A NEW METHOD FOR MONAURAL SPEECH SEPARATION USING IDEAL BINARY MASK

Guide: **Dr. R. Rajavel**, Associate Professor

Project by :
M. Naveen Narayanan - 312214106066
S. Somasundar - 312214106104

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

OUTLINE

- Speech Separation
- Objectives of the Project
- Literature Survey
- Weintraub Speech Separation System
- Proposed Speech Separation System
- Experimental Results and Discussion
- Conclusion
- Work plan
- Reference

SPEECH SEPARATION

- Speech separation is the process of separating the target speech signal from acoustic mixture
- The acoustic mixture may be an another speech or environmental noise or both
- Speech separation can be used in speech/speaker recognition, voice communication, air-ground communication, hearing aids, etc.
- Various methods have been adopted for speech separation, mainly: Spectral Subtraction, Subspace analysis, Hidden Markov Modelling and Computational Auditory Scene Analysis

CASA and IBM

- CASA is the study of auditory scene analysis (ASA) by computational means
- The Ideal Binary Mask (IBM) has been proposed as a computational goal of CASA
- IBM is basically a binary matrix, in which 1 indicates speech dominant T-F units and 0 indicated noise dominant T-F units
- IBM is defined as

$$M(t,f) = \begin{cases} 1 & \text{if } s(t,f) - n(t,f) > 0, \\ 0 & \text{Otherwise} \end{cases}$$

where s(t,f) - target speech energy and

n(t,f) - interference energy in a T-F unit

OBJECTIVES OF THE PROJECT

- To implement the Weintraub speech separation system and the proposed speech separation system using Matlab
- Compare the computational complexity of these systems and show that the proposed system is less complex
- Also to show that the proposed system improves speech quality and intelligibility

LITERATURE SURVEY

S.No	TITLE	AUTHORS	YEAR OF PUBLICATION	INFERENCE
1.	A theory and computational model of auditory monaural sound separation	M. Weintraub	1985	This research provides a complete analysis of the human auditory system. A conceptual approach was introduced to mimic the knowledge and information used by human auditory system to separate sounds. A computer model was developed to separate the speech of two simultaneous talkers
2.	Frequency analysis and synthesis using a Gammatone filterbank	V. Hohmann	2002	This paper describes an efficient implementation of the 4th-order linear Gammatone filter based on an impulse-invariant, all-pole design. A linear auditory filter bank was constructed from these filters. This filter has been used in several applications involving computational auditory peripheral filtering

LITERATURE SURVEY

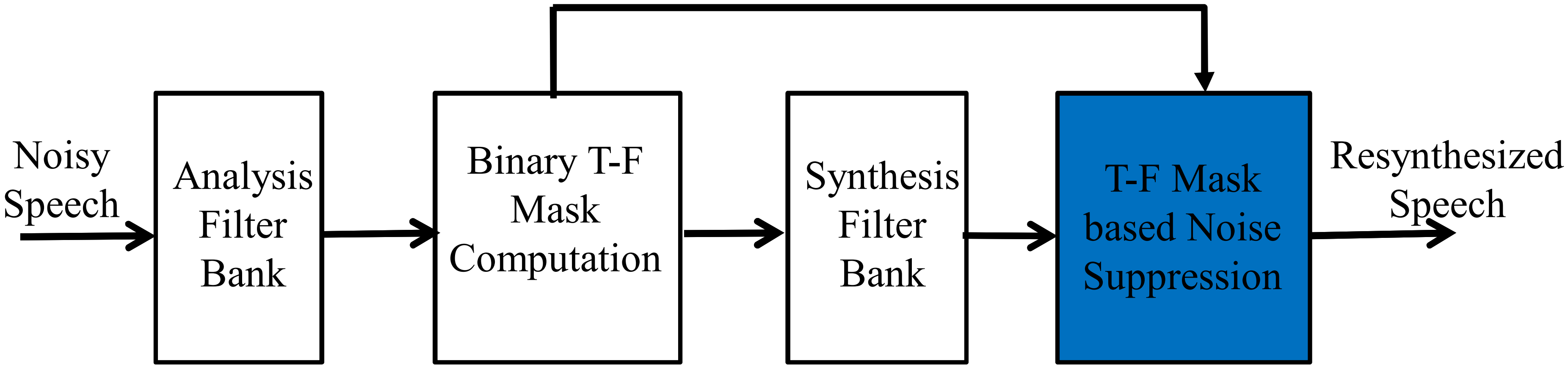
S.No	TITLE	AUTHORS	YEAR OF PUBLICATION	INFERENCE
3.	A comparison of several computational auditory scene analysis (CASA) techniques for monaural speech segregation	Jihen Zeremdini, Mohamed Anouar Ben Messaoud and Aicha Bouzid	2015	This paper mainly focused on the different CASA stages and the IBM for monaural speech segregation. It describes the several methods that used CASA to separate composite speech such as Hu and Wang, Zhang and Liu, Zhao and Shao, Li and Guan approaches etc. Finally an evaluation and a comparison was presented for the different monaural speech segregation methods
4.	Time-Frequency Masks for Monaural Speech Separation: A Comparative Review	Belhedi Wiem, Ben Messaoud Mohamed Anouar, Bouzid Aichl	2016	This paper focusses the masking effect on Computational Auditory Scene Analysis (CASA) based systems for single channel speech separation (SCSS). The CASA system employed in this study is Hu and Wang model. The new proposed system uses a soft mask instead of the hard binary mask and they have proved that the soft mask provides a better separation quality.

