© 2013 FIZ Karlsruhe

io-port 05345813

Yapanel, Umit H.; Hansen, John H.

A new perceptually motivated MVDR-based acoustic front-end (PMVDR) for robust automatic speech recognition.

Speech Commun. 50, No. 2, 142-152 (2008).

Summary: Acoustic feature extraction from speech constitutes a fundamental component of automatic speech recognition (ASR) systems. In this paper, we propose a novel feature extraction algorithm, perceptual-MVDR (PMVDR), which computes cepstral coefficients from the speech signal. This new feature representation is shown to better model the speech spectrum compared to traditional feature extraction approaches. Experimental results for small (40-word digits) to medium (5k-word dictation) size vocabulary tasks show varying degree of consistent improvements across different experiments; however, the new front-end is most effective in noisy car environments. The PMVDR front-end uses the minimum variance distortionless response (MVDR) spectral estimator to represent the upper envelope of the speech signal. Unlike Mel frequency cepstral coefficients (MFCCs), the proposed front-end does not utilize a filterbank. The effectiveness of the PMVDR approach is demonstrated by comparing speech recognition accuracies with the traditional MFCC front-end and recently proposed PMCC front-end in both noise-free and real adverse environments. For speech recognition in noisy car environments, a 40-word vocabulary task, PMVDR front-end provides a 36% relative decrease in word error rate (WER) over the MFCC front-end. Under simulated speaker stress conditions, a 35-word vocabulary task, the PMVDR front-end yields a 27% relative decrease in the WER. For a noise-free dictation task, a 5k-word vocabulary task, again a relative 8% reduction in the WER is reported. Finally, a novel analysis technique is proposed to quantify noise robustness of an acoustic front-end. This analysis is conducted for the acoustic front-ends analyzed in the paper and results are presented.

Keywords: acoustic feature extraction; robust speech recognition; noise-robustness analysis
doi:10.1016/j.specom.2007.07.006