Chapter 9
Robust Feature Vector Set Using Higher Order Autocorrelation Coefficients

Poonam Bansal
Guru Gobind Singh Indraprastha University, India

Amita Dev
Ambedkar Institute of Technology, India

Shail Bala Jain
GGSIP University, India

ABSTRACT

In this paper, a feature extraction method that is robust to additive background noise is proposed for automatic speech recognition. Since the background noise corrupts the autocorrelation coefficients of the speech signal mostly at the lower orders, while the higher-order autocorrelation coefficients are least affected, this method discards the lower order autocorrelation coefficients and uses only the higher-order autocorrelation coefficients for spectral estimation. The magnitude spectrum of the windowed higher-order autocorrelation sequence is used here as an estimate of the power spectrum of the speech signal. This power spectral estimate is processed further by the Mel filter bank; a log operation and the discrete cosine transform to get the cepstral coefficients. These cepstral coefficients are referred to as the Differentiated Relative Higher Order Autocorrelation Coefficient Sequence Spectrum (DRHOASS). The authors evaluate the speech recognition performance of the DRHOASS features and show that they perform as well as the MFCC features for clean speech and their recognition performance is better than the MFCC features for noisy speech.

DOI: 10.4018/978-1-4666-1743-8.ch009

INTRODUCTION

Cognitive informatics is an emerging interdisciplinary field in the cognitive and information sciences that aims to forge links between a diverse range of disciplines spanning the natural and life sciences, informatics and computer science (Wang, 2003). It has given introduction to the data-driven/machine-learning approach to building spoken language systems (in which large corpora of annotated speech recordings are used to capture the variability of speech) coupled with the relentless increase in available computing power. Any real time practical spoken language system must be robust and flexible in real-world environments (Pols, 1999; Saroka & Braida, 2005). To make it so, exact feature extraction during processing of speech signals plays a vital role. This paper discuss about that aspect of robust system.

The main approaches considered to improve the performance of automatic speech recognition (ASR) systems could be roughly divided into three main categories, namely, robust speech feature extraction, speech enhancement and model-based compensation for noise. In the case of speech enhancement, some initial information about speech and noise is needed to allow the estimation of noise and clean up of the noisy speech. Widely used methods in this category include spectral subtraction (SS) (Beh & Ko, 2003) and Wiener filtering (Lee, Soong, & Paliwal, 1996). In the framework of model-based compensation, statistical models such as Hidden Markov Models (HMMs) are usually considered. The compensation techniques try to remove the mismatch between the trained models and the noisy speech to improve the performance of ASR systems. Methods such as parallel model combination (PMC) (Gales & Young, 1995; Gales & Young, 1996), vector Taylor series (VTS) (Acero, Deng, Kristjansson, & Zhang, 2000; Kim, Un, & Kim, 1998; Moreno, 1996; Moreno, Raj, & Stern, 1996; Shen, Hung, & Lee, 1998) and weighted projection measure (WPM) (Mansour & Jaung, 1989a) can be classified into this category.

Use of the autocorrelation domain in speech feature extraction has recently proved to be successful for robust speech recognition. Among the techniques introduced that exploit the autocorrelation properties are Short-time Modified Coherence (SMC) (Mansour & Jaung, 1989b) and One-Sided Autocorrelation LPC (OSALPC) (Herando & Nadeu, 1997). Extracting appropriate speech features is crucial in obtaining good performance in ASR systems since all of the succeeding processes in such systems are highly dependent on the quality of the extracted features. Therefore, robust feature extraction has attracted much attention in the field.

Noise-robust spectral estimation is possible with algorithms that focus on the higher order autocorrelation coefficients such as autocorrelation mel-frequency cepstral coefficient (AMFCC) method (Shannon & Paliwal, 2006). Moreover, as the autocorrelation of noise could in many cases be considered relatively constant over time, a high pass filtering of the autocorrelation sequence, as done in relative autocorrelation sequence (RAS) (You & Wang, 1999), could lead to substantial reduction of the noise effect. Furthermore, it has been shown that preserving spectral peaks is very important in obtaining a robust set of features for ASR (Padmanabhan, 2000; Srope & Alwan, 1998; Sujatha, Prasanna, Ramakrishnan, & Balakrishnan, 2003). Methods such as peak-to-valley ratio locking (Zhu, Iseli, Cui, & Alwan, 2001) and peak isolation (PKISO) (Strope & Alwan, 1997) have been found very useful in speech recognition error rate reduction. In differential power spectrum (DPS) (Chen, Paliwal, & Nakamura, 2003), as an example, differentiation in the spectral domain is used to preserve the spectral peaks while the flat parts of the spectrum, that are believed to be more vulnerable to noise, are almost removed.

In this paper we propose a novel Differentiated Relative Higher Order Autocorrelation sequence spectrum (DRHOASS) method for computing MFCC feature vector set. In this method, removal