ANALYSIS OF OCCURRENCE OF PAUSES AND THEIR DURATIONS IN JAPANESE TEXT READING

Hiroya Fujisaki *Sumio Ohno* *Seiji Yamada*

Science University of Tokyo
2641 Yamazaki, Noda, 278-8510, Japan

ABSTRACT

Pauses play important roles both for the intelligibility and the naturalness of speech. Their occurrences and durations in text reading are influenced by syntactic structures of the text as well as by physiological constraints of respiration on the part of the speaker. The present paper describes some of the preliminary findings on Japanese text reading, especially on the effects of the syntactic role of the preceding phrase on the rate of occurrence and the duration of a pause at a syntactic boundary.

1. INTRODUCTION

Previous studies by the present authors as well as by others [1-3] have shown that pauses occur in speech generally at syntactic boundaries. Their occurrence is unconditional at certain boundaries such as between sentences, but is more or less arbitrary at other boundaries. Their durations tend to be longer at the end of larger syntactic units, but have statistical variations. Pauses serve to facilitate comprehension on the part of the listener, and allow the speaker to inspire when necessary. A good speaker is generally capable of fulfiling these two requirements. Finding the characteristics of pauses and the rules/strategies of their insertion is therefore necessary to ensure naturalness and ease of comprehension in text-to-speech synthesis.

The present paper describes our efforts toward analysis and classification of the types of pauses, their durational characteristics and the rates of occurrence. At first, pauses in text reading are classified broadly into three categories on the basis of the locations where they occur: 1) between paragraphs, 2) between sentences, and 3) within a sentence. Pauses of the third category are further subdivided into finer categories according to the syntactic roles of uninterrupted sentence segments (henceforth 'phrases' for short) that immediately precede them. These phrases are to be regarded as candidates for the insertion of pauses in speech production. The rate of occurrence of a pause at the end of each of the 'candidate' phrase types, calculated on the basis of syntactic analysis of the text used for reading, provides a data for the statistical rules/strategies for pause insertion in text-to-speech synthesis.

2. A BROAD CLASSIFICATION OF PAUSES AND THEIR DURATIONAL CHARACTERISTICS IN TEXT READING

A text generally has a hierarchical structure consisting of paragraphs, which consists of sentences containing phrases. In text reading, pauses are always inserted between paragraphs, as well as between sentences within each paragraph, while they are inserted more of less arbitrarily between phrases within a sentence. Thus a speaker may choose to insert a pause at a deep syntactic boundary such as the one between a subordinate clause and a main clause, or at a location necessary to disambiguate a syntactic ambiguity, or if the preceding uninterrupted sentence segment is too long, or if he/she is simply out of breath. This leads to a natural classification of pauses in text reading into the following three broad categories:

1) Pauses between paragraphs (P1)
2) Pauses between sentences in a paragraph (P2)
3) Pauses within a sentence (P3)

In order to obtain quantitative knowledge on the durational characteristics of these three types of pauses, we measured the durations of pauses in reading a text in Japanese. The text was a short story of 15 sentences constituting 4 paragraphs. The text was read by a native male speaker five times at an average speech rate of 8.0 *morae*/s.

In order to let the speaker decide where and how long he should pause, no punctuation marks were provided, but a space was inserted at the end of each sentence, and the beginning of a paragraph was indicated by indentation. The speech signal was digitized at 10kHz with 16-bit precision. Pauses were detected and their durations were measured automatically using thresholds in short-time power. The results were visually checked and aligned with the text to ensure high accuracy.

The measured durations were further analyzed statistically. The results indicated that the rates of occurrence of durations of each of the three pause categories can be approximated by a Gaussian probability density function on a logarithmic scale of pause duration, as shown in Fig. 1. The

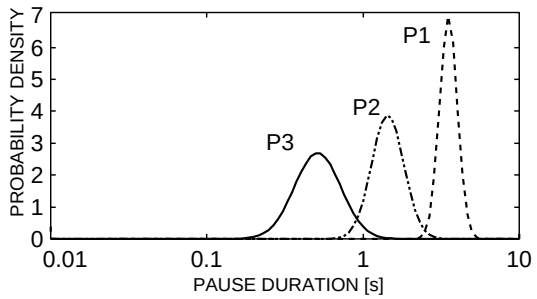

Figure 1: Distributions of pause durations in reading a short story, approximated by Gaussian probability density functions on the logarithmic scale of duration.

