Dynamic Analysis of Automatic Emotion Recognition Using Generalized Additive Mixed Models

Damien Dupré[1], Adam Booth[2], Andrew Bolster[2], Gawain Morrison[2] and Gary McKeown[1]

Abstract. The increasing ease of access to large participant populations via online recruitment platforms offers considerable promise to study aspects of emotion and participants' relationship to brands, artistic media and advertising content. Large participant samples can be quickly and easily gathered, and their facial expressions can be assessed using video captured by webcams that are ubiquitous in modern computing environments. However, a long-standing issue is the degree of reliability of the data provided by automatic facial expression recognition systems. We evaluated the facial expressions of 836 participants triggered by an online sad video. Thanks to the use of this sizeable sample, this experiment has a high statistical power and showed, as expected, that Sadness expressions are significantly more frequently recognized compared to other facial expressions (i.e. Happiness, Surprise and Disgust). However, by using Generalized Additive Mixed Models (GAMM) to analyse the time-series we show that dynamic statistical analyses can avoid biased interpretations. Technical sensitivity, posture interpretation, emotion idiosyncrasy, intensity of spontaneous expressions and social context of the emotion elicitation are open questions that must be answered to perform reliable online facial expression recognition.

Keywords. Emotion, Facial Expression, Automatic Recognition, Time-series analysis.

1 INTRODUCTION

One of the enduring debates within the psychology of human emotion concerns the extent to which facial expressions represent an automatic readout of felt emotions or social and communicative motivations [1], [2]. This debate has continued to dominate discourse in the field of emotion to a large extent over the past two decades. However, within the world of Affective Computing it is largely the former view that emotions are reflections or indices of felt emotions that has prevailed. The theoretical importance of these two viewpoints becomes of crucial importance when moving from the theoretical world into the practical uses and applications of facial expressions in technologies that are oriented towards gathering information concerning one's emotional state.

There are many commercial applications where an understanding of the emotional impact on individuals is important. These range from an understanding of the emotions of interlocutors when they interact with embodied conversational agents to the desire to be aware of the emotional impact that a given piece of artistic content or advertising may have upon individuals in the worlds of film-making and marketing. The former of these examples actively seeks to be social and communicative in nature, whereas the latter is much more passive. It is not the case that viewing artistic or marketing content is entirely non-social – typically the content contains a narrative that is highly social in nature, which is what typically grabs our attention and engages us. However, the passive nature of this content means it may be better described as implicitly [1], rather than explicitly, social. Importantly, any person watching this type of content on a screen is not actively involved in the social interaction, and at best can be considered a distant observer. They are also conscious of the real disconnect of their current situation from the social events occurring in the on-screen drama and narrative; the most compelling of artistic productions of this kind may go some way to allowing an observer to suspend their disbelief and become immersively engaged in the content, particularly in an emotional manner. Indeed, this is probably the primary goal of many creative artists and content creators.

An opportunity for the fast and efficient emotion assessment for these kinds of commercial materials exists through the utilisation of crowdsourced recruitment of participants and testing them from a distance. Several crowdsourcing services are currently available such as Amazon's Mechanical Turk, Cint, Crowdflower or Prolific Academic [3], [4]. These platforms will recruit large panels of participants quickly and efficiently. However, whilst the benefits of such panels are considerable for the fast and efficient conduct of research, they raise an important question: do participants express strong facial expressions when they are alone in front of their computer? This question is particularly important in terms of emotion studies. Indeed, the recording of facial expressions from modern computers is relatively easy, while the implementation of automatic emotion recognition systems on web-based application is becoming increasingly common. However, aside from the improvements in these systems, it is important to critically examine the meaning of the data provided.

This paper addresses the question of how to interpret automatic emotion recognition results. It presents an online crowdsourced experiment in which a primarily sad piece of emotional content is presented to a large participant sample. Facial expression analysis is conducted to evaluate the emotional response using both an overall and a time-series perspective.

