ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 1,January 2013

COMPARISON AND ANALYSIS OF NAM AND ACOUSTIC SPEECH USINGWAVELET TRANSFORM

RADHIKA.R[1], VANITHA LAKSHMI.M[2], AADITYA VELAVA.V.A[3]

PG Scholar, Dept of PG studies in Engineering, S.A Engineering College, Chennai, India[1]

Assistant professor, Dept of PG studies in Engineering, S.A Engineering College, Chennai, India[2]

PG Scholar, Dept of PG studies in Engineering, S.A Engineering College, Chennai, India[3]

ABSTRACT: ***Non-Audible Murmur (NAM) is a very quietly uttered speech received through the body tissue with the use of special acoustic sensors (i.e., NAM microphone) attached behind the talker's ear. Analysis of NAM speech has been made only using HMM (Hidden Markov Model) and GMM (Gaussian Markov Model). In this paper, study of analysing acoustic speech and NAM speech will be made using Wavelet transform. Since, wavelets have the great advantage of being able to separate the fine details in a signal. It allows complex information such as music, speech, and patterns to be decomposed into elementary forms at different positions and subsequently reconstruct it with high precision. Upon analysing the Normal speech, NAM speech will also be analysed using the same wavelet transform. The accuracy of words obtained from the two methods will be compared to determine the efficiency of using Wavelet transform in recognizing NAM speech.***

Keywords- NAM (Non-Audible Murmur), HMM (Hidden Markov Model), GMM (Gaussian Markov Model), NAM Microphone

I.INTRODUCTION

Non-audible murmur (NAM) is an unvoiced speech that can be received through the body tissue using special acoustic sensors (i.e., NAM microphones) attached behind the talkers ear. By attaching the NAM microphone behind the talker's ear, the microphone can able to capture an inaudible murmur (NAM speech), which cannot be heard by listeners near the talker.

Privacy, robustness to environmental noise, and a tool for sound-impaired people are the advantages of NAM microphone, when it is applied in a speech recognition system.Although the NAM signal is of poor quality, the signal envelope is similar to that of normal speech, and therefore speech recognition is possible.

Figure 1: Normal speech waveform

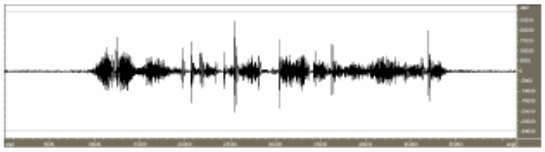


Figure 2: NAM speech waveform

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 1,January 2013

Figures 1 and 2 show normal-speech and NAM-speech waveform. For the recording of normal speech, a close-talking transmission some components are lost, therefore the NAM speech is of lower quality compared to the normal speech.

By using HMM (Hidden Markov Model) for the analysis of NAM speech, the HMM distances of NAM speech are reduced compared to that of normal speech. Because of spectral reduction, NAM's unvoiced nature, and the type of articulation, NAM sounds become similar, causing a larger number of confusions when compared with normal speech. Reduction in HMM distance of NAM speech thus decreases the recognition performance. So in addition to audio features, parameters extracted from facial shape were also used in NAM recognition to increase the performance which is quite complicated. Hence we analyze NAM speech using wavelet transform.

II.WAVELET TRANSFORM

Wavelet transform has ability to analyze different speech quality problems simultaneously in both time and frequency domain. The wavelet transform is useful in extracting the features of various types of speech signal.

Here we use discrete wavelet transform, since it overcomes the disadvantage of generating large amount of wavelet coefficients as of in continuous wavelet transform.

A.DISCRETE WAVELET TRANSFORM

The DWT analyzes the signal at different frequency bands with different resolutions by decomposing the signal into a coarse approximation and detail information. DWT employs two sets of functions, called scaling functions and wavelet functions, which are associated with low pass and highpass filters, respectively.

Half band lowpass filtering removes half of the frequencies, which can be interpreted as losing half of the information. Therefore, the resolution is halved after the filtering operation. However, the subsampling operation after filtering does not affect the resolution, since removing half of the spectral components from the signal makes half the number of samples redundant. Half the samples can be discarded without any loss of information. The lowpass filtering halves the resolution, but leaves the scale unchanged. The signal is then subsampled by 2 since half of the numbers of samples are redundant. This doubles the scale. This procedure can mathematically be expressed as

microphone was used. For the NAM speech, a NAM microphone was used. In the case of the NAM signal only the low frequency components are appeared and due to the body

$$y[n] = \sum_{k=-\infty}^{\infty} h[k].x[2n-k] \quad (1)$$

The decomposition of the signal into different frequency bands is simply obtained by successive highpass and lowpass filtering of the time domain signal. The original signal x[n] is first passed through a halfbandhighpass filter g[n] and a lowpass filter h[n]. This constitutes one level of decomposition and can mathematically be expressed as follows:

$$y_{high}[k] = x[n].g[2k-n] \quad (2)$$

$$y_{low}[k] = x[n].h[2k-n] \quad (3)$$

whereyhigh[k] and ylow[k] are the outputs of the highpass and lowpass filters, respectively, after subsampling by 2.

