ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

Mat Lab Based Synthesis of Speech & Speaker Reorganization Using Data Driven Approach

Leela Kumari [#1], Aakash Dadhich [*2], Rahul Guha [*3]

Assistant Professor, Dept. of CSE, Dr. Radhakrishnan Institute of Technology, Rajasthan, India [#1]/[*3]

Sri Balaji College of Engg., Jaipur, India [*2]

ABSTRACT: If you could talk to your Personal Computer then it would provide a comfortable and natural form of communication. It will reduce the typing amount, and it would be effortless, and allow you to move away from the terminal or screen. It is not necessary that you will not have to be in the line of sight of the terminal. As it will help you in many cases to tell you that who was speaking. If i want to use voice as a new medium on a computer workstation, it is natural explore how speech recognition can contribute to such an environment. Now we will review the state of speech and speaker recognition, focusing on current technology applied to personal workstations. The objective of this paper is to provide explanation of the speech and speaker recognition data input of the user. An algorithm of speech and speaker recognition which efficiently determines the optimum coordination of H.M.M has been successfully designed with the help of MATLAB processing.

KEYWORDS: An automation system, Speech & speaker identification, Speech processing.

I. INTRODUCTION

Speech recognition, the capability to find out the words spoken, and recognize the actual speaker, it is the ability to identify who is saying them, will become commonplace applications of speech processing technology. All the forms of the recognition of speech are available for personal workstations. Now a day the interest in speech recognition has increased more, and improving its performance. Speech recognition has proved very useful for certain applications, such as telephone voice-response systems for the selection of services, digit recognition for cellular phones, and data entry while walking around a railway yard or clambering over a jet engine during an inspection. Nonetheless, comfortable and natural communication in a general setting (no constraints on what you can say and how you say it) is beyond us for now, posing a problem too difficult to solve. Fortunately, we can simplify the problem to allow the creation of applications like the examples just mentioned. Some of these simplifying constraints are discussed in the next section. Speaker recognition is related to work on speech recognition. Instead of determining what was said, you determine who said it. Deciding whether or not a particular speaker produced the utterance is called *verification,* and choosing a person's identity from a set of known speakers is called *identification.* The most general form of speaker recognition (text- independent) is still not very accurate for large speaker populations, but if you constrain the words spoken by the user (text-dependent) and do not allow the speech quality to vary too wildly, then it too can be done on a workstation. Speech processing comes as a front end to a growing number of language processing i.e., it is a diverse field with many applications. In this Paper, I have focus on phonemes and allophones in order to try to provide a deeper insight into the kinds of issue involved in the processing of speech.

II. RELATED WORK

This thesis entails the design of a speaker recognition code using MATLAB. I give a brief survey of ASR, starting with modern phonetics, and continuing through the current state of *Speech Recognition*. A simple computer experiment, using MATLAB, into *speech recognition* is described in some detail. I experimented with some training and testing data. A code was constructed to compare a known speech file to unknown speech files and choose the top matches. Signal processing in the time and frequency domain yields a powerful method for analysis. MAT Lab's built in functions for frequency domain analysis as well as its straightforward programming interface makes it an ideal tool for speech analysis projects. For the current thesis, hence domain signals. Degradation of signals by the application of Gaussian noise is performed. Background noise was successfully removed from a signal by the application of a 3rd order Butterworth filter. A code was constructed for only ten recordings. Only 2 sec are allotted for the execution of the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

program and checking of data base. Authentication of the user can be determined by the threshold value being set by the standard variance.

1.2 Modules of speaker recognition

A speaker recognition system is mainly composed of the following four modules:

1.2.1 Front-end processing: It is the "signal processing" part, which converts the sampled speech signal into set of feature vectors, which characterize the properties of speech that can separate different speakers. is generally accomplished by digital signal processing hardware on the computer's *sound card*. Front-end processing is performed both in training- and recognition phases.

