ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

Automatic Recognition and Detection of Words in Speech- A Review

Kavya K S

P.G. Student, School of Computer Sciences M G University, Kottayam, Kerala India

ABSTRACT: Speech is the primary and the most convenient means of communication between people.Automatic Speech Recognition provides a way for natural communication between human and machine. It has become an interesting and challenging area. It helps to provide a capability to a machine for responding properly to spoken language.The main aim of this study is to understand the various approaches used for the recognition and detection of speech so that we can find out the methods that gives better accuracy and performance.

KEYWORDS: Automatic Speech Recognition,DTW algorithmHMM, Keyword Spotting,MFCC .

I. INTRODUCTION

Speech is the vocalized form of communication used by humans, which is based upon the syntactic combination of items drawn from the lexicon. Each spoken word is created out of the phonetic combination of a limited set of vowel and consonant speech sound units (phonemes). These vocabularies, the syntax that structures them, and their sets of speech sound units differ, creating many thousands of different, and mutually unintelligible, human languages.Speech is achieved by compression of the lung volume causing air flow which may be made audible if set into vibration by the activity of the larynx. This sound can then be made into speech by various modifications of the supralaryngeal vocal tract.

Speech recognition (SR) is the inter-disciplinary sub-field of computational linguistics that develops methodologies and technologies that enables the recognition and translation of spoken language into text by computers. It is also known as automatic speech recognition (ASR).ASR is technology that allows a computer to identify the words that a person speaks into a microphone or telephone and convert it to written text [9]. The aim of ASR research is to allow a computer to recognize speech in real-time as human understands [2]. The words that are spoken by any person independent of vocabulary size, noise, speaker characteristics and channel conditions is a major challenge in ASR. The accuracy rate of speech recognition depends on the feature extraction techniques as well as the training methods used.A speech interface would support many valuable applications, i.e., telephonedirectory assistance, spoken database querying for noviceusers, hands busy applications in medicine, officedictation devices, or automatic voice translation intoforeign languages [8].ASR is technology that allows a computer to identify the words
that a person speaks into a microphone or telephone and convert
it to written text

Keyword spotting has become an important branch of audiomining. Keyword spotting is well suited to data mining tasks,process large amount of speech such as real time monitoringand to audio document indexing. The task of locating theoccurrences of a given keywords K in a speech utterance istermed as keyword spotting (KWS). It plays an importantrole in audio indexing and speech data mining applications. The method is applied mainly for a large amount of spoken documents must be searched to learn whether they contain some specific words. The fast detection of the words andinformation about the exact location eliminate a lot of humanwork in such tasks like audio data mining [3]. Hence,Keyword spotting is well suited to data mining tasks forprocess large amount of speech because keyword spottingrequires significantly less processing power than transcription,and can therefore run at considerably faster speeds.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

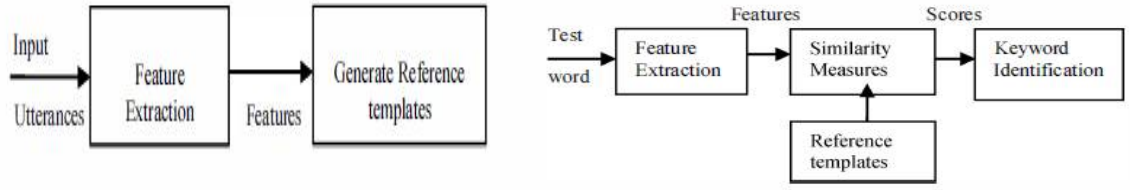
Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

This Paper is organized as follows. Section II describes. The major researches works carried out in the automatic speech recognition and the word spotting, Finally, Section IIIpresents conclusion of the review done in this work.

