To appear in the Journal of the Phonetic Society of Japan 21(1). This is not a final version.

On the Input Information of the C/D model for Vowel Devoicing in Japanese

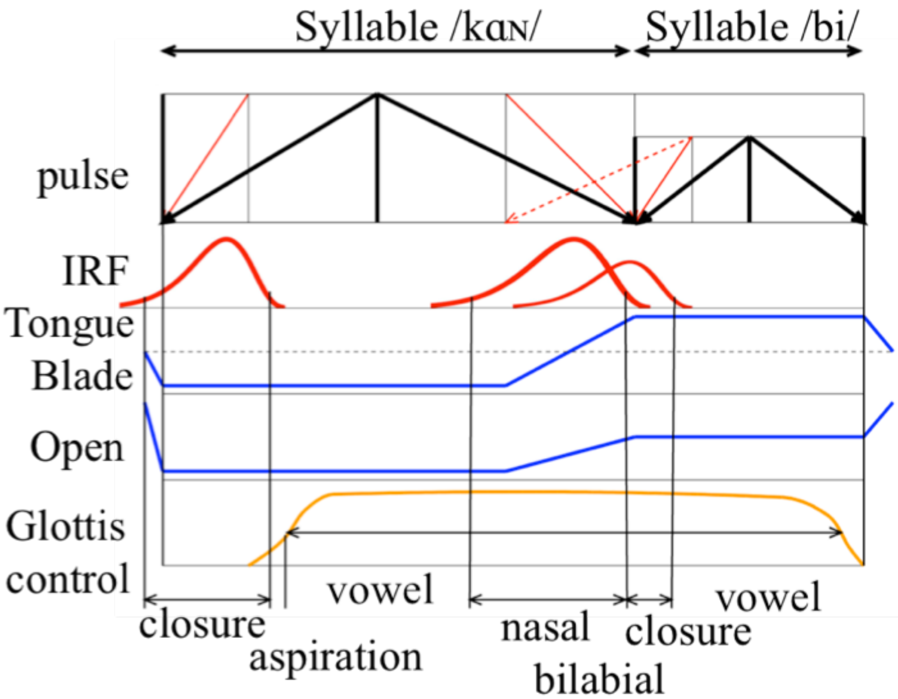
Michinao F. Matsui
(translated by Shigeto Kawahara)

Abstract: The C/D model (Fujimura 1992, 2007) is an explicit framework that calculates the continuous physical gestures and other quantitative phonetic information of speech sounds from inputs that solely consist of qualitative phonological information. Fujimura proposes that the inputs to the C/D model are syllables, themselves consisting of sets of unary and underspecified phonological features, instead of a set of binary features used in many standard phonological theories. This paper slightly revises this original formulation by Fujimura and instead proposes that syllables can be further defined in terms of "mora sets", in addition to unary features that define the qualitative characteristics of the Impulse Response Functions (IRFs). The argument is developed based on discussion of vowel devoicing in Japanese.

Keywords: the C/D model, Impulse Response Function (IRF), vowel devoicing, fricative, moras and syllables

1. Aims of the paper: inputs to the C/D model

Fujimura (1992, 2007) proposed the Converter/Distributer model (henceforth the C/D model), which takes as its input categorical, qualitative information containing phonological, semantic and paralinguistic information. It converts this qualitative information to quantitative outputs with articulatory (and acoustic) information. One characteristic of this model is that it posits the global "base state (base function)", determined by the syllable structures, and consonantal gestures are locally superimposed onto this base function (Kitahara 2008; Kawahara 2015).



[xxx INSERT FIGURE 1 HERE]
Figure 1. A C/D model diagram of ‘kaɴbi’

Consider Figure 1, which is a C/D model representation of ‘kaɴbi’ ([kɑm·bʲi]) that means “tantalizing” in English. The syllable pulses determine the duration of each syllable, while the syllable nuclei ([ɑ] and [i]) determine the general structure of the oral cavity (the tongue body and the jaw in Figure 1). Consonantal gestures of [k], [m], and [b] exert only local influences, modelled with Impulse Response Functions (IRFs). The C/D model uses syllables, rather than phonemes, as its input. For the case of ‘kaɴbi’ in Figure 1, the input consists of {low, stop^{O}, dorsal^{O}, nasal^{C}} (for the first syllable) and {high, palatal, stop^{O}, labial^{O}, voiced^{O}} (for the second syllable), where each feature bundle represents each syllable, and each feature is privative, rather than binary. Superscripted ^{O} and ^{C} represent “Onset” and “Coda”, respectively. The use of privative (atomic) features reflects the C/D model’s assumption that the input information represents concrete articulatory commands, although one may argue that binary features may better capture the aspect of language as a system of differences (Saussure 1917). Note, however, that not all the features must be fully specified in the C/D model either (i.e. features can be underspecified; see section 5 on this issue). The C/D model takes the quantitative input information, together with prosodic and speech style information, *converts* them into quantitative targets based on syllable pulse triangles, and *distributes* the information from atomic features, ultimately calculating quantitative articulatory commands (and their acoustic consequences).

As of today, the research on the C/D model largely focuses on the strengths of syllable pulses, and there is not much discussion on what the inputs should look like in the C/D model, except in the original work by Fujimura himself. The paucity of this discussion may partly come from the fact/assumption that the inputs of the C/D model can simply be outputs of phonology, and therefore they should simply follow from the results of phonological theory. However, there are issues that are specific to the C/D model, independent of the phonological theory that one embraces. For example, Fujimura (2007: 149) argues that the input of ‘kaɴbi’ is {low, stop^{O}, dorsal^{O}, nasal^{C}; high, palatal, stop^{O}, labial^{O}, voiced^{O}}, which consists of multiple sets, each representing one syllable, separated by a semicolon. However, although it makes sense to use a set of feature bundles *within* a syllable (see section 4.2), it does not seem appropriate to use a set to represent information *across* multiple syllables, as syllable chains are affected by factors such as morphological linearity. This problem can be amended by introducing the notion of a list structure, as in <{low, stop^{O}, dorsal^{O}, nasal^{C}}, {high, palatal, stop^{O}, labial^{O}, voiced^{O}}>. One advantage of this method is that this list notation is compatible with Head-driven Phrase Structure Grammar (HPSG) (Pollard & Sag 1994, Gunji 1994). This way, we can model the interaction between phonetic information and semantic/syntactic information, which allows us to address "the nature of linguistic structure in general", a goal expressed in the subtitle of Fujimura's (2007) book. As illustrated by this example, examining the input information for the C/D model explicitly reflects the philosophical nature of the C/D model, and it is crucial to explore this issue to examine the entire architecture of the C/D model itself. Therefore, this paper takes high vowel devoicing in Japanese as an example, and investigates what the inputs to the C/D model should look like. To preview the conclusion, this paper argues that (i) input structures to the C/D model need to contain moras within syllables, and (ii) featural underspecification naturally follows from the nature of atomic features, which are actually "phonological particles" and (iii) the natures of phonological particles can have influences on the phonetic patterning.

2. Characteristics of devoiced vowels following voiceless fricatives

2.1. Two patterns of vowel devoicing

In general, close vowels in Tokyo Japanese are devoiced between two voiceless consonants. Kawakami (1977) and Maekawa (1989) argue that there are two kinds of devoicing, as in (1):

(1) a. For /si/ (its pronunciation is [ɕi] with the non-devoiced vowel in general), /su/ ([sɯ̈]), /hi/ ([çi]), /hu/ ([ɸu]), /ti/ ([tɕi]), /tu/ ([tsɯ̈]), "devoiced vowels" are simply a prolongation of the preceding fricatives, and no vowels exist (see also Baba 1998).

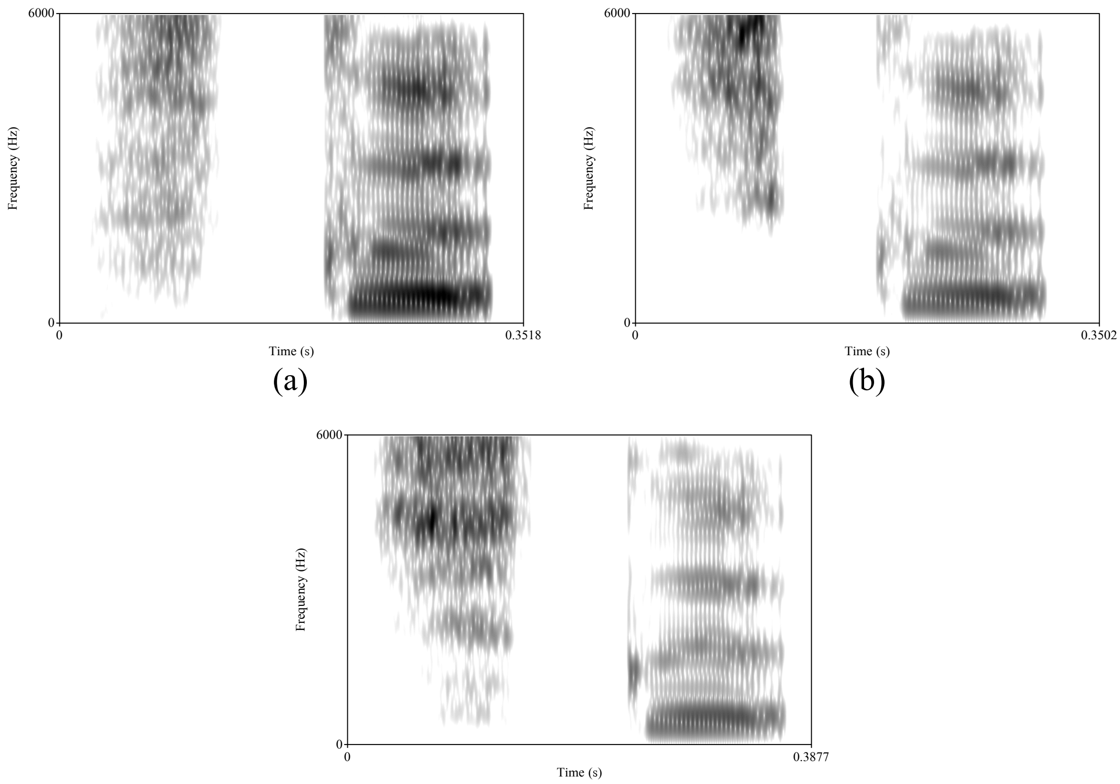
b. For /ki/ ([kʲi]), /ku/ ([kɯ]), /pi/ ([pʲi]), /pu/ ([pu]), /syu/ ([ɕɯ]), /tyu/ ([tɕɯ]) and others, there are in fact devoiced vowels.

