LRDE-EPITA Systems description

Réda DEHAK [1], Charles DELEDALLE [1] and Najim DEHAK [2]
[1] Laboratoire de Recherche et développement de l'EPITA, Le Kremlin-Bicêtre, France
[2] Centre de Recherche en Informatique de Montréal, Canada
NIST Speaker Recognition Workshop, San Juan, Puerto Rico, June 26-27, 2006

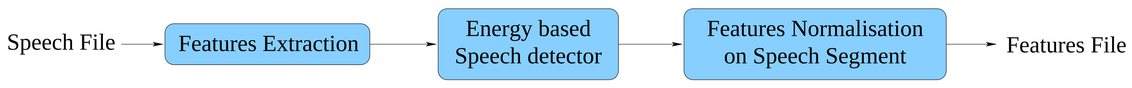
OVERVIEW

LRDE has presented **three systems** in 1 train/test conditions (1conv/1conv)

- **LRDE1** System : GMMs system (stat of the art)
- **LRDE2** and **LRDE3** Systems : Support Vector GMMs System

- GMM toolkit: **BECARS** available at http://www.tsi.enst.fr/becars
- SVM toolkit: **LIBSVM** available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

Features Extraction



- Feature Extraction : Two different sets of parameters (MFCC and LFCC)
- Speech Detector : Based on 3 gaussian
- Feature Normalization : Two different algorithms (CMS and feature warping)

Data description

World database : Gender dependent GMM with 512 or 2048 Components

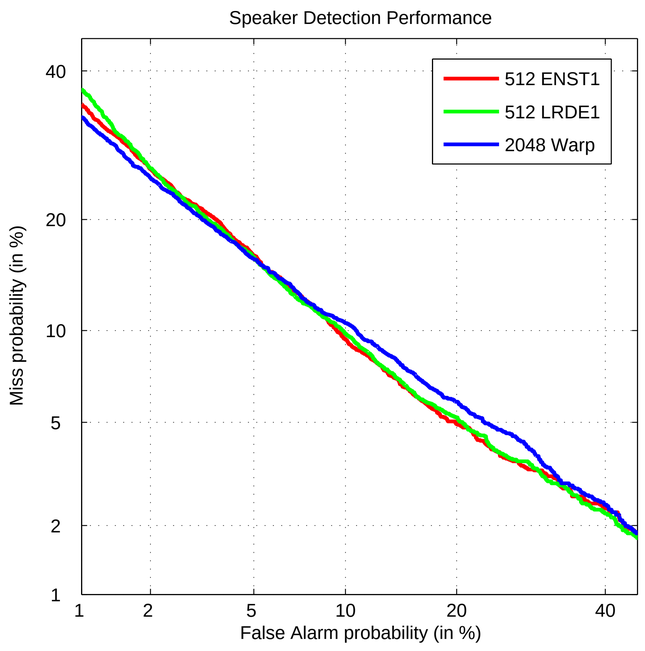
- LRDE1 and LRDE2 systems:
  - Male : 749 segments-1side extracted from Nist 2003, Fisher Part 1 and Part 2 databases
  - Female : 800 segments-1side extracted from Nist 2003, Fisher Part 1 and Part 2 databases
- LRDE3 system:
  - Male : 200 segments-1side extracted from Nist 2003 and Nist 2004 databases
  - Female : 283 segments-1side extracted from Nist 2003 and Nist 2004 databases

Impostors dataset

- LRDE1 and LRDE2 systems:
  - 373 impostors for each gender extracted from Fisher database
- LRDE3 system:
  - 195 males impostors extracted from Nist 2003 and Nist 2004 databases
  - 294 females impostors extracted from Nist 2003 and Nist 2004 databases

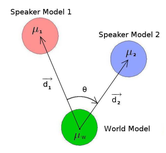
LRDE1 system

- Acoustic features : 33 coefficients
  - 16 LFCC + δ + δenergy
  - mean and standard deviation normalization or gaussian feature warping
- GMM with 512 Components
- Use a MAP adaptation to estimate the mean of client and impostor models
- Decision score is based on a normalized log likelihood ratio of the 20 best gaussian component
- use of ZT-Norm with 373 impostors models for each gender



