SYLLABLE COUNTING AND MORA UNITS IN SPEECH PERCEPTION

Donna Erickson*, Reiko Akahane-Yamada¹, Keiichi Tajima², and Kaori F. Matsumoto³
*The Ohio State University, Columbus, Ohio; ¹Kanazawa University, Kanazawa, Japan; ²ATR Human Information Processing Research Laboratories, Kyoto, Japan.

1. INTRODUCTION

In English the syllable is the basic rhythm-bearing unit to which stress is assigned. In Japanese, it may be the subsyllabic moraic unit e.g., [1]. In English, syllables can be described to a first approximation in terms of number of sonority peaks associated with a single jaw opening/closing gesture [2]. In Japanese, syllables can be described in terms of number of moras [3]. A Japanese syllable consists of one mora of the form (C)V, or two moras of the form (C)VV or (C)VC. The second C is restricted to be a moraic nasal or the first part of a geminate consonant. Additionally, tautosyllabic consonant clusters are not permitted phonotactically in Japanese. These constraints apply to both native Japanese words and loanwords, although loanwords sometimes violate them (e.g., /taan/ "turn" is a three-mora syllable of the form CVVC).

An additional complicating factor in Japanese listeners' abilities to hear English syllables is the conventional method of writing English words using the mora-based katakana script. Given the phonotactic constraints against consonant clusters in Japanese, consonant clusters in an English word are often split into singleton consonants with intervening epenthetic vowels (e.g., "stress" --> /su.to.re.su/). Each consonant in a cluster is written as a mora-bearing kana character in the loanword. In addition, English vowels can be one or two moras in length in Japanese loanwords. Lax vowels tend to be treated as one mora (e.g., "love" --> /ra.bu/). However, tense vowels, diphthongs, and lax vowels followed by voiceless obstruents often tend to be treated as two-mora units (e.g., "you" --> /yu.u/, "out" --> /a.u.to/,”tip” --> /ti.pu/). Thus, to a first approximation, it is likely that monosyllabic English words are felt by Japanese listeners to have as many moras as there are consonants in the word, together with whether the vowel is treated as one or two moras. Number of moras in the loanword may therefore be a measure of phonological length of the English word as perceived by Japanese listeners. Phonological length may in turn affect Japanese listeners’ ability to perceive syllables in English words.

Given the differences in what constitutes a syllable unit in the two languages, along with the tradition of transcribing English words using a mora-based kana script, Japanese listeners may have difficulty perceiving syllables as single units in English words. The following questions are investigated: (1) Do Japanese listeners have difficulty perceiving syllables in English words, (2) what factors contribute to Japanese listeners' perceptions of syllables, and (3) to what extent is physical length (i.e., acoustic duration) vs. phonological length (in terms of number of mora in the loanword) a factor in perceiving syllables in English words?

2. METHODS

A set of 275 English words varying in number of syllables, consonants, and vowels were recorded by a native speaker of American English, and presented through headphones to 5 American/Canadian-English listeners and 15 Japanese listeners. All the Japanese listeners had no experience with an English-speaking environment, but received for several years the ordinary English education in Japan, in which listening and speaking skills were not emphasized. Subjects were first instructed to count the syllables in each word and to indicate the number by pushing the appropriate key on the keyboard. Japanese listeners were given some additional instructions: "There are 'units' in speech. In Japanese, we tend to count the hiragana character as a speech unit. For instance, the word 'hiragana' has four units: hi-ra-ga-na. In English, a syllable generally has one vowel, optionally surrounded by consonants." Prior to the task, both English and Japanese listeners were given a practice test with feedback using a list of ten words ranging from one to three syllables.

