

Reconciling CV Phonotactics and High Vowel Deletion in Japanese

James Whang
Saarland University

1 Introduction

High vowel devoicing is a highly productive process in standard modern Japanese where /i, u/ lose their phonation when between two voiceless obstruents (e.g. /tsukue/ → [tsukue] ‘desk’). Despite being a popular topic of study, there is no consensus on how to represent the process, stemming largely from two factors. First, there is disagreement on whether the process is in the domain of phonetics or phonology. If the process is phonetic, resulting from factors such as gestural mistiming, it should be gradient and allow devoiced vowels to range from partial devoicing to complete deletion (Beckman, 1996). If the process is phonological, it must be planned and thus apply categorically (Tsuchida, 1997). The available empirical evidence suggests that both views are partially correct. The process has seemingly gradient properties in that it ranges from loss only of phonation (Faber & Vance, 2000; Jun & Beckman, 1993; Tsuchida, Kiritani & Niimi, 1997; Varden, 2010) to complete deletion (Pinto, 2015; Shaw & Kawahara, 2018b; Whang, 2018), resulting in surface clusters. However, the process also seems to apply categorically and obligatorily with the loss of phonation at minimum, regardless of speech rate or register (Kondo, 2005). Furthermore, although devoiced vowels do range from unphonated to completely deleted, the variation between the two extremes is categorical, where oral vowel gestures are either fully present or fully absent (Shaw & Kawahara, 2018b). It seems to be the case therefore that although high vowel devoicing might have begun as a fast-speech, phonetic process (Hasegawa, 1979; Kuriyagawa & Sawashima, 1989), it has been phonologized in most contexts (Fujimoto, 2015).

The second factor contributing to the disagreement on the representation of high vowel devoicing in Japanese is that the language is also well-known for having a strong phonotactic preference for CV structure. Japanese disallows consonant clusters except across syllable boundaries, and even then only the first half of geminates (e.g. [gɛp.pu] ‘burp’) and place-assimilated nasals (e.g. [aŋ.ko] ‘red bean paste’) may surface as codas (Ito, 1986). The preference is evident not only in native Japanese words, but also in how loanwords have been adapted, where heterorganic clusters are repaired through (high) vowel epenthesis (e.g. [arupusu] ‘the Alps’). Psycholinguistic studies have also shown that this preference for CV structure is so strong that Japanese listeners report hearing illusory high vowels between heterorganic consonant clusters (Dupoux, Kakehi, Hirose, Pallier & Mehler, 1999; Dupoux, Parlato, Frota, Hirose & Peperkamp, 2011). This strong CV preference, of course, conflicts with the surface clusters that often result as a consequence of high vowel devoicing and has led a number of scholars to propose that devoicing cannot result in deletion (Hirayama, 2009; Kondo, 2005; Tsuchida, 1997). The present paper proposes instead that there is no inherent conflict between the CV phonotactics and CC clusters that result from high vowel deletion in Japanese because phonotactic repairs and high vowel devoicing occur at different phonological levels.

2 Proposal and motivation

The prevalent assumption in the literature is that devoicing in Japanese is a postlexical process (for a detailed discussion, see Hirayama, 2009), but the question of how late in the phonological grammar devoicing applies and how to represent this late application remains unanswered. The current paper proposes specifically that structural and phonetically-driven phonological processes (i.e. phonotactic evaluation and high vowel devoicing, respectively) apply at separate levels (Boersma, 2009; Hayes, 1999; Zsiga, 2000), and crucially that high vowel devoicing applies *after* phonotactic repairs. As the name ‘high vowel devoicing’ suggests,

* Thanks to Frans Adriaans, Lisa Davidson, Gillian Gallagher, Maria Gouskova, Shigeto Kawahara, and the attendees of AMP 2018 for feedback on versions of this work.

there must be a high vowel that can be targeted by the process. Crucially, the process targets both underlying and epenthesized high vowels, which suggests that high vowel devoicing must occur after epenthesis. This can be most clearly seen in Sino-Japanese compounds. Take the two compounds in (1) below, for example. The word in (1a) begins with a CV root that has an underlying high vowel. The word in (1b), on the other hand, begins with a CVC root¹, which does not have an underlying high vowel. Regardless of their underlying representation, however, the high back vowels in these words are both devoiced.

