I J C T A, 9(28) 2016, pp. 157-164

RECOGNITION OF NAM SPEECH - INDIAN ENGLISH ALPHABETS USING SELF-DESIGNED NAM MICROPHONE WITH HMM-VITERBI

Mathavan Suresh Anand* Nagarajan Mohan Kumar and Angappan Kumaresan*****

Abstract: In this paper, a new approach is presented to recognize NAM speech. Non-audible murmur (NAM) is an unvoiced speech that can be received through the body tissue by a special microphone (NAM microphones) attached behind the speaker's ear. Body conducted speech is articulated sound &difficult to recognize. We propose this method to recognize body conducted speech with our self-designed NAM microphone to detect and translate it to text. As part our work, we proposed and implemented NAM speech recognition of Indian English alphabets to text. This NAM to Text conversion is implements and tested using HMM-Viterbi. We extract the data by NAM microphone and then it passes the data into a HMM-Viterbi classifier to recognize Indian English alphabets. We conducted a testing with real time audio samples and achieved average accuracy of 62.57%.

Keywords: NAM, microphone, HMM, viterbi, body conducted speech.

1. INTRODUCTION

Speaking is the basic of communication in day to day life. In recent times the speech communication is improved a lot with the recent technologies. Long time ago landline phone and now cellular phone is tremendously changed way of people to talk with each other whenever and wherever they want.

Even though mobile phones are made voice communication promising in several circumstances.But there are some times we face difficulties in communicating with hard speaking (dumb) people. Our work addresses this issue. We propose a method to recognize body conducted speech (NAM Speech) to Text. What is NAM Speech? Non-audible murmur (NAM) is an unvoiced speech that can be received through the body tissue by a special microphone (NAM microphones) attached behind the speaker's ear. As part our work, we proposed and implemented NAM speech recognition of Indian English alphabets to text.

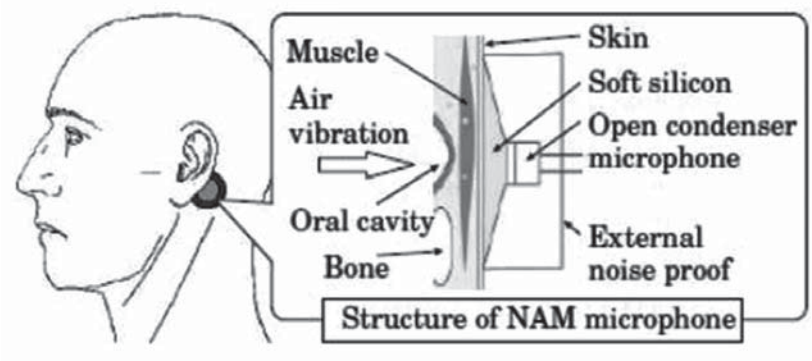


Figure 1. NAM microphone attached to the user

* Research Scholar, Anna University, Chennai.
Assistant Professor, Department of Computer Science and Engineering,Sri Sai Ram Engineering College, Anna University, Chennai.
** Professor, Department of Computer Science and Engineering,Sri Sai Ram Engineering College, Anna University, Chennai.
*** Assistant Professor, Department of Electronics and Communication Engineering,SKP Engineering College, Anna University,Tirvannamalai.

Non-Audible Murmur (NAM) microphone has been Originally developed by Nakajima.[1] We designed our own NAM microphone which is inspired by Nakajima's stethoscope based NAM microphone. The NAM microphone captures an inaudible murmur (NAM speech), which is not able heard by normal person. A tool for speech impaired people with NAM microphone used in a speech recognition system. Even though the NAM signal is of poor quality, the signal envelope is similar to that of normal speech, and therefore speech recognition is possible.

Figure 2. Self-Designed NAM microphone attached to the user

2. RELATED WORK

Speech recognition is the broad area to discuss. We are deal with this, so we should know about various happenings in the area of NAM speech recognition. Speech signal is recognized from the NAM speech by using wavelet transform in different decomposition level and the accurate speech signal is obtained by comparing the standard deviation values [2].

Normal speech and NAM speech has been analyzed using the wavelet transform and 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 [3].

