ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

Word Level Translator for Tribal Language

Babitha T. B.[1], Riyas K.S.[2], Jayan A. R.[3]

PG Student [CSP], Dept. of ECE, Govt. Engineering College Wayanad, Mananthavady, Kerala, India[1]

Dept. of ECE, Rajivgandhi Institute of Technology, Govt. Engineering College, Kottayam, Kerala, India [2]

Dept. of ECE, Govt. Engineering College Sreekrishnapuram, Mannampatta, Palakkad, Kerala, India [3]

ABSTRACT: Speech is an efficient way to express the ideas, thoughts etc with each other. Speech processing have many applications in our day to day life and speech recognition is one among them. Speech to text conversion is one of the speech recognition method. Different Languages can be used as the input of many speech recognition methods. Tribal language is a folk language possessing no literary specifications. Majority of tribes in kerala state comes from the ’Paniya’ Tribal sect. The ’Paniyaan’ means ’worker’ as they were supposed to have been the workers of non tribes. Here we are taking ’Paniya’ Language as the input of the system. This paper presents a word level translator for ’Paniya’ Tribal language. Database contains the audio files from 5 male and 5 female tribal speakers. Here we have to extract the characteristic features from the recorded audio files and the mapping of Tribal words to Malayalam text is also done with the help of MATLAB. The classifier Artificial Neural Network (ANN) is used for pattern classification. Experiments shows that this method provides a better accuracy of 97.4 %.

KEYWORDS: Paniya Language, Speech Recognition, MFCC, ANN, MATLAB.

I.INTRODUCTION

Speech is the fundamental source of communication among the human beings. Human respiratory and articulatory systems includes different organs and muscles generate speech. Coordinated action of these speech production organs creates speech. There are many speech processing applications are there and speech recognition is one among them. Speech recognition means it is the process of converting an acoustic signal to a set of words, which are captured by a microphone or a telephone. Through this work, we present a conversion system, to convert the particular Tribal language. Speech to Text conversion take input as an audio file and then it is converted into text form which is displayed on desktop. Actually, speech processing is the study of speech signals and the various kinds of methods which are used to process them. Indian Tribal people plays a key role in constructing the cultural heritage of India. Tribes of Kerala are perhaps the most unique among all the south Indian Tribes. Wayanad is the home land for many Tribal communities like Paniyas, Adiyas, Kattunayakans, Kurichians, Kurumas, Ooralis, Uralikurumas etc[1]. Majority of Tribes in Kerala state comes from Paniya Tribal sect. ’Paniya’ is the language of Paniya tribal people[2]. So here we take the ’Paniya’ language as the input of the system. Formation of proper speech database is one of the basic requirement for developing a speech recognition system in any language. Here a new database has been created for Tribal language because there are no standard set of databases available in ’Paniya’ Tribal language. After making the database, a word level translator has been created under these modules. The different modules are
(1) Preprocessing of speech signals
(2) Extracting features from the signals
(3) Conversion of Tribal words to Malayalam text

II. METHODOLOGY

Overview of speech to text conversion system is illustrated in Fig.1. For translating a Tribal word to Malayalam text, there are mainly five steps. They are database creation, preprocessing, feature extraction, recognition and conversion. These steps are done with the help of MATLAB and ANN as classifier.

