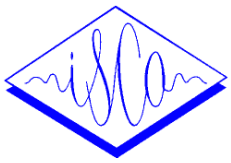

First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

PAUSE RULE FOR JAPANESE TEXT-TO-SPEECH CONVERSION USING PAUSE INSERTION PROBABILITY

Kazuhiko IWATA, Yukio MITOME and Takao WATANABE

C & C Information Technology Research Laboratories, NEC Corporation
4-1-1 Miyazaki, Miyamae-ku, Kawasaki, 213 JAPAN

ABSTRACT

A pause rule for Japanese text-to-speech conversion technique is proposed, which can determine natural pause locations. In order to insert several pauses at appropriate *bunsetsu* boundaries (which resemble "phrase" boundaries in English), the probabilities (pause insertion probabilities) that words are followed or preceded by pauses are used. The pause insertion probabilities are obtained by statistically analyzing a large number of sentence utterances. It was found that the probabilities differ from each other, according to the parts of speech for the words adjacent to the pauses. By the rule, adequate pauses are inserted at the *bunsetsu* boundaries whose pause insertion probabilities are high. An evaluation experiment for the rule was carried out, using 200 sentences. The result indicates that the pause locations, determined by the rule, are as natural, in 93% of the sentences, as those determined by humans. The rule is adopted by a Japanese text-to-speech conversion system.

1. INTRODUCTION

In order to develop a high quality text-to-speech conversion system, it is necessary to realize not only articulate speech but also natural prosody. Pauses are one of the most important prosodic features. In addition to generating natural intonations and rhythms, inserting several pauses at appropriate *bunsetsu* boundaries and controlling pause lengths are important. Several previous researches explored the features of the pause [1]-[4]. When a sentence is spoken, some pauses are inserted (1) in order to clarify the syntactic structure for the sentence, so that a listener can correctly understand the meaning of the sentence, and (2) in order to breathe, when uttering a long sentence. Pauses are generally inserted between two adjacent *bunsetsu* which have a weak relationship on the syntactic structure. Pause lengths become long when the relationship between two *bunsetsu* is weak and when the lengths of the *bunsetsu* are long. Consequently, the syntactic structure information is useful to determine natural pause locations and lengths.

In conventional pause rules, the syntactic structure for the sentence is used for inserting natural pauses [4], [5]. A syntactic structure analysis, however, requires a large amount of computation. Moreover, it is not always easy to obtain the syntactic structure correctly. Consequently, the determined pause are sometimes unnatural.

On the other hand, some Japanese function words, such as *kaku-joshi* (nominative particles) and *setsuzokushi* (conjunctions), often cause syntactic boundaries to be formed after or before them. Parts of speech for words can be obtained by a morphological analysis with a smaller amount of computation than when using the syntactic analysis. Information, concerned with parts of speech for words, is expected to be useful for a text-to-speech conversion system to determine natural pause.

In this paper, a relationship between the pause location and the part of speech is discussed as a preliminary experiment. To utilize the part of speech information for the determination of the pause locations, the pause insertion probabilities are introduced, which are the probabilities that words are followed or preceded by pauses. The probabilities for all parts of speech were investigated using a large number of sentence utterances. On the basis of the investigation results, a pause rule was proposed, which determines pause locations using the pause insertion probabilities. Section 2 describes the analysis results for the pause insertion probabilities. The determination method of pause locations and its evaluation result are discussed in Section 3 and Section 4, respectively.

2. PAUSE INSERTION PROBABILITIES

To verify that the part of speech information is useful to determine pause locations, the pause insertion probabilities for each part of speech were investigated. A speech database, which was used in the investigation, comprises speech signals for a large number of sentence utterances spoken by a male speaker and texts written in Kanji characters (Chinese ideographs). All pauses which appeared in the utterances were searched for manually and the locations for the pauses were written into the texts. The texts were broken down into words by the morphological analysis system, which was used for the Japanese text-to-speech conversion system [6]. Two kinds of pause insertion probabilities were considered. One was the probability that a word was followed by a pause (P_f). The other was the probability that a word was preceded by a pause (P_p). On the basis of the morphological analysis results, the number of total appearances of the part of speech (N_{total}), the number of appearances of the part of speech

followed by a pause (N_f) and the number of appearances of the part of speech preceded by a pause (N_p) were counted.

