Nat Lang Linguist Theory (2018) 36:175–217
DOI 10.1007/s11049-017-9368-x

Boundary phenomena and variability in Japanese high vowel devoicing

Oriana Kilbourn-Ceron[1] · **Morgan Sonderegger**[1]

Received: 2 June 2016 / Accepted: 19 April 2017 / Published online: 10 May 2017

Abstract Devoicing of high vowels (HVD) in Tokyo Japanese applies in two environments—between voiceless consonants, and between a voiceless consonant and a "pause"—and applies variably as a function of a number of factors. The role and definition of "pause" in this process, in terms of a physical pause or prosodic position (word or phrase boundary), remains unclear, as does what is expected when these environments overlap, and why HVD appears to be categorical in some environments and variable in others. This paper addresses three outstanding issues about HVD—the role of "boundary phenomena" (prosodic position and physical pauses), the relationship between the two environments, and the sources of variability in HVD—by examining vowel devoicing in a large corpus of spontaneous Japanese. We use mixed-effects logistic regression to model how boundary phenomena affect the likelihood of devoicing and modulate the effects of other variables, controlling for other major factors, including a measure of gestural overlap. The results suggest that all boundary phenomena jointly affect devoicing rate, and that prosodic phrase boundaries play a key role: variability in HVD looks qualitatively different for phrase-internal and phrase-final vowels, which are affected differently by word frequency, speech rate, and pause duration. We argue the results support an account of HVD as the result of two overlapping vowel devoicing processes, each widely-attested cross-linguistically: devoicing between voiceless consonants, and devoicing before prosodic phrase boundaries. Variability in the application of these two processes can then be partially explained in terms of aspects of phonetic implementation and processing: gestural overlap (Beckman 1996), which often plays a role in reduction processes, and the locality of production planning (Wagner 2012), a recent explanation for variability in the application of external sandhi processes.

✉ O. Kilbourn-Ceron
oriana.kilbourn-ceron@mail.mcgill.ca

[1] McGill University, 1085 Dr. Penfield, Montreal, QC H3A 1A7, Canada

Keywords Phonological variability · Prosodic boundaries · Corpus phonology · Vowel devoicing

1 Introduction

1.1 Categorical and variable devoicing

A highly salient feature of Standard Japanese is the devoicing of the high vowels *i* and *u* (henceforth *high vowel devoicing*: HVD). Since the earliest descriptions of this alternation, the environment has been described disjunctively: high vowels are devoiced when they appear between two voiceless consonants, or when preceded by a voiceless consonant and followed by a pause (Han 1962; McCawley 1968). McCawley (1968) gives the following rule:

Rule 1: $V_{[+high]} \rightarrow V_{[-voice]} / C_{[-voice]}\_ \{C_{[-voice]}, \#\}$

Although the generalization captured by this rule remains the starting point for standard descriptions of HVD (Vance 2008; Labrune 2012; Fujimoto 2015), a distinction is normally made between the two environments. Labrune (2012) states that for a high vowel between voiceless consonants, hereafter referred to as "C̥_C̥," devoicing is "almost compulsory." Nielsen (2015) similarly describes HVD as "almost obligatory in the Tokyo dialect, except in some environments where complete devoicing is often blocked." By contrast, Vance (2008:210) notes that devoicing before a pause, hereafter referred to as "C̥_#," is "much less consistent" than in the C̥_C̥ environment. Hence, we are faced with the puzzle that HVD is compulsory, yet sometimes variable.[1] The cause of this difference in variability, and more generally what conditions how often HVD applies in a particular context, remains an open question in the literature. Addressing this question is one goal of this paper.

This question connects to a broader issue: by what mechanism can the "same [phonological] process" be categorical (or nearly so) in some environments, and variable in others? This puzzle has been of interest for HVD in particular, where previous work has ascribed categorical versus variable application to "phonological" versus "phonetic" devoicing (see Sect. 2.1); many other cases intuitively involve 'more variability' at some sort of boundary, e.g. Hungarian vowel harmony, English and Navajo phonotactic restrictions (Hayes and Londe 2006; Martin 2011). Addressing this issue in the case of HVD is relevant for understanding other such cases cross-linguistically, and how to account for them in a formal analysis.

1.2 Devoicing and overlapping environments

The distribution of devoiced vowels in Japanese follows several constraints that are attested cross-linguistically, as shown in typological reviews of non-modal vowels in general (Gordon 1998) and devoicing in domain-final positions in particular

[1] Note that the literature refers to cases of non-application of Rule 1 either as "blocking" or "variability." We use the term "variability" for any non-categorical application of Rule 1.

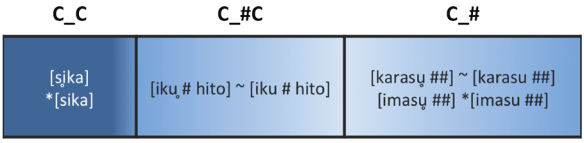


Fig. 1 Schematic representation of the high vowel devoicing environments in Japanese. Darker shade represents more categorical application of HVD. Glosses: *shika* 'deer', *iku hito* 'person who is going', *karasu* '(it's a) crow', *imasu* 'be (animate, formal)'

(Barnes 2006, Sect. 3.6.1). Many languages show a pattern like Japanese, where high vowels but not non-high vowels undergo devoicing; however, the inverse pattern is unattested (Gordon 1998). The environments for vowel devoicing in a given language can make reference to adjacent voiceless consonants, to position within a prosodic domain, or both. For example, Turkish (Jannedy 1995) and Montreal French (Cedergren and Simoneau 1985) devoice high vowels only in the C̥_C̥ environment, a subset of Japanese HVD environments. In other languages, vowel devoicing is conditioned by final position without reference to segmental context: Ainu (Crothers et al. 1979) and Woleaian (Sohn 1975) have devoicing in word-final position, while languages like European French (Smith 2003), Oneida (Michelson 1999), and Greek (Dauer 1980; Kaimaki 2015) devoice vowels at the end of larger, phrasal domains. In the surveys of both Gordon (1998) and Barnes (2006), devoicing at the end of "smaller" and "larger" domains are always in an implicational relationship within a language: devoicing at a smaller domain edge (e.g. word) implies devoicing at larger domain edges (e.g. utterance).

For Japanese, the environment for HVD takes into account both segmental conditions and domain position. But while the segmental conditions on HVD are clear, the role of domain position (i.e. the meaning of "#" in C̥_#) is not well-understood. In a recent review article, Fujimoto (2015) uses the terms *pre-pausal*, *word-final* and *phrase-final* to describe the C̥_# context, though she notes that "[f]urther investigation is essential in order to clarify the details of word/phrase-final devoicing" (186).

Describing two separate environments for this alternation obscures the fact that the environments can and often do overlap, as schematized in Fig. 1. On the left is the C̥_C̥ environment, where devoicing seems to be obligatory if all the relevant segments are within the same word and no factors blocking devoicing are present (e.g. a pitch accent; see Sect. 2.1). The right hand side gives an example of utterance-final devoicing which is variable for most words, though obligatory for a small set of high frequency verb particles (Maekawa and Kikuchi 2005; Vance 2008; Oi 2013). The overlap between the two environments is shown in the middle, tentatively labelled as variable. But it is not totally clear what is expected in a case where, for example, a word ending in a voiceless consonant and high vowel is followed by a short pause and then another word that begins with a voiceless consonant.

Would such vowels show categorical devoicing, since they are inter-consonantal, or would the devoicing be variable, since the pause precedes the following consonant? Surprisingly little of the substantial literature on Japanese devoicing has directly addressed this issue. A central goal of this paper is to understand what happens when these environments overlap, and more generally, what the relationship is between the two devoicing environments.

Addressing these questions in the case of HVD connects to the broader issue of how to analyze (variable) phonological processes that can apply in *overlapping* environments. Many devoicing processes cross-linguistically fit this description, as do many sandhi phenomena, which can often apply both across or within words (Kaisse 1985; e.g. North American English flapping). Should such cases be analyzed as two distinct processes with overlapping environments, or one process with complex conditioning factors?

1.3 Boundary phenomena

Crucial to understanding the relationship between the two environments for HVD is a definition of what exactly constitutes the C̥_# environment. This question also has rarely been addressed, and as far as we know has not been investigated empirically. Previous work suggests that "physical silent pause" is not sufficient to characterize C̥_# devoicing, although high vowels do become devoiced before some pauses (see Sect. 2.4). Setting aside disfluencies, all cases of pre-pausal devoicing in natural speech are at a *word boundary*. A number of factors come into play at word boundaries which would not affect word-internal vowels; any of these could be responsible for the effect of "pause." A number of candidates for such boundary phenomena affecting devoicing rate are raised in the literature—prosodic boundaries, pauses, word boundaries—and will be reviewed further below. The relative role of these boundary phenomena is not clear. Hence, another goal of this paper is to clarify how these boundary phenomena affect devoicing rate, and in doing so to help characterize the C̥_# devoicing environment.

Addressing this goal for the HVD case is relevant for the more general issue of how boundary phenomena affect variable (phonological) processes, and how to account for these effects formally. Many variable processes are said to be conditioned by prosodic boundaries (e.g. Nespor and Vogel 1986), a physical pause (Stevens 2012), or word boundaries (Kiparsky 1985)—but it is often difficult to tease these effects apart given how frequently different kinds of boundary phenomena co-occur (see Sects. 2.3, 2.4). Clarifying the empirical picture of how different boundary phenomena affect a variable process crucially informs theoretical accounts. If it turns out that only a single kind of boundary is relevant, this can be accommodated in existing theoretical treatments using a formal object that indexes the boundary, e.g. Optimality Theory constraints referring to faithfulness in "pre-pause position" (Coetzee and Pater 2011) or alignment at prosodic word boundaries (Nagy and Reynolds 1997). If different boundary phenomena have distinct or interacting effects, a theoretical treatment becomes more complicated.

1.4 Summary

We address the study's three goals related to Japanese vowel devoicing—the source of variability, the relationship between the two devoicing environments, and the role of boundary phenomena—by conducting a multivariate statistical analysis of devoicing in a large corpus of spontaneous speech (Maekawa et al. 2000). The analysis models how different possible correlates of "pause" affect devoicing rate, while controlling

for a number of other factors which condition HVD. To address the relationship between the two devoicing environments, the analysis also examines how the effect of other factors depend on the position of the devoiceable vowel.

The results show that, in accordance with native speaker intuitions, devoicing is nearly categorical, but only under certain conditions. HVD is most consistent word-internally, and also at sufficiently "large" domain edges: phrase boundaries that are followed by longer pauses. In other conditions, HVD applies variably. We find that how other factors affect the rate of application of HVD is modulated by the position of the vowel within prosodic phrases: in particular, speech rate and frequency have qualitatively different effects for vowels at the edge of sufficiently "small" domains versus larger domains. This finding leads us to suggest that devoiced vowels in Japanese may be best understood as the result of two different devoicing processes which apply in different, but sometimes overlapping environments. We suggest that some of the variability in these processes can be understood by reference to two sources in phonetic implementation and processing: gestural overlap, which has been previously discussed in the context of HVD (Jun and Beckman 1993; Beckman 1996), and the *locality of production planning* (Wagner 2012), which has not.

In the remainder of this paper, we first present a review of previous findings on variability in high vowel devoicing in Japanese, and outline specific research questions (Sect. 2). We then describe the data (Sect. 3) and methods (Sect. 4) of our corpus study addressing these questions, and present the results (Sect. 5). We conclude with interpretation of these results and discussion (Sect. 6), including with reference to the broader issues discussed above beyond the Japanese HVD case.

2 Background

Vowel devoicing in Japanese is the subject of a long literature, which comes from many different perspectives (e.g. Han 1962; McCawley 1968; Hasegawa 1979; Yoshida and Sagisaka 1990; Vance 1992; Jun and Beckman 1993; Beckman 1996; Kondo 1997; Tsuchida 1997; Varden 1998; Maekawa and Kikuchi 2005; Hirayama 2009; Varden 2010; Ogasawara 2013; Nielsen 2015: see Fujimoto 2015 for a recent review). This section gives a brief summary of previous work on high vowel devoicing, focusing on aspects of importance for this paper: variability and the factors affecting variability, especially the role of word boundaries, prosodic information, and pauses.

2.1 High vowel devoicing

In Japanese, it is generally assumed that the high vowels /i/ and /u/ have devoiced allophonic variants, [i̥] and [u̥].[2] Textbook descriptions (e.g. Vance 2008; Fujimoto 2015)

[2]There is some debate over whether HVD should be described as "devoicing" or "deletion," or whether both occur (see Fujimoto 2015, Sect. 4.4). This paper is agnostic to this issue, as we abstract away from the phonetic realization of the vowel and focus on the factors conditioning the probability of application of HVD. We follow Fujimoto (2015) and most previous work in referring to HVD as "devoicing," for convenience.

Table 1 Examples of words typically pronounced with devoiced vowels in Standard Japanese from Vance (2008)

Preceded and followed by voiceless consonant			V → V̥/C̥_C̥
a.	*sika*	[ʃi̥ka]	'deer'
b.	*kusa*	[ku̥sa]	'grass'
Voiceless consonant followed by pause			V → V̥/C̥_#
c.	*ikimasu*	[ikimasu̥]	'(I will) go'
d.	*karasu*	[karasu̥] ~[karasu]	'(It's a) crow'

and pronunciation manuals (NHK 1991:Japanese Pronunciation Accent Dictionary) give the generalization that the high vowels should be devoiced when they are preceded by a voiceless consonant and followed either by another voiceless consonant or by a pause. Examples of typically devoiced vowels are given in Table 1.

However, not all high vowels in the C̥_C̥ and C̥_# environments are devoiced. The most important factor is the restriction on devoiced vowels in adjacent syllables: if vowels in consecutive syllables are both in an HVD environment, generally only one of the vowels is devoiced. Also, for some speakers, the presence of a pitch accent or high tone may block HVD. But modulo these blocking factors, HVD is considered compulsory in standard (Tokyo area) Japanese (e.g. Hirayama 1985: cited in Fujimoto 2015). This assumption underlies phonological analyses of HVD in the literature, where devoicing is analyzed as categorical assimilation of laryngeal features from surrounding consonants, either [−voice] (e.g. Han 1962; McCawley 1968) or [+spread glottis] (Tsuchida 1997, 2001). The blocking effects can also be handled in a phonological analysis treating HVD as a categorical phenomenon, for example as proposed in Tsuchida (2001) and Kondo (2005).

However, other work argues that a number of factors gradiently affect devoicing rates in a way that is not easily captured in a categorical phonological analysis. Phonetically-oriented studies of devoicing argue that categorical phonological accounts are belied both by the gradient influence of phonetic context on the rate of devoicing, and by the range of possible realizations of devoiced vowels, including partial devoicing and total deletion (Jun and Beckman 1993; Beckman 1996).

Beckman (1996) proposes that devoicing of high vowels is due to gestural overlap—the encroachment of the glottal gestures for surrounding voiceless consonant—rather than a phonological change within the vowel itself (e.g. to [−voice]). In this account, varying articulatory conditions are naturally predicted to affect the likelihood of vowels being produced without voicing as the competing glottal gestures are compressed or change in magnitude for independent reasons. For example, it has consistently been found that vowels preceded by fricatives are devoiced at higher rates than those preceded by stops (see Sect. 2.5). Beckman (1996) suggests that this pattern is predicted by the articulatory differences between stops and fricatives.

