The LIMSI 2005 Speaker Recognition System

Claude Barras, Cheung-Chi Leung and Jean-Luc Gauvain

Spoken Language Processing Group
LIMSI-CNRS, France
http://www.limsi.fr/tlp

INTRODUCTION

Task condition

- 1 conversation (4-wire) for training and test

Main focus

- Generic system for landline and cellular data
- Take both audio channels into account
- Test corrective training

Primary system

- A standard GMM-UBM system, incl. feature selection, channel mapping, feature warping and T-norm
- Development using landline and cellular data from SRE’00-04.

LIMSI SRE'04 BASELINE SYSTEM

Front-end

- 31 features: 15 cepstrum + 15 Δ cepstrum + Δ energy
- Feature warping
- Speech activity detection: Viterbi decoding with a 2 state HMM

Models

- 2 gender-dependent UBM with 1024 Gaussians
- Cellular training data from SRE'01
- MAP adaptation of UBM means

Scoring

- Log-likelihood ratio with 20 top Gaussians scoring
- Gender-dependent T-norm using cellular data from SRE'01 (discard T-Norm speaker with lowest scores)

LIMSI SRE’05 PRIMARY SYSTEM

Front-end

- reordering: compute Δ before other normalizations
- more features: add $\Delta\Delta$ cepstrum + $\Delta\Delta$ energy
- speech detection: use word boundaries of BBN ASR instead of SAD + further filtering of 10% low energy frames

Cellular/landline system

- Use SRE’00 and SRE’01 cellular and landline UBM training data
- Feature mapping for channel compensation
- Separate UBM training for each source

Scoring

- Perform T-norm using SRE’02 + SRE’04 eval data

FRAME SELECTION

Use ASR information

- Make use of BBN ASR word boundaries for SRE’04 and SRE’05 data (keep baseline SAD for SRE’00 and SRE’01 data)
- For SRE’05 data, also exclude speech from other side

Test data	SRE’04 all		SRE’04 c’mon		SRE’05 c’mon	
System	*MDC*	*EER*	*MDC*	*EER*	*MDC*	*EER*
SAD	56.1	15.6	53.0	13.7	49.1	14.0
ASR	49.2	13.5	46.4	13.4	47.1	12.5
2 sides ASR	-	-	-	-	47.9	12.8

- about 12% reduction of MDC for SRE’04,
  4% MDC improvement for SRE’05 (about the same resulting MDC)
- slight degradation from using opposite side!
  echo cancellation appears good enough for the task (and the system)

TRAINING DATA

Extend baseline system with non-cellular data

UBM

- Training data:
  - SRE’00 landline data (1000 speakers)
  - SRE’01 cellular (234 spk)
- MLE training of 2 gender-dependent UBM:
  - process separately cellular, landline electret and landline carbon data.
  - train 3 models with 512 Gaussians each and fuse them
  - subsample data (saturate at 1000 frames per Gaussian)

TRAINING DATA (cont’)

T-norm

- SRE’02 cellular (330 spk) + SRE’04 mixed (616 spk)
- tests on SRE’04 with a round-robin scheme

Results

Test data	SRE’04 all		SRE’04 c’mon		SRE’05 c’mon	
System	*MDC*	*EER*	*MDC*	*EER*	*MDC*	*EER*
Cellular data	47.8	12.9	45.1	12.7	49.8	12.7
Mixed data	44.5	11.9	42.3	11.5	43.1	11.3

- 6-7% reduction of MDC for SRE’04,
  13% MDC improvement for SRE’05
- SRE’04 not subject to T-norm length mismatch like SRE’02 data

CHANNEL COMPENSATION

Feature mapping (Reynolds et al)

- Train a gender-specific root model
- MAP adapt (mean-only) to 3 channel conditions:
  cellular, landline carbon, landline electret
- Train UBM on data after feature mapping

Results

Test data	SRE’04 all		SRE’04 c’mon		SRE’05 c’mon	
System	*MDC*	*EER*	*MDC*	*EER*	*MDC*	*EER*
Raw features	44.5	11.9	42.3	11.5	43.1	11.3
Mapped features	42.3	10.8	37.4	10.2	42.7	11.0

- about 5% reduction of MDC for SRE’04 and 10% for common condition
- less than 1% MDC improvement for SRE’05

