

Speech Enhancement Based On DWPT And Nonnegative Matrix Factorization

Farsana P M¹, Robin Abraham²

¹ PG Scholar, Ilahia College Of Engineering And Technology,Ernakulam,, Kerala, India

² Asst.Professor,Head of Department,Ilahia College Of Engineering And Technology ,Ernakulam, Kerala, India

Abstract - Enhancing of speech affected by noise is the key factor of speech enhancement. The main objective of speech enhancement is to advance the perceptual aspects of speech for example overall quality, intelligibility and degree of listener fatigue. For the speech enhancement (SE) techniques, a spectrogram is frequently chosen than the corresponding time-domain raw data, since it exposes more compact presentation together with conspicuous temporal information over a long time period. In this proposed method discrete wavelet packet transform (DWPT) with level five and nonnegative matrix factorization (NMF) are used for speech enhancement. With five level DWPT thirty two sub bands are available for further processing of the noisy speech signal. And the noise estimation is carried out by nonnegative matrix factorization (NMF). When comparing with the existing system having DWPT level three, the proposed system gives better performance. Performance evaluation is done by using signal to noise ratio.

Key Words: SE,DWPT,NMF.

1. INTRODUCTION

Speech is the most natural and the most effective way of communication between human. During the speech communication, the signals contains some noise so when processing the digital speech signals; speech signal enhancement is very important step. Speech enhancement aims to improve speech quality by using various algorithms. It can be at least clarity and intelligibility, pleasantness, or compatibility with some other method in speech processing. Intelligibility and pleasantness are difficult to measure by any mathematical algorithm. Usually listening tests are employed. However, since arranging listening tests may be expensive, it has been widely studied how to predict the results of listening tests. Since humans are capable of detecting and understanding speech at low signal-to-noise ratios without prior knowledge of the speech, the noise or the environment. hence in speech enhancement research, new ideas are often stimulated by study of auditory processes. An important and much-studied feature is that of the filterbank present in the cochlea, within the ear, which splits incoming signals into a large number of band-limited signals prior to further processing. To increase the quality and intelligibility of speech signals [1],several enhancing

techniques are developed and simulated. Noise tracking and signal gain estimation are the two stages needed for speech enhancement. The power of background noise is estimated [2], [3], in the noise-tracking stage. In signal gain estimation stage, to determine the gain factor on the noisy spectrogram to predict the clean speech component, the estimated noise power information is utilized. Advances in enhancement techniques includes nonnegative matrix factorization (NMF) [4],[5]. In our proposed system discrete wavelet packet transform (DWPT) and nonnegative matrix factorization (NMF) are used for enhancing the noisy speech. Discrete wavelet packet transform (DWPT) decomposes the incoming signal into a number of sub bands based on the level of decomposition. This paper organized as follows. In section 2 gives a short description of our proposed system and details of each stage. The simulated results are reported in section 3. This paper concludes with a discussion in section 4.

2. SYSTEM DESCRIPTION

Aim of our proposed system is to remove noise from the mixture and outputs the clean speech signal. For that firstly the incoming noisy speech signal is decomposed into thirty two subbands by using discrete wavelet packet transform (DWPT) [6]. Squaring and framing processes are then applied to each subband signal. An NMF-wise noise estimation scheme is further used to estimate noise with respect to each subband for the subsequent signal gain estimation. All subband signals are updated via the individual gain and then passed through the inverse DWPT (IDWPT) [6],[7] to produce the enhanced time-domain signal. in the title or heads unless they are unavoidable.

The block diagram of our proposed system is shown in Fig 1.The detailed description of each block is given as follows.