This tension between the obligatoriness of HVD in many cases and its variability in others is at the core of much debate over the extent to which HVD is 'phonological' or 'phonetic,' and has led to proposals that both phonological and phonetic mechanisms are necessary to account for HVD (Tsuchida 1997; Varden 1998;

Nielsen 2015). Tsuchida (1997) proposed that HVD is phonological in environments where it is categorical, but due to gestural overlap in variable cases. Nielsen (2015) showed that both phonetic and phonological factors must be taken into account to predict the realization of HVD in consecutive devoicing environments, arguing that HVD is driven by both types of factors.

Distinguishing between phonological and phonetic vowel devoicing is a challenge in many different languages (Gordon 1998). In the Japanese case, this debate is complicated by the ambiguous meanings of 'phonological' and 'phonetic': in previous work, these are often used as shorthand for 'categorical' and 'variable,' following one longstanding criterion, but variable processes are now routinely addressed in phonological theory (Coetzee and Pater 2011; Coetzee and Kawahara 2013), notably for Japanese (e.g. Kawahara 2011). In this study, we do not directly address the question of which mechanisms underlie HVD in Japanese, but we do take into account both phonological and phonetic factors which have not previously been investigated, and delimit some conditions under which HVD is categorical versus variable, potentially offering some new insights for this debate.

2.2 Word boundaries

The literature on high vowel devoicing offers evidence that word boundaries affect variability, although their role has not often been the focus of direct investigation. Vance (1992) argues that one of the factors which disfavors devoicing is the presence of a morphological boundary between a potential target of HVD and the following voiceless consonant: in compound words containing consecutive devoicing environments in the NHK (1991) pronunciation dictionary, if one of the target vowels is followed by a morphological boundary, it is the other vowel that devoices. Varden (1998) reported a similar result from a production experiment. In words containing a consecutive devoicing environment, speakers devoiced the word-final vowel less often than the penultimate vowel in the same word. For example, in the first word in the sentence *Tsuki to hoshi ga kakureta*, the first vowel in *tsuki* was devoiced more often than the second.[3]

As part of a larger study on sociolinguistic effects in HVD, Imai (2004) investigated the effect of different morpheme boundary types, distinguishing between five possible cases: morpheme internal, pause, bound morpheme boundary, compound word boundary, and word boundary. A logistic regression analysis (using Goldvarb software) showed that the morpheme-internal and bound morpheme cases were most likely to devoice, followed by pause and then compound and word boundaries. However, Imai's Table 4.20 shows that when vowels in consecutive devoicing environments are excluded, devoicing rates are more similar for morpheme internal (78%) and word boundary (71%) cases than for bound morphemes (66%) and compound boundaries (35%).

In sum, these results from consecutive devoicing studies suggest that morpheme and word boundaries have some inhibitory effect on HVD, relative to presumably

[3]Note that Varden (1998) interprets this result as a linear order effect, but due to his stimuli construction, linear order is not distinguishable from word boundaries.

categorical application within a morpheme. That being said, previous work agrees that HVD is possible across both compound-internal morpheme boundaries and word boundaries of all types, regardless of syntactic constituency (Kaisse 1985; Vance 1992; Kondo 1997).

Turning to the C̥_# environment in particular: word boundaries are closely tied to this "pre-pausal" environment, since examples given in the literature almost always involve pauses that follow word boundaries. This means that the effect of pause is confounded with the effect of a word boundary (e.g. (c) and (d) in Table 1). One study where this is not the case is Vance (1992), who gives the example of syllable-by-syllable pronunciations of words with devoicing environments. He states that if words are pronounced in this way, devoicing of word-internal vowels is blocked. If this is so, it constitutes evidence that word boundaries are at least a necessary condition for C̥_# devoicing to apply. Whatever the most accurate characterization of the C̥_# devoicing environment turns out to be, it will likely be a subset of vowels at word boundaries.

Word boundaries may also be important for devoicing in that they modulate the effect of other factors. While consonant manner and speech rate effects have been reported in many studies (see Sect. 2.5), Kondo (1997) found that consonant manner and speech rate effects were not statistically significant when considering only word-internal single devoicing environments. Hence, these types of effects may be dependent on the presence of a word boundary. More broadly, there is a running question throughout the literature on HVD as to the 'level' at which devoicing applies, closely corresponding to the debate on the 'phonetic' versus 'phonological' nature of HVD discussed above. Vance (1992), in the context of Lexical Phonology, discusses a possible distinction between lexical and post-lexical applications of high vowel devoicing. Within this framework, only post-lexical process/rules should be affected by speech rate and pauses (Mohanan 1982; Kaisse 1985).[4] In this study, we compare how devoicing rates are affected by pauses and speech rate in different prosodic positions, including a direct comparison between word-internal and word-final vowels. If it is the case that the effect of pauses and speech rate differs between these two environments, it would lend support to the view that there are two qualitatively different processes that underlie the pattern of high vowel devoicing in Japanese.

In sum, previous work suggests that word boundaries are related to variability in two ways: inhibiting C̥_C̥ devoicing, and as a necessary condition for the variable C̥_# devoicing. A focus of this paper is devoicing variability in those cases where C̥_C̥ and C̥_# environments overlap, a perspective which has not generally been considered in previous work.

2.3 Prosodic organization

We begin with a brief review of the prosodic organization of utterances in Japanese, with reference to the X-JToBI system of prosodic annotation (Maekawa et al. 2002) which will be relevant for our corpus study. We then review findings and comments from the literature on how prosodic information might influence HVD.

[4]Note that only post-lexical applications could apply across words; hence C̥_# devoicing must result from postlexical rule application, while C̥_C̥ devoicing could result from lexical or postlexical rule application.

Table 2 Prosodic constituency and corresponding break index annotation for *Sankaku no yane no mannaka ni okimasu* 'I will place it right in the centre of the triangle roof' (Venditti 2005:176)

Accentual phrase	{ }	{ }	{ }	{ }
Intonation phrase	[ ]	[ ]	[ ]	[ ]
Tones	%L H*L L%	H*L L%	H- L%	L%
Break Indices	1 2	1 3	1 3	3
	sa'Nkaku no	*ya'ne no*	*maNnaka ni*	*okima'su*
	triangle-GEN	roof-GEN	middle-LOC	put

Above the level of the word, it is commonly argued that Japanese utterances are organized into two hierarchical groupings, although theoretical treatments differ as to the relationships between these levels (e.g. Beckman and Pierrehumbert 1988; Ito and Mester 2012). Here we call these levels the accentual phrase (AP) and the intonation phrase (IP), following Beckman and Pierrehumbert (1988), Venditti et al. (2008). These groupings reflect the syntactic constituency of the utterance, but are not necessarily isomorphic to it. For example, the utterance in Table 2 is organized into four APs, which are in turn grouped into three IPs.

These groupings are reflected in both the Tone and Break Index annotations in the X-JToBI system. The Break Index annotations are marks of "degree of perceived disjuncture between words," which listeners judge on the basis of several cues such as pausing, segmental lengthening, F0 lowering or resetting, and creaky voice quality (Venditti 2005:184–185). In X-JToBI, each word boundary is assigned a number from 1 to 3, with 3 indicating the highest degree of disjuncture. As shown in Table 2, the Break 2 and 3 annotations are typically associated with AP and IP boundaries, respectively.

As for the Tone annotations, the location of tonal targets is constrained by prosodic phrasing, hence these annotations offer some information about the prosodic organization of the utterance. The typical contour of an AP is an initial rise, marked in Table 2 by the *%L H-* annotation, followed by a gradual decline to a final low target, *L%* (Beckman and Pierrehumbert 1988; Venditti 2005). The AP also constrains the placement of lexical pitch accents, so that a single AP may contain at most one pitch accent. The IP is the domain to which boundary pitch movements (BPMs) are anchored, for example to signal a question or surprise (see Venditti et al. 1998 for a detailed description of BPMs). The IP is also the domain of F0 downstep, so that each AP within a single IP becomes lower in pitch range, until F0 is "reset" at the beginning of a new IP (Beckman and Pierrehumbert 1988).

While the effect of tones on HVD, especially pitch accents, has been investigated in several studies (Kuriyagawa and Sawashima 1989; Hirayama 2009; Oi 2013), the effect of phrasal boundaries per se on HVD has not been systematically tested. The mentions of prosodic boundary effects on HVD in the literature relate to the definition of the C̥_# environment. For example, Kondo (1997), based on evidence from production experiments, suggests that the C̥_# environment should instead be characterized as "utterance-final."

In the current paper, we will focus on Break Indices as the operationalization of prosodic phrase boundaries. However, information from Tone annotations will also

be included in the model as a control, since previous literature suggests that high tones, particularly pitch accents, may block devoicing for some speakers. With this in mind, we now discuss prosodic phrase boundaries in particular.

2.3.1 Phrase boundaries

Little previous work has addressed the effect of phrase boundaries on HVD per se, but phrase boundaries could plausibly decrease or increase devoicing rate.

The idea that stronger phrase boundaries may have an inhibitory effect on HVD seems plausible from a gestural overlap perspective, since phrase boundaries in Japanese (and many other languages) are associated with final lengthening (e.g. Takeda et al. 1989; Wightman et al. 1992; Den 2015), in line with articulatory strengthening at phrase boundaries cross-linguistically (Fougeron and Keating 1997). If HVD is due to overlap of adjacent laryngeal gestures, producing a longer vowel should make it more likely that the vowel's voicing gesture will have time to be realized. Using this logic, Den and Koiso (2015) attribute the negative relationship between devoicing rate and mora duration in utterance-final position to final lengthening. This same logic would apply to *any* possible HVD site—the less gestural overlap obtains, the less likely the voicing gesture will be fully realized. Thus, we include a rough measure of gestural overlap among the variables in our model: *Mora deviation*, defined as the difference between the current Mora's duration and its average duration in the corpus (Wightman et al. 1992). ("Mora" is capitalized for reasons explained below.)

However, there is also good reason to think that phrase boundaries would *increase* devoicing rate. Domain-final vowel devoicing is very common cross-linguistically (e.g. the Greek, French, and Oneida cases discussed above), and has clear phonetic motivation in the drop of subglottal pressure at utterance/phrase endings (Gordon 1998; Barnes 2006). In Japanese in particular, it has been suggested that IP boundaries are the triggers for C̥_# devoicing (Kondo 1997; Hirayama 2009; Fujimoto 2015). Also, prosodic phrasing is well-established as a unit for tonal organization in Japanese, so it seems plausible that segmental processes such as HVD would also be triggered by prosodic phrase boundaries.

To our knowledge, whether prosodic boundaries (e.g. IP) affect HVD has not been *empirically* tested. It is particularly difficult to assess whether it is a prosodic boundary per se which affects devoicing rate, or another boundary phenomenon. Phrase boundaries always coincide with word boundaries, and often with pauses, which are a major cue to intonation phrase boundaries (Venditti 2005). The occurrence of prosodic phrase boundaries are highly correlated with the occurrence and length of pauses, making it difficult to distinguish their relative contributions to devoicing rate. With the large corpus of spontaneous speech used in the present study, we are able to investigate the effect of a prosodic boundary, which we operationalize as Break Indices, while also controlling for pauses and other possible confounding factors (e.g. final lengthening, as assessed by Mora deviation). Given that we restrict our data to tokens which are followed (and preceded) by voiceless consonants, we are also able to investigate the interaction between prosodic boundaries and C̥_C̥ devoicing, a novel empirical contribution to the HVD literature.

2.4 Pause

The term "pause" is traditional and often used in the description of C̥_#. Taking this description at face value, how does an actual physical pause affect devoicing? On the one hand, the very use of the term "pause" to define an environment for HVD suggests that a pause may promote devoicing. On the other hand, the few studies addressing the effect of a pause reach the opposite conclusion: Vance (1992) states that pauses *block* devoicing from applying where it otherwise would have, as in a syllable-by-syllable pronunciation of a word containing a C̥_C̥ environment. Kondo (1997), comparing between repetitions of the same item in a production experiment, also found a negative effect: repetitions in which a pause was present after the devoiceable vowel had *lower* devoicing rates. Den and Koiso (2015), examining a subset of the spontaneous speech dataset used in this paper (Corpus of Spontaneous Japanese), found that devoicing occurs frequently before pauses (defined as silence of at least 200 ms), but that pause *length* does not significantly affect devoicing rate. In sum, the role of pauses in promoting or blocking HVD is unclear.

However, as noted above, word boundaries, phrase boundaries and especially utterance edges are highly correlated with pauses—especially in laboratory experiments, due to the short length of test items (single words or sentences). By investigating HVD in a large corpus of spontaneous speech, we will be able to tease apart the influence of boundaries (of words and prosodic units) and pauses, and delineate their respective roles in HVD.

2.5 Other factors

We now turn to some major factors that affect the rate of HVD rate: surrounding consonant articulation, speech rate and style, and lexical frequency and idiosyncrasy. These will be used in our model both as controls, and to investigate the relationship between the C̥_C̥ and C̥_# environments.

2.5.1 Consonantal context

At a basic level, consonantal context is the most important factor in high vowel devoicing, in that the presence of voiceless consonants defines the C̥_C̥ and C̥_# environments.

The *manner* of the surrounding voiceless consonants also influences HVD. In terms of the preceding consonant, there is less devoicing after plosives than after fricatives, both in single-word productions (Kondo 1997) and in spontaneous speech (i.e. the Corpus of Spontaneous Japanese (CSJ): Maekawa and Kikuchi 2005). The effect of the following consonant is the reverse, with less devoicing before fricatives than before plosives (Nielsen 2015; Maekawa and Kikuchi 2005; Kuwabara and Takeda 1988; Lovins 1976; cf. Han 1962). The effect of a preceding or following affricate is inconsistent across studies, but generally patterns with either plosives or fricatives. The preceding and following consonant effects are not independent: a high vowel flanked by voiceless fricatives is generally *less* likely to devoice than other combinations of obstruents, in both laboratory experiments (Kondo 1997;

Tsuchida 1997; Hirayama 2009) and in the CSJ (Maekawa and Kikuchi 2005). Given the important effects of consonant manner on devoicing rate, we include in our model the manner of the preceding and following consonant.

2.5.2 Speech rate and style

Speech rate and speaking style have intuitively opposite effects on HVD: Hasegawa (1979) observed that devoicing is more likely to occur in faster speech, but *less* likely to occur in casual speech. This observation was confirmed by Martin et al. (2014), a recent corpus study comparing child-oriented, adult-oriented and read speech: high vowels devoiced significantly less in adult-oriented (i.e. conversational) speech than in read speech, but significantly more in faster speech compared to slower speech.

In contrast, Kondo (1997) found no significant effect of speech rate effect when it was tested explicitly in a production experiment, where subjects read test words embedded in paragraphs at slow, normal and fast speaking tempi. However, devoicing rates were very high for all three conditions (81–97%), as expected for a formal speech style. It may be that speaking rate effects are relatively small and are more easily observable in spontaneous speech (as in Martin et al. 2014), in which devoicing is more variable, rather than read speech. The current dataset is expected to show a small positive speech rate effect, given that it examines spontaneous speech. While we do track the effect of speech rate, we do not explicitly control for speech style, as the speech contained in the CSJ is almost all from formal settings (academic presentations, simulated public speaking).

