Conversational Biometrics Group
IBM Research

IBM detection systems NIST-2005 SRE

Jiri Navratil, Jason W. Pelecanos, Ganesh N. Ramaswamy, Mohamed K. Omar, Ran D. Zilca

Conversational Biometrics Group
IBM T. J. Watson Research Center
Yorktown Heights, New York
http://www.research.ibm.com/CBG

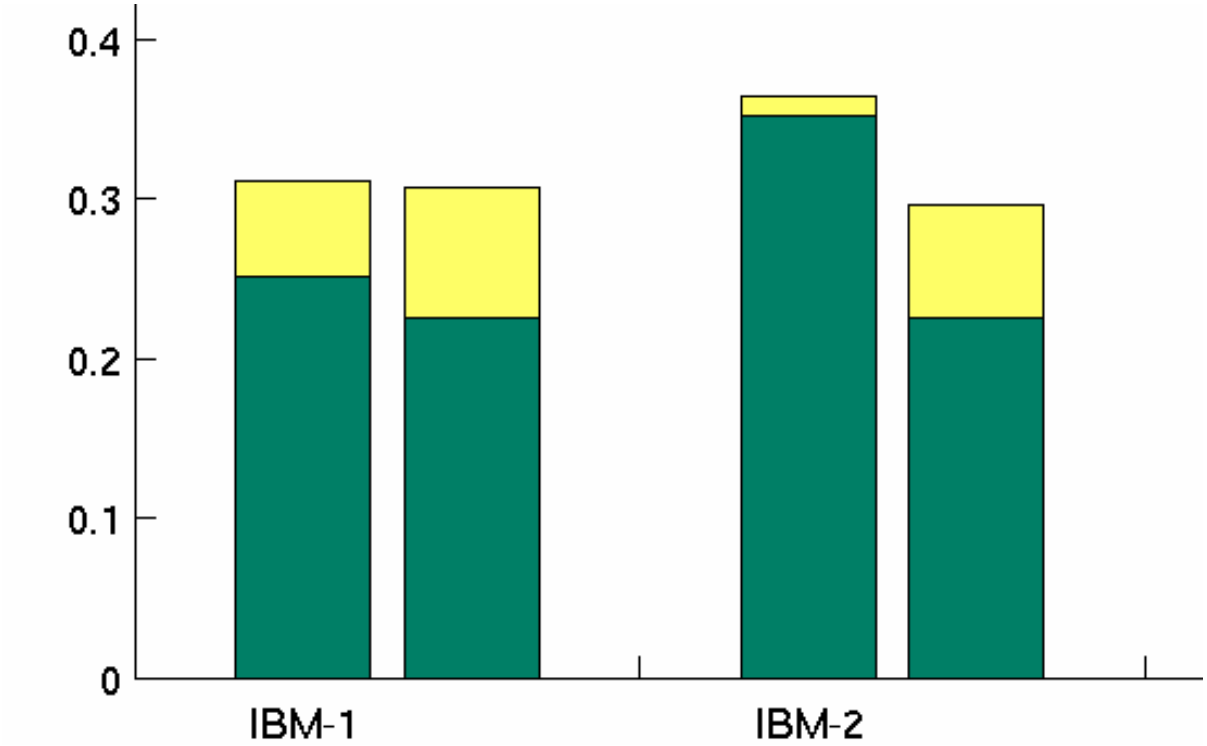
NIST-2005 Speaker Recognition Workshop
June 7-8, 2005

