PERFORMANCE ENHANCEMENT OF DCT BASED SPEAKER RECOGNITION USING WAVELET DE-NOISING

Munmuni Patir[1], V.K Govindan[2]

[1]PG Scholar, Dept. of Computer Science & Engineering, National Institute of Technology Calicut, Kerala, India
[2]Professor, Dept. of Computer Science & Engineering, National Institute of Technology Calicut, Kerala, India

Abstract

Presence of noise in the speech signal is one of the major problems in Speaker Recognition. The speaker recognition performance gradually degrades as the intensity of noise increases. The system gives high accuracy when the speech signal is noise free, but in real life scenario getting a noise free speech signal is challenging. Hence, elimination of the noise from speech signal is an important aspect in speaker recognition process. This work uses wavelet based denoising of the recorded speech signal in order to enhance the performance of speaker recognition. In this paper, wavelet based denoising technique has been applied to the DCT based speaker recognition system which was proposed in our previous work. Additive white Gaussian noise has been added to the speech signal and performance analysis of the system has been done using different SNR value.

Keywords: Wavelet denoising, AWGN, Speaker Recognition, Thresholding, DCT, Feature Extraction.

---***---

1. INTRODUCTION

Major challenges that have been faced in Automatic Speaker Recognition are variations in speaker's vocal tract, background noise, quality of the microphone. The presence of noise in recorded speech signal in speaker recognition system degrades the performance of the system; hence, noise removal plays an important role in speaker recognition. The noise free enhanced speech signal provides better performance in recognizing speakers.

In our previous work speaker recognition system was developed by proposing a modification to MFCC where DCT is used instead of Fourier transformation of MFCC [1].In this work noise removal method is applied to the DCT based speaker recognition.

In recent years, denoising has been widely used in signal and image processing in order to improve the signal quality. It has become an essential pre processing step to achieve better results in any image processing task [2][3][4].

Wavelet denoising techniques has been found to be one of the powerful technique for getting high quality image and signal [2]. Nonlocal means filtering techniques are also used in many image and signal processing applications [5] [6] [7].

In this proposed work, Wavelet denoising technique is applied on recorded speech in order to enhance the speech signal. Denoising is performed before the feature extraction to enhance the speaker recognition performance.

1.1 Wavelet-Based Denoising

Wavelet denoising which is widely used nowadays for eliminating noise makes use of wavelet transformation. As wavelet representation is better than Fourier transform representation, wavelet Denoising leads to significant improvements in signal quality.

Wavelet based noise removal techniques perform better as the shape of the signal is preserved when compared to denoising based on Fourier Transforms. The steps involved in wavelet denoising process are as follow [2]:
• Transformation of the noisy signal to obtain wavelet coefficients at different decomposition levels; this provides the noisy Wavelet coefficients.
• Selection of a threshold for different levels, and the types of thresholding techniques such as hard thresholding or soft thresholding for removal of noise.
• Transforming back to the spatial domain for obtaining the noise free signal Denoised signal is obtained by applying inverse wavelet transform to the wavelet coefficients obtained after the thresholding method.

1.1.1 Wavelet Thresholding

The wavelet coefficients after the wavelet transformation which have low energy are removed by performing thresholding as their contribution to the wavelet spectra is very low and can be considered as noise [3].

Thresholding can be done either by soft thresholding or hard thresholding [2].

Hard thresholding is define by the following equation (1)-

$$w_{ht} = \begin{cases} w, & |w| \geq t \\ 0, & |w| < t \end{cases} \qquad (1)$$

In Soft thresholding, above the threshold value, the wavelet coefficients shrink by that threshold value, it is also known as wavelet shrinkage.

Equation (2) defines the soft thresholding.

$$w_{st} = \begin{cases} [sign(w)](|w| - t), |w| \geq t \\ 0, |w| < t \end{cases} \qquad (2)$$

In the above Equation, w = wavelet coefficient value, and t= threshold value.

In this method, the wavelet coefficients below the threshold value are considered to be noise, and the shrinkage of magnitudes of all the coefficients are done by that threshold amount for eliminating the contribution due to noise [3].

In this work, Soft thresholding method has been used. Soft thresholding does not lead to severe distortion of the speech signal compared to the hard thresholding [8]. It has been also found that generally for image denoising also soft thresholding gives better results [3].

