

# CAN JAPANESE LISTENERS PERCEIVE A DIFFERENCE BETWEEN AN UNDERLYING GLIDE AND AN EPENTHETIC GLIDE?

Haruka Tada, Isabelle Grenon

The University of Tokyo  
harukatada0111@g.ecc.u-tokyo.ac.jp, grenon@boz.c.u-tokyo.ac.jp

## ABSTRACT

The word /kijo:/ ‘appointment [to a position]’ in Japanese contains an underlying glide while the word /kio:/ ‘past illness’ contains an epenthetic glide, both surfacing as a second formant (F2) transition following the steady state of the preceding vowel. This study evaluates whether Japanese listeners can perceptually distinguish the presence of an underlying glide from an epenthetic glide, which would be a case of incomplete neutralization. For this investigation, we manipulated the duration of the steady state of the /i/ vowel, and the duration of the following F2 transition. Eighteen Japanese listeners completed a two-alternative forced-choice identification task with the manipulated stimuli. The results suggest that Japanese listeners rely neither on the length of the F2 steady state nor on that of the F2 transition to distinguish /kio:/ from /kijo:/ in perception. Hence, these sub-phonemic cues do not appear to contrast an underlying glide from an epenthetic glide in Japanese.

**Keywords:** Speech perception, glide, Japanese, incomplete neutralization.

## 1. INTRODUCTION

There are a number of words in English that are pronounced the same, called homophones. Some words, however, may only appear to be homophonous. A famous example is the contrast between ‘writer’ and ‘rider’ in American English. Although the two words are almost pronounced the same due to the flapping of intervocalic /t/ and /d/, the duration of the vowel preceding the underlying /d/ is generally longer than the vowel preceding the underlying /t/ [9, 13]. Not only the words can be distinguished in production by some sub-phonemic cue (vowel duration), but also American English listeners are sensitive to this cue in order to distinguish the two words in perception [8].

The Japanese language also has words that are potentially homophonous, such as /kio:/ ‘past illness’ and /kijo:/ ‘appointment (to a position)’. The word /kijo:/ contains an underlying glide /j/ that can be measured as a second formant (F2) transition.

While /kio:/ does not contain any underlying glide, an epenthetic glide may be inserted in the sequence /io/ [10], which is similarly realized as an F2 transition. While the words can be distinguished through context, it is possible that some acoustic cues may be used by native Japanese listeners to distinguish the presence of an underlying glide from an epenthetic glide, which would be a case of incomplete neutralization, as discussed in the next section.

An F2 steady state of less than 50 ms is perceived in Japanese as the presence of a glide whereas a steady state of more than 50 ms is generally perceived as the presence of a vowel [5]. Hence, the length of the steady state may play a role in the perception of /kio:/ and /kijo:/ in Japanese, with a longer steady state leading to the perception of /kio:/ (i.e., to a sequence without underlying glide). Furthermore, a distinction in the shape of the F2 transition was observed between sequences of two vowels versus diphthongs in Spanish [1], and between glide transition and diphthongs in Sindhi [6]. Accordingly, it is possible that the shape (i.e., length) of the F2 transition also plays a role in distinguishing the sequence /io/ from /ijo/ in Japanese. Hence, the current study evaluated Japanese listeners’ sensitivity to the duration of the F2 steady state, and the duration of the F2 transition for distinguishing the words /kio:/ and /kijo:/.

## 2. INCOMPLETE NEUTRALIZATION

Incomplete neutralization is when two nearly identical words are contrasted above chance level in production and perception. An example of incomplete neutralization is the case of word-final devoicing in German, where the underlying voiced stop /d/ can be distinguished both in production and perception from its underlying voiceless counterpart /t/ [7]. Another possible example of incomplete neutralization is the case of the English words ‘prints’ and ‘prince’ where the closure of the epenthetic /t/ in the /ns/ sequence of ‘prince’ has been documented to be produced with a shorter duration than the closure of the underlying /t/ in ‘prints’ [4]. It is questionable whether underlying segments consistently differ at the sub-phonemic

level from their epenthetic counterpart, and whether listeners are sensitive to such differences. Hence, the current study investigated whether incomplete neutralization may occur in the *perception* (only) of an underlying versus epenthetic glide in Japanese.

