

Feature Extraction from Non-Audible Murmur (NAM) for the Vocally Handicapped using Wavelet Transform

Sheena Christabel Pravin
Assistant Professor
Rajalakshmi Engineering
College

Samyuktha Sundar
Student
Rajalakshmi Engineering
College

Krithika Aravindan
Student
Rajalakshmi Engineering
College

ABSTRACT

Non audible murmur is a body conducted silent speech through which the vocally handicapped can communicate. We propose a method of acquisition of Non Audible Murmur (NAM), (i.e., inaudible speech produced without vibrations of the vocal folds) from the vocally handicapped using the MEMS accelerometer, followed by its de-noising and Statistical Feature Extraction. The murmur is acquired by placing the sensor bonded to the surface of the skin over the soft-cartilage bone behind the ear. The resulting electrical signal is de-noised using Discrete Wavelet Transform (DWT). Statistical Features are extracted from the detailed coefficients of the de-noised murmur.

General Terms

Speech signal processing, Wavelet Transform

Keywords

NAM, MEMS accelerometer, DWT, De-noising, Feature Extraction, Vibration sensor.

1. INTRODUCTION

We live in an age where each new day sees technology reaching new heights. But growth that empowers life only for a selected segment of people, while leaving other sections of the society untouched, does no good. Technology for the needy: this is the need of the hour. Speech processing for the vocally handicapped is an area with an enormous scope for development. The vocally handicapped are not able to produce well-defined speech; however, they do produce murmurs from the motion and interaction of the speech organs such as tongue, palate, lips, etc. *Silent speech interfaces* have recently been studied as a technology to enable speech communication to take place without the necessity of emitting an audible acoustic signal [1].

The body-conducted murmur is usually not audible to the human ear; in other words it is a Non-Audible Murmur (NAM). This forms the basis of our work. Some body-conductive microphones, such as throat microphone [2] and non-audible murmur (NAM) microphone, are also effectively used as silent speech interface [3].

A serious drawback of body-conducted speech is that severe degradation of speech quality is caused by essential mechanisms of body conduction. Consequently, naturalness and intelligibility of body-conducted speech are much poorer than those of natural voices. To use it in human-to-human speech communication, its speech quality improvements are essential [4]. To enhance body-conducted speech, it should be de-noised for feature extraction, which is carried out in this paper.

The statistical features are extracted from the de-noised signal. De-noising the murmur signals using the Discrete Wavelet Transform removes the residual noise artifacts. The

Wavelets were developed as an alternative to Fourier analysis of signals [5]. Real and complex wavelets are widely used for signal analysis, compression and de-noising. Some of the well-known wavelets are Mexican Hat, Haar wavelet, Daubechies wavelet, Biorthogonal wavelet. The translation of the wavelets focus on the lower frequency components where the power of the signal is concentrated. The scaled versions of the mother wavelet bring different frequency resolutions. Thus, we can zoom in a particular frequency range. The important property of the mother wavelet is that it has zero mean and it is normalized. The mother wavelet is limited in time domain (i.e.) it has values in a defined range and zero elsewhere.

In the previous work on murmur recognition, a new interface to recognize NAM is proposed in [2]. Microphone is attached to skin based on stethoscope principle. New acoustic models-Hidden Markov Model used Julius – Japanese dictation toolkit [6]. In [7], new speech support system using NAM recognition for the disable is proposed for cases with speech disorders due to removal of the pharynx. The system retrieves speech from body-conducted speech. A healthy subject is chosen to demonstrate the feasibility of making body-conducted speech clear using the sub-word unit transfer function. The estimation of sub-word unit sequence was carried out for speech recovery. In [8], Maximum *a Posteriori* (MAP) criterion based Hidden Markov model (HMM) is proposed to improve the performance of the Voice Activity Detection (VAD). In [9], the issue of degraded speech quality due to body-conducted recording is addressed by statistically converting body-conducted speech into normal speech.

