Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 7, July 2015, pg.358 – 363

RESEARCH ARTICLE

SPEECH ENHANCEMENT USING HIDDEN MARVOK MODEL (HMM)

BarinderPal Singh
M.Tech Student
Information Technology
Chandigarh Engineering College
Landran, Mohali

Dr. Shashi Bhushan
Professor & Head of Department
Information Technology
Chandigarh Engineering College
Landran, Mohali

Mr. Karan Mahajan
Assistant Professor
Information Technology
Chandigarh Engineering College
Landran, Mohali

***ABSTRACT:** Speech enhancement used for better communication system and removal of all degradations or unwanted background noises .The background noises can be babble noise, training noise, traffic noise. This paper proposes Hidden Markov Model (HMM) for speech enhancement. The proposed model is based upon the combination of the hidden markov model (HMM) with the non-negative matrix factorization (NMF). The supervised sparse non-negative matrix factorization (S-SNMF) has been used for the purpose of feature description from the given speech signal. The proposed model results have been obtained in the form of signal noise, signal to noise ratio and the signal log-likelihood ratio (before and after the speech signal enhancement procedure). The proposed model has been proved itself better in the terms of given resulting parameters.*

***KEYWORDS:** HMM, non-negative matrix factorization, NMF, speech signal enhancement, quality improvement.*

1. INTRODUCTION

The speech signal deterioration is caused by various factors like disorders in production organs, different sensors (microphones),acoustic non-speech and speech background, channel and reverberation effect and disorders in perception organs. Recently researches has found methods to enhance speech, mostly related to speech distorted by background noise occurring at the source or in transmission, clicks, and other non-stationary interferences . Most cases assume noise whose pertinent features change slowly (i.e., locally stationary over analysis frames of interest), so that it can be characterized in terms of mean and variance (i.e., second-order statistics), either during non-speech intervals (pauses) of the input signals or via a second microphone (called reference microphone) receiving little speech input. In ideal environment there should be no degradation in quality or intelligibility of original speech but in practical scenario there is degradation in quality and/or intelligibility and/or human subjects have impaired speech production and perception systems. So the main aim of speech enhancement is to enhance quality and intelligibility. In some cases when inputs are from multiple sources are available then it become very difficult for speech enhancement systems to improve intelligibility. Thus most speech enhancement methods raise quality, while minimizing any loss in intelligibility. Some certain aspects of speech are more important than others. The auditory system is more sensitive. Thus speech enhancement algorithms often focus on accurate modelling of peaks in the speech amplitude spectrum, rather than on phase relationships or on energy at weaker frequencies. Voiced speech, with its high amplitude and concentration of energy at low frequency, is more perceptually important than unvoiced speech for preserving quality. Hence, speech enhancement focuses improving the periodic portions of speech. Good representation of spectral amplitudes at harmonic frequencies and especially in the first three formant regions is paramount for high speech quality. All enhancement algorithms introduce their own distortion and care to be taken to minimize distortion

1.2 Speech Enhancement Techniques

The speech enhancement methods varies depending upon type of degradation. The speech enhancement techniques can be divided into two basic categories: (i) Single channel and (ii) Multiple channels based on speech acquired from single microphone or multiple microphone sources respectively. However, single channel (one microphone) signal is available for measurement or pick up in real environments and hence focus is here on single channel speech enhancement methods. The principal degradations that we are concerned with are:

(a) Additive acoustic noise – such as the noise added to the speech signal when recorded in an environment with noticeable background noise, like in an aircraft cockpit.

(b) Non-linear distortion such as arises from clipping – such as when inappropriate gain is applied at the signal input stage.

(c) Additive broadband electronic noise

(d) Electrical interference

(e) Codec distortion – distortion caused by the coding algorithm due to compression.