Broadly speaking, (1a) includes fricatives and affricates, and (1b) includes stops. The interesting exceptions are /syu/ ([ɕɯ]) and /tyu/ ([tɕɯ]), and it is possible that the vocalic difference between /i/ and /u/ following [ɕ] affects the calculation of vowel devoicing, making them exceptional to the generalizations in (1). To address this hypothesis, the next section reports acoustic, EPG (ElectroPalatoGraphy) and perception studies on [sɯ̥], [ɕi̥], and [ɕɯ̥] with the devoiced vowel.

3.Articulatory nature of devoiced vowels following voiceless fricatives

3.1. Acoustic characteristics of devoiced [sɯ], [ɕi], [ɕɯ]

Beckman & Shoji (1984) and Tsuchida (1997) report that there were differences in spectral characteristics between [ɕi] and [ɕɯ], but we actually observe temporal differences as well. Figures 2(a)-(c) show a spectrogram of /suku/ ([sɯ̥̈kɯ]), /siku/ ([ɕi̥kɯ]), and /syuku/ ([ɕɯ̥kɯ]) produced by a female speaker of Tokyo Japanese. The frication portions of the syllable /su/ and /si/ are stable across their constriction, but on the other hand, in the syllable /syɯ/, the frication frequencies decrease toward its end, showing some transition. This observation implies that even with /syu/ ([ɕɯ̥kɯ]) with a devoiced vowel, the tongue moves toward the palate for the onset consonant, and also that even though the vocalic interval is acoustically realized with nothing but frication, the backing gesture of the vowel [ɯ] remains. On the other hand, we do not seem to observe clear evidence for the remaining vocalic gesture for /su/ ([sɯ̥̈kɯ]) and /si/ ([ɕi̥kɯ]). In order to confirm these hypotheses, I used EPG to more directly examine the articulation of devoiced vowels.



[xxx INSERT FIGURE 2 HERE]

(a) /suku/ ([sɯ̥̈kɯ]) (b) /siku/ ([ɕi̥kɯ]) (c) /syuku/ ([ɕɯ̥kɯ])

Figure 2. Spectrograms of devoiced high vowels following voiceless fricatives.

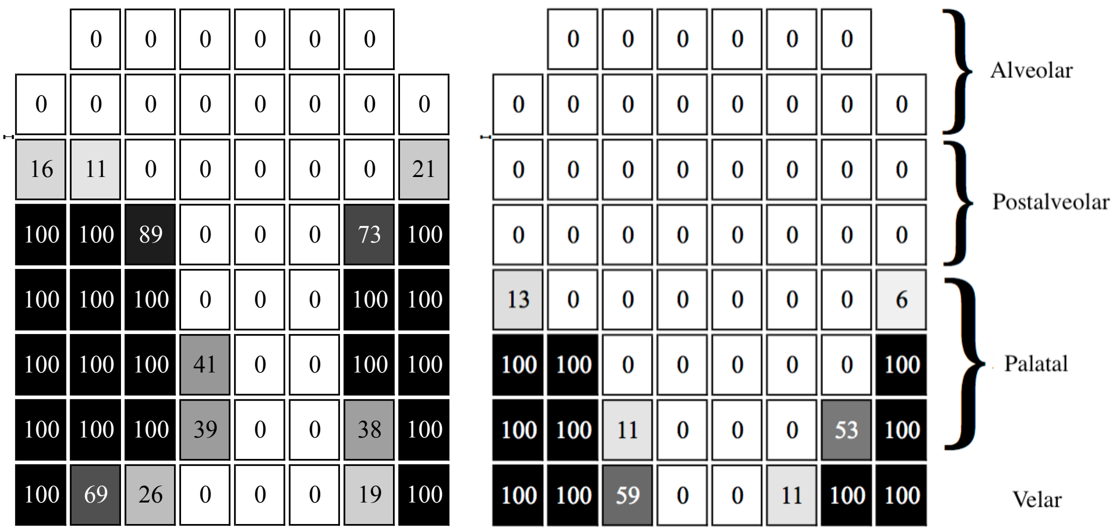
3.2. EPG measurement of devoiced vowels

The EPG study used WinEPG (Articulate Instruments Ltd) and Stars system (courtesy of EPG research group, Dr. Ichiro Yamamoto), whose sampling period is 10 ms. All the artificial palates were created by Dr. Yamamoto, and they are standardized so that cross-speaker comparison is possible. More specifically, the electro-nodes are arranged in such a way that the first two lines represent alveolar, the next two lines represent postalveolar, the next three lines represent hard palate, and the final line represents velar (see Figure 3). In terms of vertical dimension resolution, the first line has 6 points, and the remaining lines have 8 points.

The participants of the experiment were two speakers of Kansai Japanese (one male and one female), who were used to speaking with an EPG artificial palate. The stimuli include both real words and nonce words which contained a fricative-vowel sequence in their second syllable (e.g. /asuka/ ([ɑsɯ̥̈kɑ]) “personal name” and /asika/ ([ɑɕi̥kɑ]) “sea otter”). The only EPG study on Japanese devoiced vowels that I know of is Nakamura (2003), but the study is intended to examine the coarticulatory nature of devoiced vowels, and does not report the articulatory nature of devoiced vowels themselves. Therefore, this experiment had two conditions (one voiced and one voiceless), and in order to standardize the effects of vowel-to-vowel coarticulation (Öhman 1966), all vocalic conditions except [o] were included. The common features of the devoiced second syllables across these vocalic conditions were extracted. The speakers were instructed to repeat the stimuli three times with no accent; in cases when the speakers produced a stimulus with accent, they were not forced to produce that stimulus again, however.

3.3. The EPG data of devoiced high vowels following fricatives

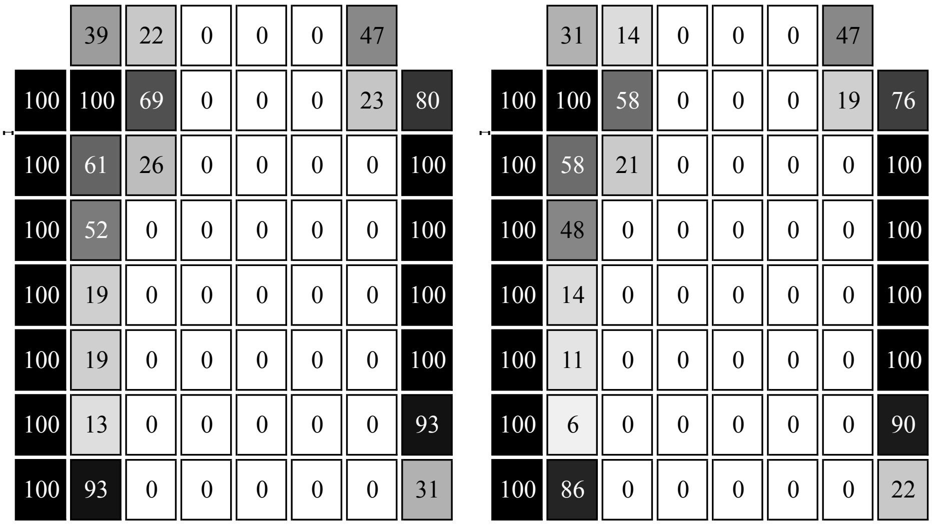
The patterns of the two speakers were comparable, and therefore, their data are pooled together. In the EPG figures that follow, each cell represents the percentages of that particular cell being contacted. Figure 3 shows the articulation of [i] and [ɯ] that followed voiceless fricatives, but were not devoiced. We can observe that [i] shows contact at the regions ranging from postalveolar to hard palate, whereas [ɯ] shows lateral contact from hard palate to soft palate. Figures 4 and 5 show devoiced [sɯ̥] and [ɕi̥], which show no particular differences between the first half of the syllable and the second half of the syllable. This observation is compatible with the characterization in (1a) that devoiced vowels are simply a prolongation of the preceding fricatives, and also compatible with the acoustic characteristics shown in Figures 2(a) and (b).



(a) (b)

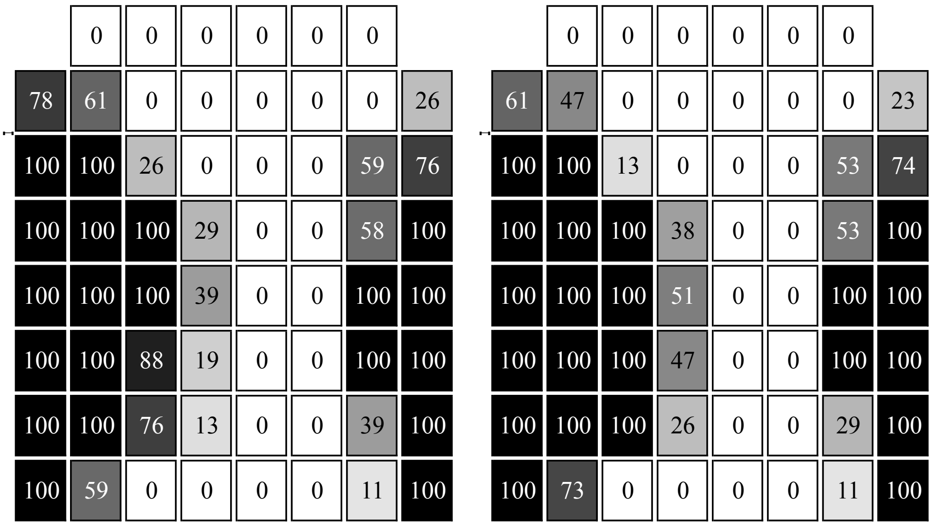
[xxx INSERT FIGURE 3 HERE]

Figure 3: Average percentages of tongue contact for non-devoiced [i] and [ɯ]



(a) (b)

[xxx INSERT FIGURE 4 HERE]
Figure 4: Average percentages of tongue contact for the syllable /su/ ([sɯ̥̈]); (a) showsthe first half of the syllable; (b), the second half of the syllable.



(a) (b)

[xxx INSERT FIGURE 5 HERE]
Figure 5: Average percentages of tongue contact for the syllable /si/ ([ɕi̥]); (a) shows the first half of the syllable; (b), the second half of the syllable.

