International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Enhancement of Speech Intelligibility using Binary Mask Based on Noise Constraints

Ramesh Nuthakki, Sreenivasa Murthy A

Abstract: ***The primary aim of this paper is to examine the application of binary mask to improve intelligibility in most unfavorable conditions where hearing impaired/normal listeners find it difficult to understand what is being told. Most of the existing noise reduction algorithms are known to improve the speech quality but they hardly improve speech intelligibility. The paper proposed by Gibak Kim and Philipos C. Loizou uses the Weiner gain function for improving speech intelligibility. Here, in this paper we have proposed to apply the same approach in magnitude spectrum using the parametric wiener filter in order to study its effects on overall speech intelligibility. Subjective and objective tests were conducted to evaluate the performance of the enhanced speech for various types of noises. The results clearly indicate that there is an improvement in average segmental signal-to-noise ratio for the speech corrupted at -5dB, 0dB, 5dB and 10dB SNR values for random noise, babble noise, car noise and helicopter noise. This technique can be used in real time applications, such as mobile, hearing aids and speech–activated machines.***

Keywords: speech intelligibility, noise estimation, Speech enhancement, objective and subjective performance measures, spectrograms.

I. INTRODUCTION

Speech communication is the important medium of oral communication. Speech often gets degraded in daily life because of the poor listening conditions such as background noise, electronic transmissions etc. Because of the presence of the back ground noise, the expected quality and intelligibility of the speech signal gets affected [1-5].

Recent studies have proved that large gains can be attained using the ideal binary mask technique [7] in speech intelligibility. The binary mask technique is proposed to retain the time frequency (T-F) units where the target speech subdues the masker signal and removes the T-F units where the masker signal subdues. The capability of the binary mask technique in improving speech intelligibility is indicated by using a Bayesian classifier. The removal or retaining of T-F bins using in binary mask is dependent on the noise spectrum overestimation or underestimation criterion. By taking the gain function into consideration a separate mask can be created by applying constraints on the two types of distortions. The size of the spectral amplitudes varies from the original spectral amplitudes with the application of the gain function. As a result, the attenuation and amplification distortions take place. It has been proved that the amplification distortion is more damaging compared to attenuation distortion. The obtained enhanced speech which contains attenuation distortion is proved to be more intelligible to that of noisy speech. For constructing a speech that contains only attenuation (or amplification) distortion, binary mask has to be applied to the enhanced speech spectrum. The contribution of each distortion initiated by the noise spectrum over estimation or under estimation is taken into consideration. The results made it clear that the new binary mask technique is capable of improving speech intelligibility for distinct background noises even at very low SNR levels [2]. This paper is organized as follows. Section II describes the binary mask scenarios based on noise constraints. Section III shows overall intelligibility and quality measures and section IV gives a conclusion.

Manuscript published on 30 September 2019
* Correspondence Author
Ramesh Nuthakki*, Research Scholar, Department of Electronics and Communication Engineering, University Visvesvaraya College of Engineering, Bengaluru, India. nuthakki.ramesh@gmail.com
Dr. Sreenivasa Murthy A, Professor, Department of Electronics and Communication Engineering, University Visvesvaraya College of Engineering, Bengaluru, India. uvceasm@gmail.com

II. BINARY MASK SCENARIOS BASED ON NOISE CONSTRAINTS

Numerous studies have proved that high gains in speech intelligibility can be attained using the binary mask technique. The binary mask takes the values of zero and one. It is one of the selection criteria on the basis of noise overestimation or noise underestimation constraints. When a binary mask is applied to the time-frequency representation of a mixture signal, it eliminates portions of a signal that are assigned to value ‘zero’ and preserves the values that are assigned to value ‘one’. Studies have shown that the speech synthesized from the binary mask is highly intelligible even in reverberant conditions [1].

A. Evaluation of noise and speech spectra

A binary mask was discussed in this module depending on the noise constraints. By regulating the distortions introduced by noise spectrum estimate, a new time frequency mask was constructed and is applied to the enhanced spectrum. The clean speech signal is considered as c(n) and zero mean noise process as d(n). Then the degraded speech signal y(n) can be written as follows :

$$y(n) = c(n) + d(n) \quad (1)$$

The steps involved in the construction of the binary mask are clearly shown in the Fig. 1.

The noisy speech sentences were divided into 20-ms frames with 50% of overlap connecting the adjoining frames. Each single speech frame is Hann-windowed and a 320-point FFT was measured. Let $Y(k, n_i)$ denote the noisy spectrum at time frame index n_i and frequency bin k. Then, the estimate of the speech magnitude spectra $\hat{C}(k, n_i)$ can be obtained by multiplying $Y(k, n_i)$ with the parametric gain function $G(k, n_i)$ as follows:

$$\hat{C}(k, n_i) = G(k, n_i) . Y(k, n_i) \quad (2)$$

Each noisy frame was multiplied by the gain function to obtain the estimated clean speech signal. After applying the parametric Weiner gain to each frame, the frames were summed up to form the final output. This output is known as the Weiner processed signal. Similarly the estimate of the noise spectral magnitude $\hat{E}(k, n_i)$ was computed in equation no. (5). Then the binary mask criterion was applied to the Wiener processed signal. The formation of the binary mask is given in the section 2.2.