- (1) a. /ku+too/ → [k^hutoo] ‘hard fight (difficult+battle)’
 b. /kak+too/ → [kak^hutoo] ‘definitive answer (certain+reply)’

Additionally, having structural constraints apply after devoicing would prohibit any deleted forms from being produced. If structural evaluation and devoicing apply at the same level, deleted forms would be repaired through vowel epenthesis, but these vowels would then be targeted for deletion, and so on *ad infinitum*. The fact that deleted outputs are allowed further suggests that devoiced/deleted forms are not phonotactically evaluated.

Shown below in Figure 1 is an adaptation of the representational levels proposed in the OT learnability literature (Apoussidou, 2007; Boersma, 1998; Tesar & Smolensky, 1996), which the remainder of this paper assumes. Noted next to each level are types of constraints that are active in the evaluation of the respective forms. The /underlying/ form is the form on which a phonological grammar (of production) operates. Similarly, the ⟨surface⟩ form, traditionally indicated with square brackets, is the output of a phonological grammar after phonotactic evaluation. The [overt] form represents both the acoustic signal and the articulatory implementations necessary to produce the acoustic signal. Although Boersma (2007, 2009) separates the [overt] form further into [auditory] and [articulatory] forms, he assumes perfect sensorimotor knowledge between the two forms for the sake of simplicity, making the two forms indistinguishable for the purposes of phonological derivation. The current paper, therefore, uses the simple [overt] label from Tesar (1997) that combines the two forms.

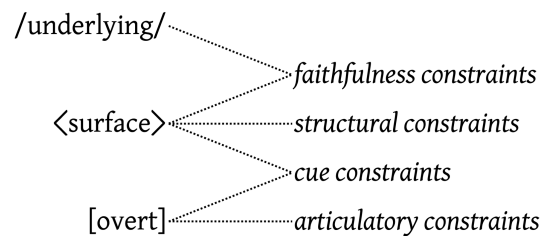


Figure 1: Minimally necessary levels of representation for a bidirectional model of phonology.

During production, an /underlying/ form is the initial input. The /underlying/ to ⟨surface⟩ form derivation remains unchanged from classic Optimality Theoretic analyses, where faithfulness constraints interact with markedness constraints. In her 1986 thesis and related works that followed, Ito proposes that the surface CV preference in Japanese can be captured by the structural markedness constraints in (2) below, of which only (2a) and (2b) are undominated in modern Japanese. Syllabification also occurs at this level, allowing the constraints below to refer to onsets and codas:

- (2) Markedness (structural) constraints:
- a. CODACONDITION: penalize coda segment that has its own place feature.
 - b. *COMPLEX: penalize every tautosyllabic cluster.
 - c. NOCODA: penalize every coda consonant.
 - d. ONSET: penalize every syllable that lacks an onset.

After phonotactic evaluation, the optimal ⟨surface⟩ candidate serves as the input for the next derivational step, where cue constraints and articulatory constraints evaluate the [overt] form. This is the level in which

¹ Although CVC roots often surface as CVCV, the second vowel is always high and almost completely predictable, and thus is analyzed as epenthetic. For a more detailed discussion on this issue, see Ito & Mester (2015).

high vowel devoicing applies. Cue constraints and articulatory constraints necessary to capture the devoicing process will be discussed in the following section.

During perception, the derivation process is simply the reverse, where an [overt] form is the initial input. Since the output during perception is a ⟨surface⟩ form, cue constraints and structural constraints are active during the derivation. Articulatory constraints are assumed to be inactive since the listener does not have direct access to the articulatory patterns of the speaker.

3 Phonological derivation in proposed framework

This section shows how the three phonological levels presented in Figure 1 and their respective constraint types function in the grammar. Perception of [overt] clusters is discussed first to illustrate how the interaction of cue constraints and structural constraints can yield a ⟨surface⟩ form output with epenthesis. The production grammar is then discussed, first illustrating how high vowels get epenthesized to repair /underlying/ clusters, then how the same cue constraints active during perception interact with articulatory constraints to yield devoiced/deleted [overt] outputs.

3.1 Perception Below in (3) is a tableau showing how a non-native, illegal cluster is perceived by a Japanese listener. Building on observations that phonetic implementation is largely language-specific, phonetic cues that are relevant to a given language are included as part of the phonological grammar in the form of *cue constraints* (Boersma, 1998, 2009), which are active in the evaluation of both ⟨surface⟩ and [overt] forms. Because the output is a surface form, structural constraints such as CODACONDITION and *COMPLEX are active, and cue constraints are also active to help interpret acoustic cues in the overt form input. Since the listener does not have direct access to the articulatory implementation the speaker used to produce the acoustic signal, *articulatory constraints* are assumed to be inactive during perception. The cue constraint *⟨ ⟩ [b] penalizes auditory cues for [b] (e.g. burst, formant transitions, etc.) that are not represented in the surface form. The cue constraint *⟨u⟩ [] penalizes ⟨u⟩ in the surface form that does not have corresponding auditory cues in the overt input.