Panikos Heracleous experimented for detecting NAM speech in Japanese vocabulary. The authors have previously conducted experiments for recognizing NAM speech using a stethoscopic and silicon microphone and they have attained a high level of accuracy for around 20000 Japanese vocabulary task. They have also done further research on NAM using hidden markov models (HMM) [4][11]. Chafic Mokbel has introduced a robust speech recognition system. The author's work includes a framework clustered for online adaptation of hidden markov models (HMM) for real-time scenarios. HMM parameters can be adapted to new conditions based on two approaches which are spectral transformation and Bayesian adaptation [5]. The author proposed a unified framework in this paper in which both of these approaches were used as particular cases. They have implemented the general adaptation algorithm within the speech recognition system and the whole system was evaluated and found to perform better.

Acoustic characteristics of body transmitted speech is sensitive to recording conditions location of NAM microphone, significant degradation of the conversion performance is often caused in practical situations by acoustic mismatches between training and conversion processes. To lessen this problem, Daisuke Miyamoto et al [6] proposed unsupervised acoustic compensation methods for body transmitted voice conversion.

NAM microphone used to detect various types of body-conducted speech such as extremely soft whisper, normal speech, and so on. Tomoki Toda [7] et al had developed two main technologies: one is body-conducted speech conversion for human-to-human speech communication. Second is body-conducted speech recognition for man-machine speech communication.

Analysis of NAM speech is made using distance measures between hidden Markov models (HMMs). It has been shown that owing to the reduced spectral space of NAM speech, the HMM distances are also reduced when compared with those of normal speech. [4] In a NAM microphone, body transmission and loss of lip radiation act as a low-pass filter. As a result, higher frequency components are attenuated in a NAM signal.

Tomoki Toda et al [8] proposed that Voice conversion (VC) methods from NAM to normal speech (NAM-to-Speech) and to a whispered voice (NAM-to-Whisper) are proposed, where the acoustic features of body-conducted unvoiced speech are converted into those of natural voices in a probabilistic manner using Gaussian mixture models (GMMs). A single conversion model capable of converting both NAM and BCW is effectively developed in our proposed VC methods.

Chen-Yu Yang et al [9] were conducted experiments on automatic recognition of continuous whispered speech. In training and test conditions are matched, the NAM microphone is found to be more robust against background noise than the close-talking microphone. In mismatched conditions (noisy data, models trained on clean speech), they found that Vector Taylor Series (VTS) compensation is particularly effective for the NAM signal.

Denis Babani et al proposed modified SAT methods capable of using a larger amount of normal speech data by transforming them into NAM data. The experimental results demonstrate that the proposed methods yield an absolute increase of approximately 2% in word accuracy compared with the conventional method. [10] When the training and test conditions are matched, the NAM microphone was found to be more robust against background noise than the close-talking microphone (word accuracy of 76.3% vs 66.5%).

3. NAM SPEECH RECOGNITION

Recognizing normal speech with the help of NAM microphone done by Panikos Heracleous et al [4] where separate GMMs (Gaussian mixture model) used for recognizing the normal and NAM speeches. They used separate HMMs (hidden markov model) for recognizing NAM and normal speeches individually. Denis Babani et al designed a special acoustic model [10] and NAM microphone which has speaker adaptive training (SAT) based on constrained maximum likelihood linear regression (CMLLR) in NAM acoustic model training.

In our work we propose the use of self-designed NAM microphone, which is inspired by stethoscope based like nakajima's. We are using Hidden Markov Model (HMM) for NAM speech recognition. Because it is ideal procedure is to train speaker-independent NAM HMMs.

HMM can used to find state sequence which maximizes probability of observation sequence. Each state is representation of one event and observations are made on these states.

$$O = O_1,O_2,O_3,O_4\ldots\ldots\ldots O_T \tag{1}$$

Let the probability of observation sequence is denoted in above equation (1).Then HMM model is represented as given in the equation (2) and equation (3). The state sequence for the model is given in the equation (4)

$$P(O|\lambda) = \Sigma P(O, \square| \lambda) \tag{2}$$

$$P(O|\lambda) = \Sigma a_{q0q1} b_{q1}(O_1) * a_{q1q2} b(O_2) \ldots\ldots.. a_{qT\text{-}1qT} b(O_T) \tag{3}$$

$$Q = q_0, q_1, q_2 \ldots\ldots q_T \tag{4}$$

For decoding the speech signal from source and trained data viterbi algorithm is used. Find the state sequence Q which maximizes P(O, Q | l).The probability by Viterbi is given at the following equations.