In order to get a clarity about the impact of these two distortions on speech intelligibility, we have taken into consideration the parametric Weiner filter for certain parameters γ and β. If we take γ = 1 and β =1, we get the square root Weiner filter. If we take γ = 1 and β = 2, we get the Weiner filter. By changing the parameters γ and β, we get different types of Weiner filters with distinct attenuation characteristics. To study the effect of these two parameters on attenuation, we can write the equation as follows [1-3].

$$G(k, n_i) = \sqrt{\left(\frac{SNR_{prio}(k,n_i)}{\gamma + SNR_{prio}(k,n_i)}\right)^{\beta}} \quad (3)$$

The Parametric Weiner gain function G(k, n_i) is demonstrated in terms of priori SNR, $\hat{C}(k, n_i)$ indicates the estimate of clean speech magnitude spectrum at frame index n_i and frequency bin k. In this paper we use wiener algorithm as the gain function, since it is capable of reducing the estimation errors. Unlike the other noise reduction algorithms, wiener algorithm is found to be accurate in terms of quality and intelligibility [1].

SNR_{prio} can be evaluated using the recursive equation as

$$SNR_{prio}(k, n_i) = \alpha . \frac{\hat{C}^2(k, n_i-1)}{\hat{P}_D^2(k, n_i-1)} + (1-\alpha).\max\left[\frac{Y^2(k, n_i)}{\hat{P}_D^2(k, n_i)} - 1, 0\right] \quad (4)$$

Where α is the smoothing constant whose value is 0.98, $\hat{P}_D^2$ (k, n_i) represents the estimates of the background noise variance. In order to estimate the noise variance, we use the noise estimation algorithm suggested in [6, 14]. The noise spectrum magnitude $\hat{E}(k, n_i)$ is estimated as follows

$$\hat{E}(k, n_i) = G_E(k, n_i)\, Y(k, n_i) \quad (5)$$

$$G_E(k, n_i) = \sqrt{\left(\frac{1}{\gamma + SNR_{prio}(k,n_i)}\right)^{\beta}} \quad (6)$$

Where G_E(k, n_i) indicates the noise equivalent Parametric wiener gain function[1-2,17].

B. *Formation of the binary mask*

The estimated noise spectrum $\hat{E}(k, n_i)$ is first calculated, then a binary mask was constructed by limiting the distortions. When $\hat{E}(k, n_i) > E(k, n_i)$, the noise over estimation distortion occurs and when $\hat{E}(k, n_i) <$ E(k, n_i) the noise under estimation distortion occurs. The processed speech comprises of both the distortions. To find out the effect of noise over estimation or under estimation distortion alone on speech intelligibility, constraints were applied on the estimated speech spectral magnitude. For each single T-F unit, the estimated noise spectrum $\hat{E}(k, n_i)$ is determined against the original noise magnitude spectrum E(k, n_i). Only the T-F units that are fulfilling the constraints were retained and the rest were removed. Then the modified magnitude spectrum $\hat{C}_M(k, n_i)$ is calculated as follows

$$\hat{C}_M(k, n_i) = \begin{cases} \hat{C}(k, n_i) & \text{if } \hat{E}(k, n_i) > \text{E(k, } n_i\text{)} \\ 0 & \text{else} \end{cases} \quad (7)$$

After this, to the above selection of T-F units, an inverse IFFT was applied to the modified spectrum $\hat{C}_M(k, n_i)$ using the phase of the noisy speech spectrum. Finally the overlap-add-technique was used for the reconstruction of the enhanced speech signal.

III. OVERALL INTELLIGIBILITY AND QUALITY MEASURES

A. *Objective Measures*

The objective measure used in this paper is average segmental SNR(segSNR). Although there are many objective measures we have chosen this because of the accuracy it provides when compared with other parameters. It is one of the extensively used objective measure. Higher the value of segSNR, the enhanced speech signal carries more signal power as against noise power. The objective measures for speech quality can be performed by dividing the speech signal into 20 ms frames. After that the distortion measure was calculated between the original and the processed speech signal. The speech distortion was estimated by equating the distortion measures of each frame in the time domain. In this paper we have considered the SSNR in the time domain as an objective measure. Clearly table- 1 shows the improvement in SSNR values for random noise, babble noise, car noise and helicopter noise. During the computation of SNR_{seg}, the signal energy at the intervals of silence was relatively low leading to large negative SNR_{seg} values. The average segmental SNR[1,15-16] is written as:

$$SSNR = \frac{10}{M} \sum_{m=0}^{M-1} 10 \log_{10} \frac{\sum_{i=im}^{im+i-1} C(i)^2}{\sum_{i=im}^{im+i-1} \left(C(i) - \hat{C}(i)\right)^2} \quad (8)$$

Where C and $\hat{C}$ denotes the clean and enhanced speech signal, M denotes the number of frames and i denotes the frame length.

The average segmental SNR is evaluated by positioning the clean and the enhanced speech signals for various types of noises and for different input SNR values. This measure yields good results using parametric wiener filter for stationary as well as non-stationary types of noises.

