Comparing Human And Machine Emotion Recognition In Mimicked And Suppressed Facial Expression Conditions

Liina Juuse, Kristian Pentus, Kairi Kreegipuu, Jüri Allik
Institute of Psychology, University of Tartu, ESTONIA
Contact: *liina.juuse@ut.ee*

Introduction

- Increasing integration of machine learning (ML) into affective science raises critical questions about the relative **sensitivity** and **accuracy** of human versus machine emotion recognition
- Automated systems (e.g., FaceReader, DeepFace) are increasingly used in affective science and industry[1,2], yet their validity under non-spontaneous or contextually complex emotional displays is uncertain
- Human observers rely on context, dynamics, and social priors, whereas ML models are trained on posed, prototypical expressions → limits generalizability.
- How sensitive and accurate are humans vs. machines in decoding **mimicked** or **suppressed** facial expressions?

Methods

Sample 1:

- Participants (N = 119; 45 men; *M* = 25.02; *SD* = 6.34)
- Multimodal recorded data: facial recordings, EEG, skin conductance, self-report, personality (EPIP–NEO[7]), and affect (PANAS[8])

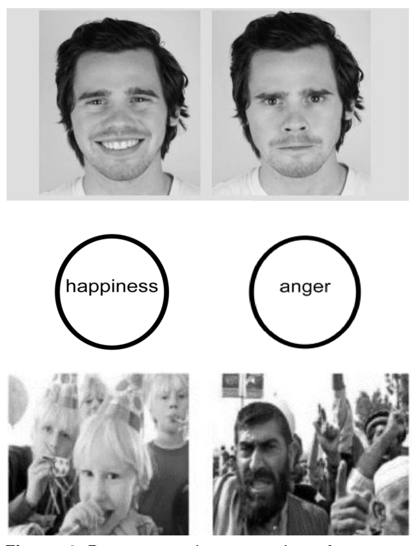

Figure 1. Representative examples of emotion stimuli by stimulus tyoe (faces, words, images). Note: Illustrative only; actual experimental stimuli not shown.

Sample 2:

- Observers (N = 223; 45 men; *M* = 31.86; *SD* =14.13) evaluated emotional expressions
- Measures: accuracy of emotion recognition (0/1, %); empathy scores[9]; professional skill mapping; PANAS[8] affective scores

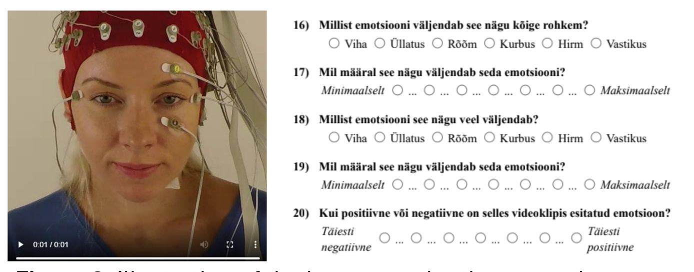

Figure 2. Illustration of the human evaluation procedure: facial expression recordings presented to observers (left), and corresponding emotion rating scales (right).

Sample 3:

- Machine learning & automated analysis of emotional facial expressions model)

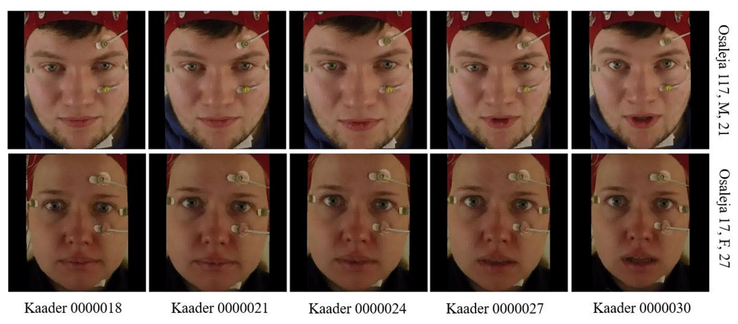

Figure 3. Illustrative examples of facial expression frames analyzed by automated methods (FaceReader, DeepFace).

FaceReader (rule-based facial expression analysis software)

DeepFace (deep-learning–based face recognition/emotion analysis

- Significant Emotion x Target x Source x Condition interaction: $F_{(10,6995)} = 19.58$; $p < .001$; $\eta^2_p = 0.027$.

