How Multiple Concurrent Users React to a Quiz Agent Attentive to the Dynamics of Their Game Participation

Hung-Hsuan Huang*
Takuya Furukawa,
Hiroki Ohashi,
Toyoaki Nishida
Graduate School of Informatics, Kyoto University, Japan
{huang, furukawa, ohashi, nishida}@ii.ist.i.kyoto-u.ac.jp

Aleksandra Cerekovic,
Igor S. Pandzic
Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia
{aleksandra.cerekovic, Igor.Pandzic}@fer.hr

Yukiko Nakano
Department of Computer and Information Science, Seikei University, Japan
y.nakano@st.seikei.ac.jp

ABSTRACT

This paper presents a quiz game agent who is attentive to the dynamics of multiple concurrent participants. The attentiveness of this agent is meant to be achieved by an utterance policy that determines the nature of the utterance and whether, when, and to whom to utter. Two heuristics are introduced to drive the policy: the interaction atmosphere (AT) of the participants and the participant who tends to lead the conversation (CLP) at a specifi time point. They are estimated from the activeness of the participants' face movements and acoustic information during their discussion of the answer. In order to the inherent drawback of a 2D agent that makes it difficult for multiple concurrent users to distinguish the focus of its attention, a physical pointer is also introduced. This system is then evaluated using questionnaire investigation and video data analysis. The joint results of the experiments indicated that the methods for estimating AT and CLP worked. The participants pay more attention to the agent and participate in the game more actively if the indication of the pointer is more comprehensive.

Categories and Subject Descriptors

H.1.2 [User/Machine Systems]: Human factors; H.5.2 [User Interfaces]: Evaluation/methodology; I.2.1 [Applications and Expert Systems]: Games, Natural language interfaces

General Terms

Design, Experimentation, Human Factors

Keywords

Virtual agent applications and empirical studies, Multimodal interaction, Verbal and nonverbal expressiveness, Agents in games and virtual environments

*Dr. Huang is now a postdoctoral researcher in Seikei University.

Cite as: How Multiple Concurrent Users React to a Quiz Agent Attentive to the Dynamics of Their Game Participation, Huang, H.H., Furukawa, T., Ohashi, H., Cerekovic, A., Pandzic, I., Nakano, Y., and Nishida, T., *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. 1281-1288

1. INTRODUCTION

Embodied conversational agents (ECAs) are lifelike virtual characters that can engage in face-to-face conversations with human users in daily life situations. Because of this inherent characteristic, ECAs are ideal candidates for the interfaces of public services where the users are not expected to be skillful in operating complex computer systems. Making ECAs go public is an emerging challenge. For example, Max [5] is a guide agent in a computer museum. He can perform real-time feedback behaviors from the visitors' keyboard inputs and track multiple visitors with the skin color feature. Sgt. Blackwell [8] is a virtual human exhibited in a design museum and answers questions from the visitors for free.

The presence of multiple concurrent users is virtually a must-happen situation in public exhibitions but is seldom addressed in previous works. Traum [11] provided the principal literature on general issues in realizing multiparty human-agent interactions. Contrary to dyadic dialogs that involve only a speaker and addressee, in multiparty dialogs, the distinction among conversation participants' roles, addressee, overhearer, and speaker is necessary. Since there are potentially more interlocutors to acquire dialog turns from and transfer dialog turns to, managing the communication fl w in a multiparty dialog is more complex. The main difficulty was posed by the possibility of the users interacting with each other, which is difficult for the agent to understand.

ECA awareness of multiple concurrent users could be an important factor in making them more humanlike. Owing to a slightly different use of the concept, *attentiveness* from regular English, we redefin it as the following: *An agent is considered to be attentive to its users if it not only tries to achieve its goal but also takes users' benefits into consideration in deciding its actions, and only if the users notice that and react positively.* The theme of this paper is the proposal of an approach for realizing a quiz agent who is attentive to multiple concurrent users. We further formalize the quiz game context as the following: The agent's goal is to make the game proceed and allow more participants to join it in a limited period of time. The attitude of the agent toward the participants is set to be fair, i.e., the agent is not evaluated according to the score of the participants; it does not try to help the participants on its own, and it does not try to mislead the participants, either. On the other hand, the participants (users) are supposed to want to enjoy the game and do not want to be disturbed. The attentive quiz agent is then designed by the following principles:

1. If the participants do not answer the quiz for a long time, the agent tries to make the game proceed by urging them to answer or indicating the availability of a hint.

2. If the participants are inactive in the quiz game, the agent tries to stimulate them.

3. When the agent makes an utterance expecting positive reactions from the participants, the participant who is most likely to have the greatest influenc on the other group members is chosen as the addressee.

4. The agent avoids making an utterance at an annoying time, e.g., when the participants are actively engaged in discussion.

In the fiel of ECA research, the ultimate goal is to realize humanlike communicative abilities using a computer. As in the case of other scientifi research works, evaluations are required to verify how effective an ECA system is. However, the human likeness of an artifact is a personal opinion and it is difficult to measure the same objectively, and thus nowadays, ECA researches usually involve subject evaluations. These empiric works can be typically classifie into the following categories.

