# EFFICIENT SPEECH RECOGNITION USING CORRELATION METHOD

**Meenal Raheja***

**Ashish Oberoi***

**Mamta Oberoi****

## ABSTRACT

*Speech recognition applications are becoming more and more useful nowadays. Various interactive speech aware applications are available in the market. But they are usually meant for and executed on the traditional general-purpose computers.The Speech is most prominent & primary mode of Communication among of human being. The communication among human computer interaction is called human computer interface. Speech has potential of being important mode of interaction with computer. Correlation represents the similarities between two signals. Here I use correlation to Recognize the uttered words (Especially in the same voice). First of all have to record the uttered words, sample it and save it in a wav file and use the function to check whether it has already in the database. The input voice signal will be correlated with voice signals already recorded in previous sessions. All recorded audio patterns must be placed in the working directory. If both voice signals match an "allow signal" will be generated. Otherwise an "access denied" signal will generated instead.*

*Department of Computer Engineering, M.M. University, Mullana, Ambala, Haryana , India

**Department of Statistics, MLN College,Yamuna Nagar, Haryana, India

## 1. INTRODUCTION

Speech Recognition is the process by which a computer (or other type of machine) identifies spoken words. Basically, it means taking to our computer, AND having it correctly recognize what you are saying. The speech is primary mode of communication among human being and also the most natural and efficient form of exchanging information among human in speech. So, it is only logical that the next technological development to be natural language speech recognition. **Speech Recognition** can be defined as the process of converting speech signal to a sequence of words by means Algorithm implemented as a computer program. Speech processing is one of the exciting areas of signal processing. The goal of speech recognition area is to developed technique and system to developed for speech input to macine based on major advanced in statically modeling of speech.

A **speech recognition** device connected to a speech recognition device that generates a recognition result by executing correlation processing which matches input voice data with an acoustic model and a word dictionary by using conversion rules for conversion between a first-type character string expressing a sound and a second-type character string for forming a recognition result, the speech recognition learning device comprising: a character string recording unit that records, in association with each other, a first-type character string generated in a process in which a recognition result is generated by the speech recognition device, and a second-type character string corresponding to the first-type character string; an extraction unit that extracts, from a second-type character string corresponding to a word recorded in the word dictionary, character strings each constituted by a series of second-type elements that are constituent units of the second-type character string, as second-type l character string candidates-

- selects a second-type learned character string, from among the second-type learned character string candidates extracted by the extraction unit, that matches at least part of the second-type character string recorded in the character string recording unit.
- extracts, from the first-type character string recorded in the character string recording unit in association with the second-type character string, a portion that corresponds to the second-type learned character string, as a first-type learned character string.
- data indicating a correspondence relationship between the first-type learned character string and the second-type learned character string.

A typical speech recognition system starts with a preprocessing stage, which takes a speech waveform as its input, and extracts from it feature vectors or observations which represent the information required to perform recognition. This stage is efficiently performed by software. The second stage is recognition. Word-level acoustic models are formed by concatenating phone-level models according to a pronunciation dictionary. These word models are then combined with a language model, which constrains the recognizer to recognize only valid word sequences. The decoder stage is computationally expensive. Although there exist software implementations that are capable of real time performance.Firstly, there exist real telephony-based applications used for call-centers where, the speech recognizer is required to process a large number of spoken queries in parallel. Secondly, there are non-real time applications, such as off-line transcription of dictation, streams in parallel may offer a significant financial advantage. Thirdly, the additional processing power offered by an FGPA could be used for real-time implementation of the “next generation” of speech recognition algorithms, which are currently being developed in laboratories.

A speech recognition system can be used in many different modes :-

**Speaker dependent / Independent system :-** A speaker dependent system is a system that must be trained on a specific speaker in order to recognize accurately what has been said. To train a system, speaker is asked to record predefined words or sentences that will be analyzed and whose analysis results will be stored. This mode is mainly used in dictation system where a single speaker is using the speech recognition system.

