Consequences of High Vowel Deletion for Syllabification in Japanese

Jason A. Shaw and Shigeto Kawahara
Yale University and *Keio University*

1 Introduction

Japanese is well-known as a language without consonant clusters, allowing only homorganic nasal-consonant clusters and geminates (e.g., Ito, 1986). In fact, not only does Japanese have no words with non-homorganic consonant clusters, Japanese speakers resort to epenthesis when they borrow words with consonant clusters from other languages; for example, the English word *strike* is pronounced as [s**u**t**o**raik**u**] when borrowed into Japanese, in which the original, monosyllabic word becomes a four-syllable word with three epenthetic vowels (shown in bold).[1] This phonotactic restriction is claimed to condition perceptual epenthesis as well, such that Japanese listeners report hearing vowels between non-homorganic consonant clusters (Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Dupoux, Parlato, Frota, Hirose, & Peperkamp, 2011). Moreover, the Japanese orthographic system is such that each letter represents a combination of a consonant and a vowel; i.e., there is no character that exclusively represents an onset consonant. All of these observations lead to the oft-stated characterization that "Japanese is a strict CV-language".

However, Japanese is also known to devoice high vowels between two voiceless obstruents, which results in apparent consonant clusters (e.g. [ɸu̥soku] 'shortage') (throughout this paper, devoiced vowels are shown with underline). Some researchers argue that these high vowels are simply devoiced—not deleted—and therefore, Japanese does not have consonant clusters after all (Beckman, 1996; Faber & Vance, 2000; Jun & Beckman, 1993). Other researchers argue that acoustically, there is no evidence for the presence of vowels at all; they therefore conclude that these vowels are entirely deleted (Beckman, 1982; Beckman & Shoji, 1984).[2] Beckman and Shoji (1984: 63), for example, state that "[w]hen the waveform of a devoiced syllable is examined, however, neither its spectral nor its temporal structure indicateks the presence of a voiceless vowel." A recent articulatory study by Shaw and Kawahara (2018) used ElectroMagnetic Articulography (EMA) to address this issue—mere devoicing vs. wholesale deletion—by examining whether the devoiced vowels retain their lingual articulation. They found that at least some devoiced tokens lack vowel height targets altogether, suggesting that these high vowels are not merely devoiced. They also found that those tokens that lack vowel height targets show patterns of temporal variation consistent with consonant-to-consonant (C-C) coordination. That is, the flanking consonants appear to be timed directly to each other instead of to an intervening vowel, i.e., consonant-to-vowel (C-V) coordination, providing further evidence that there is no vowel in the surface representation of these tokens. These results mean that Japanese, as a consequence of high vowel deletion, has consonant clusters (e.g. [ɸsoku]), contrary to the "CV-language" characterization often given to Japanese.

Based on this recent result reported in Shaw and Kawahara (2018), this paper addresses how such consonant clusters, arising from high vowel deletion, are syllabified. We compare two hypotheses regarding this question, (1) the resyllabification hypothesis and (2) the consonantal syllable hypothesis, and present

* This research is supported by the JSPS grants (#26770147 and #26284059, and especially #15F15715). Many thanks to Jeff Moore and Chika Takahashi for their help with the EMA data acquisition and analysis. We received helpful comments from the audience at AMP 2017. Although the current analysis is based on the EMA data obtained for a study reported in Shaw and Kawahara (2018, to appear), the temporal stability analysis reported in this paper, or its connection to the phonological behavior of devoiced vowels, was not reported in these previous studies. Remaining errors are ours.

[1] Personal thanks from the second author to Ann-Michelle Tessier, who caught his Japanese reading of *Wurmbrand*, i.e. [urumuburando], which was a little "off" from how that name would actually be pronounced in English or German.

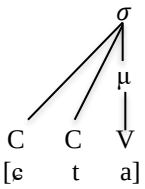
[2] Yet others argue that vowels are merely devoiced in some environments and deleted altogether in other environments (Kawakami, 1977; Whang, 2017)

evidence for the latter. Our argumentation is based on two kinds of evidence. The first is a phonological consideration; we show that phonological processes that are sensitive to syllable structure, such as prosodic truncation and pitch accent placement, are unaltered by high vowel deletion. The other one is a phonetic consideration; patterns of temporal stability in speech production are inconsistent with the resyllabification hypothesis. In addition to addressing a specific question in Japanese phonology, our results bear on more general theoretical issues, including how different syllable structures manifest themselves in articulatory timing patterns (Browman & Goldstein, 1988; Hermes, Mücke, & Auris, 2017; Hermes, Mücke, & Grice, 2013; Marin, 2013; Marin & Pouplier, 2010; Shaw & Gafos, 2015; Shaw, Gafos, Hoole, & Zeroual, 2009), and the independence of prosodic and segmental levels of representation. The convergence of the phonetic and phonological evidence bolsters the claim that syllable structure corresponds to characteristic patterns of gestural timing in in speech.

2 The two hypotheses

We are not the first to consider the question of how consonant clusters resulting from high vowel deletion are syllabified in Japanese, and there are two separate hypotheses in the literature. These two hypotheses are illustrated in Figure 1. The first hypothesis (H1), shown on the left side of Figure 1, is that the consonant that preceded the deleted high vowel is resyllabified into the following syllable, resulting in a complex syllable onset. Kondo (1997) argues for this view based on the observation that devoicing of two consecutive vowels is often prohibited. On Kondo's account, consecutive vowel devoicing is blocked by a constraint against tri-consonantal onsets. This constraint can only function to block consecutive devoicing if the devoiced vowels are also deleted. Matsui (2017) on the other hand argues that it is possible for Japanese to have consonantal syllables, as in the right side of Figure 2. His argument is primarily based on linguo-palatal patterns obtained using ElectroPalatoGraphy (EPG). He found that the pattern of lingual contact typically observed for Japanese /u/ is absent in devoiced contexts. Moreover, when devoiced /u/ is preceded by /s/, the contact pattern characteristic of /s/ extends temporally throughout the syllable.[3] Thus, in terms of linguo-palatal contact, it appears that /u/ is replaced by a consonant. Matsui (2017) discusses this result in the context of the C/D model of articulation (Fujimura, 2000), which crucially assumes that a syllable can remain even after high vowel deletion.

