INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
SHORT ABSTRACT OF THESIS

Name of the Student	:	K. T. Deepak
Roll Number	:	10610204
Programme of Study	:	Ph.D.

Thesis Title: **Foreground Speech Segmentation and Enhanecement**

Name of Thesis Supervisor(s)	:	Prof. S. R. M. Prasanna
Thesis Submitted to the Department/ Center	:	Department of Electronics and Electrical Engineering
Date of completion of Thesis Viva-Voce Exam	:	14 – 11 - 2016
Key words for description of Thesis Work	:	Zero band filtering, glottal closure instants (GCIs), foreground speech, background noise, spoken query system, automatic speech recognition (ASR)

SHORT ABSTRACT

Speech enhancement is one of the active areas of research and a challenging task when the signal is recorded in natural environments. In a typical recording scenario using a single microphone, it is safe to assume that the desired speaker is closer to the microphone sensor, relative to other interfering acoustic sources. In this work, the speech signal from close speaking person is regarded as foreground speech and rest of the interfering sources as {\it background noise}. Due to the close proximity of the desired speaker to the microphone, compared to other background sources, there are differences in the signal characteristics.

When the speech signal is recorded in natural environments, the production characteristics tend to vary depending on the levels of interfering sources. The objective of this thesis work is to exploit such unique characteristics of speech production to temporally segment foreground speech from rest of the background and further enhance it. The high signal to noise ratio (SNR) regions of foreground speech are robust to interfering noise. The high SNR region around glottal closure instants (GCIs) in the time domain and vocal tract information in the spectral domain is used to derive certain features to segment and enhance foreground speech.

This thesis proposes a new method to extract GCIs that does not need an estimation of pitch and is evaluated for its robustness under different degraded conditions. The method is used to study the nature of signal with reference to varying speaker to microphone distance and define a typical foreground speech recording scenario.

Foreground speech segmentation and enhancement methods are proposed. The methods exploit the high SNR regions in temporal and spectral domains. The temporal domain processing involves the derivation of gross and fine weight functions. Spectral processing involves enhancement of spectral peaks that represent formants and further exploit the perceptual result to minimize the effect of spectral distortion.

In a mobile based spoken query system, there is no control on recording environment from which a user can access such system. Due to other interfering sources, there can be degradation in the performance of automatic speech recognition (ASR) and thereby impact the usefulness of spoken query system to the users. This thesis demonstrates the effectiveness of foreground speech segmentation and enhancement as a pre-processing module to the spoken query system.