



Emotions play a crucial role in human-human communication with a complex socio-psychological nature, making emotion recognition a challenging task. In this dissertation, we study emotion recognition from audio and visual cues in video clips, utilizing facial expressions and speech signals, which are among the most prominent emotional expression channels. We propose novel computational methods to capture the complementary information provided by audio-visual cues for enhanced emotion recognition.

The research in this dissertation shows how emotion recognition depends on emotion annotation, the perceived modalities, modalities' robust data representations, and computational modeling. It presents progressive fusion techniques for audio-visual representations that are essential to improve their performance. Furthermore, the methods aim at exploiting the temporal dynamics of audio-visual cues and detect the informative time segments from both modalities. The dissertation presents meta-analysis studies and extensive evaluations for multimodal and temporal emotion recognition.

ESAM A.H. GHALEB
BIMODAL EMOTION
RECOGNITION THROUGH
AUDIO-VISUAL CUES

BIMODAL EMOTION RECOGNITION THROUGH AUDIO-VISUAL CUES

ESAM A.H. GHALEB

