Perception of the moraic obstruent /Q/: a cross-linguistic study

Makiko Sadakata[1], *Mizuki Shingai*[2], *Alex Brandmeyer*[1] *and Kaoru Sekiyama*[2]

[1]Donders institute for Brain, Cognition and Behaviour, Centre for Cognition, The Netherlands
[2] Division of Cognitive Psychology, Kumamoto University, Japan

m.sadakata@donders.ru.nl, m1_tsu_5@yahoo.co.jp, a.brandmeyer@donders.ru.nl
sekiyama@kumamoto-u.ac.jp

Abstract

Japanese natives segment speech into morae. The current study tested whether this extends to perception of geminate consonants: do Japanese natives rely on the moraic obstruent /Q/, and if so, which acoustic feature is perceived to characterize /Q/? Based on an informal interview, we hypothesized that Japanese natives rely on /Q/ that is represented as a silent duration. If so, it should be more difficult for them to distinguish 'geminate fricative consonants /ss/' and 'a silent duration plus singleton fricative consonant /_s/'. Two experiments with Japanese and Dutch natives compared discrimination and categorization accuracy of pseudo words including /ss/ and /_s/. Japanese natives discriminated them well while they poorly categorized them. Dutch natives performed both tasks relatively well. These results are in line with our hypothesis. This provides further support for the language specific listening.

Index Terms: geminate consonant, moraic obstruent, perception

1. Introduction

Geminate-singleton consonant contrasts occur in several languages, such as Finnish, Italian and Japanese. Geminate consonants (e.g., /kk/, /ss/) are often called *long-consonants* as opposed to singleton *short-consonants* (e.g., /k/, /s/) [1]. Indeed, the durational information provided by phonemes largely accounts for the acoustical characteristics of this contrast [2, 3]. and provides a crucial perceptual cue for native listeners [1, 4]. However, the terms '*long-*' and '*short-consonants'* may not be the best way to describe the abstract representations used by native listeners. For example, it has been noted that Japanese natives add an extra mora when *producing* geminate consonants [3], which is referred to as the *moraic obstruent /Q/* [5]. According to this perspective, Japanese natives would divide a geminate consonant into two parts (/Q/ plus a short consonant), rather than treating it as a single long part (a long consonant) while speaking. The current study investigates whether this also holds when Japanese natives *perceive* geminate consonants, and if so, which acoustic feature is perceived to characterize /Q/.

Geminated consonants are often produced by an abrupt suspension of articulator movements as well as by sustaining an oral closure or constriction during one mora length [5]. All these articulations result in a silent duration at the first part of a geminate consonant. One exception is the fricative consonants, which contain sustained frication during this part. Acoustically speaking, there are two kinds of geminate consonants: silent ones and fricative ones. Takahashi [6] reported the frequency of occurrence of these cases found in the Iwanami Japanese dictionary (5th ed). Among the 2,467 words that included geminates, 76 % were silent geminates and 24% were fricative geminates. This suggests that the majority of geminate consonants in Japanese are likely to include a silent duration when produced.

Interestingly, not many Japanese natives are aware that there are two types of geminate consonants: they often falsely believe that there is a silent duration when they produce and perceive fricative geminate consonants just as in non-fricative geminate consonants (informal interviews). This observation seems to indicate that Japanese natives are relying on an abstract representation of a silent duration when producing and perceiving any geminate consonants, including fricative ones. If this is the case, for Japanese ears, "a silence followed by a fricative singleton consonant" should sound very similar to "a fricative geminate consonant". Such a finding would provide empirical evidence that Japanese natives rely on the abstract representation of the moraic obstruent /Q/ that is characterized by a silent duration when *perceiving* geminate consonants.

In order to test this hypothesis, we carried out two cross-linguistic experiments with Japanese and Dutch native listeners. Each experiment consisted of discrimination and categorization tasks using pseudo words that contained either 'a geminate fricative consonant /ss/' or 'a silent duration and a fricative consonant /_s/'. We predicted that all listeners would be able to discriminate /ss/ from /_s/. This would indicate that they could hear a silence embedded in speech. Additionally, we predicted that these two types of stimuli would be confusing for Japanese natives in the categorization task. This means that both /ss/ and /_s/ are represented in a similar way for Japanese natives. Finally, we predict a different response pattern for Dutch natives from Japanese natives: /ss/ and /_s/ should be perceived as distinct, as the Dutch language does not make use of the same abstract representation of /Q/ as Japanese.

