NIST SRE 2005

LIA System Description

J-F Bonastre, C. Fredouille, N. Scheffer

(jean-francois.bonastre,corinne.fredouille,nicolas.scheffer)@lia.univ-avignon.fr

http://www.lia.univ-avignon.fr/heberges/ALIZE/

THALES-LIA System Description

B. Ravera, F. Capman

(bertrand.ravera,francois.capman)@fr.thalesgroup.com

http://www.thalesgroup.com

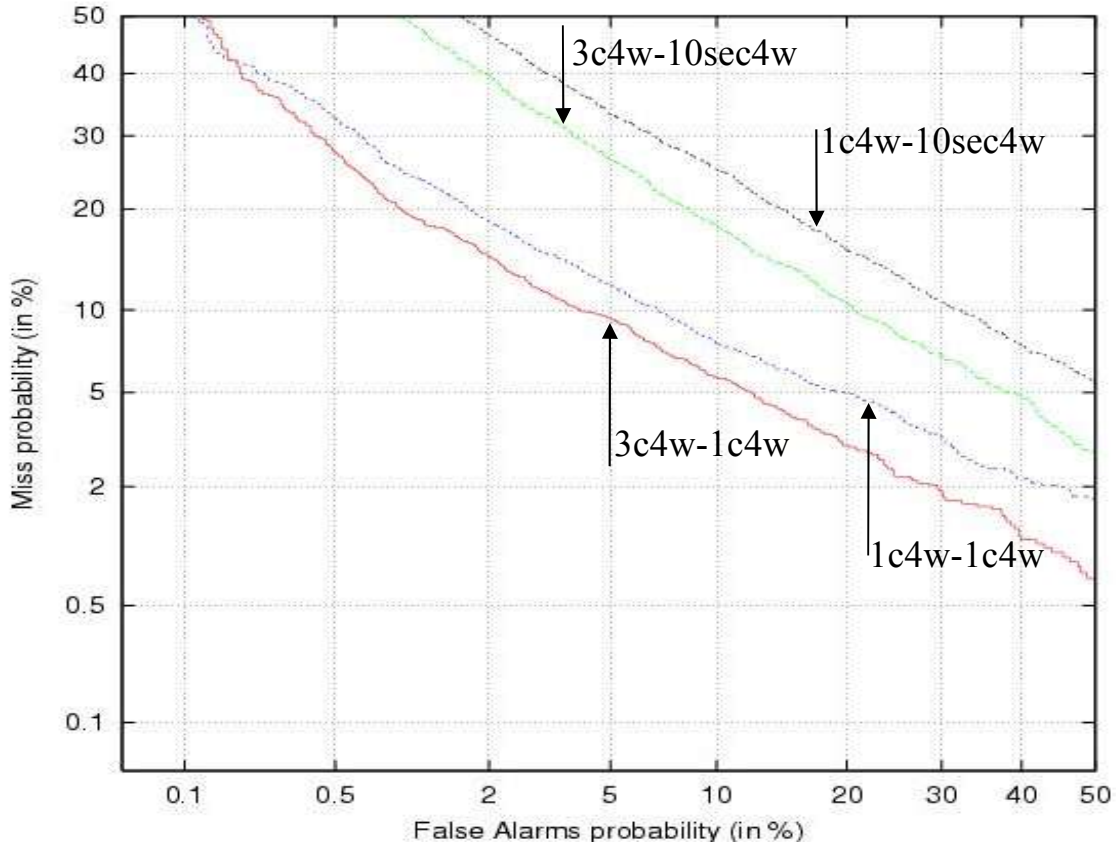
Outline

- Overview of LIA systems
- Commonalities
- GMM-based ASR system
- AES system
- Thales Communication System

