Improvement in Multi-User Communication Systems Using an Avatar's Facial Expression Features*

Takashi KUSUMI*[1], Hidetsugu KOMEDA*[1],*[2] and Takatsugu KOJIMA*[1]

*[1] *Graduate School of Education, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501, Japan*
*[2] *Currently with University of Wisconsin-Madison*

We investigated how an avatar's facial expression function affects users' motivation to engage in conversation and ability to understand a partner's emotion during online chat communication and attempted to improve the chat system based on the usability ratings of this function. In Experiment 1, 15 pairs of university students chatted in English using both an emoticon-input facial expression system and a non-facial expression system. The evaluation of the systems' usability indicated that the emoticon-input system was not superior to the non-facial expression system. The emoticon-input system was thus modified to an icon-input system to improve its usability. In Experiment 2, 16 pairs of students chatted in Japanese using both the icon-input facial expression system and the non-facial expression system and guessed their partner's emotional states during the chat. Participants found it easier to guess their partner's emotions when using the facial expression system. We discuss the outcomes and limitations of our experiments for future research.

Key words : avatar, computer-mediated communication (CMC), emotion, Multi-User Dungeon (MUD), usability

1. INTRODUCTION

Our purpose was to implement emotional expression features in educational systems that use three-dimensional (3-D) Multi-User Dungeons or Multi-User Dimensions (MUDs) over a network system. We also (a) assessed the effect of an avatar's facial expression function on users' motivation to engage in conversation and ability to understand a partner's emotion during online chat communication and (b) improved the system based on an evaluation of the usability of the emotional expression function.

We used an MUD system in which each user chats while being represented by an avatar in virtual Internet space. Examples of MUD use in higher education include PennMOO at the University of Pennsylvania and Athena University, a virtual university. SchMOOze University is particularly famous in the field of foreign language education, allowing to chat English learners freely using their keyboards (Takizawa 1995); it is used in Japan at schools such as Chubu University

(Awaji 2000). Unlike these MUDs, which employ text chat, the Three-Dimensional Interactive Education System (3D-IES, Nomura Research Institute) that we used is an educational MUD system that employs 3-D graphics to accentuate the feeling of live chat through virtual space and avatars to represent participants (Hamabe 2000). This system offers virtual spaces such as Rosenberg Place for the study of German or the Champs-Elysees for the study of French. These spaces allow students to move freely within a realistic environment that is suitable for studying a foreign language and to engage in conversation or perform skits. Suzuki (2000), Okano (2000), and Kusumi, Koyasu, and Nakamura (2002) have used this system to conduct English and German classes in general education and academic English classes in University; they observed results such as improved student motivation and performance.

Habitat is a Japanese MUD that has been used since the 1990s as a communication space with amusing characteristics. At start-up, it used two-dimensional graphics and later evolved to Habitat 2, which uses 3-D graphics; it operates today under the name of J Chat (G-Search 2006). Sakamoto, Isogai, Kimura, Tsukamoto, Kasuga,

and Sakamoto (2000) have used the anonymity of Habitat to conduct experimental research on training programs to overcome shyness. Researchers in Japan and overseas have investigated various aspects of communication systems that incorporate avatars (Cassell *et al.* 2000; Takeuchi and Katagiri 2000).

Our research focused on the educational use of communication systems that use 3-D MUDs. The use of MUDs can yield three main advantages. One is that students have the freedom to create an avatar to represent them, allowing them to maintain their anonymity and talk enthusiastically without the fear of making mistakes. These benefits are especially significant for shy students who are unable to speak confidently in real life (Adachi 1999). Second, users can flexibly set realistic scenes and roles for the characters that appear in them and can communicate at a high level of realism through not only verbal, but also non-verbal, communication using the expressions and actions of avatars. This may enable users to transfer their experiences in the virtual world to a real-life environment, similar to the way that foreign language students can acquire a higher level of skill when they shift from a pattern-based practice to a more communicative approach. Third, the use of MUDs can enable studies to be conducted collaboratively with students in distant areas (or even in other countries) beyond the walls of the classroom and can help to develop computer-mediated communication (CMC) skills.

Educational practices are beginning to incorporate the advantages of 3-D MUDs (e.g., Renninger and Shumar 2002), but little research has been conducted on (a) how adding expressive features to avatars could affect a user's motivation to engage in conversation or ability to understand another participant's emotions during communication or (b) what improvements to the system interface are required to make it more user-friendly. Researchers have evaluated how the use of avatar expression may benefit communications related to mediation or negotiating in conference systems (Yuasa 2003) and mental health (Rizzo *et al.* 2001), but little research has been conducted in the educational field.

