International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075, Volume-8 Issue-5S March, 2019

Posteriori Regularization based Non-Negative Matrix Factorization approach for Speech Enhancement

Ravi Kumar Kandagatla, Potluri Venkata Subbaiah

Abstract: The paper proposes, a speech enhancement method for reducing additive Gaussian noise using iterative posterior regularized Non-negative matrix factorization (NMF). Here, regularization for NMF criterion is obtained by assuming the prior distribution of the Discrete Fourier Transform (DFT) spectral magnitudes of speech follows Nakagami, Weibull distribution and DFT spectral magnitudes of coefficients follows as Rayleigh distribution. In this paper, different prior distributions, Nakagami, Weibull and Rayleigh are used and the estimates of distribution statistics are changed adaptively to provide regularization. The results for different priors are compared using different objective performance measures Perceptual Evaluation of Speech Quality (PESQ) and Signal to Distortion Ratio (SDR).

Keywords: Speech enhancement, Noise reduction, Non-negative Matrix Factorization, Weibull distribution, Iterative Posterior regularization.

I. INTRODUCTION

The aim of Speech Enhancement is to reduce noise and to reduce listener fatigue. Generally the clean speech signal is communicated under adverse noise (background, environmental) conditions. Background noises at different places like airport, train and crowd of people have different spectro-temporal characteristics. Especially in mobile communications, it is important to enhance the corrupted speech signal under noisy conditions. It is difficult task to enhance the degraded speech under Non-stationary noise conditions. Under such conditions the template based approach like NMF plays significant role. NMF approach efficiently handles the speech separation even under non-stationary environments [1], [2].

The traditional speech enhancement methods process the speech magnitudes by assuming speech spectral coefficients follow Gaussian distribution. Later researchers proposed several methods by assuming distributions like, Gamma,

Rayleigh, Laplacian It is noted in [3], that, modeling the magnitudes of the speech and noise spectral coefficients using different prior distributions is useful in speech Enhancement . In this paper, natural regularization of NMF [4], uses priors listed in [3].

An NMF applications includes vast areas like source separation, pattern recognition, classification [4], spectrogram analysis. In this paper, regularized NMF based speech enhancement is proposed, which uses Weibull and Rayleigh distributions (provide better fit for speech and noise magnitudes) for magnitudes [3], and thus the name Weibull Rayleigh NMF (WR-NMF). The proposed method uses NMF method for reducing noise because of its ability to well represent the audio, speech data. From [6], it is noted that NMF is used for noise reduction and especially for monaural noisy signal enhancement. Also, in [7] noise suppression filters are used for speech enhancement.

It is noted that the separation of noise and speech is difficult when the speech and noise having overlap spectrograms. Thus in this approach the prior probabilities are assumed and which in turn used to build a posterior model. The assumed priors are adaptively update using the model developed.

Different NMF methods involve addition of right penalty / regularization to the log-likelihood function of the assumed distribution data. Regularization is the key step to include a constraint regarding user annotation in the given latent model [5], and it is done by using suitable regularization functions. In this paper, an iterative procedure is introduced to overcome the differentiation between the sources. This is done by choosing suitable models that match with the observed signals. Here a posterior model is build for both speech and noise contributions, where penalties / priors are iteratively updated. The priori distributions of the magnitudes and speech are modeled using Weibull and Rayleigh respectively.

This paper is organized as follows: In section II, the basic notations and concept of NMF is discussed. Section III, provides discussion about the posteriori regularized NMF and the proposed algorithm. In section IV, the proposed method results are compared with existing methods and in section V the conclusion is given.

II. NON NEGATIVE MATRIX FACTORIZATION

Let x(n), s(n) and w(n) be the noisy speech signal, clean speech and noise signals respectively. Here ‘n’ indicates the sample number. Assume the noise is additive Gaussian and

Revised Manuscript Received on March 08, 2019.

