Recoverability-driven coarticulation: Acoustic evidence from Japanese high vowel reduction

Running title: *Recoverability and Japanese high vowel reduction*

James Whang[a)]

Department of Linguistics

New York University

10 Washington Place

New York, NY 10003

[a)]email: james.whang@nyu.edu

Abstract

High vowel reduction in Japanese, where unaccented /i,u/ in a C_1VC_2 sequence reduce when both C_1 and C_2 are voiceless, has been studied extensively, but factors that contribute to the degree of reduction is still debated. This study examines the effects of predictability on the degree of coarticulation between C_1 and the target vowel. Native Tokyo Japanese speakers (N=22) were recorded in a sound-attenuated booth reading sentences containing lexical stimuli. C_1 of the stimuli were /k, ʃ/, after which either high vowel can occur, and /ʧ, ɸ, s, ç/, after which only one of the two is possible. C_2 was always a stop. C_1 duration and center of gravity (COG), the amplitude weighted mean of frequencies present in a signal, were measured. Duration results show that reduction lengthens only non-fricatives, while it has either no effect or a shortening effect on fricatives. COG results show that vowel coarticulatory effects begin earlier for /k, ʃ/ than for /ʧ, ɸ, s, ç/. Coarticulatory information therefore seems to increase when the vowel is unpredictable but decrease or remain unchanged when the vowel is predictable.

PAC Number(s): 43.70.Fq, 43.70.Mn

I. INTRODUCTION

A. Background

The current study investigates the effects of recoverability—by way of phonotactic predictability—on the degree of high vowel reduction in Japanese. High vowel reduction is considered to be an integral feature of standard modern Japanese (Imai, 2010), so much so that dictionaries exist with explicit instructions for reducing environments (Kindaichi, 1995, pp.25–27). High vowel reduction (more commonly referred to as *devoicing*) is typically described as involving phonemically short high vowels /i/ and /u/, which are reduced in C_1VC_2 sequences when the vowels are unaccented and both C_1 and C_2 are voiceless obstruents. For example, while the /u/ in /kúʃi/ 'free use' and /kuʃi/ 'skewer' are both between two voiceless obstruents, only /kuʃi/ 'skewer' undergoes reduction because the vowel is unaccented. Likewise, the /u/ is unaccented in both /kuki/ 'stem' and /kugi/ 'nail', but only /kuki/ 'stem' undergoes reduction because the /u/ is flanked by two voiceless stops. The likelihood of reduction depends largely on the manner of the flanking consonants, where reduction rates can be as low as 60% when C_1 is a fricative or affricate and C_2 is a fricative, but can be nearly 100% elsewhere (Maekawa and Kikuchi, 2005; Fujimoto, 2015). Although not the focus of this study, non-high vowels can also reduce between voiceless obstruents but at much lower rates (<25%; Maekawa and Kikuchi, 2005), and unaccented high vowels optionally also reduce utterance finally after a voiceless fricative or affricate.

Despite the productivity of high vowel reduction in Japanese and the amount of interest the phenomenon has received in phonetics and phonology, there still is debate over whether the reduction process results in simple devoicing or in complete deletion of the target vowel. The lack of consensus regarding how reduced vowels are manifested acoustically stem in part from a lack of theoretical, experimental, and terminological consistency. For the purposes of this study, high vowel reduction is assumed to be a process that applies late in the phonological grammar, after lexical processes such as *rendaku* (Ito and Mester, 2003) and structural processes such as syllabification and phonotactic restrictions have applied (Boersma, 2009; Zsiga, 2000). Reducible high vowels, therefore, are phonologically present but can surface with varying degrees of reduction – as devoiced on one end, retaining all gestures except phonation, and deleted on the other (e.g., /ki̲ta/ → [ki̲̥ta] ‘north’; /hu̲koː/ → [ɸ_koː] ‘unhappy’). Furthermore, this study aims to test the hypothesis that the degree of reduction is dependent on the vowel’s recoverability (Varden, 2010).

Recoverability refers to the ease of accessing the underlying form—stored mental representations—from a given surface form—actual, variable output signals (e.g., [kæt̲̚, kæt̲ʰ] → /kæt̲/ ‘cat’; Mattingly, 1981; McCarthy, 1999; Chitoran et al., 2002). Recovery can be achieved by interpreting information explicitly present in the acoustic signal (coarticulatory cues) or by prediction based on context (phonotactic predictability). However, recoverability can be compromised if neither coarticulatory cues nor phonotactic predictability are sufficient. Varden (2010) states what seems to be a prevalent assumption in the Japanese high vowel reduction literature, which is that since high vowels trigger allophonic variation on preceding /t, s, h/ (i.e., /t/ → [ʧi, tsu]; /s/ → [ʃi, su]; /h/ → [çi, ɸu]), the underlying vowel is easily recoverable even if the vowel were to be phonetically deleted because the reduced vowel is predictable. For example, [ɸku] ‘clothes’ can only be analyzed as /huku/ because [ɸ_k] is a reducing context, where the vowel to be recovered can only be one of /i, ɯ/, and the mere presence of [ɸ] narrows the choice down to /u/ because [ɸ] can only occur as an allophone of /h/ preceding /u/ in non-loanwords. Because the context alone is sufficient for recovery, retaining the supralaryngeal gestures to provide more coarticulatory cues of the target vowel (e.g., [ɸu̥ku]) does little to improve recoverability. What Varden is proposing then is that little to no coarticulatory cues are necessary when phonotactic predictability is high, which also leads to the reverse prediction that more coarticulatory cues should be necessary if phonotactic predictability is low.

A number of studies have proposed similar recoverability-conditioned coarticulation, where speakers seem to preserve or enhance the phonetic cues of a target segment in situations where the target segment would be less perceptible, such as when a phoneme inventory contains acoustically similar phonemes (Silverman, 1997) or in word-initial stop-stop sequences, where the closure of the second stop would obscure the burst of the first (Chitoran et al., 2002). However, whether the amount of C_1V coarticulation is similarly modulated by phonotactic predictability in Japanese has not been tested systematically.

The consonants included in the current study and the vowels that can follow each of the consonants are summarized in Table I below. The current study investigates reduction after /t, s, h/, which are targets of allophonic variation before high vowels, and /k, ʃ/[1], which are not. The allophonic variations of /t, s, h/ make high vowels that follow highly predictable, and thus more likely to delete. High vowels after /k, ʃ/ are less predictable due to a lack of allophony, and thus more likely to devoice.

TABLE I: Consonants used in stimuli and high vowels that can follow. "–" means that the vowel is not phonologically possible in this context (in non-loanwords).

		i	u
High predictability	ʧ	✓	–
	s	–	✓
	ɸ	–	✓
	ç	✓	–
Low predictability	k	✓	✓
	ʃ	✓	✓

It should be noted that /s/ and /ʃ/ are contrastive before all vowels except /i/, where the contrast is neutralized to [ʃ]. /ʃ/ additionally cannot precede /e/ in non-loanwords. Furthermore, while [ʧ], [s], and [ɸ, ç] are allophones of /t, s, h/, respectively, they are also semi-phonemic and can precede other non-high vowels in Sino-Japanese words and loanwords. The bilabial stop /p/ is excluded because it almost never occurs word-initially outside of loanwords and mimetic words. When /p/ does occur in Yamato and Sino-Japanese words, it is usually the result of /h/ allophony after codas (e.g., /kaN+hai/ → [kampai] 'cheers (dry+cup)') or part of a suffix which begins with a geminate (e.g., /kodomo+ppoi/ → [kodomoppoi] 'childish (child+ish)'). Furthermore, the affricate [ts], which is another allophone of /t/ that occurs before /u/, is also not included to keep

[1]Although the post-alveolar fricative is more accurately an alveo-palatal fricative /ɕ/, the IPA symbol for the post-alveolar fricative is used throughout for the sake of readability and to enhance differentiation from the palatal fricative [ç] which is an allophone of /h/.

the number of stimuli balanced between high and low predictability tokens.

B. Previous studies

There are primarily three ways in which reduced high vowels are argued to be manifested acoustically: (i) by lengthening the burst/frication noise of C_1 which carries coarticulatory cues of a devoiced vowel (Han, 1994), (ii) by devoicing the vowel and coloring the C_1 burst/frication noise with the retained oral gestures without necessarily lengthening C_1 (Beckman and Shoji, 1984), and (iii) by deleting the vowel altogether (Vance, 2008). Each of the proposed manifestations has contradicting evidence in previous literature as discussed below. Since there is disagreement regarding whether the target vowels are devoiced or simply deleted, the term *reduction* is used throughout this study as a general term to refer to a lack of phonation associated with a target vowel, instead of the more common term "devoicing".

Although it is commonly argued that C_1 is longer in reduced syllables than in unreduced syllables, the empirical evidence is not unanimous. Part of the problem in the lack of consensus regarding the effects of vowel reduction on C_1 duration in Japanese is that there are differences in the methodologies and stimuli among the studies. For example, Varden (1998) examines /k,t/ (where /t/ → [ʧi, ʦu]) and reports that the burst and aspiration of C1 in reduced syllables are significantly longer than the consonant portion of their corresponding unreduced CV syllables. On the other hand, studies that focus on /s/ (→ [ʃi, su]; Beckman, 1982; Beckman and Shoji, 1984; Faber and Vance, 2000) consistently report that there is no significant difference in duration between /s/ in reduced and unreduced syllables. In other words, studies that investigate fricatives find no lengthening effect while studies that investigate stops and affricates find lengthening effects.

Additionally, studies that report lengthening effects generally assume that Japanese is mora-timed and that moras are roughly equal in duration. Based on these assumptions, the duration results of individual C_1 are often collapsed (Tsuchida, 1997; Kondo, 2005) or C_1 in reduced contexts are compared to different segments in unreduced contexts (Han, 1994). These practices are justified if moras in Japanese are indeed equal in duration, but as Warner and Arai

(2001a,b) argue, the apparent rhythm in Japanese and the compensatory lengthening effect in relation to mora-timing might be epiphenomenal, stemming from a confluence of factors that result from the phonological structure of Japanese.