## 2. METHOD

### 2.1. Participants

Eighteen native Japanese speakers (14 females) participated in this experiment. They were aged 19 to 52 years old ( $M = 26$ ). All participants were from the Kanto region (around Tokyo), and reported no history of hearing or speech impairment. None of the participants reported to speak fluently any other language than Japanese, or to have been abroad for more than 12 weeks. They received a monetary compensation for their participation.

### 2.2 Stimuli

#### 2.2.1 Recorded stimuli

In order to decide the duration values of the F2 steady state and the F2 transition for the stimuli, we looked at /kio:/ and /kijo:/ samples produced by two female Japanese speakers from the Kanto area. Table 1 summarizes the duration characteristics of the F2 steady state and the F2 transition from a total ten /kio:/ and ten /kijo:/ samples (each speaker produced five of each). The data in Table 1 suggest an overlap in both the duration of the F2 steady and of the F2 transition between the presence of an epenthetic glide and that of an underlying glide in the samples collected. The current experiment, however, is only concerned with whether the words are contrasted in *perception*.

**Table 1:** The duration characteristics of the F2 steady state (top) and F2 transition (bottom) in ten recorded samples each of /kio:/ and /kijo:/ by two native Japanese speakers.

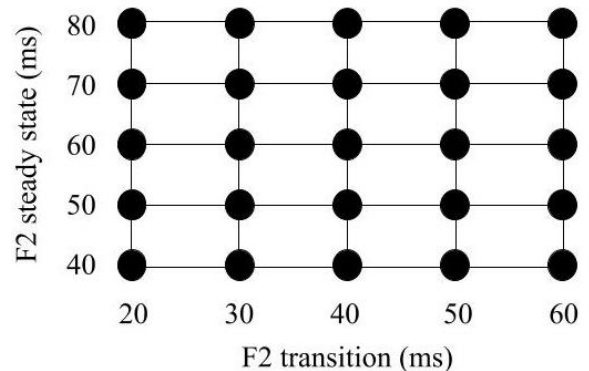
F2 steady state in milliseconds	/kio:/	/kijo:/
min	34	36.4
max	60.5	78.8
average	48.31	59.05
SD	9.08	15.9

F2 transition in milliseconds	/kio:/	/kijo:/
min	18.1	35.9
max	73.2	74.3
average	48.3	51.17
SD	18.2	11.5

