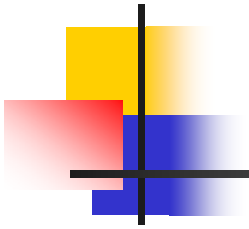
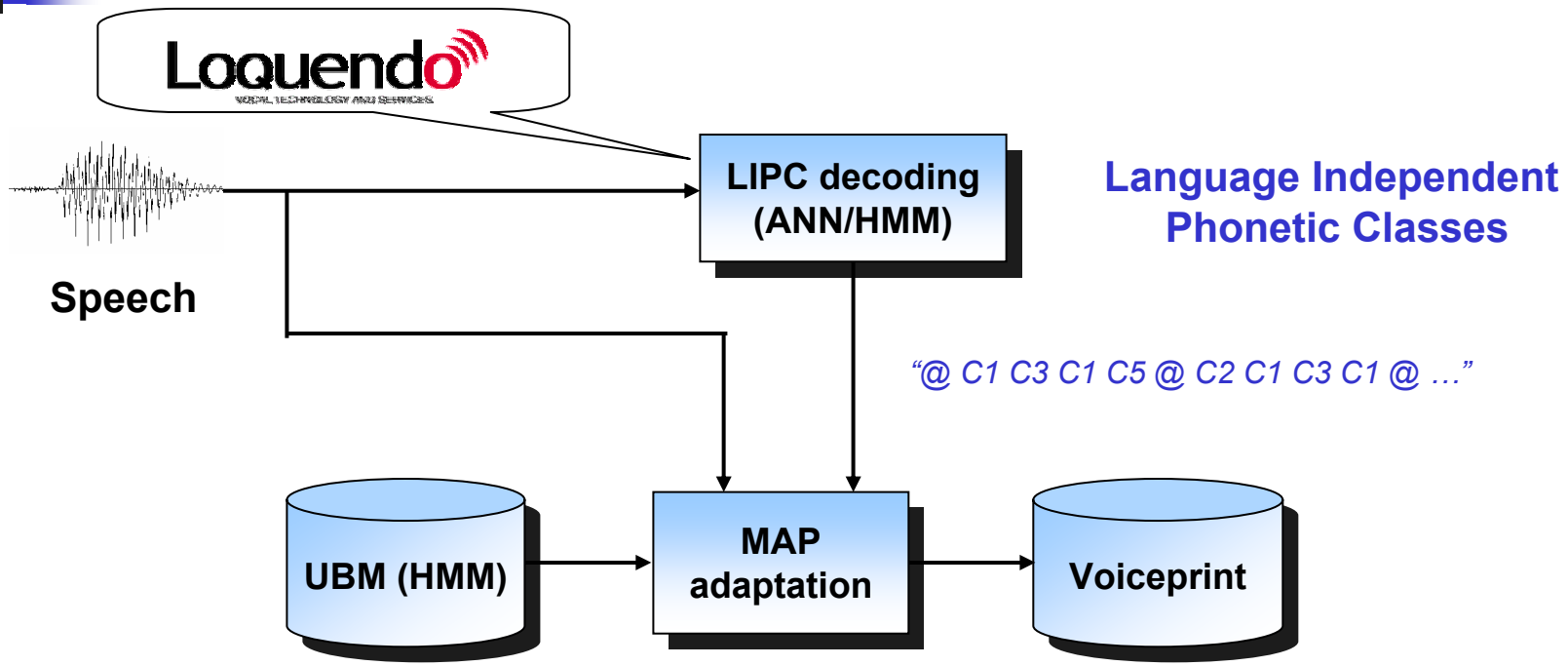


SRE-05 Evaluation

Emanuele Dalmasso and Pietro Laface

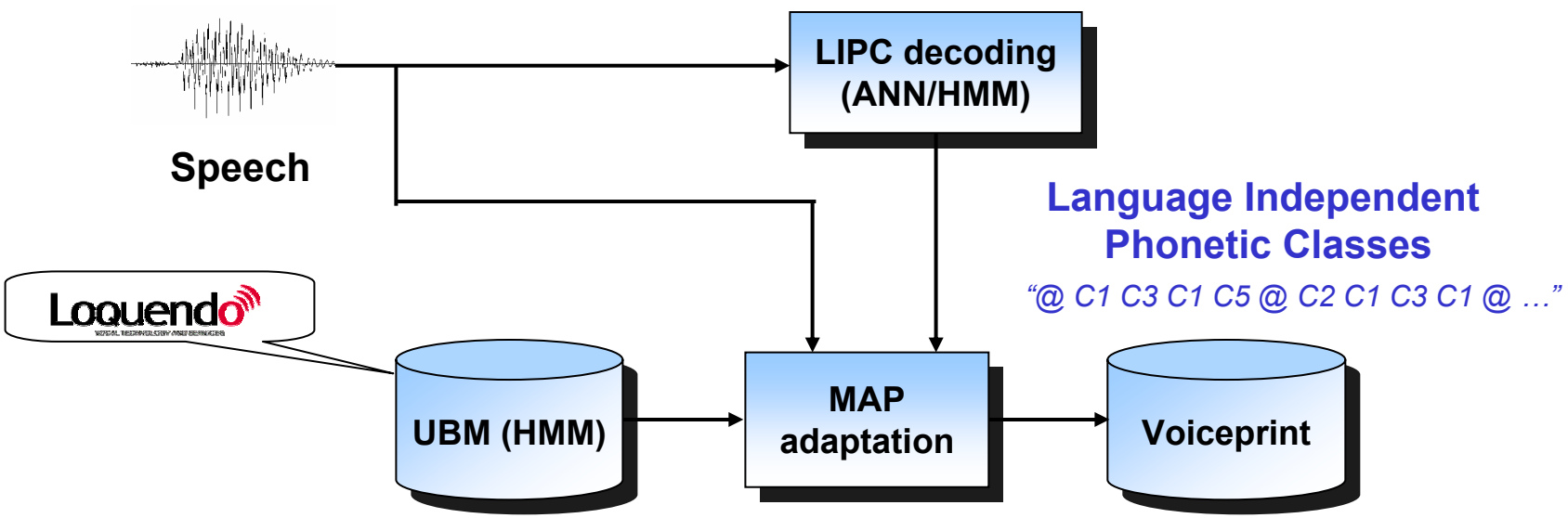
Dipartimento di Automatica e Informatica
POLITECNICO DI TORINO

