International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-5, January 2020

Enhancement of Speech Intelligibility using Binary Mask Based on channel selection criteria

Ramesh Nuthakki, Sreenivasa Murthy A, Naik D C

Abstract: ***Most of the existing noise reduction algorithms used in hearing aid applications apply a gain function in order to reduce the noise intervention. In the present paper, we study the effect of the two types of speech distortions introduced by the gain functions. If these distortions are properly controlled large gains in intelligibility can be obtained. The sentences were corrupted by various kinds of noises i.e. babble noise, car noise, helicopter noise and random noise and processed through a noise-reduction algorithm. Subjective tests were conducted with normal hearing listeners by presenting the processed speech with controlled distortions. The method proposed by Kim et al uses the wiener filter. Here in this paper, we have used the parametric wiener filter. The experimental results clearly indicated improvement in intelligibility at 0dB, -5dB, +5dB and 10dB input signal-to-noise (SNR) values in short-time objective intelligibility (STOI) and Segmental signal-to-noise ratio (SSNR) objective measures.***

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

I. INTRODUCTION

It's a great enthusiasm to improve the speech quality and intelligibility of a noisy audio. Speech enhancement is very important in various applications such as mobile communications, speech recognition systems, voice over IP and in hearing aids. The most common thing about these applications is that they are mostly used in noisy environments such as shopping malls, railway stations and market places. For example, a person speaking on a mobile phone is likely to get disturbed by natural sounds from the back ground [7]. The back ground noise is considered as a disturbance in most cases but it could be pleasant in some exceptional cases such as the musical background in a restaurant [1,18]. This noise could be pleasing to normal hearing listeners but for the hearing impaired people it is considered as a disturbance. In such situations more sophisticated speech enhancement algorithms are required to lighten the problem. The basic aim of the speech enhancement algorithms [1] is to identify the noise and it should be attenuated effectively without damaging the target signal. Most of the noise reduction algorithms are found to improve speech quality and hearing comfort. But in contrast, very few algorithms were developed that can improve speech intelligibility. For more than three decades it has been found to be difficult to develop an algorithm that can substantially improve speech intelligibility for normal hearing as well as hearing impaired listeners. Recent studies conducted with normal hearing listeners [1-2,5,9-10] shows that the algorithms that are modified to function in noisy environments have proved to be effective in improving speech intelligibility. Usually the noise reduction algorithms used in commercial hearing aids includes two continuous stages of processing as shown in Fig.1. In the first stage, it does the signal detection and analysis in order to identify the presence of speech and noise in each band. To estimate the modulation rate [5], modulation depth, or/and SNR in each frequency band, detectors are used. Depending on the estimated modulation rate or SNR of the each band ascertained in the first stage, the mixture envelope is subjected to gain reduction in the second stage. The amount of gain reduction is commonly inversely proportional to the SNR estimated in each channel. The Weiner gain algorithm applies a gain to the spectral envelopes in proportion to the estimated SNR in each frequency bin. The spectral bins with high SNR receive a high gain (close to 1) whereas the spectral bins with low SNR receive a low gain (close to 0). Normally, the gain functions used in most noise reduction algorithms introduces two types of distortions namely amplification and attenuation distortions [5,7]. The amplification distortion occurs when the target signal is overestimated and attenuation distortion occurs when the target signal is under estimated. The effect of these distortions cannot be considered to be equal, in reality there has to be a correct balance between these two distortions.

In the present study, we estimate the impact of these two types of distortions on the intelligibility of noise-suppressed speech using wide band speech corrupted by either steady noise or competing talker. Through a conventional noise reduction algorithm, the wide band speech is processed besides controlling the two types of distortions. We finally combine the signals containing either only amplification distortion or only attenuation distortion. Generally the processed signal contains both the distortions, but the individual contribution of each distortion on speech intelligibility is not known. If these distortions can be controlled large gains in intelligibility can be obtained.

Manuscript received on December 10, 2020.
Revised Manuscript received on December 20, 2020.
Manuscript published on January 30, 2020.
* 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
Naik D C, Research Scholar, Department of Electronics and Communication Engineering, University Visvesvaraya College of Engineering, Bengaluru, India. Chethan.naik24@gmail.com

II. APPLYING CONSTRAINTS ON THE ESTIMATED MAGNITUDE SPECTRA

As referred earlier, most of the noise-reduction algorithms used for hearing aids includes a gain reduction process where the mixture spectrum is multiplied by a gain function to suppress the background noise. The quantity of the gain reduction depends on the identified modulation rate or estimated SNR among others. When a gain function is applied to the mixture spectra, it produces either amplification distortion or attenuation distortion. In order to study the effect of gain-induced distortions introduced by the noise-reduction algorithms on speech intelligibility, we need to demonstrate a relationship between distortions and intelligibility or otherwise we need to adopt a suitable intelligibility measure. Such intelligibility measure perhaps gives a valuable information as to whether we must design algorithms that would reduce the amplification distortion or attenuation distortion or both and also up to what extent. For our study, we consider an intelligibility measure proposed by Ma et al [19] to equate highly (ϒ= 0.81) with the intelligibility of the noise-suppressed speech. The intelligibility measure is indicated as the frequency-weighted segmental SNR measure ($fwSNR_{seg}$) and is calculated using the following equation.

