The SVIST System for The NIST 2010 SRE

0

Changli, Chaoqunliu, Shenshenlin Shanghai Voice Information and Science Technology



Overview

- Task participated
 - core-core
 - 8conv-core
 - I0sec-I0sec
- Systems
 - JFA(Joint Factor Analysis) + PLP feature
 - TVCD(Total Variability Cosine Distance) + LPCC feature
 - GMM-SVM-NAP + MFCC feature
 - Eigenvoice + MFCC&PLP feature (10sec-10sec)



Overview

- Primary System for core-core and 8conv-core
 - JFA-PLP +TVCD-LPCC+GMM-SVM-NAP-MFCC
- Alternate System for core-core and 8conv-core
 JFA-PLP+GMM-SVM-NAP
- Primary System for 10sec-10sec
 - Eigenvoice-MFCC + Eigenvoice-PLP+GMM-SVM-MFCC
- Alternate System for 10sec-10sec
 - Eigenvoice-PLP+GMM-SVM-MFCC



Feature Extraction

- MFCC, PLP, LPCC all extracted by HTK tools
- Simply energy based VAD & ETSI VAD
- RASTA, Feature Warpping
- For interview data, interviewer's speech is removed by our vad tag
 - simply energy based vad on interviewer's channel.
 - About 20% interviewer's data removed

JFA

- Gender dependent UBM
 - NIST SRE 2004 telephone data
- 2048 Gaussians + 36 dimensions PLP
- 300 speaker factors
 - NIST SRE 2005, 2006, 2008 telephone data
- 250 channel factors
 - 100 Telephone: NIST SRE 2006 and 2008 telephone corpora
 - 50 Microphone : NIST SRE 2005, 2006 and 2008 microphone corpora.
 - 100 Interview: NIST SRE 2008 and followup interview corpora
 - TZNORM
 - Gender & channel dependent
 - 500 tnorm speakers
 - 400 znorm utterances

Total Variability cosine Distance

- 2048 Gaussians + 36 dimensions LPCC
- Total Variability Matrix 400 dimensions
 NIST SRE 2004, 2005, 2006, 2008 telephone data
- LDA transform
 - NIST SRE 2004, 2005, 2006, 2008 telephone data
- WCCN
 - Gender & Channel dependent WCCNs
 - The same as channel spaces training in JFA
- Cosine distance & TZNORM

GMM-SVM-NAP

- 512 Gaussians + 60 dimensions MFCC
- GMM mean supervectors from MAP
 - Normalized mean vectors
- NAP
 - Gender & channel dependent
 - The same as channel spaces training in JFA
- TZNORM

Eigenvoice system

- 512 Gaussians + MFCC & PLP
- JFA configuration eigenvoice step
- TZNORM



TZNORM

- Condition based
 - Tnorm model –same training condition as the hypothesized speaker Znorm test speech- same testing condition as the trial speech
- Tnorm: cohort models selection for each speaker

Fusion

- Development database(from SRE 08)
 - We remove some abnormal data from the development database
 - stimulate the training and testing conditions
- Linear fusion for each training and testing conditions

Development Experiments Core-Core



Fusion-Female

Mindcf(*I0 ⁻³)	JFA-PLP	TVCD- LPCC	GMM- SVM- MFCC	Fusion
mix_mix	0.611024	0.393552	0. 491535	0. 323139
mix_tel	0.865184	0.846471	0.812794	0.6614
tel_mic	0.851852	0.66127	0.955026	0.600476
mic_mic	0. 491997	0. 471792	0. 435185	0.366898
tel_tel	0. 302816	0.280479	0. 420995	0. 226786

Development experiments on interview-telephone condition (female)



Fusion-Male

Mindcf(*I 0 ⁻³)	JFA-PLP	TVCD- LPCC	GMM- SVM- MFCC	Fusion
mix_mix	0.287754	0.189428	0.226985	0.128816
mix_tel	0.843325	0.750511	0.782985	0. 5591
tel_mic	0.674732	0. 484553	0.731642	0.458537
mic_mic	0.458931	0.47384	0. 424613	0.341772
tel_tel	0.339843	0.282862	0.273145	0.247531

Development experiments on interviewtelephone condition (male)



SRE 10 Results Core-Core

Mindcf(* I 0 -3)	Male+Female
interview_interview_same_mic	0.2291
interview_interview_different_mic	0.3173
interview_nvephonecall_tel	0.2865
interview_nvephonecall_mic	0.3041
nvephonecall_nvephonecall_different_tel	0.3821
nvephoencall_hvephonecall_different_tel	0.7596
nvephoencall_hvephonecall_mic	0.4507
nvephoencall_lvephonecall_different_tel	0.3577
nvephoencall_lvephonecall_mic	0.2342



Conclusion

- Matched training data can greatly improve the system performance
- Interviewer's data need to be removed
- Development data has to be verified
- Multi-vad and Milti-feature are beneficial



References

- Kenny, P., Ouellet, P., Dehak, N., Gupta, V., and Dumouchel, P.A Study of Inter-Speaker Variability in Speaker Verification
- Dehak, N., Kenny, P., Dehak, R., Dumouchel, P and Ouellet, P. Front-End Factor Analysis for Speaker Verification
- W. M. Campbell, D. E. Sturim, D. A. Reynolds, SVM based speaker verification using A GMM supervector kernel and NAP variability compensation.
- Kenny P, Eigenvoice Modeling with sparse training data

•Thank you!