Enabling Continuous Emotional Status Display in Mobile Text Chat

Jackson Feijó Filho[(✉)], Wilson Prata, and Thiago Valle

Nokia Technology Institute, Av. Torquato Tapajós, 7200 - Col. Terra Nova,
Manaus, AM, Brazil
{jackson.feijo,wilson.prata,thiago.valle}@indt.org.br

Abstract. Nonverbal communication is essential to human interaction. All this information is absent in mobile text conversations. This work describes an effort to extend text chat in mobile devices by continuously displaying emotional status and inserting automatically detected facial expression responses, such as a smile, to conversations. Acknowledged related work, regarding the examination of nonverbal communication through text messaging are considered and distinguished from the present solution. The implementation of a mobile phone application with the debated feature is described and user studies are reported. Context of application, conclusions and future works are also discussed.

1 Introduction

Nonverbal communication has been intriguing researchers for very long. In early 1870 s, scientific literatures on nonverbal communication and comportment were published in Charles Darwin's book "The Expression of the Emotions in Man and Animals" [5]. Darwin claimed that all mammals, both humans and animals, presented emotion through facial expressions. Affirming questions such as: "Why do our facial expressions of emotions take the particular forms they do?" and "Why do we wrinkle our nose when we are disgusted and bare our teeth when we are enraged?" [6].

Humans depend on visual signals when interacting within a social setting, and in an online conversation scenario, the lack of eyesight due to physical remoteness hinders the quality of social interactions. Since essential nonverbal communication is comprised of facial gestures and expressions, text chat users are at a disadvantage when engaging in interpersonal communication.

2 Problem Scenario

Nonverbal indications carried in a conversation is often conclusive for disambiguation. In [8], long-distance relationships and video chat technology is well debated. In one of the interviews from this work, some subjects have described video as a way to see their partners' facial expressions and body language. In some cases, this helped avoid miscommunications.

"I always apparently sound pretty harsh when I'm talking …even when I'm joking it doesn't sound like I'm joking…I would sometimes upset her without even knowing I upset her and, of course, without intending…" – P3.

Moreover, from the work in [18] where emotional feedback is discreetly inserted to the conversation timeline, experiments and interviews have shown that an isolated insertion of an emotional response is not always suitable to represent the continuity of the nonverbal expression. That suggests that another form of continuous representation of emotions is still lacking in the interface proposed in [18].

3   Related Work

There has been substantial work on the recognition and conveying of emotions in computer-mediated communications.

3.1   Video Chat

As described in [9], the use of video chat successfully transmits consistent non-verbal elements in conversations. In this work, we will not consider video chat scenarios as related work, since we are trying to intervene in online text chat scenarios. This distinction is appropriate due to the fact that video chat typically requires higher performance systems with more expensive hardware, faster networks and more processing power (compared to lower-end platforms with minimum requirements to exchange text through a server). Not only structurally, video-chatting establishes a different social ritual [9]. In addition, video-chat may often convey extra imagery such as background rooms, make-up, clothing and hair style, as opposed to transmitting only processed facial expressions, e.g. smile.

3.2   Conveying Emotion in Text Chat

Numerous existing approaches have been trying to derive affect from written language on a semantic basis, making use of common sense knowledge [11]. The work in [10] is an example of approaches that automatically identify emotions in texts.

The work in [12] attempts to detect deceptive chat-based communication using typing behavior and message cues. It is an example of deriving nonverbal information in chat texts with a non-semantics approach. It postulates that deception impacts response time, word count, lexical diversity, and the number of times a chat message is edited. It focuses on detecting deception, unlike our work, which is attempting to convey emotions.

The works in [10] and [13] distinguish themselves from the others because they attempt to assimilate natural language and emotions by performing the recognition of these emotions based on physiological user input. They o at collecting additional information on the user's emotional state by recording and analyzing physiological feedback. Therefore, while interacting with the application, the user is monitored by means of bio sensors, measuring skin conductivity, heart rate, respiration and muscle activity [10].