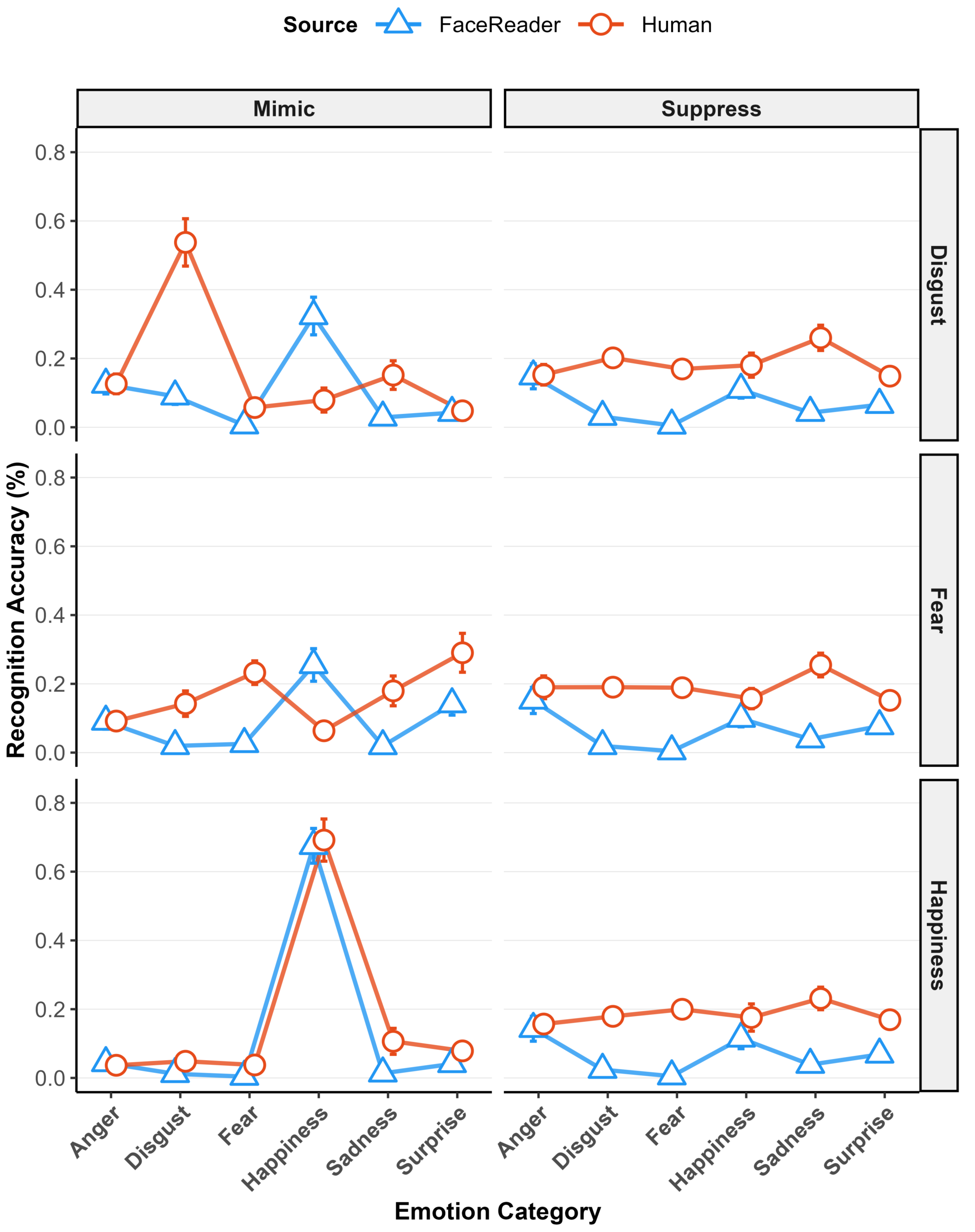

Figure 4. Recognition accuracy for each emotion category, separated by expression condition (*Disgust*, *Fear*, *Happiness*). Human observers (red circles) and FaceReader (blue triangles) are compared. Error bars represent 95% confidence intervals.

Results

Main results so far:

Mimic condition:

- Humans: 54.4% overall accuracy; strongest for happiness (82.7%), weakest for fear (19.2%)
- FaceReader: 37.4% overall accuracy; strongest for happiness (96.9%), weakest for fear (4%)

Suppress condition:

- Humans: 18.7% overall accuracy (~chance level); the same for happiness (18.5%), disgust (18.6%) and fear (18.2%)
- FaceReader: 6.6% overall accuracy; the same for happiness (6.6%), disgust (6.7%) and fear (6.5%)
- Humans significantly outperformed FaceReader for recognition of mimicked fear and disgust, but not for happiness
- Both human evaluation and automated detection poor for fear/surprise confusions and for mimic condition
- Further analyses are needed to evaluate whether different algorithmic approaches (e.g., deep learning vs. rule-based models) yield improved performance on the same stimulus material

Conclusion

FaceReader failed to generalize to non-spontaneous, strategically modulated facial expressions (with the exception of happiness, though with caveats)

Both humans and AI struggled with suppressed and mimicked expressions (though humans were significantly better in the mimicked condition)

Findings underscore the systematic confusion of emotions in humans and the need for context-sensitive automated models

ADDITIONAL INFO

References
[1] Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automatic facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics, 7*(4), 227–236. DOI:10.1037/npe0000028
[2] Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). AffectNet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing, 10*(1), 18–31. DOI:10.1109/TAFFC.2017.2740923
[3] Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science, 20*(5), 286–290. DOI:10.1177/0963721411422522
[4] Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science, 338*(6111), 1225–1229. DOI:10.1126/science.1224313
[5] Ekman, P., & Friesen, W. V. (1975). Unmasking the face: A guide to recognizing emotions from facial clues. Englewood Cliffs, NJ: Prentice-Hall.
[6] Porter, S., & ten Brinke, L. (2008). Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. *Psychological Science, 19*(5), 508–514. https://doi.org/10.1111/j.1467-9280.2008.02116.x
[7] Mõttus, R., Pullmann, H., & Allik, J. (2006). Toward more readable Big Five personality inventories. *European Journal of Psychological Assessment, 22*(3), 149-157. DOI:10.1027/1015-5759.22.3.149
[8] Allik, J.; Realo, A. (1997). Emotional experience and its relation to the Five-Factor Model in Estonian. *Journal of Personality, 65,* 625–647. DOI: 10.1111/j.1467-6494.1997.tb00329.x
[9] Realo, A., Allik, J., Nõlvak, A., Valk, R., Ruus, T., Schmidt, M., & Eilola, T. (2003). Mind-reading ability: Beliefs and performance. *Journal of Research in Personality, 37*(5), 420–445. DOI:10.1016/S0092-6566(03)00021-7

The research was funded by the Estonian Research Council grant PRG11511, TK218U2, and approved by the Research Ethics Committee of the University of Tartu.