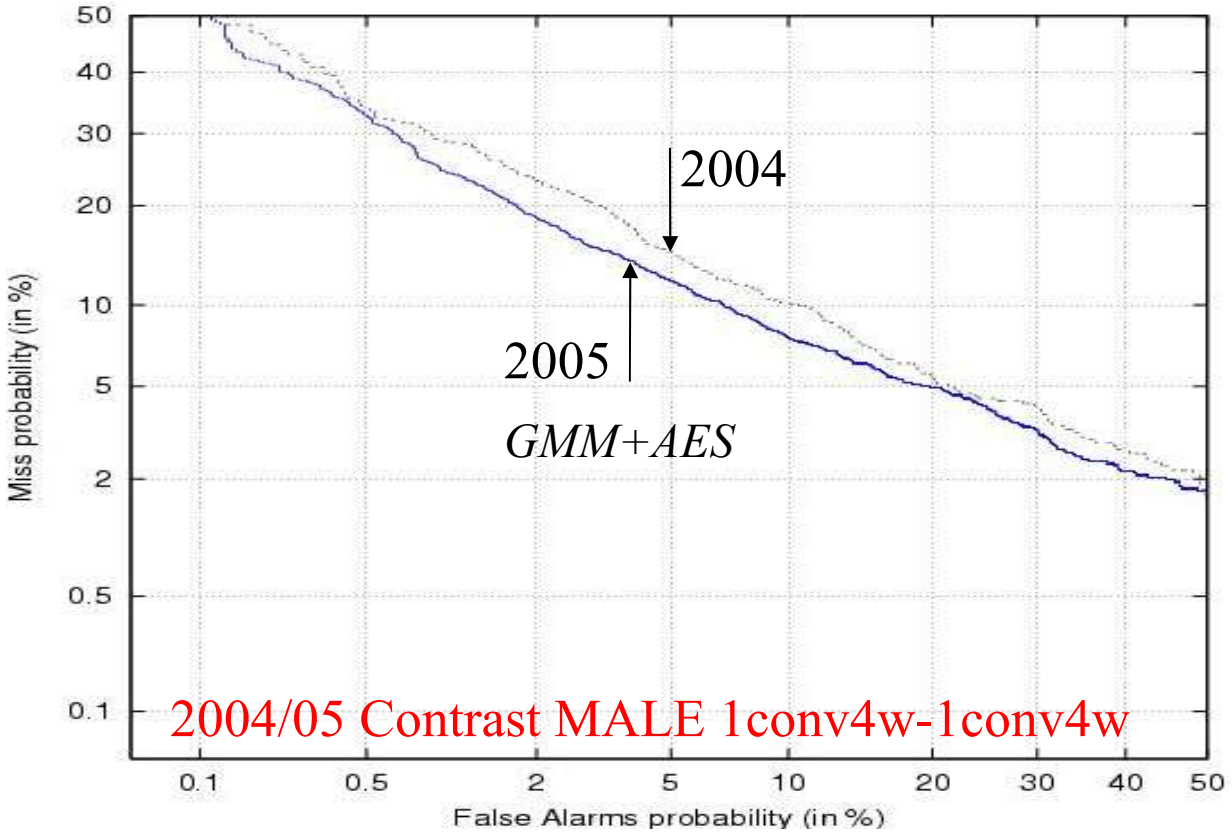
Overview of LIA Systems

- Based on ALIZE
- All LIA soft available with open software license
- 2 approaches
  - Standard UBM/GMM (GMM)
  - NGRAM Sequences (AES)
- Systems:
  - 1conv4w-1conv4w: GMM + AES, GMM, AES
  - 3conv4w-1conv4w: 3*GMM + AES, GMM-1, AES
  - 1conv4w-10sec4w: GMM
  - 3conv4w-10sec4w: 3*GMM + GMM-1, GMM-1

Performance vs systems/tasks



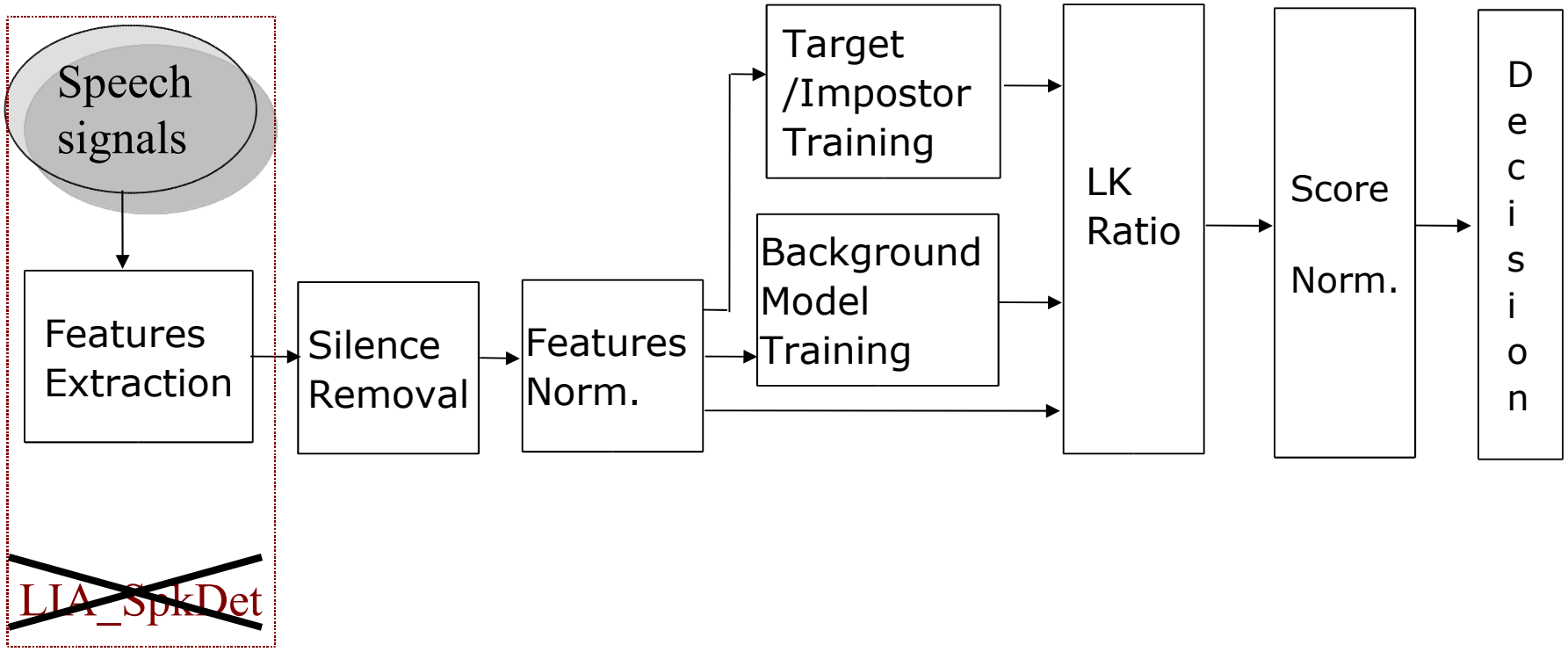
Comparison with LIA04 system



LIA-Commonalities (1)

- LIA_SpkDet Package
- Speech parameterization:
  - 3 energy component-based frame removal system
  - Morphological filter + channel overlapping pruning
  - 16 LFCC + telephone bandpass filter + derivatives
  - CMS + reduction to 1-variance (file by file)

LIA-Commonalities (2)
LIA_SpkDet overview



SPro (GPL) | LIA_SpkDet | LIA Util

LIA-Commonalities (3)
LIA_SpkDet05 main new features

- Frame pruning
  - Morphological filter
  - Channel overlapping prunning
- Segmental processing
  - Label-file based processing
  - Segmental LLR computation, normalization and decision
- Feature mapping
- Multiple models processing
  - By label GMM models
- Frame weighting for LLR computation
- Multiple weighting functions for fusion

