ENST Systems description

Najim Dehak
Eduardo Soto-Sanchez
Patrick Perrot
Gérard Chollet
dehak,esanchez,perrot,chollet@tsi.enst.fr

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

Outline

- Introduction
- Data description
- ENST 1 system
- ENST 2 system
- ENST 3 system
- New systems
- Conclusions and perspectives

Introduction

- ENST has presented three systems in 2005 in 1 train/test conditions :
  - 1conv/1conv → core test
- Our systems have the same modelling but differ in pre and post processing
- Models are obtained by Becars
- Becars results from a collaboration between the University of Balamand and ENST in the context of the Cedre project, a French-Libanese cooperation framework

data description

- World database:
  - Female (Nist 2004):
    - 343 segments – 1side
      - Channel A: 130 segments
      - Channel B: 213 segments
  - Male (Nist 2004):
    - 303 segments – 1side
      - Channel A: 130 segments
      - Channel B: 173 segments
- Normalisation Impostors dataset:
  - 60 impostors for each gender from NIST2004.
- Channel extraction
  - Speech tools: ch_wav, a free software available at http://festvox.org/docs/speech_tools-1.2.0/x44.htm

Modeling and decision

- UBM
  - Gender dependent GMM with 512 Components
  - Used 20% of data for initialisation
- Client models
  - Use a MAP adaptation to estimate the mean of client models
- Decision
  - Decision score is based on log-likelihood ratio