Outline

- Intro
- GMM system
- SVM system
- Fusion

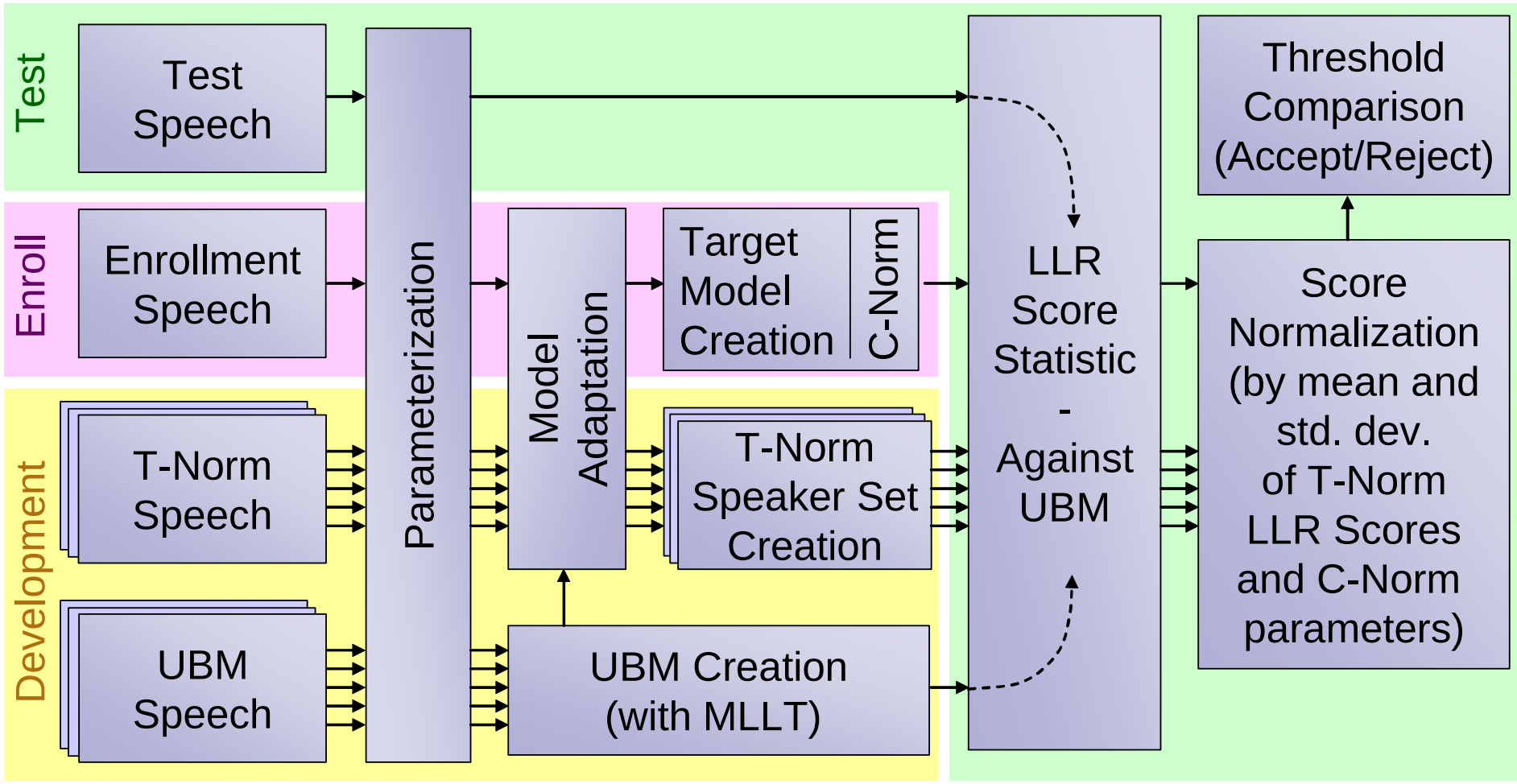
Introduction

- Two systems in the 1conv4w-1conv4w
- System = score-fused GMM/SVM (linear, ANN)