Ravi Kumar Kandagatla, Research Scholar, JNTU Kakinada, Assistant Professor, LBRCE Mylavaram,India,

Potluri Venkata Subbaiah, Professor of Dept. ECE, VRSCE, Vijayawada India,

he magnitude of noisy speech spectrogram is mathematically represent as

$$\left|X(f,t)\right| = \left|S(f,t)\right| + \left|W(f,t)\right| \quad (1)$$

Where t represents the frame number, which is in processing and f represents the frequency bin, $\left|S(f,t)\right|$ represents the magnitude of speech and $\left|W(f,t)\right|$ represents the magnitude of noise. The observed signal magnitudes in matrix form can be represented as $X \in R_{+}^{f\times t}$. NMF is used to factor matrix X into the product of two non negative matrices, basis vector matri x $W = [w_1, w_2 ......, w_R] \in R+^{f\times R}$, and the weight matrix $H = [h_1, h_2 ......, h_R] \in R+^{R\times t}$ as

$$X = WH = \sum_{z=1}^{R} w_z h_z^T \quad (2)$$

Where R denotes the number of latent components. In this paper the NMF technique using KL divergence is used for weight and bases updation.

III. PROPOSED SPEECH ENHANCEMENT USING POSTERIORI REGULARIZED NMF

The step by step process of proposed speech enhancement technique is shown in flow chart Fig.1.

From the central limit theorem, it is obvious that noise DFT coefficients distribution is assumed to be Gaussian. In NMF approach we are modeling the magnitudes of noise and thus the magnitudes of Gaussian distribution is turned out to be Rayleigh. Thus in this paper, the noise PDF coefficients are modeled by Rayleigh prior distribution.

The Rayleigh PDF is given as

$$f(x;\sigma) = \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}}, \; x \geq 0 \quad (3)$$

Here the parameter σ is estimated by using Maximum likelihood parameter estimation and it would be given as

$$\sigma^2 = \frac{1}{2N}\sum_{i=1}^{N} x_i^2 \quad (4)$$

Apply negative logarithm on both sides

$$-\log\left(f(x;\sigma)\right) = -\log\left(\frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}}\right) = -\log\left(\frac{x}{\sigma^2}\right) + \frac{x^2}{2\sigma^2} \quad (5)$$

Thus, the penalty/regularization considered with the noise magnitudes prior distribution as

$$\Lambda_N = \Lambda_{S_1} = -\log\left(f(x;\sigma)\right) = \frac{x^2}{2\sigma^2} - \log\left(\frac{x}{\sigma^2}\right) \quad (6)$$

Coming to speech signal, the magnitudes are better fit by assuming powerful Weibull distribution from [3]. It can be shown that the smoothed KL divergence error is less than the other distributions. Thus magnitudes of speech signal is modeled using Weibull distribution with adjustable scale parameter as

$$f(x|\chi,\delta) = \frac{\delta}{\chi}\left(\frac{x}{\chi}\right)^{\delta-1} \exp\left(-\left(\frac{x}{\chi}\right)^{\delta}\right), \; x \geq 0 \quad (7)$$

With the estimated scale parameter as

$$\chi = \sqrt{\frac{\lambda_s}{\Gamma(1+2/\delta)}} \quad (8)$$

Where $\lambda_s(k) = E\left[\left|X_k\right|^2\right]$, (here k is frequency bin which is omitted in formula).

The regularization/penalty term for the speech sample is obtained by taking the negative logarithm on both sides as

$$\Lambda_{SP} = \Lambda_{S_2} = -\log\left(f(x|\chi,\delta)\right) = \left(\frac{x}{\chi}\right)^{\delta} - \log\left(\frac{\delta x^{\delta-1}}{\chi^{\delta}}\right), \; x \geq 0 \quad (9)$$

Different distributions and their log-likelihood functions are given in Appendix-A.