2.5.3 Lexical frequency and idiosyncrasy

To our knowledge, the only examination of frequency effects is in Maekawa and Kikuchi (2005:218), who found a small *negative* correlation between devoicing rate and word frequency in the CSJ (empirical correlation, without controlling for other factors). This effect was found for high vowels which were preceded by a voiceless consonant, but with any kind of following segment (or lack thereof). The directionality of this frequency effect is surprising if HVD is seen as a reductive process resulting from gestural overlap, which is expected to be greater for higher-frequency words (Jurafsky et al. 2001; Pluymaekers et al. 2005); frequency and devoicing rate would then be expected to have a positive correlation.

One aspect of Maekawa and Kikuchi's data points to a positive trend: they highlight two morphemes which stood out as outliers from the negative trend, the verbal particles *desu* (polite form of copula *da*) and *masu* (an auxiliary verb of politeness). These items were among those with the highest frequency, and they also showed extremely high devoicing rates. This pattern accords with native speaker intuitions about these morphemes, as well as the findings of Oi (2013), who specifically tested utterance-final devoicing for lexical words, and found that lexical words were devoiced about 80% of the time, while the particle *masu* was always devoiced for all 10 speakers in the study. One suggested explanation for *desu* and *masu* in particular is that these functional morphemes appear almost exclusively sentence-finally. Hence, they could be much more affected by C̥_# devoicing than other types of words which rarely appear at the ends of utterances.

The case of these three morphemes means that lexical identity is another factor confounded with the boundary phenomena discussed above (e.g. IP boundary, pause). Analyzing HVD in spontaneous speech allows us to address our research questions while controlling for the high devoicing rates of certain words. In addition, by including word frequency in our multivariate model of HVD in the CSJ, we can assess the existence and directionality of a frequency effect, when other factors (such as lexical identity) are controlled for.

2.6 Summary and research questions

We have seen that many factors have been found to affect HVD rate, including consonant manner, high tones, speech rate and style, word boundaries and pauses; prosodic domain edges may also play a role. This paper focuses on three of these factors, which are confounded—word boundaries, prosodic position, and pauses—to address three research questions, in a corpus of spontaneous speech consisting of tokens in C̥_# and C̥_C̥ environments and their intersection.

First, *how do word boundaries affect devoicing rate, and modulate the effect of other factors*? Second, *how do prosodic phrase boundaries affect devoicing rate, and modulate the effect of other factors*? Previous work predicts an inhibitory effect of a word boundary on devoicing rate, and gives reasons to think that phrase boundaries (especially IP boundaries) could either increase or decrease devoicing rate. Whether and how word and phrase boundaries modulate the effects of other factors on devoicing rate will help to understand the relationship between C̥_# and C̥_C̥ devoicing; we consider speech rate, word frequency, Mora deviation, and pauses in particular. Third, *how does a physical pause (presence and duration) affect devoicing rate*? Previous work does not give a consistent prediction on how pauses should affect devoicing rate.

We address these three research questions in a dataset which was selected to best address them, and complements previous work. First, because the three 'boundary phenomena' are highly correlated, we examine HVD variability in a very large dataset of spontaneous speech (Maekawa et al. 2000), where the high degree of variation allows us to tease their effects apart, while controlling for other factors affecting devoicing rate (consonantal context, etc.), in a single statistical model.

Second, in order to understand the relationship between the C̥_C̥ and C̥_# environments, we considered only high vowel tokens which were preceded and followed by voiceless consonants (where the following consonant may occur following a word boundary or pause, in the C̥_# environment). That is, we excluded tokens in the C̥_# environment followed by a voiced segment. This exclusion allows us to understand what happens when the environments overlap, and to delimit the role of boundary phenomena by eliminating a confounding variable (following segment voicing) which could account for any observed difference between HVD application across versus within words. This restriction also means our conclusions about C̥_# position are in fact only based on a subset of the relevant data. We discuss the implications of this in Sect. 6.4.

Third, in order to focus on the effects of boundary phenomena, we only consider tokens from single-devoicing environments. Previous work on HVD variability

has largely focused on consecutive devoicing environments and lab-elicited speech—precisely *because* speakers seem to apply HVD near-categorically in single devoicing environments in laboratory speech—and it remains unclear how much variability there is in natural speech in single devoicing environments.

Thus, our study contributes a new perspective on HVD variability by examining spontaneous speech, vowels preceded and followed (eventually) by voiceless consonants, and (only) single devoicing environments.

3 Data

The source of data for this study is the Corpus of Spontaneous Japanese (Maekawa et al. 2000), a corpus of audio recordings primarily from two styles of spontaneous speech monologues: academic presentation speech and simulated public speaking. We draw from the "Core" subset of the data which, in addition to being orthographically transcribed and morphologically tagged, includes segmentally-aligned manual phonetic transcription and X-JToBI labels (Maekawa et al. 2002) to mark prosodic information. This subset contains about 44 hours of speech from 201 speakers.[5]

From the XML annotation files, we extracted all tokens of short high vowels[6] and information about whether the vowel was devoiced, immediately adjacent segments, prosody, and other factors expected to affect devoicing rate.

In the segmental phonetic transcription, vowels are transcribed as either voiced or devoiced; we used this manual annotation as our binary measure of *devoicing*. Devoicing was determined by the human labellers preparing the corpus by using information from "the wide-band spectrogram, speech waveform, extracted speech fundamental frequency, peak value of the autocorrelation function, in addition to audio playback" (Maekawa and Kikuchi 2005:208).

Word and phrase boundaries were derived from the Break Index (BI) annotations in the CSJ. These annotations involve information about the strength of a break (None/1/2/3), as well as other information (e.g. the occurrence of a pause or a "boundary pitch movement"). We collapsed BI annotations into four categories, which closely correspond to word and prosodic phrase boundaries: *None* tokens had no BI marked at the right edge of the vowel, so they are within the same word as the consonants that precede and follow them. Tokens with BI *1, 2* or *3* are at the right edge of a word. BI *1* tokens are word-final, but not final in their accentual or intonation phrase. Tokens with BI *2* are accentual phrase but not intonation phrase final,

[5] A small part of the "Core" subset (∼ 5%) consists of (spontaneous) dialogues and read speech. We found that all results reported in this paper are qualitatively the same if the read speech data (3.3% of our dataset) is excluded. Thus, we report results without excluding this data, and interpret our findings as representative of spontaneous Standard Japanese.

[6] Japanese has a phonological length distinction in vowels, and only phonologically short vowels are said to be affected by devoicing. This is corroborated by Maekawa and Kikuchi (2005) who found less than 0.5% of long high vowels and 1.2% of short non-high vowels were devoiced in the CSJ, compared to 24.3% of short high vowels.

Table 3 Summary of Break Index annotations in relation to word/phrase position of vowel token

Break Index	Position of vowel token	Number of tokens
None	word-internal	15355
1	word-final, phrase-internal	23811
2	final in accentual phrase	2361
3	final in intonation phrase	3120

while BI *3* tokens are final in their intonation phrase.[7] Table 3 shows the definition and number of tokens for each BI category.

Tone annotations were also extracted in order to control for effects of high tones. Annotations for pitch accents and other tones in the CSJ are aligned with "the corresponding F0 event" (Venditti 2005). We considered tone labels to be part of a token if their timestamps were within the start and end times of the token vowel.

Pause duration following the token was defined as the time difference between the end of the CV Mora and the beginning of the next segment. This interval sometimes included a manually annotated "pause" in the CSJ (200 ms or longer), and sometimes did not, i.e. for brief silences or other non-speech. 2634 tokens (5.9%) were followed by a pause.

As a measure of final lengthening, which is associated with larger prosodic phrase boundaries, we used a measure based on the duration of the CV sequence containing the target vowel. The duration of the vowel itself was not used because the left boundaries of devoiced vowels are often unclear, and are indicated as such in the CSJ annotations. (For example, in many [s̥u] tokens there is no clear acoustic landmark differentiating the fricative and (devoiced) vowel portion.) Our use of the duration of a larger unit than the vowel itself which can be more reliably measured follows other work examining vowel devoicing (e.g. Torreira and Ernestus 2011 for French). In the CSJ XML annotations, segments are hierarchically organized into Mora units, which include a vowel segment and its onset consonant for all tokens where HVD can apply.[8] (To avoid confusion of "mora" as referring to physical duration with the abstract weight unit used in phonological theory, we capitalize Mora throughout this paper to emphasize that it is the physical duration of a CV sequence that is referred to.) For each token, we recorded the duration of the Mora containing it. From this value we subtracted the average duration of that particular CV Mora across the CSJ corpus. This gave a measure of *Mora deviation*, a positive value if the Mora was longer than average and negative it was shorter. For example, a token of /u/ preceded by /k/ would be in a /ku/ Mora, and the difference between the duration of that Mora and the average duration of all /ku/ Moras would yield its value for Mora deviation.

We extracted two measures of *speech rate* to be included in the model. We first calculated raw speech rate as the number of phones per second in the inter-pausal unit according to the CSJ annotation (where pauses of >200 ms are manually anno-

[7]Note that Intonation Phrase boundaries (BI 3) in this dataset include "utterance" boundaries as well as "intermediate phrase" boundaries, in the terminology of Beckman and Pierrehumbert (1988) (Igarashi et al. 2006:348).

[8]Note that Japanese has moras which are not CV units (Labrune 2012), but only CV-type moras contain vowels in C̥_# and C̥_C̥ environments.

tated). Raw speech rate was used to calculate *speaker speech rate*, the average rate over all the speaker's utterances, and *local speech rate*, the difference between an utterance's speech rate and the speaker's average. Using separate speaker-level and observation-level speech rate predictors, following Snijders and Bosker (1999), allows us to differentiate between devoicing occurring more often for faster speakers, versus faster utterances (within a speaker). Both variables are in units of phones per second, such that an increase in the variable corresponds to faster speech.

The data was restricted to tokens of high vowels that were preceded and followed by voiceless obstruents ($n = 52809$). To focus on the single devoicing environment, we excluded tokens that were adjacent to other potential devoicing sites (i.e. "consecutive devoicing environments," see Sect. 2.1; $n = 7102$, 13.45% of tokens). Remaining tokens that were part of disfluencies were also excluded ($n = 984$, 5.36% of tokens). Finally, 76 tokens were excluded whose prosodic annotations reflected pathological cases or probable coding errors.[9] The final dataset contains 44647 tokens for analysis, of which 91.17% were devoiced.

4 Methods

The data was analyzed using mixed-effects logistic regression, a type of multivariate statistical model, which predicts the outcome (whether a vowel was devoiced) as a function of a number of variables (e.g. Gelman and Hill 2007; Baayen 2008). Mixed-effects logistic regression has been applied to HVD data in particular by Nielsen (2015). The advantage of using a multivariate model is that it allows the comparison of several effects at once, and the possibility of comparing their relative effect size. A mixed-effects model in particular also allows the inclusion of both fixed effects, which are the factors of interest discussed above, and random effects, which account for differences in baseline HVD rates and effect sizes within different speakers or words. The dependent variable for this study is the binary outcome of devoicing (1) or no devoicing (0) based on the phonetic transcription in the corpus. Hence, positive coefficient estimates indicate an increase in the likelihood of devoicing. More precisely, each coefficient gives the estimated effect of a factor of interest on the *log-odds* of devoicing.

4.1 Model terms

We now turn to the variables which are included in the statistical model as fixed or random effects, and how they are related to our research questions.

Word and phrase boundaries The four-level Break Index (BI) is the independent variable of primary interest, as it lets us examine the effect of word and phrase boundaries. This variable was included in the model as a four-level categorical variable with Helmert contrast coding. With this type of coding, the estimated coefficients

[9] These were: all remaining tokens whose (collapsed) *Break Index* was not 1, 2, 3, or None followed by no physical pause.

have interpretations that directly address our first and second research questions. The first coefficient will compare the devoicing rate in word-internal tokens (BI *None*) versus word-final tokens (BI *1/2/3*). The second coefficient estimates the difference in devoicing rate among word-final tokens which are phrase-internal (BI *1*) versus phrase-final (BI *2/3*). The final coefficient compares tokens at accentual phrase versus intonation phrase edges (BI *2* v BI *3*). *Break Index* is included as a main effect in the model, as well as in a number of interaction terms, discussed below.

Pauses To address our third research question, how *pause* affects the rate of devoicing, pause duration was included in the model. Because the distribution of pause duration is highly skewed, with the vast majority of tokens showing no pause or a short pause, it was not possible to include pause duration as a continuous variable.[10] Instead, pause duration was discretized into a four-level factor, called *Pause*, which allowed comparison between tokens with and without following pauses, and allowed for non-linear effects of pause duration. The first level corresponded to tokens with no pause. Tokens that did have a following pause were categorized in to three bins (levels 2–4) of roughly equal size (using the `cut2` function in R; Harrell Jr. et al. 2015) according to pause duration: less than 85 ms, 85–463 ms, and over 463 ms. The four-level factor was coded such that the intercept corresponded to no pause, and the three contrasts corresponded to Helmert contrasts: no pause vs. pause, short vs. medium/long pause, medium vs. long pause.

An interaction of pause duration with break index was included in the model, to allow for the possibility of different pause effects at different boundaries. However, because there were almost no word-internal tokens that were followed by a pause,[11] *Pause* and *Break Index* are not independent, and the model structure must somehow take into account that there can be no *Pause* effect for word-internal tokens. We did this by excluding the main effect of pause duration. Intuitively, the interaction terms describe the *Pause* effect when *Break Index* is 1, 2, or 3.

Mora deviation Mora deviation was included in the model to control for final lengthening as a confound for phrase boundaries, and to capture the effects of gestural overlap. Exploratory plots suggested a nonlinear effect of mora deviation on devoicing rate, of a roughly quadratic shape (in log-odds space). *Mora deviation* was thus coded as a nonlinear spline with three knots (using `rcs` in the `rms` R package; Harrell 2014), which corresponds to a curve with a single "bend," and included in the model as a main effect and in interactions (see below). The two spline components correspond approximately to linear and nonlinear effects, of a continuous variable. Before coding as a spline, *Mora deviation* was centered and divided by two standard deviations (Gelman and Hill 2007).

Interactions Our first and second research questions address how word and phrase boundaries modulate the effect of other variables. The model includes interactions of

[10] The distribution is highly skewed because within-word environments always show no pause, and are much more frequent than cross-word environments. Thus, discretizing pause duration is necessitated by our focus on *both* devoicing environments and the intersection between them.

[11] Such tokens exist in the corpus, but were excluded from analysis as they were determined to be mostly disfluencies.

Break Index (corresponding to phrase boundaries) with four variables in particular: *local speech rate*, *lexical frequency*, *Mora deviation*, and *Pause*. Interactions with speech rate, frequency, and Mora deviation are of interest in that differences in their qualitative effects depending on *Break Index* would bear on the relationship between C̥_C̥ and C̥_# devoicing.[12] The interaction with *Pause* is partially necessitated by the structure of the data (pauses do not occur for BI = *None*, as discussed above). The possibility of the effect of *Pause* differing at different boundary types (BI = *1*, *2*, *3*) emerged in exploratory data analysis, and will turn out to be crucial for interpreting our results.

