The Phonological Substance of Q and the Bimoraic Foot in Japanese

Hiromi OTAKA*

The first objective of this paper is to refute the view, traditional in Japanese linguistics, that the phenomenon of sokuon (hereafter referred to as “Q”, following Bloch, 1950) is the first part of a geminate consonant. Sokuon is the sound transcribed in the small version of the syllable “tsu” (osci) in the Japanese syllabary. In past studies of Q, most Japanese phonologists have focused on its duration in order to confirm its value on the timing tier; however, there has been little effort to verify its phonological status on the melody tier. Some researchers (e.g., Jinushi, 1963; Amanuma, Otsubo, & Mizutani, 1973) have even regarded Q as a syllabic consonant by itself, due to confusion between a syllable and a phonetic mora (known in Japanese as a haku ‘beat’). Irrespective of these approaches, in this paper Q will be regarded phonologically as a null phoneme equivalent to a rhythmic “rest” in music (i.e., as a pause of one mora in timing organization), which is manifested phonetically by various sounds depending on the following consonant, due to regressive assimilation.

The second purpose of this paper is to investigate the rhythmic system of Japanese, which consists of a limited number of basic units determined by syllable type, in order to explain vowel lengthening before a mora phoneme1) and vowel shortening after a mora phoneme.

It has been reported that Japanese vowels lengthen before geminate consonants (Fukui, 1978; Maddieson, 1985; Han, 1994; Ham, 2001; Aoyama, 2000; 2002) and shorten after them (Fukui, 1978; Takada, 1985; Aizawa, 1985; Campbell, 1999; Gordon, Munro, & Ladefoged, 2000; Ofuka, 2003). Usually, in syllable-timed languages, vowels are shortened in closed syllables, including syllables that are closed by a geminate consonant. This is known as closed syllable vowel shortening or CSVS (Maddieson, 1985). However, many counterexamples to CSVS can be

* Professor Linguistics, School of Economics, Kwansei Gakuin University
1) Mora phonemes are, for example, the second mora of the diphthong in /taida/ (‘laziness’), the nasal consonant in /riNgo/ (‘apple’), the second mora of the long vowel in /dooro/ (‘road’) and the first mora of the geminate stop in /katta/ (‘bought’).
found in Japanese. Maddieson (1985) attributes this deviation from CSV to the fact that Japanese is not actually a syllable-timed but instead a mora-timed language, in which the first part of a geminate consonant derives from a CV syllable (p.214). Thus, Maddieson rejects the general assumption that Q is the coda of the syllable that contains the preceding vowel.

Consider the phenomena of mora phonemes within the framework of the mora-timed rhythm of Japanese. How does mora-timed rhythm relate to the change in vowel duration before as compared to after a geminate consonant? In syllable-timed languages, vowel shortening before a geminate consonant is part of an overall timing strategy that seeks to equalize the temporal interval between syllables (Ham, 2000, p.169). In a mora-timed language, on the other hand, a vowel is not shortened before Q; instead, it is lengthened by up to 100% as compared to a vowel elsewhere (Ofuka, 2003). Why?

**Brief review of Past Findings on Mora Duration in Japanese**

Let us now briefly review past findings on the duration of Japanese syllable constituents worth one or two moras, represented as Q, N, V(C), (C)V, (C)VV, (C) VQ, and (C)VN. Many Japanese phonologists regard Q as a temporal unit and claim that it is the length of one phonetic mora, the same as the duration of a CV syllable. Thus, Q is called a mora phoneme. Another mora phoneme is hatsuon, which means a nasal that occurs after a vowel, hereafter denoted with N (e.g., /hoN/ ‘book’). The claim that Q and N are mora phonemes has been supported based on native Japanese speakers’ intuitions that Japanese rhythm is “mora-timed” (Han, 1962; Hirata, 1990; Vance, 1987; Kubozono, 1989, etc.).

However, no one has rigorously proved the validity of this claim through acoustic experiments. The major reason past studies have not been successful in doing so is that it is not feasible to demarcate clearly the boundary between the end of Q and the beginning of the following homorganic consonant, because there are no acoustic cues between them that are discernible when looking at sonograms of the sound. As a result, instead of dealing with Q in isolation, phoneticians of the past analyzed it using minimal pairs consisting of single vs. geminate consonants for the purpose of comparison, for example, /kata/ ‘shoulder’ vs. /kaQta/ ‘bought’ (Aizawa, 1985; Hirata, 1990; Ofuka, 2003).

Phonologists have claimed different ratios of duration between Japanese geminate and single consonants. For example, Han (1960, 1992) reported a mean of 2.6:1 and often ratios as high as 3:1, the same ratio claimed by Homma (1981). However, Toda (1996) reported a mean ratio of only 2.4:1. Although the findings have varied among phonologists, all of them found that the duration of a geminate
consonant is not only twice as long as a single consonant, but also includes an additional mora. Therefore, these phonologists support the claim that the length of Q is one mora.

