Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 38 (2012) 3122 – 3126

Procedia Engineering

www.elsevier.com/locate/procedia

International Conference on Modelling, Optimisation and Computing (ICMOC 2012)

Applications of Speaker Recognition

Nilu Singh[a], R. A. Khan[b], Raj Shree[b], a*

[a]*Rajiv Gandhi University, Rono Hills, Doimukh, Arunachal Pradesh and 791112, India*
[b]*Babasaheb Bhimrao Ambedkar University, Vidya Vihar, Raebareli Road, Lucknow and 226025, India*

Abstract

For the communication, speech is one of the natural forms. A person's voice contains various parameters that convey information such as emotion, gender, attitude, health and identity. Speaker recognition technologies have wide application areas, The aim of this paper is to provide the some specific areas where Speaker Recognition techniques can be used. Here we discuss three main areas where Speaker Recognition Technique can be used. They are authentication, surveillance and forensic speaker recognition.

Keywords: Speaker Recognition; Authentication; surveillance; forensic speaker recognition.

1. Introduction

Automatic speaker recognition, which is basically the intersection of two areas of computer science, first natural language technologies and the second one is biometrics. Speech signal contains information related to human being such as linguistic information (e.g. language message) and speaker information (e.g. emotional, high anticipation, stressed, inaccuracy and physiological characteristics). From the speech perception point of view, it also conveys information about the environment in which the speech was produced and transmitted [6]. The general area of speaker recognition is divided into three specific tasks. These are authentication, surveillance and forensic speaker recognition. Depending on the applications, the general area of speaker recognition again divided into three specific categories. Those are identification, detection/verification, and segmentation and clustering [8].

* Nilu Singh. Tel.: +91-995-605-2275 ;
E-mail address: nilu.chouhan@hotmail.com .

doi:10.1016/j.proeng.2012.06.363

The goal of the speaker identification task is to determine which speaker out of a group of known speakers produces the input voice sample. In speaker verification, the goal is to determine whether a person is who he/she claims to be according to his/her voice sample. This process is called speaker detection. It is also considered as a true-or-false binary decision problem. Speaker segmentation and clustering techniques are used in multiple-speaker scenarios. In many speech recognition and speaker recognition applications, it is often assumed that the speech from a particular individual is available for processing [4].

Automatic speaker recognition is example of a pattern recognition problem that finds some kind of patterns within some real-world sensor data. For all problems of pattern recognition, a training phase and a testing phase is required. In training phase a user enrols by providing voice samples to the system. The system extracts speaker-specific information from the voice samples to build a voice model of the enrolling speaker. In the testing phase, a user provides a voice sample that is used by the system to measure the similarity of the user's voice to the model of the previously enrolled user and, subsequently, to make a decision. For example, in speaker authentication system, valid users of the system need to be enrolled that will store voice sample in a database. Speech samples of the user are required for the training phase. During the later recognition process, the system compares another recorded speech signal to the training utterance. The expected output of the system can be the name of one of the training speakers, or a rejection of voice [5].

2. What is Speaker recognition?

Speaker recognition could be categorized as speaker identification and speaker verification. In Speaker identification, the obtained features are compared with all the speakers' features which are stored in a voice model database and in Speaker verification the obtained features are only compared with the stored features of the speaker she/he claimed to be. Speaker verification could be either Text-dependent or Text-independent [5]. Text-dependent that is when the same text is spoken on both training and test phases. On the other hand, in Text-independent phase, there is no restriction of voice sample. It could be differ in training and test phases. In real life, Text-independent systems are more commercially attractive than Text-dependent systems because it is harder to mimic an unknown phrase than a known one. Automatic speaker recognition use spectrum-related features based on very short time slices of speech. Speaker models based on such information suffer from a lack of robustness to channel mismatches, and fail to capture longer-range characteristics of how a person talks, including the speaker's word patterns, and patterns in speech prosody such as the timing, pausing, and intonation of speech [1].

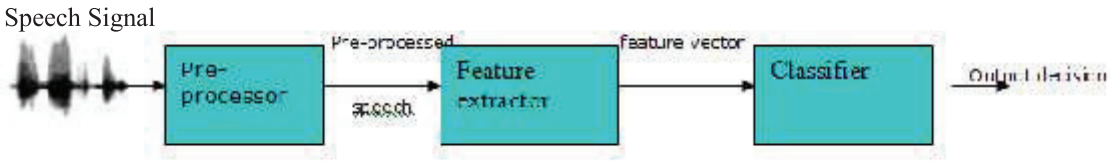

Fig. 1. Block diagram of Speaker Recognition system

Anatomical structure of the vocal tract is unique for every person and hence every person speech has different features. That's why voice information available in the speech signal can be used to identify the speaker. Recognizing a person by her/his voice is known as speaker recognition. Since differences in the

anatomical structure are an intrinsic property of the speaker, voice comes under the category of biometric identity [3]. The selection of appropriate features along with methods to estimate (extract or measure) them is known as feature selection and feature extraction. After the features are extracted a test template is formed. In the classification the test and reference templates are compared and a measurement of similarity is calculated between the two templates. If this measurement is within a threshold value, then the identity claim is accepted, otherwise it is rejected. Based on the threshold value, two types of results are possible, False acceptance and False rejections, In Speaker Recognition cases where voice verification is used to control access to secure buildings, False acceptances (where a false claim is accepted) might result in theft or fraud. Therefore, in order to eliminate them the threshold values need to be high. But this might result in many false rejections (where owner of the identity is rejected) which are undesirable. On the other hand a low threshold value while not giving any false rejections could result in false acceptances. In some classification methods two threshold values are used, that is low and high. If the measurement is below the low value then the claim is accepted and if it exceeds the high value then the claim is rejected. If it is in between the two values then further classification is performed [5][6].