On the other hand, consider the devoiced [ɕɯ̥], whose articulatory patterns are shown in Figure 6. Like [sɯ̥̈] and [ɕi̥], we do not observe clear tongue contact; however, we also observe more extended tongue contact in the second half of the syllable. This change in articulation from Figure 6(a) to Figure 6(b) can probably be the cause of the transition of the frication energy that is observed in Figure 2(c). Further, compare Figure 6(b) with Figure 3(b)—although the tongue contact patterns are slightly different from alveolar to postalveolar, their patterns are similar between hard palate to velar. From this, I tentatively conclude that the articulatory movement that we observe in the latter part of the [ɕɯ̥] syllable is due to the backness feature of [ɯ]. This conclusion is compatible with the claim in (1b) that devoiced [ɕɯ] still contains the vowel [ɯ], even after it is devoiced. However, since the oral constriction is strong even during the second portion of [ɕɯ̥], this results in frication. Therefore, it seems more reasonable to consider this entire syllable as containing “fricative vowels”, a notion that has been proposed by several researchers before (Soli 1981; Whalen 1983; Tsuchida 1994). If Japanese has fricative vowels, as in Figure 6b, we should also entertain the possibility that Figure 4b and Figure 5b involve fricative vowels as well (Goad & Shimada 2014).

	0	0	0	0	0	0	
83	69	0	0	0	0	0	34
100	100	23	0	0	0	58	79
100	100	100	29	0	0	69	100
100	100	100	39	0	0	100	100
100	100	73	14	0	0	100	100
100	100	69	8	0	0	26	100
100	44	0	0	0	0	6	100

(a)

	0	0	0	0	0	0	
44	29	0	0	0	0	0	31
83	68	0	0	0	0	44	66
92	77	13	0	0	0	39	52
100	100	56	30	0	0	100	100
100	100	83	55	0	0	100	100
100	100	100	41	0	0	100	100
100	100	100	13	0	0	100	100

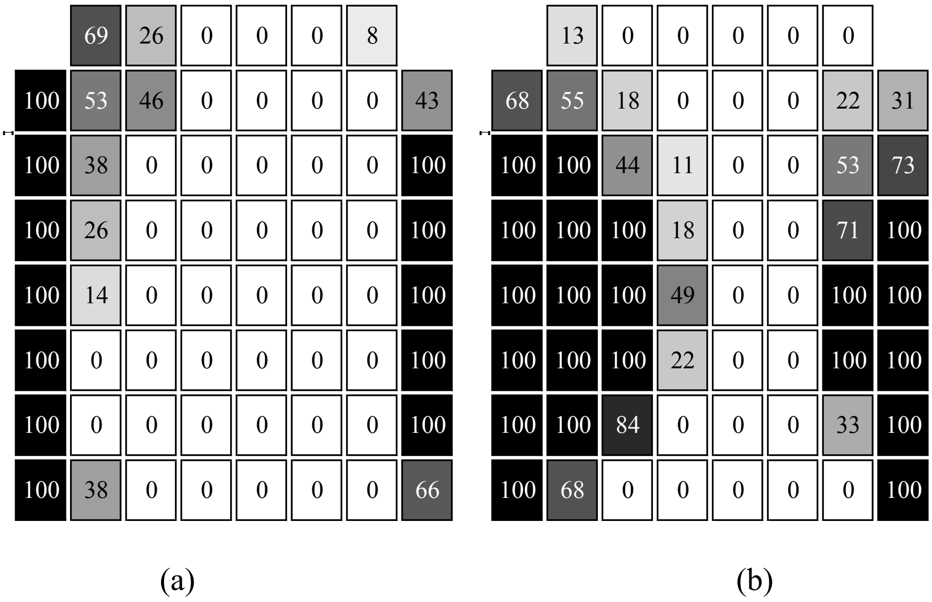
(b)

[xxx INSERT FIGURE 6 HERE]

Figure 6: Average percentages of tongue contact for the syllable /syu/ ([ɕɯ̥]); (a) shows the first half of the syllable; (b), the second half of the syllable.

3.4. EPG data on voiced fricatives and following vowels

To further validate the conclusion that we reached in the previous section, I further explored the EPG patterns of [zɯ] and [ʑi] by examining words like /sizisi/ ([ɕiʑiɕi]) “pronouns” and /kuzusu/ ([kɯzɯsɯ]) “to destroy”, in which the fricatives are surrounded by identical vowels. Figure 7 shows the second part of the [zɯ] and [ʑi] syllables, and exhibit almost the same tongue contact pattern with [ɯ] and [i], respectively. We know that these syllables are not devoiced, and therefore, Figure 7 suggests that voiced fricatives may have the same tongue contact pattern as the surrounding vowels; i.e. Japanese speakers can produce fricatives with the same tongue configuration as the surrounding vowels. This sort of articulatory data, as far as I know, has not been reported in the literature, and it offers an important piece of new data to the field.



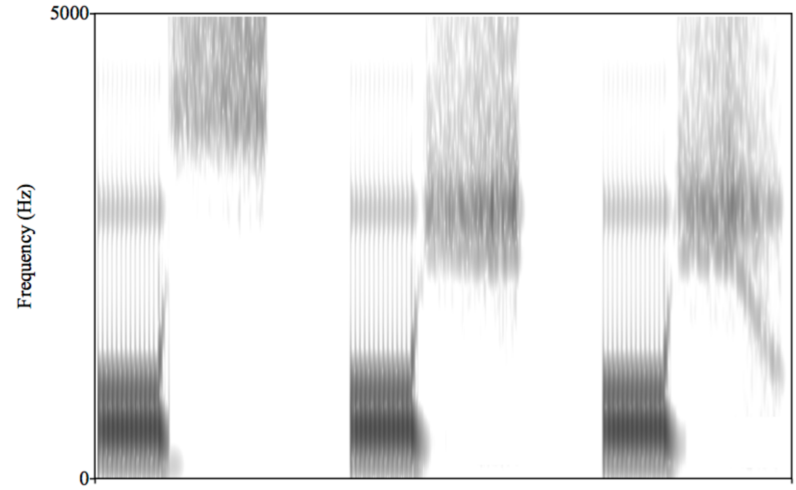
[xxx INSERT FIGURE 7 HERE]

Figure 7: The latter portions of (a) [zɯ] and (b) [ʑi], preceded by [ɯ] and [i], respectively.

This conclusion adds more credence to the idea that articulatory patterns illustrated in Figure 4(b) and Figure 5(b) involve fricative vowels, as they look very similar to those in Figures 7(a) and (b). Especially [ɕi] in Figure 5(b) resembles the palatal type contact of [i] that we observe in Figure 3(b), and we consider the more extensive contact observed in Figure 5(b) to be the result of the stronger constriction that is necessary to create frication. One remaining problem is [sɯ]. Consider Figure 4(b) again, which does not resemble [ɯ] in Figure 3(b)—this problem is addressed later in section 5. In the next section, I further pursue the idea that [s] and [ɕ] can function not just as voiceless fricatives, but also as fricative vowels. I tackle this hypothesis from the perspective of speech perception.

4. Perception of devoiced vowels/fricative vowels
4.1. Previous studies on the perception of fricatives

Tsuchida (1994) reports a perception experiment on the syllable containing a fricative and a devoiced vowel, and concludes that what is important for the perception of the devoiced vowels is the perception of the preceding fricatives, which themselves depend on frequency properties of fricatives. Consider Figure 8. For example, high frequency noise such as those in Figure 8(a) is perceived as [s], and low frequency noise such as those in Figure 8(b) is perceived as [ɕ]. Matsui (2013) shows that so long as the energy distribution of high frequency noise stays constant, Japanese listeners do not perceive /syu/ ([ɕɯ]). Instead, in order for them to perceive /syu/ ([ɕɯ]), it is necessary that the syllable contains descending transitional energy, as in Figure 8(c). This perceptual feature of /sy/ ([ɕɯ]) seems to be compatible with the acoustic characteristics that we observed in Figure 2, and also the stability of articulation of [sɯ] and [ɕi] in Figures 4-6 and the backward articulatory movement of [ɕɯ] in Figure 7.



(a) (b) (c)

[xxx INSERT FIGURE 8 HERE]

Figure 8: Acoustic characteristics of those stimuli that cause the perception of /osu/ ([osɯ̥̈]), /osi/ ([oɕi̥]) and /osyu/ ([oɕɯ̥]).

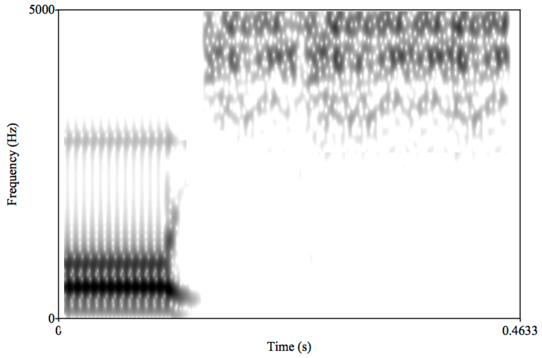
From these data, it seems safe to conclude that even if the [ɕɯ] syllable is devoiced, some sort of "back vowel" remains, and that the descending transition of frication energy illustrated in Figure 8(c) provides a perceptual cue to the remaining vocalic gesture. On the other hand, recall that [sɯ] and [ɕi] show very stable energy distributions (Figures 8(a), (b)), and therefore do not offer perceptual cues to the following vowels. Nevertheless, Japanese listeners perceive /ɯ/ and /i/, respectively, given the stimuli like Figure 8(a) and Figure 8(b). We can entertain two possibilities about why this perception occurs.

(2) a. The perception of these vowels occurs as perceptual illusion (Dupoux, Kakehi, Hirose, Pallier, Mehler 1999). Given the syllables with no vocalic signals (Figures 8(a), (b)), listeners identify the fricatives from a bottom-up manner, and knowledge of Japanese phonotactics and coarticulation consequently determines the following vowel in a top-down manner.

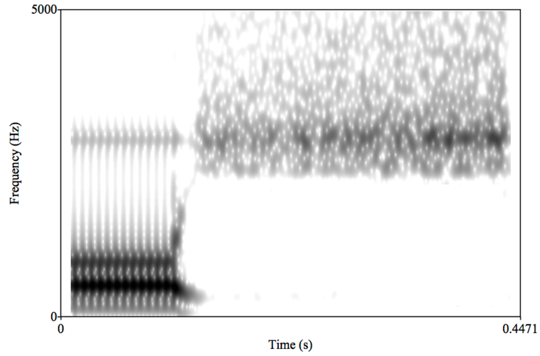
b. Japanese /ɯ/ and /i/ are "fricative vowel" allophones following voiceless fricatives. Figures 8(a) and (b) actually instantiate these fricative vowels. Thus, both the consonants and vowels can be perceived from the bottom-up manner, based only on the incoming acoustic signals.