$$fwSNRseg = \frac{10}{T}\sum_{ni=0}^{T-1}\frac{1}{\sum_{f=1}^{F}W(f,n_i)} \quad X \quad \sum_{f=1}^{F}W(f,n_i)\log_{10} SNR_{ESI}(f, ni) \qquad (1)$$

where W(f, n_i) is the weight placed on the f^{th} frequency band and time index frame n_i. F is the number of frequency bands and $fwSNR_{seg}$ is the frequency weighted segmental SNR. T denotes the total number of time frames in the signal and $SNR_{ESI}(f, n_i)$ represents the signal-to-residual spectrum ratio.

$$SNR_{ESI}(f, n_i) = \frac{|C(f,n_i)|^2}{(|C(f,n_i)-|\hat{C}(f,n_i)|)^2} \qquad (2)$$

where |c(f, n_i)| denotes the clean magnitude spectrum and |ĉ(f, n_i)| denotes the signal magnitude spectrum estimated by the noise reduction algorithm. By applying the gain function to the noise speech spectrum, |ĉ(f, n_i)|was computed. Here we consider $SNR_{ESI}(f, n_i)$ as a local metric evaluating the regularized distance between the true spectrum envelope and the processed spectrum. It is evident that when the noise-suppressed magnitude spectrum |ĉ(f, n_i)| is closer to the true magnitude spectrum |c(f, n_i)|, the value of $SNR_{ESI}(f, n_i)$ metric is higher and therefore the value of $fwSNR_{seg}$ measure will be higher [eq.(1)]. On the other hand $SNR_{ESI}(f, n_i)$ metric can be demonstrated as a function of the ratio of the estimated spectra to the true magnitude spectra.

$$SNR_{ESI}(f, n_i) = \frac{1}{(1-\frac{|\hat{C}(f,n_i)|}{|C(f,n_i)|})^2} \qquad (3)$$

The values of $SNR_{ESI}(f, n_i)$ has been divided into different regions on either the ratio $\frac{|\hat{C}(f,n_i)|}{|C(f,n_i)|}$ is smaller or greater than 1 or smaller or greater than 2. This figure gives us the perception about the contributions of the two distortions on the value of the SNR_{ESI}. Based on the distortions introduced, the figure was divided into three regions.

Region I. In this region, $|\hat{C}(f,n_i)| \leq |C(f, n_i)|$, indicating only attenuation distortion.

Region II. In this region, $|C(f, n_i)| < |\hat{C}(f,n_i)| \leq 2|C(f, n_i)|$, indicating amplification distortion ranging from 0 to 6.02 dB.

Region III. In this region, $|\hat{C}(f,n_i)| > 2|C(f, n_i)|$, indicating amplification distortion in excess of 6.02 dB.

The three regions are clearly shown in Fig.2.From the figure we can conclude that the combinations of Regions I & II denoted as Region 1+2, we will get the following constraint

$$|\hat{C}(f,n_i)| \leq 2\,|C(f,n_i)| \qquad (4)$$

In Fig.2, the relationship between the two distortions and their possible effect on the speech intelligibility is clearly shown. As per the figure, the envelope distortions should be confined within the Regions 1 & 2 in order to get higher values for the SNR metric since this measure accepts large values in Region I & II [always positive or 0]. The hypothesis made here is that when SNR_{ESI} metric obtains large values over all bands, it leads to overall $fwSNR_{seg}$ increase resulting in high intelligibility. Contrarily, the amplification distortion which is more than 6dB [in Region III] can be detrimental to speech intelligibility [Because the SNR_{ESI} metric accepts small values in Region III, and in dB, it is –ve] and thereafter it should be minimized. Taking these two observations into consideration, we can state that for improving speech intelligibility in noise-reduction algorithms, the amplification distortions should be restrained in such a way that they should not exceed 6dB which means they need to be confined within Regions I & II. In our further study we check the theory that if the envelope distortions initiated by the gain function are confined to fall within the Regions I & II, considerable improvement in intelligibility can be expected.

III. INFLUENCE OF DISTORTIONS ON SPEECH INTELLIGIBILITY

In this module, the noise corrupted sentences were first processed through a conventional noise reduction algorithm i.e parametric wiener filter algorithm and we observe the two types of distortions initiated by the gain function. Then the signal was synthesized appropriately either by amplification distortion alone, attenuation distortion alone or both the distortions. We strictly confine the distortions to fall within one of the three regions or combinations thereby as shown in the Fig.2. The synthesized signals are given to the normal-hearing listeners for identification. For the listening tests eight normal hearing listeners were taken. IEEE Sentences were used for this test as they are phonetically balanced and also have relatively low word-context predictability. The sentences were recorded at a sampling rate of 16 kHz in a sound proof room.

The sentences were corrupted by speech-shape noise (SSN) and a single-talker masker at 10, 5, 0dB,-5dB SNR levels.

