Estimation of Pause Length Set by Storyteller at Sentence Boundary in Nursery Tale: Design of Synthesized Speech to Support Listener's Comprehension

Hideji Enokizu† Kazuhiro Uenosono‡ Seiichi Komiya‡
†Shibaura Institute of Technology
3-7-5 Toyosu, Koutou-ku, Tokyo, 135-8548 Japan
‡Graduate School of Engineering, Shibaura Institute of Technology
3-7-5 Toyosu, Koutou-ku, Tokyo, 135-8548 Japan
†enokizu@shibaura-it.ac.jp, ‡{m704101, skomiya}@shibaura-it.ac.jp

Abstract: In the present study, we examined the pause length which the storyteller sets at each sentence boundary in the story text to support children's comprehension and proposed one method of estimating the pause length preliminarily to apply to the speech synthesis. While reading the story text, the reader constructs the situation model that is the mental microworld described in each sentence. Unlike the reader who can read the text at his own pace, however, the listener must hear the story at the storyteller's reading pace. Therefore we assumed that the experienced storyteller can set the pause length at each sentence boundary necessary for the listener to construct the situation model. If the listener is a young child, this assumption is true. Then we focused on the relationship between the mental operations for constructing the situation model and the pause length set by the experienced storyteller to help the listener comprehend the story. Specifically, we devised a knowledge representation with multilayered frame structure corresponding to the situation model to predict the mental operations for constructing the situation model. The knowledge representation was named as the situation frame. The results indicated that the storyteller controlled the pause length on the basis of several operations. The formula, that can estimate the pause length from the operation needed to construct the situation frame, seems to provide us the reasonable method to determine the pause length in the synthesized speech preliminarily.

Key-word: Speech synthesis, Pause length at sentence boundary, Story comprehension, Situation model, Event-indexing model, Mental operation, Multilayered frame representation, Multiple linear regression analysis

1 Introduction

Through our experience, we know that the prosodical cues of speech sound in telling the story are closely relevant to the listener's comprehension. We can obtain several findings that imply the relations between the development of stories and the pause length at the sentence boundary, the character's emotion and the pitch, and the story structure and the speed. Although such interesting findings were obtained, the synthesized speech sound curries little prosodical cues to us. The reasons seem to come from the technical problems as well as the economical problems. For example, the pause length in the synthesized speech is generally determined in two ways. First, fixed pause length is determined arbitrarily by the engineer designing the system. It is too easy to support various listeners with different skill of language understanding. Second, the pause length preliminarily set by the professional narrator is inserted in each sentence boundary. Although the narrator's pause length may be set with consideration for the diversity of the listener's skill, such way takes much time and money.

In order to overcome such problems, we focused on the pause length at the sentence boundary among the prosodical cues. The pause length at the sentence boundary is especially focused on in the researches of the listener's comprehension that involves the semantic processing of each sentence and throughout the story. It was found that the radio announcer usually sets about 1400msec pause at each sentence boundary when he or she reads the news article and the weather report [1]. However, such pause length seems not to be necessarily suitable for listener to comprehend other genre of the text. We can obtain some findings that the experienced storyteller sets the pause of varied time lengths in reading aloud story texts [2] [3]. Thus, these evidences clearly indicate that the constant pause length at sentence boundary is not appropriate to support the listener's comprehension. In fact, it was suggested that the appropriate pause length, which can facilitate the listener's comprehension, is varied

according to the story structure [4]. It was also demonstrated that the storyteller increased the pause length at the sentence boundary which occupies critical place in the text structure [5]. Although the researchers examined the relation between the surface structure of the text and the pause length, processes of the listener's comprehension were not discussed substantially [6] [7].

Cognitive linguists assume the language as a set of cues by which the listener's attention on the fictional situation is manipulated [8]. Furthermore, psycho-linguists saw the language as a set of processing instructions on how to construct a mental representation of the described situation [9]. From their perspective, the storyteller reads the story to help the listener construct the mental representation of the microworld described by the sentence. Particularly, experienced storyteller may be able to estimate the pause length necessary to comprehend each sentence through a long time learning consciously or unconsciously. Such interaction of the storyteller and the listener is shown in Figure 1.

Figure 1. Story teller narrates for children to construct the mental representation of the situation described in story. It is just like our everyday communication made about the situation that we encountered.

In the present study, we proposed a technique for estimating the pause length at each sentence boundary just like the storyteller in order to apply to the synthesized speech. We addressed following tasks to accomplish our purpose step by step.

First, we discussed scientific studies of the story comprehension to understand what the microworld is and how it is constructed. Second, we gave specific structure and contents to the microworld to predict series of the mental operations necessary to construct it. Third, we conducted one experiment to measure the pause length at each sentence boundary when experienced storyteller read aloud several stories. Then we derived the regression formula that the pause length as an objective variable and series of mental operations as explanatory variables. Finally, we conducted two verification experiments to determine whether derived regression formula can estimate the pause length appropriate for the listener's comprehension. These step-by-step processes are explained in greater detail below.

2 Story Comprehension

2.1 What is Story Comprehension?

The discourse psychologists assume that story comprehensions is a series of transformation of the mental representation. Most of them especially adopt van Dijk and Kintsch's distinction among the surface code, the textbase, and the situation model [10]. The surface code preserves the exact wording and syntax of clauses. In a way, it may imply the concept of each word and the relation among such concepts. The textbase contains the explicit text propositions in a refined form that preserve the meaning of each clause. Also it includes a little number of inferences needed to maintain the local text coherence. The situation model is the mental microworld composed of the setting, the people, the object, and action that are involved in the event described by each clause and whole text.

