2001

A SPEAKER ODYSSEY

The Speaker Recognition Workshop

Poster Session

June 21, 2001 (Thursday): 1400 - 1600

William M. Campbell, Charles C. Broun
Text-Prompted Speaker Recognition with Polynomial Classifiers

Marcos Faundez-Zanuy
On The Model Size Selection For Speaker Identification

R. Stapert, **J.S.D. Mason**
Speaker Recognition and the Acoustic Speech Space

Sachin S. Kajarekar, **Hynek Hermansky**
Speaker verification based on broad phonetic categories

Hassan Ezzaidi, Jean Rouat , Douglas O'Shaughnessy
Combining pitch and MFCC for speaker identification systems

Jason Pelecanos, Sridha Sridharan
Feature Warping for Robust Speaker Verification

Ozgur Devrim Orman, Levent M. Arslan
Frequency Analysis of Speaker Identification

Raphael Blouet, Frederic Bimbot
A Tree-based approach for score computation in speaker verification

Text-Prompted Speaker Recognition with Polynomial Classifiers

W. M. Campbell, Charles Broun

Motorola Human Interface Lab

- **What?** Using polynomials instead of traditional classifiers for speaker recognition.
- **Why?** Interesting architecture advantages:
  – Low computational complexity – scalable
  – Simple training
  – Accurate
  – *A posteriori* probabilities eliminate background normalization.
- **How?**

ON THE MODEL SIZE SELECTION FOR SPEAKER IDENTIFICATION

Marcos Faundez-Zanuy

- **MODEL SIZE SELECTION TRADE-OFF**
    - Is a critical fact on pattern recognition, polynomial fitting, etc
    - ➢ If the number of parameters is small, there is not enough precision to model the data.
    - ➢ If the model has a lot of parameters there is an overfit, so the model is unable to generalize and manage mismatch situations.
    - Usually, the same model size is used for all the speakers, and this is the unique optimized parameter.

We propose to use a different model size for each speaker

Speaker Recognition and the Acoustic Speech Space

R. Stapert, John Mason

Abstract:

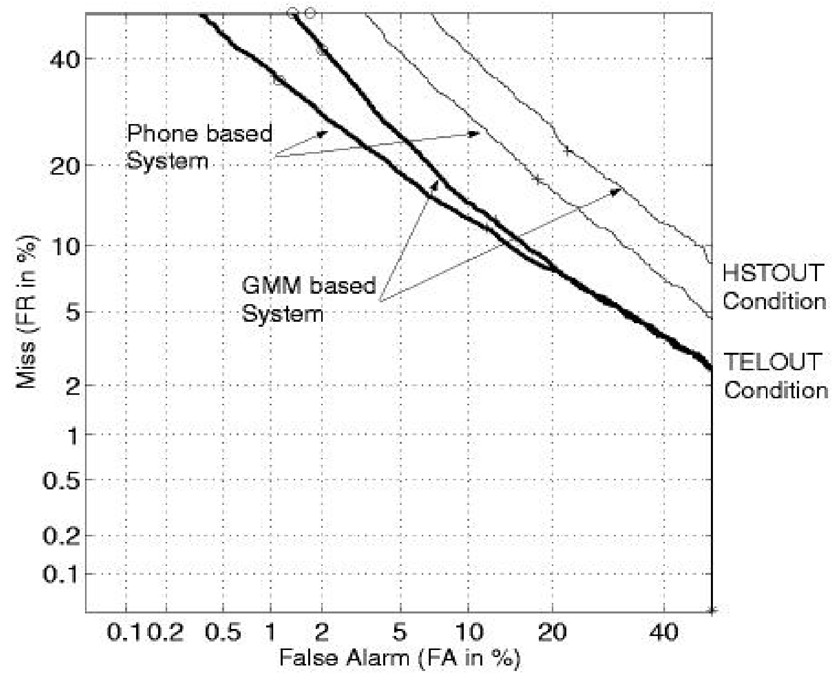
The hypothesis that for a given amount of training data a speaker model has an optimum number of components is examined. This is investigated with regard to Gaussian mixture models with and without world model adaptation. Results show that maximising the number of components in a speaker model can improve speaker recognition results. Comparisons with vector quantisation indicate that sensible use of out-of-class data is essential for optimising a recognition system.

