Volume 4, No. 3, March 2013
Journal of Global Research in Computer Science
ISSN-2229-371X
RESEARCH PAPER
Available Online at www.jgrcs.info

VOICE RECOGNITION USING GAUSSIAN MIXTURE MODEL

Nikhil D. Karande[1], Rohit V. Kumbhar[2], Abhijeet L. Jadhav[3], Sharad G. Bhosale[4], Swapnil S. Patil[5]
[1]Department of Computer Science & Engineering, Bharati Vidyapeeth's College of Engineering, Kolhapur, Maharashtra, India
nikhilkarande18@gmail.com[1]
[2]Department of Computer Science & Engineering, Bharati Vidyapeeth's College of Engineering, Kolhapur, Maharashtra, India
rohitkumbhar167@gmail.com[2]
[3]Department of Computer Science & Engineering, Bharati Vidyapeeth's College of Engineering, Kolhapur, Maharashtra, India
abhi.j.shine@gmail.com[1]
[4]Department of Computer Science & Engineering, Bharati Vidyapeeth's College of Engineering, Kolhapur, Maharashtra, India
bhosalesharad@gmail.com[4]
[5]Department of Computer Science & Engineering, Bharati Vidyapeeth's College of Engineering, Kolhapur, Maharashtra, India
Patil.swapnil513@gmail.com[5]

Abstract: Speech conveys several levels of information. On a primary level, speech conveys the words or message being spoken, but on a secondary level, speech also reveals information about the speaker. In this article we present an overview of our research efforts in an area-automatic speaker recognition. We base our approach on a statistical speaker-modelling technique that represents the underlying characteristic sounds of a person's voice. Using these models, we build speaker recognizers that are computationally inexpensive and capable of recognizing a speaker regardless of what is being said. Performance of the systems is evaluated for a wide range of speech quality; from clean speech to telephone speech, by using several standard speech corpora.

Keywords: MFCC, artificial neural: ann , MandelEllis, perceptron

INTRODUCTION

Tasks that are easily performed by humans, such as face or speech recognition, prove difficult to emulate with computers.Speaker recognition technology stands out as one application in which the computer outperforms the humans.For over six decades, scientists have studied the ability of human listeners to recognize and discriminate voices. By establishing the factors that convey speaker-dependent information, researchers have been able to improve the naturalness of synthetic and recorded speech and assess the reliability of speaker recognition for forensic science applications. Soon after the development of digital computers, research on speaker recognition turned to developing objective techniques for automatic speaker recognition, which quickly led to the discovery that simple automatic systems could outperform human listeners on a similar task.

Over the last three decades, researchers have developed increasingly sophisticated automatic speaker recognition algorithms, and the performance of these algorithms in more realistic evaluation speech corpora has improved. Today, task-specific speaker-recognition systems are being deployed in large telecommunications applications. For example, in 1993 the Sprint Corporation offered the Voice phone card calling card, which uses speaker recognition to allow access to its long-distance network. The general task of automatic speaker recognition is far from solved, however, and many challenging problems and limitations remain to be overcome. In this article we present an overview of the research, developments, and evaluation of automatic speaker recognition systems at Lincoln Laboratory. In this project we are going to implement the mechanism which will recognize the voice of particular speaker among the group of people. We record the conversation among the group of people and extract the features of voices of those speakers. These extracted features are stored in database. Afterwards to find the particular speaker we do the training and testing on input voice depends on speaker is known or unknown respectively and gives output.

METHODOLOGY

This task of speaker recognition involves two tasks: identification and verification, as shown in Figure 1. In identification, the goal is to determine which voice in a known group of voices best matches the speaker. In verification, the goal is to determine if the speaker is who he or she claims to be. In speaker identification, the unknown voice is assumed to be from the predefined set of known speakers. For this type of classification problem-an N alternative, forced-choice task-errors are defined as misrecognitions (i.e., the system identifies one speaker's speech as coming from another speaker) and the difficulty of identification generally increases as the speaker set (or speaker population) increases.

Applications of pure identification are generally unlikely in real situations because they involve only speakers known to the system, called enrolled speakers. However, one indirect application of identification is speaker-adaptive speech recognition, in which speech from an unknown speaker is matched to the most similar-sounding speaker already trained on the speech recognizer. Other potential identification applications include intelligent answering machines with personalized caller greetings and automatic speaker labeling of recorded meetings for speaker-dependent audio indexing. Speaker verification requires distinguishing a speaker's voice known to the system from a potentially large group of voices unknown to the system. Speakers known to the system who claim their true identity are called *claimants;* speakers, either known or unknown to the system, who pose as other speakers are called *impostors.*

There are two types of verification errors: false acceptances-the system accepts an impostor as a claimant; and false rejections-the system rejects a claimant as an impostor. Verification forms the basis for most speaker-recognition applications. Current applications such as computer log-in, telephone banking, calling cards, and cellular-telephone fraud prevention substitute or supplement a memorized personal identification code with speaker verification.

