



SpeakerTagger: a Speaker Diarization System

Naoki Eto Jordan Cazamias Ye Yuan

Speaker Diarization:

The process of distinguishing who is speaking in a conversation and when

Data/Tools used

- 72 hours of Linguistic Data Consortium's (LDC) Speech Corpus of UC Berkeley's ICSI meetings
- python, shell, sph2pipe
- Yaafe, sklearn

Performance Metric:

Diarization Error Rate (DER)

$$DER = E_{Spkr} + E_{Miss} + E_{FA} + E_{Ovlp}$$

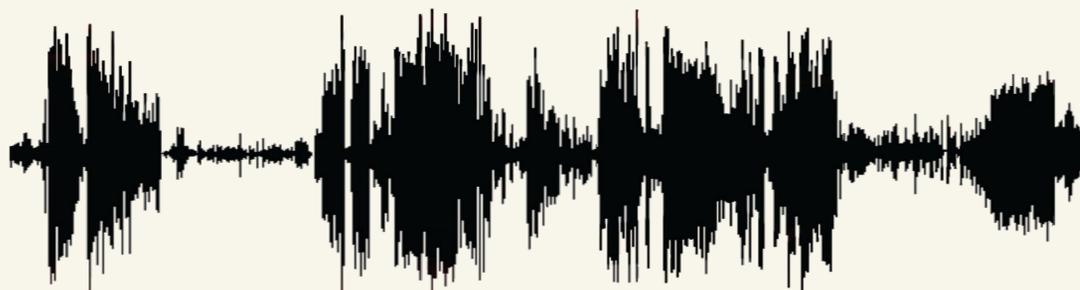
E_{Spkr} **Speaker Error Rate**
Speech segment assigned to wrong Speaker ID

E_{Miss} **Speaker Miss Rate**
Speech segment incorrectly labeled as non-speech

E_{FA} **False Alarm Rate**
Non-speech segment incorrectly labeled as speech

E_{Ovlp} **Overlap Rate**
Segment with overlapping speakers not correctly assigned to all speakers

Input: Raw audio recording of a conversation



1) Feature Extraction

Break up raw data into discrete chunks, apply MFCC



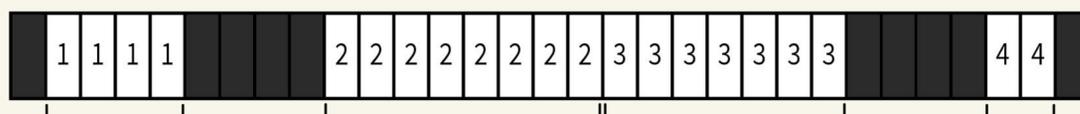
2) Speaker Activity Detection

Eliminate chunks where no one is speaking



3) Segmentation Phase

Separate different speakers, only looking locally



4) Clustering Phase

Assign similar speaker segments to a global Speaker ID



Final Output:

A timeline showing when each speaker is speaking



Baseline Implementation:

Kmeans with Segment Lengths

- There are people who tend to talk longer than others
- Difficult to distinguish between people who give one word or one sentence responses
- DER: 67% with 3 speakers (effectively random)

Other Features under Kmeans

Feature	Diarization Error Rate (DER) with 3 speakers
MFCC	68%
Magnitude Spectrum	51%
Loudness (using Bark Scale Critical Bands)	49%
Energy	49%

Gaussian Mixture Models (GMM)

Feature	Diarization Error Rate (DER) with 3 speakers
MFCC	50%
Magnitude Spectrum	51%
Loudness (using Bark Scale Critical Bands)	51%
Energy	51%

Further Work

- Clustering segments that have a short duration of time is difficult
- Use Bayesian Information Criterion to determine if two short duration segments are likely from the same speaker
- Real-time system using Vector Quantization and other computational speedup methods