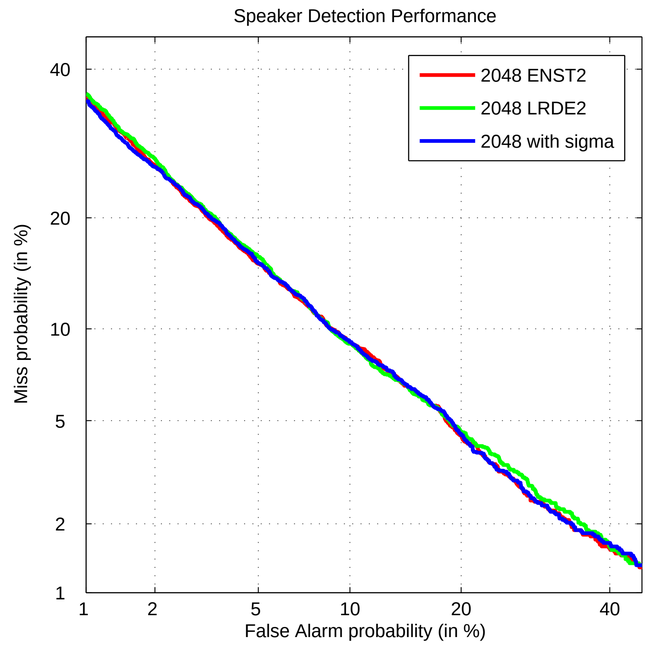
LRDE2 system

- Acoustic features : 33 coefficients
  - 16 LFCC + δ + δenergy
  - mean and standard deviation normalization
- GMM with 2048 Components
- Use a MAP adaptation to estimate the mean of client, test, and impostor models
- Optimal hyperplan decision boundary learning
- Decision score is based on distance between test models and Decision boundary



$$D(\lambda_1, \lambda_2) = \sum_{i=1}^{2048} w_i \theta^2_{\overrightarrow{d_1^i}, \overrightarrow{d_2^i}} \left(\|\overrightarrow{d_1^i}\| - \|\overrightarrow{d_2^i}\| \right)^2$$

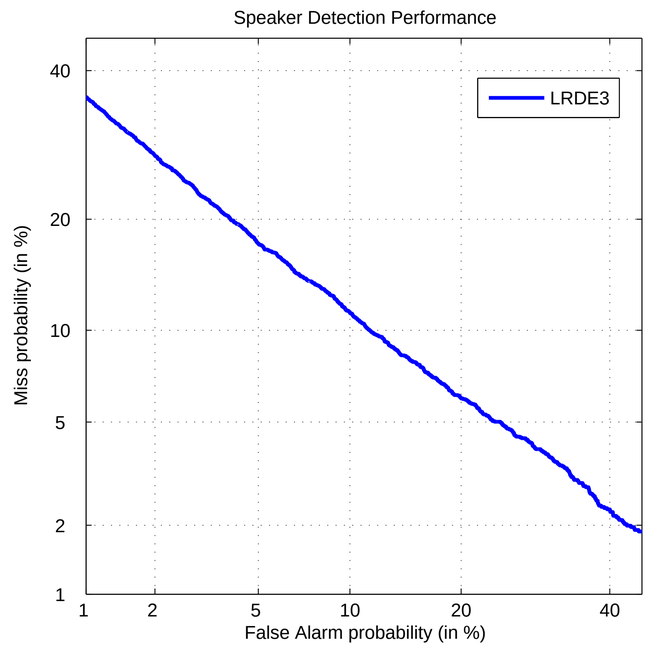
$$K(\lambda_1, \lambda 2) = e^{-D(\lambda_1, \lambda_2)}$$



LRDE3

- Acoustic features : 31 coefficients
  - 15 MFCC + δ + δenergy
  - CMS on speech data
- GMM with 512 Components
- Use a MAP adaptation to estimate the mean of client, test, and impostor models
- Optimal hyperplan decision boundary learning
- Decision score is based on distance between test models and Decision boundary

$$D(\lambda_1, \lambda_2) = \sum_{i=1}^{512} w_i (\mu_1^i - \mu_2^i)^t \Sigma_i^{-1} (\mu_1^i - \mu_2^i)$$



PERSPECTIVES

- SVM score normalization
- Feature Mapping and Canal Compensation