2 EMOTIONS TRIGGERED BY MEDIA

There remains considerable contention concerning theoretical conceptions of emotions [5], as well as the functions of facial expressions and their relation to felt emotion [1], [2]. Despite this, however, a consistent model of emotions exists that considers them as being the result of internal and external spontaneous reactions initiated by an event [6], [7]. These

[1] School of Psychology, Queen's Univ. Belfast, BT7 1NN, UK. Email: {Damien.Dupre, G.McKeown}@qub.ac.uk.
[2] SensumCo. Ldt., BT9 6RY, UK. Email: {Adam,Bolster, Gawain}@sensum.co.

appraisal models are well-suited as a theoretical basis for the current study, as they place an emphasis on emotions as they relate to a specific event and context. The event refers here to the media and the narrative content within that media which lead to the triggering of specific emotional episodes.

An extensive literature review was conducted to investigate the influence of media such as movies or TV commercial on emotions. In that perspective, several video databases were created and validated for use in an emotion elicitation task. The most well-known one is composed of 16 videos able to elicit Amusement, Anger, Disgust, Fear and Sadness [8], [9] or [10], for an extensive review. The videos have been shown to have an impact on viewers' self-reported emotions, physiological changes and facial expressions.

2.1 Crowdsourced systems for automated facial expressions recognition

With the introduction of software development kit (or SDK) for automatic facial expressions recognition, it is possible to embed these libraries online for use on crowdsourcing platforms by using an application programming interface. Major companies have their own SDKs such as Apple (using Emotient SDK), Facebook (using FacioMetrics SDK), Microsoft (Project Oxford), Qualtrics (using Affectiva SDK) or RealEyes. Moreover, some companies directly provide their SDK. This is the case for Face++, Kairos, CrowdEmotion and Visage Technologies to name a few.

Even if some systems focus on the dimensional aspect of emotions [11], most of these automatic facial expression recognition systems are based on the Facial Action Coding System (FACS [12], [13]), in which facial movements are categorised according to sets of muscle actions termed Action Units (AU). The FACS method is agnostic to any relationship with felt emotions and only seeks to provide an objective classification of the observed facial muscle movement (Figure 1).

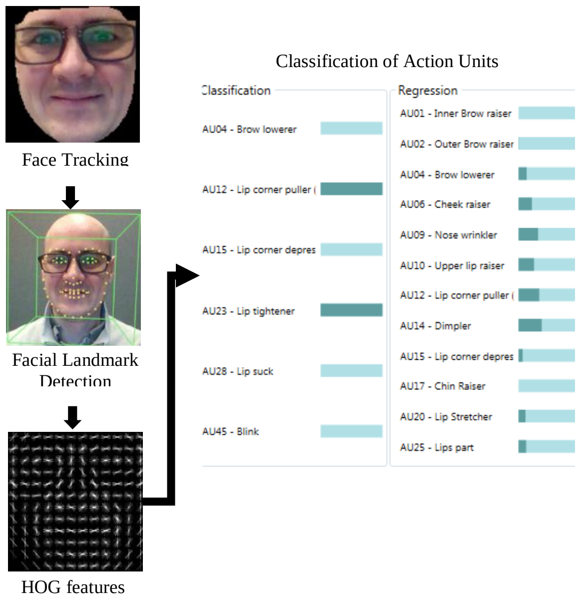

Figure 1. Example of automated FACS coding analysis using facial landmark detection with Histograms of Oriented Gradients (HOG) features (using OpenFace v0.1 [14]).

Related systems, such as EmFACS [15], exist which seek to associate these muscle movements with the underlying emotional states typically associated with a given expression. Therefore, most studies taking an affective computing perspective, as well as many within scientific psychology settings, have used facial expression analysis to detect happiness, surprise, sadness, anger, disgust or fear, with the assumption that expressions provide facial readouts of felt emotions derived from the basic emotions approach. Based on EmFACS recognition, significant efforts have been conducted to create accurate facial expression analyses of natural emotions [16]. For example, challenges such as Facial Expression Recognition and Analysis challenge (FERA) [17], [18] or Audio/Visual Emotion Challenge (AVEC) [19] aim to develop the most efficient algorithms to recognise emotions both in laboratory settings and "in the wild".