II. RELATED WORK

1. Chandra, E. [3] in this paper focused on Keyword Spotting System for Tamil Isolated Words using Multidimensional MFCC and DTW Algorithm. Tamil month names are recorded in noisy environment by a female speaker. Each month name in Tamilhas been repeated 25 times so that totally 300 words are recorded using AUDACITY speech software. A preprocessing is made for not only noise reduction, but also normalization. Then 12D, 26D and 39D MFCCs are calculated for allutterances and are stored as templates. Finally, DTW is used as a pattern matching algorithm due to its speed and efficiency in detecting similar patterns. Experiments have been conducted with multiple MFCCs to achieve best word identifier. The system is designed for Tamil language but allits modules except the spoken word references developed are language independent. This paper shows detailed clarification about audio mining and the keyword spotting. Audio mining also called audio searching technique is used to search audio files for occurrences of spoken words or phrases. The major advantage of audio mining is theability to process and search audio data thousands of timesfaster for large files than a human could - because the humanwould have to listen to each word of the audio. There are two main phases in audio mining system: training and template matching. During the training phase, atraining vector is generated from the speech signal of eachword spoken by the user.The training vectors extract the spectral features for distinguishing different classes of words.Each training vector can serve as a template for a single wordor a word class. These feature vectors are stored in a databasefor subsequent use in the template matching phase. During thekeyword matching phase, the user speaks any word that is tobe identified in the trained templates. Feature vector isgenerated for that word and compared with the trainedtemplates using pattern matching algorithm.



(a) (b)

Fig 1.Block Diagram Of Audio Mining System (a)Feature Extraction Phase (b)Feature Matching Phase

The preferred technique for feature extraction is MFCCwherein the features are generated by transforming the signalinto frequency domain. In general, cepstral features are morecompact, discriminable, and most importantly, nearlydecorrelated and therefore, they can provide higher baselineperformance over filter bank features. In phase 2 DTW is used. Dynamic time warping is an efficient algorithm optimizes to find the nonlinear alignment path between two sequence. their distance is obtained by time scaling one of the signalsnonlinearly so that it aligns with the other. For keywordspotting, a prototype of the keyword is stored as template andcompared to each of word in the incoming speech utterance. The experimental results shows that the number of occurrences ofgiven keywords are identified and maximum of 89.7% averageaccuracy of keyword spotting is obtained with 39 dimensionalMFCC. The 12, 26 and 39 dimensional MFCC are used forkeyword spotting and it is found that 39 dimensional MFCCare more efficient and has greater accuracy than 13 and 26dimensional MFCCs.

2.K. Anoop , P. Vivek and V. L. Lajish [1] in this work propose a keyword spotting approach for content based video indexingand retrieval system for classifying and indexing broadcast news videos in Malayalam. here used auditory modality to identify keywords to analyze the content of the news videos.A template based keyword spotting approach is then proposed for the purpose. Log EnergyFilter Coefficient (LEFC) feature is the major component in the keyword spotting. Theaudio signals are segmented into 10ms frames and appalled the filter bank for extraction ofeight dimensional LEFC feature. In feature matching stage, the Dynamic Time Warping(DTW) of the template and the test audio data feature vector is performed. The similarity distancesmeasures are computed and it is normalized by the

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

sum of energy in each template.Best matches from all templates are stored according to the distance measure and finally thedesired keyword will be spotted based on this distance measure.