System description – Enrollment



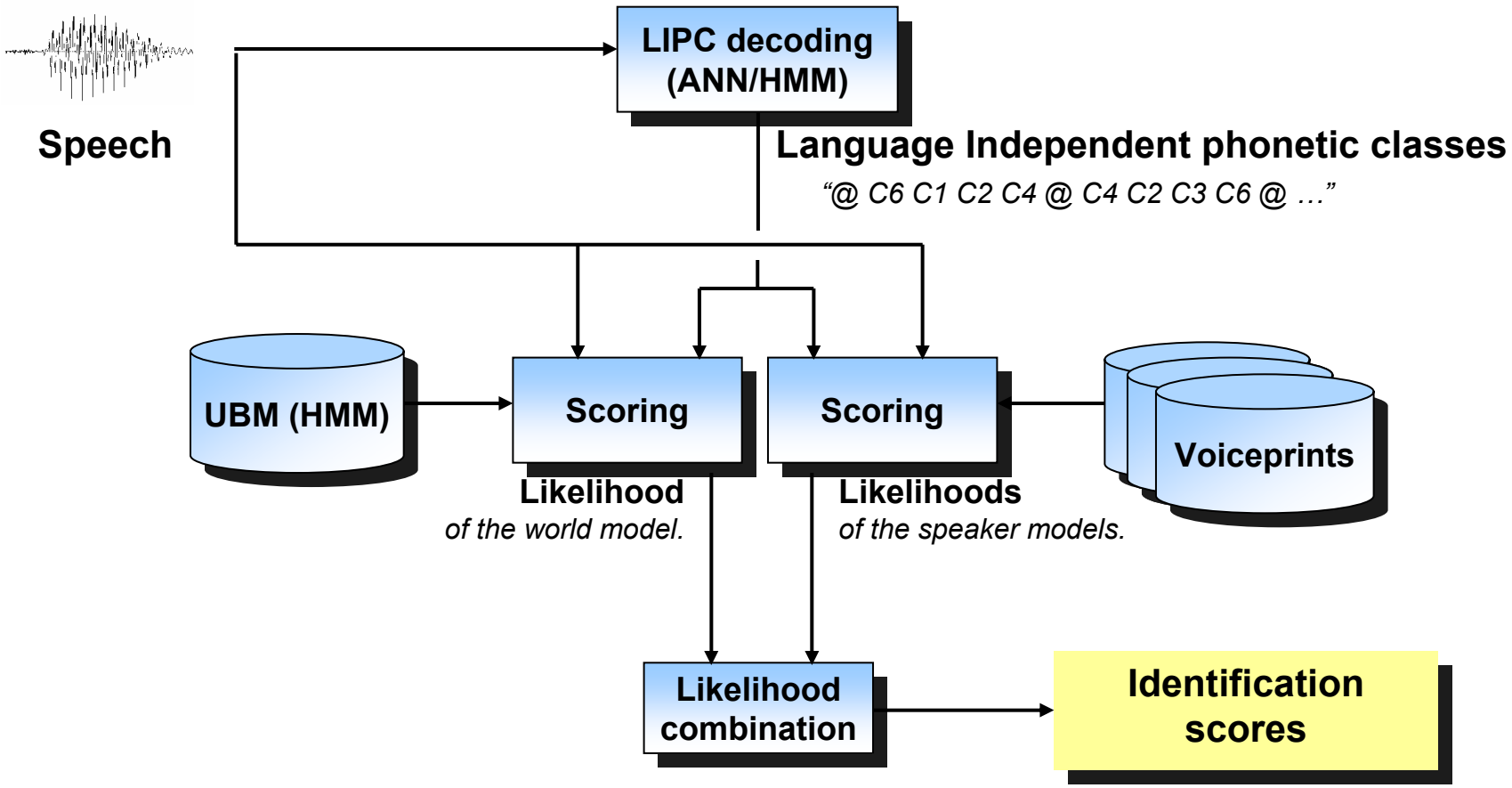
- **Phonetic decoding of the utterance producing phonetic class segments**

System description – Enrollment



- The UBM and the voiceprints are phonetic Gaussian Mixture HMMs
- Gender independent UBM trained on 20 hours of speech of 10 different languages

System description – Testing



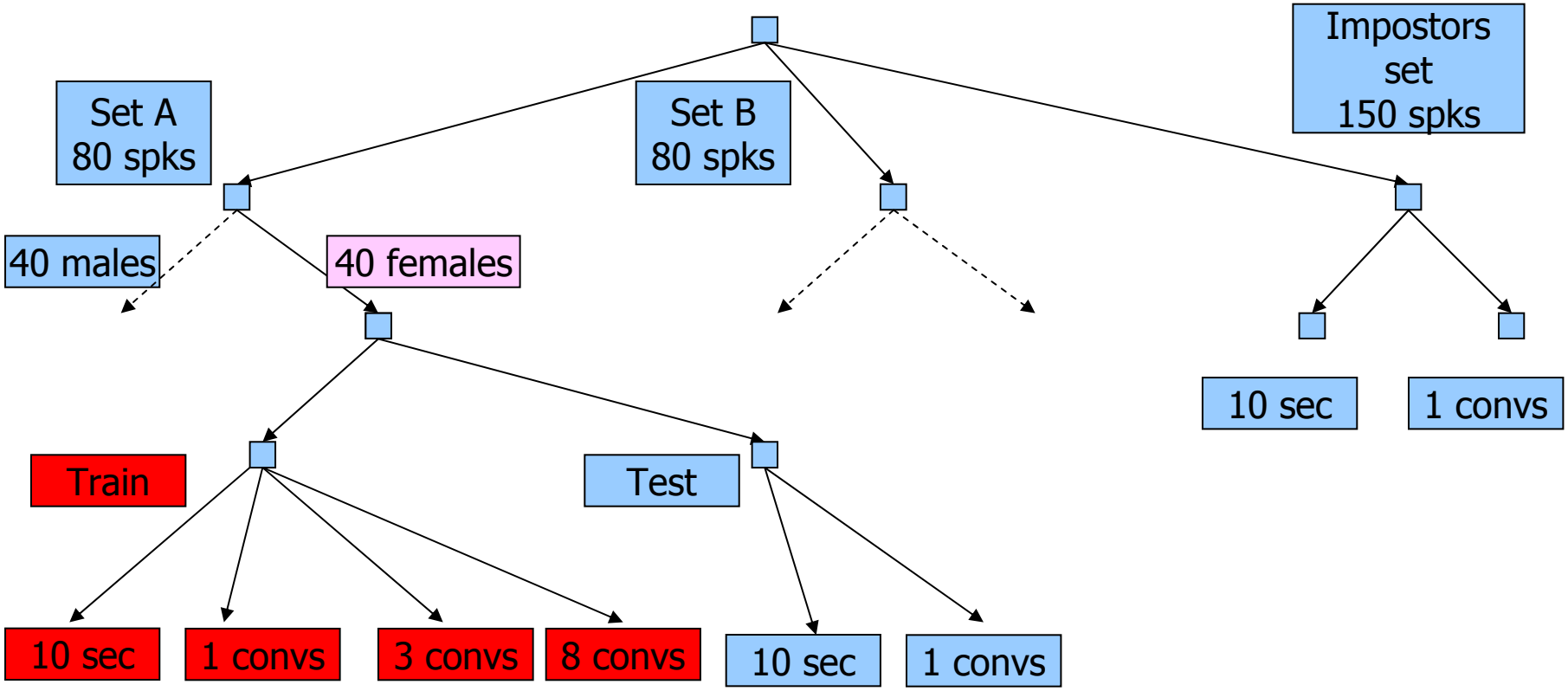
Features and models

- 18 cepstral and 18 delta cepstral parameters
- Feature warping to a Gaussian distribution
- 3 state models left to right HMM for each unit
  - one silence state

Development Setup

- NIST SRE04 Corpus divided in three sets:
  - One set (A) for training speaker models and for the true speaker detection trials
  - One set (B) for collecting normalization statistics
  - One set with impostors data only
- Training and normalization sets swapped to increase the number of tests

Partition of the Development Corpus (SRE04)



Trained models

- For conditions 10 sec, 1 conv, and 3 convs
  - 2 models per speaker
- For condition 8 convs
  - 1 model per speaker
- Z-norm and T-norm statistics collected on the complementary set of speakers
  - Z-norm performed on the same conditions of the test
  - T-norm performed on the same conditions of the training