2. LITERATURE REVIEW

Cohen et. al. [1] proposed an improved MCRA noise variance estimator improvements and MCRA approach showed a higher performance compared to weighted averaged method. The ability of the algorithms was tested by comparing the spectrograms of enhanced speech for a signal recorded in a car by suddenly turning on the defroster in full. **Berdugo, B. et. al.** [2] proposed a new approach called minima controlled recursive averaging (MCRA) for noise estimation. The noise estimate was done by calculating the average of past spectral values of noisy speech which was controlled by a time and frequency dependent smoothing factors. These smoothing factors were calculated based on the signal presence probability in each frequency bin separately. This probability was in turn calculated using the ratio of the noisy speech power spectrum to its local minimum calculated over a fixed window time**. R. Martin et. al.** [3] described that Gaussian statistical model provides a good approximation for the noise DFT coefficients. For speech signals, however, where typical DFT frame sizes used in mobile communications are short (10ms -40ms) this assumption is not well fulfilled. It is valid only if the DFT frame size is much longer than the span of correlation of the signal under consideration. **Cohen et. al**.[4] presented methods that incorporated the fact that speech might not be present at all frequencies and at all times. It estimate of the probability that speech is absent at a particular frequency bin. In this research, MMSE magnitude estimator under the assumed Laplacian model and uncertainty of speech presence has been described & considered a two-state model for speech events. According to this two state model, either speech is present at a particular frequency bin (hypothesis H1) or not (hypothesis H0**). R. Martin et. al.** [5] proposed a new estimator, in which the real and imaginary parts of the clean signal were estimated in the MMSE sense conditional on the real and imaginary parts of the observed noisy signal. This estimator, however, is not the optimal spectral amplitude estimator but clean signal & noise were modeled by a combination of Gaussian, Gamma and Laplacian distributions**. C. Breithaupt et. al**. [6] described that the real and imaginary part of the speech coefficients are better modeled with Laplacian and Gamma densities. This observation led researchers to derive a similar optimal MMSE STSA estimator but based on more accurate models, Laplacian and/or Gamma. However, the derivation of such an estimator is complicated leading some people to seek alternative techniques to compute the MMSE STSA estimator. **Malah et. al.** [7] derived the MMSE STSA estimator, based on modeling speech and noise spectral components as statistically independent Gaussian random variables. Authors analyzed the performance of the proposed STSA estimator and compared it with a STSA estimator derived from the Wiener estimator. Authors also examined the MMSE STSA estimator under uncertainty of signal presence in the noisy observations. **Y. Ephraim et. al.** [8] derived a short-time spectral amplitude (STSA) estimator for speech signals which minimizes the mean square error of the log-spectra (i.e., the original STSA and its estimator) and examined it in enhancing noisy speech. This estimator is also compared with the corresponding minimum mean-square error STSA.

3. PROPOSED METHODOLOGY

3.1 Hidden Markov Model

The hidden markov models are the most successful methods among the stochastic methods produced for the speech enhancement. The speech is divided into the phenomes (the sub-features extracted from the speech signal), and then enhanced using the given HMM algorithm. The HMM is equipped of the data training module before its deployment on any of the speech signal enhancement applications.

Algorithm 1: HMM Model

1. Initialize the process of speech signal with signal acquisition.
2. The following recursive application of HMM is used by using following algorithm design:
   a. For time t=1,2,3….N
   b. For states s=1,2,3…..M
   c. Compute the following
   d. $\delta_t(i) = max[\delta_{t-1}(j).\ aji\ ]\ .\ f_i(\mathbf{x}(t));$ *max(.) performed over all j;*
   e. $\psi_t(i) = arg\ \{max[\delta_{t-1}(j)\ .\ a_{ji}]\};$ *max(.) performed over all j;*
   f. End step b (States)
   g. End step a (Time)
3. Retrieve the most likely states in the final form.
4. Retrieve the sequence of the shortlisted states.
5. Return the enhanced speech signal after restructuring the states back in the original waveform for the given signal.

SUPERVISED SPARSE NON-NEGATIVE MATRIX FACTORIZATION

The non-negative matrix factorization is used to extract the features from the given non-negative form data. The non-negative matrix factorization (NMF) working like the following in the proposed model simulation:

Algorithm 2: Supervised sparse non-negative matrix factorization (S-SNMF)

1. Initialize the W and H factors. Methods for choosing, or seeding, the initial matrices W and H for various algorithms.
2. Uniqueness. Sufficient conditions for uniqueness of solutions to the NMF problem can be considered in terms of simplicial cones.
3. Updating the factors. Devising efficient and effective updating methods when columns are added to the data matrix A.

