Recoverability-driven gestural coordination:
Acoustic evidence from 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 whether the target vowel is devoiced (oral gesture is maintained) or deleted completely (oral gesture and voicing are both lost) is still debated. This study examines the effects of predictability on the degree of vowel reduction. Native Tokyo Japanese speakers (N=8) 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 /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 has no lengthening effect on C_1. COG results show that vowel coarticulatory effects are present in all consonants when a full vowel follows, as well as for /k, ʃ/ in reduced stimuli. In contrast, no such effect is found for /ɸ, s, ç/ in reduced tokens, suggesting a complete lack of vowel gestures. Predictable vowels, therefore, seem to delete, while unpredictable vowels devoice.

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

I. INTRODUCTION

A. Background

High vowel reduction is considered to be an integral feature of standard modern Japanese (Imai, 2010), so much so that dictionaries with explicit instructions for reducing environments exist (Kindaichi, 1995; NHK, 1985). The phenomenon is commonly described as involving phonemically short high vowels /i/ and /u/ (i.e., not /iː, uː/), which are reduced (generally assumed to be devoiced) in C_1VC_2 sequences when they are (i) unaccented and (ii) 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, namely /k/. Although an unaccented high vowel may also be reduced if the vowel is preceded by a voiceless fricative or affricate and is followed by a pause, as in /káʃu#/ → [káʃu̥#] 'singer)' (Han, 1962; Vance, 1987), this last case will not be discussed in the current study.

Despite the prevalence of high vowel reduction in Japanese and the amount of interest the phenomenon has received in the field of phonetics and phonology, how the reduced vowels are manifested acoustically is still debated. A survey of the literature on Japanese high vowel reduction reveals that 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, (ii) by devoicing the vowel and thus coloring the C_1 burst/frication noise with the retained oral gestures, and (iii) by deleting the vowel altogether. Each of the proposed manifestations has contradicting evidence as discussed below. It should be noted that although high vowel reduction is more commonly referred to as high vowel "devoicing" in Japanese, the term *reduction* is used throughout the paper as a general term to refer to a lack of phonation associated with a target vowel since there is disagreement regarding whether the target vowels are really devoiced or simply deleted.

Han (1962, 1994), whose primary focus was acoustic manifestations of mora-timing, reports that the frication noise of C_1 /s/ (where /s/ → [ʃi, su]) in a reduced syllable is significantly longer than other inherently long sounds (i.e., /h, k/, where /h/ → [çi, ɸu]) in unreduced syllables containing a voiced vowel, which Han attributes to mora-preservation. Kondo (2005), who looked at /k, t/ (where /t/ → [ʧi, t͡su]) also reports that the burst and aspiration of C_1 in reduced syllables were significantly longer than the consonant portion of their corresponding unreduced CV syllables. On the other hand, contradictory results are also reported where no significant difference in duration was found between consonants containing a reduced vowel and consonants preceding an unreduced vowel (Beckman, 1982; Beckman and Shoji, 1984; Faber and Vance, 2000). A possible explanation for the opposing results is differences in experimental methods. Beckman (1982) and Beckman and Shoji (1984) compare burst/frication duration of C_1 in reduced moras to their unreduced counterparts. Likewise, Faber and Vance (2000) compares [ʃi̥, su̥] to the unreduced counterparts [ʃi, su]. Han (1994), on the other hand, compares [ʃi̥, su̥] to other "inherently long sounds", which were /h, k/ in utterances where these sounds did not constitute a whole mora. While such comparisons may be reasonable under a working assumption that moras in Japanese are roughly equal in duration, onset consonants typically bear no metrical weight, and thus it is unclear whether onsets can contribute to a mora or even that acoustic duration is an appropriate measure (Faber and Vance, 2000). Furthermore, since Kondo (2005) does not look at fricatives but /k, t/, the apparent lengthening effect may simply be due to different articulatory properties of stops and affricates versus fricatives.

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 (Fujimoto *et al.*, 2002; Sawashima, 1971; Sawashima *et al.*, 1978; Tsuchida *et al.*, 1997; Yoshioka *et al.*, 1982) 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. Since there is no laryngeal activity associated with C_2 apart from the carry-over from that of C_1, 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 voicing, the presence of formant-like structures in the burst/frication noise of C_1 are reported, which is taken as evidence of retained oral vowel gestures. A more recent acoustic study by Varden (2010) also 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; Martin, 1987; Vance, 1987, 2008; 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]).

B. Overview

As the survey of the literature above has shown, there is little consensus on how reduced high vowels are manifested acoustically, which seems to stem mainly from two issues that this study aims to resolve. The first issue is that there is a lack of experimental and terminological consistency across the studies. While it is true that there seems to be conflicting evidence for each of the three ways in which reduced vowels have been argued to be manifested, the experimental designs and methods often differed from each other, making it difficult to directly compare the results. Specifically, in the case of C_1 lengthening, Han (1994) compared /s/ to /h, k/ to argue for lengthening, while the opposing side compared C_1 from reduced syllables to their unreduced counterparts (Beckman, 1982; Beckman and Shoji, 1984; Faber and Vance, 2000). Since the contribution of onset consonants to mora timing remains unclear (Faber and Vance, 2000), the current study compares C_1 from corresponding reduced and unreduced syllables. Similarly, the debate over whether a reduced vowel devoices or deletes has generally relied on the presence of apparent formant-like structures riding on C_1 burst/frication noise, but as Beckman and Shoji (1984) showed, spectral measurements can reveal significant differences despite inconsistently present formants. In the current study, center of gravity (COG) – the amplitude weighted mean of frequencies present in the signal (Forrest *et al.*, 1988) – is measured for C_1 to verify the presence of vowel coarticulatory effects. Related to the differing methodologies is that the terms "devoiced" and "deleted" are often defined differently or imprecisely depending on the study, where the term "deleted" in one study is actually synonymous to the term "devoiced" in another. For the purposes of this paper deletion is defined as the loss of both voicing and oral gestures associated with a target vowel; and devoicing is defined as the loss of voicing but the retention of oral gestures which results in a significant and perceptible acoustic effect on the burst/frication noise of the preceding consonant, C_1.

The second issue the current study aims to address is the question of how recoverability affects the production of reduced vowels. Recoverability is a notion that refers to the ease in which the underlying form – stored mental representations – can be accurately accessed from a given surface form – actual, variable output signals (e.g., [kæt̠̚, kæt̠ʰ] → /kæt̠/ 'cat') (Chitoran *et al.*, 2002; Mattingly, 1981; McCarthy, 1999). Recovery can be direct if acoustic information is explicitly present in the signal (direct recovery) or predicted from the context (predictability). However, recoverability can be compromised if information in the signal or predictability from context is insufficient. Varden (2010) states what seems to be a prevalent assumption in the literature on Japanese high vowel reduction, which is that since high vowels trigger allophonic variation for /t, s, h/ (i.e., /t/ → [ʧi, t͡su]; /s/ → [ʃi, su]; /h/ → [çi, ɸu]), the underlying vowel is easily recoverable even if the vowel were to be "deleted altogether" unlike in cases where the vowel is unpredictable. To give a concrete example, the simple presence of [ɸ] can and should be analyzed as /hu/, since [ɸ] can only occur as an allophone of /h/ preceding /u/ in non-loanwords.

