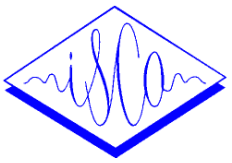


6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

PERCEPTION OF IDENTICAL VOWEL SEQUENCES IN JAPANESE CONVERSATIONAL SPEECH

Yuki Hirose†

and

Kazuhiko Kakehi+

†The University of Electro Communications

+ Nagoya University/ CIAIR

ABSTRACT

Sequences of more than two identical vowels across word boundaries can occasionally be found in Japanese speech. Our previous study (Kakehi and Hirose, 1998) investigated how hearers perceive such vowel sequences. We used isolated sentences as test materials, provided that the duration of such vowel sequences increases in proportion to the number of morae. We found that hearers can detect the number of morae in the vowel sequence by using cues such as the pitch pattern, speech rhythm, and duration. The present study attempts to examine cases in which the vowel sequence occurs as part of natural conversation (part of planned dialogs). The materials were designed so that the target vowel sequences do not have any distinctive pitch movement. In the recorded sets of conversation that we examined, the relationship between the duration and the number of morae in the vowel sequence was roughly proportional. A series of perception tests investigated whether hearers can correctly detect the number of morae in such utterances, while the pitch movement in the target was kept stable. The results indicated that the duration of vowel sequences does not serve as a sufficient cue to detect the number of morae.

1. INTRODUCTION

The Japanese language has a distinctive phonological contrast between short and long vowels. Short vowels take one moraic unit while long vowels take two, and they are in contrastive distribution, as shown by the minimal pair "obasan" (aunt) vs. "obaasan" (grandmother). This makes Japanese listeners sensitive to vowel length, as reported in Dupoux, Kakehi, Hirose, Pallier and Mehler (1999): Japanese listeners could distinguish minimal pairs which contrasted in the number of [u] vowels that were contained in nonwords, and could do so significantly more successfully than native speakers of French (in which vowel length is not contrastive). Sequences of identical vowels usually lack segmental cues which native listeners can normally rely on, such as changes in the spectrum or in the power envelope. In such cases, what other acoustic cues to native speakers use to distinguish each mora?

Fujisaki et al. (1997) showed that the duration of sequences of identical vowels increases in proportion to the number of morae when short sentences or phrases which contained the target vowel sequence were pronounced in isolation. Their study suggests that vowel duration is a very reliable cue in mora segmentation of a speech signal including identical vowel sequences. However, duration might not be sufficient in identifying each vowel segment in sequences of more than two identical vowels as part of a longer conversation, since speech rate and other suprasegmental characteristics are expected to vary in these contexts.

Kakehi and Hirose (1998) conducted a study using sequences of 2 to 6 identical vowels occuring across word boundaries. They reported that the duration of a vowel sequence is roughly proportional to the number of morae in that sequence in a sentential context, but that hearers are also helped by suprasegmental information such as the pitch pattern and speech rhythm. In natural conversation, speech rate tends to be faster and more variable. The present study investigates the relationship between the duration and the number of mora in vowel sequences that occur in an even larger unit, i.e., as part of a planned dialogue. Two perception tests were also conducted to examine how well hearers can segment the vowel sequence and detect the number of mora in conversation.

2. EXPERIMENT 1: The relationship between the duration and the number of morae

In Experiment 1, we employed materials (identical vowel sequences) shown in (1a-f) embedded in a phrase as part of short dialogues shown in (2) (not single sentences in isolation). Speakers read the dialogues as if in a natural conversation.

1.

a. Shimaneken no Matsue e jisho o okutta
("I sent a dictionary to Matsue, Shimane.")

b. Shimaneken no Matsue e ejiten o okutta
("I sent an illustrated encyclopedia to Matsue, Shimane.")

c. Shimaneken no Matsue e eeji bakari no shinbun o okutta
("I sent a newspaper written entirely in English to Matsue, Shimane.")

d. Shimaneken no Matsuee e eeji bakari no shinbun o okutta

("I sent a newspaper written entirely in English to Matsuee, Shimane.")

e. Shimaneken no Matsue e eeeejiten o okutta
("I sent an English-English dictionary to Matsue, Shimane.")

f. Shimaneken no Matsuee e eeeejiten o okutta
("I sent an English-English dictionary to Matsuee, Shimane.")

2. The conversation text (translated into English here except for the critical part) read between the experimenter and the subject.

Experimenter: "Isn't it irritating that when you send something to Japan from some country and it arrives with its package torn open or domaged?"

Subject: "I know, that is really shocking. Sometimes it's even worse. What's inside can be lost."

Experimenter: "That's so unfair, because we are always paying the postage."

Subject:

Konoaida mo ne, gaikoku no yuubinkyoku kara okusan ga hataraiteiru, shimaneken no

{matsue e jisho
/ matsue e ejiten
/ matsue e eeji bakari no shinbun
/ matsuee e eeji bakari no shinbun
/ matsue e eeeejiten
/ matsuee e eeeejiten}

wo okutta koto ga arundesukedo, tsuita toki niwa nakami wa boroboro ni natteita rashiindesu. (English translation below)

(English translation: "Just the other day, a friend of mine sent from abroad

{an dictionary to Matsue.Shimane
/ an illustrated encyclopedia to Matsue, Shimane
/ a newspaper written entirely in English to Matsue, Shimane
/ a newspaper written entirely in English to Matsue, Shimane
/ an English-English dictionary to Matsue, Shimane
/ an English-English dictionary to Matsuee, Shimane}

where his wife lives but it was all damaged by the time it arrived.")