While it is conceptually plausible that the presence of an underlying vowel can be signaled solely by C_1 lengthening, especially if mora preservation is the reason behind it, much of the literature arguing for compensatory lengthening also assumes that reduced vowels are devoiced rather than deleted. A number of articulatory studies looking at /k, t, s/ as C_1 found that the glottis is wider when the vowel in a C_1VC_2 sequence is reduced than when it is not, and that there is only one activity peak for the laryngeal muscles aligned with the onset of C_1 in reduced sequences, resulting in a long frication or a frication-like burst release for stops (Fujimoto et al., 2002; Tsuchida et al., 1997; Yoshioka et al., 1982). Since there is no laryngeal activity associated with C_2 apart from the carry-over from C_1 and because the abduction peak for the glottis was found to be larger than the sum of two voiceless consonants, these results are interpreted to mean that the glottal gesture is being actively controlled to spread the feature [+spread glottis] from the first consonant to the second. As a consequence of this spreading, the intervening high vowel is devoiced. Despite the lack of a laryngeal gesture associated with phonation, the presence of formant-like structures in the burst/frication noise of C_1 are often reported, which is taken as evidence of retained oral vowel gestures. For example, a recent acoustic study by Varden (2010) reports visible formant structures apparent in the fricated burst noise of [ki̥, ku̥], which are interpreted to be the result of oral gestural overlap that allows consistent identification of the underlying devoiced vowel.

In contrast, Ogasawara (2013) reports a lack of visible formant structures in the burst/frication noise of /k, t/ in most reduced cases and argues that this provides support for the claim that high vowel reduction results in deletion rather than devoicing (Hirose, 1971; Yoshioka, 1981). The lack of apparent formant structures in the burst/frication noise of C_1, however, seems to be an inadequate criterion for measuring the presence of vocalic oral gestures. While Beckman and Shoji (1984) also report that the presence of formant-like structures on the frication noise of

/ʃ/ is inconsistent, spectral measurements of [ʃ] showed a small yet noticeable influence of reduced vowels on the aperiodic noise of the preceding fricative, where the mean frequency of [ʃu̥] was lower than [ʃi̥] by approximately 400 Hz, suggesting a coarticulatory effect of a reduced vowel. Perceptually, this difference was enough to aid the listeners in identifying the underlying vowel above the rate of chance (77% for [ʃi̥] and 67% for [ʃu̥]).

C. Possible effects of predictability on coarticulation

There are three main possibilities with respect to the question of how predictability affects coarticulation. The first is that high vowel reduction is blind to predictability and is driven primarily by Japanese phonotactics, which has a strict CVCV structure that disallows tautosyllabic clusters (Kubozono, 2015). If this is the case, then no difference between low predictability and high predictability C_1 would be found, where the reduced vowel never deletes completely but always devoices instead, coloring the burst or frication noise of C_1 to signal the presence of the target vowel (Beckman and Shoji, 1984; Varden, 2010). The second is that the degree of coarticulation between C_1 and the following vowel is not systematic but rather a consequence of how the reduced vowel happened to be lexicalized for the speaker. Ogasawara and Warner (2009) found in a lexical judgment task that when Japanese listeners were presented with unreduced forms of words where reduction is typically expected, reaction times were longer than when presented with reduced forms. This suggests that the reduced forms, despite their phonotactic violations, can have a facilitatory effect on lexical access due to their commonness, making vowel recovery unnecessary (Cutler et al., 2009; Ogasawara, 2013). The third and last option, which this study proposes, is that high vowel reduction is constrained by recoverability. In this case, increased coarticulation would be observed either by lengthening or spectral changes of C_1 burst/frication when the predictability of the target vowel is unreliable from a given C_1 to aid phonetic interpretability as in the case of /k, ʃ/, but not when predictability is high, as in the case of [ʧ, s, ɸ, ç]. This last outcome would also be compatible with the idea that reduced forms are lexicalized as such (Ogasawara and Warner, 2009), but with the caveat that the degree of reduction is dependent on predictability from context.

While this study does not explore sociolinguistic factors that affect high vowel reduction, it is worth noting that men have been reported to reduce more than women (Okamoto, 1995) and that reduction rates are higher overall in younger speakers (Varden and Sato, 1996). However, Imai (2010) found that while younger speakers did tend to reduce more, this was only true for men. Young female speakers were actually shown to reduce the least among all age groups. Based on these findings, Imai proposes that high vowel reduction might be being utilized actively as a feature of gendered speech. If high vowel reduction is being utilized as a sociolinguistic feature, then the process could not be a purely phonological or a phonetic process, and thus a balanced number of men and women were recruited to investigate any gender-based differences.

II. MATERIALS AND METHODS

A. Participants

Twenty-two monolingual Japanese speakers (12 women and 10 men) were recruited in Tokyo, Japan. All participants were undergraduate students born and raised in the greater Tokyo area and were between the ages 18 and 24. Although all participants learned English as a second language as part of their compulsory education, none had resided outside of Japan for more than six months and have not been overseas within a year prior to the experiment. All participants were compensated for their time.

B. Materials

The stimuli for the experiment were 160 native Japanese and Sino-Japanese words with an initial C_1iC_2 or C_1uC_2 target sequence. The stimuli were controlled to be of medium frequency (20 to 100 occurrences, which is the mean and one standard deviation from the mean, respectively) based on the frequency counts from a corpus of Japanese blogs (Sharoff, 2008). Any gaps in the data were filled with words of comparable frequency based on search hits in Google Japan (10 million to 250 million). Since high vowel reduction typically occurs in unaccented syllables, an accent dictionary of standard Japanese (Kindaichi, 1995) was used as reference to ensure that none of the stimuli had a target vowel in an accented syllable.

The stimuli were divided into *low predictability* and *high predictability* groups. Predictability, for the purposes of this study, refers specifically to the predictability of vowel backness, given high vowels. Examples of the reducing stimuli are shown in Table II below.

TABLE II: Example of reducing stimuli by C_1 and vowel.

stimulus type	C_1	*V*	*example*	*gloss*
low predictability	k	i	kiki	‘handedness’
		u	kuki	‘twig’
	ʃ	i	ʃikoː	‘thought’
		u	ʃukoː	‘plan’
high predictability	ʧ	i	ʧikʲuː	‘earth’
	s	u	sukuː	‘rescue’
	ɸ	u	ɸukoː	‘unhappy’
	ç	i	çiteː	‘denial’

As shown above, for the low predictability group, C_1 was either /k, ʃ/ after which both /i, u/ can occur. For the high predictability group, C_1 was one of [ʧ, s, ɸ, ç], after which only one of the high vowels is possible. The two groups were further divided into *reducing* and *non-reducing* contexts. The difference between reducing and non-reducing tokens was that C_2 was always a voiceless stop (i.e., [p, t, k]) for reducing contexts as shown above, but a voiced stop for non-reducing tokens (i.e., [b, d, g]). Since high vowel reduction typically requires the target vowel to be flanked by two voiceless obstruents, it was expected that reduction would not occur in the non-reducing contexts. The C_1 and C_2 combinations resulted in fricative-stop, affricate-stop, or stop-stop contexts. These contexts were chosen for two reasons: (i) these are contexts in which high vowel reduction is reported to occur systematically and categorically (Fujimoto, 2015), and (ii) the C_2 stop closure clearly marks where the previous segment ends. There were 10 tokens per C_1V combination within each context, for a total of 160 tokens (80 reducing and 80 non-reducing).

It should be noted that while both /i, u/ can follow /ʧ/ in Japanese, only /ʧi/ is included in the stimulus set because /ʧ/ is rarely followed by short /u/ in Japanese. A search of the Shogakukan (2013) dictionary revealed that of the 6,041 entries that begin with /ʧ/, 38% are followed by /i/ compared to only 1% that are followed by the short vowel /u/, of which only 5 words were in potentially reducing contexts. In other words, when /ʧ/ is followed by a reducible high vowel, the phonotactic distribution of the language heavily favors the vowel /i/, making the environment highly predictable.

C. Design and procedure

All tokens were placed in the context of unique and meaningful carrier sentences of varying lengths. No tokens were included more than once in the experiment, and no two carrier sentences were identical. All carrier sentences contained at least one stimulus item, and the sentences were constructed so that no major phrasal boundaries immediately preceded or followed the syllable containing the target vowel. An example carrier sentence, which was actually uttered by a weather forecaster in Japan, is given below with glosses.

(1)	manatsu-no	ʃigaiseɴ-ni-wa	ki-o-tsuke-maʃoː
	midsummer-LNK	ultraviolet rays-DAT-TOP	be careful.VOL

‘Let’s be careful of midsummer’s ultraviolet rays’

The carrier sentences were presented one at a time to the participants on a computer monitor as a slideshow presentation. The participants advanced the slideshow manually, giving the participants time to familiarize themselves with the sentences. They were also allowed to take as many breaks as they thought was necessary during the recording. All instructions were given in Japanese, and participants were prompted to repeat any sentences that were produced disfluently. All participants were recorded in a sound-attenuated booth with an Audio-Technica ATM98 microphone attached to a Marantz PMD-670 digital recorder at a sampling rate of 44.1 kHz at a 16 bit quantization level. The microphone was secured on a table-top stand, placed 3-5 inches from the mouth of the participant.

D. Data Analysis

Once the participants were recorded, the waveform and spectrogram of each participant were examined in Praat to (a) code each token for reduction, (b) to measure the duration of C_1

and the following vowel, and (c) to measure the center of gravity of C_1 burst/frication noise. The spectrogram settings were as follows: pre-emphasis was set at +6 dB, dynamic range was set at 60 dB, and autoscaling was turned off for consistency of visual detail. Because visual inspection alone is an inadequate method for determining the presence of vowel coarticulation on C_1 (Beckman and Shoji, 1984), tokens were simply coded for “reduction”, a term used to collectively refer to devoicing and deletion of the vowel. The criteria used for reduction status are described in the following section.

1. Reduction analysis

Vowels in reducing environments were coded as unreduced if there was phonation accompanied by formant structures between C_1 and C_2. Vowels were coded as reduced when there was no phonation between C_1 and C_2. Below in Figure 1 are examples from the same female speaker. On the left is an unreduced vowel in the word [kuki] ‘twig’, which shows clear phonation and formant structures between C_1 and C_2. On the right is a reduced vowel in the word [ku̥teɴ] ‘period’, where there is neither phonation nor formant structures between C_1 and C_2.

