**INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI**
**SHORT ABSTRACT OF THESIS**

| | | |
|---|---|---|
| Name of the Student | : | Rohan Kumar Das |
| Roll Number | : | 126102026 |
| Programme of Study | : | Ph.D. |

Thesis Title: Speaker Verification using Sufficient Train and Limited Test Data

| | | |
|---|---|---|
| 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 | : | 21-09-2017 |
| Key words for description of Thesis Work | : | speaker verification, limited data, short utterances, source information, text-constrained model, kernel discriminant analysis, practical systems. |

## SHORT ABSTRACT


The thesis focuses on speaker verification (SV) from the perspective of application oriented systems and identifies a framework of sufficient train with limited test data as the favorable one. Three different directions are highlighted that have scope towards improving the system performance with limited test data based scenario. These directions are investigated in detail and a combined system is proposed including the conducted explorations.

The source features provide information about the glottal excitation in the form of pitch period, strength of excitation, glottal signal shape, etc. Since the glottis and associated muscle structure are unique for each individual, the information represented by the source features is expected to be distinct for each speaker and can be utilized for SV. Three source features namely mel power difference of spectrum in subbands (MPDSS), residual mel frequency cepstral coefficient (RMFCC) and discrete cosine transform of integrated linear prediction residual (DCTILPR) are explored and their significances for limited test data cases are demonstrated. The source features are found to capture different attributes of source information, which on fusion provides comparable performance to the conventional mel frequency cepstral coefficient (MFCC) based vocal tract features.

The lexical match between train and test sessions is found to be beneficial for having an improved performance. With this motivation a text-constrained model based SV framework has been proposed. In this framework, the explicit utilization of speech content is made in order to have a better performance with limited test data based SV. The work is extended to include speaker-specific phonetic information in terms of the vocal tract constriction (VTC) feature. It captures the level of constriction produced in the vocal tract while producing different sound units, which has definite speaker information.

In the back-end of the SV system, it is necessary to have a suitable pattern recognition approach to handle limited test data. In this regard, kernel discriminant analysis (KDA) has been explored for the discussed SV framework. It maps the data points into higher dimensional space and performs discriminant analysis. The KDA has successfully

demonstrated its importance, especially for limited test data condition. Finally, a combined SV framework is proposed including the stated explorations aiming towards improving the performance in such a scenario.

Further, as the limited test data based SV framework is from the view of practical systems, there comes a lot of issues while going for deployment. Concerning this, few issues are investigated and attempted to address, which are beneficial for improving the SV performance. Mismatch in speaking rate, session variability and template aging issues are considered in this work.