User Study of AffectIM, an Emotionally Intelligent Instant Messaging System

Alena Neviarouskaya[1], Helmut Prendinger[2], and Mitsuru Ishizuka[1]

[1] University of Tokyo, Department of Information and Communication Engineering, Japan
lena@mi.ci.i.u-tokyo.ac.jp, ishizuka@i.u-tokyo.ac.jp
[2] National Institute of Informatics, Japan
helmut@nii.ac.jp

Abstract. Our research addresses the tasks of recognition, interpretation and visualization of affect communicated through text messaging. In order to facilitate sensitive and expressive interaction in computer-mediated communication, we previously introduced a novel syntactical rule-based approach to affect recognition from text. The evaluation of the developed Affect Analysis Model showed promising results regarding its capability to accurately recognize affective information in text from an existing corpus of informal online conversations. To enrich the user's experience in online communication, make it enjoyable, exciting and fun, we implemented a web-based IM application, AffectIM, and endowed it with emotional intelligence by integrating the developed Affect Analysis Model. This paper describes the findings of a twenty-person study conducted with our AffectIM system. The results of the study indicated that automatic emotion recognition function can bring a high level of affective intelligence to the IM application.

Keywords: Affective sensing from text, affective user interface, avatar, emotions, online communication, user study.

1 Introduction and Motivation

The essentialness of emotions to social life is manifested by the rich history of theories and debates about emotions and their nature. Recently, the task of recognition of affective content conveyed through written language is gaining increased attention by researchers interested in studying different kinds of affective phenomena, including sentiment analysis, subjectivity and emotions. In order to analyse affect communicated through written language, researchers in the area of natural language processing proposed a variety of approaches, methodologies and techniques [2,8-11].

Advanced approaches targeting at textual affect recognition performed at the sentence-level are described in [1,3,4]. The lexical, grammatical approach introduced by Mulder et al. [4] focused on the propagation of affect towards an object. Boucouvalas [1] developed the Text-to-Emotion Engine based on word tagging and analysis of sentences. An approach for understanding the underlying semantics of language using large-scale real-world commonsense knowledge was proposed by Liu et al. [3], who

H. Prendinger, J. Lester, and M. Ishizuka (Eds.): IVA 2008, LNAI 5208, pp. 29–36, 2008.

incorporated the created affect sensing engine into an affectively responsive email composer called EmpathyBuddy.

Peris et al. [7] argues that online chats may stimulate rather than inhibit social relations, and chat users seem to find a media for rich, intense, and interesting experiences. The motivation behind our research is to enrich social interactivity and emotional expressiveness of real-time messaging, where a machine is used as a communication channel connecting people and transmitting human emotions. Here, a key issue is to provide the automation of multiple expressive means so that the user does not have to worry about visual self-presentation as in standard Instant Messaging (IM) systems, but can focus on the textual content of the conversation. While constructing our Affect Analysis Model we took into account crucial aspects of informal online conversation such as its specific style and evolving language [5].

The remainder of the paper is structured as follows. In Section 2 we shortly introduce the developed Affect Analysis Model. We describe the developed IM application integrated with the Affect Analysis Model and analyse the results of a user study in Section 3 and Section 4, respectively. In Section 5 we conclude the paper.

2 Rule-Based Approach to Affect Sensing from Text

We proposed a rule-based approach to affect sensing from text at a sentence-level (details are given in [6]). The algorithm for analysis of affect in text consists of five stages: (i) symbolic cue analysis, (ii) syntactical structure analysis, (iii) word-level analysis, (iv) phrase-level analysis, and (v) sentence-level analysis. The salient features of this algorithm are: (1) analysis of nine emotions and five communicative functions on the level of individual sentences; (2) the ability to handle the evolving language of online communications; (3) foundation in affect database; (4) vector representation of affective features of words, phrases, clauses and sentences; (5) consideration of syntactic relations in a sentence; (6) analysis of negation, modality, and conditionality; (7) consideration of relations between clauses in compound, complex, or complex-compound sentences; and (8) emotion intensity estimation.

An empirical evaluation of the Affect Analysis Model algorithm [6] showed promising results regarding its capability to accurately classify affective information in text from an existing corpus of informal online communication. In a study based on blog entries, the system result agreed with at least two out of three human annotators in 70% of the cases.