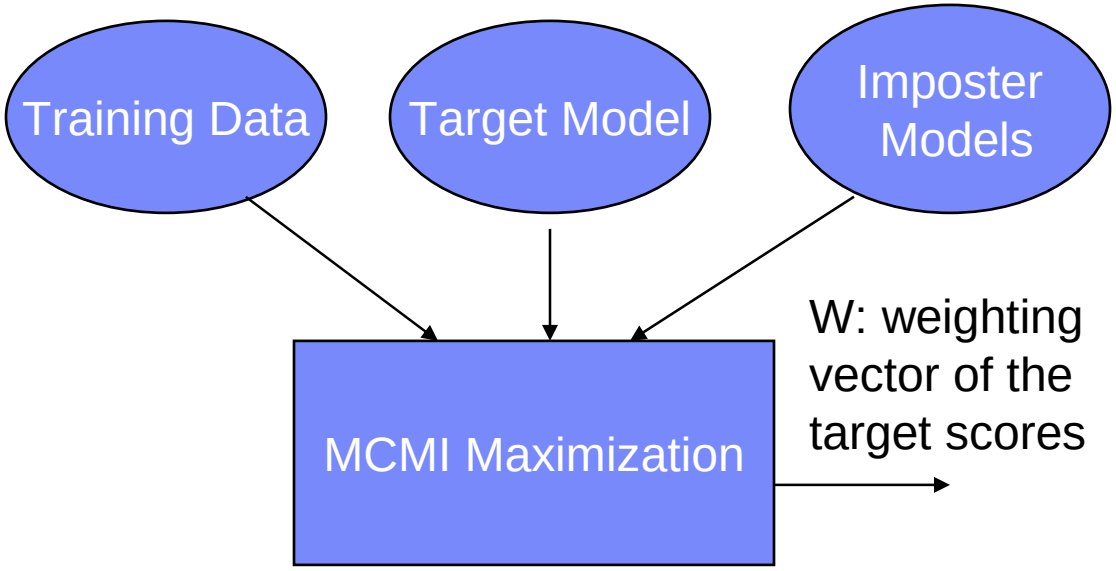
(2005 Common Evaluation Condition)

GMM Overview: Basic System



GMM Overview: MMI Variant

- Maximum Conditional Mutual Information criterion
- Target-dependent scaling of variances in each Gaussian



$$\hat{W} = \arg\max_{w} I(S, P)$$

$$where \quad S = W^{T} F, \; P \; is \; the \; speaker \; ID,$$

$$and \quad F \; is \; the \; model \; score \; vector.$$

Omar, K.M, et al., "Maximum conditional mutual information modeling for speaker verification," submitted to Interspeech 2005

GMM System Performance

- T-Norm: Cellular + Landline speakers from 2001 and 2004 NIST eval
- C-Norm: 5 gender-matched channels
- 1-iteration MMI
- Symmetric scoring (linear fusion of forward and reverse system)

	T-Norm	T-Norm + C-norm	T+C-Norm + MMI	+ Symmetric Scoring
NIST04 Core Task	M: 40.8 /3.0% F: 48.6 /14.3%	M: 39.8 /11.5% F: 47.7 /11.7%	M: 39.2 /10.8% F: 43.6 /11.2%	M: 38.0 /11.0% F:41.9 /10.8%
NIST04 Common Task	M: 36.9 /11.8% F: 43.7 /12.1%	M: 34.3 /11.8% F: 39.2 /9.8%	M: 36.3 /11.8% F: 36.1 /9.6%	M:32.8 /11.8% F:35.6 /9.3%