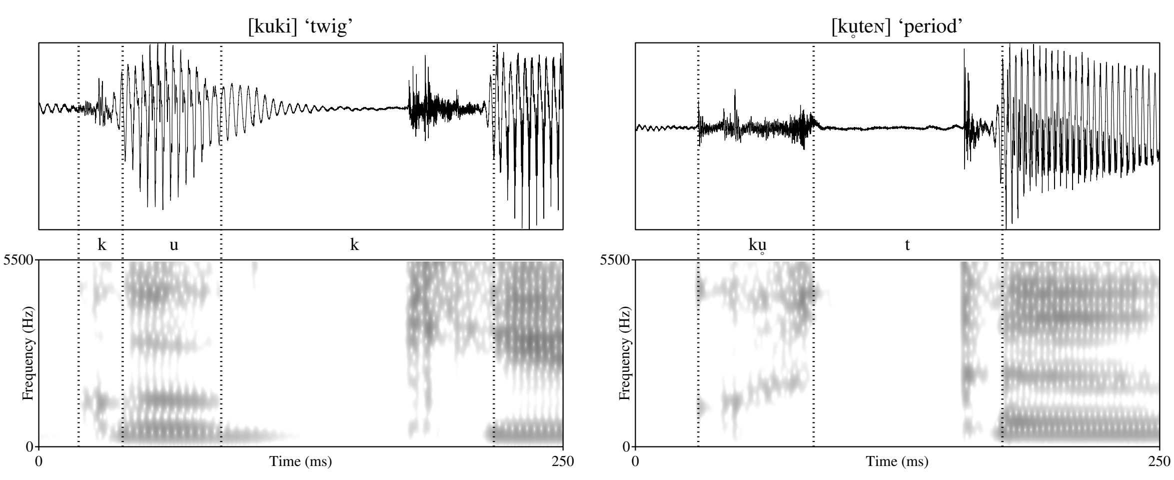

FIG 1: Waveform and spectrogram of unreduced (left) and reduced (right) vowels in reducing environments, showing landmarks for C_1, vowel, and C_2 duration.

The coding criteria were similar for non-reducing tokens. A vowel was coded as unreduced if phonation and formant structure were both present between C_1 and C_2. Otherwise, a vowel was coded as reduced. Below in Figure 2 are examples from another female speaker. On the left is an unreduced vowel in the word [ʃugeː] ‘handicraft’, where there is a clear formant structure accompanying phonation. On the right is a rare case of a reduced vowel in a non-reducing word [ʃu̥daika] ‘theme song’, where there is phonation between C_1 and C_2 but no formant structure.

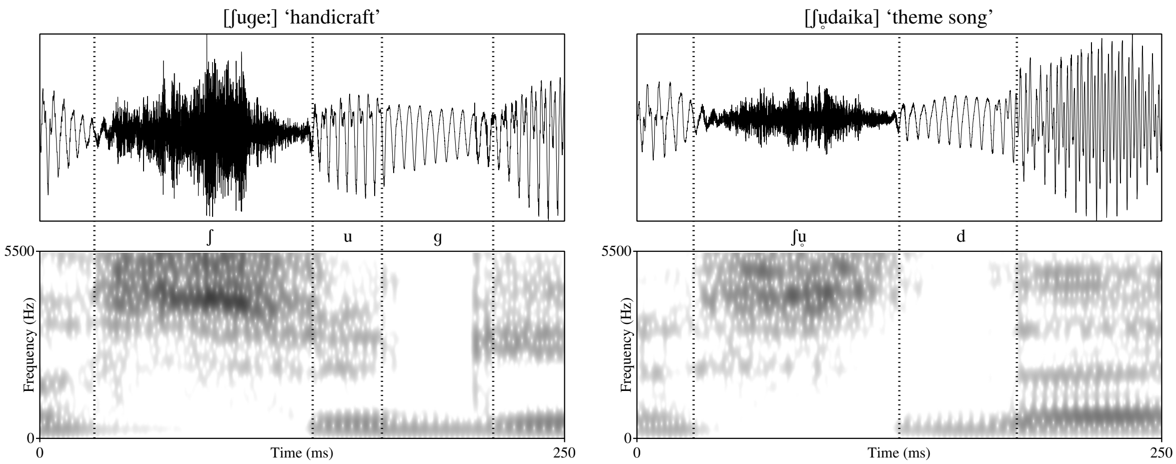

FIG 2: Waveform and spectrogram of unreduced (left) and reduced (right) vowels in non-reducing environments, showing landmarks for C_1, vowel, and C_2 duration.

2. Duration analysis

Once all tokens were coded for reduction status, duration measurements were taken to investigate how reduction affects the gestural timing of C_1 and the target high vowel. For [k] and [ʧ], duration measurements excluded the silence from closure and included only the aperiodic burst energy. For fricative C_1, duration measurements included the entire aperiodic frication noise. For tokens coded as reduced, C_1 measurements were assumed to include the reduced vowel because the vowel could not be isolated from C_1 reliably. For unreduced tokens, C_1 was measured from the onset of burst/frication noise to the onset of vowel F2. For both duration and center of gravity analyses, only reduced tokens in reducing environments and unreduced tokens in non-reducing environments were included.

3. Center of gravity analysis

Center of gravity (COG), which is the amplitude weighted mean of frequencies present in the signal (Forrest et al., 1988), was also calculated for C_1 to investigate the degree of coarticulation between C_1 and the target vowel. COG measurements are used based on Tsuchida (1994), who found that Japanese listeners rely primarily on C_1 centroid frequency (i.e., COG) to identify reduced vowels. COG measurements are known to be particularly sensitive to changes in the front oral cavity (Nittrouer et al., 1989), so the effects of increased coarticulation between a vowel and C_1 on COG values are expected to differ by the backness and roundedness of the vowel as well as C_1 place of articulation. The predicted effects of vowel coarticulation on each C_1 are discussed in detail in §III-C together with the results.

Before measuring COG values, the sound files were high pass filtered at 400 Hz to mitigate the effects of f0 on the burst/frication noise. The filtered sound files were then down-sampled to 22,000 Hz. The COG values measured therefore were taken from FFT spectra in the band of 400 to 11,000 Hz (Forrest et al., 1988; Hamann and Sennema, 2005). With the exception of /k/, two center of gravity (COG) measurements were taken from 20 ms windows for each C_1: one starting 10 ms after the beginning of C_1 burst/frication (COG1) and one ending 10 ms before the end of C_1 burst/frication (COG2). The 10 ms buffers were used to mitigate the coarticulatory effects of segments immediately adjacent to C_1. Comparisons of COG1 and COG2 between reduced and unreduced tokens allow for inference of how early or late vowel coarticulation effects begin in the consonant. Comparison of ΔCOG (COG2 $-$ COG1) also allows testing of how the trajectory of coarticulatory effects differs between reduced and unreduced tokens.

For /k/, COG measurements were taken from a single 20 ms window at the midpoint of the burst. Two COG measurements could not be taken from /k/ because /k/ duration in unreduced tokens were too short for two measurements. /k/ tokens shorter than 20 ms were excluded from analysis, which resulted in the loss of five tokens, or 0.6% of the /k/ data. Since the vocalic gesture of the following vowel most likely begins during the stop closure for /k/ (Browman and Goldstein, 1992; Fowler and Saltzman, 1993), the single COG measurement is assumed to be equivalent to the COG2 measurements of other consonants.

III. RESULTS

Statistical analyses were performed by fitting linear mixed effects models using the *lme4* package (Bates et al., 2015) for R (R Core Team, 2016). In order to identify the maximal random effects structure justified by the data, a model with a full fixed effects structure (i.e., with interactions for all the fixed effects) and the most complex random effects structure was fit first. If the model did not converge, the random effects structure was simplified until convergence was reached while keeping the fixed effects constant (Barr et al., 2013). The simplest random effects structure considered was one with random intercepts for participant and word with no random slopes.

Once the maximal random effects structure was identified, a Chi-square test of the log likelihood ratios were performed to identify the best combination of fixed effects. A complex model with all interaction terms was fit first, which was then gradually simplified by removing predictors that did not significantly improve the fit of the model, starting with interaction terms. The simplest model considered was a model with no fixed effects and only an intercept term. Because the fixed and random effects were slightly different for each of the analyses performed, the structure of the final model will be described in the respective sections below.

A. Reduction rate

1. Overall reduction rates and analysis

Reduction rates were at or near 100% for reducing tokens, while non-reducing tokens had significantly lower reduction rates, as shown in Table III below.

TABLE III: Reduction rate by C_1V and context.

stimulus type	C_1	*V*	*reducing*	*non-reducing*
low predictability	k	i	1.000	0.077
		u	0.959	0.032
	ʃ	i	1.000	0.086
		u	0.973	0.073
high predictability	ʧ	i	1.000	0.191
	ç	i	1.000	0.015
	ɸ	u	1.000	0.042
	s	u	1.000	0.214
overall			0.992	0.091

A mixed logit model was fit using the *glmer()* function of the *lmer* package for the overall reduction rate with reducing context, predictability, gender, and their interactions as predictors. Vowel was not included as a predictor because it is redundant for high predictability tokens since only one vowel is allowed. Random intercepts for participant and word were added to the model. By-participant random slopes for context and predictability as well as by-word random slopes for gender were also included in the model. The final model retained the full random effects structure. The following predictors were removed from the fixed effects structure of the final model as they were not significant contributors to the fit of the model: three-way interaction ($p =$ 0.999378), context:gender interaction ($p = 0.901798$), and predictability:gender interaction ($p =$ 0.062329). The function for the final model, therefore, was as follows:

model $=$ *glmer(reduction* $\sim$ *context + predictability + gender + context:predictability* $+$ *(1 + context + predictability* $|$ *participant)* $+$ *(1 + gender* $|$ *word), family = binomial(link = 'logit'), data = non-loanwords)*

The results of the final model showed that the difference in reduction rates between reducing and non-reducing contexts was significant ($p < 0.0001$) and that men were more likely to reduce than women ($p = 0.0175$). Predictability and context:predictability interaction did not have significant effects ($p = 0.2374$ and 0.7237, respectively).

