www.ioirp.com
International Journal of Innovative Research in Technology, Science & Engineering (IJIRTSE)
ISSN: 2395-5619, Volume – 2, Issue – 4. April 2016

Efficient Acoustic Noise Cancellation In Non-Audible Murmur Using Wavelet Transform

R.Elavarasi
PG Scholar
Department of Electronics and Communication Engineering,
Priyadarshini Engineering College,
Vaniyambadi – 6005,Vellore, INDIA
Email: elavarasir25@gmail.com

Dr.M.Umadevi
Professor
Department of Electronics and Communication Engineering,
Priyadarshini Engineering College,
Vaniyambadi – 6005,Vellore, INDIA

***Abstract*----In this paper, we present statistical approaches to enhance body-conducted unvoiced speech for silent speech communication using wavelet transform. A body-conducted unvoiced speech is called non-audible murmur, NAM microphone is effectively used to detect very soft unvoiced speech which is emitted outside almost inaudible. Analysis of NAM speech has been made using hidden Markov model (HMM) and Gaussian mixture model (GMM). In this paper study of analyzing NAM speech using haar and db2 as it is used to extract the features of various types of speech signal. Wavelet transform is capable of revealing aspects of data that other speech signal analysis technique such the extract features are then passed to classifier for the recognition of speech. The experimental results show that NAM is effectively converted to Normal Speech which improves intelligibility.**

Keywords: Non Audible Murmur, HMM, GMM, VC, Wavelet Transform

I. INTRODUCTION

SPEECH communication plays a very important role in our daily life. It is the most popular method of human communication. In recent decades, the style of speech communication has considerably changed with the advancement of information technology. For instance, the explosive spread of cell phones has enabled people to talk with each other whenever they want and has brought a more convenient style of speech communication to us. Although cell phones have made speech communication possible in various situations, there are actually some instances where we face difficulties in speech communication. For ex-ample, we would have trouble privately talking in a crowd; speaking itself would sometimes annoy others in quiet environments such as in a library; and we may lose our voice if subjected to surgery to remove speech organs such as the larynx due to laryngeal cancer. Many barriers still exist in speech communication. The development of technologies to overcome these inherent problems of speech communication is essential to make our speech communication more universal[1].

Recently, silent speech interfaces have attracted attention as a technology to support new speech communication styles. They enable speech communication to take place without the necessity of emitting an audible acoustic signal. There have been several attempts to explore sensing devices as alternatives to the air-conductive microphone, such as the throat microphone, electromyography (EMG), ultra-sound imaging and so on[1]. These sensing devices are useful to detect soft speech in a private conversation and also effective as a speaking aid for the vocally handicapped. In addition, they are also effective for noise-robust speech communication. As one of the microphones to detect body-conducted speech, the non-audible murmur (NAM) microphone has been developed by Nakajima Inspired by a stethoscope, the NAM microphone was originally developed to detect extremely soft murmur called NAM, which is so quiet that people around the speaker barely hear its emitted sound. Although NAM is a truly different medium from natural voices, it can be used easily by anyone whose speech organs function reasonably well. Placed on the neck below the ear, the NAM microphone is capable of detecting air vibrations in the vocal tract from the skin through only the soft tissues of the head. High-quality body-conductive recording of various types of speech, such as a very soft murmur as NAM, a whispered voice, soft voices, and normal speech, is possible from this position because the conduction through obstructions, such as bones whose acoustic impedance is different from that of soft tissues, is avoided. It is also robust against external noise owing to its noise-proof structure like in other body-conductive microphones[2][5][6].

II. BODY CONDUCTED UNVOICED SPEECH

Non-Audible Murmur (NAM) is the term given to the low amplitude sounds generated by laryngeal air flow noise and its resonance in the vocal track. NAM sound radiated from the mouth can barely be perceived by nearby listeners, but a signal is easily detected using a high sensitivity contact microphone attached on the skin through the soft tissue in the or-facial region. The NAM microphone is designed to detect tissue vibration in or-facial region during speech while being insensitive to environmental noise[2][3][4].

