International Journal of Information & Computation Technology.
ISSN 0974-2239 Volume 4, Number 17 (2014), pp. 1791-1797

http://www. irphouse.com

Comparative Evaluation Single Channel Speech Enhancement Algorithms with and without Phase Spectrum Compensation

Princy Sharma[1], Sachin Singh[2], M.Tripathy[3], R.S. Anand[4]

[1, 2, 3, 4] *Department of Electrical Engineering, Indian Institute of Technology Roorkee Roorkee – 247 667, Uttarakhand (India)*

Abstract

Removal of interference noise from noisy speech signal has been an area of interest for many years. Most of the work done in this area mainly concentrates on enhancement of spectral amplitude. This paper presents a method for removal of interference noise by modifying the phase spectrum in conjunction with enhancement of spectral amplitude. The method of speech enhancement using Non-negative Matrix Factorization (NMF) with phase spectrum compensation is presented. Also interference noise cancellation using Adaptive noise cancellation (ANC) followed by phase spectrum compensation is presented. These methods are compared with the standard NMF and ANC methods for noise removal. The performances of the methods are compared using various objective measures like SNR, PESQ, Cepstral Distance, etc. The analysis shows the improvement in performance by using phase spectrum compensation along with standard methods of speech enhancement when compared with methods without phase spectrum compensation.

Index Terms- Speech enhancement, ANC, Non-Negative Matrix Factorization (NMF), Phase spectrum compensation (PSC), Signal to Noise Ratio, Perceptual Evaluation of Speech Quality (PESQ), Cepstral Distance.

Introduction

Speech enhancement constitutes of removal of interfering background noise from a noisy speech. The background noises are generally additive in nature. The use of speech processing system is very common in hands free telecommunication, speech recognition system, hearing aid, etc.. These systems get affected by background interfering noise and hence, noise removal is required for improving the effectiveness of these speech processing systems by improving the intelligibility and quality of speech [1]. In literature many methods have been proposed for background noise removal. These include spectral subtraction [2], minimum mean square error (MMSE) spectral amplitude estimation [3], wiener filtering [4], etc. In this paper we propose the method of speech enhancement using NMF with Phase spectrum compensation. Another method of noise reduction using ANC followed by phase spectrum compensation is presented.

The remaining paper is organized as follows: In section II the review of speech de-noising using ANC is presented. The adaptive algorithms used in ANC for noise cancelling is also briefly discussed here. NMF method for speech enhancement is discussed briefly in section III. In section IV method of speech de-noising using NMF and ANC with phase compensation is presented. Section V presents the results obtained using the methods described. Finally section VI concludes the study.

Speech Enhancement using ANC

The basic configuration of a noise cancelling system is shown in Fig.1 [5] [6]. It requires use of two signal inputs. The first input also known as the primary input is the degraded speech signal, $s_0 + n_0$. The secondary input, n_1, is basically a noise which is in some way correlated to the noise present in the primary input. This is passed through the adaptive filter to generate an estimate of the noise in primary input. This estimated noise is then subtracted from the primary input, which results in an error signal, e, which is an estimate of the clean speech signal. The error signal, e, is used in some form as the objective function which is minimized to adjust the coefficients of the adaptive filter.