- statistical analysis on the language usage of the users from a large log corpus [5, 8]
- evaluations on participants' perception of the appearance of the agent using paper- or web-based questionnaires [10, 13]
- direct evaluations of the ECA system using rating questionnaires [4]
- cognition tests using the techniques of statistical psychology [9]

Nevertheless, very few of them actually verifie the exact effectiveness of the ECA's behavior in influencin the participants' reactions during their interaction with the agent. This paper presents a prototype of the attentive quiz agent and an evaluation methodology involving combined analyses on regular questionnaires and video data in comparing the agents with and without attentiveness. The video data are analyzed quantitatively and qualitatively with regard to the participants' reactions to the agents' behaviors. From the analysis results, the proposed quiz agent was shown to be attentive.

2. RELATED WORKS

Most of the contemporary ECA research works that address multiparty interaction issues focus on multi-agent/single-user configu rations. For example, a car presentation team consisting of a salesman agent and a customer agent [1], a tactical training system for soldiers who are going to be deployed abroad [12], and a cellular phone presentation system with two salesman agents who estimate the user's interest focus from his/her gaze pattern [3].

In multi-user configurations the conversation situation is more unpredictable and thus more difficult to realize. Gamble [7] is a dice game where an agent interacts with two human players. The round-based game rules fi ed the system's scenario and resulted in basically three dyadic interactions. To prevent unreliable speech recognition in public exhibitions, Max [5] used a keyboard to acquire inputs from the museum visitors; however, the limitation of this method is that it allows Max to interact with the visitors only on a one-on-one basis. It counts the number of multiple visitors standing in front of him on the basis of skin color features, but is not able to precisely track the visitors if they stand closely. In a more recent work [2], the authors introduced a virtual receptionist setup. This agent merged multimodal information from a face detector, microphone array, and speech recognizer to handle dynamic engagement if there is more than one person in its view field The method they used is the straightforward solution, namely, tracking each conversation participant, understanding what everyone said, and making the agent engage in the conversation with a management mechanism of dialog moves.

3. THE ATTENTIVE QUIZ AGENT

This project has been launched in collaboration with the National Food Research Institute (NFRI) of the Japanese government. In order to disseminate their research results and promote the importance of food safety among the public, this institute holds several exhibitions every year. The firs prototype developed for NFRI was a relatively simple quiz agent who is not attentive to the game participants.

The quiz game proceeds as follows. The agent reads out the question of a quiz, while the text of the question and the answer choices are shown on the screen as well. The participants press one of the graphical buttons shown on a touch panel to answer the quiz. The agent then announces the correct answer and comments on the performance of the participants. After that, the agent proceeds to the next question. In each session, there are 10 questions and if the participants press the hint button, the agent explains the hint for the current question.

This game kiosk was displayed in four exhibitions, where 290 groups (860 people) of visitors played the quiz. We derived the following finding through our observations of the game participants' interactions with the agent.

- Most of the game participants come in groups and answer the quizzes as a collaborative task.
- The participants usually answer the questions after a group discussion. The atmosphere of this discussion changes dynamically, i.e., participants sometimes engage in discussions enthusiastically and sometimes deliberate individually.
- There is usually one participant leading the discussion and coordinating the fina answer of a certain question.
- The participants guffaw or exclaim when the announced correct answer is surprising or when the agent says or does something silly, e.g., a strange and unnatural pronunciation by the text-to-speech engine or an awkward gesture.

3.1 Attentive Utterance Policy

Considering the listed phases of the quiz game, two situations need to be improved. First, during the period after the agent issues the quiz question and before the participants answer it, the agent simply stands without doing anything. Second, the agent issues the next question directly after commenting about the answer to the current question. The utterance policy is then designed to address these two situations according to the conversation of the participants.

After issuing a question and before the participants answer it: If the participants keep interacting with each other actively, the agent does nothing. If the activity is initially high but declines later, in order to make the quiz game progress and stimulate activity among the participants, the agent urges the participants to answer or reminds them about the availability of a hint (*Urge utterances* hereafter). However, because Urge utterances have to be designed on

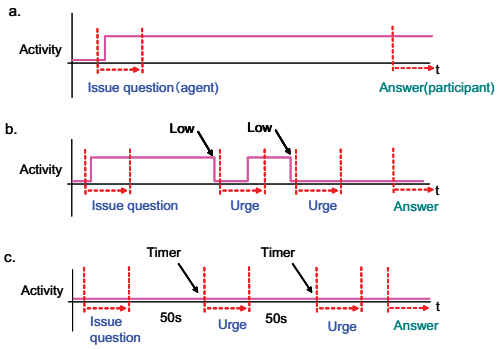

Figure 1: Utterance policy: after issuing the question. (a) The activity is always high. (b) The activity is high at first but declines. (c) The activity never becomes high

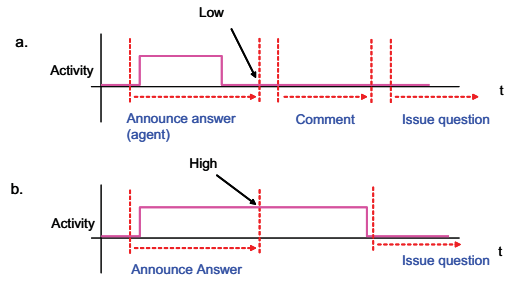

Figure 2: Utterance policy: after answer announcement. (a) The activity is low when the answer announcement ends. (b) The activity is high when the announcement ends.

the basis of the quiz question, the variations are limited. They are uttered by the agent at most twice in the period of one quiz. If the interactions among the participants are never active, Urge utterances are triggered by a 50-second timer. The relationships between time, participants' activity, and the behaviors of the agent are shown in Figure 1. Since reactions (pressing the hint button or answering the question) are expected from the participants when the agent urges them, the addressee of an Urge utterance is set to be the participant leading the group at that time.