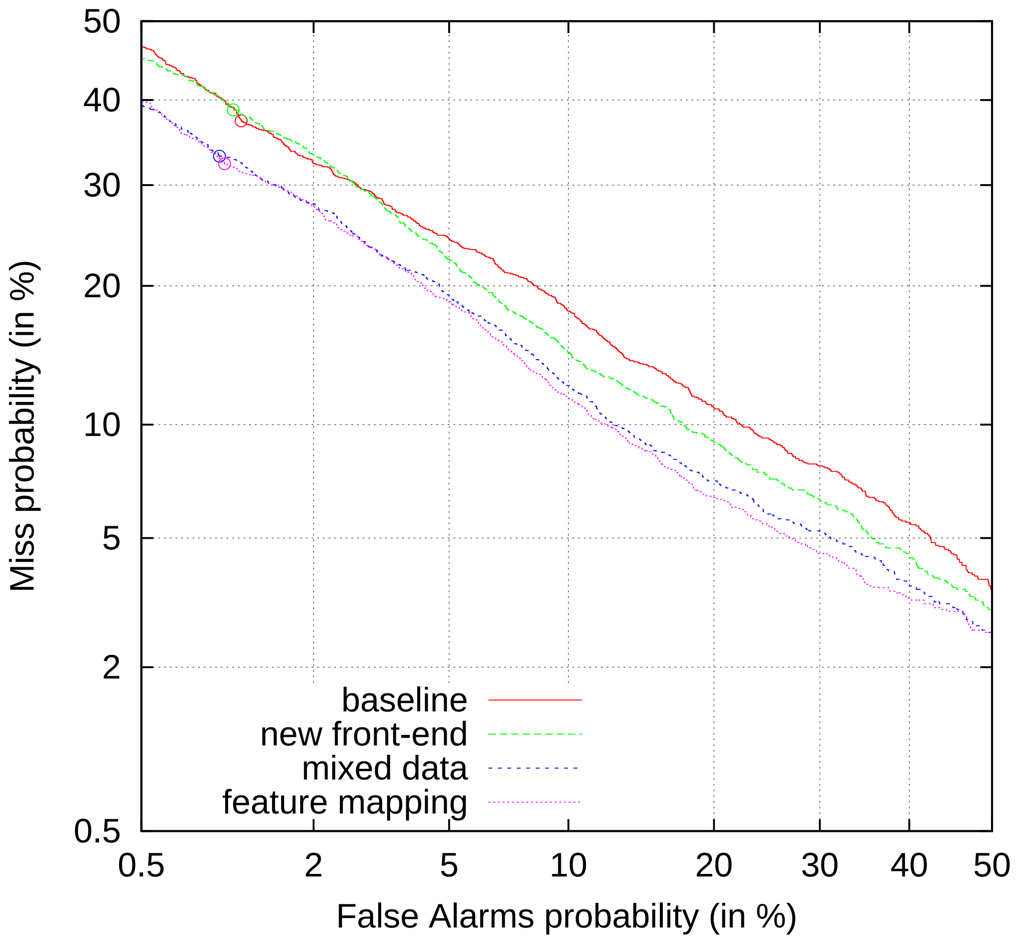
PRIMARY SYSTEM PERFORMANCES

Performances summary

Test data	SRE'04 all		SRE'04 c'mon		SRE'05 c'mon	
System	*MDC*	*EER*	*MDC*	*EER*	*MDC*	*EER*
'04 baseline	56.1	15.6	53.0	13.7	**49.1**	14.0
'05 primary	42.3	10.8	37.4	10.2	**42.7**	11.0

- 25-30% MDC reduction obtained for the primary system on SRE'04 data
- Resulted in only 13% MDC reduction on SRE'05 data
  (possible overfitting to SRE'04 data in system configuration?)
- 20-30% relative reduction of EER

SYSTEM DET



50
40
30
20
10
5
2
0.5
Miss probability (in %)
0.5
2
5
10
20
30
40
50
False Alarms probability (in %)
baseline
new front-end
mixed data
feature mapping

SUMMARY

Primary GMM system

- front-end gains mainly from use of ASR for frame selection
- as expected, matching training data helps a lot!
- limited impact of feature mapping on evaluation data

Contrastive system

- Simple approach for discriminative training:
  negative MAP adaptation weight to nearest impostors of target speaker

Conclusions

- significant improvements compared to LIMSI SRE'04 system
  (13% relative reduction of MDC)
- ...but an increased gap with the best systems!