An additional analysis was performed on just the non-reducing subset of the data because reducing tokens reduced essentially 100% of the time and had no between-participant differences to test statistically. First, a mixed logit model was fit to the low predictability non-reducing tokens with gender, C_1, vowel, and their interactions as predictors. Random intercepts for participant and word were included in the model. By-participant random slopes for C_1 and vowel, and by-word random slopes for gender were also included. /ʃ/ tokens as produced by female participants were the baseline. However, none of the predictors were significant contributors to the fit of the model, and a Chi-square test showed the fit of the intercept-only model was not significant different from more complex models. In other words, /k, ʃ/ had similar reduction rates in non-reducing contexts regardless of vowel or gender.

Second, a mixed logit model was fit to the high predictability non-reducing tokens with gender, C_1, and their interaction as predictors. Random intercepts for participant and word were included in the model. By-participant random slopes for C_1 and by-word random slopes for gender were also included. The interaction term was not a significant contributor to the model ($p = 0.07828$), and thus was removed from the final model. /ʧ/ tokens as produced by female participants were the baseline. The results showed that male participants were more likely to reduce than women ($p = 0.011490$). C_1 did not have a significant effect ($p = 0.171173$, 0.092141, and 0.516625 for /ɸ, ç, s/ respectively).

2. Summary of reduction rate results

The analysis of reduction rates showed that there is an effect of context on overall reduction rates. At essentially 100%, reduction rates are significantly higher in the reducing environments than in non-reducing environments. Male participants were also shown to be more likely to reduce than female participants, but the difference did not come from reducing tokens. Separate analyses of low and high predictability tokens revealed men reduced more in high-predictability environments, where reduction was not actually phonologically conditioned (e.g., /ɸugoːri/ → [ɸu̥goːri] ‘unreasonable’).

B. Duration

Previous studies that report lengthening effects of reduction on C_1 generally have focused on /k, t/ (Varden, 1998), while studies that report a lack of such effect focused on /s, ʃ/ (Beckman and Shoji, 1984; Vance, 2008). There are two confounded differences between /k, t/ and /s, ʃ/ that may be contributing to the contrary results: manner and inherent duration. /k, t/ are non-continuants while /s, ʃ/ are continuants, but it is also the case that the burst of the former are inherently much shorter than the frication noise of the latter. This means that the contrary results could be due to either or both of these differences. The allophones of /h/—[ɸ, ç]—are therefore crucial in teasing apart the two factors because [ɸ, ç] are fricatives but are also similar in duration as the frication portion of [ʧ] in Japanese.[2]

1. Overall duration results and analysis

Duration results are shown in Figure 3 and Table IV below. The results suggest that overall C_1 burst/frication durations are not different between women and men. Reduction seems to have a lengthening effect only on non-fricative obstruents (i.e., /ki, ku, ʧi/). For fricatives, reduction seems to have no effect on /ɸu/ and a shortening effect on others (i.e., /çi, su, ʃu, ʃi/).

[2] An analysis of consonant durations in the Corpus of Spontaneous Japanese revealed that there is no significant duration difference between [ʧ] and [ɸ] in unreduced contexts (∼65 ms; $p = 0.891$), and between [ʧ] and [ç] in reduced contexts (∼75 ms; $p = 0.475$).

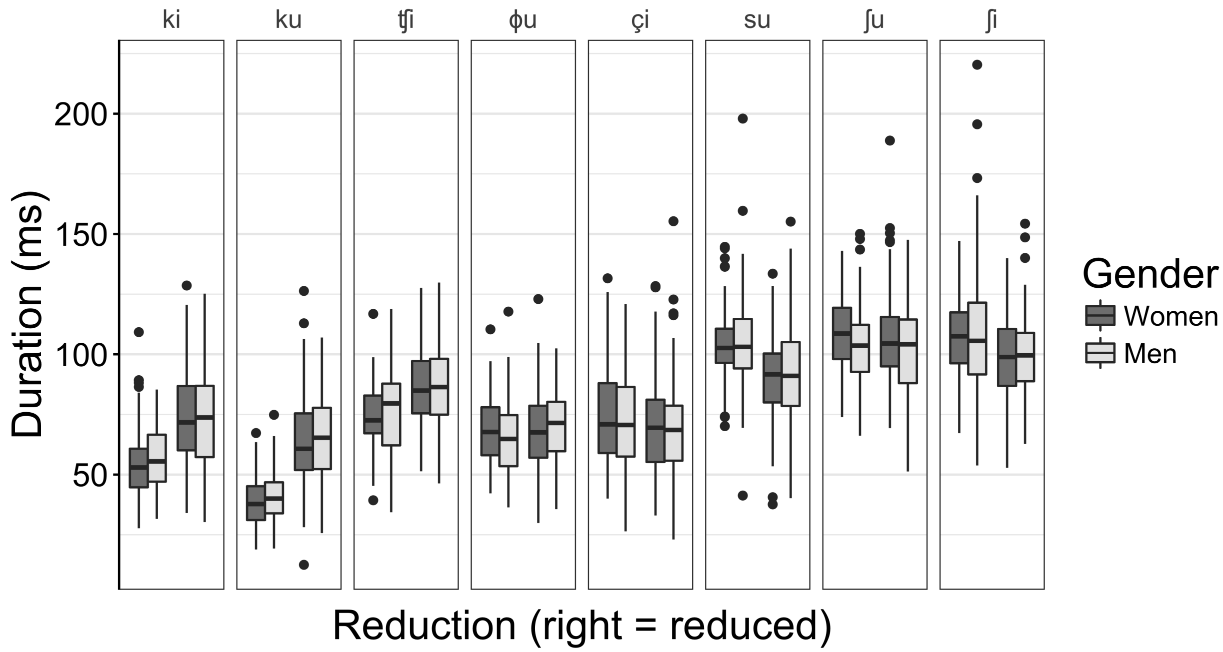

FIG 3: C_1 duration in ms by C_1V, gender, and reduction.

TABLE IV: C_1 mean duration (*standard deviation*) in ms. Lengthening effect in bold.

		unreduced		reduced	
	C_1V	*female*	*male*	*female*	*male*
low predictability	**ki**	**55 (*13*)**	**57 (*13*)**	**74 (*20*)**	**74 (*20*)**
	ku	**39 (*11*)**	**41 (*11*)**	**65 (*17*)**	**65 (*17*)**
	ʃi	107 (*16*)	108 (*26*)	99 (*18*)	99 (*17*)
	ʃu	109 (*14*)	104 (*16*)	106 (*19*)	102 (*20*)
high predictability	**ʧi**	**74 (*14*)**	**76 (*17*)**	**86 (*15*)**	**86 (*18*)**
	çi	75 (*21*)	74 (*19*)	71 (*18*)	71 (*19*)
	ɸu	69 (*14*)	66 (*15*)	69 (*15*)	71 (*14*)
	su	104 (*15*)	105 (*21*)	91 (*17*)	92 (*20*)

A linear mixed effects regression model was fit to the overall duration results with reduction, gender, C_1, and their interactions as predictors. Again, vowel was not included as a predictor because it is only meaningful for /k, ʃ/ tokens. Random intercepts for participant and word were added to the model. By-participant random slopes for context and C_1 were also included in the model, as well as by-word random slopes for gender. Because there is currently no consensus on how to accurately calculate *p*-values for mixed effects models, absolute *t* values greater than 2 were regarded as significant (Baayen et al., 2008).

The final model retained the full random effects structure. The following non-significant predictors were removed from the final model: three-way interaction ($p = 0.3040$), reduction:gender interaction ($p = 0.9266$), gender:C_1 interaction ($p = 0.6081$), and gender ($p = 0.5797$). The final model therefore retained reduction, C_1, and their interaction as predictors. The function for the final model was as follows:

*model = lmer(duration ∼ reduction * C1 + (1 + context + C1 | participant) + (1 + gender | word), control=lmerControl(optimizer="bobyqa"), REML = F, data = non-loanwords)*

The final model's results are summarized below in Table V. Unreduced /k/ tokens are the baseline.

TABLE V: Linear mixed effects regression model results for overall C_1 duration.

	Estimate	Std. Error	*t*	
(Intercept)	47.365	2.264	20.917	*
reduced	22.068	3.106	7.106	*
ɸ	20.464	3.516	5.819	*
ç	26.808	3.746	7.156	*
ʧ	27.399	3.634	7.539	*
s	55.317	3.751	14.749	*
ʃ	59.454	3.155	18.844	*
reduced:ɸ	-20.396	4.877	-4.182	*
reduced:ç	-25.340	4.964	-5.105	*
reduced:ʧ	-10.514	4.895	-2.148	*
reduced:s	-33.451	4.903	-6.823	*
reduced:ʃ	-27.009	3.983	-6.781	*

The results show that reduction indeed has a lengthening effect of 22 ms on /k/. The intercept estimates for C_1 predictors show that all other C_1 are significantly longer than the /k/ baseline. The negative values of the estimates for the reduction:C_1 interaction predictors also show that reduction has a smaller lengthening effect on all other C_1 relative to the /k/ baseline.

The model above only shows how other C_1 differ from /k/. In order to explore whether reduction actually had significant effects on the individual C_1, differences of least squares means were calculated from the final model using the *difflsmeans()* function of the *lmerTest* package

(Kuznetsova et al., 2016). The results showed that reduction had a significant lengthening effect on /ʧ/ (11.6 ms, $p = 0.007$). The fricatives on the other hand showed varying effects. Reduction had a non-significant lengthening effect of 1.7 ms on /ɸ/ ($p = 0.691$) and non-significant shortening effects of 3.3 ms on /ç/ ($p = 0.447$) and 4.9 ms on /ʃ/ ($p = 0.114$). However, reduction had a significant shortening effect of 11.4 ms on /s/ ($p = 0.008$).

A separate linear mixed effects regression model was fit to low predictability tokens (i.e., /k, ʃ/) to investigate the effects of vowel type. Since the overall model above already showed that reduction had a lengthening effect on /k/, the baseline was set to /ʃ/. Reduction status, C_1, vowel type, and their interactions were included as predictors. Random intercepts by participant and word were included. By-participant random slopes for reduction, C_1, and vowel type were also included, as well as by-word random slopes for gender. The final model retained the full random effects structure. The three-way interaction term and the reduction:vowel interaction were not significant contributors to the model ($p = 0.7549$ and 0.1262, respectively) and were removed from the fixed effects structure of the final model.

