Challenges of Emotion Detection Using Facial Expressions and Emotion Visualisation in Remote Communication

Eylül Ertay
Cornell Tech, Cornell University
New York, NY, USA
ee249@cornell.edu

Hao Huang
The University of Melbourne
Melbourne, Australia
hhh1@student.unimelb.edu.au

Zhanna Sarsenbayeva
The University of Melbourne
Melbourne, Australia
zhanna.sarsenbayeva@unimelb.edu.au

Tilman Dingler
The University of Melbourne
Melbourne, Australia
tilman.dingler@unimelb.edu.au

ABSTRACT

Emotion detection research has gained significant attention across different scientific disciplines, including Human-Computer Interaction and Ubiquitous Computing. In this work, we provide an overview of challenges that researchers and designers face when designing and deploying emotion detection systems. We discuss these challenges in the context of an online video conferencing tool we designed to detect emotions from facial expressions. Finally, we propose ideas on how to better design emotion inference systems and how to provide visual feedback on emotions.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; **Empirical studies in HCI**;

KEYWORDS

Emotion Detection; Affective Computing; Face Detection; Remote Chat; Video Conferencing Tool

ACM Reference Format:
Eylül Ertay, Hao Huang, Zhanna Sarsenbayeva, and Tilman Dingler. 2021. Challenges of Emotion Detection Using Facial Expressions and Emotion Visualisation in Remote Communication. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3460418.3479341

1 INTRODUCTION

Emotions are an integral part of human behaviour, and play a cardinal role in influencing interpersonal relationships [25]. When engaging in conversations with others, people express their feelings to each other using different types of cues, such as facial expressions, gestures, voice tone and level, as well as posture. Previous work has highlighted that being empathetic, i.e., understanding emotions and other people's feelings is an important skill in everyday life [36]. This has become particularly important during the COVID-19 pandemic, as we all wear face masks in public spaces and reading each other's emotions behind the mask has become a challenge. In addition, communication has increasingly moved from face-to-face to online settings.

Prior works have shown that emotions play a vital role in people's perception, decision making, creativity, memory retrieval, and social interaction [31]. Many HCI and UbiComp researchers have therefore created a range of technological solutions to detect, express, model, and communicate emotions to support people in their daily lives [31]. As a result, the field of Affective Computing has substantially progressed in the past few decades. Nevertheless, there still exist challenges that need to be addressed to further improve Affective Computing research.

In the scope of this paper, we provide an overview of currently existing challenges in emotion detection systems in terms of detection accuracy, ground truth measurement, reliability of data collection, and emotion visualisation. We further discuss these examples in the context of an online video-conferencing tool used in a case study. Finally, we provide suggestions for future research that can potentially take Affective Computing and emotion detection technology to the next level.

2 RELATED WORK

In a remote setting, there are many factors that negatively affect accurate inference of the emotions of interlocutors. Based on a literature analysis and a preliminary case study, we grouped the challenges of detecting and visualising emotions in remote video conversations. In the following subsections, we present a literature overview of the following categories: accuracy of emotion detection, ground truth, data collection reliability, and emotion visualisation.

2.1 Accuracy of Emotion Detection

Previous work has shown strong evidence of universality when it comes to facial expressions, as most of them describe specific emotions [49]. For many people, the most common way to read others' feelings is to observe their facial expressions, gestures and body language; muscles, lips, mouth, nose, and all other facial mimics indicate how an individual is feeling at a specific time [16]. Detecting

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA

ACM ISBN 978-1-4503-8461-2/21/09…$15.00
https://doi.org/10.1145/3460418.3479341

one's feelings by analysing their facial expressions has been a growing area of research. Over the past decade, machine learning has made emotion extraction from facial expressions significantly more accessible as many software tools and APIs perform video-based, automatic emotion inference [55]. Most emotion detection systems using facial expressions recognise six basic emotions: anger, disgust, fear, happiness, sadness, and surprise [10, 55], and scholars have successfully used such systems to conduct emotion detection research [3, 40, 41].

Following the COVID-19 pandemic, majority of social interactions moved to online environments. This makes it particularly interesting to investigate how emotions are conveyed when people communicate through video calls. Previous research has found that in Computer-Mediated Communication (CMC), the body language cues are not as available to the listeners as opposed to having a face-to-face conversation [50]. The lost information can also impact the automatic emotion detection process. For example, a key problem when detecting facial expressions from a video stream is that rigid movement of the face and body can reduce the accuracy of extracting facial features [17]. When a face is not fully visible in the video stream, which might be due to connection issues or occluded camera lenses, an optimal emotion detection system should be able to construct an entire face to be able to more accurately predict emotions [34].

Most emotion detection software using facial expressions further have a range of shortcomings as they are not indifferent to changes in appearances of people, such as people of different genders, ethnicities, ages; or simply wearing different accessories such as glasses, or having facial hair [34].

