ENST-IRCGN
System Description
SRE2006

Patrick Perrot(1)(2), Réda Dehak(3), Gérard Chollet(1)
(1) CNRS - LTCI, GET-ENST-TSI - Departement France
(2)Institut de Recherche Criminelle de la Gendarmerie Nationale (IRCGN), Rosny sous bois, France
(3)Ecole pour l'Informatique et les Techniques Avancées, Paris, France
perrot@tsi.enst.fr, reda@lrde.epita.fr, chollet@tsi.enst.fr

I. Overview

The Ecole Nationale Supérieure des Télécommunications (ENST) and the Forensic Research Institute of French Gendarmerie (IRCGN) have been proposed three different systems to the NIST-SRE2006. The first and the second systems are based on Becars and the third results on a score fusion of three different cores. The submissions proposed are all on the 1conv4w1conv4w task. The different works have been realized with the collaboration of LRDE-EPITA, UNIFR-INT of Fribourg and UPMC-PRC of Paris.

II. System Description

System n° 1: ENST_IRCGN_01

The primary system is a basic application of Becars. This one results from a collaboration between ENST and the university of Balamand. It uses an UBM (Universal Background Model) composed by a subset from the Nist and Fisher databases by using a Gaussian Mixture Model learning via the EM (Expectation-Maximization) algorithm. This UBM is adapted from a MAP adaptation to the target speaker. This system is similar to the lrde-epita primary system, based on the same core.
As pre-processing, the first task consists in removing the silence frame, then a global normalization of the mean and standard deviation is done on speech segments for all features. The SPRO software described and proposed in [5] is used in this phase.

A scoring phase is applied for each speech segment by using the 20 th top Gaussian of the UBM's and the target model log likelihood ratio. Finally, a score normalization (ZTnorm) is applied in order to compensate the variability caused by different speakers and various test segments.

Feature extraction: LFCC based acoustic representation
The signal is described by the 33 following features:

- 16 LFCC
- 16 delta LFCC
- 1 delta energy

a mean and standard deviation normalization is done on the features.

UBM: Universal Background model

The database used to build the universal background model is composed by a subset from NIST-03 and Fisher database.

GMM dimension: 512
Male : Nist-03, Nist-04 and Fisher
800 segments (one side)
Computation 155h
Female : Nist-03, Nist-04 and Fisher
883 segments (one side)
Computation 192h

Adaptation:

The UBM adaptation is realized by the application of the MAP algorithm. The aim of this adaptation in this system is to evaluate target and impostor models.
Computation: 12 h

Impostors: male: 449 (from NIST and Fisher database)
Female: 486 (from NIST and Fisher database)

This set is used for score normalization

Scoring: As previously said, a log likelihood of the 20^{th} top gaussian's of the UBM and target models are calculated to determine the score.
Compute time: 8h for scoring – 122h for normalization

System n° 2: ENST-IRCGN_02 :

This system is similar to the system presented in [2], excepted that we used a feature normalization instead of a feature warping. The scoring phase of this system is based on a combination of support vector machine (SVM) with Gaussian mixture models. A kernel is used to compute a similarity between GMM target models and GMM impostors models.

Feature extraction: LFCC based acoustic representation
The signal is described by the 33 following features:

- 16 LFCC
- 16 delta LFCC
- 1 delta energy

Universal Background model

GMM dimension : 2048
Male: Nist-03, Nist-04 and Fisher
800 segments (one side)
Computation 264h
Female : Nist-03, Nist-04 and Fisher
883 segments (one side)
Computation 300h

Adaptation :

- 65h for targets and impostors models
- 202h for tests models

Impostors: male: 373 (from Fisher database)
Female: 373 (from fisher database)

This set is used to evaluate the discriminative boundary between target GMM super vector and impostors GMM super vectors.

Scoring :
ENST-IRCGN 02 : 18h for kernels computation and score evaluation.

System 3: ENST_IRCGN_03:

This system results from a score fusion of three different applications. The first one is ENST_IRCGN_01, the second one, UPMC_PRC system, is described in [3], and the third one UNIFR-INT System, in [4]. The score are fused using the SVM technique. The SVM combiner uses a linear kernel and is trained on the NIST2005.

Thanks for Christophe Charbuillet from UPMC and Asmaa El Hannani from UNIFR-INT
for their contribution to the third system.

References:

[1]: “Becars: a free software for Speaker verification” in Proc. Speaker Odyssey Toledo 2004, Raphael Blouet , Chafic Mokbel , Hoda Mokbel , Eduardo Sánchez Soto , Gérard Chollet , Hanna Greige
[2]: Nist Speaker Recognition Evaluation 2006 – LRDE EPITA – R. Dehak, C. Deledalle, N. Dehak. in Proc. NIST Speaker recognition Workshop – PuertoRico 2006
[3]: UPMC_PRC system description - C. Charbuillet, Bruno Gas, Mohamed Chetouani, Jean-Luc Zarander in Proc. NIST Speaker recognition Workshop – PuertoRico 2006
[4]: “UNIFR-INT System Description” Asmaa El hanani, D. Petrovska Délacretaz in Proc. NIST Speaker recognition Workshop – PuertoRico 2006
[5] G. Gravier - http://irisa.fr/metiss/guig/spro