Historically prevalent theory of the mental representation such as the text base and the situation model has premised the amodal symbol system that is conceptual and propositional [11] [12] [13] [14]. However, the perspective on the discourse comprehension was influenced by the development of the cognitive linguistics and the brain science? In particular, theoretical implications and empirical evidences obtained by neuroimaging (e.g., fMRI) researches in the brain science dramatically impact the conventional perspective on mental operations underlying the cognitive processes common to various human activities. Based on the findings given by the brain science, Barsalou has recently presented alternative to the amodal symbolic system in the form of Perceptual Symbol System [15] [16]. His theory assumes that modal symbolic system in the brain supports diverse forms of the mental simulation across different cognitive processes (e.g., perceptual, motor, and affective simulations). Furthermore, it was integrated with the situated cognition. Consequently, he proposed that simulations typically embody the concepts involved in the situation model such as characters, objects, actions, events, and mental state [17].

Perceptual Symbol System obviously impacted on the research of the situational model in discourse comprehension.

2.2 Situation Model

In order to catch on the core step in the present study, the situation model has to be introduced definitely. In two consecutive sections, we discuss the situation model more specifically.

Until the early 1980s, many cognitive psychologists viewed the discourse comprehension as the construction and retrieval of the mental representation of the text itself. This perspective was changed by two seminal books published in 1983 [18] [10]. Their ideas about the mental representation of situation described verbally, which has become to be known as the mental model [18] or the situation model [10], has still underlay on the amodal symbol system such as concepts and propositions. However, the great shift has recently occurred from ideas of the language role and empirical evidences for the situation model as the modal representation.

Rather than treating the language as information to analyze syntactically and semantically and then store as the concept and the proposition in memory, the language is now seen as a set of processing instruction on how to construct a mental representation of the described situation [9]. This perspective might reflect on the salient rise of the cognitive linguistics as described earlier. In addition to these new perspective of the language role, the empirical evidence of the brain science and theoretical suggestions given by Perceptual Symbol System made the notion of the discourse comprehender as an immersed experiencer in the mental microworld described in each sentence [19][20][21]. Because this view suggests both the speaker and the listener actually experience a sequence of events in a discourse as if they were actors or observers, it is consistent with the notion of the comprehension as the mental simulation. Additionally it is also consistent with other view such as the embodied cognition. According to the idea of the embodiment, the cognition is grounded in perception and action and relies on the use of perceptual and motor representations rather than of abstract, amodal, and arbitrary mental representations such as propositional networks or feature lists.

The empirical evidences amassed recently indicated that the visual representation is usually activated during discourse comprehension. The visual representation includes visual representation of the object shape [22] [23], orientation [24], and motor direction [25]. These evidences suggest that the situation model is the mental representation which permits both the speaker and the listener to simulate the entities and the event described by each sentence perceptually.

Additionally many researchers obtained the empirical evidence that motor representations are also activated during discourse comprehension. In the area of the brain science, Pulvermüller found that when participants read the word for an action, the motor system becomes active to represent its meaning [26]. Other researchers have assessed whether physical actions affect the sentence comprehension using behavioral measures. Klatzky et al. showed the priming effect of a motor action on the time to judge the sensibility of a simple phrase describing an action [27]. Similarly, comprehension is facilitated when the action to make a response is consistent with text meaning [28] and also when the action to control text presentation is consistent [29]. These findings also implied that the situation model is the mental representation which permits both the speaker and the listener to do motor simulation of the action described by each sentence.

In consideration of those findings, it should be specifically examined how to represent the situation model as the mental representation. The way to represent the situation model, which will be proposed later, should reflect the empirical evidence given by the recent studies mentioned above.

2.3 Processes for Constructing Situation Model

To the present, several researchers have proposed the models that explain how to construct the situation model. For example, the structure-building framework [18], the construction-integration model [30], the landscape model [31], and the event-indexing model [32] [33] [34] are representative ones. On the basis of comparison among them, the event-indexing model was selected as the theoretical foundation. The determinative reasons for selecting it come from the fact that it specified the cognitive processes that are likely to predict the time for reading each sentence of the story text.

The event-indexing model makes claims about both on-line comprehension and the mental representation resulting on the reader's long-term memory. According the model, on reading the story, each event described in each sentence is decomposed into five indexes: time, space, causality, intentionality (i.e., character's goal), and character/object. These

dimensions correspond to the dimensions listed by Chafe [35]. Then the coherent situation model is constructed by connection with shared indexes between the currently processed event and the constructed situation model until now.

Incoming events can be more easily connected with evolving situation model to the extent that they share indexes with the current state of the situation model. Thus, the event-indexing model makes the general prediction that the processing load during comprehension varies as a function of the number of situational indexes shared between the currently processed event and the current state of the situation model. This processing load hypothesis was supported by several studies [32] [36] [37].

3 Situation Frame

3.1 What is Situation Frame?

Zwaan and Radvansky tried to use the frame concept to represent the situation model as one type of knowledge representation [34]. Since the frame is originated to represent the knowledge of visual world by Minsky [38], it seems to be appropriate to represent the situation model. They assumed that the establishment of the spatio-temporal frame is obligatory during the construction of the situation model so that it grounds situations in space and time. And furthermore, they referred to the entity frames (the character and the object frames) associated with physical and mental attributes. Although they didn't indicate so with certainty, it is appeared that they regarded the situation model as the knowledge representation with the hierarchical structure setting spatio-temporal frame at a top phase.

