ORIGINAL ARTICLE

**Koji Yamashita · Hidekazu Kubota
Toyoaki Nishida**

Designing conversational agents: effect of conversational form on our comprehension

Abstract We have developed a broadcasting agent system, public opinion channel (POC) caster, which generates understandable conversational form from text-based documents. The POC caster circulates the opinions of community members by using conversational form in a broadcasting system on the Internet. We evaluated its transformation rules in two experiments. In experiment 1, we examined our transformation rules for conversational form in relation to sentence length. Twenty-four participants listened to two types of sentence (long sentences and short sentences) with conversational form or with single speech. In experiment 2, we investigated the relationship between conversational form and the user's knowledge level. Forty-two participants (21 with a high knowledge level and 21 with a low knowledge level) were selected for a knowledge task and listened to two kinds of sentence (sentences about a well-known topic or sentences about an unfamiliar topic). Our results indicate that the conversational form aided comprehension, especially for long sentences and when users had little knowledge about the topic. We explore possible explanations and implications of these results with regard to human cognition and text comprehension.

1 Introduction

Agent technology is flourishing, and various agents have appeared on the scene. For example, Rea (Real Estate Agent) is a sales agent for real estate

K. Yamashita (✉)
Keihanna Human Info-Communication Research Center,
National Institute of Information and Communications Technology,
3-5 Hikaridai, Seika, Soraku, 619-0289, Kyoto, Japan
E-mail: koji@nict.go.jp · Tel.: +81-774-986897 · Fax: +81-774-986963

H. Kubota · T. Nishida
Graduate School of Informatics,
Kyoto University, Yoshida-Honmachi, 606-8501, Sakyo-ku,
Kyoto, Japan

(Cassel et al. 2000), and STEVE is a pedagogical agent for procedural training (Rickel and Johnson 1998). The word agent is commonly defined as meaning autonomous software that interacts intelligently with users or other agents. As such, it acts as a mediator between users and information. Conversational form should be an effective interface for an ideal mediator because it is primitive and familiar to users and does not require particular skills, such as keyboard typing. Conversational form for software agents can be used in two ways. One is for information provided from agents to users, and the other is for information exchanged between users and agents. In the first case, the main goal is that agents provide information to users in an easy-to-understand way. This is one-way communication because there is no continuous interaction between users and agents. In the latter case, agents provide information through interactions with users, interactively confirming the user's intention and comprehension. This is a bidirectional approach, and it may be the ideal for many situations. In this study, we have focused on information provision by conversational agents in a broadcasting system (i.e., one-way communication).

Information provided using the forms of everyday conversation is often found in magazines, newspapers, and so on. The conversation allows us to picture a scene differently from when a single speaker reads a text. A listener's comprehension is thought to be facilitated by conversational form because inserted pauses, the speaker's rhythm, and topic control by two or more speakers generate cues, which help the listener understand what is being said. More specifically, our comprehension of a topic can be deepened if we ask questions and discuss the topic with others. Is comprehension of a topic similarly affected when agents use conversation in providing information? If so, what kind of presentation is most effective? In this paper, we report on two experiments to evaluate a conversational agent that uses a method for transforming a text into a conversational form (Kubota et al. 2002a, b). In the next section, we provide a system summary and our transformation rules. In experiment 1, we evaluated our transformation rules for conversation in relation to sentence length. In experiment 2, we examined the relationship between conversational form and knowledge level. In the final section, we conclude this paper by exploring possible explanations and implications of the results.

2 Our approach to conversational agents

The public opinion channel (POC) caster is a conversational agent used by the (POC; Nishida et al. 1999; Azechi et al. 2000), which is an automatic interactive broadcasting system to support knowledge creation and facilitate knowledge circulation in communities on the Internet. The POC gathers the opinions of community members, edits broadcasting programs based on the collected opinions, and then broadcasts the edited programs as TV or radio programs. The POC elicits knowledge from each community and facilitates its circulation by automating this cycle. The POC system consists of the POC Server (Fukuhara et al. 2003) and POC clients.