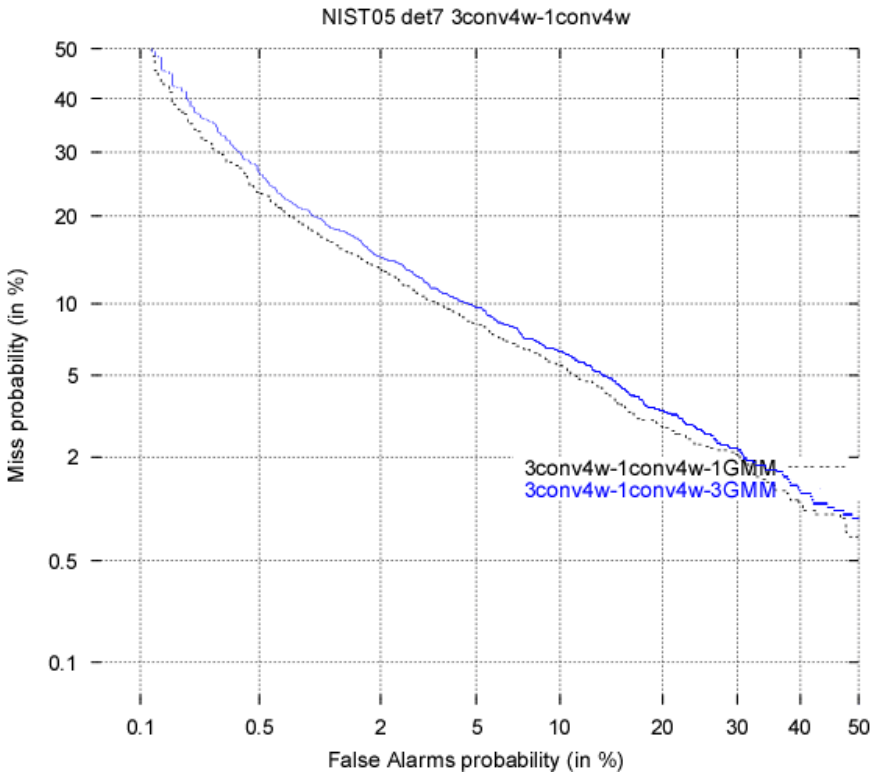
LIA-Commonalities (4)

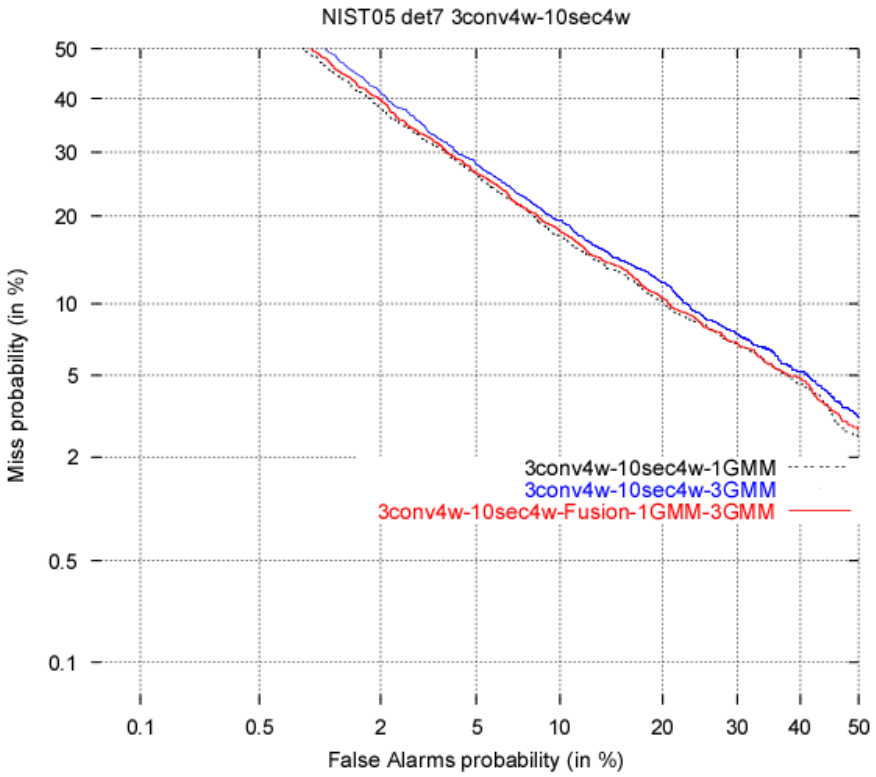
- Database:
  - DEV on male (+female?) NIST 2004 1side-1side
  - Gender dependent world models:
    - Data from 1999 to 2003 evaluation campaigns
    - 288 male (7h) and 439 female (10h) speakers
    - Channel dependent data: CDMA + GSM + Landline
  - Tnorm population:
    - World data + 2004 evaluation campaign
    - 322 speakers (equally gender balanced)

LIA-GMM-based ASR system (1)

- 2048 components, 0.5 variance flooring
- Tnorm score normalization
- Gender dependent thresholds tuned on DEV
- Training on 3conv, two strategies (MAP Factor 14):
  - GMM-1: Classical training on all the data
  - 3*GMM : Fusion of n *1conv systems (n=1 to 3)

