Affective Communication in Online Chat Using Physiological Sensors and Animated Text

Hua Wang [*1,2], Helmut Prendinger[*1], Mitsuru Ishizuka[*1], Takeo Igarashi[*3,4]

Abstract - We present a chat system that uses animated text associated with emotional information to show the affective state of the user. The system obtains the affective state of a chat user from a physiological sensor attached to the user's body. This paper describes preliminary experiments and provides examples of possible applications of our chat system. Observations from informal experiments comparing our animated chat system with a conventional system suggest that an online interface that conveys emotional information enriches online communication between remote users.

Keywords : Physiological information, animated text, kinetic typography, affective computing, online chat

1. INTRODUCTION

Online communication has become a major tool worldwide, but it still has many shortcomings. The online interfaces in use today, such as chat systems, are far more primitive than face-to-face conversation, making it difficult to convey many basic social cues, such as emotions.

The successful deployment of computer-supported communication systems requires the consideration of social factors, such as getting to know other people's feelings, emotions, *etc*. Birdwhistell's [1] linguistic analogy suggests that the information conveyed by words amounts to only 20-30% of the information conveyed in a conversation. These observations demonstrate the importance of non-verbal information, such as emotions, which are essential for human cognition and influence different aspects of peoples' lives. For example, when we are excited, our perception is biased toward the selection of happy events, while the opposite is true for negative emotions. This indicates that online communication would benefit from knowledge of the other's affective state.

There are many attempts to enhance computer-mediated human-to-human interaction by communicating affective state [2]. The most primitive technique is to use the so-called "smiley" in text based communication. Other attempts include the detection of affective state from typing speed or typing pressure [3].

*1: Graduate School of Information Science and Technology, The University of Tokyo
*2: Industrial Engineering Department, University of Toronto
*3: Department of Computer Science, The University of Tokyo
*4: PRESTO/ JST

This paper proposes a chat system that obtains a user's affective state using physiological information of the user and displays the derived affective state by using animated text. Other researchers have already investigated individual techniques. The Conductive Chat system [4] is an instant messenger client that measures a user's GSR (galvanic skin response) and reflects that information in the text the user types. For instance, if GSR data becomes higher, the size of static texture message will become bigger. Bodine et al. [5] applied kinetic typography to Instant Messaging.

We combine these two approaches in order to communicate affective state more effectively and efficiently. By using animated text, users can express their affective state more naturally compared to static textual messages. By using physiological signals, users can express their affective state without manual control, and it is even possible to communicate affect that users are not aware of by themselves.

We use Lang's two-dimensional emotion model [6] to implement the mapping from the physiological sensor information to the emotionally modulated animations. Based on psychophysiological evidences that the GSR increases with arousal, we assign GSR data to the arousal dimension of the model. On the other hand, since valence is difficult to obtain using physiological sensors, the system uses the user's manual control to indicate the valence of the current affective state.

2. RELATED WORK

2.1 Chat Interface

A number of graphical chat interfaces have been developed in recent years. Microsoft Comic Chat [7] allows users to chat in the context of a comic book. Participants are in the role of characters in a story, and their conversations are expressed as lines in this story. Since users cannot control the qualities

associated with their words, all sentences are expressed with uniform loudness and quality of voice.
In Microsoft V-chat [8], a virtual chat environment provides a user with a virtual place that appears real. Conversations are conducted using text chat alone; neither pitch nor loudness can be conveyed.

ChatScape [9] is an informal on-line communication tool that uses short pieces of text and still images. The images provide visual hints to communicate context. However, the system remains limited to still images and cannot provide dynamic environmental context.

Chat Circles [10] uses circles to graphically represent participants. This method represents social activity by altering the size of the circles and their colors. Again, however, plain text is used for the actual conversation, thereby limiting the quality of information conveyed.

