NM STATE UNIVERSITY

New Mexico State University
Klipsch School of Electrical Engineering

EE589 Digital Speech Processing
Spring 2011 – Project #6
Due: 8.55am Thu. Apr. 21

Name: ________________________

Grade: ________________________

Project

The goal of this project is to construct a Gaussian Mixture Model-Universal Background Model (GMM-UBM) Speaker Verification (SV) system. Skeleton codes for this project may be found at

`http://www.ece.nmsu.edu/~pdeleon/EE589/Projects.html`

Reference papers and a link to the Voicebox toolbox used in this project may be found at

`http://www.ece.nmsu.edu/~pdeleon/EE589/Related_Links.html`.

The project consists of two codes, train.m and test.m, which implement training and test stages. In the training stage, feature vectors $\mathbf{x}_t^{\text{train}}$ are extracted from each speaker's training signal. The collection of feature vectors from all speakers are then concatenated and a GMM-UBMλ_{UBM} is computed. Finally, the individual speaker model is computed by adapting the GMM-UBM using the speaker's feature vectors. In the test stage, the system extracts feature vectors $\mathbf{x}_t^{\text{test}}$ from a speaker's speech signal and computes the log-likelihood ratio (LLR) using the model of the claimed identity and the GMM-UBM.

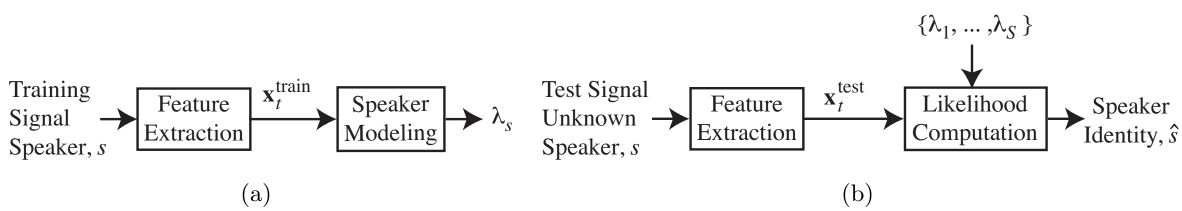

Figure 1: (a) Training and (b) testing stages of a Speaker Identification system.

We will simulate system performance by computing the LLR for every combination of test signal and claimant model. If there are S test signals (one for each speaker) and S speaker models (one for each claimant), there there are a total of S^2 tests. S of these tests will be true claims and $S(S-1)$ of these tests will be false claims (imposters).

References

[1] D. Reynolds and R. Rose, "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models," *IEEE Trans. Signal Proc.*, vol. 3, no. 1, Jan. 1995.

[2] F. Bimbot, et al. ,"A Toutorial on Text-Independent Speaker Verification", *EURASIP Journal on Applied Signal Processing*, vol. 4, pp. 430–451, 2004.

Report

Please submit a printed report with your results including commentary and plots. Please `mailto: pdeleon@nmsu.edu` a zip file containing your report and all MATLAB code to recreate your results.

Notes

We will use speech signals from the TIMIT-sample corpus in the SV simulation

`http://www.ece.nmsu.edu/~pdeleon/Teaching/EE589/TIMIT-sample.zip`

Students are encouraged to discuss detailed, technical aspects with each other and Prof. De Leon. Students are encouraged to utilize existing MATLAB functions in this project. However, students must write all other required codes on an *individual* basis.

1. Problems

Feature Extraction

1. As discussed in lectures, the feature vectors for the SV system are composed of 13 mel-frequency cepstral coefficients (MFCCs), 13 Δ-MFCCs, log-energy, and Δ-log energy. The call syntax for the mel-cepstrum function in Voicebox is

   ```
   c=melcepst(s,fs,w,nc,p,n,inc,fl,fh)
   ```

   For this project, what are the values for the input parameters: `fs, w, nc, p, n, inc`? Use these values in the train_skel.m and test_skel.m code outlines.

Gaussian Mixture Models (GMMs)

2. In this project, we will use a GMM to statistically model the distribution of feature vectors. Information about the computation of a GMM can be found in Section 14.3.3 in Quatieri and

   `http://en.wikipedia.org/wiki/Mixture_model`

   The call syntax for the GMM function in Voicebox is

   ```
   [m,v,w,g,f,pp,gg]=gaussmix(x,c,l,m0,v0,w0)
   ```

   which returns the weights, w, mean vectors, m, and covariance matrices, v for the component densities. For the questions below, use the following values for the number of component densities in the GMM: (a) $m_0 = 2$ and (b) $m_0 = 1$; default values for the other parameters are adequate.

   (a) Generate 500 samples drawn from a gaussian distribution (randn) with $\mu = 0$, $\sigma^2 = 1$. Also generate 500 samples drawn from a gaussian distribution with $\mu = 4$, $\sigma^2 = 2$. Concatenate the above two random vectors into a 1000×1 vector and feed to gaussmix.m (assume two 1-D component densities). What parameters (weights, means, and variances) are returned? How do your parameters compare with the parameters used to generate the random vectors?

   (b) Concatenate the two random vectors in (a) into a 500×2 matrix and feed to gaussmix.m (assume one 2-D component density with a diagonal covariance matrix). What parameters (weights, means, and variances) are returned? How do your parameters compare with the parameters used to generate the random vectors?

Training Stage

3. Fill out the train_skel.m code outline. This code will extract feature vectors from the speakers' training signals, compute the UBM, λ_{UBM} and adapt the UBM for each speaker to produce the speaker model λ_s.

Test Stage

4. Fill out the test_skel.m code outline. This code will extract feature vectors from the speakers' test signals and compute the log-likelihood ratio (LLR) for each of the S^2 tests.

Measuring System Accuracy

5. Determine the Equal Error Rate (EER) and the decision threshold for EER. By examining the confusion matrix (matrix of LLR values) and noting the decision threshold, determine which speaker(s)/claimant model(s) cause(s) false acceptance(s) or false rejection (s). List these.