4.2. Perception experiment: The effects of fricative duration on the perception of geminates without following vowels

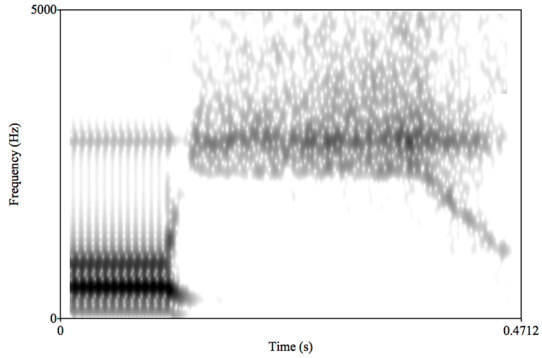
To tease apart the two hypotheses formulated in (2), I conducted the following experiment on the perception of geminates. Previous research has shown that various factors, such as the ratio between the consonant duration and preceding mora duration (Hirata 2007), affect the perception of geminates. No matter what, though, consonantal duration needs to be long in order for listeners to perceive geminates. Let us now consider Figure 9, which simply lengthened the stable portion of the frication in Figure 8.



(a)



(b)



(c)

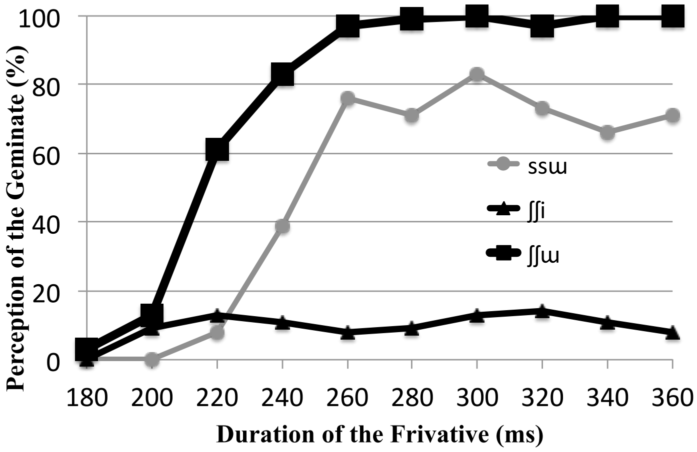
[xxx INSERT FIGURE 9 HERE]

Figure 9: Acoustic characteristics of those stimuli that cause the perception of /osu/ ([osɯ̥̈]), /osi/ ([oɕi̥]) and /osyu/ ([oɕɯ̥]). Frication intervals are lengthened.

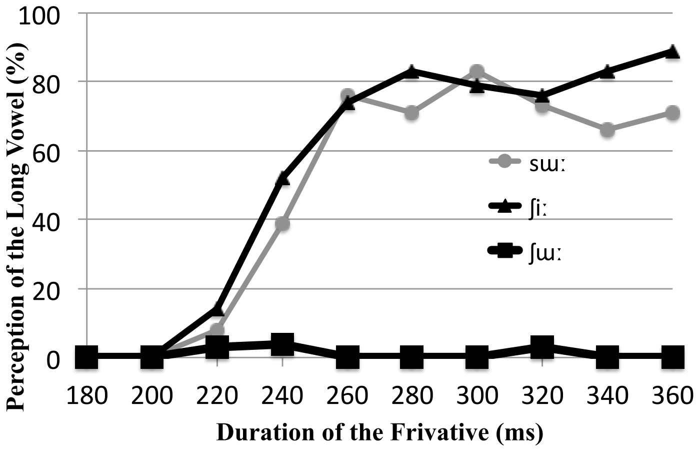
Hypothesis (2a) predicts that acoustic frication should be perceived as "consonantal fricatives", and therefore all the stimuli in Figures (a), (b), (c) can induce geminate percepts, /oQsu/ ([ossɯ̥̈]), /oQsi/ ([oɕɕi̥]), and /oQsyu/ ([oɕɕɯ̥]). On the other hand, Hypothesis (2b) predicts that given the stimuli in Figures 9(a) and (b), some portions of the frication should be perceived as "consonantal fricatives" with the remaining portion perceived as "fricative vowels" and therefore, it is not necessarily the case that they would result in a geminate percept. However, given the acoustic signal in Figure 9(c), the descending frication portion should be perceived as a "fricative vowel", and the preceding frication portion should be perceived as a "consonantal fricative", resulting in geminate percepts.

The actual experiment had two conditions. One is the presence/absence of transition (10 Hz/ms transition at the end vs. no transition). The other one is the consonantal duration, from 180 ms to 360 ms in 20 ms increments. The preceding vowel was fixed as [o] at 120 ms. The participants were 39 speakers of Japanese witha forced choice task, in which the given options were [osɯ, ossɯ, osɯɯ, oɕi, oɕɕi, oɕii, oɕɯ, oɕɕɯ, oɕɯɯ]. The results of this test are shown in Figure 10. The line with circles represent the continuum based on Figure 9(a), the line with triangles represent the continuum based on Figure 9(b), and the line with squares represents the continuum based on Figure 9(c). Figure 10(a) shows the geminate response percentages, while Figure 10(b) shows the long vowel response percentages. We can observe from these figures that simply prolonging frication like Figure 9(a)(b) does not result in geminate percepts. Instead, as the frication gets longer, we observe more long vowel percepts like /osuR/ ([osɯɯ]) and /osiR/ [oɕii]. On the other hand, the stimuli like Figure

9(c), which involves a transition, result in geminate percepts, but not long vowel percepts. These results support Hypothesis (2b) and reject Hypothesis (2a).



(a)



(b)

[xxx INSERT FIGURE 10 HERE]

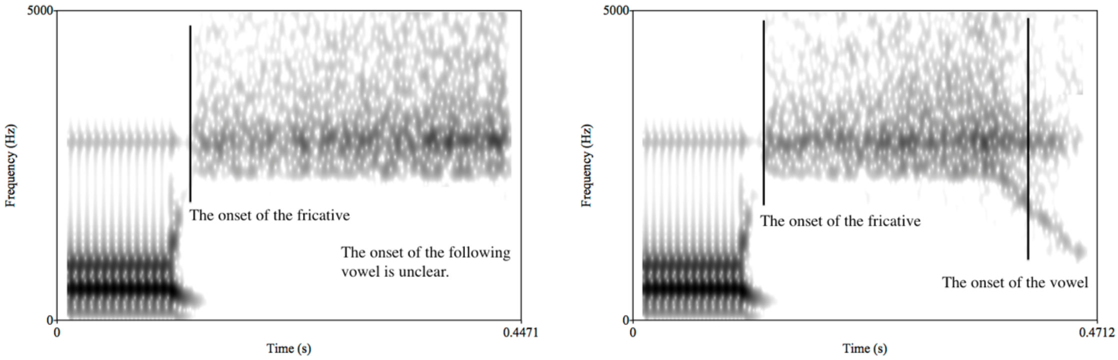
Figure 10: The duration of frication and its resulting percepts ((a): geminate response percentages, (b): long vowel response percentages)

Once we accept Hypothesis (2b), it has a significant implication for one of the assumptions of the C/D model, namely, that the input to the C/D model is a set of consonants and a vowel, unordered within a syllable. Hypothesis (2b) suggests that acoustic characteristics shown in Figures 2(a) and (b) and articulatory characteristics in Figures 8(a), (b) and Figures 9(a) and (b) form one information-package consisting of an onset fricative consonant and a following fricative vowel.

4.3. Syllable weight and the perception of geminates and long vowels

Let us also recall that when frication duration is lengthened, it results in the percept of either geminates or long vowels. This result suggests that Japanese listeners use durational information to identify whether incoming acoustic signals form a light syllable or a heavy syllable, before deciding on the actual segmental content. Let us reconsider the perceptual process of Figures (a)-(c) from the perspective of syllable weight.

First of all, for all types of the stimuli, when frication duration is long, listeners perceive heavy syllables. Japanese heavy syllables can include a coda nasal, the first part of a geminate, or a long vowel. Since the stimuli in Figure 9 include no cue to nasality (nasal pole or anti-formant), we can exclude coda nasals. Now in Figures 9(a) and (b), consonantal fricatives and fricative vowels form one long frication, and therefore, the onset of the (fricative) vowel is unclear, as shown in Figure 11(a). Since it is crucial to identify the onset of the following vowel to determine the duration of the consonant, the geminate percepts are inhibited, and the only remaining option is a long vowel. On the other hand, for [ɕɯ], as shown in Figure 11(b), the transition of frication energy serves as a cue to the onset of the following vowel. As a result, the transition makes it possible to determine the onset and the offset of the consonant interval, resulting in the geminate percept.



[xxx INSERT FIGURE 11 HERE]

(a) (b)

Figure 11: [ɕ] with the following devoiced vowel. The impossibility (a) and possibility (b) of determining the boundary between the fricative and the vowel.

This conclusion actually has independent evidence from patterns of loanword adaptation. As shown by Kubozono, Ito & Mester (2009), word-final /s/ is not borrowed as a geminate (miss → [misɯ], *[missɯ]). /ʃ/ is not borrowed as a geminate when the epenthesized vowel is [i] (brush → [buraɕi], *[buraɕɕi]), but it is borrowed as a geminate when the epenthesized vowel is [ɯ] (dash → [daɕɕɯ], *[daɕɯ]).

Also, the current result sheds new light on the distinction discussed in (1). While both groups retain their syllabicity, (1a) can be considered as a group of sounds in which the boundary between the onset fricative consonant and the following fricative vowel is indistinguishable, and (1b) can be considered as a group in which such a boundary is clear. Moreover, the conclusion that we reached above—that the perception of syllable weight (light vs. heavy) precedes whether that syllable contains a long vowel or a geminate—offers an important insight into the C/D model. That is, the inputs to the C/D model should be syllables with weight distinctions; i.e. they should have moras. Although the relationships between syllables and moras in Japanese are controversial (Kubozono & Homma 2002; Labrune 2012; Kawahara 2016), this paper proposes to use moras as units of syllable weight (section 5.3.)

5. General discussion: Inputs to the C/D model

5.1. Summary: Devoiced vowels and syllable structure

We have examined the articulatory, acoustic and perceptual characteristics of devoiced vowels following alveolar consonants in Japanese. In (3) I offer a summary of what has been discussed and concluded. I will reconsider these observations from the perspective of the C/D model, and develop a discussion on the C/D model, focusing on how the model should deal with vowel devoicing.

(3) a. Devoiced vowels are allophones of vowels, which are best characterized as fricative vowels. /ɯ/'s fricative vowel allophone is similar to /s/, while /i/'s fricative vowel allophone is similar to /ɕ/.

b. As Fujimura (2007) argues, the input to the C/D model should be syllables rather than phonemes. Syllables are comprised of a set of a consonant and a vowel. Henceforth, this paper refers to this set as a "syllable set".

c. A syllable set contains mora as units of syllable weight, and a set of syllable sets is represented as a list.

