The LIMSI 2006 Speaker Recognition System

Cheung-Chi Leung, Marc Ferras,

Claude Barras 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 characteristics

- Constrained-MLLR modeling
- SVM classifier
- Two-way matching: forward and backward approaches
- Fusion of systems

Primary system

- Score-level fusion of 6 sub-systems:
  - MFCC-GMM forward and backward
  - MFCC-SVM forward and backward
  - MLLR-SVM forward and backward
- Development using landline and cellular data from NIST SRE'00-05

LIMSI SRE'06 MFCC-GMM SYSTEM

Two-way matching similar to '05 SDV approach

- Forward: extracted features of test speech are matched with statistical models of training speech (i.e. conventional approach)
- Backward: extracted features of training speech are matched with statistical models of test speech

Front-end

- 47 features: 15 cepstrum + 15 Δ + 15 $\Delta\Delta$ + Δ / $\Delta\Delta$ energy
- Feature mapping:
  - 5 channel conditions (instead of 3):
    carbon, electret, gsm, cdma and tdma
  - Training data: SRE'00, 01 and 02 data
    (last year, only SRE'01 dev and SRE'00 training data were used)
- Feature warping

LIMSI SRE'06 MFCC-GMM SYSTEM (Cont)

Front-end (Cont)

- Speech activity detection:
  - Use word boundaries of BBN ASR + further filtering of 10% low energy frames;
  - If transcription is not available or contains too little information, 2-state HMM speech detector is used

Speaker modelling

- Gender-dependent 1536-component UBMs is formed by merging three 512-component GMMs
  - Landline electret GMM: 347 segments from SRE'00
  - Landline carbon GMM: 653 segments from SRE'00
  - Cellular GMM: 234 segments from SRE'01 + 3000 from SRE'03
- MAP adaptation of UBM means

LIMSI SRE'06 MFCC-GMM SYSTEM (Cont)

Scoring

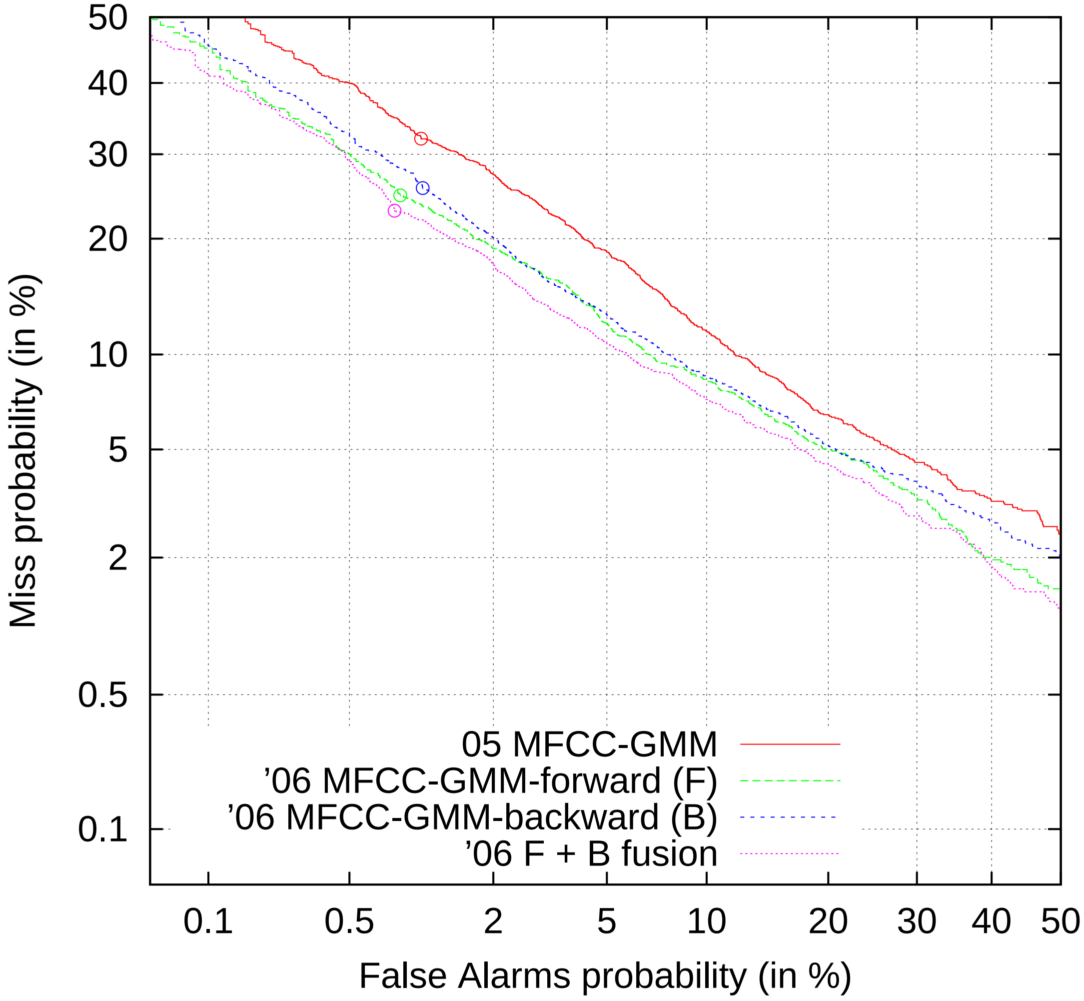
- Log-likelihood ratio with 20 top Gaussians scoring
- Perform T-norm using Fisher corpus
  (last year, SRE'02 + SRE'04 eval data was used)