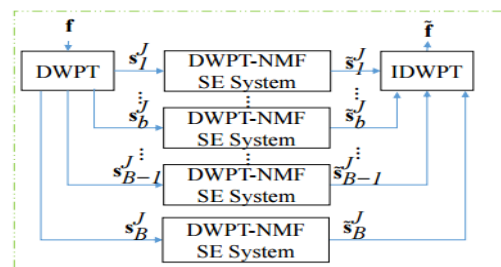


Fig -1: The procedure of DWPT speech enhancement Framework

2.1 Discrete Wavelet Packet Transform

The wavelet packet method is a generalization of wavelet decomposition that offers a richer signal analysis. Wavelet packet atoms are waveforms indexed by three naturally interpreted parameters: position, scale (as in wavelet decomposition), and frequency. The DWPT/IDWPT are performed by a set of well-defined low/high-pass filters together with a down/up sampling process. They serve as a distortion less analysis/synthesis for an arbitrary signal. Fig. 2 depicts the flowchart of the concatenation of the DWPT/IDWPT with level 2. For the left side of Fig. 2, a time signal f is decomposed into two subband signals carrying information of low- and high-frequency components. The length of each subband signal is half of that of the original full band signal due to the factor-2 down sampling operation. The decomposition operation is then applied again to each of the two sub band signals. The DWPT process can be formulated as $s_b^J = DWPT_{b^J} \{f\}$, $b= 1, 2, 3, \dots, 2^J$ where s_b^J denotes the b -th subband signal from a level- J DWPT.

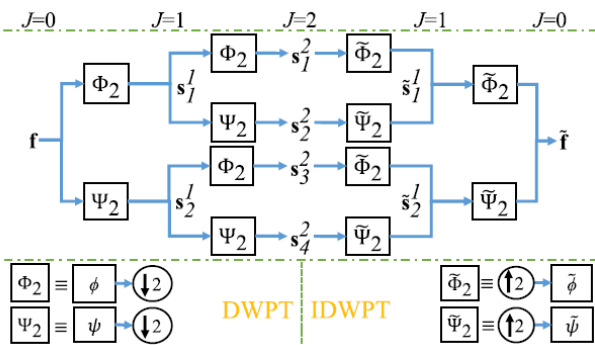


Fig - 2: An example of level-2 (J=2) DWPT and IDWPT

In our proposed system DWPT with decomposition level five is used. The number of subbands obtained by the decomposition is depends on the level. If the level is n then 2^n subbands are obtained after DWPT. Hence here, thirty two subbands are obtained after decomposition.

2.2 DWPT-NMF SE System

Nonnegative matrix factorization (NMF) is a dimension-reduction technique based on a low-rank approximation of the feature space. The NMF algorithm approximates a data matrix V by a product of another two matrices, the basis matrix W and the encoding matrix H , viz., $V \approx WH$, where all of the three matrices contain nonnegative entries only. In NMF-based SE methods, the data matrix V to be processed is usually the magnitude spectrogram of a speech utterance. In our proposed system the data matrix contains sampled values of the noisy speech signal. Which is then factorized via NMF to obtain a basis matrix W and encoding matrix H . The processes inside the DWPT-NMF SE system shown in the following figure.

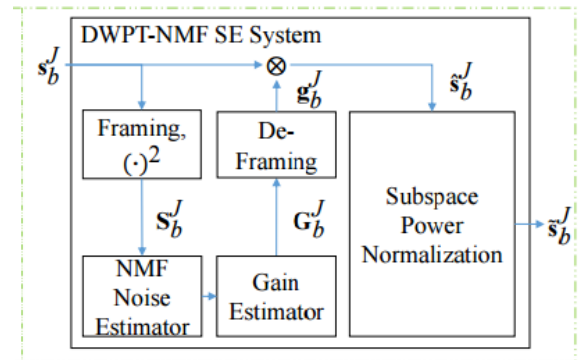


Fig - 3: The DWPT - NMF based speech enhancement

The SE procedures for each sub band signal depicted in Fig. 2 are clarified here. The subband signal s_b^J is first segmented into overlapped frames without further windowing. These frames are arranged in sequence to be columns of a matrix. This matrix is further squared in an element-wise manner to produce a nonnegative matrix S_b^J for the subsequent processing. All subband signals are individually enhanced via NMF-wise compensation. Finally, IDWPT is used to these updated subband signals to reconstruct the enhanced signal f .