P_1: Pauses between paragraphs
P_2: Pauses between sentences
P_3: Pauses within a sentence

mean pause duration is largest for P1, followed by P2, and smallest for P3. The standard deviation (on the logarithmic scale), on the other hand, is largest for P3, followed by P2, and smallest for P1.

Preliminary analysis of pause durations and their locations indicated that they are relatively closely correlated in P3, but not in P2 or P1, leading us to conduct a further investigation to be described in the following sections.

3. ANALYSIS OF PAUSES OCCURRING IN A SENTENCE

3.1. The Speech Material

For the purpose of investigating various factors affecting pause insertion in text reading, we adopted the text of radio news, where each piece of news constitutes a single paragraph containing several sentences, and the sentences are rather long. The speaker was the same as for the short story described in the preceding section, and the text was read eight times at an average speech rate of 8.3 *morae*/s. Since each piece of news constitutes a single paragraph, measurement of P1 was considered to be irrelevant. Figure 2 shows the statistical distributions of P2 and P3, approximated by Gaussian probability density

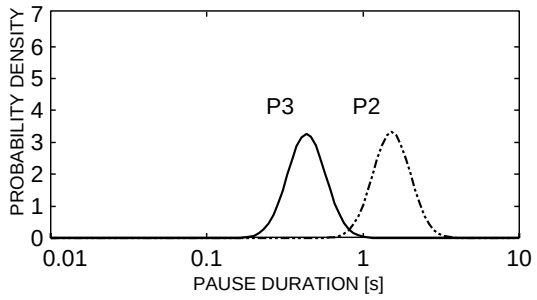

Figure 2: Distributions of pause durations in reading radio news, where each piece of news constitute a single paragraph.

functions on the logarithmic scale of pause duration, in the same way as in Fig. 1. In the following, we shall concentrate on the characteristics of P3 (i.e., pauses within a sentence).

3.2. Classification of Phrases That Precede Pauses of Category P3

Here we classify the syntactic roles of the phrases that precede a P3 into the following four broad categories.

A: Verb phrases and conjunctive adverbs
B: Subject phrases
C: Adverb phrases and object phrases
D: Adjective phrases

The corresponding pause categories are to be denoted by P_A, P_B, P_C, and P_D respectively.

Figure 3 indicates the means and the standard deviations of the durations of these four categories, and Table 1 shows the results of a *t*-test on the significance of differences in the means of these pause categories. The relatively large standard deviations for these four categories suggest the need for further categorization of these phrases on the basis of their syntactic and semantic functions. The rate of occurrence, calculated over all possible syntactic boundaries of the four types are: 89%, 44%, 19%, and 9%, respectively.

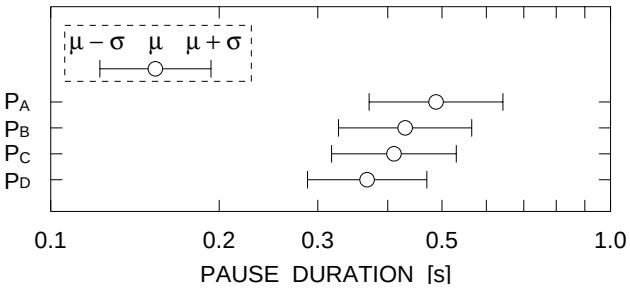

Figure 3: Mean (μ) and standard deviation (σ) of duration of pauses within a sentence, classified into four categories (P_A, P_B, P_C, and P_D) according to the syntactic role of the immediately preceding phrase.

Table 1: Results of a *t*-test on significance of differences in mean duration of the four pause categories (P_A, P_B, P_C, and P_D).

	P_B	P_C	P_D
P_A	+++	+++	+++
P_B		−	++
P_C			−

+++ at 1 % level + at 10 % level
++ at 5 % level − not significant

Table 2 shows a further categorization of the above-mentioned four phrase types, and Fig. 4 shows the corresponding means and standard deviations of the pauses (on the logarithmic scale) that follow these phrase types. The results of a *t*-test on significance of differences in their means are shown in Table 3, suggesting that (P_{A1}, P_{A2}), (P_{C2}, P_{C3}, P_{C4}), (P_{C5}, P_{C6}), and (P_{D1}, P_{D2}) can each be treated as a single category as far as pause durations are concerned.

Table 2: Further classification of phrases that precede pauses within a sentence.

