

# Abstract

Affectiva AFFDEX, Emotient FACET, openSmile and openVokaturi are four emotion recognition systems with different architectures and trained on different training datasets. The training datasets of these four systems contain only a small amount of Asian data, and there are no reports about the performance of the systems on Asian data. In this study, 45 speech videos and audios of Chinese politicians and 212 speech videos and audios of politicians from other countries were collected and processed using the four systems. In addition, two video datasets, RAVDESS and CASME II, and one audio dataset, EMO-DB, with emotion labels were used as baselines in this study. In the dataset of Chinese politicians, the speech texts were transcribed by using Amazon Transcribe. The emotions in the texts were identified through textual sentiment analysis with two Chinese sentiment lexicons, HowNet and National Taiwan University Sentiment Dictionary (NTUSD). The consistency and difference of the emotion recognition results were tested by using statistical methods including the Mann-Whitney U test, Spearman correlation coefficient, McNemar test, Cohen's Kappa, etc. Ultimately, it was found that the consistency of the results from the systems on Chinese data was weaker than that on data from other countries. The consistency among the results of facial emotion recognition, speech emotion recognition, and textual sentiment analysis was very weak or even inconsistent. Finally, this study also validated the correlation between emotion and action unit (AU) and acoustic features using the Spearman correlation coefficient. The results of the experiment revealed correlations between Joy and AU12, correlations between Anger and AU6, AU4, AU23 and AU24, and differences in fundamental frequency (F0) in different languages and different genders, which validated previous findings. Ultimately, future work is proposed, including using approaches such as machine learning to improve the accuracy of textual sentiment analysis and implementing the fusion of different emotion analysis results based on fuzzy logic to improve the accuracy of emotion recognition.