If it is true that the degree of high vowel reduction is dependent on the vowel's predictability, this also leads to the prediction that if predictability alone is not reliable enough, sufficient acoustic cues of the reduced vowel should be retained in order to disambiguate between the possible vowels and allow for direct recovery. Similar recoverability-conditioned gestural coordination has been proposed by Chitoran *et al.* (2002) and Silverman (1995), as well as other related studies. However, whether vowels in contexts with higher predictability do in fact reduce more in Japanese has never been tested systematically. The current study therefore investigates /ɸ, s, ç/, after which only one of the two high vowels can follow, and /ʃ, k/, after which both high vowels can follow. The consonants included in the current study and the vowels that can follow each of the consonants is summarized in Table I below. It should be noted that while [ʃ] is an allophone of /s/ preceding /i/, /ʃ/ is also phonemic and can precede all vowels in Japanese, with the exception of /e/ outside of loanwords. The bilabial stop /p/ is excluded due to its rareness outside of loan and onomatopoetic words. Furthermore, the affricates [t͡s, ʧ], which are allophones of /t/ before [u, i] respectively, are also not included for the sake of simplicity.

TABLE I: Stimuli consonants and possible following vowels. "–" means that the vowel is not phonologically possible in this context (in non-loanwords).

		i	u
Predictable	ɸ	–	✓
	s	–	✓
	ç	✓	–
Unpredictable	k	✓	✓
	ʃ	✓	✓

There are three possible outcomes to the question regarding the effects of a target vowel's predictability. The first is that high vowel reduction is blind to predictability and driven primarily by Japanese phonotactics, which has a strict CVCV structure (Shibatani, 1990) and thus disallows tautosyllabic clusters. If this is the case, then the vowel would never delete completely but always devoice instead, coloring the burst or frication noise of C_1 to signal the presence of the target vowel (Varden, 2010; Beckman and Shoji, 1984). The second is also that predictability has no effect, but in the sense that variation between deletion and devoicing is random. Ogasawara and Warner (2009) found in a lexical judgment task that when Japanese listeners are presented with unreduced forms of words where reduction is typically expected, reaction times are 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 (Ogasawara, 2013; Cutler *et al.*, 2009). The third and last option is that high vowel reduction is constrained by recoverability. In this case, devoicing would be favored when the predictability of the target vowel is unreliable from a given C_1 to allow direct recovery as in the case of /k, ʃ/, while deletion would be favored when predictability is high, as in the case of /ɸ, s, 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 not entirely random, but dependent on predictability from context.

Lastly, while this paper 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; Yuen and Hubbard, 1998) 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 may be being actively utilized as a feature of gendered-speech. If this is true, then high vowel reduction could not be a purely phonological or a phonetic process, and thus an equal number of men and women were chosen to participate in this study to investigate any gender-based differences.

II. MATERIALS AND METHODS

A. Participants

Eight native Japanese speakers participated in this study. All were undergraduate and graduate students recruited in New York City, whose length of residence in the United States ranged from one to five years. There were four female and four male speakers. All speakers were born and raised in the greater Tokyo area and did not start learning English until between 12-15 years of age as part of their compulsory secondary education. The participants ranged in age from 22-25 years, with the exceptions of one 38-year-old female and one 31-year-old male. All participants were compensated for their time.

B. Materials

The stimulus set in this study was a 45-word subset of a larger study. All stimuli were non-loan, lexical words, 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. 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.

There were 25 target and 20 control tokens. Examples of the stimuli are shown in Table II below. A complete list of the stimuli can be found in the Appendix. All stimuli contained a word-initial C_1iC_2 or C_1uC_2 sequence in which the vowel is unaccented. Regardless of the stimulus type, C_1 was either a voiceless stop or a fricative (i.e., [k, ʃ, ɸ, s, ç]). In target tokens, C_2 was always a voiceless stop (i.e., [p, t, k]), resulting in stop-stop and fricative-stop contexts. These two contexts were chosen for two reasons: (i) these are contexts in which high vowel reduction is reported to occur systematically and categorically (Fujimoto, in press), and (ii) the C_2

stop closure clearly marks where the previous segment ends. Five of the target tokens were ʃ-initial tokens with voiceless geminate stops as C_2. Because reduction rates are reported to be lower when C_2 is a geminate than when C_2 is a singleton (Maekawa and Kikuchi, 2005), the geminate tokens were excluded from the overall reduction rate calculations. However, a post-hoc linear regression analysis of C_1 duration and center of gravity (COG) for these five tokens revealed that neither the reduced nor the unreduced geminate tokens were significantly different from their reduced target and unreduced control counterparts acoustically (Duration: $F(3,80) = 2.485$; ns. COG1: $F(3,80) = 1.046$; ns. COG2: $F(3,80) = 0.5893$; ns.). The five geminate tokens therefore were included for duration and COG measurements. The control tokens were similar to the target tokens except that C_2 was always a voiced rather than a voiceless stop (i.e., [b, d, g]). Because 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 control tokens.

TABLE II: Example of stimuli by type and context

stimulus type	*context*	*example*	*gloss*
target	stop-stop	kikeɴ	‘danger’
		kuki	‘twig’
	fricative-stop	ʃika	‘deer’
		ɸukoː	‘unhappy’
control	stop-stop	kiguː	‘coincidence’
		kugi	‘nail’
	fricative-stop	çige	‘facial hair’
		sugoi	‘amazing’
geminate	fricative-stop	ʃippai	‘failure’
		ʃukkiɴ	‘going to work’

C. Design and procedure

The carrier sentences were presented one at a time to the participants on a computer monitor as a slideshow presentation and followed the common Japanese orthographic practice of mixing native Japanese syllabic scripts (*hiragana* and *katakana*) and Chinese characters (*kanji*). Only the Chinese characters in the official *jouyou kanji* list (The Agency for Cultural Affairs, 2012) were used. *Jouyou kanji* is a list of 2,136 characters announced by the Japanese Agency for Cultural

Affairs as appropriate for daily use. High school graduates are expected to be able to read and write all 2,136 characters. Reading aids in *hiragana* above Chinese characters, commonly known as *hurigana*, were also used for both stimuli and non-stimuli words. There was one filler sentence at the beginning to help the participants settle in before producing any of the target words.

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. The experimenter remained with the speaker during the recording and prompted the speaker to repeat any sentences that were produced with errors. All participants were recorded in a sound-attenuated booth with a Shure Beta 85A microphone attached to a Marantz PMD-670 digital recorder. The microphone was placed 3-5 inches from the mouth of the speaker.

D. Data Analysis

Once the participants were recorded, the waveform and spectrogram of each participant were examined in Praat (Boersma and Weenink, 2013) 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. Because visual inspection has been shown to be an inadequate method for determining whether the vowel is devoiced or deleted, the following visual criteria were used to code each token simply for “reduction”, a term used throughout this paper when the distinction between devoicing and deletion is unnecessary or cannot be made reliably. For control tokens, which always had a voiced, oral C_2, vowels were coded as unreduced when the aperiodic burst/frication noise of C_1 was followed by a periodic waveform with clear phonation containing formant structures as shown in Figure 1 below.

Control tokens were coded as reduced when no formant structures were apparent between the burst/frication noise of C_1 and the stop closure of C_2. Below in Figure 2 is an example of a reduced control token, /ku̥gi/ ‘nail’. Note that there is no phonation that coincides with a formant structure between the C_1 burst and C_2 closure.

