PAPER

Effects of Conversational Agents on Activation of Communication in Thought-Evoking Multi-Party Dialogues***

Kohji DOHSAKA[†*a)], ***Member***, **Ryota ASAI**[††**], ***Nonmember***, **Ryuichiro HIGASHINAKA**[†††], **Yasuhiro MINAMI**[†], ***and*** **Eisaku MAEDA**[†], ***Members***

SUMMARY This paper presents an experimental study that analyzes how conversational agents activate human communication in thought-evoking multi-party dialogues between multi-users and multi-agents. A thought-evoking dialogue is a kind of interaction in which agents act to provoke user thinking, and it has the potential to activate multi-party interactions. This paper focuses on quiz-style multi-party dialogues between two users and two agents as an example of thought-evoking multi-party dialogues. The experimental results revealed that the presence of a peer agent significantly improved user satisfaction and increased the number of user utterances in quiz-style multi-party dialogues. We also found that agents' empathic expressions significantly improved user satisfaction, improved user ratings of the peer agent, and increased the number of user utterances. Our findings should be useful for activating multi-party communications in various applications such as pedagogical agents and community facilitators.

key words: *multi-party interaction, dialogue systems, human-agent interaction, human-robot interaction*

1. Introduction

Conversational interfaces such as conversational agents and dialogue systems have been typically used in situations where a single user interacts with a single agent or system [1]–[3]. However, a new area of research on conversational interfaces has emerged that deals with multi-party interactions [4]–[7]. Multi-party conversational interfaces have been applied to several tasks: training decision making in team activities [4], collaborative learning [5], and coordinating and facilitating interactions in casual social groups [6], [7].

The main advantage of such multi-party dialogues over those that are two-party is that multi-party cases encourage group interactions and collaborations by human users. This advantage can be exploited to foster human collaborative activities in more social settings and to build and maintain social relationships among people. However, unless users actively engage in interactions, these qualities of multi-party dialogues cannot be adequately exploited. It is thus valuable to identify factors that will enhance user engagement in multi-party dialogues.

We set our objectives from this viewpoint to activate communication in multi-party dialogues between multi-users and multi-agents. What we mean here by the activation of communication is the enhancement of user engagement with multi-party dialogues. We exploited a new style of dialogue called thought-evoking dialogue as the first step toward achieving this research objective and experimentally investigated what impact conversational agents had in activating communication in thought-evoking multi-party dialogues.

A thought-evoking dialogue is an interaction in which agents act on the willingness of users to provoke user thinking and encourage involvement in the dialogue. It has the potential to activate multi-party interactions. Previous work has proposed a quiz-style dialogue system of presenting information (hereinafter called a quiz-style dialogue system) [8], which is regarded as a kind of thought-evoking dialogue system. This system conveys content as biographical facts about famous people through quiz-style interactions with users by creating a "Who is this?" quiz and individually presenting hints. The hints are ordered based on the level of difficulty of naming people experienced by users. As users have to consider hints in solving a quiz, the quiz-style dialogue system provokes their thoughts that bring together hints and combine them to come up with reasonable answers. The provocation of such kinds of thoughts is not necessarily equivalent to the activation of communication since users can exert such thoughts while they are not exhibiting their willingness to engage in interactions. However, by comparing a two-party quiz-style dialogue system with a read-out system that unilaterally presented biographical facts to users in encyclopedic order, this previous work demonstrated that a quiz-style dialogue system enhanced users' willingness to engage with the system. It also revealed that the quiz-style dialogue system led to better user memorization of biographical facts conveyed in the dialogues than the read-out system.

We focused on quiz-style multi-party dialogues to present information (hereinafter called quiz-style multi-party dialogues) in this research as examples of thought-evoking multi-party dialogues and our experimental evalu-

Manuscript received November 13, 2013.
Manuscript revised March 23, 2014.

[†]The authors are with NTT Communication Science Laboratories, NTT Corporation, Kyoto-fu, 619–0238 Japan.
[††]The author is with the Graduate School of Information Science and Technology, Osaka University, Suita-shi, 565–0871 Japan.
[†††]The author is with NTT Media Intelligence Laboratories, NTT Corporation, Yokosuka-shi, 239–0847 Japan.
[*]Presently, with Akita Prefectural University.
[**]Presently, with Denso Corporation.

[***]This paper is an extended version of a paper presented at SIGDIAL 2009.

a) E-mail: dohsaka@akita-pu.ac.jp

DOI: 10.1587/transinf.E97.D.2147

ation of how a peer agent's presence and agents' emotional expressions activated communication. We measured the activation of communication in terms of the number of user utterances, user satisfaction, and user opinions about the agent.

A peer agent is a conversational agent that acts as a peer of human users and participates in interactions in the same way that users do. We were interested in the peer agent's role in quiz-style multi-party dialogues since the positive effects of a peer agent on users have been demonstrated in the educational domain [9], [10], which is a promising area of application for quiz-style dialogues. Users in the educational domain can not only benefit from direct communication with a peer agent but also from observing dialogues between the peer agent and a tutor. Learning by observing others who are learning is called vicarious learning or observational learning, and previous work has reported that it has positively affected learners' performance [11]–[13]. Chi *et al.* analyzed various modes of observation and revealed that the more actively engaged the observers are in interactions with a peer, the more effective the learning becomes [14]. This indicates that the activation of communication (the enhancement of user engagement) would lead to better learning. However, detailed experimental investigations into the effects of a peer agent on the activation of communication have not been reported in multi-party dialogues between multi-users and multi-agents to the best of our knowledge, which are our main concern in this paper.

The topic of emotion has gained widespread attention in human-computer interactions [15]–[18]. The impact of an agent's emotional behaviors on users has also recently been studied [10], [19], [20]. However, these previous studies addressed scenario-based interactions in which a user and an agent acted with predetermined timing. This paper investigates what impact agents' emotional expressions has on the activation of communication in multi-party dialogues in which multiple users and agents can make utterances with more flexible timing.