$$VP_t(i) = MAX_{q0\ldots\ldots.qt-1}\ P(O_1O_2\ \ O_t, q_t = i \mid \lambda) \tag{5}$$

$$VP_t(j) = MAX_{i=0\ldots\ldots.N}\ VPt-1(i)\ a_{ij} b_j(O_t) \quad \text{where } t > \tag{6}$$

$$P(O, Q \mid \lambda) = VP_T(S_N) \tag{7}$$

HMM viterbi determine from an observation sequence the most likely sequence of underlying hidden states that might have generated it. The best score along the a single path at time t, which accounts for the first t observations and ends in state S_i can be expressed in equation (8) and Initialization and recursive of Viterbi algorithms are given in equations (9) and (10) respectively.

$$\delta t(i) = \max P(q_1 q_2 \ldots q_t = i, O_1, O_2 \ldots O_t|\lambda) \tag{8}$$

The complete procedure for finding the best state sequences follows: (ψ is the variable that track the of the argument which maximized)

$$\begin{aligned} &\delta_1(i) = \pi_i b_i(O_1) \quad 1 \le i \le N \\ &\psi_1(i) = 0 \end{aligned} \tag{9}$$

$$\begin{aligned} &\delta_t(i) = \max\left[\delta_{t-1}(i)a_{ij}\right] b_j(O_t) \quad 2 \le t \le T \\ &\psi_t(i) = \arg\max\left[\delta_{t-1}(i)a_{ij}\right] \quad 1 \le i \le N \end{aligned} \tag{10}$$

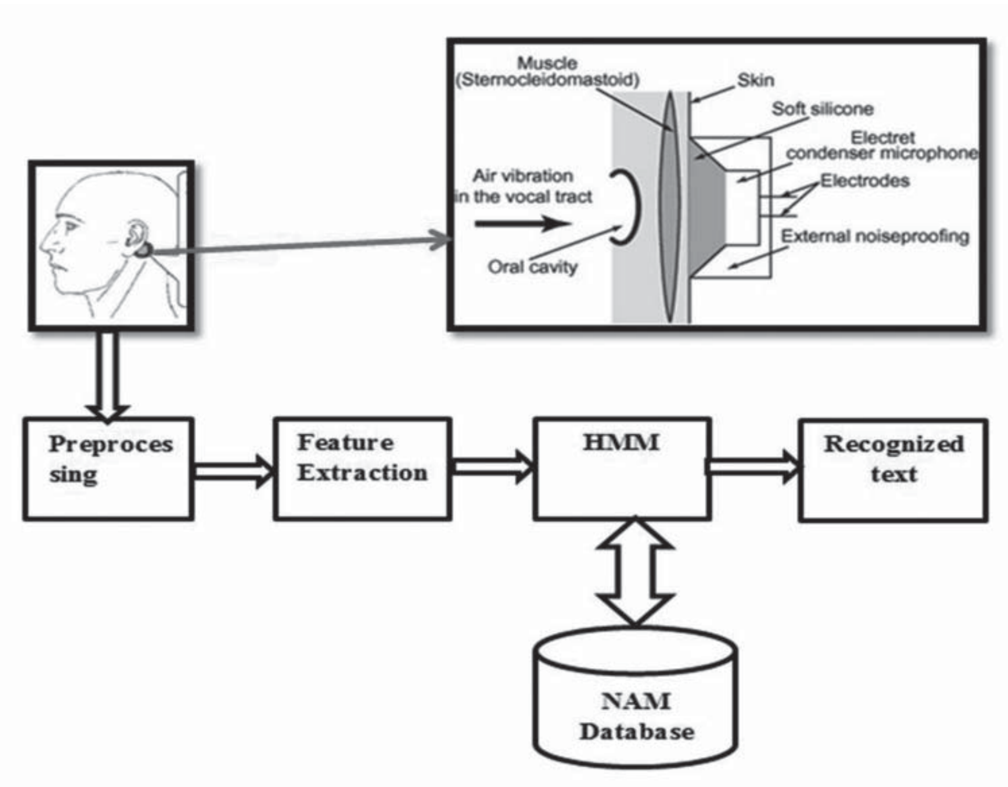
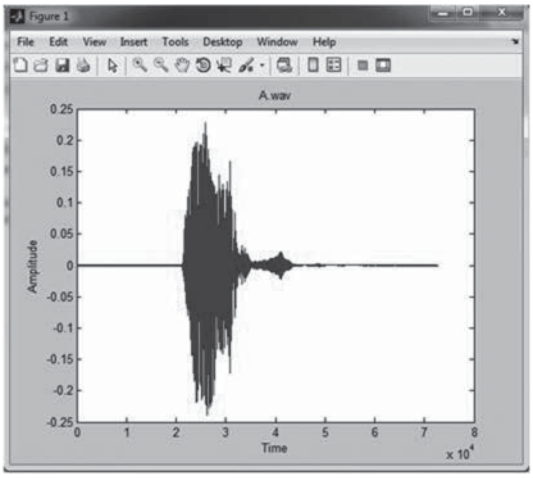


