Suprasegmental Cues for the Segmentation of Identical Vowel Sequences in Japanese

Kazuhiko Kakehi[+]

and

Yuki Hirose[†]

[+]Nagoya University
[†]The Graduate School and University Center, The City University of New´York

ABSTRACT

This paper investigates how hearers cope with a sequence of more than two identical vowels --a common occurrence in Japanese speech. In the segmentation of identical vowels there are no spectral cues and very small power envelope change in usual utterances containing identical vowels. We consider the effects of suprasegmental information such as duration, pitch pattern and rhythm of speech as important cues, and examine how, and to what extent a hearer can successfully make use of such information to segment each mora in a consecutive vowel series with and without the preceding sentential context.

1. INTRODUCTION

The Japanese language is characterized by the phonological contrast of short and long vowels. Short vowels take one moraic unit while long vowels take two, and they can make difference in meaning within a word (ex., "obasan" (aunt) vs. "obaasan" (grandmother)). Such distinctions are known to be very hard for non-Japanese speakers. In a cross-linguistic study (Kakehi et.al, 1996) it is shown that Japanese speakers distinguish between minimal pairs contrasting the number of [u] vowels contained in tokens, significantly more successful than native speakers of French, in which vowel length is not contrastive. The above study was targeted to test the ability to distinguish between one and two vowels. However, in Japanese, hearers can often encounter a sequence of more than two identical vowels across word or phrase boundaries.

One might predict perception of each vowel segment would be very difficult (not only for the automatic speech recognition machine but also for humans) because of the lack of spectral cues and prominence changes in power envelope. In such cases, how do native speakers of the language distinguish each mora?

Fujisaki et.al (1997) showed that time duration of sequences of identical vowels increases in proportion to the number of mora. Their study suggests that vowel duration is a very reliable cue in mora segmentation of the speech signal including identical vowel sequences. However, time duration might not be sufficient for identifying each vowel segment in sequences of more than two identical vowels. We consider suprasegmental information such as pitch pattern and the rhythm of speech, as suggested by Fujisaki et.al (1973) as necessary cues for successful identification.

2. EXPERIMENT 1: PERCEPTION OF IDENTICAL VOWEL SEQUENCES IN NONWORDS

In Experiment 1, the subjects were asked to detect the number of mora in the vowel sequence under the condition that (i) no pitch contour accompanied the stimuli , and (ii) pitch contour was present in the stimuli.

Subject Eleven college students participated in the experiment. All of them were native speakers of Japanese who have not lived in any foreign country.

Stimuli Two sets of stimuli consisting of 18 kinds of non-words, varying in the number of vowels from one to four, was used in the experiment (hence 72 per set). The stimuli were uttered by a female former radio broadcaster. These non-words were created by changing the vowel segment of existing words or phrases (see below).

Example: non-word stimuli set

1. **u**nchoo
2. **uu**nchoo
3. **uuu**nchoo
4. **uuuu**nchoo

Original: existing words/ phrases

1'. **i**nchoo (director/ president)
2'. **ii**nchoo (chairperson)
3'. **iii**nchoo (good director/ president)
4'. **iiii**nchoo (good chairperson)

In one set of stimuli, all the non-words were deliberately read without a pitch contour. In the other set, each item was read with an intonation contour adopted from the tonal pattern of the corresponding word in the following procedure. First, the speaker read the list of existing words (such as 1' ~ 4') with appropriate pitch patterns. Then she was asked to read the list of non-words (such as 1 ~ 4), in which only the initial vowel

segment was changed, maintaining the pitch patterns of the original words.

Procedure The subjects listened to the randomized list of stimuli through the headsets. Pace of presentation of each stimulus was controlled by subjects; the words were presented each time the subject clicked the mouse to proceed to the next stimulus. The subjects were asked to detect the number of the target vowels (the first part of each word) in each stimulus and choose the answer from the choice of 1,2,3, or 4 on the answer sheet.