In [10], a modified MAP criterion is proposed based on a two-state HMM to improve the performance of the Voice Activity Detection (VAD). A soft frame margin discriminative training with semi-definite programming optimization (SFME-SDP) is proposed in [11] for robust speaker model training when only limited amounts of training data are available. In [12], a spectral conversion method is proposed for Voice Conversion (VC). To combat the issue of excessive smoothing by statistical modeling, a conversion method based on maximum likelihood estimation of the spectral parameter is proposed.

This paper is organized as follows. Section II explains our method of acquisition of murmurs from the vocally handicapped and the hardware and software tools used in the process. Section III deals with de-noising the samples acquired, using Discrete Wavelet Transform, followed by feature extraction from the signal. Section IV details the results obtained and their analysis. Section V concludes the paper and takes a look at future scope of the paper.

2. DATA ACQUISITION

Murmur samples are taken from vocally handicapped students using a tri-axial accelerometer – ADXL335. It is a Micro Electro-Mechanical Systems (MEMS) sensor. The ADXL335

is a small, thin, low power, complete 3-axis accelerometer with signal conditioned voltage outputs. It has an on-board 3.3 V regulator as mentioned in [13]. The ADXL335 is available in a small, low profile, 4 mm × 4 mm × 1.45 mm, 16-lead, plastic lead frame chip scale package. The V_{CC} takes up to 5V in and regulates it to 3.3V with an output pin. The sensor is placed at the soft cartilage behind the ear lobe for maximum receptivity of the murmur conducted through the body (see Figure 1).

The accelerometer is powered by a voltage regulated 9V battery with a regulator that sends out a constant 5V input to ADXL335 sensor and is connected to the computer via audio jack. The murmurs for English consonants and vowels were recorded using the experimental setup shown in Figure 1.

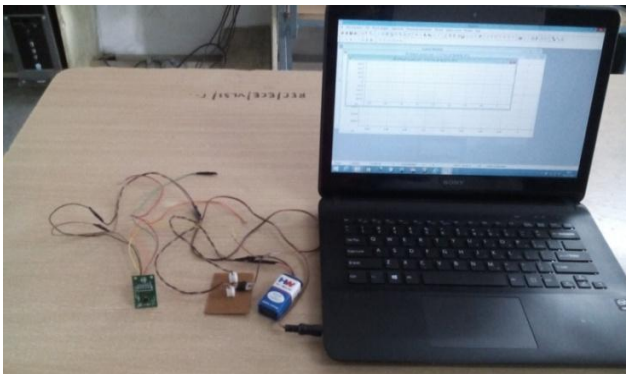


Figure 1. Experimental setup

The summary of the recorded murmurs of English Vowels and Consonants from 55 candidates are tabulated as follows:

Table 1. Summary of murmur recordings

Letters Recorded	Total minutes of recording	No. of samples taken
Vowels: a/e/i/o/u/	16.5	55
Consonants: b/s/f/	8.25	55

3. DE-NOISING THE RECORDED SIGNAL

The murmurs of the vowels and consonants were recorded using the MEMS sensor, sampled at a rate of 22050 samples per second. The recorded signals are de-noised using Wavelet transform. Fast Fourier Transform (FFT) gives only the frequency localization of the signals. While Short Time Fourier Transform (STFT) can be used to partially solve the problem of time localization, it has frequency resolution problems due to constant size window. Wavelet Transform, however, provides both time and frequency localization and is hence best suited for de-noising speech signals. The Continuous Wavelet Transform is both time and memory consuming [14], hence the Discrete Wavelet Transform is chosen. The general method for signal analysis is to decompose the signal to level N using a suitable wavelet. The detailed and approximate co-efficients can be obtained by scaling and dilating the Scaling function and the Wavelet function. The other way to find the wavelet coefficients is to introduce Subband Coding [15]. Here, the Haar wavelet. The Discrete Wavelet Transform is implemented using the filter banks to perform multi-resolution analysis. The sampled

signal is high-pass and low-pass filtered to get the detailed and approximate co-efficients respectively. These detailed co-efficients basically correspond to details in the data set. The approximate co-efficients obtained in stage I of the filter is fed as input to a stage II and so on till stage III (see Figure 2).