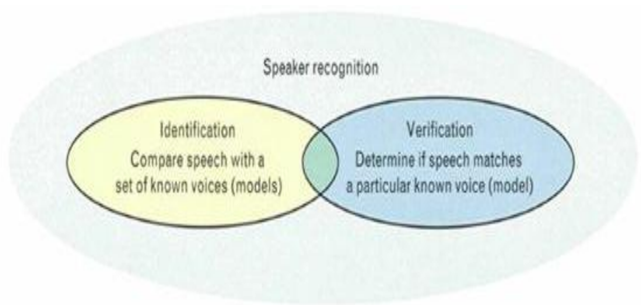

Figure 1: Speaker recognition task

PROPOSED WORK

For over six decades, scientists have studied the ability of human listeners to recognize and discriminate voices [1]. By establishing the factors that convey speaker dependent information, researchers have been able to improve the naturalness of synthetic and recorded speech [2] and assess the reliability of speaker recognition for forensic science applications [3]. Soon after the development of digital computers, research on speaker recognition turned to developing objective techniques for automatic speaker recognition, which quickly led to the discovery that simple automatic systems could outperform human listeners on a similar task [4]. Today, task-specific speaker-recognition systems are being deployed in large telecommunications applications. For example, in 1993 the Sprint Corporation offered the Voice Phone Card calling card, which uses speaker recognition to allow access to its long-distance network.

With the advancement of automated system the complexity for integration & recognition problem is increasing. The problem is found more complex when processing on randomly varying analog signals such as speech signals. Although various methods are proposed for efficient extraction of speech parameter for recognition, the MFCC method with advanced recognition method such as HMM is more dominant used. This system found to be more accurate under low varying environment but fails to recognition speech under highly varying environment. This need to the development of an efficient recognition system which can provide is efficient varying system. Research and development on speaker recognition method and technique has been undertaken for well over four decade and it continues to be an active area. Approaches have spanned from human auditory [5] and spectrogram comparisons [5], to simple template matching, to dynamic time-warping approaches, to more modern statistical pattern recognition [6], such as neural networks and Hidden Markov Model (HMM‟ s) [7].

SYSTEM IMPLEMENTATION

We proposed a system which implements following things which are shown in below figure 2.

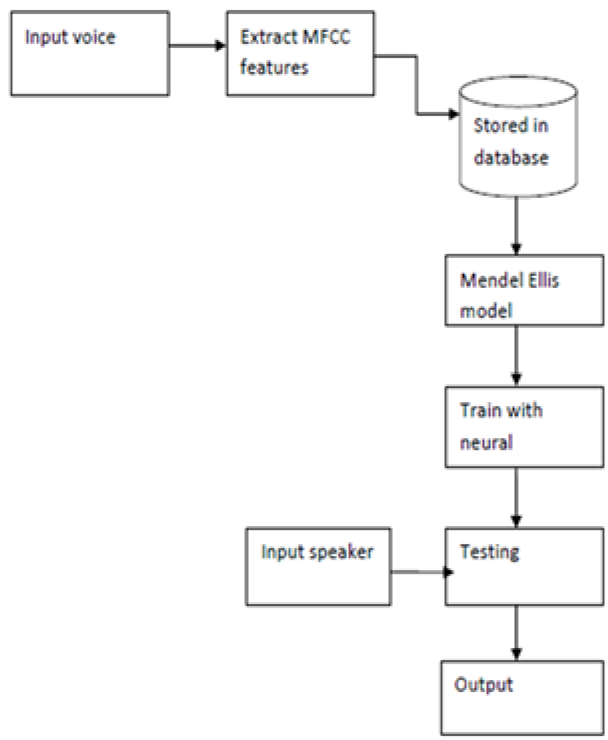

Figure 2: System flow and architecture

Input Voice:

We take input as voice of person. This recorded file is saved in .wav file format. For further feature extraction process.

MFCC:

MFCC (Mel Frequency Cepstral Coefficient) which is used for extraction of features from recorded voice input. These extracted features are different for each person. These features are stored in database.

Database:

In database we stored features of single voice input in separate model which is separate for all extracted features of single input. Then this model is subjected to MndelEllis model to calculate mean.

MandelEllis model:

This is algorithm, which we uses for calculating mean of features of single model. Then after this calculated mean is again stored in database for ease training and testing of voices.

Training with neural:

In this we train the model to identify the voices of known speakers which are known to system. For this we compare the calculated mean with stored mean of models to identify speaker. And as said speaker is known to system we got claimed speaker.