In this research, we implemented two input methods of avatar expression features in a communication system: emoticon input and icon-input. We explored (a) how these affected the motivation of students to engage in conversation during the communication process and their ability to understand their partner's
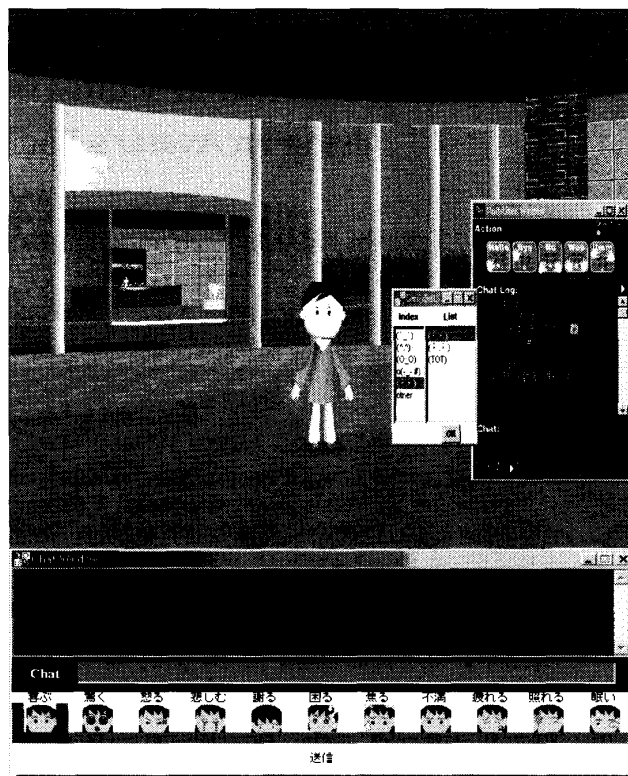

Fig. 1.   A Three-Dimensional Interactive Communication System (3D-ICS) display and its input systems (Nomura Research Institute).
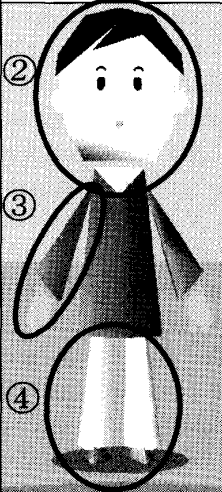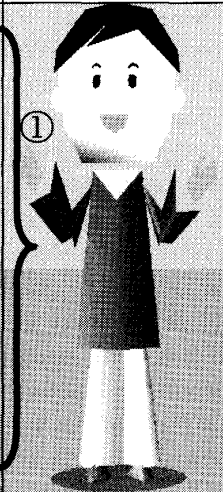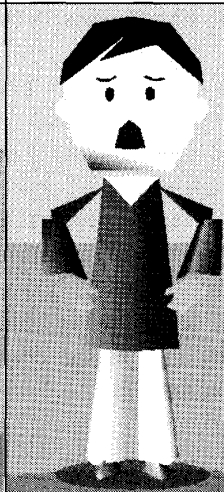
The emoticon-input system (top) displays a chat window on the right. The user enters text in the field at the bottom and selects an emoticon face from the middle window (Experiment 1).

The icon-input system (bottom) displays a chat window at the top. The user enters text in the middle field and selects icons from the bottom. The position on the Happy bar indicates the level of intensity (Experiment 2).

emotions and (b) how to improve the system, based on ratings of the usability of the features.

We used the Three-Dimensional Interactive Communication System (3D-ICS, Nomura Research Institute), a program that incorporates the emotional expression features of avatars to the 3D-IES mentioned above (Figure 1). Both 3D-IES and 3D-ICS are software packages (Hamabe 2000) that were developed based on the technologies of the Community Place Browser (Lea *et al.* 1997). They each consist of a teacher terminal, student terminals (3-D multi-user clients), and a central server (3-D multi-user server). The server processes the interactive communication according to requests from both the teacher and student terminals and provides a 3-D virtual space. To exchange text and images, basic 3-D virtual space models that use virtual reality modeling language (VRML) and communications software programs are stored in the client terminals, and only changes in the 3-D environment (participant actions, movements, or conversations) are

Table 1. Frequency of emoticon use and examples of expressions and actions

Emotional category	Normal	Happy	Surprised	Angry	Sad
Emoticons (Total frequency of use)[1]	(·_·)　　　(74)	(^.^)　Weak　(73) o(^o^)o Moderate (25) \(^o^)/Strong (28)	(O_O)　　Weak　(8) w(° o°)w Moderate (9) \(|,o_|)/　Strong　(1)	o(-_-#)　Weak　(6))˘ε´(　　Moderate (0) (▼▼)　　Strong (0)	(; ;)　Weak　(11) (;.;)　Moderate (5) (TOT)　　Strong (1)
Example of change in avatar expressions and acrtions[2]					

Note: [1] Numbers in parentheses indicate frequency.
　　　[2] Numbers in circle indicate movable parts of avatar.

exchanged on the server. This method achieves a multi-user environment and reduces the network load.

The 3-D multi-user client is a web browser that displays 3-D space written in VRML that participants use to login to a virtual 3-D space. In contrast, the 3-D multi-user server maintains connections (information such as the location and attributes of each avatar) from multiple clients and manages the transmission of information shared in the 3-D space. This enables clients to chat while observing the expressive actions of each other's avatars (Hamabe and Matsuda 2002).

