Current and New Research Perspectives on Dynamic Facial Emotion Detection in Emotional Interface

Tessa-Karina Tews[1], Michael Oehl[1], Helmut Faasch[1], and Taro Kanno[2]

[1] Leuphana University Lueneburg, Institute of Experimental Industrial Psychology, Wilschenbrucher Weg 84a, 21335 Lueneburg, Germany
tews@leuphana.de, {oehl,faasch}@uni.leuphana.de
[2] The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan
kanno@sys.t.u-tokyo.ac.jp

Abstract. In recent years there has been an increasing interdisciplinary exchange between psychology and computer science in the field of recognizing emotions for future-oriented Human-Computer and Human-Machine Interfaces. Although affective computing research has made enormous progress in automatically recognizing facial expressions, it has not yet been fully clarified how algorithms can learn to encode or decode a human face in a real environment. Consequently, our research focuses on the detection of emotions or affective states in a Human-Machine setting. In contrast to other approaches, we use a psychology driven approach trying to minimize complex computations by using a simple dot-based feature extraction method. We suggest a new approach within, but not limited to, a Human-Machine Interface context which detects emotions by analyzing the dynamic change in facial expressions. In order to compare our approach, we discuss our software with respect to other developed facial expression studies in context of its application in a chat environment. Our approach indicates promising results that the program could accurately detect emotions. Implications for further research as well as for applied issues in many areas of Human-Computer Interaction, particularly for affective and social computing, will be discussed and outlined.

Keywords: Emotional Interfaces, Affective Computing, Facial Expression, Human Machine Interface.

1 Introduction

Recently, there has been a promising increase in interdisciplinary exchange between psychology and computer science in the field of recognizing emotions [1-2]. Therefore, our study attempts to present previous developed affective communication methods and compare them with a possible usage in reliable chat emotional interface software assisted by visible muscle movements in terms of emotion detection from mimics, i.e., the user's facial expression. The psychology driven idea behind detecting a user's facial expression is based on knowledge of previous psychological studies in the field of facial expression detection [2-3].

M. Kurosu (Ed.): Human-Computer Interaction, Part II, HCII 2014, LNCS 8511, pp. 779–787, 2014.

From the beginning chat was predominantly used for casual internet communication. In the meantime it has become a common and well-established form of communication. Among its advantages are its easy access, its low-tech affordances, high usability and fast and flexible communication processes. A variant of chat, SMS messaging, has met the public with unexpected popularity. Nevertheless, due to its media properties, chat-based communication frequently suffers from deficits due to incoherence of contributions, lack of coordination and problems of awareness concerning social awareness as well as awareness of context and available knowledge [4]. These shortcomings of conventional chat-based communication pose severe problems. In this paper we focus on affective or emotional aspects auf chat-based communication. We claim that these restrictions can be overcome and propose that extending the medium ‘chat’ with appropriate strategies of affective computing embedded in the chat environment can actually improve the chat discourse. In this context the effects of emotions in the chat environment have not been researched comprehensively yet. Computer-mediated communication usually does not contain information about the emotional state of a user during typing. Reports on the influence of emotions are mostly based on observations or interviews and do not compare to empirical methods. During interviewing, computer-mediated communication studies point out several comments in logs that participants were frustrated with the interview agent’s responses [5].

As a concrete example for implementing and discussing our approach in chat-based environments, we refer to a chat-based interview agent developed by Kanno and colleagues [6-8]. This chat-based interviewer agent was developed to deviate knowledge from the user. The concept is to create a kind of chat program that automatically generates questions and responses to the answers from an interviewee. The basic technique behind this is so-called artificial non-intelligence, i.e., the agent makes responses based on keywords identification and relatively simple rules like ELIZA [9]. Also a set of concepts was implemented and an interview guide as a database for this interviewer agent. The first prototype was developed and tested [10]. However, human users sometimes found the responses from the agent unfitting. To overcome the deficiencies of chat-based interview agents, our psychology driven approach might be applied as a possible chat-based emotional interface to reduce the number of unexpected responses and hence the irritations of the user, i.e., the interviewee.

This paper is about how emotions can be detected, and what kind of methods can be applied to extract users’ responses or affective states in a possible emotional interface within a chat environment. Further, this paper will give a brief overview of recent developments of affective methods in order to recognize the user’s response. After the development section, we will introduce our method of processing the facial responses, e.g., raised eyebrows, of the user and show early results [3] and [11]. Furthermore, our first results lead to a discussion on additional applications and limitations that frames an attempted approach of emotion detection in chat-based environments.