Of considerable interest among phonologists pursuing the temporal value of Q is the relative duration of the vowel (RDV), which is the ratio of the duration of the vowel that precedes a geminate consonant to the duration of the following singleton or geminate consonant in a CV(C)CV sequence as a phonetic cue to geminate consonant perception. The RDV is considered to be a phonetic cue for perception of Q as opposed to a single consonant, based on the assumption that the underlying temporal value of Q is one mora (Fujisaki & Sugito, 1977; Hirata, 1990; Takagi & Mann, 1994; Uchida, 1996; Toda, 1996; Arai & Kawagoe, 1996; Tsurutani, 2001). However, the RDV can actually more or less vary depending on the intrinsic duration of the concerned vowel\(^2\) and the syllable structure in which the vowel is involved (e.g., V or CV).

The effect of the RDV on geminate consonant perception was investigated at both the word and sentence levels by Hirata (1990). The results of her experiment showed that the RDV was an effective cue when the target words were spoken in citation form (i.e., not in a sentence), while speaking rate (i.e., the average durations of the moras in the carrier sentence) was influential when the target words were spoken in sentence contexts. That is, the slower the speaking rate, the more often the stimulus was perceived as an utterance without a Q.

Tsurutani (2001) claimed that the RDV should be at least 1.56 to 1.69 in order for native Japanese listeners to distinguish a geminate consonant from a single consonant in perception. This confirmed Fujisaki and Sugito (1977)’s claim of 1.69 for the RDV, which was also the same as Toda (1996)’s claim. Thus, the RDV is currently thought to be an effective cue for Q perception, based on the assumption that Q’s weight is one phonetic mora. In other words, in terms of timing organization, there seems to be a duration ratio close to 1:1 between CV and Q.

In addition, Fukui (1978), Takada (1985), Aizawa (1985), Campbell (1999), Gordon et al. (2000), and Ofuka (2003) all found that in Japanese, a vowel that comes before a geminate consonant is longer than a vowel before a single consonant, and a vowel that appears after a geminate consonant is shorter than a vowel after a single consonant. In fact, Ofuka (2003) reported based on her empirical findings that a vowel that is before a geminate consonant is almost twice as long as a vowel that is before a single consonant, and that a vowel that comes after a geminate consonant is about 10 milliseconds shorter than the same vowel.

---

2) According to Han (1962), order of intrinsic duration among Japanese consonants is as follows: voiceless consonants ≥ nasals ≥ voiced consonants ≥ approximants.
after a single consonant. The kind of the vowel sound does not matter here. According to Gordon et al. (2000), vowels are longer before a geminate consonant than before a single consonant by 1.1%–8.7%.

However, there is still no consensus among phonologists about whether or not a shortened vowel after a geminate consonant is an acoustic cue for listeners to perceive Q from a given utterance. Watanabe and Hirafuji (1987) are against this interpretation, but Hirata (1990) is in favor of it. Thus, this is still an unanswered question relating to the perception of Q.

It may also be interesting to inquire whether speakers of languages other than Japanese use the same internal mechanism to control the timing of geminate consonants. Consider Italian, which is a syllable-timed language as opposed to a mora-timed language. Based on the finding of a significant difference in duration between Japanese and Italian geminate consonants observed through a production experiment, Otaka (2009) claimed that Japanese geminate consonants undergo a time-measuring procedure during production that is different from what occurs with Italian geminate consonants, supporting the hypothesis that geminate consonant duration in Japanese is different than in Italian. Otaka (2009) found that the ratio between the duration of Q and the duration of CV in Japanese is much closer to 1:1 than the corresponding ratio in Italian. This is because Japanese has a mora-timed rhythm, while Italian has a syllable-timed rhythm, which implies that Q in Japanese has a duration of one mora, making it phonologically different from the first half of a geminate consonant in Italian, for which this is not the case.

Next, let us consider the case of Spanish. Hoequist (1983) reported that the duration ratio of bimoraic closed to monomoraic open syllables (e.g., /saN/ vs. /sa/) is 1.8:1 in Japanese but only 1.66:1 in Spanish, while a closed bimoraic syllable /saN/ was slightly longer than an open bimoraic syllable /saa/ in Japanese (1.8 vs. 1.7 on the basis of 1.0 for /sa/). This led him to reject the traditionally favored claim that syllable duration in Japanese is dependent solely on the number of moras in a syllable (Block, 1950; Hattori, 1955; Han, 1962; Fukui, 1978, etc.). This leads to another question: Why is the duration of a bimoraic syllable such as CVN or CVQ not twice as long as that of CV, which is also bimoraic?

**Theoretical Validation of Q Based on the Obligatory Contour Principle**

In this section, it will be argued that the traditionally favored view that Q is the first half of a geminate consonant violates the *obligatory contour principle* (OCP) in the case of words containing two identical geminate consonants, as in *sekkekkyuu* ‘red blood cell’.