2.2 Animated text

Animated text, also known as kinetic typography, refers to text messages that dynamically change their appearance such as location and shape. The earliest known use of animated text occurred in films and TV advertising. There, animated text is used because of its capability to convey emotive content, and thus meets the goals of advertising. Ford *et al*. [11] showed that animated text was effective in conveying a speaker's tone of voice, affect and emotion. Ishizaki [12] demonstrated that, in some cases, animated text could explicitly direct or manipulate the attention of the viewer. Lee *et al*. [13] improved kinetic typography libraries. Cues of activity might also be explored with personal dynamic fonts such as Gromala's BioMorphic Typography [14]. The Kinedit System (Forlizzi *et al.* [15]) provides a way to create and control kinetic typography and provides users with convenient methods to control the size, font, and placement of textual characters.

2.3 Affective Computing Interface

Recently, there have been several attempts to build user models that consider emotion and affective state. Klein *et al*. [16] reported an interactive system that responds to the user's self-reported frustration while the user is playing a computer game. Prendinger *et al*. [17] derived the user's emotional state by monitoring physiological signals when interacting with a character-based interface. The study demonstrates that the use of an interface agent with empathic behavior may significantly decrease user stress. Vyzas *et al*. [18] studied the physiological changes that occurred in a human actor when she intentionally induced eight different emotional states. Their classifiers achieved an 80% success rate when discriminating among eight emotions. The project of Vital Signs [19] used a realistic task in which one person teaches another person how to write several Chinese characters, whereby users wear SC (skin conductance) and BVP (blood volume pulse) sensors on one hand and HR (heart rate) sensor around the waist.

There have been some studies on the detection of affective state, but none has applied this information to online communication systems, such as online chat. Our approach detects emotion in real-time and applies it to an online chat interface. By showing the animated chat associated with the users' affective information, users can determine their chat partners' affective states. One might argue that one can easily communicate affective information in video chat. We focus on text messages because they still have a number of advantages over videos, such as simplicity, small data size or privacy concerns.

3. DETECTION AND PRESENTATION OF AFFECTIVE STATE

Figure 1 depicts an overview of our online chat system. The system estimates the affective state of the user by using data from physiological sensors and manually specified animation tags. Then, the estimated affective state is presented to other users as animated text[1].

A detailed description of our system is presented below.

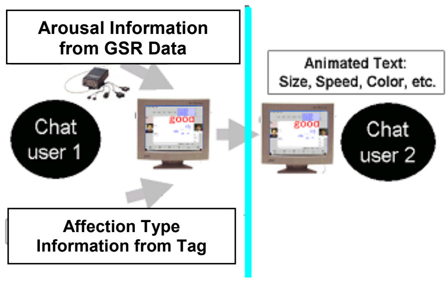


Figure 1 System overview

Lang [6] proposed a two-dimensional emotion space where one axis represents affective arousal and the other represents affective valence (Figure 2). Arousal indicates the intensity of an emotion, and often increases with stress. Emotional valence shows whether the emotion is positive or negative. Negatively valenced emotions include fear or sadness, and positively valenced ones are happiness, excitement, etc. (see the discussion in Affective Computing [2]). Ideally, we would like to determine the affective state of the user from

[1] The communication of the affective state should be two-way, but our current implementation is one-way because we presently have only one signal encoder unit.

physiological sensors only, without user intervention. However, as we will discuss in the next paragraph, it is difficult to obtain valence information from physiological sensors. Therefore, we provide a manually specified animation tag to obtain valence information (emotion type: happy, sad, etc.) and use physiological data to detect the emotional arousal (intensity of emotion: size, speed change, etc.)

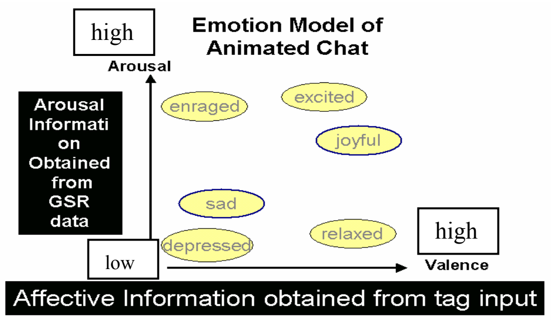


Figure 2 Lang's 2D model in our system

Arousal and valence can be determined from galvanic skin response (GSR) data and valence from blood volume pulse (BVP). BVP increases with negatively valenced emotions, such as fear and anxiety, and decreases with relaxation (Picard [2]). However, BVP is too sensitive to obtain meaningful data and thus it is difficult to detect affective valence. Since GSR data are more robust against noise, we decided to use the GSR sensor to detect the arousal and to use manual control to specify the valence. GSR sensors that measure skin conductivity (SC) are attached to the middle and index fingers of the user's non-dominant hand, and the signals are recorded with the ProComp+ unit [20]. SC varies linearly with the overall level of arousal and increases with anxiety and stress.