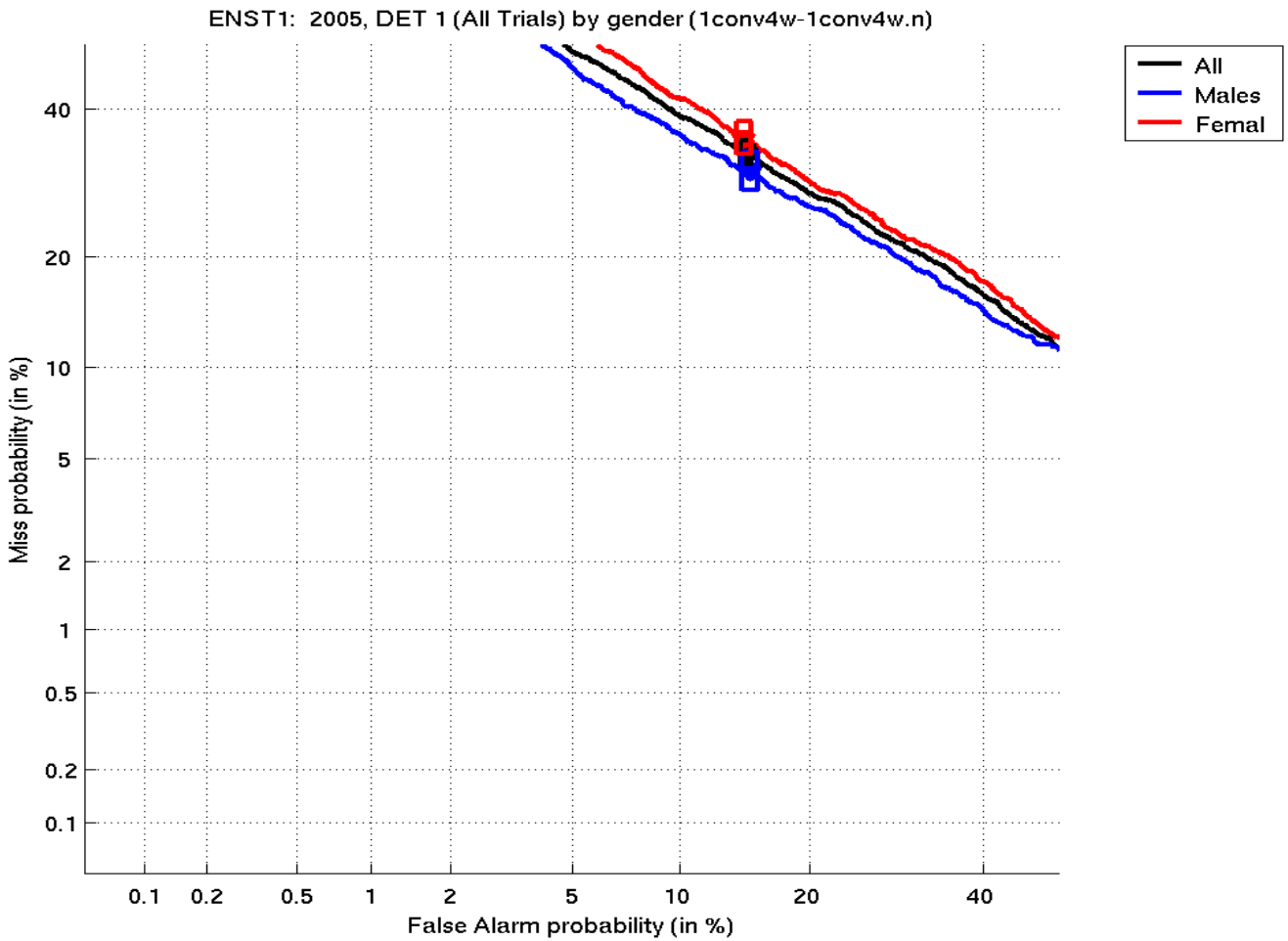
ENST 1

- Pre-processing:
  - Silence removal based on bi-gaussian
  - features extraction
  - Cms (cepstral mean substraction)

- *Acoustic parameters:* 39 coefficients
  - 12 MFCC + energy+ Δ + Δ Δ
  - CMS (windows size 1500 vectors)
  - Features extraction by Spro

- Normalisation:
  - Decision score is based on un-normalised log-likelihood ratio