The OCP is a major principle in phonology, as rich in implications as the
principle of economy.\textsuperscript{3} It functions as a constraint prohibiting the succession of two identical sounds within the same phonological word. It is why, for example the consonants /tr/ and /bl/ in the English words \textit{February} and \textit{probably} are usually omitted when pronounced, especially in casual speech, with the results of [\textit{f\textsc{b}ju\textsc{r}\textsc{a}r}] and [\textit{pr\textsc{b}l\textsc{i}}], respectively. The reason for the difficulty experienced in uttering tongue twisters such as English \textit{she sells seashells by the seashore} without stuttering can also be explained by the OCP.

Let us now examine the differences in phonotactics between Japanese and Italian geminate consonants. In Italian (Pickett, E. R, Blumstein, S. E., & Burton, M. W., 1999), the same geminate consonant does not occur more than once within a word due to the constraint of the OCP. For example, there are such Italian words as \textit{ingannerebbero} ‘will betray’, \textit{videocassette} ‘videocassette’, \textit{bellezza} ‘beauty’, and \textit{cosiddetto} ‘as it were’. However, this is not the case for Japanese geminate consonants, as seen from examples such as \textit{sekkekkyuu} ‘red blood cell’, \textit{hakkekkyuu} ‘white blood cell’, \textit{kekkindkuin} ‘tubercle bacillus’, \textit{ikkakki} ‘kind of demon with one horn’, \textit{makkakka} ‘very red’, and \textit{mottette} ‘take it’. This indicates that Q is not likely a unit on the melody tier, while the first consonant of geminate consonants in Italian is.

Therefore, we postulate that Q should be regarded as a phonological unit stipulated not on the melody tier, but solely on the timing tier. In other words, Q is a temporal unit of the mora-timed Japanese rhythm, and its acoustic realization is determined by the status of the consonant that follows it, through regressive assimilation. For example, below is an illustration of the prosodic structure of the Japanese prosodic word \textit{katta} ‘bought’, which contains Q medially. In the figure below, the symbol \(\mu\mu\) in parentheses represents a bimoraic foot, a single \(\mu\) without parentheses represents a mora, and \(\sigma\) indicates a syllable. Note that the syllable structure of CVQ has the coda Q with the lowest possible sonority, that is, as a null phoneme.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{prosodic_structure.png}
\caption{The prosodic structure of \textit{katta} ‘bought’}
\end{figure}

\textsuperscript{3} The “principle of economy” in language functions in two different ways: to achieve a given effect with reduced effort or to enhance an effect with a given amount of effort.
In fact, it is not a new claim that the phonological substance of Q depends on regressive assimilation. Several Japanese phonologists have claimed this in the past, including Hamada (1954), Kobayashi (1969), and Numamoto (1997). However, these scholars did not elaborate further regarding the substance of Q before a sound change. In fact, the present paper is the first to claim that the phonological substance of Q is solely a temporal unit, analogous to a musical “rest” with a duration of one mora in the rhythmic organization of the prosodic word.

Since Q is basically a rhythmic unit, it sometimes occurs before a vowel for emphasis, as in /aQa/ ([aʔa] ‘oh,’ ‘No’), like a glottal stop in English, or even at the end of an utterance, as in /koraQ/ ([koraʔ] ‘No!’). This shows that Q cannot be regarded as the coda consonant of the preceding syllable on the phonological as well as the phonetic level, because if this was the case, then we could not explain the occurrence of these phenomena.

In other words, Q typically occurs before a voiceless stop because it is easiest for a person to insert a pause as a rest in this position as opposed to before a voiceless fricative or a voiced stop. (The time-measuring mechanism of Q within the bimoraic foot CVQ will be discussed in the “Questions on Japanese Rhythm” section below.)

Q as a Temporal Unit Based on the Recognition of Musical Rhythm

In this section, the traditional view that Q has a length of one mora in the Japanese rhythm structure is supported from the point of view of musical rhythmic theory.

As mentioned above, Q is assumed by many Japanese phonologists to have a temporal length of one mora in both production and perception. In spite of the lack of decisive acoustic data to support this claim, we can argue that if Q is really a rhythmic unit in Japanese, then its temporal ratio to the preceding or following CV syllable should be 1:1.

According to Kajikawa (2003), newborn babies already have the ability to recognize differences in rhythm. This seems to be because durational ratios between sounds in natural languages, like those in music, are always simple. That is, they are always made up of whole numbers; the ratio can never be 1:1.3, for example. Some deviation from the ratio designated on the phonological level may possibly occur on the physical level, but the resulting sequence will still be perceived by the listener as 1:1, despite the fact that it consists of two sounds physically different in duration, as long as the speaker intended to produce the sequence at 1:1 and as long as the difference in duration between the first and second sounds is not too great, that is, not over 1:1.5. (In cases where the difference is too great to round down, the ratio
will be perceived as 1:2 instead, rounded up.)