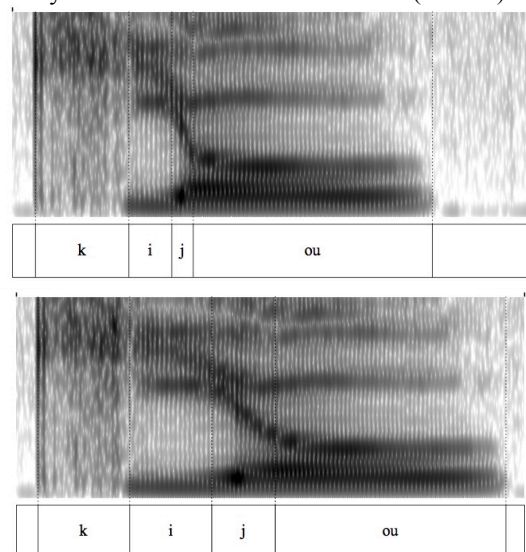
#### 2.2.2 Manipulations

Twenty-five stimuli were created from a /kijo:/ token produced by a 22 years old female Japanese speaker from the Kanto area. The duration of the F2 steady state and F2 transition were manipulated in Praat [3] using a script [11]. First, the F2 steady state was manipulated to vary from 40 ms to 80 ms in 5 equal steps of 10 ms. Then, the F2 transition was varied from 20 ms to 60 ms in 5 steps of 10 ms. The 25 manipulated stimuli are schematized in Figure 1 below, with spectrograms of 2 of the resulting stimuli presented in Figure 2. The duration of the initial stop release (89 ms), and the duration of the steady state of the word-final vowel (225 ms) were kept constant across all 25 test tokens. The resulting /kio:-/kijo:/ stimuli varied in total length from 374 ms to 454 ms.

**Figure 1:** The 25 manipulated test stimuli were varied in terms of duration of the F2 transition (x-axis) and duration of the F2 steady state (y-axis).



**Figure 2:** Spectrogram of the token with 40 ms of F2 steady state and 20 ms of F2 transition (top), and spectrogram of the token with 80 ms of F2 steady state and 60 ms of F2 transition (bottom).



### 2.3. Procedure

The perception experiment was done in a quiet room with the same Audio-Technica ATH-S100 headphones and using the same Macbook Air laptop computer. The task consisted of a two-alternative forced-choice identification task. The Japanese participants were instructed (in Japanese) to listen to the stimuli presented one at a time and to decide whether each word they heard was /kio:/ or /kijo:/ by pressing the appropriate key on the computer keyboard (the choices that appeared on the computer screen were written in Japanese: 既往 for /kio:/ –起用 for /kijo:/). The stimuli were presented randomly 4 times (4 times x 25 stimuli = 100 test tokens). The first round of 25 stimuli was considered a practice session and discarded from the analyses. Stimulus presentation and data collection were controlled using ExperimentMFC in Praat [2]. The experiment lasted approximately 10 minutes with no break.

### 3. RESULTS AND DISCUSSION

The research question addressed by this study was whether Japanese listeners could perceive a difference between an underlying glide and an epenthetic glide as in the words /kijo:/ and /kio:/ respectively. We looked at the use of two potentially relevant cues that were shown cross-linguistically to play a role in speech perception: The duration of the F2 steady state and the duration of the F2 transition. Figure 3 presents the percentage of /kio:/ responses as a function of the duration of the F2 steady state. The roller-coaster pattern suggests that this cue is not used systematically by native Japanese listeners to contrast the presence of an epenthetic glide from the presence of an underlying glide in perception.

**Figure 3:** Percentage of /kio:/ responses per change in the duration of the F2 steady state.

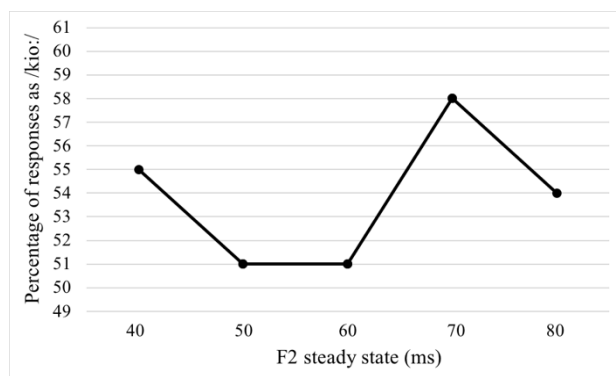
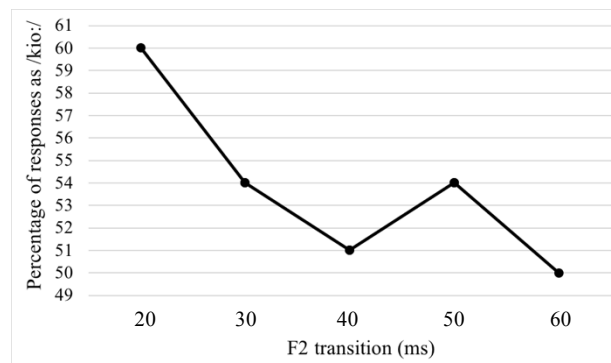


Figure 4 shows the percentage of /kio:/ responses as a function of the duration of the F2 transition. Although there is a trend towards a decrease in /kio:/

responses as the duration of the F2 transition increases, this tendency was not statistically significant.