B. *Subjective Measures*

For this test we have taken a set of 10 listeners, 5 male and 5 female and they were asked to listen to the enhanced and noisy speech signals randomly. These subjective tests were mostly based on parameters like background quality (BAK), signal quality (SIG) and overall signal quality. They were asked to give scores from 1 to 5 for the above parameters. It has been found that, at -5dB, 0dB, 5dB, 10dB SNR levels the speech signals were corrupted by various kinds of noises such as babble noise, random noise, car noise and helicopter noise etc. The total scores given by the listeners are collected and shown in the form of tables 2 and 3. The listening tests showed a clear improvement in the speech quality [8-10, 15] for random noise, babble noise, helicopter noise and Car noise.

C. Spectrograms

A spectrogram is a visual representation of sound. It displays the amplitude of the frequency components of the signal over time. It gives an account of speech signal's relative energy concentration in frequency as a result of time and it displays the time-varying properties of the speech wave form. The red regions are associated to the energy signal. The voiced regions are indicated by the striped display and the unvoiced are fully covered in. The magnitude of the spectrogram is indicated through colors such as red which specifies the high energy and blue, the low energy respectively. Figures 2 [(a),(b),(c) & (d)] , 3 [(a),(b),(c) & (d)] and 4 [(a),(b),(c) & (d)] represents the spectrograms of the helicopter noise, babble noise and car noise. From these spectrograms it is clearly evident that in terms of both voiced/unvoiced segments, the pitch and formants for the most part are well recovered. There is an improvement in speech intelligibility for the helicopter noise.

Table- 2: Subjective measures using wiener filter (γ=1, β=1, Kim's approach)

Noise	I/P SNR(dB)	BAK	SIG	OVL
Random Noise	**10**	**3.8**	**4.0**	**4.1**
	5	**3.7**	**4.1**	**4.2**
	0	**3.6**	**3.7**	**3.9**
	-5	**2.1**	**2.3**	**2.6**
Babble Noise	**10**	**4.0**	**4.1**	**4.2**
	5	**3.8**	**4.0**	**4.1**
	0	**3.8**	**3.7**	**3.9**
	-5	**2.5**	**2.7**	**2.9**
	10	**4.1**	**4.2**	**4.3**
Helicopter Noise	**5**	**3.6**	**4.0**	**4.1**
	0	**3.1**	**3.3**	**3.7**
	-5	**2.2**	**3.0**	**3.5**
Car Noise	**10**	**4.0**	**4.1**	**4.1**
	5	**3.3**	**3.6**	**4.0**
	0	**3.2**	**3.6**	**3.7**
	-5	**2.3**	**2.7**	**3.4**

Table- 3: Subjective measures using parametric wiener filter

Noise	I/P SNR(dB)	BAK	SIG	OVL
Random Noise	**10** (γ=5, β=0.2)	**3.9**	**4.5**	**4.8**
	5 (γ=4, β=0.2)	**3.9**	**4.3**	**4.7**
	0 (γ=1.5, β=0.5)	**3.8**	**4.0**	**4.1**
	-5 (γ=4.2, β=0.4)	**2.2**	**2.3**	**2.9**
Babble Noise	**10** (γ=0.7, β=0.3)	**4.3**	**4.5**	**4.8**
	5 (γ=0.7, β=0.3)	**3.9**	**4.3**	**4.7**
	0 (γ=3, β=0.2)	**3.9**	**4.2**	**4.3**
	-5 (γ=0.4, β=0.9)	**2.6**	**2.7**	**3.0**
Helicopter Noise	**10** (γ=2.5, β=0.3)	**4.3**	**4.6**	**4.8**
	5 (γ=2.5, β=0.3)	**3.9**	**4.3**	**4.7**
	0 (γ=3, β=0.5)	**3.8**	**4.2**	**4.5**
	-5 (γ=5.5, β=0.4)	**3.7**	**4.1**	**4.4**
Car Noise	**10** (γ=2.5, β=0.3)	**4.1**	**4.3**	**4.7**
	5 (γ=2.5, β=0.3)	**3.9**	**4.1**	**4.6**
	0 (γ=3.5, β=0.4)	**3.2**	**3.7**	**4.1**
	-5 (γ=3.5, β=0.4)	**2.3**	**2.9**	**4.0**

D. Intelligibility hearing Tests

In order to assess the intelligibility [11-13] of the processed speech, listening tests were conducted. The sentences were taken from Indian English data base and IEEE. Noisy speech was generated by adding helicopter noise, car noise, random noise and babble noise at -5dB, 0dB, 5dB, 10dB SNRs. Before conducting the test, each listener was made to listen to a set of noise-corrupted sentences. After that they were asked to identify the words from the estimated speech signal. The performance was evaluated by counting the number of words identified correctly and is shown in Fig.5 and Fig.6.