2. Experiment 1

2.1. Participants

Ten native Japanese speakers recruited from the Kyoto-city University of Arts (music students) and ten native Dutch speakers recruited from the participant pool of the Max Planck Institute for Psycholinguistics took part in the study. The Dutch natives all completed multiple training sessions in order to learn to perceive Japanese geminate-singleton contrasts prior to the experiment [7].

2.2. Stimuli

Figure 1a provides acoustic information of the stimuli. The three pseudo words were synthesized from natural speech recorded by a female speaker, /asu/ (/a/=90ms, /s/=150ms, /u/=90ms), /assu/ (/a/=90ms, /ss/=240ms, /u/=90ms), and /a_su/ (/a/=90ms, /_/=90 ms, /s/=150ms, /u/=90ms) using the application Praat [8]. The pitch accent patterns of two vowels were high-low. For both Experiment 1 & 2, all recordings were first low-pass filtered at 5000 Hz and average sound levels were normalized to 70 dB.

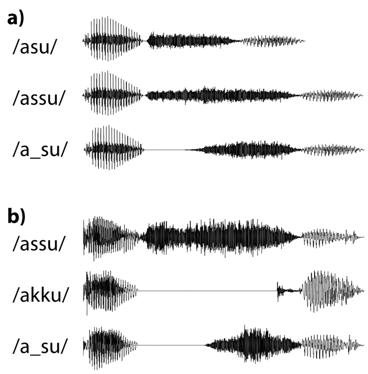

Figure 1: *Waveforms of stimuli for Experiment 1 (a) and Experiment 2 (b). Experiment 2 used multiple token words and multiple voices.*

2.3. Procedure

2.3.1 Discrimination test

A 4 interval 2 alternative forced choice (4I2AFC) task was used for the discrimination test, presenting /assu/ as standard and /a_su/ or /asu/ as stimuli to be discriminated (Deviant: /a_su/, Distracter: /asu/) [9]. Each trial included presentation of four words (Inter-stimulus interval ISI=500 ms). The task was to discriminate a deviant stimulus that was presented in either the second or the third position of the sequence. Each deviant stimulus was presented 40 times, resulted in 80 trials in total.

2.3.2 Categorization test

The categorization test consisted of a label-learning task and a 3 alternative forced choice (3AFC) task. During the label-learning task, participants were presented with six repetitions of three example categories, /a_su/-/assu/-/asu/. Each sound was presented along with a visual number (1, 2 or 3), which in turn was associated with a labeled key on the keyboard (ISI=2000ms). During the 3 AFC task, one of the three stimuli was presented per trial (90 trials) and participants pressed the key 1, 2 or 3 to indicate their judgments.

2.4. Results

2.4.1 Discrimination test

The low error rates shown in Figure 2 indicate relatively high discrimination accuracy. A two-way ANOVA with Native-language (JP and NL) and Condition (Deviant and Distracter) indicated no significant effects (Native-language $F(1,17)=0.98$, n.s., Condition $F(1, 17)=1.45$, n.s., interaction $F(1,17)=3.26$, n.s.). This suggests that all participants correctly discriminated /a_su/ from /assu/ and /asu/ from /assu/.

2.4.2 Categorization test

Figure 3 presents the error rates from the categorization test. A mixed model ANOVA with Native-language (JP and NL) and Condition (/a_su/, /assu/, /asu/) indicated no significant main effects (Native-language $F(1,18)=.074$, n.s., Condition $F(2,36)=3.14$, n.s.) but significant interaction between them ($F(2,36)=8.90$, $p=.003$). Further analysis indicated simple main effects indicating that Japanese showed significantly higher error rates for /a_su/ and /assu/ as compared to /asu/ ($p<.05$ and $p<.01$, respectively). The main source for the /a_su/ error (94%) was due to their miss-categorization of /a_su/ as /assu/. Likewise, the majority of /assu/ error (87%) was due to the miscategorization of /assu/ as /a_su/. There were no significant differences among stimuli types of error rates by Dutch natives.