DOI:10.15662/IJAREEIE.2016.0510034

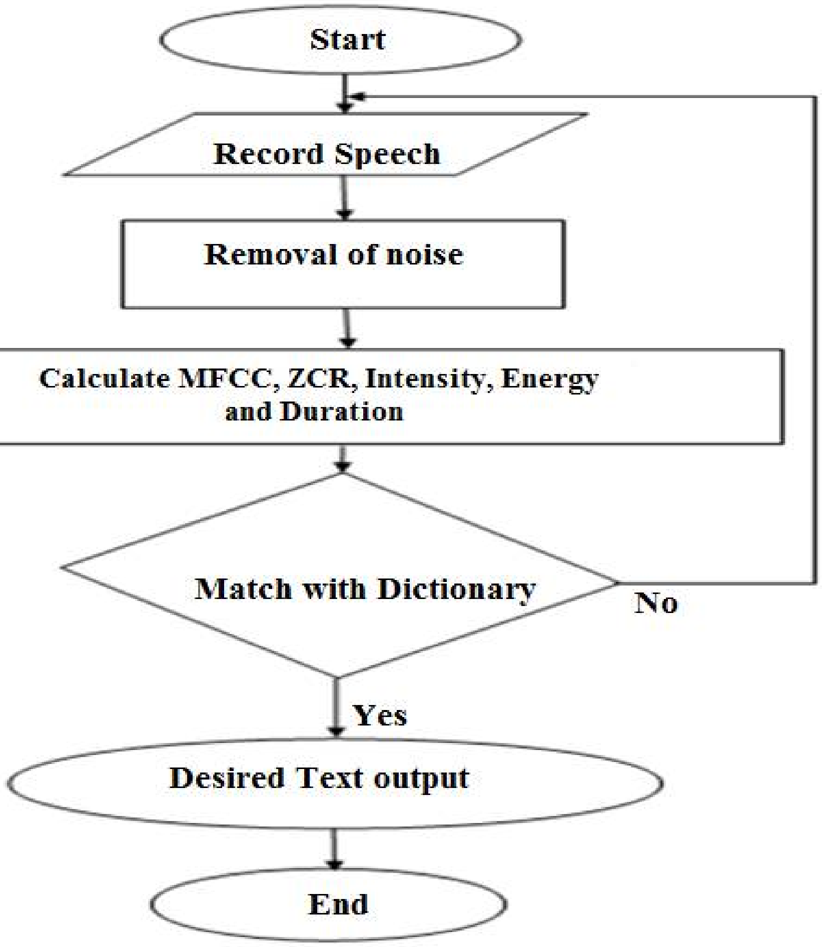


Fig. 1. Flowchart of Speech to text conversion

A. DATABASE CREATION

Creation of database is the initial stage for any speech recognition applications. As a part of this work, selected 10 Paniya words. These are the widely spoken words among them. They are collected from 10 Tribal speakers. 5 male and 5 female speakers were asked to speak the words five times, with a view to get most characteristic features for different words. Thus our database constitutes 500 samples. The samples are recorded by using Sound Forge Pro tool 10. The recordings are carried out in a studio in order to avoid the disturbances from surrounding environment.

B. FEATURE EXTRACTION

After the database creation we have to pre-process the signal for extracting the features. Windowing and Frame blocking are the major two processes in the preprocessing stage. In this work, the noise and silence part in each audio files are removed manually, by using the audio editing tools like Wave surfer, Praat, and Audacity. The selected features are

(1) MFCC (Mel Frequency Cepstral Coefficient)
(2) Energy
(3) Duration
(4) Intensity
(5) ZCR (Zero Crossing Rate)

1) MFCC (Mel Frequency Cepstral Coefficient): MFCC is an important feature, which is obtained using the MATLAB

tool. Figure below shows the schematic diagram of the MFCC extraction. In framing, the speech signals are modified as a number of frames that are overlapped and examined each frame independently.

