NIST 2005
Speaker Recognition Evaluation
ETI Submission

Workshop participants: Ciano Frost and Thomas Ostergaard

This presentation and all information contained herein is classified:

"Confidential Commercial"

Excellence through specialization

Presentation Outline

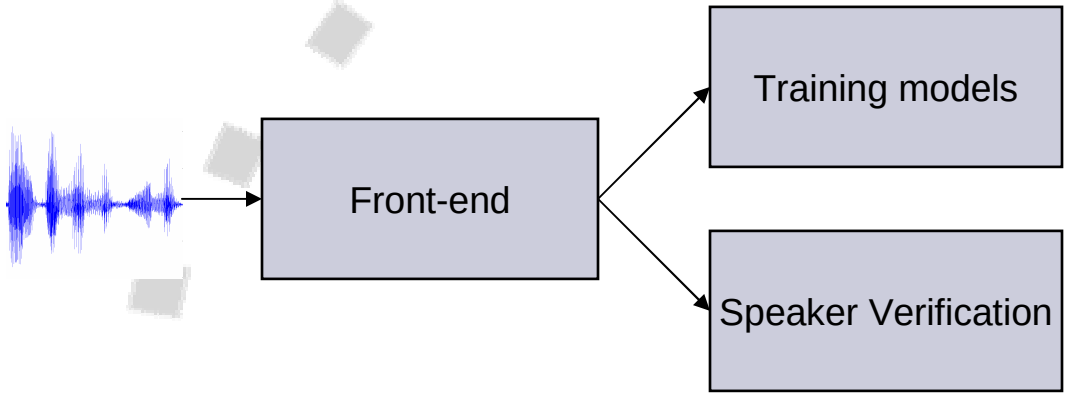
Presentation Outline:

- Introduction
- System description
  - Pre-processing and acoustic features
  - Training models
  - Speaker recognition (world models and score normalization)
- System performance
- Conclusion

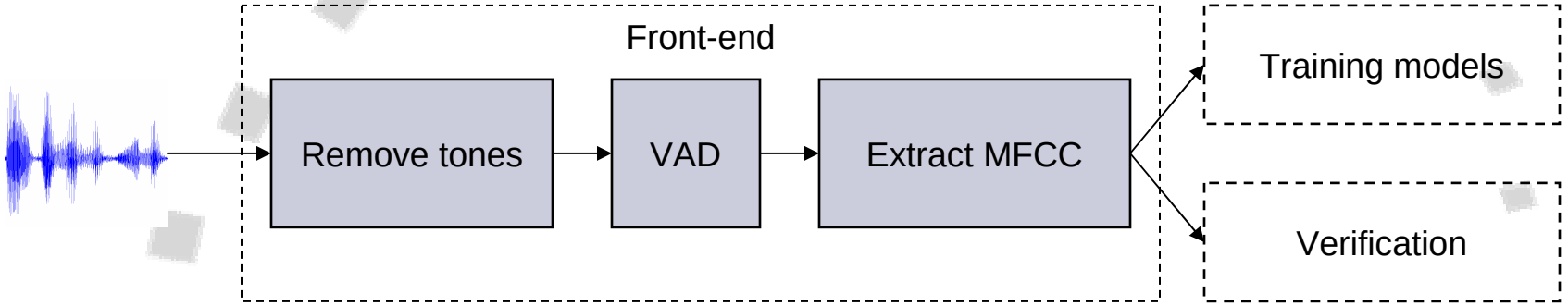
Introduction

- First time ETI participates in NIST Speaker Recognition Evaluation task.
- We have only submitted results for the mandatory 1conv4w-1conv4w.

