

Abstract

Mohamed E.Tamazin

Enhanced Automatic Speech Recognition System Based on Enhancing Power-Normalized Cepstral Coefficients

Many new consumer applications are based on the use of automatic speech recognition (ASR) systems, such as voice command interfaces, speech-to-text applications, and data entry processes. Although ASR systems have remarkably improved in recent decades, the speech recognition system performance still significantly degrades in the presence of noisy environments. Developing a robust ASR system that can work in real-world noise and other acoustic distorting conditions is an attractive research topic. Many advanced algorithms have been developed in the literature to deal with this problem most of these algorithms are based on modeling the behavior of the human auditory system with perceived noisy speech. In this research, the power-normalized cepstral coefficient (PNCC) system is modified to increase robustness against the different types of environmental noises, where a new technique based on gammatone channel filtering combined with channel bias minimization is used to suppress the noise effects. The TIDIGITS database is utilized to evaluate the performance of the proposed system in comparison to the state-of-the-art techniques in the presence of additive white Gaussian noise (AWGN) and seven different types of environmental noises. In this research, one word is recognized from a set containing 11 possibilities only. The experimental results showed that the proposed method provides significant improvements in the recognition accuracy at low signal to noise ratios (SNR). In the case of subway noise at SNR = 5 dB, the proposed method outperforms the mel-frequency cepstral coefficient (MFCC) and relative spectral (RASTA)-perceptual linear predictive (PLP) methods by 55% and 47%, respectively. Moreover, the recognition rate of the proposed method is higher than the gammatone frequency cepstral coefficient (GFCC) and PNCC methods in the case of car noise. It is enhanced by 40% in comparison to the GFCC method at SNR 0dB, while it is improved by 20% in comparison to the PNCC method at SNR -5dB.