Volume 5, Issue 4, 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

An Intelligent Mobile Application Concept by using Speech Recognition

Jay Patadia, Alpa Reshamwala
Department of Computer Engineering, NMIMS
Maharashtra, India

Abstract— ***We are living in the digital world and use of the mobile application is now part of our life. Nowadays, hands-free computing by using our voice is in demand. This type of Application gives ease of access to use our mobile phones and some of this application have intelligence too so it can do more tedious task also. These types of applications are salubrious (beneficial) for a disabled person. This paper contains an approach to implementation of an intelligent application and some useful concept by using speech recognition. To achieve this some basic techniques such as Hidden Markow model (HMM), Artificial Neural network (ANN), Hybrid Model (ANN and HMM) and MFCC and other useful algorithm for extracting appropriate features from speech is also discussed in this paper. Finally, it concludes that this type of application may be beneficial for the user and it gives a better mobile application experience in their life.***

Keywords— ***speech recognition techniques, feature extraction in speech recognition, Intelligent application.***

I. INTRODUCTION

In today's day to day life use of hands-free mobile devices is in new trend and demand so, the developer of mobile application are developing it in different varieties. There are many applications based on speech recognition in market and they are giving good result with some restriction. All those application which works only for one functionality as far as intelligence level is concerned. Our objective is to provide intelligence to existing speech recognition system so that it will give desirable results.

After giving intelligence to this type of system, it will be useful to mankind for their daily usages like a helper, friend, Personal Assistant and guide. This type of system also needs depth knowledge about the area where it is going to work. It would also be useful for disable people, Businessmen and teenager etc. For disable people, it works for operating mobile in a very simple way to speech to operate the task. For a businessman it will work as a personal assistant and for teenagers it will work as good friend or tutor.

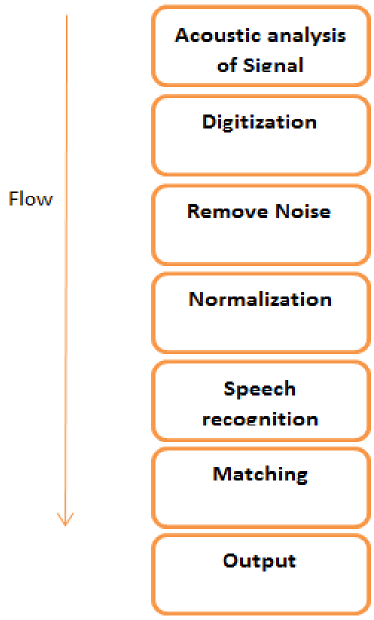

Figure-1 simple Flow diagram

Now in core of speech recognition system contains two modules [5].

1. Speech identification - for voice identification of user.
2. Speech recognition – To recognize word what they are speaking.

The figure-1 shown here gives the basic flow diagram for speech recognition:
Explanation for each step of flow diagram is given below:

- Acoustic analysis of signal- analysis of signal
- Digitization – convert analogue to digital signal
- Remove noise – remove noise for better performance or use noise cancelation mic
- Normalization – normalise all the data for ease of extraction
- Speech recognition – recognize speech as per the application concern by different algorithm and techniques.
- Database generation – generate training and testing phase database
- Matching – match query input with generated database for generating output
- Output – check generated output is appropriate or not

Followed by the artificial neural network, Hidden Markow Model and Hybrid model is used for speech recognition as well as words. After that it also shows MFCC for feature extraction technique in speech recognition. AdaBoost techniques are used for classification of emotion in speech and KFCG for vector quantization of different user. According to all these techniques and algorithm it shows appropriate inference which is discussed in this paper.

II. LITERATURE SURVEY

Literature survey is based on speech recognition technique which I found after referring research paper and it contains these sub points are as listed below:

- Hidden Markow Model
- Artificial Neural Network
- Hybrid Model

Hidden Markow Model

Hidden Markov Model (HMM) is a powerful statistical tool for modelling generative sequences that can be characterized by an underlying process generating an observable sequence [3]. HMMs have found application in many areas interested in signal processing. It consist Markow assumption and naïve bays approach for find sequences of words. Figure-2 shows **O** as observation sequence and **a** shows sequence state priority and likelihood which are being calculated.
Where, p is probability, O as observation sequence, S as state sequence and k is number of sequence.
The transition probabilities and the probability distributions along with their weights are the parameters of HMM. During training phase, these are optimized with respect to the training data to increase the likelihood of the model having being generated in the data[16].