We classified agent emotional expressions into those that were empathic and self-oriented similarly to work done by Brave *et al.* [19]. Brave *et al.* [19] showed that user opinions about the agent were influenced by agent empathic expressions in contrast with agent self-oriented expressions in scenario-based black-jack interactions, but we examined the impact of agent empathic and self-oriented utterances on the activation of communication in multi-party dialogues that enabled more flexible turn-taking. Previous studies [21], [22] have indicated that agents' empathic expressions have a positive psychological impact on users, but they only examined two-party cases. Although Traum *et al.* [4] and Gebhard *et al.* [23] exploited the role of agent emotions in multi-party dialogues, they did not adequately examine what effects agent emotions had on activating communication through experiments.

This paper presents an experimental study in which we analyzed how agents activated human communication in quiz-style multi-party dialogues between two users and two agents. Our findings should be useful for activating human communication in various applications such as educational agents and community facilitators. We dealt with disembodied agents in this work and focused on their linguistic behaviors. We believe that our results should be useful for designing embodied conversational agents by using other modalities.

Section 2 presents an overview of our quiz-style multi-party dialogue system and Sect. 3 explains the design of the experiment. Section 4 presents the results and Sect. 5 discusses them. Section 6 concludes the paper.

2. Thought-Evoking Multi-Party Dialogue System

We implemented a quiz-style multi-party dialogue system between multi-users and multi-agents. The system was a Japanese keyboard-based dialogue system with a chat-like interface. The users could make utterances at any time they wanted. User utterances were completed and displayed on the chat window when the enter key was pressed.

Our experiment dealt with cases where two users and two agents were engaged in a dialogue. The two agents were a quizmaster and a peer. The quizmaster agent created a "Who is this?" quiz about a famous person and presented hints individually to the users and the peer agent who guessed what the correct answer might be.

Figure 1 shows a sample dialogue. Mary and Jane are human users. *Whowho* is the quizmaster agent, and *Kinoko* is the peer agent. Quizmaster agent *Whowho* presents hints in lines 1, 7, 8, and 13. Users Mary and Jane and peer agent *Kinoko* provide answers in lines 3, 9, and 14.

The hints were automatically created using biographical facts (in Japanese) of people in Wikipedia[†] using a previously reported method [24].

1	*Whowho*	Who is this? First hint: Graduated from the University of Tokyo.
2	Mary	I don't know.
3	*Kinoko*	Yoshida Shigeru.
4	*Whowho*	No, not even close!
5	Jane	That's very difficult.
6	*Kinoko*	Difficult for me, too.
7	*Whowho*	Second hint: Born in Ushigome, Edo.
...		
8	*Whowho*	Third hint: Novelist and scholar of British literature.
9	Mary	Murakami Haruki.
10	*Whowho*	Close!
11	*Kinoko*	You are close. Excellent.
12	Jane	Well then, who is it?
13	*Whowho*	Fourth hint: Familiar with Haiku, Chinese poetry, and calligraphy.
14	Mary	Natsume Soseki.
15	*Whowho*	That's right. Wonderful.
16	*Kinoko*	Mary, excellent. I'm happy for you.
17	Jane	Mary, that's the right answer. Good job.

Fig. 1 Sample dialogue.

[†]http://ja.wikipedia.org/

2.1 Dialogue Acts

The users and the two agents carried out several dialogue acts based on the context of the dialogue.

Present-hint: The quizmaster agent presented hints individually (lines 1, 7, 8, and 13) in the sample dialogue in Fig. 1.

Give-ans: Users and the peer agent provided answers (lines 3, 9, and 14).

Show-difficulty: Users and the peer agent offered opinions about the difficulty of the quiz (lines 2, 5, 6, and 12).

Evaluate-ans: When the answer was wrong, the quizmaster agent evaluated it based on the person-name similarity score and uttered "very close!," "close!," "a little closer!," "a little too far," "far," or "not even close!" (lines 4 and 10). The person-name similarity score was derived based on Matsubayashi and Yamada [25], which calculated the collocation probability of people's names by Fisher's exact test using Wikipedia articles and visualized the names as a distribution on a two-dimensional map [8].

Complete-quiz-with-success: When the correct answer was given, the quizmaster agent informed the participants in the dialogue that the current quiz was completed (line 15).

Complete-quiz-with-failure: If all the hints were generated and no correct answer was given, the quizmaster agent provided the right answer, and the current quiz was completed.

Feedback-on-wrong-ans: Users and the peer agent provided feedback when their own or the other's answers were incorrect during the current quiz (line 11).

Feedback-on-success: Users and the peer agent provided feedback when their own or the other's answers were correct and the current quiz session was completed (lines 16 and 17).

Feedback-on-failure: Users and the peer agent provided feedback when the current quiz was completed without the correct answer.

Address-hearer: Users and the two agents specified an intended addressee by uttering the other's name (lines 16 and 17).

Backchannel: The users and the peer agent uttered backchannel responses.

When a user utterance was input, the system separated it into word tokens using a Japanese morphological analyzer and converted it into dialogue acts using hand-crafted grammar. The system could recognize 120,000 proper names of people.

2.2 Utterance Generation

Surface realization forms were prepared for each dialogue act by the agents. Agent utterances were generated by randomly selecting one of the forms.

Table 1 Examples of agent expressions. EMP indicates empathic expressions, SELF indicates self-oriented expressions, and NONE indicates neutral expressions when neither emotion was present.

Dialog act	Emotion	Expressions
Show-difficulty	EMP	Difficult for me, too.
Show-difficulty	SELF	I don't remember. That's so frustrating.
Show-difficulty	NONE	I don't know.
Feedback-on-success	EMP	You're right. I'm happy for you.
Feedback-on-success	SELF	I'm really glad I got the answer correct.
Feedback-on-success	NONE	You're right / I'm right.
Feedback-on-failure	EMP	Too bad you didn't know the right answer.
Feedback-on-failure	SELF	I'm disappointed that I didn't know the right answer.
Feedback-on-failure	NONE	I/You didn't know the right answer.

