Effects of Quiz-style Information Presentation on User Understanding

Ryuichiro Higashinaka, Kohji Dohsaka, Shigeaki Amano, and Hideki Isozaki

NTT Communication Science Laboratories, NTT Corporation
2-4, Hikaridai, Seika-cho, Kyoto 619-0237, Japan
{rh,dohsaka,amano,isozaki}@cslab.kecl.ntt.co.jp

Abstract

This paper proposes quiz-style information presentation for interactive systems as a means to improve user understanding in educational tasks. Since the nature of quizzes can highly motivate users to stay voluntarily engaged in the interaction and keep their attention on receiving information, it is expected that information presented as quizzes can be better understood by users. To verify the effectiveness of the approach, we implemented read-out and quiz systems and performed comparison experiments using human subjects. In the task of memorizing biographical facts, the results showed that user understanding for the quiz system was significantly better than that for the read-out system, and that the subjects were more willing to use the quiz system despite the long duration of the quizzes. This indicates that quiz-style information presentation promotes engagement in the interaction with the system, leading to the improved user understanding.

Index Terms: information presentation, quiz, user understanding

1. Introduction

Interactive systems, such as spoken dialogue systems [1, 2], need to accurately convey information and make users understand it to the fullest extent possible. For better understanding, users must be highly motivated to stay voluntarily engaged in the interaction and keep their attention on receiving information.

This paper proposes quiz-style information presentation of contents as a means to improve user understanding. In quiz-style information presentation, the contents are first converted into a quiz and then presented to a user until the user correctly guesses the answer. We specifically deal with the task of presenting descriptive texts about entities. We suppose that there are N statements describing X, and a quiz is created by ordering the N statements according to how difficult they make it to name X. The statements are presented one by one as hints to make the user guess what X is, with the system giving feedbacks as to how close the guesses are.

Figure 1 shows how a quiz-style information presentation can be realized, for example, about Natsume Soseki, a famous Japanese novelist, from his biographical facts found in the encyclopedic website goo.[1] Since users have to ponder the hints in order to come up with reasonable answers, it is expected that users will stay highly engaged in the interaction, leading to a possible improvement in understanding.

To verify the effectiveness of quiz-style information presentation, we compared two information presentation systems. One system simply reads out statements and the other poses quizzes

[1] http://dictionary.goo.ne.jp/

Entry of Natsume Soseki in an encyclopedia

Natsume Soseki Novelist and scholar of British literature. Real name: Kinnosuke. Born in Ushigome, Edo. Graduated from the University of Tokyo. Master of early-modern literature along with Mori Ogai. After the success of "I Am a Cat", quit all teaching jobs and joined Asahi Shimbun. Published masterpieces in Asahi Shimbun. Familiar with Haiku, Chinese poetry, and calligraphy. Works include "Botchan", "Sanshiro", etc.

↓

Quiz-style ordering of descriptive statements

1. Graduated from the University of Tokyo.
2. Born in Ushigome, Edo.
3. Novelist and scholar of British literature.
4. Familiar with Haiku, Chinese poetry, and calligraphy.
5. Published masterpieces in Asahi Shimbun.
6. Real name: Kinnosuke.
7. Master of early-modern literature along with Mori Ogai.
8. After the success of "I Am a Cat", quit all teaching jobs and joined Asahi Shimbun.
9. Works include "Botchan", "Sanshiro", etc.

↓

Quiz-style information presentation

S1	Who is this? First hint: Graduated from the University of Tokyo.
U1	Yoshida Shigeru?
S2	No, not even close! Second hint: Born in Ushigome, Edo.
U2	I don't know.
S3	OK. Third hint: Novelist and scholar of British literature.
U3	Murakami Haruki?
S4	Close! Fourth hint: Familiar with Haiku, Chinese poetry, and calligraphy.
U4	Mori Ogai?
S5	Very close! Fifth hint: Published masterpieces in Asahi Shimbun.
U5	Natsume Soseki?
S6	That's right!

