On the Use of Spectral and Iterative Methods for Speaker Diarization

Stephen Shum, Najim Dehak, Jim Glass

MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA

{sshum, najim, glass} @ csail.mit.edu

I. Introduction

- Speaker Diarization
  - ✓ "Who is speaking when?"
  - ✓ Model Selection + Clustering + Re-segmentation

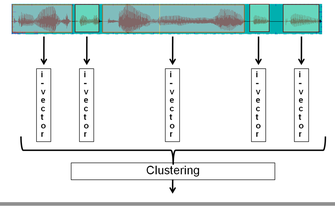
	Previous Approach [1]	Current Approach
Model Selection	Given (K = 2)	**Spectral Estimation**
Clustering	K-means	**K-means or Spectral ?**
Re-segmentation	GMM-Viterbi	GMM-Viterbi + **Iterative Optimization**
Final Pass Refinements [1]	"i-vector K-means"	"i-vector K-means"

II. Speaker Representation

- From GMMs to Factor Analysis
  - ✓ Model a speaker's distribution of acoustic features (AF) using a Gaussian Mixture Model (GMM).
  - ✓ Create a speaker **supervector** by concatenating all mixture mean components in a GMM.
    - ➢ 20 dim (AF) x 1024 mix (GMM) ≈ 20,000 dim
- Total Variability Subspace [2]
  - ✓ Assume all pertinent speaker variabilities lie in some low-dimensional subspace $\boldsymbol{T}$ of the supervector space

$$\boldsymbol{M} = \boldsymbol{m} + \boldsymbol{Tw}$$

  - ✓ $\boldsymbol{w}$ is 100-dimensional ***i-vector***
  - ✓ Use cosine distance to compare two i-vectors

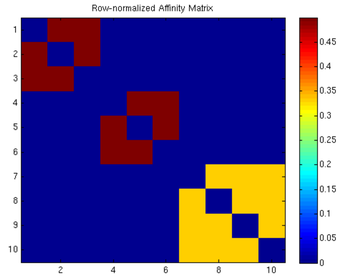


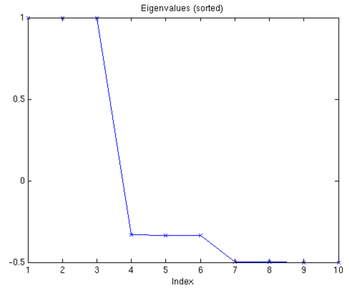
III. Algorithm Details

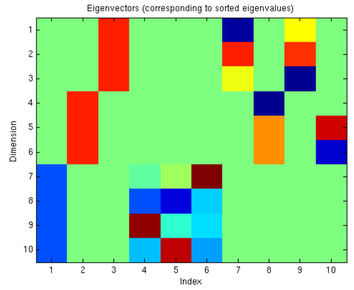
- Viterbi Re-segmentation
  - ✓ Apply the Viterbi algorithm at the acoustic feature level to re-formulate segment boundaries and re-assign frames to each speaker cluster.
- Final Pass Refinements [1]
  - ✓ Extract a single i-vector for each respective speaker
  - ✓ Re-assign each segment i-vector to the speaker whose i-vector is closer in cosine distance
  - ✓ Essentially K-means, where the "means" are computed as i-vectors.

IV. Spectral Clustering Explained

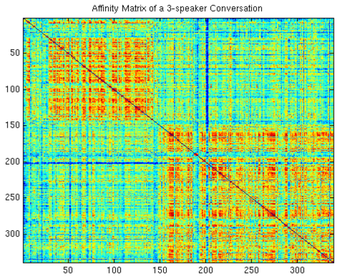
- Use the K largest eigenvectors of normalized affinity matrix (N x N) to project data onto lower-dimensional space (i.e. K-dimensions) before running K-means. [3]