2.2 Ground Truth Collection

When validating the accuracy of automatic emotion-detection systems, it is paramount to measure a ground truth for emotions. Studies have used and compared different manners of collecting ground truth, including structured and non-structured questionnaires as well as self-assessments [21]. The Positive and Negative Affect Schedule (PANAS) [52], Self-Assessment Manikin (SAM) [24], and the photographic affect meter (PAM) are common examples of methods using self-reports [32]. As opposed to automated emotion detection systems, which usually detect six basic emotions [10, 55], subjective methods such as PANAS and PAM cover a wider set of emotions [21]. Measuring emotions through self-reports, however, has several limitations, including users not being able to assess their emotions correctly at all times, or not wanting to share their emotions truthfully [1].

The majority of previous research on emotion classification relies on surveys for ground truth, which is quite subjective and hard to fully depend on in a research setting [51]. Self-reports for emotions outline information about individuals' internal state, and are bound to fluctuate significantly depending on the psychological context that people are in [26, 27]. These emotional states can change quickly and, if they do, they are also called short-term emotions [22]. In our case study, we dealt with such momentary data for asynchronously detecting motions i.e., with a delay, for each second. We found that measuring the right emotion at the right time was a consequential challenge. We collected the self-reports through pop-up questionnaires on users' screens but the distraction caused by the participants' context could have resulted in less accurate data.

2.3 Data Collection Reliability

As automatically collected emotion data is based directly on the video stream of participants, each user will be present in a different environment. Thus, video components, such as ambient light and video as well as audio quality will differ. This can potentially affect the reliability of the video, resulting in a lower accuracy when predicting individuals' emotions.

2.3.1 *Ambient Light.*

One of the most prominent challenges of emotion detection is having different illumination conditions, as facial expressions may appear divergent under varying lighting conditions [45]. Ambient light has previously been shown to cause situational impairments in mobile interaction [42]; however, limited research has been conducted in order to study the effect of ambient light on the accuracy of detecting emotions from facial expressions. Tan *et al.* [46] have studied ways of enhancing face recognition under difficult lighting conditions, showing that accurately identifying facial features prove to be more challenging under dark conditions. Not being able to correctly recognise key facial points can, therefore, result in less reliable prediction of emotions. Moreover, if the luminosity is too high or the brightness is too low in a video, parts of the user's face can be unclear or not visible enough [55]. This constrains the emotion detection software from producing the most accurate deductions.

2.3.2 *Video Stream.*

In nearly all video conferencing applications, having technical issues with video collection from time to time is close to inevitable. Problems range from having a lag between transmissions due to unstable networks, freezing of video, as well as users having low resolution cameras. These technical difficulties pose challenges when collecting data from real-time video streams, where the main prediction of emotion is directly based on the video [18]. Yang *et al.* [55], for example, found that when using several facial emotion detection software, low quality images produced less accurate results and these tools had a large room for improvement when it comes to image quality. Additionally, preprocessing images when detecting emotions is a step of utmost importance [15]. However pre-processing a real-time video stream is often not feasible, and this step is therefore not implemented in most emotion detection softwares.

2.4 Visualising Emotions

How to represent emotions accurately to communicate them to others has been a scarce research question in Human-Computer Interaction. Previous studies have used models including sound and thermal stimuli as well as visual stimuli, such as animated shapes and colours [4, 30, 39, 44, 53]. Among those different forms of representations, visual feedback has been shown to induce emotions as a recognised method [56].

When investigating different methods to provide visual feedback in order to convey emotions, colour is a highly feasible option due to its affect on different parts of the human nervous system that are responsible for emotional arousal [19]. Xin *et al.* [54] explored the influence of culture and background on detecting colour emotions. Ohene-Djan *et al.* [29] proposed a visualisation interface called Real-Time Emotional Spectrum Tool (REST), which enables users to input a real-time, continuous flow of their predominant emotion and visualising it through a colour triangle.

Another visualisation method we considered in our work and in the case study presented is emojis. Emojis can be described as pictographs represented by Unicode characters and have been used as a way of expressing feelings in online platforms [35], even inducing emotions through electrical muscle stimulation [6]. Conventional emojis, however, has become an indispensable component of conversations for most people. In a technologically mediated world with a vast increase in social media use, people are becoming more used to adding emojis to text messages in order to enhance the emotions suggested by their sentences [5, 28]. This makes emojis a good candidate to give feedback on emotions due to their wide usage and universality. Nonetheless, emojis are prone to ambiguity as individuals might perceive them in different ways [35]. Robb *et al.* [35] used emojis as a tool for giving feedback on emotions and found them not as successful as image-based feedback.

Representing emotions through visualisations is a challenging task since emotions are delicate and personal. Thus, it is difficult to find a way to objectively visualise them as different techniques can be more effective for different, people making it onerous to have a standardised and effective visualisation for each emotion. Therefore, we explored the two above-mentioned methods of visualisation due to their commonality and ease of understanding.