A. Signal Processing

For this experiment, we have chosen the parametric Weiner algorithm as the computation process is less and it is easy to implement. It has been proved by Hu and Loizou [10] to be equally effective both in terms of speech-quality and intelligibility compared to other advanced noise-reduction algorithms. The corrupted sentences were segmented into 20 ms frames with a 50% overlap between the adjacent frames. Each speech frame was Hann-windowed and a 500-point discrete Fourier transform was calculated. Let us denote $Y(f, n_i)$ as the noisy spectrum at the time frame n_i and frequency band f. Then the estimate of the signal magnitude spectrum $|\hat{C}(f, n_i)|$ is obtained by multiplying $|Y(f, n_i)|$ with parametric Weiner gain function $G(f, n_i)$ as follows:

$$|\hat{C}(f, n_i)| = G(f, n_i).|Y(f, n_i)| \quad (5)$$

The parametric Weiner gain[1,4] function is computed based on the following equation

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

Where SNR_{prio} is the a priori SNR estimated[16] using the following recursive equation

$$SNR_{prio}(f, n_i) = \alpha . \frac{\hat{C}^2(f, n_i - 1)}{\hat{P}_D^2(f, n_i - 1)} + (1-\alpha).\max\left[\frac{\gamma^2(f, n_i)}{\hat{P}_D^2(f, n_i)} - 1, 0\right] \quad (7)$$

Where the estimate of the background noise power spectrum is denoted by $\hat{P}_D^2(f, n_i)$ and α is the smoothing constant whose value is set to α= 0.98. To estimate the noise power variance, $\hat{P}_D^2(f, n_i)$ in Eq. 7, the noise estimation algorithm introduced in [8] was used here.

Complying with equation (5), an inverse DFT was applied to the processed magnitude spectrum $|\hat{C}(f, n_i)|$ applying the phase of the noisy speech spectrum. The overlap-and-add technique was lastly used to synthesize the noise suppressed signal. The parametric Weiner filter used in this study was evaluated from the mixture envelopes and it is different from the Weiner filter used in the study by Levitt et al[5]. In eq(5), no constraints were imposed on the two types of distortions that can occur while applying the parametric Weiner gain function to the corrupted speech. Because of this, the parametric Weiner processed sentences acted as one of the two control conditions. For the other conditions we have considered the knowledge of the clean speech spectrum. This was required so as to apply the constraints and analyze the effect of these two distortions on speech intelligibility. Therefore in order to impose the constraints, the estimated magnitude spectrum $|\hat{C}(f, n_i)|$ was weighted against the real speech spectrum $|C(f, n_i)|$ for each time-frequency unit (f, n_i). The T-F units satisfying the constraints were withheld and the T-F units which oppose the constraints will be removed. In order to implement the Region I constraint, the modified magnitude spectrum $|\hat{C}(f, n_i)|$, was computed as follows:

$$|\hat{C}_M(f, n_i)| = \begin{cases} |\hat{C}(f, n_i)| \text{ if } |\hat{C}(f, n_i)| \leq |C(f, n_i)| \\ 0 \quad \text{else} \end{cases} \quad (08)$$

To the above selection of T-F units in the Region I, an inverse DFT was applied to the modified spectrum $|\hat{C}_M(f, n_i)|$ using the phase of the noisy speech spectrum. Lastly the overlap and add technique was used to reconstruct the noise-suppressed signal. The percentage of bins falling in the three regions are shown in the Table 1. Almost half of the bins fall in Region 1 that is indicated by attenuation distortion whereas the other half falls in Region3 characterized by amplification distortion which is in excess of 6.02dB. In Region 2, a very small percentage of bins are found to fall which is indicated by low amplification distortion that is less than 6.02dB.

B. Methods and Procedure

Ten normal hearing listeners were employed for the listening experiments [3,6-7]. The listeners partake in overall 96 conditions (= 4 SNR levels x 4 types of noises x 6 processing conditions). The speech is processed with speech enhancement algorithm for different SNR levels with processing conditions including 1) No constraints imposed, 2) Region I Constraints, 3) Region II constraints, 4) Region I + II constraints, and 5) Region III constraints. The listeners were also granted unprocessed stimuli. The processed speech files were also given to the listeners. The tests were conducted in a sound proof room. To get familiarize with the listening procedure, each listener was made to listen to a set of noise sentences before the test. One sentence list was used for each condition. These sentences were selected from the IEEE data base. The presentation of the sentences were completely randomized. The listeners were asked to identify and write down the words they heard and the intelligibility was estimated by counting the number of words identified correctly by the listeners. All these sentences were phonetically balanced. At 10dB, 5dB, 0dB and -5dB SNR values, a different noisy speech was generated by adding various noises like Helicopter, babble, Car and random. The spectrograms are shown in the figures 3 and 4. From the spectrograms, it is clear that both in terms of voiced segments and unvoiced segments, the pitch and formants are completely recovered. This indicates that there is an improvement in speech intelligibility for helicopter noise.

IV. OVERALL INTELLIGIBILITY AND QUALITY MEASURES

A. *SSNR Objective Measures*

In this paper we have used the average segmental SNR [1,3]. The reason for choosing this method is the accuracy it provides compared to other parameters. It is the most widely used objective measure. When the value of Seg.SNR is higher, the enhanced speech signal possess more signal power than the noise power. The speech signal is divided into 20ms frames and the distortion measure was calculated between the original and the processed speech signals. Then the speech distortion was calculated by equalizing the distortion measures of each frame in the time domain. We have considered the SSNR measure in time domain in this paper.