(3) Perception of obstruent coda in Japanese

	[ebzo]	CODACONDITION	*COMPLEX	*⟨ ⟩ [b]	*⟨u⟩ []
☞	a. ⟨e.bu.zo⟩				*
	b. ⟨eb.zo⟩	*!			
	c. ⟨e.bz.o⟩		*!		
	d. ⟨e.zo⟩			*!	

The faithful candidate (b) is eliminated because the coda ⟨b⟩ in the first syllable differs in place from the following onset ⟨z⟩, violating the undominated structural constraint CODACONDITION. Candidate (c) obeys CODACONDITION by shifting the ⟨b⟩ to the second syllable, but this results in an onset cluster which violates the undominated structural constraint *COMPLEX. Although candidate (d) violates neither structural constraint, it violates the cue constraint *⟨ ⟩ [b] by deleting an audible overt [b] in the surface output. Candidate (a) violates *⟨u⟩ [], but it obeys all the higher-ranked constraints and thus is chosen as the optimal candidate, leading the Japanese listener to perceive [ebzo] as containing a high back vowel.

3.2 Production Below in (4) are Sino-Japanese words from (1), reformatted based on the phonological levels in Figure 1.

- (4) a. /kək+too/ → ⟨ka.ku.too⟩ → [ka.ku.too] ‘definitive answer (certain+reply)’
 b. /kū+too/ → ⟨ku.too⟩ → [kū.too] ‘hard fight (difficult+battle)’

Although the word ‘definitive answer’ in (4a) does not contain an underlying high vowel, it still undergoes high vowel devoicing because ⟨u⟩ first gets epenthesized at the surface level as shown in tableau (5) below. The tableau also includes the faithfulness constraints MAX-IO (every underlying segment must have a corresponding surface segment) and DEP-IO (every surface segment must have a corresponding underlying segment).

(5) Vowel epenthesis during production

	/kaktoo/	CODACONDITION	*COMPLEX	MAX-IO	DEP-IO
☞	a. ⟨ka.ku.too⟩				*
	b. ⟨kak.too⟩	*!			
	c. ⟨ka.ktoo⟩		*!		
	d. ⟨ka.too⟩			*!	

The faithful candidate (b) is eliminated because the coda ⟨k⟩ in the first syllable differs in place from the following onset ⟨t⟩, violating the undominated CODACONDITION. Candidate (c) obeys CODACONDITION by shifting the ⟨k⟩ to the second syllable, but this results in an onset cluster which violates the undominated *COMPLEX. Although candidate (d) violates neither structural constraint, it violates the faithfulness constraint MAX-IO by deleting the underlying /k/ in the surface output. Although candidate (a) violates DEP-IO, it obeys all the higher-ranked constraints and thus is chosen as the optimal candidate.

Since the epenthesis vowel in the ⟨surface⟩ form is high and between voiceless segments, it becomes a target for devoicing. There are a number of ways formalize devoicing using a combination of articulatory and cue constraints. The necessity of articulatory constraints during production is obvious since the speaker must produce an output using articulators. However, cue constraints are also necessary during production because the speaker must have an acoustic target for what is being produced as well (Boersma, 2009). The current paper proposes the following possible formulation of an articulatory constraint, which prohibits a short closed glottis (voicing) gesture associated with a vowel between two spread glottis (voiceless) gestures: *[s.g.][short c.g., vowel][s.g.]. The constraint can include more articulatory detail, such as an explicit duration for [short c.g.], but the threshold for what constitutes a short glottal gesture may vary by speaker and speech style. For the current discussion, it seems sufficient to simply point out that regardless of the threshold, it is high vowels that are the most likely to devoice because they are inherently the shortest vowels. Given such an articulatory constraint, it must outrank cue constraints such as *⟨u⟩ [], which was introduced in tableau (3), as well as other articulatory constraints such as *[vowel, s.g.], which prohibits voiceless vowels. This is shown in tableau (6) below with surface candidate (a) from tableau (5) as the input. Since both devoiced and deleted forms are possible, the ranking of the two lower-ranked constraints are assumed to vary depending on the speaker.