Figure 3. Architecture NAM speech recognition using NAM microphone

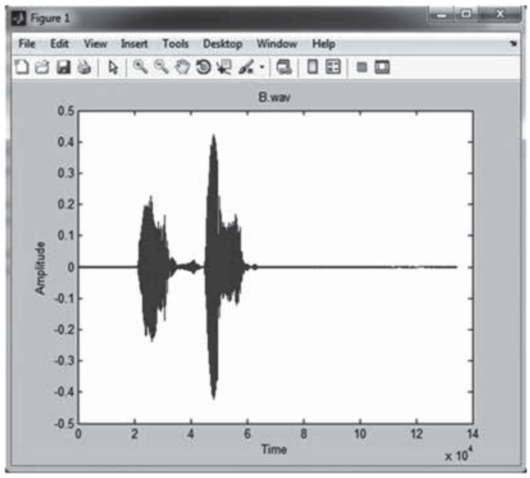
Pre-processing of Speech Signal is necessary to get better results.. They are Noise Removal, Endpoint Detection, Pre-emphasis, Framing, Windowing, Echo Cancelling etc. Noise removal filter is applied to remove unwanted signal s from speech signal. The body conducted speech is very soft and smooth, so noise removal is very much needed in recognition.

In NAM speech recognition feature extraction involves analysis of speech signal. Generally the feature extraction techniques are two types one is temporal analysis and second is spectral analysis. In our work we are using spectral analysis in which spectral representation of speech signal is used for analysis. Feature vectors are generated and given to train HMM model. Each alphabet is trained with HMM.NAM speech signals are trained and stored in as NAM speech database. HMM viterbi algorithm are used recognize NAM speech signal are recognized. Viterbi classifier gives the maximum probability of matching signal in speech recognition. In our work we implemented NAM speech recognition by using HMM-viterbi classifier.

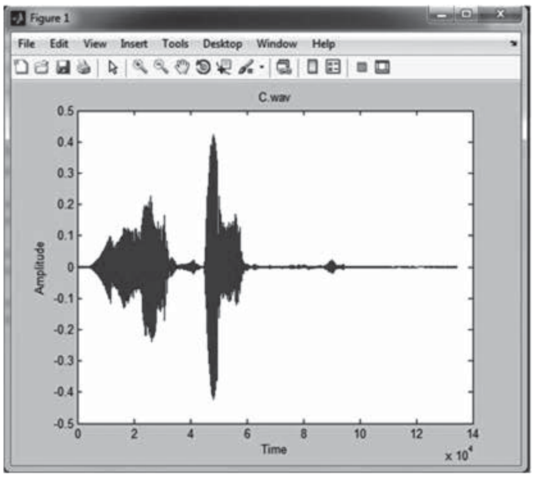
In the following Figure 4 samples of Indian English Audio Alphabets of NAM speech waveforms are shown. In our experiment we collected audio samples for all 26 alphabets from hundred individuals. In the Figure 5 HMM match random samples for above said are shown. HMM match uses Viterbi classifier for each audio-alphabet is tested with all possible random trained databases and gives maximum possible match. Depending upon sample size time taken to train and test differs. For 26 alphabets it took 66s for each alphabet audio sample. In the Table 1 the average recognition rate is given.