1.2.2 Speaker modelling: This part performs a reduction of feature data by modelling the distributions of the feature vectors.

1.2.3 Speaker **database:** The speaker models are stored here.

1.2.4 Decision logic: It makes the final decision about the identity of the speaker by comparing unknown feature vectors to all models in the database and selecting the best matching model. DTW is a non-linearly approach to normalizing time-scales i.e., it includes:

1.2.4.1 Classification

A library of feature vectors is provided – a "training" process usually builds it. The classifier uses the spectral feature set (feature vector) of the input (unknown) word and attempts to find the "best" match from the library of known words. Simple classifiers, like mine, use a simple "correlation" metric to decide. While advanced recognizers employ classifiers that utilize *Hidden Markov Models* (HMM), *Artificial Neural Networks* (ANN), and many others. The most sophisticated ASR systems search a huge database of all "possible" sentences to find the best (most probable) match to an unknown input.

6.3 Problem Formulation

Speech is a complicated signal produced as a result of several transformations occurring at several different levels: semantic, linguistic, articulator, and acoustic. Differences in these transformations appear as differences in the acoustic properties of the speech signal. Speaker-related differences are a result of a combination of anatomical differences inherent in the vocal tract and the learned speaking habits of different individuals. In speaker recognition, all these differences can be used to discriminate between speakers .The break up of the speaker recognition as a forensic investigation problem is given by the following two sub-problems:

6.3.1 Speaker Identification

Given a set of n suspected speakers {S1,., Sn} and a disputed anonymous speech segment , the investigator is asked to identify the speaker S ، {S1, ., Sn} of the anonymous speech.

6.3.2 Generic Speaker Verification

The general approach to ASV consists of five steps: digital speech data acquisition, feature extraction, pattern matching, making an accept/reject decision, and enrolment to generate speaker reference models. A block diagram of this procedure is shown in Figure 8.3. Feature extraction maps each interval of speech to a multidimensional feature space. (A speech interval typically spans 10 to 30 ms of the speech waveform and is referred to as a frame of speech.) This sequence of feature vectors x*i* is then compared to speaker models by pattern matching. This results in a match score *zi* for each vector or sequence of vectors.

