6th International Conference on Recent Trends in Engineering, Science & Management
Shri Guru Teg Bahadur Khalsa College Anandpur Sahib, Punjab (ICRTESM-17)
8th January 2017 , www.conferenceworld.in ISBN: 978-93-86171-21-4

ANALYZING AND IMPLEMENTATION OF SPEECH RECOGNITION AS ADVANCED BIOMETRIC SYSTEM

[1]Dr. Gagandeep Jagdev, [2]Tarsem Singh

[1]Dept. of Computer Science, Punjabi University Guru Kashi College, Damdama Sahib (PB).

[2]Research Scholar (M.Tech. CE), Yadavindra College of Engineering, Talwandi Sabo (PB).

ABSTRACT

The process of converting spoken words into text is known as speech recognition. Speech Recognition makes its place among most discussed techniques of biometrics. The major challenge in the field of speech recognition is that it can be altered by dialects, accents and mannerisms. The most natural form of human communication is speech and its processing has been one of the most exciting areas of signal processing. Because of advancements in speech recognition technology, today it has been made possible that computer understands human voice commands and human languages. The primary goal of speech recognition is to develop a system for speech input to machine. Technically stated, speech recognition is the ability of a machine or a program to identify words and phrases in spoken language and convert them into machine-readable format. The central theme of this research paper is to discuss different approaches involved in speech recognition system along with implementation involving creating new files in templates and matching the input file with the already existing audio files present in the database, which on exact match can provide authentication to any particular application.

Keywords – Acoustic-phonetic approach, speech recognition, template-based approach, knowledge-based approach.

I. INTRODUCTION

When we make calls in big companies, it is not a person who answers the phone, instead it is an automated voice recordings that answers and instructs people to press buttons to move through option menus. Technology has even advanced further today. There is no need to press buttons, user can just speak some words as instructed by a recording to fulfill the requirement. All this is a kind of speech recognition program and comes under automated phone system. These programs fall into two categories [5, 6].

- Small-vocabulary/many-users

  These systems are preferred for automated telephone answering. In this category, usage is restricted to small number of predetermined commands and inputs, for instance, basic menu options or numbers. There

is always a possibility that users can speak with great deal of variation in accent and speech patterns and system will understand them most of the time.

- Large-vocabulary/limited-users

  This system is preferred in environment with limited users. These systems are trained to perform best with small number of primary users. The percentage of accuracy in this case is above 85 percent. The vocabulary of such system is in tens of thousands. If a person other than primary user attempts to use the system, the accuracy rate can fall drastically as system is not trained for working with the voice of an outsider.

Speech recognition has also proved boon for people with disabilities who can't do typing. If a user has lost the utilization of his/her hands or for visually impaired users where it is not convenient to use Braille keyboard, the system permits personal expression through dictation and by other computer tasks.

Speech recognition systems comes under the choice between discrete and continuous speech. It has been observed that when the words are spoken separately with a pause between each one, are understood by the system more effectively as compared with the continuous speech.

A computer system has to go through several complicated steps in converting speech to on-screen text. When humans speak, we create vibrations in the air. The analog-to-digital converter (ADC) translates this analog wave into digital data. This is done by taking samples and digitizing the sound by taking accurate measurements of the wave at frequent intervals. The system filters the digitized sound in order to eradicate unwanted noise and often to separate it into different bands of frequency. It also normalizes the sound and adjusts it to a constant volume level. Sound also needs to be temporarily aligned. People do not always speak at same speed. So, there is a need to adjust the sound to match the speed of the template sound samples stored in system's memory [1, 2].