4. IMPLEMENTATION AND RESULTS

In this section we will compare the results of the input speech signal and after applying HMM algorithm. The original Speech of the signal and the HMM approaches are shown. The results have been obtained in the form of various performance parameters. At very first the original signal has been obtained both before and after enhancement. The signal enhancement is the technique which is used to improve the quality of the speech signal. The speech signal enhancement has been performed by using the amalgamation of the hidden markov model with the non-negative matrix factorization. The non-negative matrix factorization has returned the feature extracted from the speech signal which is further used as the core sample to improve the quality of the speech signal using the HMM model.

The speech signal has been obtained before and after the enhancement. There are the minute visible changes in the signal after the speech signal enhancement and so in the case of signal noise. The signal noise has been also calculated before and after the speech signal enhancement.

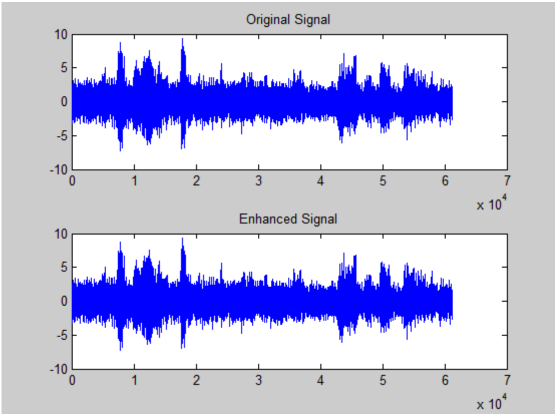

Figure 1: The speech signal before and after enhancement.

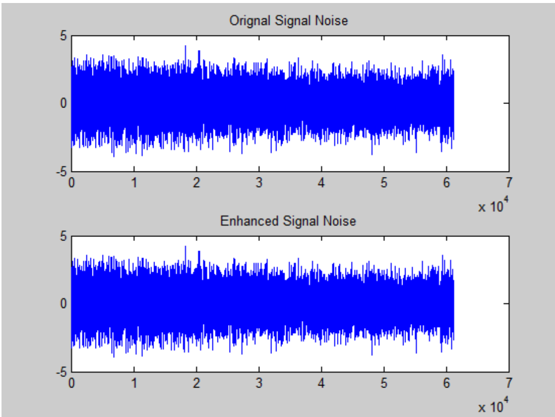

Figure 2: The signal noise before and after the quality enhancement

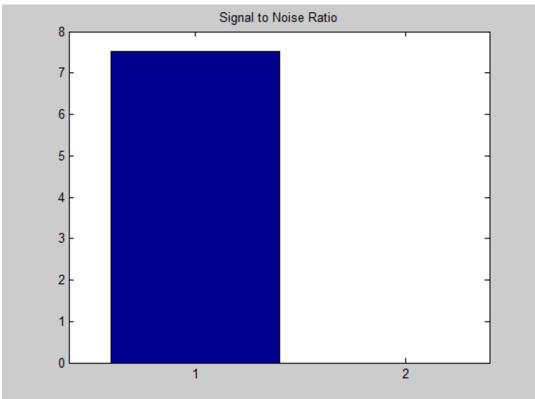

Figure 3: The plot of signal to noise ratio

The signal to noise ratio (SNR) is the quality parameter, which shows the quality difference between the original and enhanced signal. The proposed model has been performed better while evaluated on the basis of signal to noise ratio. The lower bound likelihood has been also measured before and after the signal enhancement, where we have found the enhanced signal of the better quality. The proposed model have been proved itself efficient in the case of the lower bound signal likelihood.