3 CASE STUDY – WEB APPLICATION WITH BUILT-IN EMOTION DETECTION

In online video calls, a special form of CMC, important gestures and vital body language cues are often obscured [48]. Thus, a substantial challenge for Human-Computer Interaction researchers is to understand how these cues that are lost over remote conversations or video calls [48] affect people's ability to understand each others' emotions. Basing our research questions on these premises, the case study presented here explores how to enhance people's understanding of each other's emotions by providing visual feedback on their interlocutor's emotions.

3.1 System Overview

In this case study, we built a website for peer-to-peer video conferencing with built-in emotion detection. The inference of emotions is based on users' facial expressions. The tool continuously analyses and detects the emotions of two people conversing with each other using the video streams. It produces a real-time feedback loop where each participant is constantly provided with information about their partner's emotions. To achieve this, we implemented three major parts which constitute our system:

- An extendable video conferencing application
- An emotion detection component using face recognition
- The visualisation of detected emotions and creation of a real-time feedback loop between users

With the increasing use of video chatting applications, such as Zoom and Skype, we aimed to create a similar function with built-in emotion detection from facial expressions. Since the above-mentioned applications are not open-source, we developed a website where users could connect via their browser. We used WebRTC to establish peer-to-peer communication between browsers, which is compatible with the most common browsers, such as Chrome and Firefox. Another key factor we considered when choosing a deployment platform was the ability to add an emotion detection API to the video conferencing tool. WebRTC allowed us to use MediaStream and RTCPeerConnection interfaces [8] to establish a connection between multiple browsers, and transfer video streams.

3.1.1 RTCPeerConnection. The RTCPeerConnection makes up the central part of our web application. It creates a peer-to-peer connection between multiple browsers. When two clients join the same room the server uses web sockets to relay the video stream between participants [13]. Clients exchange messages using the Session Description Protocol (SDP) [20].

Once client A receives an answer from client B, they exchange and access each other's public IP addresses to establish the connection. We use the Network Address Translation (NAT) [9], which is widely used and allows direct communication between clients. WebRTC uses Session Traversal of User Datagram Protocol [33] through NAT to query the public IP addresses from a Session Traversal Utilities for NAT (STUN) server [7, 37].

3.1.2 MediaStream & Face and Emotions Detection. Our system uses the MediaStream API for managing the synchronized video and audio streams. The website we built obtains the MediaStream object and receives the local stream from the users' web camera, which it uses to run the emotion detection software. For that we used *face-api.js*, a JavaScript module that performs face landmark detection and emotion assessment. The module offers a package named 68 Point Face Landmark Detection Models, which maps the facial structure by detecting 68 facial landmarks [2], a common facial point detection technique [38]. The *face-api.js* module detects the facial landmarks and returns 7 values per second asynchronously. For each of the six basic emotions (anger, disgust, fear, happiness, sadness and surprise) as well as a neutral state, the algorithm calculates a probability. The emotion with the greatest probability is used as the basis for the visual feedback. Although *face-api.js*'s accuracy has not been thoroughly validated, we determined that its accuracy would be sufficient for the purposes of our case study.

3.2 Procedure

We conducted a pilot study, primarily to investigate how people respond to different visualisation techniques for embodying emotions, and to find out how these visualisations and a real-time feedback loop between two interlocutors would impact their ability to identify their partner's emotions.

We, therefore, recruited 12 participants (4 women, 8 men, with an average age of $M = 23.42$, $\sigma = 2.56$) through a local student association at the University of Melbourne. We obtained ethics approval from the university's ethics board.

The experiment took place over a Zoom call. The first step of the procedure was the briefing, where we asked participants to sign a consent form, which allowed us to collect their emotional data. The second step involved getting the participants to fill in a preliminary survey, through which we collected demographic information; namely age, gender, profession and native language. We also collected data regarding how empathetic the participants would consider themselves, asking them how expressive they are and how they would rate their skills of inferring other people's feelings. For the ground truth, we relied on self-reports. As a baseline measurement of the participants' feelings before and after the experiment, we asked them to fill in the Positive and Negative Affect Schedule (PANAS) [52], and calculated their scores in order to compare them to the results after the experiment.

Before initiating a conversation among the two participants, we asked them to introduce themselves to each other. After the introduction, participants navigated to the website where the experiment would be carried out. We notified the participants that a questionnaire with 2 questions would pop up once every 3 minutes throughout their conversation. This questionnaire would touch base on the participants' emotions by asking how they are feeling and what they think the other participant is feeling.

To initiate the conversation between participants, we provided them with general topics to talk about. Since we aimed to collect data from an ongoing discussion, possibly a heated conversation which will lead to a rapid change of emotions to be able to study the relevance of participants' emotional change, we chose both controversial and personal topics. We selected questions that are common enough for participants from different backgrounds to discuss and controversial enough to allow a different range of opinions among participants. We asked these questions as we were further interested to see how people performed in correctly recognizing each other's emotions.

