	0.2308	0.3457	0.1683	0.1686	0.2278
	STOI	0.6665	0.6604	0.6794	0.7160	0.6963	0.6693	0.7009	0.7645
MULTI-BAND SS	MOS	0.3836	0.4254	0.4350	0.4026	0.3720	0.4166	0.4576	0.5019
	SII	0.2803	0.2061	0.1762	0.2921	0.3168	0.1976	0.1897	0.2013
	STOI	0.6354	0.6446	0.6513	0.6601	0.6766	0.6722	0.6655	0.7026
Berouti SS	MOS	0.3615	0.3965	0.4143	0.3810	0.3583	0.3967	0.4336	0.4965
	SII	0.3029	0.2160	0.1972	0.3464	0.3288	0.2270	0.2111	0.2113
	STOI	0.6390	0.6565	0.6558	0.6715	0.6877	0.6863	0.6836	0.7099
Parametric SS	MOS	0.3646	0.3774	0.3882	0.3632	0.3519	0.3997	0.4209	0.5409
	SII	0.4004	0.2304	0.1722	0.2609	0.3505	0.2064	0.1887	0.2300
	STOI	0.6611	0.6593	0.6690	0.7180	0.6750	0.6837	0.6939	0.7323
Scalart SS	MOS	0.3654	0.3846	0.4356	0.3544	0.3625	0.3922	0.4464	0.4883
	SII	0.4082	0.2434	0.2242	0.3793	0.3752	0.3204	0.2615	0.2364
	STOI	0.6681	0.6700	0.6846	0.7093	0.6899	0.7104	0.7142	0.7176
WIENER	MOS	0.5135	0.5314	0.5138	0.5406	0.5189	0.5276	0.4832	0.5050
	SII	0.1617	0.1125	0.1058	0.3176	0.1584	0.1861	0.1203	0.1029
	STOI	0.7822	0.7935	0.7945	0.7996	0.7915	0.7987	0.7740	0.7959
MMSE	MOS	0.5726	0.5839	0.6065	0.5715	0.5512	0.5808	0.5219	0.5813
	SII	0.2010	0.1215	0.1143	0.3204	0.1704	0.2279	0.1299	0.1277
	STOI	0.8022	0.7998	0.7803	0.8242	0.7956	0.8300	0.7727	0.7987
ICS	MOS	0.2291	0.2244	0.2230	0.2328	0.2302	0.2204	0.2189	0.2190
	SII	0.0235	0.0190	0.0172	0.0322	0.0133	0.0159	0.0223	0.0140
	STOI	0.4736	0.4808	0.4899	0.4040	0.4231	0.3694	0.4777	0.4358
IdBM	MOS	0.7203	0.7133	0.7210	0.6921	0.7191	0.7211	0.7212	0.7524
	SII	0.4206	0.2577	0.2800	0.8159	0.3915	0.4366	0.3061	0.2438
	STOI	0.8796	0.8560	0.8631	0.9019	0.8866	0.8892	0.8650	0.8783
WPT IdBM	MOS	0.7099	0.7194	0.7174	0.7951	0.7370	0.7893	0.7444	0.7374
	SII	0.4012	0.2461	0.2661	0.7871	0.3763	0.4190	0.2907	0.2360
	STOI	0.9054	0.8880	0.8922	0.9191	0.9010	0.9167	0.8972	0.8923
WPT FUZZY	MOS	0.3343	0.4112	0.4199	0.3946	0.3718	0.4046	0.4640	0.4903
	SII	0.3613	0.2292	0.2328	0.4590	0.3513	0.3955	0.2649	0.2158
	STOI	0.6515	0.6586	0.6720	0.7071	0.6811	0.7433	0.7202	0.6816
WPT FUZZY modified Wiener	MOS	0.3484	0.4129	0.4188	0.4392	0.3456	0.5058	0.5165	0.5262
	SII	0.2722	0.1910	0.2005	0.4171	0.3254	0.3658	0.2138	0.1985
	STOI	0.7027	0.6980	0.7235	0.7640	0.7342	0.7738	0.7748	0.7403
Phase ratio based method	MOS	0.3738	0.3735	0.3773	0.4538	0.3581	0.5097	0.4676	0.4389
	SII	0.3376	0.2180	0.2261	0.4505	0.3613	0.4148	0.2532	0.2066
	STOI	0.7339	0.7344	0.7503	0.7853	0.7464	0.8020	0.7995	0.7605

Appendix B shows the intelligibility scores obtained by different speech enhancement methods for seven major Indian languages. The commonly used intelligibility measure parameters such as MOS, SII and STOI are taken for intelligibility scores calculation. The intelligibility scores are measured in the range of 0 to 1. The higher value (near to 1) shows speech signal of good intelligibility and vice versa.

The seven major Indian languages are taken for intelligibility analysis. The results of intelligibility measure parameters are given in Tables from B1 to B7. WPT modified Wiener and Phase ratio based speech enhancement methods show the better intelligibility improvement in all Indian languages.