According their suggestion, we regarded the situation model as the multilayered frame representation containing the constituents of the event described in the story. Figure 2 shows the structure and components of the multilayered frame representation named as the situation frame. The situation frame will be explained in a little detail below.

3.2 Mental Operations for Constructing Situation Frame

The reader extracts the meaning from each sentence of the story using lexical and grammatical knowledge. Such meaning was assumed to be composed of the predicate, the character, who involved the personalized animal, object, the instrument, the source, the goal, and the time concepts and so on according to the case frame representation [39]. In the present study, the components of the meaning were inputs for constructing the situation frame. It is explained below how to generate several kinds of frame involved in the situation frame and to insert the value into slots equipped each frame with.

Scene Frame When the predicate concept is the verb relevant to transfer of the character from one space to another space and the goal and/or the time concepts refer to specific values, the scene frame is generated. The scene frame has the space, the time, the character, and the objects slots. On generating the scene frame, the space and/or the time concepts are inserted into their correspondent slot. And when a character and/or an object concept have the specific value at first time, it is inserted into its correspondent slots. The values in the character and the object slots are also the reference point of their correspondent frames.

Character Frame When a character concept is first inserted into the character slot in the scene frame, the character frame is generated. At the same time, the value is inserted into the name slot. Then, the value of the predicate concept referred to the physical or psychological attribute of the character is inserted into

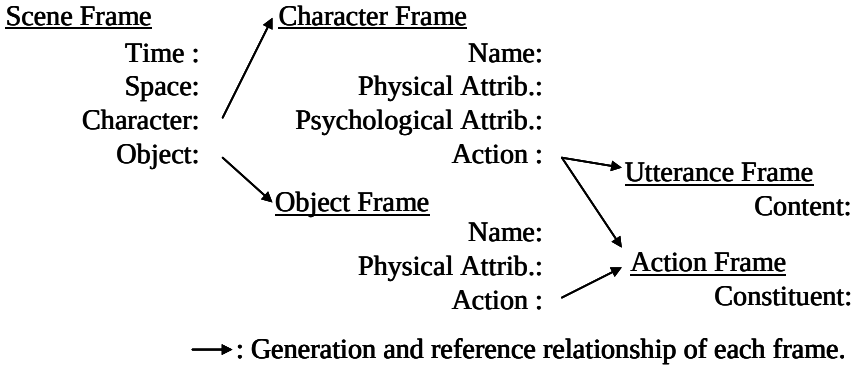

Figure 2. Configuration of multilayered situation frame.

the correspondent slot after the value of the character concept is matched against the value that has been inserted into the character slot of the scene frame already. At the same time when the value of the predicate concept referred to the action accompanied with the physical movement or the utterance is inserted into the action slot, the frame according to it is generated.

Object Frame When an object concept is first into the object slot, the object frame is generated. And the object concept is inserted as the value in to the name slot of the object frame. Once the object frame is generated, the value of the object concept is matched against the value in the object slot and inserted into the name slot. And the predicate concept referred to the physical attribute of the object is inserted into the slot of the physical attribute. On the other hand, when the predicate concept is relevant to the action, its value is inserted in to the action slot and instigates to generate the action frame. If the action frame is identical with the action frame generated from the character frame at the same moment, their frame are unified.

Action and Utterance Frames As mentioned above, the action frame is generated by the predicates concept referred with the action accompanied with physical movement, and the utterance frame is generated by the predicate referred to the action relative to the utterance (e.g., say, claim, murmur, and so on). On the utterance frame, the contents are written simultaneously.

Remember that we intend to devise the modal representation of the events described in the story, listened to in the conversation, or directly observed in daily life. Therefore, the scene, the character, and the object frame are the perceptual representation. On the other hand, the action and the utterance frame is the motoric representation.

4 Experiment

4.1 Method

Storyteller and Stories A middle-aged woman was asked to be the storyteller for reading five fairy stories. For nearly two decades, she participated in the volunteer activity that tries to contribute to the early childhood education by reading the story for children. She is the experienced storyteller with standard narrative style rather than the professional storyteller with distinct and artistic narrative style. On her request, we gave her the story texts in advance. Then she read aloud the stories to several undergraduate students in our laboratory.

Five stories were chosen from two illustrated books published for reading to the children. These stories were all classical Japanese fairy stories as follows; “The Stone Buddhas”, “A Crane’s Gratitude”, “The Rolling Rice Ball”, “The Princess of the Moon”, and “Click Click Mountain”[40][41].

Measuring Pause Length at Sentence Boundary The storyteller’s speech sound of reading each story was recorded with the digital recorder (TOSHIBA DMR-SX2). Her speech sounds were stored onto the USB memory in WAVE file format. Using the SUGI speech analyzer (ANIMO ANMSW-SSA0101), two cooperators measured the time length of the pause at final period of each sentence in millisecond. Their measured time lengths of the same pause were averaged to be input data for the explanatory variable in the multiple linear regression analysis.

Coding of Mental Operations for Constructing Situation Frame It is assumed that constructing the situation frame is equivalent to the transforming from the sentence meaning to the situation model. As mentioned before, on constructing the situation frame, the several kinds of the mental operation are needed. In order to estimate what kinds of the mental operation are carried out, the case-frame representation of each sentence meaning was manually transformed to the situation frame. In this way, we coded the mental operations, which are needed to transform the meaning of each sentence into the situation frame, according to the coding system shown in Table 1. Total number of coded sentences was sixty seven.