It has been claimed by music theoreticians that humans can recognize rhythm (i.e., the relationship between two or more sounds in durational ratio) by one or two or all the following three methods: (1) multiplying a beat: 1:1, 1:2, 1:3 (Cowell, 1930), (2) dividing a beat: 1:1, 1:½, 1:⅓, 1:¼ (Messiaen, 1954), and (3) changing the duration of the basic beat (Carter, 1997). Otherwise, humans could not identify regularity in musical rhythm. In analogy, we can postulate that the rhythmic patterns in language are made up of simple ratios that unite sounds (syllables and moras) in the same way as in music, although there are likely to be deviations from musical rhythm when syllables manifest as physical sounds.

However, we need to be careful about possible differences in length between the phonological (or psychological) and phonetic (or physical) realizations of sounds uttered to a certain rhythm. Even in music, a sequence of notes produced to the rhythm of 1:2, for example, is likely to deviate from this exact ratio to some extent, depending on how stress is allocated within a measure. This kind of deviation can be seen in linguistic rhythm because of the variation of segmental sounds in intrinsic duration, as mentioned in Footnote 2 above.

**Confirming Vowel Length Before and After Q/N**

We will now quantitatively test the validity of Ofuka’s (2003) claim that in Japanese, the vowel before a geminate consonant is almost twice as long as the vowel before a single consonant, and the vowel after a geminate consonant is 10 ms shorter than than the vowel after a single consonant. In addition, vowel length before and after N will also be measured. This is because not only Q but also N constitutes a geminate consonant in Japanese (as for example, in kanna /kaNna/ ‘plain’). 4)

Six native-Japanese-speaking informants participated in this production experiment. All were university students born and raised in the Kansai region of Japan, 20 to 22 years old (three men and three women). They were asked to read the following three Japanese sentences at a normal speed without pausing between words.

1. Byooki ni kakatta ka to omoimashita. (‘I wonder if you got sick’.)

4) Some researchers do not consider the nasal sequence /Nn/, which is manifested phonetically in Japanese as [nn], [mm], or [ŋŋ], to be a geminate consonant (Hattori, 1951, 1955; Kuroda, 1965; Nakajo, 1989; Tsuzuki, 1997), but others do (Maddieson, 1985; Murata, 1993; Otaka, 2009). In this paper, /Nn/ is regarded as a geminate consonant based on the definition adopted in Trask (1996, p.154), who regards geminate consonants simply as sequences of two identical consonants.
These are the target words:

Sentence 1: *kakattaka* (CVCVQCVCV: ‘if [you] got sick’)

Sentences 2 and 3: *kana* (CVCV: ‘syllabaries’), *kan* (CVN: ‘can’) and *kanna* (CVNCV: ‘plain’)

The first sentence is intended to test the validity of Ofuka’s claim mentioned above, while the second and third sentences are meant to examine whether or not the same phenomena occurs with N as with Q. The second and third sentences are the same except for the order of the target words. Because one of the target words, /kaN/ (CVN: ‘can’), is followed by a voiced consonant /d/ in Sentence 2, the nasal coda of this syllable may lengthen as compared to its counterpart, which is followed by a voiceless consonant /t/. The reason why the target /kaN/ is placed before /t/ in Sentence 3 is so that the durations of all the target words can be placed under the same conditions for comparison (in that all of the target words are followed by a voiceless consonant /t/).

The utterances produced by the six informants were recorded on a computer using a dynamic microphone (Sony EOM 717), and the durations of each segment C, CV, V, QC, Nn, and N of the target words were measured using sound analysis software (SpeechStation 2 by Sensimetrics, Malden, 2003).

The average durations of C, V, CV, QC, Nn, and N are shown in the three tables below, in milliseconds. Table 1 shows the duration of target word /kakaQtaka/ in Sentence 1, and Tables 2 and 3 show the durations of target words /kaN/, /kana/, and /kaNna/ in Sentences 2 and 3. (These target words contain the nasal [n].) “M1” indicates that the syllable is located in the first moraic position within the word, “M2” that it is located in the second moraic position, etc.  

From the data in Table 1, it can be seen that the average duration of the first CV syllable /ka/ (M1) in the sequence /kakaQtaka/ (CVCVQCVCV) was 142 ms, including both the closure and VOT durations of the plosive /k/, while the duration

---

5 ) Because /Qt/ in the third moraic position in Table 1 contains /t/, which is the onset of the fourth moraic syllable /tal/, it is not marked as M3.

6 ) Voice onset time (VOT) is the time which elapses between some articulatory event, most often the release of a plosive, and the point at which the vocal folds begin to vibrate. The closure duration of /k/ is the implosive part of the three articulatory phases (i.e., implosive, halting, and explosive) that /k/ involves.
of the second syllable /ka/ (M2) followed by Q was 151 ms, which is 9 ms (6%) longer. However, for the fifth mora /ka/ (M5), which is preceded by /kakata/, the duration was only 121 ms. This is 20% shorter than the duration of the syllable /ka/ (M2) before Q.