LIA-GMM-based ASR system (2)
3Conv: GMM-1 vs 3*GMM

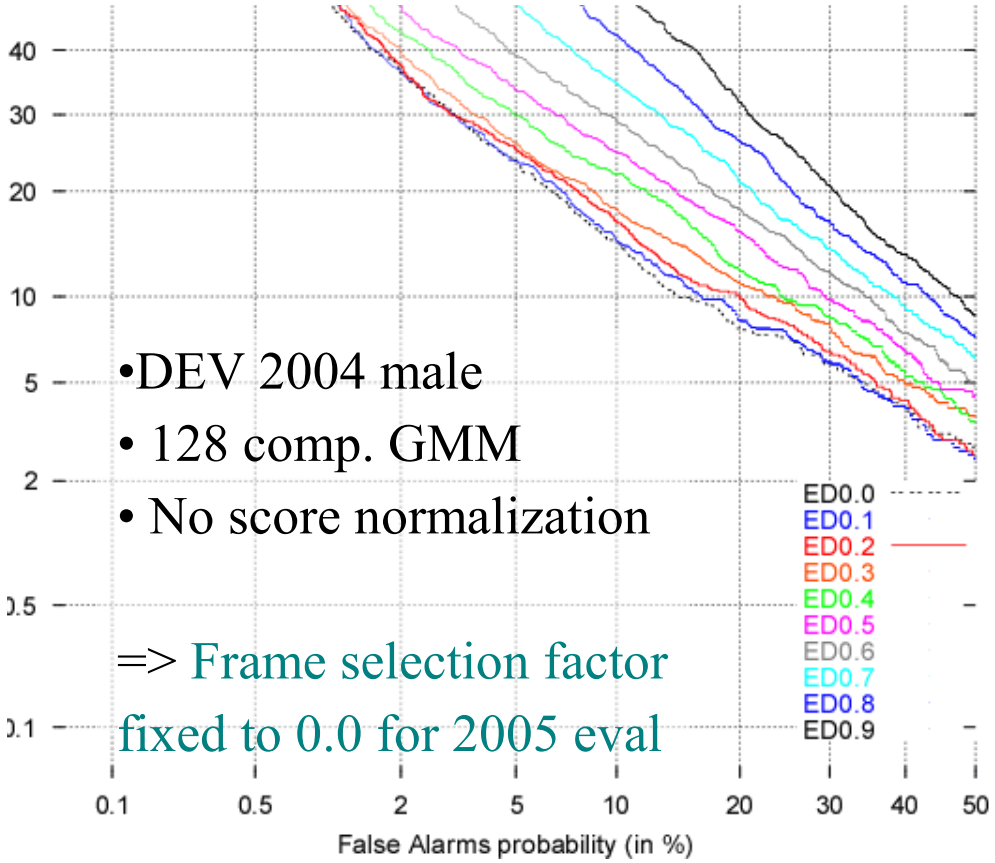




LIA-GMM-based ASR system (3)
2005 Frame removal process

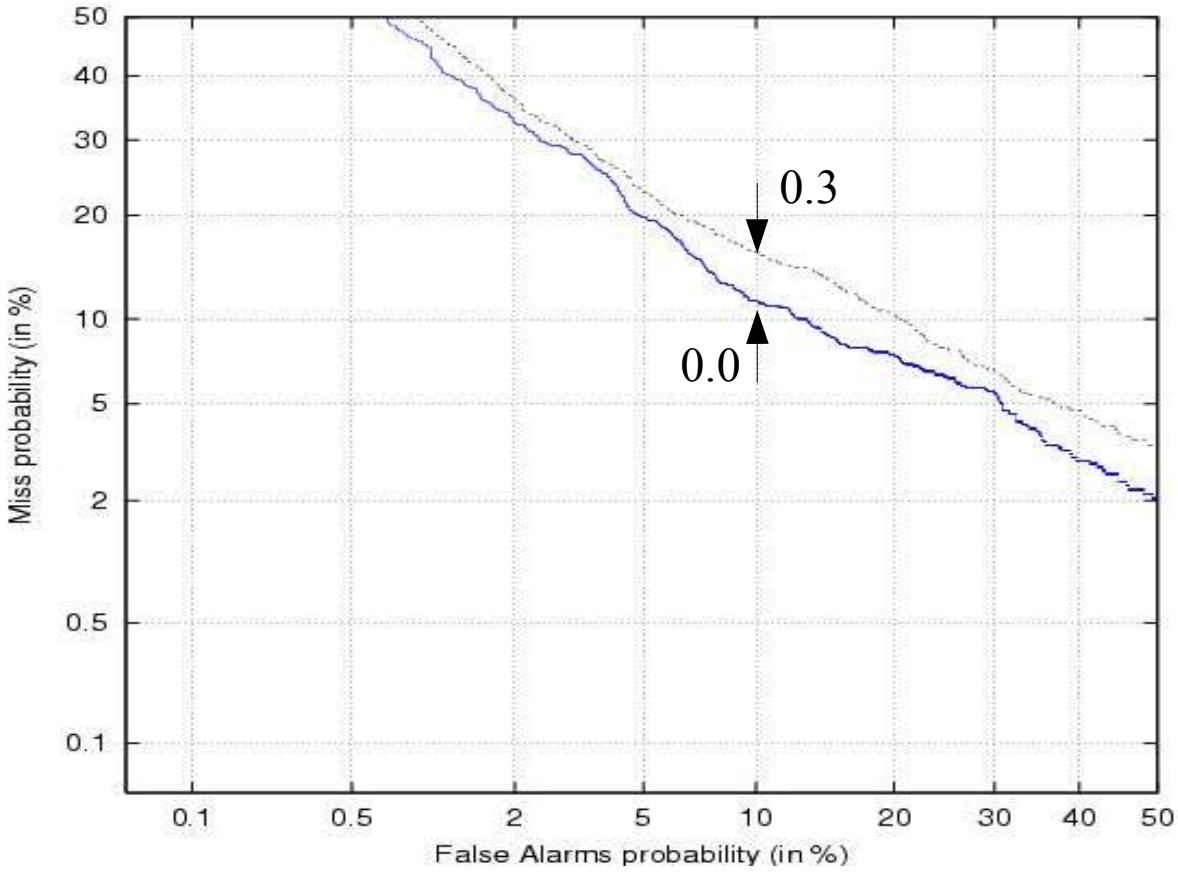
- 3 components
  GMM energy detector
- Mini loss-likelihood decision
  for the central component
  + selection factor

- Morphological filter
- Channel overlapping pruning
  (gain was not measured)