The POC caster presents information to the community by transforming opinions from members into a conversation between agents. First, a user inputs a keyword on a certain topic. The POC caster sends a query to the POC Server, and receives in return a set of opinions related to the keyword. The opinion set is arranged in a manner the user specifies: chronological order, reverse chronological order, or at random. The POC caster transforms each opinion from plain text into a conversational form, and broadcasts it to users by using speech synthesis with captions. When an opinion has been submitted with a still picture, the picture is displayed in the center of the screen. Figure 1 shows the POC caster display. On the screen, two agents with face photos and animated human bodies introduce each community member's opinions in turn. The two agents provided information as follows: (1) Agent 1 reads out the title, and this allows us to get a summary of the topic. (2) Agent 2 introduces the text of a previous opinion sentence (The original text form is divided into two parts by a period). (3) Agent 1 comments on the previous text based upon certain rules. (4) Agent 2 proceeds to the later parts of the text.

The original text is transformed into a conversational form by applying transformation rules constructed based on the end of each opinion sentence and the position of each sentence in the text. The end of a sentence can indicate two types of expression: (a) declarative expressions to indicate information related to interests, clarify a situation, or indicate hearsay; (b) interrogative expressions. In Japanese, we can grasp the meaning of a passage by focusing our attention on the statement's end because that is often where the verb is located. Sentences stating opinion could be initial, middle, or final ones. If there are only two

Fig. 1 The POC caster Screenshot

Table 1 The POC caster transformation rules

Rule 1	*Presentation of the context*	
	Objective	To facilitate listener's understanding of the context by inserting words
	End of sentence	Words providing context
	Processing	Inserting words after the sentence to provide information
	Inserted words	"What is that?"[a] or "Give me more details"[b]
Rule 2	*Repetition of question*	
	Objective	To promote listener's understanding by repeating questions asked
	End of sentence	Question
	Processing	Inserting words after the sentence represents asking a question
	Inserted words	"What do you think about that?"[c]
Rule 3	*Simple response*	
	Objective	Allowing time for the listener to understand by inserting a simple response
	End of sentence	No use
	Processing	Inserting simple responsive words after the second sentence
	Inserted words	"Yes" or "Uh-huh"[d]

[a]"Sore ha nandesuka?" in Japanese
[b]"Motto kuwashiku osiete kudasai" in Japanese
[c]"Dou omoimasuka?" in Japanese
[d]"Hai" or "Un" in Japanese

sentences, no middle sentence exists, and if there is only one sentence, it is treated as the final one. Using these cues, we have implemented three transformation rules for conversational form (see Table 1). Rule 1 shows that the following sentence gives the listener a detailed description by presenting contextual information. Rule 2 makes the listener pay attention to the topic by repeating questions from the previous sentence. Rule 3 allows the listener to gain time to understand the meaning by inserting a simple response (for details on the technical aspects, see Kubota et al. 2002a, b).

3 Experiment 1

In this experiment, we evaluated the transformation rules of POC caster. The main purpose of the experiment was to examine whether the conversational form generated by POC caster, compared to a text representation, could aid sentence comprehension in relation to sentence length. We hypothesized that conversational form would enhance comprehension of the presented sentence. More specifically, when the presented sentence was short, we expected the effect of conversational form to be lower than when the sentence was long. Moreover, we explored the effect of words inserted to generate the conversational form.

3.1 Methods

3.1.1 Design and participants

A 2×2 within-subjects design was used with one factor of representational form (single speech versus conversation) and a second factor of sentence length (long versus short). Twenty-four people (21 males and 3 females) between 21 and 41 years of age ($M = 24.92$) participated, and were paid for their cooperation.

3.2 Materials

Fifty-long sentences (160–200 characters) and 50 short sentences (50–100 characters) were selected from 300 sentences taken from opinions submitted to POC, newspaper articles, or dictionaries. Thirty-long sentences and 30 short sentences were pooled as a stimulus set. Sentences were excluded from this set if there was a significant difference between the time needed for participants to read the sentence from text and the time needed for the speech synthesis system to read it aloud. We applied the transformation rules to 16 sentences and generated conversational sentences such as inserted words that aided understanding of the context (e.g., "What is that?", Rule 1) and simple responses to allow time for understanding (e.g., "Yes.", Rule 3).

3.3 Procedure

Participants were tested individually in a single session lasting about 60 min. Each session included instruction, a practice trial, experimental trials, and a questionnaire. The participants were seated in front of a personal computer (Pentium III, 700 MHz, Windows 98), and were given written instructions. During each trial, the stimulus sentence was provided through headphones, and the participant judged the sentence comprehensibility on a seven-point scale ("1: easy" to "7: difficult") by responding with a mouse click as quickly as possible without making errors. Participants were instructed not to allow the speech synthesis quality to affect their judgment.