Table 1
Codes Given to Mental Operations to Construct Situation Frame

- GSm: Generation of Scene Frame
  - WTS: Writing Time Concept in Time Slot
  - WSS: Writing Space Concept in Space Slot
  - WGSCn: Writing Character Concept in Character Slot and Generation of Character Frame
  - WGSOn: Writing Object Concept in Character Slot and Generating Object Frame
- WGSCn:
  - WCPH: Writing Physical Attribute Concept in Physical Attribute Slot
  - WCPS: Writing Mental Attribute Concept in Mental Attribute Slot
  - WGCA: Writing Action Concept in Action Slot and Generating Action Frame
  - WGCU: Generation of Utterance Frame and Writing Utterance Content
- WGSOn
  - WOPH: Writing Physical Attribute Concept in Physical Attribute Slot
  - WGOA: Writing Action Concept in Action Slot and Generation Action Frame

Other Codes
- GSH: Preparation of Generating Scene Frame
- UF: Binding of Character Frame or Object Frame

Note m:0=previously given frame, 1=generated frame during comprehension
nは、0=original frame, 1=recalled frame, 2=changing or adding frame name

For example, the codes given to the sentences

involved in one paragraph of "The Rolling Rice Ball" and the pause length measured at each sentence boundary are presented in Table 2.

Table 2
Sequence of Codes Given to Each Sentence and Pause Length at Its Final period

A long ago, hardworking Grandpa lived somewhere. (1357.3)
[GS0 WST WSS WGSC0 WCPS WGCA]
Every day、Grandpa hold the rice balls made by Grandma
[WGSC0 WGCA WGSO0 WGOA WGCA WGOA]
and went to work the fields. (1830.7)
[WGSO0 WGCA WGOA GSH]
Over the course of plowing a field,
[GS0 WSS WGSO0 WGCA WGOA]
time for lunch came. (1550.0)
[GS0 WST]
Grandpa sit down on the stubble
[WGSC1 WGCA WGSO0 WGOA]
and opened the lunchbox. (2093.3)
[WGSO0 WGCA WGOA]+[WGCU]
" Then I am ready to eat, Grandma. "

Note the pause length (in msec.) is in parentheses ( ) and the sequence of codes is in brackets [ ]

4.2 Results

We set the pause length as the objective variable and the mental operations as the explanatory variables to conduct the multiple regression analysis. Sixteen kinds of the mental operation carried out to construct the situation frame are shown in Table 3. On the basis of the number of each mental operation when the case-frame representation of each sentence was transformed into the situation frame, the value of each explanatory variable was determined.

Table 3
Explanatory Variable of Multiple Regression Analysis

Variables associated with generation of frame	
Scene frame :	GS0, GS1, GSH
Character frame :	WGSC0, WGSC1, WGSC2
Object frame :	WGSO0, WGSO1, WGSO2
Conversation :	WGCU
Action (State):	WGCA, WGOA
Variables associated with insertion of value in slot	
Slots of scene frame :	WST, WSS
Slots of character frame :	WCPS
Slots of object frame :	WOPH

When the full regression method was used, the results of the multiple regression analysis indicated that sixteen explanatory variable can predict the pause length at each sentence boundary with considerable accuracy ($F(16,50)=4.660$, $p<.001$; determination coef., .599; multiple correlation coef., .774; Durbin-Watson ratio, 2.172). Specifically, GSH, WGSO0, and WGCU had significant effect on the pause length. Figure 3 presents the relationship between the theoretical value and the observation value.

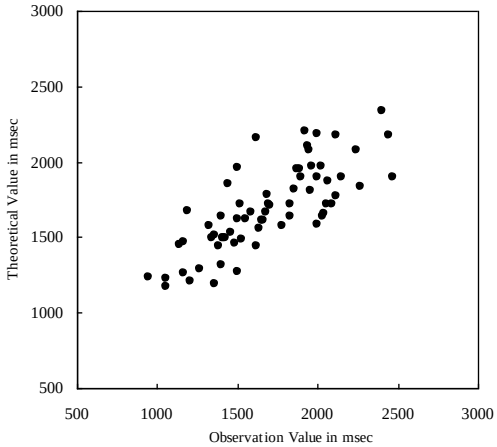

Figure 3. Distribution of observation value and theoretical value given by the multiple regression equation derived from the full regression method.

These results, however, might show the inflated predictable power of the multiple regression formula because of the number of the explanatory variables.

Consequently, the multiple regression analysis was conducted again using the stepwise regression method. The results indicated that GS1, GSH, WGSO0, WGSO2, and WGCU influenced on the pause length significantly as shown in Table 4. The result of ANOVA ($F(5,61)=15.352$, $p<.001$) and the measures of prediction accuracy indicated that only five explanatory variables were sufficient to predict the pause length. Table 5 shows several measures of prediction accuracy specifically. Figure 4 shows the relationship between the relationship between the the-

Table 4
Results of Multiple Regression Analysis with Stepwise Regression Method

Explanatory Variable	Partial Regression Coef.	F-value	T-value	Probability	Partial Correlation Coef.	Simple Correlation Coef.
WGSO0	223.72	16.765	4.094	0.001**	0.454	0.509
GSH	487.89	34.544	5.877	0.001**	0.601	0.372
WGC	318.41	20.570	4.535	0.001**	0.502	0.178
GS1	310.55	12.782	3.575	0.001**	0.416	0.066
WGSO2	414.38	3.904	1.976	0.053+	0.245	0.196
Constant	1230.85	363.757	19.072	0.001**		

oretical value and the observation value. These results indicate that the multiple regression formula with five variables and constant can predict the pause length at each sentence boundary with considerable accuracy.