Values are DCF x 10e-03 / EER for Male and Female subsets

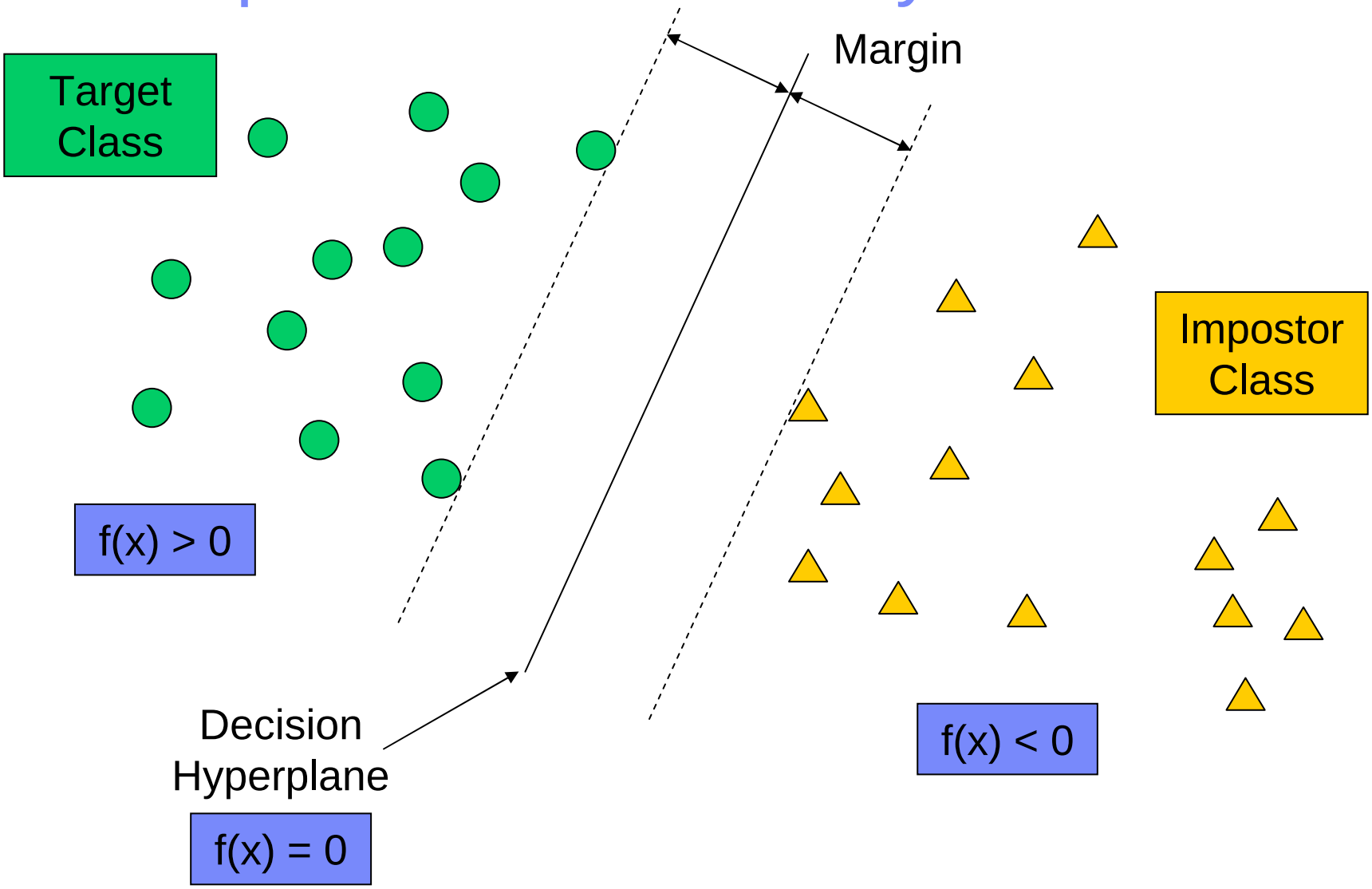
Base SVM System Description

- SVM system implemented using SVMTorch from IDIAP (Collobert and Bengio)
- Implementation of the GLDS Kernel (Campbell, et al)
- Feature Mapping and C-Norm (Reynolds)

SVM Developments

- Feature Warping for SVMs
- Full covariance SVM modeling
- Feature mapping using LPCCs
- Reciprocal Scoring
- C-Norm
- SVM enlarged delta window