A total of 32 sentences (16 long and 16 short) were selected at random from a stimulus set for each participant. The trials consisted of two blocks of 16 sentences (eight long and eight short) with a 2 min break between blocks. The sentence order was random in each block for each participant. Participants were

Table 2 Average rating scores for representational form in absolute rating

| | Sentence length | | | |
| | Long | | Short | |
	Poor	Rich	Poor	Rich
Single speech—conversation	3.78 (0.75)	3.80 (0.59)	4.56 (0.93)	4.21 (0.77)
Conversation—single speech	4.30 (0.88)	4.79 (0.73)	4.10 (0.63)	4.41 (0.71)

$N = 23$. SD values are given in *parentheses*

Table 3 Average rating scores for inserted words as a function of sentence length

	Sentence length	
	Long	Short
Poor context word	4.47	4.91
	(0.77)	(0.87)
Rich context word	5.11	5.63
	(0.75)	(0.63)

$N = 23$. SD values are given in *parentheses*

asked to judge comprehensibility three times per trial. In the absolute rating, they judged each of two sentences. The sentence order was random for each participant. In the relative rating, participants compared two sentences and judged whether the first or second sentence was easier to understand. When the single speech sentence was judged, the stimulus presentation was done in the order of the single speech sentence and then the conversational sentence, and participants rated the comprehensibility for the single speech sentence compared to that for the conversational one. The conversational form included two kinds of sentence. One included words intended to help the participant understand the context (rich context), the other consisted of simple responsive words used to obtain time for understanding (poor context). Each sentence type appeared in an equal percentage of the time.

3.4 Results and discussions

For all of the analyses reported in this paper, the statistical rejection level for significance was set at $p < 0.05$ unless otherwise indicated. Data from four participants were excluded from all following analyses—one because of a computer problem, and three because of their failure to comply with the instructions. In consequence, we analyzed data from 23 participants for the absolute rating and 20 for the relative rating. Response time (RT) was measured through the participants' mouse clicking, but it is not reported here because we are not fully confident of its reliability. The RT results we obtained, however, were consistent with the patterns of the rating score results.

Table 4 Average relative rating scores for sentence judgment as a function of sentence length and inserted words

	Sentence length			
	Long		Short	
	Poor	Rich	Poor	Rich
Single speech—conversation	3.78	3.80	4.10	4.41
	(0.75)	(0.59)	(0.63)	(0.71)
Conversation—single speech	4.30	4.79	4.56	4.21
	(0.88)	(0.73)	(0.93)	(0.77)

$N = 20$. SD values are given in *parentheses*

Absolute rating. Average rating scores are summarized in Table 2. Results from an analysis of variance (ANOVA) revealed a significant main effect of sentence length, $F_{(1,22)} = 34.36$, and a significant interaction between the sentence length and the representational form, $F_{(1,22)} = 19.32$. A posteriori Tukey tests showed that the short sentences were easier to understand than long sentences, both in single speech and conversational form. Moreover, the tests indicated that for a long sentence the conversational form was easier to understand than single speech, whereas there was no significant difference for a short sentence. Table 3 shows the effect of inserted words (rich versus poor). The ANOVA results showed a significant main effect of sentence length, $F_{(1,22)} = 15.09$, and that of inserted words, $F_{(1,22)} = 43.45$. That is, the enhancement of comprehension was greater when the sentence was short ($M = 5.27$) than when it was long ($M = 4.79$), and it was greater when inserted words were of rich context ($M = 5.37$) rather than poor context ($M = 4.69$).

