The Acoustic Characteristics of Japanese Identical Vowel Sequences in Connected Speech

Tsutomu Sato
English Department, Meiji Gakuin University, Tokyo, Japan
John A. Maidment
Department of Phonetics and Linguistics, University College London, UK

ABSTRACT
It has been pointed out that a word which contains an identical vowel sequence may not be distinguished, especially in faster speech, from a word which has a long vowel and phonologically the same mora-count. However, this indication has been made without any objective evidence and it is argued that a quantitative study is needed to investigate the acoustic characteristics of identical vowel sequence in comparison with those of a long vowel. This paper reports the results of production experiments for this purpose done by 10 native speakers of Japanese. The results will be discussed from the viewpoints of glottal stop insertion between an identical vowel sequence and durational ratio comparison. More specifically, the actual number of glottal stop insertions on citation and sentence levels will be clearly shown and the acoustic features of the glottal stop such as its duration will be investigated. In addition, the ratios of the whole token, consonant + vowel(s) and the vowel part of identical vowel sequence examples will be made clear and compared with those of long vowel examples in both citation and sentence reading forms.

1. INTRODUCTION
It has been considered that phonologically the following words which respectively contain an identical vowel sequence (IVS, hereafter) and a long vowel (Long V) are both four-mora [1]:

(1)a. {sato} 'foster home' + {oya} 'parent' --> sato oya  'foster-parent'  / sa to o ja/
b. {sato-} 'sugar'     +  { ya} 'shop' --> sato- ya   'sugar shop'   / sa to : ja/
(+... morpheme boundary,  + ...the second element of a long vowel,  : ... a length phoneme)

Phonetically, it has been pointed out that glottal stop insertion occurs between the identical vowels of (1a) in slow and careful speech [2], although the glottal stop disappears in faster speech and both words may not be distinguished [3][4]. However, this claim is based on impressionistic observation or introspection and lacks quantitative evidence. Kasuya and Sato [5] compare the acoustic characteristics of IVS and Long V and find that lower sound pressure and changes of fundamental frequency and frequency spectra structure are observed between the IVS in the example of ko 'small' + oni 'demon', kooni in careful speech. This feature of IVS, according to [5], is the acoustic reflex of the morpheme boundary, which does not exist in the case of Long V in the example of kōsan 'surrender'. They also mention that the acoustic difference between IVS and Long V can not be seen in many cases in conversational style and reading style and both phonetic forms can not be distinguished from each other, as [3] and [4] point out.

The observation of acoustic characteristics of IVS by [5] is of importance, but the actual data are not shown in their paper and how the changes occur is not certain. Besides, no remarks are made about specific speech rate differences in conversational or reading styles. Therefore this study attempts to clarify the acoustic characteristics of Japanese IVS in comparison with those of Long V in both citation form readings, and in sentence readings uttered at different speech rates.

2. EXPERIMENTAL PROCEDURE
2.1. Tokens and Test Sentences
The following sets of a 2 mora word with a short vowel, a 3 mora word with IVS and a 3 mora word with Long V were used as test tokens:

(2)

[i]

shin  [SiN]    'pencil lead'
shii  [Si i N]  'the death'
shii  [Si: N]  'scene'
kei  [ke ri]    'settlement'
keeri  [ke e ri] 'fur collar'
ke-ri ['ke: ri] 'accounting'
kato ['kato] 'transition'
kaato ['ka a to] 'mosquito bite'
ka-to ['ka: to] 'cart'
kou ['kou] 'crust'
koou ['ko o u] 'response'
ko-o ['ko: o] 'love & hatred'
suri ['suri] 'pickpocket'
suuri ['su u ri] 'vinegar seller'
su-ri ['su: ri] 'mathematical modeling'
suri to keiji ga deau
'A pickpocket and a detective meet.'
suuri to bin ga mieru
'A vinegar seller and a bottle can be seen.'
su-ri to shiki ga ukabu
'A mathematical modeling and a formula occur to me.'