A body-conductive microphone called Non Audible Murmur (NAM) microphone is effectively used to detect very soft unvoiced speech. However, body-conducted unvoiced speech is difficult to use in human-to-human speech communication because it sounds unnatural and less intelligible due to the acoustic change caused by body conduction. To overcome this, voice conversion (VC) methods can be introduced, where the acoustic features of body conducted unvoiced speech are converted into natural voices.

Fig. 1.1 setting position of NAM microphone

Fig.1.2 setting position of MEMS

Figure1.1 shows the setting position of NAM microphone. NAM can be sensed by the ear alone sing a stethoscope placed beneath the chin while whispering. A small stethoscope equipped with a microphone thus appeared to be a simple sensor for use in many situations where speaking aloud is not desirable.The best location for placing the NAM microphone was empirically determined to be on the skin below the mastoid process on a large neck muscle. All living organisms contain biological sensors with functions similar to those of the mechanical devices described. Figure1.2 shows the setting position of MEMS sensor. NAM can be sensed by the throat alone sing a MEMS sensor placed beneath the chin while whispering. A small sensor equipped with a microphone jack thus appeared to be a simple sensor for use in many situations where speaking aloud is not desirable. This sensor consists of axis x, y, z using 3pin connectors which is connected to power supply board along with 12V step down transformer. When the deaf and dump people try to deliver their speech during the time vibration was generated in vocal track. It was sensed by MEMS sensor and recorded as wave file through the microphone jack. The wave file was saved with the help of sigview.

III WAVELET TRANSFORM

Wavelet transform has the ability to analyze different speech quality problems simultaneously in both time and frequency domain. The wavelet transform is used to extract the features of various types of speech signal. Here we use discrete wavelet transform, because it can overcomes the disadvantage of generating large amount of wavelet coefficients as of in the continuous wavelet transform.

A. Discrete Wavelet Transform

The DWT analyze the signal at different frequencies with different resolutions. It can be possible by decomposing the signal into a coarse approximation and detail information. DWT has two set of functions, called scaling functions and wavelet functions. Those two functions are associated with low pass and high pass filters. Half band low pass filtering can be used to remove half of the frequencies, which can be interpreted as loss half of the information. Therefore, the resolution is reduced after the filtering operation. However, the sub sampling after filtering operation does not affect the resolution since half of the spectral components can be removed from the signal makes half the number of samples are redundant. Half of the samples can be discarded without any loss of information. The low pass filtering reduces half of the resolution, but the sample leaves the scale can be remain unchanged. The signal is then subsample by 2 since half of the numbers of samples are redundant. This leads to double the scale.

One important property of the discrete wavelet transform is the relationship between the impulse responses of the high pass and low pass filters. The decomposition of the signal into different frequency bands is obtained by successive

high pass and low pass filtering of the time domain signal. The signal x[n] is first passed through a half band high pass filter g[n] and then to low pass filter h[n].

Thus the algorithm can be summarized as:

- Framing input speech signal
- Forward WT of a frame
- Thresholding wavelet coefficients
- Inverse WT
- Keep center part of the frame.