The pause insertion probabilities were determined as :

$$P_f = \frac{N_f}{N_{total}} \times 100 \quad (1)$$

$$P_p = \frac{N_p}{N_{total}} \times 100 \quad (2)$$

These probabilities were obtained for every part of speech.

A punctuation mark "、" called *tôten* (which corresponds to a comma in English) is generally put at *bunsetsu* boundaries, whose two adjacent words have a weak relationship in regard to the syntax. Hence, the *tôten* is a good cue to find boundaries where pauses should be inserted. The pause insertion probabilities were also investigated considering whether or not *tôten* exist at *bunsetsu* boundaries.

Table 1 shows the pause insertion probabilities that words are followed by pauses (P_f). The table reveals that the probabilities differ from each other, depending not only upon the parts of speech but also upon the existence of the *tôten*. Nouns with *tôten* are followed by pauses in 84.0 % probability, while those without *tôten* are in low probability, 2.5 %. For the total appearances of the nouns, the pause insertion probability is 18.2 %. Pauses tend, in high probability, to follow "ば (*ba*)" which is one of *setsuzoku-joshi* (conjunctive particles), "は (wa)", one of *kakari-joshi* (nominative particles), adverbs and conjunctions. In particular, the *setsuzoku-joshi* "ば (*ba*)" and the *kakari-joshi* "は (wa)" have considerably high pause insertion probabilities, not only when it is followed by a *tôten* (100 % and 95.5 %, respectively), but also when it is not (52 % and 43.7 %, respectively). *Ren'yôkei*, which qualify verbs and adjectives, are almost always followed by pauses, when they are followed by *tôten*.

Table 2 shows the pause insertion probabilities that words are preceded by pauses (P_p). When conjunctions and adverbs are preceded by *tôten*, they are preceded by pauses in high probability.

On the contrary, the pause insertion probabilities for other parts of speech are considerably low, in comparison with the parts of speech appearing in Tables 1 and 2. Parts of speech have individual pause insertion probabilities. This result suggests that the information concerned with the parts of speech and the existence of the *tôten* can be utilized, instead of the syntactic structure, to determine pause locations.

3. PAUSE RULE USING PAUSE INSERTION PROBABILITIES

A pause rule was constructed, on the basis of the investigation results. Figure 1 shows the flow of the pause location determination and Figure 2 shows an example of the pause insertion process. The rule inserts pauses into an input sentence as follows.

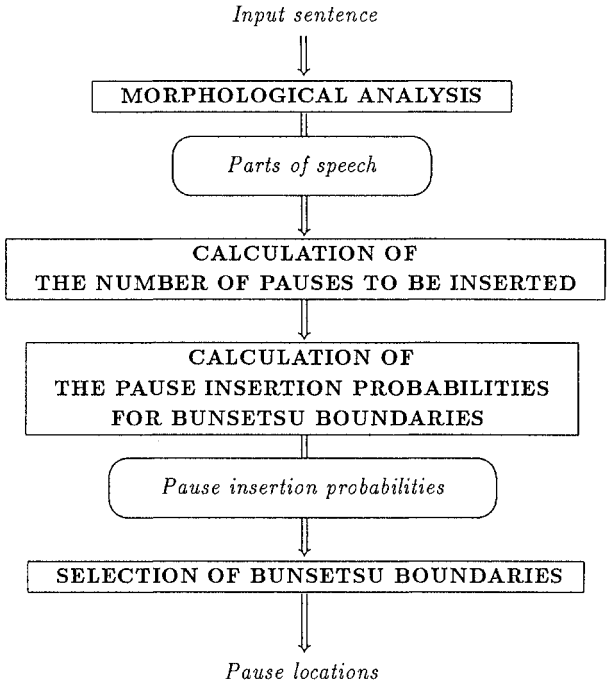

Figure 1. Pause location determination process.