2 Emotion Recognition Method Development

The first step to recognizing affective aspects of communication in Human-Computer Interaction, such as in a chat-based interviews, involves detecting relevant emotional information, e.g., raised eyebrows, from non-relevant. During this process, the main problem is how to quantify this information in order to enable a computer to recognize the meaning in the data [12]. To relate our approach to other studies, we present several approaches and a variety of methods for analyzing communications: Affective Dialog Systems [12-14], Sentence-Based Emotion Recognition [15], and Multimodal Chat Emotion Recognition [16]. All these methods will be described in the following section.

2.1 Affective Dialog System (DS)

Finding relevant information is the basis of affective communication. Affective Dialog Systems can classify the information and they are an important tool for studying affect and social aspects in online communication. An Affective Dialog System is a social intelligence model, i.e. agents that handle affective responses with the help of psychological theories of personality, emotion, and Human-Computer Interaction [12].

Morishima and colleagues [12] argued that agent's socially appropriate affective communication provides a new dimension for collaborative learning systems. In our case, for example, an interview chat-agent can more efficiently interview the user.

Turkle [13] pointed even out that online communication has a huge effect on users' social and psychological perceptions and behavior and even their self-concepts. Skowron and colleagues [14] indicated that an affective system can influence the user in terms of chatting enjoyment, dialog coherence, and realism. Furthermore, the variants of the affective system strengthen the chatting enjoyment and emotional connection. For this reason, an Affective Dialog System can be an important input in the field of developing a dynamic questionnaire. However, the direct feedback, such as facial expressions, will not be included in affective text-based research and could be a drawback during interviewing a user.

2.2 Sentence-Based Emotion Recognition

To further classify emotions in the context of computer-mediated communication, a promising approach is textual emotion recognition. Krcadinac and colleagues [15] present an approach that analyses on the sentence level based on the standard Ekman emotion classification [2]. The developed algorithm reads a text sentence in a chat as an input and sorts it to the six emotional states defined by Ekman (i.e., happiness, sadness, anger, fear, disgust, and surprise). To study emotions in computer-mediated communication, a keyword-spotting method was developed by Krcadinac and colleagues based on a free, open source library software system Synesketch which includes a WordNet-based word lexicon; a lexicon of emoticons, common abbreviations and; colloquialisms, and a set of heuristic rules. During their study each of the

214 participants needed to rate 20 sentences randomly taken from the corpus to one for each emotional status.

The results indicated a high accuracy (~80%) that can lead to further promising future research and applications. However, this approach has two drawbacks one is that the corpus is quite reduced to a basic sentence level with relatively unambiguous emotional type and cannot compared with a fluent chat communication, such as in a dynamic questionnaire. Second, it cannot recognizing neutral as a separate type. In our research we try to include our feature extraction method so that a communication can be analyzed in direct feedback of the participants face.

2.3 Multimodal Chat Emotion Recognition

However, as the use of emoticons and text-only analyses suggests, communication without nonverbal analysis, such as facial expressions, can be monotonous [16]. Another possible way to include emotions in the chat environment is to create a 3D Avatar by extracting the facial actions of each participant with real-time facial expression analysis techniques and research on synthesizing facial expressions and text-to-speech capabilities. Chandrasiri and his colleagues have created a piece of software that creates a 3D facial animation of agents [16]. Their system includes visual, auditory, and primary interfaces to communicate as one multimodal chat interaction. Participants can represent themselves as predefined agents. During the experiment, for example, a user showed facial expressions while typing text in the chat. The represented 3D agent will speak the message aloud while it repeats the recognized facial expression and also replay the synthesized voice with proper emotional pronunciation.

The biggest advantage of the software is that the visual data exchange requires only low bandwidth and, therefore, works in real time. The disadvantage is that the software need a person-specific initialization and several interfaces. Furthermore, in recent years videoconferencing tools have become more popular in our daily lives, e.g., Skype and MSN. The user might be more convinced to use simple live stream video communication software rather than an avatar software for getting more direct facial expression feedback.