A1	Verb phrases containing quotations
A2	Verb phrases constituting a part of a compound sentence
A3	Other verb phrases
A4	Conjunctive adverbs
B1	Subject phrases preceding quotations
B2	Other subject phrases
C1	Phrases containing quotations only
C2	Adverb phrases ending with nouns
C3	Adverb phrases expressing reasons and purposes
C4	Adverb phrases preceding subjects and predicates
C5	Other adverb phrases
C6	Object phrases
D1	The first one of two appositive phrases
D2	Phrases ending with a conjunctive
D3	Adjective phrases

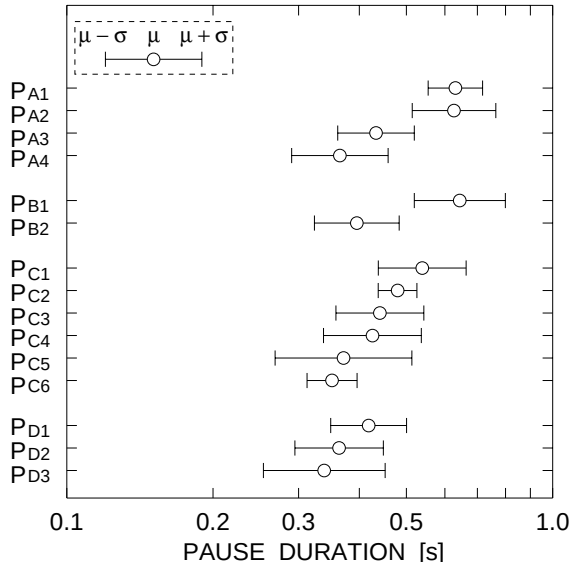

Figure 4: Mean (μ) and standard deviation (σ) of duration of pauses within a sentence, classified into finer categories according to Table 2.

Table 3: Results of a *t*-test on significance of differences in mean durations of the pauses classified into finer categories according to Table 2.

	P_{A2}	P_{A3}	P_{A4}
P_{A1}	−	+++	+++
P_{A2}		+++	+++
P_{A3}			+++

	P_{B2}
P_{B1}	+++

	P_{D2}	P_{D3}
P_{D1}	−	++
P_{D2}		++

	P_{C2}	P_{C3}	P_{C4}	P_{C5}	P_{C6}
P_{C1}	−	+	−	++	+++
P_{C2}		−	−	++	+++
P_{C3}			−	+	+++
P_{C4}				++	+++
P_{C5}					−

+++ at 1 % level + at 10 % level
++ at 5 % level − not significant

4. RATE OF OCCURRENCE OF PAUSES AND THE LENGTH OF PRECEDING PHRASES

From the viewpoint of application to text-to-speech synthesis, it is necessary to know, not only how long a pause should be when it is inserted, but also where it should be inserted. The following analysis was conducted in order to obtain basic data to answer the latter question.

In the first place, the news text was parsed into the four phrase types. The number of actual occurrence of inter-sentential pauses at the end of each of these phrases was then counted. Since our preliminary observations showed that the number is generally correlated with the length of the phrase, the counts were pooled for each phrase length expressed in terms of the number of *morae*. Because the number of phrases of a given length was rather small, the results were grouped for phrase lengths quantized in 5-*mora* steps, and the rate of occurrence of pauses were calculated for each of the groups.

The results are shown in Fig. 5, where the four panels correspond to the four phrase types. As mentioned earlier, the overall occurrence rate is highest for phrase type A, followed by types B, C, and D. In panel A, the very high rate of occurrence for phrase lengths below 5 *morae* corresponds to conjunctive adverbs (A4), which, in spite of their short lengths, are almost always followed by a pause in spoken Japanese. Examples of these conjunctive adverbs are: '*mata*' (also/furthermore), '*korede*' (thus), etc. The occurrence rate increases quite rapidly with phrase length,