LIA-GMM-based ASR system (4)
2005 Frame removal process

- Effect on 2048 component GMMs

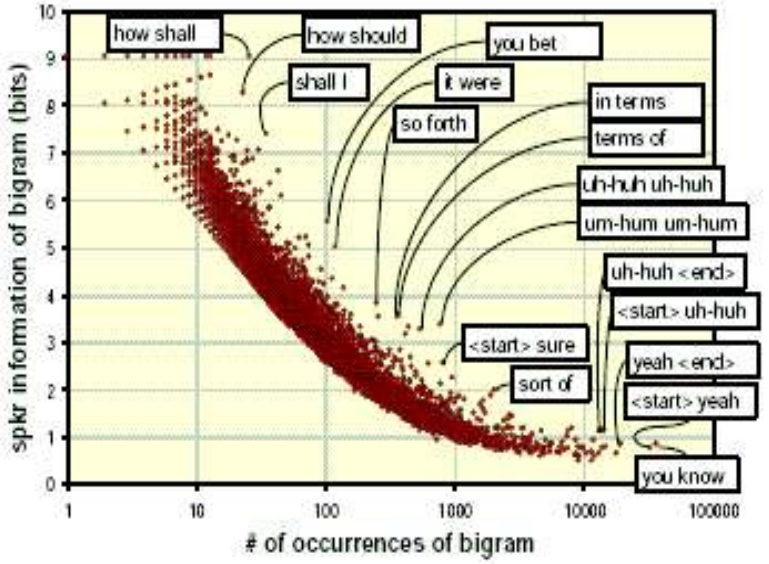


LIA
Acoustic Event Sequence System

- **High Level** based approach:
  - Good Performance on large database
  - Relies on an *a priori structure of speech*
  - *Independent* from acoustic modelling
- **Cepstral GMM** approach:
  - Does not model *signal temporal order*
  - *Efficiency* can not be argued

LIA-AES (2): Approach

- Example on **idiolectal** features:
  - Dictionary is composed of words (language dependent)
  - *Speaker independent*
  - Speaker differences appear in the analysis of *word sequences*



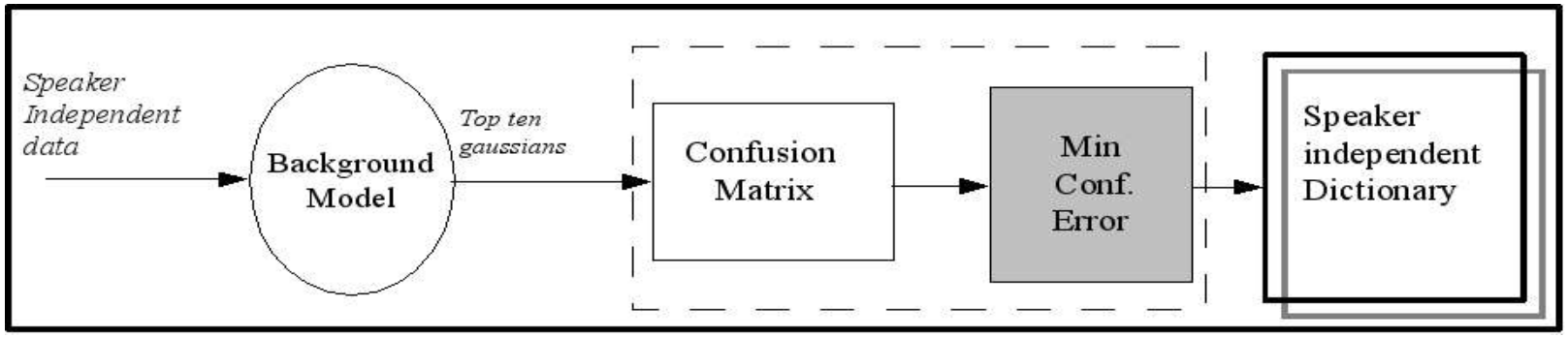
Doddington, EUROSPEECH 2001

LIA-AES System (3): Idea

- *Generalization* of this approach to non linguistics features
- Take benefit of the *acoustic modelling*
- **Methodology :**
  - Build a *speaker & language independent* dictionary based upon acoustic modelling
  - Model speakers by analysing *sequences* of the dictionary symbols (acoustic events)

LIA-AES System (3): Dictionary

- Generating *speaker independent* acoustic events
- Use a GMM with a maximum of information (*world model*) to produce a proper structure



LIA-AES System (4)

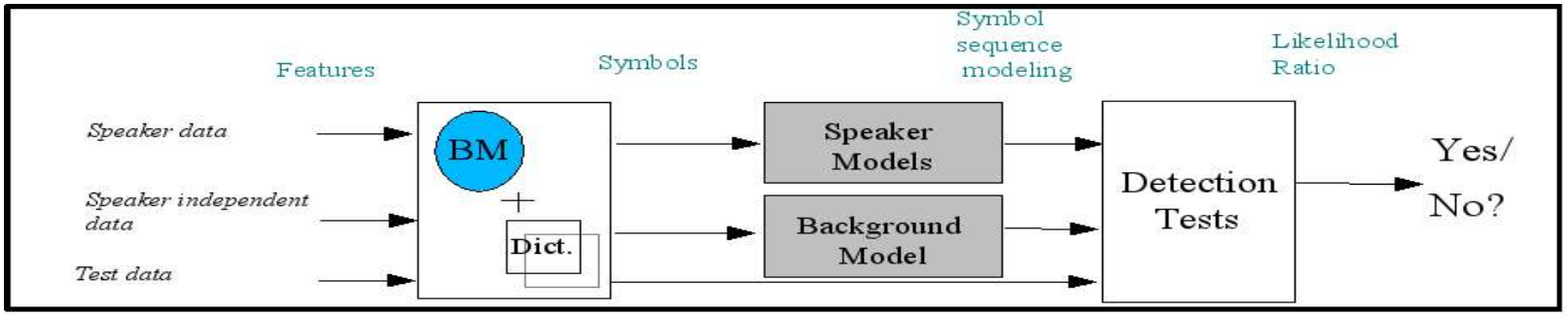
- **Speaker modeling: Bag of Ngram approach**
  - *Speaker-specific symbol sequences* (Token Ngrams)
  - A codebook is generated based upon the background model
  - A "language model" is computed for each speaker
  - Impostors are built from the background model data