Target tokens were coded as unreduced if there was any voicing between C_1 and C_2, and the

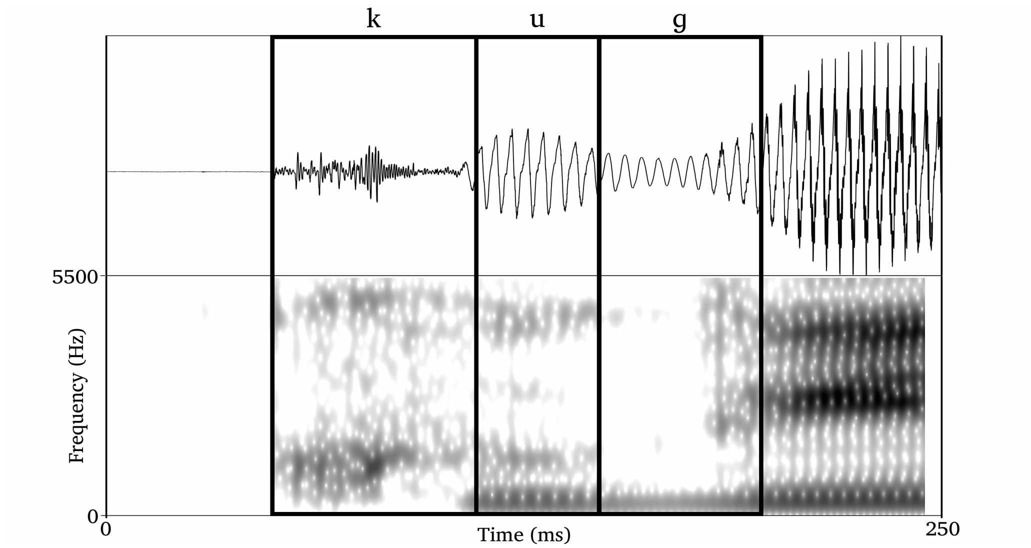

FIG 1: Spectrogram of [kugi] 'nail' with a voiced vowel, showing landmarks for C_1, vowel, and C_2 duration.

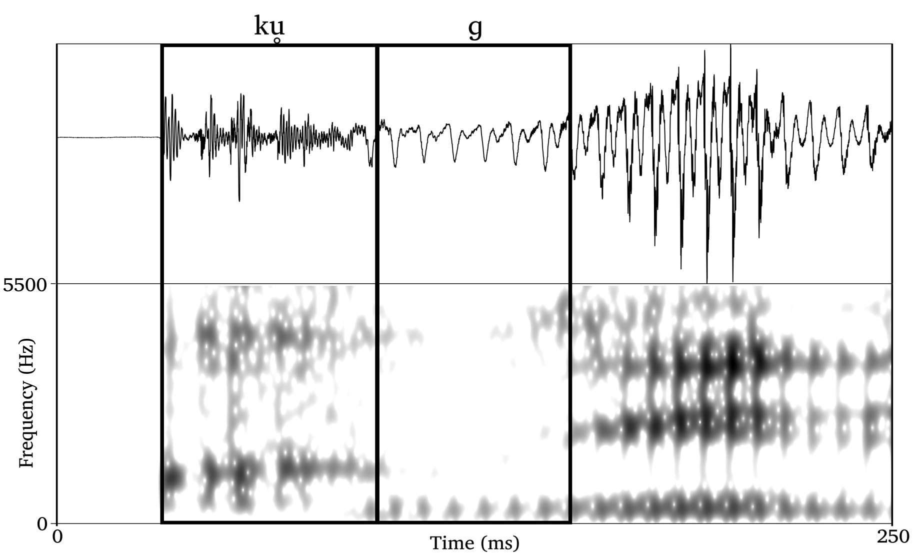

FIG 2: Spectrogram of [ku̥gi] 'nail' with a reduced vowel, showing landmarks for C_1 + vowel and C_2 duration.

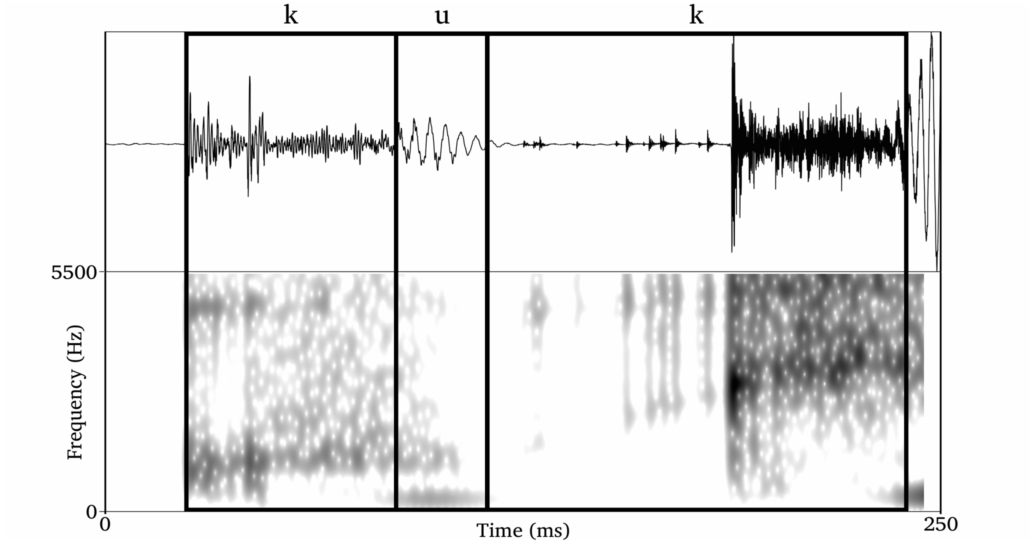

FIG 3: Spectrogram of [kuki] ‘twig’ with a voiced vowel, showing landmarks for C_1, vowel, and C_2 duration.

voiced portion was segmented separately as a vowel even if no formant structure was apparent. This criterion was based on articulatory results which showed that only one laryngeal gesture is present in reducing environments, where the feature [+spread glottis] from C_1 spreads to C_2 with no intervening voicing gesture for the vowel (Tsuchida, 1997). In other words, the presence of any voicing between C_1 and C_2 suggests that an attempt to voice the vowel was being made, interrupting the single [+spread glottis] feature from spreading from C_1 to C_2. An example of an unreduced target token is shown in Figure 3 below.

Target tokens with no voicing between C_1 and C_2 were coded as reduced. An example is shown in Figure 4 with the word [ku̥kaku] ‘a block’.

Once all tokens were coded for reduction status, duration measurements were taken to investigate whether vowel reduction results in lengthened C_1 burst/frication noise. For stop C_1, duration measurements included only the aperiodic burst energy and excluded the silence from closure. For fricative C_1, the measurement included the entire high frequency aperiodic energy. 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 F2 (for control tokens) or the onset of

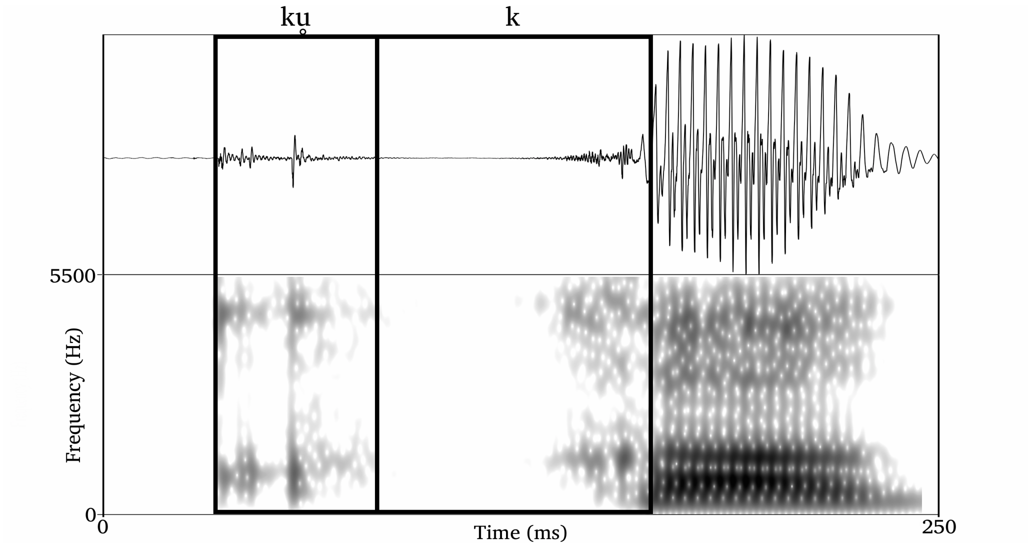

FIG 4: Spectrogram of [ku̥kaku] 'a block' with a reduced vowel, showing landmarks for C_1 + vowel, and C_2 duration.

voicing (for target tokens).