LITERATURE SURVEY

S.No	TITLE	AUTHORS	YEAR OF PUBLICATION	INFERENCE
5.	Single Channel Speech Enhancement Using Ideal Binary Mask Technique Based on Computational Auditory Scene Analysis	Abrar Hussain, Kalaivani Chellappan, Siti Zamrat	2016	In this paper, ideal binary mask which is inspired by the computational auditory analysis is used to analyze and synthesize the input speech signals and masker signals in the time-frequency domain. Synthesized signals are evaluated for speech quality measurement in terms of segmental signal-to-noise ratio. This study uses Malay language based speech as input speech signals

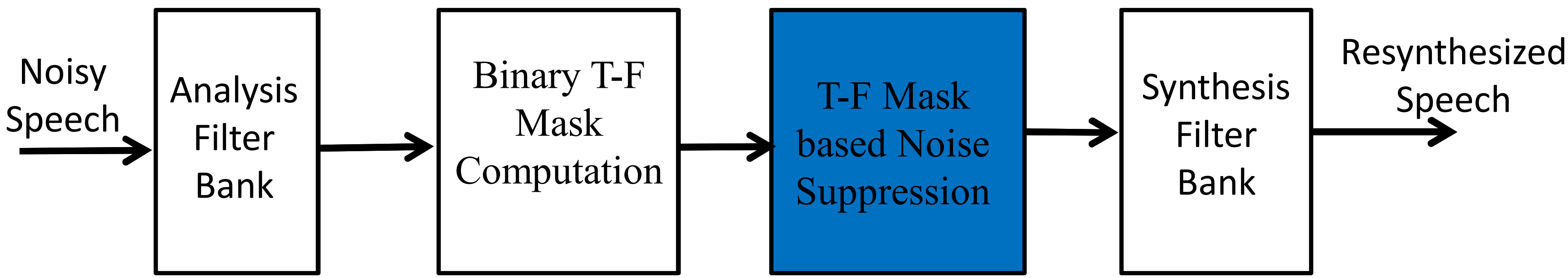
WEINTRAUB SPEECH SEPERATION SYSTEM

Speech separation system proposed by Weintraub [Ref1]



PROPOSED SPEECH SEPERATION SYSTEM

Proposed system to reduce computational complexity



For expanded diagram

EXPERIMENTAL RESULTS

- Speech and Noise Database: IEEEFemale.wav [IEEE]

  Speech: "The drip of the rain made a pleasant sound"

  Noise: Speechshapednoise.wav (Noisex-92)
- Input SNR = 5dB
- Sampling Frequency = 8000Hz
- Length of Speech signal = 21982 samples
- Number of Channels = 128
- Number of Frames per Channel = 274
- Number of samples per Frame = 160
- Length of Impulse response in Gammatone Analysis and Synthesis Filter Bank = 1024 samples