5.2. Feature correspondence and underspecification

Let us begin with the discussion of what elements a syllable set can contain, or put differently, how the characteristics of segments should be expressed in the C/D model. As stated in section 1, this question is an abstract problem that concerns both phonetics and phonology. I believe that the C/D model should capture two aspects of languages: (i) a binary opposition (or a feature system) that would capture the aspect of language as a system of differences, and (ii) unary features that would more directly reflect articulatory commands. Whether categories in language are clearly discrete, crisp entities or could involve more continuous distributions is a topic that is actively discussed in Cognitive Linguistics (Tayler 2004; Jaeger 1980; Miller 2006). The philosophy of the C/D model forces us to use features that are able to capture both aspects. One of the notions that satisfy this condition is "phonological particle/element". This paper thus assumes that syllables in the C/D model can be represented by a set of phonological particles. However, as long as syllables can be represented with unary categories with internal features, it is not necessary to use phonological particles.

Charette (1991) argues that phonological particles have a bundle of distinctive features, each of which is either hot or cold. Moreover, each segment has a head element or a dependent/operator element. (4) and (5) show distinctive features of representative particles, as well as how Japanese segments should be represented by sets of these particles. Hot features, which are strong, are shown in bold, and head features are underlined. When phonological particles are combined, head and hot features are the strongest, non-head hot features next, head cold features next, and non-head cold features are the weakest. From these principles, we can deduce the binary features of each segment as in (6), and we can relate the input information to the C/D model with binary distinctive features. In the current model, [±palatal] (or [±front]) is used instead of [∓back] to express tongue position. There are several reasons for this postulation: (i) it is not clear whether Japanese [ɯ] is back, (ii) [i] keeps its frontness even when it is devoiced (Figure 5), and (iii) front vowels are prohibited after [j] or palatalized consonants, which implies the existence of OCP(palatal/front).

(4) `A` (openness) = [-palatal, **-high**, +low…]
`I` (frontness) = [**+palatal**, +high, -low…]

U (w-ness) = [+labial, +cont, -palatal, +high…]
@(soft palate) = [-palatal, +dorsal…]
R (coronality) = [**+coronal**, +voiced…]
N (voiced, nasality) = [**+voice**, +nasal, -noise…]
? (closure) = [-voice, **-cont**, -nasal,…]
h (frication) = [-voice, **+noise**, +s.g.,…]…

(5) [u]={U}, [i]={I}, [e]={A, I}, [t]={R, ?}, [d]={N, R, ?}, [n]={N, R}, [s]={R, h}, [z]={h, R, N}, [ts]={R, h, ?}, [ɕ]={R, I, h}, [ç]={h, I},…

(6)

dep	head	*segment*	dep	head	*segment*
{A,	I} →	[e]	{R,	?} →	[t]
-palatal	**+palatal**	**+palatal**	+voice	-voice	-voice
-high	+high	**-high**	**+coronal**		**+coronal**
+low	-low	-low	+cont	**-cont**	**-cont**
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

The strength of binary features is relevant only when we calculate the combination of phonological particles, and what matters in matching phonological features is whether distinctive feature specifications agree or not. For example, given the consonant [ɕ], which is [-voice, +cor, +noise] and the vowel, which is [+palatal, +high…], the input to the C/D model should have a syllable set that has all these binary features, i.e., {R, h, I} or {R, I, h}. Similarly, for the [sɯ] syllable (consonant: [-voice, +cor, +noise,…]; vowel: [-palatal, +high,…]), the inputs to the C/D model should have either {R, h, U} or {R, h}. This is because since [ɯ] has [-palatal], it can correspond to either U or R. On the other hand, the [ɕɯ] syllable has a consonant that is [-voice, +cor, +palatal, +noise] and a vowel [-palatal, +high,…]. The entire syllable cannot share the same [palatal] specification, and therefore the syllable set can only be {R, I, h, U}. How this sort of input specification bears on the generation of C/D model diagrams will be discussed in section 5.5.

5.3. Inputs to the C/D model: Expressing syllable sets with internal moras

Next let us reconsider (3) in the light of the C/D model. Mathematically speaking, sets can consist of simple elements like {a, b…}, but sets can also take sets as their elements, as in {{a}, {a,b}…}. Unless we are talking about multiple sets and focusing on elements, we do not distinguish {a, b…} and {a, a, b,…}. However, {{a}, {b}} and {{a}, {a, b}} are distinct sets. Similarly, let us assume that syllable sets for the input to the C/D model can take sets of phonological particles as their elements. These sets, which serve as elements of syllable sets, correspond to a mora (henceforth referred to as "mora sets").

For example, take the word [kambi]. The initial syllable [kam] can have as its input {{low, stop, dorsal}, {nasal}}, or {{A, ?, @}, {N}}. {A, ?, @} is a set, and therefore, the elements within it are unordered. Nevertheless, the vocalic particle A determines the global articulation of the syllable, whereas ? and @ are consonantal features that impose only local effects. As shown in Figure 1, consonants realize themselves with Impulse Response Functions (IRFs), at the left edge of a syllable triangle, which itself is determined based on the syllable pulse. As a result, the order between the onset consonant and the vowel naturally follows. Similarly, the mora sets {$\underline{\texttt{A}}$, $\underline{\texttt{?}}$, @} and {N} are elements within the syllable set, and hence they are unordered. However, since the mora set containing A offers a vocalic, global baseline, and whereas the mora set {N} contains only a consonantal element, the latter can be distributed at the right edge of the syllable triangle, hence realizing itself as a coda nasal.

In English, unlike Japanese, onset consonants are dominated directly by a syllable node, whereas vowels and coda consonants are represented as mora sets. For example, the input representation of the word "can" ([kæn] should be {@,?,{I,A},{R,N}}. For this case, there are two consonant elements ? and N, with the latter contained in a mora set: in addition, the former is realized as onset and the latter as coda. To provide another example to illustrate the difference between Japanese and English, the same syllable [bi] is represented as {{$\underline{\texttt{I}}$, $\underline{\texttt{?}}$, N, U}} in Japanese and {{$\underline{\texttt{I}}$}, $\underline{\texttt{?}}$, N, U}} in English. The difference is whether an onset consonant is contained in a mora set (as in Japanese) or not (as in English).

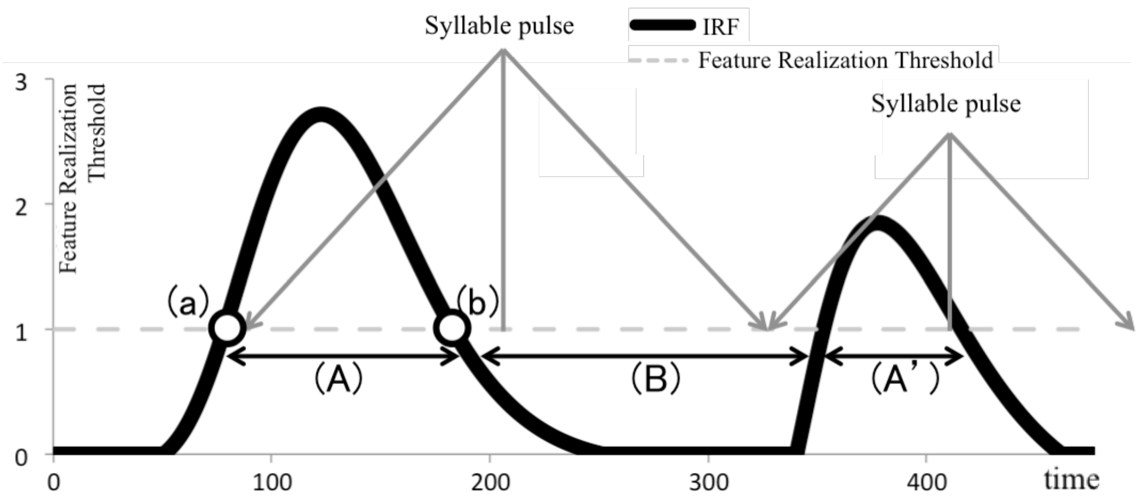
By representing syllable sets with mora sets, we can also explicitly specify the syllable weight, which would influence the generation of syllable triangles. For example, syllable sets containing a one mora set, such as {{a, b…}} and {a, b,…{c, d}…}} are light syllables, and those containing two mora sets such as {{a, b,…},{c, d,…}} and {a, b,…{c, d,..},{e, f,…}} are heavy syllables. This way we can eliminate the need to specify onset/coda distinctions, as done in the original C/D model with superscripted ^{O} and ^{C} (except that we still need to specify p-fix and s-fix). For example, in Japanese, elements that are contained within a mora set with a vocalic element are realized as onset, and other consonantal elements are realized as coda. In English, consonantal elements that are directly contained in a syllable set are realized as onset consonants, and those that are contained in mora elements are realized as coda consonants.

Japanese coda consonants (coda nasals and geminates) are interpreted as "syllable concatenators" in the C/D model (Fujimura & Williams 2008). In the current model, feature spreading across a syllable boundary serves as the actual commands for syllable concatenators. Following the HPSG notation, we can use tag expressions representing token similarity, $\boxed{1}$ and $\boxed{2}$. For example, [kakki] is represented as <{{$\underline{\texttt{A}}$, @, $\underline{\texttt{?}}$},{$\boxed{1}$}},{{$\underline{\texttt{I}}$,@,$\boxed{1}$$\underline{\texttt{?}}$}}>. In Japanese, these syllable concatenators are included within a mora set, coda nasals and geminates bear their own rhythm or duration, unlike ambisyllabic [t] in English.

Also introducing mora sets allows us to explain why in Japanese onset [k] is released whereas the first half of the geminate [k] is not released. More specifically, release is initiated by a vocalic element that co-exists with the consonantal element (in that sense, Japanese plosives are strictly speaking "stops" rather than "plosives", as they do not have inherent release). On the other hand, coda [k] does not have a vocalic element within the same mora set, and hence is not released; as a result, it is realized as having a longer consonantal closure. This nature of geminates in Japanese follows from the particle ?, which has [-voice, **-cont**,…] as its internal element. In the C/D model, the element ? not only creates oral closure via an IRF on a syllable diagram, but also it inhibits vocal fold vibration. This nature of ?, which inhibits vocal fold vibration, plays a fundamental role in causing vowel devoicing. The next section explores this IRF function using voice onset time (VOT) as an example.