At every level of decomposition, the filtering and subsampling will result in half the number of samples (and hence half the time resolution) and half the frequency band spanned (and hence doubles the frequency resolution).

While computing the DWT, discard all values in the DWT coefficients that are less than a certain threshold and save only those DWT coefficients that are above the threshold for each frame.

The reconstruction in this case is very easy since halfband filters form orthonormal bases. The above procedure is followed in reverse order for the reconstruction. The signals at every level are upsampled by two, passed through the synthesis filters g'[n], and h'[n] (highpass and lowpass, respectively), and then added. The interesting point here is that the analysis and synthesis filters are identical to each other, except for a time reversal.

B.ALGORITHM

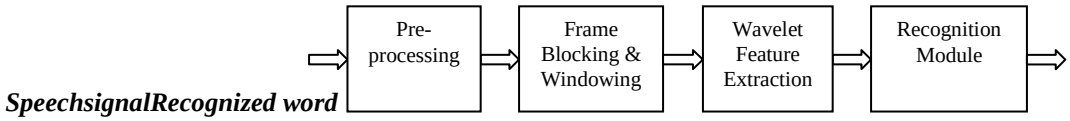
Thus the algorithm can be summarized as

- Framing input speech signal
- DWT of a frame
- Thresholding wavelet coefficients
- Inverse DWT

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 1,January 2013

III.BLOCK DIAGRAM



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/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 input speech signal is segmented into frames of 20~30 ms with optional overlap of 1/3~1/2 of the frame size. Usually the frame size (in terms of sample points) is equal to power of two in order to facilitate the use of FFT. If this is not the case, we need to do zero padding to the nearest length of power of two. 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 = 20 ms. Additional, if the overlap is 160 points, then the frame rate is 16000/(320-160) = 100 frames per second.

C.Hamming windowing

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 s(n)*w(n), where w(n) is the Hamming window defined by:

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \qquad (4)$$

D.Feature extraction

In speaker independent speech recognition, a premium is placed on extracting features that are somewhat invariant to changes in the speaker. So 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.