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. The purpose of the COG measurements was to look for any vowel coarticulation effects on the burst or frication noise. Before measuring COG values, the sound files were high pass filtered at 200 Hz to mitigate the effects of f0 on the burst/frication noise. The filtered sound files were then down-sampled at 20,000 Hz. As a result, the COG values measured were taken from FFT spectra in the band of 200 to 10,000 Hz. Midpoints were calculated automatically via Praat, and separate COG measures were taken for the first half (COG1) and the second half (COG2) of the burst/frication noise. The prediction here was that there is an inverse relationship between COG and vowel overlap, since greater coarticulation with a vowel would weaken the strength of oral constrictions, which would consequently lower the amplitude of the higher frequencies. If there is a significant drop between COG1 and COG2, this would mean that the vowel overlaps more with the latter portion of C_1. Alternatively, if there is no drop, this would suggest that the degree of overlap has not increased. Simply, more vowel overlap equals lower COG value.

III. RESULTS

All statistical analyses were performed by fitting mixed effects models using the lme4 package (Bates *et al.*, 2011) for the R software (R Core Team, 2013). Random intercepts by subject and word were included in all models, unless otherwise noted.

A. Reduction rate

The overall reduction rate by stimulus type and context (excluding geminate tokens) are shown in III below. One fricative-stop target token was removed from the analysis because one male participant misread the word /ʃitᴇɴ/ 'point of view' as the synonymous but orthographically different word /kᴀɴtᴇɴ/ 'point of view'. The lower rate of reduction for the stop-stop context was due to two female speakers. The older of the two female participants (and also the oldest among the group) was 38 years-old and did not reduce three of ten stop-stop target tokens. The other female participant was 22 years old and did not reduce one of ten stop-stop target tokens.

TABLE III: Reduction rate by C_1-C_2 manner and stimulus type.

	target	control
stop-stop	95% (n = 80)	5% (n = 80)
fricative-stop	100% (n = 79)	13% (n = 80)

The overall difference in reduction rates between men and women were not significant, but the difference between control and target tokens was significant as shown in Table IV below.

TABLE IV: Mixed logit model results for difference in reduction rate by gender and token type.

	Estimate	Std. Error	z	p	
intercept	-1.4644	1.3958	-1.049	0.2941	
women	-1.4470	0.8453	-1.712	0.0869	*n.s.*
target	9.0121	1.7465	5.160	2.47e-07	***

B. Center of gravity

In order to verify whether a vowel gesture overlaps with the consonant preceding it, center of gravity (COG) measurements were taken for the first half (COG1) and the second half (COG2) of the burst/frication noise of C_1. As the results will show, the degree of overlap for reduced vowels on C_1 seems to depend on the predictability of the vowel after certain consonants. The term *devoiced*, therefore, is used strictly to refer to cases where vowel coloring is evident on the burst/frication noise of C_1, since the coloring suggests that only the voicing is lost and vocalic oral gestures are still being made. The term *deleted* is used when there is no vowel coloring on C_1, since the lack of coloring suggests that both the voicing and vowel gestures are absent.

A mixed effects linear regression model was fit for COG using the *lmer()* function in the lme4 package (Bates *et al.*, 2011) for the R software (R Core Team, 2013). Since there is currently no widely accepted method of accurately calculating *p*-values for mixed effects models, the results were considered significant if $|t| > 2$ (Gelman and Hill, 2007). Reduction, gender, and interactions were included as predictors for all models, as well as random intercepts by subject and word and random slopes for reduction by subject and word. In addition to reduction and gender, vowel type was also included as a predictor for /ʃ, k/ since both /i, u/ can follow them. For /ɸ, s, ç/, vowel as a predictor was irrelevant because only one vowel could follow after the fricatives, where they will always be [ɸu, su, çi].

1. High-predictability contexts: /ɸ, s, ç/

For high-predictability contexts, none of the predictors had a significant effect on COG1 ($|t| < 2$), as shown in Tables V-VII. For all models, unreduced tokens as produced by men were the baseline.

TABLE V: Results of a mixed effects linear model for COG1 of /ɸ/.

	Estimate	Std. Error	*t*	
intercept	1533.9	811.1	1.891	
reduced	824.9	748.3	1.102	*n.s.*
women	-178.2	1027.8	-0.173	*n.s.*
reduced:women	-73.5	874.5	-0.084	*n.s.*

TABLE VI: Results of a mixed effects linear model for COG1 of /s/.

	Estimate	Std. Error	*t*	
intercept	6796.1	491.3	13.832	
reduced	-939.1	1096.2	-0.857	*n.s.*
women	582.7	691.1	0.843	*n.s.*
reduced:women	323.6	614.8	0.526	*n.s.*

TABLE VII: Results of a mixed effects linear model for COG1 of /ç/.

	Estimate	Std. Error	*t*	
intercept	2636.2	492.3	5.354	
reduced	965.9	523.3	1.846	*n.s.*
women	892.8	616.3	1.449	*n.s.*
reduced:women	-725.6	655.5	-1.107	*n.s.*

In the case of COG2, reduction had a significant effect, where reduced tokens tended to have higher COG2 values than unreduced tokens. Gender and the interaction between gender and reduction did not have a significant effect, as shown in Tables VIII-X.

TABLE VIII: Results of a mixed effects linear model for COG2 of /ɸ/.

	Estimate	Std. Error	*t*	
intercept	1024.00	541.45	1.891	
reduced	2259.88	854.34	2.645	*
women	83.75	707.16	0.118	*n.s.*
reduced:women	0.25	690.47	0.000	*n.s.*

TABLE IX: Results of a mixed effects linear model for COG2 of /s/.

	Estimate	Std. Error	*t*	
intercept	5273.9	598.2	8.816	
reduced	1211.8	456.8	2.653	*
women	1328.6	842.1	1.578	*n.s.*
reduced:women	-833.9	640.8	-1.301	*n.s.*

TABLE X: Results of a mixed effects linear model for COG2 of /ç/.

	Estimate	Std. Error	*t*	
intercept	2664.50	468.31	5.690	
reduced	1544.87	756.88	2.041	*
women	369.16	628.09	0.588	*n.s.*
reduced:women	-62.29	815.34	-0.076	*n.s.*

Table XI below summarizes the mean COG1, COG2, and ΔCOG (= COG2 – COG1) values for each of the fricatives with standard deviations in parentheses. As the ΔCOG column in Table XI shows, the general pattern is that COG falls (negative ΔCOG) when the vowel is unreduced, which suggests a higher degree of vocalic gestural overlap in the second half of the fricatives. In contrast, COG rises (positive ΔCOG) when the following vowel is reduced. Since the presence of an unreduced vowel lowered COG2 relative to COG1, the lack of a similar effect in reduced tokens suggests that there is no intervening oral vowel gesture before the stop closure. The raising effect of reduction on COG2 is most likely the result of a narrowing of the oral constriction as the tongue nears a full closure for the following C_2 consonant, which was always a stop. This further suggests that there is no intervening vowel gesture. In other words, the vowel is being deleted in these cases.

TABLE XI: Mean COG values and standard deviations for /ɸ, s, ç/ in Hz.