2.2 Sensum Insights: A web based platform for media content analysis

Sensum Insights is a multi-modal research tool that allows the collection of responses to events, media and situations by measuring, aggregating and integrating a variety of data streams collected from participants in either remote or laboratory conditions. The platform therefore offers utility in media testing, the monitoring of audience reaction to live events, and the analysis of large data sets related to performances and content. It is particularly well suited to the gathering of affective and emotion related information. The multi-modal nature of the platform means it can avail of and integrate social signal measurements such as: facial expressions and acoustic and prosodic cues of emotions; biometric signals such as heart rate, breath rate, eye-tracking and galvanic skin response; and more classical psychological measurement techniques such as implicit association tests and survey responses.

In the current study, the FacioMetrics system [20] was incorporated into Sensum Insights to provide real-time facial coding. The FacioMetrics system detects and automatically tracks 64 face landmarks for each frame recorded. Then, the movement of each landmark is analysed by Scale-Invariant Feature Transform (SIFT) descriptors in localized regions [21]. Using support vector machine (SVM) classifiers to associate landmark movements with AUs, the FacioMetrics system can extract facial expression information and categorise 4 emotion labels: Disgust, Happy, Sadness, and Surprise. It can also categorise more functional labels such as Focus and Attention. The FacioMetrics system uses person-specific classifiers for FACS recognition and was trained on posed and on spontaneous facial expressions (CK+, GEMEP-FERA, RU-FACS and GFT databases) [22].

3 METHOD

To test the hypotheses, the Sensum Insights platform was used to create an experimental design which allowed the evaluation of the influence of brand logo, definition quality and event video streaming quality. The dependent measure was provided by the classification of the participants' facial expression of emotions.

The web-based experiment presented either a high or low quality viewing experience whilst using Facial Coding to assess the experience.

Online participants viewed the content on their PC or Laptop. They were instructed to watch the content in a room with correct lighting and viewing conditions.

3.1 Participants

To test their facial expressions, 836 participants (442 males and 394 females, age M = 36.4, SD = 11.5) were recruited to take part online around the UK. Those who participated under insufficient lighting conditions were removed from the final data pool, leaving a final total of 774.

3.2 Material

An evocative short film that had been shown in previous pilot studies to produce feelings of sadness was used. Each participant was exposed to a short sci-fi commons piece which lasted for 1 minute and 30 seconds, called "Tears of Steel"[3]. It comprised of some action and drama elements. It typically creates a sad but bittersweet emotional connection with the viewer.

3.3 Protocol

Respondents were required to set up their viewing space with correct lighting before passing a calibration test. This was to ensure a good quality data set.

4 RESULTS

The analysis of the data obtained was performed in two steps. Firstly, we examined the overall automatic recognition of facial expression in order to identify which emotion was triggered. Secondly, we looked at the evolution of the automatic recognition during the video. By using Generalized Additive Mixed Models (GAMMs), we were then able to describe the events present in the time-series [23]. The GAMMs analysis appeared to be particularly relevant for emotion time-series [24]. GAMMs allow a researcher to spline multi data streams on a reference index, in order to identify the curve with the maximum likelihood estimation, whilst taking into account the autocorrelation of these data streams.

General facial expression recognition. Firstly, it was necessary to evaluate the overall recognition of the different facial expressions recognised by the FacioMetrics system.

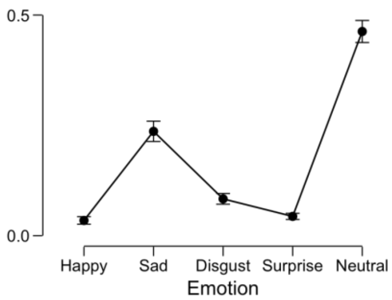

Figure 2. Means of the overall facial expression recognition for each emotion label for the whole sample studied. Error bars represent 95% confidence intervals. The FacioMetrics recognition score goes from 0 (no recognition at all) to 1 (full prototypical expression displayed all along the video).