Experimenter: "You mean, even the inside? "

The stimuli design re-examines whether the proportionality between the number of morae and the duration of vowel sequence holds in natural conversation, after the target phrases are normalized for speech rate.

Stimuli Six dialogues were prepared which were all identical except for the critical phrase (see above). The critical phrase contained the target vowel sequence, which varied the number of the [e] vowels (number of morae) from 2 to 7. In order to control the pitch information, the stimulus sentences were created so that none of the lexical items in the target sequence carried a lexical accent and thus exhibited no rise or fall in pitch when read in Kansai Japanese.

Subjects The dialogues were read by two adult female speakers of Kansai Japanese, the experimenter and the informant. The informant was assigned the part that contained the target sequence.

Procedure The set of dialogues was repeated 5 times and recorded onto a DAT tape by using SONY TCD8 portable DAT recorder. Before the recording, the informant was instructed that she could take a breath only at points marked by "," in the orthography, in order to avoid unwanted prosodic breaks within the target vowel sequences.

Data Treatment Four of the recorded sets (excluding the first one) were used in the analysis. The utterances were digitized and analyzed using the SP4-Win software. The duration of the [e] target vowel portion was measured. The speech rate in each utterance was normalized based on the 7-morae portion (/shimaneken no ma/) which preceded the target within the same intonational phrase. The duration of this portion of each utterance was measured and divided by the number of mora contained within it (i.e. 7 morae); this was taken to be the average duration per mora in the utterance (defined as T0), and was used as the measure of duration of the target vowel sequence.

Results and Discussion

Figure 1. The duration of the target vowel sequence (measured in T0)

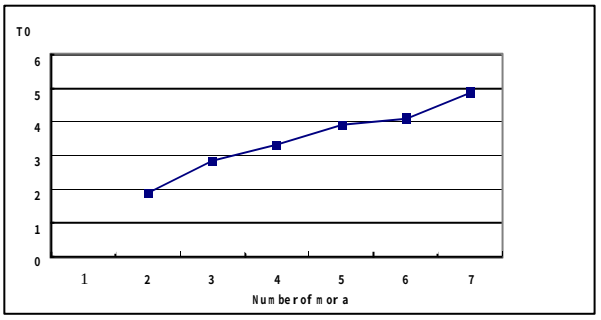


Figure 1 shows that the duration and number of morae in the vowel sequence is largely proportional across target sentences. No substantial dip in power envelope within the target vowel sequence was observed, presumably due to the instruction given to the subject to avoid prosodic breaks within the critical portion of the sentence.

Based on the results of Experiment 1, Experiments 2 and 3 were designed to examine whether durational information is sufficient in mora identification.

3. EXPERIMENT 2

The subjects listened to the target vowel sequence, which was preceded and followed by a few morae cut out from the original utterance collected in Experiment 1. They were asked to detect the number of [e] vowels contained in the vowel sequence.

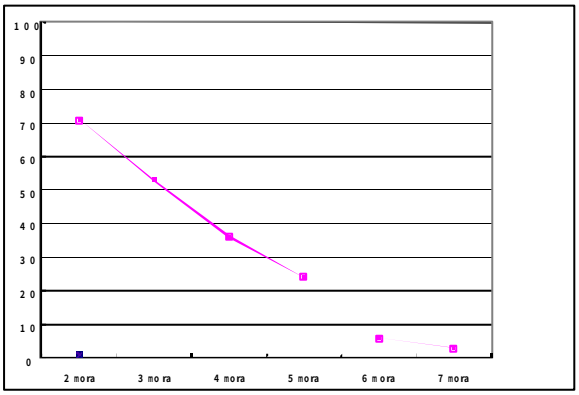
Stimuli All sentences containing the target vowel stimuli recorded in Experiment 1 were normalized for speech rate (see above). The target vowel [e] regions (the target [e] sequence flanked by three morae preceding and one mora following: /ma tsu e ... ji/) were then cut out from each of the 4 repetitions of the 6 lengths collected in Experiment 1. Each token was presented twice, resulting in 48 tokens altogether.

Subjects Twelve undergraduate students recruited from The University of Electro-communication participated in the experiment.

Procedure The experimental stimuli were auditorily presented to subjects in a randomized order. The subjects were asked to detect the number of target vowels in each stimulus by marking their choice (2, 3, 4, 5, 6 or 7) on an answer sheet.

Results and Discussion

Figure 2. Error rates in detection of number of morae.



As the error rates in Figure 2 illustrate, the responses were highly erroneous overall, and degraded as the number of morae exceeded 4. The results indicate that durational information alone is not sufficient for the hearer to correctly detect the number of the morae in spoken sentences, even though the correlation between duration and the number of mora alone can potentially signal the number of mora in the vowel sequence.