The avatars used in this study were formed from combinations of geometrical objects: spheres and cylinders. They can represent actions and facial expressions by stretching and shrinking (in various aspect ratios), moving, and turning. Movable parts include: ① body (head, right/left arms, right/left legs, and torso); ② face (hair, eyebrows, eyes, mouth, and facial outline); ③ arms (upper arm, lower arm, and hands); and ④ legs (feet and thighs; Table 1). On button entry, the normal state (as indicated in row 2) changes for 1.3 s (surprised), 2.0 s (anxious), 2.5 s (happy), and 2.6 s (angry, sad, sorry, troubled, discontented, tired, embarrassed, and sleepy; Table 1). The intensity of the emotion and the degree of action are set so that weak emotions are represented only through facial expressions, moderate emotions are represented through facial expressions and actions involving the upper body, and strong emotions are represented through facial expressions and actions involving the entire body.

We used this system to conduct two experiments. In Experiment 1, participants chatted in English using 3D-ICS, in which avatars express emotion using text-based emoticons. We evaluated (a) how this system affected the communication process, as indicated by the number of utterances and the motivation of participants to engage in conversation, and (b) the usability of the emotional expression feature. In Experiment 2, we directed participants to chat in Japanese in a more common environment, using 3D-ICS, in which avatars express emotions using icons. We evaluated how this program affected the communication process, as indicated by the number of utterances and the motivation of participants to engage in conversation, and the ability of participants to understanding their partner's emotions. We then assessed the usability of emotional expression through icons. Here, the term "usability" refers to the ease with which participants used the emotional expression features of avatars that we had installed and the ease with which participants could identify the emotions expressed by avatars. We improved the system based on the usability ratings. Our participants were university students because the

system was developed for use in college classes for foreign language students and because most students are familiar with using emoticons when e-mailing and chatting online.

2. EXPERIMENT 1:
EVALUATION OF THE EMOTICON-INPUT SYSTEM

In Experiment 1, users could express emotion through their avatars by selecting an emoticon from a menu bar and inputting it at the end of a chat text. This system was based on the widespread practice of adding an emoticon at the end of a text chat or e-mail to express emotion, which is difficult to represent with text alone. We then evaluated the usability of the emotional expression feature by comparing conditions with and without avatar expression.

2.1. Methods
2.1.1. Participants

A total of 30 participants (16 males, 14 females), all of whom were undergraduate or graduate students at Kyoto University, were grouped into same-sex pairs; no participants were familiar with their partner. Participants ranged in age from 19 to 30 and from sophomore undergraduates to graduate students; all possessed the ability to chat in English. Each pair was separated by a partition that ensured that they could not see each other, and all participants wore headphones so that they could not hear any voices. This was to ensure the anonymity of participants and to prevent conversations in the virtual world from affecting relationships in the real world after the experiment.

2.1.2. Experimental system

Participants could see a virtual world, with the front view of their partner's avatar and the back of their own avatar in a window at the top-center of the monitor (Figure 1). When participants typed into the text field at the bottom of the window (Figure 1, right-hand side) and pressed Enter, an emoticon window appeared (Figure 1, middle). Participants could select one of 11 categories of emoticon (happy, surprised, angry, sad, sorry, troubled, anxious, discontented, tired, embarrassed, and sleepy) and one of three levels of intensity (weak, medium, and strong). The system required that participants select an emoticon at the end of every chat sentence and then changed the avatar's facial expression and action accordingly.

The system's 11 categories of emotion

correspond to the four basic emotions as defined in psychology (happy, surprised, angry, and sad) plus seven emotions and moods (sorry, troubled, anxious, discontented, tired, embarrassed, and sleepy) that are frequently used in Internet chat sessions.

The procedure for selecting emoticons was as follows. We referred to websites with multiple emoticons and selected candidates that seemed to be used frequently. We then selected different emoticons for each level of emotional intensity (weak, moderate, and strong), especially for the four basic emotions. We set selection criteria whereby the design of the emoticons would go from simple to intermediate to complex as the level of emotion increased. We also referred to websites with multiple emoticons and assigned appropriate emoticons that are customarily used for strong emotions. Based on these criteria, three people, including the first author and the system developer, decided on the assignment of emoticons through consensus. Emoticons were assigned for each of the seven supplementary emotions and moods. A total of 18 adults (nine males, nine females) participated in a pilot experiment in which they matched the 11 emotions with the selected emoticons. Emoticons that were correctly identified 44-93% of the time were used in the subsequent experiment. Two emoticons were correctly identified $\leq$ 50% of the time: ((´ 0 `)~3 (tired, 44%) and (╱_-) (sleepy, 45%). We wanted to include these emoticons because we thought it was important for the system to represent as many basic emotions as possible and because no other appropriate emoticons represented these two supplementary emotions and yielded a higher percentage of correct answers. Because we were unable to confirm that there were no problems in their identification, we tried to rectify the problem so they could be used in the experiments, using the following procedure. The experimenter presented participants who understood the emoticons with a list of emoticons, emotion categories, and levels of intensity, together with a description of each emoticon; e.g., (´ 0 `)~3 (sighing); (╱_-) (rubbing one eye). Participants were asked to match these with actions through trial runs. The experimenter allowed participants to refer to the list during the experiment.

2.1.3. Experimental design

The experiment included two conditions: with and without facial expression. In the "With Facial

Expression" condition, avatar expressions and actions changed corresponding to emoticon selection. The "Without Facial Expression" condition supported text input only, with no change in avatar expressions or actions. Participants used both conditions, and the order in which the conditions were presented was balanced.