**Isolated word recognition :-** This is the simplest speech recognition system and the less greedy in terms of CPU requirements. Each word is surrounded by a silence so that the word boundaries are well known. The system does not need to find the beginning and the end of each word in a sentence. The word is compared to a list of words models

**Continuous speech recognition :-** Continuous speech recognition is much more natural and user-friendly. It assumes the computer is able to recognize a sequence of words in a sentence. But this mode requires much more CPU and memory, and the recognition accuracy is really inferior compared with the preceding mode.

**Keyword spotting :-** This mode has been created to cover the gap between continuous ans isolated speech. Recognition system based keyword spotting are able to identify in a sentence a word or a group of words.

**Speech Recognition Techniques:** The goal of speech recognition is for a machine to be able to "hear,” understand," and "act upon" spoken information. The earliest speech recognition systems were first attempted in the early 1950s at Bell Laboratories, Davis, Biddulph and

Balashek developed an isolated digit Recognition system for a single speaker. The goal of automatic speaker recognition is to analyze, extract characterize and recognize information about the speaker identity. The speaker recognition system may be viewed as working in a four stages:

1. Analysis
2. Feature extraction
3. Modeling
4. Testing

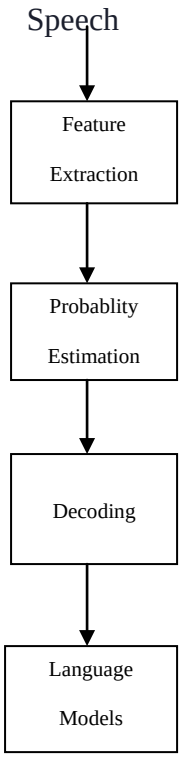


**Figure 1. Speech Recognition Process.**

## 2. THE PROPOSED METHOD

In Proposed method, the input audio files are used to record the uttered words ,and then sample it and save the file in a wav form. Finally check the database whether the file exist or not.

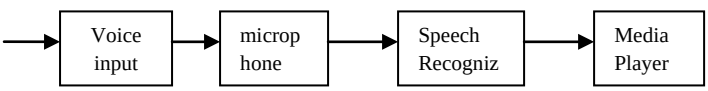


**Figure 2. Proposed model for Speech Recognition**

2.1. Voice Input

The input is human voice which, as explained before, is sampled at rate of 16,000 per second. It should be given in live mode. But because of some conflicts in the channel settings of the sound card and that used by the software, we are not able to do it in live mode. We are running the recognizer in batch mode, instead, i.e. taking input in the form of a pre-recorded audio file (in RAW format).

2.2. Microphone

The microphone that we are using for recognition is built onto the PXA27x platform itself. It has got its own advantages and disadvantages:

Advantages: Nothing to plug in And User's hands are free.

Disadvantages: Low accuracy unless the user is close to the monitor. And not good in a noisy environment.

2.3 Speech Recognizer:

Platform speed directly affected our choice of a speech recognition system for our work. Though all the members of the SPHINX recognizer family have well-developed programming interfaces, and are actively used by researchers in fields such as spoken dialog systems and computer-assisted learning, we chose the PocketSphinx [9, 10] as our speech decoder which is particularly meant for embedded platforms. It is a version of open-source Sphinx2 speech recognizer which is faster than any other SR system. We cross-compiled PocketSphinx on XScale and executed the various test scripts present in it. Both the digits and word recognition scripts are up and running on PXA27x. The voice input is supposed to be given from its microphone. PXA27x microphone is set to accept Steror input Only. PocketSphinx speech decoder takes voice input in MONO format only. Due to limitations in the code, we were not able to solve this problem, and hence we are using pre-recorded audio files for this purpose. The SR process decodes these input files, identifies the command and generates output accordingly.In our application, two input files have been used - ready.raw and stop.raw. The first one is having the utterance - PLAY - that when recognized by SRS, fires a command to the bash asking it to play a media file. It sends the Mplayer command - mplayer inc_160_68.avi Voice Input Microphone Speech Recognizer Mplayer XScale LCD (media file playing)The Mplayer starts playing this particular file on the pxa27x LCD display. This is done by creating a child process so that the speech recognition system keeps running to take further inputs. The second command has the utterance - STOP - that kills the playing process.

## 3. RESULT

3.1 Analysis of results

The proposed algorithm is implemented in MATLAB environment. Firstly Uttered words are recorded. Then sampled at the rate of 16000 Per second, and save it as the wav file. Finally check whether it is in database.