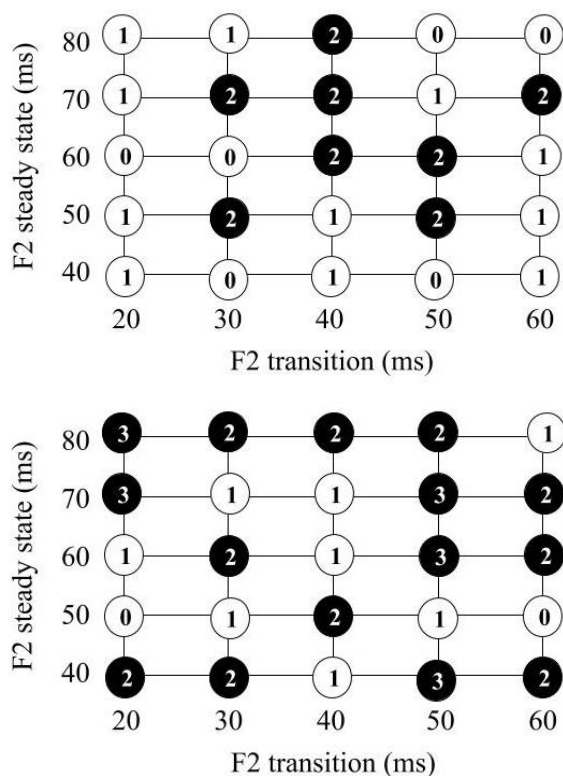
**Figure 4:** Percentage of /kio:/ responses per change in the duration of the F2 transition.



Generalized Linear Mixed-Effects Model with F2 steady state and F2 transition as the independent variables and the presence of /kio:/ responses as the dependent variable (where a /kio:/ response was coded as 1 and a /kijo:/ response as 0) confirmed no main effect of F2 steady state ( $\beta=-0.04878$ ,  $z=-0.509$ ,  $p=0.6106$ ), nor a significant effect of F2 transition ( $\beta=-0.16066$ ,  $z=-1.682$ ,  $p=0.0925$ ) or any effect of the interaction between F2 steady state and F2 transition ( $\beta=0.02385$ ,  $z=0.829$ ,  $p=0.4073$ ) in the current study. Hence, no statistical support for a perceptual difference between an underlying glide and an epenthetic glide in Japanese was found on the average data.

The responses of each individual participant were compiled separately and the identification patterns were visually inspected as a preliminary investigation into individual data. The purpose of this investigation was mainly to rule out the possibility that some individuals would show a systematic pattern, but that this pattern was canceled out in the overall data by other individuals showing, for instance, the opposite pattern. Figure 5 illustrates two representative set of results, one by a male listener (top) and one by a female listener (bottom). The number of responses out of three trials for each token is indicated in each circle, and the circles are presented in black when the number of /kio:/ responses for a token is 2 or 3 out of 3 (that is, the token is identified mainly as /kio:/). As illustrated with the examples in Figure 5, this pilot investigation suggests no clear pattern in any of the individual data either, though this would understandably need to be confirmed with a more thorough investigation (which was beyond the scope of the current study).

**Figure 5:** Sample of individual response pattern for a male (top) and a female (bottom) Japanese listener. A black circle indicates that this token was generally perceived as /kio:/. The number in each circle represent the number of times this token was identified as /kio:/ out of 3 trials.