Hypothesis 1 (H1): Resyllabification



Hypothesis 2 (H2): Consonantal syllable

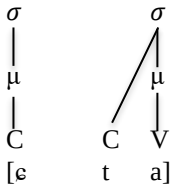


Figure 1: Two hypotheses regarding the syllabification of consonant clusters created via high vowel deletion, as in /ɕuta/→[ɕta]

This paper provides further evidence for H2, drawing on a confluence of phonological and phonetic evidence.

3 Phonological considerations

We begin with phonological considerations that favor the consonantal syllable hypothesis (H2). As observed by Tsuchida (1997) and Kawahara (2015), devoiced vowels count toward the bi-moraic requirement of some morphophonological processes. Japanese has many word formation processes that are based on a bimoraic foot (Poser, 1990) and devoiced vowels count toward satisfying this requirement. Examples in (1)

[3] Similar "voiceless syllables" have been proposed for Lushootseed (Urbancyzk, 2001) and even English (Kaisse & Shaw, 1985), although see Davidson (2006) for arguments to the contrary for English.

include loanword truncation, (1a), hypocristic formation, (1b), and mimetics, (1c). For each word formation process, two examples are provided to show the bimoraic requirement, and three examples are shown to illustrate that devoiced vowels count toward the prosodic requirement.

(1) Devoiced vowels count toward bimoraic templatic requirements

a. loanword truncation:	[demonsutoreeɕoɴ] → [demo] 'demonstration'
	[rokeeɕoɴ] → [roke] 'location'
	[sutoraiki] → [s̲uto] 'strike'
	[ripurai] → [rip̲u] 'reply'
	[sukuriiɴ-ɕotto] → [s̲uk̲u-ɕo] 'screenshot'
b. hypocoristic formation:	[tomoko] → [tomo] (personal name)
	[sumiko] → [sumi] (personal name)
	[kumiko] → [k̲uko] (personal name)
	[tɕi̲kako] → [tɕi̲ka(-tɕaɴ)] (personal name)
	[satɕi̲ko] → [satɕi̲] (personal name)
c. mimetics:	[buru-buru] 'shivering'
	[doɴ-doɴ] 'stomping'
	[ɸu̲ka-ɸu̲ka] 'fluffy'
	[s̲uka-s̲uka] 'empty'
	[ɕi̲to-ɕi̲to] 'rainy'

The patterns in (1) show that the moras of the devoiced (and possibly deleted) high vowels remain. If they did not, then the bimoraic loanword truncation for, e.g., [sutoraiki] would be *[s̲utora] instead of [s̲uto]; the hypocristic for, e.g, [tɕi̲kako] would be *[tɕi̲ka:] or *[tɕi̲kka] instead of [tɕi̲ka]; and, similarly, the mimetic for 'rainy' would be *[ɕi̲tto-ɕi̲tto] or *[ɕi̲to:-ɕi̲to:] instead of [ɕi̲to-ɕi̲to]. To further corroborate this observation, Hirayama (2009) showed that moras of devoiced vowels count in *haiku*, whose rhythm is based on mora counts, in the same way as voiced vowels. To the extent that onset consonants do not project a mora (e.g., Hayes, 1989; c.f., Topintzi, 2010), then, this observation supports H2 in Figure 1. At the very least, the patterns in (1) show that the moras of devoiced vowels remain. If these cases of devoiced vowels also contain variable deletion, as in Shaw and Kawahara (2018), then the mora must be docked to the remaining consonant, as in H2.

Phonologically, some evidence suggests that syllables of devoiced vowels remain as well. Ito (1990) observes that the morphophological truncation pattern in (2a) cannot result in monosyllabic outputs, and that a light syllable is appended in such cases, as in (2b). Ito and Mester (1992) formalize this pattern as a result of a branching condition at the prosodic word level; a PrWd must branch at the level of the syllable. As shown in (2c), a syllable hosted by a devoiced vowel satisfies this prosodic branching requirement. If devoiced vowels in this context are also deleted, then the syllabic requirement is being satisfied by the final consonant in the word. This supports the syllabic consonant analysis, as in H2.

(2) Patterns of truncation in Japanese (based on Ito 1990 with some additional examples added)

a. Bimoraic truncation:	[ookesutora] → [oke] 'orchestra'
	[hiraasaru] → [riha] 'rehearsal
	[rokeeɕoɴ] → [roke] 'location'
b. No monosyllabic outputs:	[daijamondo] → [dai.ja] 'diamond'
	[paamanento] → [paa.ma] 'permanent'
	[kombineeɕoɴ] → [kom.bi] 'combination'
	[ɕimpoʑium] → [ɕim.po] 'symposium'
c. Devoiced vowels count:	[maikuroɸooɴ] → [mai.kɯ̥] 'microphone'
	[ampuriɸaiaa] → [am.pɯ̥] 'amplifier'
	[paŋkutɕaa] → [paŋ.kɯ̥] 'puncture'
	[wam.pii.su] → [wam.pi] 'one piece'

Another piece of phonological evidence comes from patterns of pitch accent placement. Kubozono (2011) argues that the Japanese default accent pattern, which is observed in loanwords and nonce word pronunciation, generally follows the Latin Stress Rule: (i) place the accent on the penultimate syllable if it is heavy (3a), (ii) otherwise place the accent on the antepenultimate syllable (3b).[4] The presence of devoiced vowels does not disrupt this pattern (3c). In cases of vowel deletion, the final consonant must still count as a syllable.