The web application presented two links for selection: one option provided feedback to the user in colours and another option provided emotion feedback using emojis. These two feedback techniques were the independent variables in our experiment. Using a within-subjects study design we counterbalanced the order of presentation. Some experiments started the discussion with colour for the feedback, and others started the initial discussion with emoji. To help users better understand the meaning of each colour and emoji, we provided an instruction diagram at the bottom of the screen in our tool. This instruction was given in an array of images, where each colour and emoji was matched to its corresponding emotion. This chart is illustrated in Figure 1. We installed the instructive scale on the website to help users easily understand the feedback.

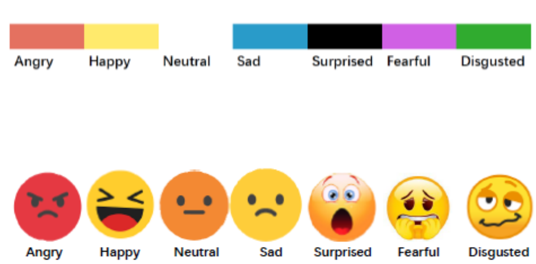

Figure 1: Colour and emoji mapping scale

Figure 2 presents a screenshot of our application depicting the view of the user on the left-hand side (User L). The screenshots, including Figure 2 and Figure 3 demonstrate the system, but not the actual study participants. For the demonstration purposes, two authors of the paper were able to connect to each other through the system, where one author (User L) was in Istanbul, Turkey, and the other author, on the right-hand side (User R) was in Melbourne, Australia. The screenshot shows a snippet from using the emoji to provide feedback on user emotions. It can be seen from Figure 2 that the emoji on the bottom left corner clearly identified User R's emotion –'disgusted' through his facial expression.

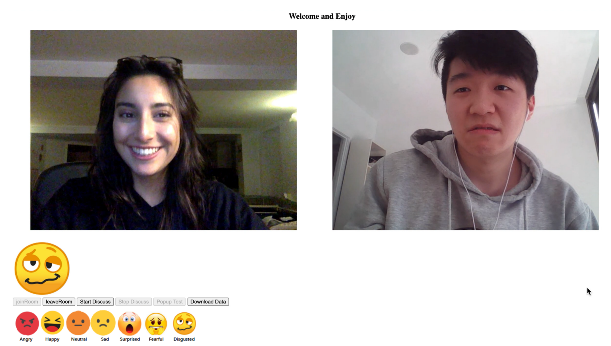

Figure 2: Two users conversing with the emoji feedback provided

Figure 3 on the other hand shows a glimpse of our system where colour is used to convey feedback on the users' emotions. The screenshot was taken by User L, and it can be clearly seen that the background colour is yellow, indicating that User R's 'happy' emotion was identified from his facial expression.

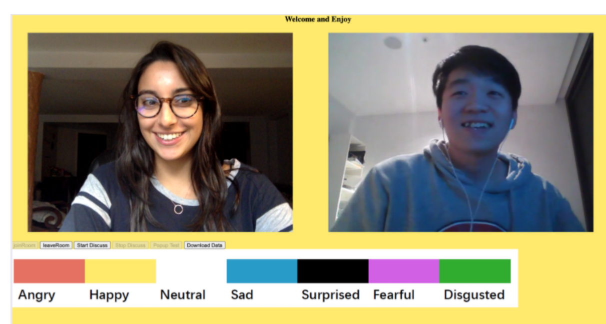

Figure 3: Two users conversing with the colour feedback provided

3.3 Results

From 12 participants, we collected 153,876 data points in total. 153,567 of these data points consisted of the emotion records detected through *face-api.js*, and 309 data points were emotion records obtained from the questionnaires. We removed 13,335 of these data points due to one of the participants not consenting to allow their data to be used for the purposes of the study and due to a failed data download in case of another participant. This resulted in 140,541 viable data points for analysis.

We found that, when participants received feedback on the other person's feelings through colour, they were able to correctly identify the other's feelings 67.6% of the time. In contrast, when they received the feedback in terms of emoji, they were able to correctly identify the other's feelings 66.7% of the time.

To see if the visualisation method variable and participants' level of accuracy in identifying others' emotions were independent of each other, we ran chi-square tests. The results did not reveal a significant main effect of the method of visualisation on participants' accuracy level of correctly identifying their partner's feelings $(F(1,74) = 0.576, p > 0.05)$.

We found that the automatic emotion recognition predicted the correct emotion of participants 54.1% of the time. However, we also found that among all the wrong predictions, 73.5% of them was due to the system predicting 'neutral' instead of the participants' real emotion, resulting in a lower accuracy.

