The perception Boundary between Single and Geminate Stops in 3- and 4-mora Japanese Words

Shigeaki Amano\(^1\), Yukari Hirata\(^2\)

\(^1\) Faculty of Human Informatics, Aichi-Shukutoku University, Japan
\(^2\) Department of East Asian Languages and Literatures, Colgate University, U.S.A.

psy@asu.aasa.ac.jp, yhirata@colgate.edu

Abstract

The perception boundary between single and geminate stops was examined by regression analyses in 3- and 4-mora Japanese words spoken at various speaking rates. It was found that the perception boundary is well predicted by a linear function with duration of stop closure and durations of word or disyllable which contained the single and geminate stops. However, we conclude that the disyllable duration was a better variable than the word duration because it provides a more consistent explanation for the perception boundary regardless of word length and speaking rate variations. The results support a relational acoustic invariance theory.

Index Terms: speech perception, single and geminate stops, phoneme boundary

1. Introduction

There is an apparent mismatch between variable speech signals and native speakers’ uniform perceptions of linguistic units, such as phonemes. To account for this many-to-one mapping from speech signals to a phoneme, the acoustic invariance theory [1] was proposed. By extending the original acoustic invariance theory, Pickett et al. [2] proposed a relational acoustic invariance theory. They claimed that acoustic invariance exists not in an “absolute” form but in a “relational” form. For evidence of their theory, Pickett et al. [2] showed that, even though the duration of Italian single and geminate stops overlapped across different speaking rates, the durational ratio of the stop closure to the preceding vowel remained stable and distinguished the two stop categories. To find an invariant perceptual cue that is uniquely associated with a phoneme, Amano and Hirata [3] investigated the perception boundaries between single and geminate stops in disyllabic Japanese words. They created stimulus continua from 2-mora words with a single stop—(C\(_1\)V\(_1\).C\(_2\)V\(_2\))—to 3-mora words with a geminate stop—(C\(_1\)V\(_1\).Q.C\(_2\)V\(_2\) (where Q represents a geminate stop)—and measured participants’ response ratios in single vs. geminate stops. They found that the perception boundaries between single and geminate stops were represented in a linear function with an intercept defined by the dependent variable of stop closure duration and the independent variable of word duration. This finding supported the relational acoustic invariance theory.

However, Amano and Hirata [3] note that it is unclear whether their finding holds true for longer words such as 3-mora words with a single stop (C\(_1\)V\(_1\).C\(_2\)V\(_2\).C\(_3\)V\(_3\)) and 4-mora words with a geminate stop (C\(_1\)V\(_1\).Q.C\(_2\)V\(_2\).C\(_3\)V\(_3\)). They pointed out the possibility that the perception of the single vs. geminate distinction does not require the duration of the entire “word” but only that of the two syllables between which the single vs. geminate distinction occurs (i.e., the portion C\(_1\)V\(_1\).C\(_2\)V\(_2\) in the 3-mora word and the portion C\(_1\)V\(_1\).Q.C\(_2\)V\(_2\) in the 4-mora word). In other words, the duration of a disyllable, which includes single or geminate stops, might be the variable that has a relational invariant cue for the perception of these stops.

The aim of this study is to verify this possibility by investigating the perception of single and geminate stops in 3- and 4-mora Japanese words spoken at various speaking rates. The results of this study will help to provide evidence for or against the relational acoustic invariance theory [2].

2. Method

2.1. Participants

Forty monolingual native Japanese speakers (20 males and 20 females) participated in the experiment. Their average age was 24.9 years (Min. = 20, Max. = 30, S.D. = 0.50). They were paid for their participation.

Table 1: Properties of original materials from which stimulus continua were made.

<table>
<thead>
<tr>
<th>Original word</th>
<th>Quantity</th>
<th>Pitch accent type (L: low; H: high)</th>
<th>Meaning</th>
<th>Closure duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ikaku/</td>
<td>single</td>
<td>LHH</td>
<td>intimidation</td>
<td>94 46 31</td>
</tr>
<tr>
<td>/ikkaku/</td>
<td>geminate</td>
<td>LHHH</td>
<td>a corner</td>
<td>398 145 106</td>
</tr>
<tr>
<td>/itoku/</td>
<td>single</td>
<td>LHH</td>
<td>dignity and virtue one advantage</td>
<td>90 46 29</td>
</tr>
<tr>
<td>/ittoku/</td>
<td>geminate</td>
<td>LHHH</td>
<td>one advantage</td>
<td>476 137 93</td>
</tr>
<tr>
<td>/jitoku/</td>
<td>single</td>
<td>LHH</td>
<td>take by oneself ten virtues</td>
<td>80 36 8</td>
</tr>
<tr>
<td>/jittoku/</td>
<td>geminate</td>
<td>LHHH</td>
<td></td>
<td>481 156 94</td>
</tr>
<tr>
<td>/kaketsu/</td>
<td>single</td>
<td>LHH</td>
<td>approval</td>
<td>80 54 26</td>
</tr>
<tr>
<td>/akketsu/</td>
<td>geminate</td>
<td>LHHH</td>
<td>hemoptysis</td>
<td>380 151 87</td>
</tr>
</tbody>
</table>

