International Journal of Science and Engineering Applications
Volume 7–Issue 08,175-176, 2018, ISSN:-2319–7560

Speech to Text Conversion System for Myanmar Alphabet

Zaw Win Aung
Technological University (Loikaw)
Loikaw, Myanmar

Abstract: This paper is aimed to implement the speech to text conversion system for Myanmar alphabet. The Myanmar alphabet consists of 33 characters from 'ka' to 'ah'. The proposed system is software architecture which allows the user to speak against the computer in Myanmar language and the corresponding character is printed on the screen in the Microsoft Office Word Document Format. The system is emphasized on Speaker Independent Isolated Word Recognition System. The proposed system directly acquires and converts speech to text. This system contains two main modules: feature extraction and feature matching. Mel Frequency Cepstrum Coefficients (MFCC) is applied for feature extraction which extracts a small amount of data from the voice signal that can later be used to represent each character. Feature matching involves the actual procedure to identify the unknown character by comparing extracted features from the voice inputs of a set of known characters. In this system, Vector Quantization (VQ) approach using Linde, Buzo and Gray (LBG) clustering algorithm, which reduces the amount of data and complexity, is applied for feature matching. To implement this system MATLAB programming language is used.

Keywords: Speech to text; Myanmar alphabet; isolated word recognition; Myanmar character; Myanmar language

1. INTRODUCTION

There is a widespread need for transcription services converting audio files into written text for various purposes: meeting minutes, court reports, medical records, interviews, videos, speeches, and so on. Written text is easier to analyze and store than audio files, and apart from this, there are many circumstances one could imagine for needing to transcribe human speech: those who are deaf still need to listen to certain audio files; people with limited ability to type, such as those who are paralyzed or suffer from Carpal Tunnel Syndrome, still need to draft documents; and so on. Speech-to-Text (STT) system is a system for conversion of speech into text. It takes speech as input and divides it into small segments. These small segments are sounds, known as monophones. It extracts the feature vectors of the monophones and matches them with stored feature vectors and most likely or higher matched character is returned to the editor for printing.

A System-on-Programmable-Chip (SOPC) based Speech-to-Text architecture has been proposed by Murugan and Balaji[1]. This speech-to-text system uses isolated word recognition with a vocabulary of ten words (digits 0 to 9) and statistical modeling (HMM) for machine speech recognition. They used Matlab tool for recording speech in this process. The training steps have been performed using PC-based C programs. The resulting HMM models are loaded onto a Field programmable gate array (FPGA) for the recognition phase. The uttered word is recognized based on maximum likelihood estimation.

An architecture for Hindi Speech Recognition System using Hidden Markov Model Toolkit (HTK) has been proposed by Kumar and Aggarwal[2]. The proposed system was built as a speech recognition system for Hindi language. Hidden Markov Model Toolkit has been used to develop the system. The proposed architecture has four phases, namely, preprocessing, feature extraction, model generation and pattern classification. The system recognizes the isolated words using acoustic word model. The system was trained for 30 Hindi words. Training data was collected from eight speakers. The developer reported the accuracy of 94.63%.

Phonetic Speech Analysis for Speech to Text Conversion has been given by Bapat, and Nagalkar[3]. Their work aimed in generating phonetic codes of the uttered speech in training-less, human independent manner. The proposed system has four phases, namely, end point detection, segmenting speech into phonemes, phoneme class identification and phoneme variant identification in the class identified. The proposed system uses differentiation, zero-crossing calculation and FFT operations.

2. IMPLEMENTATION

The proposed speech to text conversion system is simulated in MATLAB with speech signal as input and produces the corresponding text as output. The database consists of 165 speech samples which were collected from the same speaker. Each speech sample is about 1 second long. The speaker is asked to utter Myanmar character from 'ka' to 'ah' five times in a training session and one time in a testing session later on. The same microphone is used for all recordings. Speech signals are sampled at 8000 Hz.

In the training phase, feature vectors are calculated from the input speech signal by MFCC feature extraction algorithm. Finally, the codebook or reference model for each speech signal is constructed from the MFCC feature vectors using LBG clustering algorithm and store it in the database. In the identification phase, the input speech signal is compared with the stored reference models in the database and the distance between them is calculated using Euclidean distance. And then, the system outputs the speech ID which has minimum distance as identification result and the corresponding character is printed on the screen in the Microsoft Office Word Document Format. Figure 1 and Figure 2 show the training and testing phases of speech to text conversion system.