Controls A number of other variables expected to affect devoicing rate based on previous work (Sect. 2.5) were included in the model as controls, as main effect terms. Terms were included for *Preceding consonant manner* and *following consonant manner*, coded using sum contrasts as factors with the levels *stop*, *affricate* and *fricative*, with *stop* as the base level. Based on previous findings that vowels between two fricatives have very low devoicing rates, we also included a term for the interaction between these two factors. The presence of a high tone associated with the vowel was included, as a binary predictor (of high tone presence), which was converted to a numerical variable and centered.[13]

A continuous *lexical frequency* measure was included in the model: frequency was defined as a word's count in the CSJ divided by the total number of words in the CSJ; this measure was then log-transformed.

Finally, the model includes both measures of speech rate described above, *speaker speech rate* and *local speech rate*. Frequency and speech rate predictors were centered and divided by two standard deviations (Gelman and Hill 2007).

Coding and model interpretation The coding of variables included in the model results in a straightforward interpretation of model coefficients, which will be important in interpreting our results. Holding the *Pause* contrasts at zero corresponds to a token with no pause, while all other variables have been centered, or coded using Helmert or sum contrasts, where the intercept corresponds to averaging across factor levels. Hence, the interpretation of the intercept in the statistical model reflects the estimated devoicing rate for word-internal cases with no pause, with all other variables held at their mean values. All fixed effect coefficients can be interpreted as the estimated effect of one or more predictors, holding other variables at their mean values.

Random effects Previous research has reported differences in devoicing rates across both speakers and lexical items, and any spontaneous speech corpus is inherently unbalanced, such that certain words and speakers have much more data than others. These facts must be controlled for in the statistical model, or the effects of interest will be unduly influenced by a small group of speakers or words. For example,

[12] We included only *local speech rate* in interactions, and not *speaker speech rate*, to limit model complexity, and since *local speech rate* corresponds more closely to measures of speech rate used in previous work on HVD (e.g. Kondo 1997).

[13] Exploratory analysis suggested possible differences in the effect of pitch accents (H*), phrasal (H-) and boundary tone-associated H tones on devoicing rate, but due to the low number of tokens bearing a high tone in the dataset, these differences were collapsed into a single binary predictor of high tone presence.

high-frequency verbal particles (e.g. *desu*) are highly prone to devoicing (potentially skewing the estimate of overall devoicing rate), and occur disproportionately often in phrase-final position (potentially skewing the estimate of e.g. the *Break Index* effect). In a mixed-effects model, these issues are mitigated by the inclusion of random-effect terms. The model reported here includes by-speaker and by-word intercept terms, which directly account for differences between speakers and words in overall devoicing rate. We also included by-speaker random slope terms, which account for differences between speakers in effect size, for all fixed-effect terms of interest for our research questions: all terms involving *Break Index* or *Pause*, as well as main effects of variables involved in any interactions with *Break Index* (i.e. *local speech rate*, *lexical frequency*). These terms result in more accurate p-values and coefficient estimates for the fixed-effect terms of interest (Barr et al. 2013).[14] The model does not include random slopes corresponding to fixed-effect terms *not* of interest for our research questions (such as surrounding consonant manner), in order to limit model complexity. The coefficients and p-values for these terms are thus less reliable (Barr et al. 2013). Finally, correlation terms between random effects were excluded, to aid model convergence.

4.2 Model construction

A mixed-effects logistic regression was fit using the `glmer` function in the `lme4` package (Bates et al. 2015) package in R (R Core Team 2013). The inclusion of the full random effect structure described above led to non-convergent models. Analysis of the distribution of the data, guided by `glmer` warning messages, suggested that convergence issues were due to sparsity in certain parts of the data, reflecting collinearity between the presence of medium and long pauses and the type of *Break Index*. In particular, longer pauses are relatively rare at BI *1* or *2*, occurring mostly at BI *3* (94%, $n = 1756$).

In order to arrive at a convergent model, random-effect and fixed-effect terms flagged by `glmer` as unstable were iteratively removed, until a well-conditioned model was achieved. The fixed and random effect terms removed for the final model were two of those comparing medium versus long pauses: one estimating the difference between BI *1* and BI *2* and *3* (in the effect of medium vs. long pauses on devoicing rate), and the other estimating the difference between BI *2* and BI *3* (same). Hence, in the final model, the difference between medium and long pauses (in devoicing rate) was only estimated as a single effect across all word-final tokens (*Break Index* = *1*, *2*, and *3*), which will be important for interpreting the results.

5 Results

Here we report the results of the statistical model of devoicing rate. The model's estimates for the fixed-effect terms are shown in Table 4. We first discuss the results

[14] It would have also been preferable to include by-word random effect terms corresponding to the fixed effects of interest for our research questions, e.g. for *Break Index*. Adding these terms resulted in unstable models, presumably due to the high number of word types relative to the size of the dataset; we thus did not include them in the final model.

for control predictors, then turn to predictors relevant for our research questions: *Break Index*, *Pause*, and interactions between *Break Index* and *Mora deviation*, *lexical frequency* and *speech rate*.

To aid in interpretation of the model's results, we use partial effect plots (in addition to reporting model coefficients): these show the predicted effect of varying one or more predictors, while others are held constant, with predictions transformed into probability space (instead of log-odds). Model predictions in these plots were computed using the fixed effect coefficient estimates. Errorbars on model predictions correspond to two standard errors.

We do not discuss the model's random effect terms, which are shown in the Appendix.

5.1 Control predictors

The estimates for the effect of consonant manner are consistent with previous findings. Compared to the mean devoicing rate, a fricative preceding the token increases the likelihood of devoicing ($\hat{\beta} = 0.75,\ p < 0.001$), while a fricative following decreases the likelihood ($\hat{\beta} = -0.99,\ p < 0.001$). The effects of affricates are not as clear, with a preceding affricate slightly decreasing odds of devoicing relative to the mean rate, and a following affricate being not reliably different ($\hat{\beta} = -0.34$, $p = 0.033$; $\hat{\beta} = 0.16,\ p = 0.297$). There is also a significant interaction between preceding and following consonant manners. We do not discuss these terms in detail, but note that the negative coefficient for the interaction between terms for a preceding and following fricative ($\hat{\beta} = -0.49,\ p < 0.001$) suggests that vowels flanked by fricatives on both sides have particularly low devoicing rates, as expected (Tsuchida 1997).

The presence of a high tone strongly decreases the likelihood of devoicing ($\hat{\beta} = -4.35,\ p < 0.001$), again consistent with previous findings discussed in Sect. 2.3. The large effect of tone confirms that devoicing of vowels associated with a high tone is indeed highly dispreferred, but due to the small number of H-associated tokens in our data set, it was not possible to distinguish between pitch accents, phrasal high tones, and other high tones.

For the speech rate predictors, among main-effect terms, only the main effect of the speaker's mean speech rate reaches statistical significance, with a higher likelihood of devoicing for faster-talking speakers ($\hat{\beta} = 0.45,\ p = 0.006$). Neither local speech rate ($\hat{\beta} = -0.04,\ p = 0.714$) nor lexical frequency ($\hat{\beta} = 0.28,\ p = 0.402$) reached significance as main effects. However, terms in the interactions between *Break Index* and these two variables do reach significance. These interactions will be discussed below.

5.2 Break indices

The coefficients for this predictor address our first two research questions, comparing word-internal, word-final and phrase-final (AP or IP-final) vowels. Figure 2 shows the predicted probabilities for each value of *Break Index* with no pause following, and all other variables held constant at average values.

Table 4 Fixed effects for the statistical model: coefficient estimates, standard errors, z-values, and significances (assessed using a Wald test). Main-effect terms are shown first, followed by interaction terms

Fixed effects	β	se(β)	z	Pr(z)
(Intercept)	5.88	0.3	19.62	< 0.001
Break Index				
1, 2, 3 − None	−2.45	0.29	−8.41	< 0.001
2, 3 − 1	−1.97	0.24	−8.05	< 0.001
3 − 2	−1.21	0.32	−3.77	< 0.001
Mora deviation				
linear	0.48	0.27	1.75	0.081
nonlinear	−2.7	0.3	−8.85	< 0.001
Lexical frequency	0.28	0.33	0.84	0.402
Speech rate within utterance	−0.04	0.12	−0.37	0.714
Speech rate average by speaker	0.45	0.16	2.74	0.006
High tone *non-high − high*	−4.35	0.29	−15.02	< 0.001
Manner of previous phone				
fricative	0.75	0.14	5.39	< 0.001
affricate	−0.34	0.16	−2.13	0.033
Manner of following phone				
fricative	−0.99	0.1	−9.96	< 0.001
affricate	0.16	0.15	1.04	0.297
Pause : Break Index				
No Pause − Pause : *1, 2, 3 − None*	−0.01	0.37	−0.02	0.986
No Pause − Pause : *2, 3 − 1*	−2.69	0.81	−3.32	< 0.001
No Pause − Pause : *3 − 2*	−0.47	0.75	−0.63	0.529
Short Pause − Medium/Long Pause : *1, 2, 3 − None*	−0.39	0.55	−0.7	0.481
Short Pause − Medium/Long Pause : *2, 3 − 1*	−3.96	1.26	−3.14	0.002
Short Pause − Medium/Long Pause : *3 − 2*	0.08	1.12	0.07	0.944
Medium Pause − Long Pause : *1, 2, 3 − None*	−2.01	0.37	−5.43	< 0.001
Break Index : Lexical Frequency				
1, 2, 3 − None : *Frequency*	0.17	0.37	0.45	0.652
2, 3 − 1 : *Frequency*	1.25	0.32	3.93	< 0.001
3 − 2 : *Frequency*	1.49	0.52	2.85	0.004
Break Index : Speech Rate within utterance				
1, 2, 3 − None : *Speech Rate*	0.2	0.18	1.1	0.27
2, 3 − 1 : *Speech Rate*	0.51	0.24	2.17	0.03
3 − 2 : *Speech Rate*	−0.05	0.34	−0.15	0.884

First of all, the rate of devoicing for word-internal vowels is very high, essentially at ceiling (Intercept: $\hat{\beta} = 5.88$, predicted probability: 99.72%). Regarding the effect of word boundaries, the model confirms that, all else being equal, vowels followed by a word boundary (in any phrasal position) are significantly *less* likely to devoice

Table 4 (*Continued*)

Fixed effects	β	se(β)	z	Pr(z)
Break Index : Mora deviation				
1, 2, 3 − None : linear	−2.18	0.44	−4.94	< 0.001
2, 3 − 1 : linear	−0.11	0.61	−0.18	0.857
3 − 2 : linear	−1.97	0.94	−2.09	0.037
1, 2, 3 − None : nonlinear	−0.03	0.44	−0.07	0.946
2, 3 − 1 : nonlinear	−0.29	0.57	−0.51	0.611
3 − 2 : nonlinear	3.12	0.86	3.61	< 0.001
Previous phone manner : Following phone manner				
Preceding fricative: Following fricative	−0.49	0.13	−3.84	< 0.001
Preceding fricative: Following affricate	0.75	0.2	3.77	< 0.001
Preceding affricate: Following fricative	−0.08	0.16	−0.5	0.615
Preceding affricate: Following affricate	−0.84	0.25	−3.4	< 0.001

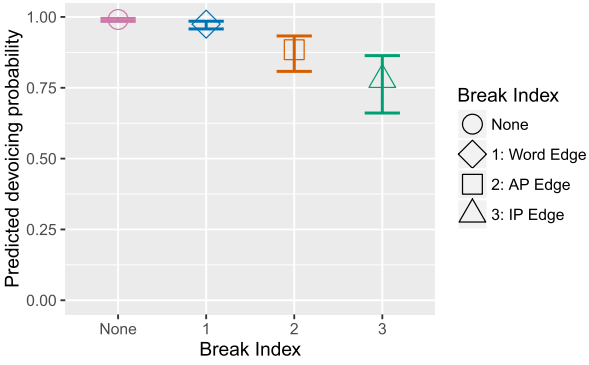


Fig. 2 Predicted probability of devoicing for a high vowel that is (**a**) word-internal, (**b**) at a word boundary, but phrase-internal, (**c**) at an accentual phrase (AP) boundary, (**d**) at an intonation phrase boundary; in all cases, the prediction assumes no following pause, and others variables held constant at mean values. *Shapes* represent the predicted probabilities, and *bars* show the 95% confidence intervals

than vowels that are within the same word as their following consonant (*Break Index* 1/2/3 − None: $\hat{\beta} = -2.45$, $p < 0.001$). This finding, on the effect of word boundaries for single devoicing environments, is consistent with the results of Varden (1998), who found that in *consecutive* devoicing environments, a word-internal vowel was more likely to be devoiced than a word-final one.

Among word-final vowels, the model finds a reliable difference between devoicing rates for phrase-internal vowels compared to vowels at the edge of an accentual phrase or intonation phrase (*Break Index* {2, 3} − 1: $\hat{\beta} = -1.97$, $p < 0.001$). Among vowels at prosodic phrase edges, vowels at the edge of an intonation phrase are *less* likely to devoice than vowels at the edge of an accentual phrase (*Break Index* 3 − 2: $\hat{\beta} = -1.21$, $p < 0.001$).

In sum, the main effect of the *Break Index* predictor confirms that, when no pause follows and other predictors are controlled, the 'higher' the boundary (greater *Break Index* value: None < word boundary < AP < IP), the less likely devoicing becomes.

5.3 Pause

The effect of *Pause* was included in the model only as an interaction with *Break Index*, since there are no word-internal tokens that are followed by a pause. When considering all word-final tokens jointly, the model does not find a significant difference in devoicing rate depending on the presence/absence of a pause ($\hat{\beta} = -0.01$, $p = 0.986$), or on the difference between a short/longer pause ($\hat{\beta} = -0.39$, $p = 0.481$). There is a significant difference between tokens followed by medium and long pauses, with long pauses associated with higher rates of devoicing ($\hat{\beta} = -2.01$, $p < 0.001$). Since the model only compares medium and long pauses across all values of *Break Index* jointly (see Sect. 4.2), it is not possible to say whether this effect is similar at all types of boundaries, but examination of the empirical data for each *Break Index* value suggests that it is driven by tokens at IP boundaries (*Break Index* = 3, which contains the most data for medium–long pauses).

The model also compares the differences in the effect of *Pause* among vowels in different prosodic positions. The presence of a pause has a *smaller* effect on the probability of devoicing following phrase-internal word-final vowels, relative to following phrase-final vowels ($\hat{\beta} = -2.69$, $p < 0.001$). There is also a difference in the effect of short pauses (<85 ms) and longer pauses (>85 ms): tokens followed by short pauses have a higher devoicing rate than tokens followed by long pauses, for *Break Index* 1 (phrase-internal word boundary); but if the token is at a phrase boundary (*Break Index* 2 or 3) then it is *longer* pauses that have higher devoicing rates than short pauses ($\hat{\beta} = -3.96$, $p = 0.002$).

The larger pattern expressed by these coefficients can be seen in the prediction plots in Fig. 3. The effect of a pause is strikingly different between the phrase-internal and phrase-final vowels. At phrase-internal vowels (left panel), an increase in pause duration has a consistently negative effect on devoicing rate, at least for null/short/medium pauses.[15] For phrase-final vowels, an increase in pause duration is associated with an *increase* in the probability of devoicing.

In sum, the relationship of pause length to devoicing rate looks qualitatively different in different prosodic positions. Pauses have an inhibitory effect on devoicing for phrase-internal vowels, but a facilitatory effect for phrase-final vowels.