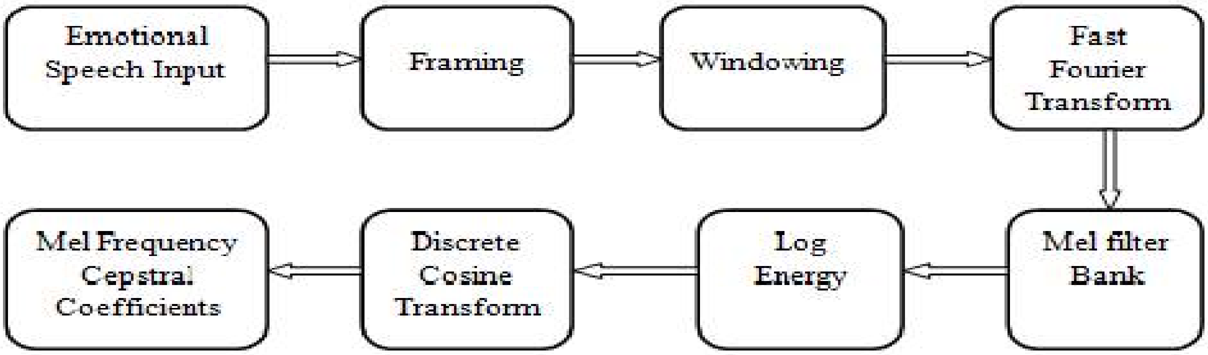


Fig. 2. Schematic diagram of the MFCC extraction

Feature extraction is based on partitioning speech into frames . There are 256 samples per frame and frame shift of 100 samples. The frame length of signal is 20 to 30 ms. Thereafter, each speech sample is multiplied by a window function as w(n). Here we are chosen the hamming window. For each frame, this window minimizes the discontinuities at the edges. This constitutes the windowing process. Windowing can be expressed as

$$yl(n) = xl(n)w(n);\ 0 \leq n \leq N - 1$$

Where N is the number of Samples in each frame
$xl(n) = l^{th}$ frame of speech
$yl(n) = l^{th}$ windowed frame
w(n) = Hamming Window
Hamming window is defined as below,

$$w(n) = 0.54 - 0.46\cos(2\pi n|N - 1),\ 0 \leq n \leq N\text{-}1$$

In frequency domain the calculations are more accurate. So here we are taking the FFT (Fast Fourier Transform) of each frame in order to convert the signal in time domain to frequency domain. Our speech signal follows non- linear scale and in FFT the frequency range is so wide. In mel frequency filtering, it is perceptual scale that helps to simulate the way human ear works[5]. It corresponds to better resolution at low frequencies and less at high. Mel scale is nonlinear scale which is inspired by this different perception rate of human ear. The Mel scale frequency F_{mel} be calculated from the linear frequency of input speech signal f_{lin}.

$$\text{Fmel} = 1125 \times log\left[1 + (\frac{flin}{700})\right]$$

Then calculate the logarithm of each filter output. This is for avoiding the phase information, making feature extraction is in-sensitive to speaker variations. The logarithm of filter bank energies results MFSC (Mel Frequency Spectral Coefficients). This MFSC's are highly correlated to each other. With few vectors we can represent the most of the information content in speech signal. For this we take DCT (Discrete Cosine Transform) and the results are known as MFCC.

2) Energy: The energy content in the speech is the most important feature useful to detect voiced and unvoiced regions. Energy computes the quantity of speech signal noted at a time. In case of a speech segment, the energy is higher as compared with the that of a non speech segment. For each frame, the energy is computed as

$$En = 1/N \sum_{m} s(m)^2$$

3) Intensity: Intensity is the average power transfer over one period of wave. It can be simply computed by using the audio editing tool called Praat.

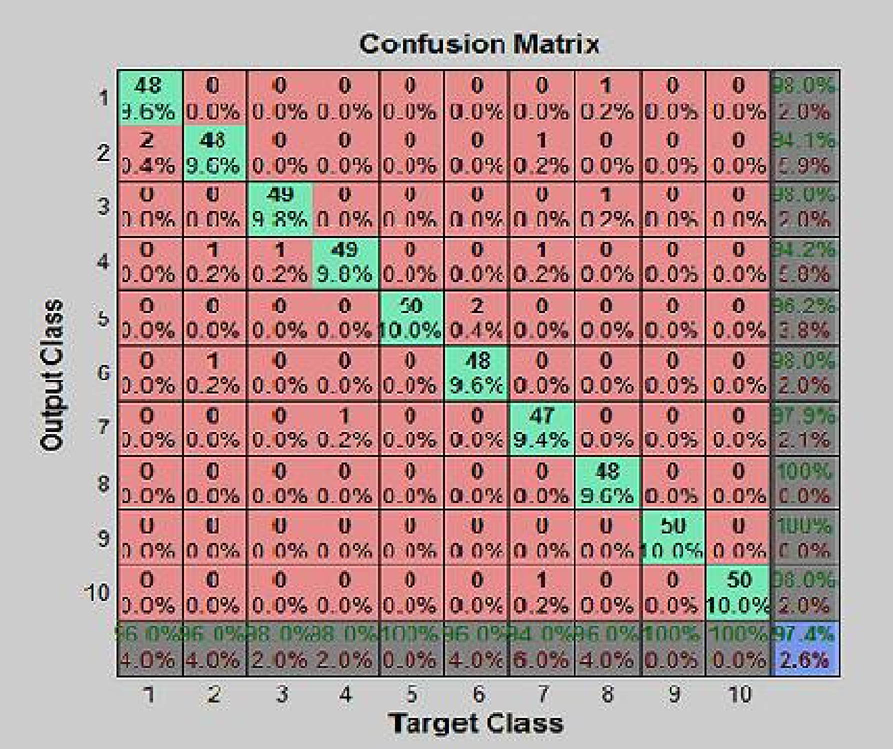
4) Duration: Duration is the length of spoken words in ms. It is important for differentiating various words in Tribal Language.