3 Instant Messaging Application Integrated with the Affect Analysis Model

The AffectIM, an Instant Messaging system with emotional intelligence, was developed as a web-based application running in the Internet browser. Within our research project, we could design only two avatars, one male and one female. So the graphical representative is automatically selected by the system according to the user's sex.

The main window of AffectIM system while online conversation is shown in Fig. 1. From the list of friends displayed in the left frame, the user selects the person (available online), whom he or she wishes to communicate with. The central frame allows user to type and to send the messages. It displays the conversation flow in three modes: plain, transcribed, and with emotions. Further, it displays emotional avatars (own – to the left of conversation field, and friend's – to the right). Two buttons located under the avatar animation refer to the visualization of emotion distribution (either in a color bar or pie graph) and emotion dynamics (line graph). Since the language of online communication is constantly evolving, AffectIM also provides the functionality to add new abbreviations, acronyms, and emoticons to the Affect database (see two buttons located to the left from the input text field).

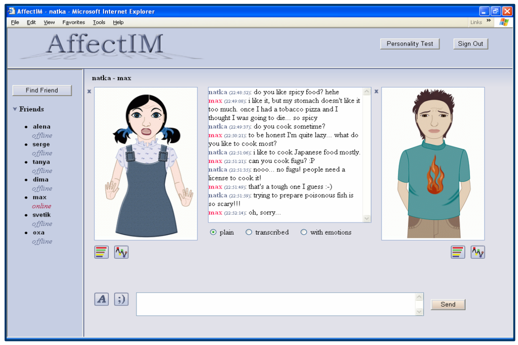

Fig. 1. AffectIM interface

4 User Study of the AffectIM System

The purpose of the user study was to evaluate "richness of experience" and "affective intelligence" of our AffectIM system. We hypothesized that user experience and effectiveness of the communication of emotions may benefit from introduction of automatic emotion recognition function and emotionally expressive avatars to IM application, as opposed to manual selection of emotional behavior or uninformed, random display of affect. Our hypotheses are tested by considering the following dimensions regarding users' experience: (1) Interactivity. (2) Involvement (engagement). (3) Sense of copresence. (4) Enjoyment. (5) Affective intelligence. (6) Overall satisfaction. In addition to these main criteria, we asked participants to give us feedback on some general questions.

4.1 Experimental Design, Subjects and Procedure

The experiment was designed as a within-subjects experiment in pairs. In particular, we compared three AffectIM interfaces using different configuration conditions.

For the user study, we prepared three versions of the system:

1. Automatic (A-condition). In this interface, affect sensing from text is performed based on the developed Affect Analysis Model, and the recognized emotions are conveyed by the avatar expressions.
2. Manual (M-condition). During this condition, no automatic emotion recognition from text is performed; however, users may select emotion (and its intensity) to be shown by avatars using "select pop-up menus".
3. Random (R-condition). Here, the avatars show a 'quasi-random' reaction. First, we process each sentence using the Affect Analysis Model, and then we apply

two rules: (1) if the output is emotional, we run two functions that randomly select the emotion out of nine available emotions and its intensity, correspondingly; (2) for the case of "neutral" output, we set the function that generates "neutral" emotion with the probability of 60% or "random" emotion with the probability of 40%.

It is important to note that in each of three interfaces five communicative functions are automatically recognized and shown by avatars. In other words, the occurrence of communicative behavior is not varied across the three experimental conditions.

Twenty university students and staff (10 males, 10 females) took part in our study. All of them were computer literate, and 19 persons had prior experience with computer based chat or Instant Messaging system.

Each pair of participants was composed by male and female subjects. Before the IM session, all participants were given instructions and their AffectIM IDs and passwords. Each pair of participants was asked to have online conversations through three interfaces given in random order. After each interface condition, users filled the corresponding page of the questionnaire in and commented on their experience.

After the participants completed the IM communication about the three topics and corresponding questionnaire, they were asked to answer some general questions about their experience with the IM system.

4.2 Analysis of Results

The average duration of sessions on each interface was 10.1 minutes (minimum 8 and maximum 12.5 minutes), excluding the time needed to fill out the questionnaires.

The 11 questions on main criteria were answered based on 7-item agreement Likert scale. Since our study involved each subject being measured under each of three conditions, we analyzed data using statistical method ANOVA (two-factor ANOVA without replications with chosen significance level $p < 0.05$).

The **interactivity** was measured using statement "*The system was interactive*". Subjects tended to consider the condition, which allowed them to manipulate the expressed emotion manually, as most interactive. However, ANOVA resulted in no significant difference in interactivity among three interfaces.