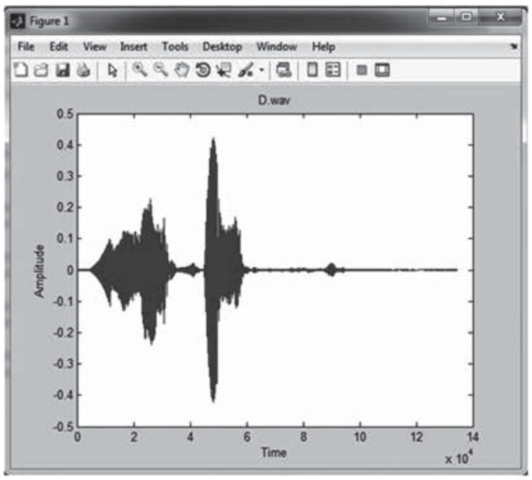
(a)



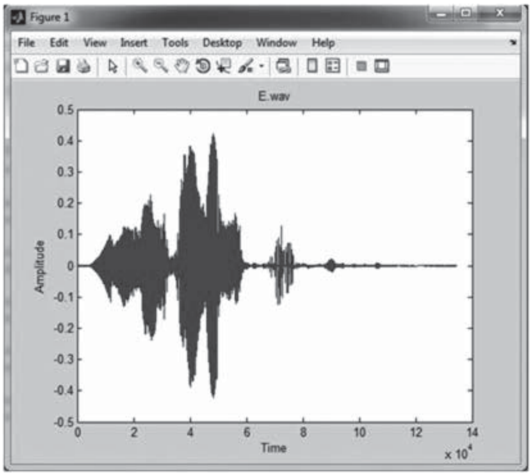
(b)



(c)

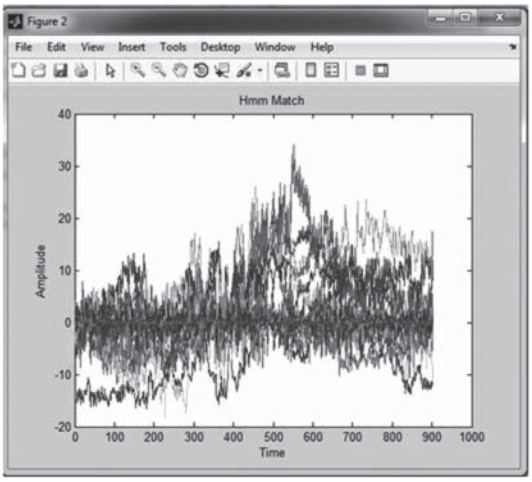


(d)

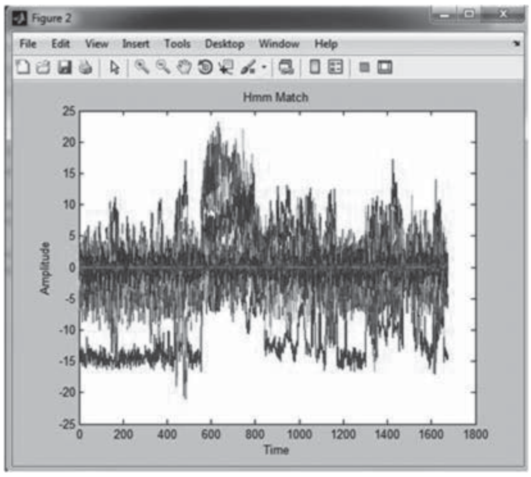


(e)

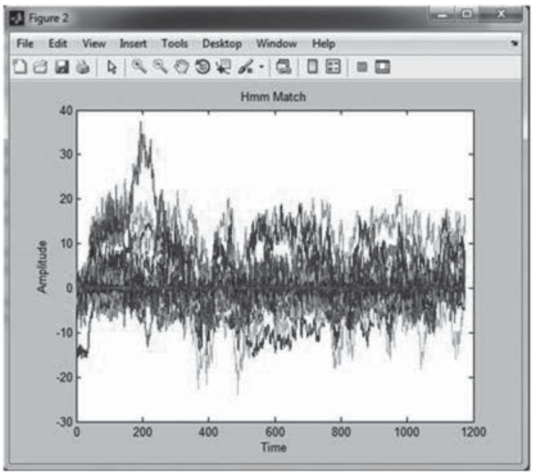
Figure 4. (a) (b) (c) (d) & (e) NAM speech waveform for alphabets ‘A’,’B’,’C’,’D’&’E’ respectively using NAM microphone