2.1.4. Tasks

Chat. Two topics were selected to encourage casual conversation: a conversation about where to go for dinner, and a conversation about travel plans. Participants were given written and verbal instructions to use as many different emotional expressions as possible, to adopt a contrary response instead of agreeing immediately, and to engage in heated exchanges. The chat time was set at 15 min, and combinations of experimental conditions were balanced. These chats were conducted in English.

Evaluations. After the chat under each condition was complete, the experimenter requested participants to evaluate the system by rating 13 items using a five-point scale (1 = *disagree*, 5 = *agree*); these 13 items are listed below, grouped into five general categories.

a. Virtual reality (2 items): Was able to chat actively in English while using the avatar in the virtual world; The experience was like conversing in English in the real world. The reliability coefficient (α) was 0.40 and 0.30 for With Facial Expression and Without Facial Expression, respectively.

b. Anonymity (2 items): Was not embarrassed to say impertinent things because I knew nothing of my partner or those around me; Was not embarrassed to say impertinent things because neither my partner nor anybody around me knew who I was (αs = 0.88, 0.86).

c. Emotional communication (2 items): Was able to communicate my feelings easily to my partner; Was able to understand my partner's feelings easily (αs = 0.82, 0.70).

d. Motivation to engage in conversation (3 items): The English conversation became lively; The English conversation was fun; Was able to concentrate on the content of our English conversation (αs = 0.83, 0.82).

e. Other (4 items): The method of input was tedious; Felt confident about actual English conversation; This system promotes my English communication skills; This system promotes my online chat skills.

Participants were then asked to provide written comments on a free-form answer sheet about characteristics of the With Facial expression and Without Facial Expression conditions: their usability, ease of English conversation, how much fun it was, and the effectiveness of the system for learning a foreign language. Participants were also asked what improvements they thought might make the system better in terms of use. They were also requested to use a five-point scale (1 = *disagree*, 5 = *agree*) to rate 26 items about shyness (e.g., I feel uneasy among people; questions were based those of Imai and Oshimi, 1987), three items about computer skills (e.g., I am good at blind-touch; questions were based those of Sakamoto *et al.* 2000), and two items about their use of emoticons (I use emoticons when texting on my cell phone; I use emoticons in e-mails).

2.2. Results and discussion

We analyzed the chat logs of all participants by condition. In total, we analyzed 363 utterances from the Without Facial Expression condition and 365 utterances from the With Facial Expression condition. Each chat utterance included an emoticon at the end; for example,

Don't worry (^.^)
Really? (?_?)
O.K. I'm fine, too. \(^o^)/

Table 1 provides a partial list of the frequency of each emotion category expressed using emoticons. The most frequent emotion category was "Normal (no action)," followed by "Happy (weak)." This was due to the fact that the conversational topics mostly produced utterances that did not require special emotional expression or that related to comfortable subject matter concerning food and travel. One survey (Ministry of Internal Affairs and Communications, 2003) found that topics discussed in online communities mostly relate to hobbies or have no specific theme, so we consider that the utterances produced in these conversations reflect the characteristics of daily chat with a real-life partner.

We compared the With Facial Expression and Without Facial Expression conditions (Table 2). There were no significant differences in the number of utterances, but the Without Facial Expression condition produced more words per utterance than the With Facial Expression condition.[1] Because the two conditions did not differ in the number of utterances, it seems that

178   T. KUSUMI *et al.*

Table 2. Comparison of means (*SDs*) from the With Facial Expression and Without Facial Expression conditions in the emoticon input system.

Measure	With Facial Expression	Without Facial Expression	*t* value (*df* = 29)
Number of utterances	12.1 (6.39)	12.2 (7.46)	0.12 *p* = .900
Number of words	68.3 (21.4)	74.9 (26.0)	2.11 *p* = .044
Sense of virtual reality[†]	2.98 (0.79)	2.77 (1.06)	1.17 *p* = .252
Anonymity[†]	3.05 (0.96)	3.58 (1.09)	2.52 *p* = .017
Emotional conveyance[†]	3.13 (1.52)	2.82 (0.90)	1.42 *p* = .166
Motivation to converse[†]	3.61 (0.78)	3.38 (0.98)	0.99 *p* = .330

Note: [†]Mean of five-point scale (1 = *disagree*, 5 = *agree*).

communication was possible with fewer words because the system allowed communications using action. Users particularly tended to use an avatar's facial expressions instead of using emotional terms such as "happy" and "sad." This is discussed later in Section 3.2. and Table 3. We examined the utterances and words used in both conditions by testing differences between means to see if any differences appeared in the number of words used by gender, high or low levels of shyness, and high or low levels of self-evaluated computer skills. As a result, no significant differences were observed. The only item that differed between the two conditions was Anonymity, which had a higher mean under the Without Facial Expression than the With Facial Expression condition (Table 2). The Pearson correlations between "Motivation for conversation" and the number of utterances and words, respectively, were 0.04 and 0.27 for the With Facial Expression condition and 0.08 and 0.08 for the Without Facial Expression condition (all *ps* > 0.1).