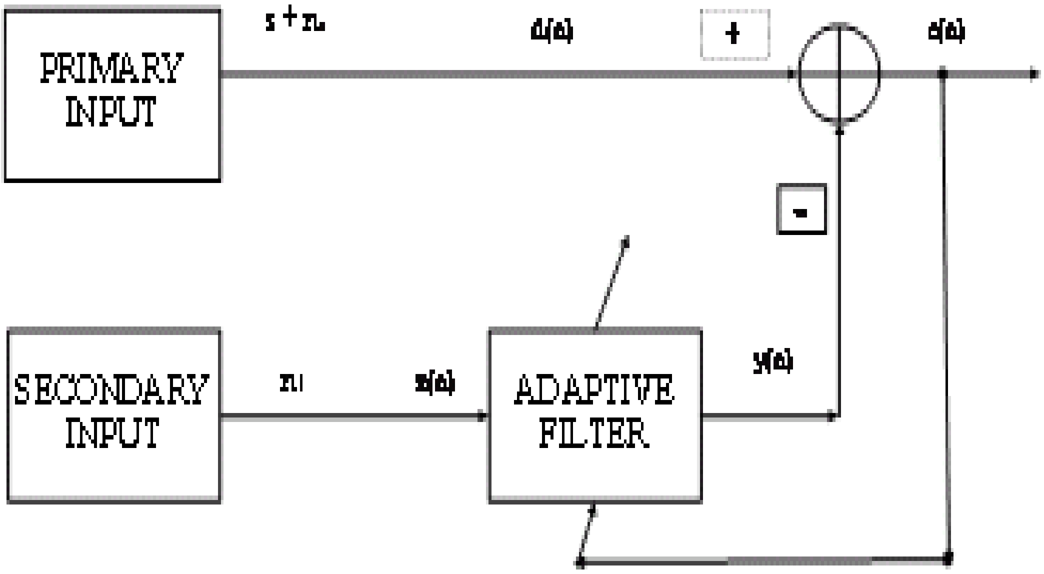

Fig. 1: Adaptive Noise Cancellation configuration

The adaptive algorithms used in this work are Least Mean Square (LMS), Normalized Least Mean Square (NLMS), and Recursive Least Square (RLS) algorithms [7] [8].

Speech Enhancement using Non-Negative Matrix Factorization (NMF)

NMF method decomposes a non-negative matrix, $V \in R^{+}_{F\times T}$, into two matrices W and H such that the elements of the two matrices are non-negative [9].

$$V \approx W.H \tag{11}$$

where $W = R^{+}_{F\times K}$ and $H = R^{+}_{K\times T}$ and normally $K \leq \min(F, T)$. The columns of V correspond to the data vector. The columns in W capture the prominent pattern of data and are known as basis vectors. The rows of H represent the gain of the corresponding basis vector. The factorization is done by iterative minimization of the KL divergence cost function [10].

The W and H matrix are updated using the multiplicative update equation [11] as follows:

$$W \leftarrow W \otimes \frac{(\frac{V}{W.H}).H^T}{H^T.1} \tag{12}$$

$$H \leftarrow H \otimes \frac{\left(\frac{V}{WH}\right).W^T}{W^T.1} \tag{13}$$

The speech enhancement comprises of two stages: training and de-noising. In training stage, the SBVs of speech and noise are determined. This can be done by performing NMF on clean speech and noise separately. The KL divergence between the magnitude spectra, $|V_{speech}|$ and $|V_{noise}|$, and their corresponding factored matrices, $W_{speech}.H_{speech}$ and $W_{noise}.H_{noise}$, is minimized using the learning rule as described above. In de-noising stage W_{speech} and W_{noise} are kept fixed and are concatenated to form a single set of SBVs and is termed as W_{mix}. NMF is performed on the magnitude spectrum of noisy speech signal ($|V_{mix}|$), updating only activation matrix H_{mix}. The magnitude spectrum of de-noised speech is reconstructed as $|V_{speech}| \approx W_{speech}.H_{speech}$ where H_{speech} is the rows of H_{mix} corresponding to the activation coefficients of W_{speech}. Finally the spectrogram is recovered using phase of original noisy speech and de-noised speech signal is transformed into time domain.