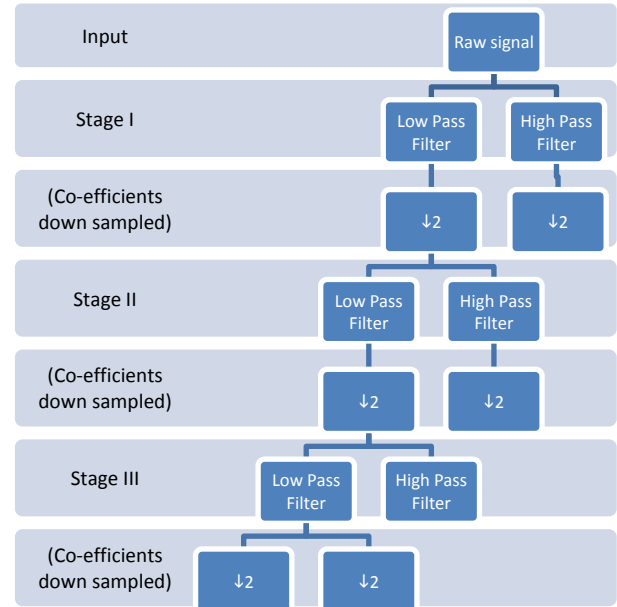


Figure 2. DWT Flow-graph

The Haar Wavelet Transform is defined as an orthogonal matrix. The 2-point single-scale DWT of the time-domain signal $\{x(0), x(1)\}$ and its inverse DWT (IDWT) are defined as:

$$\begin{bmatrix} X_\theta(0,0) \\ X_\psi(0,0) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} x(0) \\ x(1) \end{bmatrix} \text{-----(1)}$$

(or)

$$X = H_{2,0} x \text{-----(2)}$$

$$\text{and } x = H_{2,0}^{-1} X = H_{2,0}^T X \text{----(3)}$$

Here X , $H_{2,0}$, and, x represent, respectively, the coefficient, transform, and input matrices. In the DWT, the transformed domain is the time-scale domain and the transformed values have time and scale as independent variables. Scale indicates the frequency band of the corresponding signal component. A higher scale number (the detailed coefficients) corresponds to a higher frequency band [15].

4. RESULTS AND DISCUSSIONS

The Non Audible Murmurs for vowels and few consonants in the English Language were recorded at 22KHz sampling rate placing the MEMS accelerometer bonded to the surface of the skin over the soft-cartilage bone behind the ear using the and de-noised. The results of the same are as follows:

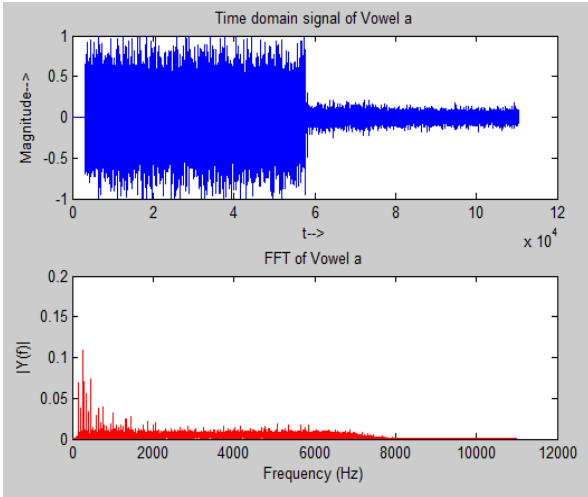


Figure 3. Time domain and Frequency domain representation of vowel 'a'