Performance on development

- Evaluation on SRE'05 Eval

System	*MDC (x10)*	*EER*
'05 MFCC-GMM	0.426	10.9
'06 MFCC-GMM-forward (F)	0.334	8.95
'06 MFCC-GMM-backward (B)	0.363	9.04
'06 MFCC-GMM fusion (.5F+.5B)	0.311	8.35

MFCC-GMM SYSTEM PERFORMANCE ('06 vs '05)



SVM-based systems

- Two systems with identical modelling set-up
  - MFCC-SVM: MFCC-derived features
  - MLLR-SVM: Constrained MLLR transforms as features

Modelling

- Gender-independent linear-kernel SVM models (using SVMTorch from IDIAP)
- 3198 impostor speakers (1376 male, 1822 female)
  - NIST SRE'99, SRE'00, SRE'01, SRE'02, SRE'04 training data
- Two-way matching as for the GMM system (2 models/trial)

MFCC-SVM System

Front-end

- GMM's cepstral features
  - 15 MEL-PLP cepstrum coefficients + $\Delta/\Delta\Delta$ cepstrum + $\Delta/\Delta\Delta$ energy (47-d features)
  - Feature mapping, feature warping
- Mean 1st, 2nd and 3rd order monomial expansion with with-in segment variance normalization (one 20824-d feature/segment)
  - Up to 3rd order moment estimation
  - High dimensional features ease SVM work

$$\mathbf{x} = (x_0, x_2, \ldots, x_{m-1}),\ \Phi(\mathbf{x}, 1) = \mathbf{x}$$
$$\Phi(\mathbf{x}, 2) = (x_0^2, \ldots, x_i x_j, \ldots, x_{m-1}^2)$$
$$\Phi(\mathbf{x}, 3) = (x_0^3, \ldots, x_i x_j x_k, \ldots, x_{m-1}^3)$$
$$\Phi(\mathbf{x}) = (\Phi(\mathbf{x}, 1), \Phi(\mathbf{x}, 2), \Phi(\mathbf{x}, 3))$$
$$\overline{\Phi_{vn}(\mathbf{x})} = \frac{\overline{\Phi(\mathbf{x})}}{\sigma}$$

MFCC-SVM System (Cont)

- Feature dimension reduction via Kernel PCA using a 2nd order cumulative homogeneous polynomial kernel (one 3197-d feature/segment)

$$\boldsymbol{\kappa}(\mathbf{x_i} \cdot \mathbf{x_j}) = \mathbf{x_i} \cdot \mathbf{x_j} + (\mathbf{x_i} \cdot \mathbf{x_j})^2$$

- Min-Max feature normalization in the range $[-\frac{1}{3197}, \frac{1}{3197}]$

Performance on SRE’06 Dev

System	*MDC (x10)*	*EER*
’06 MFCC-GMM-forward	0.334	8.95
MFCC-SVM Forward (F)	0.289	7.78
MFCC-SVM Backward (B)	0.296	7.86
MFCC-SVM fusion (.5F+.5B)	0.263	7.07

Note: SRE’06 Dev. stands for SRE’05 Eval data with specifi c test-target trial index

MLLR-SVM System

Motivation

- Focused on speaker modelling instead of just spectrum characteristics.