1.1.2 Determination of Thresholds

Finding the appropriate threshold value also is an important step in wavelet denoising. There are various way to determine the thresholds if the wavelet coefficients are known.

Following are two mostly used methods to determine the threshold [9]:

- For fixed thresholding based denoising, the threshold t may be computed using Equation (3), where σ = Median-of-coefficients/0.6745, and n =total coefficients :

$$t = \sigma\sqrt{2\log(n)/n} \qquad (3)$$

- For Global Thresholding, the thresholding may be computed using Equation (4):

$$t = 2\sqrt{\log(n)} \qquad (4)$$

2. RELATED WORK

In recent years many noise removal methods have been proposed for speech enhancement. As speech signal is mostly affected by many kinds of noises such as random noise, background noise, the removal of the noise became an important step in speaker recognition in order to increase the robustness of the system. Some of the important works dealing with noise removal is briefly reviewed in the following:

Rozeha A. Rashid et al. [10], in their work dealing with voice recognition, made use of band-pass filter, called digital FIR band-pass filter, to get rid of noise speech signal.

In [11], Radu et al. has proposed an improved method of spectral subtraction in which the ability of human ear to mask noise is considered into account in order to remove the background noise. After spectral subtraction, the remaining residual noise is masked. For spectral subtraction, they have used over subtraction factor technique to effectively remove noise.

Sumithra et al. [12] proposed an adaptive wavelet shrinkage method for suppressing the noise. Wavelet subband coefficient threshold computation is done automatically in this method. The thresholds computed are proportional to the noise contamination.

Khaled Daqrouq et al. [13] proposed a multistage convolution approach in which Reverse Biorthogonal Wavelets are used. The authors observed better quality improvements in the enhancement of the speech signal.

Khaled et al. [14] compared three different methods for enhancement of speech signal; that are- adaptive linear neuron (ADALINE), Feed Forward Neural Network Enhancement (FFNN), and wavelet transformation based ADALINE Enhancement Methods. In the 3rd method, the desired sub-signals are generated by using DWT which is the input to the neural net (ADALINE). They concluded that this method gives better enhancement and can be used in speaker identification.

Lina Liu and Jishun Jiang [15] used denoising method based on the stationary wavelet transformations. In their work, they perform denoising by using wavelet transformation as well as stationary wavelet transformation and concluded that the latter method has better effect on elimination of AWGN in the signal.

Zhou Dexiang and Wang Xianrong [16] used wavelet denoising in their work in speech recognition and proposed and improved HMM algorithm.

From the above survey, it has been found that many denoising methods for speech signal enhancement has been proposed by researchers in the recent years to enhance speech signal quality. As the recognition of speakers highly depends on the quality of speech signal, the enhancement of recorded signal can plays a vital role in the efficiency of speaker recognition system. This paper proposes to study the effect of wavelet based denoising on the performance of DCT based speaker recognition system.

3. EXPERIMENTAL RESULTS

In our previous work [1], the speech signal was recorded through microphone. Although the speech signal contain some amount of random noise, background noise etc., it was directly taken for feature extraction by using a modified MFCC method proposed. As MFCC method involves Fourier transformation some denoising is done automatically by Fourier transformation process.

In this work, wavelet denoising has been done before the feature extraction step. Wavelet denoising enhanced the quality of the recorded speech signal and provided better recognition rate. Table 1 show the experimental result of efficiency in recognition when denoised speech signal is taken into account.

Table-1: Matching out of 10 times where the speech signal was denoised by using wavelet denoising.