Some agent dialogue acts could be generated with emotional expressions. Agent emotional expressions were categorized into those that were empathic and self-oriented [19]. Agent self-oriented emotional expressions (self-oriented expressions) were oriented to their own state, and agent empathic expressions were oriented to the other's state and were congruent with the other's welfare. As explained in Sect. 3.1, we prepared different experimental conditions to determine the presence/absence of agents' empathic and self-oriented expressions. Based on the conditions, we controlled the agents' emotional expressions. Table 1 lists examples of agents' empathic, self-oriented, and neutral expressions.

2.3 Dialogue Management

The system maintained a dialogue state in which the history of the participant's dialogue acts was recorded with the time for each act. We prepared preconditions for each dialogue act by the agents. For example, the quizmaster agent's *Evaluate-ans* could be executed after the users or the peer agent provided an incorrect answer. The peer agent's *Feedback-on-success* could be executed after the quizmaster agent carried out *Complete-quiz-with-success*. We also used the following turn-taking rules:

1. Either agent must talk when neither the users nor the agents make utterances within a given time (4 sec.).
2. Agents must not talk for a given time (0.5 sec.) after the others have talked.
3. The quizmaster agent must move to the next hint when neither the users nor the peer agent have provided a correct answer within a given time (30 sec.).

The system chose the next speaker and its dialogue act based on the dialogue state, the preconditions for the dialogue acts, and the turn-taking rules.

3. Experiment

3.1 Experimental Conditions

We prepared five systems under different experimental conditions of (0), (1), (2), (3), and (4) based on the presence/absence of the peer agent and agents' empathic and self-oriented expressions to evaluate what effects the presence of the peer agent and the agents' emotional expressions had. These are summarized in Table 2. The peer agent was absent under condition (0), and only the quizmaster agent was present. Both the quizmaster and peer agents were present under the other conditions. Neither empathic nor self-oriented expressions were exhibited under conditions (0) and (1). Only empathic expressions were exhibited under condition (2). Only self-oriented expressions were exhibited under condition (3). Both empathic and self-oriented expressions were exhibited under condition (4).

We evaluated what effects the presence of the peer agent had by comparing conditions (0) and (1). We evaluated what effects agents' empathic and self-oriented expressions had by comparing conditions (1), (2), (3), and (4).

3.2 Measures

We measured the activation of communication from three perspectives on the number of utterances, user satisfaction, and user opinions about the peer agent.

The number of utterances has been used to measure the activeness of interaction [26]. We regarded the number of user utterances as an objective measure to evaluate the activation of communication. We counted the number of user utterances per quiz hint and investigated how they were influenced by the peer agent's presence and agents' emotional expressions.

User satisfaction and opinions about the peer agent were subjective measures based on questionnaires (ten-point Likert scale). They have been used to measure the quality of interactions [19], [27]. Table 3 summarizes the questionnaires we administered in the experiment. We expected that a high level of user satisfaction and positive opinions about the peer agent would lead to the activation of communication.

User satisfaction was evaluated from different perspectives with three questions: Q1, Q2, and Q3. Q1 focused on user willingness to engage in the dialogue, Q2 focused on user experience of the dialogue's pleasantness, and Q3 focused on user satisfaction with the system. We evaluated user satisfaction with averages for the ratings of Q1, Q2, and Q3. Using the averaged ratings of Likert questions allowed us to apply such parametric statistical tests as multi-factor ANOVA since the summed or averaged responses to Likert questions tended to follow a normal distribution.

User opinions about the peer agent were evaluated in terms of how users perceived the peer agent's intimacy (Q4), its compassion (Q5), its amiability (Q6), and its encouragement (Q7). We evaluated user opinions about the peer agent with the averaged ratings for these items. Previous studies revealed that empathic behaviors exhibited by an agent improved user opinions about the agent in a black-jack scenario [19] and in a social dialogue between a single user and an agent [22]. We examined these items in multi-party dialogues with flexible turn-taking.

Besides our main objective of evaluating the activation of communication, we also evaluated user memorization of biographical facts conveyed in the dialogues as a secondary objective. User memorization was measured with memory tests to examine how well the participants had memorized the biographical information conveyed during the dialogues. It is known that a two-party quiz-style dialogue system leads to better memorization than that in a read-out system [8]. We evaluated user memorization to examine how the advantage of a quiz-style dialogue was influenced by a peer agent's presence and agents' emotional expressions in multi-party interactions. It is possible that these factors may increase user memorization since they might enhance user attention to listen to conveyed biographical information. It is also possible that these factors may decrease user memorization since they might distract his/her attention.

3.3 Procedure

We recruited and remunerated 64 Japanese adults (32 males

Table 2 Experimental conditions based on presence/absence of peer agent and agent empathic and self-oriented expressions.

Condition	Peer agent	Empathic	Self-oriented
(0)	Absent	Absent	Absent
(1)	Present	Absent	Absent
(2)	Present	Present	Absent
(3)	Present	Absent	Present
(4)	Present	Present	Present

Table 3 Questionnaire items to evaluate user satisfaction (Q1, Q2, and Q3) and user opinions about the peer agent (Q4, Q5, Q6, and Q7).

	Questionnaire items
Q1	Did you want to converse with this system again? (Willingness to engage in dialogue)
Q2	Was the dialogue enjoyable? (Pleasantness of dialogue)
Q3	Did you feel satisfied using the dialogue system? (Satisfaction with system usage)
Q4	Was the peer agent friendly? (Agent's intimacy)
Q5	Did you feel that the peer agent cared about you? (Agent's compassion)
Q6	Was the peer agent amiable? (Agent's amiability)
Q7	Did the peer agent encourage you? (Agent's encouragement)

and 32 females) to participate in the study. The mean age of the male group was 32.0 and that of the female was 36.2 (male group: SD=9.2, min=22, and max=59 and female group: SD=9.6, min=20, and max=50). The participants were divided into 32 pairs of the same gender: 16 pairs of males and 16 pairs of females. The participants in each pair were unacquainted.