E. Results and Discussion

From the subjective results shown in the table 2 and 3, it is evident that there is an improvement in intelligibility for the random noise, babble noise, helicopter noise and Car noise at 0dB, -5dB, +5dB and 10 dB SNR values respectively. The results are obtained by considering the mean percentage of words identified correctly by the normal hearing listeners. When $\hat{E}(k, n_i) > E(k, n_i)$, the intelligibility has improved with parametric wiener filter. The unprocessed speech scores are indicated by UN. From the figures 5 and 6, it is clear that when the proposed noise constraints were applied, the performance at -5dB, 0dB,+5dB and 10 dB SNR levels improved from 21%,50%,65% and 80% with unprocessed speech to 88%,92%,94% and 97% respectively using wiener filter and reduced to nearly zero with noise underestimated constraints. Similarly in parametric wiener filter too, the performance at SNR levels -5dB, 0dB, +5dB and 10dB improved from 21%,50%,65% and 80% with the unprocessed speech to 96%,98%,99% and 100% respectively and reduced to closely zero with noise underestimated constraints. When we compare the two figures 5 and 6, with SNR levels -5dB, 0dB, +5dB, +10dB there is an improvement in word count from 88% to 96%, 92% to 98%, 94% to 99% and 97% to 100%. These results demonstrate that the proposed binary mask yields good in parametric wiener filter against wiener filter for the random noise, babble noise car noise and helicopter noise.

Table- 1: Objective measures using parametric wiener and wiener Filter

Noise	I/P SNR (dB)	γ	β	Seg.SNR(dB) using Parametric Wiener Filter	Seg.SNR(dB) using Wiener Filter (γ=1, β=1, Kim's approach)
Random Noise	10	5	0.2	17.8190	14.5425
	5	4	0.2	13.0516	9.8073
	0	1.5	0.5	12.9847	12.8214
	-5	4.2	0.4	8.7793	8.6038
Babble Noise	10	0.7	0.3	14.8044	11.0596
	5	0.7	0.3	11.0111	5.0853
	0	3	0.2	5.2701	3.0390
	-5	0.4	0.9	1.0528	0.4857
Helicopter Noise	10	2.5	0.3	22.7748	21.4692
	5	2.5	0.3	18.4329	15.6423
	0	3	0.5	14.5684	12.6549
	-5	5.5	0.4	10.4940	8.3977
Car Noise	10	2.5	0.3	16.9248	13.5615
	5	2.5	0.3	14.6185	12.5752

	0	3.5	0.4	13.0179	10.1321
	-5	3.5	0.4	11.1581	7.2035

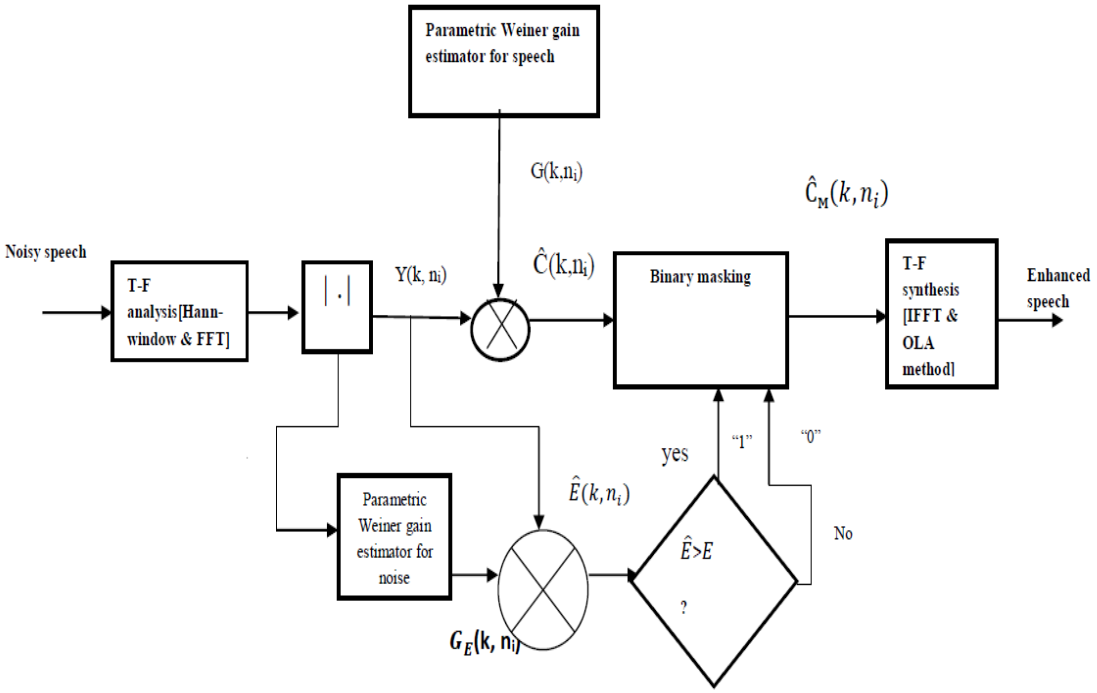


Fig. 1. Steps used for constructing the Proposed binary mask depending on noise constraints

