INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
SHORT ABSTRACT OF THESIS

Name of the Student	:	**RAMESH KUMAR BHUKYA**
Roll Number	:	**126102001**
Programme of Study	:	Ph.D.

Thesis Title:

Approaches for Robust Text-Dependent Speaker Verification Under Degraded Conditions

Name of Thesis Supervisor(s)	:	**Prof. S. R. Mahadeva Prasanna**
Thesis Submitted to the Department/ Center	:	**Electronics and Electrical Engineering**
Date of completion of Thesis Viva-Voce Exam	:	**09-11-2019**
Key words for description of Thesis Work	:	Text-Dependent Speaker Verification, Speech Enhancement, MEMD, IMFs, Hilbert Spectrum, End Point Detection

SHORT ABSTRACT

The objective of this thesis work is to develop a robust text-dependent speaker verification (TDSV) system by using robust techniques for achieving better system performance under clean and degraded speech conditions. To achieve this, three different directions are explored for a TDSV task. The existing TDSV system employs energy based end point detection, mel frequency cepstral coefficients (MFCCs) as features and dynamic time warping (DTW) for template matching. The same is treated as baseline system in this work. The performance of the baseline system affected depending on operating conditions in practice. The work attempts to improve the performance by providing robustness at different levels.

In practice, the speech signal is affected by the acoustic degradation present in the recording environment. This results in poor performance at different stages. One way is to first enhance the speech signal and then perform TDSV. The first novel contribution proposes combined temporal and spectral speech enhancement for enhancing speech regions embedded in background noise. The efficacy of the proposed framework is demonstrated by comparing the performance with the baseline system.

The spectral or cepstral based features, mainly MFCCs are used in the baseline system. In the next exploration, the goal is to develop new features. A new approach for feature extraction based on modified empirical mode decomposition (MEMD) is attempted. The Hilbert spectrum (HS) based features are extracted from the intrinsic mode functions (IMFs) of MEMD and used as features for TDSV.

The energy based end point detection is not enough under practical conditions. The third exploration proposes a new end-point detection method using speech-specific knowledge to remove non-overlapping speech and non-speech background degradation. The vowel-like regions (VLRs), dominant resonant frequency (DRF), foreground speech segmentation (FSS), glottal activity detection (GAD), obstruent (OBS) region detection, and speech duration knowledge (SDK) have been used to detect begin and end points accurately. Due to the speech-specific knowledge exploited, the algorithm is robust under different practical conditions.

In the final working chapter of the thesis, a robust TDSV system is developed by exploiting the evidences from these three different proposed approaches. A combined system is developed in a sequential manner, where robust end point detection is performed on the enhanced speech and then HS of the IMFs obtained from MEMD features are extracted from the regions between the detected end points. The combined method significantly improves the system performance.