Abstract: In this paper the model of automatic Gujarati speaker recognition systems is presented. Automatic Speaker recognition is the process of recognizing a person from a spoken phrase by computer. These systems are generally operates in two modes: to identify a particular person or to verify a person’s claimed identity. The basic components of the automatic Gujarati speaker recognition systems and design tradeoffs are discussed.

Keywords: Access control, authentication, computer security, identification of persons, speaker recognition, speech processing, verification.

1. INTRODUCTION
Speech processing is a diverse field with many applications. Figure-1 shows areas of speech processing and how speaker recognition relates to the other fields [2]. This paper will emphasize the speaker recognition applications shown in the boxes of Figure 1.

The difference between speaker identification and speaker verification is given in Table-1. The user speaks the phrase into a microphone. This speech signal is analyzed by a system and makes decision to accept or reject the user’s identity claim or reports insufficient confidence and request additional input before making the decision.

Table 1: Difference Speaker Verification and Speaker Identification

<table>
<thead>
<tr>
<th>Speaker Verification</th>
<th>Speaker Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is defined as deciding if a speaker is who he claims to be</td>
<td>It is defined as if a speaker is a specific person or is among a group of persons</td>
</tr>
<tr>
<td>A person makes an identity claim (e.g., entering an employee number or presenting his smart card)</td>
<td>A person does not make identity claim</td>
</tr>
</tbody>
</table>

The model for Gujarati Automatic speaker verification system is shown in Figure-2. The user, who previously registered in the system presents smart card containing his identification information and user then speaks a phrase into the microphone. Before verification session, users must register in the system. During registration, voice models are generated and stored on a smart card for use in later verification sessions.

Table 2 lists some of the human and environmental factors that contribute to verification and identification errors. These factors are outside the scope of algorithms and are better corrected by means other than algorithms (e.g., better microphones). These factors are important because, no matter how good a speaker recognition algorithm is, human error ultimately limits its performance.
Table 2: Sources of verification error

<table>
<thead>
<tr>
<th>Error Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misspoken or misread prompted phrases</td>
</tr>
<tr>
<td>Extreme emotional states (e.g., stress or duress)</td>
</tr>
<tr>
<td>Time varying (intra- or intersession) microphone placement</td>
</tr>
<tr>
<td>Poor or inconsistent room acoustics (e.g., multipath and noise)</td>
</tr>
<tr>
<td>Channel mismatch (e.g., using different microphones for enrollment and verification)</td>
</tr>
<tr>
<td>Sickness (e.g., head colds can alter the vocal tract)</td>
</tr>
<tr>
<td>Aging (the vocal tract can drift away from models with age)</td>
</tr>
</tbody>
</table>

2. MOTIVATION

Speaker verification and Speaker identification are the most natural and economical methods for authorize the use of computer and communications systems and multilevel access control. The cost of a speaker recognition system might only be for the software for the recognition algorithm.

Biometric systems automatically recognize a person using distinguishing characteristics. In Speaker recognition you perform a task to be recognized. Speaker-recognition systems can be made somewhat robust against noise and channel variations [7, 9], ordinary human changes, and mimicry by humans and tape recorders [6].

3. PROBLEM FORMULATION

Speech is a complicated signal created as a result of several transformations occurring at several different levels such as semantic, linguistic, articulatory, and acoustic and appeared as differences in the acoustic properties of the speech signal. In speaker recognition, all these differences can be used to discriminate between speakers.

4. MODEL FOR GUJARATI SPEAKER VERIFICATION

The Gujarati Speaker Verification consists of five steps:
1. Digital speech data acquisition
2. Feature extraction,
3. Pattern matching,
4. making decision, and
5. Enrollment to generate speaker reference models.

A block diagram of this procedure is shown in Figure 3 as suggested in [2].

Figure-3: Model for Gujarati speaker verification system.

Feature extraction maps each interval of speech to a multidimensional feature space. This feature vectors \( x_i \) is then compared to speaker models by pattern matching. This results in a match score \( z_i \). The match score measures the similarity of the computed input feature vectors to models of the speaker. Finally, a decision is made to either accept or reject the claimant according to the match score.

4.1. Speech Signal Acquisition

Initially, the acoustic sound pressure wave is transformed into a digital signal suitable for voice processing. A microphone is used to convert the acoustic wave into an analog signal. This analog signal is conditioned with anti-aliasing filtering. The conditioned analog signal is then sampled to form a digital signal by an analog-to-digital (A/D) converter.

4.2. Feature selection and measures

The speech signal can be represented by a sequence of feature vectors. The selection of appropriate features and methods to estimate them are known as feature selection and feature extraction, respectively.

In speaker verification, the goal is to design a system that minimizes the probability of verification errors. Thus, the underlying objective is to distinguish between the given speaker and all others. For an overview of the feature selection and extraction methods, please refer to [2].

4.3. Pattern matching
The pattern-matching involves computing a match score that represents similarity between the input feature vectors and some model. Speaker models are built from the features extracted from the speech signal. To register users, a model of the voice extracted features is generated and stored on an encrypted smart card. 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:

(a) **Stochastic models:** The pattern matching is probabilistic and results in a measure of the likelihood of the observation given the model.

(b) **Template Model:** The pattern matching is deterministic.

### 4.4. Classification and Decision

A verification decision is made whether to accept or reject the speaker or request another utterance. If a verification system accepts an impostor then it is a false acceptance error. If the system rejects a valid user then it is a false rejection error.

The result of decision process can be accept user, continue session, session time-out, or reject user. The decision making procedure is a sequential hypothesis-testing problem [14]. For a brief overview of the decision theory involved, please refer to [2].

### 5. IMPLEMENTATION

The proposed model will be implemented using Modular Audio Recognition Framework (MARF) [15]. It contains a collection of algorithms for Sound, Speech, and Natural Language Processing arranged into a uniform framework to facilitate addition of new algorithms for preprocessing, feature extraction, classification, parsing, etc. implemented in Java. MARF is also a research platform for various performance metrics of the implemented algorithms.

### 6. CONCLUSIONS

Speaker recognition is the use of a computer to recognize a person from a spoken phrase. It can be used to identify a particular person or to verify a person’s claimed identity. In this paper speech processing and general model for Gujarati speaker recognition were discussed. It also discussed sources of verification errors.

### REFERENCES


Authors:

Dr. Himanshu N. Patel received the B.E (IC) from Bhavnagar University, M.C.A degrees from IGNOU and PhD in Computer Science from Sardar Patel University in 1998, 2004 and 2012, respectively. He has qualified SET and NET Examination conduct by UGC in 2006 and 2008 respectively. He has 8 years of teaching experience and currently working as Assistant Professor at Anand Institute of Information Science, Anand. His publication includes 2 papers in international journal, 2 papers in national journals and 10 papers in national level conferences, His research interest includes the areas of Speech recognition, Open source technology and object oriented technology.

Dr. Paresh V Virparia is working as Professor in the Department of Computer Science of Sardar Patel University, Vallabh Vidyanagar. He has completed his MCA in 1989 and Ph.D. in 2002 from Sardar Patel University. He is recognized Ph.D. guide in Computer Science at various universities. Two students have completed their Ph.D. under his guidance. Currently, SEVEN students are doing Ph.D. under his guidance. Two students have completed their M.Phil. (Comp. Sc.) under his supervision. His publication includes 12 papers in International Journal, 8 papers in National Journals, and 30 papers in national conferences/seminars. His research interest includes the areas of Computer Simulation & Modeling, Networking, and IT enabled services.