		COG1	COG2	ΔCOG
ɸ	unreduced	1445 (1508)	1066 (1024)	-379 (543)
	reduced	2233 (742)	3326 (994)	1093 (1193)
s	unreduced	7091 (1098)	5967 (1404)	-1124 (1027)
	reduced	6310 (1521)	6733 (1004)	422 (1138)
ç	unreduced	3081 (1195)	2830 (1404)	-251 (1021)
	reduced	3686 (692)	4363 (611)	677 (1006)

Although these results could have stemmed from the fact that the three high predictability contexts all involved fricatives, the results from /ʃ/ suggest that it is not the manner of the consonant but the predictability of the following vowel that matters as discussed below.

2. Low predictability contexts: /ʃ, k/

For low-predictability contexts, the overall picture shows that while vowel type has a significant effect, reduction does not. Where each of the effects are evident (COG1 or COG2), however, was different for the two consonants. For all models, unreduced tokens with the vowel /i/ as produced by men were the baseline.

In the case of /ʃ/, none of the predictors had a significant effect on COG1, except for the interaction of vowel type and gender, suggesting that women had a greater lowering effect on COG1 by the vowel /u/ than men. The results of the mixed effects linear model are shown in Table XII and Figure 5 below.

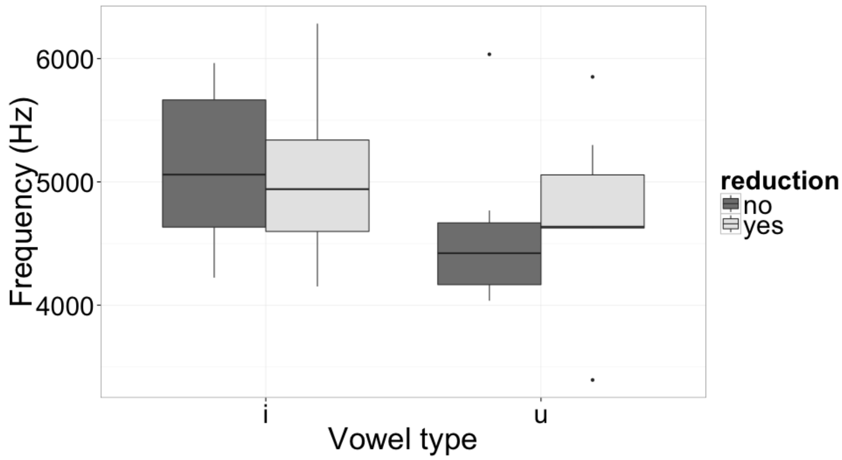

FIG 5: COG1 of [ʃ] frication by vowel type and reduction status.

TABLE XII: Results of a mixed effects linear model for COG1 of /ʃ/.

	Estimate	Std. Error	*t*	
(intercept)	4869.89	283.60	17.172	
vowel-u	-105.64	222.39	-0.475	*n.s.*
reduced	50.43	202.95	0.248	*n.s.*
women	367.31	393.01	0.935	*n.s.*
u:reduced	-16.43	361.59	-0.045	*n.s.*
u:women	-740.31	266.38	-2.779	*
reduced:women	-106.78	257.28	-0.415	*n.s.*
u:reduced:women	393.78	365.48	1.077	*n.s.*

For COG2, vowel type had a significant effect but none of the other predictors did. The results of the mixed effects linear model are shown in Table XIII and Figure 6 below.

TABLE XIII: Results of a mixed effects linear model for COG2 of /ʃ/.

	Estimate	Std. Error	*t*	
(intercept)	4723.865	242.739	19.461	
vowel-u	-674.365	334.876	-2.014	*
reduced	-108.982	170.442	-0.639	*n.s.*
women	87.425	307.108	0.285	*n.s.*
u:reduced	210.482	542.393	0.388	*n.s.*
u:women	-36.425	264.684	-0.138	*n.s.*
reduced:women	-9.327	174.622	-0.053	*n.s.*
u:reduced:women	-69.173	364.821	-0.190	*n.s.*

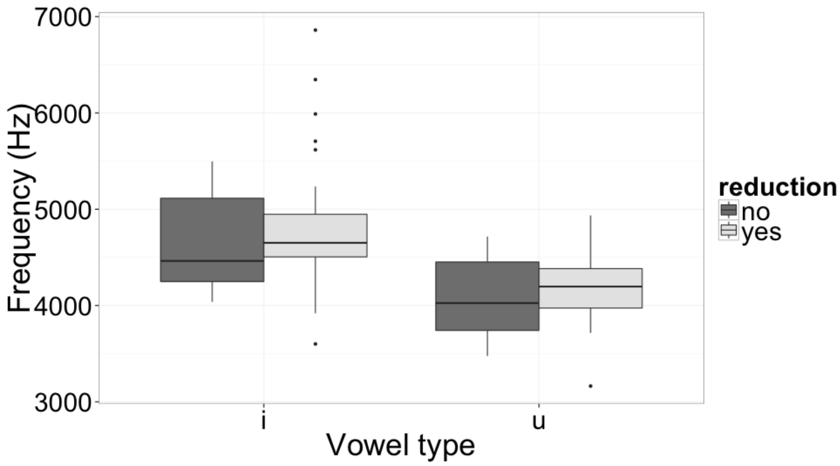

FIG 6: COG2 of [ʃ] frication by vowel type and reduction status.

To summarize the key results of /ʃ/, reduction had no significant effect and the vowel /u/ had a significant lowering effect regardless of reduction status, although only for women in COG1 and for both men and women in COG2. The fact that COG values for reduced and unreduced tokens are not significantly different suggests that a comparable degree of vocalic oral gesture is being made throughout the consonant regardless of reduction status. COG2 was lower than COG1 for both reduced and unreduced tokens, unlike the high-predictability fricatives, where COG2 was lower than COG1 in unreduced tokens only and higher in reduced tokens. The fact that COG2 lowers even in reduced tokens suggests that vowel reduction results in devoicing rather than deletion. Given these results, where oral gestures for the vowel seem to be made regardless of reduction, the overall lowering effect of /u/ is unsurprising, since the back vowel /u/ has a larger front oral cavity than the front vowel /i/ and thus has a lowering effect on high-frequency burst/frication noises.

For /k/, vowel type had a significant effect on both COG1 and COG2 similar to /ʃ/, but reduction also had a significant effect on COG2, much like the high-predictability consonants /ɸ, s, ç/. As shown in Table XIV below, COG1 is lower when the vowel is /u/, but reduction has no significant effect within each vowel type (Figure 7).

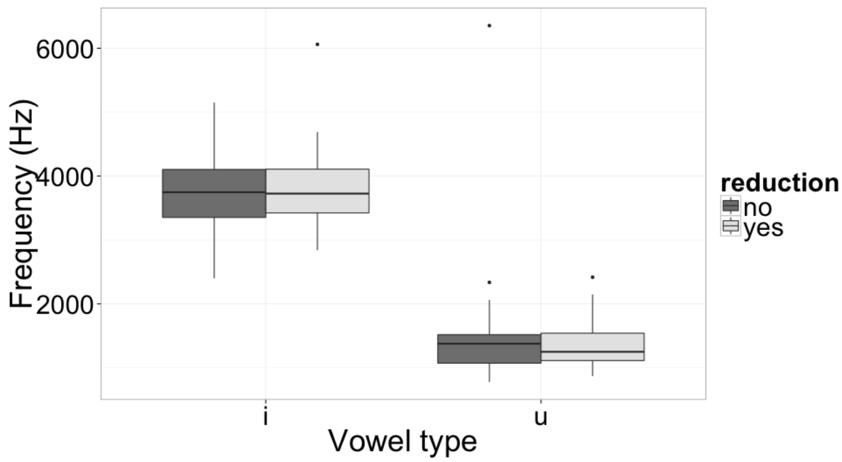

FIG 7: COG1 of [k] burst by following vowel and reduction status.

TABLE XIV: Results of a mixed effects linear model for COG1 of /k/.