After announcing the answer and before the next question: If the activity of the participants becomes low while the agent is announcing the answer, the agent comments about the answer and cheers up or praises the participants (*Comment utterances* hereafter). If the participants are actively conversing when the answer announcement ends, the agent suspends the issue of the next question or the fina summary (*Proceed utterances* hereafter) until the participants calm down. The relationships between time, participant activity, and the behaviors of the agent are shown in Figure 2. Figure 3 is an interaction that actually occurred in the subject experiment (see section 4), and shows how the attentive policy works.

3.2 Participant Status Estimation

In order to implement the utterance policy described in the last section, it is necessary to measure how active the participants' conversation is and identify the person who is most likely to lead the conversations in the group at a certain moment. We then defin two heuristics, *Interaction Activity (AT)* and *Conversation Leading Person (CLP)*, as follows:

Interaction Activity (AT): It indicates whether the users are active in their interactions. *High* and *low* are the two possible measured statuses. AT is high when all of the members of the participant group react to an utterance made by one of them with successive utterances and intense face movements. AT is low otherwise.

Conversation Leading Person (CLP): It is the participant who is most likely to lead the group at a certain time point. It is estimated by calculating who spoke the most and initiated the most AT in the group.

The computation of AT and CLP is reset at the beginning of each quiz on the basis of the assumption that participant activity depends heavily on the quiz. The intensity of face movements is approximated from the face orientation information measured by a webcam and Omron's OkaoVision [1] face detection library. C_t indicates how much attention each participant paid to the screen at a certain time point t; it is computed from N sampling data by the following equation.

$$V(t) = (x_t, y_t) \qquad where \; -\frac{\pi}{2} \le x_t, y_t \le \frac{\pi}{2}$$

$$V_{max} = (x_{max}, y_{max})$$

$$f(V(t)) = \left\{ \begin{array}{ll} 1 & if \quad -V_{max} \le V(t) \le V_{max} \\ 0 & if \quad -V_{max} > V(t) \quad or \quad V_{max} < V(t) \end{array} \right\}$$

$$C_t = \frac{\sum_{k=0}^{N} [(N-k)^2 \times f(V(t-k))]}{\sum_{k=0}^{N} (N-k)^2} \qquad where \quad t \ge N$$

Here, $V(t)$ is the face orientation of a participant at time t (0 when the direction is toward the camera), while x_t and y_t represent the angle in horizontal and vertical directions within the range $\pm\pi/2$. V_{max} is the threshold to judge whether the participant is looking at the screen at t (the angles in horizontal and vertical directions: x_{max} and y_{max}). $f(V(t))$ denotes whether the participant is looking at the screen; $f(V(t)) = 1$ when (s)he is looking at the screen and $f(V(t)) = 0$ otherwise. When C_t is lower than the value of α, the participant is considered to be not paying attention to the screen (the agent) and showing intense face movements.

These parameters are adopted with the assumption of using the system in the experiment space shown in Figure 4; the number of participants is fi ed at three. Because the width of the screen is nearly the same as that of the whole space, and its height (1.8 m) is assumed to be higher than most participants, the participants are assumed to face the screen orthogonally when they are looking at it. Therefore, x_{max} and y_{max} are set in the middle of 0 and $\pm\pi/2$, that is, $\pm\pi/4$ is used to distinguish the directions of the screen and the other participants. The other parameters are adopted according to the empiric results. When $N = 12$ and $\alpha = 0.7$, the appropriate results could be obtained.

Whether the participants are speaking or are engaged in a conversation is detected only through acoustic information. A two-second silence is used as a threshold to partition speaking segments from the voice streams of the microphone attached to each participant. The information from all participants is combined to detect whether a conversation takes place in the case that their successive utterances do not break for longer than two seconds. A conversation sequence is judged to have high AT if any one of the participants, except the current speaker, has active face movements (Figure 5). The changing AT status is used to further partition the conversation segments; the participant who is the starting point of each AT period is considered to have initiated the AT once.

[1] http://www.omron.com/r_d/coretech/vision/okao.html

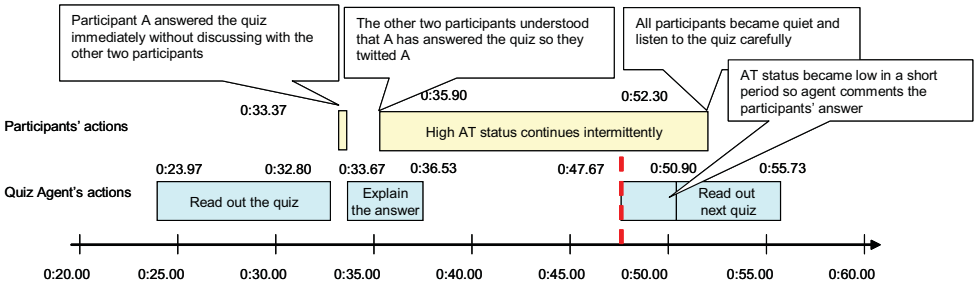

Figure 3: One example showing how the attentive quiz agent's utterance policy works

