문학석사 학위논문

Aggressive Reduplication in Japanese high vowel devoicing

일본어 고모음 무성음화에서 나타나는 적극 중첩

2023년 2월

서울대학교 대학원

언어학과 언어학전공

변 한 영

Aggressive Reduplication in Japanese high vowel devoicing

지도 교수 전 종 호

이 논문을 문학석사 학위논문으로 제출함
2023년 1월

서울대학교 대학원
언어학과 언어학전공
변 한 영

변한영의 문학석사 학위논문을 인준함
2023년 1월

위 원 장 ______________________ (인)

부위원장 ______________________ (인)

위 원 ______________________ (인)

Aggressive Reduplication in Japanese high vowel devoicing

by

Hanyoung Byun

Submitted in partial fulfillment
of the requirements for the degree of
Master of Arts
at
Seoul National University
February 2023

Thesis Supervisor ___________________
Jongho Jun

Thesis Committee ___________________
Chairman, Ho-Young Lee

____
James Whang

Abstract

Aggressive Reduplication in Japanese high vowel devoicing

Byun, Hanyoung
Department of Linguistics
The Graduate School
Seoul National University

In this study, I investigate the phonological factors that contribute to the variable pattern of Japanese high vowel devoicing. High vowel devoicing, whereby high vowels lose voicing between voiceless consonants, has often been considered to be a phonetic process; it has been claimed that this process occurs because voicing of high vowels fails to be achieved due to an overlap of the glottal opening gestures of the surrounding voiceless consonants (Jun et al. 1998). Such an account assumes that the application of high vowel devoicing is solely conditioned by the articulatory characteristics of the surrounding consonants, and the temporal organization of the target vowel with the surrounding consonants. As such, it predicts that Japanese high vowel devoicing will be insensitive to the phonological structure of Japanese.

Contrary to this prediction, the current study shows that the structural knowledge of similarity between adjacent syllables plays an important role in

the application of high vowel devoicing. Using a large-scale dataset from a Japanese speech corpus (Corpus of Spontaneous Japanese; Maekawa et al. 2000), I report the following phonological tendencies that affect the rate of high vowel devoicing:

(i) Devoicing is less likely if the target vowel is preceded by a fricative or affricate and followed by a fricative. (*Matching manner condition*)
(ii) In the matching manner condition, devoicing is even less likely if the syllable following the target vowel contains another high vowel. (*Matching height condition*)
(iii) Devoicing is less likely if the target vowel is accented.
(iv) Devoicing is less likely if the target vowel is followed by a geminate.
(v) In the environment where the devoicing context occurs consecutively, devoicing of two vowels in adjacent syllables tends to be avoided.

I then provide a formal analysis of the observed tendencies within the framework of Optimality Theory (Prince and Smolensky 1993/2004). First, I claim that devoicing is derived by a typical ranking of constraints for allophonic variation: Context-sensitive Markedness (DEVOICE: "No voiced short high vowel between voiceless consonants") >> Context-free Markedness (SONVOI: "No voiceless sonorant") >> Faithfulness (IO-ID(vce): "No change of the input [voice] in the output"). Additionally, I propose OCP-V̥ ("No voiceless vowels in adjacent syllables") to account for the tendency of consecutive devoicing to be avoided.

The matching manner and matching height conditions among the tendencies above can be generalized by making reference to the syllable structure. /$C_1V_1C_2V_2$/ sequences, where V_1 is a target high vowel and C_1 and C_2 are voiceless consonants, are parsed into $[C_1V_1]_{\sigma 1}[C_2V_2]_{\sigma 2}$ in Japanese. Given this, it can be said that devoicing is suppressed if the onsets of the adjacent syllables (i.e., C_1 and C_2) match in [+continuant] (e.g., ***suso*** 'hem'). Affricates show a position-specific behavior, patterning with fricatives in prevocalic position (i.e., C_1) as [+continuant], and with stops in postvocalic position (i.e., C_2) as [−continuant]. In addition to the matching manner condition, devoicing is further suppressed if the height of the nuclei (i.e., V_1 and V_2) agrees in [high] (e.g., ***susi*** 'sushi'). This additive effect of the matching height condition to the matching manner condition suggests that devoicing rates decrease as the degree of similarity between two adjacent syllables increases.