Speech enhancement using NMF with Phase spectrum compensation

The phase spectrum compensation for speech enhancement was first proposed in [12]. The short time Fourier Transform (STFT) of noisy speech x(n) is given by

$$X(n,k) = |X(n,k)|e^{j\angle X(n,k)} \tag{14}$$

The STFT phase spectrum is modified such that there is large cancellation in noise components during synthesis operation using inverse short time Fourier Transform (ISTFT). Phase spectrum compensation function is calculated as

$$(n,k) = \lambda\psi(k)\left|\hat{D}(n,k)\right| \tag{16}$$

where λ is a empirically determined constant, $\psi(k)$ is the anti-symmetry function and $|\hat{D}(n,k)|$ is an estimate of magnitude spectrum of noise. The anti-symmetry function is given by,

$$\psi(k) = \begin{cases} 1, \text{if } 0 < k < N/2 \\ -1, \text{if } 0 < k < N/2 \\ 0, \text{otherwise} \end{cases} \quad (17)$$

The phase spectrum compensation function is added to the noisy speech spectrum as given

$$X(n,k) = X(n,k) + (n,k) \quad (18)$$

The enhanced speech spectrum is then estimated as

$$\hat{S}(n,k) = |X_M(n,k)|e^{j\angle X(n,k)} \quad (20)$$

where $|X_M(n,k)|$ is the clean speech magnitude spectrum estimated using NMF. $\hat{s}(n)$ is calculated using overlap add procedure with ISTFT.

Speech enhancement using ANC with PSC:

In this method, the noisy speech is first operated upon by ANC. The magnitude and phase spectrum of the resultant speech is then calculated using STFT. The phase spectrum is then modified using (18) and (19). This modified phase is then combined with the magnitude spectrum to reconstruct the estimated speech using ISTFT and overlap-add method.

Results:

Clean speech segments and noise signals, for the test purpose, are taken from NOISEX-92 database. In this work, for ANC, adaptive filters of length 8 taps are considered. In case of LMS algorithm the step size is taken as μ=0.4. In RLS algorithm the forgetting factor is taken as λ=0.9. K=35, basis vectors are considered while enhancing speech using NMF algorithm.

Table 1 shows the results thus obtained. It shows the improvement in SNR obtained using all the methods. From the table it can be interpreted that SNR improvement using phase spectrum compensation in conjunction with NMF and LMS is better than with only NMF and LMS respectively. All the SNRs are in dB.

Fig.2. shows the PESQ score of the enhanced speech obtained after using the different methods described. For different SNR of noisy speech, the PESQ score obtained using RLS is highest as compared to other methods. PESQ score using NMF and LMS with phase spectrum compensation is higher as compared to the PESQ score obtained using only NMF and LMS respectively.

Fig. 3 shows the cepstral distance between the enhanced speech and clean speech. It also shows better performance of methods using phase compensation.