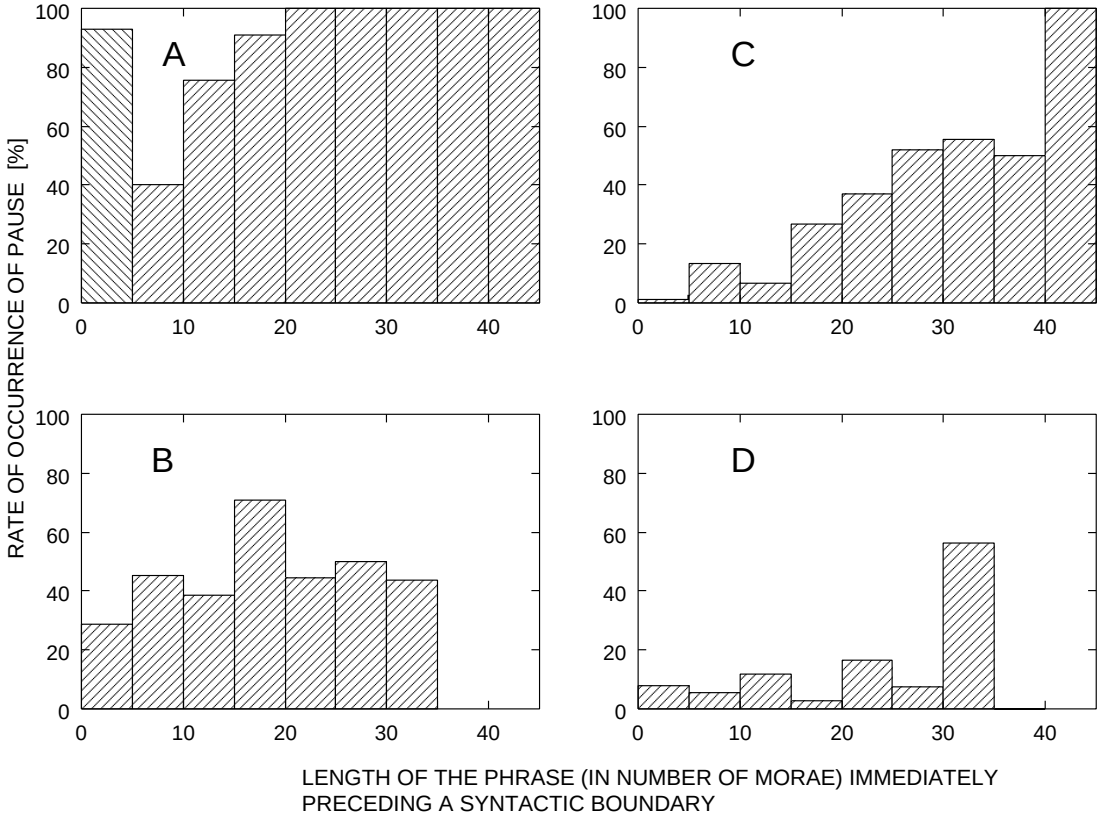

Figure 5: Rate of occurrence of pause at a syntactic boundary versus length (in number of *morae*) of the immediately preceding phrase. The four panels correspond to the following phrase types.

A: Verb phrases and conjunctive adverbs
B: Subject phrases
C: Adverb phrases and object phrases
D: Adjective phrases

and reaches 100% at about 20 *morae*. Type C phrases also show a clear dependency of the occurrence rate upon phrase length, though the rate of increase is smaller than for type A. On the other hand, the dependency is much smaller in types B and D, but there is a sudden increase in the occurrence rate in type D at phrase lengths beyond 30 *morae*.

5. SUMMARY

This paper has described a part of the preliminary results of our on-going research on the analysis and modeling of the human processes involved in the production and perception of temporal characteristics of speech, and has focused on the process of pause insertion in text reading. Although much is left to be investigated, the results shown here clearly indicate the effects of some of the important factors in the process of pause insertion by a human speaker in text reading. Work is under way toward elucidating the influences of such factors as the overall speech rate, speaking style, text type and speaker, deriving rules for pause insertion and pause length assignment in text-to-speech synthesis, as well as evaluating their relevance in the quality of synthesized speech.

6. REFERENCES

1. Fujisaki, H. and Omura, T. "Characteristics of durations of pauses and speech segments in connected speech," *Annual Report, Engineering Research Institute, Faculty of Engineering, University of Tokyo*, vol. 30, pp. 69–74, 1971.
2. Fujisaki, H., Ohno, S. and Yamada, S. "Analysis of the global timing in text reading," *Report of Fall Meeting, Acoustical Society of Japan*, vol. 1, pp. 303–304, 1996.
3. Fujisaki, H., Ohno, S. and Yamada, S. "Timing of pauses in news text reading — Factors affecting interphrasal pause durations," *Report of Spring Meeting, Acoustical Society of Japan*, vol. 1, pp. 203–204, 1998.