I claim that these similarity-driven blocking effects arise due to an effort to preserve the voicing identity between the vowels in adjacent, self-similar syllables. To formalize this claim, I provide an analysis adopting Zuraw's (2002) Aggressive Reduplication. As much as reduplicative identity is argued to be driven by correspondence between a base and reduplicant (McCarthy and Prince 1995), the Aggressive Reduplication account proposes that correspondence between word-internal substrings is imposed by the constraint REDUP, and correspondence constraints (*κκ*-CORR) that are invoked by REDUP prevent any disruption of self-similarity between the correspondent strings. Under this account, the similarity-driven blocking effects are explained by the effects of correspondence constraints operating between adjacent syllables, which disprefer a mismatch of [voice] between the correspondent vowels when the target vowel devoices. In addition, the ranking of *κκ*-CORR over REDUP derives the results where correspondence structure is only posited in self-similar syllables, and thus blocks devoicing only in those environments. I discuss alternative similarity-related theories such as the Obligatory Contour Principle (McCarthy 1986) and Agreement by Correspondence (Rose and Walker 2004), and argue that only Aggressive Reduplication successfully accounts for the similarity-driven effects in Japanese high vowel devoicing, since it allows correspondence beyond the segmental level.

On top of the above constraints, I provide additional constraints for the effects of pitch accent and geminacy. Based on the psycholinguistic and perceptual salience of accented syllables, I claim that pitch accent impedes devoicing since accented syllables require greater positional faithfulness (IO-ID(vce)/σ́; "Do not change the input [voice] of an accented syllable in the output", following Beckman 1998). I further claim that the inhibitory effect of geminates is based on pre-geminate vowel lengthening in Japanese (Kawahara 2015). Based on the P-map hypothesis (Steriade 2001, 2009), I assume that the perceptual difference in voicing between a lengthened vowel and its voiceless counterpart is greater than that between a short vowel and its

voiceless counterpart. As such, I propose another positional faithfulness constraint, IO-Id(vce)/_Gem, which bans a voicing change of a vowel before a geminate.

Finally, to account for variation, I employ Anttila's (1997) Partially Ordered Constraints approach. Based on this, I assume some parts of the constraint ranking are fixed, while others can change at each production.

The current study finds that structural knowledge such as similarity relations across syllables plays a crucial role in Japanese high vowel devoicing, which has often been treated as a phonetic or post-lexical process. This suggests that Japanese speakers are sensitive to the (dis)similarity of vowel voicing, even when this information is purely allophonic.

Keywords: Japanese high vowel devoicing, Aggressive Reduplication, similarity, correspondence, variation

Student number: 2018-29432

Table of Contents

1. Introduction

Phonology and phonetics are often assumed to be distinct components of grammar, governed by independently motivated principles. While phonological processes are claimed to be regulated by structural knowledge of symbolic representations, phonetic processes are claimed to be determined by the articulatory properties or the temporal organizations of the gestures involved.[1] Under this assumption, phonological and phonetic processes are qualitatively different, and the principles governing one part of the grammar cannot affect the other part.

For example, Davidson (2006) claims that English pretonic schwa elision (e.g., *potato → [pt]ato*) is a process that results from an overlap of two surrounding consonantal gestures, rather than a phonological deletion of the underlying vowel. This claim is corroborated by the observation that the outcome of schwa elision is acoustically different from that of a phonological deletion, and that schwa elision is insensitive to the phonotactic legality of English. Therefore, the pattern of English schwa elision seems to support the division between phonetics and phonology.

Japanese high vowel devoicing is another case that is often analyzed from a phonetic point of view. In many varieties of Japanese, including Tokyo

---

[1] There are explicit arguments against such a distinction. For example, Articulatory Phonology (Browman and Goldstein 1992) rejects the featural representations of phonological grammar and argue that the basic units of phonological grammar are articulatory gestures. On the other hand, Flemming (2001) proposes a unified model of phonetics and phonology, in which phonetic and phonological representations are incorporated into auditory dimensions based on a constraint-based grammar. See also Hayes, Kirchner and Steriade (2004) for a review of the relevant literature.

Japanese, there is an allophonic process whereby short high vowels, /i, u/, undergo devoicing between voiceless consonants, e.g., /sika/ → [ʃi̥ka] ‘deer’. This process is gradient, such that many linguistic factors affect its occurrence (see Fujimoto 2015 for an overview). Although traditional literature (e.g., McCawley 1968) has described high vowel devoicing as a rule-based phonological process, as illustrated in (1), recent studies such as Jun et al. (1998) and Beckman (1996) argue that high vowel devoicing is not a phonological process at all; it results from a failure to achieve the voicing of high vowels between the glottal opening gestures of the surrounding voiceless consonants, which is incorporated under the general principle of *gestural overlap*.

(1) Japanese high vowel devoicing rule

$V_{[+\text{high}, -\text{long}]} \rightarrow [-\text{voice}] \,/\, C_{[-\text{voice}]}\_\_C_{[-\text{voice}]}$

Under the gestural overlap account, high vowel devoicing does not involve a change of the underlying [voice] feature of the vowel. Rather, the application of high vowel devoicing is claimed to be solely determined by the articulatory characteristics of the surrounding consonants, and the temporal organization of the target vowel with the surrounding consonants. As such, Japanese high vowel devoicing is predicted to be insensitive to the phonological structure of Japanese, as with the case of English schwa elision.