Therefore, our analysis of usability items did not indicate that the With Facial Expression condition was better than the Without Facial Expression condition in the emoticon input system. There was also the problem of a low α coefficient for "virtual reality." To improve our emoticon input system, we conducted an additional analysis, focusing on single items and testing the differences in the mean ratings between With Facial Expression and Without Facial Expression conditions. For the item "Was able to understand my partner's feelings easily," participants indicated that feelings were easier to understand in the With Facial Expression condition than in the Without Facial Expression condition (*Ms* = 3.23, 2.73, *t* (29) = 2.19, *p* = 0.037). Participants also indicated they "Felt confident about actual English conversation" in the With Facial Expression condition more than in the Without Facial Expression condition (*Ms* = 2.70, 2.40, *t* (29) = 2.34, *p* = 0.020). A further comparison of the Pearson correlation coefficients between items in the With Facial Expression and Without Facial Expression conditions revealed a relationship between the ratings "Our English conversation became lively" and the number of utterances and words in the With Facial Expression condition (*rs* = 0.41, 0.46, *ps* = 0.02, 0.01), but did not reveal a similar relationship in the Without Facial Expression condition (*rs* = 0.07, 0.02, *ps* = 0.71, 0.93). The results for the With Facial Expression condition also showed a stronger correlation between "The conversation became lively" and "This system promotes English communication skills" (*rs* = 0.47, 0.23, *ps* = 0.01, 0.22) and between "Conversation was lively" and "Felt confident about actual English conversation" (*rs* = 0.41, 0.13, *ps* = 0.03, 0.49) than for the Without Facial Expression condition. This indicates that in a system that includes facial expression, a participant's self-evaluation of "lively conversation" is related to his or her evaluation of the ability to transfer that skill to real life and of the system's ability to help a user to improve English skills. Both conditions also yielded ratings of ≥ 3.5 (on a five-point scale) on the following items: "Was able to chat actively using the virtual world avatar," "Was not embarrassed because I knew nothing of those around me and nobody knew who I was" and "English conversation was fun."

Experiment 1 did not indicate that emotional expression features in an emoticon system produce better usability values than conditions without facial expression. A supplementary analysis involving the evaluation of correlations among individual items by condition suggested a relationship in the With Facial Expression condition between understanding a partner's emotion and the motivation to engage in conversation. The fact that the emoticon system used emoticons that were unfamiliar to the participants may have affected the level of difficulty that participants experienced in identifying emotions or their evaluations of usability. This fact suggests that there is a need to improve the system. In addition, the experiment only involved conversations about food and travel topics, limiting the frequency and variety of emotional expression. The scale for "virtual

reality" also had a low level of reliability. Therefore, in Experiment 2, we improved the input system for emotions, provided participants with scenarios and roles that produced a variety of emotions, and incorporated participants' understanding of their partner's emotion and emotional expression features into their usability ratings. We also improved the usability scales and increased the number of utterances from participants. In addition, we asked participants to chat in Japanese so that data were gathered from a more natural state of communication.

3. EXPERIMENT 2: EVALUATION OF THE ICON-INPUT SYSTEM

For Experiment 2, we improved the emotion input system from emoticon-input to icon-input to allow participants to select emotions more easily during chat. Participants were also given a story setting and a scenario instead of being allowed to chat freely, and we asked them to evaluate the usability of the system within such constraints. In other words, we asked participants to act as characters in a given story.

3.1. Methods
3.1.1. Participants
The participants were 32 undergraduate or graduate students (16 males, 16 females) at Kyoto University, ranging in age from 19 to 30. They were grouped into pairs and, as in Experiment 1, were unfamiliar with their partner.

3.1.2. Experimental system
In Experiment 2, we revised the 3D-ICS system from the first experiment, replacing the emoticons with icons representing emotions. First, a designer created these "emotional icons" based on the avatar faces used in this research. Next, three people, including the first author, agreed on and selected icons from three candidates per emotion category based on legibility, distinctiveness, and design integrity.

The system for inputting emotional icons (Figure 1, bottom) was displayed under the virtual environment window (Figure 1, top). The 11 types of emotional icons represented the same emotions as used in the emoticon input system in Experiment 1. Participants were asked to select one of the 11 emotional icons using a function key every time they entered text, thereby changing the avatar's expression and action.

Participants could select from three levels of

intensity for each emotion (weak, medium, and strong, as indicated on the red bar in the background) by pressing the key the desired number of times. Participants were allowed to familiarize themselves with the system before the experiment began.

The virtual environment window (Figure 1, top) showed the avatars of the participant and his/her partner. Avatars faced each other at a 30-degree angle so that users could see the facial expressions of both.

3.1.3. Experimental design
Experiment 2 had two conditions: with and without facial expressions. In the "With Facial Expression" condition, avatar expressions and actions changed when an emotional icon was selected; the "Without Facial Expression" condition only used text entries, and avatar expressions and actions were unchanged. Each participant was subjected to both conditions, and the order of conditions was counterbalanced.

