

BEET: Behavior Engagement Emotion Trigger – A AI-based model to maximize audience engagement

Wael Badawy

Abstract— Engagement and measuring is a key performance for any content. The wide spread of the Internet and the adoption in our daily life, audience has many options to explore and audience developed the power of rejection. This power of rejection reduces the time of exposure and engagement to any contents including the in person engagement. The access to the online tools and portals provide an escape route from human-to-human interaction. A new tool to maximize audience engagement to a content that can influence a decision becomes a key performance and an asset for exposure. It enables a better delivery of the Contents, an increased interaction with the content, simplify the process of creating an engaging content, and maximize the exposure time. This paper presents a prototype, and evaluation of the performance of an automated Content Behavior Emotion Engagement Model “BEEM” that uses Deep Learning and Big Data analysis to discover the Behavior Emotion trend of the audience. BEETAIR is a new innovative framework that will transform the media market through Behavior Emotion Engagement Trigger Analyzer, and Intelligent Recommender. It uses The Artificial Intelligent-based Dialogue Generator for Maximum Audience Engagement.

Keywords—*BEET, Behavior Engagement Emotion Trigger, Communication, Content Engagement.*

I. INTRODUCTION

Measuring quality of content and its audience related engagement factor is a key performance of marketing, media and journaling. The effective content to impact audience has to increase its engagement level. Affective Computing (AC) has been a popular area of research for several years where machines detect and understand human affective states, such as emotions, interests and the behavior. It is assumed that to become more user-friendly and effective, systems need to become sensitive to human emotions. Nonverbal information is considered to complement the verbal message providing a better interpretation of the message. It is claimed that 70–90% of communication between humans is nonverbal. The studies conducted by Albert Mehrabian in 1967 established the 7%–38%–55% rule, also known as the “3V rule”: 7% of the communication is verbal, 38% of the communication is vocal and 55% of the communication is visual [1].

The first study in the field of emotion detection was born during the sixties/seventies. The most prominent example is that of mood rings [2]. The principle is simple; rings contain thermotropic liquid crystals that react with body temperature. When a person has stressed, his mood ring take on a darker color.

The scientific publications of Rosalind Picard (MIT) have introduced a great progress in this field since the nineties [3, 4]. He is one of the pioneers of affective computing. In his book “Affective Computing”, Picard proposed that emotion can be modeled using the nonlinear sigmoid function. Over the last 20 years, the development of technology has allowed the implementation of relatively good system market and efficient such as ambient intelligence (AMI), virtual reality (VR) and augmented reality (AR).

Nowadays, in the automotive field for example, an on-board computer that is able to detect confusion, interest or fatigue can increase safety. The AutoEmotive (MIT Media Lab) is a prototype equipped with sensors and a camera placed on the steering wheel [5]. This vehicle measures the level of stress and fatigue of the driver. When the need arises, he puts a background music, changes the temperature and light in the vehicle interior, or still proposes to follow a less stressful journey.

A multimodal system is widely adopted and several multimodal datasets include sentiment annotations. Zadeh et al. introduced the first multimodal dataset (MOSI) with opinion-level sentiment intensity annotations and studying the prototypical interaction patterns between facial gestures and spoken words when inferring sentiment intensity. A multimodal dictionary using language-gesture study is proposed in a speaker-independent model for sentiment intensity prediction [6]. For other examples of data sets we can cite ICT-MMMO [7] and MOUD [8] datasets. Intra-modality dynamics is modeled through three Modality Embedding Subnetworks, for language, visual and acoustic modalities, respectively [9]. LSTM-based network to extract contextual features from the video for multimodal sentiment analysis is shown in [10]. A multimodal sentiment analysis framework, which includes sets of relevant features for text and visual data, as well as a simple technique for fusing the features extracted from different modalities [11].

Multimodal emotion analysis has the following challenge: (1) model the interactions between language, visual and acoustic behaviors that change the observation of the expressed emotion (named the inter-modality dynamics). (2) Multimodal emotion analysis (named intra-modality dynamics) is to efficiently explore emotion, not only on one but also on highly expressive nature modality (ex.-spoken language where proper language structure is often ignored, video and acoustic modalities which are expressed through both space and time).