Note that if the duration of the closure of /k/ is excluded from the calculation of CV length (i.e., if only the VOT and the vowel duration of /a/ are counted), then the duration of /a/ (M4) followed by Q (87 ms) is 32% longer than the duration of /ka/ (M2) preceded by Q (66 ms). This implies that the claim made by Ofuka (2003) is valid, but that the difference in duration between the two vowels does not seem to be as great as she had reported (32%, not 100%, as Ofuka claimed).

In addition, it also turned out that the CV syllable (M1) that was followed by a bimoraic foot (CVQ) was longer than its counterpart (M5), which was located between CV syllables (as the third /ka/ in the target /kakaQtaka/).

Table 1 Durations (ms) of Segments in /kakaQtaka/ ([kakattaka]): \(n=6\)

<table>
<thead>
<tr>
<th></th>
<th>/ka/(M1)</th>
<th>/ka/(M2)</th>
<th>/Qu(/tt/)</th>
<th>/a/(M4)</th>
<th>/ka/(M5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV</td>
<td>142</td>
<td>151</td>
<td></td>
<td></td>
<td>122</td>
</tr>
<tr>
<td>V</td>
<td>80</td>
<td>88</td>
<td></td>
<td>60</td>
<td>66</td>
</tr>
<tr>
<td>CC</td>
<td>62</td>
<td>63</td>
<td>157</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>62</td>
<td>63</td>
<td></td>
<td></td>
<td>56</td>
</tr>
</tbody>
</table>

Table 2 Durations (ms) of Segments in /kana/ (Left), /kanNa/ ([kanNa]: Middle), and /kaN/ ([kan]: Right); \(n=6\)

<table>
<thead>
<tr>
<th></th>
<th>/ka/(M1)</th>
<th>/na/(M2)</th>
<th>/ka/(M1)</th>
<th>/Nn/</th>
<th>/a/(M3)</th>
<th>/ka/(M1)</th>
<th>/N/(M2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV</td>
<td>146</td>
<td>136</td>
<td>183</td>
<td></td>
<td></td>
<td>186</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>66</td>
<td>86</td>
<td>82</td>
<td></td>
<td>64</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td></td>
<td></td>
<td>130</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>80</td>
<td>50</td>
<td>101</td>
<td></td>
<td></td>
<td>104</td>
<td>112</td>
</tr>
</tbody>
</table>

Table 3 Durations (ms) of Segments in /kana/ (Left), /kaN/ ([kan]: Middle) and /kaNna/ ([kanNa]: Right); \(n=6\)

<table>
<thead>
<tr>
<th></th>
<th>/ka/(M1)</th>
<th>/na/(M2)</th>
<th>/ka/(M1)</th>
<th>/N/(M2)</th>
<th>/ka/(M1)</th>
<th>/Nn/</th>
<th>/a/(M3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV</td>
<td>148</td>
<td>134</td>
<td>187</td>
<td></td>
<td>178</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>61</td>
<td>80</td>
<td>90</td>
<td></td>
<td>85</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>122</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>87</td>
<td>54</td>
<td>94</td>
<td>93</td>
<td>93</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
From the results shown in Tables 2 and 3, then, it turns out that the same results mentioned in relation to Q above were also found for vowels before and after N. As shown in Table 2, the duration of the first syllable /ka/ (M1) of /kana/ was 146 ms, including both closure and VOT of the plosive /k/, while the duration of M1 in /kaNna/ was 173 ms, which is an increase of 18%. Moreover, if compared based only on vowel duration, this would amount to an increase of 38%. On the other hand, the duration of /a/ (M3) before /Nn/ ([nn]) in /kaNna/ was 64 ms, but the duration of word-final /a/ (M2) in /kana/ was 86 ms in this case, the former is 28% shorter than the latter.

The same thing can be said about the data in Table 3. Since the average mora duration for the target utterances in Sentence 2 was the same as in Sentence 3 (136 ms), the target words in Sentences 2 and 3 were comparable in terms of segment duration.

Now let us look at the data in Table 3 in order to re-examine the validity of Hoequist (1983) by comparing the two target words /kana/ and /kaN/ in terms of duration under the same condition (before the voiceless consonant /t/). The duration of /kana/ was 282 ms, while that of /kaN/ was 277 ms, which supports Hoequist (1983)’s claim. However, the difference between the durations of the closed and open syllables was not as great as Hoequist reported; the ratio of CVN:CV in our data was 1.24:1, where he claimed that it was 1.8:1. More significantly, in our data, the vowel lengthening in M1 before N (90 ms) as compared to its counterpart before CV (61 ms in M1), was exactly the same result as in the case of the vowel before Q for M2, as seen in Table 1.