3 Software

In relation to the above mentioned approaches, this paper suggests a new psychology driven dynamic approach to detect emotions. In 1979 Bassili [17] suggested that even with minimal information about the spatial arrangement of features, participants can recognize facial expressions. Another interesting approach was presented by Kaiser and colleagues [18]. Here, in order to reduce the complexity, small dots were placed on the participant’s face which were then detected and analyzed on videotapes. The dots allow the algorithm to determine the underlying muscle activity.

Earlier approaches attempted to detect a set of typical emotions, such as happiness, surprise, anger, and fear [19]. In contrast to those studies, we focus on the detection of emotion or no emotion present with the less complex feature dots tracking method [3], [11], and [20]. Thus, we developed a computer program that might be able to detect ten blue dots placed on the participant's face. The positions of the points were derived from earlier psychological studies, investigating facial muscle movement with the help of Electromyography (EMG) [21], as well as observation of human mimics [22]. Our software locates the blue dots by searching each frame of the video, line by line. We selected blue as the color for the dots, since blue is present only minimally in the color-spectrum of the human face.

Within the chat environment, our analysis might help to improve the interview communication of a chat agent and a user by detecting the unexpected responses of the chat agent with a facial expression detection.

3.1 Software Development

In our previous studies we recorded several videos of the participants' face [3] and [11]. The study was performed with N = 59 (40 female) participants with an average age of M = 23.39 years (SD = 4.51), who acted the emotions of anger, happiness and no expression. The recorded videos were rated for further analyses by independent raters with regard to the emotional content of the facial expressions.

In a further step the facial expression videos of the participants were analyzed and the area (A-D) and distances (1-3) between the selected blue dots were calculated (see Figure 1).

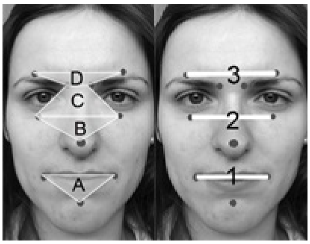

Fig. 1. Faces with dots and their areas (A-D) and distances (1-3)

Then the algorithm summed up each area and distance and then calculated the arithmetic mean with their variance. Consequently, the variance clearly reveals the motion of each area or distance. The following table shows the ideal state of the development of each emotion over time (see Table 1).

For example, the area of “A” shrinks during expressing the emotion anger (closed mouth) and grows during showing the emotion anger (open mouth), since the participant opens their mouth. The preliminary results in Tables 2-3 result from how similar the values correspond to the ideal values in Table 1. An up-arrow means the area or distance is growing larger and a down-arrow means the area or distance is getting smaller over time. The minus signs represent no detectable dot movements, i.e., the participant is in a neutral state. Additionally, we split angry emotions into open and closed mouth types for a more efficient analysis. Preliminary results showed significant changes in the mouth area enclosed by the selected blue dots when participants experienced anger with an open and closed mouth. Finally, the results of the variation of each area and distance are displayed in tables.

Table 1. Dynamic variation of each area (A-D) and distance (1-3) in relation to each emotion

	A	B	C	D	1	2	3
Anger (open mouth)	↑	↑	—	↓	↓	—	↓
Anger (closed mouth)	↓	↓	↑	↓	↓	↓	↓
Happy	↑	↓	↑	↓	↑	↓	↕
Neutral	—	—	—	—	—	—	—

3.2 Software Testing

We reported first software testing results in Tews [3] and [11] within psychological facial expression studies, mostly in an automobile context. Due to technical dropouts only n = 10 subjects' data could be analyzed. Because of the small database, we refrained from using inferential statistics and our results are only displaying descriptive statistics. The tables 2 and 3 show the results of the angry with open (oM) and closed mouth (cM) and no facial expressions video analysis. We excluded the results of the happiness condition, because they were similar to the angry emotions results. To standardize the results, the average of all emotion feature values are set to 100% so that the deviation from each emotion can be expressed as a percentage. The emotion values are shown on the x-axis, describing how strongly each emotion was expressed. The participants' relations are shown in separate columns and displays each participant. For the angry facial expression videos, the raw data of the participants does not show any clear results (see Table 2).

Table 2. Results of the angry facial expression videos

	P0	P1	P2	P3	P4	P5	P6	P7	P8	P9
Anger (oM)	34	38	8	1	0	2	4	2	8	8
Anger (cM)	35	33	8	1	28	2	0	1	8	2
Happy	29	27	14	1	28	2	2	1	14	12
Neutral	2	2	70	97	44	94	93	96	70	78

The values were broadly spread because of the mimic activities caused by the emotions. In contrast to participants (P0) and (P1), participant (P3) showed only few emotional expressions (see Table 2).