[3] https://www.youtube.com/watch?v=bjYbA1bWjeE

As shown in Figure 2, the participants displayed mostly neutral expressions. However, we also observed a homogeneous dispersion of the mean which indicated no specific trends among the participants (M = 0.44, SD = 0.34). We also observed similar representation of Disgust (M = 0.08, SD = 0.15), Happy (M = 0.03, SD = 0.09) and Surprise (M = 0.02, SD = 0.07). Overall, they were weakly recognised, which suggested that their natural expression was brief and subtle. It was therefore expected that we would observe low levels of the appearance of Action Units and facial muscle movements associated with these emotions when watching the "Tears of Steel" film. The results were somewhat different, however, when observing the level of facial muscle movement and expressions associated with sadness. We saw a broader distribution of the participants compared to the other emotions. However, there were still relatively low absolute levels of the sadness facial expression (M = 0.24, SD = 0.30). Given the broad distribution of the scores in this measure, it was also prudent to pay attention to the median as a measure of central tendency. Sadness labelled facial expressions had a median score of 0.08, which was much weaker. Finally, it is interesting to note that means and distribution of Attention and Focus are very different. Due to their involvement in the task, the FacioMetrics system recorded mostly participants' faces displaying Attention (M = 0.79, SD = 0.31). Contrary to Attention, the FacioMetrics system did not record high levels of faces displaying Focus (M = 0.07, SD = 0.06).

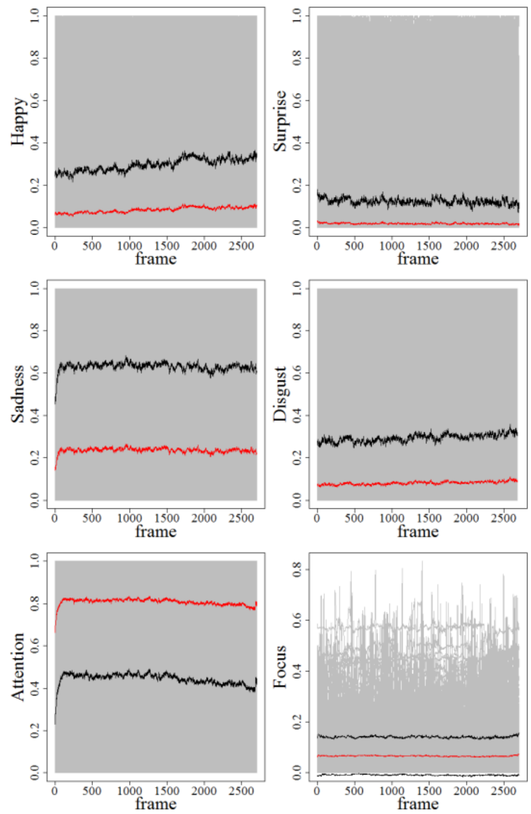

Figure 3. Temporal evolution of the facial expression according the time (frame) for each label. The grey lines represent each participant, the red lines are the mean and the black lines are the standard deviation.

When we assessed the ability to distinguish between emotions using a repeated measures ANOVA (with Greenhouse-Giesser correction), we found that it was possible to distinguish emotional classifications, $F_{(1.89,\ 1461.8)} = 445.8$, $p < 0.001$, $\eta^2 = 0.37$. Then, given the overall analysis of the facial expression, we concluded that there was evidence of efficient triggering of emotion by the video. Whilst the Neutral recognition was very high, this was not particularly surprising, as it was expected that Sadness would not be triggered all along the video, but rather only at certain key moments. Therefore, it was considered particularly important to examine the time-series evolution in order to draw conclusions on the identity of the emotion triggered and at what specific points in the video.

Time-series analysis of facial expression recognition. In addition to the general analysis of facial expression recognition, a first step was to observe the descriptive statistics of the results. Therefore, we plotted the temporal evolution of the facial expression recognition according to the time for each label recognised (Figure 3).

