TUL NIST 2010 SRE System

Jan Silovsky

SpeechLab, Faculty of Mechatronics, Informatics and Interdisciplinary Studies,
Technical University of Liberec, Czech Republic

Highlights

- Three monolithic systems (no fusion applied):
  - JFA (primary)
  - UBM-GMM
  - I-Vectors
- All systems are gender-dependent
- Limited use of SRE08 data
- No optimization for new weights of detection cost function
- No special compensation of high or low vocal effort
- Results submitted for:
  - core-core condition
  - extended core-core condition

Feature extraction and segmentation

- Feature extraction:
  - 19 MFCC + c_0
  - feature warping (3 s window)
  - augmented by $\Delta \rightarrow$ 40 features
  - CMS applied over whole utterance
- Segmentation based on:
  - ASR transcripts provided by NIST
  - energy detector

Background data utilization

- Data resources

		SRE04	SRE05	SRE06	SRE08
UBM		•	•	•	•
Calibration		•	•	•	•
JFA	*v*	•	•	•	
	u	•	•	•	
	d	•	•	•	
UBM-GMM	*u*	•	•	•	
I-Vectors	*t*	•	•	•	
	LDA	•	•	•	
	WCCN	•	•	•	
Z-norm		•	•	•	
T-norm		•	•	•	
S-norm		•	•	•	

- Score normalization

Gender	Channel	# recordings
T-norm		
female	tel	141
	mic	64
male	tel	104
	mic	50
Z-norm, S-norm		
female	tel	249
	mic	79
male	tel	192
	mic	69

- UBM training

Gender	Channel	# recordings	# hours
female	tel+mic	22872	1931
male	tel+mic	16899	1432

- Hyper-parameters estimation

Gender	Channel	# recordings	# speakers
eigenvoices (*v*)			
female	tel+mic	11989	819
male	tel+mic	9010	593
eigenchannels (*u*)			
female	tel	8691	704
	mic	2688	98
male	tel	6218	490
	mic	2224	82
diagonal matrix (*d*)			
female	tel+mic	528	89
male	tel+mic	328	68
total variability (*t*), LDA, WCCN			
female	tel+mic	3916	881
male	tel+mic	2800	618

JFA system (primary)

- Full joint factor analysis model
- 1024 Gaussians $\rightarrow$ 40960-dimensional supervectors
- Decoupled estimation of hyper-parameters
  - estimation order:
    1. *v* (200 eigenvoices)
    2. *u* (100+100 eigenchannels)
    3. *d*
  - 7 iterations of maximum likelihood estimation and 2 iterations of minimum divergence estimation performed
- Dot-product linear scoring used
- ZT-norm applied

Alternate systems

- UBM-GMM system
  - MAP adapted models (relevance factor 16)
  - 512 Gaussians
  - 50+50 eigenchannels
    - Eigenchannel adaptation applied in both training and scoring
  - ZT-norm applied
- I-Vectors system
  - Raw cosine kernel distance
  - 1024 Gaussians
  - 300 total factors
  - Compensation techniques:
    - LDA 300 $\rightarrow$ 200
    - WCNN
  - S-norm applied

Calibration

- Based on Linear Logistic Regression (LLR)
- Performed in two stages, conditioned to:
  1. channel-type (gender-dependent)
  2. gender
- Decision threshold set to 6.9

Development experiments

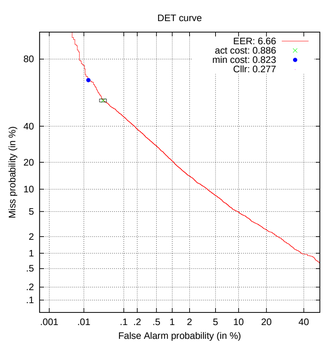
- SRE08, short2-short3

System	EER [%]				
	int-int (same mic)	int-int (dif mic)	int-tel	tel-tel	tel-tel (eng)
JFA	1.7	4.9	7.1	6.2	3.2
UBM-GMM	1.6	6.7	7.2	7.0	3.9
I-Vectors	1.7	7.1	7.6	8.4	5.8

Results

Condition	System		int-int (same mic)	int-int (dif mic)	int-Nvephncall (tel)	int-Nvephncall (mic)	Nvephncall-Nvephncall (dif tel)	Nvephncall-Hvephncall (dif tel)	Nvephncall-Hvephncall (mic)	Nvephncall-Lvephncall (dif tel)	Nvephncall-Lvephncall (mic)
core-core	JFA	*EER*[%]	4.6	6.7	5.3	15.0	4.2	8.1	15.4	2.0	9.3
		DCF_{act}	3.477	0.886	0.888	1.190	0.896	0.892	1.271	0.574	0.999
		DCF_{min}	0.746	0.823	0.693	0.851	0.852	0.781	0.877	0.418	0.855
		C_{llr}	0.282	0.277	0.282	0.856	0.181	0.351	0.887	0.096	0.450
	UBM-GMM	*EER*[%]	6.0	8.6	5.3	18.0	5.1	10.8	18.6	2.3	11.7
		DCF_{act}	10.401	1.044	0.893	2.318	3.812	1.050	2.186	0.819	3.526
		DCF_{min}	0.827	0.925	0.639	0.918	1.034	0.970	0.969	0.759	0.993
		C_{llr}	0.526	0.318	0.244	0.808	0.249	0.620	0.775	0.117	0.414
	I-Vectors	*EER*[%]	4.7	9.3	7.2	17.0	8.2	10.6	18.1	4.4	12.1
		DCF_{act}	1.146	0.914	1.094	0.885	0.853	0.989	0.930	0.919	0.879
		DCF_{min}	0.628	0.869	0.872	0.849	0.812	0.983	0.911	0.804	0.855
		C_{llr}	0.514	0.343	0.440	0.572	0.301	0.824	0.594	0.205	0.547
coreext-coreext	JFA	*EER*[%]	4.8	6.6	6.7	15.4	4.6	8.4	15.8	3.9	9.3
		DCF_{act}	3.691	0.955	0.804	1.317	0.761	1.042	2.841	0.736	1.063
		DCF_{min}	0.794	0.814	0.720	0.854	0.754	0.991	1.009	0.694	0.887
		C_{llr}	0.297	0.277	0.308	0.856	0.182	0.340	0.916	0.162	0.454

- JFA, int-int (dif mic)



- JFA, Nvephncall-Nvephncall (dif tel)