Specifically in [13] modifications in the display of the text are used to communicate emotions (animated text). Our work differs from these because it relies solely on low-resolution mobile phones front cameras, not requiring extra hardware. Also, it restricts the recognition of emotions to processing images of faces through mobile phones.

The accessibility proposal in [5] approaches the present matter by implementing a computer vision system with facial recognition and expression algorithms to relay nonverbal messages to blind users. This project differs from our work by focusing on visually impaired audience. In addition, unlike our solution, robust processing power is required for demanding algorithms if compared to our software. Lastly it is pertinent to mention that it relies on extra hardware, for example, using a high speed IEEE-1394b camera.

4   Continuous Emotional Status Display in Mobile Text Chat

Our solution attempts to intervene in mobile online text chat by detecting users' emotional reactions during the conversations. Initially, we implemented the chat application to be integrated with the emotion detection feature.

4.1   Mobile Chat Application

The mobile chat application was developed for Windows Phone, using the XMPP protocol [14] (formerly and also known as Jabber) as not only it is defined in open standards but it is also widely positioned across the Internet.

This standard is used by very popular chat platforms like Google Talk, Messenger and Facebook Chat, which made it simpler for test subjects to use our solution using their own everyday accounts.

4.2   Mobile Emotion Detection Software

In [18], the detection of emotions is performed only when new messages have arrived, in order to increase the possibility of a facial expression to be a reaction to the most recent message and minimize server requests.

Experiments and interviews have shown that these isolated emotional feedback ignores the continuity of nonverbal cues. Furthermore, it disregards the paralleled nature of verbal and nonverbal communications. That suggests that another form of continuous representation of emotions is still lacking in the interface proposed in [18].

To address this continuity feature, the image capturing and emotion recognition attempts to be executed every 2 s, providing a more fluid perception of nonverbal status.

This has enabled the discrete structure of the system to be maintained, and the image processing can still be executed on the server side. This is important because lower-end devices, with low memory and low processing power can be used with good performance.

In [15], Krishna et al. uses a Principal Component Analysis (PCA) algorithm for facial recognition. For the work in [18], our system used a commercial solution (Rekognition;

Orbeus, Inc.) [16] for face and emotion detection. Presently, we have implemented a more simple web service that accepts requests containing face images, and returns a JSON with emotion information. We have used Intel's Perceptual SDK to develop this server side system.

System Structure. As shown in Fig. 1, the user chats from the mobile, using our application. As soon as a conversation starts, the face images start being collected, converted to base64, and sent via http to the server to be processed. The server sends back a JSON feed, which is parsed to retrieve emotions status. The emotions results range from 0 to 1, providing the detected level of "happy", "surprised" and "calm". If the level one of these emotions are above a pre-determined threshold, the software automatically adds "<name-of-user > smiles" or "<name-of-user > is surprised", to the conversation thread. In parallel, the client software on the phone compares the higher emotion feedback, to continually update the status.
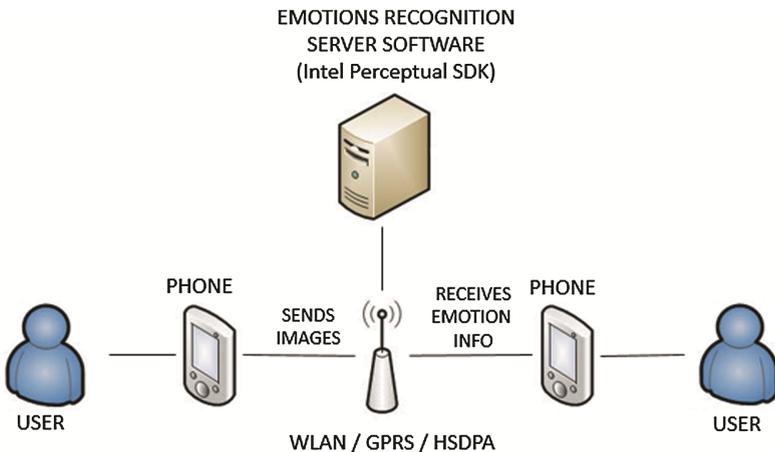

EMOTIONS RECOGNITION
SERVER SOFTWARE
(Intel Perceptual SDK)