5) ZCR(Zero Crossing Rate): It is defined as the number of times at which the signal crosses the amplitude of zero. It is simply computed by implementing an algorithm with 'for' loop and 'if else if ' condition in MATLAB.

C. CLASSIFICATION

The relation between multiple features are evaluated by using a suitable classifier. In this work ANN (Artificial Neural Network) is used as a classifier[6]. For classifying different Tribal words from speech, the extracted features are given to ANN classifier. For creation, training, and simulation of the network, a toolbox called MATLAB. Neural Network is used. Network created with an input layer, one hidden layer and an output layer. Hidden layer have 10 neurons. Training and Testing are two important stages in classification. After network is created, it can be trained for recognizing words. From the total speech samples 75% is used for training the network. Type of training used is supervised training in which user has provided desired output for each input pattern. The results obtained during each time of training of neural network are completely different because of different initial conditions. To obtain good accuracy, retrain the ANN classifier several times until better performance obtained.

Trained neural network can now be evaluated with the testing samples. 25% of database is used for testing the network. Result analysis can be done from confusion matrix. A confusion matrix represents the performance of classifier in a matrix form. For a recognition having m classes, then the confusion matrix having a size of m by m. The elements along the diagonal of the confusion matrix indicates the correctly classified classes. For a good classifier, the diagonal elements of confusion matrix should be large, and the other values are smaller or a value of zero. Figure 3 shows the confusion matrix of classifier output.



Confusion Matrix

Output Class
Target Class

	1	2	3	4	5	6	7	8	9	10	
1	48 9.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	98.0% 2.0%
2	2 0.4%	48 9.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	94.1% 5.9%
3	0 0.0%	0 0.0%	49 9.8%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	98.0% 2.0%
4	0 0.0%	1 0.2%	1 0.2%	49 9.8%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	94.2% 5.8%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	50 10.0%	2 0.4%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	96.2% 3.8%
6	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	48 9.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	98.0% 2.0%
7	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	47 9.4%	0 0.0%	0 0.0%	0 0.0%	97.9% 2.1%
8	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	48 9.6%	0 0.0%	0 0.0%	100% 0.0%
9	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	50 10.0%	0 0.0%	100% 0.0%
10	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	50 10.0%	98.0% 2.0%
	96.0% 4.0%	96.0% 4.0%	98.0% 2.0%	98.0% 2.0%	100% 0.0%	96.0% 4.0%	94.0% 6.0%	96.0% 4.0%	100% 0.0%	100% 0.0%	97.4% 2.6%

Fig. 3. Confusion Matrix

D. CONVERSION

For converting the Tribal words to Malayalam word as in text, we have to create a dictionary. The dictionary contains the feature values of various Tribal words that they usually spoken. Values in the dictionary are in array format and each array is coded to the Malayalam words. Then for each test audio file we are calculating the feature values and compare this with the values in the closed set of dictionary. If a match occurs the Malayalam word corresponding to that Tribal word is displayed as text.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

III. SIMULATION RESULTS

The performance of proposed method is measured using the well known parameters like accuracy, sensitivity, precision, specificity and F-measure. They provided a measure of the performance of our classifier. Obtained results are given in Table 1. The proposed method obtained an overall accuracy of 97.4 %. The figure 4 shows the converted output of Tribal word 'kiyaanku' in Malayalam as 'Kizhangu'.