To evaluate the results we have waveform of all uttered words(‘One, two, three, four, five’). Show as below-

Speechrecognition(‘one.wav’)

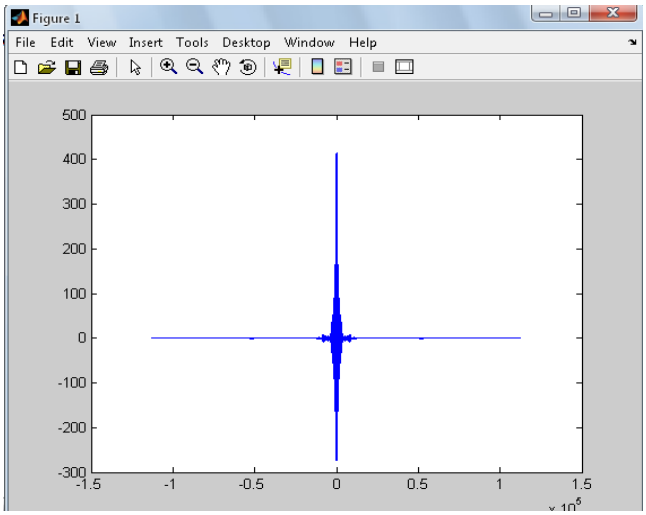


Figure 3. Waveform For 1st audio file

Speechrecognition(‘two.wav’)

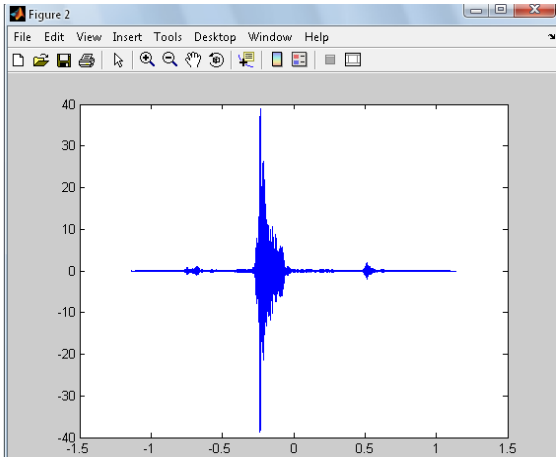


Figure 4. Waveform For 2nd audio file

Speechrecognition('three.wav')

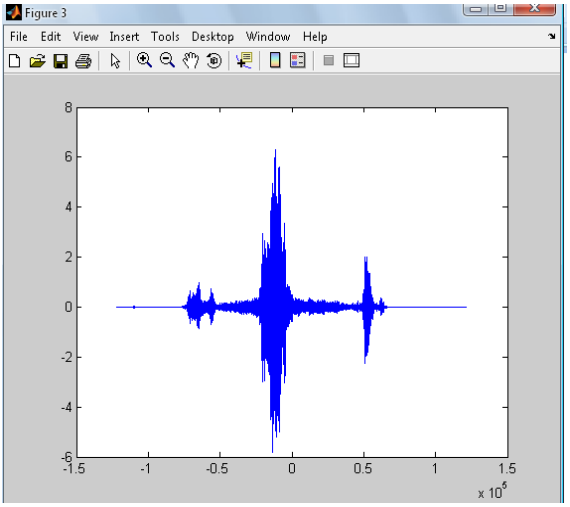


Figure 5.Waveform for 3rd audio file

Speechrecognition('four.wav')

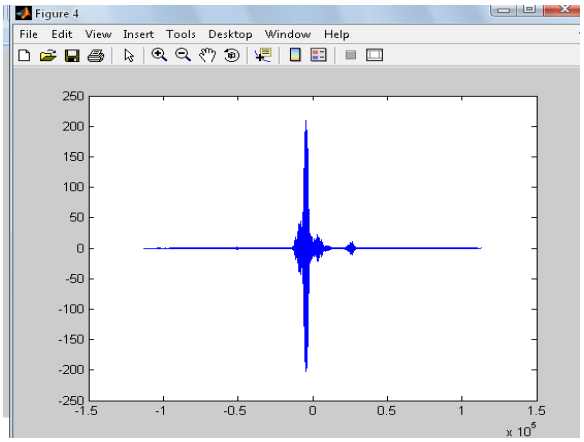


Figure 6. waveform for 4th audio file

Speechrecognition('five.wav')