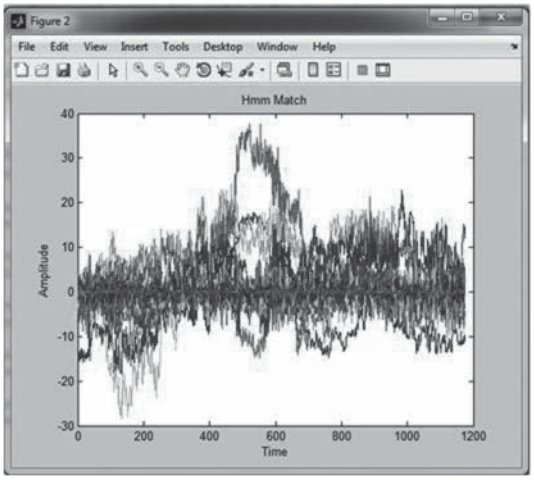
(a)



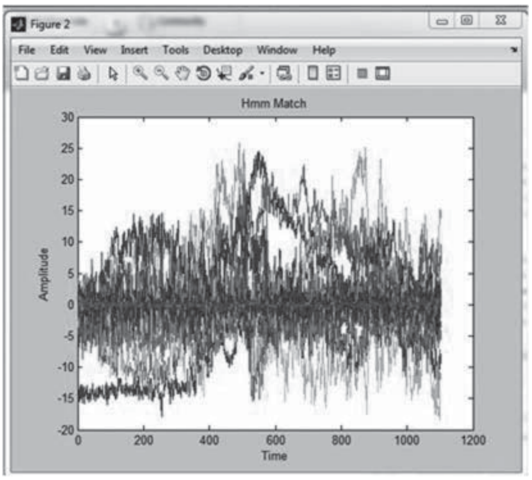
(b)



(c)



(d)



(e)

Figure 5. (a) (b) (c) (d) & (e) HMM match waveform for alphabets 'A','B','C','D'&'E' respectively

Table 1.
Recognition Rate of NAM speech samples

	NAM Speech Recognition Indian English Alphabets																										
Alphabet	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	Overall
Recognition Rate in %	62	67	61	59	64	61	58	59	60	63	62	61	60	63	59	58	66	64	65	68	64	67	64	63	62	67	62.576

The table 1.NAM speech recognition rate of each alphabet and overall result is shown. Each audio sample is trained and tested. The result of NAM speech is summarized in the above shown table. And the result of recognition is given in the graph in Figure 6.