Speaker Id	Number of matches out of 10 utters
1	10
2	7
3	9
4	9
5	8
6	6
7	9
8	9
9	7
10	7
11	9
12	10
13	9
14	7
15	9
16	8
17	9
18	9
19	8
20	9
Average Match percent (%)	84

Experiments have also been done by adding Additive White Gaussian Noise (AWGN) to the speech signal, and the performance of the system developed in the previous work [1] has been analyzed for different Signal to Noise Ratio (SNR) values of the AWGN. Figure 1, Figure 2 and Figure 3 shows speech signal with AWGN with different SNR value. Table 2 shows the efficiency percent of the system with respect to SNR values of the AWGN.
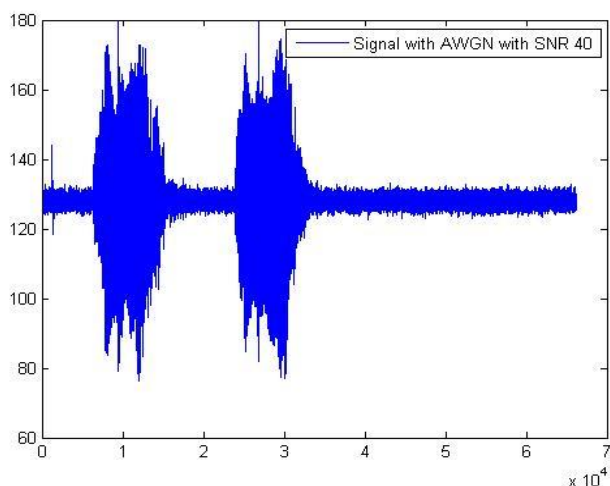

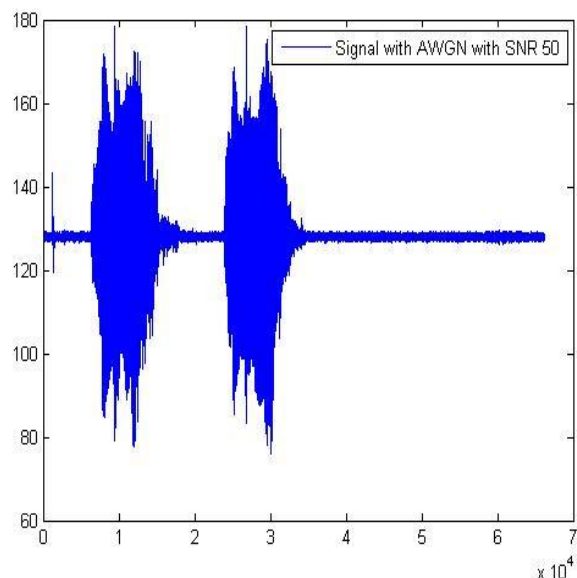**Fig-1:** Speech signal with Additive White Gaussian Noise of SNR value 40

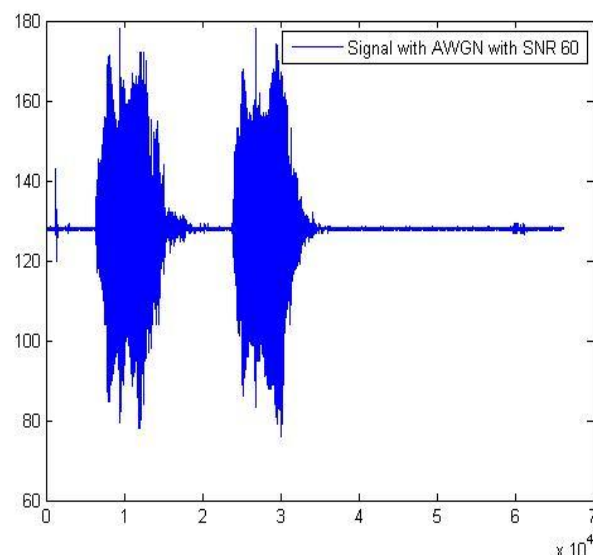**Fig-2:** Speech signal with Additive White Gaussian Noise of SNR value 50

Fig-3: Speech signal with Additive White Gaussian Noise of SNR value 60

Table-2: Performance analysis of the DCT Based Speaker recognition system when Gaussian noise is added to speech signal.

SNR values	40	45	50	60	65	Without AWGN added
Efficiency in %	0	62	66	75	79	81

4. CONCLUSION AND FUTURE WORK

A modification to 'DCT based MFCC in speaker recognition' which was proposed in our previous work has been done by using wavelet denoising method. It was observed that the enhancement in quality of recorded signal provided better system performance. When denoising was applied, the recognition efficiency increased from 81% to 84%, which is a significant improvement. The performance of the system is analyzed by adding Additive White Gaussian Noise of different SNR values to the speech signal, and it has been found that the efficiency of the system degrades as the noise level increases. As the SNR value decreases the performance decreases, and for SNR less than 45, the performance was totally unacceptable. For SNR above 65, the performance was found to be in the acceptable range.