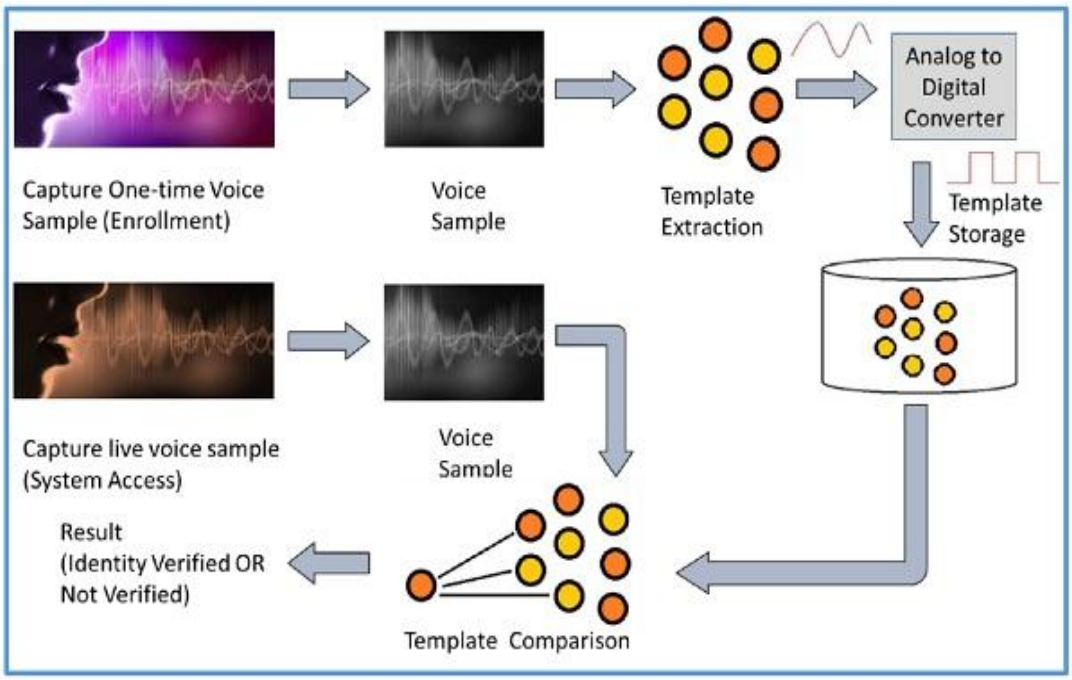


Fig. 1 Working of Speech recognition system

Thereafter the signal is divided into small chunks as short as few hundredths of a second, or even thousandths in the case of plosive consonant sounds (consonant stops produced by obstructing airflow in the vocal tract, like

"p" or "t". These segments are then matched to know phonemes in the appropriate language. A phoneme is the smallest element of a language. There are nearly 40 phonemes in English language. Fig. 1 shows the detailed functioning of speech recognition system.

Speech recognition deals with speech variability and account for learning relationship between corresponding word and specific utterance. There have been two popular trends in the field of speech recognition over the recent years. First one is academic approach and second is pragmatic which makes use of technology and offers simple low-level interaction with machine, replacing switches and buttons. While the former approach has always made promises for future, the second approach is in already in use [7].

II. TYPES OF APPROACHES IN SPEECH RECOGNITION

There are basically three different approaches to speech recognition. These are described as under [3, 4].

- **Acoustic-Phonetic Approach**

In this system tries to decode the speech signal in a sequential manner based on relations between phonetic symbols and acoustic features of the speech waveform. The steps involved in this approach are mentioned as under.

- In first step an appropriate spectral representation of the speech signal is provided and is classified as parameter measurement process.
- The next is feature detection stage where the spectral measurements are converted to a set of features describing acoustic properties of the various phonetic units.
- In last step, the recognizer makes an attempt to determine the best matching word or sequence of words.

- **Pattern Recognition Approach**

In this speech patterns are used directly without any feature determination and segmentation. This approach works in two steps, training of speech patterns and recognition of patterns. A sequence of measurements is made on the input signal to define the test pattern. The unknown test pattern is then compared with each sound reference pattern and a measure of similarity between the test pattern and reference pattern is computed. Finally, the decision rule decides which reference pattern best matches the unknown test pattern based on the similarity scores from the pattern classification phase.