Questions Regarding Japanese Rhythm Based on the Use of Divisive/Additive Rhythm

The findings presented in the previous section lead to the following questions:
1. Why are Japanese vowels likely to lengthen when they are before Q or N, but to shorten when they are after a mora phoneme?
2. Why is the duration of the bimoraic syllable CVN shorter than the duration of two CV moras?

In order to answer the first question, we will assume that the bimoraic syllables

---

7) The average mora durations for the utterances in Sentence 2 and 3 were calculated as the sum of the durations of the three target words (/kana/, /kaNna/, and /kaN/) plus the two syllables of /to/ ‘and’ divided by nine (because there were a total of nine moras in the target part /ka.na.to. ka,N,na.to.ka.N/.

8) As seen in Table 3, the duration of /kaN/ as a whole is 184 ms, and the duration of the /ka/ of /kana/ is 148 ms. Thus, the ratio between /kaN/ and /ka/ is 184:148, which is 1.24:1.
(C)VN and (C)VQ are full-fledged rhythmic units that are twice as long as (C)V syllables. The former can be called a “long unit” and the latter a “short unit”, because of the ratio 2:1 between them in terms of moraic value. However, phonetically speaking, this ratio cannot reliably amount to exactly 2:1, because the syllables are independent units. In other words, the duration of bimoraic syllables cannot be determined by doubling that of monomoraic syllables, that is, not through an additive rhythm based on a CV mora, but instead through a divisive rhythm. The difference in time-measuring mechanisms between these methods is comparable to that between series of two different kinds of musical notes with the same temporal value overall, as shown in (b) and (c) below.

a. ↓  b. ♩ ♩  c. ♩ ♩

The temporal value of a quarter note, shown in (a), is equal to the temporal value of two eighth notes, seen in (b) and (c). However, there is a fundamental difference between (b) and (c) in terms of time-measuring mechanism. The rhythmic pattern (a) is one single sound, but pattern (b) has an additive rhythm, whereas pattern (c) has a divisive rhythm (Sachs, 1953). As a result, the actual duration of each eighth note in (b) will tend to be slightly different from that of the corresponding note within the domain of a rhythmic unit called a duplet (two musical notes played over the duration of one beat⁹), as in (c). The duration of an eighth note is repeated twice as a unit in pattern (b), but the whole duration, worth one quarter note, is divided into two parts in (c).

Measuring time in rhythmic organization is comparable to measuring the length of a straight line in geometry; therefore, our argument about rhythm can also be illustrated with the help of geometry, as follows. A line is regarded as a collection of points aligned on a flat geometrical space, and the line’s length is determined by measuring the distance between its two end points. When two lines of equal length AB and CD are connected together into one straight line, as shown in Figure 2 below, four points need to be employed to measure and unite the lines in a certain ratio AB:CD. In this case, since the two lines have equal length, AB:CD = 1:1.

---

⁹ The definition of “duplet” in linguistic rhythm here is somewhat different from that in music theory. In music, a duplet is a type of “tuplet” — i.e., a note-grouping of two, which fits into the length of three of its note-type. For example, an eighth-note duplet spans three eighth note beats (one quarter-note or one beat). However, both are the same in that one beat (or the basic rhythmic unit) is divided into two equal sounds.
On the other hand, when a line AC is divided into two lines of equal length, another point (B) must be placed in the middle of line AC to divide the line AC into two equal halves, as illustrated in Figure 3 below.

Note that in the case of divisive rhythm (illustrated in Figure 3), only three points are utilized to produce the rhythmic pattern. The ratio 1:1 should manifest more accurately in additive rhythm (Figure 2) than in divisive rhythm (Figure 3) because the phonetic mora functions as the basic duration unit, equivalent to lines AB and CD in Figure 2.

This is analogous to the method of measuring distance using our feet while walking across the ground. It is easier for a person to measure a distance that is longer than his or her foot as opposed to shorter. In the latter case, the distance (for example, one-third of a foot or half a foot) has to be guessed in comparison with the length of a foot. In this way, divisive rhythm is less accurate than additive rhythm in measuring temporal duration.

Regarding the phonetic mora as the basic temporal unit in Japanese, it will be clear that the normal duration of a mora varies among individual speakers, just like the length of a person’s foot varies among different people. (All other things being equal, a mora for slow speakers is longer than one for fast speakers.) However, it can be assumed that for each individual, the phonetic mora has an absolute/regular durational value in general, just like each person’s feet have a fixed length, unless affected by some kind of unusual emotion such as anger, fear and surprise. Therefore, there is some variation in the duration of a mora among different people, but for a certain person, the length of a mora is fixed. In contrast, a line in geometry does not have an absolute length unless that length is defined in terms of a unit of measurement, for example as a certain number of centimeters or inches.

