

Title : Speaker Verification Using gaussian mixture models

Students:

G . Jashwanth Reddy

Kurra Lokesh

The basis for both the identification and verification systems is the GMM used to represent speakers. More specifically, the distribution of feature vectors extracted from a person's speech is modeled by a Gaussian mixture density.

First, the speech is segmented into frames by a 20 ms window progressing at a 10 ms frame rate.

A speech activity detector (SAD) is then used to discard silence/noise frames.

Next, mel-scale cepstral feature vectors are extracted from the speech frames.

Verification system must decide if the input voice came from claimed speaker, with a well defined model, or not the claimed speaker which is not defined in the models.

Consider an input feature vector X . There are two possibilities

H_0 : X is the claimed speaker.

H_1 : X is not the claimed speaker.

We have to apply a likelihood ratio test to an input utterance to determine if the claimed speaker is accepted or rejected. i.e., given X and a speaker model λ we should find if the ratio $\Pr(X \text{ is from } \lambda) / \Pr(X \text{ is not from } \lambda) > \text{some threshold } (\theta)$. If true then we accept that X is from the claimed speaker.