(3) Japanese accentuation patterns and the lack of influence by high vowel devoicing

a.	[ɸu.re'n.do] 'friend'	[pu.ra'a.to] 'Praat'
	[pu.ra'i.zu] 'prize'	[pu.re'e.su] 'place'
	[ɸu.ro'o.zuɴ] 'frozen'	[maa.ma.re'e.do] 'marmalade'
b.	[re'.ba.noɴ] 'Lebanon'	[se'.ku.taa] 'sector'
	[do'.ku.taa] 'doctor'	[pa'.ku.tɕii] 'coriander'
	[ɕi'.na.moɴ] 'cinnamon'	[ga'.va.doɴ] (proper name)
c.	[pu.ro'.se.sɯ̥] 'process"	[bu.ra'n.kɯ̥] 'blank'
	[su.pa'i.sɯ̥] 'spice'	[su.pa'i.kɯ̥] 'spike'
	[ku.re'e.pɯ̥] 'crepe'	[pu.ra'i.sɯ̥] 'price'

There is evidence from compound accentuation patterns and statistical distributions in native words that Japanese strongly disfavors accent on final syllables (Kubozono 2011). Given this dispreference, take words like [pu.ra'i.sɯ̥] and [ku.re'e.pɯ̥]. If the final syllables are lost due to high vowel deletion, it would be natural to expect that accent shifts away to the word-initial syllables, which does not occur. This lack of accentual shift also supports the view that the syllables of deleted high vowels remain phonologically.

In addition, devoiced syllables can bear pitch accents in modern Japanese (Vance, 1987). For example, Japanese accented verbs predictably bear accent on the penultimate syllable; when the penultimate syllables in verbs are devoiced, accent remains on that syllable (e.g. [kakɯ̥'su] 'hide'; [tsɯ̥'ku] 'to arrive'). Since accent bearing unit in Japanese is a syllable (Kubozono 2011 and many others), this observation too shows that devoiced vowels keep their own syllables. If, besides being devoiced, the vowel is also deleted in some of these cases, it must be the case that the remaining consonant supports the presence of the syllable.

All of these observations converge on one conclusion: (morpho)phonological processes that make reference to prosodic structure in Japanese do not treat devoiced vowels and voiced vowels differently. To the extent that devoiced vowels are deleted (Beckman, 1982; Beckman & Shoji, 1984; Shaw & Kawahara,

[4] Some four-mora words can be unaccented (Ito & Mester, 2016).

2018), then the general conclusion should be that moras and syllables remain after the deletion of these vowels, which is consistent with H2 in Figure 1.

In the next section, we further corroborate this conclusion from the perspective of articulatory coordination. In particular, we build on previous research findings that different syllable structures show different articulatory stability patterns (Browman & Goldstein, 1988; Byrd, 1995; Goldstein, Chitoran, & Selkirk, 2007; Hermes et al., 2013; Shaw & Gafos, 2015).

4 Temporal stability analysis

4.1 *Approach* The following analysis is of ElectroMagnetic Articulograph (EMA) data obtained for the study reported in Shaw and Kawahara (2018). The general idea of the analysis is, as illustrated in , to evaluate patterns of temporal stability in syllable-referential intervals across CV and CCV sequences. Previous studies, beginning with pioneering work by Browman and Goldstein (1988), have shown that languages that parse word-initial consonants tautosyllabically, i.e., as complex syllable onsets, tend to exhibit a specific pattern of temporal stability across CV and (C)CCV sequences (Hermes et al., 2013; Honorof & Browman, 1995; Marin, 2013; Marin & Pouplier, 2010). This includes results for English (Honorof & Browman, 1995; Marin & Pouplier, 2010), Romanian (Marin, 2013), and rising sonority clusters in Italian (Hermes et al., 2013). Specifically, as illustrated schematically in the right side of Figure 2, in these languages the center-to-anchor interval is more stable across CV and CCV sequences than the left edge-to-anchor interval or the right edge-to-anchor interval. In contrast, languages that enforce a heterosyllabic parse of initial CCV sequences, e.g., Moroccan Arabic and Tashlhiyt Berber (Dell & Elmedlaoui, 2002), tend to exhibit a different stability pattern. As illustrated schematically in the left side of Figure 2, these languages tend to show right edge-to-anchor stability (for Berber, see: Hermes et al., 2017; for Arabic, see: Shaw, Gafos, Hoole, & Zeroual, 2011).

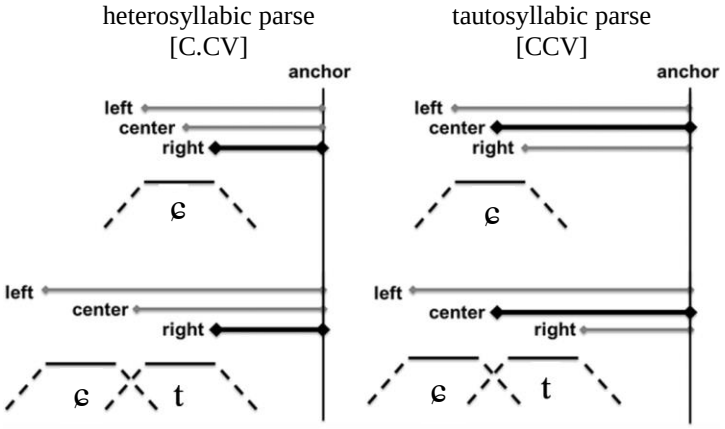


Figure 2: Illustration of stability indices: heterosyllabic parse vs. tautosyllabic parse.

The different patterns of temporal stability illustrated in Figure 2 can be derived from distinct coordination topologies organizing the relative timing of consonant and vowel gestures (Gafos, Charlow, Shaw, & Hoole, 2014; Shaw & Gafos, 2015). The key assumption linking syllable structure to patterns of temporal stability is an isomorphism between the arrangements of segments into syllables and the network of coordination relations that makes up the coordination topology. Specifically, onset consonants are assumed to enter into a relation of temporal coordination with the syllable nucleus, an assumption adopted from Browman and Goldstein (2000). Relevant coordination topologies are illustrated in Figure 3. Gestures are represented as vertices and coordination relations between them are represented as edges, a schema which follows the representational formalism developed in Gafos (2002). Different types of coordination relations are color-coded. The relation between adjacent consonants, i.e., C-C coordination, is shown in red; the relation between an onset consonant and a vowel, i.e., C-V coordination, is shown in blue. For completeness,

a yellow edge is also included—this indicates a relation between a vowel and possible post-vocalic segment, i.e., V-C coordination, although it does not play a role in the current analysis.