Testing:

It might happen that the speaker is not known to system then we output as unknown speaker but before we give output, we store that calculated mean for further use of system. If that user again uses the system, it got output as claimed person.

RESULTS AND DISCUSSION

Earlier projects were fail to identify speaker among the group of peoples and were not suitable to perform well if size of input voice is large. In our project we overcomes these difficulties in more extent. Identification accuracy for a population size was computed by performing repeated speaker-identification experiments on fifty sets of speakers randomly selected from the pool of 630 available speakers and averaging the results. This procedure helped average out the bias of a particular population composition. Population sizes of 10, 100,200,300,400,500,600, and 630 were used.

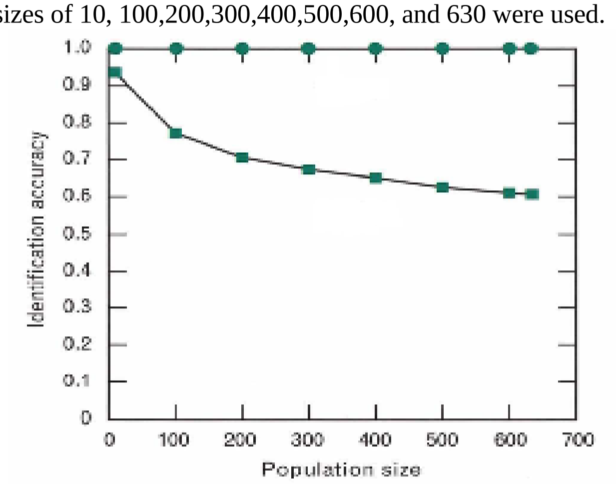

Figure 3: Results for identification accuracy

CONCLUSION

By doing the proposed work of implementation of first part of project we conclude that we have reviewed the research, development, and evaluation of automatic speaker-recognition systems. Starting from the speaker-dependent vocal-tract information conveyed via the speech spectrum, we outlined the development of a statistical speaker-model approach to represent the underlying characteristic vocal-tract shapes of a person's voice. With a text-independent assumption, this statistical speaker model leads to the Gaussian mixture speaker model that serves as the basis for our speaker identification and verification systems. The Gaussian mixture model provides a simple yet effective speaker representation that is computationally inexpensive and provides high recognition accuracy on a wide range of speaker recognition tasks.

REFRENCES

[1]. F. McGehee, "The Reliability of the Identification of the Human Voice," J General Psychology 17, 249 (1937)

[2]. W D. Voiers, "Perceptual Bases ofSpeaker Identity," J Acoust. Soc. Am. 36, 1065 (1964).

[3]. G.R. Doddington, "Speaker Recognition-IdentifyingPeople by Their Voices," Proc.IEEE73, 1651 (1985).

[4]. AE. Rosenberg, "Listener Performance in Speaker-Verification Tasks," IEEE Trans. Audio Electroacoust. AU-21, 221 (1973).

[5]. Allen, J.B.: How do humans process and recognize speech IEEE Trans. on Speech and Audio Processing, vol. 2, no. 4, pp.567--577 (1994.)

[6]. Kim W. Kang, S. and Ko, H. : Spectral subtraction based on phonetic dependency and masking effects, IEEE Proc.-Vision, Image and Signal Processing, 147(5), pp 423--27 (2000)

[7]. R.J. Elliott, L. Aggoun and J.B.: Moore Hidden Markov Models: Estimation and Control", Springer Verlag, (1995)

Short Bio Data for the Author

N. D. Karande received the B.E. degree in Computer Science and Engineering from Bharati Vidyapeeth's College of Engineering, Kolhapur, India in 2006. He has completed his MTech in Computer Science and Technology at Shivaji University, Kolhapur, India. In 2009, he is working as Professor at Bharati Vidyapeeth's College of Engineering, Kolhapur, India. He has published various papers in the area of Network Security and Natural Language Processing.

R. V. Kumbhar currently persuing the B.E. degree in Computer Science and Engineering from Bharati Vidyapeeth's College of Engineering, Kolhapur, India He has completed his diploma in Computer Science from D.Y.Patil College of Engineering, at Shivaji University, Kolhapur, India.

A.L. Jadhav currently persuing the receiving the B.E. degree in Computer Science and Engineering from Bharati Vidyapeeth's College of Engineering, Kolhapur, India. He has completed his H.S.E. in B.N.N. College at Bhiwandi, Thane.

S. G. Bhosale currently persuing the B.E. degree in Computer Science and Engineering from Bharati Vidyapeeth's College of Engineering, Kolhapur, India. He has completed his diploma in Computer Science and Engineering.

S. S. Patil currently persuing the B.E. degree in Computer Science and Engineering from Bharati Vidyapeeth's College of Engineering, Kolhapur, India. He has completed his diploma in Computer Science and Engineering.