Since a bimoraic foot can be regarded as a full-fledged rhythmic unit, whose length is the same as that of two CV moras, and since the duration of the two elements is determined through divisive rhythm, the duration of a vowel coming before Q or N will tend to deviate from the standard duration of a mora for that speaker. However, the question of why the first segment lengthens compared to the
second segment within a foot is still unanswered. We might be able to postulate a difference in function between the *downbeat* (the first beat of a measure or foot) and the *upbeat* (the second beat of a measure or foot) and apply it to the two constituents of the bimoraic foot, but it would require further experimental research to make a firm conclusion regarding the validity of this approach.

Figure 4 below illustrates the rhythmic structure of the target word /kakaQtaka/. The foot /kaQ/ is indicated in parentheses as a full-fledged rhythmic unit in Japanese (i.e., \( \overline{\gamma} \)). Note that in a bimoraic foot like CVQ or CVN, Q or N must occur in the second position. That is why there are no words in Japanese that can begin with Q or N.

There is one more question left to answer: Why does the duration of a CV syllable tend to differ depending on its location within a word? In the previous section, a monomoraic syllable before a bimoraic syllable (such as M1, which is CV) was seen to be longer than a monomoraic syllable after a bimoraic syllable (such as M5, which is CV). This apparently runs counter to the workings of mora-timed rhythm, and so this question is a very perplexing one, because a CV syllable and the preceding or following bimoraic syllable (CVQ or CVN) are supposed to combine into one sequence through additive rhythm, as discussed earlier. Thus, one would expect that the duration of the CV syllable would become more stable regardless of its location within the utterance.

However, as it turned out (as shown in the previous section; see Tables 1, 2, and 3 above) that a CV syllable before a foot was somewhat longer, and one after a foot shorter, than one not adjacent to a foot. This seems to indicate that an unknown rhythmic phenomenon caused the CV mora to deviate from the norm of the mora-timed rhythm. One plausible cause of this phenomenon is assimilation from the adjacent bimoraic foot, as a long rhythmic unit. In other words, a CV mora coming before CVQ or CVN tends to be lengthened because the following CV is lengthened, whereas a CV mora after CVQ or CVN tends to be shortened because the preceding mora phoneme Q/N is shortened within the foot. This will be a focus

---

10) The symbol “\( \overline{\gamma} \)” in music indicates “eighth rest” whose temporal value is equal to “\( \overline{\gamma} \).”
of further research on my part.

**Conclusion**

The first objective of this paper was to present a new view of the phonological substance of *sokuon* or Q (Bloch, 1950) in Japanese. Q has long been regarded by many Japanese phonologists as the first part of a geminate consonant. However, in this paper, Q is newly defined as a *null phoneme*, equivalent to the concept of a rhythmic *rest* in music and manifested as a consonant with the same quality as the following consonant due to regressive assimilation. In other words, the substance of the first half of a geminate consonant is intrinsically unspecified in the prosodic word, and is only filled by the following consonant.

On the phonological level, Q is a rhythmic rest worth one mora, stipulated only on the timing tier, in contrast to the first half of a geminate consonant in a language like Italian. This argument is validated based on the observation of the difference in phonotactics between Japanese and Italian geminate consonants in light of the constraint from the Obligatory Contour Principle (OCP). Italian geminate consonants are consonant phonemes on both the melody tier and the timing tier, but Q in Japanese geminate consonants is stipulated only on the timing tier. Thus, it phonologically functions as the null phoneme coda of the preceding (C)V on the syllable level in the prosodic hierarchy.

The second objective of this paper was to explore the basic Japanese rhythmic units determined by syllable types in order to explain coherently the Q/N-related phenomena reported by past researchers in terms of the mora-timed rhythm. It has been claimed by past researchers that the vowel before Q gets lengthened (Fukui, 1978; Maddieson, 1985; Han, 1994; Ham, 2001; Aoyama, 2000, 2002), but the vowel after Q gets shortened (Fukui, 1978; Takada, 1985; Aizawa, 1985; Campbell, 1999; Gordon et al., 2000; Ofuka, 2003). In addition, Hoequist (1983) reported that the duration of closed syllables such as CVN are 10% shorter than the duration of CVCV, even though both have two moras.

In the previous section, the validity of these claims was examined through a production experiment in which target words containing N (/kana/, /kaN/, /kaNna/) were examined, because not only Q but also N can constitute a bimoraic foot in Japanese. It turned out that vowels do indeed become lengthened before N as well as Q within bimoraic feet, and shortened after. It was also discovered that the overall duration of CVN (e.g., /kaN/) is much shorter than that of CVCV (e.g., /kana/).

Moreover, the experiment showed that the duration of a CV syllable (M1) followed directly by a bimoraic foot (/kaQ/) also lengthened, as did that of a CV
syllable within a bimoraic foot (/kaQ/). (See Table 1). These syllables are much longer than the CV syllable (M5), which was located after another CV syllable (M4).