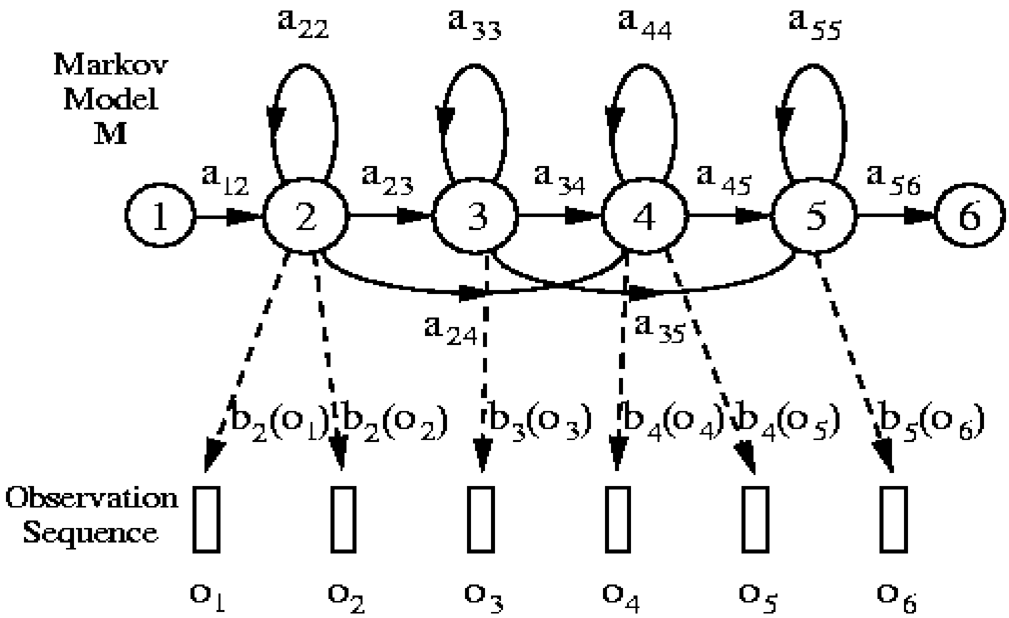

Figure-2 Hidden Markow Model [16]

Artificial Neural Network

Artificial Neural network comes from human's brain neurons concept. In human brain all neurons are connected with each other and share information from one to another in ANN and every node passes information from one node to another node which are connected together. Figure – 3 shows input layer, hidden layer and output layer. We can also show that all input nodes are connected with hidden layer and all hidden layer nodes are connected with output layer.

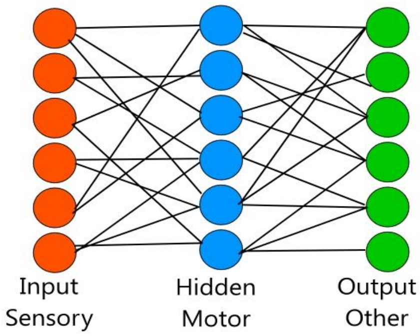

Figure-3 Artificial Neural Network

In this process our main task is to get desired output from given input but in this case we don't know which the desired one is. So it takes concept of high weighted sum priority to select proper node from many connected node. All this process can be done by Back propagation Algorithm. For computation this process is very hard and time taking so selection of desired node is very tedious task but it can be done by error checking and generate nearer desired output.

Hybrid Model

Artificial neural network and hidden Markow model are used in one model is call hybrid model. In this model, ANN rejected data will become the input of HMM. In this calculation of new cut-off value is pre-decided and in the processing step, the calculation will be done with input data. During processing, the dataset will be justified by cut-off value and if not rejected then the analysis of the frequency, time and amplitude will be done and extraction of new feature will be calculated [3].

III. RELATED ALGORITHM

Mel Frequency Cepstral Coefficient (MFCC)

Feature extraction by Mel Frequency Cepstral Coefficient (MFCC) is based on human hearing perception [6][7]. MFCC extracts features frame by frame in speech recognition. It has two types of filter which are spaced linearly at low frequency below 1000 Hz and logarithmic spacing above 1000Hz. A subjective pitch is present on Mel Frequency Scale to capture important characteristic of phonetic in speech. It has seven computational steps as shown in figure-3.

- Pre-emphasis

It refers to a filter within some frequency band and function of this step is to emphasize the higher frequencies and to increase energy of the signal.