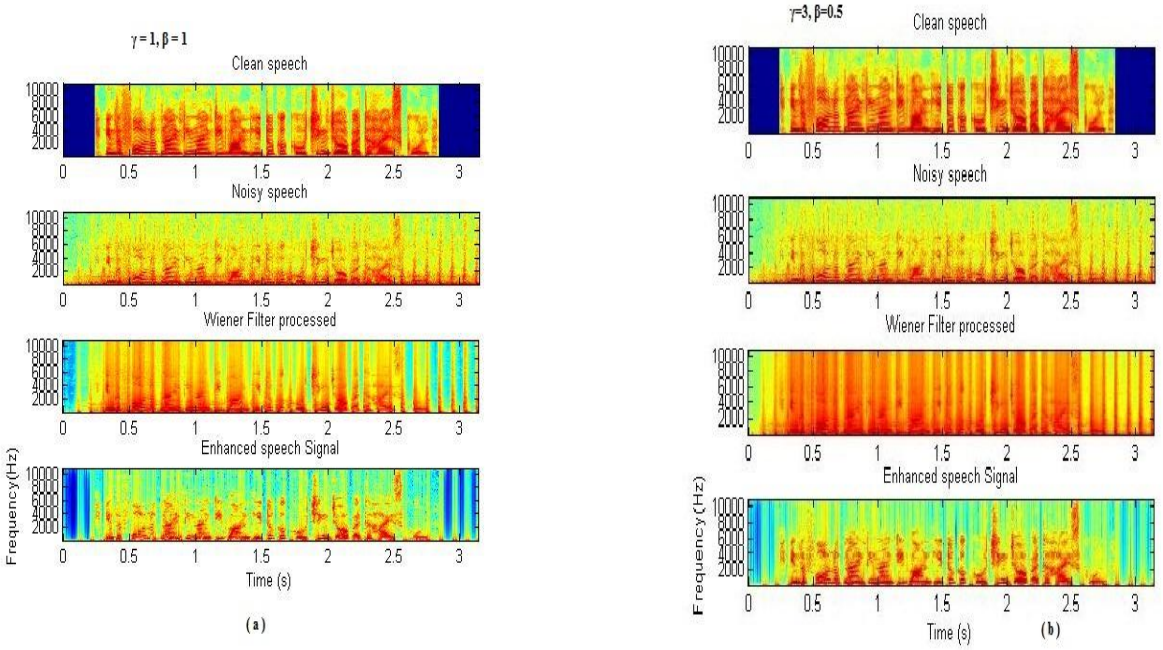


Fig. 2. Spectrograms showing the helicopter noise (SNR= 0 and -5 dB)

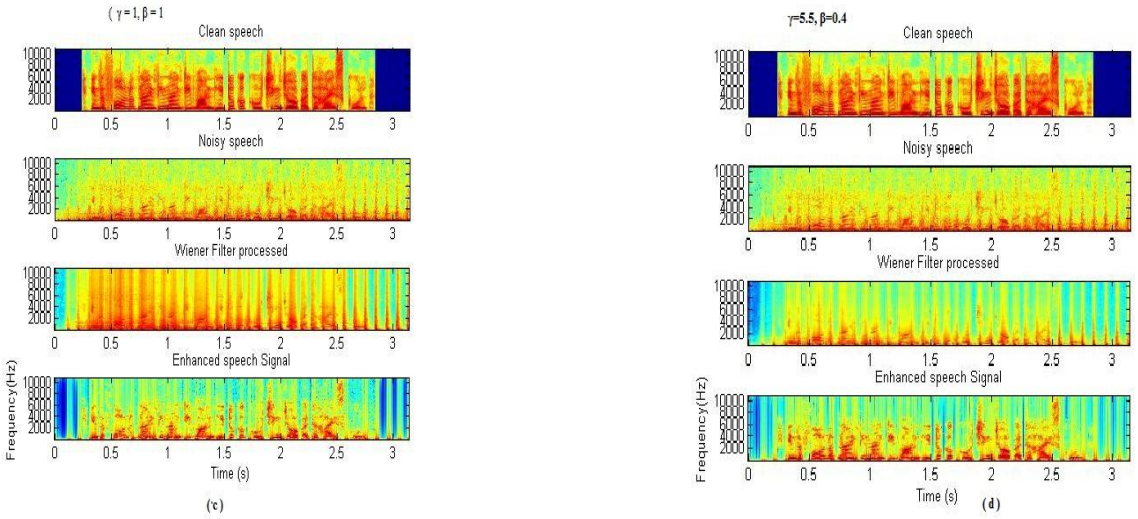


Fig. 3. Spectrograms showing the babble noise (SNR= 0 and 5 dB)