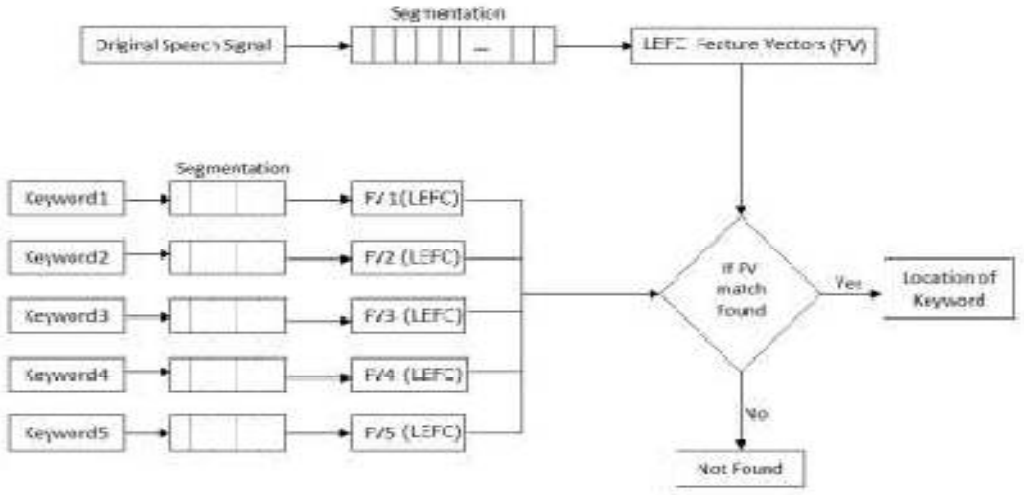


Fig 2.Keyword spotting using template matching

The final results shows that, template based keyword spotting using the LEFC feature can be effectively used for the speaker depended keyword spotting, while it will be less effective with the speaker independent keyword spotting.

3.Kurian, Cini, and Kannan Balakrishnan.[4] This paper presents a report onthe development of a speaker independent, continuoustranscription system for Malayalam. The system employsHidden Markov Model (HMM) for acoustic modeling and MelFrequency Cepstral Coefficient (MFCC) for feature extraction.It is trained with 21 male and female speakers in the age groupranging from 20 to 40 years.

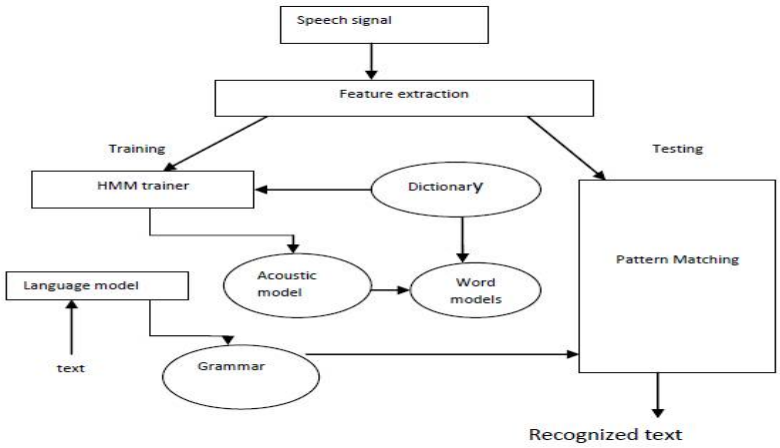


Fig 3.System Architecture of transcription system

During training phase the HMM trainercreates acoustic models from three components; feature vectors,language models, and dictionary. Word models which are builtfrom language models and pronunciation dictionary are used

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

forpattern matching in the testing phase along with the syntax andsemantics of the language. The simulation results shows that the transcription system for Malayalam achieved while working with mediumsize vocabulary.

4.Satori, Hassan, et al. [6] in this developed Arabic voice recognition system usingentirely Arabic environment. Both training andrecognizing process use Arabic characters. Here both training and test werebased on CMUSphinx4 system, which is HMM-based,speaker-independent, continuous recognition systemcapable of handling large vocabularies. This approach for modeling Arabic sounds in The CMUSphinx system consisted of generated and trainedacoustic and language models with Arabic speech data.The dictionary adopted in the experiments was made
using Arabic characters.The training process consists of: convert the audiodata to a stream of feature vectors, convert the text intoa sequence of linear triphone HMMs using the pronunciation dictionary, and findthe best state sequence or state alignment through thesentence HMM for the corresponding feature vectorsequence. For each senone, gather all the frames in thetraining corpus that mapped to that senone in theabove step and build a suitable statistical model for thecorresponding collection of feature vectors. Thecircularity in this training process is resolved using theiterative Baum-Welch or forward-backward training algorithm. The final results the recognition rate calculated that is found with better accuracy.

