Abstract

Speech enhancement algorithms have been proposed over the years to reduce the effect of various background perturbations. These are based on different estimation rules and distributions of speech. This dissertation focuses on investigations of novel Log Gabor Wavelet (LGW) based Long-Term Squared Spectral Amplitude (LT-SSA) estimators that uses Maximum a Posteriori (MAP) criterion. At first, a new algorithm, based on LGW and Long term Cepstral Mean Subtraction (LT-CMS) is presented to reduce the effects of both noise and distortion. This is followed by development of LGW and MAP estimator that automatically determines the appropriate shrinkage rule from the statistics of the decomposition of noisy speech signal. The probability density function (pdf) of the filtered speech coefficient is modeled with Generalized Laplacian Distribution (GLD), which allows a high approximation accuracy for Laplace distributed real and imaginary parts of the speech coefficients. Next, the drawbacks associated with the previous approach are addressed by introducing a new method that uses inter-scale dependency between the coefficients and their parents. This is modeled by the family of Spherically Invariant Random Processes (SIRPs) and corresponding joint shrinkage estimators are derived by MAP estimation theory. Finally, the estimators are refined with the Speech Presence Uncertainty (SPU) probability. The performance evaluations are shown in terms of objective, subjective and composite measures on NOIZEUS database for eight different background perturbations. The proposed methods are compared against seven existing methods from the category of

spectral subtractive, Wiener-type, wavelet based and statistical model based algorithms. The results validate the superiority of the proposed methods in almost all noisy conditions and under different Signal to Noise Ratios (SNR). An experiment is conducted to find the suitability of proposed enhancement algorithm on Automatic Speech Recognition (ASR) task in a standardized robust word and sentence recognition problems. The classifier is implemented by means of a Hidden Markov Model (HMM) and database used is noisy speech of TIMIT database. The results indicate that the proposed bivariate estimators perform better than other methods. Overall, the present study demonstrates that the assumed distribution of the Log Gabor Wavelet coefficients can have a significant effect on the quality of the enhanced speech signals.

Keywords: Speech Enhancement, Bayesian Estimation, Log Gabor Wavelet, Long term Cepstral Mean Subtraction, Generalized Laplacian Distribution, Spherically Invariant Random Processes, Speech Presence Uncertainty, NOIZEUS database, Automatic Speech Recognition, TIMIT database.