Figure 1: A procedure for realizing a quiz-style information presentation dialogue. S stands for system utterance and U for user utterance in the dialogue.

made from the statements. In the task of presenting biographical facts, experimental results showed that the quiz-style system leads to significantly better user understanding (measured by memory tests) and that the users would be more willing to use the quiz-style system than the read-out system, indicating that quiz-style information presentation is a viable alternative.

Section 2 describes the quiz-style information presentation in detail. Section 3 describes the experiment we performed to verify the effectiveness of quiz-style information presentation in comparison with the read-out method. Section 4 summarizes and mentions future work.

2. Quiz-style Information Presentation

We focus on quiz-style information presentation because we believe that the strong engagement of users in receiving information is important for understanding. Previously, Sugiyama et al. proposed a method for delivering written texts by spoken dialogue [3]. The method first tokenizes the texts into units suitable for colloquial information presentation (called utterance units) and then presents them one by one on user demand. Since the method was shown to improve user understanding due to the involvement of users, it is likely that quiz-style information presentation, which also requires user involvement, could improve user understanding. From the perspective of memory, it has been suggested that the cognitive load in receiving information correlates with memory fixation [4]. Since quizzes can be cognitively expensive (e.g., when they are difficult), they could lead to better understanding.

From the above features of quizzes, one may think it to be obvious that the quiz-style method would result in better user understanding than a read-out method. However, for the following reasons, this is not as obvious as it may seem:

(i) The read-out method presents information in an order that promotes high coherence because the contents are generally prepared for reading, which makes it easy for users to understand the contents. In contrast, the quiz-style method would break up the discourse, which may make it difficult for users to structurize the contents for understanding.

(ii) The read-out method tells the user what X is before its N descriptive sentences are read (See Sec. 1 for our task setting). Therefore, users would be able to associate each sentence with X from the beginning. In contrast, the quiz-style method reveals X at the end of the interaction, which may make it difficult for users to associate N statements with X.

(iii) According to [4], high cognitive load would lead to better memory. Still, it is difficult to say whether the quiz-style method requires higher cognitive load than the read-out method, considering that the read-out method may require more concentration to keep listening to the contents.

(iv) The quiz-style method would involve feedback utterances, such as those shown in the dialogue of Fig. 1. Since such utterances would increase the time it takes to convey the information than the read-out method, users may get tired of receiving the contents.

If we want quiz-style information presentation to be an alternative to the read-out method, we also need to devise a way to create a quiz automatically. Since we create quizzes by ordering N descriptive statements, the task is similar to sentence ordering in multi-document summarization [5] or ranking definitions in definitional question answering [6]. The great body of research in these fields indicates that the quiz-style ordering is feasible. Since feedback utterances, such as "close!" and "that's right!" in Fig. 1, can also be implemented calculating the collocation of person names with conventional co-occurrence metrics (e.g., pointwise mutual information) from corpora, we consider that quiz-style information presentation is a realistic approach.

Since our aim here is to verify the effectiveness of quiz-style information presentation, we do not deal with the automatic creation of quizzes in this paper. Although several systems aim at automatic quiz generation [7, 8], they do not regard quizzes as a means of information presentation. We also consider that quiz-style information presentation would be useful for educational systems, such as tutoring systems [9].

3. Experiment

We performed an experiment to verify the effectiveness of quiz-style information presentation. We created a quiz-style information presentation system (hereafter the quiz system) and compared it with a read-out system as a baseline using human subjects. The task is to present biographical facts. Subjects performed a memory test, from which we examined whether the quiz-style information presentation improves user understanding.

3.1. Systems

We prepared the read-out and quiz systems. Both systems work as stand-alone applications and present descriptive statements about people by speech. Texts being read are also simultaneously shown in a separate window for assistance but they disappear soon (1.0 second) after the end of each statement. The systems produce utterances using pre-recorded voice.