5.4. The internal structure of voicelessness of obstruents and IRFs

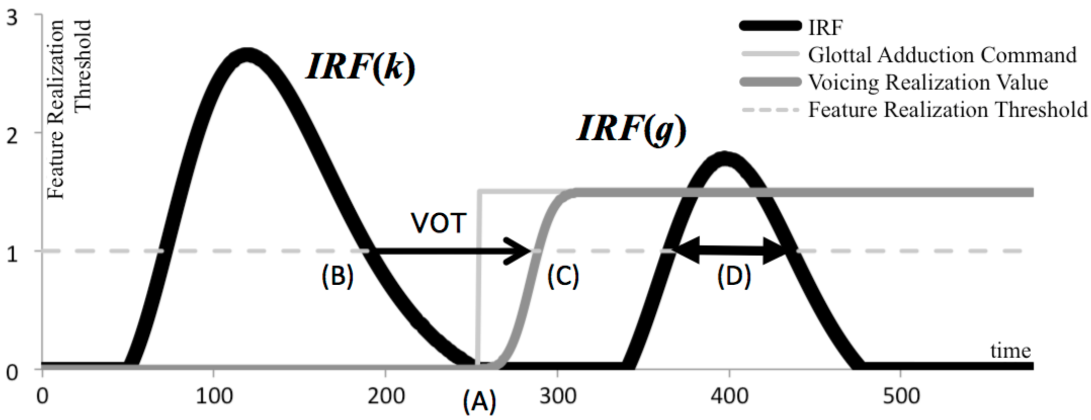
Languages differ in terms of VOT ranges that they use. In Japanese, voiced stops have negative or slightly positive VOT, and voiceless stops have fairly long VOT (Takada 2011). The IRF of the closure phonological particle ʔ heavily influences the quantitative nature of VOT. Let us consider C/D model diagrams of [kigi] "trees" and [giki] "feeling of obligation". Figure 12 shows a schematic illustration, in which two syllable pulses represent the positioning of the two syllables, and within them, the IRF of each stop is distributed. The C/D diagrams used in this paper are normalized in such a way that features have a "realization threshold"; any feature that is above 1 is phonetically realized. The points shown as (a) and (b) in Figure 12 represent the closure and release of the stop consonant, respectively. The interval (A) represents its closure interval, and the interval (B) includes the release, the formant transition, and the vocalic portion.



[xxx INSERT FIGURE 12 HERE]

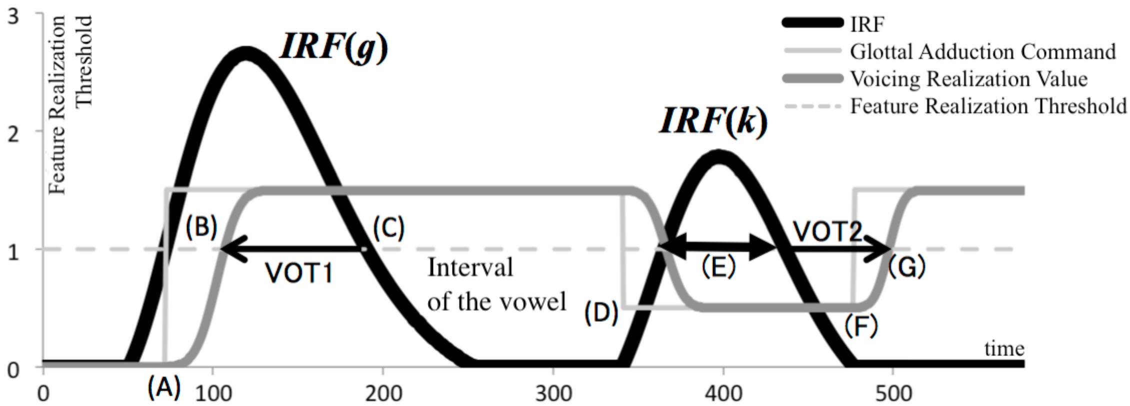
Figure 12: A C/D model diagram of syllable pulses and IRF for stops

The distinction between voiced/voiceless, or the control of vocal fold vibration behind that distinction, can be regulated with the IRF function. Figure 13 is a C/D model diagram of [kigi]. The closure element ʔ contained within the word-initial [k] has a [-voice] feature. This feature inhibits the glottal adduction command (the step function shown with the grey line in Figure 13) at the level of zero (which corresponds to glottal abduction). Only when this inhibition command ceases at the end of the IRF ((A) in Figure 13), does the glottal adduction become [+voice], which is shown in the figure by going over the realization threshold line. Since actual glottal vibration is delayed with respect to the glottal adduction command, we observe actual voicing at point (C). Thus, the interval between the consonantal release (B) and the realization of voicing (C) corresponds to [k]'s VOT. Since the consonantal release (B) must precede the end of the IRF (A), and since it thereby precedes (C), VOT for voiceless consonants must be positive. On the other hand, since the second consonant [g] inhibits the [-voice] feature of the closure element ʔ and also has N which encourages spontaneous voicing, the glottal adduction command never goes below 1 after (A). As a result, voicing continues starting with [i] following [k] through the closure of [g] (D), and the following vowel [i].



[xxx INSERT FIGURE 13 HERE]
Figure 13: A C/D model diagram of [kigi] (IRF and glottal command only)

By way of comparison, Figure 14 shows a C/D model diagram of [giki]. The element N of the word-initial [g] exerts active a glottal adduction command at the closure point (A). As a result, at the point (B) which precedes the release point (C), the voicing command goes above the threshold and negative VOT is achieved (shown as “VOT 1” in Figure 14). On the other hand, since the second consonant [k] does not have a voicing element, the closure element ? inhibits vocal fold vibration, and the glottal adduction command goes below 1 at the onset of [k]’s IRF (D). As a result, no voicing occurs during the closure interval of [k] (E), and [k] is realized as a voiceless consonant. The inhibition of voicing finishes at the end of [k]’s IRF (F), and hence voicing restarts at point (G). The interval between the end of (E) and (G) becomes [k]’s VOT (shown as “VOT 2” in Figure 14).



[xxx INSERT FIGURE 14 HERE]
Figure 14: A C/D model diagram of [giki] (IRF and glottal command only)

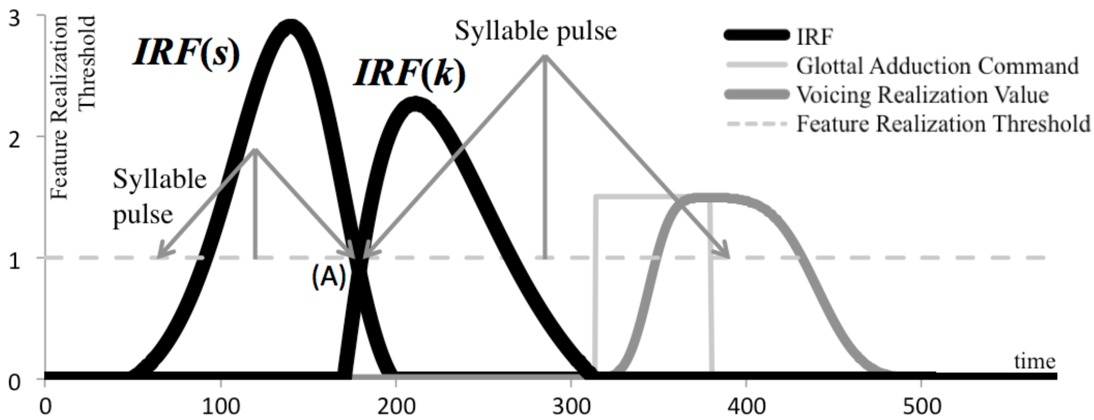
Note that VOT for [k] is shorter in Figure 14 than in Figure 13. This difference is caused by the difference in magnitude of the syllable pulses of [k]. In [giki], [k]’s IRF has a weak pulse, and results in shorter VOT (Fujimura 2002). The same logic applies to [g]. In Figure 14, as [g]’s IRF becomes weaker, the interval between (B) and (C) gets shorter, and the VOT of [g] gets closer to zero. However, since voiced obstruents have a voicing feature N, it can make a glottal adduction command during its IRF, resulting in a slightly positive IRF. To summarize, the feature [-voice] has elements ? and h, whose quantitative influences on the IRF can explain the range of VOT values that voiced and voiceless consonants show.

5.5. The effects of IRF causing vowel devoicing

The thesis that the elements ʔ and h contained in the [-voice] feature can directly inhibit vocal fold vibration sheds new light on the mechanism of vowel devoicing. This section thus explores those syllable diagrams that contain devoiced vowels. As stated in section 5.2., [sɯ], [ɕi] and [ɕɯ] should be represented as syllable sets containing mora sets in (7).

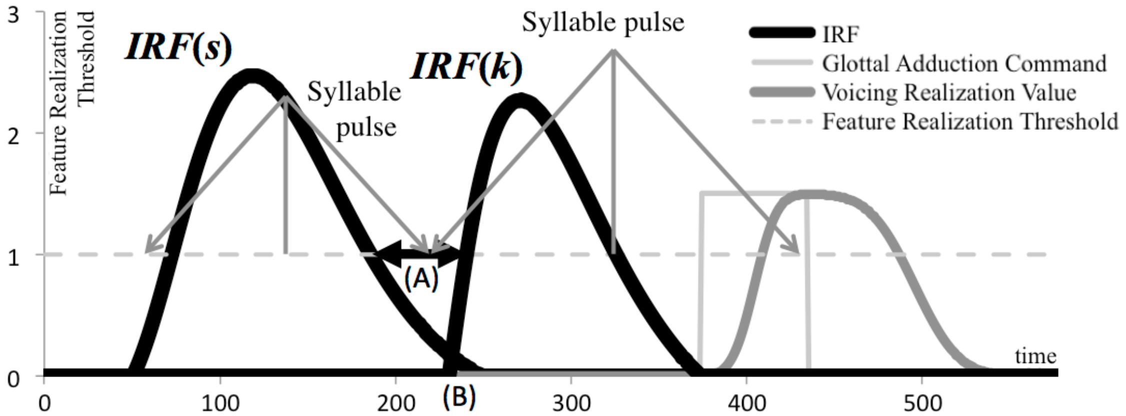
(7) a. [sɯ]: (i) {{R, h, U}} or (ii) {{R, h}}
b.[ɕi]: (i) {{R, h, I}} or (ii) {{R, I, h}}
c, [ɕɯ]: {{R, I, h, U} only

Among these, (7a-ii) and (7b-ii) are input representations in which the whole syllable is covered with frication. For those syllables, as shown in Figure 15, the IRF of the first syllable and the IRF of the second syllable necessarily overlap. As a result, even after point (A), the inhibition effect of ʔ and h continues, and therefore the glottal adduction command is made within the first syllable. In other words, the input representations for (7a-ii) and (7b-ii) result in obligatory vowel devoicing no matter what the speed of the utterance is.



