

Accuracy of automatic emotion recognition from voice

Damien Dupré^{*,a}, Gary McKeown^b

^a*The Insight Centre for Data Analytics, University College Dublin, Republic of Ireland*

^b*School of Psychology, Queen's University Belfast, United Kingdom*

Abstract

Voice is one of the main communicative sources of evidence in interpreting the expression of emotion (Poria, Cambria, Bajpai, & Hussain, 2017). Affective computing aims to create systems and algorithms that automatically analyse people's emotional state. Consequently, several companies such as Affectiva, Beyond Verbal and Audeering have developed automatic systems to analyse the vocal expression of emotions. However, little is known about the accuracy of such systems.

To evaluate the accuracy of automatic emotion recognition from voice, we processed vocal expressions from the GEMEP database with "SensAI Emotion" developed by Audeering. The GEMEP database contains audio-video recordings from 10 actors performing 17 different emotional scenarios in which they vocalize the same non-words with an emotional tonality (Bänziger & Scherer, 2010). SensAI Emotion analyses emotions from speech and renders a value for 23 affective states and for valence and arousal dimensions (Eyben, Scherer, & Schuller, 2018).

In terms of category recognition, the accuracy of SensAI labelling GEMEP vocal expressions of emotion is 6.67%. However, this low result is partly due to the high number of different affective state labels recognized. To bypass this label matching bias, we compared the recognition accuracy for valence and arousal dimensions. The results show an accuracy of 0.56 (CI95%[0.46,0.65]) for valence and 0.73 (CI95%[0.64,0.81]) for arousal recognition.

Vocal automatic emotion recognition is a growing research area in affective computing. The categorical recognition of emotion remains a challenge due to the diversity of affective states. However, the accuracy of a system like SensAI Emotion provides promising results in the recognition of valence and arousal.

References

Bänziger, T., & Scherer, K. R. (2010). Introducing the geneva multimodal emotion portrayal (gemep) corpus. *Blueprint for Affective Computing: A Source-*

*Corresponding Author

Email addresses: damien.dupre@ucd.ie (Damien Dupré), g.mckeown@qub.ac.uk (Gary McKeown)

book, 271–294.

Eyben, F., Scherer, K. R., & Schuller, B. W. (2018). Method for automatic affective state inference and an automated affective state inference system. Google Patents.

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, 37, 98–125.