***Template Based Approach*:** Template based approach to speech recognition have provided a family of techniques that have advanced the field. A collection of prototypical speech patterns is stored as reference patterns representing the dictionary of candidate s words. Recognition is then carried out by matching an unknown spoken utterance with each of these reference templates and selecting the category of the best matching pattern. Each word must have its own full reference template; One key idea in template method is to derive a typical sequences of speech frames for a pattern (a word) via some averaging procedure, and to rely on the use of local spectral distance measures to compare patterns. Another key idea is to use some form of dynamic programming to temporarily align patterns to account for differences in speaking rates across talkers as well as across repetitions of the word by the same talker.

ISBN: 978-93-86171-21-4

***Stochastic Approach*:** Stochastic modeling entails the use of probabilistic models to deal with uncertain or incomplete information. In speech recognition, uncertainty and incompleteness arise from many sources; for example, confusable sounds, speaker variability, contextual effects, and homophones words. Thus, stochastic models are particularly suitable approach to speech recognition. The most popular stochastic approach today is hidden Markov modeling. A hidden Markov model is characterized by a finite state Markov model and a set of output distributions. A template based model is simply a continuous density HMM, with identity covariance matrices and a slope constrained topology. Although templates can be trained on fewer instances, the lack the probabilistic formulation of full HMMs and typically underperforms HMMs. Compared to knowledge based approaches; HMMs enable easy integration of knowledge sources into a compiled architecture. A negative side effect of this is that HMMs do not provide much insight on the recognition process. As a result, it is often difficult to analyze the errors of an HMM system in an attempt to improve its performance. Nevertheless, prudent incorporation of knowledge has significantly improved HMM based systems.

- **Artificial Intelligence Approach (Knowledge Based Approach)**

The Artificial Intelligence approach is a hybrid of the acoustic phonetic approach and pattern recognition approach. Knowledge based approach uses the information regarding linguistic, phonetic and spectrogram. Knowledge based approach uses the information regarding linguistic, phonetic and spectrogram. Some speech researchers developed recognition system that used acoustic phonetic knowledge to develop classification rules for speech sounds. The basic idea is to compile and incorporate knowledge from a variety of knowledge sources with the problem at hand.

Real-world applications

Despite its limitations, present speech recognition technology can be a very useful tool for a variety of applications, as long as designers and users fully understand the boundaries and weaknesses of such systems. It is regrettable that the desire to hype up a new product or generation of speech recognition engine sometimes leads to blatantly misleading statements or misrepresentation of the realities of speech recognition and its role in real-world delivery [1, 2].

- Speech recognition is already used for live subtitling on television, as dictation tools in the medical and legal profession, and for off-line speech-to-text conversion or notetaking systems. For all these applications, human editing of the output is needed to achieve really good levels of accuracy. In addition, there are an increasing number of small vocabulary or specialized command and control applications and voice command in smartphones, to home automation.
- One can also use speech recognition software in homes. A variety of software products enables users to dictate to their computers or electronic gadgets and convert their words to text in a word processing or e-mail document. It can be used to control things such as heating, lighting or appliances. It also forms a key part of spoken dialogue systems.
- People can communicate with a speech recognizer using a phone, or by wearing a pin-on microphone (that may be wireless). Microphones may also be attached to the walls or placed on a table, although this makes the task of recognition harder.

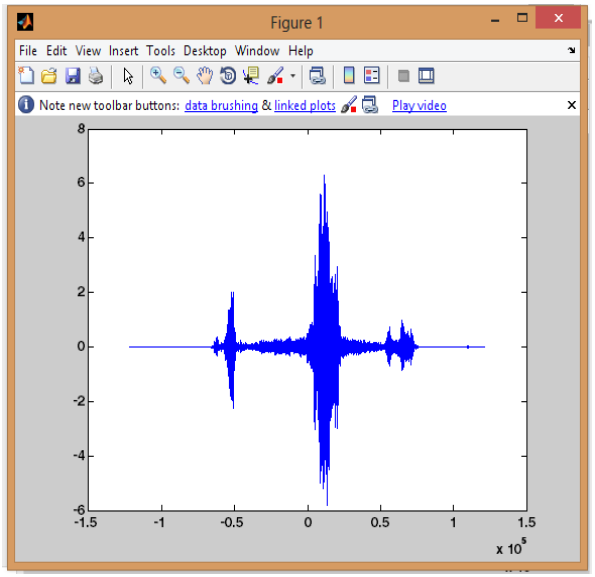
What is not feasible with the current state of science and technology, is to produce a system that converts free, natural speech into text in a fully reliable manner or with at least human-level accuracy.