Speaker Verification using Broad Phonetic Categories

Hynek Hermansky, Sachin S. Kajarekar

- Categories- Vowels, diphthongs, glides, nasals, stops, and silence
- Analysis of phone-specific speaker variability
  - Vowels, diphthongs, fricatives and nasals are most speaker-specific sounds
- Speaker Verification System
  - Hidden Markov models used for modeling the categories
  - SI model used for obtaining labels
  - Use vowels, diphthongs, fricatives and nasals for testing

NIST 2000 Speaker Verification Evaluation



Phone-based system outperforms GMM-based system in both matched-handset and mismatched-handset condition

Combining pitch and MFCC for speaker recognition systems

Hassan Ezzaidi, Jean ROUAT and Douglas O'Shaughnessy+
Université du Québec à Chicoutimi, DSA, Canada
+ INRS-Télécommunications, Université du Québec, Montréal, Canada

- A model of joint probability is proposed to retain the dependence between the vocal source and the vocal tract.
- Three strategies are compared:
  - 1. The baseline system operating on all voiced and unvoiced speech segments;
  - 2. The voiced system considers only the voiced segments;
  - 3. The Pitch Dependent Vocal Track Models includes the pitch information with the standard MFCC.
- Two pattern recognizers are used : GMM and LVQ-SLP.
- Results show an increase in the identification rates (specifically for short time duration test).

Feature Warping for Robust Speaker Verification

Jason Pelecanos, Sridha Sridharan

Abstract:

We propose a novel feature mapping approach that is robust to channel mismatch, additive noise and to some extent, non-linear effects attributed to handset transducers. These adverse effects can distort the short-term distribution of the speech features. Some methods have addressed this issue by conditioning the variance of the distribution, but not to the extent of conforming the speech statistics to a target distribution. The proposed target mapping method warps the distribution of a cepstral feature stream to a standardised distribution over a specified time interval.

We evaluate a number of the enhancement methods for speaker verification, and compare them against a Gaussian target mapping implementation. Results indicate improvements of the warping technique over a number of methods such as Cepstral mean Subtraction (CMS), modulation spectrum processing, and short-term windowed CMS and variance normalisation.

Frequency Analysis of Speaker Identification

Özgür D. Orman[1,2], Levent M. Arslan[2]
[1]TÜBİTAK-UEKAE
[2]Boğaziçi University, İstanbul

- Subband based representation of speaker acoustics

  Is there a better subband representation of speaker ?
- Vector Ranking (VR) criteria for ASI performance evaluation
- Identification Performance Index (IPI)
- Comparision of VR and F-ratio results
- Significance of frequency bands in speaker discrimination

  An acoustic feature set which is well defined for ASR (MFCCs) might not be optimal for ASI.
- New filterbank for ASI
- Better performance than MFCCs based ASI

  6 % performace increase

A Tree-Based Approach for Score Computation inSpeaker Verification

Raphaël BLOUET and Frédéric BIMBOT

IRISA (CNRS & INRIA) - Campus Universitaire de Beaulieu - Rennes - France
{rblouet,bimbot}@irisa.fr

- An original approach for Speaker Verification
  - State-of-the-art :

    consists in estimating a client and non-client *probability density function* and in computing a decision score based on the likelihood ratio test.
  - In our approach the training process consists in a direct estimation of local density ratio.
- Realization
  - The feature space is first split in disjoint regions. Then a constant score is assigned to each of the region.
  - The CART algorithm is an efficient tool that provides the feature space partition. Given a criterion, it recursively finds a sub-optimal split of the feature space.
  - The score function is derived from ML estimation of local densities.
- Result
  - Results on the NIST 2000 Speaker Verification Evaluation are presented
- Several tracks to improve the performance are given