It is argued in Kakehi and Hirose (1999) that durational information is used in combination with other cues such as pitch information contained in the target vowel sequence, and some top-down semantic information which would have been provided if the hearer could access to the words preceding the vowel sequence in the sentence. Since pitch information was purposely suppressed in the current experimental stimuli design, we focus on the latter factor.

4. EXPERIMENT 3

The subjects listened to the same stimuli from Experiment 2, but this time they were shown the original clause in which the target vowel sequences were embedded.

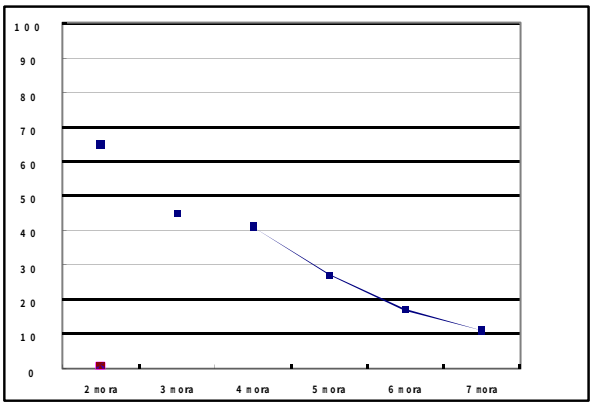
Stimuli The same stimuli as in Experiment 2 were used.

Subjects Twelve undergraduate students at The University of Electro communication served as subjects. None of them had participated in the previous experiments.

Procedure The procedure was the same as in Experiment 2, except that the entire set of original sentences (see 1a-f) were presented to the subjects both in written and spoken format. The subjects were told that the all of the stimuli they would hear were part of one of these sentences.

Results and Discussion

Figure 3. Error rates in detection of number of morae.



As can be seen in Figure 3, the error rate did not show any substantial improvement: an analysis of variance (ANOVA) found no reliable difference between the patterns of error rates in the two experimental results.

The fact that the hearers knew the original sentences containing the target vowel sequence does not appear to help them to improve their performance in detecting the number of mora in the sequence. A follow-up experiment may be considered to examine whether their performance improves when the preceding phrases are presented not only in a written format but also as part of the presented auditory stimuli.

5. DISCUSSION

The results of the three experiments suggest that the duration of vowel sequences does not serve as a sufficient cue for the hearers to detect the number of morae in a sequence of segmentally identical vowels, even though one could use such information as suggested by the results of Experiment 1. This poor performance of hearers in detecting the number of morae might be explained by the notion of a 'difference limen' (DL) (or

difference or change in the stimulus that is at the threshold of detectability) for the duration of a filled sound, which has been studied by many researchers in psychoacoustics.

In the utterances collected in Experiment 1, the duration of the identical [e] vowel sequence varies from 0.14 sec to 0.71 sec, depending on the number of morae in the sequence (2-7). Abel (1972) reports that the Weber ratio (or the change in magnitude of a stimulus required to produce a change in sensation 50% of the time / the magnitude of a stimulus) for duration of filled sound is almost constant in non-speech sound at 0.1 where the sound duration is from 0.1 to 1 sec. Although Abel's data from non-speech sound is not directly comparable to our data, it is suggestive in explaining the subjects' performance in Experiments 2 and 3. The ratio of the duration of a certain number of morae to the durational difference between vowel sequences with successive number of morae is far greater than the Weber ratio when the number of morae is as low as 2 or 3, while the ratio is almost twice for the 7 morae cases. Therefore, this suggests that the discrimination of the number of morae may be easier when the number of mora is as small as 2 or 3, but should become difficult as the number of morae increases. Although the perception of duration in non-speech sound is different from that of the speech sound embedded in continuous speech, the experimental results may be explained with recourse to this Weber ratio.

Based on the findings from the present experiments, further studies are currently ongoing to investigate how and how much the pitch movement and semantic information can improve hearers' ability to detect the number of mora in conversational contexts.

REFERENCES

Abel, S. M. "Duration discrimination of noise and tone bursts," *Journal of Acoustical Society of America*, 51, 1219-1223, 1972.

Dupoux, E., K. Kakehi, Y. Hirose, C. Pallier, and J. Mehler, "Epenthetic Vowels in Japanese: a Perceptual Illusion?" *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 25, No. 6, 1568-1578. (1999)

Fujisaki, H., Nakamura, K. and Imoto, T. "Auditory perception of duration of speech," paper presented at Symposium on Auditory Analysis and Perception of Speech, Leningrad, 1973.

Fujisaki, H., Ohno, S., Tomita, O., and Yagi, T., "The influence of moraic phonemes upon segmental and prosodic features of Japanese (2) - sequence of segmentally identical vowels-" . "An Analog Electronic Cochlea," *IEEE Trans. ASSP* 36: 1119–1134, 1988.

Kakehi, K. and Y. Hirose "Suprasegmental Cues for the Segmentation of Identical Vowel Sequences in Japanese." *Proceedings of the 1998 International Conference on Spoken Language Processing*, Vol 6, 2283-2286. (1998)