5.4 Mora deviation

Mora deviation strongly affects the likelihood of devoicing. As shown in Fig. 4, evoicing is progressively less likely for longer Moras, and this holds across prosodic positions (when other variables are held constant). The regression terms are difficult to interpret directly, but their significance can be evaluated jointly: a likelihood ratio test (comparing the full model with one where all terms involving *Mora deviation* are excluded) shows that information about *Mora deviation* contributes significantly to explaining the variation in the data ($\chi^2(8) = 2325$, $p < 0.0001$). To visualize the predicted effect of *Mora deviation*, the model-predicted probabilities of devoicing as

[15] The high standard errors of the <463 ms and >463 ms points, presumably due to the small number of phrase-internal tokens followed by appreciable pauses, prevent us from concluding there is an effect of increasing pause duration from medium to long pauses, in either direction.

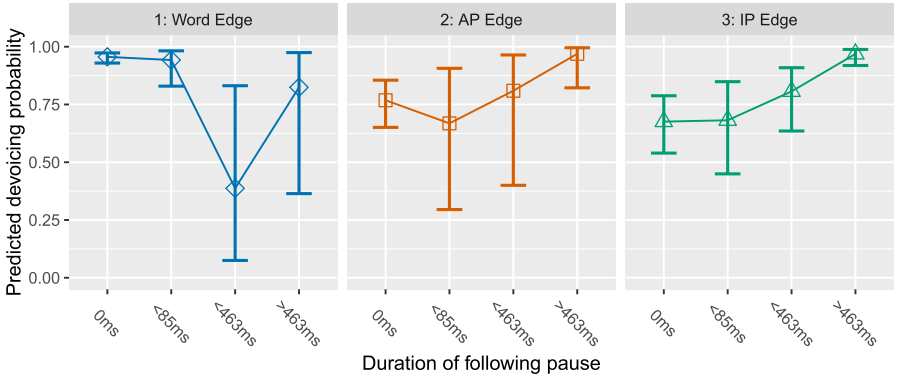
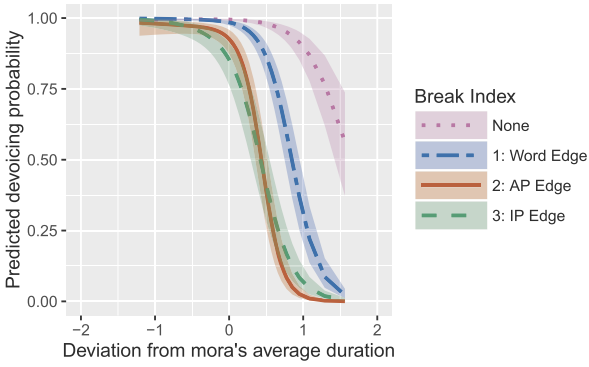


Fig. 3 Predicted probability of devoicing for a high vowel at a word boundary as duration of the following pause increases, by prosodic position (*Break Index*), with other predictors held constant at mean values. *Shapes* represent the estimated probabilities, and *bars* show the 95% confidence intervals

Fig. 4 Predicted probability of devoicing for a high vowel by the degree of Mora deviation, by prosodic position (*Break Index*), with other predictors held constant at mean values. *Lines* represent the estimated probability, and *shading* shows 95% confidence intervals



a function of *Mora deviation* for each prosodic position, with other variables held constant, are shown in Fig. 4.

For word-internal vowels, devoicing is predicted to be at ceiling until the duration of the Mora is around the mean value (represented by 0 on the x axis in Fig. 4), and from there ranges to about 50% at its lowest value. This agrees with previous work on consecutive devoicing environments which found that a Mora is significantly shorter when it is produced with a devoiced vowel (Kondo 2005). For word-final vowels (*Break Index* $= 1, 2, 3$), the probability of devoicing ranges from almost 100% to almost 0% across the range of observed *Mora deviation* values, as shown in Fig. 4.

Some of the interaction terms with *Break Index* were statistically significant. The slope of the estimated linear effect was significantly different between word-internal and word-final position, with devoicing probability being less affected by *Mora deviation* in word-final position ($\beta = -2.18$, $p < 0.001$). In addition, the effect of Mora deviation differs between IP-final vowels and AP-final vowels ($\beta = -1.97$, $p = 0.037$), such that Mora deviation has a stronger effect on AP-final vowels (steeper slope in Fig. 4). However, none of the interaction terms change the qualita-

tive shape of the effect of *Mora deviation*, which is similarly negative across prosodic positions.

In sum, the duration of the Mora has a significant negative correlation with probability of devoicing. Interpreting higher *Mora deviation* as a proxy for more final lengthening and less gestural overlap, this pattern suggests that devoicing is less likely when there is more final lengthening, and more likely when there is more gestural overlap. The effect is qualitatively similar across all prosodic positions, in contrast with the effect of *Pause* described above.

5.5 Lexical frequency

The main effect of *lexical frequency* does not reach significance ($\hat{\beta} = 0.28$, $p = 0.402$), suggesting that word frequency does not play an important role in determining devoicing rates, averaging across prosodic positions. This is in contrast to an empirical plot of word frequency by devoicing rate of our data, which suggested a slightly negative effect, similar to the negative effect found by Maekawa and Kikuchi (2005) for the same corpus (although their analysis was for high vowels preceded by a voiceless consonant, but with any following environment). The fact that the model does not find a significant effect, in contrast to plots of the empirical data, suggests that the trend is primarily an artefact of other factors (variables which may be confounded with frequency, or lexical idiosyncrasies).

However, there are significant terms for the interaction of *lexical frequency* with *Break Index*, suggesting that word frequency does affect devoicing rate for some prosodic positions. Figure 5 shows the predicted frequency effect for each prosodic position, illustrating the pattern captured by these interaction terms. For word-internal vowels, the devoicing rate is at ceiling. Among word-final tokens, the frequency effect is slightly negative at phrase-internal word boundaries, versus slightly–greatly positive at phrase-final word boundaries ($\hat{\beta} = 1.25$, $p < 0.001$): thus, we again see a qualitative difference among word-final vowels depending on whether they are phrase-internal or phrase-final. The frequency effect is significantly larger (= more positive) at IP boundaries than at AP boundaries ($\hat{\beta} = 1.49$, $p = 0.004$). Both of these terms point to the broader pattern in Fig. 5: the effect of frequency is essentially restricted to IP-final vowels, where there is a strong *positive* effect: devoicing is more frequent for more frequent words.

A frequency effect in phrase-final position is expected under our account of phrase-final devoicing as a phonetically-motivated reduction process, discussed further below. However, we do not have a good explanation for why the frequency effect is essentially restricted to IP-final vowels. This may be due in part to high-frequency words which devoice near-categorically and occur disproportionately in IP-final position (e.g. *desu*, *masu*), though the by-word random intercept term should mitigate such effects of individual words.

5.6 Speech rate

Two measures of speech rate were included in the model, average talker speech rate and local deviation from the talker's average speech rate. The *average speech rate*

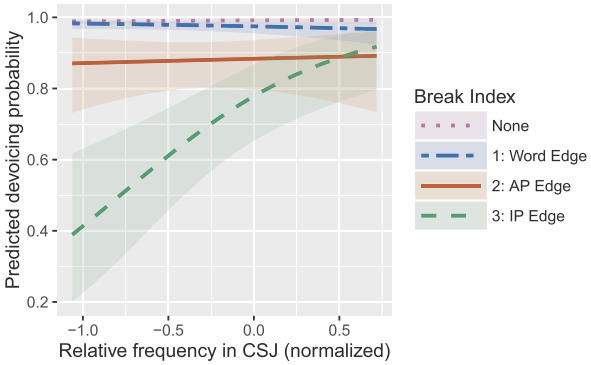


Fig. 5 Predicted probability of devoicing for a high vowel by relative lexical frequency (log-transformed and normalized), by prosodic position (*Break Index*), with other predictors held constant at mean values. *Lines* represent the estimated probability, and *shading* shows 95% confidence intervals

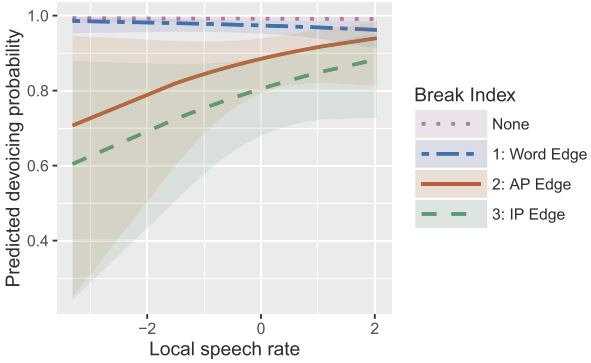


Fig. 6 Predicted probability of devoicing for a high vowel by local speech rate (phones/second, normalized), by prosodic position (*Break Index*), with other predictors held constant at mean values. *Lines* represent the estimated probability, and *shading* shows 95% confidence intervals

significantly increases the probability of devoicing ($\hat{\beta} = 0.45$, $p = 0.006$), suggesting that faster talkers devoice vowels more readily. The main effect of *local speech rate* has a fairly small coefficient estimate, and does not reach significance ($\hat{\beta} = -0.04$, $p = 0.714$), suggesting that how fast a speaker is talking relative to their norm has little effect on devoicing rate, averaging across prosodic positions.

However, as with the lexical frequency effect, there is a significant interaction of *local speech rate* with *Break Index*, suggesting that the speech rate effect differs qualitatively by prosodic position. Figure 6 shows the predicted rate effect for each prosodic position, illustrating the pattern captured by these interaction terms.

For word-internal vowels, the devoicing rate is at ceiling regardless of speech rate. Among word-final vowels, the speech rate effect is significantly greater for phrase-final vowels than for phrase-internal vowels ($\hat{\beta} = 0.51$, $p = 0.03$). This results in the pattern apparent in Fig. 6: phrase-final vowels tend to devoice more in faster speech, while phrase-internal vowels are not greatly affected by speech rate, if anything showing a tendency to devoice *less* in faster speech. Thus, we again see a qualitative split by prosodic position, depending on whether the vowel is internal or at the edge of a prosodic phrase.

6 Discussion

The results of the mixed-effects regression show that in the single devoicing environment, the devoicing rate for high vowels surrounded by voiceless consonants is affected by a number of factors—notably prosodic position, which both directly affects devoicing rate and modulates other variables, in a way which suggests a qualitative split between phrase-internal and phrase-final environments. These results bear on the three questions raised at the outset about Japanese vowel devoicing: the role of boundary phenomena, the relationship and characterization of the two environments in which devoicing applies, and the sources of variability. We first discuss our findings with respect to our research questions, which focused on boundary phenomena: how do word and prosodic boundaries affect devoicing rate (including modulating other factors), and what is the role of physical pauses? We then turn to the broader issues of how to characterize HVD, and sources of variability in its application.

6.1 Boundary phenomena

6.1.1 Word boundaries

Our first research question was how word boundaries affected the rate of high vowel devoicing and modulated the effects of other factors. The results confirmed a difference in devoicing rate between word-internal and word-final vowels, with a significantly lower probability of devoicing expected for vowels followed by word boundaries. This finding may seem unsurprising, given that external sandhi processes cross-linguistically usually apply more consistently within words than across words, but to our knowledge this study is the first to demonstrate and quantify this effect for Japanese high vowel devoicing. Furthermore, the word boundary effect exists after controlling for confounding factors that could be correlated with word boundaries, such as domain-final lengthening and high tones, by including appropriate terms in the statistical model. This seemingly intuitive and simple inhibitory effect of word boundaries points to a new question for HVD, and external sandhi processes more generally: why should word boundaries per se have an inhibitory effect on process application, when other factors are held constant?

Another interesting result is that the estimated baseline rate for word-internal vowel devoicing is so high that it is not appreciably lowered by most inhibiting factors.[16] For example, Figs. 5 and 6 show that when other factors are held constant, the probability of devoicing a word-internal vowel stays more or less at ceiling for any value of speech rate or lexical frequency (99.72% at mean values, 98.32% with frequency and speech rate at −2.5 standard deviations away from their mean value). Hence, these subtle effects are predicted to be masked for word-internal vowels. Even the relatively large effect size of *Mora deviation*, illustrated in Fig. 4, is not predicted to completely block word-internal vowel devoicing at its most extreme value, with the lowest predicted probability reaching only about 50%. This is in striking contrast

[16]The one exception is the presence of a high tone, confirming the intuition that this blocks devoicing for some speakers (Han 1962; Lovins 1976; Hirayama 2009; Nielsen 2015).

to word-final vowels, where devoicing is predicted to be almost totally absent for the most lengthened Moras. These results confirm textbook statements (e.g. NHK 1991; Vance 2008) and native speaker intuitions that high vowel devoicing is obligatory, but with the qualification that this holds for word-internal devoicing environments.

On the other hand, for word-final vowels the devoicing rate is estimated to be reliably slightly lower (Fig. 2). This makes the devoicing rate at word boundaries more susceptible to the influence of even relatively small effects like *local speech rate* and *lexical frequency*, as well as large effects like *Mora deviation*. Importantly, this difference in susceptibility is not due to the effects of word frequency, Mora duration, or speech rate actually differing between word-final and word-internal vowels—the relevant model terms are not significant (frequency, speech rate) or do not change the effect's direction (*Mora deviation*). Rather it is simply due to the much higher baseline devoicing rate for word-internal vowels.

On the whole, the results show that the presence of a word boundary is correlated with a decrease in devoicing rate, all else being equal. The effects of *Mora deviation*, *local speech rate*, and *lexical frequency* do not differ qualitatively if we compare their effects on word-internal versus word-final vowels.

6.1.2 Prosodic phrase boundaries

Our second research question was how prosodic phrase boundaries affected the rate of high vowel devoicing and modulated the effects of other factors. In examining the effect of prosodic phrase boundaries, we discuss both the presence/absence of a phrase boundary (either accentual or intonation phrase) at a word edge, and the difference between AP-final and IP-final tokens.

The statistical analysis shows that word-final devoicing rates differ significantly depending on whether a phrase boundary follows. In the absence of a pause, the presence of an accentual or intonation phrase boundary significantly decreases the probability of devoicing compared to a word-final vowel that is not followed by a phrase boundary. Among vowels that are followed by a phrase boundary, the stronger intonation phrase boundary is associated with significantly less devoicing than a weaker accentual phrase boundary. The overall pattern (Fig. 2) is that as Break Index increases, devoicing rate decreases. What is driving this effect, and how does it fit in with current accounts of HVD?

Consider first the C̥_C̥ environment. Taking the view of HVD as a reductive process, the decrease in devoicing at stronger boundaries fits in with the cross-linguistic tendency to see less reduction at stronger prosodic boundaries (Wightman et al. 1992; Keating 2006). Since phrase boundaries are associated with segmental lengthening, this would also fit nicely with a gestural overlap account of HVD: the phrase-final vowel is lengthened, so the gestures of the surrounding consonants are less likely to overwhelm the vowel's voicing gesture. However, our model included *Mora deviation* as a separate factor, which accounts for this kind of temporal overlap. Indeed, our model estimates that as the Mora becomes longer (relative to its expected duration), the rate of devoicing declines sharply, so gestural overlap may play a role, but the effect of prosodic boundaries cannot be simply attributed to the temporal alignment of gestures. For example, if we consider two identical word-final vowels, both

surrounded by the same consonants and *of the same duration*, the vowel followed by an accentual phrase boundary is more likely to be devoiced than the one followed by an intonation phrase boundary. A gestural overlap analysis of devoicing would have to be augmented to account for these effects. One possibility is that higher level prosodic boundaries are associated with some increase in magnitude (rather than timing) of the voicing gesture for the vowel, which leads to devoicing rates even lower than would be expected from simply articulating the vowel more slowly. In sum, our results show that prosodic boundaries have an effect on HVD above and beyond the potential confound of final lengthening, but overall it makes sense that a stronger boundary would disrupt the interaction between a word-final vowel and following voiceless consonant.