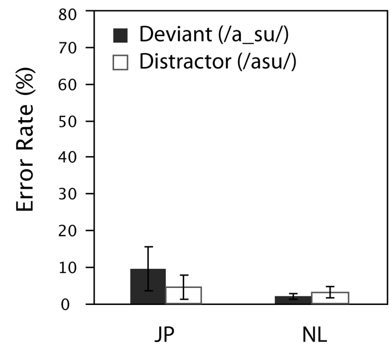

Figure 2: *Mean error rates for the discrimination test for 2 conditions by Japanese and Dutch participants.*

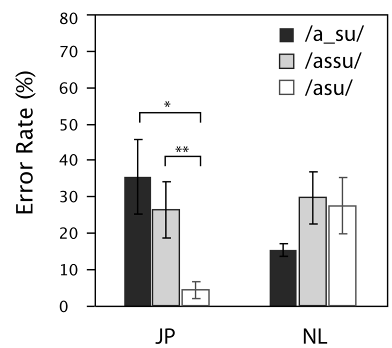

Figure 3: *Mean error rates for the categorization test for 3 conditions by Japanese and Dutch participants. (**p<.01, *p<.05)*

2.5. Discussion

The results of the two tests confirmed our hypotheses to some extent. Firstly, all participants were able to discriminate /a_su/ from /assu/, indicating that they were able to detect a difference between these two sounds due to the embedded silence in speech. Secondly, Japanese natives showed higher error rates when categorizing /a_su/ and /assu/ than /asu/, and these errors were mainly due to the confusion between these two stimuli. Finally, Dutch natives did not show the same pattern of results in the categorization task as Japanese natives did.

There are number of issues to be addressed, however. First, the particular way in which /a_su/ stimulus was composed may have influenced the results. It is therefore useful to test whether this result pattern holds when tested with more variable stimuli. Second, the task difficulty may also have contributed to the relatively higher error rates of /a_su/ and /assu/ for Japanese natives. The 3 AFC categorization test is not easy to perform, especially with limited experience in an experimental setting. Using an easier task would be desirable when further testing this effect. Third, all Dutch participants took part in 5-day training to learn to perceive /assu/ and /asu/, which also administered the 4I2AFC and 3AFC task. Such a difference in experience with these tasks may account for the different result patterns, in particular the relatively low error rates for categorization of /a_su/ by Dutch participants: the sound of /a_su/ may have been easy to categorize for them because it was a new type of sound (as compared to /assu/ and /asu/ that they have been trained on). Finally, all Japanese participants were musicians. Musicians tend to outperform non-musicians when performing speech perception tasks [10]. In our experiment, Japanese participants did not perform better at any tasks as compared to the Dutch non-musicians, indicating that the known benefit of musical experience was not confirmed. Nevertheless, the degree of musical experience should be controlled for.

3. Experiment 2

Experiment 2 addressed the above-mentioned issues. First, the stimuli included multiple voices, tokens, and words in order to increase the variability. Second, a less demanding categorization task [11] was employed. Lastly, we controlled for participants' prior exposure to the experimental procedure as well as for musical experience.

As in Experiment 1, we predict that all participants will perform the discrimination task well. Predictions for categorization test are summarized in Table 2. Critically, we predict that Japanese natives will categorize /assu/ and /a_su/ in the same group but Dutch natives will not.

3.1. Participants

Two groups of sixteen native speakers of Dutch and Japanese took part in the two tests included in this experiment. Japanese participants were students of the University of Kumamoto, while Dutch participants were students of the Radboud University of Nijmegen. For both groups, half were individuals who had followed more than 5 years of formal musical training and who were still actively playing musical instruments at the time of the experiment (musicians). Another half were non-musicians, which was defined as individuals who completed less than 3 years of musical training and had followed training other than music (mostly sports) for at least 5 years.

3.2. Stimuli and design

Figure 1b provides acoustic details on the stimuli. The three types of stimuli were constructed with the following parameters: V1-/_s/-V2 (e.g.,/a/=90ms, /_/=100ms, /s/=160ms, /u/=100ms), V1-/ss/-V2 (/a/=90ms, /ss/=260ms, /u/=100ms), and V1-/kk/-V2 (/a/=90ms, /_/=220ms, /k/=40ms, /u/=100ms). Six combinations of V1 and V2 were chosen in order to compose 18 pseudo words (Table 1). V1 and V2 were always different in order to increase variability. Stimuli consisted of six female voices and one male voice. The pitch accent patterns of two vowels were high-low. These voices were recorded by Japanese native spekers, and then used to synthesize the stimuli using Praat [8].