The tables clearly shows an improvement in SSNR values for helicopter noise. While computing SNR_{seg}, the signal energy at silence intervals was comparatively low resulting in large negative SNR_{seg} values. The average segmental SNR is written as follows.

$$SSNR = \frac{10}{N}\sum_{n=0}^{N-1} 10\log_{10}\frac{\sum_{i=in}^{in+i-1} C(i)^2}{\sum_{i=in}^{in+i-1}\left(C(i) - \hat{C}(i)\right)^2} \quad (9)$$

Where C and $\hat{C}$ denotes the clean and enhanced speech signal, N 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. The SSNR results are mentioned in the table 2. This measure yields good results using parametric wiener filter for stationary as well as non-stationary types of noises.

B. *STOI Objective Measures*

The suggested method is a result of the clear and processed speech indicated by c and y respectively. This method [20-21] is designed for a sample-rate of 10000Hz so as to cover the relevant frequency range for speech intelligibility. The signals at other sample rates should be resampled. Also it is presumed that the clean and the processed signal are both time-aligned. First both the signals were segmented into 50% overlapping in order to obtain a T-F-representation. The frames were Hann-windowed with a length of 256 samples, each individual frame is zero-padded upto 512 samples and Fourier transformed. An one-third octave band analysis is carried out by grouping DFT-bins. A total of 15 one-third octave bands were used with the lowest center frequency is set equal to 150Hz. Let $\hat{C}(f, n_i)$ represent the f^{th} DFT-bin of the n_i^{th} frame of the clean speech. The average of the j^{th} one-third band, specified as a T-F unit is determined as

$$C_j(n_i) = \sqrt{\sum_{f=f_1(j)}^{f_2(j)-1} |\hat{C}(f, n_i)|^2} \quad (10)$$

Where f_1 and f_2 indicate the one-third octave band edges rounded to the nearest DFT-bin. Similarly, the processed speech denoted by $Y_j(n_i)$ can also be attained. The intermediate intelligibility measure for one TF-unit, let us say $d_j(n_i)$ depends on the region of N consecutive TF-units from both $C_j(n_i)$ and $Y_j(n_i)$ where $n_i \in M$ and $M=\{(n_i - N+1), (n_i - N+2), \ldots\ldots n_i - 1, n_i\}$. Initially, a local standardization procedure was applied by weighting all the T-F units from $Y_j(n_i)$ with a factor $\alpha = \left(\sum_{n_i} C_j(n_i)^2 \middle| \sum_{n_i} Y_j(n_i)^2\right)^{1/2}$ in such a way that its energy equals to the clean speech energy inside the TF-region. After that $\alpha Y_j(n_i)$ is sniped to lower bound the SDR, which is defined as $SDR_j(n_i)$ Eq.

$$SDR_j(n_i) = 10\log_{10}\left(\frac{C_j(n_i)^2}{(\alpha Y_j(n_i) - C_j(n_i))^2}\right) \quad (11)$$

Therefore,

$$\bar{Y} = \max(\min(\alpha Y, C+10^{-\beta/20}C), C-10^{-\beta/20}C) \quad (12)$$

Where $\bar{Y}$ denotes the standardized and clipped TF-unit where as β denotes the lower SDR bound. The frame and one-third octave band indices are left out for notational convenience. The intermediate intelligibility measure is explained as an estimate of the linear correlation coefficient between the clean and modified processed TF-units,

$$dj(n_i) = \frac{\sum_{n_i}\left(C_j(n_i) - \frac{1}{N}\sum_l C_j(l)\right)\left(\bar{Y}_j(n_i) - \frac{1}{N}\sum_l \bar{Y}_j(l)\right)}{\sqrt{\sum_{n_i}\left(C_j(n_i) - \frac{1}{N}\sum_l C_j(l)\right)^2 \sum_{n_i}\left(\bar{Y}_j(n_i) - \frac{1}{N}\sum_l \bar{Y}_j(l)\right)^2}} \quad (13)$$

Where $l \in M$. Subsequently the ultimate OIM is specified by the average of the intermediate intelligibility measure over all bands and frames,

$$d = \frac{1}{JM}\sum_{j,n_i} dj(n_i) \quad (14)$$

Where M represents the total number of frames and j represents the number of one-third octave bands.We have taken different values of $N \in [20,30,40,50,60]$ and $\beta \in [-\infty, -30,-20,-15,-10]$ for this experiment. When β = 15 and N = 30, maximum correlation was obtained.

The STOI is evaluated by positioning the clean and the enhanced speech signals for various types of noises and for different input SNR values. The STOI results are mentioned in the table 2. This measure yields good results using parametric wiener filter for stationary as well as non-stationary types of noises.

C. Quality Subjective Measures

For the subjective tests [4,11,17] we have taken a total of 10 listeners, 5 female and 5 male. All the subjects were asked to listen to an enhanced and noisy speech signals in a random way. The subjective tests are dependent on certain parameters like background quality (BAK), overall signal quality and signal quality (SIG). For the above mentioned parameters, the subjects were asked to give scores anywhere from 1 to 5. At 0dB, -5dB, +5dB and 10dB SNR values, it has been identified that the speech signals were corrupted by different kinds of noises such as car noise, babble noise, random noise and helicopter noise. The scores thus obtained are taken and shown in the table 3. Clearly these tests shows an improvement in speech quality for the random, car, babble and helicopter noises.