An input sentence, written in Kanji characters, are analyzed and broken down into words by the morphological analysis system. Parts of speech for the words are derived from a word dictionary (*see* Figure 2 (a) *and* (b)).

The number of pauses (N_{pause}) which should be inserted into the sentence is determined. One of roles which pauses play is to allow the speaker to draw breath at normal intervals. More pauses appear when a sentence is longer. Therefore, the N_{pause} is calculated according to the number of moras for the sentence. The average number of moras for a breath group is used as a threshold to determine the N_{pause}. It was obtained from the speech database, beforehand.

At each *bunsetsu* boundary, the pause insertion probability P_f for the word which precedes the boundary and the probability P_p for the word which follows the boundary are obtained from Table 1 and Table 2, respectively (*see* Figure 2 (c)). When the word does not accompany a *tôten*, the probability for all appearances of the part of speech (which is shown as "Total" in Table 1 and Table 2) is used, instead of the probability for the appearances without a *tôten*. The reason is that *tôten* do not always appear in a sentence. A score of each *bunsetsu* boundary, which is used to determine pause locations, is defined as the sum of the probability P_f and the probability P_p (*see* (d)). The *bunsetsu* boundaries, whose scores are in the first N_{pause} rank, are selected as the boundaries where the pauses should be inserted. In this example,

Table 1. Pause insertion probabilities that words are followed by pauses (P_f).
(With *tôten* : probability when a word is followed by a *tôten*. Without *tôten* : probability when a word is not followed by a *tôten*. Total : probability for the total appearance of the word.)

Part of Speech	With *tôten*	Without *tôten*	Total
Noun	84.0	2.5	18.2
Conjunction	70.0	37.5	50.0
Adverb	87.5	6.5	11.4
kaku-joshi "が (*ga*)"	94.4	12.3	19.5
kaku-joshi "と (*to*)"	88.2	6.8	17.2
kakari-joshi "は (*wa*)"	95.5	43.7	53.1
setsuzoku-joshi "ば (*ba*)"	100.0	52.0	70.0
ren'yôkei of adjective	100.0	2.7	10.0
ren'yôkei of verb and auxiliary verb	97.6	11.5	53.5

Table 2. Pause insertion probabilities that words are preceded by pauses (P_p).
(With *tôten* : probability when a word is preceded by a *tôten*. Without *tôten* : probability when a word is not preceded by a *tôten*. Total : probability for the total appearance of the word.)

Part of Speech	With *tôten*	Without *tôten*	Total
Noun	91.4	15.0	26.3
Conjunction	100.0	6.7	17.7
Adverb	95.4	36.3	56.9
Prefix	100.0	14.3	21.7

(a)	彼女	は	その		花屋	で	赤い		花	を	買い	ました。
	kanozyo	*wa*	*sono*		*hanaya*	*de*	*akai*		*hana*	*o*	*kai*	*masita*
(b)	*n.*	*par.*	*par.*		*n.*	*par.*	*adj.*		*n.*	*par.*	*v.*	*auxil. v.*
(c)		P_f	P_p	P_f	P_p	P_f	P_p	P_f	P_p	P_f	P_p	
		53.1	54.6	0.0	26.3	20.0	16.7	0.0	26.3	4.7	5.3	
(d)		107.7		26.3		36.7		26.3		10.0		
(e)		< *pause* >		—		—		—		—		
(f)	ka"nozyowa ;		sono /		hana"yade //		akai /		hana"o /		kaima"s_ita.	

Figure 2. A pause insertion process example.

(a) The input text. (*She bought the red flower at the flower shop.*)
The boundaries represented by "||" are *bunsetsu* boundaries.

(b) The parts of speech for the words.
(*n.* : noun, *par.* : particle, *adj.* : adjective, *v.* : verb, *auxil. v.* : auxiliary verb.)