Front-end

- Contrained-MLLR transforms as features
  - Ax+B, affine transform of means and variances to maximize likelihood of a GMM (UBM) model
  - Constrained mean and variance allows transformation of the input features directly.
  - No need for ASR:
    the transform is global rather than specific to a phonetic class

MLLR-SVM System (Cont)

- Process:
  1. Train UBM model on background speakers' cepstral features
  2. Estimate C-MLLR transforms (one 2256-d feature per segment)
  3. Apply the transforms on the cepstral features
  4. Go back to 1 for UBM re-estimation. Iterate 4 times.
- Use last iteration's C-MLLR transforms as features (A and B in vector form)
- Min-Max feature normalization in the range $[-\frac{1}{2256}, \frac{1}{2256}]$

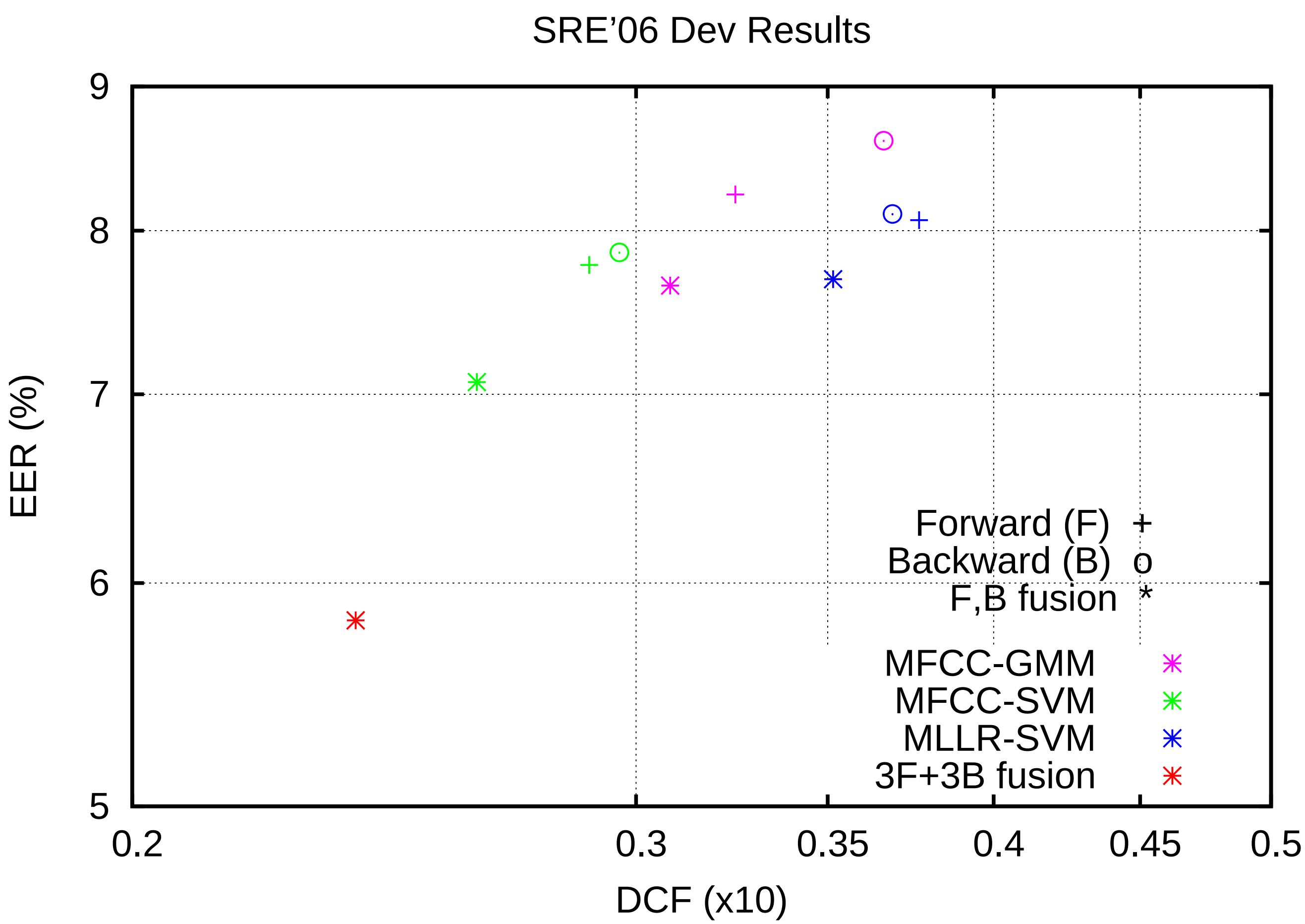
Performance on SRE'06 Dev

System	*MDC (x10)*	*EER (%)*
'06 MFCC-GMM-forward	0.334	8.95