Developments - SVM Analysis



SVM Analysis

- SVM Kernel Evaluation

$$f(X) = \sum_{i=1}^{N} w_i c_i K(X, X_i) + d$$

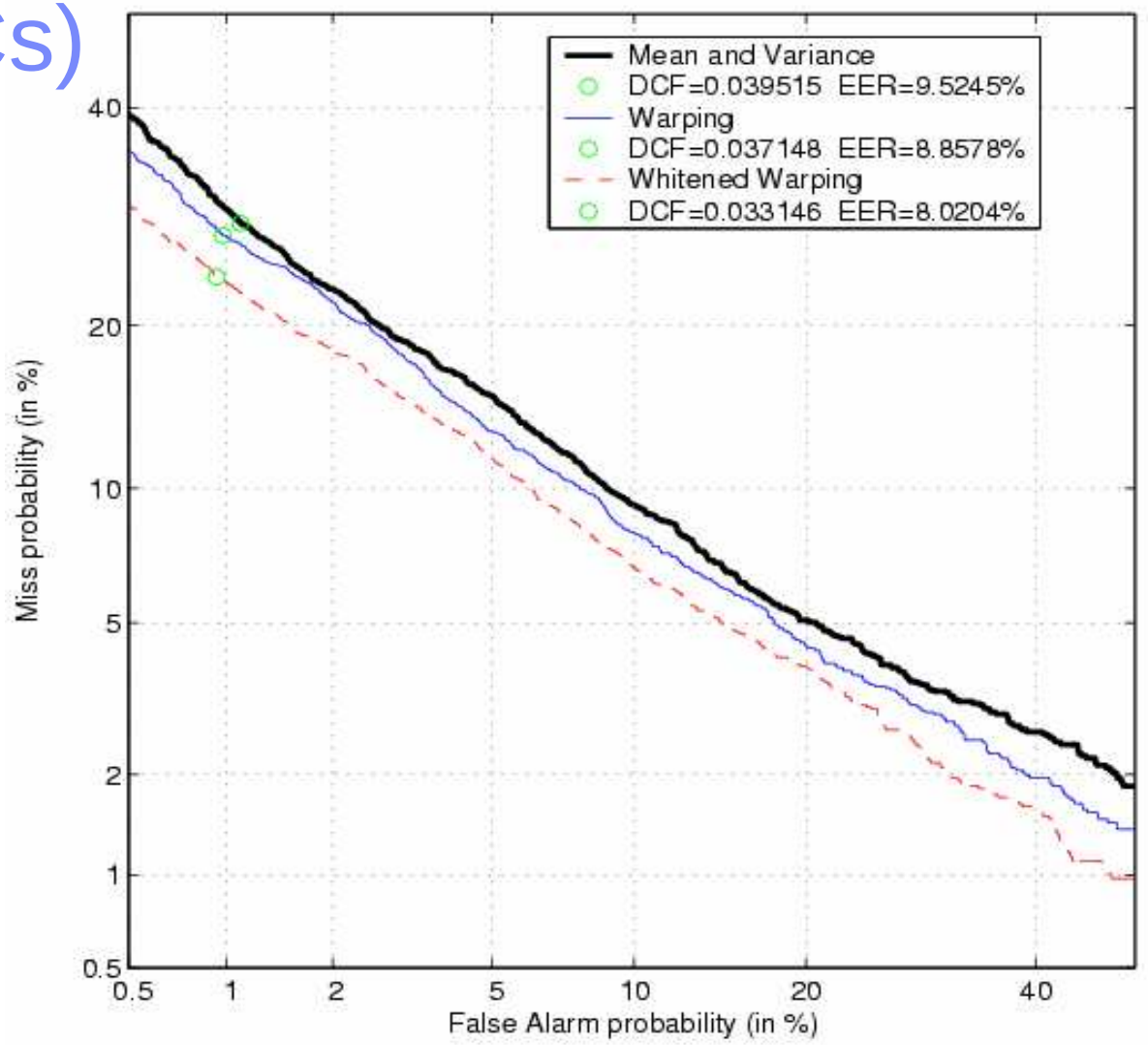
$$K(X, X_i) = \bar{b}(X)' \boldsymbol{R}^{-1} \bar{b}(X_i)$$