Copyright © 2011 ISCA

2233 28 – 31 August 2011, Florence, Italy
2.2. Stimuli

The original material consisted of four Japanese word pairs in which one of the words contained a single stop and the other contained a geminate stop (Table 1). These were embedded in the carrier sentence *sorewa ___ dato omoimasu* ("I think that it is ___"). Each word pair had the same segmental composition and pitch-accent type except for the stop quantity. The word pairs were pronounced by a Japanese male speaker at the slow, normal, and fast speaking rates. They were digitally recorded with 16-bit quantization at a 48 kHz sampling frequency. There were a total of 24 tokens of original material (4 original word pairs x 2 stop quantities x 3 speaking rates).

The stimulus continua were generated by modifying the closure duration of each original word in 18 steps. That is, the closure duration in the original word was shortened or lengthened within its range with the step sizes shown in Table 2. The range and step size were decided after a preliminary listening by the authors, so that a stimulus at each end point of the continuum would be clearly perceived as either a single or a geminate stop. A total of 432 stimuli were created.

### 2.3. Procedure

Stimuli were presented diotically to participants through headphones at a comfortable level in a quiet room. The 432 stimuli were presented four times in a randomized order, and thus there were a total of 1,728 trials for each participant. The randomized order differed for each participant. The 1,728 trials were broken into eight blocks of 216 trials.

Upon hearing each stimulus, two response buttons were displayed on a computer screen. A word with a single stop or a geminate stop was shown on each button in Japanese hiragana orthography.

The participants’ task was to make a two-alternative forced choice between a single-stop and geminate-stop word by clicking the button that corresponded to the stimulus they heard. The participants were then asked to click a button marked “next” to hear the next stimulus.

Before the experiment, the participants were given two sets of 12 practice trials. The experiment was self-paced, but the computer prompted the participants to take a three-minute break between any two consecutive blocks. It took the participants between 100 and 130 minutes to complete the experiment.

### 2.4. Analysis

The response ratio to a geminate stop was calculated by dividing the number of geminate stop responses by the total number of responses for each stimulus. For each word at each speaking rate, a logistic function was fitted to the response ratio of the geminate stop as a dependent variable with closure duration as an independent variable (Figure 1). The closure duration at the perception boundary was obtained as the point that gave a 50% response ratio on the fitted logistic function (Table 3).

Regression analyses with a linear function, with and without an intercept term, were performed using the closure duration at the perception boundary (Table 3) as a dependent variable and one of the following durations as an independent variable, as follows:

1. word (*C*1*V*1.(Q).*C*2*V*2.*C*3*V*3),
2. disyllable including a closure (*C*1*V*1.(Q).*C*2*V*2),
3. two morae following a closure (*C*2*V*2.*C*3*V*3),
4. mora preceding a closure (*C*1*V*1),
5) consonant preceding a closure (C1),
6) vowel preceding a closure (V1),
7) first mora following a closure (C2V2),
8) first consonant following a closure (C2),
9) first vowel following a closure (V2),
10) second mora following a closure (C3V3),
11) second consonant following a closure (C3), and
12) second vowel following a closure (V3).

3. Results

Figure 2 shows the goodness of fit (Akaike’s information criterion, AIC, and Bayesian information criterion, BIC) for regression functions with each independent variable, with and without an intercept. For both AIC and BIC, values tended to be smaller when regression functions were with than without an intercept. Moreover, AIC and BIC values with an intercept were smallest when the independent variables were the word and the disyllable (AIC: 113.4 for word, 114.4 for disyllable; BIC: 115.7 for word, 116.8 for disyllable). These results indicate that both the word and the disyllable durations are good variables for predicting the perception boundary.