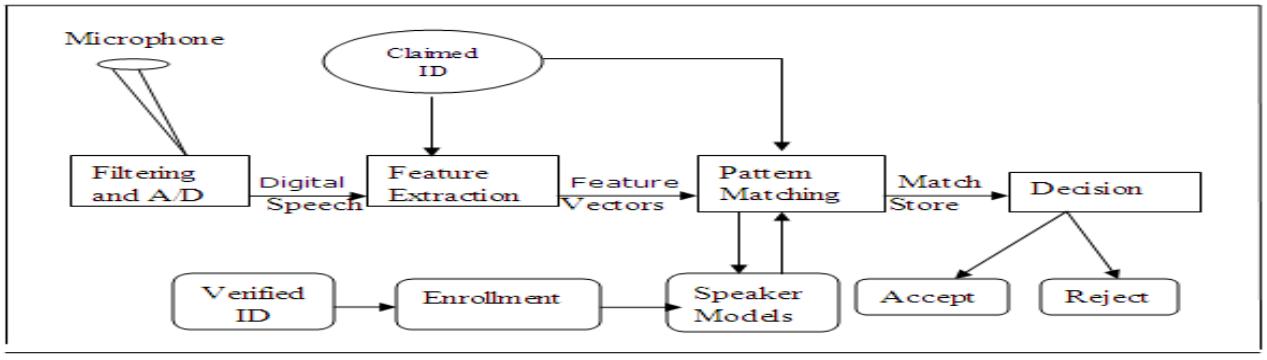


Fig: 1:- Speech recognition automation architecture

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

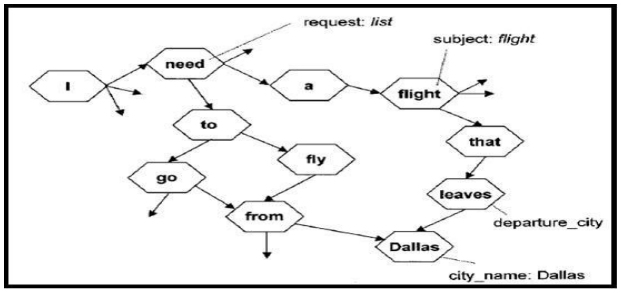


Fig.1.1- Finite state network

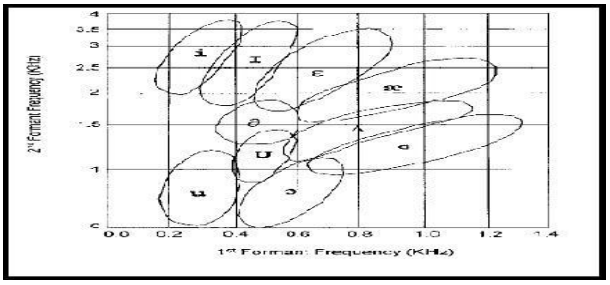


Fig. 1.2 - Distribution of vowels in F1-F2 plane

III. A SPEECH RECOGNITION ALGORITHM

The algorithm must compute a measure of goodness-of-fit between the pre-processed signal from the user's speech and all the stored templates or speech models. A selection process chooses the template or model (possibly more than one) with the best match.

Two major types of pattern matching in use are template matching by dynamic time warping and hidden Markov models. Artificial neural networks applied to speech recognition have also had some success, but this work is still in the early stages of research. Moreover, linguistic knowledge incorporated into the pattern-recognition algorithm can enhance performance.

The pattern-matching task of speaker verification involves computing a match score, which is a measure of the similarity between the input feature vectors and some model. Speaker models are constructed from the features extracted from the speech signal. To enrol users into the system, a model of the voice, based on the extracted features, is generated and stored (possibly on an encrypted smart card). Then, to authenticate a user, the matching algorithm compares/scores the incoming speech signal with the model of the claimed user. There are two types of models: stochastic models and template models. In stochastic models, the pattern matching is probabilistic and results in a measure of the likelihood, or conditional probability, of the observation given the model. For template models, the pattern matching is deterministic. The observation is assumed to be an imperfect replica of the template, and the alignment of observed frames to template frames is selected to minimize a distance measure d. The likelihood *L* can be approximated in template-based models by exponentiation the utterance match scores-

$$L = \exp(-a\, d)$$

Where *a* is a positive constant (equivalently, the scores are assumed to be proportional to log likelihoods). Likelihood ratios can then be formed using global speaker models or cohorts to normalize *L*.The template model and its corresponding distance measure is perhaps the most intuitive method. The template method can be dependent or independent of time. An example of a time-independent template model is VQ modeling. All temporal variation is ignored in this model and global averages (e.g., centroids) are all that is used. A time-dependent template model is more complicated because it must accommodate human speaking rate variability.

IV. LITERATURE SURVEY

Development of speech recognition systems began as early as the 1960s with exploration into voiceprint analysis, where characteristics of an individual's voice were thought to be able to characterize the uniqueness of an individual much like a fingerprint. The early systems had many flaws and research ensued to derive a more reliable method of predicting the correlation between two sets of speech utterances. Speaker identification research continues today under the realm of the field of digital signal processing where many advances have taken place in recent years. The first speech recognition system having creditable performance was built in 1952 at Bell labs using acoustic features to recognize digits spoken by single speaker. But these techniques and results could not be extrapolated towards larger and more sophisticated systems. The research has been carried out with acoustic phonetic approach in mid 1970's. The advanced research project agencies support of speech understanding research has lead to significantly increased level of activity in this area since 1971. Several isolated and connected speech recognition systems have been developed and demonstrated.. For speech signal processing fast Fourier transform, cepstral analysis, and linear predictive coding were started to be used. New techniques for pattern matching like DTW (dynamic time warping) and HMM were invented.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

Linde, Buzo and Gray presented an efficient and intuitive algorithm for the design of vector quantizes. Gaussian mixture techniques were also used to provide a possible useful extension to the popular EM (expectation maximization) algorithm [6]. To improve generalization capabilities of HMMs in many practical classification problems large margin HMM based classifiers are used.