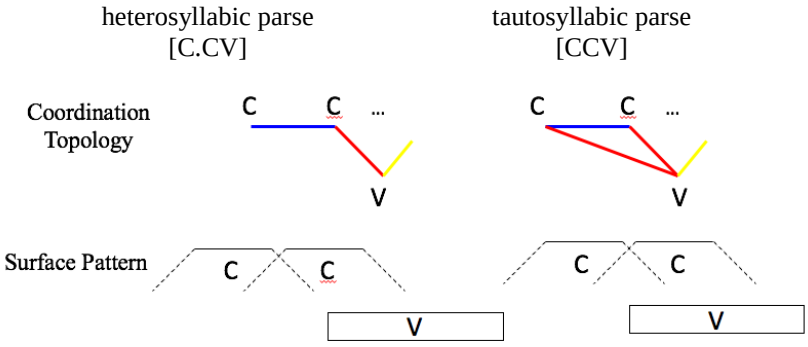


Figure 3: Relation between syllable parse, coordination topology and surface timing pattern for heterosyllabic (left) and tautosyllabic (right) parses of consonant clusters.

Under a heterosyllabic parse of initial consonants (Figure 3: left), the initial consonant is not contained in the same syllable as the following vowel—it is not a syllable onset—and, therefore, it is timed only to the following consonant (and not to the following vowel). In contrast, under a tautosyllabic parse (Figure 3: right) both pre-vocalic consonants are syllable onsets and, therefore, both enter into a coordination relation with the following vowel. Under the assumptions adopted here, complex onsets result in a coordination topology that, unlike the heterosyllable parse, places competing constraints on the temporal organization of gestures. That is, to satisfy the pattern of relative timing imposed by C-V coordination, the onset consonants would have to be temporally overlapped in time, a violation of C-C coordination. Satisfying C-C coordination, on the other hand, would entail a violation of C-V coordination. Although proposals differ in the technical details of how such competition is resolved (Browman & Goldstein, 2000; Gafos, 2002; Goldstein, Nam, Saltzman, & Chitoran, 2009), the surface timing patterns shown at the bottom of Figure 3 derive from the coordination topologies shown in the center of the Figure. It is therefore possible to recover a syllabic parse from the pattern of relative timing in articulatory movements. We make use of this mapping to bring in phonetic data bearing on the syllabification of consonant clusters in Japanese. Specifically, we pursue a stability analysis, evaluating the stability of intervals across CV and CCV sequences (Figure 2) to assess whether consonant clusters in Japanese resulting from targetless vowels syllabify like sequences in Arabic (i.e., $C_1.C_2V$, according to H2) or sequences in English (i.e., C_1C_2V, according to H1).

4.2 *Method* The stimuli are listed in Table 1. They contained five dyads, the members of which differ in whether they contain a devoicable high vowel (first column) or not (second column); in addition, the stimuli included singleton controls (third column).

Table 1: The list of the stimuli

Voiced vowel	Deleted (devoiced) vowel	Singleton control
[masuda] (personal name)	[mastaa] ‘master’	[bataa] ‘butter’
[yakuzai] ‘medication’	[haksai] ‘white cabbage’	[dasai] ‘uncool’
[ɕudaika] ‘theme song’	[ɕtaisee] ‘subjectivity’	[taisee] ‘system’
[ɸuzoku] ‘attachment’	[ɸsoku] ‘shortage’	[kasoku] ‘acceleration’
[katsudoo] ‘activity’	[katstoki] ‘when winning’	[mirutoki] ‘when looking’

Six native speakers of Tokyo Japanese (3 male) read items in the carrier phrase [okee___to itte] ‘ok, say ___’, where the underlined blank indicated the position of the target word. Items were randomized within a block, and 10-15 blocks were recorded per participant. For additional methodological details, such as EMA

sensor attachments and post-processing routines, see Shaw and Kawahara (2018). The first author and one research assistant inspected the acoustics of the produced tokens and found that all devoicable vowels were actually devoiced.

In order to assess whether the devoiced vowels were deleted or not, Shaw and Kawahara analyzed tongue dorsum trajectories from the vowel preceding /u/, e.g., [a] in [k**a**tsudoo] or [e] from the carrier phrase in [**e**#ɕudaika], to the following vowel, e.g., [o] in in [katsud**o**o] or [a] in [ɕud**a**ika]. A sample illustration is given in Figure 4, which plots tongue dorsum height trajectories from the preceding vowel /e/ in the frame sentence /ooke**e**/ through /u/ and onto the following vowel /a/. Thus, the blue lines represent tongue dorsum movement across the underlined portion of /e#ɕudaika/, whereas the red lines represent tongue dorsum movement across the underlined portion of /e#ɕutaisee/. A rise in tongue dorsum height between /e/ and /a/, corresponding to the intervening /u/ is expected if there is an articulatory target for /u/. We observe from Figure 4 that when the /u/s are devoiced (red lines), the tongue dorsum does not substantially rise between /e/ and /a/, at least not in some tokens, indicating a lack of /u/ target. To assess this quantitatively, Shaw and Kawahara (2018) apply an analytical technique involving machine classification of the trajectories based on competing phonological hypotheses: (i) a vowel present scenario, for which the voiced vowels (Table 1: column one) provided the training data and (ii) a vowel absent scenario, which was simulated. The simulations are guided by the assumption of phonetic interpolation, i.e., if there is no /u/ target, then the tongue dorsum will move from /e/ to /a/. The technique for simulating trajectories based on phonetic interpolation of flanking targets (including the hypothesized vowel absent scenario) is described in detail in Shaw and Kawahara (to appear). The outcome of the classification yields a posterior probability that the trajectories contains a vowel target. Shaw and Kawahara (2018) found that the posterior probability of a vowel target was very high for some tokens and very low for others (with few intermediate values). They conclude that the data support an optional process of deletion.