(6) Devoicing epenthetic vowel during production

	⟨ka.ku.too⟩	*[s.g.][short c.g., vowel][s.g.]	*⟨u⟩ []	*[vowel, s.g.]
☞	a. [ka.k.too]		*	
☞	b. [ka.k̥u.too]			*
	c. [ka.ku.too]	*!		

The high vowel devoicing process for ‘hard fight’ in (4a) is simple. Because there are no phonotactic violations that result from concatenating the two Sino-Japanese roots, the faithful candidate is chosen as the optimal ⟨surface⟩ form, syllabified as ⟨ku.too⟩. The surface form is then evaluated by the same cue and articulatory constraints as in tableau (6) above, resulting in two possible overt outputs: [k.too, k̥u.too].

Next, consider the input ⟨e.bu.zo⟩, shown in tableau (7) below.

(7) Non-devoicing high vowel during production

	⟨e.bu.zo⟩	*[s.g.][short c.g., vowel][s.g.]	*⟨u⟩ []	*[vowel, s.g.]
☞	a. [e.bu.zo]			
	b. [e.b.zo]		*!	
	c. [e.b̥u.zo]			*!

In this case, the articulatory *[s.g.][short c.g., vowel][s.g.] does not eliminate any candidate because the high vowel is not between voiceless segments. Therefore, the effects of the lower-ranked cue constraint *⟨u⟩ [] and articulatory constraint *[vowel, s.g.] emerge, the former eliminating candidate (b), which fatally deletes a surface ⟨u⟩ in the overt output, and the latter eliminating candidate (c), which fatally contains a voiceless vowel in the overt output.

4 Summary and conclusion

This paper demonstrated that the apparent phonotactic violations that result from high vowel devoicing/deletion can be reconciled by employing a more sophisticated representation of phonological levels. Based on the observation that high vowel devoicing targets both underlying and epenthetic vowels, devoicing was argued to be a late, overt level process that occurs after necessary phonotactic repairs are made at the surface level. Since vowels only need to be represented at the surface level for high vowel devoicing to apply, the process is unconcerned with whether the target vowel is underlying or epenthetic.

Although not the focus of this paper, the separation of ⟨surface⟩ and [overt] levels also helps explain how underlying obstruents can be produced as their allophones even in cases of high vowel deletion. Consider the examples in (8), where (8a) is a native Japanese word and (8b) is a Sino-Japanese compound. Even though the underlying high vowels (underlined> are ultimately deleted in the overt form, the initial obstruents still undergo allophony because it is triggered at the surface level by highly ranked phonotactic constraints such as *si and *hu (Ito & Mester, 1995).

- (8) a. ‘tongue’
 /sita/ → ⟨ɕi.ta⟩ → [ɕ_.ta, *s_.ta]
 b. ‘unhappy (un+happy)’
 /hu+koo/ → ⟨ɸu.koo⟩ → [ɸ_.koo, *h_.koo]

Lastly, since syllabification occurs at the surface level before devoicing/deletion applies, the proposed theoretical account predicts that onset consonants that become stranded due to high vowel deletion should not resyllabify. Using the word ‘hard fight’ again as an example, once the word is syllabified as ⟨ku.too⟩, the overt form after devoicing applies would retain the structure even in the case of vowel deletion, leading to the overt form [k.too]. This prediction is supported by a recent study by Shaw & Kawahara (2018a), which showed that stranded onset consonants do not show c-center effects that would be expected in the case of resyllabification.