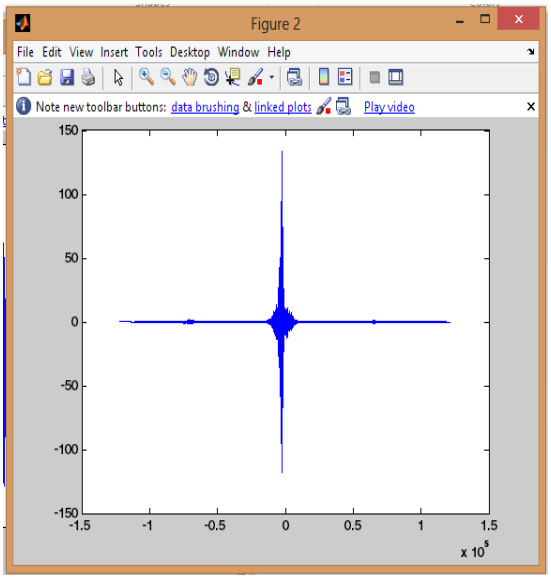
III. IMPLEMENTATION

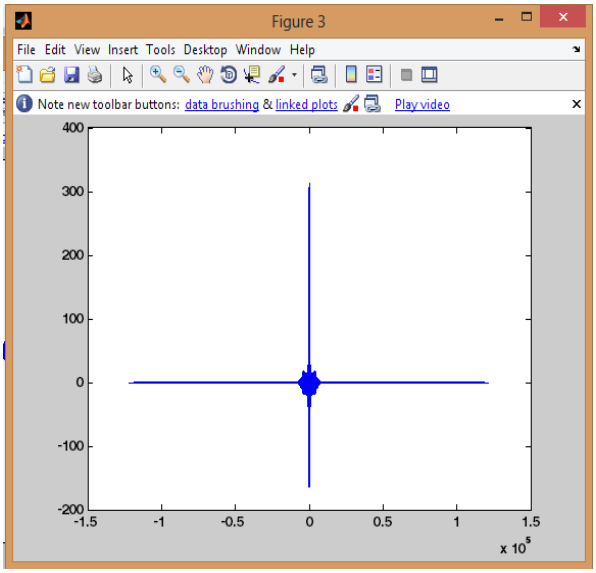
The research work is implemented using Matlab 7.8.0 (R2009a). The entire coding has been distributed in two Matlab files.

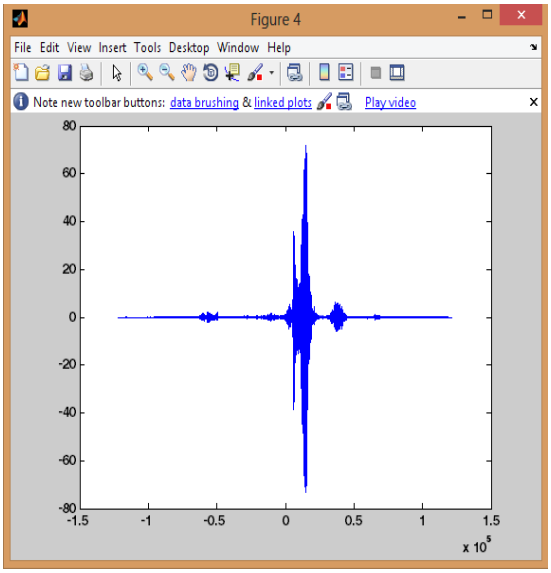
In the first module, system has been implemented that recognizes speech of a user and generates an audio file which can be added in the template further. The coding comprises the feature of adjusting the time period and frequency of the audio file as required.

