NORMALIZATION OF MODULATION FEATURES FOR SPEAKER RECOGNITION

Tharmarajah Thiruvaran[1,2], *Eliathamby Ambikairajah*[1,2], *Julien Epps*[3,2]

[1]School of Electrical Engineering and Telecommunications,
The University of New South Wales, Sydney NSW 2052 Australia.
[2]National Information and Communication Technology, Australia (NICTA),
Australian Technology Park, Eveleigh 1430, Australia.
[3]UNSW Asia, 1 Kay Siang Road, Singapore 248922.

ABSTRACT

Modulation features are emerging as a more recent alternative to more conventional magnitude-based features for speech processing applications such as speaker recognition. Frequency Modulation (FM) based features are one such example and in this paper their normalization using feature warping is examined in detail. Evaluations of different FM feature and warping configurations on the NIST 2001 Speaker Recognition corpus show that feature warping is not an effective normalization technique for FM features, despite its well-known effectiveness for Mel frequency cepstral coefficients (MFCC). This study further suggests a closer investigation on feature dependency of the existing system when using new features.

***Index Terms*—** Feature Warping, FM, Speaker Recognition.

1. INTRODUCTION

Speaker recognition is the task of identifying speakers based only on their speech signals. For a real world application, this system should be robust to environmental noise, channel effects and handset mismatch. This can be achieved at the front-end through a variety of means, such as utilising additional features or by normalizing the existing features to cancel the undesirable effects. FM features have recently received research attention as an additional feature for use with MFCCs [1, 2]. Such features are extracted based on the AM-FM model of speech signal [3]. In this model, the vocal tract resonances are characterised as AM-FM signals, and the speech signal is represented as the sum of all resonances. Conventional features such as MFCCs carry only information related to spectral magnitude, while the FM feature carries only information related to phase, specifically the phase variation.

Previous applications of FM features have suggested poorer effectiveness than magnitude-based features in speaker recognition systems developed principally for magnitude-based features. In order to extract the maximum information from FM features, we need to analyze the effectiveness of each component on an existing speaker recognition system when applying FM features.

To address this problem, in this paper we study the behaviour of FM features in an existing system at the feature normalization stage, because feature normalization techniques are performed on features to increase the robustness of the features at the front-end. There are many feature normalization techniques available for various speech processing front-ends based on conventional features. One such normalization is feature warping [4]. Because this normalization provides improved performance over other normalizing techniques such as cepstral mean subtraction and variance normalization [5], we investigate feature warping in conjunction with FM features herein.

2. FM FEATURE EXTRACTION

Several methods are available for the demodulation of FM from speech signal. Among them, the popular Digital Energy Separation Algorithm (DESA) [3] is used for FM extraction in our experiment. This method is based on an energy operator known as the Teager Energy Operator, and is defined as shown in equation (1).

$$\Psi(s(n)) = s^2(n) - s(n-1)s(n+1) \qquad (1)$$

The DESA algorithm for FM extraction is given in equation (2) with $y(n)$ defined as per (3).

$$FM \approx \arccos(1 - \frac{\Psi[y(n)] + \Psi[Y(n+1)]}{4\Psi[s(n)]}) \qquad (2)$$

$$y(n) = s(n) - s(n-1) \qquad (3)$$

In this paper, speech is analysed using a band pass filter bank with center frequencies spaced according to the critical band specifications. For this filter bank design Gabor filters are used, since they have the optimal time and frequency sensitivity and the side lobes are relatively insignificant [3]. The DESA algorithm is used to demodulate the FM from the

1-4244-0882-2/07/$20.00 ©2007 IEEE