RESULTS

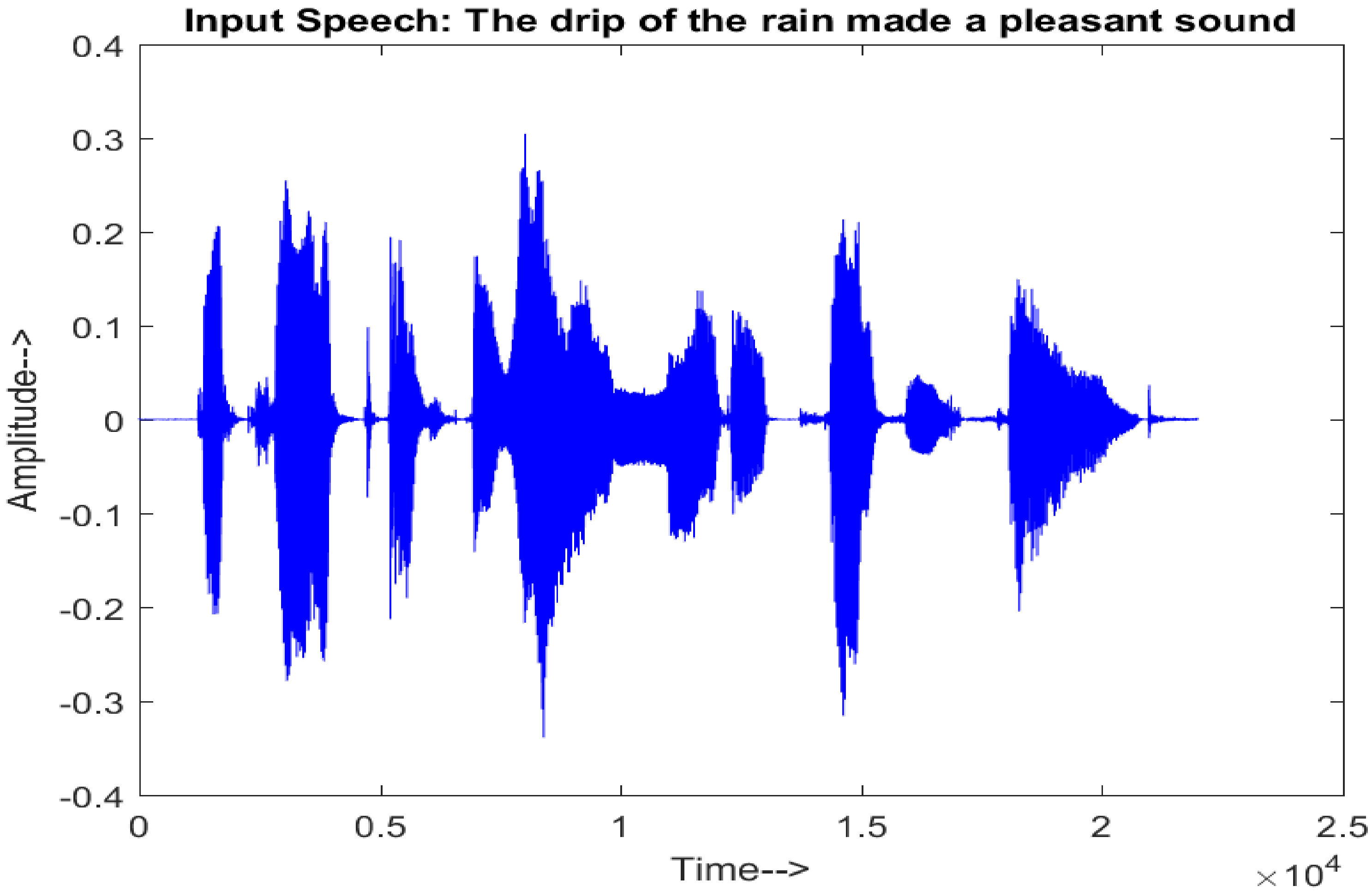
Speech

Noisy Speech at input SNR 5dB

