Title: Phase-Aware Speech Enhancement Methods using Statistical Signal Processing and Deep Learning

Name: Suman Samui

(Roll no. 13AT91R03)

Abstract

This thesis has primarily focused on effectively utilizing the areas of statistical signal processing and deep learning to solve the problem of single-channel speech enhancement. The statistical estimation based approaches for speech enhancement mainly process the signal using a time-frequency representation, most frequently in the short-time Fourier transform (STFT) domain. In the majority of these techniques, the short-time spectral amplitude (STSA) is modified (enhanced) by applying a gain function which can be derived from the assumed statistical spectral distribution of speech and noise signals, while the noisy spectral phase has been left unchanged, mainly because of the earlier trend towards unimportance of phase in audio perception. However, many recent psycho-acoustical studies and experiments have shown the positive impact and usefulness of the phase spectrum for improving the speech intelligibility and perceived quality in the context of speech enhancement. Being motivated by these observations, in this thesis, we have proposed various speech enhancement techniques which exploit the properties of spectral phase. We have termed these proposed techniques as phase-aware speech enhancement methods. Firstly, we have derived the gain function of a phase-aware parametric Bayesian STSA estimator under the generalized Gamma speech prior assumption. In addition, the performance of the proposed psycho-acoustically motivated Bayesian STSA estimator is evaluated with different types of STFT domain phase estimation methods. Furthermore, a new phase-aware spectral subtraction method, namely Multi-Band Complex Spectral Subtraction has been proposed for enhancing speech under low SNR conditions. Next, we have examined the intricate theoretical details of the learning algorithm necessary to train a deep neural architecture for developing a Time-Frequency masking based speech enhancement system. It has been mainly established that significant performance improvement can be achieved if the deep neural network is pre-trained by using Fuzzy Restricted Boltzmann Machines rather than using regular Restricted Boltzmann Machines and phase-sensitive training target is selected for DNN training. Next, leveraging the temporal information in speech, a single-channel speech enhancement framework using Recurrent Temporal Restricted Boltzmann Machines has been proposed to jointly model all the sources within a mixture as targets to a deep recurrent neural network.

Keywords: Speech enhancement, Statistical signal processing, Bayesian estimator, Deep learning, Restricted Boltzmann machine, Phase-aware speech enhancement, Time-frequency masking.