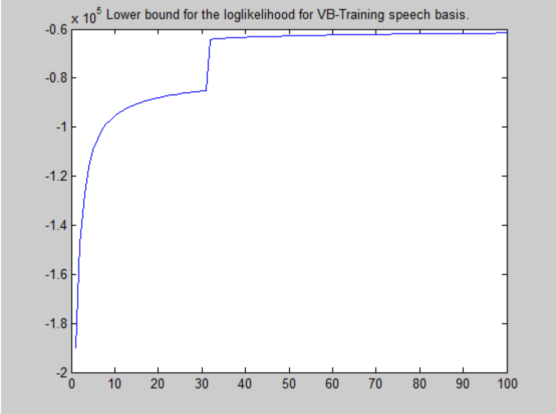

Figure 4: The lower bound likelihood after enhancement.

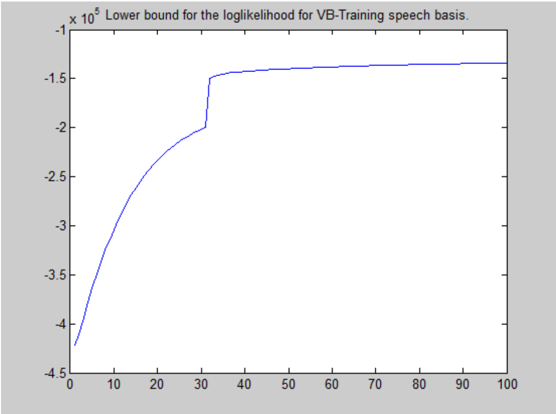

Figure 5: The signal likelihood before the signal enhancement.

5. CONCLUSION

The proposed model has been designed for the speech signal enhancement using the combination of hidden markov model with the non-negative matrix factorization (HMM-NMF). The supervised non-negative matrix factorization model has been used for the sparse matrix formation. The proposed model has been improved by preparing the combination of the S-SNMF (Supervised-Sparse matrix non-negative matrix factorization technique). Both of the techniques combined under this research model, where the S-SNMF has been used for the feature description (or sampling) from the given noise signal, which is further used by the hidden markov model (HMM) for the signal quality enhancement. The experimental results have been obtained signal to noise ratio and lower bound likelihood for before and after the signal enhancement. The experimental results have been shown the effectiveness of the proposed model. The proposed model has shown the significant improvement on the basis of the signal to noise ratio and lower bound signal likelihood.

In the future, the proposed model can be enhanced using the combination of the other effective techniques along with the proposed model such as linear component analysis or log-spectral amplitude for the improvement in the signal enhancement model.

REFERENCES

[1] Cohen, I., ―Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging‖, IEEE Trans. on speech and audio processing, vol. 11, no. 5, pp. 466-475, Sept. 2003..

[2] Berdugo, B. and Cohen, I., ―Noise estimation by minima controlled recursive averaging for robust speech enhancement‖, IEEE Signal Proc. Letters, vol. 9, no. 1, pp. 12-15, Jan. 2002.

[3] R. Martin, ―Speech enhancement using a minimum mean-square error short-time spectral amplitude estimation,‖ in Proc. IEEE Int. Conf. Acoust Speech, Signal Processing, pp. 504—512, 2002.

[4] I. Cohen, ―Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator,‖ IEEE Signal Processing Lett., vol. 9, pp. 113-116, Apr. 2002.

[5] R. Martin and C. Breithaupt, ―Speech enhancement in the DFT domain using Laplacian speech priors,‖ in International Workshop on Acoustic Echo and Noise Control (IWAENC), pp. 87–90, Sept. 2003

[6] C. Breithaupt and R. Martin, ―MMSE estimation of magnitude-squared DFT coefficients with super-gaussian priors,‖ in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, pp. 848-851, 2003.

[7] Malah, D., Cox, R.V. and Accardi, A.J., ―Tracking speech-presence uncertainty to improve speech enhancement in non-stationary noise environments,‖ Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2, pp.789-792, 15-19 Mar 1999

[8] Y. Ephraim,‖ Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator‖, IEEE Transactions On Acoustics, Speech and Signal Processing 0096-35 18/8S/0400-0443, 1985