- Framing

This step divides large signal into smaller frames with considering some overlapping. Since signal is not stable so have to make it stable for some time period for extracting feature, this time period is call as frame rate too.

- Windowing

Humming windowing is used for windowing and in this step signal is multiply by window function. It joints signal at the edges of the previous frame and avoid discontinuity in the beginning as well as in the end.

- DFT(Discrete Fourier Transform)

In this step we are using fast Fourier Transform because it is computationally efficient algorithm of DFT and it was used because each and every number of N sample frames has to be converted from time domain to frequency domain.

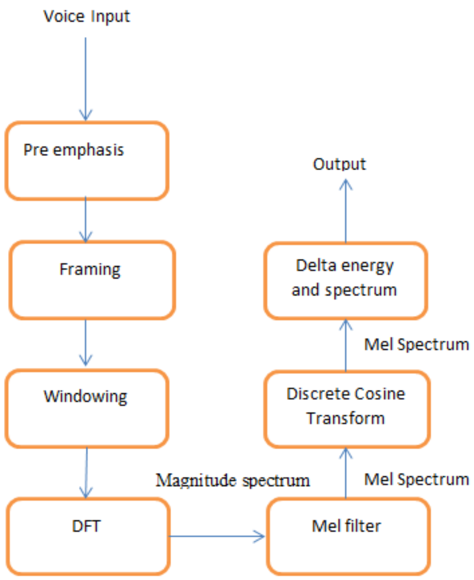

Figure-3 MFCC steps

- Mel Filter

Mel filter bank consists of triangular filters which are used to calculate the weighted sum of the filter spectral components [7]. Mel scale is approximated by output and given frequency is computed as a Mel.

- Discrete cosine Transform(DCT)

DCT is used in this step to convert Mel log spectrum into time domain. After that we are getting result as a MFCC. The set of coefficient is called as acoustic vector. DCT is applied to the log spectral-energy vector of such frame.

- Delta Energy and spectrum

The MFCC feature vector describes only the power spectral envelope of a single frame, but it seems like speech where it would also have information in the dynamics i.e. they are the trajectories of the MFCC coefficients over time. It turns out calculating the MFCC trajectories and appending them to the original feature vector increases ASR performance by quite a bit (if we have 12 MFCC coefficients, we would also get 12 delta coefficients, which would combine to give a feature vector of length 24).

And other algorithm like AdaBoost [4], Kekre's fast codebook generation (KFCG) [13-14], LBG are also useful for classification of speech data and for extracting feature from speech from that KFCG is really very useful for generating codebook using transform domain [13].

IV. INFERENCE

It contains scenario and concept for developing intelligent application for mobile. Its shows what type of techniques are useful for what task and which training is required for making is better for performance.
Below table shows that,

Table-1 Concept application requirement on the basis of reviewed paper

User	Task	Useful technique	Training data type
Disable	Speech to text And text to speech	ANN and HMM [5]	Synthesiser working
	Emergency Call to doctor	ANN HMM MFCC (feature extraction)	Doctor knowledge (some medicine database and useful spontaneous solving task training)
	Search to internet	ANN for recognition word Classification	Speech to text and search on internet
	Easy command to operate mobile	HMM and MFCC [7]	HMM training data set for understand words
Business	Remember Dates and remind	Hybrid model MFCC And basic understanding of cognitive science	Basic Cognitive approach to work and learn type
	User Identification	HMM and MFCC	User voice sample dynamic training and Specific feature extraction required
Teenager	As taking like friend	Hybrid model and feature extraction by MFCC for emotion Database high with intelligence	Need NLP knowledge , Emotion extraction is done by affected science[4]

Along with this knowledge this paper also focused on emotion extraction from the speech which is really difficult task. It also required more computation for make it working properly. In paper [4] it gives knowledge about AdaBoost algorithm and also gives better results also same as in paper [8] it gives knowledge about affected science base valence axis technique is also very useful for emotion extraction and using this knowledge we can make better technique and also this type of applications.
Making this type of application is needed much knowledge about test bids training to train data so it will easily recognize and according to that this application works fast and effectively as well.