The experiment had a within-participants' design. Each pair of participants successively engaged in dialogues using the five systems under different experimental conditions. Five "Who is this?" quizzes about famous people were presented in the dialogues with each system. The order in which the systems were used was counter-balanced to prevent the order effect. After they had completed a dialogue with each system, the participants filled out the questionnaires and took a memory test about the information they had acquired during the dialogue. The questionnaires on user opinions about the peer agent were only used when it was present (conditions (1), (2), (3), and (4)). Section 3.4 explains how we created the quizzes and memory tests.

The participants were informed before they started the experiment that they would have to fill out questionnaires and take a memory test after the dialogue with each system. They were also told that the agents were computer programs and not human participants. Each pair of participants was seated in separate rooms in front of a computer display with a keyboard and a mouse during the experiment, and they could only communicate with each other through the system.

A pair of participants took 18 minutes on average to complete a dialogue with each system. There were an average of 7.5 hints that were actually presented in a quiz.

3.4 Creating Quizzes and Memory Test

The quiz subjects in the "Who is this?" quizzes were chosen so that the level of difficulty of the quizzes was approximately the same in all the systems. We first sorted people in Wikipedia in descending order for this purpose by their PageRank ™ score based on Wikipedia's hyper-link structure [8]. We then extracted the top 50 people and divided them from the top into five groups of 10. We next randomly selected five people from each group to make five sets (Sets-A) of five people with approximately identical PageRank scores.

After people in Sets-A had been excluded from the original five groups, we then randomly selected five people from each group to create another five sets (Sets-B) of five people with approximately identical PageRank scores.

When participants successively engaged in dialogues with the five systems in the experiment, five sets of five people in Sets-A were successively used for the quizzes and memory tests. Five sets of five people in Sets-B were not used for the quizzes. They were used for memory tests after a dialogue with each system to measure the prior knowledge of the participants.

The quiz hints for people in Sets-A and Sets-B were automatically created using biographical facts (in Japanese) about people in Wikipedia using a previously reported method [24].

We next explained how the memory tests were created. Participants in the experiment took a memory test after each dialogue. The memory test was composed of people's biographical statements where one important expression in each statement was a blank (e.g.,"Soseki Natsume graduated from [ ]"). Each hint statement with a blank was presented with three multiple choices, one of which was the correct answer. There is an example of such a hint statement with a blank with three multiple choices of (a), (b), and (c) below.

Soseki Natsume graduated from [ ].
(a) Kyoto University, (b) The University of Tokyo,
(c) Keio University

The correct answer here is (b) The University of Tokyo. The memory test after each dialogue was composed of twenty hint statements with a blank, which were composed of ten hint statements for the set of five people in Sets-A and ten hint statements for the set of five people in Sets-B. They were randomly shuffled. The five people in Sets-A were used as quiz subjects in the dialogue, and the five people in Sets-B were not used.

These hint statements with a blank with three multiple choices were created by two annotators, who were not the authors, in four steps:

1. The annotators first spotted expressions in the quiz hints for all 50 people (Set-A plus Set-B) that they considered important in characterizing the person in question (e.g., "The University of Tokyo" in "He graduated from The University of Tokyo").
2. One of the expressions in each hint statement was randomly replaced by a blank (e.g., "He graduated from [ ]"). If a hint statement did not contain such expressions or the whole statement became a blank, the hint statement was not chosen for the memory test.
3. Two hint statements with a blank were randomly chosen for all people and they were used for the memory test. Two annotators revised the hint statements to disclose whose biographical facts each hint statement was about (e.g., "Soseki Natsume graduated from [ ]").
4. Two annotators discussed making up two choices other than the correct answer for the blank in each hint statement. These three choices including the correct answer were randomly shuffled.

4. Results

4.1 User Satisfaction

Cronbach's alpha was 0.83 for questions Q1, Q2, and Q3, which justified combining these items into a single index. Therefore, we evaluated user satisfaction with the averages of the ratings for these items.

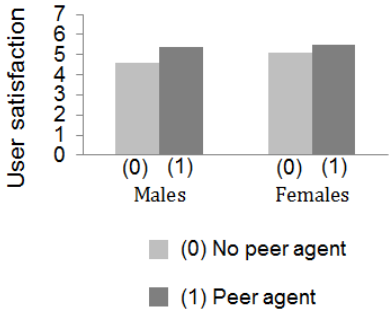

Fig. 2 User satisfaction under conditions (0) and (1).

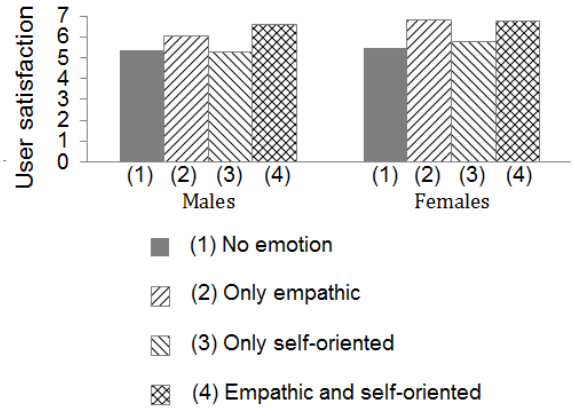

Fig. 3 User satisfaction under conditions (1), (2), (3), and (4).

We compared conditions (0) and (1) to evaluate what effect the peer agent's presence had on user satisfaction. Figure 2 shows user satisfaction under conditions (0) and (1). The F-test results indicated that variances were assumed to be equal across groups ($p > 0.2$), and the Kolmogorov-Smirnov test results demonstrated that the assumption of normality was satisfied ($p > 0.2$). We found that the peer agent's presence significantly improved user satisfaction (male group: $t(31) = 4.21$ and $p < 0.001$ and female group: $t(31) = 2.86$ and $p < 0.01$) by applying a paired t-test to both the male and female groups.