	Estimate	Std. Error	*t*	
(intercept)	3988.77	231.58	17.224	
vowel-u	-2353.94	245.30	-9.596	*
reduced	11.38	235.68	0.048	*n.s.*
women	-489.52	287.54	-1.702	*n.s.*
u:reduced	-303.26	319.06	-0.950	*n.s.*
u:women	211.49	260.57	0.812	*n.s.*
reduced:women	53.97	76.14	0.195	*n.s.*
u:reduced:women	313.76	367.12	0.855	*n.s.*

The fact that vowel type had such a large effect on COG1 is unsurprising. Because /k/ is a stop consonant, the vocalic gesture of the following vowel most likely began during the stop closure (Browman and Goldstein, 1992; Fowler and Saltzman, 1993), resulting in a high degree of gestural overlap by the time the stop is released. This, however, does not explain why the COG2 results behave like the predictable cases. As shown in Table XV below, while the lowering effect of the vowel /u/ is still significant on COG2, reduction also has a significant raising effect (Figure 8). In addition, women also had significantly lower COG2 than men (Figure 9). The interaction between vowel type and reduction status was significant, suggesting that the difference in COG2 between vowel types is smaller for the vowel /u/. The interaction between vowel type and gender was also significant, suggesting that the difference in COG2 between vowel types is smaller for

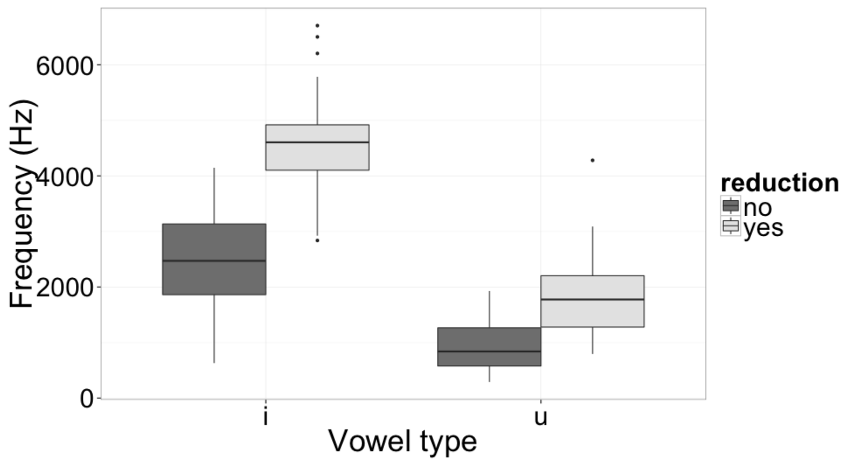

FIG 8: COG2 of [k] burst by following vowel and reduction status.

women than men. However, the interactions of reduction/gender and all three predictors were not significant, which shows that the overall pattern of COG2 raising in reduced tokens is the same regardless of vowel type and gender.

TABLE XV: Results of a mixed effects linear model for COG2 of /k/.

	Estimate	Std. Error	*t*	
(intercept)	3000.4	184.5	16.261	
vowel-u	-1947.5	226.3	-8.606	*
reduced	1752.6	347.6	5.042	*
women	-1051.2	258.6	-4.065	*
u:reduced	-962.5	324.8	-2.964	*
u:women	801.4	309.2	2.592	*
reduced:women	628.5	480.2	1.309	*n.s.*
u:reduced:women	-486.3	437.1	-1.113	*n.s.*

To summarize the key results of /k/, presented in Table XVI below are COG results broken down by vowel type. While the overall values of /ku/ are lower than /ki/, the effect of reduction mirrors that of the high-predictability fricatives where COG falls (negative ΔCOG) in unreduced tokens but rises (positive ΔCOG) in reduced tokens within each vowel type. Reasons for the observed behavior of /k/ will be discussed in further detail in §

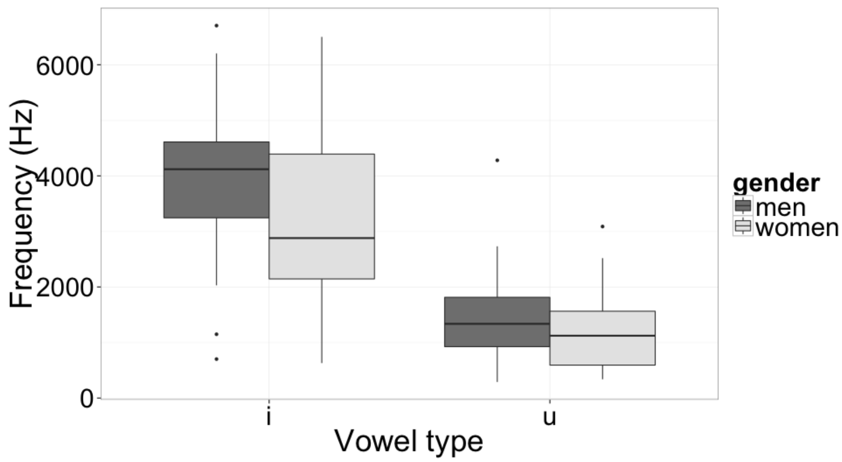

FIG 9: COG2 of [k] burst by following vowel and gender.

TABLE XVI: Mean COG values and standard deviations for /ki, ku/ in Hz.

		COG1	COG2	ΔCOG
ki	unreduced	3743 (621)	2456 (973)	-1287 (1064)
	reduced	3782 (606)	4542 (858)	759 (766)
ku	unreduced	1481 (897)	901 (410)	-580 (887)
	reduced	1385 (396)	1821 (718)	436 (515)

C. Duration

As with the center of gravity analysis, a mixed effects linear regression model was fit to analyze the duration of C_1 using the *lmer()* function in the lme4 package (Bates *et al.*, 2011) for the R software (R Core Team, 2013). Random intercepts by subject and word were included in the model, as well as random slopes for reduction by subject and word. For each of the C_1 types, reduction, vowel type, and gender were included as predictors. Control tokens containing an unreduced vowel were the baseline. Although rare, unreduced target tokens and reduced control tokens were excluded from analysis because their anomalous production of the underlying vowel may have affected the C_1 as well.

Reduction had no significant effect on overall C_1 duration ($|t| < 2$), but opposite patterns emerged for /k/ and the fricatives. In the case of /k/, the burst duration was longer for reduced tokens with an overall mean duration of 56 ± 24 ms compared to 40 ± 16 ms for unreduced

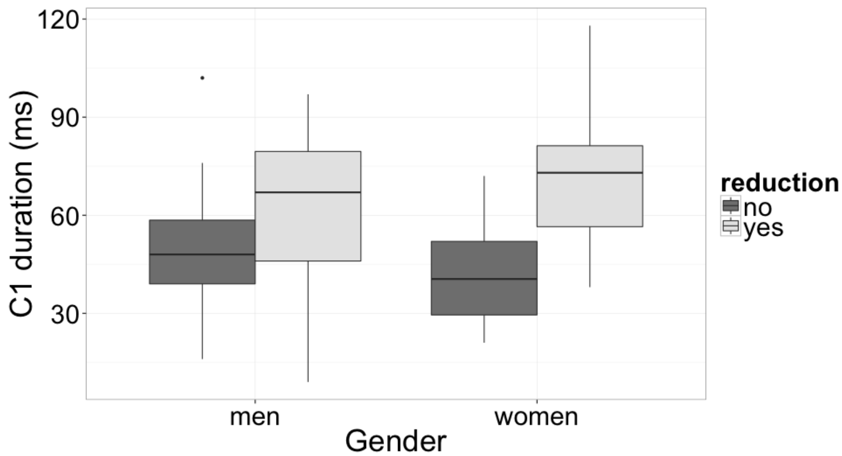

FIG 10: Duration of /k/ burst by gender and reduction status.

tokens. Furthermore, as shown in Figure 10, the duration difference was significant in women (t = 3.186) but not in men (t = 0.572).