Results and Discussion

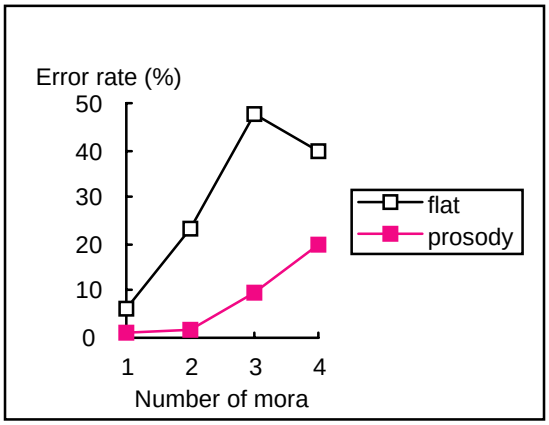

Figure 1. Error rates in detection of number of mora when the pitch contour is present ("prosody") and absent ("flat"), for actual number of mora in the stimuli (1 to 4)

As the error rates in Figure 1 illustrate, the performance was highly accurate when the number of mora was one and two, and when the stimuli were accompanied by pitch contours. When the number of mora in the vowel sequence in the stimuli exceeds two, the error rate started to increase and reached 20% for four-mora vowel sequences even when a pitch contour was present. In contrast, when there is no pitch movement, the accuracy of the performance was close to the chance level at two-mora vowel sequences, and increased to 47% for three-mora sequences. (In four-mora sequences, the error rate appears to be lower than in three-mora sequences. It could be because the choice given in the answersheet was only up to 4.) The difference between the two conditions: with pitch, and without pitch was highly significant for 2-mora and 3-mora sequences (2-mora: $F1(1,20) = 33.74$, $p < .001$, $F2(1,34) = 10.46$, $p < .005$, 3-mora: $F1(1,20) = 36.81$, $p < .001$, $F2(1,34) = 80.42$, $p < .001$), and highly significant but only by item for 1-mora and 4-mora sequences (1-mora: $F2(1,34) = 12.14$, $p < .005$, 4-mora: $F2(1,34) = 15.94$, $p < .001$) . The data suggest that lack of pitch movement results in significantly greater error rates. Durational information alone does not seem to be used effectively.

3. EXPERIMENT 2: PERCEPTION OF IDENTICAL VOWEL SEQUENCES ACROSS WORD BOUNDARIES

Experiment 1 has shown that pitch information plays an important role in recognition of multiple vowel sequences in an artificial word-context. Experiment 2 examined whether the above results from Experiment 1 hold for vowel sequences across word boundaries in spoken sentences. We also examined whether the correlation between the duration and the number of the mora in the vowel sequence holds in such cases.

Subject Ten college students participated in the experiment. Again, all of them were native speakers of Japanese who have not lived in any foreign country.

Stimuli We created a set of sentences in which all the sentences began with identical phrases prior to the target vowel sequence. The length of the vowel sequence was varied in number of mora 2,3,4 and 6. The stimulus sentence in which the vowel sequence consists of five mora was absent because such a sequence in the given environment could not be related to a meaningful word combination.

5. Shimaneken no Matsu**e e** jisho o okutta

("I sent a dictionary to Matsue, Shimane.")

6. Shimaneken no Matsu**e e e**jiten o okutta

("I sent an illustrated encyclopedia to Matsue, Shimane.")

7. Shimaneken no Matsu**e e ee**ji bakari no shinbun o okutta

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

8. Shimaneken no Matsu**e e eeee**jiten o okutta

("I sent an English-English dictionary to Matsue, Shimane.")

In order to vary the presence of pitch information, two types of stimuli sets were designed: In one set, all the stimulus items were read in one dialect of Japanese (Osaka), in which (for these particular materials) there is no rise or fall in the vowel sequence, and the other set, the items were read in another dialect of Japanese (Tokyo) and all the items which had a vowel sequence of more than two contained a rise or fall pitch contour (as shown below). A bi-dialectal male graduate student was chosen to utter the stimuli. The target vowel sequence preceded and followed by one extra mora was extracted from the original utterance ("tsu e... ji") . The list below indicates the pitch pattern of the vowel sequence part taken from the above sentences 5 ~ 8, in each dialect.

	Tokyo Japanese	Osaka Japanese
9.	**ee** (HH)	**ee** (LL)
10.	**eee** (HHL)	**eee** (LLL)
11.	**eeee** (HHLH)	**eeee** (LLLL)
12.	**eeeee** (HHLHHH)	**eeeee** (LLLLLL)

For each stimulus, the duration of the vowel sequence was measured to see whether (i) it correlates with the number of mora, and (ii) the correlation rate is different in the two conditions.