The word list consisted of 80 one-syllable words, 144 two-syllable words, 28 three-syllable words, 9 four-syllable words, 12 five-syllable words, 2 six-syllable words, and 1 fourteen-syllable word. The one-syllable words were coded with respect to vowel type and number of initial and final consonants. These were determined according to how they would be written in Japanese using the katakana script in a loanword. The vowel types were S, R, D, and SQ: ‘S’ refers to a one-mora short vowel, ‘R,’ to a two-mora long vowel, ‘D,’ to a two-mora diphthong, and “SQ,” to a two-mora sequence consisting of a short vowel (S) followed by the first part of a geminate (Q). The number of initial and final consonants was the same as that in the original English word with several exceptions. The affricates "ts" and "ch" are treated as singleton consonants since they are written as such in Japanese. The velar nasal "ng" is treated as two consonants since it is written using two kana characters ("n" and "gu"). Post-vocalic [r] is treated not as a consonant but as the second part of a long vowel (R) or a diphthong (D).
The coding for the one-syllable words was then used to count the number of moras that each English word would have when borrowed as a loanword. This provides a measure of Japanese listeners’ perception of subjective length of each word. In addition, the acoustic durations of each of the words were measured.

3. RESULTS

3.1. Overall results and the effect of number of syllables

Native English listeners correctly identified the number of syllables in an English word approximately 98% of the time whereas Japanese listeners correctly identified the number of syllables 57% of the time (Figure 1, left panel). For Japanese listeners, percent correct generally tended to decrease with the number of syllables in the word, ranging from 62% correct for two-syllable words to 32% correct for five-syllable words (Figure 1, right panel). However, two syllable words were better counted than one-syllable words.

Looking at the error pattern by number of syllables (confusion matrix in Table 1), words with up to three syllables tended to be counted by Japanese listeners as having more syllables than they actually had, and words with more than three syllables were counted as having fewer syllables than they actually had.

The ensuing discussion in this paper focuses on an error analysis of the one syllable words in terms of Japanese syllable constituents, i.e., vowel types and number of initial and final consonants. We also examined the effect of physical and phonological length on the Japanese listeners’ perception performance.

### Table 1. Confusion Matrix showing perception of syllables by Japanese listeners (%).

<table>
<thead>
<tr>
<th>syl</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>53.0</td>
<td>32.5</td>
<td>10.8</td>
<td>2.9</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6.0</td>
<td>62.5</td>
<td>23.9</td>
<td>5.9</td>
<td>1.5</td>
<td>0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.7</td>
<td>17.1</td>
<td>61.4</td>
<td>4.8</td>
<td>5.0</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.5</td>
<td>7.4</td>
<td>33.3</td>
<td>41.5</td>
<td>13.3</td>
<td>2.2</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5.0</td>
<td>20.0</td>
<td>35.6</td>
<td>31.7</td>
<td>6.1</td>
<td>1.1</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>16.7</td>
<td>20.0</td>
<td>16.7</td>
<td>33.3</td>
<td>13.3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2. Vowel type

Since Japanese syllables consist of one or two moras, as determined by the vowel type, vowel type may affect perception of syllables by Japanese listeners. A one-way ANOVA with percent correctly perceived as one syllable as the dependent variable and vowel types (S,SQ,D,R) as the independent variable indicates that the vowel type has a significant effect on the accuracy of counting syllables \( [f(3,74)=3.467, p<.05] \) (Figure 2). (Since the two words counted by one of the subjects as having zero syllables is excluded from the analysis, the total number of one-syllable words in this figure is 78, not 80.) Post-hoc analysis showed that the words with S-vowels are significantly different from both those with D \( (p=.0120) \) and SQ \( (p=.0222) \), and words with an R-vowel are significantly different from D \( (p=.0318) \) but words with S-vowels are not significantly different from those with R-vowels.