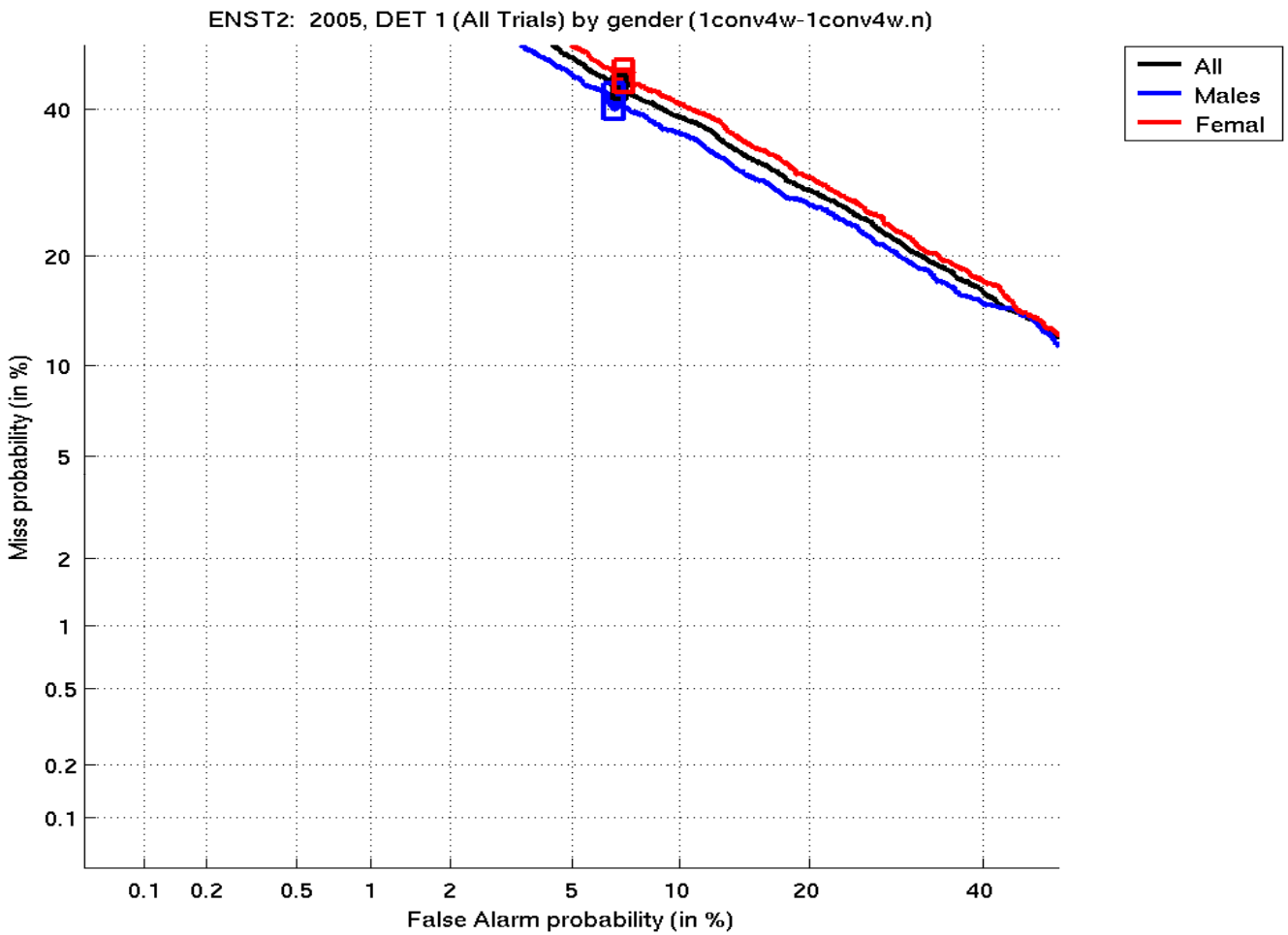
DET Curve : ENST 1



ENST 2

- Pre processing
  - The same settings as the ENST 1 system
- Normalisation
  - Decision score is based on normalised log-likelihood ratio
  - Used T-norm with 60 impostors models for each gender