V. IMPLEMENTATION

During the first experiment a program has been written in MATLAB to verify the characteristic of speech and speaker recognition. During the program first we enter the name to be verified, which should be done within two second. There is a choice of entering the data voice again also, by pressing 1 else the program will verify the voice with the stored data. During the first program we have speech the correct word, and the output is as:

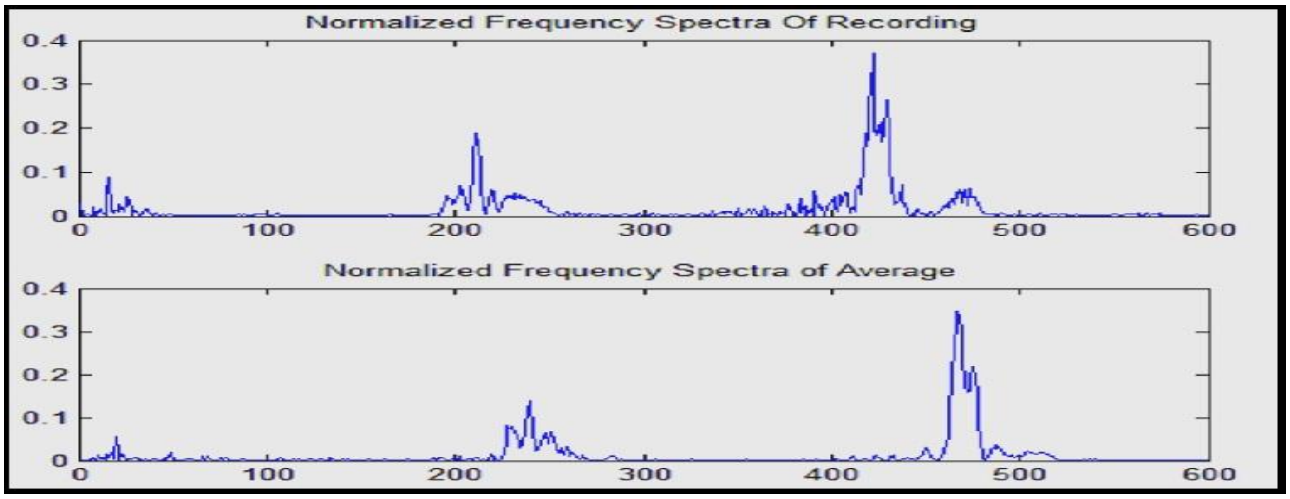
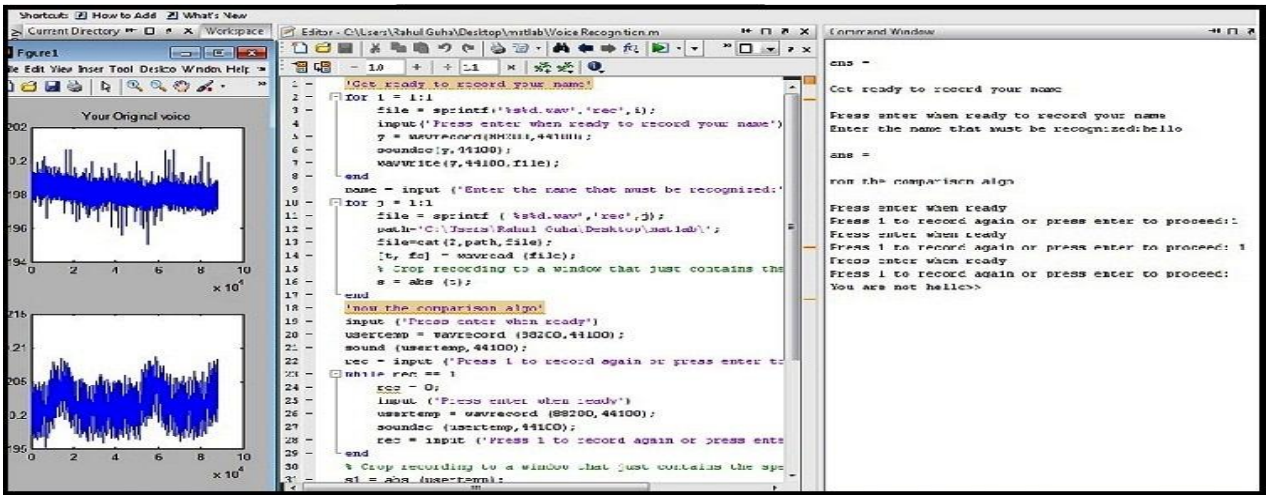