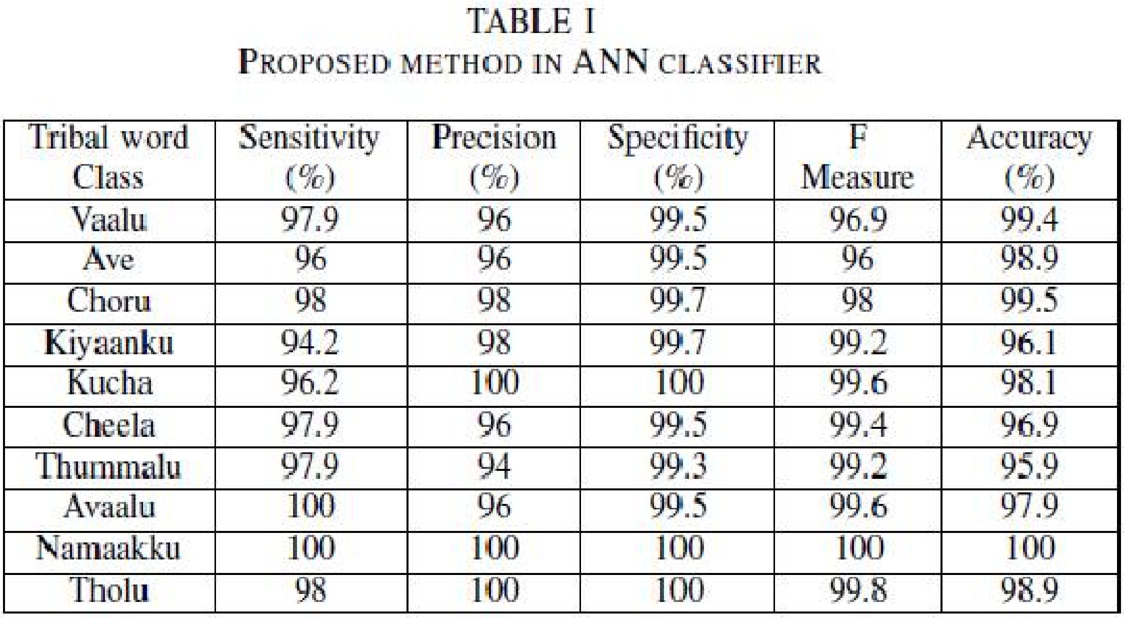
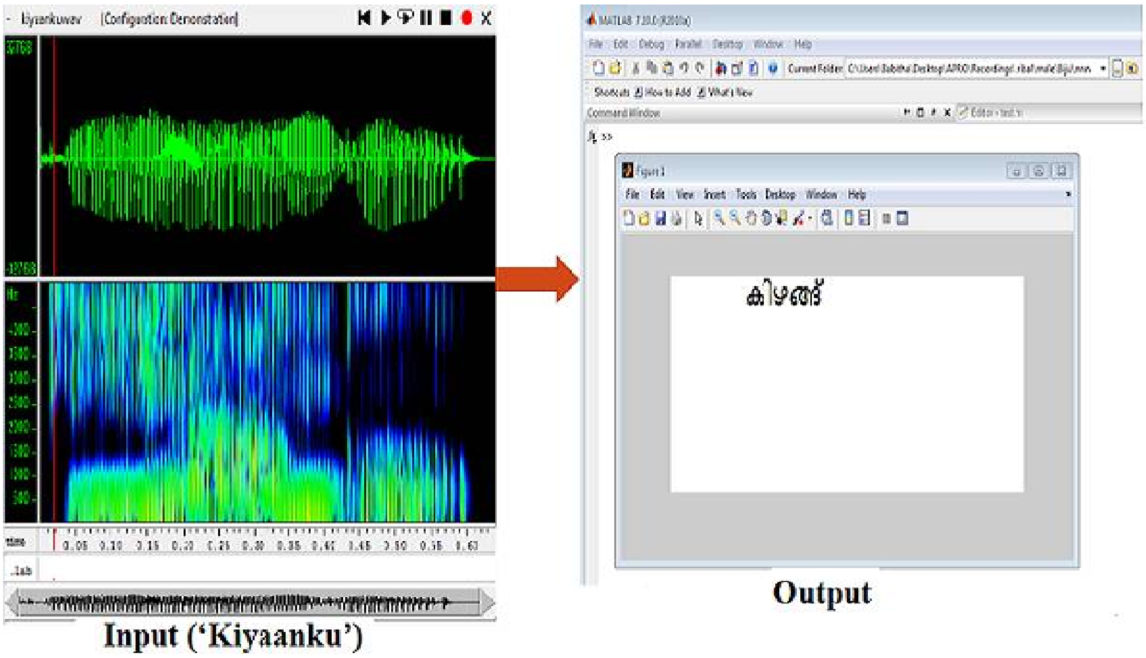