2.3 NMF-Based Noise Estimation And Gain Estimation

The noise estimation block is performed by NMF with offline and online phases. At the offline phase, the nonnegative matrices S_b^J for the clean utterances in the training set with respect to a specific subband b are concatenated and then analyzed by NMF to create a speech basis matrix W^{b_s} and encoding matrix H^{b_s} . Likewise, the noise basis matrix W^{b_n} and encoding matrix H^{b_n} associated with sub band b is created thereby using the speech free noise in the training set. For the online phase, the matrix S_b^J of the DWPT sub band signal s_b^J for the input noisy utterance is

$$S_b^J \approx W^{b_s} H^{b_s} + W^{b_n} H^{b_n} \tag{1}$$

where H^{b_s} and H^{b_n} denote the NMF encoding matrices for the speech and noise, respectively. The gain estimation as for sub band b is achieved by

$$G_b^J = \sqrt{(W^{b_s} H^{b_s})} / (W^{b_s} H^{b_s} + W^{b_n} H^{b_n}) \tag{2}$$

where $\sqrt{\cdot}$ and $/$ denote the element-wise square root and division operations, respectively.

Finally, the OLA process is applied to G_b^J to obtain a gain sequence g_b^J that has the same size as the original subband signal s_b^J . Therefore, s_b^J is modulated with g_b^J to produce a new subband signal as

$$\hat{s}_b^J = s_b^J \cdot g_b^J \tag{3}$$

where \cdot stands for an element-wise multiplication.

2.4 Sub-band Power Normalization

A power normalization procedure is applied to the enhanced sub band signal \hat{s}_b^j , to further compensate the noise effect. At the offline phase, we concatenate the DWPT sub band- b signals associated with all the clean utterances in the training set as a clean sample, and then we calculate the root mean square (rms) value for this sample, denoted by $\sigma_{b,c}$. At the online phase, the rms value of \hat{s}_b^j is computed and denoted by $\hat{\sigma}_b$. Then we obtain the power normalized sub band signal by

$$\tilde{s}_b^j = \sigma_{b,c} \frac{\sigma_{b,c}}{\hat{\sigma}_b} \hat{s}_b^j \tag{4}$$

3. EXPERIMENTAL RESULTS

In the proposed method a new feature parameter called Wavelet Packet Transform and nonnegative matrix factorization are used for speech enhancement. For decomposing the signal a five point wavelet packet tree with Db4 wavelet is used. Daubechies D4 wavelet closely matches to the audio signals and are commonly used in speech analysis. To determine the accuracy and performance of the system some evaluations are conducted on noisy speech signal. As a starting point of the analysis training set is created. Which consist of different clean speech utterances and speech free noise signals. Each signal decomposed into thirty two subbands using five level DWPT. Using these training set information clean signal is separated from the noisy speech signal.

The performance was measured by using signal to noise ratio. The existing system with DWPT level three and the proposed with level five are simulated with matlab software and their enhancing performance are evaluated. Signal-to-Noise ratio compares the level of a desired signal to the level of background noise. The higher the ratio, the less obstructive the background noise is. SNR is defined by equation

$$SNR = 10 \log \frac{\sum x(n)^2}{\sum (x(n) - \hat{x}(n))^2} \tag{5}$$

Where $x(n)$ is the original signal and $\hat{x}(n)$ is the reconstructed speech from the mixture.