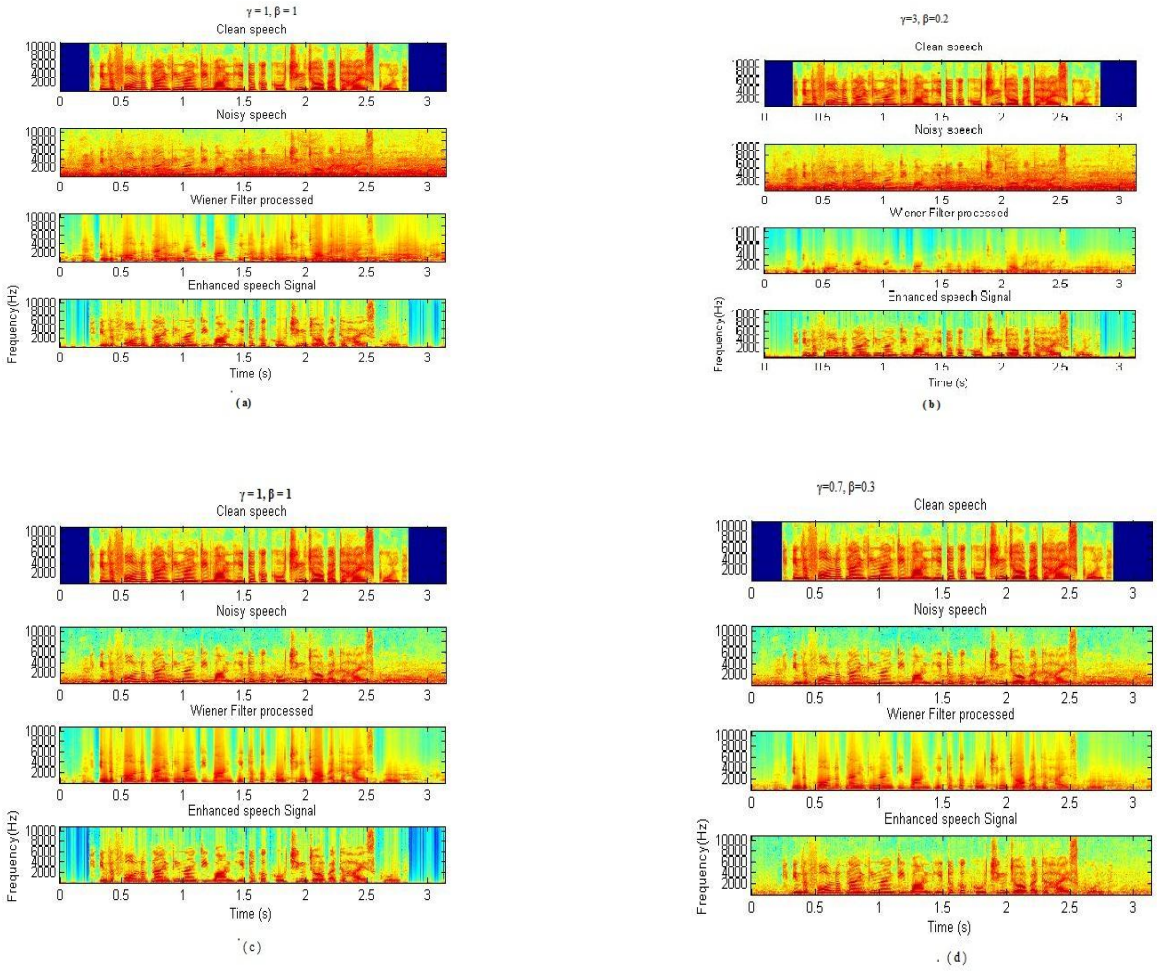


Fig. 4. Spectrograms showing the car noise (SNR= 0 and -5 dB)