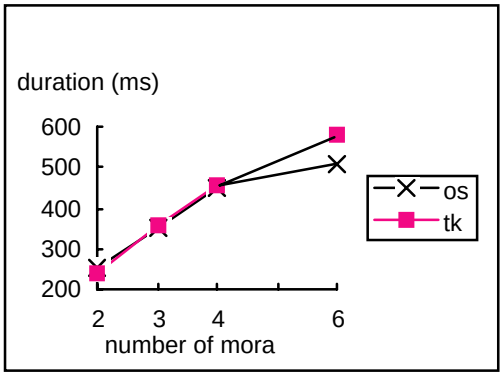

Figure 2: The duration of the vowel sequence in each stimulus (ms) as a function of number of mora, in the Osaka ("os") and the Tokyo ("tk") dialect conditions

As Figure 2 shows, the duration and the number of mora in the vowel sequence seem to be proportional at least from two to four mora. Duration increased by about 115 ms. as the number of mora increases by one, in both dialects. Fujisaki et.al (1997)'s results thus seem to hold for our stimuli sets, but in our results, duration per mora is shorter than what Fujisaki et.al reported 160 ms. However, in 6-mora sequence in our stimuli, duration per mora decreased to about 97 ms. in the Tokyo dialect condition (with pitch movement) and as short as 84 ms in the Osaka dialect condition (without pitch movement). The reason for greater decrease in the 6-mora sequence in the Osaka dialect condition may be due to the lack of pitch movement. There were virtually no power envelope changes in the vowel sequences that would indicate the mora boundaries.

Procedure: The procedure was the same as Experiment 1, but the number of vowels in the sequence varied from 1 to 6. There were two experimental sessions, one for the Tokyo dialect set and the other for the Osaka dialect set. The order of presentation of the two sets was balanced: half of the subjects heard the Tokyo dialect sets first, whereas the other half heard the Osaka dialect set first.

Results and Discussion:

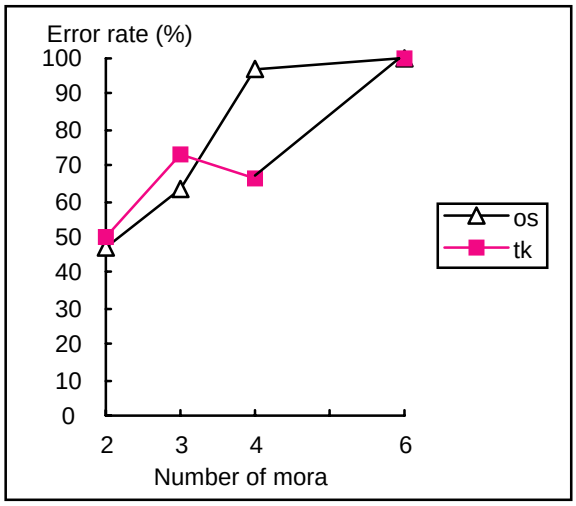

Figure 3. Error rates in detection of number of mora in the Osaka ("os") and the Tokyo ("tk") dialect conditions, for actual number of mora in the stimuli (2,3 4 and 6)

Overall, a huge error rate was observed with and without the pitch information. In every condition, the number of mora was underestimated by one or two mora. It is again implied that time duration is not a sufficient cue to detect the number of mora in natural speech, given the measured correlation between the duration and the number of mora in up to a four-mora sequence. There is no reliable difference between the two dialects for two-mora and three-mora sequences, however, a considerable difference was observed between the two dialects for the four-mora sequence ($F(1,18= 5.65$, $p < .05$). Considering that the time duration for these stimuli were almost the same in both conditions, the difference can be interpreted to reflect the presence of pitch movement: there is a pitch fall on the third vowel segment in the Tokyo dialect condition whereas there is no such pitch movement in the Osaka dialect condition. For the six-mora sequence, one might predict an advantage of the Tokyo-dialect condition because of the pitch movement on the third vowel segment, and also there was less reduction of time duration per mora compared to the Osaka-dialect condition. Nonetheless, the error rate reached 100% in both conditions.

In the previous experiment we argued for the importance of pitch information in segmenting each mora in the identical vowel sequence, but it turns out that pitch movement is not strong enough to enable the hearer to detect accurately the number of mora when the number of vowels reaches six, although the high error rate may be due to the reduced duration of each of the mora in the six-mora sequence.