$$\bar{b}(X) = \frac{1}{T} \sum_{t=1}^{T} b(\boldsymbol{x}_t)$$

$$\text{where } X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T\}$$

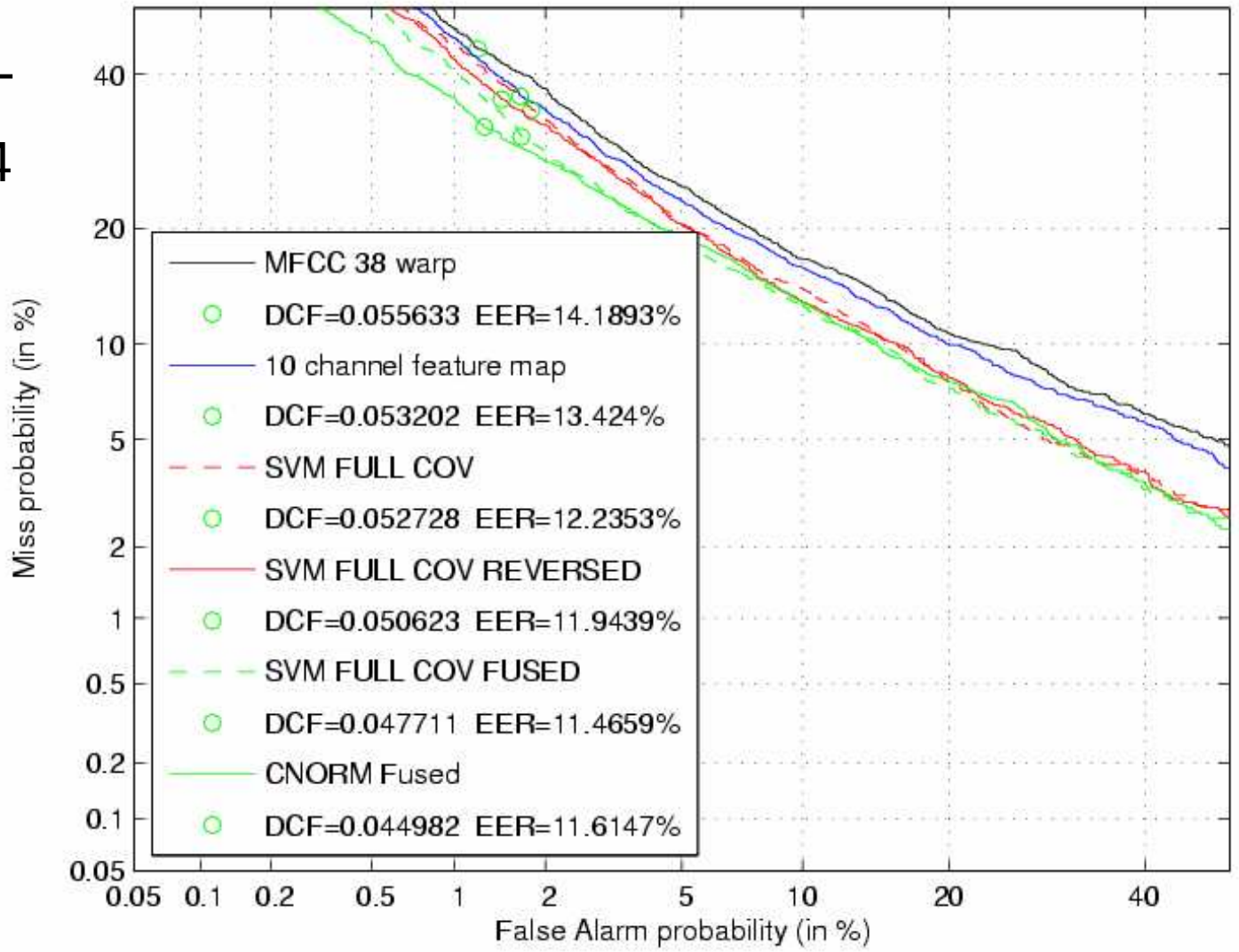
SVM Analysis – FW / Whitening (MFCCs)