System overview

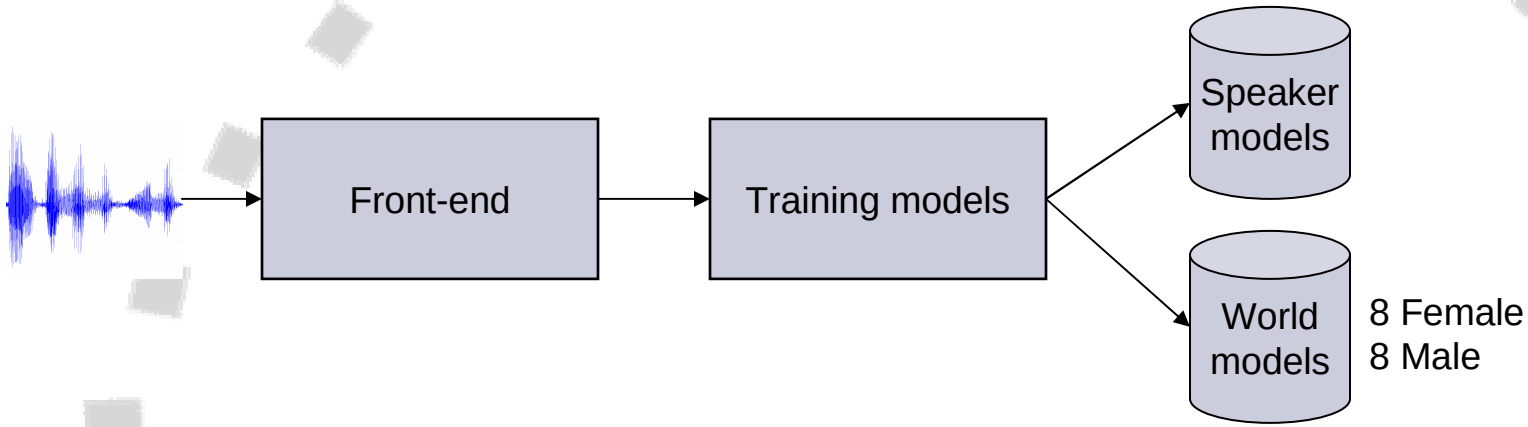


Front-end



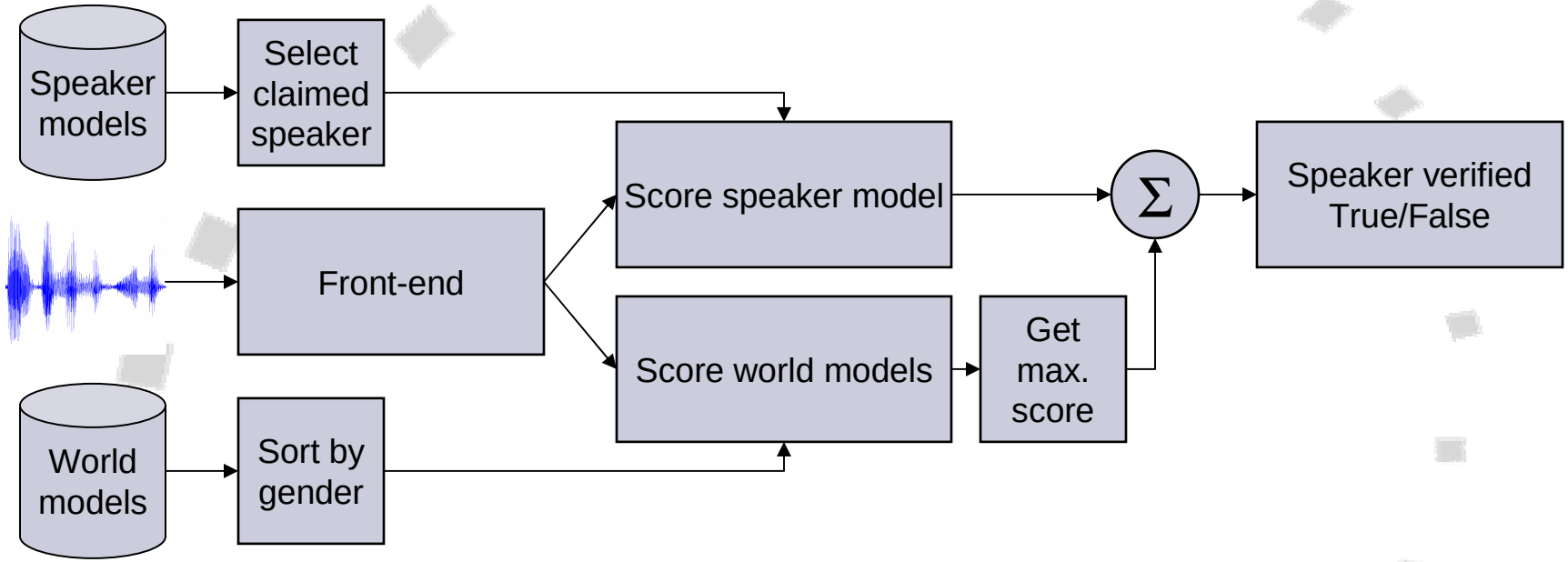
- Remove tones
  - Sinusoidal modelling
- VAD
  - Generalized log likelihood ratio
- Extract MFCC
  - 16 Coefficients, C1-C16
  - MLT Filter bank (Modulated Lap Transform)