The **involvement** was evaluated using two questionnaire items: "*I felt it was important for my conversation partner that I responded after each his/her statement*" and "*I was awaiting the replies of my conversation partner with true interest*". The ANOVA results showed that the reported involvement of all three systems does not differ significantly, showing that the level of engagement was almost the same.

The following two questionnaire items covering the aspects of space and togetherness are intended for evaluation of **sense of copresence**, or social presence: "*I felt if I were communicating with my conversation partner in the shared virtual space*", "*The system gave me the sense that the physical gap between us was narrowed*". The statistic ANOVA results for the first questionnaire item indicated the significance of the difference in sense of copresence felt in A-condition and R-condition (p (R-A) < 0.05), and showed that the A-condition gave stronger feeling of communication in the shared virtual space than the R-condition. No significant difference among three interfaces was reported on the second statement.

The level of **enjoyment** was evaluated using the statement "*I enjoyed the communication using this IM system*". The high levels of enjoyment were reported during A-condition and M-condition. However, ANOVA resulted in no significant differences among all three IM interfaces.

To evaluate **affective intelligence**, four statements (three – directly related to the system and one – indirectly related) were proposed to subjects in questionnaire: "*The system was successful at conveying my feelings*", "*The system was successful at conveying my partner's feelings*", "*The emotional behavior of the avatars was appropriate*", and "*I understood the emotions of my communication partner*". Fig. 2 shows the bar graphs of means of questionnaire results for these statements.

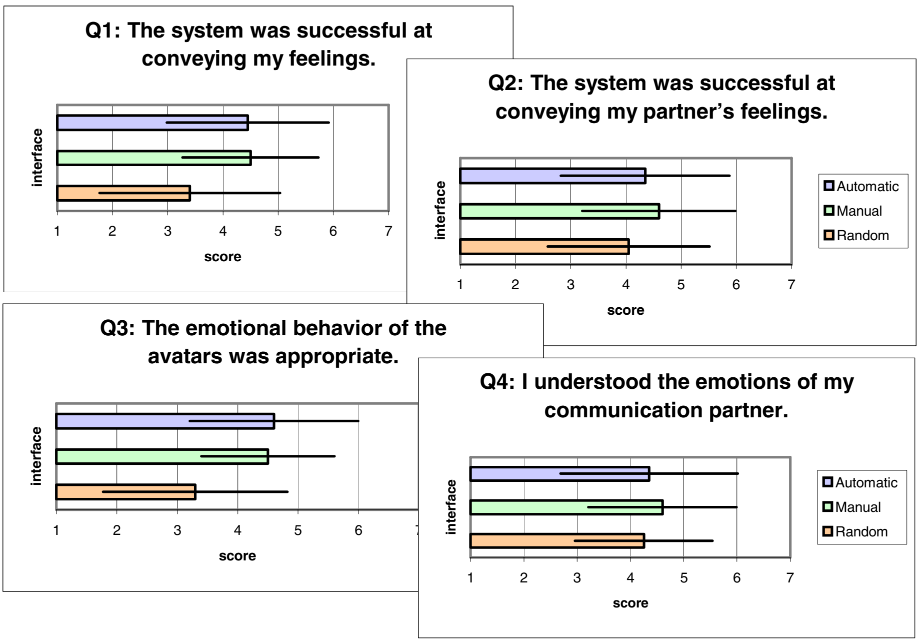

Fig. 2. Questionnaire results on affective intelligence (Q1, Q2, Q3, Q4)

As seen from the graph bar for Q1 (Fig. 2), the systems in A-condition and M-condition (with small prevalence of mean results in M-condition) were both more successful at conveying own feelings than the system in R-condition. Since M-condition is considered as a "gold standard" in communicating person's emotions, and ANOVA showed no significant difference between M-condition and A-condition, we might say that automatic emotion recognition system performed well enough to bring high affective intelligence to IM application. As was expected, significant differences were found between R-condition and M-condition (p (R-M) < 0.05), and between R-condition and A-condition (p (R-A) < 0.01).

While evaluating successfulness of the interfaces at conveying conversation partner's feelings (see Q2 in Fig. 2), the highest rate was given by subjects to M-condition, and the lowest – to R-condition. However, ANOVA for this criterion resulted in no significant difference among all interfaces. One user's comment regarding the emotional reactions of the partner's avatar was: "I concentrated too much on the reactions of my avatar and not enough on that of my partner. Reading and thinking about the answer took away the concentration on the avatar".

