Optimizing Feature Set for Speaker Verification

D. Charlet, D. Jouvet

France Télécom, CNET Lannion, LAA/TSS/RCP,
Technopole Anticipa, 2 avenue Pierre Marzin, 22307 LANNION, FRANCE
e-mail: charlet,jouvet@lannion.cnet.fr

Abstract. This paper proposes a framework for the optimization of the feature set, in an HMM-based text-dependent speaker verification system, in which we distinguish the alignment task from the scoring task. The optimization is based on the search, among a set for potential features, for the feature subset that gives the minimal experimental Equal Error Rate. We have studied and compared various heuristics to find the optimal subset. We have also extended this optimization principle to the search for an optimal weighting of the different axes of the acoustic space. The optimal weighting was found by using a genetic algorithm.

The proposed framework was applied to study cepstral coefficients and their first and second derivatives, in order to find a speaker and text-independent optimal feature set. Experiments were conducted on a large scale telephone database. The results indicate that the selection of an appropriate feature set, or the appropriate weighting of the features, could significantly improve verification performance, especially when little training data is available. Practically, it was found that cepstral coefficients of high order and first derivatives of all cepstral coefficients are the most useful for speaker verification.

1 Introduction

Speaker verification is the task that decides whether an unknown speaker is the speaker he claims to be, through the analysis of his speech signal.

A key-issue is the acoustical feature set used for analyzing the speech signal. Indeed, the feature set is required to convey as much as possible speaker-dependent information. Section 2 briefly recalls different ways to study this problem which has been a focus of interest for a long time.

The proposed optimization method is presented in section 3. It consists in selecting, among a set of potential features, the feature subset that minimizes the experimental error rate, in an particular HMM-based system. In this system, we distinguish the alignment task from the scoring task. Experimental results are reported in section 4 that shows the interest of the proposed method. Finally, section 5 proposes the determination of a discriminative weighting of the different features, through the use of genetic algorithms.

2 The Problem of the Feature Set

The most common approach to study features is to associate a measure of quality to a potential feature. The advantage is that the evaluation of a feature is independent of the verification system used, assuming the measure is relevant. A widespread measure of quality is the F_{ratio} [1]:

$$F_{ratio} = \frac{\sigma_b^2}{\bar{\sigma_w^2}} \tag{1}$$

where σ_b^2 is the variance of the means of the speakers, and $\bar{\sigma_w^2}$ the mean of the variances intra-speakers. A "good" feature must vary a lot between speaker (σ_b^2 high) and have little intra-speaker variation ($\bar{\sigma_w^2}$ low), so a good feature has a high F_{ratio}. Other measures, derived from the F_{ratio}, have been proposed [2][3].

Another approach consists in considering feature subsets rather than features studied individually. The advantage is that it directly accounts for interactions between features [4] [5]. In addition, a criterion is necessary to decide whether a feature set A is better than a feature set B. [4] and [5] use theoretical criteria such as divergence or estimation of the error probability. However, we consider that directly using the experimental error rate as comparison criterion is more attractive to reach our goal, namely to obtain the best possible system (i.e. with the minimal error rate).

3 The Proposed Optimization Method

3.1 Principle

We are interested in finding the subset that gives the minimal error rate among a set of potential features. As for speaker verification, the error rate depends on the value of the decision threshold, we consider a particular error rate, called Equal Error Rate (EER), for which the false acceptance rate is equal to the false rejection rate.

Finding the optimal subset is an optimization problem that can be divided into 2 parts. First, we have to define the evaluation function of a potential feature set; in our case, we use the EER. Then, we have to optimize this evaluation function to find the best feature set.

3.2 Evaluation of a Feature Set

The verification system we used is HMM-based and text-dependent. We want to compute, for a given speech signal X, the probability $P(M/X)$ of the speaker M. Applying the Bayesian rule, we can compute the probability that X has been emitted by the model M of the speaker (for a given keyword). For the optimal path $\hat{s}$, this probability is given by:

$$P(X, \hat{s}/M) = \pi_0 . \prod_{\tau=1}^{T} \{a_{[\tau-1][\tau]} . G_{[\tau-1][\tau]}(X(\tau))\} \tag{2}$$