Volume 3, Issue 11, November 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Isolated BODO Spoken Word Identification using Mel-frequency Cepstral Coefficients and K-means clustering

Akalpita Das*, Purnendu Acharjee, Laba Thakuria, Pran Hari Talukdar

Instrumentation and USIC

Gauhati University, India

Abstract— This paper tries to propose an approach to recognize isolated Bodo spoken words. This paper deals with a speech feature extraction using MFCC and along with K-mean clustering. We used 10 words from Bodo for experiment purpose and 10 speakers (both male and female) are used to identify these words. We have used The words (stored in a database of GU_BODO) and later used in our experiment. MFCCs are calculated for different words. We have chosen those word which are having minimum quantization distance, which is calculated between features of each word in training phase and the individual word in testing phase. In real time this system is proposed, speaker dependent and independent word recognition systems for limited number of words.

Keywords: Bodo, GU_BODO, K-means clustering, MFCC, Speech Recognition.

I. INTRODUCTION

Assam is a land of multicultural and multi-ethnic diversity. People of Assam speak various different languages other than Assamese in different parts of it. Among those Bodo is one of the prominent languages spoken by the people of Bodo community [5]. The Bodos belong to a larger group of ethnicity called the Bodo-Kachari. Racially, they belong to a Mongoloid stock of the Indo-Mongoloids or Indo-Tibetans. Mythologically, according to Dr. Suniti Kumar Chatterjee, a well-known historian, the Bodos are "the Offspring of son of Vishnu and mother Earth", who are termed as Kiratas during the epic period. They are recognized as a plain tribe in the sixth schedule of the constitution of India.

Before 1953, the Bodo language had no standard form of writing. It had a history of using Deodhai, Roman and Assamese scripts. At present, Bodos adopted the Devanagari script. But, there is a huge difference in the usage of the letters in Bodo language from the Devanagari script. Bodo language shares some common salient features with other languages belonging to the Bodo group. These features are similar in terms of phonology, morphology, syntax, and vocabulary. Bodo language is closely associated with the Dimasa language of the state of Assam and with the Garo language of the state of Meghalaya, and also with Kokborok language of Tripura. It important to note that ,among the four districts of present Bodo land , namely, Kokrajhar, Chirang, Baksa and Udalguri, the language is heard in pure form only in the district of Udalguri. The language is affected by other communities , mostly, Assamese, Bengali and Hindi speaking communities. Speech is the most basic, common and efficient form of communication method for people to interact with each other. People are comfortable with speech therefore persons would also like to interact with computers via speech, rather than using primitive interfaces such as keyboards and pointing devices . Speech Recognition is the computing task of validating a voice identity by using characteristics extracted from the voice. This can be accomplished by developing an Automatic Speech Recognition (ASR) system, which allows a computer to identify the words that a person speaks into a microphone or telephone and convert that into written text [3].

In a multilingual society like India, where there are about 1672 dialect of spoken forms, Automatic Speech Recognition (ASR) technology has a wider scope. It would be a vital step in bridging the digital divide between English speaking Indian masses and others. Since there are no input standards for Indian languages, it eliminates the keyboard mapping of different fonts of Indian languages [3].In the current Indian context; the machine-oriented interfaces restrict the computer usage to miniature fraction of the people, who are both computer literate and conversant with written English. This paper formulates a basic method to develop a basic speech recognition system for Malayalam, having the ability to recognize the isolated words in Malayalam speaker independently. Even though in this paper we are experimenting with different Malayalam words, this method can be used for other languages also. This method can be used for various speech based applications especially based on limited number of voice..

