
Project

Faculty of Engineering,
Computer Science and Psychology –
Dialogue Systems

dialogue-systems.org

07.03.2025

Leveraging transformers for speech emotion recognition in remote interactions

DESCRIPTION:

Speech is a fundamental mode of human communication, conveying not only linguistic content but also emotional and affective states. In remote interactions, where visual and contextual cues may be limited, speech emotion recognition (SER) plays a crucial role in enhancing user experience, improving human-computer interaction, and supporting affective computing applications. Traditional SER approaches rely on handcrafted features and conventional machine learning models, which often struggle to generalize across diverse speakers and environments. Recent advancements in transformer-based models have shown great potential in capturing complex temporal dependencies and contextual nuances in speech data.

This project aims to leverage transformers for speech emotion recognition using the RECOLA database, which contains multimodal recordings of spontaneous interactions. The objective is to preprocess the speech data, implement transformer models, and evaluate their performance against traditional baselines. The study will focus on optimizing model architectures, handling variability in emotional expression, and analyzing the impact of transformer-based representations on emotion classification accuracy.

The project comprises the following tasks:

- Preprocessing RECOLA speech data for using in Machine Learning
- Implementing transformer-based models for SER
- Comparing performance with traditional machine learning approaches
- Evaluating model robustness and generalization

PREREQUISITES:

- Basic knowledge in machine learning
- Python & PyTorch skills would be desirable, but not mandatory

CONTACT:

M.Sc. Jonas Jostschulte
Institute of Communications Engineering – Dialogue Systems
Room: 43.2.225 (Uni West)
Tel: 0731/50-31528
E-Mail: jonas.jostschulte@uni-ulm.de

