Leveraging Speech Breathing Parameters for the Detection of Stress in Young Adults

ABSTRACT

Stress is a response to physical, occupational, and emotional stimuli. When these stimuli exceed normal limits, they are termed stressors. Chronic or disproportionate exposure to stressors may lead to psychological disorders such as anxiety and depression. The Anxiety and Depression Association of America reports that only 36.9% of individuals with anxiety receive treatment. One primary reason for this treatment gap is the fear and stigma associated with being labelled as mentally ill. Recognizing that students are a particularly vulnerable population facing intense academic, social, and career pressures, this thesis addresses the urgent need for nonintrusive stress screening tools that overcome the limitations of traditional questionnaires and wired physiological sensors.

This thesis presents a comprehensive investigation into stress detection using speech breathing parameters. The research introduces a multimodal stress database to meet the critical demand for authentic and naturalistic data. This unique multimodal stress database captures speech, normal, and thermal video of students during real internship evaluations. This dataset covers diverse emotional states, including stress, neutral, and relaxed, and includes expert annotations, addressing limitations of acted or artificial stress datasets.

A novel breath annotation protocol is developed, combining auditory inspection with synchronized thermal and video analysis to accurately identify breath sounds embedded within speech. This multimodal approach is taken to overcome errors posed by perceptual recognition of breath sounds, thereby generating precise ground truth labels critical for model training.

Breath segment detection is tackled using two approaches: 1) traditional machine learning models leveraging cepstrogram features and voice activity detection; 2) deep learning architectures including bidirectional long short-term memory networks and convolutional neural networks. To address data imbalance and limited data, self-supervised learning is used, which uses jigsaw puzzle-based pretext tasks, improving robustness and detection accuracy. These methods are also validated on external sources such as YouTube videos.

A comprehensive set of speech breathing parameters from the detected breath segments, including temporal, energy-based, frequency-based, positional, and statistical parameters, are extracted for state anxiety classification. Comparative analysis reveals that these parameters outperform conventional speech features in identifying stress, albeit marginally.

The thesis also explores the impact of diaphragmatic breathing, a well-established stress reduction technique, on speech features. Results show significant changes in speech features consistent with relaxation, demonstrating the potential for speech-based biofeedback in stress management.

This work not only advances speech-based affective computing but also opens new avenues for non-intrusive mental health monitoring and intervention by detecting physiological parameters from communication modalities.

Keywords: Breath sound detection, diaphragmatic breathing, multimodal database, stress detection.