Relative rating. Table 4 shows the means of the relative ratings for comprehension. We conducted a three-way ANOVA with sentence length (long versus short), inserted words (rich versus poor), and judgment sentence (single speech-conversation versus conversation-single speech) as within-subject variables. The ANOVA revealed a significant main effect of sentence length, $F_{(1,19)} = 4.58$, a significant interaction between sentence length and judgment sentence, $F_{(1,19)} = 13.82$, and a significant interaction between inserted words and judgment sentence, $F_{(1,19)} = 10.10$. Subsequent analysis of the interaction between the sentence length and judgment sentence revealed that when participants were tested on single speech-conversation, their ratings were better for short sentences ($M = 4.38$) than long ones ($M = 3.79$), $F_{(1,19)} = 15.93$, and when tested on conversation-single speech, they scored better for long sentences ($M = 4.54$) than short ones ($M = 4.26$), $F_{(1,19)} = 5.10$. Also, while there was no significant difference in the judgment sentence when the stimulus was a short sentence, when the stimulus was a long sentence, the rating score was better in conversation-single speech ($M = 4.54$) than single speech-conversation ($M = 3.79$), $F_{(1,19)} = 9.15$. Subsequent analysis of the interaction between inserted words and judgment sentences showed that there was no significant difference between single speech-conversation ($M = 4.17$) and conversation-single speech ($M = 4.20$) when the inserted words were of poor context; however, when the inserted words were of rich context, rating scores for conversation-single speech ($M = 4.60$) were better than those for single speech-conversation ($M = 4.00$), $F_{(1,19)} = 8.60$. In addition, the effect of the inserted words was not significant in single speech-conversation, but was in conversation-single speech, and participants scored better when sentences with rich context ($M = 4.60$) were presented than when sentences with poor context were presented ($M = 4.20$), $F_{(1,19)} = 6.86$.

Our results are summarized as follows. The conversation generated by the transformation rules promotes comprehension more effectively when the sentence length is long than when it is short; inserting words having rich context information has a stronger effect on our comprehension than simple responses. Although, it is important that we have demonstrated when conversational form will be beneficial, note that the reliability of our results is limited for three reasons. First, the voice provided by the speech synthesis system was artificial with respect to accent, rhythm, and so on. We cannot exclude the possibility that the unnatural voice affected the participants' judgment even though we

instructed them not to pay attention to the speech synthesis quality. Second, the pauses between sentences were picked at random by the speech synthesis system and were not manually controlled. Third, there may have been methodological problems in our procedure. For example, participants might have found it difficult to separately judge the absolute and relative ratings. These problems were resolved in experiment 2.

Results are summarized as follows. The conversation generated by the transformation rules promotes comprehension more effectively when the sentence length is long than when it is short; inserting words having rich context information has a stronger effect on our comprehension than simple responses. Although, it is important that we have demonstrated when conversational form will be beneficial, note that the reliability of our results is limited for three reasons. First, the voice provided by the speech synthesis system was artificial with respect to accent, rhythm, and so on. We cannot exclude the possibility that the unnatural voice affected the participants' judgment even though we instructed them not to pay attention to the speech synthesis quality. Second, the pauses between sentences were picked at random by the speech synthesis system and were not manually controlled. Third, there may have been methodological problems in our procedure. For example, participants might have found it difficult to separately judge the absolute and relative ratings. These problems were resolved in experiment 2.

4 Experiment 2

In experiment 1, we examined the validity of the POC caster transformation rules. After the trials in experiment 1, some participants made comments such as "I could easily understand the sentence when I knew the topic very well". The effect of context should be maximized when we know the topic well. Our main purpose in experiment 2 was to investigate the relationship between a participant's knowledge level and conversational form by using a forced-choice task. The idea was that if a participant knew little about the topic, conversational form should aid understanding. On the other hand, if a participant was knowledgeable about the topic, conversational form should not strongly affect understanding.

4.1 Methods

4.1.1 Design and participants

We used a 2×2×2 mixed design with one factor of representational form (single speech versus conversation) as a within-subject variable, a second factor of knowledge level (high versus low) as a within-subject variable, and a third factor of modality (visual versus auditory) as a between-subject variable. Seventy-eight people (45 females and 33 males) between 19 and 35 years of age ($M = 22.21$, $SD = 3.55$) participated in this experiment and were paid for their cooperation. Participants were randomly assigned in equal numbers to two groups (one for the visual and the other for the auditory condition).

4.1.2 Materials

A total of 108 sentences were selected from newspaper and magazine articles within the range of 160–200 words (corresponding to the long-sentence length of experiment 1). The sentences included 36 topics (e.g., environment-related issues, fortune-telling, horse races, football, etc.). We applied the transformation rules and generated conversational sentences with inserted words to aid understanding of the context (e.g., "What is that?") and simple response words to gain time to understand (e.g., "Yes."). The voice stimuli for the auditory condition were recorded with a human male voice (Standard Japanese). The conversation stimuli consisted of male and female voices. The pauses between sentences were 1500 ms long, and pauses following a comma were 700 ms long. The speech was at a natural speed. These modifications were made to resolve the above-mentioned problems in experiment 1.