Table 5
Several Measures of Prediction Accuracy

Determination coef.	0.557
Adjusted determination coef.	0.521
Multiple correlation coef.	0.746
Adjusted multiple correlation coef.	0.722
Durbin-Watson ratio	2.101

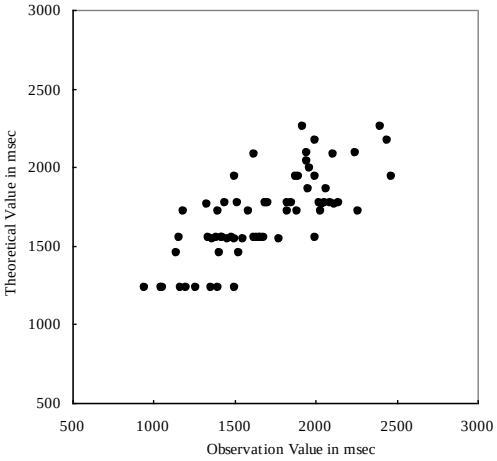

Figure 4. Distribution of observation value and theoretical value given by the multiple regression equation derived from the stepwise for- and backward regression method.

4.3 Discussion

4.3.1 Results and Processing Load Hypothesis

It is important that the significant explanatory variables, which have an influence on the pause length, referred to the frame generation. The processing load hypothesis implies that reading time of each sentence increases in proportion to the number of the generated frame when the reader comprehends the story. However, because it was not the reader but the listener that comprehends the story in present study, the storyteller must read it in consideration of such processing load of the listener. The results of the multiple regression analysis were indicated that our experienced storyteller intentionally or unintentionally controlled the pause length at each sentence boundary to help the listener comprehend the story. Then we will examine more specifically the results.

4.3.2 Effect of Scene Frame Generation

The results indicated that the generation of the new scene frame increases the pause length by 311msec. In addition, it was indicated that the generation of the scene frame based on the prospective description increases the pause length by 488msec.

These results are critically important because the scene frame is assumed to embody the concepts that compose the event described by each sentence. The preliminary generation of the scene frame, however, did not effect on the pause length. One reason of this result might come from the fact that the first sentence of the story generally described the space and the time through the story.

The structure-building model indicates that laying the foundation of each event occurring in the story requires the additional time and the processing resource [9]. In the experiment to support her model, the temporal and/or spatial cues were actually involved in each sentence that described the foundation of the event [43]. The event-indexing model also suggests that the patio-temporal framework is a critical component in construction of the situation model [34]. Therefore, it is possible that implications of these models are consistent with the indication of our results.

4.3.3 Effects of Character and Object Frame Generation

Any generation of the character frame didn't have an effect on the pause length at all. However, the generation of the object frame and the change of the object frame name have significant and marginally significant effect respectively. The generation of the object frame increases the pause length by 224msec. And the change of the object frame name also increases it by 414msec.

At first glance, these results are appeared to be strange because each character play an important role in the story. The character frames may be ready to generate so that the characters in stories used in the present study are prototypical ones coming on the folk story (e.g., grandpa, grandma, young hero, princess, and so on). So it is assumed that the character frame of each prototypical character becomes available immediately on hearing the storyteller's reading.

On the other hand, it is thought that each object frame is not easily generated so that the objects involved in the folk story are so diverse and sometimes unusual. If the storyteller was permitted to communicate with gestures, she probably used her hands, or show and draw the picture to embody the described object. When it is unusual, she was sure to do so. Therefore, it is implied that she increased the pause length to help the listener embody the object concept and simulate the event relevant to it.

Changing the name of the object frame is not so easy because not only the frame generation but also the reference (or identification) processes intervene in it. Such processing load predicted by the storyteller might induce to greater increase of the pause length.

4.3.4 Effect of Utterance Frame Generation

The utterance frame in the third layer of the situation frame is originally set on the basis of difference of the sentence type in the story. The generation of the utterance frame increased the pause length by 318msec.

Unlike the narrative sentence, the conversational sentence makes the listener to stand near the character and immerse in the microworld described in the story furthermore. Although the utterance frame is assumed to be generated when the listener hears the reading of each conversational sentence, such difference of the sentence type may be reflected the increase of the pause length. It may also suggest that the increase of the pause length on generating the utterance frame is reflected on the time that the storyteller needs to tune her voice to the character's voice quality.

4.3.5 Effects of Action Frame Generation and Value Insertion

Both kinds of multiple regression analysis didn't indicate the significant effect of the generation of the action frame. The action frame is set in the third layer and has particularly strong tie with the character frame that has little effects on the pause length. Therefore, the action frame may be generated together with the character frame on hearing the reading of each sentence. The generation of the action frame must be closely examined in the future study because the two critical indices (i.e., intention and causality) of the described event, which were not distinctly treated in the present study, are included in the action.

The insertion of value in any slots had little effect on the increase of the pause length. Since the insertion means the association with the frame, it may not take time as to need the increase of the pause length.

5 Validity of Pause Length Estimated by Formula

5.1 Two Verification Experiments

The results given by the multiple regression analysis suggested that the pause length at each sentence boundary can be predicted by some kinds of the frame generation on constructing the situation frame. Based on the formula obtained in our study, the pause length set by the experienced storyteller can be calculated with considerable accuracy.

However, the validity of the pause length estimated by the formula, which was obtained with the stepwise regression method, has to be tested. Then we performed two experiments using the synthesized speech in which the specific pause length at each sentence boundary has been set preliminarily. We defined the validity as the listenability and the comprehensibility of the synthesized speech.

5.2 Experiment 1

5.2.1 Method

Material One paragraph of "The Rolling Rice Ball" was selected as the material transformed into the synthesized speech to test the validity of the pause length. Because it involved the ten sentences, each pause length at nine sentence boundaries was manipulated preliminarily according to three experimental conditions.