The results of the final model showed that although reduction had a slight shortening effect of 5 ms and the vowel /u/ had a slight lengthening effect of 3 ms on /ʃ/, neither was significant ($t = -1.522$ and 1.061, respectively). Also, as was shown in the overall model above, reduction had a significant lengthening effect of 22 ms on /k/ ($t = 6.496$).

2. Summary of duration results

For duration, gender did not have a significant effect. Reduction, on the other hand, had opposite effects depending on manner and C_1 duration. In terms of manner, reduction had a lengthening effect on the two non-fricative C_1 /k/ and /ʧ/. Reduction actually showed a significant shortening effect for /s/ tokens, and no effect on the rest of the fricatives tested in this study /ʃ, ɸ, ç/. Particularly, the fact that /ʧ/ lengthened but not /ɸ, ç/ despite being similar in length suggests that lengthening is largely dependent on whether the consonant is a continuant or not. However, inherent C_1 duration also had an effect on the magnitude of the duration effects. Both /k/ and /ʧ/ lengthened, but the shorter segment /k/ lengthened more (22 ms vs. 12 ms), a difference that was shown to be significant (t = -2.148). For the fricatives, /s/ shortened while the shorter fricatives did not. Additionally, despite the fact that /s/ and /ʃ/ both have similar durations of $\sim$100 ms, only /s/ shortened significantly, suggesting that there may be an effect of predictability as well.

C. Center of gravity (COG)

COG is sensitive primarily to changes in the front cavity (Nittrouer et al., 1989), so the effects of gestural changes are expected to differ depending on the coarticulated vowel and the C_1 place of articulation. In general, however, increased coarticulation with the high back vowel /u/ is expected to have a lowering effect for all C_1 due to lip rounding, which would lengthen the front oral cavity. Although the high back vowel of Japanese has traditionally been regarded as unrounded (i.e., [ɯ]), a recent articulatory study by Nogita et al. (2013) showed that the high back vowel is actually closer to a high central rounded vowel [ʉ] in younger speakers. On the other hand, the high front vowel is expected to have different effects depending on how front in the oral cavity the C_1 place of articulation is, making direct statistical comparisons impractical. This section therefore analyzes each C_1 separately. /ʃ/ is analyzed first because it is the only fricative that can be tested for both vowel and reduction effects. The /ʃ/ results are then used as reference for interpreting all other C_1.

A linear mixed effects regression model was fit for all statistical analyses. Unless noted otherwise, the random effects structure for the fully complex model included random intercepts for participants and words, by-participant random slopes for vowel and reduction, and by-word random slopes for gender.

1. /ʃ/ COG results and analysis

For /ʃ/, the COG values for /u/ tokens are expected to be lower than for /i/ tokens regardless of reduction, similar to Beckman and Shoji (1984). COG2 is also expected to be lower than COG1 in unreduced tokens for both vowels, as lip rounding increases for /u/ articulation or tongue shifts back towards the palate for /i/, both of which would lengthen the front oral cavity. Given the expected lowering effect of coarticulation for both vowels, there are a three possible effects of reduction. First, reduced tokens may show increased coarticulation between C_1 and the target vowel, resulting in lower COG values. Second, reduced tokens may show decreased coarticulation, leading to higher COG values. Third, reduced and unreduced tokens may show similar degrees of C_1V coarticulation, showing no difference in COG values.

Shown below in Table VI are the COG1 and COG2 values of /ʃ/. The mean COG values are lower when the vowel is /u/ for both COG1 and COG2 as expected.

TABLE VI: COG1 and COG2 mean (*standard deviation*) in Hz for /ʃ/.

C_1V		unreduced		reduced	
		female	*male*	*female*	*male*
ʃi	COG1	5694 (*622*)	4996 (*488*)	5201 (*862*)	4738 (*619*)
	COG2	5317 (*489*)	4592 (*342*)	5317 (*926*)	4695 (*787*)
ʃu	COG1	4948 (*606*)	4403 (*384*)	4924 (*758*)	4504 (*470*)
	COG2	4469 (*508*)	4060 (*409*)	4555 (*842*)	4311 (*633*)

Male participants also have lower COG values overall, which is unsurprising given than men generally have longer vocal tracts than women. Reduction also seems to have a lowering effect on COG1 when the vowel is /i/, but not when the vowel is /u/.

The final model fit to the COG1 results of /ʃ/ retained the full random effects structure. The following non-significant predictors were removed from the final model: three-way interaction ($p = 0.5353$) and gender:vowel interaction ($p = 0.3846$). The final model's results are summarized below in Table VII, with the baseline set as unreduced /ʃi/ tokens produced by female participants.

TABLE VII: Linear mixed effects regression results: COG1 of /ʃ/.

	Estimate	Std. Error	*t*	
(Intercept)	5625.1	140.7	39.97	*
/u/	-631.0	126.6	-4.98	*
reduced	-448.3	132.7	-3.38	*
male	-606.3	159.4	-3.80	*
reduced:/u/	358.9	163.5	2.20	*
reduced:male	199.4	73.0	2.73	*

The results of the model show that COG1 is significantly lower for /ʃu/ tokens compared to /ʃi/

tokens, suggesting that coarticulation with /u/ begins early in unreduced tokens. Additionally, reduction has a significant lowering effect. Since the model's baseline was unreduced /ʃi/ tokens, lower COG1 suggests that the front oral cavity is larger in reduced [ʃi̥] tokens than in unreduced [ʃi] tokens. Coarticulation with /i/, therefore, begins earlier when the vowel is reduced. The lowering effect of reduction is significantly smaller for male participants and when the vowel is /u/, however. Differences of least squares means of the model revealed that the effects of reduction are in fact not significant for the male participants ($p = 0.429$) and when the vowel is /u/ ($p = 0.932$). In other words, reduced tokens do not show evidence of increased coarticulation in male participants and for [ʃu̥] tokens. Lastly, the results also show that COG1 is significantly lower for male participants.

For COG2, the full random effects structure was retained, and the following predictors were removed from the final model as they did not improve the fit of the model: three-way interaction ($p = 0.4223$), reduction:vowel interaction ($p = 0.5073$), reduction:gender interaction ($p = 0.2178$), and reduction ($p = 0.3771$). The results of the final model are summarized below in Table VIII, with /ʃi/ tokens produced by female participants as the baseline.

TABLE VIII: Linear mixed effects regression results: COG2 of /ʃ/.

	Estimate	Std. Error	*t*	
(Intercept)	5343.6	127.1	42.04	*
/u/	-795.4	156.9	-5.07	*
male	-752.8	116.5	-6.46	*
/u/:male	313.9	118.0	2.66	*

The fact that reduction was not a significant predictor means that the /ʃi/ tokens show comparable degrees of coarticulation with /i/ by the end of the consonant. There still was a significant lowering effect of /u/, however, suggesting that /u/ coarticulation begins early as shown in the COG1 results and remains throughout the consonant for both reduced and unreduced tokens.

Lastly, a linear mixed effects model was fit to the ΔCOG (COG2 − COG1) data to check whether the change from COG1 to COG2 were significantly different between reduced and unreduced tokens. The final model retained the full random effects structure, and the following non-significant predictors were removed from the fixed effects structure of the final model: three-way interaction ($p = 0.3216$), reduction:vowel interaction ($p = 0.8130$), reduction:gender interaction ($p = 0.6935$), and reduction ($p = 0.1653$). The results of the final model are shown below in Table IX, with /ʃi/ tokens produced by female participants as the baseline.

TABLE IX: Linear mixed effects regression results: ΔCOG (COG2 - COG1) of /ʃ/.

	Estimate	Std. Error	*t*	
(Intercept)	-124.92	155.24	-0.805	
/u/	-290.05	212.80	-1.363	
male	-100.97	85.07	-1.187	
male:/u/	244.59	92.58	2.642	*

The intercept of the final model was not significantly different from zero. Gender and vowel were also not significant, suggesting COG1 and COG2 are not significantly different from each other. The interaction term for gender and vowel was significant, but a separate analysis of the male participants showed that, like the female participants, the change in COG for /u/ tokens were not significantly different from /i/ tokens ($t = 0.564$). The non-significant results of ΔCOG seemingly contradicts the significant effect of reduction on COG1. A separate analysis of reduced and unreduced tokens revealed that while the intercept term for ΔCOG was not significantly different from zero for reduced tokens ($t = 0.403$), unreduced tokens did show a significant fall of 365.01 Hz ($t = -5.366$).

2. /ʧi, su/ COG results and analyses

COG2 is expected to be lower than COG1 for the affricate /ʧ/ as it begins with an alveolar constriction and moves back towards the alveo-palatal region for an /ʃ/-like frication. The possible effects of reduction are similar to that of /ʃi/ tokens: increased coarticulation would result in further lowering of COG values as the tongue shifts back towards the palate for /i/, while decreased coarticulation would lead to higher COG values. For /s/, increased coarticulation with /u/ would also lead to lower COG values, since lip rounding would lengthen the front oral cavity.

Shown below in Table X are the COG1 and COG2 values of /ʧ, s/. The overall pattern seems to be that reduction does not have a significant effect, suggesting that the degree of coarticulation is not different between reduced and unreduced tokens. Additionally, male participants have lower values for /ʧ/, but there seems to be no significant gender effect on /s/.

TABLE X: COG1 and COG2 mean (*standard deviation*) in Hz for /ʧ, s/.

C_1V		unreduced		reduced	
		female	*male*	*female*	*male*
ʧi	COG1	6397 (*687*)	5350 (*588*)	6185 (*751*)	5186 (*587*)
	COG2	5594 (*456*)	4803 (*406*)	5468 (*1102*)	4822 (*898*)
su	COG1	6118 (*1464*)	6154 (*879*)	6032 (*1196*)	5976 (*834*)
	COG2	6125 (*1076*)	6046 (*797*)	6026 (*1347*)	5977 (*1106*)

For /ʧ, s/, the fully complex model included reduction status, gender, and their interaction as predictors. The maximal random effects structure included random intercepts for participant and word, by-participant random slopes for reduction status, and by-word random slopes for gender.

