The effect of distal speech rate on the Japanese singleton-geminate distinction

Tim Gadanidis

March 8, 2019

1 Introduction

To date, the literature on the Japanese singleton-geminate distinction has focused on local context—specifically, the preceding and following vowels. Several authors have suggested, based on both production and perception data, that the preceding and following vowel are all that is needed for Japanese speakers to accurately classify stops as either singleton or geminate.

For example, in a production study, Hirata and Whiton (2005) found that while Japanese singletons and geminates (embedded in bi- or trimoraic words) overlap across speech rates, the ratio of closure duration to word duration optimally classified them 95.7%–98%. The authors interpret this finding as support for relational acoustic invariance, such that all that is needed to differentiate stops and geminates is the preceding and following syllables.

On the perception side, Amano and Hirata (2010) found that the ratio of closure duration to word duration optimally predicted participants' responses in a perception study. The authors interpret this as potential evidence that "we do not need information about the entire "word" duration, but [...] we only need the duration of two syllables between which the quantity distinction occurs" (p. 2057). Idemaru and Guion-Anderson (2010) conducted a similar perception study and found that while acoustically the best measure was stop-to-word ratio, speakers actually made more use of the stop-to-previous mora ratio.

While these studies (among others) all indicate that preceding (and sometimes following) syllables/moras play an important role in the Japanese singleton-geminate distinction, no study on this phenomenon (that I am aware of) has examined units larger than a single word, and none has manipulated *distal* context[1]. However, Mitterer (2018) has shown that the perceptual category boundary between Maltese singleton and geminates is affected by distal speech rate (defined as greater than one syllable away from the target contrast), indicating that Maltese listeners employ rate normalization.

This study expands on Mitterer's (2018) findings by manipulating the duration of distal context (i.e., carrier sentences) in order to see whether Japanese listeners perform distal-context rate adaptation. If Japanese listeners make use of distal context rate adaptation, they will be more likely to classify a stop as *geminate* when it is embedded in a fast carrier sentence than a slow carrier sentence due to boundary shift, thus indicating that despite the existence of apparently invariant measures such as closure-to-word ratio, distal speech rate still provides useful cues for listeners. If they do not make use of distal context, it could be considered evidence that in the presence of relational acoustic invariance, redundant cues such as distal speech rate are not necessary and thus unused by Japanese listeners.

2 Methods

2.1 Participants

Participants have yet to be recruited. Participants will be native speakers of Japanese with no diagnosed speech, hearing or (uncorrected) vision impairments. As many participants will be recruited as possible, (though that may not be very many).

2.2 Materials

A male native speaker of Japanese produced each of the words listed in Table 1, which are the same stimuli as were used by Amano and Hirata (2010), twice. In each production, the target word was embedded in the carrier sentence 〈その単語は __ と読みます〉 /sono taŋgo wa _ to jomimasɯ̥/ 'That word is read _'. This sentence is slightly longer than the carrier used by Hirata and Whiton (2005), Amano and Hirata (2010), and Idemaru and Guion-Anderson (2010)[2]. This was done to give participants more of an opportunity to adapt to distal context, if indeed they made use of it.

The second utterance of each word was selected for manipulation, except in cases where the speaker made an error in the second utterance, in which case the first was used. The stimuli were manipulated using Praat's PSOLA algorithm (Boersma & Weenink, 2017).

[1] Hirata and Lambacher (2004) found that surrounding speech rate had an effect on the Japanese *vowel* quantity distinction, but as Mitterer (2018) notes, the targets in that study were separated from the surrounding rate by one consonant rather than a full syllable. This is typically not long enough to be considered distal (cf. Heffner, Newman, & Idsardi, 2017).

[2] /soko wa _ to jomimasɯ̥/

Word	Quantity	Translation
nita	short	boil+PST
nitta	long	Nitta (surname)
tate	short	height
tatte	long	stand up
ika	short	less
ikka	long	one family
kako	short	past
kakko	long	parentheses

Table 1: Stimuli

Nine-step durational continua were made from each stimulus by lengthening or shortening the closure using a modified version of a script originally developed by Yoonjung Kang. The closure was manually marked as the interval after the end of all visible vowel formants and before the beginning of the following consonant release. The first step of the closure continuum is 20ms and the final step is 200ms. In a pilot production study based on one speaker, I found that in actual production at slow rates, geminate consonants can actually have closures exceeding 400ms, but stimuli as short as 200 or so ms are still unambiguously geminates, so there is no need to extend the continua all the way to 400ms.

At each step of each continua, the script generated three stimuli, one in which the carrier sentence is unmodified, one in which the carrier sentence is 20% longer (i.e., slower) and one in which the carrier sentence is 20% shorter (i.e., faster). The choice to use 20% was following Mitterer (2018).

Following Amano and Hirata (2010), in order to counterbalance effects of cues other than duration, stimuli were generated from both the singleton and geminate tokens. The PSOLA algorithm does not perform well when asked to elongate short singleton closures, so in order to create singleton-derived stimuli, part of the closure from the equivalent geminate stop was spliced in, and the resulting combined stop was then modified.

After these manipulations, the design features a total of 216 stimuli: 9 durational steps × 3 carrier rates × 8 words = 216 stimuli.

2.3 Procedure

The experiment will take place in a quiet room in front of a computer screen. The experiment will be implemented using PsychoPy Peirce (2007). Like Mitterer (2018), the answer options will be presented on the screen as full words, written in whichever of the various Japanese character sets is customary for that word in everyday life. For example, the word *kako* 'past' is written using a sequence of two kanji, ⟨過去⟩, while the word *nita* 'boil' is written using a mixture of kanji and hiragana, ⟨煮た⟩. The singleton option will be presented on the left side of the screen, and the geminate option on the right; participants will indicate their choice by pressing either the left or right arrow key accordingly. After the participant makes their choice, the next stimulus will be presented (i.e., the experiment will be self-paced). After every 54[3] trials, participants will have the opportunity to take a short break.

2.4 Analysis

The methodology that I use for analysis will depend on how many participants are recruited. If there are enough participants to enable the use of inferential statistics, the results will be analyzed using generalized logistic regression as implemented by the function `glmer()` in the R (R Core Team, 2018) package `lme4` (Bates, Mächler, Bolker, & Walker, 2015), using a binomial linking function. If not, I will make use of only descriptive statistics for the time being, using the package `ggplot2` (Wickham & Chang, 2008).

3 Results

TBD

4 Discussion

TBD

[3] 216 ÷ 4 = 54