3.1.1. Read-out System

The read-out system presents the descriptive statements in an encyclopedic order. It first says the name of the person to be described and then starts reading the statements one by one. Users press a button to proceed to the next statement until all statements have been read.

3.1.2. Quiz System

The quiz system presents descriptive statements in the quiz-style order (See Section 3.3 for how we ordered the statements). The system presents the statements one by one as hints and prompts users to type in a guess after each statement is read. The system utters a feedback utterance to let users know whether the typed name is close to the correct answer based on the person-name similarity score (See Section 3.3). The similarity score was mapped to one of six coarse similarity ranges defined by hand in order for the system to utter "very close!", "close!", "a little close!", "a little far", "far", or "Not even close!".

The user is allowed to press a button without guessing to proceed to the next hint when he/she cannot come up with a guess. In this case, a feedback utterance is not given. When the user types the correct name, the system applauds and reads the remaining hints in the same manner as the read-out system. If the user cannot guess the correct name after all hints have been given, the system says the answer and ends the dialogue.

3.2. Participants

We recruited 40 Japanese adults (20 males and 20 females). The participants were divided into two groups (Group-1 and Group-2), each group consisting of 10 males and 10 females (See Section 3.5 for the reason for this division). The mean ages of both groups were controlled to be 26 (Group-1: SD=5.4, Min=20, Max=38. Group-2: SD=5.5, Min=20, Max=39). They were paid for their participation.

3.3. Data

3.3.1. Biographical Facts

We collected concise biographies (in Japanese) from the goo encyclopedia. We first mined Wikipedia[2] to calculate the PageRank™of people using its hyper-link structure. After sorting the people in descending order by the PageRank score, we extracted the top-150 people for whom we could find an entry in the encyclopedia. We used the people with high PageRank scores because such people are likely to be famous and therefore enable the creation of answerable quizzes.

We divided the 150 people from the top into ten groups of 15 people, and randomly selected three people from each group to create three sets of ten people of approximately the same PageRank scores. After randomly shuffling each set, we merged the two of the three sets in alternate shifts to create a list of 20 people for presentation (List-A). The remaining set (List-B) was not used for the information presentation; it was used in the memory test in order to measure the prior knowledge of the participants. All descriptive statements for the 30 people were collected from the encyclopedia with their original orders preserved. There were 192 descriptive statements in all, with each person having approximately 6–7 statements.

3.3.2. Quiz-style Ordering

To create the quiz-style ordering of the descriptive statements, eleven annotators were asked to order the descriptive statements of the 20 people (List-A) in the quiz-style ordering individually. Here, the mean of the Kendall's coefficients of concordance for the 20 people was sufficiently high at 0.7 with a standard deviation of 0.13. Then, taking the average ranks assigned to each statement, we created a reference ordering for each person. An example of a reference ordering is the middle one in Fig. 1.

3.3.3. Person-name Similarity

The person-name similarity score used by the quiz system was derived based on [10], which calculates the collocation probability of person names by Fisher's exact test using Wikipedia articles and visualizes the names as a distribution on a two-dimensional map by the force-directed method [11]. The quiz system uses the Euclidean distance of person names on the map to tell whether the user's guess is close to the answer. We also prepared a list of aliases to normalize user inputs in order to cope with their stylistic variations before calculating the distance.

3.4. Evaluation Criteria

3.4.1. User Understanding

We prepared a memory test to see how well the participants understood the information presented by the systems. The test was created by creating blanks in the statements. Three annotators, who were not the authors, first spotted expressions that they considered important in characterizing the person in question (e.g., "University of Tokyo" in "Graduated from University of Tokyo"), and then one of the expressions was randomly replaced by a blank (e.g., "Graduated from [ ]"). If a statement did not contain such expressions or the whole statement becomes a blank because of the replacement, the statement was left as it was. The test was created for all 30 people (List-A plus List-B). The order of the people and the statements for each person were randomly shuffled. User understanding was evaluated from the rate and number of correct answers.