If we now consider the C̥_# environment, we run into a different puzzle. It has been suggested that "#" should be interpreted as the end of an intonation phrase or an utterance (Kondo 1997; Hirayama 2009; Fujimoto 2015). If we take IP boundary as the definition of "#" in the C̥_# environment, then a feature-based analysis such as Rule 1 would not immediately explain the difference in variability between C̥_C and C̥_# devoicing environments—if the process is the same in both environments, it should apply equally often in both cases. In fact, HVD at intonation phrase boundaries was much less consistent overall, with the model estimating rates between 56% and 93% depending on the manner of surrounding consonants, all other variables held constant. Again, these differences between phrase positions are found even after controlling for presence of pause and Mora deviation, so the inhibitory effect is above and beyond these correlates of prosodic boundaries. On the other hand, defining "#" as a physical pause is also clearly not right, since the effect of a pause is inhibitory for phrase-internal vowels. The environment in which we find categorical HVD, other than word-internally, is defined by the *joint* effect of a phrase boundary and a longer pause, so both factors must somehow be incorporated into the definition of C̥_#.

The model also shows that prosodic phrase boundaries strongly modulate the effects of other variables. Most strikingly, the effects of *pause duration*, *lexical frequency*, and *local speech rate* are significantly different for phrase-internal vowels and phrase-final vowels, and that the effects of these variables is qualitatively different depending on prosodic position. We return to the pause effect below (Sect. 6.1.3), and here discuss the frequency and speech rate effects.

Overall, *lexical frequency* has little effect on devoicing rate. However, phrase-internal vowels and phrase-final vowels show a qualitatively different frequency effect. As Fig. 5 shows,this effect is driven mostly by IP-final vowels, for which there is a strong positive frequency effect. This is the direction predicted for a reductive, phonetically-motivated process (e.g. Jurafsky et al. 2001; Pluymaekers et al. 2005), and consistent with a gestural overlap account of HVD.

A similar pattern emerges for the effect of *local speech rate* on HVD. For phrase-internal vowels at a word-boundary the effect is slightly negative, meaning that devoicing becomes *less* probable as speech rate increases. This is the opposite of what is expected for a reductive process, since reductions typically become more common at faster speech rates (e.g. Fosler-Lussier and Morgan 1999). Phrase-final vowels, on the other hand, show the expected pattern (for a reductive process) of higher likelihood of devoicing at faster speech rates. The positive speech rate effect is consistent

with previous findings, such as the study in Martin et al. (2014), although other studies have failed to find speech rate effects in the single devoicing environment (e.g. Kondo 1997).

It is striking that in both of these cases—as well as for the case of *Pause*, discussed below—there is a clear qualitative split between phrase-internal (*Break Index* None and 1) and phrase-final (*Break Index* 2 and 3) vowels. It would have been possible for these differences in effects to be only differences in magnitude, but still going in the same direction, as is the case for the *Mora deviation* effect. It could also have been the case that presence/absence of word boundaries modulated the frequency and rate effects, rather than phrase boundaries. The fact that the interaction terms involving phrase-internal/phrase-final differences in Table 4 are consistently significant (for *Pause*, *frequency*, and *local speech rate*) suggests that something about higher-level prosodic groupings must be invoked to explain this pattern of variability.

6.1.3 Physical pause

Part of the puzzle of high vowel devoicing we seek to address in this paper was the effect of a "pause," which ostensibly triggers devoicing, is associated with variable devoicing, and blocks devoicing. Our results on the effect of a physical pause, our third research question, show that these claims are all valid, but depend on context.

First of all, for word-final vowels that are not at any larger phrase boundary (Fig. 3, left panel), pauses inhibit devoicing: devoicing is less probable if there is a pause following, of any duration. This effect is consistent with Vance's (1992) observation that in syllable-by-syllable pronunciations of words with potential devoicing sites, the pauses between the syllables block devoicing. It also supports the intuition expressed by some authors that C̥_# devoicing is not exactly conditioned by the pause itself, but by finality in an intonation phrase or utterance (Kondo 1997; Hirayama 2009; Fujimoto 2015). The exact duration of the pause also affects devoicing rate, in a similar way: devoicing is more likely before a short pause than before a medium pause. Thus, pauses gradiently and negatively affect the likelihood of devoicing for phrase-internal word-final vowels.

For phrase-final vowels, pauses have the opposite effect (Fig. 3, middle–right panels): vowel devoicing becomes *more* probable before a pause, and more probable as pause duration increases. Thus, pauses gradiently and *positively* affect the likelihood of devoicing. In fact, with all other predictors held constant, devoicing is predicted to reach almost 100% probability for vowels which are followed by a long pause (>463 ms), but only if they are at an accentual or intonation phrase boundary.

The differences in the effect of pause once again mirrors the split seen for lexical frequency and speech rate effects: phrase-internal and phrase-final vowels are affected differently by these variables.

6.1.4 The role of boundary phenomena

Our findings on how boundary phenomena condition HVD is relevant for the more general issue of how boundary phenomena affect variable (phonological) processes. We found that prosodic boundaries and physical pauses have distinct and interacting

effects: notably, the direction of one effect (*Pause*) flips depending on the value of the other (*Break Type*). Thus, the correct characterization of the 'pre-pausal' environment is more complicated than just one boundary phenomenon (e.g. 'utterance boundary') or another (e.g. 'long pause'). An interesting question for future work is whether this empirical pattern holds for other cases where variable processes apply in 'final' or 'pre-pausal' position, especially for vowel devoicing processes, where this description is common (Gordon 1998; Barnes 2006). How to capture the observed patterns in a formal analysis is a non-trivial question, which depends on how one assumes HVD is characterized. We return to this issue below.

6.2 High vowel devoicing as two overlapping processes

This study has shown that in a large corpus of spontaneous speech, it is possible to tease apart the effect of several (often correlated) variables on HVD. The results of our analysis suggest a complex relationship between HVD variability and word and phrase boundaries, pauses, lexical frequency, and speech rate measures.

It was confirmed that, in line with native speaker intuitions, HVD is "almost compulsory" when the following voiceless consonant is within the same word or across a word boundary (phrase-internal, no intervening pause). But HVD is also nearly categorical in basically the opposite context, when the vowel is followed by a prosodic phrase boundary *and* a relatively long pause. The results also confirmed the seemingly contradictory claims that pauses trigger devoicing (cf. the traditional description of Ç_#: Han 1962; McCawley 1968) and block devoicing (Vance 1992; Kondo 1997): in fact, pauses have opposite effects on devoicing depending on whether the vowel is at the edge of a prosodic phrase or not. Prosodic position also modulates the effect of lexical frequency and speech rate in a similar way. We now discuss possible interpretations of this complex pattern of variability within existing analyses of HVD, and follow with our own proposal. We suggest that by breaking down the source of devoiced vowels into two separate processes, we can describe two sub-patterns in the distribution of devoiced high vowels, which can help explain the overall pattern of variability.

The traditional description of Japanese high vowel devoicing exemplified in Rule 1 implies that the alternation between voiced and voiceless vowels is the same qualitative process, independent of which environment is the trigger of the change.

This assumption is difficult to reconcile with the patterns of high vowel devoicing variability observed in this study. Even allowing a rule such as Rule 1 to apply variably would not go very far toward explaining why devoicing is categorical or variable in a particular prosodic context. Furthermore, our results show that the position of the vowel within the prosodic phrase affects not only the amount of variability in devoicing, but also the manner in which pauses, speech rate, and lexical frequency influence variability. In our view, this pattern suggests that two different processes underlie the alternations between voiced and voiceless vowels in Japanese.

The idea that devoiced high vowels may have different underlying sources has already been suggested in the literature, but with a different motivation. Tsuchida (1997), focusing on variability in consecutive devoicing environments, proposed that devoiced vowels in Japanese have two different underlying mechanisms depending

on context. In the C̥_C̥ environment, it is argued, devoicing is categorical and due to a phonological rule. This classification is motivated by its categorical rate of application in the C̥_C̥ single devoicing environment. The variability of devoicing in consecutive devoicing environments, and for vowels flanked by fricatives, suggests a phonetically-driven process in those cases. This dual-mechanism account is also defended by Varden (1998) and Nielsen (2015). Under these accounts, two processes are needed to account for variable and categorical application: variation within a consistent phonological context implies a phonetic process, while a phonological process should be categorical within a given context.

While the results of the present study agree with the claim that devoicing is near categorical within a word, the pattern of variability becomes more complicated as we investigate what happens to vowels at word boundaries and in different positions in a prosodic phrase. Surprisingly, we also see near categorical devoicing when there is the most disjuncture between a vowel and following consonant, namely at a prosodic phrase boundary with a long pause. Intuitively, if we think of C̥_C̥ devoicing as applying categorically word-internally, and C̥_# devoicing as applying categorically at a very strong boundary, all cases where there is variability lie in between these two extremes, where the two environments *overlap*. The picture of devoicing that emerges is not easily interpreted within a dichotomy of categorical/phonological versus continuous/phonetic.

However, we agree with the intuition that two different processes underlie devoiced vowels in Japanese. Recall that cross-linguistically, there are two attested parameters that define the environments for vowel devoicing: the segmental context, and position with a (prosodic) domain. We propose that Japanese has two separate devoicing processes that differ precisely along these parameters, corresponding intuitively to C̥_C̥ and C̥_#. The C̥_C̥ process, which we call *interconsonantal devoicing*, is sensitive to the voicelessness of the following segment, but not to finality within a domain. This process parallels devoicing in Turkish (Jannedy 1995) and Montréal French (Cedergren and Simoneau 1985) in which vowels are only devoiced between voiceless consonants.

On the other hand, what has been described as C̥_# devoicing is a separate process, which we call *phrase-final devoicing*, which *does not* make reference to the following segment's properties, but rather the position of the vowel within a larger domain—tentatively, finality in an accentual phrase (and thus also in intonation phrases or utterances). This process parallels devoicing in languages like Greek (Dauer 1980; Kaimaki 2015), in which vowels are only devoiced phrase or utterance-finally. Note that an important caveat to our characterization of phrase-final devoicing is that our data only contains vowels followed by voiceless consonants. We assume in the following discussion that the "phrase-final" characterization is correct, but come back to this caveat in Sect. 6.4.

6.2.1 Overlapping environments

Our two-process proposal for HVD connects to the broader issue of how to analyze (variable) processes that apply in overlapping environments. We argued for two overlapping processes based on their distinct phonetic sources, cross-linguistic typology

(where both kinds of devoicing processes are attested), and qualitatively different effects of non-grammatical factors (frequency, speech rate) by prosodic position. If our two-process proposal is correct, a formal analysis would be fairly straightforward: intervocalic devoicing and phrase-final devoicing could each be analyzed similarly to other cases of intervocalic devoicing or phrase-final devoicing (respectively) (e.g. Tsuchida 2001). Devoiced vowels between voiceless consonants in Japanese would then result from two different processes, analogously to other such cases, like word-final underlyingly-voiced obstruents in languages with both final devoicing and regressive voicing assimilation for obstruents (e.g. Polish, Catalan: Lombardi 1991).

In contrast, a formal analysis of our HVD data as a single process would need to account for the complex effects of boundary phenomena, in particular the fact that the effect of one variable (pause duration) on devoicing rate *reverses direction* depending on the value of another variable (Break type). In a standard constraint-based analysis of a variable process (e.g. Maximum Entropy harmonic grammar: Hayes and Wilson 2008; Coetzee and Pater 2011), the effects of these two variables would be captured by two (sets of) constraints, each of which always assumes the same directionality of an effect. For example, one constraint could penalize devoicing before shorter pauses (accounting for the pattern in phrase-final position), but the opposite effect of a pause in phrase-internal tokens would be unaccounted for. In order for the effect of one variable to 'flip' depending on the value of another variable, additional mechanisms would need to be invoked, such as weighted constraint conjunction (e.g. Shih 2016; Hayes et al. 2012). While such an analysis is certainly possible, it would leave unexplained *why* the effect of pause differs by prosodic position, which falls out naturally from the two-process proposal.

6.3 Sources of variability

Our account of HVD in terms of two processes, which differ in sensitivity to following context versus prosodic position, helps elucidate the overall pattern of variability, shown in Fig. 7, and the differing effects of pause, lexical frequency, and speech rate for phrase-internal and phrase-final vowels. The pattern of variability can be further explained by reference to two aspects of phonetic implementation and processing: gestural overlap, and the locality of production planning.

Across prosodic positions, *Mora deviation* is an important predictor of devoicing rate: devoicing becomes much less likely as Mora duration increases (as found by Den and Koiso 2015 for utterance-final vowels), even for word-internal vowels. We suggest that the strong Mora duration effect reflects gestural overlap as a major source of variability in HVD, in addition to the effect of final lengthening on gestural overlap (Byrd and Saltzman 2003). In all prosodic positions, shorter Mora duration will correlate with more gestural overlap with adjacent voiceless segment(s) making devoicing more likely. For word or phrase-final vowels, final lengthening will correlate with less gestural overlap and a longer vowel, either of which make devoicing less likely (Gordon 1998; Barnes 2006).

Turning to phrase-final devoicing in particular: the profile of variability we observe for phrase-final tokens is mostly consistent with phrase-final devoicing being

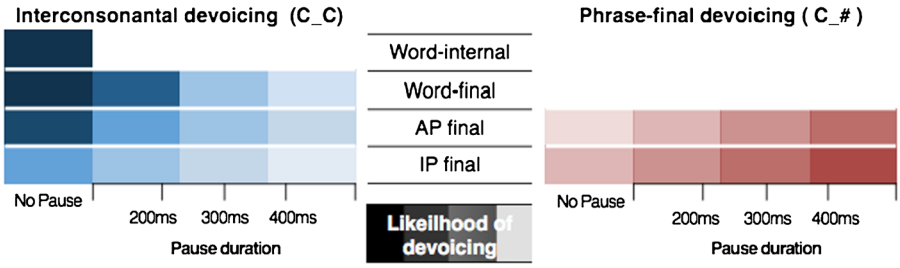


Fig. 7 Schema of the pattern of variability for our two proposed processes of devoicing in Japanese. *Darker colors* represent higher likelihood of a devoiced vowel. *Each row* represents one of the prosodic conditions investigated in this study, and *each column* represents an interval of pause durations, with the first column being the case where there is no pause at all

a phonetically-motivated process (e.g. a postlexical, or 'late' phonological process; Coetzee and Pater 2011), in particular reduction due to gestural overlap and aerodynamic factors. Phrase-finally, two kinds of phonetic factors promote devoicing: gestural overlap with the preceding voiceless segment, and decreased subglottal pressure over the course of an utterance (Gordon 1998:100). The positive effects of lexical frequency and speech rate for phrase-final vowels are consistent with the first of these sources: there should be more gestural overlap for higher-frequency words, or in faster speech, making the duration and magnitude of the voicing gesture shorter, both of which make it less likely that the aerodynamic conditions for voicing are met. The effect of pause duration makes sense assuming that a longer pause correlates with decreased subglottal pressure; there is then less likely to be sufficient pressure across the glottis to initiate voicing. Thus, the directions of the frequency, speech rate, and pause effects are consistent with a phonetically-motivated devoicing process which applies phrase-finally.