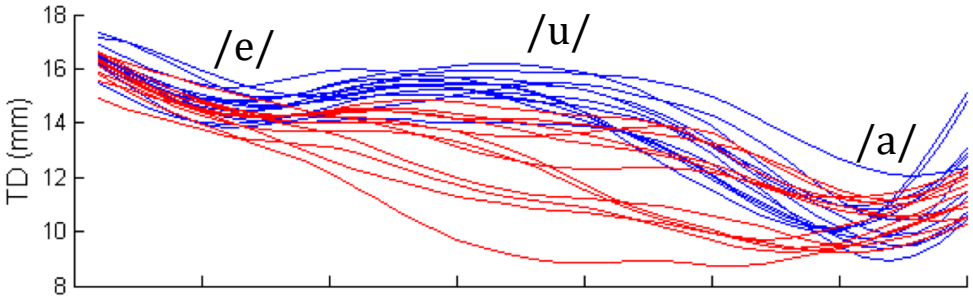


Figure 4: Sample tongue dorsum trajectories of /e#ɕudaka/ (blue lines) and /e#ɕ(u)taisee/ (red lines).

The current analysis builds on Shaw and Kawahara's (2018) results. We applied the stability analysis to the subset of tokens that had a high (>0.5) posterior probability of linear interpolation. These tokens were taken to lack a tongue dorsum target for /u/, thereby forming a consonant cluster. This amounted to different numbers of tokens from different dyads. Only [ɕtaisee], [ɸsoku] and [katstoki] exhibited sufficient numbers of such tokens. For [ɕtaisee], there 138 tokens (from five speakers) classified as deletion (lacking an /u/ target); for [ɸsoku], there 129 tokens (from four speakers); and, for [ɸsoku], there 88 tokens (from two speakers). The following analysis is based on tokens from these three words, classified as lacking an /u/ target, and an equal number of singleton controls. Since each item in Table 1 was produced in a block, we used in the analysis the singleton control from each block that also contained a case of vowel deletion. Consequently, the stability analysis below is based on 276 tokens for the [ɕtaisee]~[taisee] dyad, 258 tokens for [katstoki]~ [mirutoki], and 176 tokens for [ɸsoku]~[kasoku].

The three intervals schematized in Figure 2, left-edge-to-anchor (LE_A), center-to-anchor (CC_A), and right-edge-to-anchor (RE_A) were calculated for each token containing a consonant cluster as well as for the singleton control (Table 3: third column). The stability of these intervals across CV (singleton control) and CCV provided our phonetic diagnostic of syllable affiliation. All three of the intervals were right-delimited by a common anchor, the point of maximum constriction of the post-vocalic consonant. The landmarks that

left-delimit the three intervals were parsed in the following manner. The LE_A interval was left-delimited by the achievement of target of the first consonant in the sequence, e.g., [ɛ] in [ɛ̱taisee] (and [t] in the singleton control [ṯaisee]). The RE_A interval was left-delimited by the release of the immediately pre-vocalic consonant, e.g., [t] in [ɛṯaisee] (and also [t] in the singleton control [ṯaisee]). The third interval, CC_A was left-delimited by the mean of the midpoints between the consonants in the cluster and by the midpoint of the single onset consonant in the singleton control. The midpoint was the timestamp halfway between the achievement of target and the release. The target and release landmarks were determined from the articulatory signal with reference to movement velocity, allowing us to apply a uniform criterion for all consonants, regardless of manner or place of articulation. Specifically, we used 20% of peak velocity in the movement towards/away from consonantal constrictions. Figure 5 illustrates the parse of relevant landmarks for C1 and C2 in a token of [ɛtaisee]. The achievement of target and release of C1, which is [ɛ] in this case, is shown on the tongue blade (TB) trajectory (blue line). The parse of C2, [t], is shown on the tongue tip (TT) trajectory. As an index of interval stability across CV and CCV sequences, we computed the relative standard deviation (RSD), also known as the coefficient of variance, by dividing the standard deviation of interval duration calculated across tokens of CV and CCV by the mean interval duration across these same tokens.

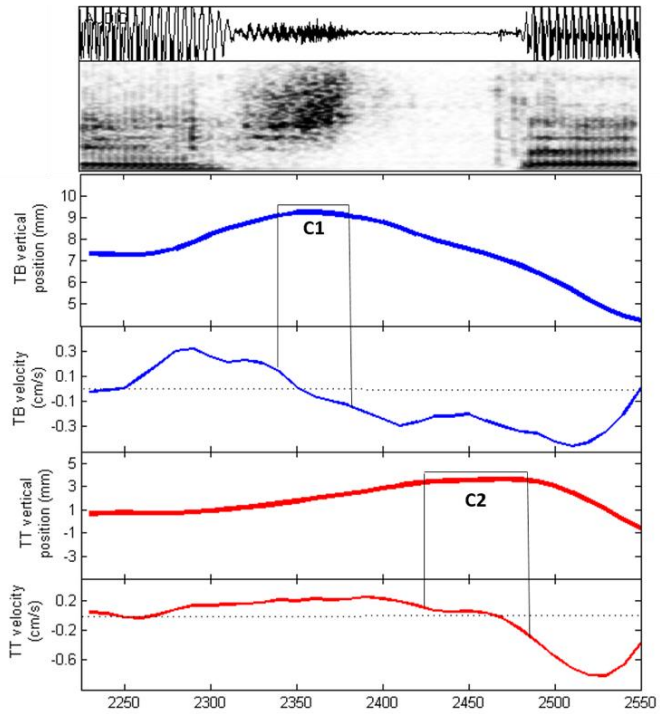


Figure 5: Illustration of how consonantal gestures are parsed in a token of [ɛtaisee]. The portion of the signal shown begins with the [e] of the carrier phrase and ends with the [a]. The panels show, from top to bottom, the audio signal, spectrogram, tongue blade (TB) height trajectory, tongue blade (TB) velocity signal, tongue tip (TT) height trajectory, and tongue tip velocity signal. The thin black lines show the achievement of target and release of the consonants, C1 and C2, and the 20% threshold of the velocity peak that was used to parse them.