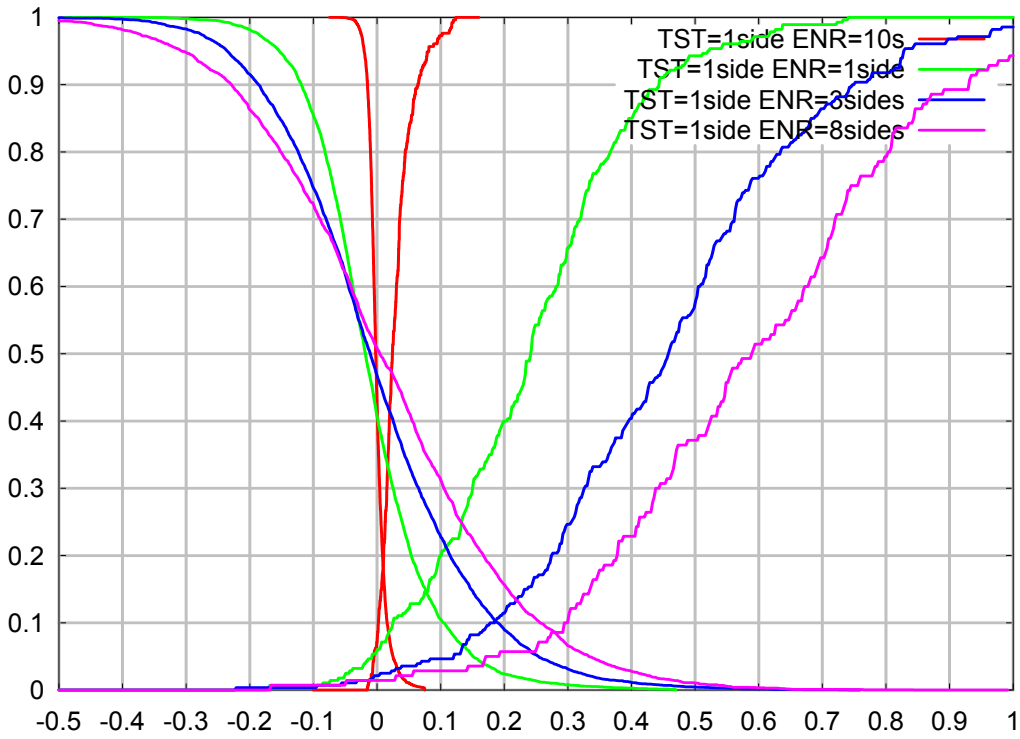
Tests

- Total number of tests for each condition, using one set (A or B) and the Impostor set of speakers:
  - ~5000 female impostor tests
  - ~3300 male impostors tests
  - 300 female true speaker tests
  - 300 male true speaker tests

  - 150 true speaker tests for the 8 convs condition

FA/FR without normalization

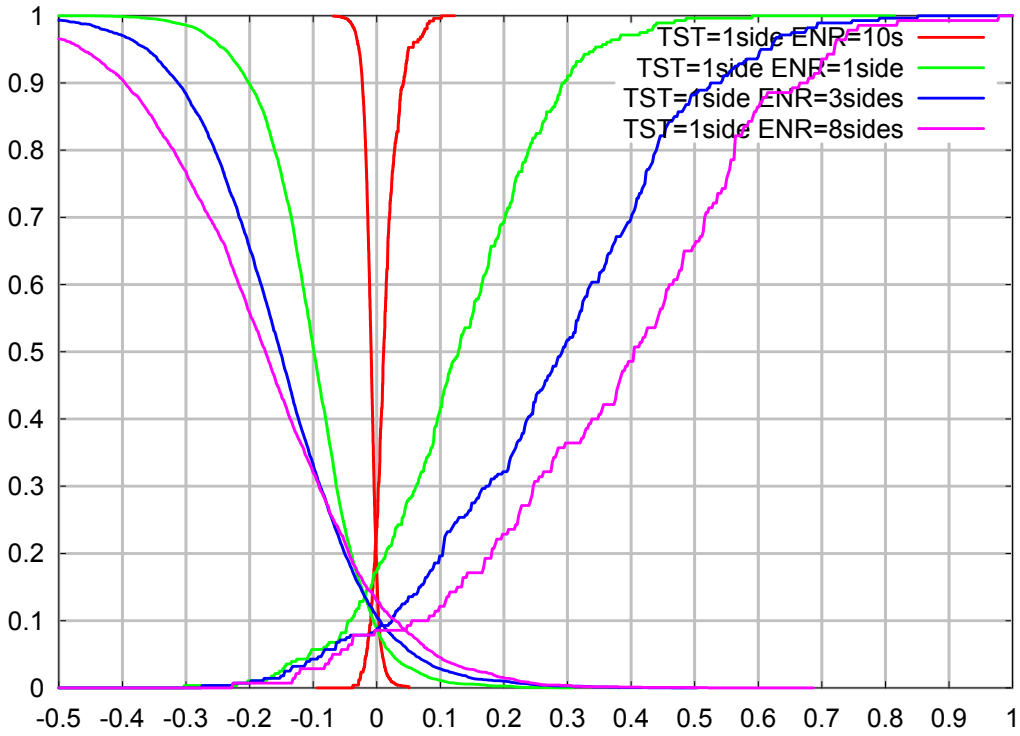
Cumulative Distributions of the Impostor and True Speaker Scores



- Increasing the number of sides
  - Lower EER
  - Decreasing (better) slopes
  - Right shift of the EER

FA/FR with World-Model adaptation

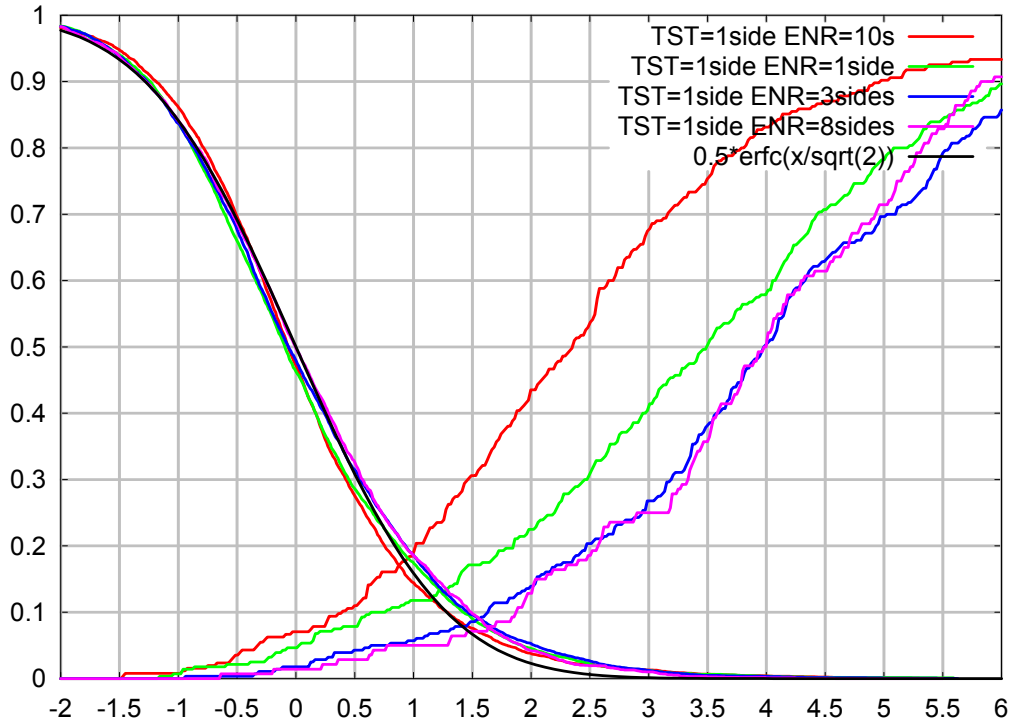
Cumulative Distributions of the Impostor and True Speaker Scores



- Control experiment
  - UBM adapted using the SRE04 data
  - Same offset ~0.0
  - Different slopes remain