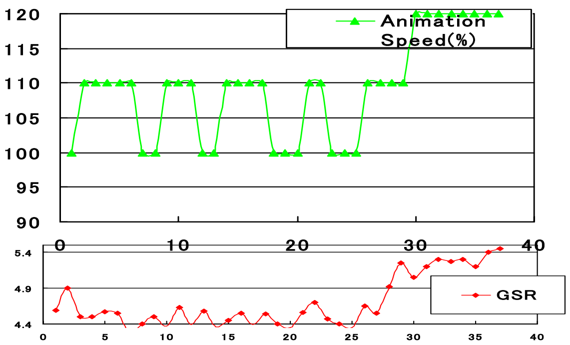


Figure 3 The mapping from GSR data to animation speed.

In addition to GSR data, we employ user-specified animation tag input to obtain affective valence information. Affective valence decides the animation type to use, such as happy or sad and emotion arousal decides the speed, size, color, and the motion of animated text. When GSR data suddenly increase and show a peak, text animation simultaneously has a higher pace, increases in size and becomes lighter in color. When GSR data decrease, the animation becomes slower, the size is reduced, and the color darkens. To be more specific, the system computes the difference between the current GSR level and the normal GSR level obtained in a calibration phase. If the difference exceeds a predefined threshold, the system increases font size and animation speed by step-by-step to filter out the noise in the GSR input (Figure 3).

We implemented 20 types of animation (Figure 4), which the user specifies by using a button, a short-cut key, or a tag embedded in a text message. The tags before sentences direct the text animations. For example, "<happy> I am happy!" presents "I am happy!" with happy motion.

Here we describe some major text animations. There are two kinds of animation. The first consists of animation used to express the feelings of a user. The second consists of animation used to convey physical information more clearly.

4. ANIMATIONS

4.1 Emotional Animations

"Chat" is used as the default case in conversations. When a user writes a new sentence, the words in the sentence scroll up from the bottom. "Happy" is used when a user wants to express the feeling of happiness or cheerfulness. Words jumping up and down at a high speed indicate happiness. "Question" is used when a user wants to ask a question or show that he or she is suspicious. First, a question mark appears to indicate that there is a question. Then, the actual question appears. "Escape" is used when users are unsure of their opinion and state something without confidence. The remarks appear in a small font from the right and then they quickly move back to the right, as if they were escaping from the screen. "Sad" is used to express sadness. The words in the sentence become progressively smaller and all the words fall down to the bottom, as if they are feeling low and sinking off the screen. "Cheer" is used to express joyful emotions. The words in the sentence become progressively bigger and brighter, as if they are cheerful. "Whisper" is used to express a whisper. The words in the sentence are distinctly tiny and the pace is slow, as if one user were whispering to another.

4.2 Structural Animations

We describe some major structural animations here. "List" is used when a user wants to show a list in a conversation. The items in the list emerge from the left of the screen. "Title" is used when a user wants to express a word as a title. The animation looks like the scene for the title in a movie, with the word appearing from the right, stopping in the center of the page briefly, and then disappearing. “Attention” is used when a user wants to attract other chat users’ attention to his or her remarks. The words are initially small, and then increase in size until they fill the screen.