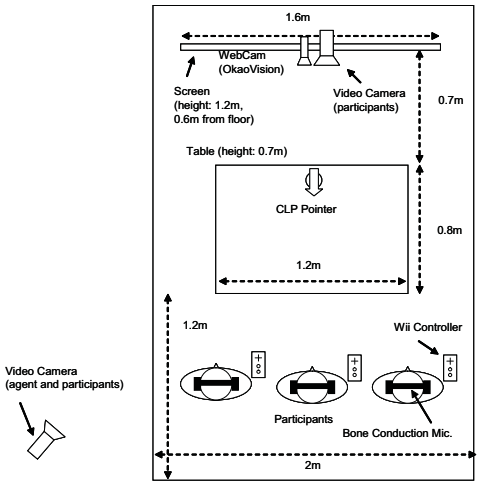

Figure 4: The experiment space layout of the attentive quiz agent

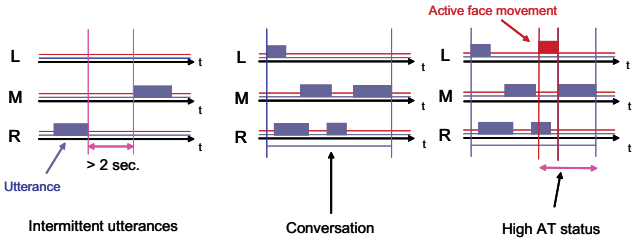

Figure 5: The criteria to judge a conversation sequence and high AT status. "L," "M," and "R" denote the three different participants

CLP is then estimated by tracking how many times each user spoke and how many times he or she initiated an AT in the group. Each participant is ranked according to these two criteria. The participant who spoke the most is assigned three points, while the one who spoke the least is assigned one point. The participant who initiated the most AT is assigned three points and the one who initiated the least AT is assigned one point. These two scores are then summed with the same weight, and the participant who has the most points is judged as the CLP at that moment. The system constantly computes the CLP and thus, there is always one CLP at any moment. There may be some periods when all of the participants are not speaking but are paying attention to the system. We assume that even when there is no conversation in progress, the participants should be thinking about the answer on the basis of their last conversation, which should be counted as being influence by the last CLP participant. In other words, we assume that even in a quiet period, there is a CLP participant (last one).

3.3 Implementation

All system functionalities are distributed into concurrently running components that are connected in the topology shown in Figure 6. As shown in Figure 7, each participant is equipped with a Nintendo Wii remote controller, so that any one of them can answer the quiz directly without the constraints of distance from the touch panel that may influenc the computation of CLP. Each one of them is also equipped with a bone conduction microphone to prevent the voices of the other participants from being mistakenly collected.

Due to the "Mona Lisa Effect" of 2D agents mentioned in [6], the users cannot correctly recognize the gaze directions of an agent, except the middle user. A physical pointer is therefore introduced to enable the quiz agent to indicate its focus of attention.

Each microphone is connected to an *Audio Processing* component that digitalizes the voice, extracts the sounds within the human voice frequency range, and determines whether that user is speaking by the voice power cue. The *Conversational Status Detection* component judges whether there is a conversation existing among the participants via the overlapping and successive relationship between the participants' utterances. A 2-second silence threshold is used to distinguish two segments.

The video information captured by webcam (640 x 480 pixels, 30 fps) is processed by the *Video Processing* component, mainly utilizing the OkaoVision face detection library. Recognized face orientations of the users are sent to the *Input Understanding* component for further processing. Because the OkaoVision library fails to recognize faces outside its range ($\pi/3$ in the horizontal direction and $\pi/6$ in the vertical direction), to compensate for this and enumerate the jitters, the CamShift method in Intel OpenCV [2] and Kalman fil ter are applied. The face direction is recognized at around 4 fps on the computer used by us.

The face movement intensity information and the conversation status information are then combined by the *Input Understanding* component to estimate AT and CLP. Current AT and CLP are used to judge when what should be done to whom by the *Dialog Manager* component; animation commands are then generated by it to drive the *Character Animator* component to render CG character animations.

[2]http://sourceforge.net/projects/opencvlibrary/

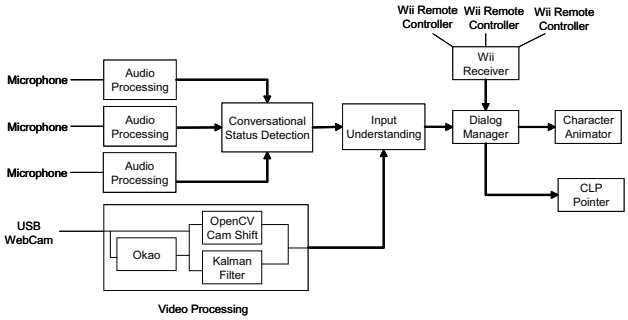

Figure 6: The system architecture of the attentive quiz agent

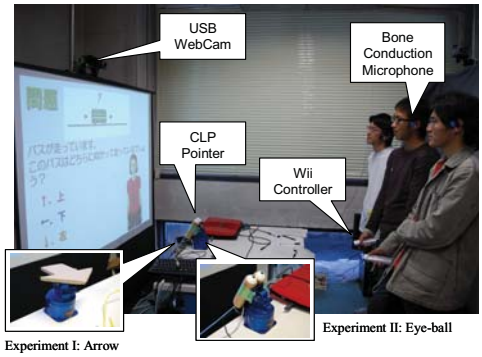

Figure 7: The sensor device configuration of the attentive quiz agent

4. EVALUATION EXPERIMENTS

A series of subject experiments was conducted to evaluate the effectiveness of the attentive quiz agent. We considered the functionalities of the CLP pointer—attracting the participants' attention and indicating the addressee of the agent's utterances. The shape of the pointer could have a great influenc on the participants' reactions. Therefore, in the evaluation experiment for the attentive quiz agent, two shapes of CLP pointers are adopted. One of them is simply an arrow, but the other one has two ping-pong balls marked with black dots on top (eyeball hereafter, Figure 7). They are investigated in two experiments, experiment I with the arrow pointer and experiment II with the eyeball pointer, respectively.

