Speech Enhancement using Segmental Non-Negative Matrix Factorization (SNMF) and Hidden Marvok Model (HMM)

Barinder Saini
M.Tech Student, Information Technology
Chandigarh Engineering College
Landran, Mohali

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

ABSTRACT

Speech Enhancement refered as to improve quality or intelligibility of speech signal. Speech signal is often degraded by additive background noise like babble noise, train noise, restaurant noise etc. Speech enhancement aims at improving the performance of speech communication systems in noisy environments. This paper proposes a segmental NMF (SNMF) speech enhancement scheme to improve the conventional frame-wise NMF-based method. In this two algorithms are derived to decompose the original nonnegative matrix associated with the magnitude spectrogram, the first algorithm is used in the spectral domain and the second algorithm is used in the temporal domain .In this paper Hidden macro model and SNMF(S) for subjective learning (SNMF-S). Then the SNMF for the objective learning (SNMF-O) will be implemented.

Keywords:- Speech Enhancement, Non negative Matrix Factorization (NMF) , Segmental Nonnegative Matrix Factorization (SNMF)

1. INTRODUCTION

Speech enhancement aims at improving the performance of speech communication systems in noisy environments. Speech enhancement may be applied, for example, to a mobile radio communication system, a speech recognition system, a set of low quality recordings, or to improve the performance of aids for the hearing impaired. The interference source may be a wide-band noise in the form of a white or colored noise, a periodic signal such as in hum noise, room reverberations, or it can take the form of fading noise. The first two examples represent additive noise sources, while the other two examples represent convolution and multiplicative noise sources, respectively. The speech signal may be simultaneously attacked by more than one noise source. Generally, speech enhancement methods can be categorized into two broad classes: unsupervised and supervised. Unsupervised methods include a wide range of approaches such as spectral subtraction (SS*)*. Speech is the fundamental, most effective, reliable and common medium to communicate in real time systems. In market due to advancement in technology many speech communication applications based devices is available, they are cheaper and easily available. However, undesired noises in environment cause undesired effects in real time speech processing systems. Human communications and intelligent machines are suffers from the degraded performance in which they takes decision based on what it receives as a speech. Earlier many researchers investigated and developed various approaches for noise reduction and speech enhancements. The speech enhancement is useful for storage and transmission of speech data, also it improves speech recognition based system performance where accurate identification of words and sentences can provide automation in most of the human-machine or machine-machine based interface. Speech enhancement can boost up the performance of speech recognition systems by keeping low word error rate (WER).

1.2 Enhancement Methods

Enhancement methods can be divided into two categories: Supervised and unsupervised. [1]

In these unsupervised methods, statistical models are used for both speech and noise, and the model parameters for clean speech are estimated without any prior information about the noise type or speaker identity. However, most difficult part of most unsupervised speech enhancement methods is estimating the power spectral density (PSD) of noise [2,3], particularly in case when the interfering noise is non-stationary. By contrast, supervised speech enhancement methods use distinct models for clean speech and noise signals, and respective samples are used for calculating parameters. An interaction model is then defined by combining speech and noise models, and finally the noise reduction task is conducted. Some examples of the supervised methods include codebook-based approaches [4] and hidden Markov model (HMM) based methods [5]. In these methods there is no need to calculate the power spectral density (PSD) of noise using a separate algorithm. The supervised approaches produce higher quality enhanced speech signals than those produced using the unsupervised methods. This is because more prior information is incorporated into the algorithm when using supervised methods than it is when using unsupervised methods .The required prior information regarding noise type (and speaker identity, in some cases) can be provided by the user or a separate acoustic environment classification algorithm [6], or obtained using a built-in classification scheme [4]. A successful supervised speech enhancement algorithm is based on the nonnegative matrix factorization (NMF) technique. With NMF, the basis spectra for clean speech and noise are first estimated using the corresponding training samples. Hidden Markov model (HMM)-based speech enhancement method, aiming at reducing non-stationary noise from speech signals. Hidden Markov model (HMM) based methods uses a prior information about both the speech and the noise/

