INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN: 2278 – 0211 (Online)

Speaker Identification System

Mr. Kadam Sudhir A.
department of Electronics & Telecommunication, Bharati Vidyapeeth's College of Engineering, Morewadi, Kolhapur, Maharashtra, India
Mr. Gurav Ramchandra K.
department of Electronics & Telecommunication, Rajarambapu Institute of Technology, Sakharale, Sangli, Maharashtra, India

Abstract:

Human speech is our most natural form of communication and conveys both meaning and identity. The identity of a speaker can be determined from the information contained in the speech signal through speaker identification. Speaker identification is concerned with identifying unknown speakers from a database of speaker models previously enrolled in the system. We are using text dependent speaker identification i.e. speaker will have to speak predefined password. The general process of speaker identification involves two main stages. The first stage extracts features from speakers. And second stage involves processing the identity of a speaker using features extracted from the speech. Several techniques available for feature extraction including Linear Predictive Coding (LPC), Mel-Frequency Cepstral Coefficients and LPC Cepstral coefficients. These features are used with a classification technique to create a speaker model. In this project we are using the Mel Frequency Cepstral Coefficients (MFCC) technique to extract features from the speech signal and compare the unknown speaker with the exits speaker in the database. For matching we are using Vector Quantization which is commonly used in speaker identification producing reliable results.

Keywords: *Human speech, Mel Frequency Cepstral Coefficients, Linear Predictive Coding (LPC), Database, Vector Quantization*

1. Introduction

Speaker recognition is a branch of biometric authentication which refers to the automatic identity recognition of individuals using certain intrinsic characteristics of the person. Biometric authentication has been an important technique for human-machine communication system in applications with security consideration. Besides the voice, there are many other physical and behavioral patterns, e.g. eyes, face, fingerprint, signature, etc., for biometric authentication. Practically, selection of a promising biometric pattern should take into account at least the following concerns: robustness, distinctiveness, accessibility, and acceptability. The judgment of a *good* biometric pattern is very complicated and depends on the specifics of the applications.

Among all the biometric authentication technologies, speaker recognition is probably the most natural and economical one for human-machine communication systems due to (1) speech data collection is much more convenient than other patterns; and (2) more importantly, speech is the dominant mode of information exchange for human beings and it tends to be the dominant mode for human-machine information exchange. The development of speech processing technology has boosted many applications of speaker recognition, especially in the following areas:

- Access control to physical facilities or data networks.
- Telephone credit card purchases or other bank transaction.
- Information retrieval, e.g. customer information for call centers and audio indexing.
- Remote monitoring.
- Forensic voice sample matching.

2. Speaker Recognition

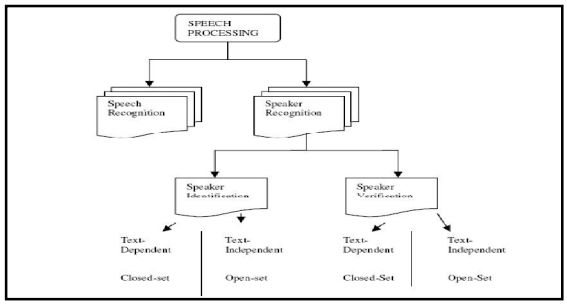

Figure 1: Speaker recognition

The verification task is to decide whether or not an unlabeled voice belongs to a claimed speaker. There are only two possible decisions: either to accept the voice as belonging to the claimed speaker or to reject it as belonging to an impostor.

3. Speaker Verification

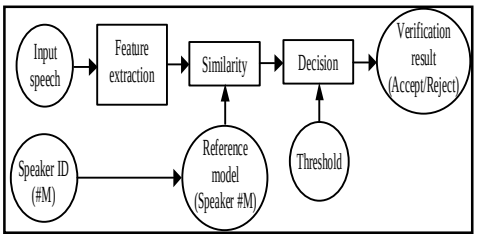

Figure 2: Speaker verification

The identification task is to classify the unlabeled voice as belonging to one of the registered speakers. The number of decision alternatives in speaker identification is the same as the population size *N*, and generally the performance is inversely proportional to *N*. Therefore, it is usually a more difficult task than verification with large *N*. Speaker identification as defined is also called *closed-set* identification. Its contrast, the *open-set* identification encompasses a possibility that the unlabeled voice belongs to none of the registered speakers. Therefore, the number of decision alternatives is *N* +1 which includes a decision that the voice belongs to an unknown speakers. The open-set identification is a combination of identification and verification. Speaker recognition can also be divided into text-dependent and text-independent recognitions. In text-dependent recognition, the system knows exactly the spoken text which could be either fixed phrase or prompted phrase. In text-independent recognition, the system does not know the text of the spoken utterance, which could be user selected keywords or conversational speech.