The emotion analysis lacks the ability to measure the engagement between the user and the content, the interaction with user to influence the user decision, and keep the user in front of the content. This paper presents a new model to measure the user behavior emotion trigger and measure the engagement level of the user. It also demonstrate a technique to personalized the content and introduce a metric to measure engagement. The reset of the paper is structure as follows: Section II presents a review of emotion and sentiment analysis. Section III presents the proposed model Section IV Experiment and performance evaluation. And finally Section V conclude this paper.

II. EMOTION AND SENTIMENT ANALYSIS

“Sentiment analysis is the field of study that analyses people’s opinions, sentiments, evaluations, appraisals, attitudes, and emotions toward entities such as products, services, organizations, and their attributes. It represents a large problem space. There are also many names and slightly different tasks, e.g., sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining, etc.” [12]

Sentiment Analysis (SA) [13] is a computational study of how opinions, attitudes, emoticons and perspectives are expressed in language. Sentiment Detection, or in its simplified form – Polarity Classification, is a tedious and complex task. Contextual changes of polarity indicating words, such as negation, sarcasm as well as weak syntactic structures make it troublesome for both machines and humans to safely determine polarity of messages.

Sentiment analysis methods involve building a system to collect and categorize opinions about a product. This consists in examining natural language conversations happening around a certain product for tracking the mood of the public. The analysis is performed on large collections of texts, including web pages, online news, Internet discussion groups, online reviews, web blogs, and social media. Opinion Mining aims to determine polarity and intensity of a given text, i.e., whether it is positive, negative, or neutral and to what extent. To classify the intensity of opinions, we can use methods introduced in [14, 15, 16, 17].

Text Mining and Social Network Analysis have become a necessity for analyzing not only information but also the connections across them. The main objective is to identify the necessary information as efficiently as possible, finding the relationships between available information by applying algorithmic, statistical, and data management methods on the knowledge. The automation of sentiment detection on these social networks has gained attention for various purposes [18, 19, 20, 21].

The aim of [22] was to report on the associations between depression severity and the variability (time-unstructured) and instability (time-structured) in emotion word expression on Facebook and Twitter across status updates. Several works on depression have emerged. They are based on social networks: Twitter [23, 24] and Facebook [25, 26].

Several authors have been interested in the use of emoticons to complete the sentiment analysis. Authors in [27] utilize Twitter API to get training data that contain emoticons like :) and :(. They use these emoticons as noisy labels. Tweets with :) are thought to be positive training data and tweets with :(are thought to be negative training data. In [28], authors present the ESLAM (Emoticon Smoothed LAnguage Models) which combine fully supervised methods and distantly supervised methods. Although many TSA (Twitter Sentiment Analysis) methods have been presented. The authors in [29] explored the influence of emoticons on TSA.

Automatic emotion recognition based on utterance level prosodic features may play an important role within speaker-independent emotion recognition [30]. The recognition of emotions based on the voice has been studied for decades [31, 32, 33, 34]. Paper in [35] focused on mono-modal systems with speech as only input channel.

Artificially influence mental and emotional states to get a better individual performance in stress-related occupations and prevent mental disorders from happening [36]. Recent research has shown that under certain circumstances multimodal emotion recognition is possible even in real time [37].

Sound signals (including human speech) is one of the main mediums of communication [38] and it can be processed to recognize the speaker or even emotion. There are some physical features applied for indexing speech, like: spectrum irregularity, wide and narrow band spectrograms, speech signals filtering and processing, enhancement and manipulation of specific frequency regions, segmentation and labeling of words, syllables and individual phonemes [37]. Moreover, the Mel-Frequency Cepstral Coefficients (MFCC) is widely used in speech classification experiments [39]. For

the reduction of leakage effect, the Hamming window is implemented. This is necessary for increasing the efficiency of frequency in human speech [38].

MPEG 7 Audio standard contains descriptors and description schemes that can be divided into two classes: generic low-level tools and application-specific tools [40]. Artificial Neural Networks (ANN), k-Nearest Neighbor (k-NN) and Support Vector Machines (SVM), decision trees, probabilistic models such as the Gaussian mixture model (GMM) or stochastic models such as Hidden Markov Model (HMM) can be applied [36].

