LIA: NIST SRE'06 Evaluation Campaign

Nicolas SCHEFFER, Alexandre PRETI, Corinne FREDOUILLE, Driss MATROUF, Jean-François BONASTRE
first.last@univ-avignon.fr, Laboratoire Informatique Avignon, France

Overview

	1c-1c	3c-1c	1c-10s
LIA_1	**3.29**/7.02	X	X
LIA_2	**3.16**/6.65	2.53/5.60	6.06/15.24
LIA_3	**3.30**/7.12	2.78/6.62	6.03/14.89

Datasets

- Development Set:
NIST SRE 2005
- Background Set (UBM & SVM):
Fisher corpus Part I (single conv speakers, 1464 speakers)
- TNorm speakers:
Gender dependant, 180/Gender, (60 cellular, 60 landline, 60 cordless)

Front end

- SPRo sofware: LFCC analysis
- 50 coeffficients: 19 static + ∂ + 11 first ∂∂ + ∂e
- Bandwidth: 300 - 3400 Hz
- Frame removal 3 component GMM (30% frames selected)
- Overlapped speech segment removed
- Avoiding short speech segment with a morphological filter

- Vectors N(0,1),
- Variance estimated File-by-File

- Gender dependant Feature Mapping
  3 conditions (landline, cellular, cordless)
  Submodels derived from root UBM by MAP (mean and variance r=14)

ALIZE/LIA_SpkDet

- LIA systems are powered by the ALIZE toolkit (LGPL licence) and LIA_SpkDet (GPL)

- ALIZE is a high level API making Speaker Recognition application development easier for researchers and students

- v2 is out with lots of improvements
http://lia.univ-avignon.fr/heberges/ALIZE/

GMM/UBM Cepstral

- 2048 Component gender dependent UBMs

- Initialization with random selection of frames

- Channel and Native speaker weighting stream learning

- 0.5 variance flooring

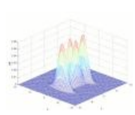
- MAP mean only adaptation of client models

- Top-ten component likelihood computation

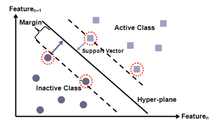
SVM/UBM Cepstral

- Hybrid GMM/SVM system using UBM-based statistics.
- Derivation of the TFLLR kernel by using Gaussian indexes as tokens

Principle: For each UBM Gaussian, its *a posteriori* probabilty on speaker data is computed. Input of the SVM is the ratio between this quantity and the Gaussian weight (*a priori*).



$$\frac{p(W_k|X)}{p(W_k)}$$



- Computationally efficient, suitable for large scale evaluation
- Fusion of two systems: "raw" features and rank normalized

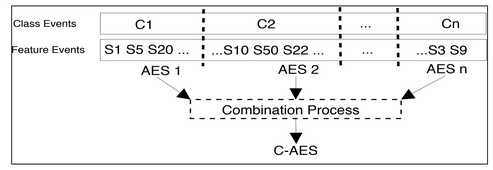
See Odyssey presentation or "UBM-driven discriminative approach for Speaker Verification, Odyssey'06, Porto Rico."

AES/C-AES High Level

- Acoustic Event Sequence system (AES)
  See LIA NIST SRE 2005 or "Speaker Verification using Acoustic Event Sequences, INTERSPEECH 2005, Lisboa, Portugal."
  Principle: Consider clusters of UBM Gaussian Indexes as Acoustic Events (speaker independent). Perform a sequence analysis (Ngram type) to model speakers. Use a SVM framework to score. (Dictionary size of 128, Variable Length Ngram analysis 2gram to 4gram)

- A Multi-class extension for AES: C-AES
  To appear in INTERSPEECH 2006 in "A multi-class framework for Speaker Verification within an Acoustic Event Sequence System."
  - Generation of an additional set of acoustic events (Class) from the original events (Feature)
  - Perform multiple AES on each Class Event
  - Integration of apriori information within the TFLLR framework

$$\phi(k,X) = p(C_k|X)\frac{p(k|X,C_k)}{\sqrt{p(k|X_W)}}, \forall k \in B$$



Results and Analysis

	male	female
LIA_1	2.80/5.09	3.69/8.49
LIA_2	2.61/4.88	3.54/7.78
LIA_3	2.95/5.62	3.53/8.10

	2005 (dev)	2006
GMM/UBM	3.05/8.05	3.33/7.12
SVM/UBM raw	3.53/10.38	4.11/9.64
SVM/UBM rank	3.53/9.78	4.16/9.50
AES	5.87/15.50	6.96/16.1
C-AES	6.52/16.53	7.25/16.9

SVM/UBM	3.35/9.73	3.90/9.13
LIA_2	2.73/7.68	3.16/6.65
GMM AES	2.94/7.77	3.84/8.27
GMM CAES	2.88/7.54	3.44/6.83
LIA_1	2.69/7.68	3.29/7.02

- Fusion is an arithmetic mean of scores after Tnorm.
- LIA_1 used to be better on the DEV set than LIA_2
- The female set degrades our system a lot, even fusion does not work
- It seems that Fisher only Tnorm is the cause of a bad decision threshold

Summary

- What worked:
  - 50 coefficients feature vectors
  - 3 channel feature mapping
  - Male set fusion with SVM/UBM
  - 5 systems combination is the best on DEV set.
- What did not:
  - Fusion on EVA set
  - Losing 1 point DCF with a bad decision
  - Not a (0,1) distribution with TNorm
  - Sequence systems degrades performance
- What is needed:
  - Rock-solid DEV/EVA set
  - Gender independent system