In each experiment, the attentive quiz agent is compared with the other agent called the "fi ed timing agent." It is exactly the same as the attentive quiz agent, except for the fact that the utterance timings are fi ed and the addressee of the CLP pointer is randomly decided. The relationship between the 2D graphical agent character and the physical CLP pointer is not explicitly specifie in the instruction, but the participants are instructed that when the pointer is pointing to one of them, it means that the 2D character is only talking to that person, and when the pointer is pointing upward, it means that the 2D character is talking to all of them. The differences between the attentive quiz agent and fi ed timing agent are detailed in Table 1.

The experiment participants are recruited from the university campus with only one prerequisite: they must enroll as three-people groups. To achieve counterbalance, the order of the internal algorithms, external appearance (color or clothes), and the quiz contents of the agents and the session order are switched in every session. Since there are three variables in this case, eight groups of participants are required in each experiment.

Table 1: The different settings of the attentive quiz agent (Attentive) and fixed timing quiz agent (Fixed) in the experiments

Utterance	Attentive	Fixed
Urge	policy	every 50 seconds (at most twice)
Comment	policy	immediately after answer announcement
Proceed	policy	immediately after comment
CLP Pointer	**Attentive**	**Fixed**
Urge	CLP	random
Otherwise	upward	upward

Eight groups (average age 21.3 years, 18 males and 6 females) of participants are chosen randomly for experiment I; the other eight groups (average 21.9 years, 21 males and 3 females) participated in experiment II. Each group plays a quiz game with an agent in one session and the comparable system in the other session. In order to facilitate active conversations among the participants, they are instructed that the reward varies according to their performance in the game. Questionnaires are administered immediately after each session.

4.1 Questionnaire Results

The questionnaire results, the Wilcoxon signed-rank test results of each experiment, and the Mann-Whitney U test results comparing the attentive quiz agent in the two experiments are summarized in Table 2. In both the experiments, the participants paid more attention to the movements of the attentive quiz agent's CLP pointer (Q10, I: p = .08, II: p<.01). This shows that the participants are conscious of the different meanings of the pointer's indications in the case of the attentive agent and fi ed timing agent. Moreover, in both the experiments, and especially in experiment II, the participants felt uncomfortable about the CLP pointer (Q12, I: p = .20, II: p = .02). A possible reason for this is that the shape of an eyeball is too offensive; hence, the participants felt that they were being looked at by somebody despite the fact that it attracts more attention. For the same reason, it seems that the eyeball pointer is more comprehensive (Q11, U test, p = .08); the participants themselves paid more attention to the pointers and thus felt that the agent paid more attention to them (Q8, II: p = .09, U test: p = .03).

In the questions related to utterance timings, no significan differences were found between the attentive quiz agent and fi ed timing agent. With regard to Q9, "The progress of the game was smooth (I: p = .11, II: p = .08)," the participants tended to feel that the game was not smooth with the attentive quiz agent. Since the fi ed timing quiz agent always makes comments immediately after it announces the correct answer and then immediately proceeds to the next question without waiting for the participants' active discussions to calm down, the participants may have an impression of the fi ed timing agent as being *faster*. If the participants mistakenly interpret the meaning of *smooth* as *fast*, it could lead them to develop an impression of the attentive quiz agent as being *not smooth*. Because the objective of the attentive quiz agent's utterance policy is not to make the quiz game progress *faster*, this may not be considered as a failure.

On the other hand, with regard to Q5, "The discussion was active (I: p = .86, II: p=.07)," and Q13, "There were silent periods in the session (I: p = .05, II: p = .68)," a positive shift was observed in the results for the attentive quiz agent from experiment I to II. Therefore, we assume that the eyeball CLP pointer seems to stimulate the participants' conversation more successfully.

Table 2: Summary of the seven-point (1 being the lowest and 7 being the highest degree) questionnaire results. M_F and M_A columns are the median of the fixed timing and attentive quiz agents. The numbers within parentheses are the values of inter-quartile deviation. The p columns are the results of the two-tailed Wilcoxon signed-rank test. The p_U column is the Mann-Whitney U test result that compares experiments I and II