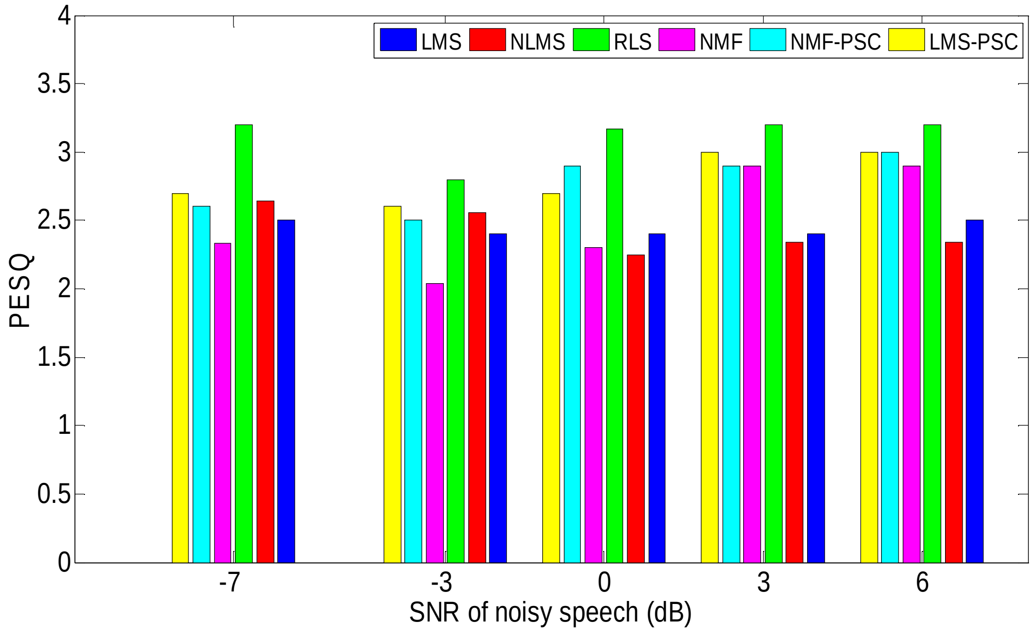

Fig. 2: Comparison of PESQ score of the enhanced speech using different methods

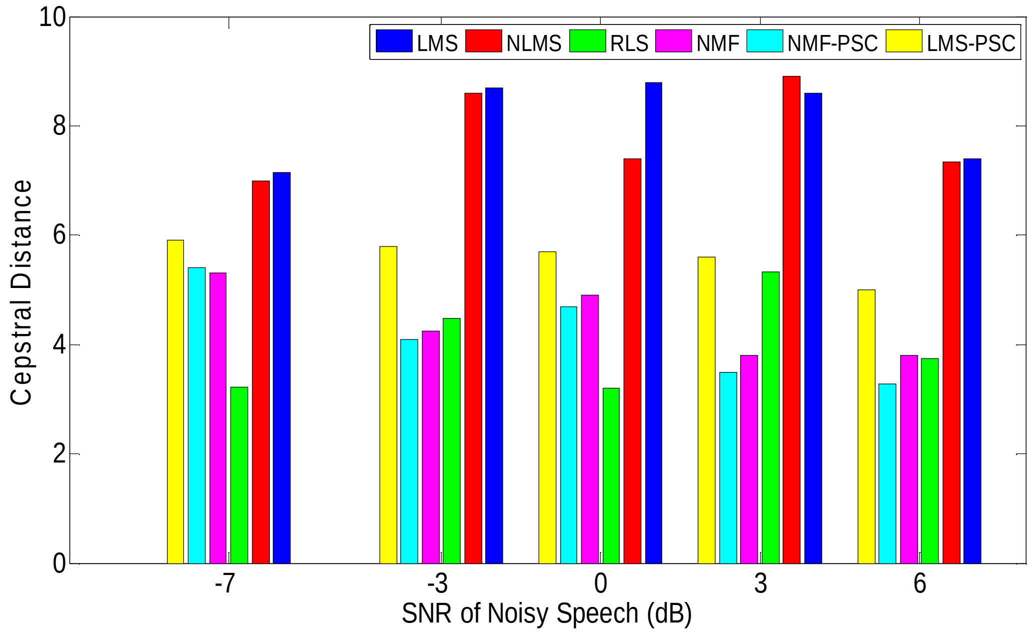

Fig. 3: Comparison of Cepstral distance between clean speech and enhanced speech

Table 1: SNR improvement using different methods

Noise Type	Noisy Speech SNR	LMS	NLMS	RLS	NMF	NMF-PSC	LMS- PSC
Airport	-7	15.6	16	21.3	10.7	12.7	17.6
	-3	12.2	13.06	16.5	8.5	10.5	18
	0	9.4	10.7	13.35	9.6	10.6154	15.05
	3	7.77	8.7	11.3	8.5	10.85	13.97
	6	5.4	6.34	8.3	6.4	10.5	11.8
Babble	-7	16	16.55	21	10.43	11	18.4
	-3	12.9	13	15.2	8.9	9.6	16
	0	9.1	10.2	12.7	6.88	8	14.2
	3	6.4	7.77	10	6.7	7.8	12
	6	5.2	6	6.25	4.5	6.7	10.8
Train	-7	16.77	17.2	19.6	13.2	14	20.15
	-3	13.66	13.8	16	12.37	13.5	19
	0	10.2	11.3	12	9.8	11.4	17.6
	3	7	9	10.6	8	12	15
	6	7.3	7.4	7	8	12	14.5
Pink	-7	16.56	16.67	21.4	11.31	12	20.18
	-3	5	12.7	17.3	9.28	10	17.34
	0	9.85	9.9	14.4	8.3	9.36	15.3
	3	6.6	7.4	11.2	7	8.4	13.8
	6	6	5.4	8	5	7.8	11.35
White	-7	16.5	16.7	20.7	10	11.5	21
	-3	12.2	12.8	16.77	7.8	8.7	16.4
	0	9.5	9.7	13.8	7.7	8.5	15
	3	5	7	10	5	8.5	12
	6	3.47	4.2	7.2	4.7	6	11

Conclusion:

The performance of the speech enhancement methods have been analyzed for different noise conditions. From the results shown it can be interpreted that RLS performs better as compared to other algorithms when used in ANC for noise cancellation. Its main disadvantage is its computational complexity. It can also be concluded that NMF and LMS with phase spectrum compensation performs better than simple NMF and LMS respectively in terms of all the parameters used for analysis. Therefore it is seen that PSC improves the performance of the conventional methods and results in better quality speech.

References

[1] B. Gold and N. Morgan, Speech and Audio Signal Processing, John Wiley & Sons, Inc., 2006.