4.1.3 Procedure

Participants were tested individually in a single session lasting about 90 min. The procedure consisted of three parts: a knowledge task, experimental trials, and questionnaire completion. In the knowledge task, participants were asked to rate their level of familiarity and interest regarding words presented on a personal computer (Pentium III, 700 MHz, Windows 98) on a five-point scale ("1: not at all" to "5: very high"). Thirty-six words were presented in the center of a computer screen one at a time and in a random order for each participant. We selected 12 words (six words representing well-known topics and six words unfamiliar topics) as theme words for each participant from the results.

Next, in the experimental trials, participants were instructed to judge, which of two sentences presented by a personal computer was easiest to understand. They were also told to press a particular key on the keyboard as quickly as possible when the word "Judgment" appeared on the screen. They were asked to press the "1" key if the first sentence was easiest to understand, and to press the "|" key if the second sentence was easiest. We used a forced-choice method. Stimuli under the visual condition were presented in the center of the computer screen, whereas those under the auditory condition were presented though headphones. Under the auditory condition, turn-switching in conversations could be clearly recognized because of differences in the speakers' voices. Under the visual condition, "A" and "B" were displayed before the sentences to

Table 5 Average probability of selecting the representational form as a function of knowledge level and modality in experiment 2

Knowledge level	Visual		Auditory	
	High	Low	High	Low
Single speech	0.67	0.33	0.58	0.42
	(0.22)	(0.22)	(0.22)	(0.22)
Conversation	0.57	0.43	0.46	0.54
	(0.17)	(0.17)	(0.22)	(0.22)

$N = 21$ for both the visual condition and the auditory condition. SD values are given in *parentheses*

Table 6 Average RT when selecting the representational form in experiment 2

Knowledge level	Visual		Auditory	
	High	Low	High	Low
Single speech	973.63	752.05	752.05	1025.03
	(483.37)	(316.79)	(316.79)	(675.65)
Conversation	1057.68	1063.97	1063.97	800.89
	(540.49)	(812.73)	(812.73)	(433.49)

$N = 21$ for both the visual condition and the auditory condition. The RTs are given in milliseconds. SD values are given in *parentheses*

visually indicate the conversational form. Two sentences were presented as one pair for each trial. The trials consisted of two blocks of 18 trials, with a 3 min break between blocks. Each sentence pair was presented in a random order within a block. The total presentation time required to display characters or recall the recorded voices was adjusted so that the time was not markedly different between the two conditions. There were two practice trials using sentences not used in the experimental trials. After the experimental trials, participants were asked to answer the following questions in an open-ended style: "What kind of reason led you to select the sentence?", and "What kind of impression did you have?"

4.2 Results and discussions

Data from 36 participants were excluded from all following analyses because of the knowledge task criteria. That left 21 participants (7 males and 14 females) for the visual condition (age: $M = 23.57$, $SD = 4.39$) and 21 participants (10 males and 11 females) for the auditory condition (age: $M = 21.76$, $SD = 3.06$). Our dependent measures were the number of times that participants selected a particular sentence and the response latency.

The probability of selection is shown in Table 5. A 2 (representational form) × 2 (knowledge level) × 2 (modality) mixed ANOVA for data with inverse sine transformation revealed a significant interaction between the representational form and the knowledge level, $F_{(1,40)} = 16.36$, and a marginally significant interaction between the representational form and modality, $F_{(1,40)} = 2.99$, $p < 0.10$. The former interaction is reported here. Post hoc analysis revealed no reliable difference in the selection of sentences when the level of knowledge was low; in contrast, for sentences where the level of knowledge was high, selection was clearly better for single speech ($M = 0.54$) than for conversation ($M = 0.35$), $F_{(1,40)} = 13.17$. Also, when we tested a simple main effect of the representational form, we found that sentences for which the participant's level of knowledge was high were selected significantly more often than those for which the level of knowledge was low under the single speech condition, $F_{(1,40)} = 15.71$; under the conversation condition, a low level of knowledge led to higher selection than did a high level of knowledge, $F_{(1,40)} = 16.52$.