3.4.2. Subjective Evaluation

To see how the participants engaged in interaction with the systems, we created a questionnaire asking for subjective evaluations of the systems. For each of the two systems, the questionnaire items ask about (1) the usability of the system, (2) the willingness to use the system again, and (3) the usefulness for learning biographical facts. The questionnaire uses a 1-5 Likert scale and also has some extra space for writing in opinions.

3.4.3. Duration

In most task-oriented dialogue systems, task completion times are frequently used to measure the efficiency of the dialogue. Although our focus here is on user understanding and not on efficiency, it would be useful to know how much time is required to present information when the quiz-style information presentation is applied. It would also be interesting to examine the relationship between the time to receive information and the results of subjective evaluations. If the quiz-style information presentation proves to create longer dialogues without aggravating willingness or usefulness, it would mean that the users are highly engaged in the dialogue.

3.5. Procedure

Each participant used both systems in an alternating manner; i.e., the quiz system was always used after the read-out system and vise-versa. To reduce the effect of the order of the systems (read-out/quiz), Group-1 started the experiment with the quiz-system, whereas Group-2 started with the read-out system. Since it is also necessary to consider the effect of the order of the people to be presented, each participant in each group started the experiment from a different person; i.e., the first participant in Group-1 started from the first person in List-A, the second participant from the second person in the list, and likewise till the *n*-th participant.

The participants were seated before a personal computer and used the systems. They listened to the descriptive statements through headphones and pressed a button or typed guesses where necessary. They were informed, before using the systems, that there would be a memory test about the information they would receive. The logs of all user actions were stored on the computer. After completing the 20 people, they took the memory tests and filled out the questionnaire.

3.6. Results

3.6.1. User Understanding

Table 1 shows the results of the memory test. The table shows the total number of correct answers for all 40 participants with three levels of leniency: accepting (L1) partial answers, (L2)

[2] http://ja.wikipedia.org/

Table 1: Results of the memory test: the total number of correct answers for all 40 participants with three levels of leniency. The ratio of correct answers to the total number of blanks (2320 for read-out and quiz and 2680 for List-B) are shown in parentheses.

Leniency	Read-out	Quiz	List-B
(L1) Partial answers	719 (0.31)	808* (0.35)	448 (0.17)
(L2) Stylistic variations	640 (0.28)	723* (0.31)	410 (0.15)
(L3) Exact answers	615 (0.27)	680 (0.29)	387 (0.14)

Table 2: Results of the questionnaire: usability, willingness, and usefulness scores averaged over the 40 participants.

	Read-out	Quiz
Usability	3.4	3.4
Willingness	2.9	3.5*
Usefulness	2.9	3.5*

stylistic variations, and (L3) exact answers. The grading was done by a human judge.

A sign test showed that the number of correct answers for the quiz-system is significantly larger than for the read-out system for both L1 and L2 ($p<0.05$; marked by * in the table), suggesting the effectiveness of the quiz-style information presentation. When we allow only exact answers, the difference is reduced to an insignificant level ($p=0.08$). This is probably because answering exactly requires some prior knowledge and is not greatly helped by the style of presentation. It can also be seen that the read-out and quiz systems are both helpful in conveying information when we compare them with the results for the people in List-B.

3.6.2. Subjective Evaluation

Table 2 shows the results of the questionnaire. The scores are averaged over the 40 participants. It can be seen that the two systems have the same level of usability. However, a pair-wise t-test revealed that the willingness and usefulness scores for the quiz system were significantly better than those for the read-out system ($p<0.01$; marked by * in the table), suggesting that the quiz system promotes a high level of engagement. It is interesting to see such a difference in willingness with the same level of usability. It is also encouraging that the participants thought the quiz system was more useful for learning.