All the test sentences are constructed to have 10-mora length and are declarative types to avoid intonational effects. The sentence structure consists of coordinated NPs and VP, i.e., six-mora NPs and one-mora nominative case 'ga' and a three-mora intransitive Verb.

2.2. Participants
10 Japanese college students, five males and five females, took part in the data collection for this study. Their ages vary from 20 to 27 and the average is 23.3 years old. Five participants were born and raised in Tokyo and the other five come from border areas with surrounding prefectures whose accent systems can be regarded as the same as the Tokyo accent system.

2.3. Recording
The recording was held in an anechoic room situated in the department of phonetics and linguistics, University College London. In the first part of the experiment, a participant was given a list of the 15 tokens which were randomised with 5 distracter words, and the participant was asked to read the tokens with a pause between them, at a comfortable rate of speech, with five repetitions. In the second part, the participant read 15 sentences listed in (3) in randomised order which were shown on a computer display for 4.0 seconds per sentence as slow rate, then shown for 2.8 seconds as a mid rate, and finally for 1.7 seconds as fast rate, with five repetitions. A Hypercard stack used for presenting sentences was written by Kevin Varden [6].

In this way, 750 tokens in citation forms (15 tokens x 5 repetitions x 10 participants) and 2250 tokens framed in sentences (15 tokens x 3 speech rates x 5 repetitions x 10 participants) were recorded.

2.4. Measurement
The duration of a whole token, the duration of the initial consonant plus vowel part, for example [Sii] in the word 'shiin', and duration of the vowel part only in a token such as [ii] in 'shiin', were measured from their waveforms and from wide-band spectrograms by using a speech analysis program called Signalyze.
3. RESULTS AND DISCUSSION

3.1. Glottal Stop Insertion

Six examples of glottal stop insertions were found in total, five examples out of 750 tokens in citation forms and only one example out of 2250 tokens in sentence readings. All examples of glottal stop insertions here occurred in the word *shiin*. Out of the five examples in citation forms, four were produced by a female participant JN and the following example is JN’s fifth token reading (T5):

![-waveform-and-spectrogram-of-glottal-stop-example-by-JN](image)

A clear separation of F1 and the weakening of upper formants at the morpheme boundary can be seen in the spectrogram. The durational values of glottal stop, token, consonant and vowels (CVV-), vowel part of the JN’s examples, and average values of whole token readings (including T1 which has no glottal stop) of *shiin* by JN, are shown in the following table:

<table>
<thead>
<tr>
<th>Glottal stop</th>
<th>Token</th>
<th>CVV</th>
<th>Vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>No</td>
<td>770.6</td>
<td>540.86</td>
</tr>
<tr>
<td>T2</td>
<td>38.91</td>
<td>837.82</td>
<td>593.92</td>
</tr>
<tr>
<td>T3</td>
<td>65.49</td>
<td>912.11</td>
<td>629.21</td>
</tr>
<tr>
<td>T4</td>
<td>22.40</td>
<td>837.82</td>
<td>597.46</td>
</tr>
<tr>
<td>T5</td>
<td>66.03</td>
<td>915.65</td>
<td>597.46</td>
</tr>
<tr>
<td>Ave</td>
<td>854.8</td>
<td>591.78</td>
<td>432.74</td>
</tr>
</tbody>
</table>

Table 1. The results of *shiin* in citation forms by JN (msec)

It is evident from the table that the duration of token, CVV and vowels of these four examples which contain glottal stops are longer than those of T1 which has no glottal stop insertion. Another example was uttered by a male participant MF and the duration of the glottal stop is 28.12 msec.