[2] K. Paliwal, K. Wójcicki and B. Schwerin, "Single-channel speech enhancement using spectral subtraction in the short-time modulation domain, " *Speech Communication,* vol. 52, no. 5, pp. 450-475, May 2010.

[3] K. Paliwal, B. Schwerin and K. Wójcicki, "Speech enhancement using a minimum mean-square error short-time spectral modulation magnitude estimator, " *Speech Communication,* vol. 54, no. 2, pp. 282-305, February 2012.

[4] M. A. A. El-Fattah, M. I. Dessouky, S. M. Diab and F. E. A. El-samie, "Speech Enhancement Using An Adaptive Weiner Filtering Approach, " *Progress in Electromagnetics Research M,* vol. 4, pp. 167-184, 2008.

[5] B.Widrow, J. McCool and C. S. Williams, "Adaptive Noise Cancelling: Principles and Applications, " *Proceedings of the IEEE,* vol. 63, no. 12, December 1975.

[6] B. Widrow and S. D. Stearns, Adaptive Signal Processing, Pearson Education, 1985.

[7] T. Bose, Digital Signal and Image Processing, John Wiley & Sons, Inc., 2004.

[8] S. Haykin, Adaptive Filter Theory, Pearson Education, Inc., 2002.

[9] N. Mohammadiha, P. Smaragdis and A. Leijon, "Supervised and Unsupervised Speech Enhancement using Non-Negative Matrix Factorization, " *IEEE Transactions on Audio, Speech and Language Processing,* vol. 21, no. October, 2013, p. 10, 2013.

[10] Z. Yang, H. Zhang, Z. Yuan and E. Oja, "Kullback-Leibler Divergence for Nonnegative Matrix Factorization, " in *Artificial Neural Networks and Machine Learning - ICANN 2011*, Espoo, Finland, Springer Berlin Heidelberg, 2011, pp. 250-257.

[11] Z.-Y. Zhang, "Nonnegative Matrix Factorization: Models, Algorithms and Applications, " in *Data Mining: Foundations and Intelligent Paradigms*, Springer Berlin Heidelberg, 2012, pp. 99-134.

[12] K. Wojcicki, M. Milacic, A. Stark, J. Lyons and K. Paliwal, "Exploiting conjugate symmetry of the short-time Fourier spectrum for speech enhancement, " *IEEE Signal Process. Lett.,* vol. 15, pp. 461-464, 2008.