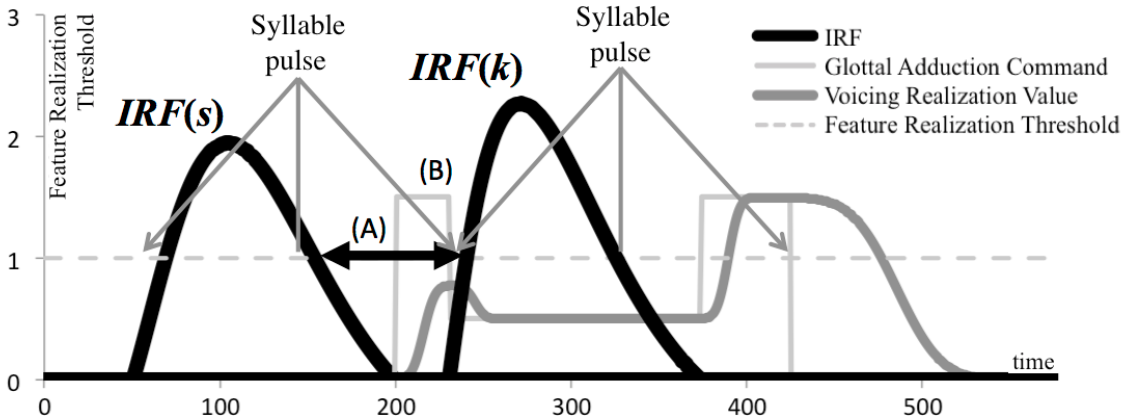
[xxx INSERT FIGURE 15 HERE]
Figure 15: A C/D model diagram of [sɯkɯ] with the input (7a-ii), or the input with an extremely weak pulse for the first syllable.

On the other hand, (7a-i), (7b-i) and (7c) have explicit vocalic elements in the input. Therefore, whether these vowels are devoiced or not depends on the nature of the IRF function of phonological elements ʔ and h. If a syllable pulse is weak, and hence the effect of the IRF becomes relatively strong, then the frication element covers the whole syllable, just as in Figure 15. Figure 16 shows a case in which the syllable pulse becomes slightly stronger, and hence the effects of the IRF become slightly weaker. Given this configuration, a transitional vocalic interval (A) arises in the initial syllable, but since the IRF's inhibition effect exerts itself at point (B), it results in a fricative vowel.



[xxx INSERT FIGURE 16 HERE]

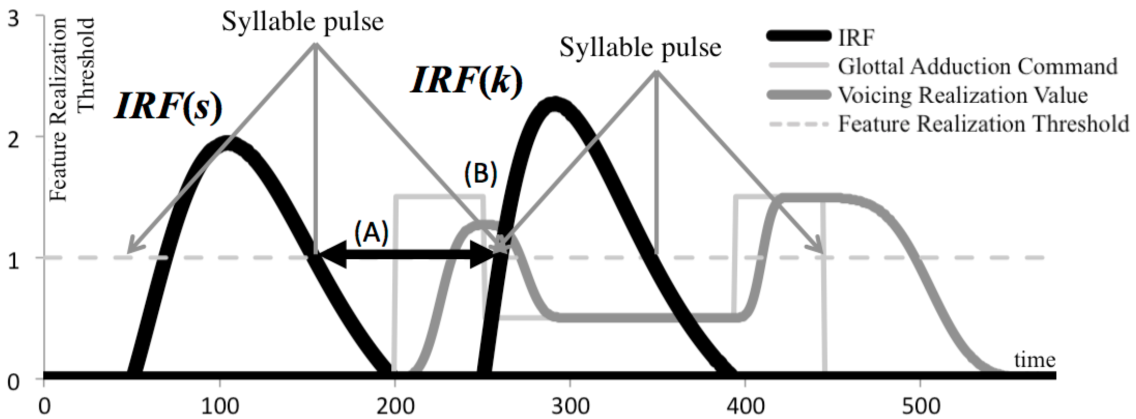
Figure 16: A C/D model diagram of [sɯkɯ] in which the syllable pulse of the initial syllable is weak.



[xxx INSERT FIGURE 17 HERE]

Figure 17: A C/D model diagram of [sɯkɯ] in which the syllable pulse of the initial syllable is slightly stronger.

When the syllable pulse gets even stronger, and the IRF's relative strength gets weaker, the glottal adduction command realizes itself at (B). However, since its time is so short,it does not reach its realization threshold. If the syllable pulse gets even stronger, as in Figure 18, the glottal adduction command (B) goes over the threshold, and the vowel is no longer devoiced. This variation concerning vowel devoicing is compatible with the observation by Fujimoto and Kiritani (2003) and Fujimoto (2004).

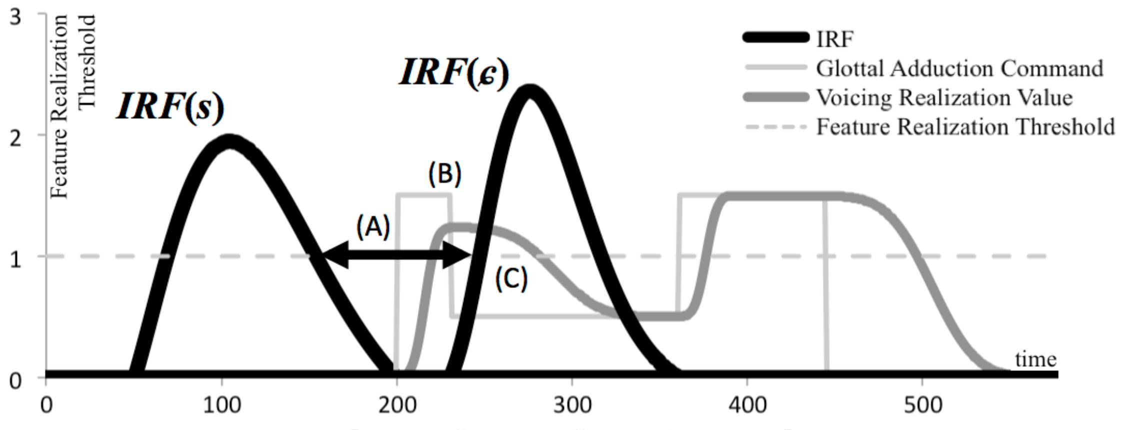


[xxx INSERT FIGURE 18 HERE]
Figure 18: A C/D model diagram of [sɯkɯ] in which the syllable pulse of the initial syllable is strong.

5.6. Differences in inhibitory effects between ʔ and h

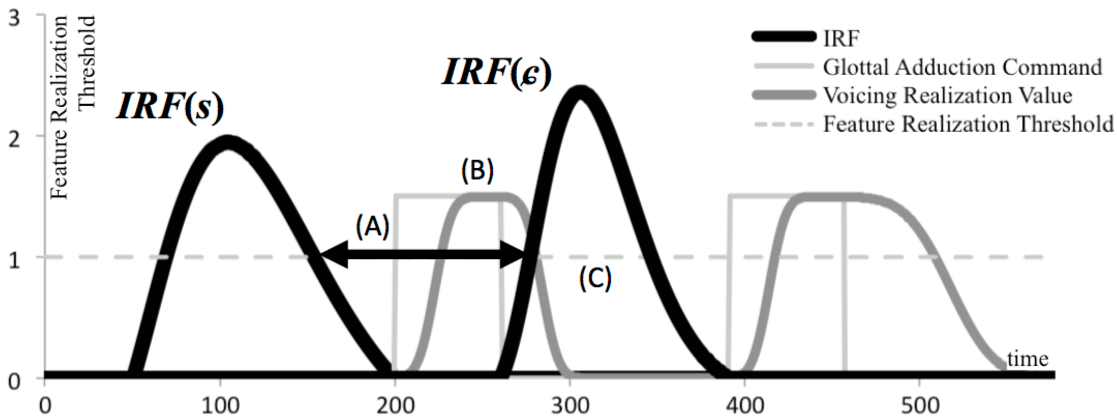
Let us further explore the potential differences in inhibitory effects between ʔ and h. Yoshida (2002, 2006) points out that the rate of vowel devoicing goes lower when followed by voiceless fricatives. We can attribute this observation to the inherent differences in inhibitory effects between ʔ and h. That is, the element ʔ makes a complete closure in the oral cavity, thereby raising the intraoral air-pressure, resulting in stronger inhibitory force. On the other hand, the fricative element h results in a weaker rise in the intraoral air-pressure, resulting in a weaker inhibitory force.

Thanks to its nature, the element ʔ can realize voicelessness by setting the glottal adduction command to 0.5, which is larger than the maximum abduction command (=0) but is smaller than the voicing realization threshold (=1). In Figures 17 and 18, the reason that the consonants in the second syllables are realized as voiceless, despite the glottal adduction command being set to 0.5, is attributed to the strong inhibitory effect of voicing due to ʔ. On the other hand, if the consonant in the second syllable is a fricative, given the same articulatory setting as in Figure 17, it would result in weak inhibitory effects, as in Figure 19. In this configuration, at point (B), the voicing realization command is not reduced, resulting in a short interval of voicing. Overall, given this configuration, merely lowering the glottal adduction command to 0.5 does not result in vowel devoicing at point (C), despite the fact that the following fricative does not have a voicing element N. Neither does it result in a completely voiceless fricative.



[xxx INSERT FIGURE 19 HERE]
Figure 19: A C/D model diagram in which the consonant in the second syllable is a fricative, and the glottal adduction command is set to 0.5.

Therefore, in order to realize the following fricative as voiceless, the glottal adduction command must be set to 0 (=complete glottal abduction). However, the articulatory plan in which the glottal adduction command changes from 0 to 1 and from 1 to 0 runs into a danger of undershooting, and hence should be avoided. In order to lower the glottal adduction command to 0, it is desirable that the vowel interval be sufficiently long, and to that end, a strong syllable pulse is required. As a result, a vowel following a fricative sometimes fails to undergo devoicing, as in Figure 20.



[xxx INSERT FIGURE 20 HERE]

Figure 20: A C/D model diagram in which the consonant in the second syllable is a fricative, and the glottal adduction command is set to 0.

This C/D model-based explanation of vowel devoicing seems compatible with the detailed physiological experiment offered in Fujimoto (2005). According to Fujimoto (2005), in cases of vowel devoicing with following voiceless stops, we do not observe glottal abduction during the voiceless stops. On the other hand, voiceless fricatives show clear glottal abduction gesture. These observations are compatible with the modeling in Figures 17 and 18 in which the IRF of the second syllable has the glottal adduction command of only 0.5, whereas in Figure 18, the IRF of the second syllable has the glottal adduction command up to 0.

