OZU NIST SRE-2010 Evaluation

Ozyegin University
A. Erdoğan, F. Yeşil and C. Demiroğlu

Feature extraction

- Bimodal energy is used as Voice Activity Detector
- Complete silence periods are removed
- Log-Energy, 12 MFCC extracted for each frame
- Delta of these features are computed as well
- Log-Energy is removed
- 25-D feature vector used eventually
- Feature warping did not seem to help
  - Removed due to time constraints

UBM Training

- EM algorithm is used to construct GMM with 32 mixtures
  - No time left for experimenting with more mixtures
- Different UBMs constructed for each gender (male/female) and channel (tel/mic)
- Variance and weight flooring is used
- 2004 NIST database is used
- Had significant problems training a UBM using the 2008 data (severe overfitting and other issues)

UBM Training

		Database	Speaker Number	Duration (hour)	
				Without Silence	Total
Mic	Female	SRE 08 mic - long	820	79.6297	174.1415
	Male	SRE 08 mic - long	615	59.1251	128.6382
Tel	Female	SRE 04 tel – 1side	368	7.6734	13.4230
	Male	SRE 04 tel – 1side	248	4.9185	8.9013

Adaptation

- MAP adaptation is used to adapt Universal Background Models
- Only mean parameters are adapted
- Eigenvoice adaptation is used prior to MAP but a significant improvement could not be obtained
  - Removed due to time constraints

Adaptation

		Database	Speaker Number	Duration (hour)	
				Without Silence	Total
Mic	Female	SRE 10 mic – core	1471	35.6841	73.6450
	Male	SRE 10 mic – core	1231	29.8706	62.5763
Tel	Female	SRE 10 tel – core	1551	31.2524	54.6013
	Male	SRE 10 tel – core	1202	22.5006	40.9291

Testing

- Same features and VAD are used as in training phase
- Each trial is scored with UBM and claimed speaker's model

$$\frac{\log p(x|\lambda_{Speaker})}{\log p(x|\lambda_{UBM})} \gtrless \tau$$

- Znorm could not be completed before the deadline
  - Hence a large gap between the optimal decision point and the performance of the submitted systems

Testing

		Database	Trial Number	Duration (hour)	
				Without Silence	Total
Mic	Female	SRE 10 mic – core-core	106744	3400.8362	8327.3985
	Male	SRE 10 mic – core-core	96754	2906.5789	6390.3820
Tel	Female	SRE 10 tel – core-core	102051	4335.8847	7645.2849
	Male	SRE 10 tel – core-core	77859	1464.0267	2645.2980

Processing Speed

- 2 HP z800 machines are used with 8 processing cores
- Processor type: Intel Xeon X5570, 2.93 GHz, 8 MB cache, 1333 MHz Memory, 6.4 GT/s QPI, 95W
- Processing times are normalized wrt to the number of processors used
- 24 GB total RAM on each machine

Processing Speed

Acoustic Models	Speed (xRT)
female mic train	0.22 xRT
female mic test	0.02 xRT
female tel train	0.16 xRT
female tel test	0.03 xRT
male mic train	0.22 xRT
male mic test	0.03 xRT
male tel train	1.45 xRT
male tel test	0.45 xRT