PHONE   SENDS   ((•))   RECEIVES   PHONE
               IMAGES   EMOTION
                                  INFO

USER   WLAN / GPRS / HSDPA   USER

Fig. 1. System structure

4.3   Screenshot

Figure is a screenshot of the application, with an on-going dialogue. It is relevant to mention that no actual face images are exchanged. This consolidates the social ritual of the text dialoguing, when people are able to have conversations unmindful of make-up, hair and overall facial aesthetics status. There is also a clear demonstration of when the "emotions feature" is on or off (Figs. 2 and 3).
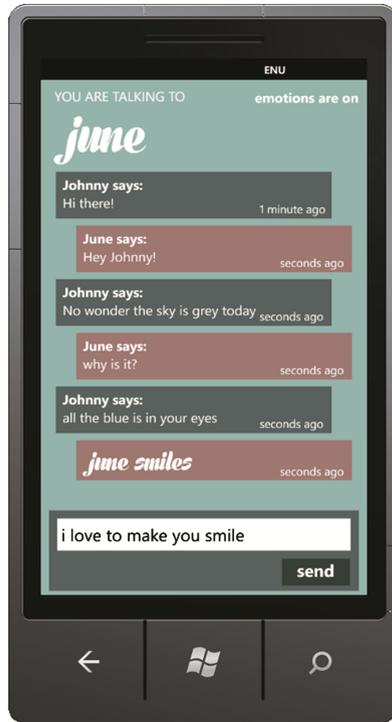
Fig. 2. Screenshot from the application in [18]

5   Interviews and Experiments

Firstly, we perceived that the test subjects were inclined to change their facial expressions for no reason, in order to check with their chat partners whether the emotion detection accurately worked. Because of that, we decided to implement a preliminary educational usability test phase, in order to get test subjects habituated with the software.

For the sake of productivity, we recruited test subjects, chat partners, who are used to talking to each other using mobile text chat.

We referred to some inquiries in the work by Wang [12] when developing the questionnaire. The subjects were asked to answer the questions after using both modes of chat. See Table 1.

6   Results and Discussion

We developed an approach to the integration of continuous nonverbal communication to mobile text chat. The motivation is to enrich mobile text chat by conveying frequently lost non-verbal communication. This benefits users by conveying information that is valuable to augment textual conversations.
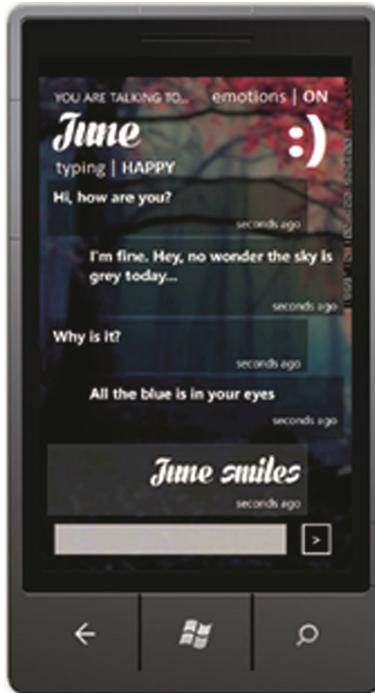
Fig. 3. Screenshot of the application

Our solution differed from the related research for it did not attempt to extract affection from the semantics of the text and did not use any extra hardware to collect images or physiological signals. Finally, our work is devoted to relaying emotions on mobile text chat scenarios, relying on lower end devices with front cameras.

We implemented a mobile text chat application based on the XMPP protocol, easily connectable with popular chat platforms, with the feature of detecting users' emotions automatically and transmitting them to chat partners.

We have also implemented the server side software that processes that face images, returning a JSON feed with emotional feedback information.

During the experiments and interviews we found that people perceived their partners' reactions as an accomplishment or a reward to what they just posted.

This has included some sort of pressure in the commitment of the conversation, in comparison to plain text chat. In some cases, subjects have mentioned that their partners inserted more humor then usual to the dialogue, in order to receive the automatically detected smiles. During the preliminary trials, we found that people tended to try to foresee their chat partner's reactions, creating frustration when no smiles when detected or causing great sense of accomplishment when it did.