FA/FR Z-Norm

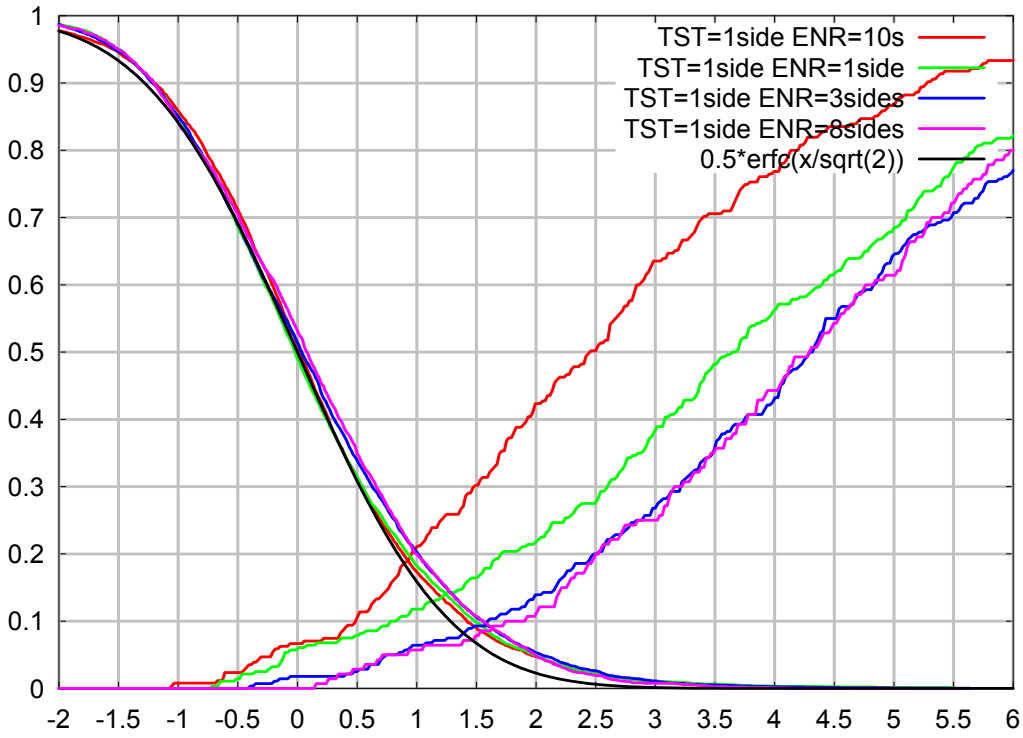
Cumulative Distributions of the Impostor and True Speaker Scores



- Z-norm
  - Normalizes the impostor scores
    - ~ same slopes
  - Still right shift of the EERs
  - DCF threshold doesn't change

FA/FR ZT-Norm

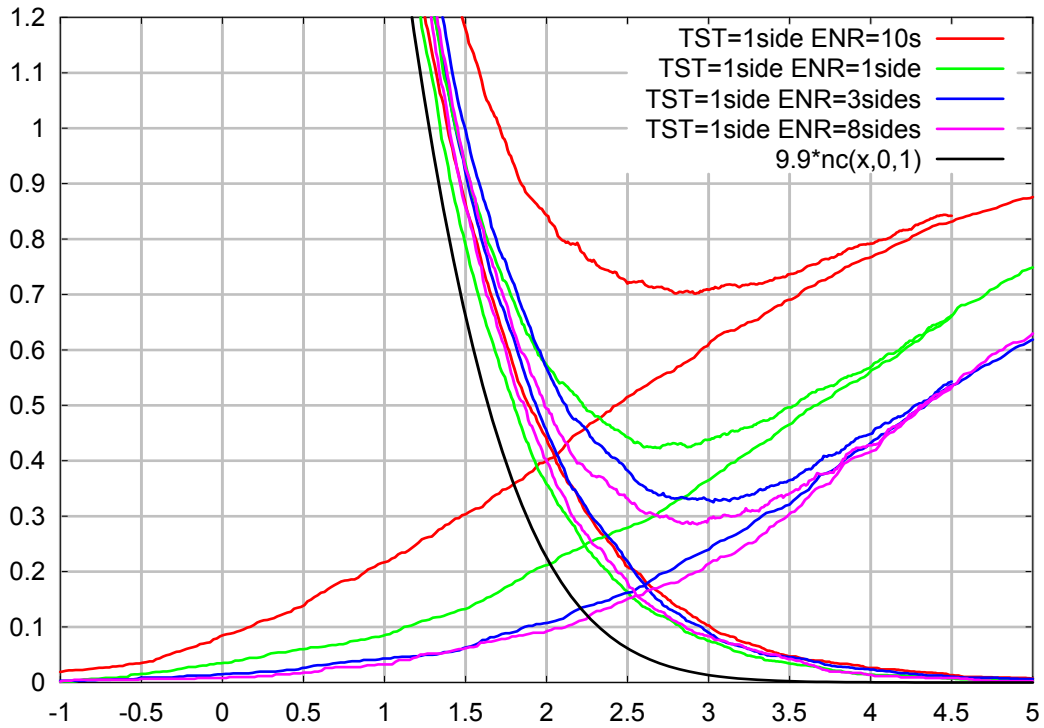
Cumulative Distributions of the Impostor and True Speaker Scores



- ZT-norm
  - Better behavior at low FA
  - More evident in DET plots

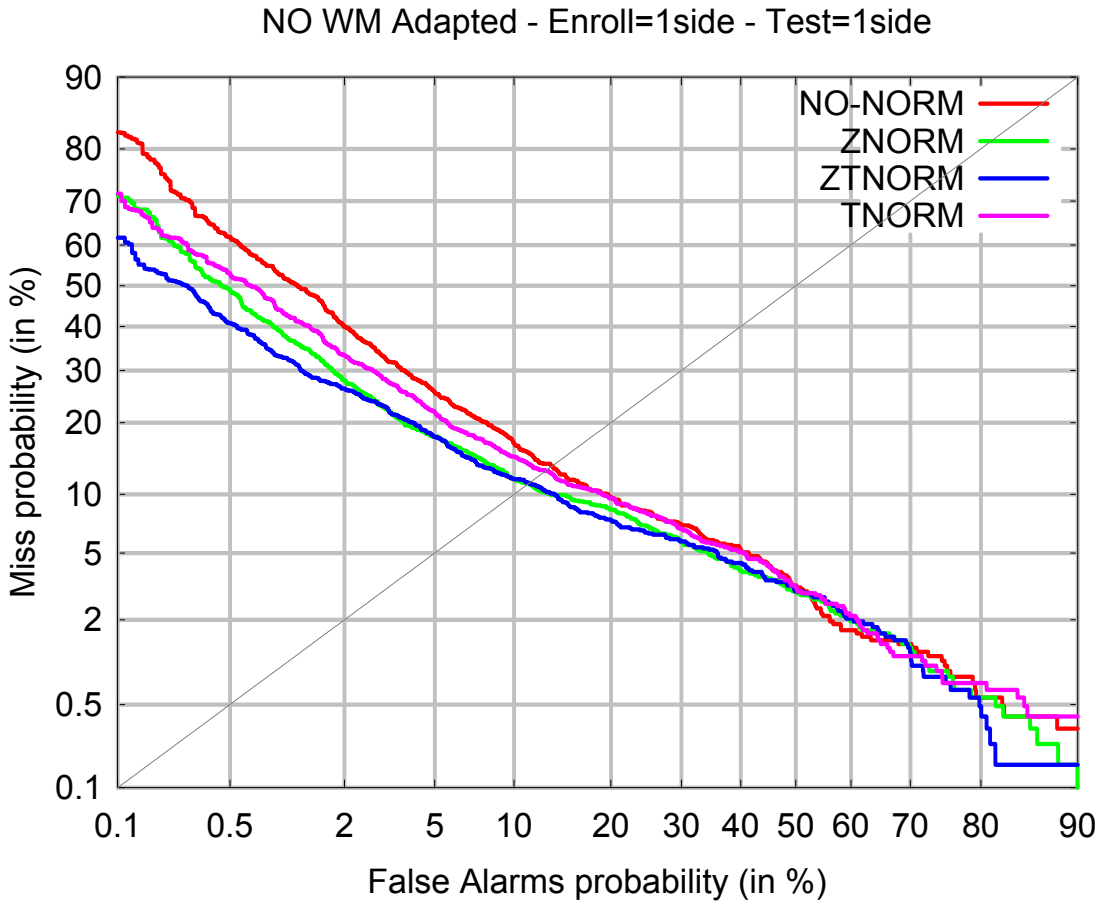
DCF using ZT-Norm

Cumulative Distributions of the Impostor and True Speaker Scores and the Corresponding DCF plots