The proposed speech enhancement involves short Time Fourier Transform (STFT) operation as first stage. The input time domain signal is applied to STFT to achieve short time stationary and for frequency conversion. The magnitudes of STFT coefficients are considered for further processing and phase is left unprocessed. The unprocessed phase is considered in reconstruction of enhanced signal in final stage.

The STFT magnitudes are used to obtain the log likelihood ratios of penalties called priors and the priors are updated using Least Mean Squares (LMS) algorithm and the updated priors are used in NMF process. Here the Kulback-Leibler divergence based NMF is considered for bases update. After obtaining updated W, H matrices, the filter mask is developed. The filter mask separates the individual sources present in the noisy corrupted signal.

This process is repeated for ‘N’ number of times until you can hear reasonably good separation. While reconstruction process of each source the noisy STFT phase is used. Here no. of sources used are two. i.e., the wanted speech signal as one source which is a clean speech signal and other is the noise which corrupted the unwanted signal (additive). The aim of this technique is to separate the clean speech and noise at the output.

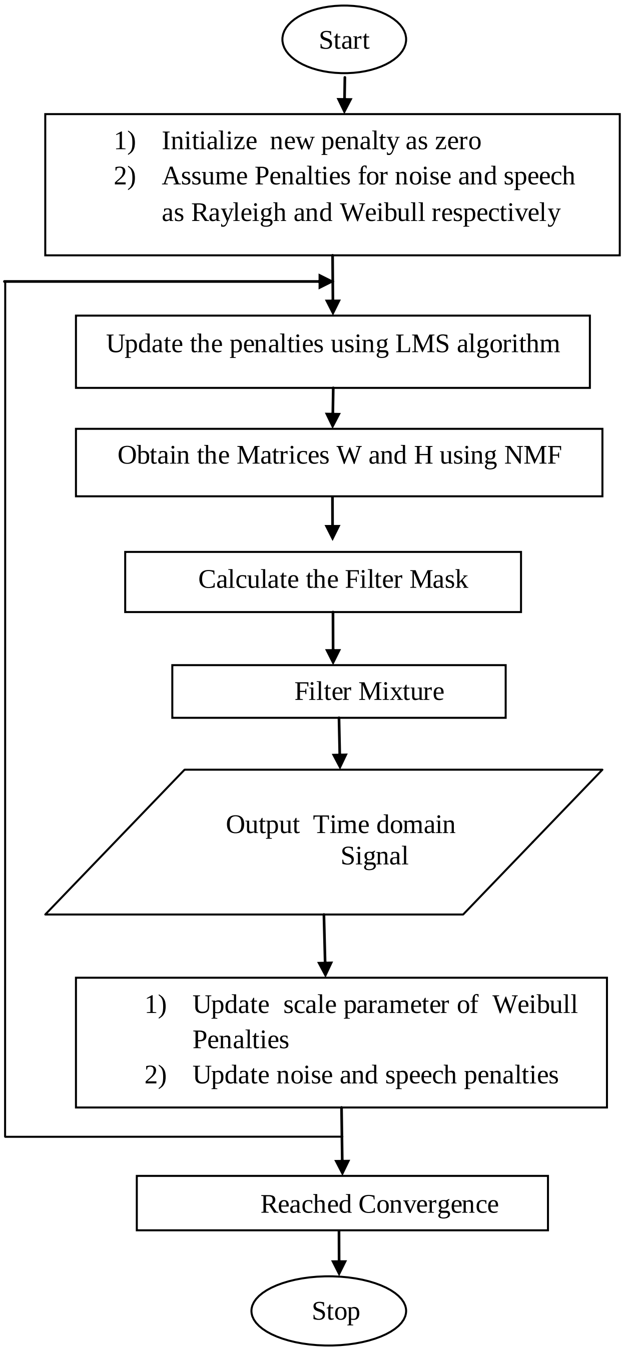

Fig. 1 Flow chart for proposed Method

Different distributions and penalties are given in Appendix.A . The proposed algorithm and weight updates using posteriori regularization is as follows