D. Intelligibility hearing Tests

Intelligibility listening tests [12,13,14,15,18] were conducted in order to assess the intelligibility of the processed speech. The sentences were collected from Indian English data base and IEEE. By adding helicopter noise, car noise, random noise and babble noise at -5dB, 0dB ,5dB and 10dB SNRs, a noisy speech was generated. Prior to the tests, each listener was made to listen to a set of noise corrupted sentences. Then the listeners were asked to identify the words from the estimated speech signal. The performance was assessed by counting the number of words identified correctly by the listeners and is shown in the figures 5 and 6.

E. Results and Discussion

The subjective results are mentioned in the table 3, figures 5 and 6. It is clear from the results that there is an improvement in intelligibility for babble noise, car noise, random noise and helicopter noise at 0dB, -5dB, +5dB and 10dB input SNR values respectively. ((10))

To obtain the results, we have taken into consideration the mean percentage of words identified correctly by the normal hearing listeners. Reg1 shows an improvement in intelligibility with parametric wiener filter.

The unprocessed speech scores are represented by UN. According to the figures 4a and 4b, when the proposed noise constraints were imposed, the performance at 0dB, -5dB, +5dB and 10dB input SNR's improved from 10%, 30%, 40% and 40% with unprocessed speech to 88%, 92%, 94% and 95% respectively using wiener filter and degraded to nearly zero with failed constraint conditions. Correspondingly in parametric wiener filter also, the performance at 0dB, -5dB, +5dB and 10dB SNR levels increased from 10%, 30%, 40% & 40% with unprocessed speech to 95%, 98%, 100% & 100% and degraded to nearly zero with failed constraint conditions. While comparing the figures 4a & 4b at SNR's 0dB, -5dB, +5dB and 10dB, it is evident that there has been an improvement in word count from 88 to 95, 92 to 98, 94 to 100 and 95 to 100. The results obtained clearly demonstrates that the proposed binary mask holds good in parametric wiener filter compared to wiener filter for the random, car, babble and helicopter noises.

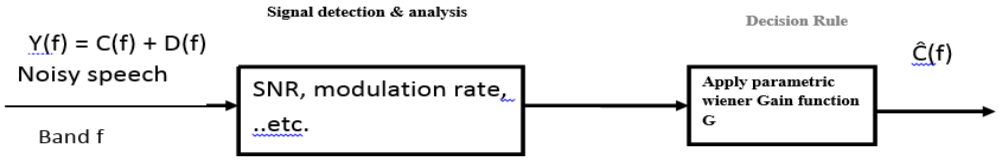

Fig. 1.Signal-processing stages involved in noise-reduction algorithms for hearing-aid applications.

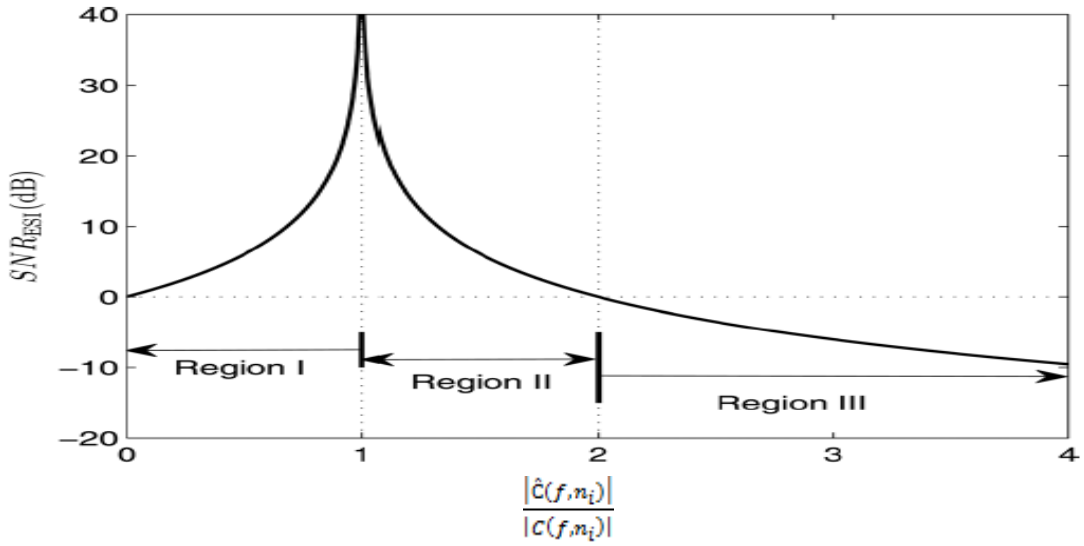

Fig. 2.Represents $SNR_{ESI}(f, n_i)$ as a function of the ratio of the estimated to clean magnitude spec

Table- 1: Percentage of bins falling in the three regions

Noise	I/P SNR (dB) levels	γ	β	Bins falling in Region I using parametric wiener (wiener filter, γ=1, β=1)	Bins falling in Region II using parametric wiener (wiener filter, γ=1, β=1)	Bins falling in Region III using parametric wiener (wiener filter, γ=1, β=1)
Random Noise	10	3	0.3	27.3747 (54.9993)	1.7861 (0.4291)	70.8392 (44.5717)
	5	1.8	0.5	36.3913 (65.4939)	1.3163 (0.8989)	62.2924 (33.6072)
	0	3	0.3	15.7254 (60.4713)	2.3463 (0.7168)	82.9181 (38.8119)
	-5	3	0.3	9.9760 (54.9993)	3.1206 (0.4291)	86.9034 (44.5717)
Babble Noise	10	3.5	0.3	70.2008 (74.6072)	1.5764 (1.4771)	28.2228 (23.9157)
	5	3.5	0.3	50.1632 (72.2931)	2.0178 (1.4015)	47.8191 (26.3054)
	0	1.8	0.5	45.2712 (64.1123)	2.5234 (1.5212)	52.2012 (34.3512)

