INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
SHORT ABSTRACT OF THESIS

Name of the Student	:	**Bidisha Sharma**
Roll Number	:	**136102017**
Programme of Study	:	**Ph.D.**
Thesis Title	:	**Improving Quality Of Statistical Parametric Speech Synthesis Using Sonority Information**
Name of Thesis Supervisor(s)	:	**Prof. S R Mahadeva Prasanna**
Thesis Submitted to the Department/ Center	:	**EEE**
Date of completion of Thesis Viva-Voce Exam	:	**01.03.2018**
Key words for description of Thesis Work	:	

SHORT ABSTRACT

This thesis aims towards improving naturalness and intelligibility of synthesized speech obtained from statistical parametric speech synthesis (SPSS). Along with the conventional source and spectral information, some additional significant features can also be derived from the speech signal to preserve its characteristics in parametric form. The sonority information represents spectral prominence, higher energy and periodicity aspects, which are related to human speech perception, that change with the varying vocal-tract constriction and glottal source amplitude during speech production. Therefore, this information is extracted from the speech signal in terms of sonority feature. It is capable to delineate the degree of sonority associated with a sound unit. The sonority feature is incorporated in the SPSS framework to use it in the studies related to this thesis.

To alleviate the over-smoothing effect from parameter sequences generated from SPSS, post-filtering mechanisms are found to be effective. By considering the fact that the characteristics of the speech parameters may extensively vary based on the broad categories of sound units, a class based dynamic post-filtering method is proposed. The excitation source (fundamental frequency and strength of excitation (SoE)) and spectral parameters (sharpness of peaks and valleys of the spectrum) corresponding to each frame are enhanced using post-filtering factors that change with sonorant sound categories. The sonorant class information is derived from a support vector machine based classifier trained using sonority feature associated with each frame. This method improves the temporal variation, fine spectral structure as well as reduces the deviation with the natural counterpart leading to improvement in synthesized speech quality.

Spectral slope is another aspect that influences on perception of synthesized speech. From the analysis of natural and synthesized speech, it is observed that the spectral slope of synthesized speech is more negative compared to that of the natural. Therefore, a novel method is proposed to modify the spectral slope of synthesized speech that

reduces the deviation in spectral tilt between natural and synthesized speech. The enhanced synthesized speech with modified flatter spectrum sounds clearer. It shows improvement in terms of naturalness, intelligibility and speaker similarity.

The voicing decision plays a significant role in excitation source generation module of SPSS. Along with the excitation source feature, the spectral prominence aspect in the sonority information makes it useful for voicing decision. The sonority feature is employed to develop a voiced/unvoiced classifier, that improves the naturalness compared to the existing methods of voicing decision. A common framework is developed that models the sonority feature and incorporates it in post-filtering and voicing decision followed by spectral tilt modification. The application of individual modules and their combination brings significant improvement to quality of synthesized speech obtained from SPSS.