IV.SIMULATED RESULTS

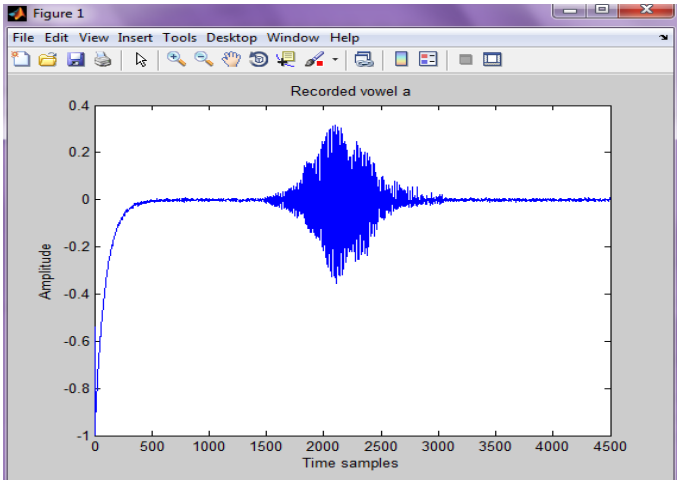


Figure 3: Recorded vowel ‘a’ using microphone

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 1,January 2013

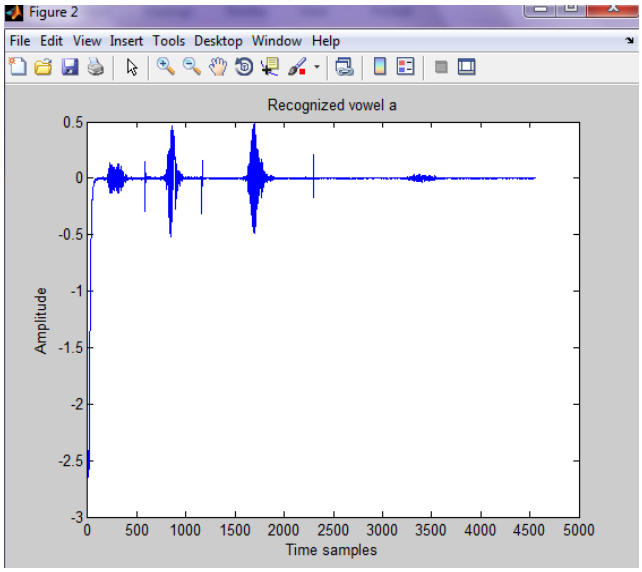


Figure 4: Recognized vowel 'a' using wavelet transform

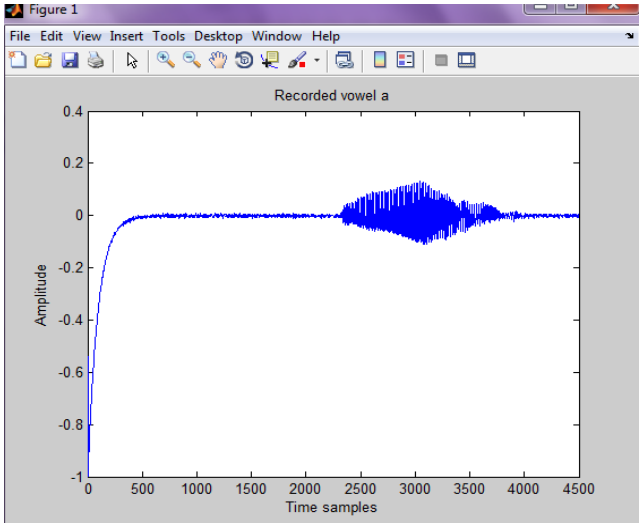


Figure 5: Recorded vowel 'a' using NAM microphone

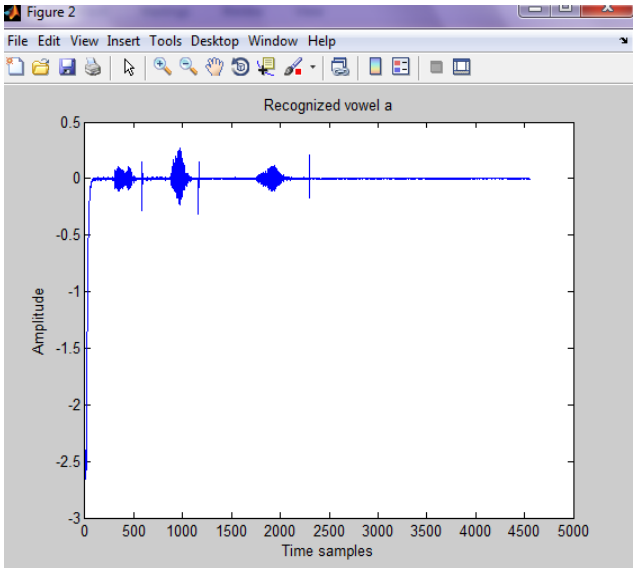


Figure 6: Recognized vowel 'a' using NAM microphone

V.CONCLUSION

In this paper normal speech and NAM speech has been analyzed using the wavelet transform. Recognition of speech using wavelet transform gives more accuracy than the speech when heard before analysis. Comparison of normal speech and NAM speech is in progress to determine the increase in accuracy.

REFERENCES

1. P. Heracleous, Y.Nakajima, A. Lee, H. Saruwatari, and K. Shikano, 'Non-Audible Murmur (NAM) recognition using a stethoscopic NAM microphone', in *Proc. Interspeech 2004-ICSLP*, 2004, pp. 1469–1472.
2. Y. Nakajima, H. Kashioka, K. Shikano, N. Campbell, .Non-Audible Murmur Recognition Input Interface UsingStethoscopic Microphone Attached to the Skin., *Proceedings of ICASSP*, pp. 708.711, 2003.
3. *PanikosHeracleous, Yoshitaka Nakajima, Akinobu Lee, Hiroshi Saruwatari, KiyohiroShikano,* 'Audible (normal) speech and inaudible murmur recognition using NAM microphone'.
4. C.H. Vithalani, "Speech analysis using wavelet transform", Proceedings of NCC-2003 P 524-528
5. *Subhradebdas, vaishalijagrit, chinmaychandrakar, m.f.queresh,,* 'Application of wavelet transform for speech processing', International Journal of Engineering Science and Technology (IJEST) Vol. 3 No. 8 August 2011.
6. V. A. Tran, G. Bailly, H. Loevenbruck, and C. Jutten, "Improvement to a NAM captured whisper-to-speech system," in *Proc. Interspeech 2008*,pp.1465–14.