IV. BLOCK DIAGRAM

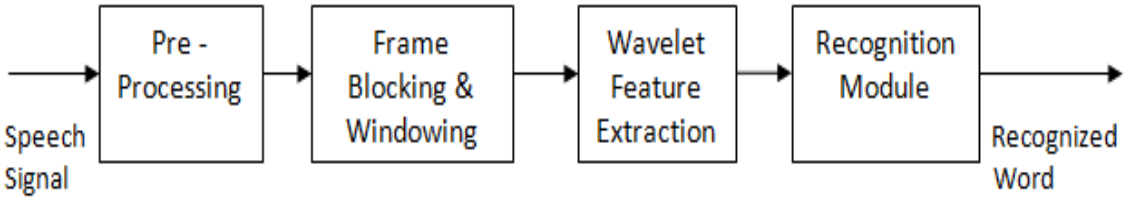

Fig. 3.5 Block diagram for speech signal recognition

A. Pre-Processing

Pre-processing of Speech Signal serves various purposes in any speech processing application. It includes Noise Removal, Endpoint Detection, Pre-emphasis, Framing, Windowing, Echo Cancelling etc. Out of these, silence/unvoiced portion removal along with endpoint detection is the fundamental step for applications like Speech and Speaker Recognition. Pre- Processing of speech signals, i.e. segregating the voiced region from the silence or unvoiced portion of the captured signal is usually advocated as a crucial step in the development of a reliable speech or speaker recognition system. This is because most of the speech or speaker specific attributes are present in the voiced part of the speech signals; moreover, extraction of the voiced part of the speech signal by marking and removing the silence and unvoiced region leads to substantial reduction in computational complexity.

B. Frame blocking

The common approaches in speech signal processing are based on short time analysis. The pre-emphases signal is segmented into frames. The digitized speech signal is blocked into overlapping frames. Frame duration typically ranges between 10-3 msec. Values in this range represent a tradeoff between the rate of change of spectrum and system complexity. The proper frame blocking is ultimately dependent on the velocity of the articulators in the speech production system. Some sounds exhibits sharp spectral transitions which can result in Spectral peaks shifting as much as 80 Hz/msec. the amount of overlap to some extent controls how quickly parameters can change from frame to frame with optional overlap of 1/3~1/2 of the frame size. If the sample rate is 16 kHz and the frame size is 320 sample points, then the frame duration is 320/16000 = 0.02 sec = 20ms. Additional, if the overlap is 160 points, then the frame rate is 16000/(320-160) = 100 frames per second.

C. Hamming windowing

Hamming window is applied to the frame to minimize the effect of discontinuities at the edges of the frame. Each frame has to be multiplied with a hamming window in order to keep the continuity of the first and the last points in the frame. If the signal in a frame is denoted by s(n), n = 0,…N-1, then the signal after Hamming windowing is denoted as s(n)*w(n), where w(n) is the Hamming window can be defined by:

$$w(n) = 0.54 - 0.46cos\left(\frac{2\pi n}{N-1}\right)$$

D. Feature extraction

Feature extraction involves analysis of speech signal. Broadly the feature extraction techniques are classified as temporal analysis and spectral analysis technique. In temporal analysis the speech waveform itself is used for analysis. In spectral analysis spectral representation of speech signal is used for analysis. Then the word can be recognized using the above extracted feature from which the isolated word can be obtained.

V. SIMULATED RESULTS

A. Simulation Result Using HAAR

Simple wavelet analysis was used in the computationally efficient source feature extraction. To clarify the effect of the proposed diagonalization, Thus the following results are analyzing the efficiency of de-noised speech signal using haar wavelet.

Step:1
Speech samples of deaf and dumb which is sensed by mems sensor as a vibration generated from the internal muscular tissues in the vocal track.

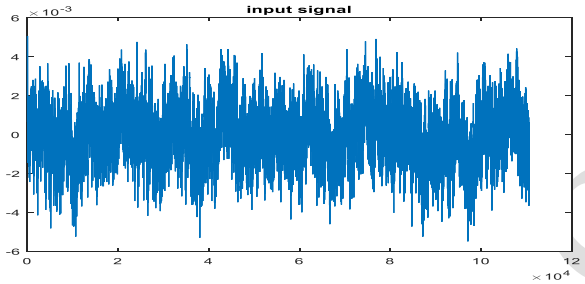

Fig 4.1 signal corrupted by noise

Step :2
After initializing the input signal it undergoes different level of decomposition using haar transform in order to separate the noise from noise corrupted signal. Here it performs 5 level of decomposition. If the decomposition level increased then its accuracy get decreased. Hence the decomposition level 5 gives best noise separation of NAM speech.