3.1.4. Tasks
Chat. We randomly assigned 16 pairs to one of the two conditions so that eight pairs participated in the With Scenario condition and eight pairs participated in the Without Scenario condition. In both conditions, participants had conversations about two topics: a conversation by siblings concerning a gift for their parents, and a conversation between concerned friends of the same sex about a love triangle problem. In the With Scenario condition, participants were provided with a printed scenario that listed the lines of dialogue and the emotional category and level to enter after each line (e.g., happy-medium). Each scenario included one of each of the 11 emotional categories. The printed scenario given to each participant listed only the 11 lines of dialogue that he or she was to enter; the partner's lines were left blank. After each user had entered the specified line, they were permitted to continue free conversation. In contrast, users in the Without Scenario condition were only provided with a setting for the situation. For example, the setting for the conversation on the love triangle problem was as follows.

A and B are close friends (boys) in the same club at the same university. A and B met Girl C at a party the other day. B is dating C, but A, who likes C, does not know that. A asks B to come to the campus that night and consults him

about his feelings for C. Act as B and talk to A.

The chat time for each topic was 15 min. The combinations of the two items and the With and Without Facial Expression conditions were counterbalanced. All conversations were conducted in Japanese.

Understanding the partner's emotion. The participants were asked to guess the emotion of their partner's character and of their own character. They were asked to select one emotion from seven categories (happy, surprised, angry, sad, anxious, discontented, and embarrassed) at two times during the chat (at 5 and 10 min into the chat). Before and after the chat, the participants were asked to judge the intensity of each of the seven categories using a five-point scale (0 = *none*, 4 = *strong*).

Evaluations. Experiment 2 used the same scales used in Experiment 1: "Virtual reality" (2 items), "Anonymity" (2 items), "Emotional communication" (2 items), "Motivation to engage in conversation" (3 items), and 13 other items. Experiment 2 also included 11 new items: "Felt like the virtual character was my other self" was added because the α coefficient for the "Virtual reality" scale in Experiment 1 was low (the score was calculated in combination with the item, "The experience was like conversing in the real world"). Participants in the With Facial Expression Condition also scored the following new items: Was able to select the desired emotion type easily; Was able to select the level of emotion easily; Chose emotion by looking at the icons of the faces; Chose emotion by looking at the emotion words above the face icons; The "Thinking" and "Under consideration" actions of the virtual character were useful in chat; Was able to express the various emotions with the actions of the virtual character; Was able to express the intensity of the emotion; Looked at the actions of the virtual character while chatting; Was able to guess the emotion type from the actions of the virtual character; and Was able to guess the intensity of the emotions. Participants were also asked to provide written answers on a free-form sheet about the same three items as in Experiment 1 (comments about using the two systems, chatting, etc.).

The experimenter also asked the participants to rate their shyness and computer skills (30 items) using a five-point scale.

3.2. Results and observations

3.2.1. Frequency of emotion-related words and their categorization

We analyzed the chat logs of all participants in the Without Scenario conditions. We counted 341 utterances in the With Facial Expression condition and 419 in the Without Facial Expression condition. For data from the With Facial Expression group, we used the True Teller (Nomura Research Institute) text-mining tool to filter out words that were not related to emotions and then conducted an automatic categorization based on the 11 emotion category buttons. Table 3 shows the categories of words that co-occurred with the emotion buttons: the Happy emotion button was used with overwhelming frequency and co-occurred frequently with words like "excited." This emotion button was probably used frequently to convey positive feelings to a partner and to make the conversation run smoothly. The next most frequently used emotion buttons tended to be those used as responses to a partner's utterance such as Troubled, Anxious, Embarrassed, and Surprised. These were most likely used to heighten the reality of a participant's response to an utterance.

In contrast, the emotion buttons for Discontented, Sorry, Angry, and Sad were rarely used. These buttons were used to produce a facial expression corresponding to, for example, an utterance of "Sorry" and to communicate the speaker's emotion accurately. However, they were used less frequently because they can only be used in limited conversations. Other rarely used buttons included Tired and Sleepy. These convey a physical state, and in common real-life chat sessions, they are used implicitly to herald an exit from the session. Because our experiments were

Table 3.   Frequency of use of emotion buttons and co-occurrence words.

Emotion button		Example of co-occurrence words
Happy	(77)	excited, good, inexpensive, joy, kind
Troubled	(61)	sorry, think, give up, bad
Anxious	(45)	think, sorry, suffer, decide
Embarrassed	(25)	in fact, opposite sex, love, embarrass
Surprised	(24)	great, definitely, apologize
Discontented	(23)	imagine, talk
Sorry	(22)	in fact, I'm sorry, feel, think
Angry	(17)	strange, can say
Sad	(16)	feeling, think, know
Tired	(11)	---------
Sleepy	(3)	---------

Note: Numbers in parentheses indicate frequency.

short and participants were aware of the finish time, this was not an issue, which may explain the rare use of these buttons.

We found that the words accompanying the emotion buttons did not include words from the emotion categories (such as "Surprised," "Angry," or "Sad"). This indicates that in the With Facial Expression condition, participants tended to convey these emotions by pressing the emotion button and omitting the actual words. This finding corresponds to the results for Experiment 1 in that participants used fewer words in the With Facial Expression condition than in the Without Facial Expression condition (Table 2).

