Tsinghua EE Telephone Speech Speaker Recognition System

Liu Jia, Li Ke, Lu Xiang-Feng, Hou Tao

Dept. EE, Tsinghua University, Beijing, PRC

Outline

- Introduction
- System Description
  - ✓ GMM
  - ✓ SVM
- Results on development data
- Results on SRE 06

Introduction

- Speech Technology Laboratory, Department of Electronic Engineering, Tsinghua University, China
- Research on
  - ✓ Speech signal processing: Speech Recognition, Speaker Recognition, Language Identification…
  - ✓ Speech Chip: 8bit MCU or 16bit DSP core
  - ✓ Microphone Array and Audio Signal Processing

Speaker Recognition System

- 3 systems are used in SRE 06
  - ✓ GMM
  - ✓ SVM
  - ✓ Fusion

Speech Features

- The same features in the two sub-systems
- Speech detect
  - ✓ ITU G.723.1 VAD algorithm is used.
  - ✓ Low-energy frames are cut.
- 19-Dimensions MFCC Features
- Short Time Gaussianization for MFCC
- 19 Delta Gaussianized MFCC

GMM System

- UBM-GMM architecture

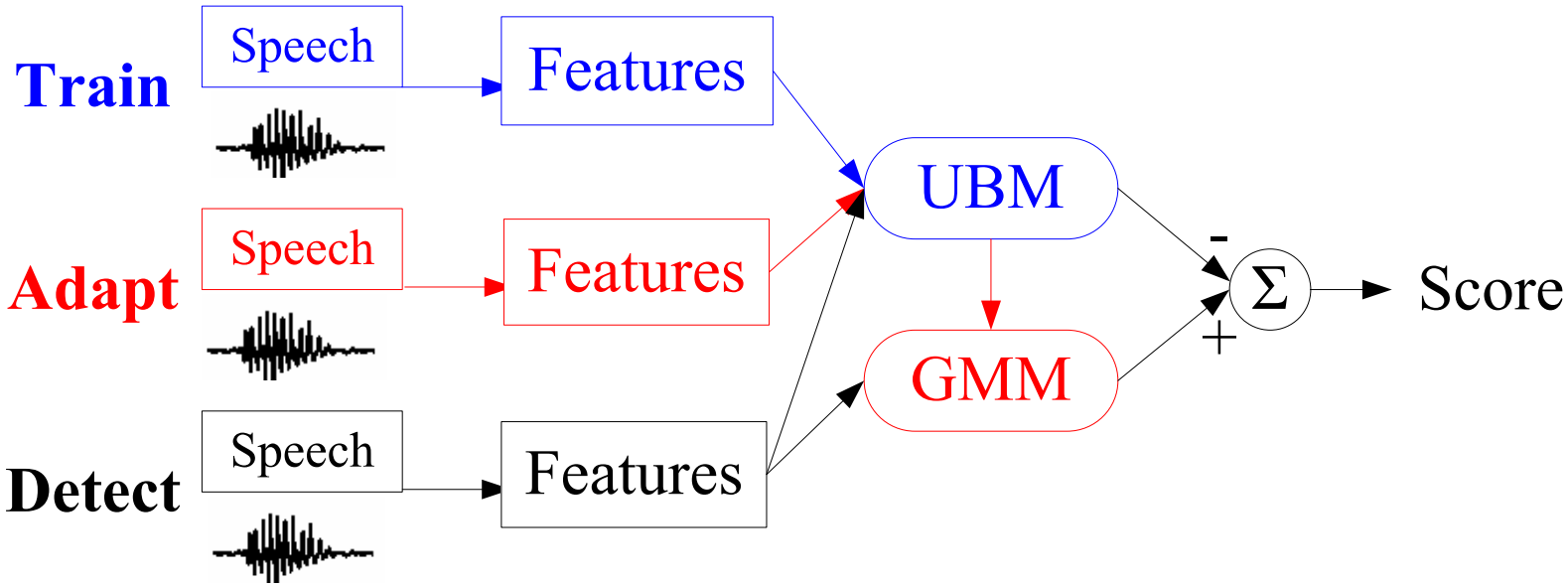


Fig 1 UBM-GMM system

UBM Training

- Gender Dependent UBM is used.
- 2048 Mixture GMM
- 2 hours channel balanced data (Cordless, Cellular and Regular)

Adapt GMM and Detection

- Adapt
  - ✓ Mean-only adaptation
  - ✓ Relevance factor is fixed at 16
- Detection
  - ✓ Log Likelihood Ratio
  - ✓ Fast score processing, top 5 components in UBM are chosen and used for computation of GMM.