Interesting results were observed for the evaluation of appropriateness of emotional behavior of avatars. As seen from the graph (Q3 in Fig. 2) and statistical data of ANOVA, results for A-condition and M-condition significantly prevailed those for R-condition (p (R-A) < 0.01; and p (R-M) < 0.01). Users' comments confirmed that during R-condition subjects sometimes couldn't understand why the avatars did not correspond to their words and reacted in "wrong" ways. Although A-condition was rated a little bit higher than M-condition, no significant difference was detected between these interfaces.

The statement "*I understood the emotions of my communication partner*" measured affective intelligence of the system indirectly, since people used to derive emotional content from text based on semantic information and their empathetic abilities. Emotional expressions of avatars may help to understand the partner's emotion clearer. As was expected, the highest rate was reported in M-condition, and the lowest – in R-condition, where participants might be confused, since sometimes emotions shown by the avatar contradict actual emotional content (see Q4 in Fig. 2). However, no significant difference was found in partner's emotion comprehension among all three interfaces. A possible explanation for such results might be that a person typically relies on his/her own affective intelligence rather than on results of artificial affective intelligence. That is why the mean for R-condition appeared relatively high.

The **overall satisfaction** from using three AffectIM interfaces was evaluated using statement "*I am satisfied with the experience of communicating via this system*". Regarding the results, average scores for A-condition and M-condition were equal (4.6), whereas less satisfaction was reported for R-condition (4.25). The results of ANOVA showed no significant difference in overall satisfaction among interfaces.

In addition to the main questionnaire items, participants were given general questions. Subjects were asked to associate nine emotion states with nine avatar expressions shown on still figures. Female avatar was shown to male subjects, while male avatar was shown to female subjects. The percentages of reported correct associations within males and females are shown in Fig. 3. As seen from the graph, all 10 female subjects correctly associated 'anger', 'joy', and 'shame' emotions, while all 10 male subjects completely agreed only on 'joy' emotion. The detected pairs of most often confused emotions are 'fear' – 'surprise' and 'guilt' – 'sadness'; and less often confused emotions are 'guilt' – 'fear' and 'sadness' – 'fear'. Some participants confused emotions in 'interest' – 'joy' and 'surprise' – 'guilt' pairs. These results suggest that during the experiment some participants faced the difficulty with correct interpretations of emotional behavior of avatars.

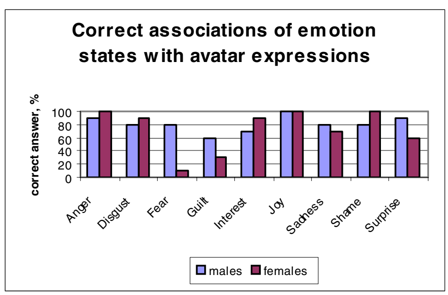

Fig. 3. Questionnaire results on emotions associated with avatar expressions

To the question "*While online, do you use emoticons or abbreviations?*", 19 subjects answered positively. We observed all automatically recorded dialogs, and found that to some degree the majority of participants used abbreviated language.

The participants' comments and the results of answers to the question "*To what degree do you think is necessary to look at a graphical representation of the other communicating person?*" suggest that there are two types of IM users: (1) some are open to new features of IM, and find animated graphical representation of a person helpful in understanding the partner's emotions and giving some sense of physical presence; (2) others tend to concentrate their attention on content, and prefer small emotional symbolic cues, like emoticons, to avatar expressions.

Participants also were asked to indicate whether manual selections of emotion state and intensity were helpful or not during M-condition. Only 30% of males and 60% of females answered positively. The result of answers to the question "*How often did you use this function, when you wanted?*" is represented as a bar graph in Fig. 4. As seen from these data, female subjects used emotion selection function more ardently than male subjects.

