Introducing a FM based Feature to Hierarchical Language Identification

Bo Yin[1,2], Tharmarajah Thiruvaran[1,2], Eliathamby Ambikairajah[1,2], Fang Chen[2,1]

[1]School of Electrical Engineering and Telecommunications,
The University of New South Wales, Sydney, NSW 2052, Australia
[2]National ICT Australia (NICTA), Australian Technology Park, Eveleigh 1430, Australia
bo.yin@student.unsw.edu.au, ambi@ee.unsw.edu.au, fang.chen@nicta.com.au

Abstract

Although relatively neglected in auditory analysis, phase information plays an important role in human auditory intelligibility. This paper investigates a Frequency Modulation (FM) based feature and its contribution to a Language Identification (LID) system, using a Hierarchical LID framework. FM components represent the phase information of a given signal in an AM-FM model. In this paper, we extract a FM-based feature using a technique which produces consistent and continuous FM components, and build a LID system on this feature with GMM based modeling. The performance is improved by combining this system with existing MFCC, Prosody based systems and a PRLM system. When compared to the baseline system without integrating a FM-based system, the proposed Hierarchical LID system shows improvements. Additionally, the proposed system outperforms the GMM fusion-based system integrating the same four primary systems, showing that the Hierarchical LID framework is more effective in integrating additional features.

Index Terms: frequency modulation, language identification, hierarchical classification, fusion

1. Introduction

Recent research in Language Identification (LID) has mainly focused on back-end classification and feature enhancement techniques. Various classifiers, learning schemes and fusion techniques have been proposed and evaluated, such as Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) classifiers for acoustic systems, and Hidden Markov Model (HMM) classifiers for phonetic systems. More recently, positive results have been achieved by SVM based hybrid classifiers which integrate another classifier in SVM by using the likelihood scores as a super vector, e.g. SVM-GMM [1]. To incorporate a number of LID systems which utilize different features and/or varied classifiers, fusion techniques have been the subject of much investigation, such that most state-of-the-art systems are fusion-based systems. By fusing the scores from different classifiers, a higher performance can be achieved than from independent LID systems, even if any one primary LID system does not exhibit superior performance to others. However, exploring alternative features other than those spectrum magnitude based features (e.g. MFCC) has drawn less attention.

Recent research has found that phase information in speech signals, along with magnitude information, contributes to auditory intelligibility [2]. Several phase-related features have been proposed and evaluated in varied speech processing research areas [3]. Due to the fact that these features are relatively unstable and it is difficult to separate the phase information accurately, there has been little progress until recently with Robust ASR and Speaker Identification [4, 5], where a Modified Group Delay Function (MODGDF) is used. MODGDF has been tested in Language Identification but did not show a significant or stable performance improvement [4]. With this as motivation, we investigated an alternative phase-related feature and introduced this feature to a Hierarchical LID system. The Hierarchical LID framework has been shown to be more effective in integrating multiple features into the LID system than commonly used fusion-based systems [6, 7].

The alternative phase-related feature comes from the Frequency Modulation (FM) components of a speech signal in the AM-FM model [8]. Examining the extraction process of the commonly used Mel-Frequency Cepstral Coefficients (MFCC), we notice that only the magnitude part of the short-time spectrum is used to produce the coefficients, by applying a Mel-scale filterbank and Discrete Cosine Transform (DCT). The phase part of the spectrum is simply discarded. This method is not unique to the MFCC process, but is practiced in other spectral magnitude-based feature extraction techniques, because classic theories did not suggest a significant contribution of phase information to human audible intelligence [2]. In this paper, a FM-based feature is introduced to carry phase-related information. By utilizing a novel all-pole model to extract FM components, this feature extraction technique produces a more consistent FM-based feature representation.

This FM based feature is modeled by a GMM based classifier with the help of Universal Background Model (UBM), Shifted Delta Coefficients (SDC) and Segmental Histogram Normalization (SHN) [9]. SDCs capture temporal variation patterns and SHN reduces channel/speaker variation. The evaluation results show performance improvement by introducing the FM-based feature, compared to a LID system using MODGDF in previous research [4].

There is more than one approach to integrate a FM feature-based LID system to existing LID systems. Fusion-based LID systems have achieved the state-of-the-art performance for their ability to integrate multiple LID systems which utilize different features or modeling techniques. Most fusion techniques such as linear weighting, GMMs, or Neural Networks accept a combination of the likelihood scores produced by primary LID systems and produce another set of likelihood scores for a final decision. However, existing fusion techniques sometimes experience difficulty in improving performance when the number of languages and features increases [6], because all language hypotheses are examined at a single level. This means the variation of the distances between languages in different feature spaces is not sufficiently considered.

The Hierarchical LID (HLID) framework has been proposed [6] to alleviate this performance degradation. This