Fig:2 :- Voice match with pre recorded voice



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

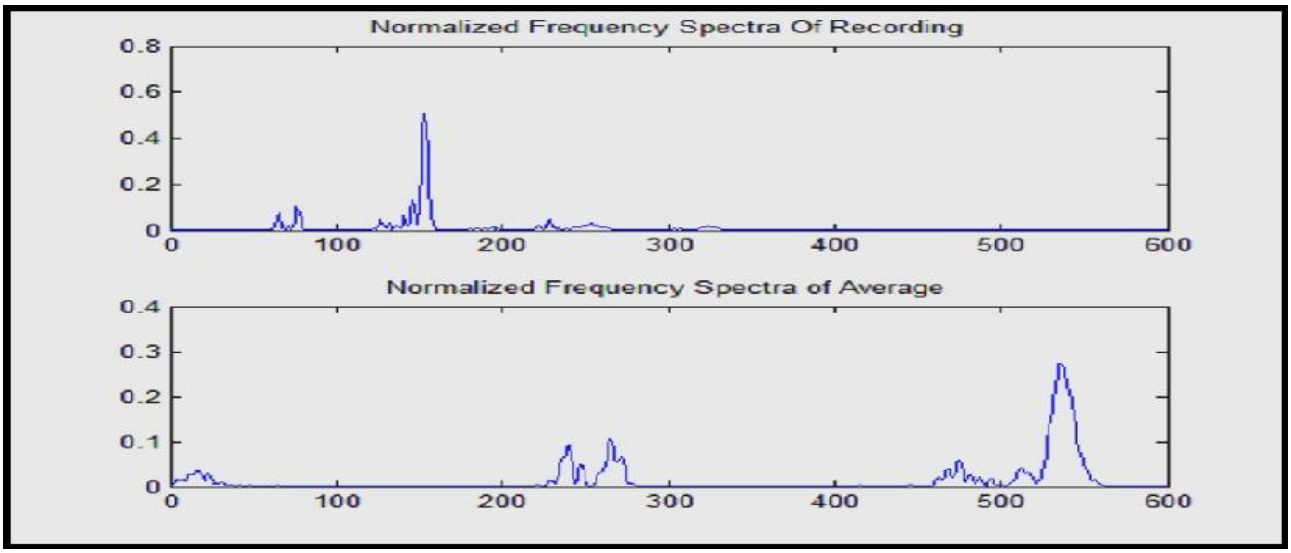


Fig: When voice does not match with pre recorded voice.

VI. CONCLUSION AND FUTURE WORK

The main objective of this thesis is to provide some explanation of the speech and speaker recognition data input of the user. An algorithm which efficiently determines the optimum coordination of H.M.M has been successfully designed with the help of MATLAB. Authentication of the user can be determined by the threshold value being set by the standard variance.