How to explain variability in application of interconsonantal devoicing, on the other hand, is a more challenging question. Interconsonantal devoicing does not show significant effects of lexical frequency or speech rate (when *Break Index* = None, 1), and is generally very consistent as long as no pause follows. However, the presence of a word boundary and the strength of the prosodic juncture between the vowel and the following consonant have gradient inhibitory effects on devoicing rates: devoicing is progressively less likely for higher *Break Index* values (Fig. 2), for longer pauses (for *Break Index* = None, 1), and for higher *Mora deviation*, which we assume in part reflects the final lengthening expected for stronger prosodic boundaries. These patterns cannot be explained solely by reference to gestural overlap, which would lead us to expect the same patterns as for phrase-final devoicing: positive frequency and speech rate effects, and positive effects of pause and boundary strength. Thus, another explanation is needed for variability in interconsonantal devoicing: why is there variability *at all*, why do higher break indices condition less devoicing, and why do pauses condition less devoicing for phrase-internal word-final vowels?

We suggest that the locality of production planning can help explain these aspects of variability in the dataset, while allowing us to maintain a simple description of the environment for both processes. We offer a brief overview of the locality of production planning hypothesis before discussing how it may explain some of the

patterns of variability found in this study, complementing those patterns which are well-explained by reference to gestural overlap.

The locality of production planning (LPP) hypothesis is a proposal developed in Wagner (2012), Tanner et al. (2017) which relates prosodic boundaries to phonological variability. It is proposed that the scope of speech planning constrains the application of phonological processes across word boundaries.

This hypothesis is based on findings in the psycholinguistics literature on speech production that at the phonological level, speech is planned hierarchically and incrementally (Sternberg et al. 1978; Ferreira 1988, 1991; Dell and O'Seaghdha 1992; Levelt et al. 1999). Higher-level information, such as number of words in an utterance, is planned before all lower-level information, such as number of syllables or segmental content, is retrieved or encoded. For example, Sternberg et al. (1978) found an asymmetry in the type of information that induced delays in initiating an utterance: the overall number of words in the utterance always increased the delay, but the number of syllables in a word only had an effect for the first word in the utterance. Wheeldon and Lahiri (1997, 2002) similarly found the overall number of words in an utterance affected latency, but the number of syllables only had an effect when considering the first word (i.e. the number of syllables in the second word did not have an effect). They furthermore showed that prosodic organization plays a significant role in production planning, with production latencies crucially depending on the number of *prosodic* rather than lexical words. In Levelt's influential model of speech production (Levelt et al. 1999), segmental information is retrieved only incrementally, in word-sized planning units. Although there is an ongoing debate in the literature as to the size of the window for phonological encoding (see Wheeldon 2013, for an overview), it is agreed that in some cases, especially in spontaneous speech, the window is fairly limited, possibly as small as a single prosodic word. Hence, it must be the case that segments early in an utterance are planned in the absence of detailed information about later segments. The LPP is premised on the idea that even the segmental details of the *very next segment* may not be always be available. This situation is predicted to be more likely if the following segment is in a separate planning unit, and should be made even more likely by any other factors which delay the retrieval and encoding of phonological material.

How does this help explain the variability of interconsonantal devoicing? The LPP hypothesis predicts that any alternation that is dependent on information from a following word (i.e. a separate planning unit) should be subject to variability. Applications of interconsonantal devoicing across a word boundary fall under this category: a word-final high vowel may have to be planned without the information that the upcoming word begins with a voiceless consonant, and hence there would be no motivation to plan a devoiced vowel. This would not be the case for word-internal applications of HVD, where the following consonant is always in the same planning unit and therefore always known at the moment of planning the vowel. Hence, the LPP explains the consistent difference in variability between word-internal and word-final vowels.

Our results also showed that among word-final vowels, there is a gradient decrease in the probability of devoicing beyond what could be attributed to temporal overlap of voicing gestures. Wagner (2012) suggests that the strength of the boundary between two words is correlated with the likelihood of their being planned within the

same window. Hence, under the LPP, it is predicted that stronger prosodic boundaries are associated with less availability of the segment following the boundary. For interconsonantal devoicing, this would lead to a decreased probability of application for higher level prosodic boundaries.

The inhibitory effect of pauses (for word-final, phrase-internal vowels) can be explained along similar lines. Pauses are associated with complexity of the upcoming phrase being planned (Sternberg et al. 1978; Ferreira 1991; Wheeldon and Lahiri 2002), so they may also track availability of the segment following the pause. Again, decreased availability of the following segment would lead to decreased application of interconsonantal devoicing.

In sum, the LPP offers an explanatory mechanism as to why a seemingly planned, phonological process may show variability in spontaneous speech. When an alternation depends on information in an upcoming word, many factors may interfere with online phonological encoding during the course of speech planning, leading to an "opaque" output from the perspective of the ultimate pronunciation (e.g. a voiced high vowel between two voiceless consonants).

6.3.1 Production planning and formal analysis

We have invoked the LPP to describe variability observed in our empirical data, without providing a formal analysis. But, related to the broader theoretical issue of how to account for processes which show near-categorical behavior in some environments and variability in others, it is worth discussing how the mechanism of LPP could be incorporated into a formal analysis of HVD, and of other such processes. We propose two options. As suggested by Wagner (2012), explaining variability in a process' application in terms of production planning could be used to maintain a *non*-probabilistic account: interconsonantal HVD could be a categorical process described in purely segmental terms (i.e. devoice high vowels between voiceless consonants), as in traditional descriptions, while the variability observed across word boundaries and as a function of various factors (prosodic boundary, pause, speech rate, frequency) would be 'factored out' to production planning. Alternatively, LPP could be incorporated into the structure of phonological grammar, as a factor restricting which phonological patterns are possible—similarly to projecting constraint scales based on perceptibility in an Optimality Theoretic analysis which restrict possible neutralization patterns (Steriade 2008). For example, it is predicted to be impossible to have a process which is "more variable" across words than within words. The choice between these options is beyond the scope of this paper.

6.4 Other issues

An important caveat to our characterization of phrase-final devoicing is that we have characterized it without considering a large subset of cases where it could apply: phrase-final short high vowels followed by a *voiced* segment. Recall that the dataset was restricted to tokens followed by a voiceless consonant, for reasons discussed above. Thus, strictly speaking we cannot show that "phrase-final devoicing" shows particular behavior without showing that all aspects of phrase-final devoicing (e.g.

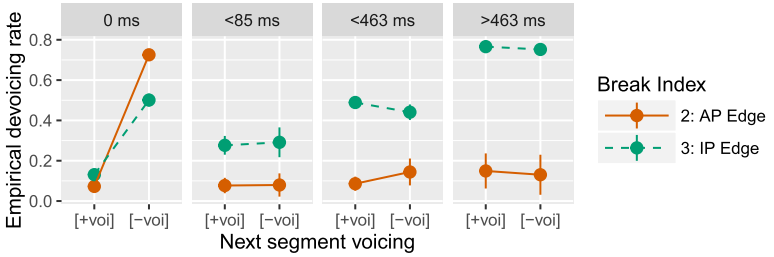


Fig. 8 Percentage of devoiced high short vowels in phrase-final position as a function of following segment voicing, by phrase type (Break Index) and duration of following pause (panel labels). Errorbars indicate ±2 standard errors

frequency effect, devoicing rate) do not depend on the following segment's voicing. As a basic check of this, Fig. 8 shows the empirical devoicing rates for all phrase-final tokens (BI = 2, 3) in the CSJ ("Core" subset), broken down by following pause duration, as a function of voicing of the following segment. When any pause is present—the positions where phrase-final devoicing is most likely to apply—there is no apparent effect of following segment voicing on devoicing rate, suggesting that our characterization of "phrase-final" devoicing is on the right track.[17] A more detailed check would need to consider the role of following segment voicing more generally, in all positions (and report a statistical analysis): within-word and across-word. In fact, the facts are complex: Maekawa and Kikuchi (2005) showed that in the CSJ, devoiced short vowels actually do occur in C̥_[+voi] context *within words*, not infrequently (10%, in the current dataset). By any treatment of HVD, devoicing in this context should be impossible, suggesting that the role of following segment voicing is an important but complex direction for future work.

Another direction for future work is the relationship between the categorical measure (constituent boundaries: AP, IP) and continuous phonetic measures (pause duration, Mora deviation) of boundary strength, and how they affect the two hypothesized devoicing processes. These three measures are strongly correlated, and a more thorough examination of their relationship could give a better understanding of the articulatory characteristics of "final" position. While we have shown that the three measures all independently affect devoicing rate in a dataset where the C̥_C̥ and C̥_# environments are pooled, their relative effects on each kind of devoicing remains unclear. If the phrase-final devoicing process is phonetically driven and interconsonantal devoicing is an "early" phonological rule, then one might expect these processes to be more/less affected by "phonetic" variables (Mora deviation, pause duration) than by categorical constituent boundaries (Break Type), respectively. A reviewer notes that it is difficult to examine this issue without considering Pause Duration as a continuous variable, rather than discretizing it into bins—as was done in this paper to better address our research questions (see Sect. 4.1). Future work examining the effect of pause duration (as a continuous measure) on phrase-final devoicing rate could reveal

[17] When no pause is present, in the left panel, there is an effect of voicing in the direction expected if both C̥_C̥ and C̥_# devoicing can apply before a voiceless segment but only C̥_# can apply before a voiced segment.

a more nuanced relationship, potentially related to a finer prosodic hierarchy than AP/IP.

7 Conclusion

This paper investigated the role that boundary information plays in the variability of Japanese HVD. We focused on teasing apart the effect of highly correlated boundary phenomena—including prosodic phrase boundaries, pauses, and final lengthening—and how these might interact with C̥_C̥ devoicing and define the C̥_# devoicing environment. By examining these factors in a large corpus of spontaneous speech and controlling for other factors known to influence HVD, we were able to pinpoint different sources of variability for HVD depending on the particular context the vowel appears in.

Our results showed that the correlated boundary phenomena have a *joint* influence on variability in HVD. All else being equal, a larger prosodic phrase boundary following the vowel was correlated with a decrease in devoicing rate. Also, the duration of a particular Mora relative to its average duration in the corpus was negatively correlated with the likelihood of HVD, which likely reflects gestural overlap and final lengthening. But the effect of a physical pause was dependent on whether the target vowel was phrase-internal, where pause inhibited HVD, or phrase-final, where pause promoted HVD. The joint effect of a phrase boundary and a long pause led to almost categorical devoicing rates. Phrase-internal and phrase-final vowels were also influenced in qualitatively different ways by speech rate and word frequency: phrase-final vowels showed a positive effect, typical of reductive processes in general, while phrase-internal vowels showed no such effects.

We proposed that there are two separate devoicing processes: interconsonantal and phrase-final devoicing, which show different patterns of variability.

Phrase-final devoicing shows telltale signs of a reductive process, namely the positive effect of speech rate and lexical frequency. This pattern could be accounted for under existing proposals of devoicing as gestural overlap and reduction. Phrase-final devoicing is also promoted by a long pause, which could also receive an articulatory explanation if it is assumed that long pauses at the end of a phrase or utterance are associated with a decrease in subglottal pressure, making it harder to initiate voicing.

Interconsonantal devoicing, on the other hand, shows a pattern of variability that is less easily explained by gestural overlap and reduction. We suggest that its variability can be better understood by reference to the *locality of production planning* hypothesis, which explains part of the variability as a consequence of limitations imposed by online speech production. The inhibitory effect of larger prosodic phrase boundaries, and negative effect of pause for phrase-internal word-final vowels, are due to these two factors correlating with later planning of an upcoming voiceless obstruent, which interferes with the planning of a devoiced vowel variant in the interconsonantal environment.

Acknowledgements A preliminary version of this work was reported in Kilbourn-Ceron (2015). We thank audiences at LabPhon 14 and ICPhS 2015, Kuniko Nielsen, Hisako Noguchi, and James Tanner for feedback on this project; Michael Wagner, Heather Goad, three anonymous reviewers, and Rachel

Walker for useful comments on manuscript drafts; and Michael McAuliffe for translation help. This work was supported by a SSHRC CGS Doctoral Scholarship (767-2012-1089) and CRBLM Graduate Scholar Stipend to Oriana Kilbourn-Ceron, and research grants from SSHRC (#430-2014-00018) and FRQSC (#183356) to Morgan Sonderegger.

Appendix: Random effects

Predictor	Variance	Standard Deviation
Word		
(Intercept)	5.139	2.267
Speaker		
(Intercept)	0.769	0.877
Break Index		
1, 2, 3 – None	0.719	0.848
2, 3 – 1	1.447	1.203
3 – 2	1.467	1.211
Lexical frequency	0.042	0.205
Speech rate within utterance	0.000	0.000
Pause : Break Index		
No Pause – Pause : 1, 2, 3 – None	3.441	1.855
No Pause – Pause : 2, 3 – 1	8.697	2.949
No Pause – Pause : 3 – 2	5.076	2.253
Short Pause – Medium/Long Pause : 1, 2, 3 – None	2.924	1.710
Short Pause – Medium/Long Pause : 2, 3 – 1	16.638	4.079
Short Pause – Medium/Long Pause : 3 – 2	7.409	2.722
Medium Pause – Long Pause : 1, 2, 3 – None	3.549	1.884
Break Index : Lexical Frequency		
1, 2, 3 – None : Lexical frequency	1.332	1.154
2, 3 – 1 : Lexical frequency	2.412	1.553
3 – 2 : Lexical frequency	10.890	3.300
Break Index : Speech rate within utterance		
1, 2, 3 – None : Speech Rate	0.623	0.789
2, 3 – 1 : Speech Rate	1.817	1.348
3 – 2 : Speech Rate	1.281	1.132

References

Baayen, R. Harald. 2008. *Analyzing linguistic data*. Cambridge: Cambridge University Press.