We compared conditions (1), (2), (3), and (4) to evaluate what effect the empathic and self-oriented expressions exhibited by the agents had on user satisfaction. Figure 3 shows user satisfaction under conditions (1), (2), (3), and (4). Three-factor ANOVA was conducted with two within-participant factors of empathic and self-oriented expressions and one between-participant factor of gender. The F-test for the homogeneity of variances ($p > 0.2$) and the Kolmogorov-Smirnov normality test ($p > 0.2$) indicated that the data met the ANOVA assumptions. As a result of ANOVA, a significant main effect was found for empathic expressions with respect to user satisfaction, i.e., $F(1, 62) = 92.7$ and $p < 0.001$. No significant main effects were found for either self-oriented expressions or gender, and there were no significant interactions.

These results revealed that the peer agent's presence and the agents' empathic expressions significantly improved user satisfaction in quiz-style multi-party dialogues.

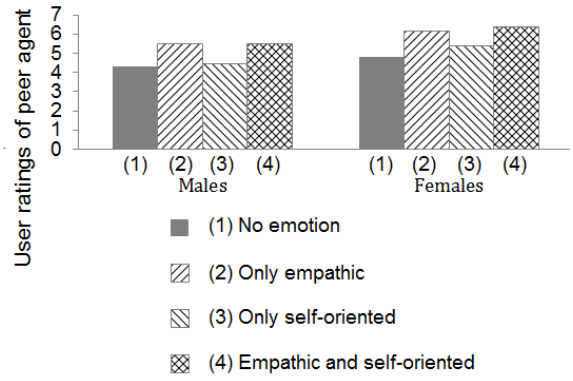

Fig. 4 User ratings of peer agent under conditions (1), (2), (3), and (4).

4.2 User Opinions about Peer Agent

Cronbach's alpha was 0.92 for questions Q4, Q5, Q6, and Q7, which justified combining these items into a single index. Therefore, we evaluated user opinions about the peer agent with the averaged ratings of these items under each experimental condition. Figure 4 shows the user ratings for the peer agent under each condition.

We compared conditions (1), (2), (3), and (4) to evaluate what effect agents' empathic and self-oriented expressions had on user ratings of the peer agent. Three-factor ANOVA was conducted with two within-participant factors of empathic and self-oriented expressions and one between-participant factor of gender. The F-test for the homogeneity of variances ($p > 0.2$) and the Kolmogorov-Smirnov normality test ($p > 0.2$) indicated that the data satisfied the ANOVA assumptions. As a result of ANOVA, a significant main effect was found for empathic expressions with respect to user ratings of the peer agent, i.e., $F(1, 62) = 77.4$ and $p < 0.001$. There was a moderate main effect for self-oriented expressions with respect to user ratings of the peer agent, i.e., $F(1, 62) = 4.38$ and $p < 0.05$. There were no significant main effects for gender, and there were no significant interactions.

These results indicated that agents' empathic expressions significantly improved user ratings of the peer agent in quiz-style multi-party dialogues.

4.3 Number of User Utterances

We compared conditions (0) and (1) to evaluate what effect the peer agent's presence had on the number of user utterances per quiz hint. Figure 5 shows the number of user utterances per quiz hint under conditions (0) and (1). The assumptions of variance homogeneity ($p > 0.2$) and normality ($p > 0.2$) were satisfied based on the F-test and the Kolmogorov-Smirnov test. We found that the presence of the peer agent significantly increased the number of user utterances per hint (male group: $t(31) = 3.11$ and $p < 0.01$ and female group: $t(31) = 5.62$ and $p < 0.001$) by applying

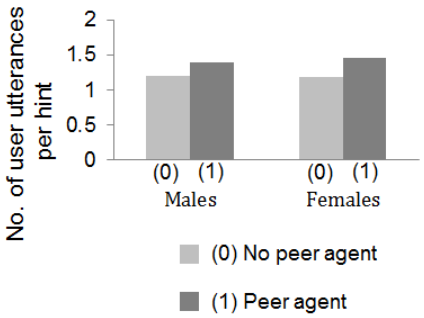

Fig. 5 Number of user utterances per quiz hint under conditions (0) and (1).

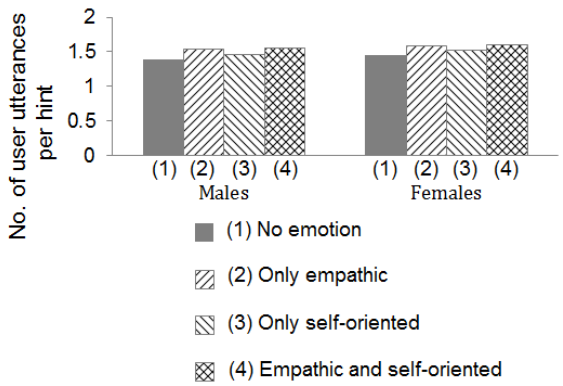

Fig. 6 Number of user utterances per quiz hint under conditions (1), (2), (3), and (4).

a paired t-test to both the male and female groups.

We classified user utterances into two categories of answer and non-answer to examine the number of user utterances in more detail. Answer utterances were those uttered as answers to a quiz and non-answer utterances were those other than answer utterances. Non-answer utterances included those indicating difficulty or ease in answering a quiz (e.g. "That's so frustrating"), those indicating feedback on the success or failure of a quiz (e.g. "I'm really glad") and those indicating that the user was thinking (e.g. "I'm thinking"). We applied a paired t-test to the number of user answer and non-answer utterances per quiz hint. As a result, the presence of the peer agent significantly increased the number of both user answer and non-answer utterances per hint. (answers in male group: $t(31) = 3.94$ and $p < 0.001$ and answers in female group: $t(31) = 4.35$ and $p < 0.001$; non-answers in male group: $t(31) = 2.47$ and $p < 0.05$ and non-answers in female group: $t(31) = 3.12$ and $p < 0.01$).