Surprisingly, all of the label recognitions displayed flat evolution according to the time for both mean and standard deviation, whereas the dispersion of individual time-series was highly heterogeneous. In order to evaluate the influence of the label recognition over the time, we performed a Generalized Additive Mixed Model analysis of the time series using restricted maximum likelihood estimation (REML) [25]. The following model was tested:

$$y = X\beta + Zu + \varepsilon$$

$$u \sim N(0,\Psi_\theta)$$

$$\varepsilon \sim N(0,\Lambda\sigma^2)$$

where y is the emotion recognition vector, X is the model design matrix according to the time (or Frames in our case), β is the β coefficient vector. Moreover, u contains a random effects vector, and Z is a model matrix for these random effects (for each participant), Ψ is the covariance matrix, and θ the unknown parameters within that covariance matrix. Λ is a matrix that is part of the error term and which can be used to model the residual autocorrelation. Finally, ε is the error term.

When examined the residual distribution for the overall model, we observed that the residuals were not normally distributed (Figure 4).

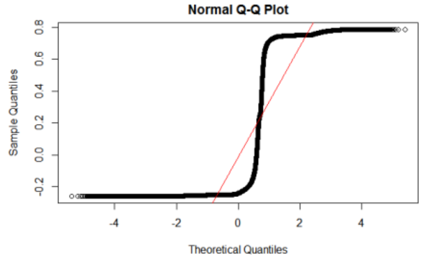

Figure 4. Residual distribution of the GAMM analysis using a Q-Q plot.

This distribution showed the limit of using parametric overall statistical analyses based on the normality distribution of the residual such as the t-test or F-test. Then, by looking at the trends of the differents emotion labels, we observed that the maximum likelihood estimation of the GAMMs revealed evolutions in the facial expression recognition (Figure 5).

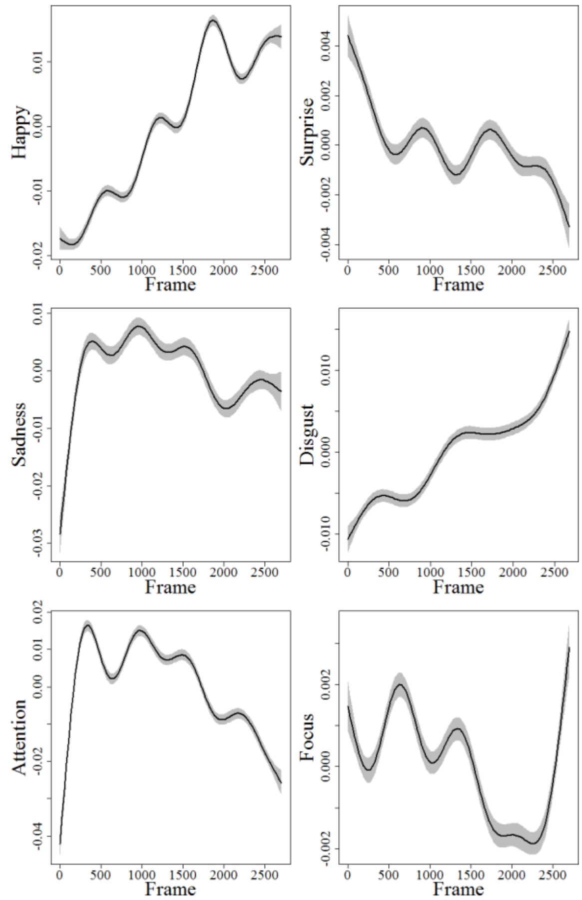

Figure 5. Evolution of the fitted values identified by the GAMM analysis for each emotion label over the video.

Even if the effect size of the trends was small, using such a sizeable participant sample provided enough statistical power for the GAMMs to be significant. Consequently, given the results provided by the GAMM analysis over the video, we concluded that the result provided by the overall descriptive analysis was not too limited to describe the automatic facial expression recognition signal. Therefore, the GAMM analysis appears to provide a highly promising methodological framework to investigate temporal evolutions of emotional facial expressions.

5 DISCUSSION & PERSPECTIVES

In this paper we analysed the emotions triggered by an online media with an automatic facial expression recognition system. The objective of the study was to accurately evaluate the emotion expressed by a substantial participant sample, which was recruited online. Even if the general analysis showed a significant expression of sadness according to the participants, the analysis of the time-series revealed a more complex pattern of facial expression evolution, indicating the necessity to take into account the temporal evolution of facial expressions.