In contrast, fricatives were generally shorter in reduced tokens, as shown in Table XVII below. The differences were not significant ($|t| > 2$), however, and gender also did not have a significant effect.

TABLE XVII: Mean duration and standard deviation of fricatives (ms).

	Unreduced	Reduced	t value
ʃ	92 (21)	86 (18)	-0.458
ɸ	60 (19)	58 (16)	0.000
s	99 (15)	82 (21)	-1.296
ç	80 (26)	64 (16)	-0.534

IV. DISCUSSION AND CONCLUSION

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 center of gravity (COG) and burst/frication duration of the consonant preceding the target vowel (C_1). COG results suggest that the phonological grammar of Japanese actively takes into account the predictability of a high vowel in a given context and does one of two things: (i) if predictability is sufficiently high, delete the vowel, or (ii) if predictability is insufficient for successful recovery, retain the oral gesture of the reduced vowel and overlap it with the preceding consonant to aid recovery. More specifically, when C_1 is /ɸ, s, ç/, the articulatory gestures associated with the vowel are not essential to provide acoustic cues for the underlying high vowel when C_2 is voiceless because the vowel that can follow the fricatives is fully predictable. In other words, retaining the oral gesture of the vowel does little to increase the likelihood of recoverability. The results for /ʃ/ in particular 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 Chitoran *et al.* (2002) and Silverman (1995), and it was also suggested by Varden (2010) for Japanese.

The results for /k/ were less straightforward in that while /u/ had a significant lowering effect compared to /i/ like in the case of /ʃ/, COG2 results revealed that within each vowel type, reduction actually had a raising effect like the high-predictability fricatives. A possible explanation for why /k/ patterns with the high predictability consonants is the large COG1 difference of 2,350 Hz between the burst noises of /ki/ and /ku/. This difference is nearly four times the COG2 difference of 640 Hz observed for /ʃ/ and nearly six times the 400 Hz spectral difference reported in Beckman and Shoji (1984), which Japanese speakers were shown to be sensitive to.

The large spectral difference can be interpreted as stemming from two possible reasons, the first of which is that the difference is due to /k/-fronting that results from coarticulation with the following /i/. Positing the presence of coarticulatory effects even in reduced tokens allows /k/ to be grouped with /ʃ/ with the vowel considered as being devoiced, although the fact that initial vowel gesture seems to significantly weaken by the second half of the burst noise is still puzzling. One possible explanation is that because /k/ is a stop, the coarticulation with the vowel began during the closure, leading a nearly maximal overlap by the time a burst is produced (i.e., lower COG1). This would mean that the gesture can only weaken towards the second half of the burst, leading to a higher COG2.

The second (and perhaps simpler) explanation is that because /k/-fronting is completely regular before /i/ and because Japanese has phonemic consonant palatalization, the sequence that has been assumed to be /ki/ throughout this paper is actually being treated as /kʲi/ underlyingly. In other words, the large difference is not due to coarticulation with the vowels *per se*, but rather because the consonant preceding /i, u/ are simply different allophones, namely [kʲ, k], respectively (an observation also made in Maekawa and Kikuchi (2005), as made evident by the transcription convention employed). If this is indeed the case, the reduced vowel after [kʲ, k] becomes predictable since the only reducible high vowel that can follow each of the consonants is /i/ and /u/, respectively. While a high back vowel can follow /kʲ/, a dictionary search shows that it is always the long vowel /uː/, which typically does not reduce. Even in the case of loanwords where /kʲ/ is followed by /u/, there is generally an alternative pronunciation as simply /kʲi/, showing again that a short high back vowel is dispreferred after /kʲ/ in the language (Shogakukan, 2013).

With respect to the issue of a lengthening effect of reduction, duration measurements showed that such an effect is unreliable at best if present. The longer duration of C_1 containing a reduced vowel as reported by (Kondo, 2005; Han, 1994) seems to be true only for the one stop consonant included in this study /k/, and even then only for women. The fricatives showed a consistent but insignificant shortening effect rather than a lengthening effect. The fact that the lengthening effect depends on the manner of the consonant and possibly the gender of the speaker suggests that C_1 lengthening is not an obligatory process. Furthermore, since the fricatives in the data showed no such effect, it is difficult to argue that the longer duration of /k/ is mora-preserving in nature. A more likely explanation seems to be that the longer burst duration of /k/ is recoverability driven, where /k/ burst containing a reduced vowel is longer because it is otherwise too short to carry the information necessary to recover an underlying vowel. There is no added duration of a reduced vowel for fricatives because the vowel information can be sufficiently carried in the long frication noise.

The analyses of reduction rate, center of gravity, and C_1 duration revealed that the effect of gender is also inconsistent. The only significant gender difference was found in /k/. As noted above, only women significantly lengthened the stop burst in reduced tokens. Also, the raising effect of reduction on the COG2 of /k/ was significantly less for women. Both of these results suggest that at least in the case of /k/ women are more reluctant to get rid of certain cues that could aid the recovery of a reduced vowel. The fact that only /k/ showed an effect could have resulted from the fact that there were more /k/-initial tokens than any of the fricatives. A follow-up study with a larger, more balanced stimulus set may yet reveal that gender has a significant effect for other consonants as well.

The current study also raises some new questions. First, the fact that complete vowel deletion can occur raises the issue of the legality of the resulting consonant clusters. Japanese is generally argued to be a strict CVCV language, where tautosyllabic clusters are disallowed (Shibatani, 1990). With reports of high vowel reduction becoming increasingly obligatory for younger speakers (although reportedly only for men (Imai, 2010)), it seems possible that due to the prevalence of the process, certain clusters have become legal in Japanese.

Second, because the participants in the current study were L2 speakers of English currently residing in the United States, their tendency to delete vowels in certain cases may have been the result of familiarity with clusters in English. Whether monolingual speakers of Japanese also tend to delete vowels that are in high-predictability contexts at the expense of violating a CVCV structure remains to be answered. Monolingual speakers may show similar behavior as the participants in this study, where high vowels reduce to the point of complete deletion in high-predictability contexts but devoice to retain acoustic cues for the reduced vowel in low-predictability contexts to aid recovery. On the other hand, since monolingual speakers lack experience with languages that allow consonant clusters, it is possible that their preference for a CVCV structure is stronger, resulting in devoicing across all contexts. Lastly, this study focused on the effects of predictability on production. Whether predictability also has an effect on how a reduced sequence is perceived is also an interesting question. Since more acoustic cues are available in low-predictability contexts, Japanese speakers may be more sensitive to them than in high-predictability contexts where such cues are unnecessary.

APPENDIX

Target tokens

manner category	*word*	*gloss*
stop-stop	kikai	‘chance’
	kikeɴ	‘danger’
	kikoeru	‘to be audible’
	kitatʃoːseɴ	‘North Korea’
	kitaku	‘return home’
	kukaku	‘a block’
	kukeru	‘to blindstitch’
	kukeː	‘rectangle’
	kuki	‘twig’
	$\text{kuk}^{\text{j}}\text{oː}$	‘trouble’
fricative-stop	çikui	‘low’
	çiteː	‘denial’
	ʃikata	‘means’
	ʃika	‘deer’
	ʃiteɴ	‘point of view’
	ɸukoː	‘unhappy’
	ɸutatabi	‘once more’
	sukuː	‘rescue’
	suteru	‘throw away’
	ʃukuhaku	‘lodging’

Control tokens

manner category	*word*	*gloss*
stop-stop	kibiʃiː	‘strict’
	kidoː	‘startup’
	kigawari	‘changing one’s mind’
	kigoː	‘symbol’
	kiguː	‘coincidence’
	kubaru	‘to distribute’
	kubi	‘unemployed person; neck’
	kudaranai	‘trivial’
	kudari	‘down-slope’
	kugi	‘nail’
fricative-stop	çibaiçiɴ	‘item not for sale’
	çidoi	‘cruel’
	çige	‘facial hair’
	ʃiboː	‘fat’
	ʃidoːʃa	‘leader’
	ɸuda	‘label’
	ɸugoːkaku	‘disqualification’
	sudeni	‘already’
	sugoi	‘amazing’
	sugu	‘immediately’

Geminate tokens

manner category	*word*	*gloss*
fricative-stop	ʃikkari	'hard working'
	ʃikke	'humidity'
	ʃippai	'failure'
	ʃippitsu	'writing (as a profession)'
	ʃukkiɴ	'going to work'

REFERENCES

Bates, D., Maechler, M., and Bolker, B. (**2011**). *lme4: Linear mixed-effects models using s4 classes*, R package version 0.999999-2.