3.2.2. Comparison of the With Facial Expression and Without Facial Expression conditions

We determined the absolute value of the difference (on a five-point scale) between actual and estimated ratings of a partner's emotional level and compared the means of the seven emotion categories (Table 4). The differences tended to be smaller in the With Facial Expression condition than in the Without Facial Expression condition; however, two emotion categories had a significantly larger difference in the Without Facial Expression condition: Anxious (Ms = 2.00, 1.25, t (30) = 2.09, p = 0.046) and Surprised (Ms = 1.88, 1.00 , t (30) = 2.49, p = 0.019). This indicates that participants can determine the emotional intensity more accurately in the With Facial Expression condition than in the Without Facial Expression condition. Thus, this system can increase the precision with which the partner's emotion can be determined.

To compare ratings on the usability of the With Facial Expression and Without Facial Expression systems, we calculated the criteria scores based on the mean ratings of "Virtual reality" (reliability coefficients for the two conditions: as = 0.61 and 0.63), "Anonymity" (as = 0.90, 0.87), "Emotional communication" (as = 0.82, 0.87), and "Motivation to engage in conversation" (as = 0.71, 0.64). As a result, the degree to which participants feel as if they were communicating in the real world ("Virtual reality") was significantly higher in the With Facial Expression condition than in the Without Facial Expression condition; "Motivation to engage in conversation" was also high, but less significantly so. We observed no significant differences between the two conditions in the mean numbers of utterances.

Because the reliability of the "Virtual reality" scale was low, we conducted a supplementary

Table 4. Comparison of means (*SD*) in With facial expression and Without Expression conditions in the emotional icon-input system

Measure	With Facial Expression	Without Facial expression	t value (df = 31)
Number of utterance[a]	13.6 (3.5)	15.8 (7.7)	1.03 p = .310
Difference in judging emotion level of partner[b1]	0.80 (0.32)	1.09 (0.42)	2.30 p = .021
Virtual reality[2]	3.38 (1.13)	2.92 (0.98)	2.44 p = .043
Anonymity[2]	3.00 (1.11)	2.94 (1.15)	0.34 p = .737
Emotional communication[2]	3.11 (0.82)	3.13 (0.82)	0.08 p = .940
Motivation for conversation[2]	2.75 (0.81)	2.42 (0.72)	1.97 p = .058

Note: [a] Without Scenario [b] With Scenario
[1] Range is 0 - 4
[2] 5-point scale (1=*disagree*, 5=*agree*)

analysis by checking the Pearson product-moment correlation coefficient between items. We found a higher correlation between ratings for "Easy to convey feelings" and "The conversation became lively" (rs = 0.43, 0.13, ps = 0.02, 0.48) and between ratings for "Feel that the virtual character is my other self" and "Can converse actively with a virtual character" (rs = 0.44, 0.07, ps = 0.02, 0.70) in the With Facial Expression condition than in the Without Facial Expression condition.

These results indicate that conditions in which participants can express emotions by manipulating the avatar's expression using emotional icons produce a higher correlation between the ease of conveying feelings and motivation to engage in conversation. Levels of shyness, computer skills, and gender did not affect the number of utterances or the evaluation of the system in either condition. This finding indicates that a learning environment that uses virtual space can increase a learner's motivation to converse, regardless of shyness or computer skills.

3.2.3. Free-answer descriptions of usability

We categorized evaluations of the system into positive (11 participants), negative (7 participants), and neither positive nor negative (14 participants). Of the positive evaluations, 10 participants commented that the system made it easy to express and understand emotions that are difficult to communicate in words. Other comments included, "was able to understand and judge my partner's emotions and mood based on the expressions of the avatar" (7 participants), and

"the system is emotionally expressive" (4 participants).

Negative comments included operational problems in entering emotions (12 participants) and, most notably, "the selection of emotions is tedious and difficult" (6 participants). Some participants identified problems in the types and content of emotions (7 participants), commenting, "it is not equivalent to the real world because it is difficult to represent subtle expressions and emotions behind an expression," or "it is not possible to create real and diverse expressions."

4. CONCLUSIONS AND FUTURE DIRECTIONS

We implemented an emotion expression function for avatars in a communication system that uses 3-D virtual space. Experiment 1 tested an emoticon-input system, and Experiment 2 tested an emotion icon-input system. Participant ratings indicated that the emoticon-input system was not advantageous in terms of usability over a system with no facial expression. After revising our system to an emotion icon-input system, we observed a higher level of precision about judging emotions compared to the system with no facial expressions. In addition, participants tended to be more motivated to engage in conversation. In other words, our findings indicate that when avatars are imbued with expression, participants tend to have a better understanding of emotions in communication and an increased motivation to enter into conversation. Our findings also highlighted the need for usability improvements to the system. In addition, Experiment 1 involved chatting about casual topics with a real-life partner, and Experiment 2 involved chatting with a partner engaged in role-playing based on scenarios that evoked emotion. Future studies should analyze the commonalities and differences between these two situations.

Future directions may include the following. First, it would be helpful to make the emotional expressions of avatars automatic, allowing students to engage in smooth conversation and establishing an educational communication system over a virtual space. This might include, for example, developing a system that can anticipate emotions from emotion words in utterances and automatically change the avatar's expression or allow users to select emotions more easily. The system could select potential emotions from keywords within a conversation, the mood of the participant, or the data logged during the

conversation. This process would require using corpus data (data about words that co-occur with emotion buttons, as shown in Table 3) tagged to emotion selection buttons to assess the speaker's emotional state. It would also require clarifying the factors underlying an emotion, the intensity of the emotion, and the rules that support these factors. Both requirements constitute a challenge for future studies.