	-5	0.8	0.4	15.8326 (54.4602)	5.5529 (2.1292)	78.6145 (43.4105)
Helicopter Noise	10	5.5	0.4	55.7670 (66.5924)	1.3250 (1.6951)	42.9079 (31.7126
	5	5.5	0.4	48.1636 (62.4432)	1.5232 (1.8524)	50.3132 (35.7044)
	0	5.5	0.4	40.4269 (57.3912)	1.7803 (2.0000)	57.7928 (40.5910)
	-5	5.5	0.4	32.3441 (51.7012)	2.2647 (2.2812)	65.3912 (46.0123)
Car Noise	10	2.5	0.3	33.7930 (61.7723)	1.6594 (1.1524)	64.5476 (37.0753)
	5	2.5	0.3	27.8285 (56.9122)	1.9916 (1.2922)	70.1799 (41.7956)
	0	3.5	0.4	31.4154 (52.0221)	1.6718 (1.5428)	66.9129 (46.4350)
	-5	3.5	0.4	25.8428 (47.3507)	2.1300 (1.8546)	72.0272 (50.7947)

Table- 2: Objective measures of Region I constraints using parametric wiener and wiener Filter

Noise	I/P SNR (dB)	γ	β	Seg.SNR(dB) using Parametric Wiener Filter. (using Wiener Filter γ=1, β=1, Kim's approach)	STOI using Parametric Wiener Filter. (using Wiener Filter γ=1, β=1, Kim's approach)
Random Noise	10	3	0.3	4.5707 (12.5140)	0.918 (0.870)
	5	1.8	0.5	10.8171 (9.8824)	0.880 (0.835)
	0	3	0.3	10.7254 (8.3282)	0.887 (0.811)
	-5	3	0.3	3.3339 (8.8072)	0.880 (0.792)
Babble Noise	10	3.5	0.3	13.4709 (9.7321)	0.901 (0.832)
	5	3.5	0.3	9.0214 (7.3420)	0.899 (0.846)
	0	1.8	0.5	1.0762 (4.2269)	0.854 (0.803)
	-5	0.8	0.4	0.6006 (0.3950)	0.747 (0.718)
Helicopter Noise	10	5.5	0.4	15.0383 (6.5301)	0.898 (0.855)
	5	5.5	0.4	8.6505 (7.0335)	0.887 (0.835)
	0	5.5	0.4	8.5946 (5.1451)	0.865 (0.816)
	-5	5.5	0.4	8.8196 (7.2561)	0.850 (0.786)
Car Noise	10	2.5	0.3	9.2352 (12.0312)	0.945 (0.894)
	5	2.5	0.3	6.9329 (7.5973)	0.931 (0.873)
	0	3.5	0.4	8.0527 (7.1106)	0.908 (0.853)
	-5	3.5	0.4	7.6511 (7.3565)	0.883 (0.813)

Table- 3: Subjective measures of R1 constraints using parametric wiener filter and wiener Filter

Noise	I/P SNR (dB) levels	γ	β	BAK { Subjective measures using parametric wiener (wiener filter, γ=1, β=1) }	SIG { Subjective measures using parametric wiener (wiener filter, γ=1, β=1) }	OVL { Subjective measures using parametric wiener (wiener filter, γ=1, β=1) }
Random Noise	10	3	0.3	4.0(3.7)	4.4 (4.0)	4.5 (4.1)
	5	1.8	0.5	3.9(3.7)	4.3(4.1)	4.7(4.2)
	0	3	0.3	3.8(3.6)	4.0(3.7)	4.1(3.9)
	-5	3	0.3	2.3(2.1)	2.4(2.3)	3.0(2.6)
Babble Noise	10	3.5	0.3	4.3(4.0)	4.5(4.1)	4.8(4.2)
	5	3.5	0.3	3.9(3.8)	4.3(4.0)	4.7(4.1)
	0	1.8	0.5	3.9(3.8)	4.2(3.7)	4.3(3.9)
	-5	0.8	0.4	2.6 (2.5)	2.9 (2.7)	3.1 (2.9)
Helicopter Noise	10	5.5	0.4	4.3(4.1)	4.6(4.2)	4.8(4.3)
	5	5.5	0.4	3.9(3.6)	4.3(4.0)	4.7(4.1)
	0	5.5	0.4	3.9(3.1)	4.3(3.3)	4.7(3.7)
	-5	5.5	0.4	3.8(2.2)	4.2(3.0)	4.4(3.5)
Car Noise	10	2.5	0.3	4.2(4.0)	4.4(4.1)	4.7(4.1)
	5	2.5	0.3	3.9(3.3)	4.2(3.6)	4.6(4.0)
	0	3.5	0.4	3.8(3.2)	3.9(3.6)	4.2(3.7)
	-5	3.5	0.4	2.4(2.3)	3.1(2.7)	4.0(3.4)