Finally, the inhibitory effects of IRF are influenced not only by the differences in phonological particles, but also by the base function defined based on the nuclei of the syllable. For example, in mid and low vowels, the oral cavity is wide-open, which means that IRFs are superimposed on the base function where voicing inhibitory effects are weakened and vowel durations are long. This offers an explanation of why vowel devoicing occurs mainly in high vowels, although we cannot leave out perceptual factors to account for the lack of devoicing in non-high vowels.

5.7. Treatments of allophones in the C/D model and remaining issues

In (8) I offer a summary of how vowel devoicing can be modeled in the C/D model. The more controversial one is (8a), because it is not compatible with the idea that "all allophonic variations are created at the phonetic implementation level" (Sproat & Fujimura 1993). Fujimura (2007) and Pinter (2008) argue that affrication of /t/ preceding high vowels in Japanese can be attributed to phonetic implementation effects, which are grounded in their articulation. Therefore, the thesis (8a) should be carefully reexamined in light of the C/D model's claim.

(8) Devoiced vowels are allophones of vowels, which are fricative vowels. There are two types:

a. Variation which has a root in input information, as in (7a-ii) and (7b-ii).

b. Variation which arises during the phonetic implementation stage due to interaction with IRFs and syllable pulses.

In Tokyo Japanese, I believe that vowel devoicing caused by (8a) is limited to a part of those in (1a). (1b) corresponds to (8b), and even among those in (1a), devoicing in [ɕi] and [çi], which retains vocalic articulation (recall Figure 5), can be explained in terms of (8b). Moreover, prevention of vowel devoicing due to following voiceless fricatives was previously attributed to OCP (spread glottis) by Tsuchida (1997), but it seems that we can explain away this observation as a result of (8b).

However, as for devoicing of [sɯ], the entire syllable is covered by frication noise, and there seems no vocalic information of [ɯ]. Therefore, in order to explain this devoicing pattern within the framework of the C/D model, in which consonantal effects are localized modelled via IRFs, requires some additional possibilities, such as those in (9). If (9a) is correct, we can maintain the thesis by Sproat & Fujimura (1993), but if (9b) is correct, that thesis may need to be reconsidered.

(9) a. /ɯ/ is phonologically underspecified. Therefore, phonology does not need to delete the information of [ɯ]. The input to the C/D model is (7a-ii) after all.

b. Phonological features are fully specified at the input level. Therefore, phonology has to derive (7a-ii) by deleting the vowel [ɯ].

This issue is related to several theoretical problems, such as phonological underspecification (Archangeli 1988), Richness of the Base in Optimality Theory (Tesar & Smolensky 2000) and lexical optimization (Ito, Mester & Padgett 1995), the nature of phonological representations, and the general question of what phonology should deal with. The question also remains regarding “grammaticalized devoicing” (Maekawa 1989) and intentional inhibition of vowel devoicing (Martin, Utsugi & Mazuka 2012). These issues are however beyond the scope of the current paper. Other issues regarding the input of the C/D model include the structure of phonological particles that consist of segments (Nasukawa 2005; Nasukawa & Phillip 2009) and more specifically, the nature of IRFs for each phonological particle, the treatment of p-fix and s-fix, the relationship between vowel devoicing and accent, and vowel devoicing and syllabification (Mimatsu, Fukumori, Sugai, Utsugi, Shimada 1999; Hirayama 2009; Kondo 1997). At any rate, to close this paper, I would like to emphasize that the C/D model is a framework that allows us to think about important and general issues in phonological theory.

Acknowledgments
I am grateful to Dr. Ichiro Yamamoto for his help with EPG, Dr. Osamu Fujimura for the discussion about the C/D model, Dr. Haruo Kubozono, Dr. Shigeto Kawahara, Dr. Mieko Takada, Dr. Kuniya Nasukawa, Dr. Toshio Matsuura, and Dr. Natsuya Yoshida for the discussion on various topics on the phonetics and phonology of voicing in Japanese. I am also grateful to two anonymous reviewers for the detailed comments that they provided on the previous version of the paper. This paper is supported by MEXT/JSPS KAKENHI Grant Number JP.26370467

References

Archangeli, D. (1988) "Aspects of underspecification theory." *Phonology*, 5, 183–207.

Baba, R. (1998) Museika boin-wa masatsuon-dearu. *Kumamotokenritsudaigaku Bungakubu Kiyoo* 50,1–22.

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

Charette, M. (1991) *Conditions on phonological government*. Cambridge University Press.

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

Fujimoto, M. & S. Kiritani (2003) Tookyoo hoogen-to oosaka hoogen-niokeru boin-no museika-no hikaki. *Journal of the Phonetic Society of Japan* 7: 1,58‐69.

Fujimoto, M. (2004) Boinchoo-to boinmuseika-tono kankee: tookyoo hoogen washa-to oosaka hoogenwasha-no hikaku. *Kokugogaku* 55: 4, 2‐15.

Fujimoto, M. (2005) Boin museikaji-no kootoochoosetsu: museika-no sukunai oosaka hoogenwasha-no baai. *Journal of the Phonetic Society of Japan* 9: 1,50‐59.

Fujimura, O. (1992) "Phonology and phonetics — a syllable-based model of articulatory organization." The Acoustical Society of Japan (E), 13, 39–48.

Fujimura, O. (2002) "Temporal organization of speech utterance: A C/D model perspective." Cadernos de Estudos Linguisticos, Instituto de Estudos da Linguagem, 43, 9–36.

Fujimura, O. (2007) *Onseikagaku genron: gengo-no honshitsu-o kangaeru*. Tokyo: Iwanami Shoten .

Fujimura, O., & Williams, J. C. (2008) "Prosody and syllables." *Phonological Studies*, 11, 65‐74.

Goad, H., & Shimada, A. (2014) "In some languages, /s/ is a vowel." *Proceedings of the 2013 Meeting on Phonology*.

Gunji, T. (1994) *Shizengengo*. Nihon hyooronsha.

Hirata, Y. (2007) "Durational variability and invariance in Japanese stop quantity distribution: Roles of adjacent vowels." *Journal of the Phonetic Society of Japan*, 11, 9–22.

Hirayama, M. (2009). Postlexical prosodic structure and vowel devoicing in Japanese. Ph. D. dissertation, University of Toronto.

Ito, J., Mester, A. & Padgett, J. (1995) "Licensing and underspecification in Optimality Theory." *Linguistic Inquiry*, 17, 49–73.

Jaeger, J. J. (1980). Categorization in phonology: An experimental approach. Ph. D. dissertation, University of California, Barkeley.

Kawahara, S. (2015) "The C/D model as a theory of the phonetics-phonology interface." *Journal of the Phonetic Society of Japan*, 19, 33-43.

Kawahara, S. (2016) Japanese has syllables: A reply to Labrune (2012). *Phonology* 33, 169-194.

Kawakami, S. (1977) *Nihongo onsei gaisetsu*. Oh-huusha.

Kitahara, M. (2008) Shohyoo: *Fujimura Osamu cho "Onseikagaku genron" Iwanami shoten 2007*. Journal of the Phonetic Society of Japan 12: 1, 63‐69.

Kubozono, H. & T. Homma (2002) *Onsetsu-to moora*. Kenkyuusha.

Kondo, M. (1997). Mechanisms of vowel devoicing in Japanese. Ph. D. dissertation, the Universityof Edinburgh.

Kubozono, H., Ito, J. & Mester, A. (2009) "Consonant gemination in Japanese loanword phonology." *Current issues in unity and diversity of languages* (pp.953–973).

Labrune, L. (2012) “Questioning the universality of the syllable: Evidence from Japanese.” *Phonology*, 29, 113–152.

Maekawa, K. (1998) Boin-no museika. M. Sugito (ed.) *Kooza: Nihongo-to nihongokyooiku*, vol 2. Tokyo: Meiji Shoin.

Martin, A., Utsugi, A. & Mazuka, R. (2014) “The multidimensional nature of hyperspeech: Evidence from Japanese vowel devoicing.” *Cognition*, 132, 216–228.

Matsui, M. (2013) Masatsuon-no yuuseisei-to museikaboin-no chikakuteki tegakari-ni tsuite. *Proceedings of 27th meeting of the Phonetic Society of Japan*, 35-40.

Miller, J. L. (2006) “On the internal structure of phonetic categories: A progress report.” *Cognition*, 50, 271–285.

Mimatsu, K., Hukumori, T., Sugai, Y. Utsugi, A., Shimada, T. (1999) Nihongo-no boin-no museika-ni tsuite: tookyoohougen-no 2 onsetsu renzoku museika-no onkyoo bunseki. *Ippan Gengogaku Ronkoo* 2,73 - 101.

Nakamura, M. (2003) “The articulation of vowel devoicing.” *Phonological Studies*, 6, 49 - 58.

Nasukawa, K. (2005) *A unified approach to nasality and voicing*. Berlin: Mouton de Gruyter.

Nasukawa, K., & Phillip, B. (2009) Strength relations in phonology. Berlin: Mouton de Gruyter.

Öhman,S.(1966)“Coarticulation in VCV utterances: spectrographic measurements.” *Journal of Acoustical Society of America*, 39, 151–168.

Pinter, G. (2008). Asymmetrical segment distributions in Japanese. Ph. D. dissertation, Kobe University.

Pollard, C. J., & Sag, I. A. (1994) *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Saussure, F. d. (1916) *Course in general linguistics*. New York: McGraw.

Soli, S. D. (1981) “Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation.” *Journal of Acoustical Society of America*, 70, 976–984.

Sproat, R., & Fujimura, O. (1993) “Allophonic variation in English /l/ and its implications for phonetic implementation.” *Journal of Phonetics*, 21, 291–311.

Takada, M. (2011) *Nihongo-no gotooheisaon-no kenkyuu*. Tokyo: Kuroshio Publisher.

Tayler, J. R. (2004) *Linguistic Categorization*. Cambridge: Oxford University Press.

Tesar, B., & Smolensky, P. (2000) *Learnability in Optimality Theory*. The MIT Press.

Tsuchida, A. (1994) “Fricative-vowel coarticulation in Japanese devoiced syllables: Acoustic and perceptual evidence.” *Working paper of the Cornell Phonetics Laboratory*, 9, 183–222.

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

Whalen, D. H. (1983) “Vowel information in postvocalic fricative noises.” *Language and Speech*, 26, 91–100.

Yoshida, N. Onseikankyoo-ga boin-no museika-ni ataeru eikyoo-ni tsuite. *Kokugogaku* 53: 3,34–47.

Yoshida N. (2006) Nihongo boin museika-no onseigakuteki kenkyuu. *Phonological studies* 9,173 - 180.