4.3 *Results*

Figure ***6*** shows box plots of interval duration for LE_A, CC_A, and RE_A intervals as calculated across CV and CCV strings in three dyads. Of course, it is always the case LE_A is the longest, followed by CC_A and then RE_A—what we are interested in is the degree of *variability* of these intervals, as, following the schema in Figure 2, this provides phonetic evidence for syllabic organization. We observe that for each of the dyads, RE_A shows the least variability (i.e. the boxplots have the smallest *width*). This suggests that vowels are timed with respect to the right edge of the CC clusters, c.f., the center of CC

clusters. All else equal, shorter intervals also tend to be less variable, a general property of timed events but also of other phonetic measurements (see, e.g., Nguyen & Shaw, 2014 who show that variability in F1 and F2 for vowels is also correlated with the magnitude of the formant measurements). To correct for the effect that mean interval duration may have on the variability of the interval, we also computed the relative kstandard deviation.

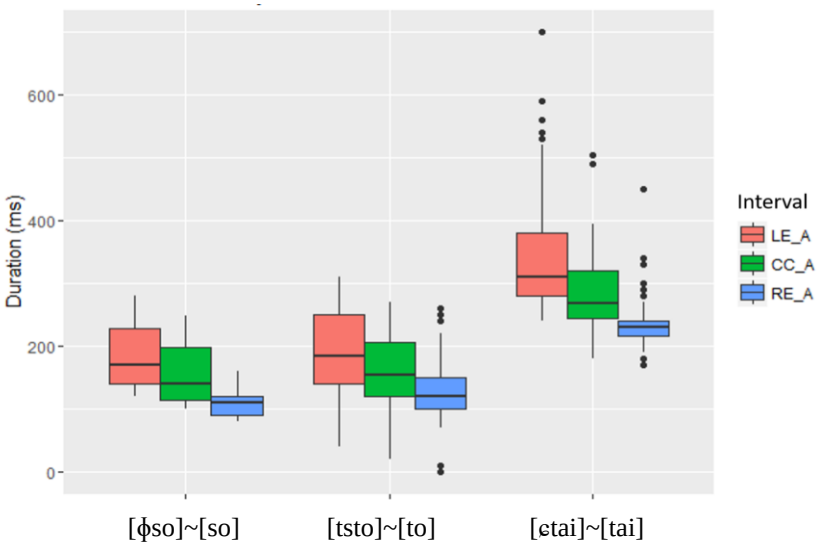


Figure 6: The durations between the three articulatory landmarks of the CC sequences and the vowel anchor.

The relative standard deviation (RSD) of the intervals in Figure 6 is shown in Table 2. Across dyads, the right-edge to anchor (RE_A) interval is the most stable (lowest RSD). On the assumptions illustrated in Figure 3, this pattern points unequivocally to simplex onsets, i.e., a heterosyllabic parse of initial clusters. Although care must be taken when interpreting stability patterns in terms of syllable structure, a point we return to in the general discussion, the pattern of RE_A stability is one of the most straightforward to interpret. The results of the stability analysis, therefore, converge on the same hypothesis as the phonological evidence. Both point to H2, a heterosyllabic parse of consonant clusters arising from high vowel deletion.

Table 2: Relative Standard Deviation (RSD)

	LE_A	CC_A	RE_A
[ɸso]~[so]	0.32	0.34	0.24
[tsto]~[to]	0.25	0.23	0.20
[ɕtai]~[tai]	0.23	0.28	0.11

5 General discussion

To summarize, the EMA study by Shaw and Kawahara (2018) showed that Japanese /u/ optionally deletes in devoicing environments, yielding consonant clusters. Both phonological and phonetic evidence reviewed here suggests that these consonant clusters are parsed heterosyllabically. The current results imply a rather surprising conclusion that Japanese allows consonantal syllables headed by a fricative or an affricate, a conclusion that is especially surprising in light of the view that considers Japanese a "strict CV-language".[5]

The current results show that Japanese consonant clusters arising from high vowel deletion behave in

[5] This reminds us of the observation made by Kawahara (2002) that otherwise prohibited marked structures can be permitted as a result of an optional process, the observation Kawahara (2002) dubbed "the emergence of the marked".

terms of articulatory stability like word-initial consonant clusters in Moroccan Arabic. The connection to Moroccan Arabic is intriguing because word-initial clusters in this language also arose diachronically from the loss of short vowels (Benhallam, 1980), and there have been similar debates about syllabification based on internal phonological evidence, see, e.g., Keegan (1986: 214) who argues for complex onsets (H1 in Figure 1) vs. Kiparsky (2003) who argues for moraic consonants (H2). Ultimately, the weight of the evidence, which includes now arguments from temporal stability in articulation (Shaw et al., 2009) and metrical patterns in verse (Elmedlaoui, 2014) points to H2, the same conclusion that we have drawn for Japanese. In both cases, higher level syllabic structure is preserved despite the loss of a vowel.

More generally speaking, then, our data presents a case in which prosodic and temporal stability are maintained despite loss of a segment. Better-known cases of prosodic structure preservation include those discussed under the rubric of compensatory lengthening (Kavitskaya, 2014). In these patterns, higher level structure preservation is more salient because it conditions segmental-level lengthening. In the Japanese case, loss of a vowel neither lengthens adjacent segments nor shortens the transitions between consonants (Shaw & Kawahara, 2018). The existence of patterns that delete segments while preserving prosodic structure supports independent representations of timing (prosodic structure) and articulation (segmental content), a dissociation which may have a neural basis (Long et al., 2016). For example, Long et al. (2016) found that perturbation of normal brain function (through focal cooling) could selectively influence speech timing or segmental content, depending on the brain region targeted. Phonologists often think about "building" prosodic structure over segments, but it may instead be that prosody provides a temporal frame into which segments are "filled in" (cf. Fujimura's C/D model 2000).