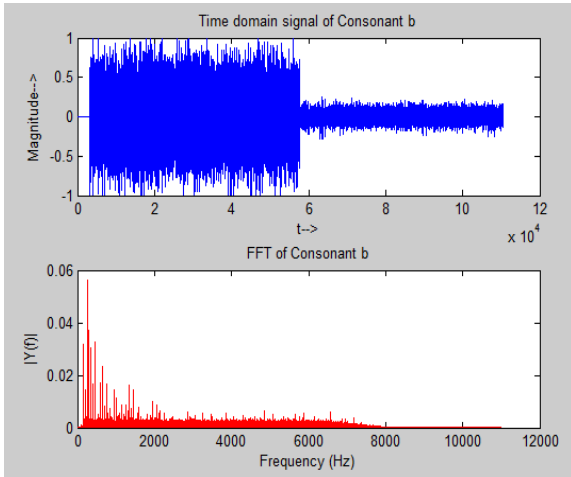


Figure 4. Time domain and Frequency domain representation of murmur 'b'

Figure 3 shows the raw signal for the vowel 'a' as recorded from the vocally handicapped using the MEMS sensor, along with its Fast Fourier Transform. Figure 4 shows the same for the consonant 'b'.

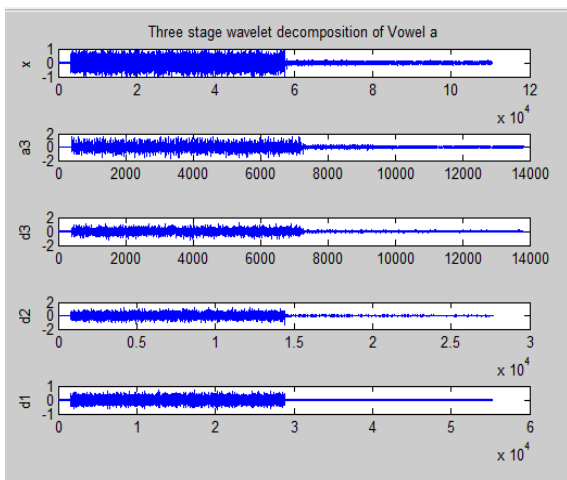


Figure 5. DWT decomposition for vowel 'a'

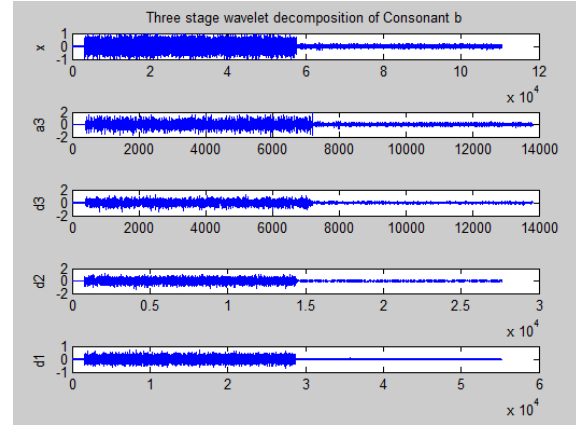


Figure 6. DWT decomposition for consonant 'b'

Figure 5 and Figure 6 show the DWT Decomposition of the murmur signals, including the detailed coefficients from stage I, II and III, and the approximate coefficient of stage III. Table 2 contains the features extracted from the coefficients of the signals obtained after decomposition. The statistical features extracted here are the maximum and minimum values, mean and standard deviation values of the coefficients of each input signals.

Table 2. Statistical Features extracted from the murmurs

Murmur signal	Coefficients	Max	Min	Mean	St. Dev.
		*10 ⁻⁷			
Vowel a	a3	0.024158	-0.00838	81522.54	0.006564
	d1	0.005805	-0.00587	1.05367	0.001364
	d2	0.008774	-0.00667	-466.406	0.0019
	d3	0.011307	-0.00959	303.036	0.003186
Consonant b	a3	0.027686	-0.00907	88291.33	0.007483
	d1	0.005524	-0.00527	-161.212	0.001522
	d2	0.007492	-0.00708	200.57	0.002098
	d3	0.010725	-0.01229	-474.152	0.003492

5. CONCLUSION

For many decades, the world of science and technology has aimed to achieve the impossible; to provide vision for the visually impaired, to provide limbs for the maimed and so on. One such miracle is to provide the ability to speak to the speech impaired. In this paper a novel method is proposed to acquire the Non Audible Murmur (NAM) from the vocally handicapped and the de-noising of murmur signals to extract the statistical features for distinguishing vowels and consonants. The DWT decomposition of Vowels and Consonants were compared and the features tabulated. The statistical features from the low frequency components of the murmurs carry more information about the signal.