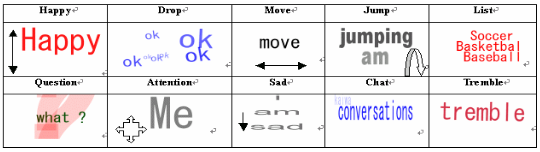
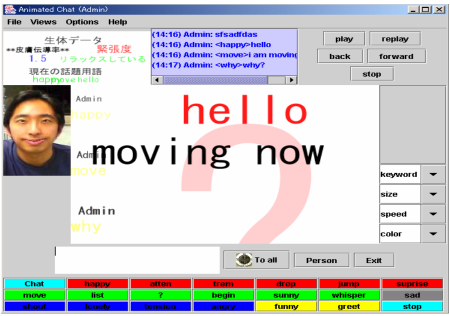


Figure 4 Examples of animations

5. IMPLEMENTATION

Our system uses a two-dimensional graphical window to display chat animation and information (Figure 5).



F

Figure 5 Animated chat system

5.1 Animation Chat Screen

This shows animations of specific words or sentences. The movement of kinetic text changes according to the emotional state detected from the tags and GSR data. (Figure 4)

5.2 Face Avatar

A user can use a face avatar, which can simulate real chat. When an emotional animation is used or changed, the pictures or avatar will change accordingly. To illustrate, when a user selects a happy animation, the face will become happy.

5.3 Scrolling Text Log

A chat log of the conversation is also displayed in plain text. When a word or sentence is accompanied by an animation, the type of animation is shown using a tag. For example, the chat log might say “<happy>I am happy”. Thus, the chat log stores all the information of the conversation, including the animation. This log allows the user to replay the conversation and see the same animation as shown earlier. It also facilitates cutting-and-pasting in order to recreate animations.

5.4 Affective State Monitor

The screen shows the physiological data obtained from the GSR and provides a means of evaluating whether the chat partner is aroused or calm. The state of arousal is measured from the GSR. When the GSR indicates that the partner is in a state of high mental tension, “Very tense” is displayed on the screen. When the user is in a state of low mental tension, “Relaxed” is shown on the screen.

6. INFORMAL TESTS AND PRELIMINARY FINDINGS

We performed two preliminary experiments using our system. The first was performed during a short conversation, and the second was conducted in an online educational setting. Detailed descriptions of each test are given below.

6.1 Short Conversations

We conducted a preliminary usability study of our system. Six subjects participated in the preliminary experiment. The subjects were college students who used both our system and the conventional Microsoft Messenger. The subjects were assigned to use both the chat systems at a random order. They were in the same building but were not supposed to see each other. They were separated into three pairs and talked with the partner. Each chatted for more than one hour. They had conversations about school courses, sports, *etc*. After a conversation, the subjects answered questionnaires and commented on the system.

we examined the GSR data and the reactions of the subjects. They were asked when they felt most involved in the chat, and we compared their answers with the GSR results. There was a good correlation between the GSR data and the user-reported tension. The subjects reported that they concentrated on the conversations gradually, and the GSR showed similar changes

This indicated that the GSR can be used to determine changes in mental tension in real time during an online conversation. The results also suggest that emotional information might be able to increase the subject’s involvement in the conversation. The emotional information from the chat partner gave users the feeling that they were not only exchanging textual information, but were also communicating their feelings to each other, which allowed them to become more involved in the conversation. One drawback that some subjects mentioned was that they did not always want others to see their GSR information. Table 1 shows some of their answers.

Table 1: Comparison of the systems 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) (6 subjects).

Question	EF	non-EF	Standard Deviation (EF)	Standard Deviation no-EF
I became very involved in the conversation	4.7	4.0	0.47	0.58

I enjoyed using the system	4.5	4.2	0.50	0.55
I feel know my partners' feelings	4.2	3.5	0.61	0.50
I was more interested in talking	4.2	4.0	0.49	0.58
It is easy to find my partner's emotions	3.8	3.5	0.37	0.50

Table 1 shows that the chat system with emotional feedback influenced the subjects' impressions of the other users' emotional states. This may make them feel more eager to talk with their partner and allow them to become more involved in the conversation.

