Cyprus University of Technology
Department of Electrical Engineering, and Computer Science and Engineering

Bachelor Thesis

Speech Emotion Recognition using Wav2Vec2 and LoRA

Andreas Shimitras

Limassol, May 2026

CYPRUS UNIVERSITY OF TECHNOLOGY

Faculty of Engineering and Technology

Department of Computer Science and Engineering

Bachelor Thesis

Speech Emotion Recognition using Wav2Vec2 and LoRA

Andreas Shimitras

Advisor: Dr Sotirios Chatzis

Limassol, May 2026

Copyrights

Copyright © 2026 Andreas Shimitras

All rights reserved.

The approval of the dissertation by the Department of Computer Science and Engineering, does not necessarily imply the approval by the Department of the views of the writer.

Acknowledgements

I would like to express my sincere gratitude to my advisor, Dr Sotirios Chatzis, for his guidance, support, and valuable feedback throughout the development of this thesis.

I would also like to thank the Department of Electrical Engineering, and Computer Science and Engineering of the Cyprus University of Technology for providing the academic environment and resources needed to complete this work.

ABSTRACT

Speech Emotion Recognition is an important research area that focuses on identifying the emotional state of a speaker from speech audio. Emotion in speech appears through acoustic cues such as pitch, intensity, rhythm, articulation, and vocal tone. However, automatic emotion recognition remains challenging because emotional expression varies between speakers, datasets, and recording conditions.

This thesis investigates the use of a pretrained Wav2Vec2 speech model for audio-based Speech Emotion Recognition. The proposed system classifies speech recordings into six emotional categories: angry, disgust, fearful, happy, neutral, and sad. The experiments use three publicly available emotional speech datasets, RAVDESS, SAVEE, and TESS, which are harmonized into a common six-class label structure. Audio preprocessing standardizes the recordings to a 16 kHz sampling rate and prepares them as mono waveform inputs.

The model uses Wav2Vec2ForSequenceClassification as the main architecture and applies Low-Rank Adaptation for parameter-efficient fine-tuning. This approach allows the model to adapt to the emotion recognition task while reducing the number of trainable parameters compared with full fine-tuning. The final model achieves approximately 70% to 71% accuracy and 68% to 69% macro F1-score on the test set. The results show stronger performance on fearful, neutral, and sad speech, while some confusion remains between acoustically similar emotions such as angry, disgust, and happy. Additional dataset-level analysis shows that many angry-class errors are concentrated in the TESS subset, highlighting the importance of per-dataset evaluation in Speech Emotion Recognition.

Overall, this thesis demonstrates that Wav2Vec2 combined with LoRA provides a practical and reproducible approach for speech-based emotion classification, while also revealing the limitations caused by dataset variation and acoustic overlap between emotion classes.

Keywords: Speech Emotion Recognition, Wav2Vec2, LoRA, Deep Learning, Audio Classification