6. FUTURE WORK

This paper has been the first step towards murmur-to-text conversion for the vocally handicapped. The statistical features extracted from the murmurs of the various English letters can be fed to suitable classifiers like the Artificial Neural Networks (ANN) for classification and interpretation of murmur by training, testing and validation processes. The letters will then be extended to words for more complete recognition. This paper is a fore-runner to a murmur recognition system for the vocally handicapped, in the English language.

7. ACKNOWLEDGMENTS

Our thanks to St. Louis Institute for the Deaf and the Blind for allowing us to take murmur samples from 55 vocally challenged students.

8. REFERENCES

- [1]. B. Denby, T. Schultz, K. Honda, T. Hueber, M. Gilbert, and S. Brumberg. Silent speech interfaces, *Speech Communication*, Vol. 52, No. 4, pp. 270-287, 2010.
- [2]. Y. Nakajima, H. Kashioka, N. Cambell, and K. Shikano. "Non-Audible Murmur (NAM) recognition", *IEICE Trans. Information and Systems*, Vol. E89-D, No. 1, pp. 1-8, 2006.
- [3]. Ashish Khare and Uma Shanker Tiwary, "Soft-thresholding for denoising of medical images - a multiresolution approach", *International Journal of Wavelets, Multiresolution and Information Processing*, 3(4):477-496, April 2005.
- [4]. Jim Lambers, 'Introductino to Wavelet Analysis'.
- [5]. Duraisamy Sundararajan, 'Fundamentals of the Discrete Haar Wavelet Transform'.
- [6]. Yoshitaka Nakajima, Hideki Kashioka, Kiyohiro Shikano. Nick Campbell, "Non-Audible Murmur Recognition Input Interface Using Stethoscopic Microphone Attached To the Skin", *Proceedings of ICASSP'03*, April 2003.
- [7]. Ishimitsu, Nakayama, "Construction of Speech Support System Using body-conducted Speech recognition for disorders", *Proceedings of ICICIC*, June 2008.
- [8]. Heracleous, et.al., "Accurate Hsidden Markov models for non-audible murmur (NAM) recognition based on iterative supervised adaptation", *Automatic Speech Recognition and Understanding, ASRU '03*, pp.73-76, Dec.2003.
- [9]. Tomoki Toda, 'Statistical Approches to Enhancement of Body-Conducted Speech Detected with Non-Audible Murmur Microphone', *Proceedings of 2012 ICME International Conference on Complex Medical Engineering July 1 - 4, Kobe, Japan, 2012*.
- [10]. Shiwen Deng, Jiqing Han, Tieran Zheng, Guibin Zheng, "A modified MAP criterion based on Hidden Markov Model for Voice Activity Detection", *Proceedings of ICASSP-2011*
- [11]. Yan Yin, Qi Li, "Soft frame margin estimation of gaussian mixture models for speaker Recognition with sparse training data", *Proceedings of ICASSP-2011*.
- [12]. Tomaki Toda, Alan W. Black, Keiichi Tokuda, "Voice Conversion Based on Maximum Likelihood Estimation of Spectral Parameter Trajectory", *Audio, Speech and Language Processing Journal*, Vol.15, Issue:8, pp.2222-2235, Oct.2007.
- [13]. <http://www.analog.com/en/products/mems/mems-accelerometers/adxl335.html#product-overview>
- [14]. Mohammed A. Salem, Nivin Ghamry, and Beate Meffert, "Daubechies Versus Biorthogonal Wavelets for Moving Object Detection in Traffic Monitoring Systems", 2009
- [15]. Ingrid Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992.