For /ʧ/, the final models for COG1, COG2, and ΔCOG retained the full random effects structure and only gender as a predictor. Neither reduction nor the reduction:gender interaction were significant contributors to the COG1 model ($p = 0.1520$ and 0.9884, respectively), the COG2 model ($p = 0.9773$ and 0.4069, respectively), and the ΔCOG model ($p = 0.6599$ and 0.4337, respectively). The fact that reduction does not play a significant role in the COG results suggest that the degree of coarticulation between /ʧ/ and /i/ does not increase for reduced tokens, unlike /ʃi/. On the other hand, the results of the final models showed that male participants had lower values for both COG1 (-1006 Hz; $t = -5.16$) and COG2 (-760 Hz; $t = -7.0$), which was also the case for /ʃi/. The ΔCOG model also had a significant non-zero intercept at -745 Hz ($t = -5.033$), showing that COG2 is significantly lower than COG1 as predicted, regardless of reduction. Male participants were also shown to have a significantly smaller degree of change, where the drop in COG was 441 Hz. A separate analysis for just the male participants showed that the smaller lowering effect was still significant at $t = -3.243$.

For /s/, the final models for COG1, COG2, and ΔCOG retained the full random effects structure but none of the predictors. The fact that an intercept-only model fit the data equally well as a model with predictors shows that COG values for /s/ do not change throughout the consonant regardless of gender or reduction.

3. COG results and analyses of /h/ allophones

Although /ɸ, ç/ can contrast with /h/ depending on the lexical stratum (Ito and Mester, 1995; Moreton and Amano, 1999), /ɸ/ and /h/ neutralize to [ɸ] before /u/, while /ç/ and /h/ neutralize to [ç] before /i/ across all strata. Because /ɸ, ç/ are essentially identical in place with their respective neutralizing vowels, changes in COG are expected to come primarily from constriction strength rather than change in the length of the front oral cavity[3], where weakening constriction lowers the amplitude of the higher frequencies and results in a lower COG value overall (Hamann and Sennema, 2005; Kiss and Bárkányi, 2006). In other words, for /ɸ/, an increased coarticulation with /u/ would result in more lip rounding and weaker constriction, both contributing to lower COG values. For /ç/, increased coarticulation with the vowel would make the fricative more vowel-like with a weaker constriction, also resulting in lower COG values.

The COG1 and COG2 results are summarized in Table XI below. The overall pattern is that COG falls for unreduced /ɸu/ tokens but rises for unreduced /çi/. On the other hand, COG measures for reduced /ɸu, çi/ tokens both rise.

TABLE XI: COG1 and COG2 mean (*standard deviation*) in Hz for /ɸ/.

C_1V		unreduced		reduced	
		female	*male*	*female*	*male*
ɸu	COG1	1926 (*559*)	2014 (*824*)	2703 (*721*)	2872 (*1055*)
	COG2	1879 (*541*)	2044 (*891*)	2931 (*913*)	2816 (*1091*)
çi	COG1	3706 (*931*)	3682 (*833*)	4009 (*701*)	3793 (*760*)
	COG2	3951 (*919*)	3723 (*852*)	4579 (*673*)	4210 (*728*)

For /ɸ/, the final models for COG1 and COG2 retained the full random effects structure and only reduction as a predictor. Neither gender nor the reduction:gender interaction were significant contributors to the COG1 model ($p = 0.5886$ and 0.6246, respectively) and the COG2 model ($p =$

[3] Although, see Kumagai (1999) whose EPG study found that palatal constriction is more fronted before reduced vowels for [ɸ].

0.4454 and 0.0916, respectively), and thus were removed from the final model. Reduction had a significant raising effect of 813 Hz on COG1 (t = 5.418) and 905 Hz on COG2 (t = 4.097).

The final model for /ɸ/ ΔCOG retained the fully complex fixed and random effects structures. The results are summarized in Table XII below. Unreduced tokens produced by female participants are the baseline.

TABLE XII: Linear mixed effects regression results: ΔCOG of /ɸ/.

	Estimate	Std. Error	t value	
(Intercept)	-46.54	155.40	-0.300	
reduced	275.26	222.93	1.235	
male	82.03	77.49	1.059	
reduced:male	-366.14	122.12	-2.998	*

The fact that the intercept is not significantly different from zero suggests that COG1 and COG2 are not significantly different from each other for female participants in unreduced tokens. Reduction and gender did not have a significant effect on how the COG values change. However, the interaction term shows that COG rises less for male participants in reduced tokens. A separate analysis for the male participants showed that the change in COG is in fact not significant for the male participants (t = -0.502).

For /ç/, the final model for COG1 retained the full random effects structure but only the intercept for fixed effects. This means that /ç/ COG1 values were unaffected by either gender or reduction at the start of the consonant. The final model for COG2, however, retained reduction and gender but not their interaction for its fixed effects structure, which showed that reduction has a significant raising effect of 580 Hz towards the end of the consonant (t = 3.704) and that COG2 was lower for male participants by 349 Hz (t = -2.300).

For ΔCOG of /ç/, the model with the full random effects structure failed to converge. The fit of a model with only by-participant random slopes for reduction did not significantly differ from a model with only by-word random slopes for gender (p = 1.000). Since reduction was the only significant predictor for both COG1 and COG2, the random effects structure with by-participant random slopes for reduction was selected for the final model. For the fixed effects structure, the reduction:gender interaction was not a significant contributor to the model ($p = 0.6981$) and was removed from the final model.

TABLE XIII: Linear mixed effects regression results: ΔCOG of /ç/.

	Estimate	Std. Error	t value	
(Intercept)	236.38	70.56	3.350	*
reduced	348.43	90.42	3.854	*
male	-185.76	71.99	-2.580	*

The intercept for the final ΔCOG model was significantly higher than zero. In other words, COG2 was higher than COG1. Reduction, however, had a significant increasing effect of 794 Hz ($t = 5.033$), showing that COG rose even more in reduced tokens. Additionally, COG rose less for male participants.

The rising COG pattern for /çi/ is somewhat unexpected, since under the prediction stated above, a rising pattern would suggest a weakening coarticulation. A separate analysis of just the male participants revealed that ΔCOG for men actually was not significantly different from zero (41 Hz; $t = 0.400$), but reduction had a raising effect of 376 Hz ($t = 2.404$). In other words, the rising effect in unreduced tokens is present only in female participants. A closer examination of /çi/ tokens as produced by female participants revealed a pattern not observed in male participants or other fricatives. For female participants, the aperiodic noise for /çi/ typically began quite diffuse, with the lower frequencies gradually being lost until the concentration of the aperiodic noise stabilized just before the onset of /i/, contributing to the overall rise in COG. Examples of a typical waveform and spectrogram of unreduced /çi/ tokens for female participants (left) and male participants (right) are shown below in Figure 4 below from the word [çidoi] ‘harsh’.

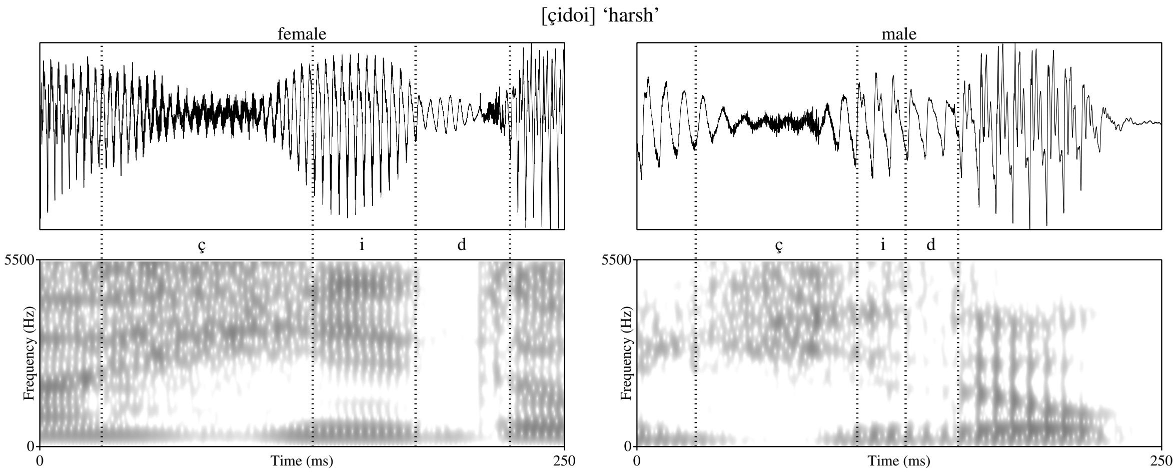

FIG 4: Waveform and spectrogram of unreduced C_1V in [çidoi], showing landmarks for C_1, vowel, and C_2 duration.

Although an articulatory study is required to verify the exact nature of /ç/ articulation in Japanese, there are two likely gestural explanations for the observed spectral pattern of /ç/ in female participants. First, the observed pattern is consistent with the lingual gesture for /ç/ beginning more [h]-like, further back in the oral cavity, then transitioning forward into the palatal place of articulation. If true, this suggests that the neutralization between /h/ and /ç/ before /i/ is actually incomplete, at least for female speakers. Second, the pattern is also consistent with /ç/ starting as a weak palatal fricative whose constriction strengthens over time. A combination of the two explanations is also consistent with the observed spectral pattern, both of which suggest that female and male speakers are using different variants of the |h| allophone. Regardless of which of the explanations is correct, the fact that the COG of /ç/ for reduced tokens continues to rise significantly beyond that of unreduced tokens, suggests that there is no anticipatory /i/ articulation that intervenes and halts the continued increase in the degree of palatal constriction between C_1 /ç/ and C_2 in reduced tokens.

4. COG results and analysis /k/

For /k/, COG measures were taken from the burst with a 20 ms window centered at the middle o the burst noise. Coarticulation with the high back vowel would lower the COG measures of /k/ due to the lengthened oral cavity through lip rounding. Increased coarticulation would lead to further lowering of COG. For the high front vowel, the COG will be higher as palatalization moves the tongue body forward, shortening the front oral cavity.