3. Application of Speaker recognition System

Speaker recognition technologies are used in wide application areas. In this research paper, we are discussing about three wide areas as well as related application areas where speaker recognition techniques can be used. These areas are authentication, surveillance and forensic speaker recognition [7]. Depending on these, the area of speaker recognition again divided into three specific categories: identification, detection/verification, and segmentation and clustering Here we list some example applications of speaker recognition technologies.

3.1. Speaker Recognition for Authentication

Speaker recognition for authentication allows the users to identify person using their voices. A person can be identified by various characteristics like signature, fingerprints, voice, facial features etc. This type of authentication methods known as biometric person authentication. In this case, the chance of misused of these type of identity problems are lesser as compared to the key or credit card can be stolen or lost, followed by the PIN number or password can be easily misused or forgotten. Each person has unique anatomy, physiology and learned habits that familiar persons use in everyday life to recognize the person. This can be much more convenient than traditional means of authentication which require to carry a key with you or remember a PIN [1][4].

3.2. Speaker Recognition for Surveillance

Security agencies have several means of collecting information. One of these is electronic eavesdropping of telephone and radio conversations. As these results in high quantities of data, filter mechanisms must be applied in order to find the relevant information. One of these filters may be the recognition of target speakers that are of interest for the service [6].

3.3. *Forensic Speaker Recognition*

It is an important application of speaker recognition. If there is a speech sample that was recorded during the crime. The suspect's voice can be compared with this in order to give an indication of the similarity of the two voices. Proving the identity of a recorded voice can help to convict a criminal or

discharge an innocent in court. Although this task is probably not performed by a completely automatic speaker recognition system, signal processing techniques can be used in this field nevertheless [1][6].
The voice characteristics used in this system during experiments are as follows-

- Not affected due to health problems
- Not susceptible to mimicry
- Differs little from one speaking environment to another
- Not able to consciously controlled by the speaker
- Robust to added noise due to recording and transmission channel

3.4. Security

It is the most obvious application of any biometric authentication techniques. Speaker recognition could be used in credit card transactions as an authentication method combined with some others like face recognition. Speaker recognition technology can provide transaction authentication facility or computer access control, monitoring, telephone voice authentication for long distance calling or banking access etc. [4][1].

3.5. Speech recognition

Speech and speaker recognition are dual research areas in the sense that speaker variability is one of the major problems in speech recognition, whereas in speaker recognition it is an advantage. Speaker recognition technology could be used to reduce the speaker variability in speech recognition systems by speaker adaptation. For instance, speech recognition system could have a speaker gating unit that recognizes who is speaking. Then, the system could adapt its speech recognizer parameters to suit better for the current speaker, or to select a speaker-dependent speech recognizer from its database [4].

3.6. Multi-speaker Tracking

In this several speakers are included in the audio recording. Also it is desired to know who is speaking in a teleconference especially when there are many attendants in the tele-conference and the attendants are not very familiar with each other. Three different type of multi-speaker tasks are recognized - speaker detection, speaker tracking, and speaker segmentation. The detection task consists of deciding whether a known speaker is present in a multi-speaker recording. In the tracking task, a given speaker's speaking intervals are located in the recording. The segmentation task consists of locating the speech intervals of each different speaker. In the most general case, there might be no prior knowledge of the speakers or their number. Applications of speaker segmentation have been proposed for segmentation of news broadcasts [4].

3.7. Personalized user interfaces

Such as voice-mail are becoming more and more popular due to the developments in speech technology in general. By recognizing the speaker, the system could adapt to his/her needs and preferences.

The above applications require robust speaker recognition techniques, e.g. in the telephone-aided services, users may call in under different kinds of acoustic conditions such as in the office, on the street etc. and use different telephone networks like land-line or cellular. In the meeting scenarios, participants may talk while moving around facing the microphone in different directions and different distances.

Mismatched conditions may be encountered at any time in these cases. Therefore robustness is one of the critical factors that decide the success of speaker recognition in these applications [1].

4. Conclusion:

In this paper the aim is to provide information about various applications of Speaker Recognition Technologies. Automatic speaker recognition applications defines which information in the speech signal is relevant, such as the linguistic information will be relevant if the goal is to recognize the sequence of words that the speaker is producing. The presence of irrelevant information like speaker or environment information may actually degrade the system accuracy. From the above discussion, if someone wants to see where Speaker Recognition can be useful, this paper will help in gathering information related to applications.

Reference:

[1] Qin Jin. Robust Speaker Recognition. *Carnegie Mellon University Pittsburgh* 2007. ISBN- 0549270736, 9780549270737.

[2] David Michael Graeme Watts “Speaker Identification - Prototype Development and Performance”. ENG4111/ENG4112 Research Project, November, 2006

[3] Tiwari, Vibha. "MFCC and its applications in speaker recognition." *International Journal on Emerging Technologies 1(1): 19-22(2010)*. ISSN : 0975-8364 19-22.

[4] Kinnunen, Tomi. "Spectral Features for Automatic Text-Independent Speaker Recognition." *LICENTIATE'S THESIS.* (2003): 1-151. Print.

[5] WEE KEONG, THANG, and DR YUAN ZHONG XUAN. "VOICE PRINT ANALYSIS FOR SPEAKER RECOGNITION." . December 21, 2003

[6] The speaker recognition homepage-algorithms, applications and links." *http://www.speaker-recognition.org/navApps.html*.

[7] Mahendran Sriganesh and Kandasamy sugumaran "Speaker Verification in Forensic Applications ." .

[8] M.S.Sinith, Anoop Salim, Gowri Sankar K, Sandeep Narayanan K V, Vishnu Soman "A Novel Method for Text-Independent Speaker Identification Using MFCC and GMM" .

.