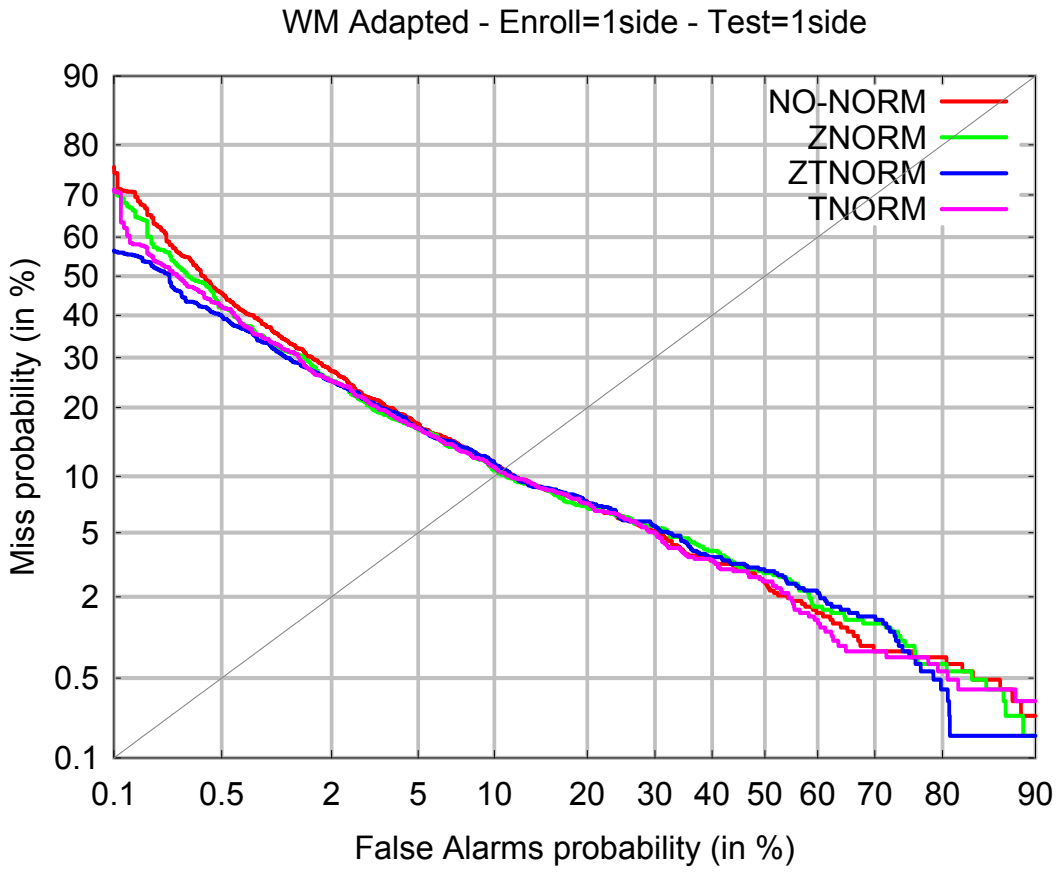
- DCF plots
  - FR scores are weighted 9.9 according to NIST specifications
  - Empirically, 2.75 is a stable threshold for all the conditions

Effects of the normalization techniques



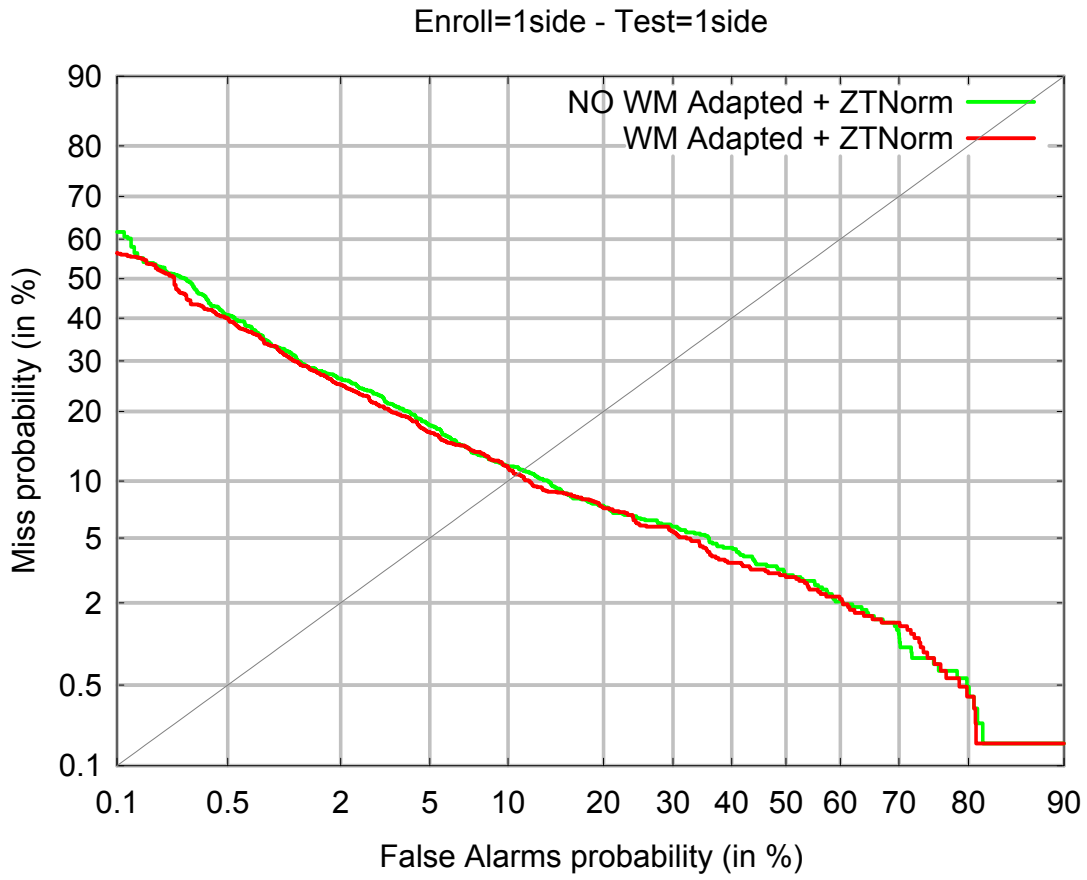
- Not adapted UBM
  - Z-norm alone most effective than T-norm
  - ZT-norm improves the DCF performance