TABLE XIV: COG mean (*standard deviation*) in Hz for /k/.

	unreduced		reduced	
C_1V	*female*	*male*	*female*	*male*
ki	4707 (*634*)	4353 (*529*)	4727 (*852*)	4378 (*614*)
ku	1871 (*682*)	1708 (*526*)	2270 (*665*)	2277 (*877*)

To analyze the /k/ tokens, a model with the most complex fixed and random effects structures were fit first. The predictors were reduction, vowel, gender, and their interactions. The random effects structure included random intercepts for participant and word, by-participant random slopes for reduction and vowel, and by-word random slopes for gender. The final model retained the full random effects structure. The following non-significant predictors were removed from the fixed effects structure of the final model: three-way interaction ($p = 0.3106$), reduction:vowel interaction ($p = 0.3754$), reduction:gender interaction ($p = 0.4037$), gender:vowel interaction ($p = 0.1042$), and gender ($p = 0.2817$). The results of the final model are summarized in Table XV below. Unreduced /ki/ tokens are the baseline. Reduction had a significant raising effect 272 Hz, and the vowel /u/ had a significant lowering effect of 2505 Hz.

TABLE XV: Linear mixed effects regression results: COG of /k/.

	Estimate	Std. Error	*t*	
(Intercept)	4431.73	122.00	36.33	*
reduced	272.35	77.73	3.50	*
/u/	-2504.55	117.29	-21.35	*

5. Summary of COG results

The mean COG values for all C_1 are summarized in Table XVI below. C_1 that showed significantly more coarticulation for reduced tokens are noted with a '+' after the C_1V label, while C_1 that showed significantly less coarticulation for reduced tokens are noted with a '-' after the C_1V label. Also, although not noted in the table below, /ʃ, k/ both showed significant lowering of COG when followed by /u/ compared to /i/.

TABLE XVI: COG1 and COG2 mean (*standard deviation*) in Hz for all C_1. '+' = decreased coarticulation in reduced tokens, '−' = increased coarticulation in reduced tokens, **bold** = significant effect of vowel.

C_1V			unreduced *female*	unreduced *male*	reduced *female*	reduced *male*
ʃi	+	COG1	5694 (*622*)	4996 (*488*)	5201 (*862*)	4738 (*619*)
		COG2	5317 (*489*)	4592 (*342*)	5317 (*926*)	4695 (*787*)
ʃu	+	COG1	4948 (*606*)	4403 (*384*)	4924 (*758*)	4504 (*470*)
		COG2	4469 (*508*)	4060 (*409*)	4555 (*842*)	4311 (*633*)
ʧi		COG1	6397 (*687*)	5350 (*588*)	6185 (*751*)	5186 (*587*)
		COG2	5594 (*456*)	4803 (*406*)	5468 (*1102*)	4822 (*898*)
su		COG1	6118 (*1464*)	6154 (*879*)	6032 (*1196*)	5976 (*834*)
		COG2	6125 (*1076*)	6046 (*797*)	6026 (*1347*)	5977 (*1106*)
ɸu	−	COG1	1926 (*559*)	2014 (*824*)	2703 (*721*)	2872 (*1055*)
		COG2	1879 (*541*)	2044 (*891*)	2931 (*913*)	2816 (*1091*)
çi	−	COG1	3706 (*931*)	3682 (*833*)	4009 (*701*)	3793 (*760*)
		COG2	3951 (*919*)	3723 (*852*)	4579 (*673*)	4210 (*728*)
ki	−		4707 (*634*)	4353 (*529*)	4727 (*852*)	4378 (*614*)
ku	−		1871 (*682*)	1708 (*526*)	2270 (*665*)	2277 (*877*)

Center of gravity measurements for /ʃ/ tokens showed that coarticulation with /u/ begins early in the consonant regardless of reduction status. These early difference in COG between /ʃi/ and /ʃu/ tokens suggest an attempt to maximize recoverability through increased coarticulation. Furthermore, coarticulation with the following vowel was also shown to begin earlier in reduced /ʃi/ tokens. This effect was not found in /ʃu/ tokens, however, but given that /ʃi, ʃu/ differences are already apparent early in the consonant, further increasing the coarticulation for /ʃu/ perhaps is not necessary for recovery. In other words, only modifying one of the /ʃ/ contexts is enough to distinguish the two vowel possibilities.

/ʃi/ results can be compared directly with /ʧi/ results, since the two consonants share a place of articulation. Unlike /ʃi/, however, COG results for /ʧi/ did not show any effect of reduction, suggesting that there was no increase of coarticulation to aid recoverability. The same was true of /s/ tokens, which showed no effect of reduction.

An increase in coarticulation was expected to lower COG values for the two allophones of /h/, and both allophones in fact had higher COG values for reduced tokens, suggesting a decrease in coarticulation for reduced tokens. In other words, vowel gestures were not maintained to the same degree as in unreduced tokens perhaps because the allophonic variation was sufficient for recovery.

Lastly, the single COG measurement for /k/ showed that there is a difference of 2500 Hz between /ki/ and /ku/ tokens, which is more than six times greater than the 400 Hz difference that Japanese listeners were shown to be sensitive to for /ʃ/ in Beckman and Shoji (1984) and more than four times greater than the difference of $\sim$600–800 Hz found in the current study.

IV. DISCUSSION

The aim of this study was to investigate the acoustic properties of high vowel reduction in Japanese – specifically, what cues in the signal allow the recovery of a reduced vowel and whether gender and predictability from context affect the availability of these cues. The cues specifically tested for were coarticulatory effects of the target vowel on C_1, measured in the form of burst/frication duration and center of gravity (COG) of C_1.

With respect to the issue of lengthening, duration measurements showed that lengthening is observable only in non-fricatives. Reduction generally had no effect on fricatives, with the exception of /s/ which shortened in reduced contexts instead. The fact that C_1 lengthening is dependent on the manner of the consonant suggests that it is not an obligatory process whose goal is to maintain mora-timing (Han, 1994). Furthermore, the fact that [ʧ] lengthened while [ɸ, ç] did not despite similar durations suggests that C_1 lengthening is not a recoverability-conditioned process.

On the other hand, reduced /s/ tokens showed significant shortening while /ʃ/ did not, despite similar durations of $\sim$100 ms. Based on the COG results, it can be safely assumed that vocalic oral gestures are actively retained for /ʃ/, but less so for /s/. When both duration and COG results are considered together, the reason for shortening in /s/ is likely due to a lack of an intervening vowel gesture, which shortens the gestural timing for C_1 to C_2 transition. In other words, there is perhaps vowel deletion rather than devoicing when a high vowel is reduced after /s/. The question arises as to why it is only /s/ that shortens and none of the other obstruent C_1. One possible explanation is that reduced /sV.C_2V/ sequences are in fact being resyllabified as [sC_2V] with the initial /s/ being treated as part of an onset cluster, whereas other obstruents lengthen or remain unchanged because they are treated as belonging to a separate syllable (e.g., /hu.ku/ $\rightarrow$ [ɸ.ku]). In light of recent articulatory work by Kawahara et al. (2016), where it was found that Japanese speakers tend to retain the oral gestures of high vowels in reducing contexts either completely or not at all, the formation of clusters seems unproblematic. As to why only /s/ resyllabifies, it may have to do with special cluster-forming properties of /s/, which has long been noted in previous works (Selkirk, 1982; Kaye, 1992; Gierut, 1999; Morelli, 1999; Barlow, 2001).

COG results suggest that there is an effect of predictability on how early vowel coarticulation begins in Japanese high vowel reduction. The high predictability tokens showed either no change in C_1V coarticulation as in /ʧ, s/ or a possible decrease in /ɸ, ç/. For these consonants, the phonetic cues associated with the vowel are not essential for the recovery of the underlying high vowel when C_2 is voiceless because the reducible vowel that can follow is fully predictable. In other words, enhancing coarticulatory cues between C_1 and the target vowel does little to increase the likelihood of recovery.

The results for /ʃ/ on the other hand show that this is not the case for low-predictability contexts. Since both /i, u/ can follow the consonant, complete deletion of the vowel in these cases would jeopardize the recoverability of the vowel. Additional articulatory effort is required to transmit the contrastive information necessary for vowel recovery. As the COG results in this study and the spectral analysis in Beckman and Shoji (1984) have shown, oral gestures alone are enough to color the burst/frication noise of C_1 for reliable recovery. By retaining and overlapping the oral vowel gesture with the preceding consonant, maximal recoverability is obtained even in the absence of phonation. The idea that overlap of gestures are coordinated in order to preserve recoverability has been proposed by Silverman (1997) and Chitoran et al. (2002), and it was also suggested by Varden (2010) for Japanese.

The results for /k/ were less straightforward. First, /u/ had a significant lowering effect compared to /i/ like in the case of /ʃ/. The large spectral difference is most likely due to /k/-fronting that results from coarticulation with the following /i/, and positing the presence of coarticulatory effects even in reduced tokens allows /k/ to be grouped with /ʃ/. However, reduction also had a raising effect, suggesting a weakening coarticulation with the target high vowels, much like the two allophones of /h/, which are high predictability fricatives. A possible explanation as to why /k/ seemingly patterns with the high predictability consonants for reduced tokens is the large COG difference of 2,500 Hz between the burst noises of /ki/ and /ku/. This difference is nearly four times the differences of $\sim$600–800 Hz observed for /ʃ/ in the current study and nearly six times the 400 Hz spectral difference reported in Beckman and Shoji (1984), which Japanese speakers were shown to be sensitive to. Given such a large spectral difference, loss of some coarticulatory cues are unproblematic for recovery, since the difference is already quite obvious.

Lastly, gender did not seem to have an effect on the acoustic results, although male participants were shown to reduce more than the female participants, which confirms what Imai (2010) also found in younger speakers. What is interesting from the reduction results, however, is where the observed difference between men and women came from. With tokens in reducing environments having reduction rates of essentially 100%, the difference in reduction rates was clearly from the non-reducing tokens. An analysis of just the non-reducing tokens showed that reduction rates were significantly different for high predictability environments but not low predictability environments. In other words, predictability also seems to affect reduction rates, although only in men.