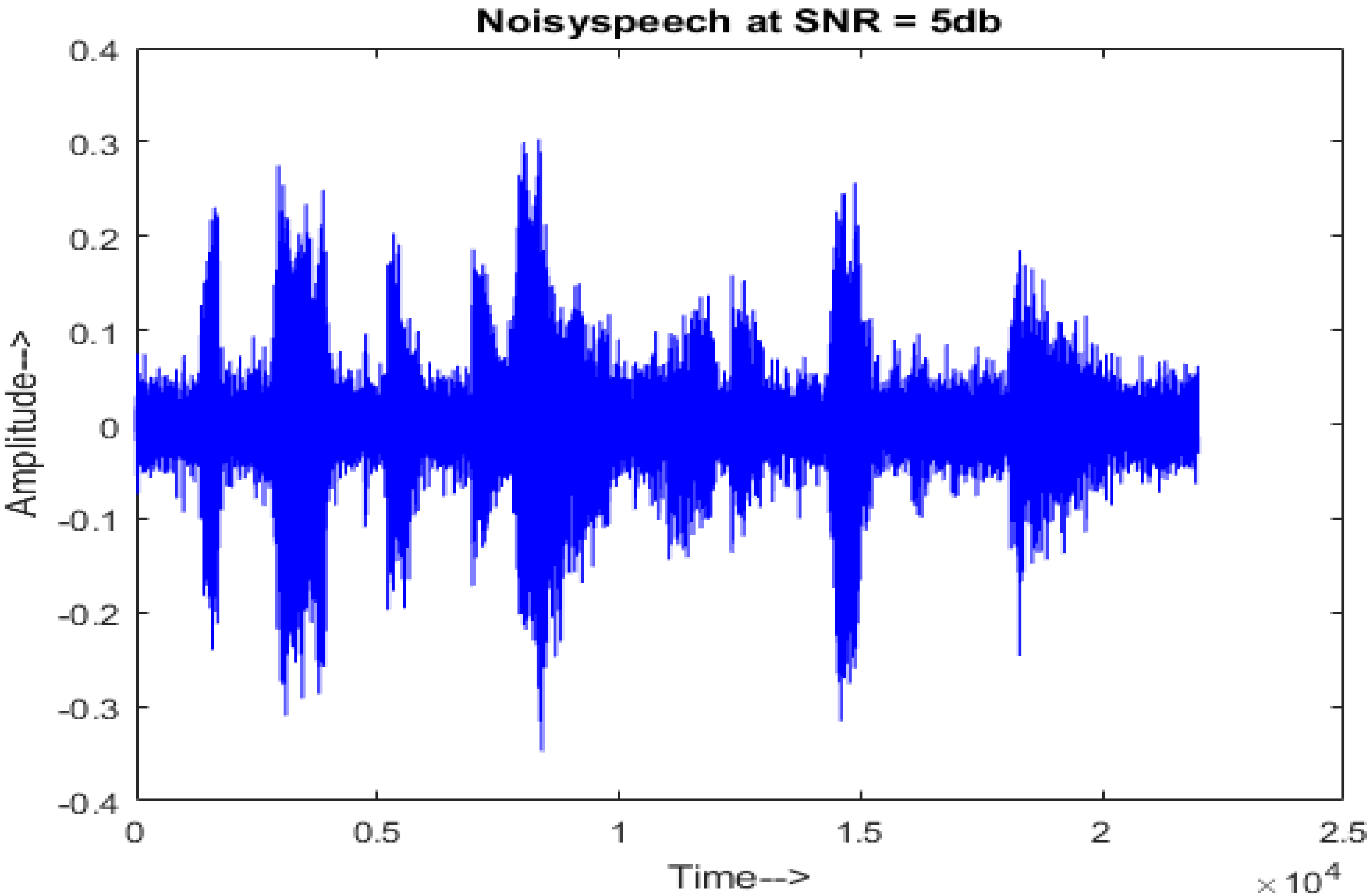
Resynthesized Speech using Weintraub System