Table 3. Results of the no facial expression videos

	P0	P1	P2	P3	P4	P5	P6	P7	P8	P9
Anger (oM)	**0**	**0**	**0**	**9**	**3**	**0**	**0**	**1**	**1**	**0**
Anger (cM)	**1**	**2**	**1**	**4**	**3**	**1**	**1**	**0**	**1**	**0**
Happy	**1**	**0**	**1**	**9**	**2**	**1**	**1**	**0**	**1**	**1**
Neutral	**98**	**98**	**98**	**78**	**92**	**98**	**98**	**98**	**97**	**99**

Compared to the emotional facial expression results, the neutral values are explicit, because of lack of movement in the face, as it is expected from a neutral face (see Table 3). The values are concentrated within the neutral emotion row. Another interesting point is the noise-induced error, especially participant (P3) shows the typical error with few values within the emotion states.

In conclusion, our study yielded the promising result that our approach was able to distinguish between an emotional state and no emotional expression. This might be used for example as a chat-based emotional interface for an interview agent to improve the communication with a user.

4 Conclusion

Chat-based internet communication has become a common form of communication. Nevertheless, chat-based communication frequently suffers from deficits due problems of awareness concerning social perceptions as well as context and available knowledge [4]. To address those problems, in this paper we focused on a possible chat-based emotional interface for an interview agent to improve the communication with a user. By reducing the number of unexpected responses, the chat-based interview agent can, for example, adapt and response to the user more dynamically for a more efficient communication. As a concrete example for implementing and discussing our approach in chat-based environments, we referred to a chat-based interview agent developed by Kanno and colleagues [6-8] and [10].

Within this paper we presented previously developed affective communication methods and compared them with a possible usage in reliable dynamic interface assisted by visible muscle movements, i.e., Affective Dialog Systems [12-14], Sentence-Based Emotion Recognition [15], and Multimodal Chat Emotion Recognition [16].

In contrast to the involved multimodal interface method, Multimodal Chat Emotion Recognition, our approach tries to reduce the complexity of affective detection by extracting the features with a simplified feature dots detection method. Though the Affective Dialog Systems and Sentence-Based Emotion Recognition methods were less involved, the main drawback was the text-only analysis without the direct feedback of the participants face.

To detect the affective expressions with our new psychology driven dynamic approach, we placed ten dots on the face of the participant. By analyzing the movement of blue dots, our software can help to distinguish the participants' facial expressions by discriminating the neutral and the emotional state. The new measurement, the dynamic variance of areas and distances was implemented to distinguish the participants' states.

Results showed that the variance of an area and distance defined by distinct dots can support the affective detection. Though our study has some limitations, our methods indicate promising results that our program could be tested in the chat environment. Our future research will also include the extended collection of data of affective expressions in the chat environment, in relation to the responses of a chat agent.

Acknowledgments. This research was supported by JSPS Pre/Postdoctoral Fellowship and FS-Nr. 2006.63 from the 'Arbeitsgruppe Innovative Projekte' (AGIP) of the Ministry of Science and Culture, Lower Saxony, Germany. Many thanks go to Mark Hammonds for his support in style and editing of English language.