MLLR-SVM Forward (F)	0.376	8.07
MLLR-SVM Backward (B)	0.368	8.11
MLLR-SVM FB (.5F+.5B)	0.351	7.69

SCORE FUSION

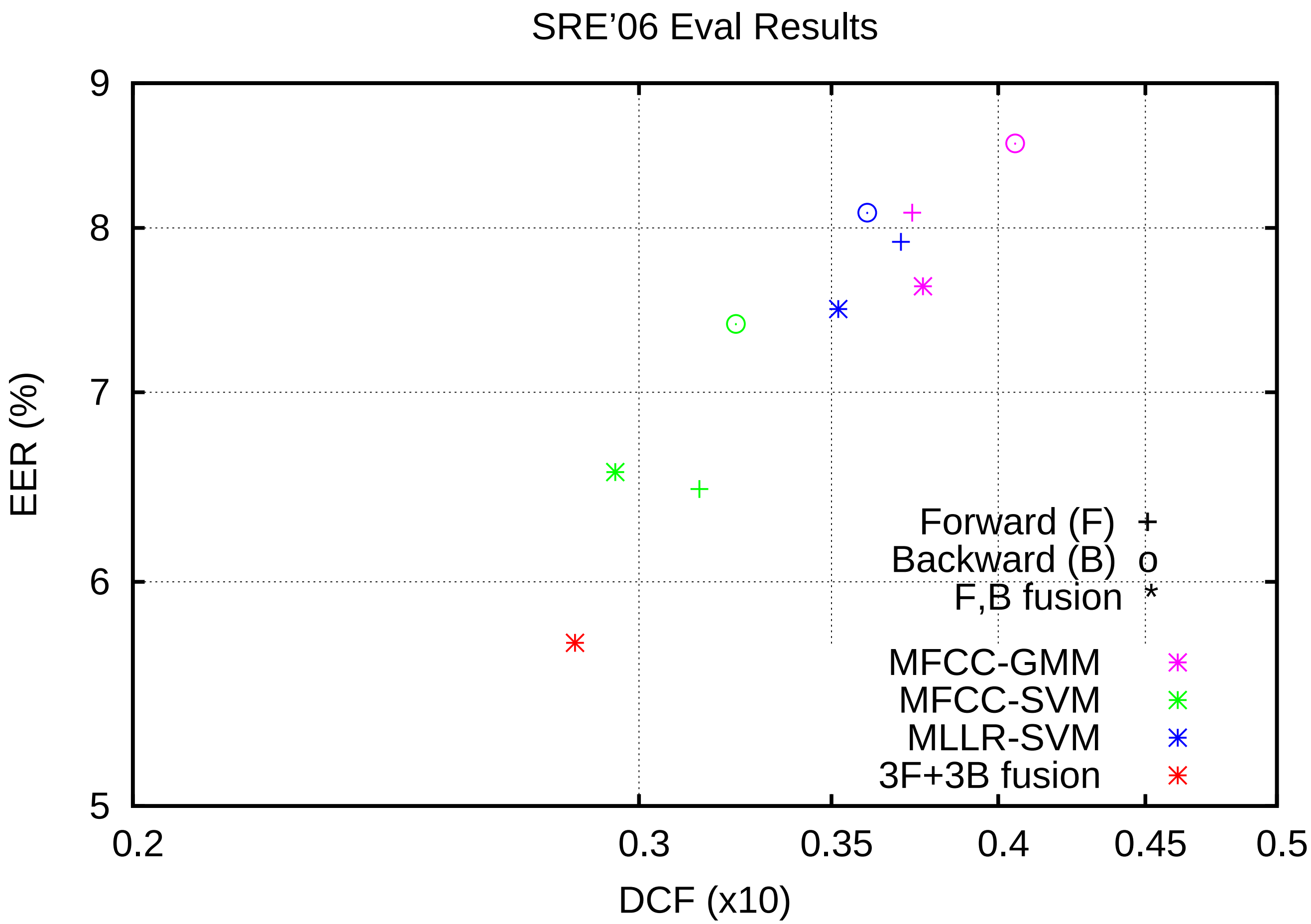
- Normalize 6 sub-system scores to zero mean and unit variance (using SRE'05 data)
- Arithmetic mean of the 6 normalized scores
- Decision threshold on the mean score is chosen using SRE'05 data
- Other score fusion classifiers we tested: MLP, SVM and Gaussian mixture
- Evaluation methodology of fusion approaches (3-fold cross validation)
  - Split SRE'05 eval data into three independent subsets (A, B and C) – a specific trial index was used to insure the independence
    - train a classifier using A and B subsets ⇒ test it on C subset
    - train a classifier using A and C subsets ⇒ test it on B subset
    - train a classifier using B and C subsets ⇒ test it on A subset
  - Combine the 3 test subsets and evaluate its performance