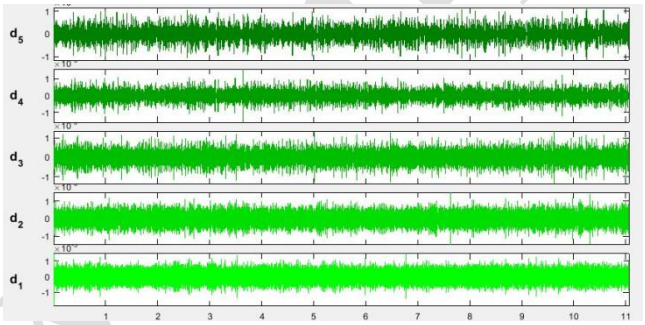

Fig.4.2 Decomposed signal using haar

Step:3
After decomposition it separates the original speech signal. Fig4.3 shows denoised speech signal of deaf and dumb using haar tansform.

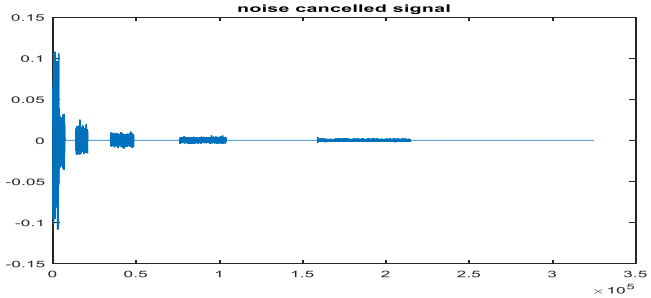

Fig4.3 Denoised signal

B. Simulation Result Using db2 wavelet

Simple wavelet analysis was used in the computationally efficient source feature extraction. To clarify the effect of the proposed diagonalization, Thus the following results are analyzing the efficiency of de-noised speech signal using haar wavelet.

Step :1
Initializing the input signal

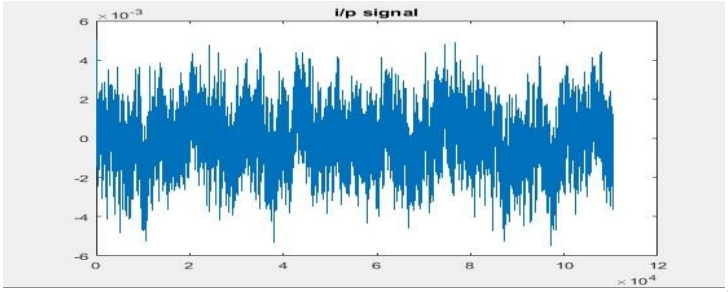

Fig4.4 signal corrupted by noise

Step:2

After initializing the input signal it undergoes different level of decomposition using db2 transform in order to separate the noise from noise corrupted signal.

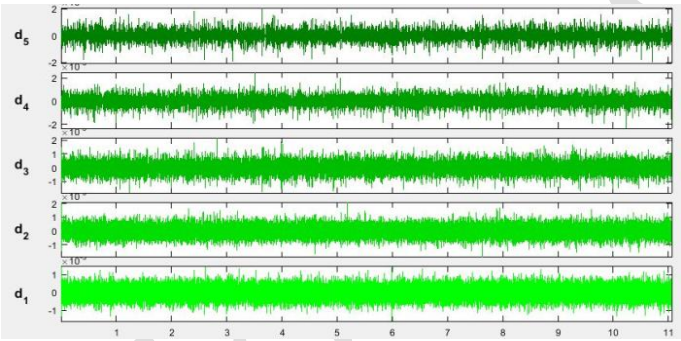

Fig.4.5 Decomposed signal using db2

Step:3

After decomposition it separates the original speech signal. Fig4.3 shows denoised speech signal of deaf and dumb using db2 transform.