With the knowledge of spoken text, the system can exploit the speaker individuality associated with specific phonemes or syllables. Thus a text-dependent system generally performs better than the text-independent system. However, it requires highly cooperatives of the speakers and can be used only for applications with strong control over user input. The text-independent system is more user-friendly and more applicable but, without the knowledge of the spoken text, also more difficult to achieve high performance. In text-independent applications, a speech recognizer which provides the

correct text knowledge can improve the speaker recognition accuracy. Although the text-independent task has been accepted as a good platform for evaluating the general technologies for speaker recognition, many commercial and industrial applications focus more on the text-dependent, or text-constraint speaker recognition.

4. Speaker Identification

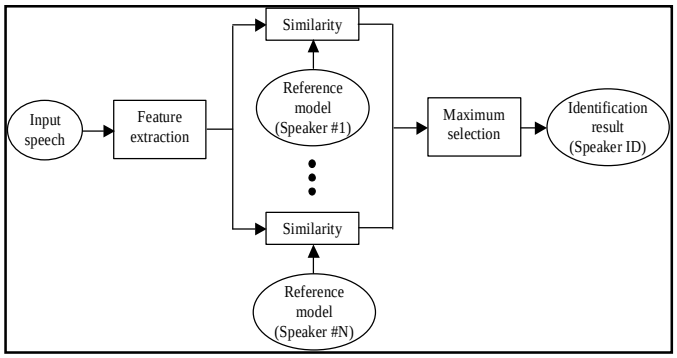

Figure 3:Speaker identification

Speaker identification involves two main stages, the enrolment stage and the identification stage. These phases involve three main parts:

• Pre-Processing.

• Feature Extraction.

• Speaker Modeling.

Speech is recorded by sampling the input, which results in a discrete-time speech signal. Pre-processing is a technique used to make the discrete-time speech signal more amendable for the processes that follow. There are five pre-processing techniques that can be used to enhance feature extraction. These include DC offset removal, silence removal, pre-emphasis, windowing and autocorrelation.

In order to create a speaker profile, the speech signal must be analyzed to produce some representation that can be used as a basis for such a model. In speech analysis this is known as feature extraction. Feature extraction allows for speaker specific characteristics to be derived from the speech signal, which are used to create a speak model. The speaker model uses a distortion measure to determine features which are similar. This places importance on the features extracted, to accurately represent the speech signal.

After extracting speaker-specific characteristics from the speech signal we need a method to classify the speaker in order to determine the author of a given speech signal.

In order for identification a speaker must first be enrolled in the system using a modeling process. Once models for the speakers have been created, a matching or classification process is then used for identification.

5. BLOCK DIAGRAM

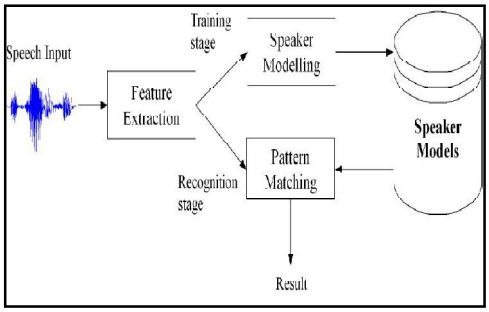

Figure 4: Block diagram

The general process of speaker identification involves two stages. The first stage is to enroll the speakers into the system. Enrolment involves determining distinct characteristics of the speaker's voice, to be used as a source in the modeling process. Speaker models are then created for each of the speakers and stored in a database. The second stage involves the identification of a speaker. Similar to the enrolment stage, this involves extracting distinct features from an unknown speaker to compare with the speaker database. The enrolment and identification processes are very similar, and both require distinct features to be extracted from the speech signal. The identification process depends on the modeling procedure used in the enrolment stage.

In order to construct a speaker identification system, there are two important aspects of the process that require further investigation; namely: feature extraction and classification. Both of these stages have a critical effect on the identification result. Feature extraction is the process of extracting distinct characteristics from the speech of an individual. Classification refers to the process of determining a speaker based upon previously stored models or information.

Pattern matching and speaker modeling are techniques used to classify and enroll speakers to an identification system. Speaker modeling constructs a model of an individual's voice based upon the features extracted from their speech signal. This is completed when speakers are enrolled in the speaker identification system to produce a

database of registered speakers. This occurs through a training stage in which the system creates the speaker model. Pattern matching uses the models in the speaker database to calculate a matching score for each model. The final result is a measure of the similarity between the features extracted from the unknown speech signal and each of the models in the speaker database.