A second direction would be to conduct research to verify the long-term educational benefits of using this system. We assessed differences in usability over short-term use in two experiments, but our analyses of students' emotional understanding during communication and changes in conversational motivation were not extensive. It will also be necessary to use communication systems that feature avatars in class work to assess the long-term educational benefits. Future studies should clarify what elements of the system heighten motivation, improve performance, and improve real communication skills among students.

A third direction would be to establish the system as a virtual communication system for foreign language education. In doing so, heightening the reality by giving avatars expressions and actions that match the culture of the foreign language could help improving students' social and conversation skills. It will also be necessary to promote the development of educational programs (e.g., Suzuki 2000) that allow learners to assign situations in virtual space such as shopping, debates, and academic presentations. This would allow users to practice role-playing using expressive avatars so that they can communicate in a way similar to a real-life setting. This practical research could be extremely beneficial.

ACKNOWLEDGEMENTS

Scientific Research on Priority Areas (A) and from a "Cognitive Assessment and Application of Communication Systems that use Virtual Space" Grant-in-Aid for Exploratory Research, from the Japanese Ministry of Education, Science, and Culture.

NOTES

1) When analyses omitted utterances that included emoticons with $\leq$ 50% correct responses in the preliminary experiment in the With Facial Expression condition, the mean number of utterances (SD) was 11.9 (6.45) and the mean number of words (SD) was 66.6 (22.0). When we evaluated the means in the Without Facial Expression condition, we found no differences in the number of utterances between the two conditions (t = 0.52, p = 0.605) and we found more words in the Without Facial Expression condition than in the With Facial Expression condition (t = 2.55, p = 0.016). These results were the same as the results when we did not omit utterances with emoticons with a low percentage of correct answers.

References

ADACHI, N. (1999) Educational use of network games (MUD). *NEW Education and Computers*, **7** : 94–95

AWAJI, Y. (2000) Text-based virtual reality saMoorai and its application to English education. Proceedings of the 6th Joint Conference on Educational Technology, **2** : 457–458

CASSELL, J., SULLIVA, J., PREVEST, S., and CHURCHILL, E. (Eds.) (2000) *Embodied Conversational Agents. Cambridge*, MIT Press, Massachusetts

G-SEARCH (2006) The 15-year history of Habitat http://www.j-chat.net/habitat/history.html

HAMABE, T. (2000) Three-dimensional interactive educational system. *Knowledge creation and integration (Nomura Research Institute)*, **7** : 12–13

HAMABE, T. and MATSUDA, M. (2002) Towards the Realization of Learning by Cyber-Communication. Technical Creation: *NRI Technical Report*, 1:20–35

IMAI, A. and OSHIMI, T. (1987) Analysis on shyness scale. Proceedings of 28th Annual Meeting of the Japanese Society of Social Psychology, **6**

KUSUMI, T., KOYASU, M. and NAKAMURA, A. (2002) Effective use of multi-media network in an ESP class: Analysis of cooperative learning A. SAKAMOTO (Ed.) *Advanced Use of Multimedia for the Higher Education Reform. Annual Report of Grant-in-Aid for Scientific Research in Priority Areas (A)*, pp. 47–50

LEA, R., HONDA, Y., MATSUDA, K. and MATSUDA, S. (1997) Community Place: Architecture and performance. **VRLM-97** : 41–50

MINISTRY of INTERNAL AFFAIRS and COMMUNICATIONS (2003) *Annual Report on Information Telecommunications 2003*

OKANO, S. (2000) Foreign language learning and three-dimensional virtual space. *Studies in German and French Literature*, **50** : 61–68

RENNINGER, K.A. and SHUMAR, W. (Eds.) (2002) *Building virtual communities: Learning and change in cyberspace.* Cambridge University Press, Cambridge.

RIZZO, A.A., NEUMANN, U., ENCISO, R., FIDALEO, D. and NOH, J.Y. (2001) Performance-driven facial animation: Basic research on human judgments of emotional state in facial avatars. *CyberPsychology & Behavior*, **4** : 471–487.

SAKAMOTO, A., ISOGAI, N., KIMURA, A., TSUKAMOTO, K. KASUGA, T. and SAKAMOTO, T. (2000) Using the Internet as the sociality trainer of shy people: An experiment of female university students. *Japan Journal of Educational Technology*, **24**(3) : 153–160

SUZUKI, Y.(2000). English classes with the help of 3-Dimensional interactive education system. *Studies in languages and cultures (Faculty of Languages and Cultures, Kyushu University)*, **12** : 105–125.

TAKEUCHI, Y. and KATAGIRI, Y. (2000) Inducing social agreement responses toward interface agent. *Transactions of Information Processing Society of Japan*, **42** : 1257–1266

TAKIZAWA, N. (1995) Learning a foreign language by MOO. *Gengo*, **10** : 111–113

YUASA, M., YASUMURA, S. and NITTA, K. (2003) A method of generating facial expressions for negotiation agents by using Bayesian network. *Transactions of Information Processing Society of Japan*, **44** : 2710–2717