References

- Apoussidou, Diana (2007). *The Learnability of Metrical Phonology*. Ph.D. thesis, University of Amsterdam.
- Beckman, Mary (1996). When is a syllable not a syllable? Otake, Takashi & Anne Cutler (eds.), *Phonological Structure and Language Processing: Crosslinguistic Studies*, Speech Research 12, Mouton de Gruyter, 95–123.
- Boersma, Paul (1998). *Functional Phonology: Formalizing the Interaction Between Articulatory and Perceptual Drives*. Holland Academic Graphics, The Hague.
- Boersma, Paul (2007). Some listener-oriented accounts of h-aspiré in French. *Lingua* 117:12, 1989–2054.
- Boersma, Paul (2009). Cue constraints and their interactions in phonological perception and production. Boersma, Paul & Silke Hamann (eds.), *Phonology in Perception*, Berlin: Mouton de Gruyter, 55–110.
- Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier & Jacques Mehler (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception & Performance* 25, 1568–1578.
- Dupoux, Emmanuel, Erika Parlato, Sónia Frota, Yuki Hirose & Sharon Peperkamp (2011). Where do illusory vowels come from? *Journal of Memory and Language* 64, 199–210.
- Faber, Alice & Timothy J. Vance (2000). More acoustic traces of “deleted” vowels in Japanese. Nakayama, Mineharu & Charles J. Quinn Jr. (eds.), *Japanese/Korean Linguistics*, vol. 9, 100–113.
- Fujimoto, Masako (2015). Vowel devoicing. Kubozono, Haruo (ed.), *Handbook of Japanese Phonetics and Phonology*, Mouton de Gruyter, chap. 4.
- Hasegawa, Nobuko (1979). Casual speech vs. fast speech. *Proceedings of the Chicago Linguistics Society*, vol. 15, 126–137.
- Hayes, Bruce (1999). Phonetically driven phonology: The role of Optimality Theory and inductive grounding. Darnell, Michael, Edith Moravcsik, Michael Noonan, Frederick J. Newmeyer & Kathleen M. Wheatley (eds.), *Functionalism and Formalism in Linguistics*, John Benjamins, Amsterdam, 243–285.
- Hirayama, Manami (2009). *Postlexical Prosodic Structure and Vowel Devoicing in Japanese*. Ph.D. thesis, University of Toronto.
- Ito, Junko (1986). *Syllable Theory in Prosodic Phonology*. Ph.D. thesis, University of Massachusetts, Amherst. Published 1988. Outstanding Dissertations in Linguistics series. New York: Garland.

- Ito, Junko & Armin Mester (1995). Japanese phonology. Goldsmith, John (ed.), *Handbook of Phonological Theory*, Blackwell, Cambridge, MA, 817–838.
- Ito, Junko & Armin Mester (2015). Sino-Japanese phonology. Kubozono, Haruo (ed.), *Handbook of Japanese Phonetics and Phonology*, Mouton de Gruyter, chap. 7.
- Jun, Sun-Ah & Mary Beckman (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*.
- Kondo, Mariko (2005). Syllable structure and its acoustic effects on vowels in devoicing. van der Hulst, Harry, Jan Koster & Henk van Riemsdijk (eds.), *Voicing in Japanese*, Mouton de Gruyter, 229–246.
- Kuriyagawa, Fukuko & Masayuki Sawashima (1989). Word accent, devoicing and duration of vowels in Japanese. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics* 23, 85–108.
- Pinto, Francesca (2015). High vowels devoicing and elision in Japanese: A diachronic approach. *International Congress of Phonetic Sciences 18*.
- Shaw, Jason Anthony & Shigeto Kawahara (2018a). Consequences of high vowel deletion for syllabification in Japanese. *Proceedings of the Annual Meetings on Phonology*, vol. 5.
- Shaw, Jason & Shigeto Kawahara (2018b). The lingual articulation of devoiced /u/ in Tokyo Japanese. *Journal of Phonetics* 66, 100–119.
- Tesar, Bruce (1997). An iterative strategy for learning metrical stress in Optimality Theory. Hughes, Elizabeth, Mary Hughes, Annabel Greenhill, Elizabeth Hughes, Mary Hughes & Annabel Greenhill (eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development*, Cascadilla Press, Somerville, MA, 615–626.
- Tesar, Bruce & Paul Smolensky (1996). Learnability in Optimality Theory (long version). Technical Report, Department of Cognitive Science.
- Tsuchida, Ayako (1997). *Phonetics and Phonology of Japanese Vowel Devoicing*. Ph.D. thesis, Cornell University.
- Tsuchida, Ayako, Shigeru Kiritani & Seiji Niimi (1997). Two types of vowel devoicing in Japanese: Evidence from articulatory data. *Journal of Acoustical Society of America* 101:5, p. 3177.
- Varden, J. Kevin (2010). Acoustic correlates of devoiced Japanese vowels: Velar context. *The Journal of English and American Literature and Linguistics* 125, 35–49.
- Whang, James (2018). Recoverability-driven coarticulation: Acoustic evidence from Japanese high vowel devoicing. *Journal of the Acoustical Society of America* 143:2, 1159–1172.
- Zsiga, Elizabeth (2000). Phonetic alignment constraints: Consonant overlap in English and Russian. *Journal of Phonetics* 28, 69–102.