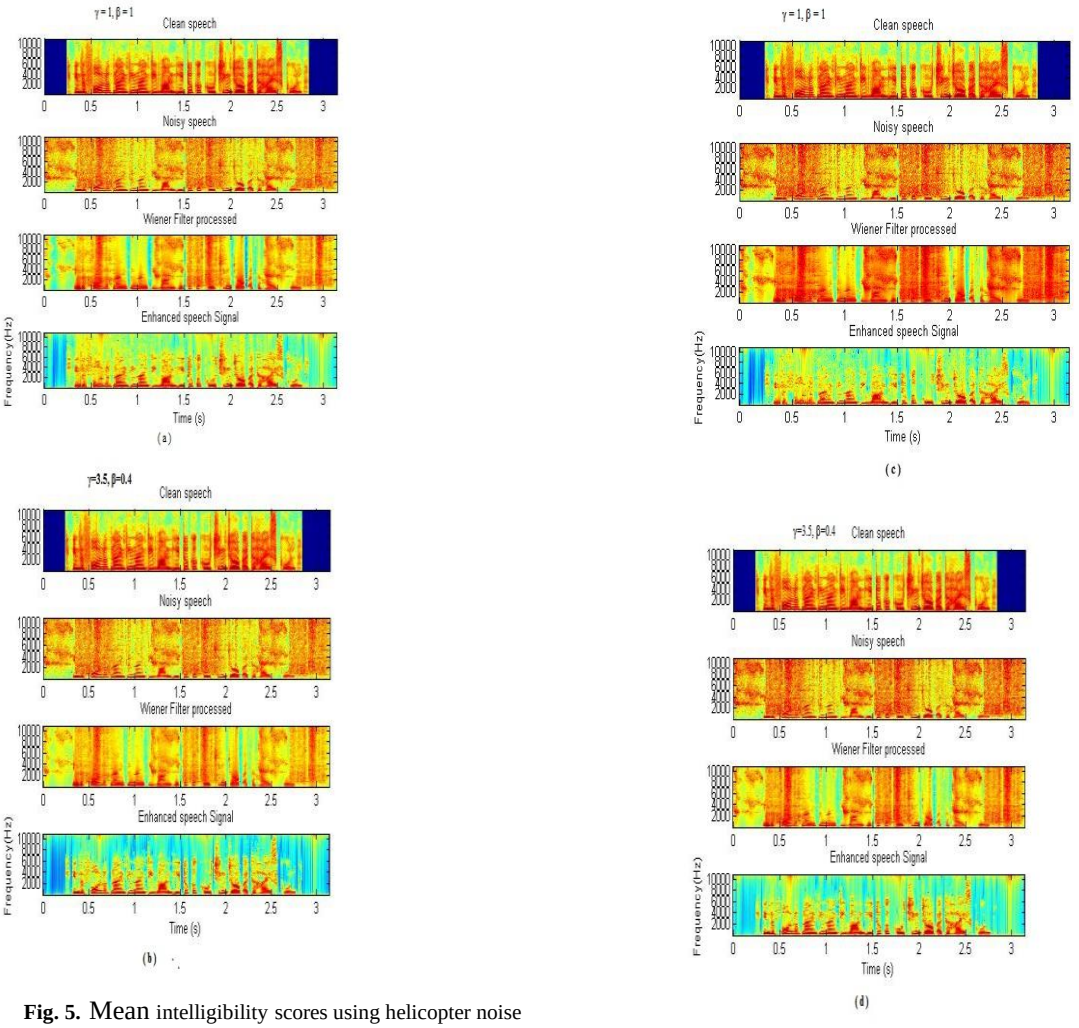


Fig. 5. Mean intelligibility scores using helicopter noise

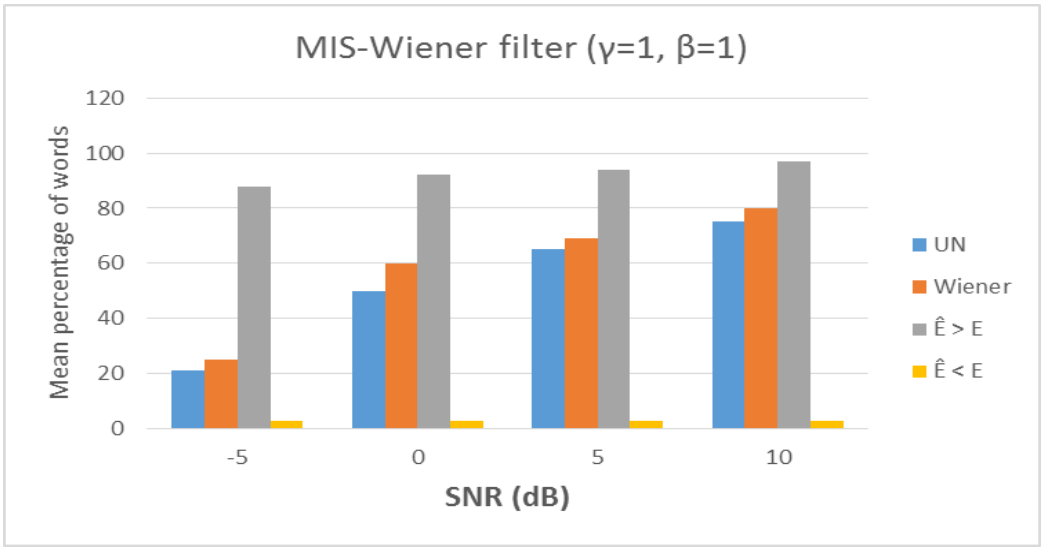
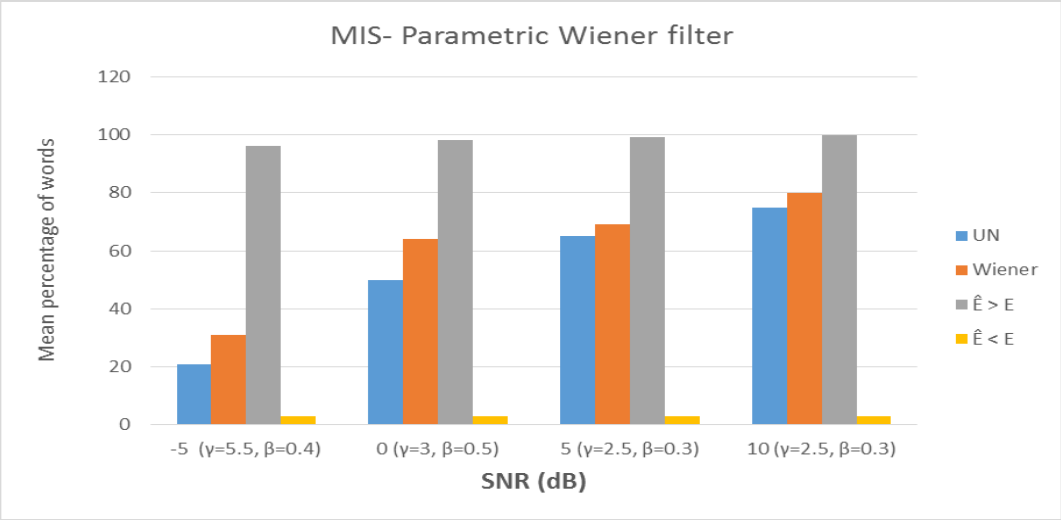


