SPEAKER VERIFICATION

PROJECT REFERENCE NO.: 39S_BE_1020

COLLEGE: P D A COLLEGE OF ENGINEERING, KALABURAGI
BRANCH: DEPARTMENT OF ELECTRONICA AND COMMUNICATION ENGINEERING
GUIDE: MR. ARUN KUMAR N KANTHI
STUDENTS: MS. ARUNDATI C SURPUR
MS. ANNAKURNA S MEKALE

KEYWORDS:
MFCC, Gaussian mixture model, LLP.

INTRODUCTION:
Speaker recognition refers to recognizing persons from their voice. No two individuals sound identical because their vocal tract shapes, larynx sizes, and other parts of their voice production organs are different. In addition to these physical differences, each speaker has his or her characteristic manner of speaking, including the use of a particular accent, rhythm, intonation style, pronunciation pattern, choice of vocabulary and so on. State-of-the-art speaker recognition systems use a number of these features in parallel, attempting to cover these different aspects and employing them in a complementary way to achieve more accurate recognition. The general field of speaker recognition includes two fundamental tasks: speaker identification and speaker verification. Speaker identification involves classifying a voice sample as belonging to one of a set of reference speakers (possible outcomes), whereas speaker verification involves deciding whether or not a voice sample belongs to a specific reference speaker (two possible outcomes the sample is either accepted as belonging to the reference speaker or rejected as belonging to an impostor).

OBJECTIVES:
Objective of this project is to develop a text-independent speaker verification system.

METHODOLOGY:
The Methodology of speaker verification includes two phases, they are Training phase & Testing phase.

Training Phase:
In the Training phase, the background model is created. The background model is basically a large pool of all sample data, just a large Gaussian mixture model. Then it can be converted the wave files into a different format so that they can be used for this analysis. The wave file is a continuous signal, which must be broken down in discrete parameter vectors. Each vector is about 10ms long, because we assume that in this duration the vector is stationary.
Training Phase

![Block diagram of speaker verification](image)

**Fig. 1:** Block diagram of speaker verification

**Database description:** The result is evaluated for three databases, they are

1. TIMIT
2. NTIMIT
3. PDA Database

<table>
<thead>
<tr>
<th>Database</th>
<th>Male speakers</th>
<th>Female speakers</th>
<th>Total</th>
<th>Sampling frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIMIT</td>
<td>438</td>
<td>192</td>
<td>630</td>
<td>16000</td>
</tr>
<tr>
<td>NTIMIT</td>
<td>166</td>
<td>407</td>
<td>573</td>
<td>16000</td>
</tr>
<tr>
<td>PDA database</td>
<td>100</td>
<td>100</td>
<td>200</td>
<td>16000</td>
</tr>
</tbody>
</table>

**Threshold calculation:**

The purpose of this system is to test whether a given voice print belongs to the person the speaker claims to be. In order to achieve this, one should need to device a method to calculate a threshold value, which would make it easy to identify the speaker.

To do this, one should use some train files from each user, then compared the each file of the speaker to the speaker model, and based on the matching of the features the threshold value will be calculated.

**Testing phase:**

Once all the threshold values and speaker models are obtained, it is time to test the remaining files. The input speech is compared with the stored reference model and verification decision is made.
RESULTS AND CONCLUSIONS:

<table>
<thead>
<tr>
<th>DATABASE</th>
<th>ACCURACY (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIMIT</td>
<td>88.25</td>
</tr>
<tr>
<td>NTIMIT</td>
<td>41.13</td>
</tr>
<tr>
<td>PDA DATABASE</td>
<td>76.5</td>
</tr>
</tbody>
</table>

In this project we have used MFCC feature vectors and Gaussian Mixture Models to implement a simple speaker verification model. The major limiting factors in performance of this system are noise and microphone variability.

SCOPE FOR FUTURE WORK:

The verification system considered in this project is based on single phoneme sound and it cannot be used in application areas, where 100% accuracy in verification is desired. Researchers are currently combining classification from several phoneme groups for higher performance and accuracy in text-independent speaker verification systems.