Figure 3 shows scattergrams of perception-boundary data when the regression functions had either word (upper panel) or disyllable duration (lower panel) as an independent variable. The regression function in the upper panel was $y = 0.244x - 25.4$, where $y$ is closure duration and $x$ is word duration. The regression function in the lower panel was $y = 0.371x - 29.3$, where $y$ is closure duration and $x$ is disyllable duration. The coefficients of determination were very high for these regression functions ($R^2 = .991$ for the upper panel, and $R^2 = .990$ for the lower panel), which indicates that both regressions were successful. These results show that closure duration and word/disyllable duration were the best variables to represent the perception boundary between single and geminate stops.

4. Discussion

The results of the regression analyses indicated that, when closure duration was a dependent variable, both word and disyllable durations were best suited among the 12 variables examined to be independent variables for the perception boundary between single and geminate stops in 3- and 4-mora words. That is, either of these variables can best represent the perception boundary from a statistical viewpoint.

Which variable is more plausibly used, along with the closure duration, for the perception of stop quantity distinction from a psychological viewpoint? We argue that the disyllable is more plausible than the word, for the following reasons.

First, the disyllable comes to its end earlier in time than the word where the speech waveform is sequentially processed from the word-initial position to the word-final position. In other words, the disyllable requires a shorter input time than the word. A shorter input time increases the chance that the perception process can make a faster decision about a
phoneme. This efficiency would be desirable for speech perception.

Second, the disyllable provides a more consistent explanation than the word about the perception process for all the 2-, 3-, and 4-mora words. Amano and Hirata [3] claimed that word duration is a good variable for the perception boundary between single and geminate stops in 2- and 3-mora words. However, because they used disyllabic words as stimuli, their claim is equivalent to saying that the disyllable duration is a good variable for distinguishing between single and geminate stops in 2- and 3-mora words. On the other hand, the current study provided direct evidence that disyllable duration is a good variable for 3- and 4-mora words. Therefore, if the disyllable is chosen as a variable, we can make a more general claim about the role of the disyllable duration for the perception of single and geminate stops in words that are 2 to 4 morae in length.

Moreover, this general claim is also supported by the result of a regression analysis using both data from Amano and Hirata’s study [3] and the present study. As shown in Figure 4, the regression analysis resulted in a very high coefficient of determination ($R^2 = .952$) for a regression function with an intercept when disyllable duration was used as an independent variable and closure duration was used as a dependent variable. This result indicates that perception boundaries were well defined when we used the two duration variables of the closure and disyllable, regardless of the speaking rate and word length.

Combining these results, we argue for the superiority of the disyllable to the word as a variable for perception boundary between single and geminate stops. We conclude that the perception process uses closure duration in relation to the disyllable duration to determine whether the stop is a singleton or a geminate in 2- to 4-mora words across different speaking rates. This means that the perception process does not need durational information about the entire word, but only needs the duration of two syllables.

In addition, we conclude that the perception process to distinguish single and geminate stops is well predicted by a linear function with the closure and disyllable durations. This linear function is regarded as a relational invariant cue for the perceptual distinction of Japanese stop quantity across speaking rates. Therefore, the findings of the current study provide support for the relational acoustic invariance theory [2].

Finally, we note a limitation of the present study, which should be addressed in the future. The present experiment used 3- and 4-mora words in which the second syllable was of the CV type only. However, in Japanese, there are other types of syllables such as CVN (where “N” represents a moraic nasal) and CV: (where “:” represents a lengthened vowel). Perception boundaries should be examined with these types of syllables as the second syllable in a stimulus word. That is, words with $C_iV_1{(Q)}C_iV_2N$ and $C_iV_1{(Q)}C_iV_2$: should be investigated to check the validity of disyllable duration for perceiving the boundary between single and geminate stops in these disyllable shapes. It is possible that only the duration of the first and second mora in a disyllable (i.e., $C_iV_1{(Q)}C_iV_2$) and not the duration of the entire disyllable contributes to the distinguishing of single and geminate stops. Future studies should examine this possibility to address the theoretical question of relational acoustic invariance.

5. Acknowledgements

This study was conducted using facilities and services at NTT Communication Science Laboratories, Kyoto, Japan. This study was supported by the Research Council of Colgate University.

6. References


Figure 4: Scattergram of perception-boundary data between single and geminate stops. Black and white markers show the data of for 3- and 4-mora words. Dark and light gray markers show the data of for 2- and 3-mora words in Amano and Hirata [3]. The legend indicates the speaking rates and original materials used for making the stimulus continuum. The number in parentheses indicates a stimulus’s length in mora. The solid line is a regression function ($y = 0.392x – 26.3$, where $y$ is the closure duration and $x$ is the disyllable duration).