INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
SHORT ABSTRACT OF THESIS

Name of the Student	:	SIBASIS SAHOO
Roll Number	:	166302009
Programme of Study	:	MS + Ph.D.

Thesis Title:
Evaluation of Out-of-Breath Speech Using Machine Learning Approaches

Name of Thesis Supervisor(s)	:	Prof. SAMARENDRA DANDAPAT
Thesis Submitted to the Department/ Center	:	EEE
Date of completion of Thesis Viva-Voce Exam	:	17/04/2024
Key words for description of Thesis Work	:	Out-of-breath speech, stressed speech, breathing pattern, deep learning

---

SHORT ABSTRACT

Stress alters the speech production mechanism. Factors like emotion, cognitive load, pathology, noisy condition (Lombard effect), physical load, sleep deprivation, etc., affect speech production. Among these, speech under emotional, noisy, and pathological conditions are investigated extensively. Little light has been shed on speech under physical load conditions, called out-of-breath speech. Such evaluation of out-of-breath conditions can be used in context-aware speech interfaces to estimate the workload level, exercise intensity of an athlete, and physical fitness of a person.

In this thesis work, we investigate speech signals under the out-of-breath condition from the perspective of the speech production system concerning respiratory changes. For the same, four new databases are created containing speech utterances for the neutral and the out-of-breath conditions. The speech-based investigations are carried out as (i) the analysis of production characteristics, (ii) the detection of out-of-breath conditions, and (iii) the estimation of breathing characteristics from speech. It is observed that the out-of-breath condition influences the excitation source characteristics more than the vocal tract. With relaxation, speakers gain more control over the voicing apparatus while producing voiced sounds. The out-of-breath condition affects the lower frequency range of the speech spectrum. Therefore, it is proposed to use lower frequency-centric warped-spectral input to the deep neural networks (DNN), which produces a better detection rate compared to the linear spectral inputs. Also, the detection performance is enhanced further by deploying the DNN model in a multi-task-learning (MTL) setup, which uses the level of exertion as an auxiliary task. Finally, the breathing pattern (and rate) from speech is estimated using a convolutional neural network (CNN). Here, the out-of-breath speech is used as the testing condition for evaluating the robustness of the model. The proposed MsCNN uses several kernel-dilated branches to track speech changes at multiple scales. The improved correlation coefficient for out-of-breath speech ensures the robustness of the proposed model.