We then compared conditions (1), (2), (3), and (4) to evaluate what effect empathic and self-oriented expressions by agents had on the number of user utterances. Figure 6 shows the number of user utterances per quiz hint under conditions (1), (2), (3), and (4). Three-factor ANOVA was conducted with two within-participant factors of empathic and self-oriented expressions and one between-participant factor of gender. The F-test for the homogeneity of variances ($p > 0.2$) and the Kolmogorov-Smirnov normality test ($p > 0.2$) revealed that the data satisfied the ANOVA assumptions. As a result of ANOVA, a significant main effect was found for empathic expressions with respect to the number of user utterances, i.e., $F(1, 62) = 18.9$ and $p < 0.001$. No significant main effects were found for either self-oriented expressions or gender, and there were no significant interactions.

We applied three-factor ANOVA to the number of user answer and non-answer utterances per quiz hint to enable a more detailed examination. As a result, there were no significant main effects for empathic and self-oriented utterances or gender with respect to the number of answer utterances. However, a significant main effect was found for empathic expressions i.e., $F(1, 62) = 7.53$ and $p < 0.01$, with respect to the number of non-answer utterances. No significant main effects were found for either self-oriented expressions or gender.

These results indicated that the peer agent's presence increased the number of both user answer and non-answer utterances in quiz-style multi-party dialogues. The agents' empathic expressions increased the number of user non-answer utterances.

4.4 User Memorization of Conveyed Information

We measured user memorization of biographical facts conveyed in dialogues by using the ratio of correct answers to the total number of blanks in the memory test. Figure 7 shows the results obtained from the memory test, where there are two kinds of ratios for correct answers (A) and (B):

- (A) is the rate of correct answers for hint statements that were about people in Sets-A used as quiz subjects and that were conveyed during the dialogue.
- (B) is the rate of correct answers for hint statements that were about people in Sets-B. These people were not used as quiz subjects.

The data set satisfied the assumptions of variance homogeneity ($p > 0.2$) and normality ($p > 0.2$) based on the F-test and the Kolmogorov-Smirnov test.

First, we compared two kinds of rates of correct answers (A) and (B) by using a paired t-test under each condition. We found that the rate of correct answers for (A) was significantly higher than that for (B) ($p < 0.01$). This was not a surprising result, and it indicated that the hint statements conveyed during dialogues remained in the participants' memory, and that participants took the memory test seriously.

We compared the rate of correct answers (A) between conditions (0) and (1) to evaluate what effect the peer agent's presence had on the results for the memory test. No significant differences between conditions (0) and (1) were found by applying a paired t-test to both the male and female groups.

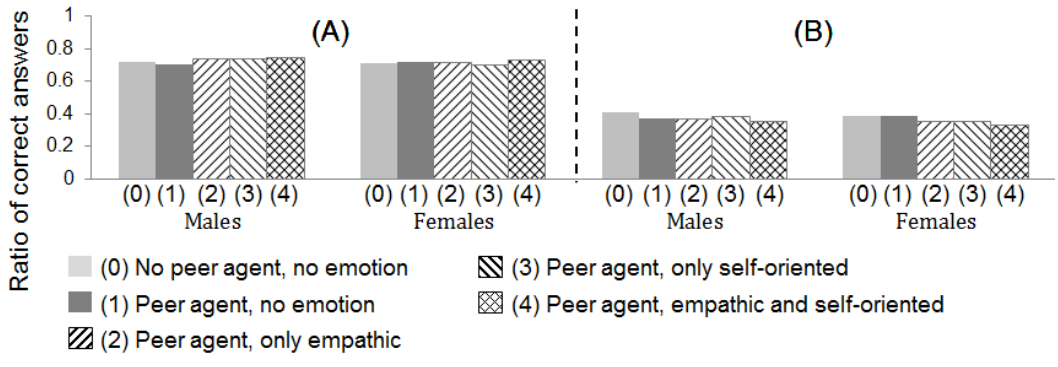

Fig. 7 Results from memory test: Ratio of correct answers to total number of blanks in memory test. (A) is ratio of correct answers for hint statements that were about people in Sets-A used as quiz subjects and that were conveyed during dialogues and (B) is ratio of correct answers for hint statements that were about people in Sets-B.

We compared the rate of correct answers (A) between conditions (1), (2), (3), and (4) to evaluate what effect agents' empathic and self-oriented expressions had on the results for the memory test. Three-factor ANOVA was conducted with two within-participant factors of empathic and self-oriented expressions and one between-participant factor of gender. As a result of ANOVA, no significant main effects were found for either empathic expressions, self-oriented expressions, or gender.

There was no evidence that the peer agent's presence and agents' emotional expressions influenced the users memorizing of conveyed information according to these results.

5. Discussion

Our study was chiefly concerned with activating quiz-style multi-party dialogues between multi-users and multi-agents. We were interested in what effect the presence of the peer agent and agents' emotional expressions had on activating communication.

We conducted experimental studies in line with this chief concern and found that the presence of a peer agent significantly improved user satisfaction and increased the number of both user answer and non-answer utterances. The peer agent's participation in a dialogue could raise user willingness to answer a quiz and utter non-answer expressions. We also found that agents' empathic expressions improved user satisfaction and user's positive ratings of the peer agent and these further increased the number of user non-answer utterances. The agent's empathy did not affect user willingness to answer a quiz but could raise user willingness to engage in interactions by uttering non-answer expressions. These results indicated that the peer agent's presence and agents' empathic expressions could raise user willingness to engage in interactions and activate communication in multi-party quiz-style dialogues.

The participants in our experiments were asked to carry out a predetermined number of dialogues, but they did not continue dialogues for as long as they wanted. It would be valuable to examine whether these factors can sustain user willingness to interact with agents for long periods in more voluntary settings. That will be one of our future directions of research.

We also found that there was no evidence that the peer agent's presence and agents' emotional expressions influenced user memorization of conveyed information during dialogues. Our findings would become more constructive if these factors could not only activate communication but also enhance user memorization. However, these results were not necessarily disappointing. At least, they indicate that the peer agent's presence and agents' emotional expressions may not have had a negative impact on user memorization. It was possible for these factors to have had a negative impact on user memorization by distracting their attention.