1.3. Segmental Nonnegative Matrix Factorization Speech Enhancement Scheme (SNMF)

Speech signals have a unique spectral-temporal structure. In the spectral domain, speech signals shows distinct characteristics in the high and low frequency bands. In particular, major speech components are primarily located at lower frequencies. In addition, speech is a time-varying signal, and its temporal information plays a crucial role in identifying the corresponding characteristics. This section develop two speech enhancement approaches based on the concept of NMF, which decompose the non-negative matrix associated with the magnitude spectrogram of speech, namely sub-band and patch processing approaches. In Spectral domain is called sub-band processing .In this it segment the full-band speech and noise magnitude spectrograms into sub-bands, and this approach is called spectral-domain SNMF

(denoted)by SNMF (S) as a short-hand notation hereafter) . Temporal domain is also called patch processing. In the patch processing, each spectrum frame was augmented into a spectrum patch with neighbouring spectrum frames, and then the standard NMF process was applied to the spectrum patches to perform speech enhancement.

2. LITERTURE SURVEY

Berdugo, B. et. al. [7] proposed a new approach called minima controlled recursive averaging (MCRA) for noise estimation. The noise estimate was calculated by averaging the past spectral values of noisy speech. The smoothing factors were calculated based on the signal presence probability in each frequency bin separately. The ratio of the noisy speech power spectrum to its local minimum calculate probability over a fixed window time .**Cohen et. al. [8]** presented methods that incorporated the fact that speech might not be present at all frequencies and at all times. Authors provided an 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 .It is two state model, either speech is present at a particular frequency bin (hypothesis H1) or not (hypothesis H(0). **Malah et. al. [9]** derived the MMSE STSA estimator, based on modelling speech and noise spectral components as statistically independent Gaussian random variables. **Ephraim et. al. [10]** 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 estimator derived previously.

3. PROPOSED METHODLOGY

The proposed model will be implemented using the MATLAB simulator. The implementation will be following the implementation of basic speech enhancement model using the Hidden Markov model (HMM) method along with segmental non-negative matrix factorization (SNMF) for subjective learning (SNMF-S). Then the SNMF for the objective learning (SNMF-O) will be implemented over the pre-developed model of SNMF-S and HMM. The proposed model will be tested using the normalized absolute error (NAE), mean squared error (MSE), peak signal to noise ratio (PSNR) and root mean square error (RMSE). The detailed result analysis will lead towards the conclusion of the proposed.

4. CONCLUSION

Speech Enhancement refered as to improve quality or intelligibility of speech signal. In this paper speech enhancement is done using SNMF. In this two algorithms are used , the first algorithm is used in the spectral domain and the second algorithm is used in the temporal domain. Further for better speech enhancement we will use HMM and SNMF(S) for subjective analysis. Then SNMF for the objective learning (SNMF-O) will be implemented.

5. REFERENCES

[1] *Hao-Teng Fan1, Jeih-weih Hung 1, Xugang Lu2, Syu-Siang Wang3 , and Yu Tsao3,"* Speech enhancement using segmental non negative matrix factorization, 2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP).

[2] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on Speech and Audio Processing*, 9(5), pp. 504–512, 2001.

[3] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Transactions on Speech and Audio Processing*, 11(5) ,pp. 466–475, 2003.

[4] S.Srinivasan, J.Samuelsson, and W. Kleijn, "Codebook driven short-term predictor parameter estimation for speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1), pp. 163–176, 2006.

[5] D. Y. Zhao and W. B. Kleijn, "HMM-based gain modelling for enhancement of speech in noise," *IEEE Transactions on Audio, Speech, and Language Processing*, 15(3), pp. 882–892, 2007.

[6] K. El-Maleh, A. Samouelian, and P. Kabal, "Frame level noise classification in mobile environments," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 237–240, 1999.

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

[8] 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.

[9] 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.

[10] 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.