6. Simulation Results

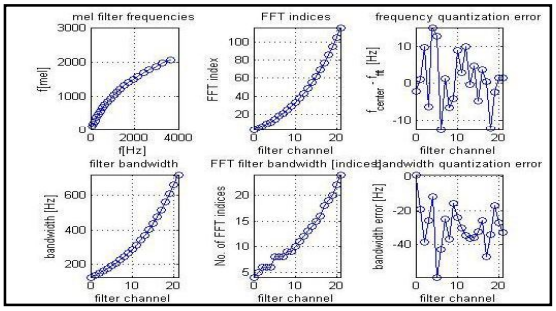

Figure 5: Filter frequency, FFT, Frequency Quantization error & B.W.

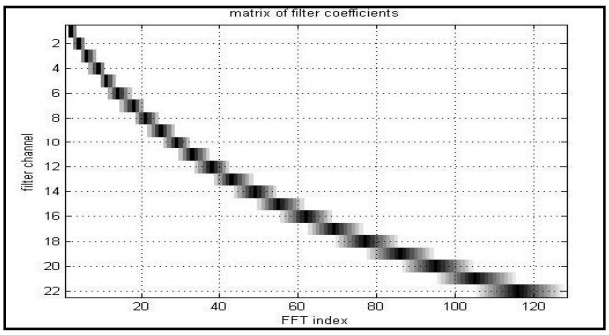

Figure 6: FFT Matrix of filter Coefficients

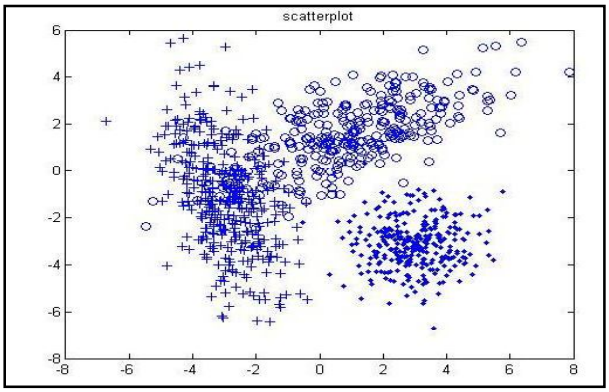

Figure 7: Scatter plot

7 . Conclusion And Future Outlook

The MFCC technique has been applied for speaker Identification. VQ is used to minimize the data of the extracted feature. The study reveals that as number of centroids increases, identification rate of the system increases. It has been found that combination of Mel frequency and Hamming window gives the best performance. It also suggests that in order to obtain satisfactory result, the number of centroids has to be increased as the number of speakers increases. The study shows that the linear scale can also have a reasonable identification rate if a comparatively higher number of centroids is used. However, the recognition rate using a linear scale would be much lower if the number of speakers increases. Mel scale is also less vulnerable to the changes of speaker's vocal cord in course of time. The present study is still ongoing, which may include following further works. HMM may be used to improve the efficiency and precision of the segmentation to deal with crosstalk, laughter and uncharacteristic speech sounds. A more effective

normalization algorithm can be adopted on extracted parametric representations of the acoustic signal, which would improve the identification rate further. Finally, a combination of features (MFCC, LPC, LPCC, Formant etc) may be used to implement a robust parametric representation for speaker identification.

8. Reference

1. Ashish Jain,Hohn Harris,Speaker identification using MFCC andHMM based techniques,university Of Florida,April 25,2004.
2. Cheong Soo Yee and abdul Manan ahmad, Malay Language Text Independent Speaker Vertification using NN - MLP classsifier with MFCC, 2008 international Conference on Electronic Design.
3. Zaidi Razak,Noor Jamilah Ibrahim, emran mohd tamil,mohd Yamani Idna Idris, Mohd yaakob Yusoff,Quranic verse recition feature extraction using mel frequency ceostral coefficient (MFCC),Universiti Malaya.
4. http://www.cse.unsw.edu.au/~waleed/phd/html/node38.html, downloaded on 3rd March 2010.
5. Stan Salvador and Pjilip Chan,FastDTW: Toward Accurate Dynamic Time Warping in Linear time space,Florida Institute of Technology,Melbourne.
6. Chunsheng Fang, From Dynamic time warping (DTW) to Hidden Markov Model (HMM), University of Cincinnati,2009.
7. Johansen F. T., et. al., The COST 249 Multilingual reference recognizer, LREC 2000.