LIA-AES System (5)

- **Detection tests:** SVM-based approach
  - *TFLLR* weighting method (Campbell 04)
  - LLR is expressed as a Kernel function
  - Target and Impostor examples are passed through a linear kernel
  - Distance to the classifier is used for verification

LIA-AES System (6)

- System overview:

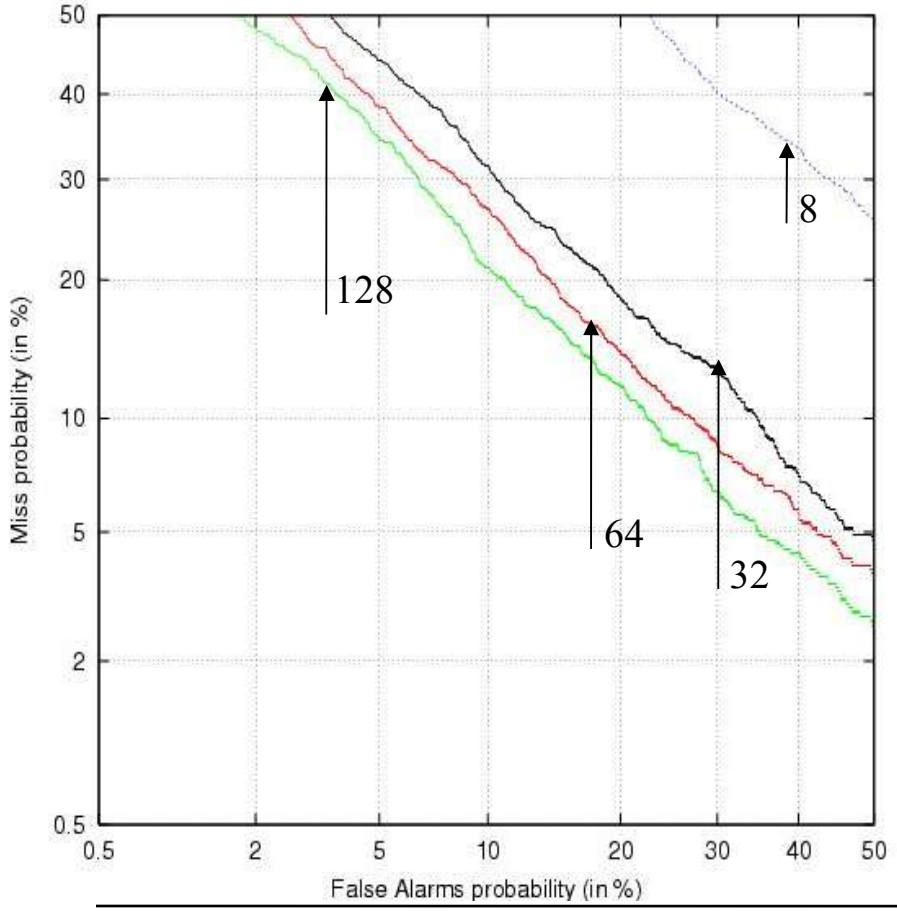


- Parameters of NIST SRE 2005 Evaluation:
  - Dictionary size 128
  - 3gram sequence analysis

LIA-AES System (7): Submission

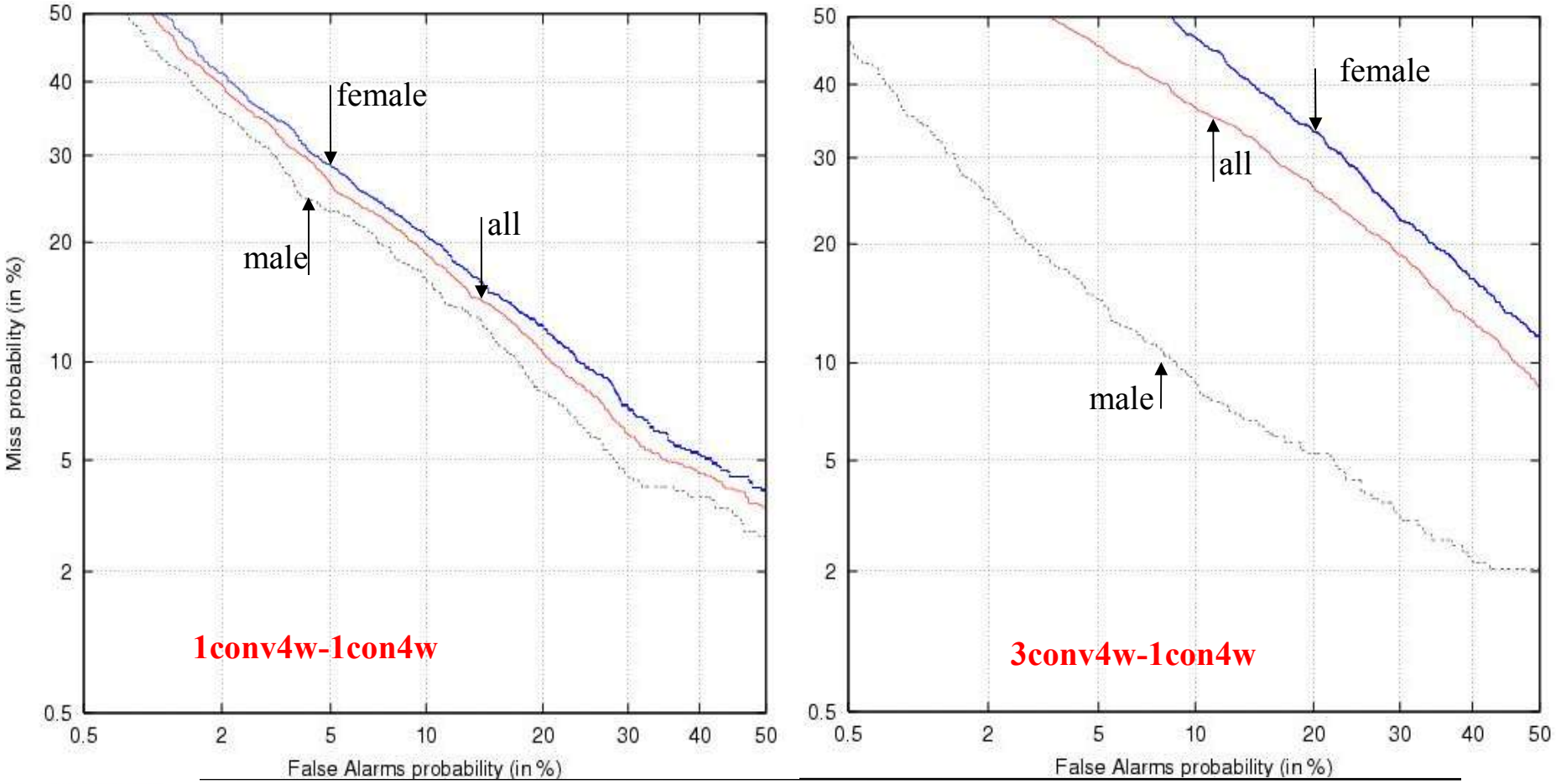
- Completed tasks:
  - Train: 1conv4w & 3conv4w
  - Test: 1conv4w
- Standalone system: LIA 3
- Fusion with the GMM system:
  - LIA Primary system: LIA 2 & 4
- Fusion is an unweighted mean of both scores