Table -1: SNR results for Enhanced and Noisy signals

	SNR	
	Existing system	Proposed system
Noisy signal	-0.6698	-0.6698
Enhanced signal	5.7991	7.4541

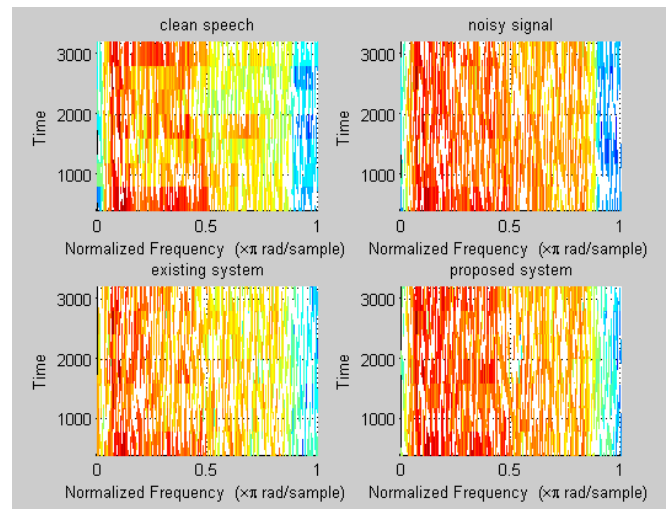


Chart -1: spectrogram

Chart - 1 shows the spectrogram of existing and proposed systems. Welch power spectral density estimate is also given in Chart - 2.

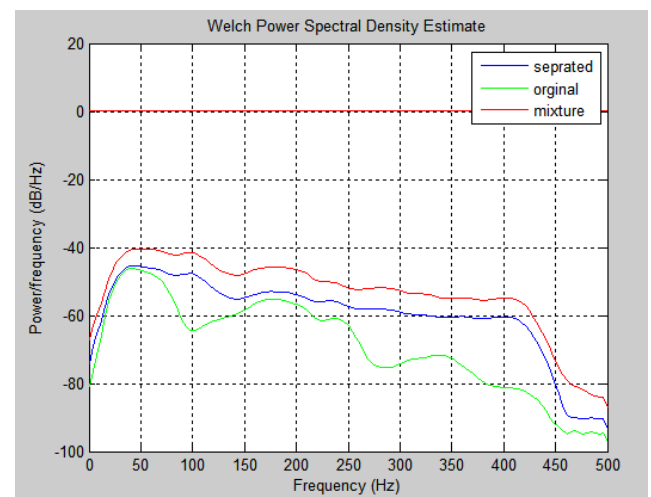


Chart -2: welch power spectral density estimate of proposed system

4. CONCLUSION

In this paper, we presented a new Speech Enhancement Based On DWPT And Nonnegative Matrix Factorization. The DWPT-NMF SE framework provides the noisy speech with a significant improvement in both quality and intelligibility. By analyzing the values of signal to noise ratio it is clear that the proposed DWPT system with level five is superior to the existing system with level three. As we implemented this system as the pre-processing stage of a digital hearing aid, real time realization is very important and hence wavelet packet can be used in the analysis phase.

REFERENCES

- [1] Devyani S. Kulkarni, Ratnadeep R. Deshmukh and Pukhraj P. Shrishrimal, "A Review of Speech Signal Enhancement Techniques" International Journal of Computer Applications (0975 - 8887) Volume 139 No.14, April 2016.
- [2] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, Jul. 2001. R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.
- [3] I. Cohen, "Speech enhancement using a noncausal a priori SNR estimator," *IEEE Signal Process. Lett.*, vol. 11, no. 9, pp. 725–728, Sep. 2004.
- [4] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 3, pp. 550–563, Mar. 2010.
- [5] N. Mohammadiha, P. Smaragdis, and A. Leijon, "Supervised and unsupervised speech enhancement using nonnegative matrix factorization," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 10, pp. 2140–2151, Oct. 2013.
- [6] M. Gokhale *et al.*, "Time domain signal analysis using wavelet packet decomposition approach," *Int. J. Commun. Netw. Syst. Sci.*, vol. 3, no. 3, p. 321, 2010.
- [7] D. D. Ariananda, M. K. Lakshmanan, and H. Nikookar, "An investigation of wavelet packet transform for spectrum estimation," *arXiv preprint arXiv:1304.3795*, 2013