(c) The pause insertion probabilities for the parts of speech.

(d) The scores of the *bunsetsu* boundaries.

(e) The determined pause location is indicated by < *pause* >.

(f) The pronunciation symbol sequence for the speech synthesizer [6].

the *bunsetsu* boundary between "は (*wa*)" and "その (*sono*)" is chosen, as shown in (e).

Figure 2 (f) shows a pronunciation symbol sequence for the speech synthesizer [6]. The synthesizer inserts a pause according to a pause symbol ';'. In this work, only the relationship between the pause location and the part of speech information is discussed. The synthesizer uses a constant pause length. The relationship between the pause length and the part of speech information should be investigated in further research.

4. EVALUATION

An evaluation experiment for the pause rule was carried out. In addition to the sentences in the speech database which was used for the investigation of the pause insertion probabilities, approximately 200 sentences which were not in the database were used for the evaluation.

The criterion used to evaluate naturalness of the pause locations was whether or not the number of the determined pauses was adequate to utter the sentences at a normal speech rate, and whether or not the positions of the pauses were appropriate. Since using pauses is not restricted to only one manner, the number and the positions of pauses were judged to be correct, when they could be considered as natural or possible.

The experiment result indicated that, the number and the positions of the pauses determined by the proposed rule were as natural as those determined by humans, in 95.4 % of the sentences in the database and in 93.2 % of the sentences not in the database. These results suggest that the pause insertion probabilities are effective for the pause rule.

5. CONCLUSION

The pause rule, which inserted pauses into a sentence using the pause insertion probabilities, was proposed. The pause insertion probabilities were obtained by statistical analysis of a large number of sentence utterances. It was found that the probabilities differ from each other, according to the parts of speech and the existence of the *tôten*. The function words, such as conjunctive particles and nominative particles, had considerably high pause insertion probabilities. Several parts of speech had 100 % pause insertion probabilities, when they accompanied the *tôten*. The proposed rule could determine natural pause locations utilizing the part of speech information, which could be obtained more correctly than the syntactic structure. The evaluation experiment results indicated that the rule was useful for a Japanese text-to-speech conversion system.

Though the determination method only for the pause locations was discussed in this work, it is interesting to control the pause length using the part of speech information. The relationship between the pause length and the part of speech should be a subject for further investigation.

REFERENCES

[1] H. Fujisaki and T. Omura, "Characteristics of Durations of Pauses and Speech Segments in Connected Speech," *Proceedings of the Autumn Meeting of the Acoustical Society of Japan*, 2-1-19, November, 1971 (in Japanese).

[2] K. Hakoda and H. Sato, "Prosodic Rules in Connected Speech Synthesis," *Transactions of the Institute of Electronics and Communication Engineers of Japan*, Vol. J63-D, No. 9, pp.715-722, September, 1980 (in Japanese).

[3] T. Uyeno, H. Hayashibe, K. Imai, H. Imagawa and S. Kiritani, "Syntactic Constructions and Prosody in Japanese : On the Pauses at Phrase Boundaries," *Transactions of the Committee on Speech Research, The Acoustical Society of Japan*, S80-97, March, 1981 (in Japanese).

[4] K. Hakoda and H. Sato, "A Pause Insertion Rule for Connected Speech," *Transactions of the Committee on Speech Research, The Acoustical Society of Japan*, S74-64, March, 1975 (in Japanese).

[5] Y. Mitome and K. Fushikida, "Japanese Speech Synthesis by Rule Using Formant–CV, VC Compilation Method," *Transactions of the Committee on Speech Research, The Acoustical Society of Japan*, S85-31, July, 1985 (in Japanese).

[6] K. Iwata, Y. Mitome, J. Kametani, M. Akamatsu, S. Tomotake, K. Ozawa, and T. Watanabe, "A Rule-Based Speech Synthesizer Using a Pitch Controlled Residual Wave Excitation Method," *Proceedings of International Conference on Spoken Language Processing*, November, 1990.