The second module deals with recognizing and plotting graphs when exact match of the file is found in the template. A correlation method has been adopted in this module. There are currently 10 different wav files in the template which can be further extended as required. The snapshots below display that the Figure 3 is the exact match in this case because it is having a least scattered graph.









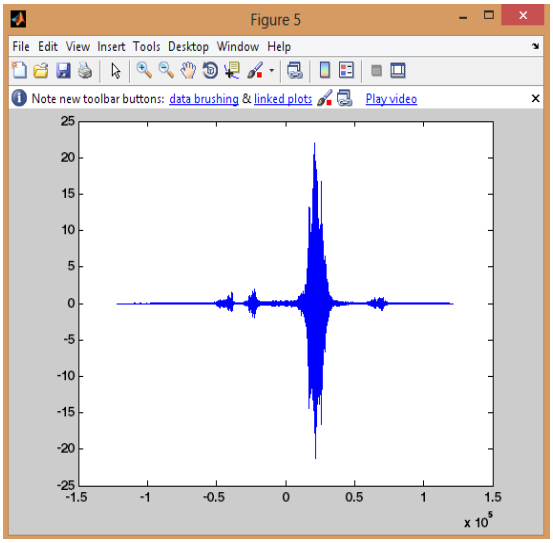


Fig. 2 Snapshots obtained after executing the source code on Matlab with Figure 3 showing the most accurate match

IV. CONCLUSION

Speech recognition has a big potential in becoming an significant factor of interaction between human and machine in the near future. This results obtained from this research paper are promising and is an effective step in analyzing the working of speech recognition. A universal translator is still far into the future, however -- it's very difficult to build a system that combines automatic translation with voice activation technology. The major problem to be handled is to make a system that can flawlessly handle barricades like slang, dialects, accents and background noise. The different grammatical structures used by languages can also pose a problem. For example, Arabic can convey the same meaning in a single letter word which is done by one complete sentence in English.

The pros and cons of speech recognition is summarized as under.

Advantages:

a. Non-intrusive. High social acceptability.

b. Verification time is about five seconds.

c. Cheap technology.

Disadvantages:

a. A person's voice can be easily recorded and used for unauthorized PC or network.

b. Low accuracy.

c. An illness such as a cold can change a person's voice, making absolute identification difficult or impossible.

There is a large possibility that at some point in the future, speech recognition may become speech understanding. Although it is a huge hop in terms of computational power and software complexity, some researchers claim that speech recognition development offers the most direct line from the computers of today to true artificial intelligence. We can talk to our computers today. In 25 years, they may very well talk back.

REFERENCES

[1] Joost Van Doremalen et. al., "Optimizing Automatic speech recognition for Low-Proficient Non-Native Speakers", EURASIP Journal on audio,speech amd music processing.

[2] Youssef Zouhir et. al., "A bio-inspired feature extraction for robust speech recognition", SpringerPlus, 2014, 3:651.

[3] Woon S. Gan et. al., "Intelligent audio,speech, and music processing applications", EURASIP Journal of audio, speech and music processing, DOI: 10.1155/2008/854716.

[4] Miss Himanshu et. al., "Literature survey on automatic speech recognition system", IJARCSSE, Volume 4, Issue 7, July 2014, ISSN: 2277128X.

[5] Preeti Saini, Parneet Kaur, "Automatic speech recognition: A review", IJETT, Volume 4, Issue 2, 2013, ISSN: 2231-5381.

[6] Suma Shankaranand et. al., "An enhanced speech recognition system", IJRDET, ISSN: 2347-6435 (online), Volume 2, Issue 3, March 2014.

[7] Shikha Gupta et. al., "A study on Speech Recognition System: A literature survey", IJSETR, Volume 3, Issue 8, August 2014.

About the author

Dr. Gagandeep Jagdev is a faculty member in Dept. of Computer Science, Punjabi University Guru Kashi College, Damdama Sahib (PB). His total teaching experience is above 10 years and has above 90 international and national publications in reputed journals and conferences to his credit. He is also a member of editorial board of four international peer reviewed journals. His field of expertise is Big Data, ANN, Biometrics, RFID, Cloud Computing and VANETS.