3.3. Procedure

3.3.1 Discrimination test

The same 4I2AFC task was used as in Experiment 1, this time presenting /assu/ as standard and /a_su/ or /akku/ as stimuli to be discriminated (Deviant: /a_su/, Distracter: /akku/). Four different female voices were randomly chosen from the six voices per trial.

2.3.2 Categorization test

The same categorization experiment paradigm as Dehaene-Lanbertz, Dupoux, & Gout was used [11]. Participants were presented five words in each trial (ISI=500ms). The first four items (standard) were either /assu/ or /a_su/, presented by four different female voices that were randomly chosen. The fifth item (test) was either /assu/, /a_su/ or /akku/, spoken by a male voice. These served as control, deviant and distracter conditions, respectively (see Table 2). Participants were asked to judge whether the test word was the same as the standard items.

Table 1. *Pseudo words used for Experiment 2*

V_1-/_s/-V_2	a_su, i_se, u_sa, u_se, o_sa, o_su
V_1-/ss/-V_2	assu, isse, ussa, usse, ossa, ossu
V_1-/kk/-V_2	akku, ikke, ukka, ukke, okka, okku

Table 2. *Experiment conditions and predictions for two groups.*

Condition	Standard	Test	JP	NL
Control	assu	assu	Same	Same
	a_su	a_su		
Deviant	assu	a_su	Same	Different
	a_su	assu		
Distracter	assu	akku	Different	Different
	a_su	akku		

3.4. Results

3.4.1 Discrimination test

A three-way ANOVA with Condition (Deviant or Distracter) as a within subject factor and Native-language (JP and NL) and Musical experience (musicians and non-musicians) as between-subject factors indicated a significant effect of Condition (F(1,18)=0.16.62, p<.0003) with higher error rate for the Deviant than the Distracter stimuli. No other effects were significant (native-language F(1, 28)=1.92, n.s., Musical experience F(1,28)=1.93, n.s.). Error rates for the discrimination test are shown in Figure 4. Results for the two musical groups are plotted together as they did not differ significantly. Taken together, this indicates that the deviant stimuli were more difficult to discriminate from /assu/ than the distracter stimuli for all participants. Nevertheless, the error rates for the deviant condition were lower than 10% (JP 9.3%, NL 9.6%), suggesting that participants performed the task relatively well.

3.4.2 Categorization test

Figure 5 shows participant error rates from the categorization test. Again, the two musical groups are plotted together because they did not differ significantly. A mixed model ANOVA with Condition (Deviant, Control, Distracter) as a within-subject factor and native-language (JP and NL) and Musical experience (musicians and non-musicians) as between-subjects factors indicated significant main effects of Condition (F(2,27)=31.03, p<.0001) and significant interaction between Condition and Native-language (F(2,27)=5.81, p<.008). No significant main effects of Native-language (F(1,28)=3.23, p<.08) or Musical experience were observed (F(1,28)=0.94, p<.34) nor were any other interaction effects. Further analyses indicated several significant simple effects, which are indicated in Figure 5. Most critically, Japanese natives showed significantly higher error rates for Deviant stimuli than Dutch natives did (p <.05).

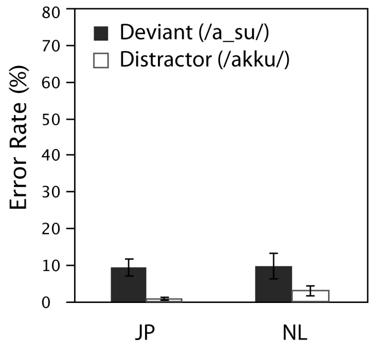

Figure 4: *Mean error rates for the discrimination test for 2 conditions by Japanese and Dutch participants.*

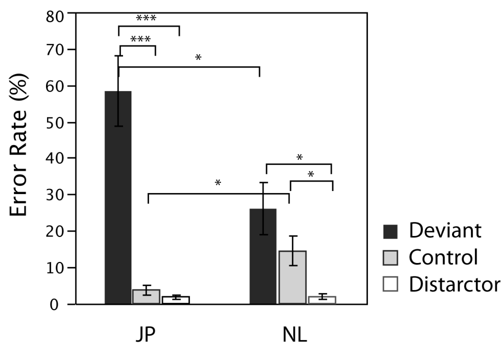

Figure 5: *Mean error rates for the categorization test for 3 conditions by Japanese and Dutch participants. (***p<.0001, **p<.01, *p<.05)*