II. LITERATURE SURVEY

There are 24 phonemes in both the languages. The phoneme consists of seventeen (17) consonant and seven (7) vowels. Types of phonemes: vowels, semivowels, dipthongs, and consonants. Phonemic is a structure of perceive sound and how it is organized in a language. The present Bodo language originated form the Sino-Tibetan family of language. The Sino-Tibetan family has many sub-branches(fig 3.1), of which the Tibeto Burman had been living in the North-Eastern part of India [75]. The Bodo sub group of Tibeto Burman family of language comprises of the following tribes: Kachari or Bodo, Tiwa, Dimasa, Garo, Rabha, Kakborok, Koch, Deuri and Moran. Bodo is the second largest major language of Assam. It is also the official language in the Bodo dominated area. Many rivers like Dihing, Dibru, Dihong,

Dikarai etc. in the North-East region was named after some Bodo words which reveles that spatial distribution patern of related ethno-cultural groups with their cultural traits and phenomenon.The literary development of Bodo language seems very slow. The Bodo language in written form was found in the late 19th century and early 20thcentury [76]. Since the early 20th century, there started a development phase of the Bodo language. It can be said that the peak point in the entire development process of the Bodo language is the Socio-political movement initiated by some Bodo organizations in the year 1913. In the year 1963, the Bodo language was introduced in the primary school levels of the Bodo dominated areas as medium of instruction. In the year 2005, the language got constitutional recognition as one of the 22 scheduled languages of India. The Bodo language has got the status of one of the official languages of the state of Assam.In spite of the urgent need for automation in all domains, the development of strategies for speech recognition in regional languages is still perceived to be cumbersome due to various issues such as non-availability of speech corpus for training purpose, complexity in the language, lack of phoneme recognizers and difficulty in creating a speech corpus with necessary transcriptions. In a multilingual society like India only little work was done in making speech recognition systems in native languages. Major work is done on Hindi [4], Marathi [5] and Tamil [6] languages. A strategy for recognizing spoken words in Bodo language is an ongoing task. Till today several feature extraction techniques have been proposed and a several experiments have been performed. In this paper some standard ideas are proposed with some modifications.

III. METHODOLOGY

The various types of feature extraction and speech recognition approaches used.

A. Speech Recognition

In speech recognition we need to recognize the speech. The identification task is a matching that needs of making a decision whether a given voice sample is produced matches the stored one. An identity set of features is given to the system, and the unknown voice sample is compared with the stored voice. If the degree of similarity between the voice sample and the template exceeds a predefined threshold, it is accepted, and otherwise rejected [14].

B. Feature Extraction

Feature extraction is a special form of dimensionality reduction, we will do that by extracting a specific features from the speech, these features carry the characteristics of the speech which are different from one speaker to another, so these features will play the major role as our mission is to identify the spoken word and make a decision that highly depends on how much we were successful in extracting a useful information from the speech in a way enables our system to differentiate between different words and identify them according to their features.

C. MFCC

Several feature extraction algorithms such as; linear predictive coefficients (LPC), linear predictive cepstral coefficients (LPCC), mel frequency cepstral coefficients (MFCC), and human factor cepstral coefficient (HFCC). We used the (MFCC) algorithm are used to extract the features. MFCCs are less susceptible to the physical conditions of the speaker's vocal tract. The feature extraction is usually a non-invertible (lossy) transformation, as the MFCC described pictorially in Figure 1

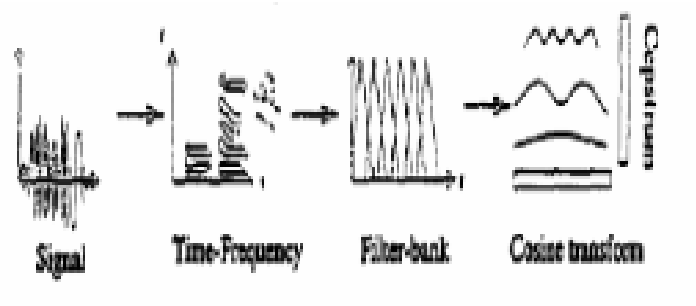


Figure 1 mel-frequency cepstrum (MFCC) calculation.