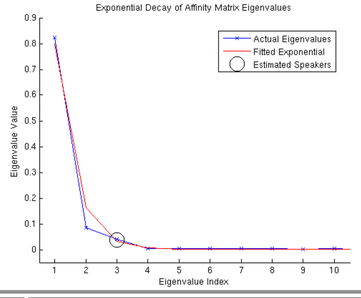


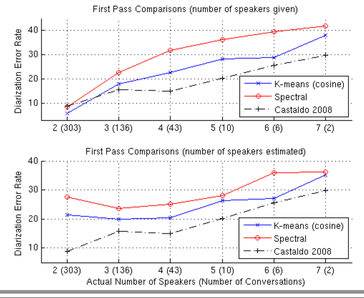




- Estimate number of speakers by fitting an exponential to decaying eigenvalues; then K_0 is the smallest value where the derivative $\geq \theta$. [4]

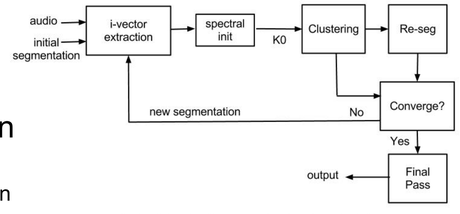


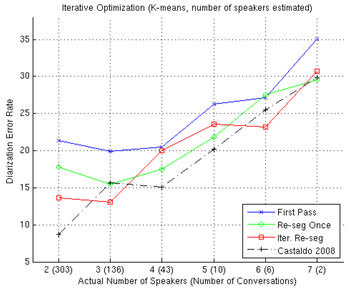




V. Iterative Optimization

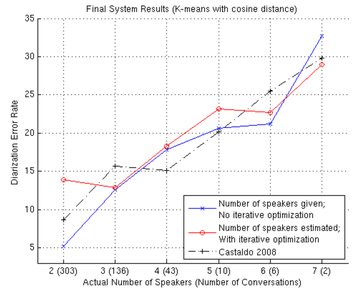
- Utilize multiple levels of information
  - ✓ Clustering on *i-vectors* provides good resolution for speaker identity.
  - ✓ Re-segmentation using *acoustic features* provides better temporal resolution for speaker changes.
  - ✓ Alternate between both until "convergence" (in the DER sense).





VI. Experiments

- Summed-channel Telephone Speech
  - ✓ 2000 NIST SRE subset of multilingual CallHome data
  - ✓ 500 recordings, 2-5 minutes each, 2-7 speakers each
  - ✓ Benchmark for comparison → Castaldo 2008 [5]



- Final System Outline
  - ✓ Model selection via spectral estimation
  - ✓ K-means clustering (cosine distance)
  - ✓ Iterative optimization of segment boundaries – using Viterbi Re-segmentation – and cluster assignments
  - ✓ Final Pass Refinements ("i-vector K-means")

VII. Future Work

- From K-means to a more probabilistic clustering approach (e.g. GMMs, etc.)
  - ✓ Consider variational methods for model selection
- Temporal modeling of conversations [6]

VIII. References

[1] S. Shum, N. Dehak, E. Chuangsuwanich, D. Reynolds, and J. Glass, "Exploiting Intra-Conversation Variability for Speaker Diarization," in *Proceedings of Interspeech*, 2011.

[2] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end Factor Analysis for Speaker Verification," *IEEE Transactions on Audio, Speech, and Language Processing*, May 2011.

[3] A. Ng, M. Jordan, and Y. Weiss, "On Spectral Clustering: Analysis and an Algorithm," in *Proceedings of NIPS*, 2001.

[4] H. Ning, M. Liu, H. Tang, and T. Huang, "A Spectral Clustering Approach to Speaker Diarization," in *Proceedings of ICSLP*, 2006.

[5] F. Castaldo, D. Colibro, E. Dalmasso, P. Laface, and C. Vair, "Stream-based Speaker Segmentation Using Speaker Factors and Eigenvoices," in *Proceedings of ICASSP*, 2008.

[6] M. Johnson and A. Willsky, "The Hierarchical Dirichlet Process Hidden Semi-Markov Model," in *Proceedings of UAI*, 2010.