NIST 2003 All Condition



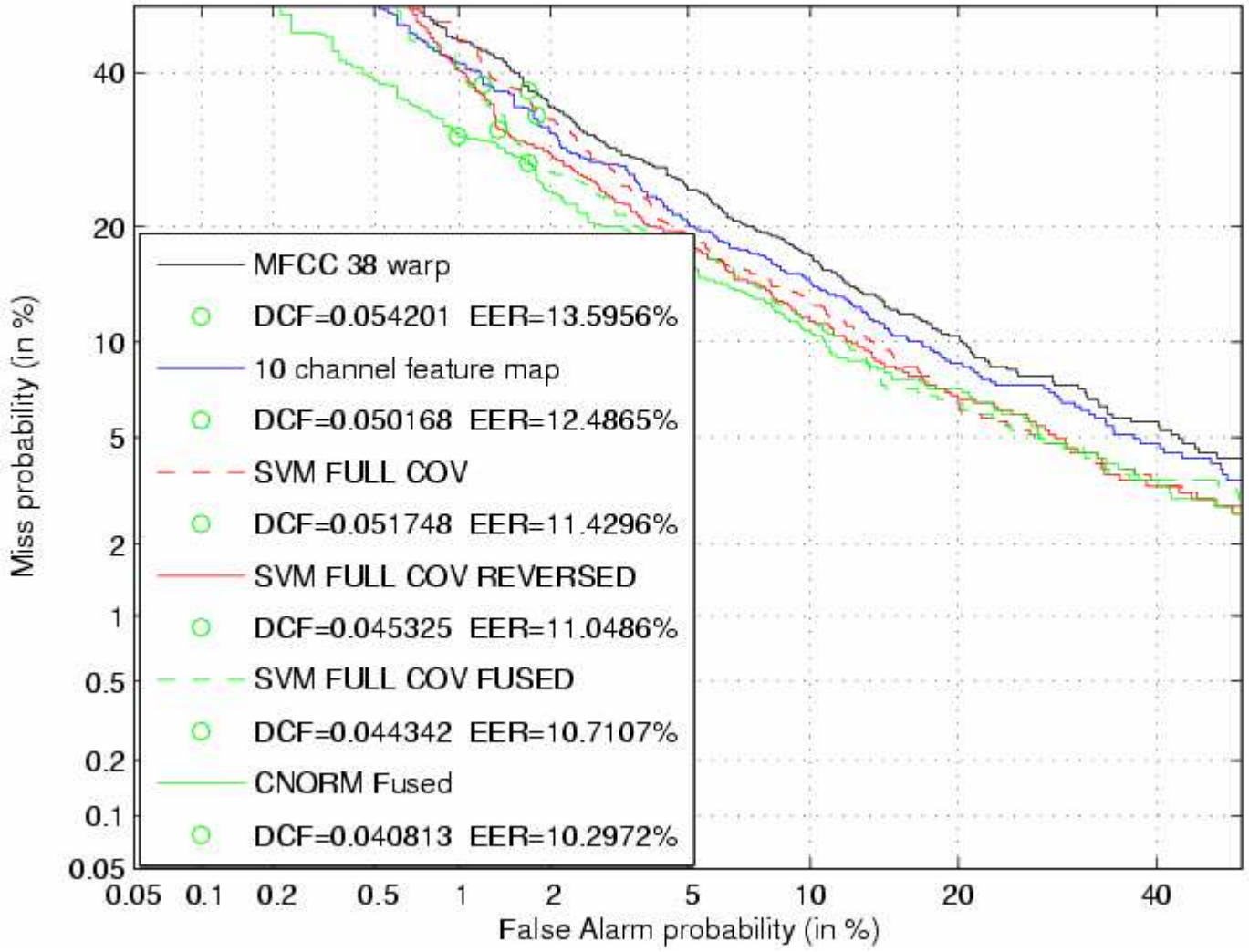
SVM Developments – MFCCs (core task)