Table 6 shows the RT of the selective responses. A 2 (representational form)×2 (knowledge level)×2 (modality) mixed ANOVA for data with loga-

rithmic transformation revealed that there was a significant second-order interaction, $F_{(1,40)} = 6.18$. We computed a 2 (representational form)×2 (knowledge level) ANOVA for each modality condition according to our interest. The interaction between the representational form and knowledge level was reliable for the auditory condition, $F_{(1,20)} = 8.58$, but there was no significant effect for the visual condition. As to the significant interaction for the auditory condition, participants with a high level of knowledge were faster to respond to a sentence of single speech than to one of conversation, $F_{(1,20)} = 3.47$, $p < 0.10$. On the other hand, participants with a low level of knowledge were faster to respond to conversation than to single speech, $F_{(1,20)} = 4.09$, $p < 0.10$. Under the single speech condition, there was no significant difference that depended on the participant's level of knowledge; under the conversation condition, though, participants with a low level of knowledge responded to sentences more quickly than those with a high level of knowledge, $F_{(1,20)} = 9.23$.

In experiment 2, we examined the relationship between conversational form and knowledge level. Our results were as follows. (1) The conversational form had a beneficial effect when participants were knowledgeable about the topic; however, conversational form had no effect when participants had little knowledge of the topic. (2) Under the conversational condition, participants with no knowledge of the topic were more likely to consider the sentence as easier to understand than the knowledgeable participants. (3) The RT was shorter with conversational form when participants had little knowledge about the topic. These results indicate that the effect of conversational form depends on the user's relevant knowledge.

5 General discussion

What kinds of information agents should give us, when they should speak, and how they should speak are important research issues. In this study, we have investigated, which representational form is needed to promote our comprehension. We have shown the beneficial effect of conversational form in a long sentence (in experiment 1) and for a person lacking relevant knowledge (in experiment 2) when agents provide information. These results support the basic hypothesis that conversational form aids comprehension. Related work includes a series of studies done by Mayer and his colleagues (Mayer et al 2003; Moreno et al. 2001; Moreno and Mayer 2002) in which the type of presentation was examined. For example, Moreno and Mayer (2002) experimentally varied whether the agent's words were presented as speech or as on-screen text. Their results showed that people learned better when words were presented as speech rather than as on-screen text. The main difference between our study and their work is in the conversational form. In that sense, we are the first to demonstrate there are specific conditions under which conversational form, compared to text form, aids comprehension.

We found in experiment 2 that conversational form allowed a person without relevant knowledge to grasp the sentence focus and the connection between sentences. This was because they could easily process the sentence since the text was divided into small parts. In the field of cognitive psychology, Kintch (1998) has proposed a "macro-structure model", which is a semantic structure of the entire sentence, and a "situation model", which is connected to knowledge. He

has also suggested a global model for text comprehension, which consists of bottom-up and top-down processing (van Dijk and Kintch 1983). According to Kintch and his colleagues, this macro-structure model means that text comprehension depends on making an exact representation (text-base) for the content of a text, from analysis of each word in a text to integrating higher meanings, and integrating this into the higher meaning of the whole text (macro-structure). This can be considered "learning of text" processing. On the other hand, the situation model is aimed at understanding the meaning of an object that is explained in a text, relating the content to our knowledge. This can be considered "learning from text" processing. These forms of processing are not thought of as alternatives, but as forms that run in parallel.

When we know a topic well, our processing can focus on checking for current information relative to our existing knowledge because the situation model can be readily formulated. On the other hand, when we lack relevant knowledge, our processing is dominated by the bottom-up processing needed for superficial understanding because the situational model cannot be formulated or will be inaccurate. Applying these models to the conversational form relationship allows a person without knowledge regarding a particular topic to grasp the issue and connection between sentences. Inserting conversational form in this case supports text-based processing, and facilitates text understanding. On the other hand, conversational form has no effect on understanding for a person who is knowledgeable because that person is able to create the situation model without help; that is, such users only have to verify the current information relative to their existing knowledge. Thus, conversational form supports text-based processing only for those who cannot formulate a sufficient situational model.