Experimental Conditions On no pause condition, all pauses were carried away from the synthesized speech. And on the constant pause condition, the pause length was 1400msec. at all sentence boundaries. Then on the estimated pause condition, the pause length estimated by the formula was set at each sentence boundary.

Participants and Procedure Ten undergraduate students participated in the experiment 1 as the listeners for hearing the synthesized speech given by computer on each experimental condition. Furthermore, immediately after each synthesized speech was heard, they assessed the listenability and the comprehensibility of it using three choices method. Specifically, they judged each measure as more, neutral, or less on the listenability and the comprehensibility.

5.2.2 Results

The results indicated that both the listenability and the comprehensibility are significantly different among three experimental condition, $\chi^2(4) = 31.25$ and $\chi^2(4) = 30.272$, all $ps < .001$ respectively. As shown in Figure 5 and Figure 6, the synthesized speech with no pause was the least listenable and comprehensible one. However, the synthesized speech with the estimated pause length was likely to be more listenable and comprehensible than one with the constant pause length.

5.2.3 Discussion

These results obviously indicated the superiority of the estimated pause length on the listenability and the comprehensibility to ones in other two pause conditions. Therefore, these evidences indicated the estimated pause length is valid to be set in the synthesized speech in the experiment 1.

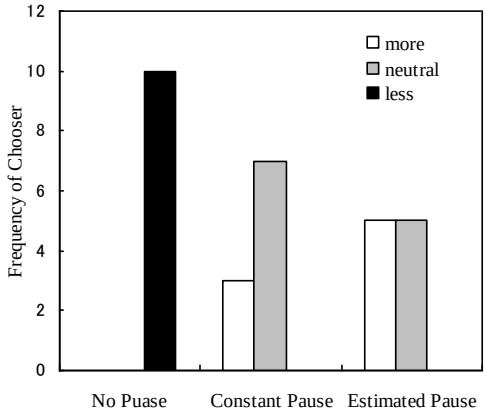

Figure 5. Listenability of synthesized speech on three pause length conditions.

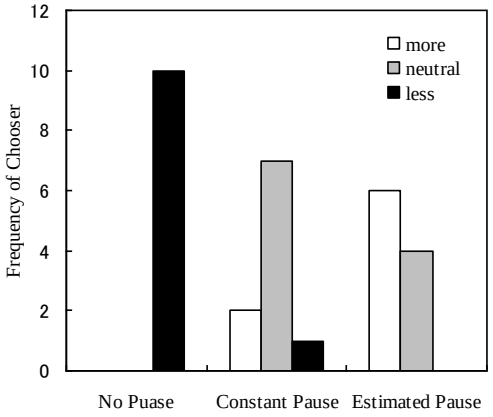

Figure 6. Comprehensibility of synthesized speech on three pause length conditions.

5.3 Experiment 2

5.3.1 Method

Material There are some parts of the story where the difference between the pause lengths set by our storyteller and estimated by the formula is larger. We selected one paragraph of them, which is involved in "The Stone Buddhas", to be the material that was transformed into the synthesized speech.

Experimental Conditions Because the paragraph was composed of seven sentences, each pause length at six sentence boundaries was preliminarily manipulated to make two experimental conditions.

On the storyteller's pause condition, each pause length in the synthesized speech was originally set by our storyteller. On another condition named as the estimated pause condition, each pause length calculated by means of the formula.

Procedure The procedure was same as the experiment 1 except ten new undergraduate students who participated in this experiment and two experimental conditions.

5.3.2 Results

The results indicated that both listenability and comprehensibility are not significantly different between two experimental conditions, $\chi^2(1) = 0.000$ and $\chi^2(2) = 1.667$, all ps<.435 respectively. Figure 7 and Figure 8 show theses results specifically. At first glance, estimated pause length appears to increase both the listenability and the comprehensibility of the synthesized speech in comparison to the pause length set by the storyteller. However, the statistical analyses based on the frequency clearly indicated that the pause length not only estimated by the formula but also set by the storyteller have same effect on both listenability and

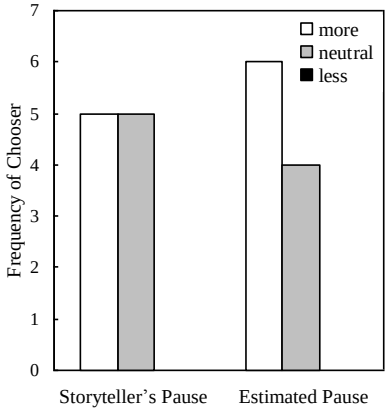

Figure 7. Listenability of synthesized speech on two pause length conditions.

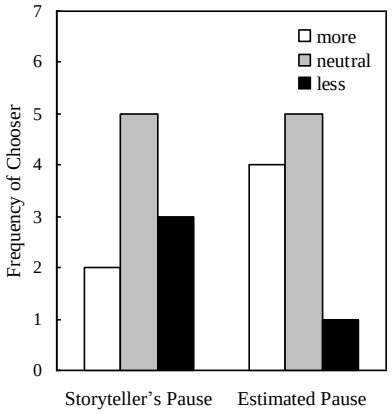

Figure 8. comprehensibility of synthesized speech on two pause length conditions.

comprehensibility.

5.3.3 Discussion

The results of the experiment 2 additionally confirmed that the pause length estimated by the formula is comparable to the pause set by the storyteller. Therefore, the pause length, which is calculated by means of the formula, is useful to generate the synthesized speech that can support young children to comprehend the story.