Beckman, M. (**1982**). "Segmental duration and the 'mora' in Japanese", Phonetica **39**, 113–135.

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

Boersma, P. and Weenink, D. (**2013**). "Praat: Doing phonetics by computer. Version 5.3.40", .

Browman, C. P. and Goldstein, L. (**1992**). "Articulatory phonology: An overview", Phonetica **49**, 155–180.

Chitoran, I., Goldstein, L., and Byrd, D. (**2002**). "Gestural overlap and recoverability: Articulatory evidence from Georgian", in *Papers in Laboratory Phonology VII*, edited by N. Warner and C. Gusshoven (Mouton de Gruyter., Berlin).

Cutler, A., Otake, T., and McQueen, J. M. (**2009**). "Vowel devoicing and the perception of spoken Japanese words", Acoustical Society of America **125**, 1693–1703.

Faber, A. and Vance, T. J. (**2000**). "More acoustics traces of "deleted" vowels in Japanese", in *Japanese/Korean Linguistics*, edited by M. Nakayama and C. J. J. Quinn, volume 9, 100–113.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). "Statistical analysis of word-initial voiceless obstruents: preliminary results", Journal of Acoustical Society of America **84**, 115–123.

Fowler, C. A. and Saltzman, E. (**1993**). "Coordination and coarticulation in speech production", Language and Speech **36**, 171–195.

Fujimoto, M. (**in press**). "Vowel devoicing", in *Handbook of Japanese Phonetics and Phonology*, edited by H. Kubozono, chapter 4 (Mouton de Gruyter).

Fujimoto, M., Murano, E., Niimi, S., and Kiritani, S. (**2002**). "Differences in glottal opening patterns between Tokyo and Osaka dialect speakers: Factors contributing to vowel devoicingevoicing", Folia Phoniatrica et Logopedia **54**, 133–143.

Gelman, A. and Hill, J. (**2007**). *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge University Press, New York, NY).

Han, M. S. (**1962**). *Japanese Phonology: An Analysis Based Upon Sound Spectrograms* (Kenkyusha, Tokyo).

Han, M. S. (**1994**). "Acoustic manifestations of mora timing in Japanese", Acoustical Society of America **96**, 73–82.

Hirose, H. (**1971**). "The activity of the adductor laryngeal muscles in respect to vowel devoicing in Japanese", Phonetica **23**, 156–170.

Imai, T. (**2010**). "An emerging gender difference in Japanese vowel devoicing", in *A Reader in Sociolinguistics*, edited by D. R. Preston and N. A. Niedzielski, volume 219, chapter 6, 177–187 (Walter de Gruyter).

Kindaichi, H. (**1995**). *Shin Meikai Nihongo Akusento Jiten [Japanese Accent Dictionary]* (Sanseido).

Kondo, M. (**2005**). “Syllable structure and its acoustic effects on vowels in devoicing”, in *Voicing in Japanese*, edited by H. v. d. Hulst, J. Koster, and H. van Riemsdijk, 229–246 (Mouton de Gruyter).

Maekawa, K. and Kikuchi, H. (**2005**). “Corpus-based analysis of vowel devoicing in spontaneous Japanese: an interim report”, in *Voicing in Japanese*, edited by J. van de Weijer, K. Nanjo, and T. Nishihara (Mouton de Gruyter).

Martin, S. (**1987**). *The Japanese Language through Time* (Yale University Press, New Haven).

Mattingly, I. G. (**1981**). “Phonetic representation and speech synthesis by rule”, in *The Cognitive Representation of Speech*, edited by J. Myers, J. Laver, and A. J., 415–420 (North-Holland Publishing Company).

McCarthy, J. J. (**1999**). “Sympathy and phonological opacity”, Phonology **16**, 331–399.

NHK (**1985**). *Nihongo Hatsuon Akusento Jiten [Japanese Pronunciation Accent Dictionary]* (Nihon Hoso Kyokai [Japan Broadcasing Corporation], Tokyo).

Ogasawara, N. (**2013**). “Lexical representation of Japanese high vowel devoicing”, Language and Speech **56**, 5–22.

Ogasawara, N. and Warner, N. (**2009**). “Processing missing vowels: Allophonic processing in Japanese”, Language and Cognitive Processes **24**, 376–411.

Okamoto, S. (**1995**). ““Tasteless” Japanese: less “feminine” speech among young Japanese women”, in *Gender articulated: Language and the socially constructed self*, edited by K. Hall and M. Bucholtz, 297–325 (Routledge, New York).

R Core Team (**2013**). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Sawashima, M. (**1971**). “Devoicing of vowels”, Annual Bulletin (Research Institute of Logopedics and Phoniatrics, University of Tokyo) **6**, 7–13.

Sawashima, M., Hirose, H., and Yoshioka, H. (**1978**). “Abductor (pca) and adductor (int) muscles of the larynx in voiceless sound production”, Annual Bulletin, Research Institute of Logopedics and Phoniatrics **12**, 53–60.

Sharoff, S. (**2008**). “Lemmas from the internet corpus”, URL `http://corpus.leeds.ac.uk/frqc/internet-jp.num`.

Shibatani, M. (**1990**). *The Languages of Japan* (Cambridge University Press, Cambridge).

Shogakukan (**2013**). “Daijisen Zoubo/Shinsouban (Digital Version)”, URL `http://dictionary.goo.ne.jp/`.

Silverman, D. (**1995**). “Phasing and Recoverability”, Ph.D. thesis, UCLA, Los Angeles.

The Agency for Cultural Affairs (**2012**). “Jouyou Kanji Hyou [List of Jouyou Kanji]”, .

Tsuchida, A. (**1997**). “Phonetics and phonology of Japanese vowel devoicing”, Ph.D. thesis, Cornell University.

Tsuchida, A., Kiritani, S., and Niimi, S. (**1997**). “Two types of vowel devoicing in Japanese: Evidence from articulatory data”, Journal of Acoustical Society of America **101**, 3177.

Vance, T. (**1987**). *An Introduction to Japanese Phonology* (SUNY Press, New York).

Vance, T. J. (**2008**). *The sounds of Japanese* (Cambridge University Press, New York).

Varden, J. K. (**2010**). “Acoustic correlates of devoiced Japanese vowels: velar context”, The Journal of English and American Literature and Linguistics **125**, 35–49.

Varden, J. K. and Sato, T. (**1996**). “Devoicing of Japanese vowels by Taiwanese learners of Japanese”, Proceedings of International Conference on Spoken Language Processing **96.2**, 618–621.

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

Yoshioka, H., Löfqvist, A., and Hirose, H. (**1982**). "Laryngeal adjustments in Japanese voiceless sound production", Journal of Phonetics **10**, 1–10.

Yuen, C. L.-K. and Hubbard, K. (**1998**). "Vowel devoicing and gender in Japanese", Presented at the LSA in New York, NY.