		Experiment I			Experiment II			I vs II
Q	Question	M_F	M_A	p	M_F	M_A	p	p_U
1	The character was friendly.	4.0(1.00)	4.0(1.50)	0.199	4.0(1.00)	4.0(1.00)	0.954	0.531
2	The character's utterances were annoying.	4.0(2.00)	3.0(1.25)	0.159	4.0(1.50)	3.5(2.00)	1.000	0.531
3	The character was passive.	2.0(1.00)	2.0(1.00)	0.268	2.0(1.00)	3.0(1.50)	0.187	0.760
4	The character's acted in response to our status.	4.5(1.13)	4.0(1.00)	0.463	5.0(1.00)	4.0(1.00)	0.299	0.643
5	The discussion was active.	6.0(1.00)	6.0(0.63)	0.855	5.0(1.13)	6.0(0.63)	**0.068**	0.200
6	I often considered the question alone.	2.0(1.00)	2.0(1.50)	0.177	3.0(1.50)	3.0(1.00)	0.501	0.983
7	The character's behaviors stimulated our discussion.	4.5(1.13)	5.0(0.50)	0.793	5.0(1.63)	5.0(1.50)	0.403	0.430
8	The character paid attention to us.	3.0(1.00)	3.5(1.00)	0.792	3.0(1.50)	4.5(1.00)	**0.087**	**0.027**
9	The game progress was smooth.	5.0(1.13)	4.5(2.00)	0.109	5.0(1.13)	4.0(1.00)	**0.081**	0.884
10	I paid attention to the movement of the pointer.	2.5(2.00)	4.0(2.00)	**0.075**	2.0(1.63)	5.0(0.88)	**0.002**	0.992
11	The indication of the pointer was comprehensive.	2.0(1.50)	2.0(1.13)	0.835	2.0(1.00)	3.0(1.00)	**0.037**	**0.080**
12	The indication of the pointer was incongruous.	3.0(1.63)	4.0(2.00)	0.204	4.0(1.13)	5.0(1.50)	**0.024**	0.588
13	There were silent periods in the session.	2.0(1.50)	3.0(2.50)	**0.053**	4.0(1.50)	4.0(1.50)	0.678	0.826
14	I would like to respond to the character's urgings.	5.0(1.00)	5.0(1.50)	0.256	4.0(1.30)	5.0(1.50)	0.632	0.892

4.2 Video Analysis Results

The video data are recorded from two cameras set up at the positions shown in Figure 4. In order to reduce the tendency caused by the subjective judgment of individual annotators, four annotators who are familiar with video annotating but are not involved in the development of this study are asked to annotate the video data (32 sessions, 5 hours and 28 minutes in total). The video data of two groups in experiment I and two groups in experiment II are selected randomly and assigned to the annotators. Every annotator is asked to annotate eight sessions with the tool iCorpusStudio [3]. The objectives and algorithms of this study were not included in the instructions for the annotators. The annotators are instructed to annotate the video data according to the following conditions:

Utterance timings: to see whether the agent makes the utterances at the appropriate timings. The short periods when the agent just starts to make *Proceed*, *Urge*, and *Comment* utterances are annotated. Since the firs question is issued immediately after a long greeting in any case, the situations when the agents are issuing the firs question are not counted. The following labels are available for timing annotations:

Smooth (S): nothing unexpected happened; the quiz game proceed smoothly.

Abrupt (A): the agent spoke to the participants at an abrupt timing and disturbed their active conversation. The participants either ignored the agent's utterances and continued their conversation, or interrupted their current conversation suddenly and paid attention to the agent.

Tardy (T): the agent talked to the participants after the following situation: the system seemed to be freezing and the participants looked confused, wondering why the game was not proceeding.

Participants' attention: to investigate whether the participants paid attention to the agent's utterances. The periods during which the agent is making *Urge* and *Comment* utterances are annotated. The short period just after the agent begins to talk is ignored in this annotation. Since the *Proceed* utterances are relatively longer and are important to the participants, they always pay attention to the agent. Therefore, the Proceed utterances are not counted here. The following labels are define for this annotation:

Listen (L): at least two participants were listening to the agent's utterance, or at least one of the participants replied to the agent, commented on the agent's utterance, as well as any other observable reaction to the agent.

Ignore (I): at least two participants were engaged in their own conversation and ignored the agent's utterances.

Conversation Leading Person: when the CLP pointer is in action, whether the participant it is pointing to is the person who is leading the conversation of the group at that time point. If it is not very clear who the CLP is at this point, then use the CLP of the whole session as the criterion. The following labels are defined

Conversation Leading Person (C): the person pointed to is the CLP at this time point.

Not Conversation Leading Person (NC): the person pointed to is not the CLP at this time point.

Unclear: the cases when the person pointed to is not observable owing to the viewpoint of the camera and the activity of the participants. These cases are not counted in the analysis.

The comparison of utterance timings between the attentive quiz agent and fi ed timing quiz agent is depicted in Table 3. According to the observation, there was nearly no difference between these two types of agents in making smooth utterances to ensure that the game proceeds (P: 70.0%:68.9%). On the other hand, in the case of Urge and Comment utterances, the attentive quiz agent tends to make a smooth impression more often (U: 72.6%:56.0%, C: 63.2%:51.7%). The difference was particularly high in Urge utterances. This difference can be attributed to the different properties of the two types of utterances. Although the total number is low, the attentive quiz agent created the impression of tardy timings of utterances more often (10:1); this coincides with the results from the questionnaires.

The investigation of the influence of different combinations of utterance timings and types on the participants' attention is shown

[3]http://www.ii.ist.i.kyoto-u.ac.jp/iCorpusStudio/index.html

Table 3: Comparison of the frequency of smooth utterance timings between the attentive quiz agent and fixed timing quiz agent. The results of experiments I and II are combined. The numbers without remarks represent the number of times.