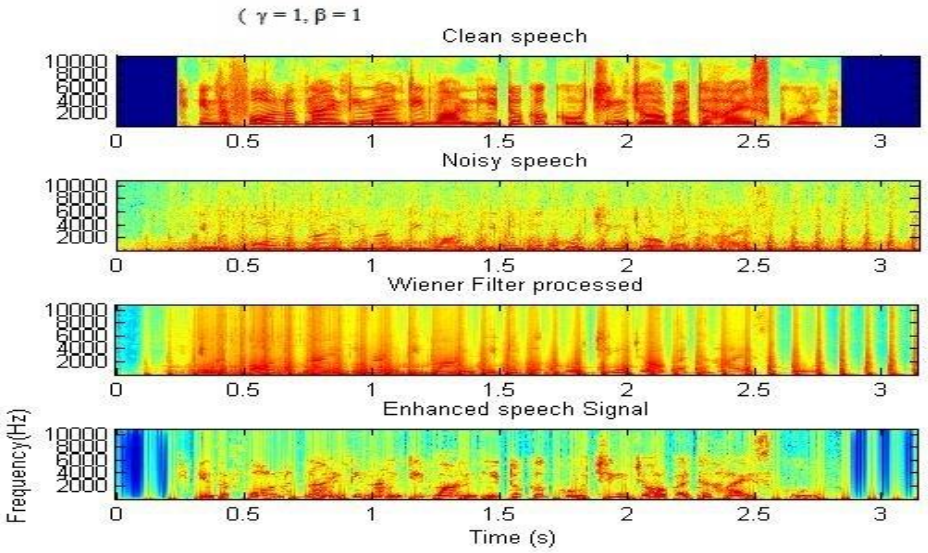

Fig.3. Spectrograms of the clean signal, noisy signal in -5 dB SNR Helicopter and signal processed by the Wiener filter after imposing the constraints in Region I

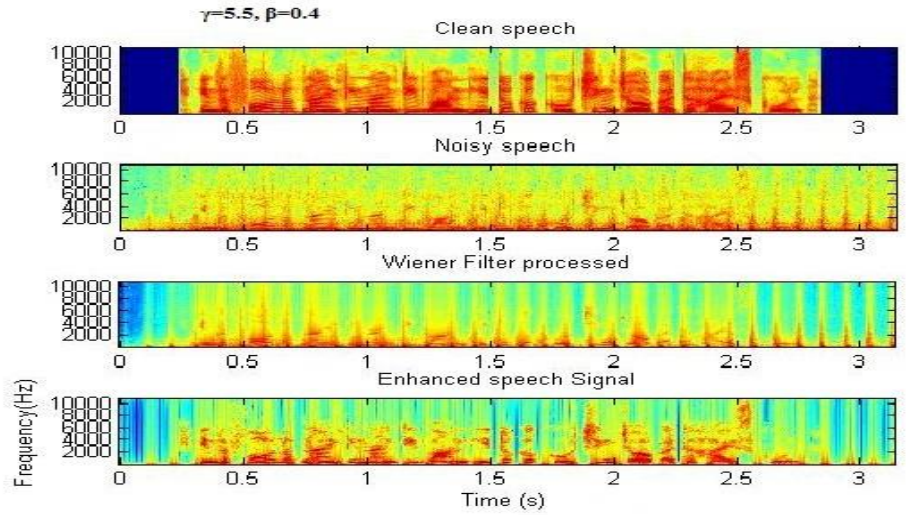

Fig.4.Spectrograms of the clean signal, noisy signal in -5 dB SNR Helicopter, signal processed by the Parametric Wiener filter after imposing the constraints in Region I