5.1 Weaknesses of emotion recognition systems

There is a current notable boom in the opportunities provided by emotion recognition systems. Emotional expressions become a gold-standard approach in terms of analysing the ways in which humans are interacting with computers (HCI) or with robots (HRI). However, emotion recognition systems are still dealing with recognition problems in the wild. Firstly, face-tracking techniques still suffer from poor recognition performance for non-frontal faces. There also exist problems accompanying some face characteristics such as wrinkles, beards or glasses, for example. Secondly, emotion recognition systems can be biased by face orientation. If the face is oriented down, this can be mistaken as a sadness expression. This probably represents a bias towards sadness recognition becomes so important when people are recorded by laptop webcams. In terms of signal treatment, the results show a false positive recognition of sadness, which may also explain why happiness and disgust levels appear to increase during a sad video.

5.2 The intensity of facial expression in implicitly social settings

We obtained evidence that the strongest facial expression configurations throughout the video clip were those with the label Neutral, which is in line with our prediction that this would be the most prevalent expression. We also found that there were greater levels of expressions with the label Sadness reported compared to any other facial expression. This would suggest that the emotion elicited during the video clip was recognised as sadness, and was discernible from the other emotion labels. However, given the dispersion of the scores, a median measure presented a much lower level of emotion label. This would suggest that only some of the participants were producing expressions that were related to sadness, or that the system may recognize sadness in the participant's expression due to their position in front of the webcam. Indeed, most of the participants were using laptops with an embedded webcam. Therefore, the overall position of the face was not effectively placed directly in front of the camera, but rather at a slight incline in order to see the screen. A future line of enquiry may therefore include the examination of individual differences in response to implicitly social stimuli.

5.3 Idiosyncrasy of facial expressions

The variability in the time-series recorded is an important problem for further online media evaluations, as it can lead to misleading analysis of triggered emotions. A first explanation comes from the technical characteristic of automatic recognition systems. Indeed, because they have a frame-to-frame analysis, their sensibility to micro-change in the face is particularly high [26], [27]. Thus, the automatic system can give a meaning to movement which may have no meaning.

Moreover, human emotions are idiosyncratic. There exists high variability in terms of how people react to a particular stimulus. The dispersion of the individual time-series demonstrates that it is very difficult to predict which expression people will spontaneously produce in a very large sample. Some people can be more expressive of certain emotions at baseline. Expressiveness can also vary according to personality and culture [28]. However, by using GAMM analysis it is possible to take into account the idiosyncrasy of how people express emotions at the individual level [29].

These sources of variability in the recognition signal are particularly important to consider, because they can lead to misunderstanding of the emotion triggered by a media source. Therefore, new statistical tools, such as the Generalized Additive Mixed Model, should be utilised in order to reduce the levels of false emotion detection resulting from the use of less sensitive and more general statistical analyses.

5.4 The type of facial expressions

Most of the emotional expressions that occur in realistic interpersonal interactions or in HCI settings are not basic emotions [30]. Studies on everyday facial expressions have begun to increasingly examine mental states [31] such as interest, boredom, anxiety, and thoughtfulness, which are the kinds of expressions most often observed in face-to-face interactions and HCI. Often these natural affective states are very difficult to elicit in a laboratory setting, and it is difficult to know when they have occurred through traditional techniques of self-report [32]. However, the majority of affective computing works which examine facial expressions are based on a basic emotion paradigm, perhaps for historical reasons or due to the likelihood of acceptance [33]. Given the issues highlighted in this paper, including the problem that low levels of sociality may lead to low levels of expressive behaviour, it may be worthwhile taking a more nuanced approach in the understanding of facial expressions, and to target systems at a broader range of expressive behaviour labels such as amused, persuaded, informed, sentimental or inspired for example [34].

