



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

Name of the Student : MRINMOY BHATTACHARJEE

Roll Number : 156102026

Programme of Study : Ph.D.

Thesis Title: ANALYSIS OF SPEECH AND MUSIC CONTENT FOR MOVIE GENRE CLASSIFICATION

Name of Thesis Supervisor(s) : Dr. PRITHWIJIT GUHA  
Prof. S. R. MAHADEVA PRASANNA (Dept. of EE, IIT Dharwad)

Thesis Submitted to the Department : ELECTRONICS AND ELECTRICAL ENGINEERING

Date of completion of Thesis Viva-Voce Exam : 30/01/2023

Key words for description of Thesis Work : Movie genre classification, speech music classification, audio signal processing, magnitude spectrum features, phase spectrum features, harmonic percussive source separation, multi-task learning, time and feature attention

---

**SHORT ABSTRACT**

Movies are a popular mode of entertainment around the world. The consistent rise in the production and consumption of movies demands more efficient automatic movie content analysis applications. Movie Genre Classification (MGC) is vital for underage censorship, search, retrieval, and targeted publicity. Current trends in MGC literature indicate a focus on short trailers instead of full movies and a multi-modal approach. The audio modality is generally used only as an auxiliary channel. However, due to its rich genre-specific information, the audio signal deserves a dedicated study in the current context. Hence, this thesis aims to perform only audio-specific MGC. The thesis has four principal contributions. First, spectral peak tracking-based magnitude spectrum features are proposed for isolated speech and music classification. Second, the underexplored phase component of the audio signals is utilized for discriminating speech and music. The third contribution involves using harmonic-percussive source-separated features and classifiers in the multi-task learning framework for identifying speech overlapped with music. Finally, the above proposals are employed for the MGC task. The spectral peak tracking-based method performs better than the other proposals and the baselines. Specific combinations of all the proposed and baseline features provide the overall best performance, even in the cross-dataset scenario. The thesis work can be extended in the future by analyzing the individual constituents of speech and music for a more nuanced representation of movie genres.