4 DISCUSSION

Based on our experience of implementing the system we faced the following challenges: Firstly, the API we used for detecting participants' emotions from facial expressions did not have perfect accuracy, resulting in discrepancies between people's actual emotions (gathered from self-reports) and automatically detected emotions. Two participants reported that they observed that the emotion displayed would not match what they thought it was supposed to be, which brings out the question of accuracy. As discussed above, this can also be due to the fact that the emotional state of an individual cannot be solely determined from their facial expressions [12].

Another difficulty we faced while conducting our study was not being able to collect reliable data from all participants at all times. Some of the participants would be present in a rather dark environment, which could have complicated the emotion detection API to correctly identify their faces and predict their emotions. Additionally, some of the participants' videos were highly pixelated and, thus, not very clear. This could have been due to having web cameras of varying qualities, as well as network issues reducing the video quality. We faced connection related problems due to some participants having unstable networks, which resulted in a lag in conversations during the experiment. This also affected the natural flow of the conversation between participants.

Moreover, we found that connection issues could also alter the collection of ground truth data for emotions. As the conversation could be hindered due to network issues, this would also affect the emotional state of the participants. In addition, answering the self-reports could evoke different emotions in participants; or simply answering the questionnaire might steer and obstruct the conversation resulting in a change of emotional state of participants. We noticed that often participants would stop talking and focus on the self-report questionnaire when it was presented on their screen. Such examples show reasons why the assessment might not be accurate to detect participants' emotions and how well they were able to perceive their partner's feelings. It also shows that developing unobtrusive ground truth collection would benefit the experiments in the foreseeable future.

Another challenge we encountered throughout our study occurred when we were investigating how participants performed in correctly recognizing their experiment partner's emotions. In order to do this, we matched the emotions that participants inferred their partner was feeling with the detected emotion of their partner and used this to calculate the accuracy. This might not have represented the real accuracy rate of a participant correctly identifying their partner's feelings, as one would take around a minute to answer the questionnaire, whereas we were only able to retain the answers on the moment it was submitted. This allows for a period of time between receiving the questionnaire and submitting the questionnaire where both participants' emotions are subject to change.

Lastly, the emotion visualisation component of our experiment proved to be challenging, as some participants were confused about the feedback despite the feedback guides and explanations about colour and emojis. It was clear from the results that there wasn't one specific emotion visualisation technique that worked well with all participants. In the exit survey, some participants expressed that they preferred seeing the emojis, while others preferred colours.

An important limitation that should be mentioned in our study is the lack of baseline measurements. In our experiments, we merely tested and gathered data when participants were having a conversation while always receiving feedback on the other's emotions in some form. We did not include a baseline step in order to collect data on how well participants were able to recognise each other's feelings when there was no feedback provided. Such measurements could have provided a baseline to be able to compare which visualisation method has performed better in a more reliable way, as well as assessing the impact on the feedback on participants' ability to empathise and understand their partner's emotions.

5 FUTURE WORK

There is vast room for further research following the baseline outlined in this paper. As the world has recently shifted to a life where wearing masks is essential and most parts of our faces stay hidden behind the masks, it is notably interesting to investigate how much information can be read from seeing different parts of people's faces. A future study we plan to conduct focuses on this research question.

We plan to investigate how the ability of people to read others' emotions in an online conversation leads to being more empathetic, while several audio visual cues are hidden throughout the conversation. For our following study, we will be implementing a similar video conferencing application, which covers different parts of users' faces while they are partaking in a call. Some parts of participants' faces and their conversation partner's faces will be occluded by our software at determined time intervals. Audio may also be interrupted at times. We will again be using self-reports to collect emotional data for ground truth, through pop-up questionnaires initiated by the software.

Similar to our previous study, the new system will detect users' emotions by analysing their facial expressions. Furthermore, we will also be collecting participants' physiological data throughout the conversation, such as blood volume pulse (BVP) and heart rate variability (HRV); skin conductance response; and skin temperature. Feedback on participants' emotions will still be present as a part of the experiment, which will be conveyed through several colour schemes and patterns. We also plan to explore other visualisation methods such as GIFs and animations and/or animated backgrounds.

We will be giving conversation prompts to participants, where these prompts will cover both lighter subjects as well as emotionally charged topics. We hope to start conversations where the variety and intensity of emotions would be high in order to answer our next research questions investigating how people can become more empathetic.

Regarding our study outlined in this paper, we did not use any evidence-based colour schemes to map emojis to emotions. This is one of the limitations of our work. Nevertheless, we plan to enhance this work further by providing a proper mapping of human emotions to colour scheme in our future study.

As mentioned in the previous section, another notable limitation of our past study was the lack of baseline measurements. In our future study that we have outlined, we will also be including a baseline measurement step in order to see how well participants are able to understand each other's feelings prior to being provided with the emotion feedback while engaging in a conversation.