Effects of the normalization techniques



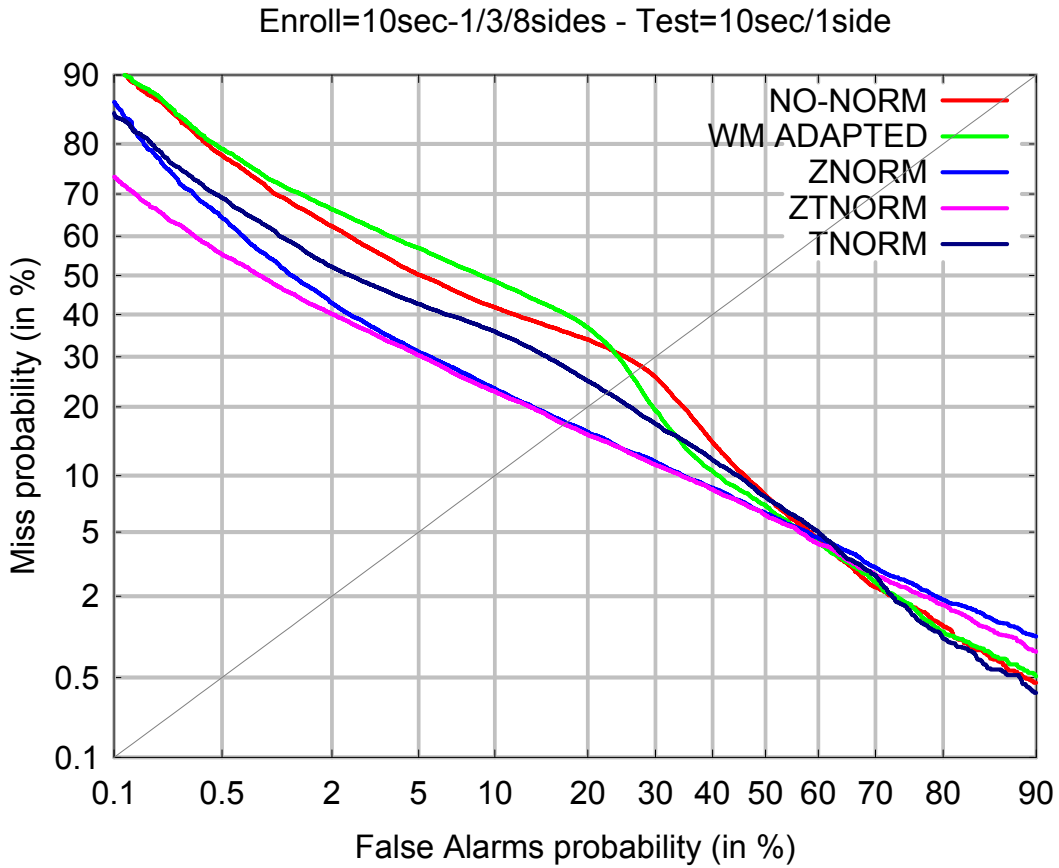
- Adapted UBM
  - Z-norm is less relevant using an adapted UBM
  - T-norm and ZT-norm still give some contribution

Adapted vs non adapted UBM



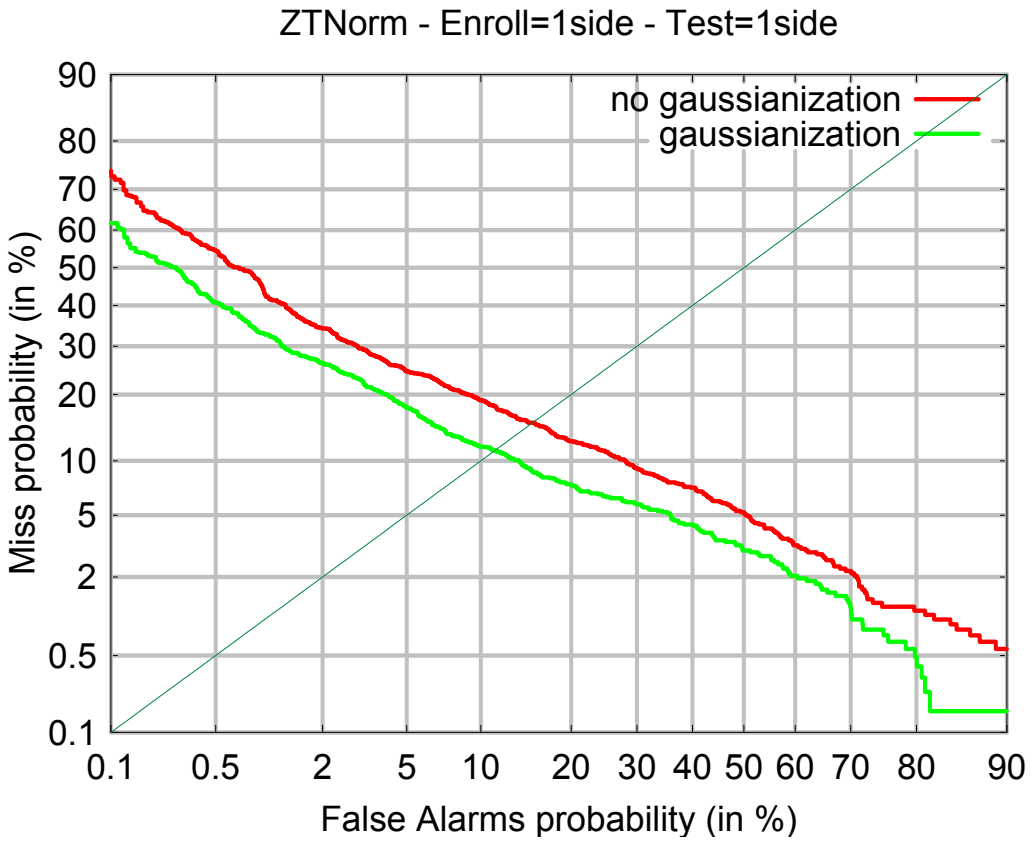
- Using ZT-norm
  - The adaptation of the UMB is not so relevant; marginal improvement
  - Our UBM is trained using corpora completely unrelated to the speaker recognition databases