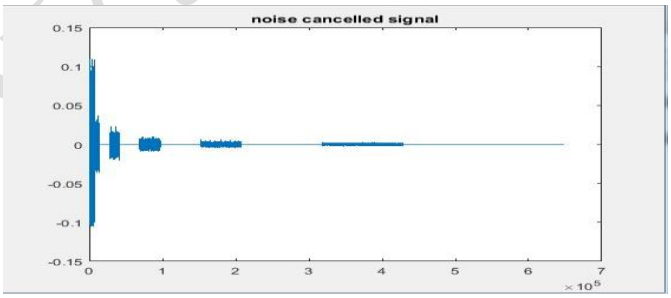

Fig4.6 Denoised signal

From the above result we are analyzing the efficiency of NAM speech after denoising by using different wavlet transform. Haar wavelet gives symmetric, smooth, and regular denoised signal whereas db2 Wavelet is asymmetric, sharp, and irregular denoised signal.

VI. CONCLUSION

In this paper statistical analysis of NAM speech has been analyzed using wavelet transform such as haar and db2. Recognition of speech using wavelet transform gives more efficiency than the speech when compared to other analysis such as hidden markov model (HMM) and Gaussian mixture model (GMM). Our proposed method is capable of separating the noise from body conducted unvoiced speech using haar and db2. From this results we conclude daubechies shows best compression and noise removal of speech signal processing. In future it is proposed to use

Gabor transform for noise suppression in NAM speech and also to implement the hardware for deaf and dumb using stellaris processor.

REFFERENCE

[1] Statistical Voice Conversion Techniques for Body-Conducted Unvoiced Speech Enhancement Tomoki Toda, *Member, IEEE*, Mikihiro Nakagiri, and Kiyohiro Shikano, *Fellow, IEEE*

[2] B. Denby, T. Schultz, K. Honda, T. Hueber, J. M. Gilbert, and J. S. Brumberg, "Silent speech interfaces," *Speech Commun.*, vol. 52, no. 4, pp. 270–287, 2010.

[3] S.-C. Jou, T. Schultz, and A. Waibel, "Adaptation for soft whisper recognition using a throat microphone," in *Proc. INTERSPEECH*, Jeju Island, Korea, Sep. 2004, pp. 1493–1496.

[4] T. Schultz and M. Wand, "Modeling coarticulation in EMG-based continuous speech recognition," *Speech Commun.*, vol. 52, no. 4, pp. 341–353, 2010.

[5] T. Hueber, E.-L. Benaroya, G. Chollet, B. Denby, G. Dreyfus, and M. Stone, "Development of a silent speech interface driven by ultrasound and optical images of the tongue and lips," *Speech Commun.*, vol. 52, no. 4, pp. 288–300, 2010.

[6] A. Subramanya, Z. Zhang, Z. Liu, and A. Acero, "Multisensory pro-cessing for speech enhancement and magnitude-normalized spectra for speech modeling," *Speech Commun.*, vol. 50, no. 3, pp. 228–243, 2008.

[7] Y. Nakajima, H. Kashioka, N. Cambell, and K. Shikano, "Non-audible murmur (NAM) recognition," *IEICE Trans. Inf. Syst.*, vol. E89-D, no. 1, pp. 1–8, 2006.

[8] T. Hirahara, M. Otani, S. Shimizu, T. Toda, K. Nakamura, Y. Naka-jima, and K. Shikano, "Silent-speech enhancement using body-con-ducted vocal-tract resonance signals," *Speech Commun.*, vol. 52, no. 4, pp. 301–313, 2010.

[9] T. Toda, K. Nakamura, T. Nagai, T. Kaino, Y. Nakajima, and K. Shikano, "Technologies for processing body-conducted speech detected with non-audible murmur microphone," in *Proc. INTER-SPEECH*, Brighton, U.K., Sep. 2009, pp. 632–635.

[10] Articulatory Controllable Speech Modification Based on Statistical Feature Mapping with Gaussian Mixture Models