---

Proposed Algorithm WR-NMF (Weibull Rayleigh –NMF)

$X \in R_{+}^{f \times t}$ % *STFT of Noisy Data* (10)

$\Lambda_S \in R_{+}^{f \times t}, s \in \{1,....N_S\}$ $\%\Lambda_S -$ *penalties,* $N_S -$ *Number of sources* (11)

$$\Lambda_{S(New)} = 0 \quad (12)$$

$$\Lambda_{S_1} = \Lambda_{N(old)} = \frac{X}{2\sigma^2} - \log\left(\frac{X}{\sigma^2}\right) \quad (13)$$

$$\Lambda_{S_2} = \Lambda_{SP(old)} = \left(\frac{X}{\chi}\right)^{\delta} - \log\left(\frac{\delta X^{\delta-1}}{\chi^{\delta}}\right) \quad (14)$$

Repeat

For all s do

$\tilde{\Lambda}_{S(old)} \leftarrow \exp\{-\Lambda_s\}$, % *Penalties update using NMF* (15)

$$\tilde{\Lambda}_S = (1-\mu)\Lambda_{s(old)} + \mu\Lambda_{s(New)} \quad (16)$$

$$\Lambda_{s(old)} = \Lambda_{s(New)} \quad (17)$$

$$X_s \leftarrow X \circ \tilde{\Lambda}_s \quad (18)$$

End For

$$\Gamma \leftarrow \sum_s \left(W_{(s)}H_{(s)}\right)^{\circ} \tilde{\Lambda}_s \quad (19)$$

For all s do

$$z \leftarrow \frac{X_s}{\Gamma} \quad (20)$$

$$W_{(s)} \leftarrow W_{(s)} \circ \frac{Z_s H_{(s)}^T}{1H_{(s)}^T} \quad (21)$$

$$H_{(s)} \leftarrow H_{(s)} \circ \left(W_{(s)}^T Z_s\right) \quad (22)$$

End For

Reconstruction

For all s do

$$M_s \leftarrow \frac{W_{(s)}H_{(s)}}{WH}, \quad \% \textit{ FilterMask} \quad (23)$$

$$\hat{X}_s \leftarrow M_s \circ X, \quad \% \textit{ Filter Mixture} \quad (24)$$

$$x_s \leftarrow ISTFT\left(\hat{X}_s, \angle X, P\right) \quad \% P-STFT \textit{ parameters} \quad (25)$$

Update χ

$$\chi = \sqrt{\frac{\lambda_s}{\Gamma(1+2/\delta)}} \quad (26)$$

Where $\lambda_s(k) = E\left[\left|X_k\right|^2\right]$

End update

$$\Lambda_{S_1} = \Lambda_{N(old)} = \frac{\hat{X}_{S_1}^2}{2\sigma^2} - \log\left(\frac{\hat{X}_{S_1}}{\sigma^2}\right) \quad (27)$$

$$\Lambda_{S_2} = \Lambda_{SP(old)} = \left(\frac{\hat{X}_{S_2}}{\chi}\right)^{\delta} - \log\left(\frac{\delta\left(\hat{X}_{S_2}\right)^{\delta-1}}{\chi^{\delta}}\right) \quad (28)$$

$$\Lambda_{s(New)} = \exp\left(-\Lambda_{s(old)}\right) \quad (29)$$

$\Lambda_{s(old)}$ *represents both* Λ_{SP} *and* Λ_N

End For

Untill *Convergence*

Return: *Time domain signals* x_s

In the proposed method, the penalties are obtained by using negative logarithm for the respective distributions as

Weibull for speech magnitudes and Rayleigh for noise magnitudes. The penalties are updated using Least Mean Square (LMS) method. The noisy data is multiplied by penalties to perform posterior regularization using (19), (20). From (21), (22), the matrices W, H are updated. From (23), the filter matrix is obtained and using (24), the observed data is filtered and ISTFT is performed on resultant data. Finally enhanced speech is obtained using phase and output.