We also find the convergence between the phonological evidence (section 3) and the phonetic evidence (section 4) to be generally encouraging, as it speaks to the potential to reach common conclusions from diverse data sources (c.f., Broselow, Chen, & Huffman, 1997). We close here by pointing out some of the key assumptions that have supported this convergence. For starters, we assumed at times that the vowel deletion observed in Shaw and Kawahara (2018) is present in other environments in which devoicing is observed, particularly in the word final environment. This is not necessarily the case. Kilbourn-Ceron & Sonderegger (2017) have recently argued in fact that the devoicing processes word-finally and between voiceless consonants come from different sources/mechanisms. The EMA data supporting vowel deletion in Shaw and Kawahara (2018) includes only vowels occurring between voiceless consonants. However, our phonological arguments assume that deletion of devoiced vowels also occurs at least some of the time in devoicing contexts word-finally. If devoiced vowels word-finally are never deleted, then the phonological arguments around data in (2) and (3) are less compelling. A related alternative, which we cannot rule out, is that the vowel gesture is preserved in just those cases in which it is required to fulfill a morpho-phonological requirement, such as the bimoraic templates in (2). Testing this hypothesis would require new EMA data. As it currently stands, the full force of our argument for converging phonological and phonetic evidence rests on the assumption that the optional deletion we have observed between voiceless consonants generalizes to other devoicing environments. A second assumption, on the side of the temporal stability analysis, is that RE_A stability reflects a heterosyllabic parse of consonants. There are by now numerous studies that have applied this phonetic heuristic, which follows from the theoretical framework summarized in Figure 3. Through computational simulation using stochastic models, Shaw and Gafos (2015) probed the range of stability patterns (expressed in terms of RSD, as we do here) that can arise from different parses of initial clusters. They found that it is not always the case that simplex onsets correspond to RE_A stability while complex onsets correspond to CC_A stability. In particular, they highlight specific conditions, those of how overall variability, under which simplex onsets are predicted to condition CC_A stability. A realistic scenario of increasing variability presents itself in language acquisition. During the acquisition of the lexicon, increasing exposure to new words and new speakers increases the overall level of temporal variability in speech experience, which can drive a shift in the aggregate statistics from RE_A stability to CC_A stability (Gafos et al., 2014). In the current case at hand, that of our Japanese data, the level of variability in the data is low enough that we can be reasonably sure that simplex onset topology (Figure 3: left) maps to RE_A stability. More importantly, the conditions under which a complex onset (Figure 3: right) parse could condition RE_A stability are exceedingly rare (given our working assumption that onset consonants are timed to the syllable nucleus). We are therefore reasonably confident of our conclusions for the Japanese data, but a more complete analysis would report as well patterns of covariation between temporal intervals predicted by the competing hypothesis (see, e.g., Shaw & Davidson, 2011; Shaw et al., 2011).

To the extent that the above assumptions are valid, the results provide support for H2, the hypothesis that Japanese consonant clusters resulting from vowel deletion are parsed heterosyllabically. This conclusion follows from converging evidence from the analysis of phonological patterns sensitive to syllable structure and an analysis of temporal stability in articulation.

References

Beckman, M. (1982). Segment duration and the 'mora'in Japanese. *Phonetica, 39*(2-3), 113-135.

Beckman, M. (1996). When is a syllable not a syllable? In T. Otake & A. Cutler (Eds.), *Phonological Structure and Language Processing* (pp. 95-124). New York: Mouton de Gruyter.

Beckman, M., & Shoji, A. (1984). Spectral and Perceptual Evidence for CV Coarticulation in Devoiced/si/and/syu/in Japanese. *Phonetica, 41*(2), 61-71.

Benhallam, A. (1980). *Syllable structure and rule types in Arabic.* (PhD), University of Florida.

Broselow, E., Chen, S.-I., & Huffman, M. (1997). Syllable weight: Convergence of phonology and phonetics. *Phonology, 14*, 47-82.

Browman, C. P., & Goldstein, L. (1988). Some Notes on Syllable Structure in Articulatory Phonology. *Phonetica, 45*, 140-155.

Browman, C. P., & Goldstein, L. M. (2000). Competing Constraints on Intergestural Coordination and Self-Organization of Phonological Structures. *Les cahiers de l'lCP, Bulletin de la Communication Parlee, 5*, 25-34.

Byrd, D. (1995). C-centers revisited. *Phonetica, 52*, 285-306.

Davidson, L. (2006). Schwa elision in fast speech: segmental deletion or gestural overlap? *Phonetica, 63*(2-3), 79-112.

Dell, F., & Elmedlaoui, M. (2002). *Syllables in Tashlhiyt Berber and in Moroccan Arabic*. Dordrecht, Netherlands, and Boston, MA: Kluwer Academic Publishers.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1568-1578.

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from? *Journal of Memory and Language, 64*(3), 199-210.

Elmedlaoui, M. (2014). What does the Moroccan Malḥun meter compute, and how? *The Form of Structure, the Structure of Form: Essays in honor of Jean Lowenstamm, 12*, 139.

Faber, A., & Vance, T. J. (2000). More acoustic traces of 'deleted'vowels in Japanese. *Japanese/Korean Linguistics, 9*, 100-113.

Fujimura, O. (2000). The C/D Model and Prosodic Control of Articulatory Behavior. *Phonetica, 57*, 128-138.

Gafos, Charlow, S., Shaw, J. A., & Hoole, P. (2014). Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics, 44*, 152-166.

Gafos, A. (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory, 20*, 269-337.

Goldstein, L. H., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. In G. Fant, H. Fujisaki, & J. Shen (Eds.), *Frontiers in Phonetics and Speech Science: Festschrift for Wu Zongji* (pp. 239-249). Beijing: Commercial Press.