It can also be said that conversational form is a kind of "advance organizer" (Ausubel 1963, 1978) in the sense that specific contextual information is added in advance. Ausubel (1963) defines an advance organizer as the "anchoring foci for the reception of new material." It allows the learner to recall and transfer prior knowledge to the new information being presented. This theory is based on the idea that learning is facilitated if the learner can find meaning in the new information. If a connection can be made between knowledge and new information, the learning experience will become more meaningful to the learner. Studies have shown that comprehension and memory of a text is promoted by presenting particular information before the text such as a title (Bransford and Johnson 1972), a summary (Bromage and Mayer 1986), a perspective (Pitchart and Anderson 1977), or an illustration (Waddili and McDaniel 1988). Do any qualitative differences added to an original text affect our comprehension? If so, what kind of difference is needed? Further research is needed to clarify this point by manipulating the nature of inserted words more precisely.

The success of text comprehension is likely to depend on the interaction between the characteristics of the learner and the properties of the text itself (e.g., amount of information provided about the topic, connection between sentences, clarity of wording). Although, we have studied the effect of one learner characteristic (knowledge level) and one text property (sentence length) in this paper, many factors remain to be examined. Further research is needed to investigate these factors and the relationships between them.

References

Azechi S, Fujihara N, Sumi K, Hirata T, Yano H, Nishida T (2000) Public opinion channel: A challenge for interactive community broadcasting. In: Ishida T, Isbister K (eds) Digital cities: technologies, experiences, and future perspectives. Springer-Verlag, New York, pp 427–441

Ausubel D (1963) The psychology of meaningful verbal learning. Grune & Stratton, New York

Ausubel D (1978) In defense of advance organizers: a reply to the critics. Rev Educ Res 48:251–257

Bransford JD, Johnson MK (1972) Contextual prerequisites for hierarchical retrieval schemes in recall of categorized word lists. J Verbal Learn Verbal Behav 11:717–726

Bromage BK, Mayer RE (1986) Quantitative and qualitative effects of repetition on learning from technical text. J Educ Psychol 78:271–278

Cassel J, Bickmore T, Campbell L, Vilhjalmsson H, Yan H (2000) Human conversation as a system framework: Design embodied conversational agents. In: Cassell J et al (eds) Embodied conversational agents. MIT Press, Cambridge, MA, pp 29–63

van Dijk TA, Kintch W (1983) Strategies for discourse comprehension. Academic Press, New York

Fukuhara T, Chikama M, Nishida T (2003) Supporting an experiment of a community support system: Community analysis and maintenance functions in the public opinion channel. In: Huysman M, Wenger E, Wulf V (eds) Communities and technologies. Kluwer Academic Publishers, Dordrecht, pp 347–367

Kintch W (1998) Comprehension: a paradigm for cognition. Cambridge University Press, Cambridge

Kubota H, Yamashita K, Fukuhara T, Nishida T (2002a) POC caster: Broadcasting agent using conversational form for Internet community. Trans Jpn Soc Artif Intelligence 17:313–321 (in Japanese with English abstract)

Kubota H, Yamashita K, Nishida T (2002b) Conversational contents making comment automatically. In: Damiani E, Howlett RJ, Jain LC, Ichalkaranje N (eds) Frontiers in artificial intelligence and applications, vol 4. Knowledge-based intelligent information engineering systems & allied technologies Part 2. IOS Press, Amsterdam, pp 1326–1330

Mayer RE, Dow GT, Mayer S (2003) Multimedia learning in an interactive self-explaining environment: what works in the design of agent-based microworlds? J Educ Psychol 95:806–812

Moreno R, Mayer RE, Spires HA, Lester JC (2001) The case for social agency in computer-based teaching: do students learn more deeply when they interact with animated pedagogical agents? Cogn Instr 19:177–213

Moreno R, Mayer RE (2002) Learning science in virtual reality multimedia environments: role of methods and media. J Educ Psychol 94:598–610

Nishida T, Fujihara N, Azechi S, Sumi K, Yano H, Hirata H (1999) Public opinion channel for communities in the information age. N Generation Computing 17:417–427

Pitchart JW, Anderson RC (1977) Taking different perspectives on a story. J Educ Psychol 69:309–315

Rickel J, Johnson WL (1998) STEVE: A pedagogical agent for virtual reality. In: Syncara KP, Wooldridge M (eds) Proceedings of the 2nd international conference on autonomous agents (AGENTS 98). ACM Press, New York, pp 332–333

Waddili PJ, McDaniel MA (1988) Illustrations as adjuncts to prose: a text-appropriate processing approach. J Educ Psychol 80:457–464