Resynthesized Speech using Modified System

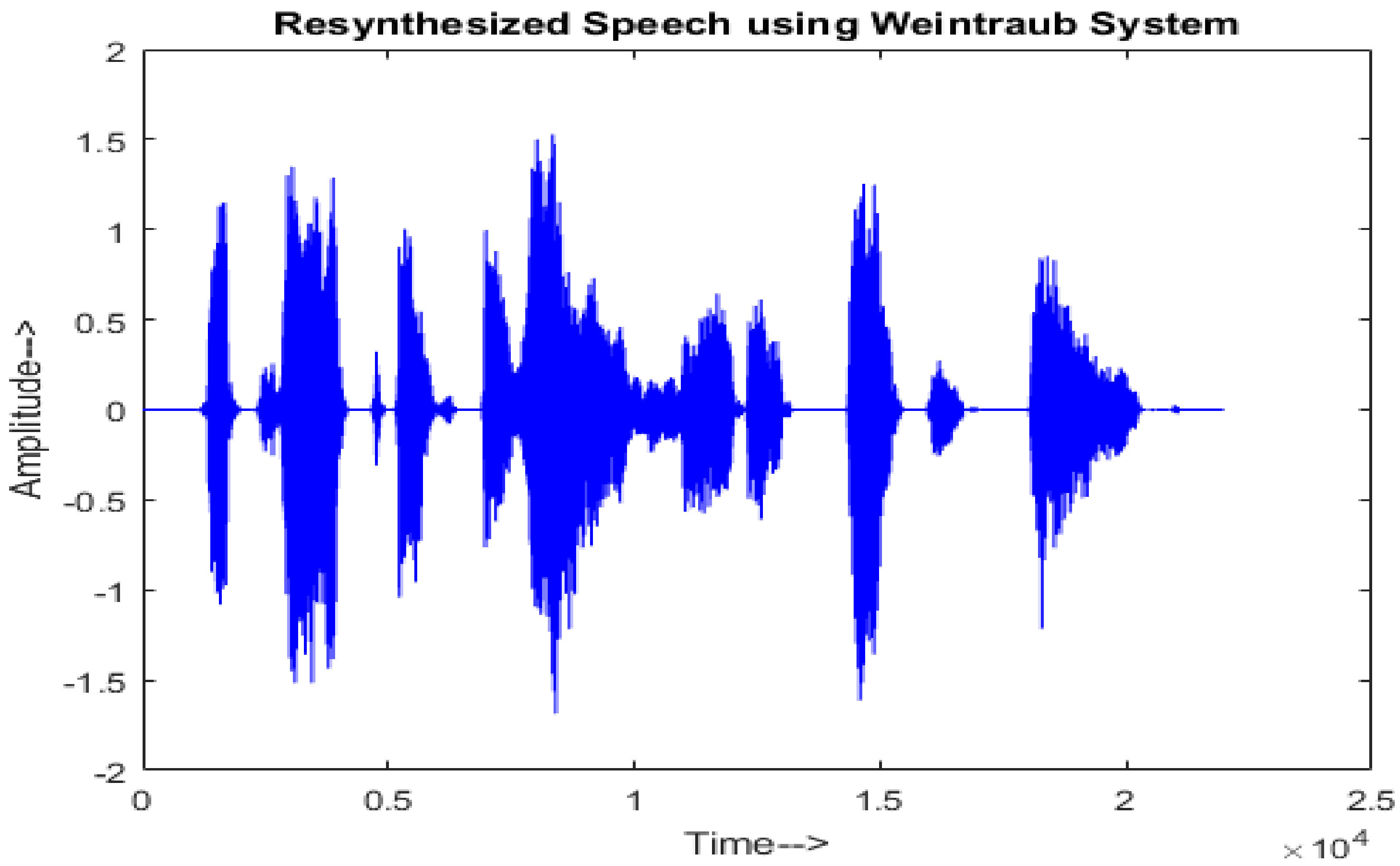
GRAPHICAL RESULTS



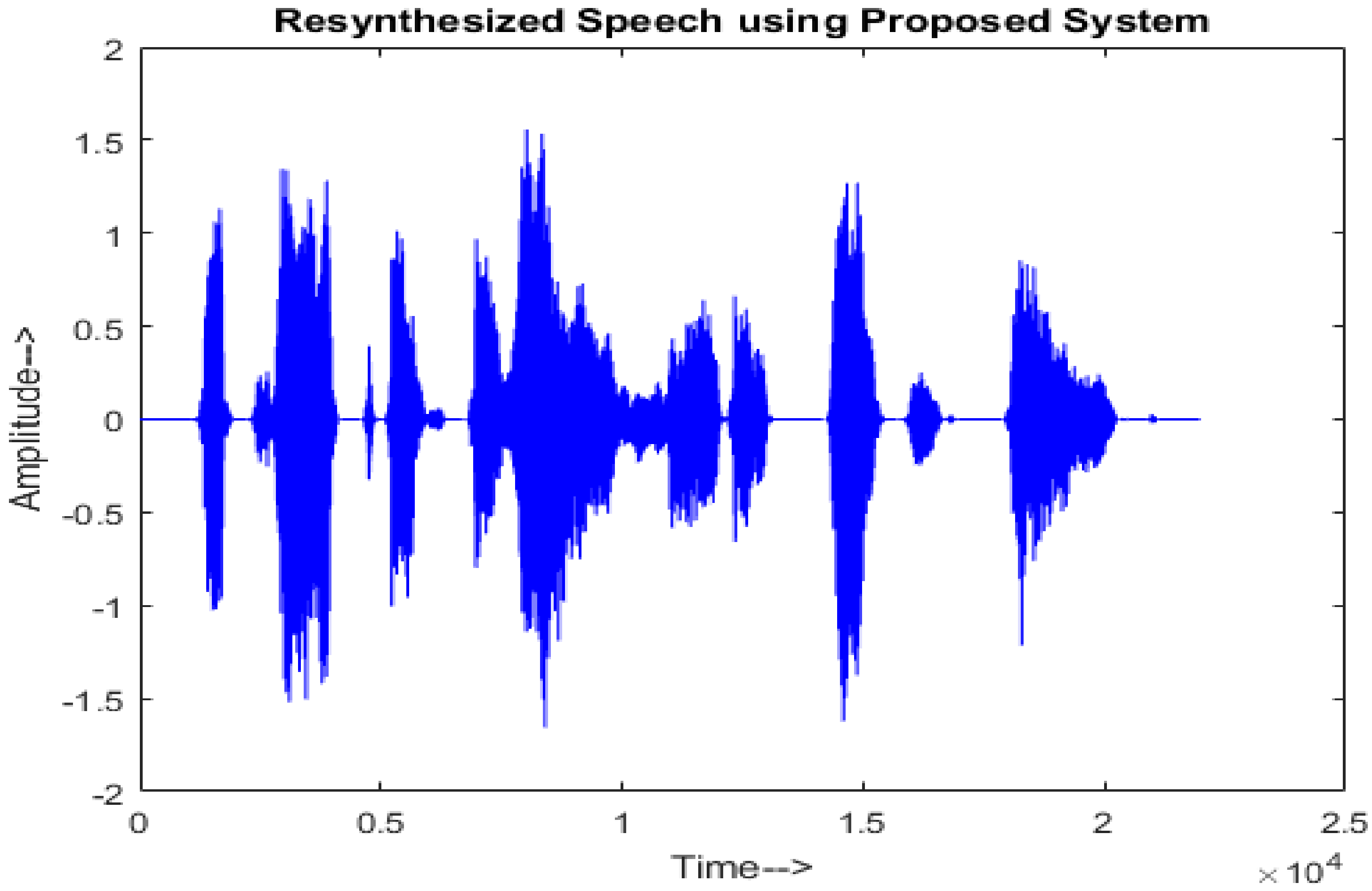
GRAPHICAL RESULTS



GRAPHICAL RESULTS



GRAPHICAL RESULTS



PERFORMANCE MEASURE QUALITY (SNR)

INPUT SNR (dB)	OUTPUT SNR (dB)		IMPROVEMENT (dB)
	Weintraub System	Proposed System	
-5	4.5874	4.9005	0.3131
0	6.6499	6.9077	0.2578
5	9.7100	9.9143	0.2043
10	13.6690	13.8812	0.2122
15	17.6510	17.8724	0.2214
AVERAGE	10.45346	10.69522	0.24176

PERFORMANCE MEASURE QUALITY (SNR)

ANALYSIS FILTER BANK*	SYNTHESIS FILTER BANK*	OUTPUT SNR (dB)	
		Weintraub System	Proposed System
1024	1024	9.6742	9.8804
1024	512	9.6742	9.8804
512	512	9.6742	9.8804
512	512/128	9.6659	9.8807

