

Speaker Recognition

William Xia, Michael S. Chen





Problem:

Given the speech data of several subjects, correctly identify new speech samples belonging to one of the subjects.

Approach:

- Find a dataset of a few people saying different phrases
- Generate and use the Mel-frequency cepstrum coefficients of the sample for feature extraction
- Generate gaussian mixture model for each subject. The models predict the subject for a given test sample by comparing its extracted features to the trained model



Results:

- Each speech sample is correctly classified by its respective model
- The Gaussian models for speakers 2 and 3 are not best modeled by their respective speakers.

Future Improvement:

- Optimize the score of a speech sample with its model over the block_size-to-num_samp ratio.
- Optimization is also possible for the number of components used in the gaussian model generation
- This system can be generalized to a larger number of speakers.



BRIEF DEMONSTRATION