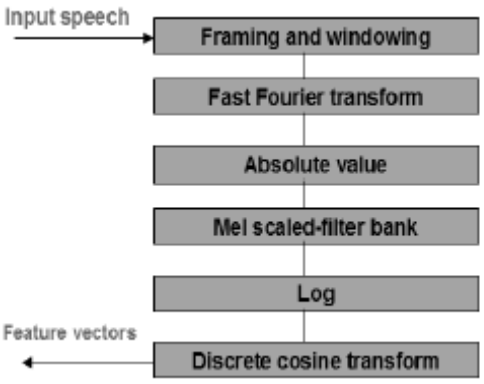


Figure 2 Block diagram for the feature extraction processes applying MFCC

D. Vector Quantization

Vector Quantization is a quantization technique used to compress the information and manipulate the data such in a way to maintain the most prominent characteristics [11]. VQ is used in many applications such as data compression (i.e. image and voice compression), voice recognition, etc. The technique of VQ consists of extracting a small number of representative feature vectors as an efficient means of characterizing the speech specific features. The clustering is done by a clustering algorithm The K-means algorithm partitions the T feature vectors into M centroids [17]. The algorithm first randomly chooses M cluster-centroids among the T feature vectors. Then each feature vector is assigned to the nearest centroid, and the new centroids are calculated for the new clusters. This procedure is continued until a stopping criterion is met, that is the mean square error between the feature vectors and the cluster centroids is below a certain threshold or there is no more change in the cluster-centre assignment. Flow diagram of the K-means algorithm is shown infigure3[7].

E. Feature Matching

In the speech recognition phase an unknown voice, represented by a sequence of feature vectors; it is compared with the database. For each distortion measure is computed, and the one with the lowest distortion is chosen. This match is selected as the identified voice.

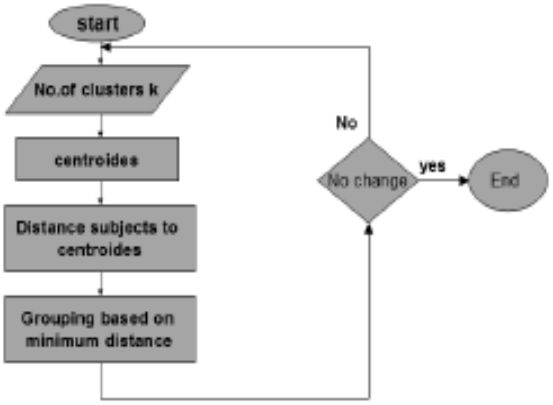


Figure 3: Block diagram for feature matching

F. Identification System

The Identification system consists of two phases, training and testing phase. In the training phase the sample words are stored in the database and in the testing phase different speakers are allowed to verify one by one. The steps in testing phase as follows in figure 4.

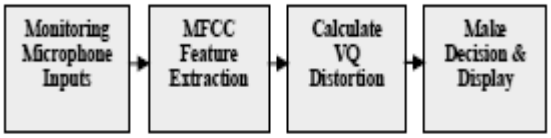


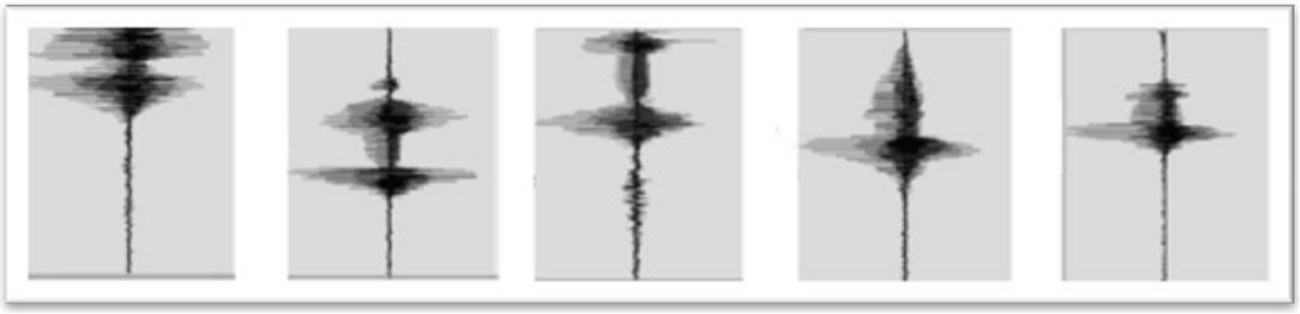
Figure 4 Block diagram for feature matching

G. Speech Database

For conducting this experiment we choose 10 Bodo words. Speech data recorded by using a high quality studio-recording microphone at a sampling rate of 8 KHz, 16bit PCM Mono.