6 General Discussion

The multiple regression analysis using the stepwise regression method suggested that the pause length at each sentence boundary can be predicted by some kinds of the frame generation on constructing the multilayered frame representation. Based on the formula obtained in our study, the pause length set by the experienced storyteller can be calculated with considerable accuracy.

The situation frame can be constructed by a series of natural language processing. The results of our study indicated that the mental operations can be predicted through such construction processes. Then if the mental operations are identified, we can estimate the pause length set by the experienced storyteller using the formula obtained in the present study. In comparison with the fact that the pause length in the synthesized speech is determined on the basis of the pause length set by a human storyteller who read the target text, our formula seemed to provide a reasonable way in order to estimate the pause length in advance.

In the verification experiment 1, it is implied that the time length estimated by the formula, which is derived by the stepwise regression method of multiple regression analysis, makes the synthesized speech more listenable and comprehensible than the pause length with a fixed value. Then, in the verification experiment 2, it is also implied that the formula can provide the pause length comparable to one set by human storyteller on the listenability and comprehensibility of the synthesized speech. These implications indicate that we were able to propose a series of the procedure that derive the formula to estimate the appropriate pause length in the synthesized speech for young children.

7 Conclusion

On reading the story to the children, the storyteller must try to carry a sequence of events to the listener in one way or another. That is, the storyteller reads the message with consideration for the listener's comprehension of it. The pause length set at each sentence boundary is likely to be one of the considerations because it is thought to help the listener extract meaning from the sentence.

In the area of the psycholinguistics, it is assumed that the listener constructs the situation model to comprehend the story. The situation model is the mental representation of the microworld described in each sentence of the story. Recently, the researches in the area of brain science indicated that the situation model is the modal representation permitting the listener to simulate it mentally. In consideration of these accumulating evidences, we devised the knowledge representation named as the situation frame to specify the situation model in this study. The situation frame has multilayered frame structure composed of the scene, the character, the object, the action, and the utterance frames.

The present study examined the pause length at each sentence boundary set by storyteller on reading the story for the listener to construct the situation model. Specifically, we asked the storyteller to read the story as if she was facing the younger children. Then, the multiple regression analysis was conducted to obtain the formula to estimate the pause length from the processing loads for constructing the situation model. We assumed that the processing loads directly reflect a series of the operation needed for constructing the situation frame. The results showed that the pause length can be predicted from the formula given by the multiple regression analysis with considerable accuracy. Therefore, the pause length at each sentence boundary can be determined preliminarily by the identification of the operations for constructing the situation frame.

References:

[1] T. Nakamura, Kansei of pause. In H. Harashima (Ed.), *Kansei information processing* (pp. 151-169), Tokyo: Oume co., 1994. (translated from Japanese).

[2] D. C. O'connell, S. Kowal, U. Bartels, H. Nundt, & D. A. van de Water, Allocation of time in reading aloud: Being fluent is not the same as being rhetorical, *Bulletin of psychonomic Society*, Vol. 27, No. 3, 1989, pp. 223-226.

[3] Nakamura, T. (1997). Relationship between speech content and pause. J*apanese Psychological Association Sixty-first conference collected papers*, 1997, p. 691. (translated from Japanese).

[4] S. Komori, C. Nagaoka, & T. Nakamura, Right pause length in speech: Examination by the use of serial judgment method, *Human Interface Symposium collected papers*, 1999, pp. 393-398. (translated from Japanese).

[5] S. Komori, T. Nagaoka, M. R. Draguma, & T. Nakamura., Effect of text structure on optimal pause length in speech. *Journal of Japanese Human Interface Association*, Vol. 4, No. 1, 2002, pp. 59-65. (translated from Japanese).

[6] M. Sugitou, Relationship among pause length speech time length, and meaning. In M. Sugitou (Ed.), *Research on Japanese speech sound, Vol. 1: Japanese speech sound* (pp. 27-42), Osaka: Izumisyoin, 1994a. (translated from Japanese).

[7] M. Sugitou, Approach to effective reading. In M.Sugitou (Ed.), *Research on Japanese speech sound, Vol. 1: Japanese speech sound* (pp. 43-60), Osaka: Izumisyoin, 1994b. (translated from Japanese).

[8] R. W. Langacker, Discourse in cognitive grammar, *Cognitive Linguistics*, Vol. 12, No. 2, 2001, pp. 143-188.

[9] M. A. Gernsbacher, *Language comprehension as structutre building*, Hisdale, NJ: Lawrence Erlbaum Associates, 1990.

[10] T. A. van Dijk, & W. Kintsch, *Strategies in discourse comprehension*, New York: Academic Press, 1983.

[11] A. Newell, & H. Simon, *Human problem solving*, Englewood Cliffts, NJ: Prentice-Hall, 1972.

[12] Z. W. Pylyshyn,(1981). The imagery debate: Analogue media versus tacit knowledge, *Psychological Review*, Vol. 88, No. 1, 1981, pp. 16-54.

[13] Z. W. Pylyshyn, Z. W. *Computation and cognition: Toward a foundation for cognitive science*, Cambridge, MA, MIT Press, 1984.

[14] W. Kintsch, *Comprehension: A paradigm for cognition*, Cambridge, MA: Cambridge University Press., 1998.

[15] L. W. Barsalou, Language comprehension: Archival memory or preparation for situated action? *Discourse Processes*, Vol. 28, No. 1, 1999a, pp. 61-80.

[16] L. W. Barsalou, Perceptual symbol system. *Behavioral and Brain Science*, Vol. 22, No. 4, 1999b, pp. 577-660.