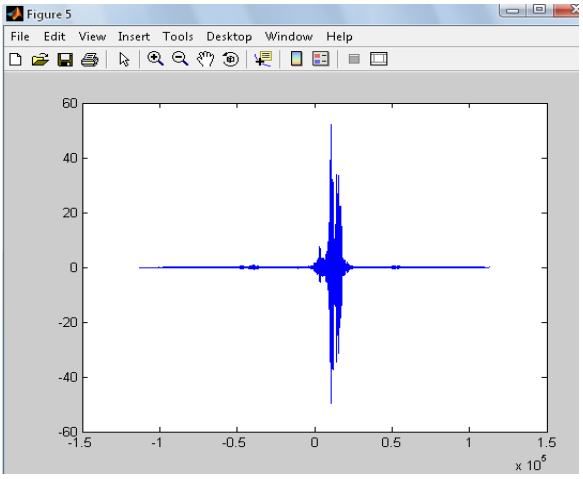


Figure 7.Waveform for 5th audio file

## 4. CONCLUSION

In this paper , speech recognition using correlated method is presented. The proposed method is based on the input audio file and sampling of wav file. The method consist of four phases. In first stage, uttered words are being recorded especially in same voice. Secondly , sample the audio file at the rate of 16000 per second. When input the audio file (as one.wav) the corresponding waveform will display on monitor and a voice is returned from the microphone. If the voice already exist in database then a allow signal is generated otherwise access denied signal generated that show the wav file is not in the database. The algorithm is simple, using only the correlated functions.

## REFERENCES

[1] Tang, K. W., Ridha Kamoua, and Victor Sutan. "Speech Recognition Technology for Disabilities Education." Journal of Educational Technology Systems volume(33) pp. 173-84,2004.

[2] H. Hermansky, B. A. Hanson, and H. Wakita, "Perceptually based linear predictive analysis of speech," Proc. IEEE Int. Conf. on Acoustic, speech, and Signal Processing," pp. 509-512, Aug.1985.

[3] An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition, S. E. Levinson, L. R. Rabiner and M. M. Sondhi; in Bell Syst. Tech. Jnl. v62(4), pp1035--1074, April 1983.

[4]. A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang, "Phoneme recognition using time-delay neural networks," IEEE Transactions on Acoustics, Speech and Signal Processing, volume. 37, pp. 328-339, 1989.

[5] Goel, V.; Byrne, W. J. "Minimum Bayes-risk automatic speech recognition".Computer Speech & Language 14 (23) volume 12 p.p.115-135 2006.

[6] J. Wu and C. Chan,"Isolated Word Recognition by Neural Network Models with Cross-Correlation Coefficients for Speech Dynamics," IEEE Trans. Pattern Anal. Mach. Intell., volume. 15, pp. 1174-1185, 1993

[7]. Janet M. Baker, Li Deng, James Glass, Sanjeev Khudanpur, Chin-Hui Lee, Nelson Morgan, Douglas O'Shaughnessy (MAY, 2009). "Research Developments and Directions in Speech Recognition and Understanding, Part 1". IEEE SIGNAL PROCESSING MAGAZINE. Retrieved May, 2010.

[8]An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition, S. E. Levinson, L. R. Rabiner and M. M. Sondhi; in Bell Syst. Tech. Jnl. v62(4), pp1035--1074, April 1983.

[9]. Computational Speech Processing: Speech Analysis, Recognition, Understanding, Compression, Transmission, Coding, Synthesis ; Text to Speech Systems, Speech to Tactile Displays, Speaker Identification, Prosody Processing : BIBLIOGRAPHY, *by Conrad F. Sabourin,2 volumes, pp. 1187 1994.*

[10]. Garrett, Jennifer Tumlin, et al. "Using Speech Recognition Software to Increase Writing Fluency for Individuals with Physical Disabilities." Journal of Special Education Technology volume(09), pp. 25-41, 2011.