Table 1: 10 words are recorded and stored in a database in the training phase.

Speaker ID	**Words in Bodo**
M1	सोनाब
M2	गोरलै
M3	अखाफोर
M4	थसे
M5	आरामि
F1	बिनि
F2	गुसथि
F3	आद्रा
F4	नडा
F5	हाया



IV. TRAINING AND RESULT

Table 2 shows the experimental results for ten different users (both male and female). The speakers are asked to speak the stored words and the corresponding identification rate described.

Table 2
Sample identification table

SpeakerID	Word1	Word2	Word3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word1 0
	सोनाब	गोरलै	अखाफो र	थसे	आरामि	बिनि	गुसथि	आद्रा	नङा	हाया
1	सोनाब	गोरलै	अखाफो र	थसे	आरामि	बिनि	गुसथि	आद्रा	नङा	हाया
2	सोनाब	गोरलै	अखाफो र	थसे	आरामि	बिनि	गुसथि	आद्रा	नङा	हाया
3	सोनाब	गोरलै	अखाफो र	थसे	आरामि	बिनि	गुसथि	आद्रा	नङा	हाया
4	सोनाब	गोरलै	अखाफो र	थसे	आरामि	बिनि	गुसथि	आद्रा	नङा	हाया
5	सोनाब	गोरलै	अखाफो र	थसे	आरामि	बिनि	गुसथि	आद्रा	नङा	हाया
6	सोनाब	गोरलै	अखाफो र	थसे	आरामि	बिनि	गुसथि	आद्रा	नङा	हाया
7	सोनाब	गोरलै	अखाफो र	थसे	आरामि	बिनि	गुसथि	आद्रा	नङा	हाया
8	सोनाब	गोरलै	अखाफो र	थसे	आरामि	बिनि	गुसथि	आद्रा	नङा	हाया
9	सोनाब	गोरलै	अखाफो र	थसे	आरामि	बिनि	गुसथि	आद्रा	नङा	हाया

The accuracy obtained by this system is shown in Table 3. Improvement can be obtained by a better approach, with the training set including utterances from a large number of speakers with variation in ages, gender and accents

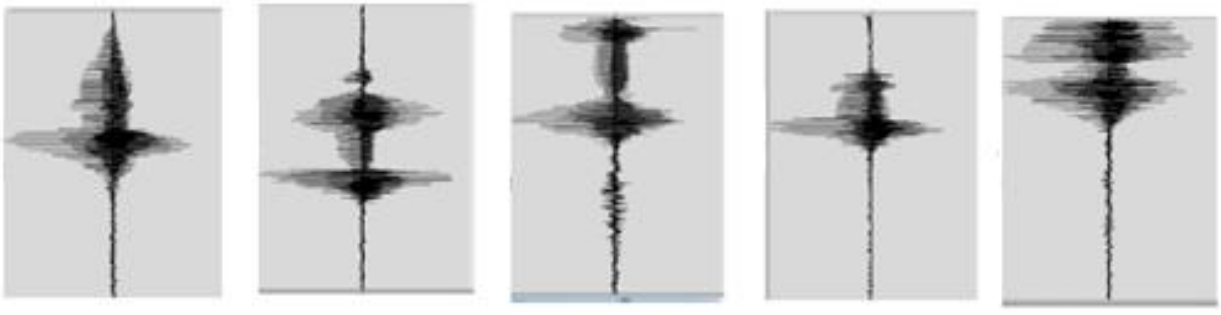


Table 3: Accuracy test results

SpeakerID	Words	Accuracy
1	सोनाब	89%
2	गोरलै	82%
3	अखाफोर	85%
4	थसे	80%
5	आरामि	86%
6	बिनि	87%
7	गुसथि	82%
8	आद्रा	89%
9	नङा	88%
	हाया	85%

The system consists of 10 stored words and 10 speakers and the experiment was conducted with almost 86 % accuracy

V. CONCLUSION

Speech recognition remains one of the most important problems in the current trends. It has numerous applications including those in, various mobile telephony, sorting applications, automatic scoring of tests containing multiple choice questions, automatic machine control, job application form sorting and various voice applications. The research in speech recognition in various languages is still going on and is a challenging task . To the best of our knowledge, little work has been done in Indian language, especially in Bodo as compared with those for non Indian languages. This paper has discussed an effective method for recognition of isolated Bodo words. This method can be also used for limited number of voice recognition in different languages irrespective of speakers. This confirms that the goal of creating a large vocabulary continuous transcription system is a realizable task in the near future.

