



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

Name of the Student : SARFARAZ JELIL

Roll Number : 156102027

Programme of Study : Ph.D.

Thesis Title: Exploration of Source and Filter Information for the Detection of Replay Attacks in Speaker Verification.

Name of Thesis Supervisor(s) : Prof. Rohit Sinha and Prof. S.R.M. Prasanna

Thesis Submitted to the Department/ Center : Electronics and Electrical Engineering

Date of completion of Thesis Viva-Voce Exam : 27/01/2025

Key words for description of Thesis Work : Replay attacks, automatic speaker verification, spoofing attacks, ASVSpooF 2017, ASVSpooF 2019.

SHORT ABSTRACT

Automatic speaker verification (ASV) is defined as the task of accepting or rejecting an identity claim of a speaker based on their speech. ASV systems are prone to different kinds of spoofing attacks where the system is presented with a spoofed speech signal instead of a speech signal from a genuine speaker. These spoofing attacks can be a serious threat to an ASV system as they can increase false acceptance rates and negatively impact the performance of the system. Hence, it becomes essential to detect these spoofing attacks and protect the security of an ASV system. This thesis deals with a specific kind of spoofing attack called the replay attack. A replay attack is performed by secretly recording the speech of a genuine user of an ASV system and playing it back to the system to obtain unauthorized access.

A speech signal is produced as a result of the convolution between an excitation source signal and the vocal tract filter. It is hypothesized that the replay attacks will affect both the source and filter components of the speech signal and hence modeling the impact on these components separately will lead to better characterization of replay attacks. Thus, the core of this work is the separation of a speech signal into its components and the design of novel features to capture the replay information present in both components. The initial focus is on utilizing the source component for replay attack detection. To this end, two pitch-synchronous handcrafted source features are proposed which are motivated by a visual analysis of the source component. The handcrafted features are naively defined to extract information only around the glottal closure instants. Thus, the replay information available in the rest of the source signal cannot be utilized resulting in poor performance. This problem is addressed with the proposal of a transform-based pitch-synchronous source feature which is extracted from the region between two adjacent glottal closure instants in the spectral domain.

Extracting features between two glottal closure instants ensures that all voiced regions of the source signal are used. However, it is found experimentally that the unvoiced regions of the speech signal also contain significant replay information. Pitch-synchronous processing of the source signal implies that features can only be extracted from the voiced regions. Two modifications are proposed to overcome the problem involved in pitch-synchronous processing. The first modification is the use of non-pitch-synchronous processing and the second is the extraction of features in

the spectro-temporal domain instead of the spectral domain resulting in the development of a novel feature. Finally, the proposed spectro-temporal source feature is augmented with a novel feature extracted from the filter component of the speech signal. The augmentation of the source feature with the filter feature results in the best replay attack detection performance in this thesis. This validates our hypothesis that decomposing a speech signal into source and filter components and modeling both components separately results in a better capture of replay information than modeling the replay attacks directly from the speech signal.