3.2. Ratio Comparison

In the following figure, the ratios of IVS token to 2 mora token and Long V token to 2 mora token are represented by different symbols. The actual ratio values, which were calculated based on the basic ratio, 1, of 2 mora token, are shown in the table below the graph:

![-ratios-of-IVS-and-Long-V-Tokens-to-2-Mora-Token](image)

The speech rate differences are shown on the horizontal axis, i.e., C (citation form), S (Slow), M (Mid), and F (Fast). The simple phonological ratios of a 2 mora token and corresponding 3 mora token i.e., an IVS token or a Long V token are 1 : 1.5. As figure 2 shows, the average ratio comparison of 2 mora and Long V tokens is 1 : 1.49 and the comparison of 2 mora and IVS tokens is 1: 1.61. Thus, it was found out that the ratio of a Long V token is very close to the phonological ratio, 1.5, but the ratio of IVS token, 1.61, is larger than the phonological ratio. As the speech rates get faster, the ratios get smaller and the ratios of IVS token are larger than those of Long V token, as was true of the citation forms.

The figure 3 shows the results of consonant + vowel(s) comparison:

![-ratios-of-consonant-vowel(s)-comparison](image)

(The comparison between koou and ko-o is excluded because the segmentation between the vowels was almost impossible in many cases.)
The phonological ratio comparison between the first mora of a 2 mora token, i.e., CV and the first and second morae of 3 mora words, i.e., CV: of a long V token or CVV of an IVS token is 1 : 2. It is clear from the figure that the average ratio of 2 mora token to Long V token in citation forms is 1 : 1.75 and the ratio of 2 mora token to IVS token is 1 : 1.87. The ratio 1.75 is around the middle of the ratios found in the previous studies, i.e., 1.72 in Y. Sato [7], 1.77 in Hoequist [8], and 1.73 to 1.77 in Kanzaki [9]. The ratio of IVS token against 2 mora token was found to be larger than the ratio of Long V token against 2 mora token and both ratios are lower than the phonological ratio, 2, although the ratios in the slow rate are larger than 2. The decrease of ratios in accordance with speech rate increase from S to F can be seen here as well, as in the case of token ratio. Most importantly, the consonant + vowel ratios of IVS token are larger than those of Long V token when framed in test sentences, too.

The average ratio of a short vowel to a long vowel is 1 : 2.66, as the figure 4 indicates:

(The comparison between koou and ko-o is excluded here as well for the same reason mentioned above.)

The ratio is exactly identical to the result of Y. Sato, whereas Kanzaki’s result is slightly lower than this value, i.e., 1 : 2.27 - 2.48. The average ratio of the vowel part of IVS token to a short vowel of 2 mora token in this study is 2.93. The ratios of S, M, and F are larger than that of C and the ratios are over 3.0. Unlike the token and consonant + vowel ratios in figures 2 and 3, the vowel ratios of IVS token and Long V token against 2 mora token do not get smaller as the speech rates get increased and even get slightly larger from S to M.

4. CONCLUSION

The glottal stop insertion seldom occurs in citation form reading, let alone in sentence reading even at a slow speech rate. The duration of token, consonant + vowels, and vowel part of the glottal stop examples get longer than those of the example without a glottal stop in the case of JN.

The ratios of IVS token are larger than those of Long V token in all the cases of token, consonant + vowel(s), and vowel comparisons at all speech rates. While the token and consonant + vowel ratios of IVS token and Long V token get smaller as the speech rates get faster, the vowel ratios do not get smaller and remain nearly consistent. One possibility of the ratio dominance of IVS token over Long V token is the effect of pitch accent differences between them. That is to say, a rising pattern of IVS token, i.e., LHH might have caused a relatively longer duration than that of Long V token which has a falling pattern, i.e., HLL. Relating to this point, some of the younger participants tend to flatten the falling pitch accent pattern and pronounce Long V tokens with a high level pattern such as HHH, as in ka-to. Considering this accentual change among young speakers, the result of this study should be reconsidered and whether the durational ratio dominance are intrinsic to IVS tokens or not will be subject to further study.

5. REFERENCES