NIST 2004



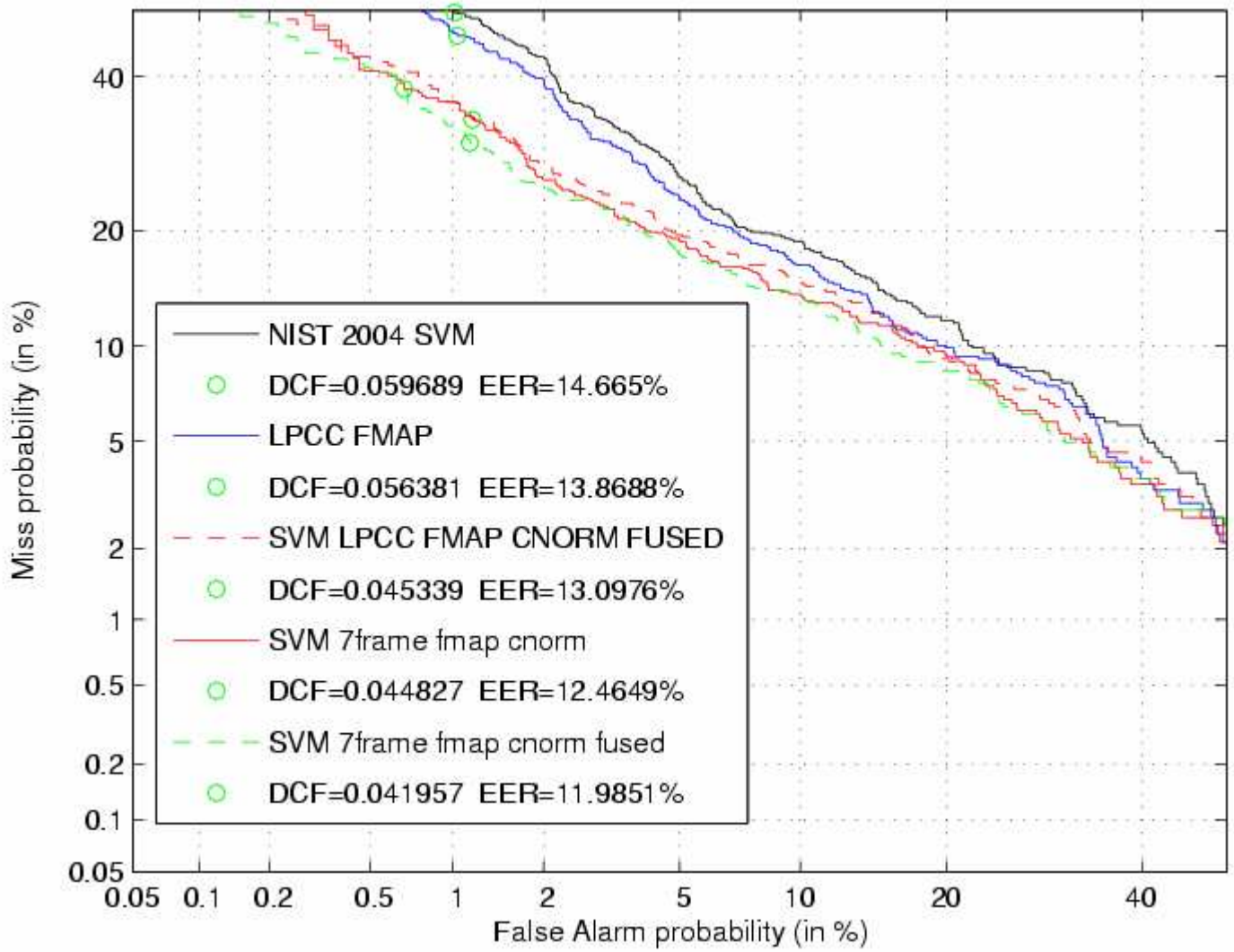
Overview – MFCCs (common task)

NIST 2004



Developments – LPCCs (common task)

NIST 2004



SVM Conclusions

- Feature warping is a useful feature normalisation technique within the SVM framework for NIST 03/04
- Empirical results suggest that full covariance modeling aids performance for MFCCs (improvement was not observed for LPCCs)
- Reciprocal scoring boosts SVM and GMM results
- For the GLDS SVM configuration, C-Norm tends to provide a consistent performance gain.
- An observation of the GLDS kernel is that it is more important to have more speaker specific information contained in each feature vector than it is to have diverse speech frames.

Fusion

- IBM-1 = Linear fusion {GMM, GMM+MMI, SVM-MFFC, SVM-LPCC}
- IBM-2 = ANN fusion, 2-Layer FF 4-3-1 Net, trained on NIST04 core task (gender cross-eval)