# Automatic Speaker Recognition System

Yuanzhong Zheng

*Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

HAJIM SCHOOL OF ENGINEERING & APPLIED SCIENCES
UNIVERSITY of ROCHESTER

## Overview

**Speaker recognition** is the process of automatically recognizing who is speaking on the basis of individual information included in speech waves. This technique makes it possible to use the speaker's voice to verify their identity and control access to services such as voice dialing, banking by telephone, telephone shopping, database access services, information services, voice mail, security control for confidential information areas, and remote access to computers.

The automatic speaker recognition system can recognize different speakers in different situation. The system extracts MFCCs from audio fragments and implements Linde–Buzo–Gray algorithm to generate a codebook for training dataset and recognizes different speakers according to dataset.

## Objectives

- Use less audio fragments for training (around only one second)
- Recognize different speakers in telephone
- Texts which speakers say do not influence the result of the system

## Principles of Speaker Recognition

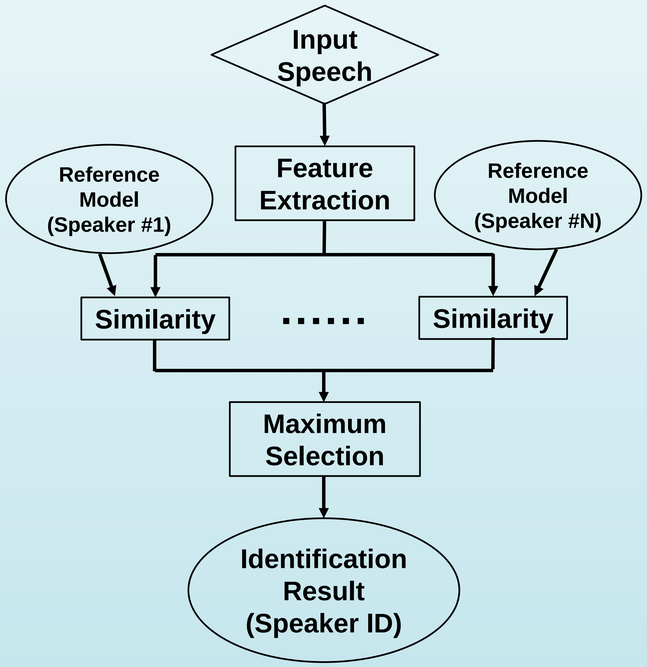


**Figure 1. Speaker Recognition**

## Speech Feature Extraction

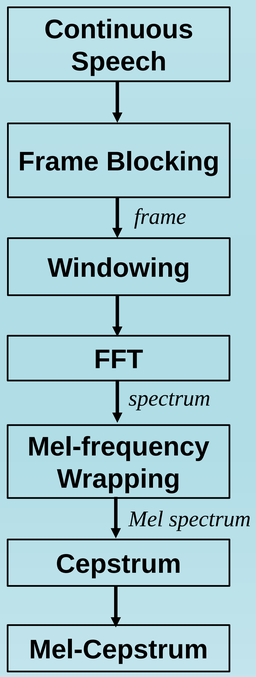


**Figure 2. Block diagram of the MFCC processor**

## Feature Matching

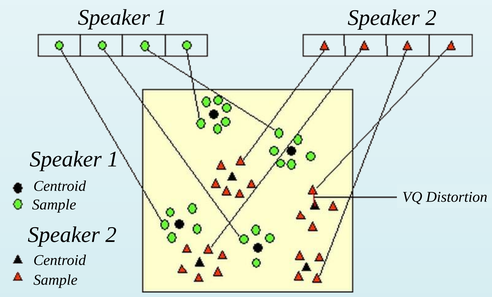


**Figure 3. Conceptual diagram illustrating vector quantization codebook formation. One speaker can be discriminated from another based of the location of centroids. (Adapted from Song et al., 1987)**

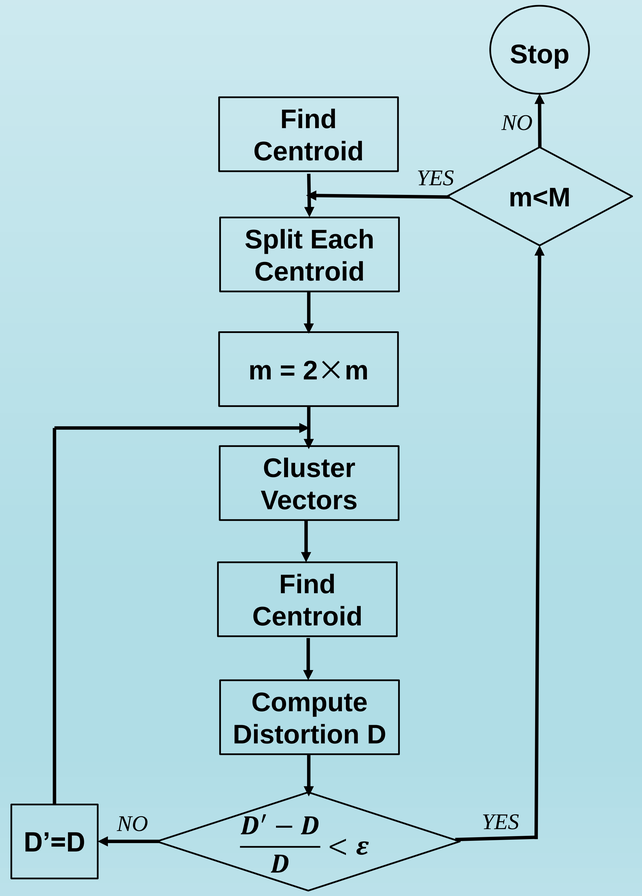


**Figure 4. Flow diagram of the LBG algorithm. (Adapted from Rabiner and Juang, 1993)**

## Result

| Test Sample | Speaker Gender | Result |
|---|---|---|
| 1 | Male | Correct |
| 2 | Male | Correct |
| 3 | Male | Correct |
| 4 | Male | Correct |
| 5 | Female | Incorrect |
| 6 | Female | Correct |
| 7 | Male | Correct |
| 8 | Male | Correct |

**Table 1. Result of the same situation-test dataset.**

| Test Sample | Speaker Gender | Result |
|---|---|---|
| 1 | Male | Correct |
| 2 | Male | Incorrect |
| 3 | Male | Correct |
| 4 | Female | Incorrect |
| 5 | Female | Correct |
| 6 | Male | Correct |
| 7 | Male | Incorrect |
| 8 | Male | Correct |

**Table 2. Result of the telephone-test dataset.**

## Future Work

- Large datasets should be used for training and test.
- Accuracy should be improved in telephone-situation
- Add the function of speaker verification in the system
- Cross-Language can be a new issue for exploring