ACKNOWLEDGEMENT

Firstly i would like express my sincere gratitude and heartfelt thanks to my guide Prof. P. H. Talukdar , Dept. Instrumentation and USIC, Gauhati University. I m very fortunate for having an opportunity to work with him from which i benefitted enormously.

REFERENCES

[1] Cini Kurian, Kannan Balakrishnan "*Automated Transcription System for Malayalam Language"*, International Journal of Computer Applications (0975 – 8887) Volume 19– No.5, April 2011

[2] Sonia Sunny , David Peter, "*Development of Speech Recognition System for Speaker Independent Isolated Malayalam Words"* , International Journal of Computer Science & Engineering Technology (IJCSET), ISSN: 2229-3345 Vol. 3 No. 4 April 2012

[3] Vimala.C, Dr.V.Radha, "*A Review on Speech Recognition Challenges and Approaches"*, World of Computer Science and Information Technology Journal (WCSIT), ISSN: 2221-0741, Vol. 2, No. 1, 1-7, 2012

[4] Poonam Bansal, Amita Dev, Shail Bala Jain , "*Optimum HMM combined with vector quantization for hindi speech word recognition* " ,IETE journal of research Year : 2008 , Volume : 54 , Issue : 4 , Page : 239-243

[5] Siddheshwar S. Gangonda, Dr. Prachi Mukherji , "*Speech Processing for Marathi Numeral Recognition using MFCC and DTW Features"* , International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 National Conference on Emerging Trends in Engineering & Technology (VNCET-30 Mar'12)

[6] AN. Sigappi1 and S. Palanivel, "*Spoken Word Recognition Strategy for Tamil Language* ", International Journal of Computer Science Issues, Vol. 9, Issue 1, No 3, January 2012,ISSN (Online): 1694-0814

[7] Allam Mousa ,Mare ,"*Text Independent Speaker Identification based on K-mean Algorithm*", International Journal on Electrical Engineering and Informatics - Volume 3, Number 1, 2011

[8] Md. Rashidul Hasan, Mustafa Jamil, Md. Golam Rabbani Md. Saifur Rahman *,"Speaker identification using Mel Frequency Ceptral Coefficients"*, 3rd International Conference on Electrical & Computer Engineering ICECE 2004, 28-30 December 2004, Dhaka, Bangladesh

[9] Vimal Krishnan V R, Athulya Jayakumar, Babu Anto.P, "*Speech Recognition of Isolated Malayalam Words Using Wavelet Features and Artificial Neural Network"* , 4th IEEE International Symposium on Electronic Design, Test & Applications

[10] http://www.spectrogramsforspeech.com/tutorials-2/wavesurfer-overview/

[11] Kayte Charansing Nathoosing *," Isolated Word Recognition for Marathi Language using VQ and HMM"*, Science Research Reporter 2(2):161-165, April 2012

[12] Tara Mohanan , "*Syllable Structure in Malayalam*" , Linguistic Inquiry, Vol. 20, No. 4 (autumn, 1989), pp. 589-625

[13] http://www.omniglot.com/writing/malayalam.htm (Visited September 2012)

[14] Daniel Jurasfky and James H. Martin, "*An introduction to natural language processing, computational linguistics, and speech recognition"*, Pearson Education, Inc. 2000
[15] Haowen Jiang, "*Malayalam: a Grammatical Sketch and a Text"*, Department of Linguistics, Rice University
[16] http://code.google.com/p/sharika (visited September 2012)
[17] http://www.bio.umontreal.ca/casgrain/en/labo/k-means.html (visited September 2012)
[18] http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/ voicebox.html (visited September 2012)