	Attentive				Fixed			
Timing	P	U	C	Total	P	U	C	Total
Smooth	112	45	67	224	104	14	78	196
Abrupt	42	17	35	94	47	10	73	130
Tardy	6	0	4	10	0	1	0	1
Smooth(%)	70.0	72.6	63.2	68.3	68.9	56.0	51.7	59.9

Table 4: Influences of different combinations of utterance timings and types on the attention of the participants. "L" and "I" indicate whether the participants listen to the agent or ignore it, respectively. The results combine the findings of experiments I and II, and the numbers without remarks represent times. The result of tardy utterances is omitted owing to too few samples (5 in total)

	Comment			Urge			Total
Timing	L	I	L(%)	L	I	L(%)	L(%)
Smooth	129	18	87.8	54	6	90.0	88.4
Abrupt	41	65	38.7	17	11	60.7	43.3

in Table 4. From these data, we can see that when the utterances are made at smooth timings, the participants tend to pay attention to the agent and listen to its utterances (C: 87.8%, U: 90.0%). Contrary to this, when the utterances are made at abrupt timings, the probability of the participants stopping their own conversations and listening to the agent decreases (C: 38.7%, U: 60.7%). The reason that Comment utterances are particularly ignored may be because participants consider them to be less important. They were often surprised if they their answer was wrong and discussed the matter even after the answer announcement.

The difference in the frequency of the agent being ignored according to different shapes of the CLP pointer is shown in Table 5. From the above, in terms of their ability to attract the participants' attention, the agents can be arranged as follows: eyeball pointer > arrow pointer > no pointer. On the other hand, the findin that the utterances made to the CLP were ignored less often implies that the hypothesis that talking to the CLP should be able to cause the group to react more easily was correct.

4.3 CLP Estimation

In order to measure the accuracy of the CLP estimation method, the question, "Who leads our group's discussion during the game?" is included in the questionnaire. The annotators are also asked to judge which participant tended to lead the discussions during the whole session. These results are compared with the estimation of the system in terms of the time ratio of each participant in Table 6.

The candidates of *correct answers* should be either the ones from the participants themselves or the ones from the annotators; however, by comparing the estimation of the system, we found that there was around a 50% possibility of coincidence. In addition to this, the comparison between the judgment of the participants and the annotators also had around a 50% possibility of coincidence. These results imply the difficulty of judging who is leading the conversation during a relatively long time (the whole session);

Table 5: The influences of the different shapes of the CLP pointer on the participants' attention and whether or not the addressee is the current CLP. "C" and "NC" indicate that the pointer points to the person who is the current CLP or does not point to the current CLP respectively. The numbers without remarks represent times. The data of "Comment" utterances that have no CLP pointer movements is listed in the last column for reference.

	Arrow		Eyeball		None
	C	NC	C	NC	—
Ignore	4	4	1	4	84
Listen	21	11	13	16	174
Ignore (%)	16.0	25.0	7.1	20.0	32.6

they also suggest that the accuracy of the estimation done by the system is of a similar level to that done by humans. On the other hand, although the social relationship among the participants can be considered to have a great influenc on their answers to the questionnaire, it was not clear how it affected the participants in this experiment.

The annotators are also required to evaluate the accuracy of the CLP estimation while the CLP pointer is in action. The accuracy was 60.4% (32 correct out of 53 samples), which is higher than that on the whole session (50.0%). The reason can be considered to be as follows: for humans, while the judgments of the CLP over short periods are relatively stable, over longer periods (e.g., the whole session), the dynamically changing discussion (CLP) creates the impression of ambiguity and thus the difficulty in CLP judgment.

4.4 Summary and Discussions

The experiment results validate the design of the attentive quiz agent in the following ways. In line with our intuition, it is observed that if the agent talks to the appropriate participant (CLP) at an appropriate (smooth) timing, the utterance can be expected to be more effective (the participants listen to it). Depending on the shape of the CLP pointer, it is possible to attract the participants' attention and to activate conversation. Further, the hypotheses of the utterance policy and the required information, AT and CLP, could be estimated at an acceptable level. Although the evaluation of AT estimation is difficult, given the fact that the attentive quiz agent could manage smooth utterance timings at higher percentages, the AT estimation method seems to work properly.

On the other hand, despite the fact that the eyeball CLP pointer is considerably more effective as a pointer device, participants found its headlike shape more offensive than an arrow pointer and hence felt more uncomfortable. This implies that using a physical pointing device with the 2D agent can be an effective way to specify the addressee of the agent's attention, but the utterance policy that treats the person who is leading the conversation as the addressee may not always be appropriate. Whom to point to and what to say at that time—these aspects should be designed in a more careful and detailed manner.

We must also state that we failed to predict the effects of some parts of this approach; all the same, we would like to discuss some finding in the experiment that did not directly prove the effectiveness of the approach but may be interesting to the readers of this paper. The attentive agent has the same qualities (which is not really a good point when compared to more sophisticated systems) as its comparable systems—namely, graphics, TTS, and nonverbal animations. The only difference was the *timings* for taking actions.

Table 6: The comparison between the CLP from the estimation of the system (S) in the presentation of the percentage of time during the whole session when each participant is judged as the CLP, the judgment of the annotators (A), and the questionnaires (Q) answered by the participants. The system always keeps the computation of CLP, so the percentages sum up to 100. The column ID denotes the 16 participant groups. "L," "M," and "R" indicate the participant who stands at the left, middle, and right positions, respectively.