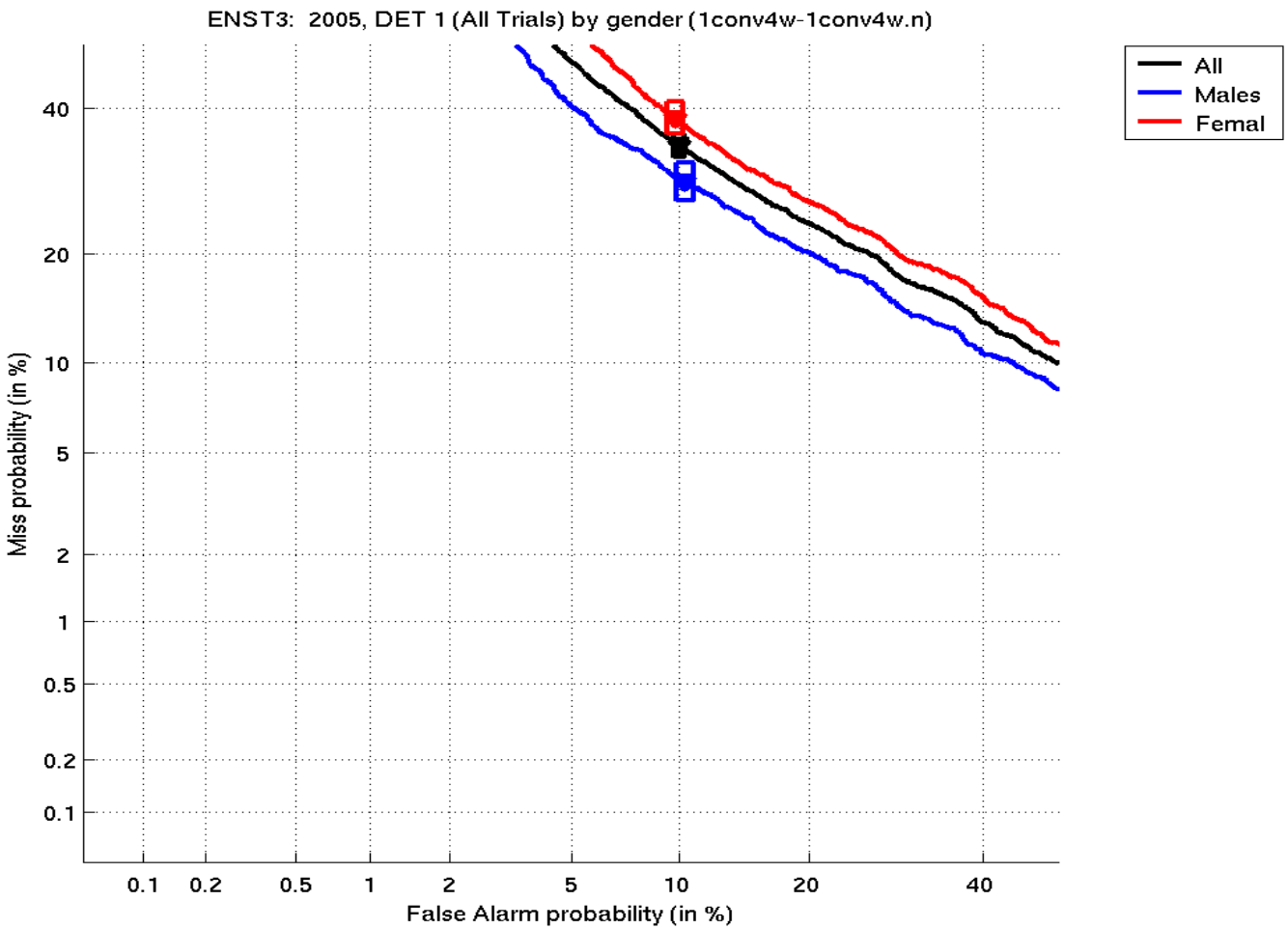
DET Curve : ENST 2



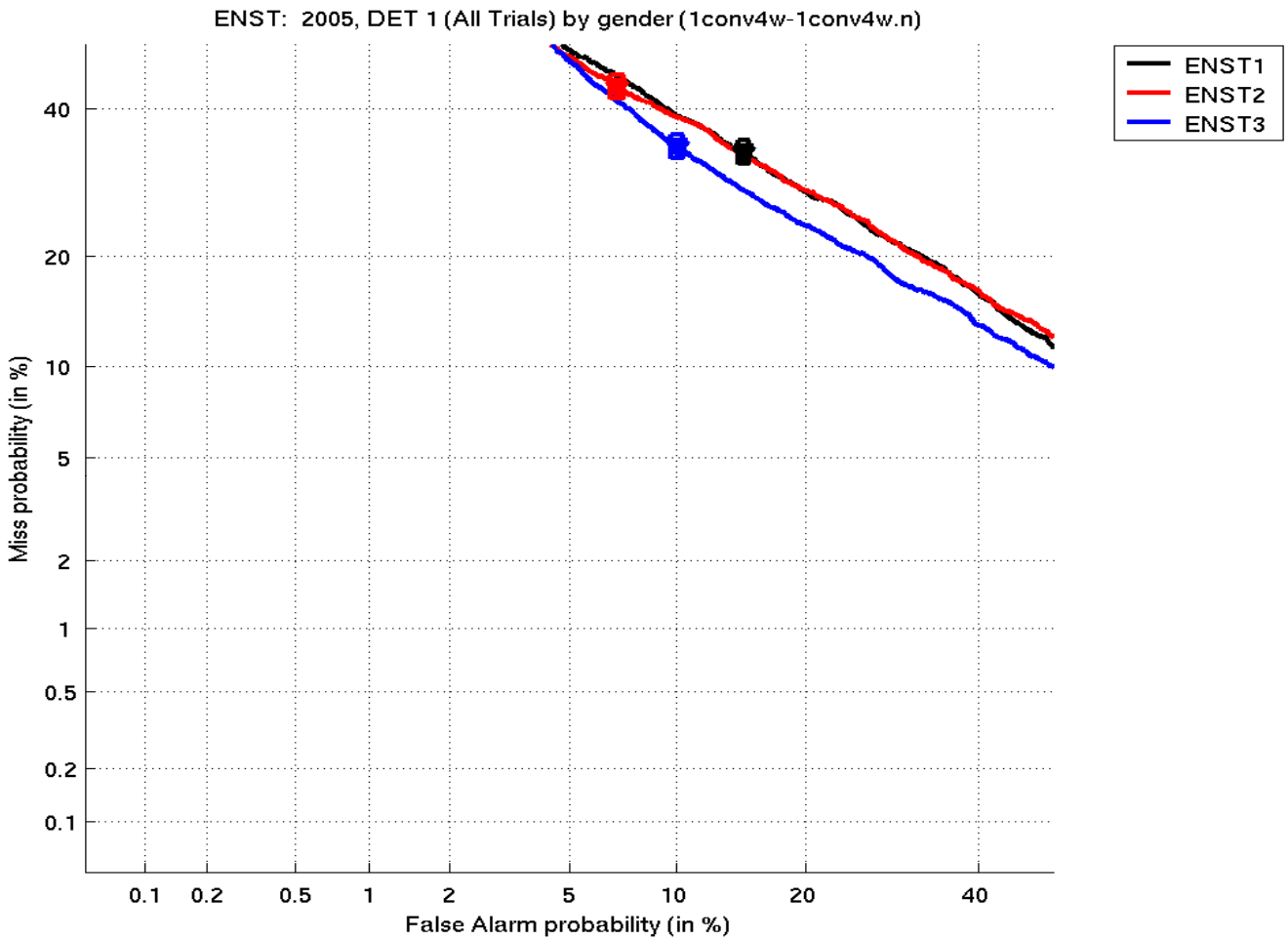
ENST 3

- Pre processing
  - Features extraction
  - Cms (cepstral mean substraction)
  - Silence removal
- *Acoustics parameters:* 39
  - 12 MFCC + energy+ Δ + Δ Δ.
  - CMS: the mean of mfcc vectors is calculated for the whole signal (silence and speech).
  - Features extraction by HTK.
- Normalisation
  - Decision score is based on un-normalised log-likelihood ratio