4. General discussion

Two series of experiments suggested that "a silence followed by a fricative singleton consonant /_s/" and "a fricative geminate consonant /ss/" are perceptually similar for Japanese native ears: while it was easy to discriminate /a_su/ from /assu/, it was more confusing when categorizing these two. The results correspond to our observation, namely, that Japanese natives are often not aware of sustained frication included in fricative geminate consonants. Taken together, these results support the hypothesis that Japanese natives seem to rely on the abstract representation of moraic obstruent /Q/ when perceiving geminates, which is perceptually characterized as a silent duration.

The pattern of results was consistent for both experiment 1 and 2 thus the effect was robust. The effect persisted when stimuli were extended to multiple words, tokens and voices, as well as when the difficulty of the task was reduced and when musicians were tested. Although musicians are known to outperform non-musicians when tested on speech perception tasks [10], no such benefit of musical experience was observed.

Importantly, the result patterns of categorization tests were different between two native language groups. Dutch natives were consistently better at categorizing /a_su/ as different category from /assu/ than the Japanese counterpart. Most likely, these two sounded much less similar to them because Dutch natives do not rely on the abstract representation of /Q/.

5. Conclusion

Previous studies have demonstrated that Japanese natives segment speech into morae [12, 13]. The results of the current study suggested that this extends to perception of geminate consonants. The results of two experiments supported the hypothesis that Japanese natives rely on the abstract representation of /Q/ that is characterized as a silent duration when perceiving geminate consonants. This provides further support for language-specific listening hypotheses [12].

6. Acknowledgements

This research was supported by a Grant-in-Aid for Scientific Research (B) 21330169 under the project leadership of Shuji Mori (Japan Society for the Promotion of Science). The authors are grateful for useful suggestions by Anne Cutler, and for help in setting up the first experiment in Japan by Minoru Tsuzaki.

7. References

[1] Kingston, J., Kawahara, S. Chambless, D., and Brenner-Alsop, E., "Contextual effects on the perception of duration", Journal of Phonetics, 37(3): 297-320, 2009.
[2] Amano, S. and Y. Hirata, "Perception and production boundaries between single and geminate stops in Japanese", Journal of Acoustical Society of America, 128(4): 2049-2058, 2010.
[3] Han, M. S., "Acoustic manifestations of mora timing in Japanese", The Journal of Acoustical Society of America, 96(1): 73-82, 1994.
[4] Tagliapietra, L. and McQueen, J. M., "What and where in speech recognition: Geminates and singletons in spoken Italian", Journal of Memory and Language, 63: 306-323, 2010.
[5] Vance, T. J., "An introduction to Japanese phonology", State University of New York Press, 1987.
[6] Takahashi, W., "Nihongo no Sokuon /Q/ ni tsuite [On Japanese consonant /Q/]", Journal of the Faculty of Education, Shinshu University, 95: 59-68, 1998.
[7] Sadakata, M. and McQueen, J. M., "The role of variability in non-native perceptual learning of a Japanese geminate-singleton fricative contrast", Proceeindgs of Interspeech 2011, 873-876, 2011.
[8] Boersma, P. and Weenink, D., "Praat, a system for doing phonetics by computer", Glot International, 5(9-10): 341-345, 2001.
[9] Gerrits, E. and Schouten M. E. H., "Categorical perception depends on the discrimination task", Perception & Psychophysics, 66(3): 363-376, 2004.
[10] Sadakata, M. and Sekiyama, K., "Enhanced perception of various linguistic features by musicians: a cross-linguistic study", Acta Psychologica, 138: 1-10, 2011.
[11] Dehaene-Lambertz, G., Dupoux, E., and Gout, A., "Electrophysiological correlates of phonological processing: A cross-linguistic study", Journal of Cognitive Neuroscience, 12(4): 635-647, 2000.
[12] Cutler, A. and Otake, T. "Mora or phoneme? Further evidence for language-specific listening", Journal of memory and language, 33: 824-844, 1994.
[13] Otake, T., Hatano, G., Cutler, A. and Mehler, J., "Mora or Syllable - Speech segmentation in Japanese", Journal of Memory and Language, 32(2): 258-278, 1993.