In our work, the simple Wavelet denoising method has been used but there are many improvements proposed for wavelet denoising that can be used to achieve better accuracy of the system. As a future work, different denoising techniques can be implemented and studied to propose a better denoising method in speaker recognition.

REFERENCES

[1]. Munmuni Patir, V.K. Govindan, "DCT Based MFCC in Speaker Recognition," International Journal of Research in Computer Applications & Information Technology Volume 2, Issue 6, November-December, 2014

[2]. Burhan Ergen, "Signal and Image Denoising Using Wavelet Transform," Advances in Wavelet Theory and Their Applications in Engineering, Physics and Technology, 2010, pp496-514[Sd-008].www.intechopen.com

[3]. Baby Vijilin,V.K.Govindan, "Optimal Threshold Selection for Wavelet Transform based on Visual Quality," International Journal of Computer Applications (0975 – 8887) Volume 81 – No 19, November 2013.

[4]. J.Dinesh Peter, and V.K.Govindan, "A Phase and intensity preserved image de-noising technique," 14th International conference on advanced computing and communication (ADCOM 2006), National Institute of technology Karnataka, India, 2006.

[5]. Dineshpeter, V K Govindan and Abraham T Mathew, " Nonlocal-Means Image Denoising Technique using Robust M-estimator," Journal of Computer science and Technology, Springer, 25(3): 623-631, May 2010.

[6]. Dineshpeter, V K Govindan and Abraham T Mathew, Robust Estimation Approach for Nonlocal-means denoising based on structurally similar patches, International Journal of Open problems in Compt. Maths., Vol. 2, No. 2, June 2009.

[7]. J. Dinesh Peter, V.K. Govindan, Abraham T. Mathew, "Robust Estimation Approach for NL-Means Filter" In Proceedings: ISVC'08, Part II, LNCS 5359, pp. 571-580, 2008, USA (Las Vegas). (LNCS-Springer).

[8]. Jeena Joy , Salice Peter , Neetha John, "Denoising Using Soft Thresholding," International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 2, Issue 3, March 2013.

[9]. P. Karthikeyan, M. Murugappan, and S.Yaacob, "ECG Signal Denoising Using Wavelet Thresholding Techniques in Human Stress Assessment." International Journal on Electrical Engineering and Informatics - Volume 4, Number 2, July 2012.

[10]. Rozeha A. Rashid, Nur Hija Mahalin, Mohd Adib Sarijari, Ahmad Aizuddin Abdul Aziz, "Security System Using Biometric Technology: Design and Implementation of Voice Recognition System (VRS)," Proceedings of the International Conference on Computer and Communication Engineering 2008.

[11]. Radu M. Udrea, Silviu Ciochina, Drago N. Vizireanu, "Reduction of Background Noise from affected Speech using a Spectral Subtraction Algorithm based on Masking Properties of the Human Ear," 0-7803-9164-0/05/$20.00 (2005 IEEE)

[12].Sumithra ManimegalaiGovindana, PrakashDuraisamyb, XiaohuiYuan, "Adaptive wavelet shrinkage for noise robust speaker recognition," Digital Signal Processing 33 (2014) 180–190.

[13]. Khaled Daqrouq,Ibrahim N, Abu-Isbeih, Omar Daoud,Emad Khalaf, "An investigation of speech enhancement using wavelet filtering method," Int J Speech Technol (2010) 13: 101–115.

[14]. Khaled Daqrouq,Ibrahim N, Abu-Isbeih, Omar Daoud,Emad Khalaf, "Speech Signal Enhancement Using Neural Network and Wavelet Transform," IEEE International Multi-Conference on Systems, Signals and Devices, 2009.

[15]. Lina Liu ,Jishun Jiang, "Using Stationary Wavelet Transformation for Signal Denoising,"IEEE International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2011

[16]. Zhou Dexiang , Wang Xianrong, "The Improvement of HMM Algorithm using wavelet de-noising in speech recognition," IEEE International Conference on Advanced Computer Theory and Engineering(ICACTE),2010.