In the above algorithm the notations used are as follows: a subscript as (s) represents the columns or rows of source and symbol $\circ$ indicates element wise multiplication.

The algorithm discussed with Weibull as speech PDF and Rayleigh as Noise PDF. The algorithm is repeated for Nakagami as Speech PDF and Rayleigh as Noise PDF.

IV. EXPERIMENTAL RESULTS

The experiments were conducted on speech samples taken from NOIZEUS speech corpus data base [8]. Each signal is sampled at 8 KHz frequency. Short-time Fourier Transform (STFT) is performed by using hanning window with a frame size 1024 (Number of samples) with 75 % overlap. Experiments are conducted on different samples with SNR 0 dB, 5 dB, 10 dB, 15 dB. The performance of the proposed method is analyzed with noisy signals corrupted in Babble and Street environment. The parameter values are assumed as $\mu = 0.001$, and $\delta = 0.680$, in [3]. The penalties of speech is considered to be Weibull distribution, where as the penalties of noise is considered to be Rayleigh distribution. The assumed speech and noise priors are updated iteratively until convergence (based on the sound listened). The proposed method is compared with different NMF Variants like Itakura-Saito IS-NMF[1], Euclidean NMF[2], Posteriori regularized NMF, PR-NMF [5]. The objective performance measures considered in this work are Perceptual Evaluation of Speech Quality (PESQ) [9], and Signal to Distortion Ratio (SDR) [10]. Higher values of PESQ, SDR indicates better performance. The results show that the proposed method performs well than the benchmark algorithms.

Speech samples corrupted by Babble noise and Street noise at different input SNRs of 0 dB, 5 dB, 10 dB, 15 dB are taken from NOIZEUS database and -5 dB is obtained by adding noisy source to clean speech of NOIZEUS database. Experimental results shows that better PESQ values are obtained by proposed method. At low SNR of -5 dB it is showing better performance. Fig. 2 and Fig. 3 shows comparison of PESQ values for different NMF algorithms with proposed method corrupted by Babble noise and Street noise respectively. Fig. 4 and Fig. 5 shows comparison of SDR values for different NMF algorithms ((Posteriori Regularized NMF (PR-NMF), Nakagami Rayleigh NMF (NR-NMF), Weibull Rayleigh (WR-NMF) with proposed method corrupted by Babble noise and Street noise. Different speech enhancement methods using statistical modeling and NMF approaches are discussed in [11], [12].

Experimental results shows that WR-NMF provides better PESQ values and NR-NMF provides better SDR values and the two proposed methods shows improved results than traditional algorithms. The Nakagami distribution and its log-likelihood function is given in Appendix. I

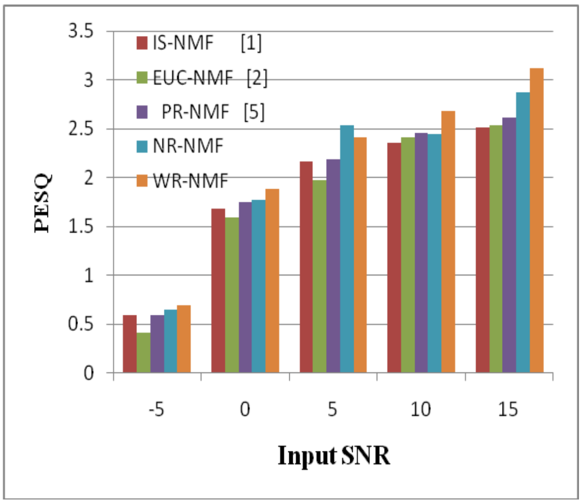

Fig. 2 PESQ values for proposed method compared with traditional methods at different input SNRs of -5 dB, 0 dB, 5 dB, 10 dB, 15 dB for Babble noise.