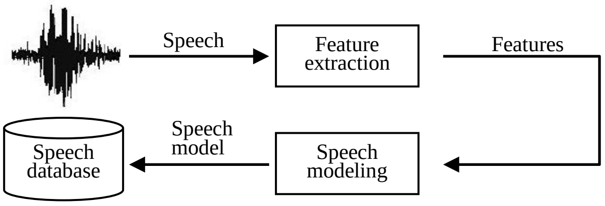

Figure 1. Training phase of speech to text system

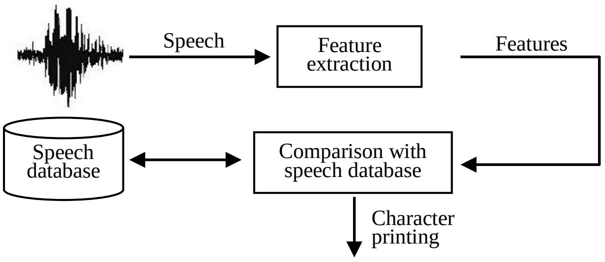

Figure 2. Testing phase of speech to text system

3. EXPERIMENTAL RESULT

This section describes the results of experiments carried out in different database sizes. In order to show the effectiveness of the proposed system, the computation time as well as accuracy of the system is computed. The training times taken by the system are shown in Table 1. The computation times and accuracy of the system are shown in Table 2.

Table 1. Computation time taken by the system in training phase

No	No: of Trained Samples	Time taken (seconds)
1	33 sample	1.67
2	66 samples	3.16
3	99 samples	4.66
4	132 samples	6.14
5	165 samples	7.67

Table 2. Computation time and accuracy of the system in testing phase

No	No: of Test Samples	No: of Samples in the Database	Time taken (seconds)	Accuracy (percent)
1	33	33	0.73	91%
2	33	66	1.89	97%
3	33	99	2.78	100%
4	33	132	5.52	100%
5	33	165	6.67	100%

4. RESULT ANALYSIS

In the training phase, including feature extraction and codebook construction, total length of training time is about 7.67 seconds for all 165 speech samples. The system is also tested with 33, 66, 99 and 132 speech samples in the database and it takes 1.67 seconds, 3.16 seconds, 4.66 seconds and 6.14 seconds respectively for training.

In the testing phase, when the system is tested by 165 speech samples in the database, the computation time taken by the system is 6.67 seconds for testing all 33 characters. On the other hand, the accuracy of the system is exactly 100 percent. In the experiments of testing by 33, 66, 99 and 132 speech samples in the database, it is found that the computation times is 0.73 seconds, 1.89 seconds, 2.78 seconds and 5.52 seconds respectively for testing all 33 characters. In the case of accuracy, the system achieves 91 percent, 97 percent, 100 percent and 100 percent respectively.

According to the experiments, it was found that most of the errors occurred among 'Ka Gyi', 'Gha Gyi', 'Na Gyi' and 'La Gyi' because these characters produce quite similar sound in Myanmar Language. The error also occurred between 'Ah' and 'Ha'. When the accuracy is taken into account, the larger the size of the database is, the higher the accuracy of the system is.

5. CONCLUSION

From this work it can be concluded that the system is reliable to use in real world applications and it is reasonably fast for working in real-time.

6. ACKNOWLEDMENTS

The author would like to take the opportunity to thank all of his colleagues who have given him support and encouragement during the period of the research. The author also would like to express his indebtedness and deep gratitude to his beloved parents, wife and son for their kindness, support and understanding during the whole course of this work and encouragement to attain his ambition without any trouble.

7. REFERENCES

[1] Bala Murugan M.T, Balaji .M, “SOPC-Based Speech-to-Text Conversion”, Nios II Embedded Processor Design Contest—Outstanding Designs 2006, Second Prize, National Institute of Technology, Trichy, 2006.

[2] Kumar Kuldeep and Aggarwal R.K., “Hindi Speech Recognition System using HTK”, J. of International Journal of Computing and Business Research, vol. 2, pp. 3-7, 2011.

[3] Bapat Abhijit V., Nagalkar Lalit K., “Phonetic Speech Analysis for Speech to Text Conversion”, in IEEE Region 10 Colloquium and the Third International Conference on Industrial and Information Systems, Kharagpur, India, 2008, pp. 1-4.