V. CONCLUSION

The results of the production experiment suggest that speakers provide more coarticulatory information on C1 burst/frication when the target vowel is unpredictable (i.e., after /k, ʃ/), supporting the results of previous studies which showed that the degree of coarticulation between segments are controlled to aid recoverability (Silverman, 1997; Chitoran et al., 2002). The results also provide novel insight into recoverability-driven coarticulation in that speakers not only increase the perceptibility of coarticulatory information when recoverability is in jeopardy (i.e., after /ʃ/) but that they also do the opposite, where speakers provide less or an unchanged amount of coarticulatory information when the normal amount of coarticulation is already highly perceptible (i.e, after /k/) or the vowel is highly predictable (i.e., after /ʧ, ɸ, s, ç/) and additional coarticulatory cues are unnecessary for recovery.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. BCS-1524133.

REFERENCES

Baayen, R. H., D. J. Davidson, and D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59:390–412.

Barlow, Jessica A. 2001. The structure of /s/-sequences: evidence from a disordered system. *J. Child Lang.* 28:291–324.

Barr, D. J., R. Levy, C. Scheepers, and H. J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* 68:255–278.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Software* 67:1–48.

Beckman, Mary. 1982. Segmental duration and the 'mora' in Japanese. *Phonetica* 39:113–135.

Beckman, Mary, and A. Shoji. 1984. Spectral and perceptual evidence for CV coarticulation in devoiced /si/ and /syu/ in Japanese. *Phonetica* 41:61–71.

Boersma, Paul. 2009. Cue constraints and their interactions in phonological perception and production. In *Phonology in perception*, ed. Paul Boersma and Silke Hamann, 55–110. Berlin: Mouton de Gruyter.

Browman, Catherine P., and Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49:155–180.

Chitoran, Ioana, Louis Goldstein, and Dani Byrd. 2002. Gestural overlap and recoverability: Articulatory evidence from Georgian. In *Papers in Laboratory Phonology VII*, ed. Natasha Warner and Carlos Gusshoven. Berlin: Mouton de Gruyter.

Cutler, Anne, Takashi Otake, and James M. McQueen. 2009. Vowel devoicing and the perception of spoken Japanese words. *J. Acoust. Soc. Am.* 125:1693–1703.

Faber, Alice, and Timothy J. Vance. 2000. More acoustic traces of "deleted" vowels in Japanese. In *Japanese/Korean Linguistics*, ed. Mineharu Nakayama and Charles J. Jr. Quinn, volume 9, 100–113.

Forrest, Karen, Gary Weismer, Paul Milenkovic, and Ronald N. Dougall. 1988. Statistical analysis of word-initial voiceless obstruents: preliminary results. *J. Acoust. Soc. Am.* 84:115–123.

Fowler, Carol A., and Elliot Saltzman. 1993. Coordination and coarticulation in speech production. *Lang. Speech* 36:171–195.

Fujimoto, Masako. 2015. Vowel devoicing. In *Handbook of Japanese Phonetics and Phonology*, ed. Haruo Kubozono, chapter 4. Mouton de Gruyter.

Fujimoto, Masako, Emi Murano, Seiji Niimi, and Shigeru Kiritani. 2002. Differences in glottal opening patterns between Tokyo and Osaka dialect speakers: Factors contributing to vowel devoicingevoicing. *Folia Phoniatrica et Logopedia* 54:133–143.

Gierut, Judith A. 1999. Syllable onsets: clusters and adjuncts in acquisition. *J. Speech. Lang. Hear. Res.* 42:708–726.

Hamann, Silke, and Anke Sennema. 2005. Acoustic differences between German and Dutch labiodentals. *ZAS Papers in Linguistics* 42:33–41.

Han, Mieko S. 1994. Acoustic manifestations of mora timing in Japanese. *J. Acoust. Soc. Am.* 96:73–82.

Hirose, Hajime. 1971. The activity of the adductor laryngeal muscles in respect to vowel devoicing in Japanese. *Phonetica* 23:156–170.

Imai, Terumi. 2010. An emerging gender difference in Japanese vowel devoicing. In *A Reader in Sociolinguistics*, ed. Dennis Richard Preston and Nancy A. Niedzielski, volume 219, chapter 6, 177–187. Walter de Gruyter.

Ito, Junko, and Armin Mester. 1995. Japanese phonology. In *Handbook of phonological theory*, ed. John Goldsmith, 817–838. Cambridge, MA: Blackwell.

Ito, Junko, and Armin Mester. 2003. Lexical and postlexical phonology in Optimality Theory: evidence from Japanese. *Linguistische Berichte* 11:183–207.

Kawahara, Shigeto, Jason Shaw, and James Whang. 2016. Targetless /u/ in Tokyo Japanese. In *Poster at the 16th Conference on Laboratory Phonology*. Ithaca, NY.

Kaye, Jonathan D. 1992. Do you believe in magic? The story of s+C sequences. *SOAS: Working Papers in Linguistics* 2:293–313.

Kindaichi, Haruhiko. 1995. *[Japanese Accent Dictionary].* Sanseido.

Kiss, Zoltán, and Zsuzsanna Bárkányi. 2006. A phonetically-based approach to the phonology of /v/ in Hungarian. *Acta Linguistica Hungarica* 53:175–226.

Kondo, Mariko. 2005. Syllable structure and its acoustic effects on vowels in devoicing. In *Voicing in Japanese*, ed. Harry van der Hulst, Jan Koster, and Henk van Riemsdijk, 229–246. Mouton de Gruyter.

Kubozono, Haruo. 2015. Loanword phonology. In *Handbook of Japanese Phonetics and Phonology*, ed. Haruo Kubozono, chapter 8, 313–362. Mouton de Gruyter.

Kumagai, Shuri. 1999. Patterns of linguopalatal contact during Japanese vowel devoicing. *The 14th Int. Cong. Phon. Sci.* 375–378.

Kuznetsova, Alexandra, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen. 2016. *lmertest: Tests in linear mixed effects models*. URL https://CRAN.R-project.org/package=lmerTest, r package version 2.0-30.

Maekawa, Kikuo, and Hideaki Kikuchi. 2005. Corpus-based analysis of vowel devoicing in spontaneous Japanese: an interim report. In *Voicing in Japanese*, ed. Jeroen van de Weijer, Kensuke Nanjo, and Tetsuo Nishihara. Mouton de Gruyter.

Mattingly, Ignatius G. 1981. Phonetic representation and speech synthesis by rule. In *The cognitive representation of speech*, ed. J. Myers, J. Laver, and Anderson J., 415–420. North-Holland Publishing Company.

McCarthy, John J. 1999. Sympathy and phonological opacity. *Phonology* 16:331–399.

Morelli, Frida. 1999. The phonotactics and phonology of obstruent clusters in Optimality Theory. Doctoral Dissertation, University of Maryland.

Moreton, Elliott, and Shigeaki Amano. 1999. Phonotactics in the perception of japanese vowel length: evidence for long-distance dependencies. *EUROSPEECH* 99:82–86.

Nittrouer, S., M. Studdert-Kennedy, and R. S. McGowan. 1989. The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *J. Speech Hear. Res.* 32:120–132.

Nogita, Akitsugu, Noriko Yamane, and Sonya Bird. 2013. The Japanese unrounded back vowel /ɯ/ is in fact rounded central/front [ʉ - ʏ]. *Ultrafest VI* .

Ogasawara, Naomi. 2013. Lexical representation of Japanese high vowel devoicing. *Language and Speech* 56:5–22.

Ogasawara, Naomi, and Natasha Warner. 2009. Processing missing vowels: Allophonic processing in Japanese. *Language and Cognitive Processes* 24:376–411.

Okamoto, Shigeko. 1995. “Tasteless” Japanese: less “feminine” speech among young Japanese women. In *Gender articulated: Language and the socially constructed self*, ed. Kira Hall and Mary Bucholtz, 297–325. New York: Routledge.

R Core Team. 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Selkirk, Elisabeth. 1982. The syllable. In *The structure of phonological representations*, ed. Harry van der Hulst and Norval Smith, volume Part II. Dordrecht: Foris Publications.

Sharoff, Serge. 2008. Lemmas from the internet corpus. URL `http://corpus.leeds.ac.uk/frqc/internet-jp.num`.

Shogakukan. 2013. [Daijisen Japanese Dictionary] (Digital Version). URL `http://dictionary.goo.ne.jp/`.

Silverman, Daniel. 1997. Phasing and Recoverability. Doctoral Dissertation, UCLA, Los Angeles.

Tsuchida, Ayako. 1994. Fricative-vowel coarticulation in Japanese devoiced syllables: Acoustic and perceptual evidence. *Working Papers of the Cornell Phonetics Laboratory* 9:183–222.

Tsuchida, Ayako. 1997. Phonetics and phonology of Japanese vowel devoicing. Doctoral Dissertation, Cornell University.

Tsuchida, Ayako, Shigeru Kiritani, and Seiji Niimi. 1997. Two types of vowel devoicing in Japanese: Evidence from articulatory data. *J. Acoust. Soc. Am.* 101:3177.

Vance, Timothy J. 2008. *The sounds of Japanese*. New York: Cambridge University Press.

Varden, J. Kevin. 1998. On high vowel devoicing in standard modern Japanese. Doctoral Dissertation, University of Washington.

Varden, J. Kevin. 2010. Acoustic correlates of devoiced Japanese vowels: velar context. *J. Eng. Am. Lit. Ling.* 125:35–49.

Varden, J. Kevin, and Tsutomu Sato. 1996. Devoicing of Japanese vowels by Taiwanese learners of Japanese. *Proceedings of Int. Conf. Spoken Lang. Processsing* 96.2:618–621.

Warner, Natasha, and Takayuki Arai. 2001a. Japanese mora-timing: A review. *Phonetica* 58:1–25.

Warner, Natasha, and Takayuki Arai. 2001b. The role of the mora in the timing of spontaneous Japanese speech. *J. Acoust. Soc. Am.* 109:1144–1156.

Yoshioka, H. 1981. Laryngeal adjustment in the production of the fricative consonants and devoiced vowels in Japanese. *Phonetica* 38:236–351.

Yoshioka, H., A. Löfqvist, and H. Hirose. 1982. Laryngeal adjustments in Japanese voiceless sound production. *J. Phon.* 10:1–10.

Zsiga, Elizabeth. 2000. Phonetic alignment constraints: consonant overlap in English and Russian. *Journal of Phonetics* 28:69–102.