5.5 Using additional information as evidence to disambiguate facial expression meaning

Human communication, of course, involves more than the simple signalling of facial expressions. Other social signals, including vocal tone, gesture and the entirety of the body itself seems to operate as a vector of emotion. Often these can suffer similar issues of low levels of expression in situations of low sociality. The results of this experiment suggest that caution is required in relying too heavily on any given evidence stream with regard to the interpretation of the social signals. However, combining a variety of these social signals may provide opportunities for researchers to increase the weight of evidence available on which to judge. Combining multi-modal streams of data appears to be a worthwhile endeavour in disambiguating the emotional state of individuals. The study of the relationship between posture and emotion was explored through gestures emitted during conversation between two individuals [35]. Whereas the participants did not display obvious facial expressions, studying the influence of body posture can be a way to increase the interpretation of emotion and communicative motives [36]. Therefore, whilst posture seems to play a particularly important role in social interactions, its analysis is also complicated by issues of context of sociality. Therefore, there may be value in the use of other modalities and sources of evidence. Notably, biometric physiological signals do not suffer from issues of low sociality and should therefore respond in accordance with subjective experience whether there are others present or not [37]. Adding this source of evidence to multi-modal streams should be considered, in order to obtain a more well-rounded picture of the response to material presented to participants. This is becoming an increasingly more viable option with the availability of smartphones, smartwatches and wearable technologies. Whilst we are currently not at the levels of availability that make the online recruitment of large numbers of people with similar sets of technology a feasible option, the time is nearing.

6 REFERENCES

[1] A. J. Fridlund, *Human facial expression: An evolutionary view*. San Diego, CA: Academic Press, 1994.

[2] B. Parkinson, "Do facial movements express emotions or communicate motives?," *Personal. Soc. Psychol. Rev.*, vol. 9, no. 4, pp. 278–311, 2005.

[3] I. H. Gleibs, "Are all 'research fields' equal? Rethinking practice for the use of data from crowdsourcing market places," *Behav. Res. Methods*, pp. 1–10, 2016.

[4] E. Peer, L. Brandimarte, S. Samat, and A. Acquisti, "Beyond the Turk: Alternative platforms for crowdsourcing behavioral research," *J. Exp. Soc. Psychol.*, vol. 70, pp. 153–163, 2017.

[5] V. Shuman and K. Scherer, "Psychological structure of Emotions," *Int. Encycl. Soc. Behav. Sci.*, pp. 526–533, 2015.

[6] K. R. Scherer, "On the nature and function of emotion: A component process approach," in *Approach to emotion*, K. R. Scherer and P. Ekman, Eds. Hillsdale: Erlbaum, 1984, pp. 293–317.

[7] K. R. Scherer, "What are emotions? And how can they be measured?," *Soc. Sci. Inf.*, vol. 44, no. 4, pp. 695–729, 2005.

[8] J. J. Gross and R. W. Levenson, "Emotion elicitation using films," *Cogn. Emot.*, vol. 9, no. 1, pp. 87–108, 1995.

[9] J. Rottenberg, R. D. Ray, and J. J. Gross, "Emotion elicitation using films," in *Handbook of emotion elicitation and assessment*, J. A. Coan and J. J. B. Allen, Eds. London: Oxford University Press, 2007, pp. 9–28.

[10] T. L. Gilman *et al.*, "A film set for the elicitation of emotion in research: A comprehensive catalog derived from four decades of investigation," *Behav. Res. Methods*, pp. 1–22, 2016.

[11] C. Georgakis, Y. Panagakis, and M. Pantic, "Dynamic Behavior Analysis via Structured Rank Minimization," *Int. J. Comput. Vis.*, pp. 1–25, 2017.

[12] P. Ekman, W. Friesen, and J. Hager, *New Version of the Facial Action Coding System*, A Human Face Publication. Salt Lake City, USA, 2002.

[13] P. Ekman and W. V. Friesen, *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto, USA: Consulting Psychologists Press, 1978.

[14] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Openface: an open source facial behavior analysis toolkit," presented at the Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on, 2016, pp. 1–10.

[15] W. V. Friesen and P. Ekman, "EMFACS-7: Emotional facial action coding system," University of California at San Francisco, 1983.

[16] S. Zafeiriou, A. Papaioannou, I. Kotsia, M. Nicolaou, and G. Zhao, "Facial Affect 'in-the-wild': A survey and a new database," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016, pp. 36–47.