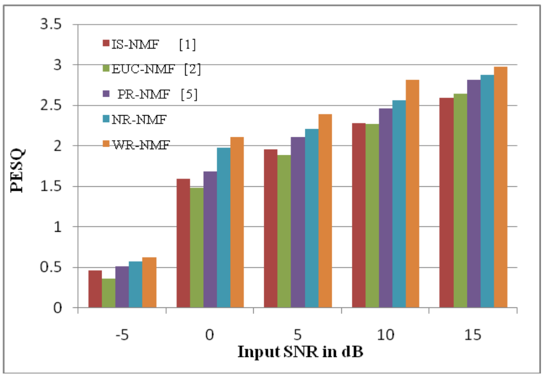

Fig. 3 PESQ values for proposed method compared with traditional methods at different input SNRs of -5 dB, 0 dB, 5 dB, 10 dB, 15 dB for Street noise

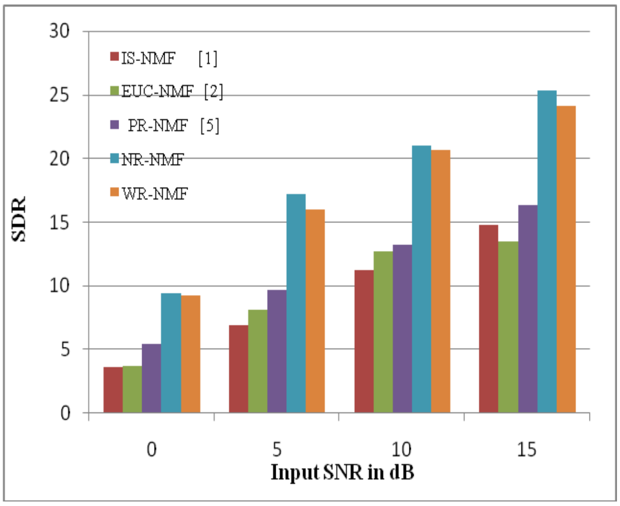

Fig. 4 PESQ values for proposed method compared with traditional methods at different input SNRs of -5 dB, 0 dB, 5 dB, 10 dB, 15 dB for Babble noise

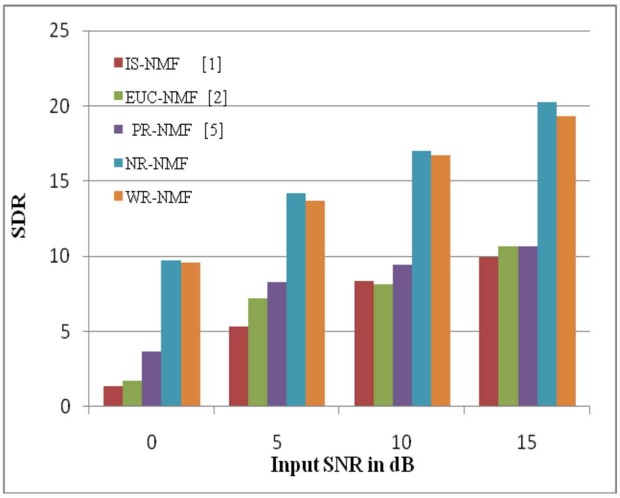

Fig. 5 PESQ values for proposed method compared with traditional methods at different input SNRs of -5 dB, 0 dB, 5 dB, 10 dB, 15 dB for Street noise

V. CONCLUSION

Speech enhancement improves quality and intelligibility. For robust noise reduction and analysis of non-Stationary signal analysis NMF technique is used. From the experiment results it shows that NR-NMF provides better SDR values and WR-NMF provides better PESQ values in case of different noise environments. Also it is noted that individually provide better performance in terms of enhanced signal quality and intelligibility than to traditional NMF approaches. Experimental results shows that the combination of NMF techniques with statistical techniques improves the quality of enhanced speech.