Fig. 6. Mean intelligibility scores using helicopter noise



IV. CONCLUSION

We have used the binary mask approach for parametric wiener gain filter using MATLAB. Subjective and objective tests were conducted for different values of γ and β for various background noises at 10dB, 5dB, 0dB and -5dB SNR values. The objective tests clearly indicate improvement in values of SSNR for random noise, babble noise, car noise and helicopter noise at 10dB, 5dB, 0dB and -5dB SNR values. The subjective results also shows an overall improvement in speech quality as well as intelligibility for random noise, babble noise, car noise and helicopter noise at 10dB,5dB, 0dB and -5dB SNR values. The results shows a significant improvement in single channel speech intelligibility even at low SNR values (-5dB).

REFERENCES

1. P.C. Loizou, Speech enhancement: Theory and Practice. Taylor & Francis Group, CRC Press,2013.
2. G. Kim and P. Loizou, "A new binary mask based on noise constraints for improved speech intelligibility" ,INTERSPEECH, 2010, Japan.
3. Ramesh Nuthakki, A. Sreenivasa Murthy, Naik D.C, "Single channel speech enhancement using a new binary mask in power spectral domain , " in Proc. of IEEE Intern. Conf -2018 (ICECA - 2018).
4. Siddala Vihari, A. Sreenivasa Murthy, Priyanka Soni and D. C. Naik, "Comparison of Speech Enhancement Algorithms" Elsevier, Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016).
5. Naik D.C, A. Sreenivasa Murthy, Ramesh Nuthakki, "Modified Magnitude Spectral Subtraction Methods for Speech Enhancement, " in Proc. of IEEE Intern. Conf -2017 (ICEECCOT-2017).
6. S. Rangachari and P. C. Loizou, "A noise-estimation algorithm for highly non-stationary environments," Speech Commun., vol. 48, pp. 220–231, 2006.
7. G. Kim and P. Loizou, "Why do speech enhancement algorithms not improve speech intelligibility?" in Proc. of IEEE Intern. Conf. on Acoust., Speech, Signal Processing, 2010, pp. 4738–4741.
8. Y. Hu and P. C. Loizou, "A comparative intelligibility study of single-microphone noise reduction algorithms," J. Acoust. Soc.Am., vol. 122, pp. 1777–1786, 2007.
9. S F Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSp-27, No. 2, pp. 113-120, Apr.1979.
10. Berouti, R. Schwartz and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," Proc.IEEE Int. Conf. Acoust., Speech, Signal Process., pp. 208-211, Apr. 1979.
11. Kim, Gibak, P. C. Loizou, "Improving Speech Intelligibility in Noise Using a Binary Mask That Is Based on Magnitude Spectrum Constraints "Volume: 17, Issue -12, Dec. 2010, IEEE Signal Processing letter.
12. Y. Hu and P. C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," Speech Commun., vol. 49, pp. 588–601, 2007.
13. 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, September 2009.
14. P. Scalart and J. Filho, "Speech enhancement based on a priori signal to noise estimation," in *Proc. of IEEE Intern. Conf. on Acoust.,Speech, Signal Processing*, 1996, pp. 629–632.
15. IEEE, "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio and Electroacoustics*, pp. 225–246, 1969.
16. Naik D.C, A. Sreenivasa Murthy, "A study of multiband spectral subtraction for speech enhancement in magnitude and power spectral domains, " JETIR, Vol.6, Issue 6, june 2019.
17. Y. Lu and P. Loizou, " speech enhancement by combining statistical estimation of speech and noise," in Proc.of IEEE Intern.Conf, on Acoust,Speech, Signal Processing,2010, PP.4754-4757.

AUTHORS PROFILE

Ramesh Nuthakki received the B.Tech degree in ECE and master's degree in Digital systems and communication engineering from R.E.C (NIT) Calicut University, Calicut in 1999. From 1999-2005 he worked as a senior engineer in VSNL(TCL), Chennai and later joined the MNC Wipro Technologies and worked as a senior software engineer from 2005-2008 and he had also been to Canada and deployed as a project lead for Nortel Networks, After that he worked in the Esteemed Organization IBM as a Associated Project Manager from 2008-2011. In the year 2012 he joined as an Asst. Professor in Atria Institute of Technology, Bangalore and is continuing till date. His area of interest includes speech signal processing, and Networking. He is member of IEEE and IETE.

Dr. Sreenivasa Murthy is a research guide/superviser in department of Electronics and Communication Engineering, University Visvesvaraya College of Engineering, Bangalore. Initially he worked in industry (BEL, Bangalore) for a period of 7 years. He is currently working as a Professor of ECE and mentoring M.Tech students and Ph D scholars. He has a teaching experience of more than three decades. He has completed Ph D in IISc, Bangalore. His area of research includes Digital Signal processing, Speech processing, Image processing, probability theory & stochastic processes and information theory & coding.