LIA-AES System: DET 1 DEV

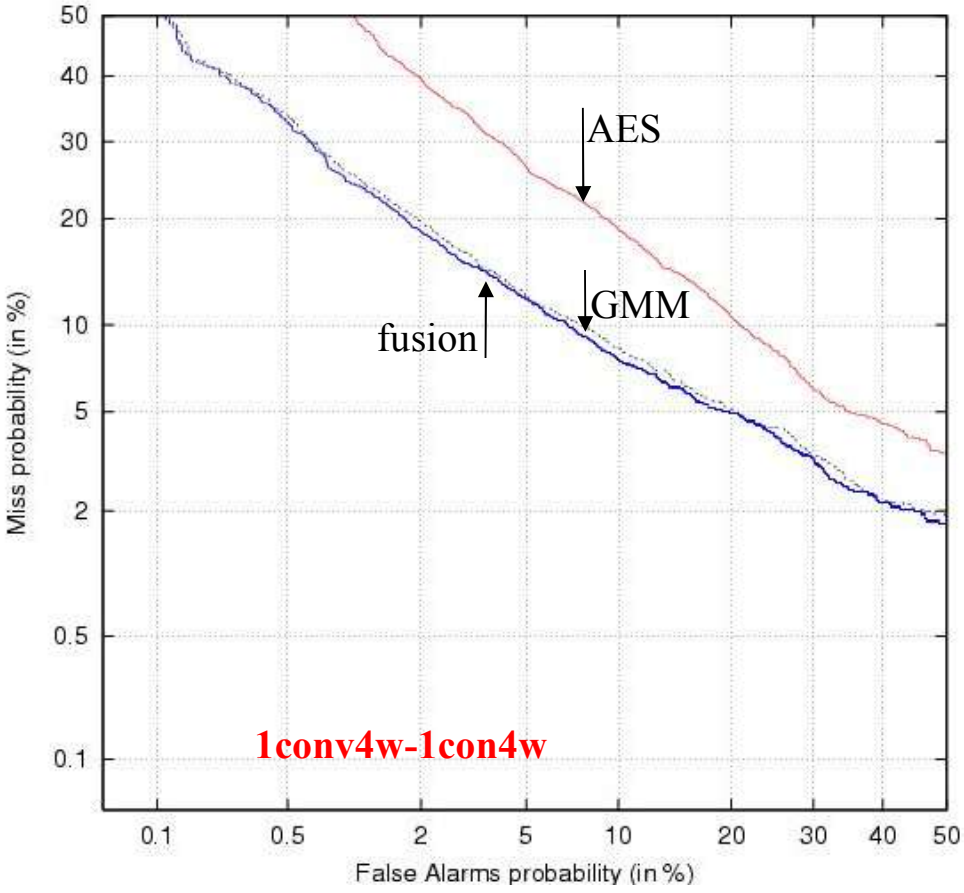


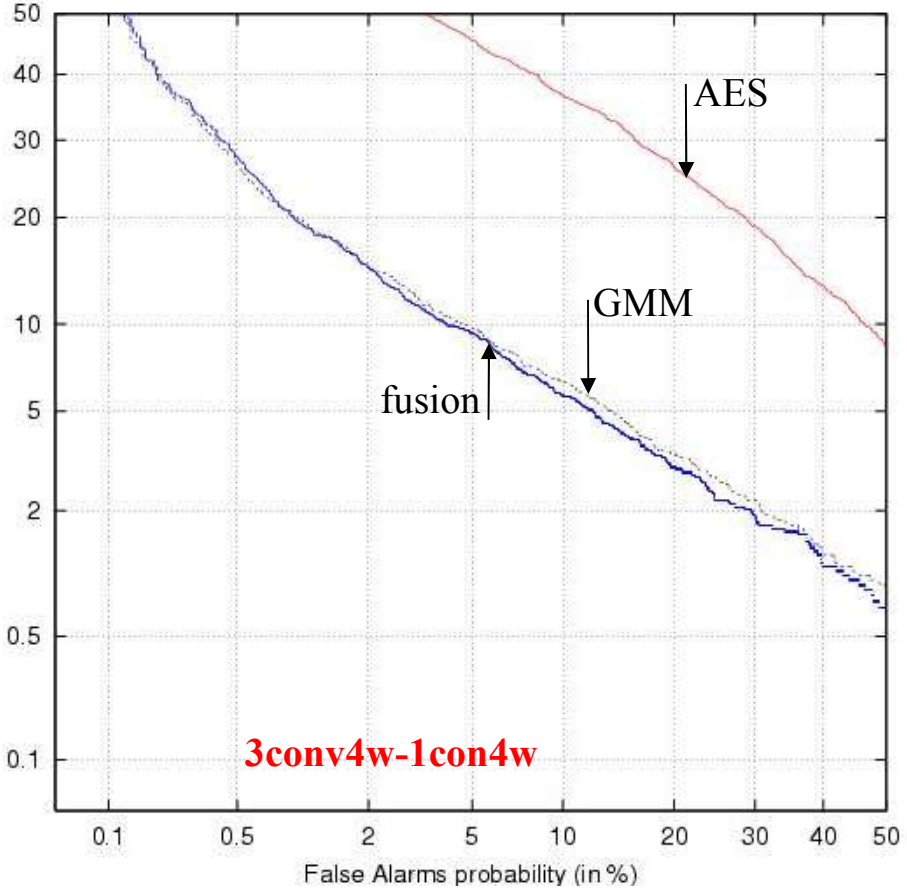
- Modifying size of the acoustic event dictionary
  - 3gram sequence analysis
  - TFLLR weighting
  - 2004 Male set