V. CONCLUSIONS

Speech recognition is a very vast area to make hands-free computing by giving them intelligence. After reviewing all research and review papers, I conclude that there is some common thing for speech recognition. The flow of generating features and how it is used in the system for proper output is discussed here. So making an intelligent application is not an easy task but after using this inference table I can say that it will develop for huge and dynamic user experience for mankind. These all technique and algorithm in this paper shows that are useful to make the speech recognition system better and by giving appropriate training to this recognition system it will definitely change the experience of using mobile application.

ACKNOWLEDGMENT

This literature review was supported by my mentor Prof. Alpa Reshamwala. I am grateful to him for sharing his pearls of wisdom with me during this literature review.

REFERENCES

[1] Sadaoki Furui, Kiyohiro Shikano, Shoichi Matsunaga, Tatsuo Matsuoka, Satoshi Takahashi, and Tomokazu Yamada, RECENT TOPICS IN SPEECH RECOGNITION RESEARCH AT NTT LABORATORIES. 1994.

[2] Raja Sukumar A and Sarin Sukuma A, and Firoz Shah A and Babu A into P Key-Word Based Query Recognition In a Speech Corpus By Using Artificial Neural Networks , Second International Conference on Computational Intelligence, Communication Systems and Networks in 2010.

[3] Niladri Sekhar Dey, Ramakanta Mohanty and K. L. chugh Speech and Speaker Recognition System using Artificial Neural Networks and Hidden Markov Model, International conference on Communication Systems and Network Technologies(IEEE) 2012

[4] Jasdeep Singh Bhalla and Anmol Aggarwal, Using Adaboost Algorithm Along with Artificial Neural Networks for Efficient Human Emotion Recognition From Speech, IEEE, 2013

[5] Suma Swamy and K.V Ramakrishnan,, AN EFFICIENT SPEECH RECOGNITION SYSTEM ,Computer Science and Engineering: An International Journal (CSEIJ), Vol. 3, No. 4, August 2013

[6] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques,JOURNAL OF COMPUTING, VOLUME 2, ISSUE 3, MARCH 2010, ISSN 2151-9617.

[7] Dalmiya C.P, Dr. Dharun V.S and Rajesh K.P," An Efficient Method for Tamil Speech Recognition using MFCC and DTW for Mobile Applications", Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT 2013).

[8] Martin Borchert and Antje Diisterhoft, "Emotions in Speech - Experiments with Prosody and Quality Features in Speech for Use in Categorical and Dimensional Emotion Recognition Environments", Proceeding ofNLP-KE in 2005.

[9] Sabine Deligne, Satya Dharanipragada, Ramesh Gopinath, Benoît Maison, Peder Olsen and Harry Printz, "A Robust High Accuracy Speech Recognition System for Mobile Applications", IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 10, NO. 8, NOVEMBER 2002.

[10] Kil-Ram Ha, Jung-Hyun Kim, Jeh-Seon Youn and Kwang-Seok Hong, "Speech Recognition-Based Mobile Geo-Mashup Application Technology", Third International Symposium on Intelligent Information Technology Application in 2009.

[11] George E. Dahl, Dong Yu, Li Deng and Alex Acero, "Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 20, NO. 1, JANUARY 2012.

[12] Saf Asghar and Lin Cong, "ROBUST SPEECH RECOGNITION FOR MOBILE APPLICATIONS", IEEE in 1999.

[13] Dr. H. B. Kekre and Vaishali Kulkarni, "Comparative Analysis of Automatic Speaker Recognition using Kekre’s Fast Codebook Generation Algorithm in Time and Transform Domain" , International Journal of Computer Applications (0975 – 8887) inVolume 7– No.1, September 2010.

[14] DR. H. B. Kekre and Vaishali Kulkarni, "Performance Comparison of Speaker Recognition using Vector Quantization by LBG and KFCG", International Journal of Computer Applications (0975 – 8887) in Volume 3 – No.10, July 2010.

[15] Ms. Rupali S Chavan and Dr. Ganesh. S Sable, "An Overview of Speech Recognition Using HMM", International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 2, Issue. 6,, pg.233 – 238 in June 2013

[16] Ashok Shigli, Ibrahim Patel and Dr. K. Srinivas Rao, "A SPECTRAL FEATURE PROCESS FOR SPEECH RECOGNITION USING HMM WITH MFCC APPROACH", National Conference on Computing and Communication Systems (NCCCS) in 2012

[17] Hemakumar G and Dr. Punitha P., "Speaker Dependent Continuous Kannada Speech Recognition Using HMM", 2014 International Conference on Intelligent Computing Applications in 2014.