Emotion analysis of speech is possible; however, it highly depends of the language. Study by Chaspari et al. showed that emotion classification in speech (Greek language) achieved accuracy up to 75.15% [41]. Similar study by Arruti et al. showed mean accuracy of 80.05% emotion recognition rate in Basque and a 74.82% in Spanish [42].

Nonverbal behavior constitutes useful means of communication in addition to spoken language [43] identifies at least six characteristics from posed facial actions that enable emotion recognition: morphology, symmetry, duration, speed of onset, coordination of apexes and ballistic trajectory. They are common to all humans confirming Darwin's evolutionary thesis. Therefore, an emotional recognition tools based on facial video is universal.

Automatic detection of emotions from facial expressions are not simple and their interpretation is largely context-driven. To reduce the complexity of automatic affective inference, measurement and interpretation of facial expressions, Ekman and Friesen developed in 1978 special system for objectively measuring facial movement; the Facial Action Coding System (FACS) [45]. FACS, based on a system originally developed by a Swedish anatomist named Hjortsjö [46] became the standard for identifying any movement of the face. Later, Ekman and Sejnowski studied also computer based facial measurements [47].

Automatic emotion recognition based on physiological signals is a key topic for many advanced applications (safe driving, security, mHealth, etc.). Main analyzed physiological signals useful for emotion detection and classification are:

- electromyogram (EMG) - recording of the electrical activity produced by skeletal muscles,
- galvanic skin response (GSR) - reflecting skin resistance, which varies with the state of sweat glands in the skin controlled by the sympathetic nervous system, where conductance is an indication of psychological or physiological state,
- respiratory volume (RV) - referring to the volume of air associated with different phases of the respiratory cycle,
- skin temperature (SKT) - referring to the fluctuations of normal human body temperature,
- blood volume pulse (BVP) - measures the heart rate,
- heart rate (HR),
- electrooculogram (EOG) - measuring the corneo-retinal standing potential between the front and the back of the human eye,
- photoplethysmography (PPG) - measuring blood volume pulse (BVP), which is the phasic change in blood volume with each heartbeat, etc.

The recognition of emotions based on physiological signals covers different aspects: emotional models, methods for generating

emotions, common emotional data sets, characteristics used and choices of classifiers. The whole framework of emotion recognition based on physiological signals has recently been described by [55].

III. THE PROPOSED MODEL

Basic Integrated Analytic System for Emotion detection. That uses the following models. A cognitive bias model, which is a systematic error in thinking that affects the decisions and judgments that people make. Some of these biases are related to memory. The way you remember an event may be biased for a number of reasons and that in turn can lead to biased thinking and decision-making. Other cognitive biases might be related to problems with attention. Since attention is a limited resource, people have to be selective about what they pay attention to in the world around them. Because of this, subtle biases can creep in and influence the way you see and think about the world.

A cognitive bias is a type of error in thinking that occurs when people are processing and interpreting information in the world around them. The human brain is powerful but subject to limitations. Cognitive biases are often a result of your brain's attempt to simplify information processing. They are rules of thumb that help you make sense of the world and reach decisions with relative speed.

Cognitive biases can be caused by a number of different things, but it is these mental shortcuts, known as heuristics, that often play a major contributing role. While they can often be surprisingly accurate, they can also lead to errors in thinking. Social pressures, individual motivations, emotions, and limits on the mind's ability to process information can also contribute to these biases. Selected attention involves the tendency to pay attention to some things while simultaneously ignoring others. This impacts not only the things that we perceive in the environment but the decisions that we make based upon our perceptions. It is often described as the ability to focus on and prioritize relevant information while filtering out irrelevant information. This prioritization transpires in attending only to a given channel in dichotic listening tasks, finding target features in visual search or identifying diagnostic features that distinguish categories.

Interpretive bias or interpretation bias is an information-processing bias, the tendency to inappropriately analyze ambiguous stimuli, scenarios and events. One type of interpretive bias is hostile attribution bias, wherein individuals perceive benign or ambiguous behaviors as hostile. For example, a situation in which one friend walks past another without acknowledgement. The individual may interpret this behavior to mean that their friend is angry with them.

