Contents

Abbreviations

ASV	Automatic Speaker Verification
BW	Baum-Welch
CMS	Cepstral Mean Subtraction
DET	Detection Error Trade-off
DP	Dynamic Programming
DTW	Dynamic Time Warping
EER	Equal Error Rate
EM	Expectation Maximization
FA	False Accept (or False Alarm)
FFT	Fast Fourier Transform
FR	False Reject
FS	Feature Selection
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
IBM	Individual Background Model
LAR	Log-Area Ratio
LPC	Linear Predictive Coefficients
MFCC	Mel-Frequency Cepstral Coefficients
ML	Maximum Likelihood
PDF	Probability Density Function
RRC	Recognition Related Criterion
SFFS	Sequential Floating Forward Sequence
SNR	Signal to Noise Ratio
STD	Standard Deviation
SVM	Support Vector Machine
UBM	Universal Background Model
VAD	Voice Active Detection
VQ	Vector Quantization

Notations

x	Scalar
$\mathbf{x}, \boldsymbol{\mu}$	Column Vector
$\boldsymbol{\Sigma}, \mathbf{O}$	Matrix
$\mathbf{a}^T, \mathbf{A}^T$	Transposition of Vector/Matrix
$\boldsymbol{\Sigma}^{-1}$	Inverse of matrix
$\lvert\boldsymbol{\Sigma}\rvert$	Determinant
$\mathrm{tr}[\boldsymbol{\Sigma}]$	Trace
X	Set
$p(\ )$	Probability density function
$\log(\ )$	Logarithm (base e)
$\sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Normal density with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$\mathrm{E}[\ ]$	Expectation operation
$\mathrm{std}[\ ]$	Standard-deviation operation
$\mathrm{erf}(\ )$	Error function
$x(n)$	Sampled signal at discrete time n
$x(t)$	Signal at continuous time t
$X(z)$	z-transform of $x(n)$

List of figures

List of tables

Abstract

Today's Automatic Speaker Verification (ASV) systems use a common feature space for all speakers. This common set of features is usually a set of cepstral and delta-cepstral coefficients, which are used for speech recognition tasks. This research is based on the assumption that every speaker has his own 'optimal' feature space, which optimally discriminates him from other speakers. The goal of this work is to demonstrate the significance of employing an individual feature space in modern Continuous Density Hidden Markov Model (CD-HMM) or Gaussian Mixture Model (GMM) based-verification systems.

In this research, a new criterion for feature selection was developed, which is suitable for speaker verification tasks and correlated with the recognition rate, named "Recognition Related Criterion" (RRC). Several feature selection procedures were implemented and tested along with the new feature selection criterion, such as k-best, forward selection, Sequential Floating Forward Sequence (SFFS), and Dynamic Programming (DP). We also implemented two speaker verification systems that combine an individual feature selection algorithm: the first is a text-dependent verification system, based on an HMM classifier and the second is a text-independent verification system, based on a GMM classifier.

The proposed HMM-based verification system was evaluated on a text-dependent database. A significant improvement over the "standard" Mel Frequency Cepstrum Coefficients (MFCC) space (12 MFCC + 12 ΔMFCC) in verification results was demonstrated with the selected individual feature space. An EER of 0.7% was achieved when the feature set was the MFCC space. Under the same conditions, the system based on the selected individual feature space (order of 24) yielded an EER of only 0.48%. It was found that the two best selection procedures are the DP and the SFFS. However, the SFFS

was more efficient than the DP in terms of calculation load. This system was also evaluated on a noisy text-dependent database. It was found that when using an individual feature space the verification system becomes very sensitive to changing environmental conditions, meaning, different Signal to Noise Ratios (SNR) for training and testing database.

The proposed GMM-based verification system was evaluated on a text-independent database. A significant improvement in verification results was demonstrated with the selected individual feature space. An EER of 6.14% was achieved when the feature set was the MFCC space. Under the same conditions, a system based on the selected feature space yielded an EER of only 4.15%.

Keywords:

Speaker Verification, Feature Selection, Individual Feature Space, Equal Error Rate (EER), Recognition Related Criterion (RRC), Hidden Markov Model (HMM), Gaussian Mixture Model (GMM).