Comparison of DETs in all conditions



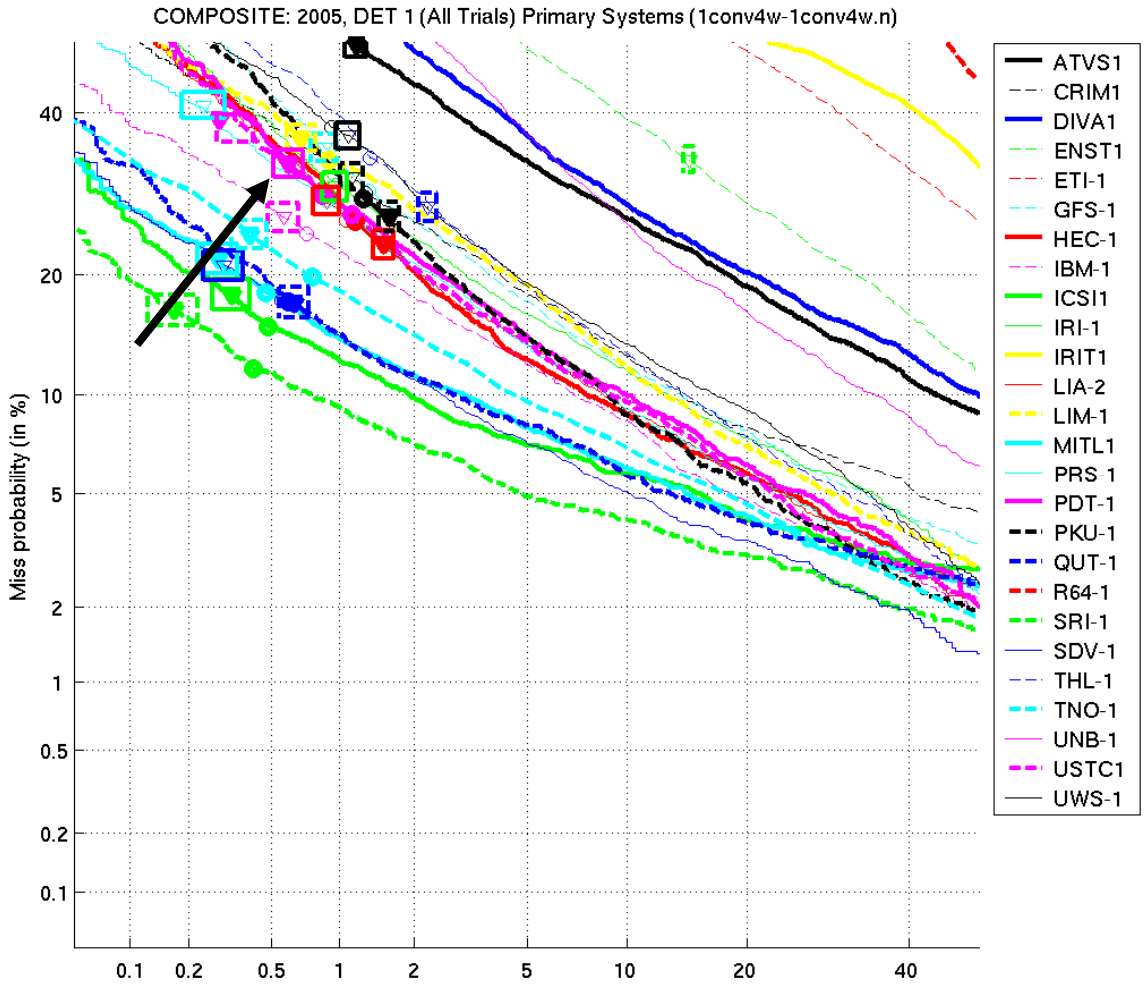
- Adapted UBM doesn't help
- T-norm alone is not sufficient
- Z-norm gives most of the improvement especially for EER
- ZT-norm is effective only for the NIST defined DCF

Effects of feature warping

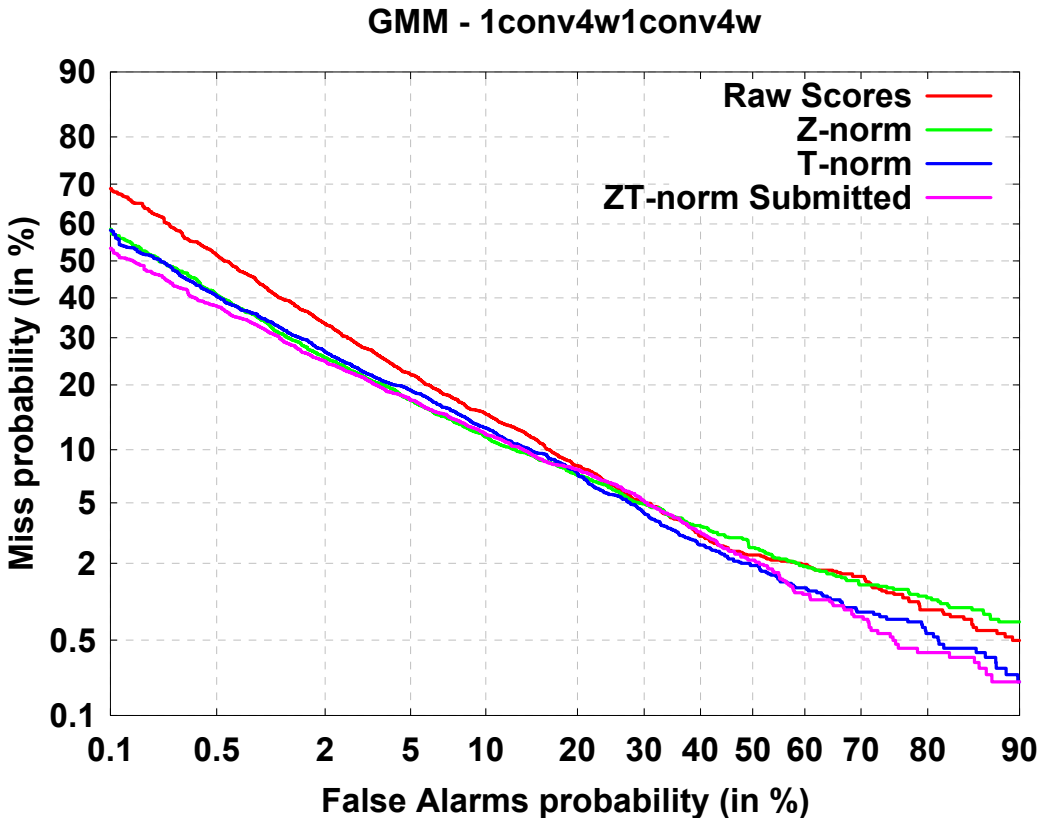


- Gaussianization
  - Substantially contributes to better modeling the Gaussian Mixtures
  - No feature mapping has been used in our tests