References

1. Reisenzein, R., Hudlicka, E., Dastani, M., Gratch, J., Hindriks, K., Lorini, E., Meyer, J.-J.C.: Computational Modeling of Emotion:Toward Improving the Inter- and Intradisciplinary Exchange. IEEE Transactions on Affective Computing 4(3), 246–266 (2013)
2. Ekman, P., Friesen, W.V., Hager, J.: The Facial Action Coding System (FACS): A technique for the measurement of facial action, Palo Alto (1978)
3. Tews, T.-K., Oehl, M., Siebert, F.W., Höger, R., Faasch, H.: Emotional human-machine interaction: Cues from facial expressions. In: Smith, M.J., Salvendy, G. (eds.) Human Interface, HCII 2011, Part I. LNCS, vol. 6771, pp. 641–650. Springer, Heidelberg (2011)
4. Oehl, M., Pfister, H.-R.: E-Collaborative Knowledge Construction in Chat Environments. In: Ertl, B. (ed.) E-Collaborative Knowledge Construction: Learning from Computer-Supported and Virtual Environments, pp. 54–72. IGI Global, New York (2010)
5. Kumar, R., Rosé, C.P., Wang, Y.C., Joshi, M., Robinson, A.: Tutorial dialogue as adaptive collaborative learning support. Frontiers in Artificial Intelligence and Applications 158, 383 (2007)
6. Ochi, Y., Kanno, T., Furuta, K.: An Interview Agent for Cognitive Task Analysis. In: Proceedings of Human-Agent Interaction Symposium 2010, 3C-1 (2010) (in Japanese)
7. Ochi, Y., Kanno, T., Furuta, K.: An Interviewer Agent for Cognitive Task Analysis. In: Proceedings of Human Interface Symposium 2011, 1441L, pp. 381-390 (2011) (in Japanese)
8. Kanno, T., Ochi, Y., Chou, T., Furuta, K.: Service Cognition Probe Techniques. In: Proceedings of the 3rd Symposium on Systems Innovation, pp. 51–53 (2011) (in Japanese)
9. Weizenbaum, J.: ELIZA—a computer program for the study of natural language communication between man and machine. Communications of the ACM 9(1), 36–45 (1966)

10. Kanno, T., Uetshuhara, M., Furuta, K.: Interviewer agent for cognitive task analysis. In: Stephanidis, C., Antona, M. (eds.) UAHCI/HCII 2013, Part I. LNCS, vol. 8009, pp. 40–49. Springer, Heidelberg (2013)
11. Tews, T.-K., Oehl, M., Faasch, H., Kanno, T.: A Survey of Dynamic Facial Emotion Detection in Emotional Car Interfaces. In: FAST-Zero 2013 Proceedings, JSAE Paper 20134654, No.TS2-5-4, JSAE - Society of Automotive Engineers of Japan, Tokyo (2013)
12. Morishima, Y., Nakajima, H., Brave, S., Yamada, R., Maldonado, H., Nass, C., Kawaji, S.: The role of affect and sociality in the agent-based collaborative learning system. In: André, E., Dybkjær, L., Minker, W., Heisterkamp, P. (eds.) ADS 2004. LNCS (LNAI), vol. 3068, pp. 265–275. Springer, Heidelberg (2004)
13. Turkle, S.: The second self. In: Simon & Schuster, New York (1984)
14. Skowron, M., Theunis, M., Rank, S., Kappas, A.: Affect and Social Processes in Online Communication—Experiments with an Affective Dialog System. IEEE Transactions on Affective Computing 4(3), 267–279 (2013)
15. Krcadinac, U., Pasquier, P., Jovanovic, J., Devedzic, V.: Synesketch: An Open Source Library for Sentence-Based Emotion Recognition. IEEE Transactions on Affective Computing 4(3), 312–325 (2013)
16. Chandrasiri, N.P., Naemura, T., Ishizuka, M., Harashima, H., Barakonyi, I.: Internet communication using real-time facial expression analysis and synthesis. IEEE Multimedia 11(3), 20–29 (2004)
17. Bassili, J.N.: Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. Journal of Personality and Social Psychology 37, 2049–2058 (1979)
18. Kaiser, S., Wehrle, T.: Automated coding of facial behavior in humancomputer interactions with FACS. Journal of Nonverbal Behavior 16, 67–83 (1992)
19. Tian, Y.-I., Kanade, T., Cohn, J.F.: Recognizing action units for facial expression analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(2), 97–115 (2001)
20. Tews, T.-K., Oehl, M., Siebert, F.W., Höger, R.: Emotional Interfaces in Cars: Cues from Facial Expressions. In: de Waard, D., Gérard, N., Onnasch, L., Wiczorek, R., Manzey, D. (eds.) Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society Europe Chapter, Human Centred Automation, pp. 111–122. Shaker Publishing, Maastricht (2011)
21. De Luca, C.J.: The use of surface electromyography in biomechanics. Journal of applied biomechanics 13, 135–163 (1997)
22. Cohn, J.F., Ekman, P.: Measuring facial action by manual coding, facial EMG, and automatic facial image analysis. In: Harrigan, J.A., Rosenthal, R., Scherer, K. (eds.) Handbook of Nonverbal Behavior Research Methods in the Affective Sciences, pp. 9–64. The Oxford University Press, New York (2005)