* Number of coefficients in the Gammatone filter bank

PERFORMANCE MEASURE INTELLIGIBILITY (STOI)

INPUT SNR (dB)	OUTPUT STOI		IMPROVEMENT
	Weintraub System	Proposed System	
-5	0.7109	0.7602	0.0493
0	0.8548	0.8595	0.0047
5	0.8833	0.8937	0.0104
10	0.9171	0.9331	0.0160
15	0.9462	0.9504	0.0042
AVERAGE	0.86246	0.87938	0.01692

COMPARISON OF COMPUATATIONAL COMPLEXITY

ANALYSIS FILTER BANK*	SYNTHESIS FILTER BANK*	COMPUTATION TIME (Secs)	
		Weintraub System	Proposed System
1024	1024	1.6775	1.3867 (0.2908)
1024	512	1.6038	1.3429 (0.2609)
512	512	1.4708	1.2992 (0.1716)
512	512/128	1.3813	1.2156 (0.1657)

* Number of coefficients in the Gammatone filter bank

COMPARISON OF COMPUATATIONAL COMPLEXITY

ANALYSIS FILTER BANK*	SYNTHESIS FILTER BANK*	THROUGHPUT (Samples per Sec)	
		Weintraub System	Proposed System
1024	1024	453	517 (64)
1024	512	486	524 (38)
512	512	511	548 (37)
512	512/128	527	554 (27)

* Number of coefficients in the Gammatone filter bank

COMPARISON OF COMPUATATIONAL COMPLEXITY

ANALYSIS FILTER BANK*	SYNTHESIS FILTER BANK*	NUMBER OF MULTIPLICATIONS	
		Weintraub System	Proposed System
1024	1024	5768060928	5768060928
1024	512	4327448576	4327448576
512	512	2886836224	2886836224
512	512/128	1924552192	1924552192

* Number of coefficients in the Gammatone filter bank

COMPARISON OF COMPUATATIONAL COMPLEXITY

ANALYSIS FILTER BANK*	SYNTHESIS FILTER BANK*	NUMBER OF ADDITIONS	
		Weintraub System	Proposed System
1024	1024	5759635712	5759635712
1024	512	4319023360	4319023360
512	512	2878411008	2878411008
512	512/128	1916126976	1916126976

* Number of coefficients in the Gammatone filter bank

CONCLUSION

- The Weintraub speech separation system has been implemented using Matlab
- The proposed speech separation system has also been implemented using Matlab
- Both these systems were tested using standard speech and noise database from IEEE corpus and Noisex-92 respectively.
- Mathematical Analysis of the system has been carried out.
- Finally the experimental results show that the proposed system improves the speech quality and intelligibility and also reduces the computational delay of the existing monaural speech separation system

WORKPLAN

	AUG-2017	SEP-2017	OCT-2017	NOV-2017	DEC-2017	JAN-2018	FEB-2018	MAR-2018
Literature Survey								
Implementation of the existing system								
Implementation of the proposed system								
Comparing the Computational Complexity and Report Preparation								

REFERENCES

1. M. Weintraub, “A theory and computational model of auditory monaural sound separation,” Ph.D. dissertation, *Dept. Elect. Eng., Stanford Univ., Stanford, CA,* 1985
2. V. Hohmann, “Frequency analysis and synthesis using a Gammatone filterbank,” *Acta Acustica united with Acustica,* Vol. 88, pp. 433 – 442, January 2002
3. D.L. Wang, G.J. Brown, “Fundamentals of Computational Auditory Scene Analysis,” in *Computational Auditory Scene Analysis,* D.L Wang and G.J Brown, *Wiley-IEEE Press*, pp. 1-38, 2006
4. D.L. Wang, “Time–Frequency Masking for Speech Separation and Its Potential for Hearing Aid Design,” *Trends in Amplification*, Vol. 12, pp.332-353, December 2008