Similar results were reported in a previous study that investigated what effect an agent's small talk had on human users [27]. This previous work revealed that a real estate conversational agent enhanced a user's trust in the agent and a user's engagement with interaction by utilizing small talk, but that the agent's small talk did not change the user's willingness to pay the agent for the real estate. The agent's small talk was similar to the agent's emotional utterances in that both were kinds of social behaviors. Both this previous study and our investigations indicated that it is not straightforward to influence the performance of tasks such as buying real estate or memorizing biographical information by using agents' social behaviors.

If the peer agent's presence and agents' empathic expressions encourage users to maintain dialogues voluntarily, these factors might also improve user memorization. This is because user memorization might be improved by conveying more information to users during longer periods of interaction.

6. Conclusion

We experimentally analyzed how conversational agents activated human communication in thought-evoking multi-party dialogues between multi-users and multi-agents. This

meant the activation of communication by enhancing user engagement. We focused on quiz-style multi-party dialogues between two users and two agents as an example of such dialogues. We investigated how a peer agent's presence and agents' emotional expressions influenced user satisfaction, user ratings of the peer agent, and the number of user utterances.

The experimental results indicated that the peer agent's presence significantly improved user satisfaction and increased the number of user utterances. We also found significant effects where agents' empathic expressions improved user satisfaction and users' positive ratings of the peer agent and they further increased the number of user utterances. These results indicated that employing a peer agent and agents' empathic behaviors will activate communication in quiz-style multi-party dialogues in that they will promote user utterances and enhance user experience of the dialogue with the agents. Our findings should be useful for a broader class of applications such as educational agents and community facilitators. For example, we could activate discussions between students in group learning and activate communication between people in day-care centers.

We also found that there was no evidence that the peer agent's presence and agents' emotional expressions influenced user memorization of information conveyed during the dialogues. As discussed in Sect. 5, this result at least indicates that these factors may not have had a negative impact on user memorization.

Many directions for future work still remain. First, it would be worthwhile to examine what kinds of factors could sustain human users' willingness to interact with agents for long periods in more voluntary settings. Second, it would also be beneficial to examine what kinds of factors would improve user memorization of information conveyed during dialogues. These future studies should be helpful especially for educational agents. Third, we could clarify the role of agents by investigating cases where four dialogue participants were all human. Fourth, it would be more credible to use user's eye gaze behaviors [28] as a measure to activate communication. Fifth, a previous study has found that the effectiveness of vicarious learning depends on how well or poorly a peer agent performs [14]. It would be interesting to examine how user memorization is affected by how well or poorly a peer agent performs in multi-party quiz-style dialogues. Finally, we plan to extend our work to deal with various modalities such as speech, gestures, body postures, facial expressions, the direction of eye gazes, and agent representations (embodied or disembodied) to investigate what effects the other modalities have in thought-evoking multi-party dialogues.

Acknowledgments

We would like to thank Professor Makoto Imase of the Graduate School of Information Science and Technology at Osaka University for his helpful advice and suggestions. We would also like to thank Drs. Junji Yamato and Keiji Hirata of NTT Communication Science Laboratories for their encouragement and support. Thanks also go to Messrs. Akira Mori, Minako Sawaki, Toyomi Meguro, and Hiroaki Sugiyama of NTT Communication Science Laboratories for the constructive discussions we had with them. This work was supported by a Grant-in-Aid for Scientific Research on Innovative Areas in the "Formation of robot communication strategies" (21118004) made available by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

References

[1] V. Zue, S. Seneff, J. Polifroni, M. Phillips, C. Pao, D. Goodine, D. Goddeau, and J. Glass, "PEGASUS: A spoken dialogue interface for on-line air travel planning," Speech Commun., vol.15, pp.331–340, 1994.
[2] J. Allen, D. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent, "Toward conversational human-computer interaction," AI Magazine, vol.22, no.4, pp.27–37, 2001.
[3] J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, eds., Embodied Conversational Agents, MIT Press, Cambridge, MA, 2000.
[4] D. Traum and J. Rickel, "Embodied agents for multi-party dialogue in immersive virtual worlds," Proc. 1st Int. Joint Conf. Autonomous Agents and Multi-Agent Systems, pp.766–773, 2002.
[5] Y. Liu and Y.S. Chee, "Intelligent pedagogical agents with multiparty interaction support," Proc. Int. Conf. Intelligent Agent Technology, pp.134–140, 2004.
[6] J. Zheng, X. Yuan, and Y.S. Chee, "Designing multiparty interaction support in Elva, an embodied tour guide," Proc. 4th Int. Joint Conf. Autonomous Agents and Multiagent Systems, pp.929–936, 2005.
[7] S. Fujie, Y. Matsuyama, H. Taniyama, and T. Kobayashi, "Conversation robot participating in and activating a group communication," Proc. 10th Annual Conf. Int. Speech Communication Association, pp.264–267, 2009.
[8] R. Higashinaka, K. Dohsaka, S. Amano, and H. Isozaki, "Effects of quiz-style information presentation on user understanding," Proc. 8th Annual Conf. Int. Speech Communication Association, pp.2725–2728, 2007.
[9] C.Y. Chou, T.W. Chan, and C.J. Lin, "Redefining the learning companion: the past, present, and future of educational agents," Computers & Education, vol.40, no.3, pp.255–269, 2003.
[10] H. Maldonado, J.E.R. Lee, S. Brave, C. Nass, H. Nakajima, R. Yamada, K. Iwamura, and Y. Morishima, "We learn better together: enhancing elearning with emotional characters," Proc. 2005 Conf. Computer Support for Collaborative Learning, pp.408–417, 2005.
[11] S.D. Craig, B. Gholson, M. Ventura, A.C. Graesser, and the Tutoring Research Group, "Overhearing dialogues and monologues in virtual tutoring sessions: Effects on questioning and vicarious learning," Int. J. Artificial Intelligence in Education, vol.11, pp.242–253, 2000.
[12] J. Lee, F. Dineen, and J. McKendree, "Supporting student discussions: It isn't just talk," Education and Information Technologies, vol.3, no.3-4, pp.217–229, 1998.
[13] K. Stenning, J. McKendree, J. Lee, R. Cox, F. Dineen, and T. Mayes, "Vicarious learning from educational dialogue," Proc. 1999 Conf. Computer Support for Collaborative Learning, pp.341–347, 1999.
[14] M.T.H. Chi, M. Roy, and R.G.M. Hausmann, "Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning," Cognitive Science, vol.32, no.2, pp.301–341, 2008.
[15] J. Bates, "The role of emotion in believable agents," Commun. ACM, vol.37, no.7, pp.122–125, 1994.
[16] R.W. Picard, Affective Computing, MIT Press, 1997.
[17] E. Hudlicka, "To feel or not to feel: The role of affect in human-computer interaction," Int. J. Human-Computer Studies, vol.59,