5.Radha, V. [5] This research work presents a speaker independent isolated speech recognition system for Tamil language. The most flexible and successful approach to speech recognition so far has been Hidden Markov Model (HMM) which is implemented in this research work. The HMM method provides a natural and highly reliable approach of recognizing speech for a broad range of applications. The system usually consists of two phases. They are called pre-processing and post-processing. Pre-processing involves feature extraction and the post-processing stage comprises of building an acoustic model, phonetic lexicon or the pronunciation dictionary and language modeling.
Initially the speech waveform is processed to produce a new representation as a sequence of vectors containing values of features or parameters. The parameter values extracted from raw speech are used to build acoustic models. The Dictionary provides pronunciations for words found in the Language Model.The pronunciations break words into sequences of sub-word units found in the Acoustic Model. The other component is called a language model, which gives the probabilities of sequences of words.The pre-processing or front-end of a speech recognizer is responsible for the extraction of the features from the speech waveform. The post processing consists of modeling of the system.
The performance of speech recognition system is generally measured in terms of word error rate, which is the ratio between misclassified words, and total number of tested words. The results shows that the obtained error rate for this research work is 0.88.

6.Moneykumar, Maya, et al [2] in this paper Malayalam Word Identification For Speech Recognition System proposed. Here also the MFCC feature extraction will be taken and the HMM model used for the final identification.
Here the word identification system designed for Malayalam language uses a syllable based segmentation approach. Instead of training the system with independent words, we use syllables and phoneme combinations for training and identification. This eliminates the problem of maintaining large set of training data, as different words common syllable as well as phone segments. New words can be added to the vocabulary without building new models for the existing syllables and phonemes corresponding to the word. The experiment results shows that syllable based word identification system for Malayalam using HTK on the Linux platform was performed using MFCC and HMM. The training was conducted for 40 vocabularies of bi-syllable words. The system was trained to identify the utterances within the training vocabulary, with a moderate accuracy.

III. CONCLUSION

The review of Automatic Recognition and Detection of Words in Speech for different spoken language has been made and the all the papers gives different innovative ideas. Major works in the recognition and detection of speech carried out in Hidden Markov Model (HMM), this gives better performance compared to other existing model. And also the Mel Frequency Cepstral Coeffients (MFCC) can be used as the better choice for the feature extraction process. In

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

future works large vocabulary speech recognition will be considered for improving the accuracy of the recognition system.

REFERENCES

[1] K. Anoop , P. Vivek and V. L. Lajish A Keyword Spotting Approach For ContentBased Indexing And Retrieval Of Malayalam News Videos, NSA-2015 Goa, India, October 7-9, 2015

[2] Moneykumar, Maya, et al. "Isolated Word Recognition System for Malayalam using Machine Learning." *12th International Conference on Natural Language Processing*. 2015.

[3] Chandra, E. "Keyword spotting system for Tamil isolated words using Multidimensional MFCC and DTW algorithm." *Communications and Signal Processing (ICCSP), 2015 International Conference on*. IEEE, 2015.

[4] Kurian, Cini, and Kannan Balakrishnan. "Automated transcription system for malayalam language." International Journal of Computer Applications 19.5 (2011): 5-10.

[5] Radha, V. "Speaker independent isolated speech recognition system for Tamil language using HMM." *Procedia Engineering* 30 (2012): 1097-1102.

[6] Satori, Hassan, et al. "Investigation arabic speech recognition using CMU sphinx system." *Int. Arab J. Inf. Technol.* 6.2 (2009): 186-190.

[7] Ghai, Wiqas, and Navdeep Singh. "Analysis of automatic speech recognition systems for indo-aryan languages: Punjabi a case study." *Int J Soft Comput Eng* 2.1 (2012): 379-385.

[8] Kurian, Cini, and Kannan Balakrishnan. "Speech recognition of Malayalam numbers." *Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on*. IEEE, 2009.

[9] Iqbaldeep Kaur, Navneet Kaur, Amandeep Ummat, Jaspreet Kaur, Navjot Kaur “Automatic Speech Recognition: A Review” IJCST Vo l. 7, Issue4, oct- dec2016.