UNIVERSIDAD POLITÉCNICA DE MADRID

research group GIAPSY ©UPM

GIAPSI 2010 Speaker Recognition System

NIST SRE 10

Álvarez Marquina, A. ; Gómez Vilda, P. ; Mazaira Fernández, L.M.

04/05/2010

Grupo de Informática Aplicada al Procesado de Señal e Imagen (GIAPSI)
Universidad Politécnica de Madrid, Campus de Montegancedo, s/n
28660 Madrid, España, tel.: +34.91.3367384, fax: +34.91.3366601
e-mail: luismiguel.mazaira@ upm.es

INDEX

Grupo de Informática Aplicada al Procesado de Señal e Imagen (GIAPSI)
Universidad Politécnica de Madrid, Campus de Montegancedo, s/n
28660 Madrid, España, tel.: +34.91.3367384, fax: +34.91.3366601
e-mail: luismiguel.mazaira@ upm.es

1. SYSTEM DESCRIPTION.

The GIAPSI group has submitted results obtained from two different systems, UPMfim_1 (primary) and UPM_fim2 (alternative):

1.1 Primary System.

- Preprocessing and Feature Extraction
  A preprocessing step for both Voice Activity Detection (VAD) and Noise Reduction (NR) has been applied to all the evaluation data. An adaptive VAD algorithm based on energy detection has been implemented and computed over 64ms Blackman window length with 13ms overlapping. The NR algorithm is a variation of the Ephraim-Malah spectral subtraction algorithm in a single channel.

  For the feature extraction step, two different set of features have been computed. A 45 feature vector has been extracted for each 32ms voiced frame (with 8ms overlapping), which contains the following information:
    - Frame energy
    - 18 MFCC + 18 ΔMFCC from voiced signal
    - 8 MFCC extracted from the reconstructed glottal pulse.
- Speaker Modeling
  A gender-dependent 1024 GMM-UBM has been trained using the SRE06 data for both microphone and telephone speech. Speaker models have been adapted from this UBM using MAP algorithm.
  ALIZE toolkit has been used to generate these models.
- Scoring.-
  Only top 16 Gaussians have been used to produce a LLK score for each test segment. TNORM normalization has been applied based on a cohort of impostors extracted from SRE06 evaluation data.

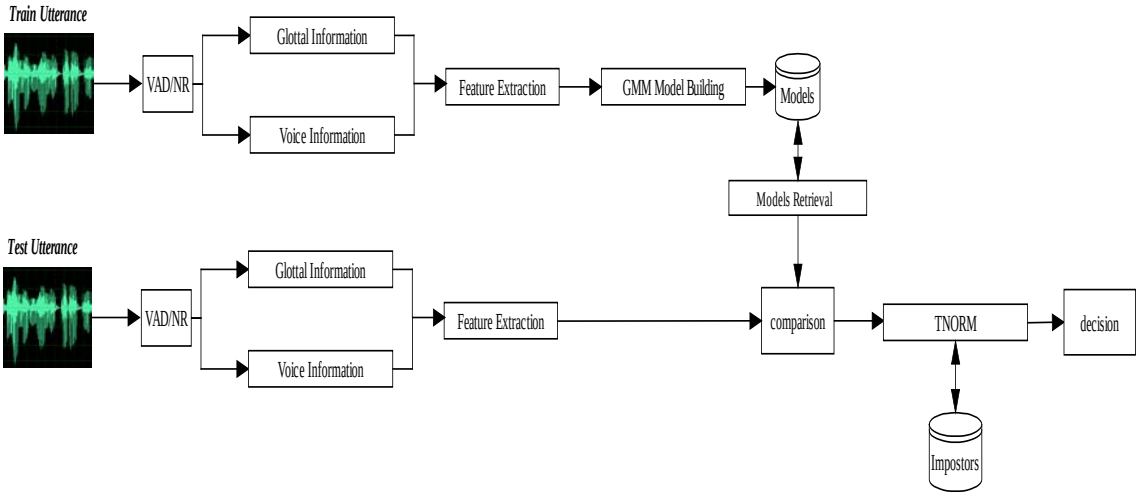

Figure 1 Model description

1.2 Alternative System.

This system ouputs a score based on a linear combination of the scores produced by two UBM-GMM systems.

This system differs from the previous one in the way features are used to create the model of the target speakers. This means that the preprocessing step is common in both systems.