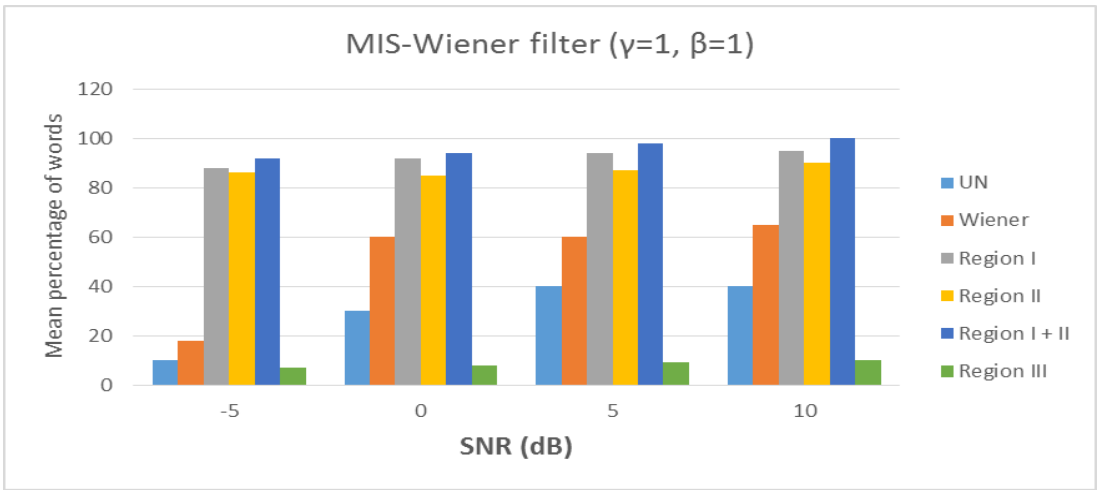

Fig.5. Mean intelligibility scores using helicopter noise

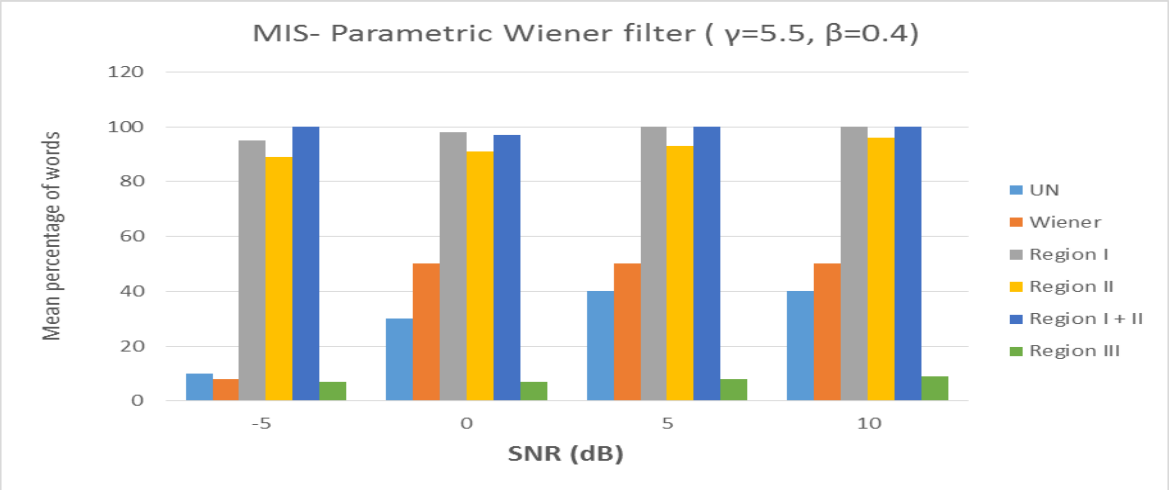

Fig.6. Mean intelligibility scores using helicopter noise

V.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 and STOI for random noise, babble noise, car noise and helicopter noise at 10dB, 5dB, 0dB and -5dB input 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. Ramesh Nuthakki, A. Sreenivasa Murthy "Enhancement of Speech Intelligibility using Binary Mask Based on Noise Constraints", IJRTE, Scopus indexed, ISSN: 2277-3878, Volume-8 Issue-3, September 2019.
5. Gibak Kim, Philips c.Loizou "Gain-induced speech distortions and the absence of intelligibility benefit with existing noise-reduction algorithms",Acoustical society of America,2nd july 2011, pp. 1581–1596.
6. 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).
7. 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).
8. S. Rangachari and P. C. Loizou, "A noise-estimation algorithm for highly non-stationary environments," Speech Commun., vol. 48, pp. 220–231, 2006.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. Y. Hu and P. C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," Speech Commun., vol. 49, pp. 588–601, 2007.
15. 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.
16. 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.
17. IEEE, "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio and Electroacoustics*, pp. 225–246, 1969.
18. 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.
19. Ma, J., Hu, Y., and Loizou, P. C. (2009). "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions," J. Acoust. Soc. Am. 125, 3387–3405.
20. Cees H. Taal, Richard C. Hendriks, Richard Heusdens "A SHORT-TIME OBJECTIVE INTELLIGIBILITY MEASURE FOR TIME-FREQUENCY WEIGHTED NOISY SPEECH", ICASSP 2010.
21. Cees H. Taal, Richard C. Hendriks, Richard Heusdens, and Jesper Jensen "An Algorithm for Intelligibility Prediction of Time–Frequency Weighted Noisy Speech", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 19, NO. 7, SEPTEMBER 2011.

AUTHORS PROFILE

Ramesh Nuthakk, Currently working as an Asst. Professor in Atria Institute of Technology, Bengaluru, India. He has completed M.Tech in Digital systems and communication engineering from R.E.C (NIT) Calicut University, Calicut in 1999. He has worked in telecom industry for over 12.5 years in esteemed organizations like,VSNL(TCL),Wipro & IBM. He has more than 8 years of teaching experience. Published more than 13 in National/International journals and conferences. He is currently pursuing his Ph.D in the department of electronics and communication engineering, UVCE, Bangalore University, Bengaluru. His area of interest includes speech signal processing, and Networking. He is a member of IETE.

A Sreenivasa Murthy - Currently working as a professor in the Department of Electronics and Communication Engineering, UVCE, Bangalore University, Bengaluru, India. He obtained his Bachelor's degree in Electronics from UVCE in 1982, worked as senior engineer in BEL, Bengaluru for 6 years. Then he joined as a lecturer in the Department of Electronics and Communication Engineering, UVCE in 1988. He completed his masters in Indian Institute of Science in 1994 and Ph.D. (speech signal processing) in 2013. He has published many papers in reputed journals and has a teaching experience of 31 years.

D C Naik - Received his Bachelor's degree in Telecommunication Engineering from Siddaganga Institute of Technology, Tumakuru in 2012 and received Master of technology in Computer Network Engineering from EWIT, Bengaluru, India in 2014. He is presently working as a full time research scholar in the department of electronics and communication engineering, UVCE, Bangalore University, and Bengaluru, India. His research interest area is speech enhancement techniques.