DET Curve : ENST 3



DET Curve : all systems



New Systems

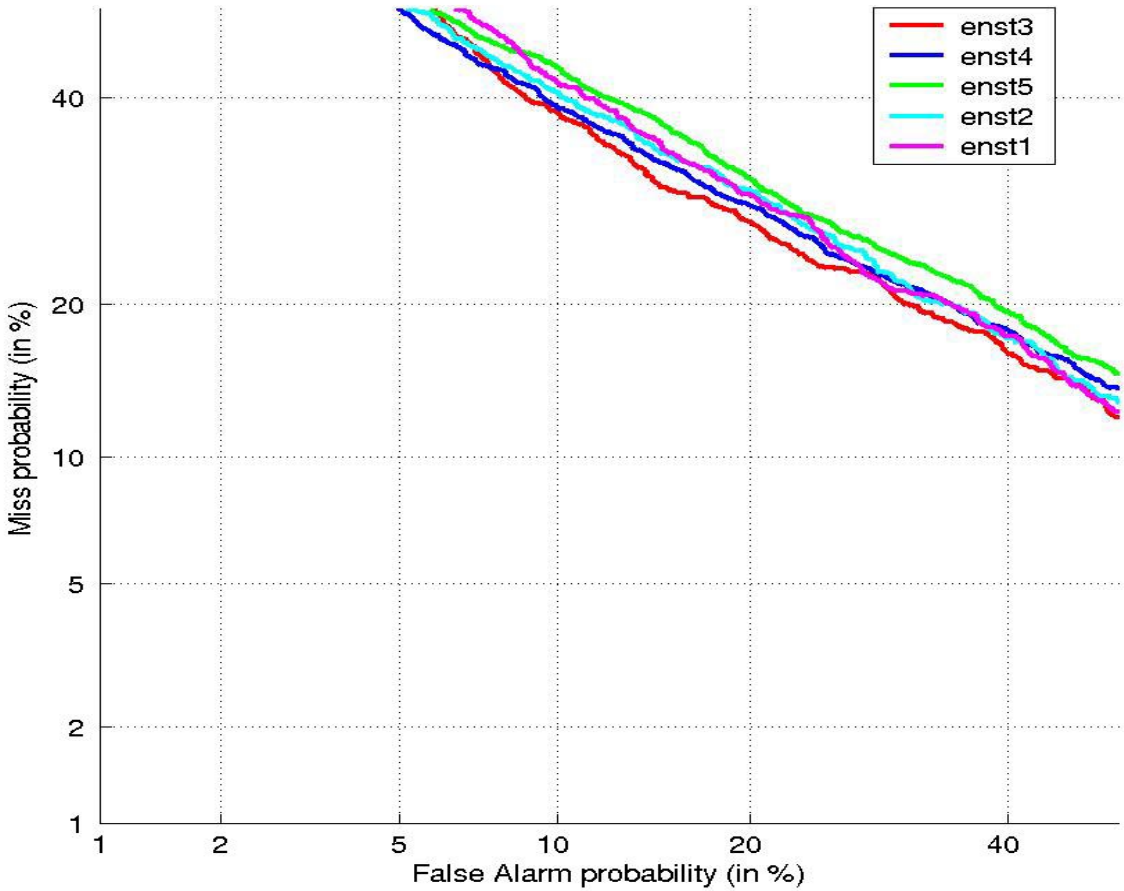
ENST 4

- Pre processing
  - Features extraction 16 LFCC + Δ
  - Cms (windows size 1500 vectors).
  - Silence removal
  - Features extraction by Spro.
- Modelling:
  - The same settings as the last system
  - Used 50% of data for initialisation
- Decision
  - Decision score is based on un-normalised log-likelihood ratio

ENST 5

- Pre processing
  - Features extraction 16 MFCC + Δ
  - Cms (windows size 1500 vectors).
  - Silence removal
  - Features extraction by Spro.
- Modelling:
  - The same settings as the last system
  - Used 50% of data for initialisation
- Decision
  - Decision score is based on un-normalised log-likelihood ratio

DET Curve : all systems



Conclusion

- Conclusion
  - Influence on results
    - Remove silence after acoustic parameters extraction
    - For CMS, mean of the whole signal should be taken
  - no influence on results
    - T-norm doesn't give better performance in our system
- Perspectives
  - More development database should be provided