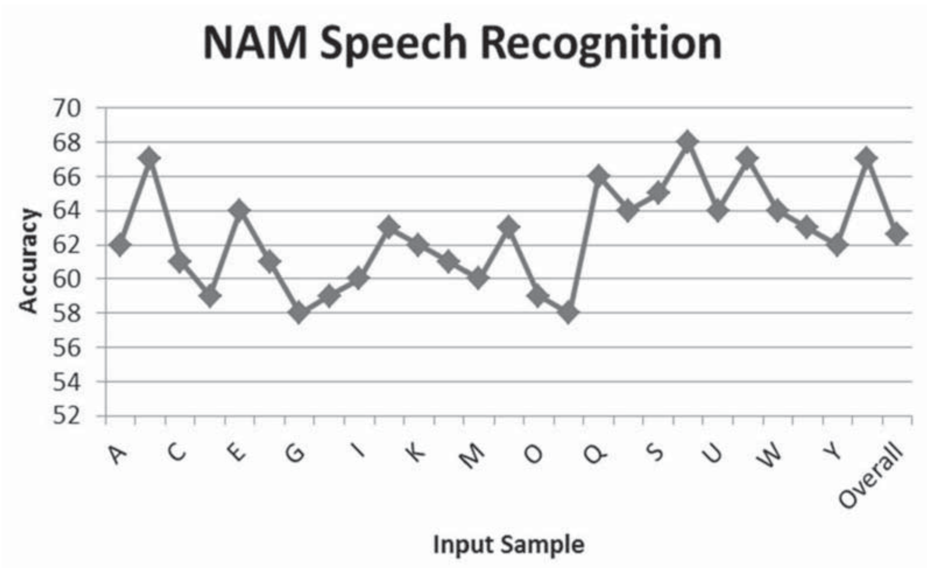


Figure 6. Performance graph for NAM speech of Indian English Alphabets

4. CONCLUSION

In our work we proposed a new approach to recognize NAM speech. Non-audible murmur (NAM) is an unvoiced speech that can be received by a special microphone (NAM microphones). Body conducted speech is articulated sound &difficult to recognize. To address this we proposed this method to recognize body conducted speech with our self-designed NAM microphone to detect and translate it to text. We collected audio samples and implemented NAM speech recognition of Indian English alphabets to text. This NAM to Text conversion is trained and tested using HMM-Viterbi. We extract the data by NAM microphone and then it passes the data into a HMM-Viterbi classifier to recognize Indian English alphabets. We conducted a testing data of hundred audio samples and achieved average accuracy of 62.57% recognition result.

5. ACKNOWLEDGEMENT

We would like to acknowledge Mr. yoshikazu nishiura , Mr. Nawaki along with Mr. Nakajima for providing us great support from NARA Institute of Science and Technology, Japan for replying our queries and guiding us in our research. I also like to acknowledge Mr. A.Kumaresan who helped me in self-design of NAM microphone from SKP Engineering College, India.

References

1. Punikos Herucleous, Yoshituka Nakajima, Akinobu Lee, Hiroshi Suruwuturi, Kiyohiro Shikuno,"ACCURATE HIDDEN MARKOV MODELS FOR NON- AUDIBLE MURMUR (NAM) RECOGNITION BASED ON ITERATIVE SUPERVISED ADAPTATION", 2003 IEEE.
2. K. Kalaiselvi, M. S. Vishnupriya,"Non - Audible Murmur (NAM) Voice Conversion by Wavelet Transform", International Journal of Science and Research (IJSR), Volume 3 Issue 3, March 2014.
3. RADHIKA.R, VANITHA LAKSHMI.M, AADITYA VELAVA.V.A,"COMPARISON AND ANALYSIS OF NAM AND ACOUSTIC SPEECH USINGWAVELET TRANSFORM", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 1, January 2013.
4. Panikos Heracleous, Viet-Anh Tran, Takayuki Nagai, And Kiyohiro Shikano, Fellow, IEEE,"Analysis and recognition of NAM speech using hmm distances and visual information" in IEEE transactions on audio, speech, and language processing, vol. 18, no. 6, august 2010
5. Chafic Mokbel,"Online adaptation of hmms to real-life conditions: a unified framework" in IEEE transactions on speech and audio processing, vol. 9, no. 4, may 2001.
6. Daisuke Miyamoto, Keigo Nakamura, Tomoki Toda, Hiroshi Saruwatari, and Kiyohiro Shikano," ACOUSTIC COMPENSATION METHODS FOR BODY TRANSMITTED SPEECH CONVERSION", 2009 IEEE.
7. Tomoki Toda, Keigo Nakamura, Takayuki Nagai,Tomomi Kaino, Yoshitaka Nakajima, Kiyohiro Shikano," Technologies for Processing Body-Conducted Speech Detected with Non-Audible Murmur Microphone", 2009 ISCA.
8. Tomoki Toda, Mikihiro Nakagiri, and Kiyohiro Shikano," Statistical Voice Conversion Techniques for Body-Conducted Unvoiced Speech Enhancement", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 20, NO. 9, NOVEMBER 2012.
9. Chen-Yu Yang, Georgina Brown, Liang Lu, Junichi Yamagishi, Simon King," NOISE-ROBUST WHISPERED SPEECH RECOGNITION USING A NON-AUDIBLE-MURMUR MICROPHONE WITH VTS COMPENSATION", 2012 IEEE.
10. Denis Babani, Tomoki Toda, Hiroshi Saruwatari, Kiyohiro Shikano," ACOUSTIC MODEL TRAINING FOR NON-AUDIBLE MURMUR RECOGNITION USING TRANSFORMED NORMAL SPEECH DATA", 2011 IEEE.
11. Panikos Heracleous," Using Teager Energy Cepstrum and HMM distances in Automatic Speech Recognition and Analysis of Unvoiced Speech", International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering Vol: 3, No:11, 2009.