We would also like to outline general suggestions that need to be followed when conducting future research on emotion detection. First of all, it is important to provide longitudinal emotion sensing, as it has been shown in Psychology research that emotions and affect could be extended from the past affect states [47]. Moreover, it has been shown in literature that human affective states may be influenced by external environmental factors, e.g., ambient noise [43], weather and other environmental challenges [14]. Therefore, when designing emotion detection systems, it is crucial to learn how to integrate contextual information (both internal and external) in order to build efficient and effective emotion detection technologies [47].

Furthermore, when designing emotion detecting systems, it is necessary to investigate and find other modeling standards apart from humans, as literature has shown that humans are not very good at distinguishing false emotions from the real ones [11]. Moreover, every individual expresses their emotions in a different way; hence, it is important to account for personal and individual characteristics when designing emotion detection systems [23]. In other words, affect detection systems should be able to learn from and model users' emotions.

Lastly, it is necessary to understand that emotion sensing should happen only when it is relevant and appropriate. As Picard [31] states, there is a time to express and forbear emotions, as well as there is a time to sense and ignore the feelings of others. While this balance exists in human life; it is missing in Affective Computing [31]. Hence, developers and designers of technology should aim to reach this balance when creating affect detection systems [23, 31].

6 CONCLUSION

In this paper, we provide a detailed outline of the challenges that might arise when using emotion recognition via facial expressions in remote communication tools. The main challenges include accuracy of emotion detection, ground truth for emotions, reliability of data collection, and visualising emotions when providing feedback on emotions. We discuss these challenges in the context of the case study we conducted at the University of Melbourne. Finally, we provide suggestions for future emotion detection to bring Affective Computing to remote communication tools.

ACKNOWLEDGMENTS

This work was partially funded by Science Gallery Melbourne. Zhanna Sarsenbayeva is supported by the Doreen Thomas Postdoctoral Fellowship Program.

REFERENCES

[1] Yusuf Albayram, Mohammad Maifi Hasan Khan, Theodore Jensen, Ross Buck, and Emil Coman. 2018. The effects of risk and role on users' anticipated emotions in safety-critical systems. In *International Conference on Engineering Psychology and Cognitive Ergonomics*. Springer, 369–388.

[2] Giuseppe Amato, Fabrizio Falchi, Claudio Gennaro, and Claudio Vairo. 2018. A comparison of face verification with facial landmarks and deep features. In *10th International Conference on Advances in Multimedia (MMEDIA)*. 1–6.

[3] Ebrahim Babaei, Namrata Srivastava, Joshua Newn, Qiushi Zhou, Tilman Dingler, and Eduardo Velloso. 2020. *Faces of Focus: A Study on the Facial Cues of Attentional States*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376566

[4] Pascal Belin, Sarah Fillion-Bilodeau, and Frédéric Gosselin. 2008. The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior research methods* 40, 2 (2008), 531–539.

[5] Henriette Cramer, Paloma de Juan, and Joel Tetreault. 2016. Sender-intended functions of emojis in US messaging. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 504–509.

[6] Tilman Dingler, Takashi Goto, Benjamin Tag, and Kai Kunze. 2017. EMS Icons: Conveying Information by Analogy to Enhance Communication through Electrical Muscle Stimulation. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (Maui, Hawaii) *(UbiComp '17)*. Association for Computing Machinery, New York, NY, USA, 732–739. https://doi.org/10.1145/3123024.3129275

[7] Sam Dutton. 2013. WebRTC in the real world: STUN, TURN and signaling. *Google, Nov* (2013).

[8] Sam Dutton et al. 2012. Getting started with WebRTC. *HTML5 Rocks* 23 (2012).

[9] K. Egevang and P. Francis. 1994. The IP Network Address Translator (NAT). *RFC* 1631 (1994), 1–10.

[10] Paul Ekman. 1992. An argument for basic emotions. *Cognition and Emotion* 6, 3-4 (1992), 169–200. https://doi.org/10.1080/02699939208411068 arXiv:https://doi.org/10.1080/02699939208411068

[11] Paul Ekman and Maureen O'Sullivan. 1991. Who can catch a liar? *American psychologist* 46, 9 (1991), 913.

[12] Samuel W Fernberger. 1929. Can an emotion be accurately judged by its facial expression alone. *Am. Inst. Crim. L. & Criminology* 20 (1929), 554.

[13] Ian Fette and Alexey Melnikov. 2011. The websocket protocol.

[14] Nico H Frijda et al. 1986. *The emotions*. Cambridge University Press.

[15] Sasanko Sekhar Gantayat and Swathi Lenka. 2021. Study of Algorithms and Methods on Emotion Detection from Facial Expressions: A Review from Past Research. In *Communication Software and Networks*, Suresh Chandra Satapathy, Vikrant Bhateja, M. Ramakrishna Murty, Nguyen Gia Nhu, and Jayasri Kotti (Eds.). Springer Singapore, Singapore, 231–244.

