Ben-Gurion University of the Negev
Industrial Engineering and Management

אוניברסיטת בן-גוריון בנגב
המחלקה להנדסת תעשיה וניהול

Project No.	Project Title	
2022-01-126	Speech Emotion Recognition from Hebrew Speakers Using ML Methods	
Academic Advisor		**Co-Advisor**
Dr. Nir Nissim		Mr. Itzik Gurowiec, Mr. Trivikram Muralidharan
Team Members		
Itay Noy	Lior Rasuli	
Itayno@post.bgu.ac.il	liorras@post.bgu.ac.il	

Abstract

The field of Speech Emotion Recognition (SER) has been greatly developed in the last 20 years and continues to propose new and designated methods to improve SER and classification quality. The ability to classify and recognize emotions from human speech is with great of importance and can improve our lives in variety of domains such as human computer interface (HCI), interpersonal communication, security, lie detector, health care systems etc.

Modern SER systems utilize machine learning (and more broadly data science) methods to extract sound characteristics and leverage them to induce accurate classification models in order to further classify a new given speech to the correct emotion. The features from the sound domain are roughly divided into three types of features: Excitation source features; Spectral features and Prosodic features, while modern methods such as Deep Learning can automatically learn and engineer some other features that combine between the information concealed in the abovementioned feature types. Along the years, data science methods have been shown to be effective at speech emotion recognition, however, based on over extensive coverage and informative taxonomy of previous studies, we have identified some gaps in the existing studies, gaps that worth exploration.

First, while every language has its own uniqueness that affects the way the emotions are being expresses through the speech, there is a lack of data science-based studies that are aimed at recognizing emotions from Hebrew speakers. Secondly, the only study that was aimed at Hebrew speakers, has not utilized data science methods and was based on non-representative collection of only 40 entities, which put in doubt the generalization capabilities and robustness of their proposed system. Third gap is the poor use of authentic Natural Emotional Speech based data collections (without the knowledge of the subject being recorded for the purpose of the study), particularly, authentic Natural Emotional in Hebrew has not been studied at all.

In this paper, we propose a novel ML and DL-based SER system aimed at SER from Hebrew speakers. The proposed system is trained on an original data collection of audio recordings from TV interviews. Our data collection was based on 355 speech samples recorded from 112 different Hebrew speaker's subjects (57 males and 55 females), from variety of 4 main professions. The expressed utterances were labeled into three main emotions: joy, anger and sadness.

We trained various machine learning models that leveraged the aggregation of extracted prosodic and spectral features suggested by previous academic works. Moreover, we utilized deep learning Convolution Neural Network (CNN) and trained it over spectrograms representation of the speech. Our experimental design was aimed at addressing multiple research questions such as "Is it possible to recognize emotions from Hebrew speakers' utterance? "; "Is there any difference between the ability to recognize emotion from male and female Hebrew speakers? "; "Does the profession of a Hebrew speaker subject affect the ability to recognize emotions? "

Our encouraging results show that emotions in Hebrew speakers can be effectively detected by our proposed system. An accuracy of 72% was obtained using Random Forest classifier based on the combination of prosodic and spectral features: 13 MFCCs, ZCR, Energy using 3-BINS aggregation, while an accuracy of 80% was obtained by the CNN based model that was trained on speech spectrograms images extracted from the audio samples. Both outperforming models obtained an accuracy higher by 40% compare to the chance level, demonstrating the scientific contribution we have achieved in our study, and the potential for further improvements in such study.

Keywords: Speech, Emotion, SER, Spectrogram, Prosodic.