SCORE FUSION (Dev)



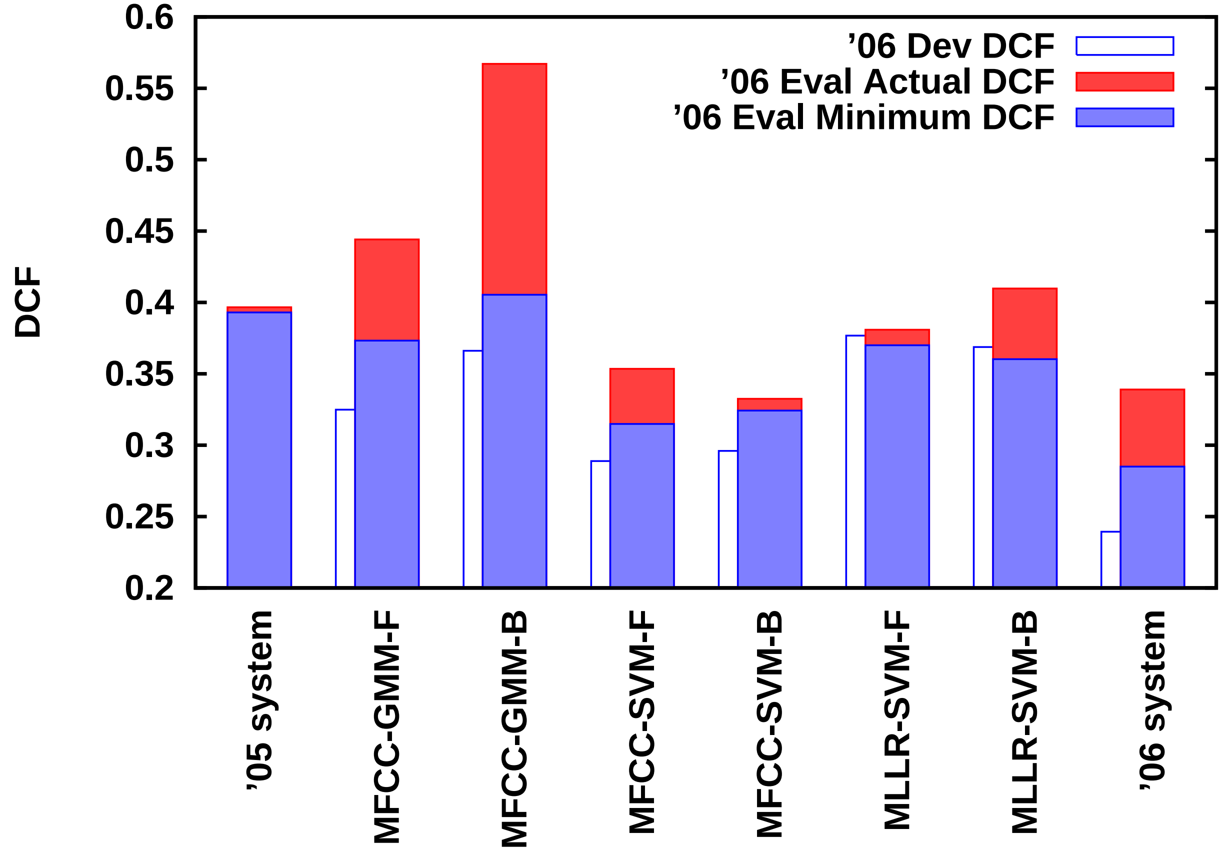
Note: SRE'05 data is used with self-made test-target trial index