The findings discussed above can be summarized as follows:

1. Within a bimoraic foot CVQ or CVN, the duration of the first mora is longer than that of a CV syllable, implying that the second mora is shorter than the first mora in CVQ or CVN.
2. A CV syllable that comes before a bimoraic foot lengthens to almost the same duration as the first moraic syllable of a bimoraic foot, whereas a CV syllable after a bimoraic foot is shorter than a CV syllable not adjacent to the foot.
3. Bimoraic feet tend to be shorter than bimoraic CVCV sequences.

In order to explain these three findings in terms of the theory of Japanese mora-timed rhythm, we postulate that in Japanese, there are only four basic rhythmic units, determined by syllable type as follows, assuming that the temporal value of a mora is represented as an eighth note:

\[ \ddot{\text{ } } = (C)V, \quad \dot{\text{ } } = (C)VR, \quad \dddot{\text{ } } = (C)VQ, \quad \dddddot{\text{ } } = (C)VN \]

Based on these four units, various rhythmic patterns are made possible for words and phrases, as exemplified below. (Incidentally, words and phrases are thought to be stored in the mental lexicon bearing these patterns, together with their meanings and sounds.)

Ex.: /ko/ ‘circle’: /\dddot{\text{ } }/, /koko/ ‘here’: /\dddot{\text{ } }\dddot{\text{ } }/, /kokoro/ ‘heart’: /\dddot{\text{ } }\dddot{\text{ } }\dddot{\text{ } }/, /kokoo/ ‘isolation’: /\dddot{\text{ } }\dddot{\text{ } }/, /koko/ ‘anxiety about the future’: /\dddot{\text{ } }\dddot{\text{ } }/, /kookoo/ ‘high school’: /\dddddot{\text{ } }\dddddot{\text{ } }/, /koQko/ ‘exchequer’: /\dddddot{\text{ } }\dddot{\text{ } }/, /koQkool/ ‘diplomatic relations’: /\dddddot{\text{ } }\dddddot{\text{ } }\dddddot{\text{ } }/, /kiN/ ‘gold’: /\dddddot{\text{ } }\dddot{\text{ } }\dddot{\text{ } }\dddot{\text{ } }/, /kiNko/ ‘safe’: /\dddddot{\text{ } }\dddddot{\text{ } }\dddot{\text{ } }\dddot{\text{ } }/, /koKiN/ ‘old and new’: /\dddot{\text{ } }\dddddot{\text{ } }\dddot{\text{ } }\dddot{\text{ } }/, /kiNkoo/ ‘suburban’: /\dddddot{\text{ } }\dddddot{\text{ } }\dddddot{\text{ } }\dddot{\text{ } }\dddot{\text{ } }\dddot{\text{ } }/, etc.

In terms of duration, the four basic units above can be divided into either “short units” — monomoraic syllables (CV) — or “long units” — bimoraic syllables (CVR, CVQ, and CVN). 11) Both of these are full-fledged rhythmic units. That is

---

11) Q by itself could be considered a short unit, but it always appears together with a preceding CV syllable within a bimoraic foot, so CVQ should instead be regarded as a basic rhythmic unit in Japanese.
why bimoraic syllables are phonetically not exactly twice as long as monomoraic syllables in the mora-timed rhythm of Japanese.

As to the question of why bimoraic feet are shorter in duration than CVCV, this may be caused by the “principle of economy”, as defined in Footnote 3 above.

Now consider why vowels get lengthened before Q or N and shortened afterward. When CV and CVR syllables are connected in a sequence, an additive rhythm is adopted, but if a CV syllable is followed by either Q or N, with the result of a bimoraic foot (CVQ, CVN), then a divisive rhythm is adopted. Thus, in the case of CVQ or CVN, the actual duration of CV is likely to deviate from the standard length of a CV mora.

Why does the CV part of a bimoraic foot lengthen? It could be because of the difference between the “upbeat” and the “downbeat” of the foot. However, this is only a guess. Further studies based on quantitative experimentation are needed to give a definite answer to this question.

Finally, the last question is why CV syllables are likely to become longer before, but shorter after, a bimoraic foot, compared to the standard mora length given to these syllables outside such a foot. A possible answer to this may be “assimilation” of duration. In other words, despite the mora-timed rhythm, the length of a CV mora is influenced by the length of the preceding and following moras. This is why vowels lengthen slightly before a bimoraic foot in which the first mora lengthens, and shortened after a bimoraic foot in which the second mora shortens.

References


Kaishaku to kanshoo [Interpretation and appreciation], 34(14).


Watanabe, S. & Nobuo, H. (1985). Nionsetsugo ni okeru musei haretsuon to senkoo boin no nagasa no kankei [The relation between the perceptual boundary of voiceless plosives and their moraic counterparts and duration of the preceding vowels], *Onsei gengo* [Phonetic aspects of languages], 1, 1–8.