A memory bias is a cognitive bias that either enhances or impairs the recall of a memory (either the chances that the memory will be recalled at all, or the amount of time it takes for it to be recalled, or both), or that alters the content of a reported memory.

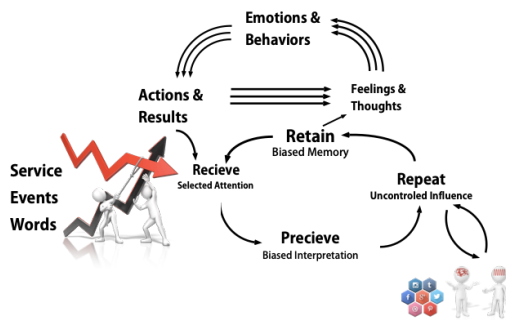


Figure 1: Interaction model

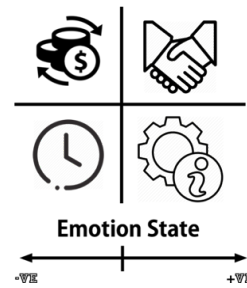


Figure 2: Behavioral Emotional Engagement Triggers

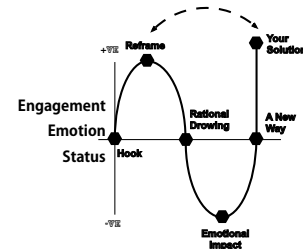


Figure 3: Audience engagement model



Figure 4: Decision Dominant Dividers

IV. EXPERIMENT AND PERFORMANCE EVALUATION

We developed an AI application to measure the impact of BEET on human engagement. The test has been applied 23,453 test cases where a personalized articles of an average of 2,000 words is customized to an audience with been profile. An audience is exposed to a random article vs. personalized article. We computed the engagement time and we determined that a none-customized article will be rejected with an average of 13.6 second and a customized article has an engagement time of an a average of 72.4 second. A performance will be presented in the final paper.

V. ACKNOWLEDGE

The authors would like to acknowledge the support of NSERC, NSF, DoD, CIDA and the University of Calgary on supporting this research. We would to acknowledge the support of MIT and UCLA for their extreme support to analyze the performance of this model.

VI. REFERENCES

1. Mehrabian, A., Ferris, S.R.: Inference of attitudes from nonverbal communication in two channels. *J. Consult. Psychol.* 31(3), 248 (1967)
2. Mood Ring Monitors Your State of Mind, *Chicago Tribune*, 8 October 1975, at C1; Ring Buyers Warm Up to Quartz Jewelry That Is Said to Reflect Their Emotions. *The Wall Street Journal*, 14 October 1975, at p. 16; and "A Ring Around the Mood Market", *The Washington Post*, 24 Nov. 1975, at B9
3. Picard, R.W.: *Affective Computing*. MIT Press, Cambridge (1997)
4. Picard, R.W., Vyzas, E., Healey, J.: Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Trans. Pattern Anal. Mach. Intell.* 23(10), 1175–1191 (2001)