[16] Jose Maria Garcia-Garcia, Victor MR Penichet, and Maria D Lozano. 2017. Emotion detection: a technology review. In *Proceedings of the XVIII international conference on human computer interaction*. 1–8.

[17] Anchal Garg and Rohit Bajaj. 2015. Review on Facial Emotion Recognition. *International Journal of Advanced Trends inComputer Applications* 1, 1 (2015), 10–13.

[18] Scott B. Greenspan, Kelsey L Gordon, Sara A. Whitcomb, and Alexandra A. Lauterbach. 2021. Use of Video Conferencing to Facilitate Focus Groups for Qualitative Data Collection. *American Journal of Qualitative Research* 5, 1 (2021), 85–93.

[19] Mitsuhiko Hanada. 2018. Correspondence analysis of color–emotion associations. *Color Research & Application* 43, 2 (2018), 224–237. https://doi.org/10.1002/col.22171 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/col.22171

[20] Mark Handley, Van Jacobson, Colin Perkins, et al. 1998. SDP: session description protocol.

[21] Mariam Hassib, Max Pfeiffer, Stefan Schneegass, Michael Rohs, and Florian Alt. 2017. Emotion actuator: Embodied emotional feedback through electroencephalography and electrical muscle stimulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 6133–6146.

[22] Jennifer M Jenkins and Keith Oatley. 2000. Psychopathology and short-term emotion: The balance of affects. *The Journal of Child Psychology and Psychiatry and Allied Disciplines* 41, 4 (2000), 463–472.

[23] Arvid Kappas. 2010. Smile when you read this, whether you like it or not: Conceptual challenges to affect detection. *IEEE Transactions on Affective Computing* 1, 1 (2010), 38–41.

[24] Peter J. Lang, Mark K. Greenwald, Margaret M. Bradley, and Alfons O. Hamm. 1993. Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology* 30, 3 (1993), 261–273. https://doi.org/10.1111/j.1469-8986.1993.tb03352.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8986.1993.tb03352.x

[25] Richard S. Lazarus. 1991. Emotion and adaptation. Oxford University Press on Demand.

[26] Heather C Lench, Sarah A Flores, and Shane W Bench. 2011. Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: a meta-analysis of experimental emotion elicitations. *Psychological bulletin* 137, 5 (September 2011), 834—855. https://doi.org/10.1037/a0024244

[27] Iris B Mauss, Robert W Levenson, Loren McCarter, Frank H Wilhelm, and James J Gross. 2005. The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion* 5, 2 (2005), 175.

[28] Hannah Miller, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson, Loren Terveen, and Brent Hecht. 2016. Blissfully happy" or "ready to fight": Varying Interpretations of Emoji. *Proceedings of ICWSM* 2016 (2016).

[29] James Ohene-Djan, A. Sammon, and Rachel Shipsey. 2006. Colour Spectrums of Opinion: An Information Visualisation Interface for Representing Degrees of Emotion in Real Time. *Proceedings of the International Conference on Information Visualisation*, 80 – 88. https://doi.org/10.1109/IV.2006.34

[30] Sébastien Paquette, Isabelle Peretz, and Pascal Belin. 2013. The "Musical Emotional Bursts": a validated set of musical affect bursts to investigate auditory affective processing. *Frontiers in psychology* 4 (2013), 509.

[31] Rosalind W Picard. 2003. Affective computing: challenges. *International Journal of Human-Computer Studies* 59, 1-2 (2003), 55–64.

[32] John Pollak, Phil Adams, and Geri Gay. 2011. PAM: A Photographic Affect Meter for frequent, in situ measurement of affect. *Conference on Human Factors in Computing Systems - Proceedings*, 725–734. https://doi.org/10.1145/1978942.1979047

[33] Jon Postel et al. 1980. User datagram protocol. (1980).

[34] Jyoti Rani and Kanwal Garg. 2014. Emotion Detection Using Facial Expressions -A Review. *International Journal of Advanced Research in Computer Science and Software Engineering* 4, 4 (2014), 465–467.

[35] David Robb, Stefano Padilla, Thomas Methven, Britta Kalkreuter, and M. Chantler. 2017. Image-based Emotion Feedback: How Does the Crowd Feel? And Why? https://doi.org/10.1145/3064663.3064665

[36] Claire L. Roether, Lars Omlor, Andrea Christensen, and Martin A. Giese. 2009. Critical features for the perception of emotion from gait. *Journal of Vision* 9, 6. https://doi.org/10.1167/9.6.15

[37] Jonathan Rosenberg, Joel Weinberger, Christian Huitema, and Rohan Mahy. 2003. STUN-simple traversal of user datagram protocol (UDP) through network address translators (NATs). (2003).

[38] Christos Sagonas, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic. 2013. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 397–403.