REFERENCES

1. C. Fevotte, N. Bertin, J-L. Durrieu, “ Nonnegative matrix factorization : with the itakura-saito divergence with application to music analysis,” Neural computation., Vol. 21, no. 3, pp. 793 – 830, Mar. 2009.
2. D. D. Lee, H. S. Seung, “ Algorithms for non-negative matrix factorization, ” in Proc..Advances in neural information processing systems (NIPS)., pp. 556 – 562, 2001.
3. Ivan Tashev, Alex Acero, “ Statistical Modeling of the speech signal, ” in International Workshop on Acoustic, Echo, and Noise Control (IWAENC)., 2010
4. M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, R. J. Plemmons, “Algorithms and applications for approximate nonnegative matrix factorization,” at Computational statistics & data analysis., Vol. 52, no. 1, pp. 155 - 173. 2006
5. N. J. Bryan, G. J. Mysore, “ An efficient posterior regularized latent variable model for interactive sound source separation, ” in International Conference of Machine Learning (ICML)., pp.208 – 216, June 2013
6. N. Mohammadiha, P. Smaragdis, A. Leijon, “ Supervised and unsupervised speech enhancement using nonnegative matrix factorization, ” IEEE Transactions on Audio, Speech and Language Processing., Vol. 21, no. 1, pp. 2140- 2151. Oct. 2013
7. K. Y. Chan, S. Nordholm, K. F. C. Yiu, R. Togneri, “ Speech enhancement strategy for speech recognition microcontroller under noisy environments,” in Neuro computing., Vol 118, pp. 279-288, March. 2013
8. NOIZEUS Database.
9. Y. Hu, P. C. Loizou, “ Evaluation of objective quality measures for speech enhancement, ” in IEEE Transactions on Audio, Speech, and language processing., Vol.16 , no. 1, pp.229-238, Jan. 2008
10. E. Vincent, R. Gribonval and C. Fevotte, " Performance measurement in blind audio source separation," in *IEEE Transactions on Audio, Speech, and Language Processing.*, vol. 14, no. 4, pp. 1462-1469, July 2006.
11. K.Ravi Kumar, P.V.Subbaiah, “Speech Enhancement using MMSE Estimation under phase Uncertainty”, in International Journal of Speech Technology, Volume 20, pp 373–385, June 2017
12. Ravi Kumar Kandagatla, P.V. Subbaiah, “Speech enhancement using MMSE estimation of amplitude and complex speech spectral coefficients under phase-uncertainty” , Speech Communication, Volume 96, pp 10-27, February 2018

APPENDIX-A

PDF $f(x;\sigma)$	**Regularization or penalty** $-\log(f(x;\sigma))$	**Parameters**
Nakagami $\frac{2}{\Gamma(\mu)}\left(\frac{\mu}{\sigma^2}\right)^{\mu} x^{2\mu-1}\exp\left(-\frac{\mu}{\sigma^2}x^2\right)$	$-\log\left(\frac{2}{\Gamma(\mu)}\right)-\mu\log\left(\frac{\mu}{\sigma^2}\right)-(2\mu-1)\log x+\left(\frac{x^2\mu}{\sigma^2}\right)$ $\mu<1$ allows Super- Gaussian	$\sigma^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2$
Wiebull $\frac{\delta}{\chi}\left(\frac{x}{\chi}\right)^{\delta-1}\exp\left(-\left(\frac{x}{\chi}\right)^{\delta}\right)$	$\left(\frac{x}{\chi}\right)^{\delta}-\log\left(\frac{\delta x^{\delta-1}}{\chi^{\delta}}\right)$ $\delta=0.68$	$\chi=\sqrt{\frac{\lambda_s}{\Gamma(1+2/\delta)}}$ $\lambda_s(k)=E\left[\lvert X_k\rvert^2\right]$
Rayleigh $\frac{x}{\sigma^2}e^{\frac{-x^2}{2\sigma^2}}$	$-\log\left(\frac{x}{\sigma^2}\right)+\frac{x^2}{2\sigma^2}$	$\sigma^2=\frac{1}{2N}\sum_{i=1}^{N}x_i^2$