SVM

- The SVM system includes 3 parts.
  - ✓ Feature Extract
  - ✓ Model Training
  - ✓ Speaker Verification.

Feature Extraction

- Feature extract part deals with the feature extraction task.
- GLDS(Generalized linear discriminant sequence) kernel is used.

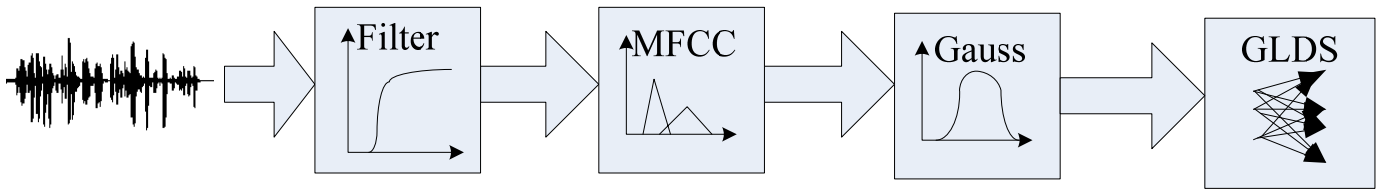


Fig2 Feature Extraction in SVM system

Model Training

- The target speaker samples and the background speaker samples are send to feature extraction in parallel order.
- Vector sequences are used for training the speaker model.

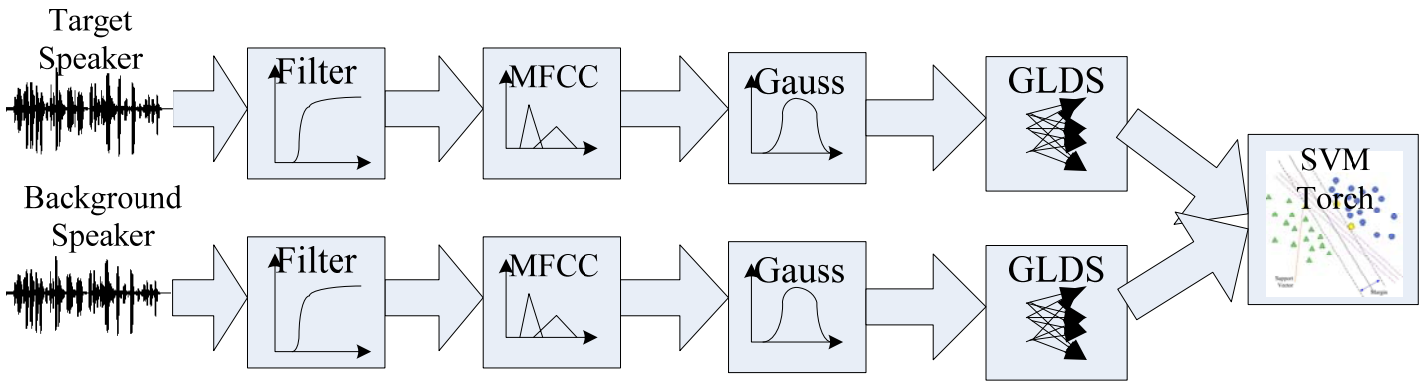


Fig3 Model Training in SVM system

Speaker verification

- Vectors of unknown speaker samples are extracted and send to the SVM test core.
- These Vectors are then evaluated against designated model to verify whether the speaker matches the true speaker.

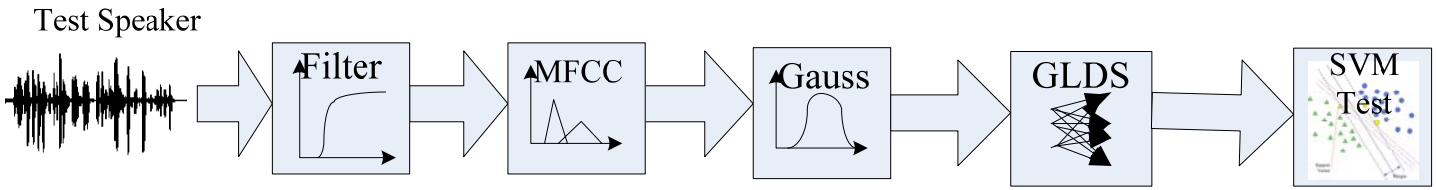


Fig4 Verification in SVM system

Fusion system

- Linear fusion of the result scores from GMM and SVM results.
- The weight coefficients are calculated using Fisher Criteria.