[39] Katri Salminen, Veikko Surakka, Jukka Raisamo, Jani Lylykangas, Johannes Pystynen, Roope Raisamo, Kalle Mäkelä, and Teemu Ahmaniemi. 2011. Emotional responses to thermal stimuli. In *Proceedings of the 13th international conference on multimodal interfaces*. 193–196.

[40] Zhanna Sarsenbayeva, Denzil Ferreira, Niels van Berkel, Chu Luo, Mikko Vaisanen, Vassilis Kostakos, and Jorge Goncalves. 2017. Vision-Based Happiness Inference: A Feasibility Case-Study. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (Maui, Hawaii) *(UbiComp '17)*. Association for Computing Machinery, New York, NY, USA, 494–499. https://doi.org/10.1145/3123024.3124438

[41] Zhanna Sarsenbayeva, Gabriele Marini, Niels van Berkel, Chu Luo, Weiwei Jiang, Kangning Yang, Greg Wadley, Tilman Dingler, Vassilis Kostakos, and Jorge Goncalves. 2020. *Does Smartphone Use Drive Our Emotions or Vice Versa? A Causal Analysis*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376163

[42] Zhanna Sarsenbayeva, Niels van Berkel, Weiwei Jiang, Danula Hettiachchi, Vassilis Kostakos, and Jorge Goncalves. 2019. Effect of Ambient Light on Mobile Interaction. In *Human-Computer Interaction – INTERACT 2019*, David Lamas, Fernando Loizides, Lennart Nacke, Helen Petrie, Marco Winckler, and Panayiotis Zaphiris (Eds.). Springer International Publishing, Cham, 465–475.

[43] Zhanna Sarsenbayeva, Niels van Berkel, Eduardo Velloso, Vassilis Kostakos, and Jorge Goncalves. 2018. Effect of distinct ambient noise types on mobile interaction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 1–23.

[44] Marc Schröder. 2003. Experimental study of affect bursts. *Speech communication* 40, 1-2 (2003), 99–116.

[45] P. Suja, Sherin Mariam Thomas, Shikha Tripathi, and V. K. Madan. 2016. Emotion Recognition from Images Under Varying Illumination Conditions. In *Soft Computing Applications*, Valentina Emilia Balas, Lakhmi C. Jain, and Branko Kovačević (Eds.). Springer International Publishing, Cham, 913–921.

[46] Xiaoyang Tan and Bill Triggs. 2010. Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions. *IEEE Transactions on Image Processing* 19, 6 (2010), 1635–1650. https://doi.org/10.1109/TIP.2010.2042645

[47] Jianhua Tao and Tieniu Tan. 2005. Affective computing: A review. In *International Conference on Affective computing and intelligent interaction*. Springer, 981–995.

[48] Cameron Hon Leng Teoh. 2012. Body Language Availability In Videoconferencing. (2012).

[49] Thanyathorn Thanapattheerakul, Katherine Mao, Jacqueline Amoranto, and Jonathan Chan. 2018. Emotion in a Century: A Review of Emotion Recognition. 1–8. https://doi.org/10.1145/3291280.3291788

[50] Joseph B. Walther. 1992. Interpersonal Effects in Computer-Mediated Interaction: A Relational Perspective. *Communication Research* 19, 1 (1992), 52–90. https://doi.org/10.1177/009365092019001003 arXiv:https://doi.org/10.1177/009365092019001003

[51] Yichen Wang and Aditya Pal. 2015. Detecting Emotions in Social Media: A Constrained Optimization Approach. In *Proceedings of the 24th International Conference on Artificial Intelligence* (Buenos Aires, Argentina) *(IJCAI'15)*. AAAI Press, 996–1002.

[52] David Watson, Lee Anna Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology* 54, 6 (1988), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

[53] Graham Wilson, Pietro Romeo, and Stephen A Brewster. 2016. Mapping Abstract Visual Feedback to a Dimensional Model of Emotion. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 1779–1787.

[54] John Xin, K. Cheng, Gail Taylor, T. Sato, and Aran Hansuebsai. 2004. Cross-regional comparison of colour emotions Part I: Quantitative analysis. *Color Research & Application* 29 (12 2004), 451 – 457. https://doi.org/10.1002/col.20062

[55] Kangning Yang, Chaofan Wang, Zhanna Sarsenbayeva, Benjamin Tag, Tilman Dingler, Greg Wadley, and Jorge Goncalves. 2021. Benchmarking commercial emotion detection systems using realistic distortions of facial image datasets. *The Visual Computer* (2021), 1447–1466. https://doi.org/10.1007/s00371-020-01881-x

[56] Shigeo Yoshida, Tomohiro Tanikawa, Sho Sakurai, Michitaka Hirose, and Takuji Narumi. 2013. Manipulation of an emotional experience by real-time deformed facial feedback. In *Proceedings of the 4th Augmented Human International Conference*. 35–42.