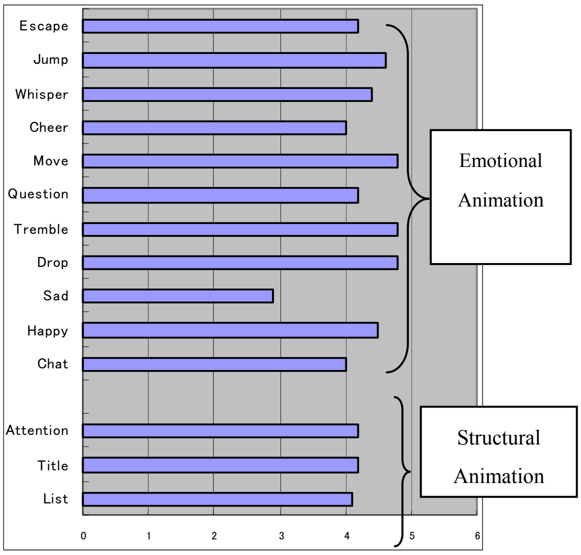


Figure 6 Evaluation of each animation

Figure 6 shows that users evaluated animation positively. Some animations, like happy and drop, were ranked very highly. Users often stated that these animations contained more dynamic movements and were capable of conveying stronger and more effective visual impact. We conclude that animation is an effective method of providing chat users with tools to communicate information in a richer visual format.

6.2 Teacher-Student Interaction in Online Education

We also tested our system in an online education setting. There is already some research on the use of facial expressions in online education. Nakatsu *et al*. [21] claim that by sharing non-verbal information during online education, teachers can interact with students more easily. We tested teacher-student interactions to determine the usefulness of emotional information in an educational setting. We used online education because it is an application in which knowledge about the user's emotion essentially contributes to the interlocutor's ability to respond appropriately, by tailoring the interaction so as to promote learning, while maintaining a high level of engagement.

One teacher and one student took part in the experiment. The physiological sensor was attached to the student, so that the teacher could observe the student's affective state during the session. The teacher presented information about the theme "cartoon history" and asked the student questions about the topic. They were in the same building but were not able to see each other. The online class lasted for about 40 minutes. The subjects answered questionnaires and commented on the system. We also analyzed the chat log that accompanied the GSR records. The teacher tried to use the physiological information to gauge the student's reactions. We found that if the teacher suddenly became very strict, the student's GSR data changed rapidly. (Figure 7)

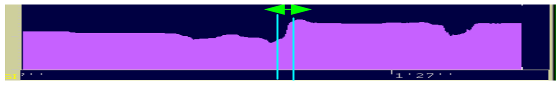


Figure 7 Sudden arousal revealed in a student's GSR (the graph plots the student's GSR every minute)

The subjects also suggested possible uses for this system. Their suggestions can be classified into two categories: 1) Real-time feedback regarding others' tension in a conversation can be useful. From the degree of tension in the conversation, the teacher can identify the student's involvement, *e.g.*, whether they are concentrating or bored. 2) Feedback can be used later to compare the student's motivation in different classes. By monitoring the changes in their students' concentration, teachers can learn which topics are of interest.

7. DISCUSSIONS AND FUTURE WORK

We also explored other approaches for obtaining information on emotion, such as typing speed and auditory feedback. Typing speed can be a supplemental when it is inconvenient to use GSR data. When users are anxious or in a hurry, their typing speed tends to increase. In our system, the text animation can become faster and thereby give other users graphical feedback. In addition, we also tried sound to show the affective state of the chat user. In our system, the features of auditory feedback are associated with mental arousal, which is detected by the GSR signal. For instance, when a user is anxious or stressed, a vibrating sound is played to indicate his or her mood. Alternatively, a peaceful sound can be played to regain their calmness.

During the preliminary trials, we found that people tended to read their chat partner's physiological data and tried to influence their partner. They felt that our system provided more feedback on emotion and that this made it more engaging than plain text chat. Most people were interested in knowing whether their partner was nervous, and most said that they enjoyed the chat with accompanying physiological information more than that without it. This system has some limitations. It is not convenient to wear and carry the sensors, and there is still much noise in GSR data. However, we believe that miniaturization and advanced technology of the physiological sensors will make them more convenient to wear and will provide more reliable results. There are also some concerns regarding privacy; some people indicated that sometimes they did not want other people to see their GSR data, and they turned off the signal in order to feel comfortable.