From the interviews, which were conducted individually and in groups, we were able to observe the rise of the sense of connection that chat partners experienced with one another, even more with the constant emotional status feedback.

Table 1. Comparison of the application with and without emotional feedback. Mean scores for questions about the interactive experience in chat systems with and without emotional feedback (EF and Non-EF). The possible responses ranged from 1 (disagree) to 5 (agree) (10 subjects).

Question	EF	Non-EF
I became very involved in the conversation	4.4	3.2
I enjoyed using the system	4.5	3.1
It was easy to understand my partner's emotions	4.8	2.9
I emulated some of my expressions	0.5	0
I felt more connected to my chat partner	4.7	2.1
I would use this application in my everyday chat conversations.	4.1	3.7

References

1. Argyle, M., Salter, U., Nicholson, H., Williams, M., Burgess, P.: The communication of inferior and superior attitudes by verbal and non-verbal signals. Br. J. Soc. Psychol. **9**, 222–231 (1970)
2. Intel Perceptual SDK. https://software.intel.com/en-us/perceptual-computing-sdk
3. Codes of professional ethics: Social Science Computing Review **10**(2), 453–469 (1992)
4. Liu, H., Lieberman, H., Selker, T.: A model of textual affect sensing using real-world knowledge. In: Johnson, W.L., Andre, E., Domingue, J. (eds.) International Conference on Intelligent User Interfaces, pp. 125–132. ACM, Florida (2003)
5. Astler, D., Chau, H.: Increased accessibility to nonverbal communication through facial and expression recognition technologies for blind/visually impaired subjects. In: ASSETS 2011 the Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 259–260. ACM, New York (2011)
6. Pease, B., Pease, A.: The Definitive Book of Body Language. Bantam Books, New York (2004)
7. Krauss, R.M., Chen, Y., Chawla, P.: Nonverbal behavior and nonverbal communication: what do conversational hand gestures tell us? Adv. Exp. Soc. Psychol. **1**(2), 389–450 (2000)
8. Power, M.R.: Chapter 11: Non-verbal communication Working Through Communication.Paper 12 (1998)
9. Neustaedter, C., Greenberg, S.: Intimacy in long-distance relationships over video chat. In: CHI 2012 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 753–762. ACM, New York (2012)
10. Bosma, W., Andre, E.: Exploiting emotions to disambiguate dialogue acts. In: IUI 2004 Proceedings of the 9th International Conference on Intelligent User Interfaces, pp. 85–92. ACM, New York (2004)
11. Barros, L., Rodríguez, P., Ortigosa, A.: Emotion recognition in texts for user model augmenting. In: INTERACCION 2012 Proceedings of the 13th International Conference on Interacción Persona-Ordenador, Article No. 45. ACM, New York (2012)

12. Detecting Deceptive Chat-Based Communication Using Typing Behavior and Message Cues. ACM Transactions on Management Information Systems (TMIS) TMIS. Homepage archive. 4 (2), Article No. 9. ACM, New York (2013)
13. Wang, H., Predinger, H., Igarashi, T.: Communicating emotions in online chat using physiological sensors and animated text. In: CHI EA 2004 CHI 2004 Extended Abstracts on Human Factors in Computing Systems, pp. 1171–1174. ACM, New York (2004)
14. XMPP Standards Foundation. http://www.xmpp.org
15. Little, G., Black, J., Panchanathan, S., Krishna, S.A.: Wearable face recognition system for individuals with visual impairments. In: 7th Annual ACM Conference on Assistive Technologies, pp. 106–113 (2005)
16. Rekognition APIs and SDK. https://rekognition.com
17. Marge, M., Joao Miranda, A.W.B., Rudnicky, A.I.: Towards improving the naturalness of social conversations with dialogue systems. In: Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pp. 91–94. Association for Computational Linguistics (2010)
18. Feijó Filho, J., Valle, T., Prata, W.: Non-verbal communications in mobile text chat: emotion-enhanced mobile chat. In: Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services. ACM (2014)