SCORE FUSION (Eval)

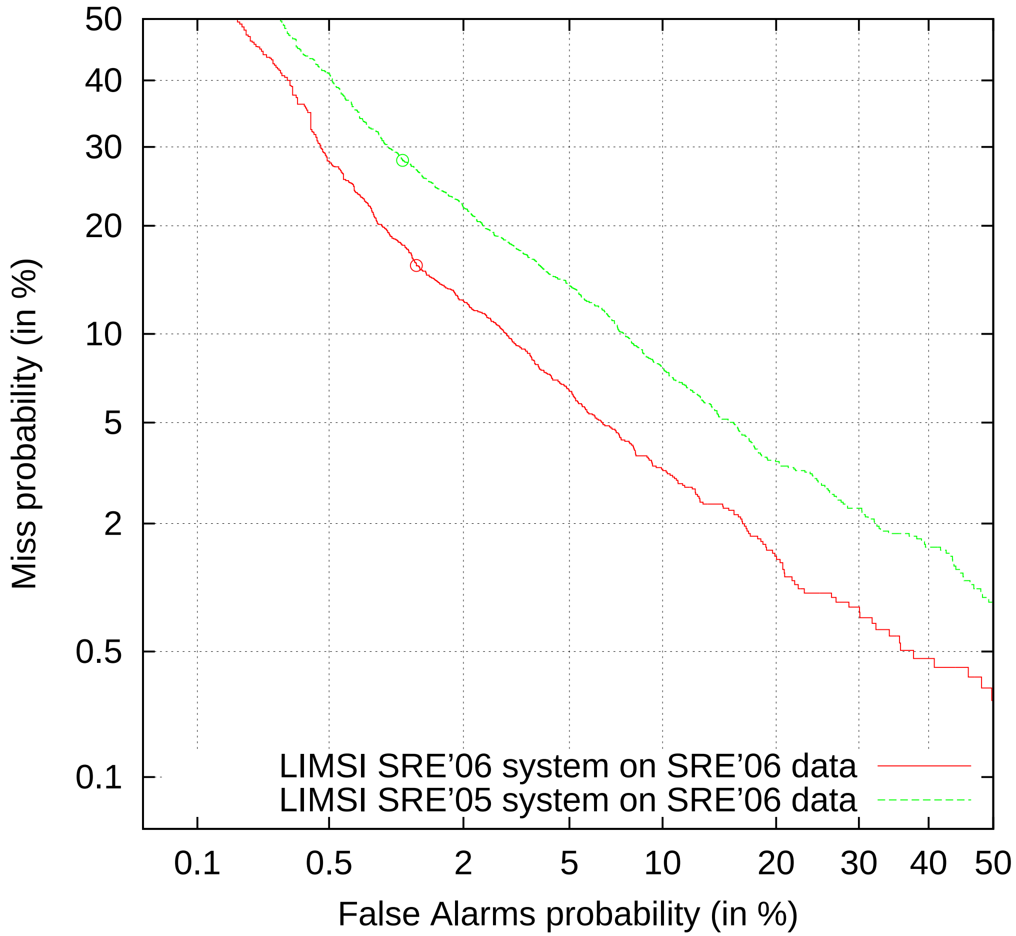


THRESHOLD SETTING

Actual vs. minimal DCF for ’05 mothballed and ’06 (sub)systems: threshold setting issue mainly for ’06 GMM system



DET CURVE ('06 vs '05 SUBMITTED SYSTEM)



SUMMARY

- Using more data to train UBMs and feature mapping models improves the GMM performance (in development using SRE'05 data)
- Use of MFCC-SVM and MLLR-SVM helps a lot!
- The simplest fusion scheme proved the most robust
- Significant MDC improvement compared to LIMSI SRE'05 system:
  0.393 $\Rightarrow$ 0.285, i.e. 27% rel. reduction
- but LESS obvious improvement in actual DC:
  0.397 $\Rightarrow$ 0.339, i.e. 15% rel. reduction only
- Large difference between actual DC and MDC this year