In the current experiment, the presence of an underlying glide in the Japanese word /kijo:/ was not found statistically distinguishable in perception from the presence of an epenthetic glide in the word /kio:/ in terms of the duration of the F2 steady state or the duration of the F2 transition. These results suggest that listeners are not necessarily sensitive to sub-phonemic cues to contrast an underlying segment from an epenthetic segment. However, it is too early to conclude that this contrast is entirely neutralized, as it still could be a case of near merger [e.g.,12] where the words are contrasted in production but not in perception. Also, a further investigation that looks not only at the cues investigated in the current study, but also at the duration of the vocalic interval (including the following vowel), the relative duration of the entire word, and the relative degree of oral construction, among possible correlates, may shed further light on this issue, and confirm whether the presence of the underlying glide contrasts in any way from an epenthetic glide in either perception or production.

#### 4. CONCLUSION

The current study investigated whether the words /kio:/ and /kijo:/ in Japanese may represent a case of incomplete neutralization, where the words are distinguished in perception through sub-phonemic cues. While /kijo:/ contains an underlying glide, the word /kio:/ features an epenthetic glide, both of which are realized acoustically as an F2 steady state of the /i/ vowel followed by an F2 transition into the vowel /o/. We hypothesized that Japanese listeners may be able to rely on either the duration of the F2 steady state or the duration of the F2 transition in order to differentiate the target words presented in isolation using a two-alternative forced-choice identification task. The results indicated no significant effect of the duration of the F2 steady state, nor a significant effect of the duration of the F2 transition, as well as no significant interaction between the two cues. Hence, the current results suggest that the words /kio:/ and /kijo:/ are neutralized in perception. It remains to investigate whether the words are also neutralized in production. If these words can be distinguished in production, it would mean that this contrast is a case of near merger.

#### 7. REFERENCES

- [1] Aguilar, L. 1999. Hiatus and diphthong: Acoustic cues and speech situation differences. *Speech Communication* 28 (1), 57-74.
- [2] Boersma, P. *ExperimentMFC* [Experiment file]. Version 7 retrieved June 2018 from <http://www.fon.hum.uva.nl/praat/manual/ExperimentMFC.html>.
- [3] Boersma, P., Weenink, D. 2018. *Praat: doing by computer* [Computer program]. Version 6.0.42, retrieved June 2018 from <http://www.praat.org/>.
- [4] Fourakis, M., Port, R. 1986. Stop Epenthesis in English. *J. of Phonetics* 14 (2), 197-221.
- [5] Nakata, K., Suzuki, J. 1959. Synthesis and perception of Japanese vowels and vowel-like sounds. *Quarterly review of the Radio Research Laboratories* 5 (21), 237-244.
- [6] Keerio, A., Dhomeja, L., Shaikh, A. A., Malkani, Y. A. 2011. Comparative analysis of vowels, diphthongs and glides of Sindhi. *Signal & Image Processing: An International Journal* 2 (4), 109-120.
- [7] Port, R., Crawford, P. 1989. Incomplete Neutralization and Pragmatics in German. *J. of Phonetics* 17 (4), 257-282.
- [8] Sharf, D. J. 1960. Distinctiveness of 'voiced T' words. *American Speech* 35, 105-109.
- [9] Sharf, D. J. 1962. Duration of post-stress intervocalic stops and preceding vowels. *Language and Speech* 5, 26-30.

- [10] Son, B. 2011. Glide epenthesis for avoiding vowel in contemporary Japanese. *Journal of Japanese Language and Literature* 78, 197-210.
- [11] Winn, M. 2014. *Create duration continuum* [Praat script]. Version August 2014, retrieved July 2018 from [www.mattwinn.com/praat.html](http://www.mattwinn.com/praat.html).
- [12] Yu, A. C. L. 2007. Understanding near mergers: The case of morphological tone in Cantonese. *Phonology* 24 (1), 187-214.
- [13] Zue, V. W., Laferriere, M. 1979. Acoustic study of medial /t, d/ in American English. *JASA* 66, 1039–1050.