The system could also be used in fields such as long-distance health care and bio-feedback. The state of a patient can be determined in real-time from data on their online state together with physiological information. Patients could also become familiar with their own mental and physical states by comparing their own information logs for different sessions or comparing them with those of other people.

Currently, we are investigating methods for recognizing patterns in the physiological signals in online chat. In addition, we are investigating how to remove noise that is generated by external factors. We hope to develop an expression library for our chat system and a sub-system that allows users to create new animations. We also hope to further identify differences in how people from different cultures perceive the animations used. We also plan to analyze video data recorded during the online chat sessions to search for correlations between the video and physiological data. In particular, by integrating animated text, physiological, and video data, we hope to create a new multimodal chat system that can demonstrate users' affective states more efficiently and conveniently.

REFERENCES

[1] Birdwhistell, R. Kinetics and Context: Essays on Body Motion Communication*,* University of Pennsylvania Press, 1970.

[2] Picard, R., Affective Computing. Cambridge: MIT Press, 1995.

[3] Yamada, Y. Hirano, T. and Nishimoto, K., Tangible Chat: Communication of con-versation situation awareness using a sense of touch in a key-board chat system, Tech. Report SIG-GW-43-18 Information Processing Society of Japan, pp.103-108.

[4] Morris, J., DiMicco, V. Lakshmipathy, A. Fiore, T., "Conductive Chat: Instant Messaging With a Skin Conductivity Channel." Poster Presentation, Conference on Computer Supported Cooperative Work (CSCW '02). New Orleans, LA, November 2002.

[5] Bodine, K., and Pignol, M. Kinetic typography-based instant message, In Proceedings of CHI03, Demonstration, 2003.

[6] Lang, P. J., The emotion probe: Studies of motivation and attention. American Psychologist, 50(5), 1995, pp 372.

[7] Kurlander, D., Skelly, T., Salesin, D. Comic Chat, *Proceedings of SIGGRAPH 96*, 1996, pp 225-236.

[8] MicrosoftVChat, http://www.microsoft.com/ie/chat/vchatmain.htm.

[9] Ayatsuka, J., Matsushita, S., Rekimoto, J. ChatScape: a visual informal communication tool in communities, *CHI 2001 Extended Abstracts,* Apr. 2001, 327-328.

[10] Viegas, F.B., Donath, J.S. Chat circles*. In Proceedings of CHI99, ACM Press*, 1999, pp 9-16.

[11] Ford, S., Forlizzi, J., and Ishizaki, S. Kinetic typography: Issues in time-based extended abstracts, In Proceedings of CHI97 Conference, Extended Abstracts, 1997, pp 269-270.

[12] Ishizaki, S. On Kinetic Typography Statements, American Center for Design Journal Vol. 12 No.1, 1998.

[13] Lee, J., Forlizzi, J., and Hudson, S. The kinetic typography engine: an extensible system for animating expressive text, In Proceedings of UIST02, 2002, pp 81-90.

[14] Gromala, D. BioMorphic Typography. [On-line] Available: http://www.lcc.gatech.edu/~gromala/ excretia/.

[15] Forlizzi, J., Lee, J., and Hudson, S. Kinedit system: Affective messages using dynamic texts, In Proceedings of CHI, ACM Press, 2003, pp 377-384.

[16] Klein, J., Moon, Y., and Picard, R. This computer responds to user frustration: Theory, design, and results. Interacting with Computers, 2002, pp 119-140.

[17] Prendinger, H., Mori, J., and Ishizuka, M. Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. International Journal of Human-Computer Studies, Vol. 62, No. 2, 2005.

[18] Vyzas, E., Recognition of Emotional and Cognitive States Using Physiological Data, Mechanical Engineer's Degree Thesis, MIT, June 1999.

[19] Lyons,M., Kluender,D., Chan,C., Tetsutani,N., "Vital Signs: Exploring Novel Forms of Body Language", Proc. of ACM SIGGRAPH 2003, 2003.

[20] ProComp UNIT, Thought Technology Ltd., http://www.thoughttechnology.com/.

[21] Nakatsu, R., Tadenuma. M., Maekawa, T., Computer technologies that support Kansei expression using the body, In Proceedings of 9th ACM Multimedia, 2001, pp 358-364.