LIA-AES System : DET 7 EVA

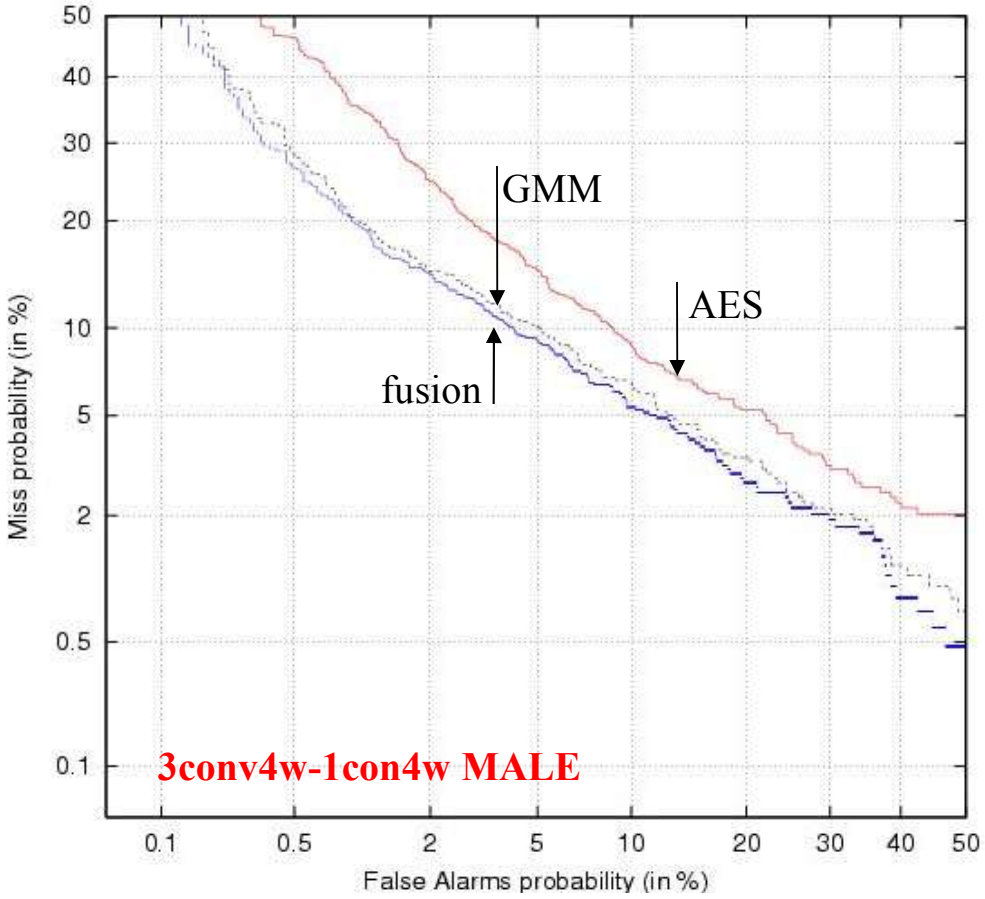


LIA Submission: Fusion DET 7





LIA Submission: Fusion DET 7



LIA-AES System: Conclusion

- Novel approach using *benefits of "low" and "high" level-based* techniques
- Modelling *other information* than the GMM
- Naive fusion seems to bring performance
  but small gains -> a better fusion is needed
- **Future works:**
  - Work will focus on different sequence lengths and new techniques to generate acoustic events
  - Application to *other tasks* (speech recognition?)

Thales Communications Submission

- Signal Processing and Multimedia Department
  - main background in speech and image compression
- First participation to NIST evaluation campaign
- Co-operation with L.I.A. (PhD, projects, …)
- Development based upon L.I.A. systems (ALIZE and Speak Det 05)
- Objective : improve our “know-how” in Speaker Verification and related technologies

Thales/LIA 1: Overview of the 1st system

- Speech frame selection
  - active frames are simply detected using an estimated histogram of frame energies
  - the applied threshold is calculated from the most energetic peak of the histogram, using a constant speech dynamic hypothesis
- Cross-channel spectral cross-correlation
  - spectral vectors are estimated using a 32 Mel-scaled filter-bank
  - normalised spectral inter-channel cross-correlation is used to discard potential double-talking frames

Thales/LIA 2: Overview of the 2nd system

- Based upon Thales/LIA 1 system
- Additional weighting module
  - Estimation of a 32-classes speaker (file) dependent codebook using Mel-scaled filter-bank energies
  - Each frame is then weighted using a Fuzzy-vector quantization criteria applied to the previous codebook (weighted spectral distance to the 3 closest codebook vectors) in order to reinforce spectrally stable frames
  - Likelihood ratio is then weighted using the highest fuzzy weight during the test process

Conclusion and Future Work

- Be in first position next year? Sure!
  - Thales ?
  - LIA ?
  - LIA+Thales ?