Barnes, Jonathan. 2006. *Strength and weakness at the interface: Positional neutralization in phonetics and phonology*. Berlin: de Gruyter.
Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3): 255–278.
Bates, Douglas, Martin Maechler, Ben Bolker, and Steven Walker. 2015. Fitting linear mixed-effects models using lme4. R package version 1.0-5. *Journal of Statistical Software* 67(1): 1–48. http://CRAN.R-project.org/package=lme4. Accessed 17 April 2017.
Beckman, Mary, and Janet Pierrehumbert. 1988. *Japanese tone structure*. *Linguistic inquiry monographs*. Cambridge: MIT Press.
Beckman, Mary. 1996. When is a syllable not a syllable? In *Phonological structure and language processing: Cross-linguistic studies*, eds. Takashi Otake and Anne Cutler, 95–123. Berlin: de Gruyter.
Byrd, Dani, and Elliot Saltzman. 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31 (2): 149–180.
Cedergren, Henrietta J., and Louise Simoneau. 1985. La chute des voyelles hautes en français de Montréal: As-tu entendu la belle syncope? In *Les tendances dynamiques du français parlé à Montréal*, eds. Monique Lemieux and Henrietta Cedergren, 57–144. Québec: Bibliothèque nationale du Québec.
Coetzee, Andries W., and Shigeto Kawahara. 2013. Frequency biases in phonological variation. *Natural Language and Linguistic Theory* 31(1): 47–89.
Coetzee, Andries W., and Joe Pater. 2011. The place of variation in phonological theory. In *The handbook of phonological theory*, eds. John A. Goldsmith, Jason Riggle, and Alan C. L. Yu, 401–434. Oxford: Wiley-Blackwell.
Crothers, John H., James Lorentz, Donald Sherman, and Marilyn Vihman. 1979. Handbook of phonological data from a sample of the world's languages: A report of the Stanford Phonology Archive. Stanford University: Department of Linguistics.
Dauer, Rebecca M. 1980. The reduction of unstressed high vowels in Modern Greek. *Journal of the International Phonetic Association* 10(1–2): 17–27.
Dell, Gary S., and Padraig G. O'Seaghdha. 1992. Stages of lexical access in language production. *Cognition* 42(1–3): 287–314.
Den, Yasuharu. 2015. Some phonological, syntactic, and cognitive factors behind phrase-final lengthening in spontaneous Japanese: A corpus-based study. *Laboratory Phonology* 6(3–4): 337–379.
Den, Yasuharu and Hanae Koiso. 2015. Factors affecting utterance-final vowel devoicing in spontaneous Japanese. In *18th International Congress of Phonetic Sciences (ICPhS)*, ed. The Scottish Consortium for ICPhS 2015. Glasgow: University of Glasgow. Paper 582.
Ferreira, Fernanda. 1988. Planning and timing in sentence production: The syntax-to-phonology conversion. PhD diss., University of Massachusetts at Amherst.
Ferreira, Fernanda. 1991. Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language* 30(2): 210–233.
Fosler-Lussier, Eric, and Nelson Morgan. 1999. Effects of speaking rate and word frequency on pronunciations in conversational speech. *Speech Communication* 29(2): 137–158.
Fougeron, Cécile, and Patricia A. Keating. 1997. Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America* 101(6): 3728–3740.
Fujimoto, Masako. 2015. Vowel devoicing. In *The handbook of Japanese phonetics and phonology*, ed. Haruo Kubozono, 167–214. Berlin: de Gruyter.
Gelman, Andrew, and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
Gordon, Matthew. 1998. The phonetics and phonology of non-modal vowels: A cross-linguistic perspective. In *Annual Meeting of the Berkeley Linguistics Society (BLS)* 24, 93–105.
Han, Mieko Shimizu. 1962. Unvoicing of vowels in Japanese. *Onsei no Kenkyuu* 10: 81–100.
Harrell Jr., Frank E. 2014. rms: Regression modeling strategies. R package version 4.2-0. http://CRAN.R-project.org/package=rms. Accessed 17 April 2017.
Harrell Jr., Frank E., et al. 2015. Hmisc: Harrell miscellaneous. R package version 3.17-1. https://CRAN.R-project.org/package=Hmisc. Accessed 17 April 2017.
Hasegawa, Nobuko. 1979. Fast speech vs. casual speech. In *Papers from the fifteenth regional meeting of the Chicago Linguistics Society (CLS)*, 126–137.
Hayes, Bruce, and Zsuzsa Cziráky Londe. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23(1): 59–104.

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39(3): 379–440.

Hayes, Bruce, Colin Wilson, and Anne Shisko. 2012. Maxent grammars for the metrics of Shakespeare and Milton. *Language* 88(4): 691–731.

Hirayama, Manami. 2009. Postlexical prosodic structure and vowel devoicing in Japanese. PhD diss., University of Toronto.

Hirayama, Teruo. 1985. Zennippon no hatsuon to akusento [Pronunciation and accent in all Japan]. In *Nihongo hatsuon akusento jiten [Japanese pronunciation accent dictionary]*, ed. Nihon Hoosoo Kyookai, 37–69. Tokyo: Nihon Hoosoo Shuppan Kyookai.

Igarashi, Yosuke, Hideaki Kikuchi, and Kikuo Maekawa. 2006. Inritsu joohoo [Prosodic information]. In *Nihongo hanashi kotoba koopasu no koochikuhoo [Construction of The Corpus of Spontaneous Japanese], Kokuritsu Kokugo Kenkyuujo [National Institute for Japanese Language (NIJAL)] Report* 124, 347–453.

Imai, Terumi. 2004. Vowel devoicing in Tokyo Japanese: A variationist approach. PhD diss., Michigan State University.

Ito, Junko, and Armin Mester. 2012. Recursive prosodic phrasing in Japanese. In *Prosody matters: Essays in honor of Elisabeth Selkirk*, eds. Tony Borowsky, Shigeto Kawahara, Mariko Sugahara, and Takahito Shinya, 280–303. Sheffield: Equinox.

Jannedy, Stefanie. 1995. Gestural phasing as an explanation for vowel devoicing in Turkish. *Ohio State University Working Papers in Linguistics* 45: 56–84.

Jun, Sun-Ah, and Mary Beckman. 1993. A gestural-overlap analysis of vowel devoicing in Japanese and Korean. Paper presented at the 1993 Annual Meeting of the LSA, Los Angeles, January 7–10.

Jurafsky, Dan, Alan Bell, Michelle Gregory, and William D. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In *Frequency and the emergence of linguistic structure*, eds. Joan Bybee and Paul Hopper, 229–254. Amsterdam: Benjamins.

Kaimaki, Marianna. 2015. Voiceless Greek vowels. In *18th International Congress of Phonetic Sciences (ICPhS)*, ed. The Scottish Consortium for ICPhS 2015. Glasgow: University of Glasgow. Paper 791.

Kaisse, Ellen M. 1985. *Connected speech: The interaction of syntax and phonology*. Orlando: Academic Press.

Kawahara, Shigeto. 2011. Japanese loanword devoicing revisited: A rating study. *Natural Language and Linguistic Theory* 29(3): 705–723.

Keating, Patricia A. 2006. Phonetic encoding of prosodic structure. In *Speech production: Models, phonetic processes, and techniques*, eds. Jonathan Harrington and Marija Tabain, 167–186. New York: Psychology Press.

Kilbourn-Ceron, Oriana. 2015. The influence of prosodic context on high vowel devoicing in spontaneous Japanese. In *18th International Congress of Phonetic Sciences (ICPhS)*, ed. The Scottish Consortium for ICPhS 2015. Glasgow: University of Glasgow. Paper 932.

Kiparsky, Paul. 1985. Some consequences of lexical phonology. *Phonology* 2(1): 85–138.

Kondo, Mariko. 1997. Mechanisms of vowel devoicing in Japanese. PhD diss., University of Edinburgh.

Kondo, Mariko. 2005. Syllable structure and its acoustic effects on vowels in devoicing environments. In *Voicing in Japanese*, eds. Jeroen van de Weijer, Kensuke Nanjo, and Tetsuo Nishihara, Vol. 84, 229–246. Berlin: de Gruyter.

Kuriyagawa, Fukuko, and Masayuki Sawashima. 1989. Word accent, devoicing and duration of vowels in Japanese. *Annual Bulletin of the Research Institute of Language Processing* 23: 85–108.

Kuwabara, Hisao and Kazuya Takeda. 1988. Analysis and prediction of vowel devocalization in isolated Japanese words. *The Journal of the Acoustical Society of America* 83(S1): 29.

Labrune, Laurence. 2012. *The phonology of Japanese*. Oxford: Oxford University Press.

Levelt, Willem J. M., Ardi Roelofs, and Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22(1): 1–38.

Lombardi, Linda. 1991. Laryngeal features and laryngeal neutralization. PhD diss., University of Massachusetts.

Lovins, Julie B. 1976. Pitch accent and vowel devoicing in Japanese: A preliminary study. *Annual Bulletin of Research Institute of Logopedics and Phoniatrics, University of Tokyo* 10: 113–125.

Maekawa, Kikuo, and Hideaki Kikuchi. 2005. Corpus-based analysis of vowel devoicing in spontaneous Japanese: An interim report. In *Voicing in Japanese*, eds. Jeroen van de Weijer, Kensuke Nanjo, and Tetsuo Nishihara, 205–228. Berlin: de Gruyter.

Maekawa, Kikuo, Hanae Koiso, Sadaoki Furui, and Hitoshi Isahara. 2000. Spontaneous speech corpus of Japanese. In *Second International Conference of Language Resources and Evaluation (LREC)* 2,

947–952.
Maekawa, Kikuo, Hideaki Kikuchi, Yosuke Igarashi, and Jennifer Venditti. 2002. X-JToBI: An extended J-ToBI for spontaneous speech. In *7th International Conference on Spoken Language Processing (ICSLP2002)*, 1545–1548. Denver.
Martin, Andrew. 2011. Grammars leak: Modeling how phonotactic generalizations interact within the grammar. *Language* 87(4): 751–770.
Martin, Andrew, Akira Utsugi, and Reiko Mazuka. 2014. The multidimensional nature of hyperspeech: Evidence from Japanese vowel devoicing. *Cognition* 132(2): 216–228.
McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.
Michelson, Karin. 1999. Utterance-final phenomena in Oneida. In *Linguistics and Phonetics* 4 *(LP'98): Item order in language and speech*, eds. Osamu Fujimura, Brian D. Joseph, and Bohumil Palek, 31–45. Prague: Charles University Press.
Mohanan, Karuvannur Puthanveettil. 1982. Lexical phonology. PhD diss., Massachusetts Institute of Technology.
Nagy, Naomi, and Bill Reynolds. 1997. Optimality theory and variable word-final deletion in Faetar. *Language Variation and Change* 9(1): 37–55.
Nespor, Marina, and Irene Vogel. 1986. *Prosodic phonology*. Berlin: de Gruyter.
NHK. 1991. *Nihongo hatsuon akusento jiten [Japanese pronunciation accent dictionary]*. Tokyo: Nihon Hoosoo.
Nielsen, Kuniko Y. 2015. Continuous versus categorical aspects of Japanese consecutive devoicing. *Journal of Phonetics* 52: 70–88.
Ogasawara, Naomi. 2013. Lexical representation of Japanese vowel devoicing. *Language and Speech* 56(1): 5–22.
Oi, Mutsumi. 2013. The interaction between accent contrast and vowel devoicing in Tokyo Japanese. Master's thesis, University of Ottawa.
Pluymaekers, Mark, Mirjam Ernestus, and R. Harald Baayen. 2005. Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America* 118(4): 2561–2569.
R Core Team. 2013. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.R-project.org. Accessed 17 April 2017.
Shih, Stephanie S. 2016. Super additive similarity in Dioula tone harmony. In *33rd West Coast Conference on Formal Linguistics (WCCFL)*, ed. Kyeong min Kim, Pocholo Umbal, Trevor Block, Queenie Chan, Tanie Cheng, Kelli Finney, Mara Katz, Sophie Nickel-Thompson, and Lisa Shorten, 361–370. Somerville: Cascadilla Proceedings Project.
Smith, Caroline L. 2003. Vowel devoicing in contemporary French. *Journal of French Language Studies* 13(2): 177–194.
Snijders, Tom A. B., and Roel J. Bosker. 1999. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks: Sage.
Sohn, Ho-min. 1975. *Woleaian reference grammar*. Honolulu: University of Hawaii Press.
Steriade, Donca. 2008. The phonology of perceptibility effects: The p-map and its consequences for constraint organization. In *The nature of the word: Studies in honor of Paul Kiparsky*, eds. Kristin Hanson and Sharon Inkelas, 151–179. Cambridge: MIT Press.
Sternberg, Saul, Stephen Monsell, Ronald L. Knoll, and Charles E. Wright. 1978. The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In *Information processing in motor control and learning*, ed. George Stelmach, 117–152. New York: Academic Press.
Stevens, Mary. 2012. A phonetic investigation into "raddoppiamento sintattico" in Sienese italian. PhD diss., University of Melbourne.
Takeda, Kazuya, Yoshinori Sagisaka, and Hisao Kuwabara. 1989. On sentence-level factors governing segmental duration in Japanese. *The Journal of the Acoustical Society of America* 86(6): 2081–2087.
Tanner, James, Morgan Sonderegger, and Michael Wagner. 2017. Production planning and coronal stop deletion in spontaneous speech. Ms. *Laboratory Phonology*, to appear.
Torreira, Francisco, and Mirjam Ernestus. 2011. Vowel elision in casual French: The case of vowel /e/ in the word *c'était*. *Journal of Phonetics* 39(1): 50–58.
Tsuchida, Ayako. 1997. Phonetics and phonology of Japanese vowel devoicing. PhD diss., Cornell University.
Tsuchida, Ayako. 2001. Japanese vowel devoicing: Cases of consecutive devoicing environments. *Journal of East Asian Linguistics* 10(3): 225–245.
Vance, Timothy J. 1992. Lexical phonology and Japanese vowel devoicing. In *The joy of grammar: A festschrift in honor of James D. McCawley*, eds. Gary N. Larson, Lynn A. MacLeod, James D. Mc-

Cawley, and Diane Brentari, 337–350. Amsterdam: Benjamins.
Vance, Timothy J. 2008. *The sounds of Japanese*. Cambridge: Cambridge University Press.
Varden, J. Kevin. 1998. On high vowel devoicing in standard modern Japanese: Implications for current phonological theory. PhD diss., University of Washington.
Varden, J. Kevin. 2010. On vowel devoicing in Japanese. *The MGU Journal of Liberal Arts Studies* 4(1): 223–235. http://hdl.handle.net/10723/83. Accessed 17 April 2017.
Venditti, Jennifer J. 2005. The J_ToBI model of Japanese intonation. In *Prosodic typology: The phonology of intonation and phrasing*, ed. Sun-Ah Jun, 172–200. Oxford: Oxford University Press.
Venditti, Jennifer J., Kazuaki Maeda, and Jan P. H. van Santen. 1998. Modeling Japanese boundary pitch movements for speech synthesis. In *The third ESCA/COCOSDA workshop (ETRW) on speech synthesis*, 317–322.
Venditti, Jennifer J., Kikuo Maekawa, and Mary E. Beckman. 2008. Prominence marking in the Japanese intonation system. In *Handbook of Japanese linguistics*, ed. Natsuko Tsujimura, 456–512. Oxford: Blackwell.
Wagner, Michael. 2012. Locality in phonology and production planning. *McGill Working Papers in Linguistics* 22(1): 1–18.
Wheeldon, Linda. 2013. Producing spoken sentences: The scope of incremental planning. In *Cognitive and physical models of speech production, speech perception, and production-perception integration*, eds. Susanne Fuchs, Melanie Weirich, Daniel Pape, and Pascal Perrier, 97–118. Bern: Peter Lang.
Wheeldon, Linda, and Aditi Lahiri. 1997. Prosodic units in speech production. *Journal of Memory and Language* 37: 356–381.
Wheeldon, Linda, and Aditi Lahiri. 2002. The minimal unit of phonological encoding: Prosodic or lexical word. *Cognition* 85(2): 31–41.
Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf, and Patti J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America* 91(3): 1707–1717.
Yoshida, Natsuya, and Yoshinori Sagisaka. 1990. Factor analysis of vowel devoicing in Japanese [in Japanese]. ATR Technical Report TR-I-0159. Kyoto: ATR Interpreting Telephony Research Laboratories.