Results on SRE05 – All trials

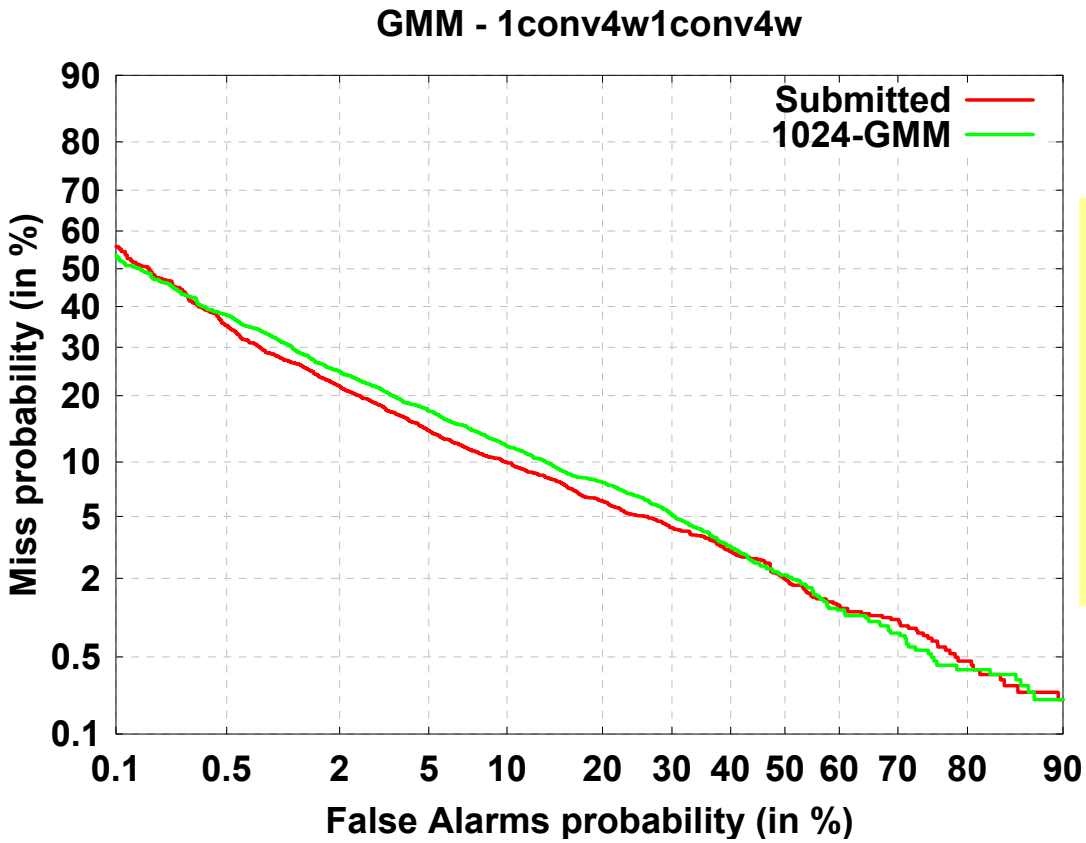


GMM – 1conv4w1conv4w



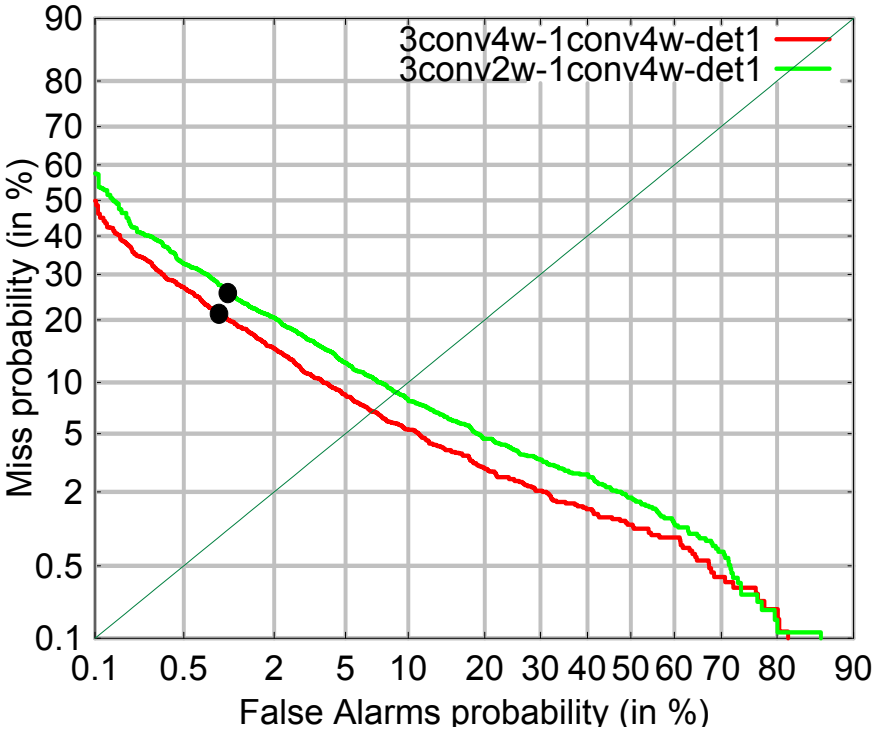
- GMM UBM
  - Trained on SRE04 data
  - Again, Z-norm gives most of the improvement
  - Small improvement with ZT-norm

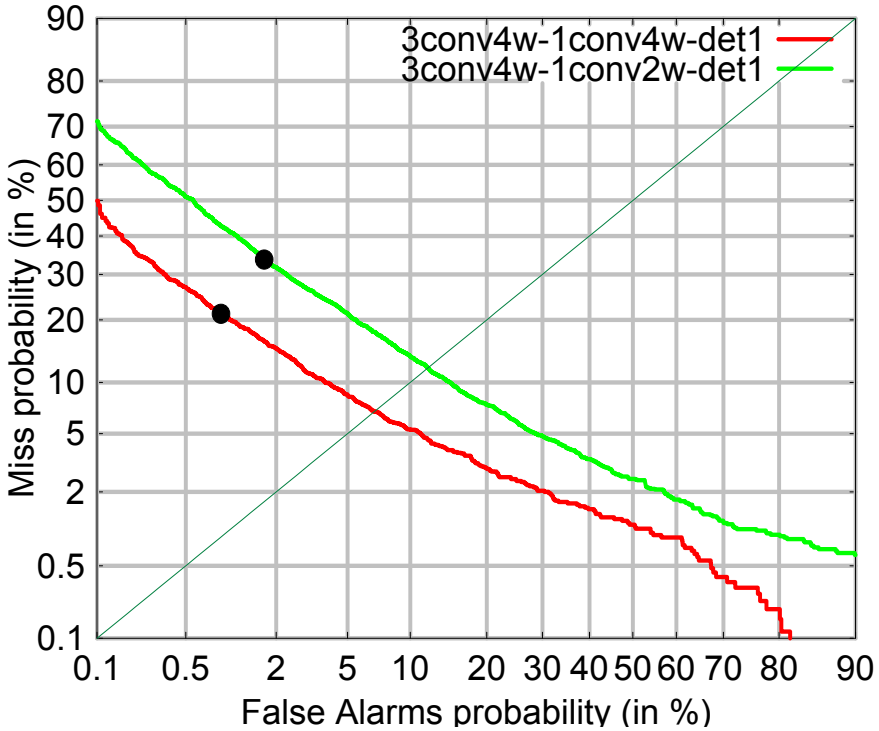
GMM vs Phonetic Based Approach



- Phonetic-based
  - Same approach for text-prompted and text-independent recognition
  - Linguistic information can be exploited
- GMM UBM
  - faster

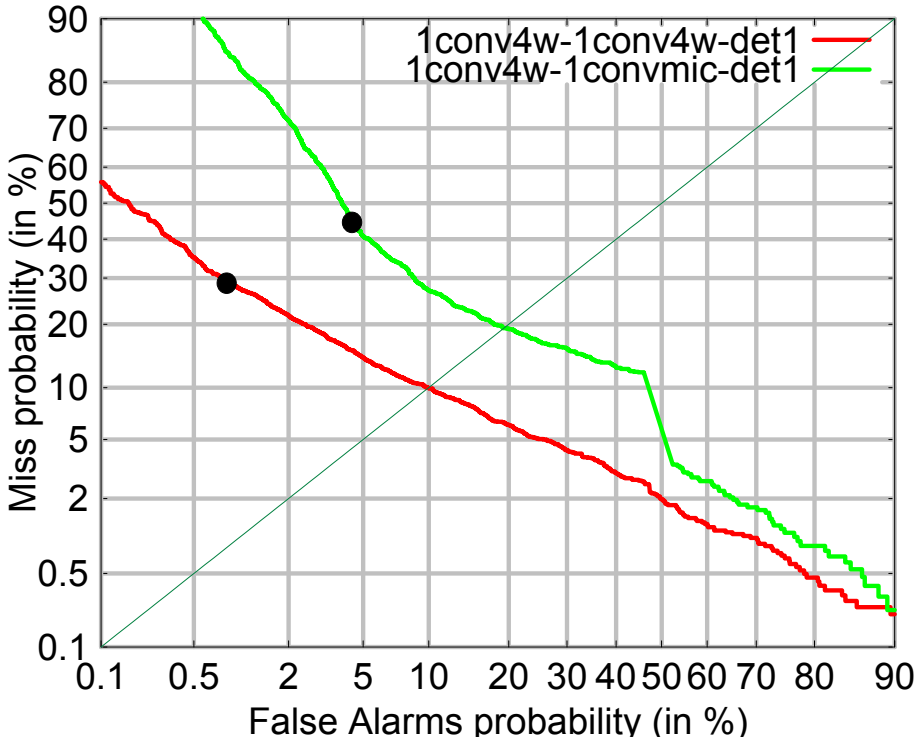
Comparison of 4w vs 2w tests





The segmentation approach is more effective in training than in testing

Post evaluation mic tests



- We didn't submit the mic test results because we didn't use any particular setting or technique for these tests
- Performance decreases, but …

Comparison with 1conv4w-1convmic.n tests