no.1-2, pp.1–32, 2003.
[18] H. Prendinger and M. Ishizuka, eds., Life-Like Characters: Tools, Affective Functions, and Applications, Springer-Verlag, 2004.
[19] S. Brave, C. Nass, and K. Hutchinson, "Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent," Int. J. Human-Computer Studies, vol.62, no.2, pp.161–178, 2005.
[20] H. Prendinger, J. Mori, and M. Ishizuka, "Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game," Int. J. Human-Computer Studies, vol.62, no.2, pp.231–245, 2005.
[21] T.W. Bickmore and R.W. Picard, "Establishing and maintaining long-term human-computer relationships," ACM Trans. Computer-Human Interaction, vol.12, no.2, pp.293–327, 2005.
[22] R. Higashinaka, K. Dohsaka, and H. Isozaki, "Effects of self-disclosure and empathy in human-computer dialogue," Proc. 2008 IEEE Workshop on Spoken Language Technology, pp.109–112, 2008.
[23] P. Gebhard, M. Klesen, and T. Rist, "Coloring multi-character conversations through the expression of emotions," Lect. Notes Comput. Science (Tutorial and Research Workshop on Affective Dialogue Systems), vol.3068, pp.128–141, 2004.
[24] R. Higashinaka, K. Dohsaka, and H. Isozaki, "Learning to rank definitions to generate quizzes for interactive information presentation," Proc. 45th Annual Meeting of the Association for Computational Linguistics (Poster Presentation), pp.117–120, 2007.
[25] T. Matsubayashi and T. Yamada, "A force-directed graph drawing based on the hierarchical individual timestep method," Int. J. Electronics, Circuits and Systems, vol.1, 2, no.17, pp.116–121, 2007.
[26] S. Kopp, L. Gesellensetter, N.C. Krämer, and I. Wachsmuth, "A conversational agent as museum guide: Design and evaluation of a real-world application," Lect. Notes Comput. Sci. (the 5th Int. Working Conf. Intelligent Virtual Agents), vol.3661, pp.329–343, 2005.
[27] T. Bickmore and J. Cassell, "Relational agents: A model and implementation of building user trust," Proc. Conf. Human Factors in Computing Systems, pp.396–403, 2001.
[28] Y.I. Nakano and R. Ishii, "Estimating user's engagement from eye-gaze behaviors in human-agent conversations," Proc. 15th Int. Conf. Intelligent User Interfaces, pp.139–148, 2010.

Kohji Dohsaka received the B.S. and M.E. degrees in Information and Computer Science from Osaka University in Osaka, Japan, in 1984 and 1986, respectively. He received the doctorate degree in Information Science from the Japan Advanced Institute of Science and Technology in Ishikawa, Japan, in 2004. He joined Nippon Telegraph and Telephone Corporation (NTT) in 1986, and he was a visiting researcher at the Massachusetts Institute of Technology, in Cambridge, MA in 2002 and an invited professor at Osaka University from 2008 to 2011. He has been a professor with Akita Prefectural University in Akita, Japan, since 2012. His research interests include dialogue systems and natural language processing.

Ryota Asai received the B.S. and M.E. degrees in Information Science and Technology from Osaka University in Osaka, Japan, in 2007 and 2009, respectively. He joined Denso Corporation in 2009. He is currently with Information & Communications Technology ENG. EIV. 3, of Denso Corporation. His research interests include dialogue systems and computer networks.

Ryuichiro Higashinaka received the B.A. degree in Environmental Information, the Master of Media and Governance, and the Ph.D. degree from Keio University, in Kanagawa, in 1999, 2001, and 2008, respectively. He joined NTT in 2001. His research interests include building question answering systems and spoken dialogue systems. He was a visiting researcher at the University of Sheffield in the UK from November 2004 to March 2006. He was a part-time lecturer at Osaka Electro-Communication University from 2006 to 2008. He has been a part-time lecturer at Keio University since 2010.

Yasuhiro Minami received the M.Eng. and Ph.D. degrees in Electrical Engineering from Keio University in Kanagawa Japan, in 1988 and 1991, respectively. He joined Nippon Telegraph and Telephone Corporation (NTT), in Tokyo, Japan, in 1991, where he worked on robust speech recognition. He was a Visiting Researcher at the Massachusetts Institute of Technology in Cambridge, U.S.A., from 1999 to 2000. Since February 2000, He has been with NTT Communication Science Laboratories, in Kyoto, Japan, since February 2000. He is currently the group leader of the Communication Environment Research Group. His interests include speech recognition, dialogue systems, and language development.

Eisaku Maeda received the B.S. and M.S. degrees in Biological Science and the Ph.D. degree in Mathematical Engineering, from the University of Tokyo in Tokyo, Japan, in 1984, 1986, and 1993 respectively. He joined the Nippon Telegraph and Telephone Co. (NTT) in 1986 and was a visiting scholar at the of Physiological Laboratory, of the University Cambridge, UK, from 1995 to 1996, an invited professor at Osaka University in Osaka, Japan, from 2002 to 2013, and an invited professor at Keio University, in Yokohama, Japan, from 2004 to 2006. He is currently a director of NTT Communication Science Laboratories and a part-time lecturer at the University of Tokyo. His research interests are in machine learning, pattern recognition, and bioinformatics.