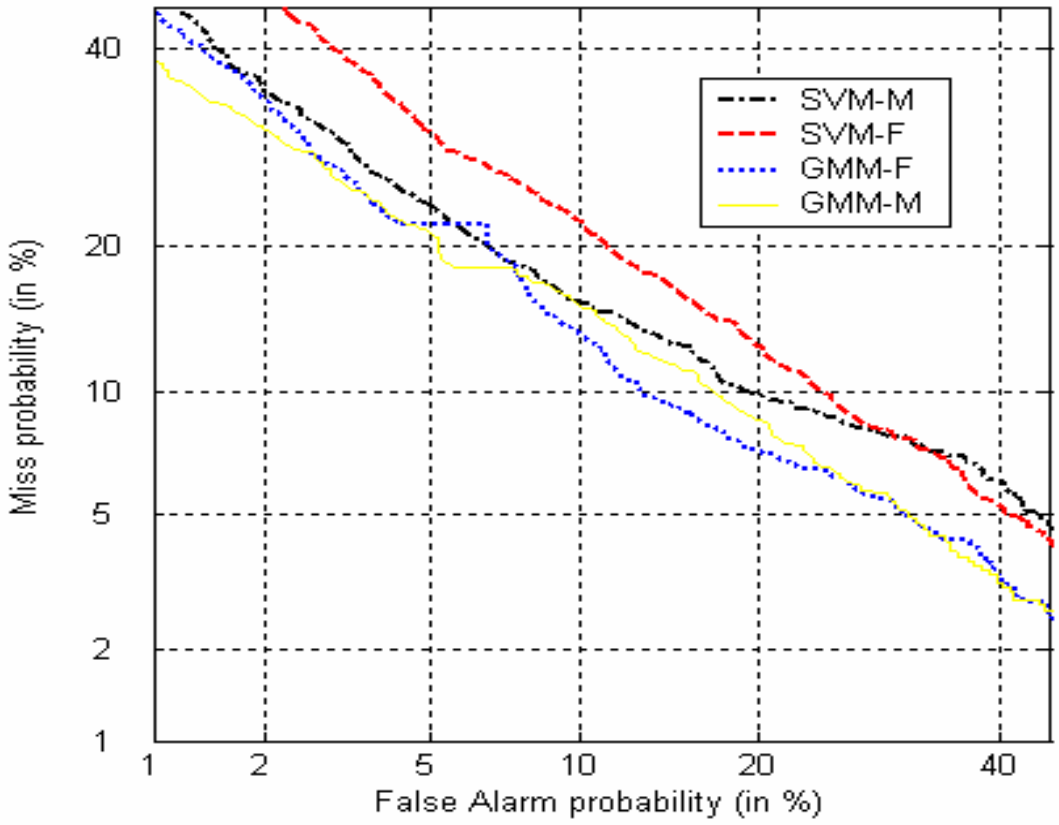
Experiments

EER	GMM	SVM	Fusion
04M	12.3	13.1	11.6
04F	10.5	15.3	11.3
05M	9.0	11.8	9.4
05F	12.7	14.0	12.1
04+05M	10.8	12.9	10.4
04+05F	11.9	14.3	11.4