4. EXPERIMENT 3: PERCEPTION OF IDENTICAL VOWEL SEQUENCES WITHIN SENTENTIAL CONTEXTS

In Experiment 2 , the stimuli were the vowel sequence preceded and followed by an additional mora, spliced out from the utterance as a whole sentence. High error rate may be due to the lack of speech rate of the entire original utterance, or the lack of

some top-down semantic information which would have been provided if the hearer could access the words preceding the vowel sequence in the sentence. Speech rate (rhythm) varies over individual speakers and situation or environments, thus it may influence speech perception. In Experiment 3, the stimuli used in Experiment 2 were modified and subjected to the same experimental procedure. Each stimulus contained the entire first part of the utterance prior to the target vowel sequence, so that listeners infer the approximate duration per mora from the preceding phrases.

Subject: Ten college students from the Osaka area who were familiar with both Osaka and Tokyo dialects participated in the experiment. We assumed that they were able to understand the experimental sentences equally well in both dialects, since students in Osaka frequently hear the Tokyo dialect (roughly equivalent the standard dialect) through the mass media.

Stimuli: The same speech set as Experiment 2 was used, but the entire part of the utterance prior to the vowel sequence ("Shimaneken no Matsue...") was included in each token.

Procedure: The same as Experiment 2, except that the newly added front part of the stimuli was included in each choice item in the answer sheet.

Results and Discussion: The error rate for the six-mora sequence in Tokyo dialect condition showed a drastic change compared to the results of Experiment 2, as shown below.

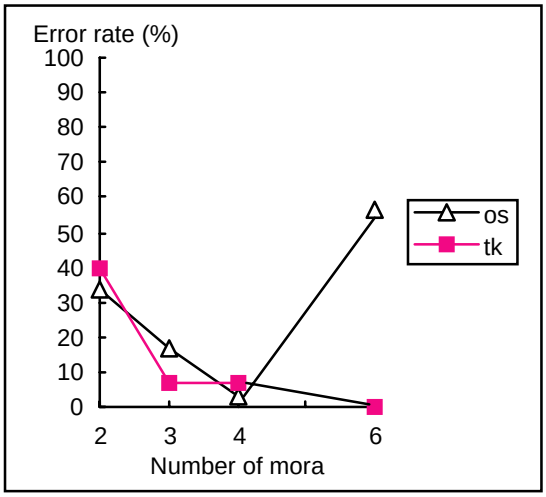

Figure 4. Error rates in detection of number of mora in Osaka ("os") and Tokyo ("tk") dialect conditions, for actual number of mora in the stimuli (2,3 4 and 6)

The error rate dropped from 100% to 0% while that of the Osaka dialect condition was 57%. The difference between the dialects for the six-mora condition was highly significant ($F(1,18) = 32.11$, $p < 001$). The fact that the duration per mora is more drastically reduced in the Osaka dialect may be responsible for the difference, but also the presence of pitch movement in the Tokyo dialect may have made the difference. However, considering that the pitch information was not useful in Experiment 2, it is suggested that when the vowel sequence is as long as six-mora, a hearer has to rely on the rhythm information obtained in reference to the speech rate of the previous part of the utterance.

For four-mora sequences, the error rates in both dialect conditions seem negligible while in Experiment 2, the subjects produced huge error rates (almost 100% error rate in the Osaka dialect condition). The contrast between Experiment 2 and 3 for four-mora sequences implies that rhythm information is the factor that plays a significant role in the perception of a sequence of four identical vowels.

Overall, it appears that the effect of rhythm information is larger when the sequence of vowels is longer: for two-mora sequence, there seems to be very little improvement in error rates in Experiment 3 compared to Experiment 2, in spite of the drastic differences for three, four and six-mora sequences.

5. CONCLUSION

The results of the three experiments suggest that duration of vowel sequences is not sufficient for detection of number of mora in sequences of segmentally identical vowels. We argue that pitch information is an important cue for speech segmentation in Japanese, but in natural speech, hearers need this information in combination with other cues such as the speech rhythm and some sentential information, especially when the number of vowels in the sequence is large.

REFERENCES

1. Fujisaki, H., Nakamura, N. and Imoto, T., “Auditory perception of duration of speech and non speech stimuli,” Ann. Bull. Research Institute of Logopedics, University of Tokyo, 7, 45-64 (1973)

2. 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 -,” presented at the 1997 spring meeting of the Acoustical Society of Japan (1997)

3. Kakehi, K., Hirose, Y., Dupoux, E. and Mehler, J., “The effect of the phonological system in a language on speech perception for speaker variabilities,” Proceedings of ASA and ASJ Third Joint Meeting, 1pSC14, 839-842 (1996)