Goldstein, L. M., Chitoran, I., & Selkirk, E. (2007). *Syllable structure as coupled oscillator modes: evidence from Georgian vs. Tashlhiyt Berber*. Paper presented at the XVIth International Congress of Phonetic Sciences, Saabrucken, Germany.

Hayes, B. (1989). Compensatory Lengthening in moraic phonology. *Linguistic Inquiry, 20*, 253-306.

Hermes, A., Mücke, D., & Auris, B. (2017). The variability of syllable patterns in Tashlhiyt Berber and Polish. *Journal of Phonetics, 64*, 127-144.

Hermes, A., Mücke, D., & Grice, M. (2013). Gestural coordination of Italian word-initial clusters: The case of "impure s". *Phonology, 30*(1), 1-25.

Hirayama, M. (2009). Postlexical Prosodic Structure and Vowel Devoicing in Japanese (2009). *Toronto Working Papers in Linguistics*.

Honorof, D. N., & Browman, C. P. (1995). *The center or the edge: how are consonant clusters organised with respect to the vowel?* Paper presented at the XIIIth International Congress of Phonetic Sciences, Stockholm, Sweden.

Ito, J. (1986). *Syllable theory in prosodic phonology.* (Ph.D.), University of Massachusetts at Amherst.

Ito, J. (1990). Prosodic minimality in Japanese. In K. Deaton, M. Noske, & M. Ziolkowski (Eds.), *CLS 26: Parasession on the Syllable in Phonetics and Phonology* (pp. 213-239). Chicago: Chicago Linguistic Society.

Ito, J., & Mester, A. (1992). *Weak layering and word binarity*. Linguistics Research Center Report. (92-09). UC Santa Cruz.

Ito, J., & Mester, A. (2016). Unaccentedness in Japanese. *Linguistic Inquiry, 47*(3), 471-526.

Jun, S.-A., & Beckman, M. (1993). *A gestural-overlap analysis of vowel devoicing in Japanese and Korean.* Paper presented at the 67th annual meeting of the Linguistic Society of America, Los Angeles.

Kaisse, E. M., & Shaw, P. (1985). On the theory of lexical phonology. *Phonology, 2*, 1-30.

Kavitskaya, D. (2014). *Compensatory lengthening: phonetics, phonology, diachrony*: Routledge.

Kawahara, S. (2002). Faithfulness among variants. *Phonological Studies, 5*, 47-54.

Kawahara, S. (2015). The phonology of Japanese accent. *The handbook of Japanese phonetics and phonology*, 445-492.

Kawakami, S. (1977). Outline of Japanese Phonetics [written in Japanese as" Nihongo Onsei Gaisetsu"]: Tokyo: Oofuu-sha.

Keegan, J. (1986). The role of syllable structure in the phonology of Moroccan Arabic. In G. Dimmendaal (Ed.), *Current Approaches to African Linguistics* (Vol. 3, pp. 209-226).

Kilbourn-Ceron, O., & Sonderegger, M. (2017). Boundary phenomena and variability in Japanese high vowel devoicing. *Natural Language & Linguistic Theory*, 1-43.

Kiparsky, P. (2003). Syllables and moras in Arabic. In C. Féry & R. Vijver (Eds.), *The optimal syllable* (pp. 143-161). Cambridge: Cambridge University Press.

Kondo, M. (1997). *Mechanisms of vowel devoicing in Japanese.* (Ph.D.), Edinburgh, Edinburgh, UK.

Kubozono, H. (2011). Japanese pitch accent *The Blackwell companion to phonology* (Vol. 5, pp. 2879-2907).

Long, M. A., Katlowitz, K. A., Svirsky, M. A., Clary, R. C., Byun, T. M., Majaj, N., . . . Greenlee, J. D. (2016). Functional segregation of cortical regions underlying speech timing and articulation. *Neuron, 89*(6), 1187-1193.

Marin, S. (2013). The temporal organization of complex onsets and codas in Romanian: A gestural approach. *Journal of Phonetics, 41*(3), 211-227.

Marin, S., & Pouplier, M. (2010). Temporal organization of complex onsets and codas in American English: Testing the predictions of a gesture coupling model. *Motor Control, 14*, 380-407.

Nguyen, N., & Shaw, J. A. (2014). *Why the SQUARE vowel is the most variable in Sydney.* Paper presented at the Proceedings of the 15th Australasian International Conference on Speech Science and Technology (SST2014). Christchurch, New Zealand.

Poser, W. J. (1990). Evidence for foot structure in Japanese. *Language, 66*, 78-105.

Shaw, J. A., & Davidson, L. (2011). Perceptual similarity in input–output mappings: A computational/experimental study of non-native speech production. *Lingua, 121*(8), 1344-1358.

Shaw, J. A., & Gafos, A. I. (2015). Stochastic Time Models of Syllable Structure. *PLoS One, 10*(5), e0124714.

Shaw, J. A., Gafos, A. I., Hoole, P., & Zeroual, C. (2009). Syllabification in Moroccan Arabic: evidence from patterns of temporal stability in articulation. *Phonology, 26*, 187-215.

Shaw, J. A., Gafos, A. I., Hoole, P., & Zeroual, C. (2011). Dynamic invariance in the phonetic expression of syllable structure: a case study of Moroccan Arabic consonant clusters. *Phonology, 28*(3), 455-490.

Shaw, J. A., & Kawahara, S. (2018). The lingual articulation of devoiced /u/ in Tokyo Japanese. *Journal of Phonetics, 66*, 100-119. doi:https://doi.org/10.1016/j.wocn.2017.09.007

Shaw, J. A., & Kawahara, S. (to appear). Assessing feature specification in surface phonological representations through simulation and classification of phonetic data. *Phonology*, 1-34.

Topintzi, N. (2010). *Onsets: Suprasegmental and prosodic behaviour* (Vol. 125): Cambridge University Press.

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

Urbancyzk, S. (2001). *Patterns of reduplication in Lushootseed*. New York: Garland Pub.

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

Whang, J. D. Y. (2017). *Shaping Speech Patterns via Predictability and Recoverability.* (Ph.D.), New York University.