3.6.3. Duration

By analyzing the stored logs, we found that, on average, the read-out system took 49.1 seconds to present information about a person and the quiz system took 85.3 seconds, taking approximately twice the time. We also found that out of 400 quiz sessions (10 quizzes $\times$ 40 participants), only 87 were correctly answered by the participants before all the hints were given, leaving 313 quizzes unanswered. The ratio of the presented hints to the total number of hints was 89%. This means that most of the descriptive statements were presented in the quiz-style without switching to the read-out mode. Since the subjective evaluation revealed that the participants would be more willing to use the quiz system than the read-out system despite the long time it took to answer the quizzes, we can say that the subjects were likely to have been deeply engaged in interaction with the quiz system, leading to the improved understanding.

4. Summary and Future Work

This paper proposed quiz-style information presentation for interactive systems as a means to achieve better user understanding in educational tasks. Experimental results showed that the user understanding for the quiz system was significantly better than that for the read-out system. Since the subjects were more willing to use the quiz system despite the long duration incurred by quizzes, it is likely that they were more engaged in receiving information, explaining the improved understanding.

There may be an argument that the long duration is the major factor in the improvement; however, what we believe is important is that the engagement was kept high during the interaction by our approach. We naturally believe that the longer interaction would result in better understanding when other conditions are the same. We also acknowledge that the approach is limited to describing entities using their encyclopedic descriptions. We will investigate question generation techniques from sentences such as [12] to overcome this limitation. Currently, we are creating a fully-automated information presentation system based on our approach, employing techniques in definitional question answering such as [6] to automatically collect descriptive statements from text archives or the web.

5. References

[1] A. I. Rudnicky, C. Bennett, A. Black, A. Chotomongcol, K. Lenzo, A. Oh, and R. Singh, "Task and domain specific modelling in the Carnegie Mellon Communicator system," in *Proc. ICSLP*, vol. 2, 2000, pp. 130–134.

[2] J. Glass and T. J. Hazen, "Telephone-based conversational speech recognition in the JUPITER domain," in *Proc. ICSLP*, vol. 4, 1998, pp. 1327–1330.

[3] A. Sugiyama, K. Dohsaka, and T. Kawabata, "A method for conveying the contents of written texts by spoken dialogue," in *Proc. PACLING*, 1999, pp. 54–66.

[4] F. I. M. Craik and R. S. Lockhart, "Levels of processing: A framework for memory research," *Journal of Verbal Learning and Verbal Behavior*, vol. 11, pp. 671–684, 1972.

[5] M. Lapata, "Probabilistic text structuring: Experiments with sentence ordering," in *Proc. 41st ACL*, 2003, pp. 545–552.

[6] J. Xu, Y. Cao, H. Li, and M. Zhao, "Ranking definitions with supervised learning methods," in *Proc. WWW*, 2005, pp. 811–819.

[7] K. Binstead and O. Takizawa, "BOKE–A Japanese punning riddle generator," *Journal of Japanese Society for Artificial Intelligence*, vol. 13, no. 6, pp. 920–927, 1998, (in Japanese).

[8] F. Bagnoli, F. Franci, and A. Sterbini, "WebTeach: an integrated web-based cooperative environment for distance teaching," in *Proc. SEKE*, 2002, pp. 519–520.

[9] D. J. Litman and S. Silliman, "ITSPOKE: An intelligent tutoring spoken dialogue system," in *Proc. HLT-NAACL*, 2004, pp. 5–8, (demonstration paper).

[10] T. Matsubayashi and T. Yamada, "A force-directed graph drawing based on the hierarchical individual timestep method," in *Proc. CESSE*, 2007, pp. 218–223.

[11] T. M. J. Fruchterman and E. M. Reingold, "Graph drawing by force-directed placement," *Software - Practice and Experience*, vol. 21, no. 11, pp. 1129–1164, 1991.

[12] T. Misu and T. Kawahara, "Speech-based interactive information guidance system using question-answering technique," in *Proc. ICASSP*, 2007, pp. 145–148.