Over three decades of research in spoken language processing have produced remarkable advances in automatic speech recognition and understanding that helps us take a big step toward natural human–machine communication. Signal-processing techniques led to a better understanding of speech characteristics, providing deep insights into acoustic-phonetic properties of a language. The introduction of a statistical framework not only makes the problem of automatic recognition of speech tractable but also paves the road to practical engineering system designs. It was found that a particular probabilistic measure, the HMM, provides a speech modelling formalism that is powerful and yet easy to implement. Coupled with a finite state representation of a language, hidden Markov modelling has become the underpinning of most of today's speech-recognition and understanding systems under deployment. To accomplish the ultimate goal of a machine that can communicate with people, however, a number of research issues are awaiting further study. Such a communicating machine needs to be able to deliver a satisfactory performance under a broad range of operating conditions and have an efficient way of representing, storing, and retrieving "knowledge" required in a natural conversation. With the current enthusiasm in research advances, we are optimistic that the Holy Grail of natural human–machine communication will soon be within our technological reach.

REFERENCES

[1] K. Lee, Automatic, Speech Recognition: the development to f the Sphinx System, Kluwer Academic Publishers, Norwell, Mass. 1989 .
[2] L. R. Bahl et al., "Speech Recognition of a Natural Text Read as Isolated Words," Proc. IEEE IntI Conf. Acoustics, Speech, and Signal Processing, April 1981, pp. 1,168-1,17 1.
[3] D.O. Kimbal et al., "Recognition Performance and Grammatical Constraints," Proc. DARPA Speech Recognition Workshop, Feb. 1986, pp. 53-59.
[4] D. O'Shaughnessy, Speech Communication: Human and Machine, Addison- Wesley, Reading, Mass., 1987.
[5] M.A. Franzini, M.J. Witbrock, and K.-F. Lee, "Speaker-Independent Recognition of Connected Utterances Using Recurrent and Non recurrent Neural Networks," Proc. Int'l Joint Conf. Neural Networks, V01.2, Washington, DC, June 1989R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.
[6] J. Mariani, "Recent Advances in Speech Processing," Proc. IEEE Intl Conf. Acoustics, Speech, and Signal Processing, Glasgow, Scotland. May 1989, pp. 429-440.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

[7] M.-W. Fung et al., "Improved Speaker Adaptation Using Text-Dependent Spectral Mappings," Proc. IEEE Int'l Conf. Acoustics, Speech. and Signal Processing, New York City, 1988, pp. 131-134.

[8] D.B. Paul, "The Lincoln Robust Continuous Speech Recognizer," Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing, Glasgow, Scotland, 1989, pp. 449- 452.

[9] H. Murveit and M. Weintraub, "1,000-Word Speaker-Independent Continuous-Speech Recognition Using Hidden Markov Models," Proc.

BIOGRAPHY

Leela Kumari is a Assistant professor in the Computer Science Department, Dr. Radhkrishnan Institute Of Technology, Rajasthan Technical University. She received Bachelors of Technology in Computer Science degree in 2010 from SBCET, Jaipur, India. Her research interests are Computer Networks (wireless Networks), Speech Recognition, Algorithms, etc.

Aakash Dadhich is a Assistant professor in the Computer Science Department, SBCET, Jaipur, Rajasthan Technical University. He received Bachelors of engineering in Computer Science degree in 2009 from Sobhasaria Engineering College, Sikar, Rajasthan, India and masters of technology from Sobhasaria Engineering College, Sikar, and Rajasthan, India in 2015. His research interests are Software Engineering, Computer Networking, Speech Recognition, Algorithms, etc.

Rahul Guha is a Associate professor in the Computer Science Department, DRIT, Jaipur, Rajasthan Technical University. He received Bachelors of engineering in Computer Science degree in 2011 from AIT, Ajmer, Rajasthan, India and masters of technology from BU, Ajmer, and Rajasthan, India in 2013. His research interests are Mobile Computing, Computer Networking, Speech Recognition, Algorithms, etc.