ID	L	M	R	S	A	Q	S/A	S/Q	A/Q
1	0.3	51.7	47.0	M	M	M	√	√	√
2	20.1	54.0	25.7	M	L	R			
3	4.8	32.9	62.0	R	M	R		√	
4	74.5	12.9	8.6	L	L	L	√	√	√
5	44.5	12.9	42.2	L	R	M			
6	21.3	30.9	47.4	R	L	R		√	
7	9.6	32.0	57.9	R	M	R		√	
8	4.2	15.4	80.1	R	L	L			√
9	56.4	33.0	10.6	L	L	n/a	√	n/a	n/a
10	34.6	56.5	8.9	M	M	L	√		
11	0.1	5.2	94.8	R	R	M	√		
12	73.9	17.4	8.7	L	L	L	√	√	√
13	16.2	25.4	58.4	R	M	M			√
14	4.1	2.4	93.5	R	M	M			√
15	1.1	17.8	81.1	R	R	R	√	√	√
16	20.6	40.6	38.8	M	M	M	√	√	√
				Coincidence (%)			50.0	53.3	53.3

Nevertheless, significan differences could be found. This suggests an alternative way of improving the lifelikeness of ECAs as opposed to realistic looking characters and animations. By controlling the timings of the behaviors of ECAs, positive impressions could be achieved. On the other hand, the effects of the CLP pointer and how it should be used in coordination with the CG character are not clear. At present, three types of settings are possible. (1) The CLP pointer has its own personality and behaves as a separate agent. (2) The CLP pointer is an external device controlled by the 2D agent. This relationship should be cognitively recognizable by the participants, for example, by showing an animation that the 2D agent is operating the pointer. (3) The CLP pointer is a part of the 2D agent. In this setting, the appearance and the movement of the pointer need to be carefully designed to prevent contradictions in the cognition of the participants. Although quantitative analysis could not be done, during the experiments, the participants were observed to react in some of the following ways—they said "good work" or "yes, you are right" or bowed to the agent. These reactions may suggest that the participants had positive impressions of the agents; however, from the high ratio of agent's utterances that were ignored by the participants (which seldom happen in human-human conversations), we could not conclude that the agents are regarded as lifelike.

5. CONCLUSIONS AND FUTURE WORKS

This paper presented our investigations into the issues involved in communication with multiple users of ECAs in the context of a quiz game. An approach featuring an attentive utterance policy that involves reacting to the participants' activeness in the game is proposed for improving the lifelikeness of the quiz agent. Two experiments are conducted to evaluate the agent's effectiveness from the participants' reactions. From the preliminary results, determining the action timings of the agent in reacting to the participants' activeness status is proved to be effective.

The ideas proposed will then be improved (e.g., using a microphone array instead of bone conduction microphones) to develop the next version of our NFRI quiz agent that is deployable in practical exhibitions. Finally, we would like to further investigate the influence of the CLP pointer as well as the transcripts of the participants' utterances on linguistic aspects.

6. REFERENCES

[1] E. André and T. Rist. Presenting through performing: on the use of multiple lifelike characters in knowledge-based presentation systems. *Knowledge-Based Systems*, 14(1-2):3–13, March 2001.

[2] D. Bohus and E. Horvitz. Models for multiparty engagement in open-world dialog. In *Proceedings of SIGdial'09*, London, UK, 2009.

[3] T. Eichner, H. Prendinger, E. André, and M. Ishizuka. Attentive presentation agent. In *Proceedings of the 7th International Conference on Intelligent Virtual Agents (IVA'07)*, pages 283–295, Paris, France, 2007.

[4] J. Gratch, N. Wang, J. Gerten, E. Fast, and R. Duffy. Creating rapport with virtual agents. In *Proceedings of the 7th International Conference on Intelligent Virtual Agents (IVA'07)*, pages 125–138, Paris, France, 2007.

[5] S. Kopp, L. Gesellensetter, N. C. Kramer, and I. Wachsmuth. A conversational agent as museum guide - design and evaluation of a real-world application. In *Proceedings of the 5th International Conference on Intelligent Virtual Agents (IVA'05)*, Kos, Greece, 2005.

[6] O. Morikawa and T. Maesako. Hypermirror: Toward pleasant-to-use video mediated communication system. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work (CSCW'98)*, pages 149–158, New York, NY, USA, 1998. ACM Press.

[7] M. Rehm. "she is just stupid"-analyzing user-agent interactions in emotional game situations. *Interacting with Computers*, 20(3):311–325, May 2008.

[8] S. Robinson, D. Traum, M. Ittycheriah, and J. Henderer. What would you ask a conversational agent? observations of human-agent dialogues in a museum setting. In *Language Resources and Evaluation Conference (LREC)*, 2008.

[9] B. Rossen, K. Johnsen, A. Deladisma, S. Lind, and B. Lok. Virtual humans elicit skin-tone bias consistent with real-world skin-tone biases. In *Proceedings of the 8th International Conference on Intelligent Virtual Agents (IVA'08)*, pages 237–244, 2008.

[10] R. ten Ham, M. Theune, A. Heuvelman, and R. Verleur. Judging laura: Perceived qualities of a mediated human versus an embodied agent. In *Proceedings of the 5th International Conference on Intelligent Virtual Agents (IVA'05)*, pages 381–393, 2005.

[11] D. Traum. Issues in multiparty dialogues. In *Advances in Agent Communication, International Workshop on Agent Communication Languages (ACL'03)*, pages 201–211, 2003.

[12] D. Traum, S. C. Marsella, J. Gratch, J. Lee, and A. Hartholt. Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents. In *Proceedings of the 8th International Conference on Intelligent Virtual Agents (IVA'08)*, pages 117–130, Tokyo, Japan, 2008.

[13] H. van Vugt, E. Konijn, J. Hoorn, and J. Veldhuis. Why fat interface characters are better e-health advisors. In *Proceedings of the 6th International Conference on Intelligent Virtual Agents (IVA'06)*, pages 1–13, 2006.