Synergy of Acoustic-Phonetics and Auditory Modeling towards Speech Enhancement

Om Deshmukh
Department of Electrical and Computer Engineering
University of Maryland, College Park.

Abstract

The problem addressed in this work is that of enhancing speech signals corrupted by additive noise. We have developed a model for speech enhancement called the Modified Phase Opponency (MPO) model. The MPO model is based on the physiologically-based Phase Opponency model that was designed to explain psychophysical detection of tones-in-noise and does not need to estimate the characteristics of the corrupting noise. The MPO-based speech enhancement scheme leads to the lowest value of the linear-predictive coefficients based objective measures and the highest value of the Perceptual Evaluation of Speech Quality (PESQ) measure compared to some of the other methods when the speech signals are corrupted by fluctuating noise. The performance of the MPO speech enhancement scheme is comparable to some of the other enhancement techniques when the global SNR and the noise type of the corrupting noise are not fluctuating. Automatic speech recognition experiments show that replacing noisy speech signals by the corresponding MPO-enhanced speech signals leads to an improvement in the recognition accuracies at low SNRs.

Combining the MPO speech enhancement technique with our Aperiodicity, Periodicity and Pitch (APP) detector further improves its performance. The APP detector was initially developed to estimate the proportion of periodic and aperiodic energy in the spectro-temporal channels in speech signals. The combined MPO-APP speech enhancement technique strikes a better balance between the amount of speech-deletions and noise-suppression as compared to the MPO technique.

Some of the ongoing research in developing MPO-APP-based noise-robust parameters for speech recognition in noise will also be discussed.