TABLE I
PROPOSED METHOD IN ANN CLASSIFIER

Tribal word Class	Sensitivity (%)	Precision (%)	Specificity (%)	F Measure	Accuracy (%)
Vaalu	97.9	96	99.5	96.9	99.4
Ave	96	96	99.5	96	98.9
Choru	98	98	99.7	98	99.5
Kiyaanku	94.2	98	99.7	99.2	96.1
Kucha	96.2	100	100	99.6	98.1
Cheela	97.9	96	99.5	99.4	96.9
Thummalu	97.9	94	99.3	99.2	95.9
Avaalu	100	96	99.5	99.6	97.9
Namaakku	100	100	100	100	100
Tholu	98	100	100	99.8	98.9



കിഴങ്ങ്

Fig. 4. Conversion Result

IV.CONCLUSION

In this work, the system could achieved a satisfactory performance. Since there is no standard datababase for the Tribal language (Paniya), developed speaker independent Tribal word database with total of 500 speech samples and experiments are performed there in. The proposed method focus only on ten mostly spoken Tribal words. This method uses MFCC, Energy, Intensity, ZCR and Duration for feature extraction. The classification and conversion has been performed using MATLAB. The experiments on our dataset shows that this method is a better method in translation of Tribal words with an accuracy of 97.4%.

REFERENCES

[1] Narayanan Nair (A. R), A textbook of ”Kattarum Avarude Kalamozhikalum”, Connemara Publication.

[2] A textbook about Tribal culture and language, ”Nayam”, an year long collective action research done in the TTC class in DIET Wayanad.

[3] Shaheena Sultana , M. A. H. Akhand and Prodip Kumer Das, ”Bangla Speech-to-Text Conversion using SAPI”, International Conference on Computer and Communication Engineering (ICCCE 2012), 3-5 July 2012, Kuala Lumpur, Malaysia

[4] Kapang Legoh , Utpal Bhattacharjee, T. Tuithung ”Development of Multi- Variability Speech Corpus of Adi Language for Speech Recognition Researches”, International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 3, Issue 10, October 2013.

[5] Ling Cen, Wee Ser and Zhu Liang Yu, ”An Approach to Extract Feature using MFCC”, IOSR Journal of Engineering (IOSRJEN), ISSN (e): 2250- 3021, ISSN (p): 2278-8719 Vol. 04, Issue 08 (August. 2014), pp. 21-25.

[6] Sonia Sunny, David Peter S and K. Poulose Jacob, ”Performance of different classifiers in speech recognition”, ISSN: 2319 - 1163,Volume: 2, Issue: 4, pp. 590 - 597.

[7] https://en.wikipedia.org/wiki/Sound Forge.

[8] https://en.wikipedia.org/wiki/WaveSurfer.

[9] Amos Gilat, MATLAB: An Introduction with Applications, 3rd ed. India: Wiley India, 2007.

[10] L. R. Rabiner, and B. H. Juang, Fundamentals of Speech Recognition, New Jersey, USA: Engle-wood Cliffs Publisher, 1993.

[11] Md. Abul Hasnat, Jabir Mowla, Mumit Khan, ”Isolated and Continuous Bangla Speech Recognition: Implementation, Performance and application perspective”.

[12] Thin Thin Nwe and Theingi Myint, ”Myanmar Language Speech Recognition with Hybrid Artificial Neural Network and Hidden Markov Model”, Proceedings of 2015 International Conference on Future Computational Technologies (ICFCT’2015), ISBN 978-93-84468-20-0, Singapore, March 29-30, 2015, pp. 116-122.