Training



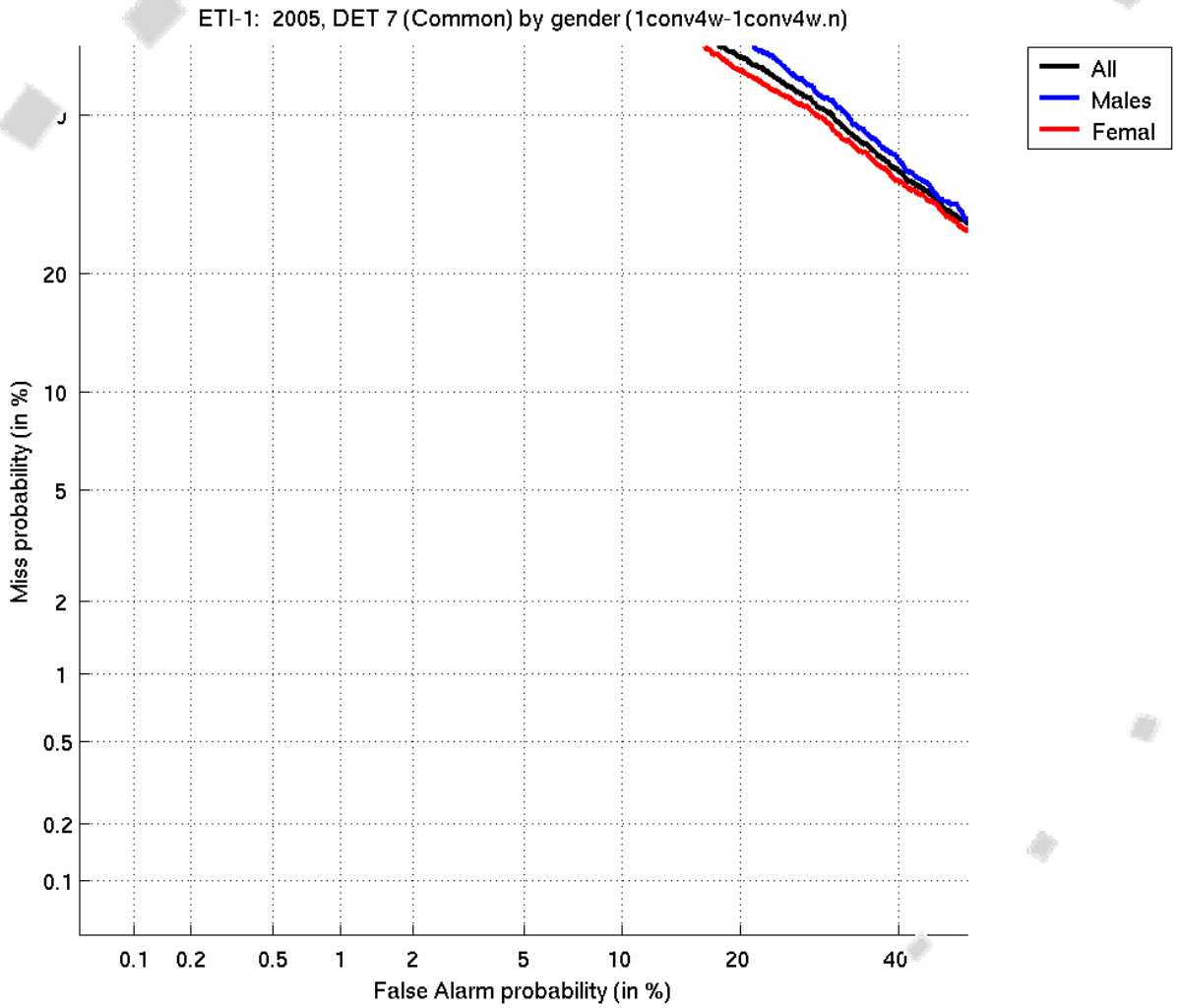
- Train GMM
  - 30 mixtures
  - EM algorithm
- World models
  - Trained on NIST SRE04 development data

Verification



- **LL score calculated for:**
  - **Speaker model**
  - **8 world models (only use same gender as claimed speaker)**
- **Normalize score**
- **Compare score against threshold**

System Performance



System Performance

- All trials were executed on a Pentium 4 @ 2.8 GHz with 512 MB RAM
- Execution time:
  - The execution time for file I/O, front-end processing and verification is 30 times faster than real time.
  - The execution time for verification alone is 59 times faster than real time.
- Memory usage:
  - 8 Mb when doing verification.

Conclusion

- We have used a ”baseline” system with no channel normalization. Which to some extend explains verification performance.
- We have tested CMS and RASTA earlier but with no improvement (probably too clean data).

- Future directions
  - Evaluate CMS and RASTA again on NIST SRE05 data.
  - Use UBM and MAP adaptation to generate better speaker models.
  - Use better score normalization. E.g. T-norm.