REFERENCES

5. Ke Hu and DeLiang Wang, "An Unsupervised Approach to Cochannel Speech Separation", *IEEE Transactions on Audio, Speech and Language Processing*, Vol.21, No.1, January 2013
6. N. Harish Kumar, R. Rajavel, "Monaural speech separation system based on optimum soft mask," *IEEE Int. Conf. on Computational Intelligence and Computing Research*, 18-20 Dec 2014
7. Jihen Zeremdini, Mohamed Anouar Ben Messaoud and Aicha Bouzid, "A comparison of several computational auditory scene analysis (CASA) techniques for monaural speech segregation", *Brain Informatics*, Vol. 2, Issue 3, pp. 155 – 166, September 2015

REFERENCES

8. Abrar Hussain, Kalaivani Chellappan and Siti Zamrat, "Single channel speech enhancement using ideal binary mask technique based on computational auditory scene analysis", *Journal of Theoretical and Applied Information Technology*, Vol. 91. No. 1, September 2016
9. Belhedi Wiem, Ben Messaoud Mohamed Anouar, Bouzid Aichl, "Time-Frequency Masks for Monaural Speech Separation: A Comparative Review", *7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT),* 2016

THANK YOU

REFERENCES

10. V.A. Mane , Prof. Dr. S. B. Patil, ”Survey of Methods and challenges in Computational Auditory sense analysis”, International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, Vol. 4, Issue 9, September 2016

LITERATURE SURVEY

S.No	TITLE	AUTHORS	YEAR OF PUBLICATION	INFERENCE
9.	Survey of Methods and challenges in Computational Auditory sense analysis	V.A. Mane , Prof. Dr. S. B. Patil	2016	This paper is a study of various literature and methods that are used to solve the cocktail party problem. This paper discusses about the evolution of computational auditory scene analysis and the various steps taken to design and define the underlying process which will do human mimicry. The paper also describes about the challenges faced by CASA to segregate unvoiced speech from non speech interference

LITERATURE SURVEY

S.No	TITLE	AUTHORS	YEAR OF PUBLICATION	INFERENCE
6.	Time–Frequency Masking for Speech Separation and Its Potential for Hearing Aid Design	D.L. Wang	2008	This article introduces the T-F masking concept and reviews T-F masking algorithms that separate target speech from either monaural or binaural mixtures, as well as microphone-array recordings. This article also surveys recent studies that evaluate the perceptual effects of T-F masking techniques, particularly their effectiveness in improving human speech recognition

LITERATURE SURVEY

S.No	TITLE	AUTHORS	YEAR OF PUBLICATION	INFERENCE
7.	An Unsupervised Approach to Cochannel Speech Separation	Ke Hu and DeLiang Wang	2013	In this paper, an unsupervised method was proposed for cochannel speech separation. The proposed system uses the CASA algorithm for voiced speech segregation. The proposed system was compared with several other methods across a range of input SNR conditions
8.	Monaural speech separation system based on optimum soft mask	N. Harish Kumar, R. Rajavel	2015	This paper proposes the optimum soft mask (OSM) to reduce the musical noise, by replacing the hard limiting weights of IBM with the variable weights. The Signal-to-Noise ratio is used as a measure to evaluate the performance of the proposed soft mask with the existing IBM in the context of monaural speech separation. The rest of the paper is organized in the following manner

COMPUTATIONAL ANALYSIS

- Length of Speech signal = 21982 samples
- Length of the Impulse response in Gammatone Analysis and Synthesis Filter bank = 1024 samples
- Number of Channels = 128
- Number of Frames per channel = 275
- The number of multiplications in Analysis Filter Bank = Length of Speech signal*Length of Impulse response*Number of Channels = 22004*1024*128 = 2881224704
- The number of multiplications in the process of speech separation = Number of Channels*Number of Frames*Length of each Frame = 128*274*160 = 5611520
- The number of multiplications in Synthesis Filter Bank = 21982*1024*128 = 2881224704

EXPERIMENTAL RESULTS

	WEINTRAUB SYSTEM	PROPOSED SYSTEM
Number of Multiplications	5768060928	5768060928
Number of Additions	5759635712	5759635712
Throughput (in samples per secs)	453	517

System Specifications: Intel® Core™ i5-3210M CPU@2.50Ghz
RAM: 4.00GB
64- bit operating System
Windows 10 Home edition

MATLAB Version: R2015a

GRAPHICAL RESULTS