5. Hernandez, J., et al.: AutoEmotive: bringing empathy to the driving experience to manage stress. In: DIS 2014, 21–25 June 2014, Vancouver, BC, Canada. ACM (2014). <http://dx.doi.org/10.1145/2598784.2602780>. 978-1-4503-2903-3/14/06
6. Zadeh, A., Zellers, R., Pincus, E., Morency, L.P.: Multimodal sentiment intensity analysis in videos: facial gestures and verbal messages. *IEEE Intell. Syst.* 31(6), 82–88 (2016). <https://doi.org/10.1109/mis.2016.94>
7. Wöllmer, M., et al.: YouTube movie reviews: sentiment analysis in an audio-visual context. *IEEE Intell. Syst.* 28(3), 46–53 (2013)
8. Perez-Rosas, V., Mihalcea, R., Morency, L.P.: Utterance-level multimodal sentiment analysis. In: *ACL*, vol. 1, pp. 973–982 (2013)
9. Zadeh, A., Chen, M., Poria, S., Cambria, E., Morency, L.P.: Tensor fusion network for multimodal sentiment analysis, arXiv:1707.07250. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 7–11 September 2017, Copenhagen, Denmark, pp. 1103–1114. Association for Computational Linguistics
10. Poria, S., Cambria, E., Hazarika, D., Majumder, N., Zadeh, A., Morency, L.P.: Context-dependent sentiment analysis in user-generated videos. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 873–883 (2017)
11. Poria, S., Cambria, E., Howard, N., Huang, G.B., Hussain, A.: Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing* 174(Part A), 50–59 (2016)
12. Liu, B.: Sentiment analysis and opinion mining. *Synth. Lect. Hum. Lang. Technol.* 5(1), 1–167 (2012)
13. Pang, B., Lee, L.: Opinion mining and sentiment analysis. *J. Found. Trends Inf. Retrieval* 2(1–2), 1–135 (2008)
14. Dziczkowski, G., Węgrzyn-Wolska, K.: RRSS - rating reviews support system purpose built for movies recommendation. In: Węgrzyn-Wolska, K.M., Szczepaniak, P.S. (eds.) *Advances in Intelligent Web Mastering. Advances in Soft Computing*, vol. 43, pp. 87–93. Springer, Berlin (2007)
15. Dziczkowski, G., Węgrzyn-Wolska, K.: An autonomous system designed for automatic detection and rating of film. Extraction and linguistic analysis of sentiments. In: *Proceedings of WIC*, Sydney (2008)
16. Dziczkowski, G., Węgrzyn-Wolska, K.: Tool of the intelligence economic: recognition function of reviews critics. In: *ICSOFT 2008 Proceedings*. INSTICC Press (2008)
17. Kepios: Digital in 2018, essential insights into internet, social media, mobile, and ecommerce use around the world, April 2018.
18. Ghiassi, M., Skinner, J., Zimbra, D.: Twitter brand sentiment analysis: a hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Syst. Appl.* 40(16), 6266–6282 (2013)
19. Zhou, X., Tao, X., Yong, J., Yang, Z.: Sentiment analysis on tweets for social events. In: *Proceedings of the 2013 IEEE 17th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2013*, 27–29 June 2013, pp. 557–562 (2013)
20. Salathé, M., Vu, D.Q., Khandelwal, S., Hunter, D.R.: The dynamics of health behavior sentiments on a large online social network. *EPJ Data Sci.* 2, 4 (2013).
21. xxi. Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., Demirbas, M.: Short text classification in Twitter to improve information filtering. In: *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 19–23 July 2010, pp. 841–842. <http://doi.acm.org/10.1145/1835449.1835643>
22. Seabrook, E.M., Kern, M.L., Fulcher, B.D., Rickard, N.S.: Predicting depression from language-based emotion dynamics: longitudinal analysis of Facebook and Twitter status updates. *J. Med. Internet Res.* 20(5), e168 (2018).
23. Wang, W., Hernandez, I., Newman, D.A., He, J., Bian, J.: Twitter analysis: studying US weekly trends in work stress and emotion. *Appl. Psychol.* 65(2), 355–378 (2016)
24. Reece, A.G., Reagan, A.J., Lix, K.L., Dodds, P.S., Danforth, C.M., Langer, E.J.: Forecasting the onset and course of mental illness with Twitter data (Unpublished manuscript). <https://arxiv.org/pdf/1608.07740.pdf>
25. Park, J., Lee, D.S., Shablack, H., et al.: When perceptions defy reality: the relationships between depression and actual and perceived Facebook social support. *J. Affect. Disord.* 200, 37–44 (2016)
26. Burke, M., Develin, M.: Once more with feeling: supportive responses to social sharing on Facebook. In: *Proceedings of the ACM 2016 Conference on Computer Supported Cooperative Work*, pp. 1462–1474 (2016)
27. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. *J. CS224N Proj. Rep.*, Stanford 1, 12 (2009)
28. Liu, K.L., Li, W.J., Guo, M.: Emoticon smoothed language models for Twitter sentiment analysis. In: *AAAI* (2012)
29. Węgrzyn-Wolska, K., Bougueroua, L., Yu, H., Zhong, J.: Explore the effects of emoticons on Twitter sentiment analysis. In: *Proceedings of Third International Conference on Computer Science & Engineering (CSEN 2016)*, 27–28 August 2016, Dubai, UAE
30. Bitouk, D., Verma, R., Nenkova, A.: Class-level spectral features for emotion recognition. *Speech Commun.* 52(7–8), 613–625 (2010)
31. Busso, C., et al.: Analysis of emotion recognition using facial expressions, speech and multimodal information. In: *Sixth International Conference on Multimodal Interfaces, ICMI 2004*, October 2004, State College, PA, pp. 205–211. ACM Press (2004)
32. Dellaert, F., Polzin, T., Waibel, A.: Recognizing emotion in speech. In: *International Conference on Spoken Language (ICSLP 1996)*, October 1996, Philadelphia, PA, USA, vol. 3, pp. 1970–1973 (1996)
33. Lee, C.M., et al.: Emotion recognition based on phoneme classes. In: *8th International Conference on Spoken Language Processing (ICSLP 2004)*, October 2004, Jeju Island, Korea, pp. 889–892 (2004)
34. Deng, J., Xu, X., Zhang, Z., Frühholz, S., Grandjean, D., Schuller, B.: Fisher kernels on phase-based features for speech emotion recognition. In: *Jokinen, K., Wilcock, G. (eds.) Dialogues with Social Robots. LNEE*, vol. 427, pp. 195–203. Springer, Singapore (2017). https://doi.org/10.1007/978-981-10-2585-3_15
35. Steidl, S.: Automatic classification of emotion-related user states in spontaneous children’s speech. Ph.D. thesis, Erlangen (2009)
36. Lugovic, S., Horvat, M., Dunder, I.: Techniques and applications of emotion recognition in speech. In: *MIPRO 2016/CIS* (2016)
37. Kukolja, D., Popović, S., Horvat, M., Kovač, B., Čosić, K.: Comparative analysis of emotion estimation methods based on physiological measurements for real-time applications. *Int. J. Hum.-Comput. Stud.* 72(10), 717–727 (2014)
38. Davletcharova, A., Sugathan, S., Abraham, B., James, A.P.: Detection and analysis of emotion from speech signals. *Procedia Comput. Sci.* 58, 91–96 (2015)
39. Tyburek, K., Prokopowicz, P., Kotlarz, P.: Fuzzy system for the classification of sounds of birds based on the audio descriptors. In: *Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2014. LNCS (LNAI)*, vol. 8468, pp. 700–709. Springer, Cham (2014).
40. Tyburek, K., Prokopowicz, P., Kotlarz, P., Michal, R.: Comparison of the efficiency of time and frequency descriptors based on different classification conceptions. In: *Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2015. LNCS (LNAI)*, vol. 9119, pp. 491–502. Springer, Cham (2015).
41. Chaspari, T., Soldatos, C., Maragos, P.: The development of the Athens Emotional States Inventory (AESI): collection, validation and automatic processing of emotionally loaded sentences. *World J. Biol. Psychiatry* 16(5), 312–322 (2015)
42. Arruti, A., Cearreta, I., Alvarez, A., Lazkano, E., Sierra, B.: Feature selection for speech emotion recognition in Spanish and Basque: on the use of machine learning to improve human-computer interaction. *PLoS ONE* 9(10), e108975 (2014)
43. Ekman, P.: Facial expression and emotion. *Am. Psychol.* 48, 384–392 (1993)
44. Jack, R.E., Schyns, P.G.: The human face as a dynamic tool for social communication. *Curr. Biol. Rev.* 25(14), R621–R634 (2015).
45. Ekman, P., Friesen, W., Hager, J.: Facial action coding system: Research Nexus. Network Research Information, Salt Lake City (2002)
46. C.H.: Man’s face and mimic language (1969).
47. Ekman, P., Huang, T.S., Sejnowski, T.J., et al.: Final report to NSF of the planning workshop on facial expression understanding, vol. 378. Human Interaction Laboratory, University of California, San Francisco (1993)
48. Afzal, S., Sezgin, T.M., Gao, Y., Robinson, P.: Perception of emotional expressions in different representations using facial feature points. *IEEE* (2009). 978-1-4244-4799
50. De la Torre, F., Chu, W.S., Xiong, X., Vicente, F., Ding, X., Cohn, J.: IntraFace. In: *IEEE International Conference on Automatic Face and Gesture Recognition Workshops* (2015).