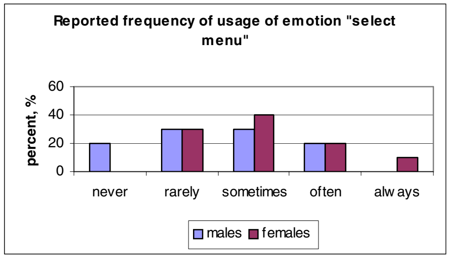

Fig. 4. Questionnaire results on frequency of usage of emotion "select menu"

The user opinions regarding the emotion "select menu" aspect were very diverse. Some users criticized the type of pop-up menu, commenting that it was difficult to use and it took long time to select. For more convenience, they proposed to replace pop-up menus by icons and spread them out. One of the subjects complained that emotion select menu disturbed the flow of the chat. Another reported problem is that since there is no preview of what the emotion expression looks like, it is unclear whether it matches the user's intention. Some subjects felt that basic emotions are too general and are not sufficient to convey emotion in many cases. Also, they suggested providing the possibility of showing more different or even mixed emotions (some state between sadness and joy). However, we think that displaying mixed emotional expressions would add more confusion and misinterpretation to the conversations.

Some subjects underlined positive aspects of manual selection of emotion states. They found this function helpful, because (1) it offered the possibility to visually express feelings and better understand them, (2) it allowed preventing inappropriate emotional reaction of avatar, and (3) guaranteed accuracy of communicated emotion. We can conclude that for sensitive conversation users would prefer manual control to avoid system mistakes that could sometimes harm the conversation.

5 Discussion and Conclusions

We implemented a web-based IM application, AffectIM, and endowed it with emotional intelligence by integrating our Affect Analysis Model. The user study conducted on AffectIM showed that the IM system with automatic emotion recognition function was successful at conveying users' emotional states during communication online, thus enriching expressivity and social interactivity of online communications. From the experiment we learned that the IM application might benefit from an integration of automatic emotion sensing with manual control of emotional behavior of avatars in one interface, which will allow users to select between two modes depending on type and sensitivity of conversation.

While analyzing the recorded conversations from our study, we detected some misspelled emotion-related words: "feiled" instead of "failed"; "promissing" instead of "promising", etc. In our future work, we plan to add correction of misspelled words to the system. Moreover, we aim to study cultural differences in perceiving and expressing emotions, and to integrate a text-to-speech engine with emotional intonations into the developed IM application.

Acknowledgments. We acknowledge and thank Dr. Ulrich Apel and Dr. Boris Brandherm for their efforts and help in organization of the user study on AffectIM. We wish also to express our gratitude to all the participants of the experiment.

References

1. Boucouvalas, A.C.: Real Time Text-to-Emotion Engine for Expressive Internet Communications. In: Being There: Concepts, Effects and Measurement of User Presence in Synthetic Environments, pp. 306–318. IOS Press, Amsterdam (2003)
2. Kim, S.-M., Hovy, E.: Automatic Detection of Opinion Bearing Words and Sentences. In: Proceedings of IJCNLP 2005 (2005)
3. Liu, H., Lieberman, H., Selker, T.: A Model of Textual Affect Sensing using Real-World Knowledge. In: Proceedings of IUI 2003, pp. 125–132 (2003)
4. Mulder, M., Nijholt, A., den Uyl, M., Terpstra, P.: A Lexical Grammatical Implementation of Affect. In: Proceedings of the 7th International Conference on Text, Speech and Dialogue, pp. 171–178. Springer, Berlin (2004)
5. Neviarouskaya, A., Prendinger, H., Ishizuka, M.: Analysis of Affect Expressed through the Evolving Language of Online Communication. In: Proceedings of IUI 2007, pp. 278–281. ACM Press, New York (2007)
6. Neviarouskaya, A., Prendinger, H., Ishizuka, M.: Textual Affect Sensing for Sociable and Expressive Online Communication. In: Paiva, A.C.R., Prada, R., Picard, R.W. (eds.) ACII 2007. LNCS, vol. 4738, pp. 220–231. Springer, Heidelberg (2007)
7. Peris, R., Gimeno, M.A., Pinazo, D., et al.: Online Chat Rooms: Virtual Spaces of Interaction for Socially Oriented People. CyberPsychology and Behavior 5(1), 43–51 (2002)
8. Strapparava, C., Valitutti, A., Stock, O.: Dances with Words. In: Proceedings of IJCAI 2007, Hyderabad, India, pp. 1719–1724 (2007)
9. Subasic, P., Huettner, A.: Affect Analysis of Text Using Fuzzy Semantic Typing. IEEE Transactions on Fuzzy Systems 9(4), 483–496 (2001)
10. Turney, P.D.: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: Proceedings of ACL 2002, USA (2002)
11. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. In: Proceedings of HLT/EMNLP 2005, Vancouver, Canada (2005)