[17] M. F. Valstar *et al.*, "Fera 2015-second facial expression recognition and analysis challenge," presented at the International Conference and Workshops on Automatic Face and Gesture Recognition, 2015, vol. 6, pp. 1–8.

[18] M. F. Valstar *et al.*, "FERA 2017-Addressing Head Pose in the Third Facial Expression Recognition and Analysis Challenge," *ArXiv Prepr. ArXiv170204174*, 2017.

[19] B. Schuller, M. Valstar, F. Eyben, G. McKeown, R. Cowie, and M. Pantic, "Avec 2011–the first international audio/visual emotion challenge," presented at the International Conference on Affective Computing and Intelligent Interaction, 2011, pp. 415–424.

[20] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," presented at the Conference on Computer Vision and Pattern Recognition, 2013, pp. 532–539.

[21] J. M. Girard, J. F. Cohn, L. A. Jeni, M. A. Sayette, and F. De la Torre, "Spontaneous facial expression in unscripted social interactions can be measured automatically," *Behav. Res. Methods*, vol. 47, no. 4, pp. 1136–1147, 2015.

[22] W.-S. Chu, F. De la Torre, and J. Cohn, "Selective Transfer Machine for Personalized Facial Expression Analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 99, pp. 1–15, 2016.

[23] S. Wood, *Generalized additive models: an introduction with R*. CRC press, 2006.

[24] G. J. McKeown and I. Sneddon, "Modeling continuous self-report measures of perceived emotion using generalized additive mixed models.," *Psychol. Methods*, vol. 19, no. 1, p. 155, 2014.

[25] S. N. Wood, "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 73, no. 1, pp. 3–36, 2011.

[26] I. A. Essa and A. P. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 757–763, 1997.

[27] S. Namba, T. Kagamihara, M. Miyatani, and T. Nakao, "Spontaneous Facial Expressions Reveal New Action Units for the Sad Experiences," *J. Nonverbal Behav.*, pp. 1–18, 2017.

[28] J. A. Russell, "A Sceptical Look at Faces as Emotion Signals," *Expr. Emot. Philos. Psychol. Leg. Perspect.*, p. 157, 2016.

[29] I. Sneddon, G. McKeown, M. McRorie, and T. Vukicevic, "Cross-cultural patterns in dynamic ratings of positive and negative natural emotional behaviour," *PloS One*, vol. 6, no. 2, p. e14679, 2011.

[30] R. Cowie, E. Douglas-Cowie, and C. Cox, "Beyond emotion archetypes: Databases for emotion modelling using neural networks," *Neural Netw.*, vol. 18, no. 4, pp. 371–388, 2005.

[31] D. E. Back, T. R. Jordan, and S. M. Thomas, "The recognition of mental states from dynamic and static facial expressions," *Vis. Cogn.*, vol. 17, no. 8, pp. 1271–1286, 2009.

[32] H. G. Wallbott and K. R. Scherer, "Cues and channels in emotion recognition," *J. Pers. Soc. Psychol.*, vol. 51, no. 4, pp. 690–699, 1986.

[33] H. Gunes and H. Hung, "Is automatic facial expression recognition of emotions coming to a dead end? The rise of the new kids on the block," *Image Vis. Comput.*, pp. 6–8, 2016.

[34] D. McDuff, "Discovering facial expressions for states of amused, persuaded, informed, sentimental and inspired," presented at the Proceedings of the 18th ACM International Conference on Multimodal Interaction, 2016, pp. 71–75.

[35] M. Kipp, M. Neff, and I. Albrecht, "An annotation scheme for conversational gestures: How to economically capture timing and form," *Lang. Resour. Eval.*, vol. 41, no. 3–4, pp. 325–339, 2007.

[36] A. Kleinsmith and N. Bianchi-Berthouze, "Form as a cue in the automatic recognition of non-acted affective body expressions," in *Affective Computing and Intelligent Interaction*, 2011, pp. 155–164.

[37] N. Hamelin, O. El Moujahid, and P. Thaichon, "Emotion and advertising effectiveness: A novel facial expression analysis approach," *J. Retail. Consum. Serv.*, vol. 36, pp. 103–111, 2017.