[17] L. W. Barsalou, Situated simulation in the human conceptual system, *Language and Cognitive Processes*, Vol. 18, No. 5, 2003, pp. 513-516.

[18] P. N. Johnson-Laired, *Mental models: Toward a cognitive science of language, inference, and consciousness*, Cambridge, MA: Harvard University Press, 1983.

[19] R. A. Zwaan, The immersed experiencer: Toward an embodied theory of language comprehension, In B. H. Ross (Ed.), *The Psychology of Learning and Motivation: Advance in Research and Theory* (pp.35-62), New York: Academic Press, 2004.

[20] R. A. Zwaan, R.A., & C. J. Madden, Embodied sentence comprehension, In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp.224-245), New York: Cambridge University Press, 2005.

[21] R. A. Zwaan, & D. N. Rapp, Discourse comprehension, In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics, 2nd edition* (pp.725-764), New York: Academic Press, 2006

[22] R. A. Zwaan, R. A. Stanfield, & R. H. Yaxley, Language comprehenders mentally represent the shape of objects, *Psychological Science*, Vol. 13, No. 2, 2002, pp. 168-171.

[23] R. A. Zwaan, & R. H. Yaxley, Lateralization of object-shape information in semantic processing, *Cognition*, Vol. 94, No. 2, 2004, pp. B35-B43.

[24] R. A. Stanfield, & R. A. Zwaan, The effect of implied orientation derived from verbal context on picture recognition, *Psychological Science*, Vol. 12, No. 2, 2001, pp. 153-156.

[25] R. A. Zwaan, C. J. Madden, R. H. Yaxley, & M. E. Aveyard, Moving ward: Dynamic representations in language comprehension, *Cognitive Science*, Vol. 28, No. 4, 2004, PP. 611-619.

[26] F. Pulvermüller, Brain mechanisms linking language and action, *Nature Reviews Neuroscience*, Vol. 6, No. 7, 2005, pp. 576-582.

[27] R. L. Klatzky, J. W. Pellegrino, B. P. McCloskey, & S. Doherty, The role of motor representations in semantic sensibility judgments, *Journal of Memory and Language*, Vol. 28, No. 1, 1989, pp. 56-77.

[28] A. M. Glenberg, & M. P. Kaschak, Grounding language in action, *Psychonomic Bulletin and Review*, Vol. 9, No. 3, 2002, pp. 558-565.

[29] R. A. Zwaan, & L. J. Taylor, Seeing, acting, and understanding: Motor resonance in language comprehension, *Journal of Experimental Psychology: General*, Vol. 135, No. 1, 2006, pp. 1-11.

[30] W. Kintsch, *Comprehension: A paradigm for cognition*, Cambridge, MA: Cambridge University Press, 1998.

[31] P. Van den Broek, K. Young, Y. Tzeng, & T. Linderholm, The landscape model of reading inferences and the online construction of a memory representation, In H. Oostendrop & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71-98), Mahwah, NJ: Erlbaum, 1999.

[32] R. A. Zwaan, M. C. Langston, & A. C. Graesser, The construction of situation model in narrative comprehension: An event-indexing model, *Psychological Science*, Vol. 6, No. 5, 1995, pp. 292-297.

[33] R. A. Zwaan, J. P. Mangliano, & G. A. Graesser, Dimension of situation model construction in narrative comprehension, *Journal of Experimental Psychology: Learning,Memory,and Conition*, Vol. 21, No. 2, 1995, pp. 386-397.

[34] R. A. Zwaan, & G. A. Radvansky, Situation models in language comprehension and memory, *Psychological Bulletin*, Vol. 123, No. 2, 1998, pp. 162-185.

[35] W. L. Chafe, The flow of thought and the flow of language, In T. Givón (Ed.), *Syntax and semantics, Vol. 12: Discourse and syntax* (pp. 159-181). New York: Academic Press, 1979.

[36] R. A. Zwaan, Five dimensions of narrative comprehension:The event-indexing model, In S.R.Goldman,A.C.Graesser, & P. van den Broek, (Eds.) *Narrative comprehension, causality, and coherence* (pp.93-110), Nahwah, NJ: Lawrence Erlbaum Associates, 1999.

[37] J. P. Magliano, R. A. Zwaan, & A. C. Graesser, **T**he role of situational continuity in narrative understanding, In S.R. Goldman & H. van Oostendorp (Eds.), *The construction of mental representations during reading* (pp 219-245), Mahwah, NJ: Erlbaum, 1999.

[38] M. A. Minsky, A framework for representing knowledge, In P.H. Winston (Ed.), *The psychology of computer vision* (pp. 211-277), New York: McGraw-Hill, 1975.

[39] C. J. Fillmore, The case for case, In E. Bach & R. T. Harms (Eds.), *Universal in linguistic theory* (pp. 1-88), New York: Holt, Rinehart & Winston, 1968.

[40] Shyuhutoseikatu-shya (Ed.), *Illustrated Book of old Japanese tales: Entertaining Book for parent and child 3-6 year old*, Tokyo: Shyuhutoseikatu-shya, 1991.

[41] K. Nishimoto (Ed.), Tales for Reading to Young Children: 2nd series, Tokyo: Popura-shya, 2000.

[42] W. Kintsch, Notes on the structure of semantic memory, In E. Tulving & W. Donaldson (Eds.), *Organization of Memory* (pp. 247-308), New York: Academic Press, 1972.

[43] K. Haberlandt, Components of sentence and word reading times, In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 219-252), Hillsdale, NJ: Erlbaum, 1984.