- Feature Extraction.-
  In this case, two different vectors are extracted for each 32ms voiced frame (with 8ms overlapping). The first vector contains:
    - Frame energy
    - 18 MFCC + 18 ΔMFCC from voiced signal

  While the second one contains:
    - Frame energy
    - 8 MFCC extracted from the reconstructed glottal pulse

- Speaker Modeling
  Two 1024 GMM-UBM have been trained using the SRE06 data for both microphone and telephone speech. The first one with the classical set of features, and the second one with the glottal features. Speaker models have been adapted from this UBM using MAP algorithm.
  ALIZE toolkit has been used to generate these models.
- Scoring.-
  Only top 16 Gaussians have been used to produce a LLK score for each test segment. TNORM normalization has been applied based on a cohort of impostors extracted from SRE06 evaluation data.
- Fusion Strategy.-
  A linear fusion strategy has been adopted in this system:
    - $S_{spk}=\alpha S_{gmm\_voice} + (1-\alpha)S_{gmm\_glottal}$.
    - α is a coefficient trained from the development data set.

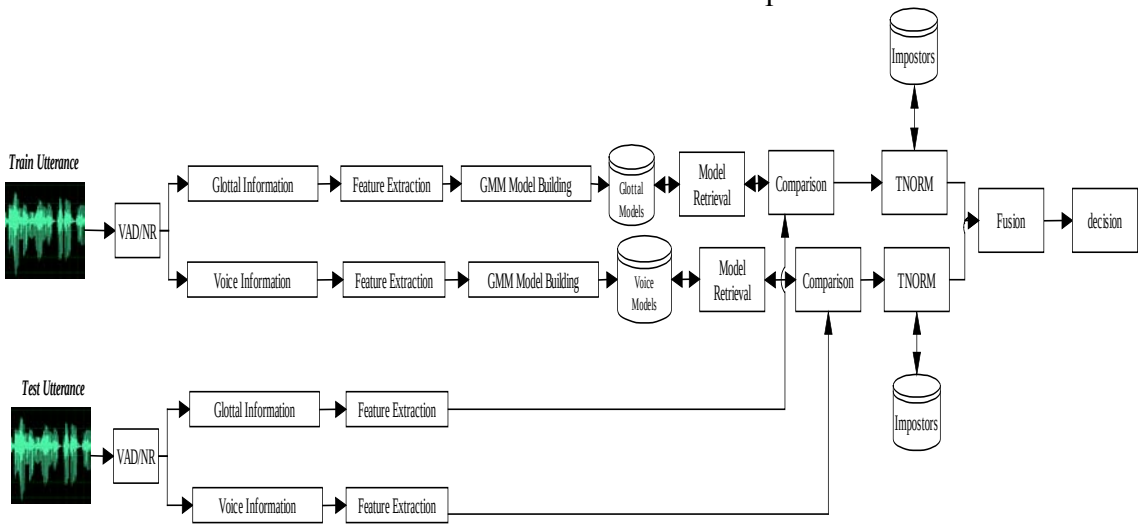

Figure 2 Alternative System Description

Grupo de Informática Aplicada al Procesado de Señal e Imagen (GIAPSI)
Universidad Politécnica de Madrid, Campus de Montegancedo, s/n
28660 Madrid, España, tel.: +34.91.3367384, fax: +34.91.3366601
e-mail: luismiguel.mazaira@ upm.es

2. CPU INFORMATION.

3 Intel Core i7 CPU 975 at 3.33GHz have been used for processing the evaluation data. Each one for each system and test conditions evaluated.

Task	Train	Test
10 sec System 1	0,14xRT	0,0012xRT
Core System 1	0,14xRT	0,0012xRT
Core System 2	0,26*RT	0,002XRT

3. EVALUATION TESTS.

3.1. Primary System.

Primary system (also kwon as System 1), has been used both for the core test condition and for the 10 sec train and 10 sec test condition.

3.2. Alternative system.

Alternative system (also kwon as System 2), has been used only to produce results for the core test condition.

Grupo de Informática Aplicada al Procesado de Señal e Imagen (GIAPSI)
Universidad Politécnica de Madrid, Campus de Montegancedo, s/n
28660 Madrid, España, tel.: +34.91.3367384, fax: +34.91.3366601
e-mail: luismiguel.mazaira@ upm.es