3.3. Number of consonants

Since consonant clusters are not permitted in Japanese, the number of initial or final consonants also might affect perception of syllables by Japanese listeners. A one-way ANOVA with percent correctly perceived as one syllable as the dependent variable and number of initial consonants as the independent variable indicates that the number of initial consonants has a
significant effect on the accuracy of counting syllables \([f(3,74)=12.789, p<.001]\) (Figure 3). Perception accuracy of words with zero initial consonants is significantly different from those with three initial consonants \((p=.0014)\), one initial consonant significantly different from both two \((p=.0049)\) and three \((p=.0001)\) and two initial consonants significantly different from three initial consonants \((p=.0024)\).

![Figure 3. Effect of number of initial consonants. Double asterisks indicate \(p<.01\).](image)

A one-way ANOVA with percent perceived as one syllable as the dependent variable and number of final consonants as the independent variable indicates that the number of final consonants has a significant effect on the accuracy of counting syllables \([f(3,74)=6.4, p=.0001]\) (Figure 4). Perception accuracy of words with zero final consonants is significantly different from words with both one final consonant \((p=.0268)\) and two final consonants \((p=.0001)\) but the other differences are not significant.

![Figure 4. Effect of number of final consonants. Double asterisks indicate \(p<.01\), a single asterisk indicates \(p<.05\).](image)

3.4. Physical length vs. phonological length

In order to examine the effect of physical length vs. phonological length on the perception of monosyllabic English words by Japanese listeners, correlation and multiple regression analyses were done. The dependent variable was the percent perceived as one syllable and the independent variables were (1) the acoustic duration of the word, and (2) the number of moras that the word would have when borrowed as a loanword. The standardized equation was the following: \% correct = -0.237*duration + -0.626*mora (R-squared = .673). Both regression coefficients are negative, indicating that as the number of moras or acoustic duration of the word increases, Japanese listeners’ correct responses to one-syllable words decrease. Moreover, the regression coefficient for number of moras is slightly larger than that for acoustic duration. This suggests that number of moras contributes more to predicting Japanese listeners’ performance than does acoustic duration. However, a correlation coefficient between duration and mora of 0.786 indicates that the two independent variables were highly correlated with each other. Thus, further work is needed to examine the effect of individual factors.

4. SUMMARY

The results from this study confirm that Japanese listeners have difficulty counting syllables in English. First, the number of syllables in the words affected the performance. Errors generally increased with the number of syllables in the word. However, one-syllable words were counted less accurately than two-syllable words. This suggests that for some words, Japanese listeners are more sensitive to the precise structure of the word than to the number of syllables in a word.

Second, vowel type also affected Japanese listeners’ perception of syllables. One-syllable words whose loanwords contain a diphthong (D) or a vowel-plus-geminate sequence (SQ) were counted as having more syllables than those words with a one-mora short vowel (S). This result suggests that Japanese listeners count one-syllable words as having more than one syllable when they perceive the word to have diphthongization or gemination with the following consonant, but not necessarily simple vowel lengthening.

Third, the one-syllable error rate increased as the number of consonants increased. This result is consistent with Japanese phonotactic rules that prohibit consonant clusters and account for vowel epenthesis between consonants in loan words in Japanese.

Finally, the physical length (duration) and/or phonological length (number of moras heard in the loanword) also contributes to Japanese listeners’ perception of English syllables.

Japanese syllable structure is different from English syllable structure. This difference contributes to the difficulty in counting syllables in English. To further test these statements about what factors contribute to Japanese listeners’ perceptions of syllables, research is underway with a controlled set of nonsense CVC words which vary according to number and voicing of initial/final consonants and vowel type. Using nonsense syllables will control for influence of familiarity and spelling in a way that was not possible using loanwords.
Exploration of the interference of L1 rhythmical units (mora-based syllables) with L2 rhythmical units (sonority-based syllables) will hopefully lead to methods that help language learners be able to perceive rhythmical units, not according to those of L1, but according to the language to be learned.

ACKNOWLEDGMENTS
The authors wish to thank Rieko Kubo at ATR for her data collection, and Naho Suganuma at ATR for her assistance in acoustic measurements.

REFERENCES