Fig5 EER on Development Data

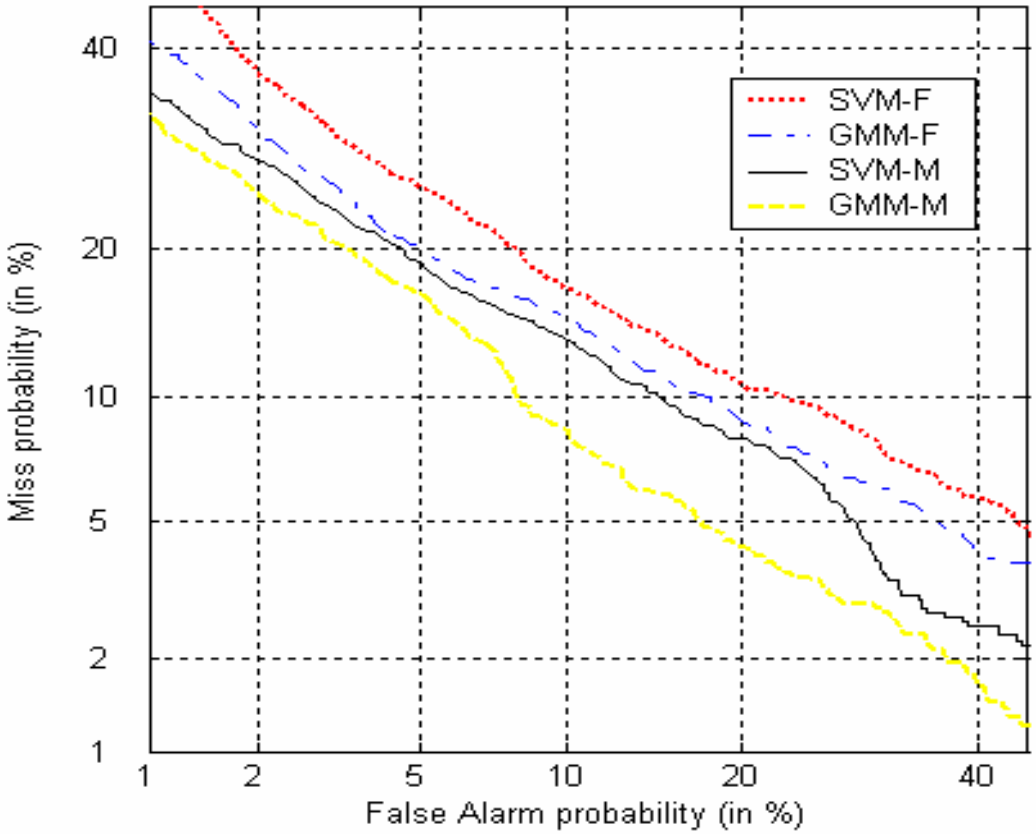
DET Curves

- NIST SRE 04



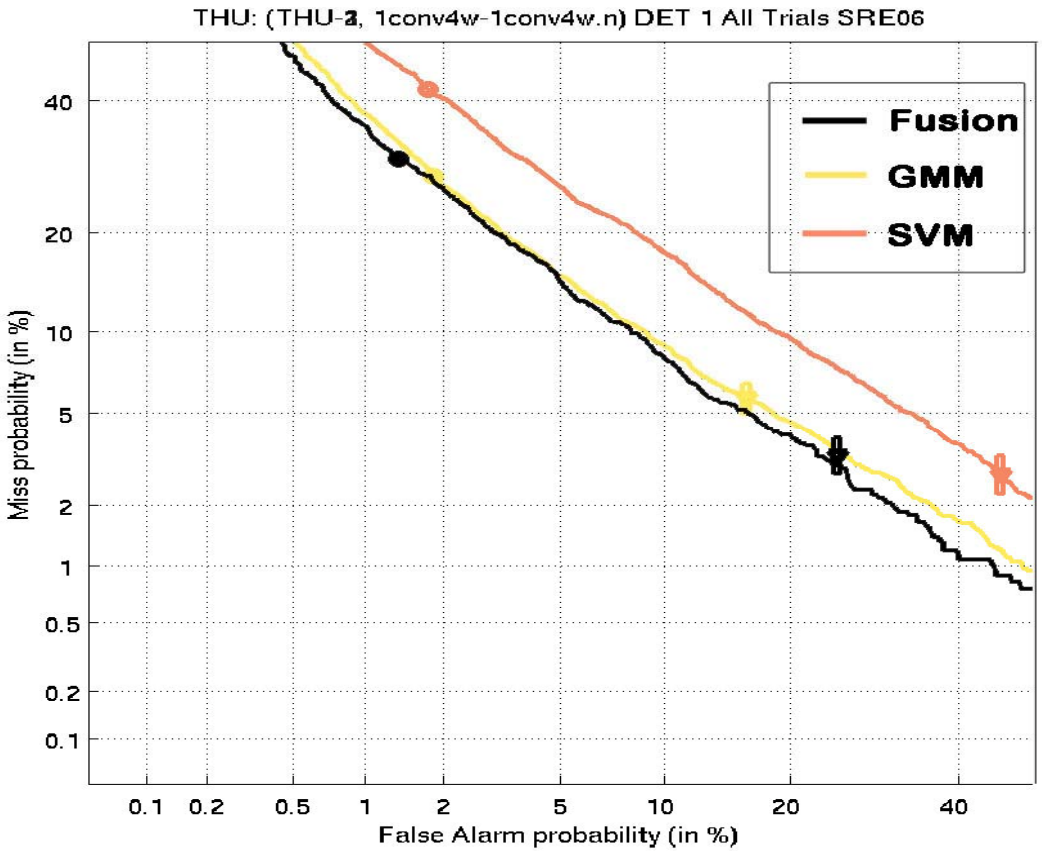
DET Curves

- NIST SRE 05



DET Curves

- NIST SRE 06



Conclusion

- GMM performs well in Nist SRE when appropriately implemented.
- SVM shows similar capability as GMM in Nist SRE.
- System Fusion utilizes information from various sources and yields decent results.
- Muti-system and integrated system will become a future trend.