However, this view leaves some crucial aspects of Japanese high vowel devoicing unexplained. For example, as reported by several studies (Yoshida 2002, Fujimoto 2005, Maekawa and Kikuchi 2005), devoicing is variably blocked when the target vowel is surrounded by fricatives (e.g., *susi* ‘sushi’). This seems to indicate that phonological similarity plays a role in the application of high vowel devoicing, which is not easily explained by the gestural overlap account. Furthermore, the vowel height of the following syllable is reported to affect the rate of devoicing (i.e., the difference between *sus**o*** ‘hem’ vs. *sus**i*** ‘sushi’; Yoshida 2002, Byun 2012), which the gestural overlap account cannot easily explain, since the laryngeal gesture of the following vowel is not immediately adjacent to that of the target vowel. Therefore, a phonetic analysis of vowel devoicing does not offer a complete explanation of the observed patterns.

In light of these issues, this study aims to thoroughly investigate the phonological tendencies observed in the variable pattern of Japanese high vowel devoicing and to provide an analysis of such tendencies in the framework of Optimality Theory (OT; Prince and Smolensky 1993/2004). Using quantitative data from the Corpus of Spontaneous Japanese (CSJ; Maekawa et al. 2000), I provide an in-depth discussion of the phonological tendencies reported by previous studies, ultimately claiming that the structural knowledge of similarity between adjacent syllables affects the occurrence of Japanese high vowel devoicing. Specifically, I show that devoicing is suppressed if the syllable containing a target vowel is phonologically similar to the following syllable (i.e., similarity between σ_1

and σ_2 in a $[C_1V_1]_{\sigma1}[C_2V_2]_{\sigma2}$ sequence where V_1 is the target for devoicing). For example, if the manner of the onsets (C_1 and C_2) matches in [+continuant], devoicing is suppressed (e.g., *suso* 'hem'). Affricates show position-specific behaviors, such that they pattern with fricatives as [+continuant] in prevocalic position, and pattern with stops as [−continuant] in postvocalic position. In addition to the matching manner condition, devoicing is further suppressed when the height of the following vowel (V_2) is high (e.g., *susi* 'sushi'). Note that if V_2 is high, the two vowels in adjacent syllables agree in height, since the target vowel is always high. In sum, three levels of similarity that affect devoicing are demonstrated: No matching manner > Matching manner but mismatching height > Matching manner and height, where devoicing rates decrease in this order.

I argue that devoicing in the matching conditions is blocked so as to maintain self-similarity between adjacent syllables. This captures the insight that devoicing of the target vowel disrupts the similarity with the following syllable, which contains a voiced vowel. This insight is then formalized by the Aggressive Reduplication account proposed by Zuraw (2002). Adopting this account, I provide an analysis whereby a reduplication-like structure between adjacent syllables is imposed by a constraint REDUP, and then a family of correspondence constraints (*κκ*-CORR) checks the identity between those syllables. An important consequence of this analysis is that disruption of self-similarity will be increasingly resisted as adjacent syllables are more similar, since these correspondent syllables violate fewer correspondence constraints. In this way, Aggressive Reduplication accounts for the additive similarity effect observed in Japanese high vowel devoicing; devoicing is increasingly suppressed as the adjacent syllables get similar. While supporting Zuraw's proposal that correspondence beyond the segmental level is necessary for phonology, the current study further shows that the Aggressive Reduplication account can be extended to non-neutralizing processes, which have often been considered to be phonetic or post-lexical processes.

The organization of this study is as follows. In Chapter 2, I introduce the corpus data used for the study, describe the phonological factors that affect devoicing, and demonstrate the statistical significance of these factors via a mixed-effects regression model. Importantly, it will be shown that devoicing is less likely the more the syllable containing the target vowel is similar to the following syllable. In Chapter 3, I give an OT analysis to account for the tendencies observed in Chapter 2. Noting the parallelism between the current case and the reduplicative identity effects in morphological reduplication, the notion of Aggressive Reduplication is introduced to account for the similarity-driven blocking effects. To account for variation, I propose that some parts of the constraint ranking are variable, and show that it is possible to deduce the probabilities of the observed rankings from the output probabilities and conditional probabilities of outputs given a specific ranking. In Chapter 4, I discuss several relevant alternative accounts for the similarity effects in phonology. It will be shown that similarity across syllables can only be captured by the Aggressive Reduplication account. Chapter 5 summarizes the findings of the study and discusses some questions remaining in light of the current analysis.

2. Data

In order to investigate phonological tendencies of Japanese high vowel devoicing, a large-scale dataset such as a corpus is ideal, given a lot of linguistic factors contributing to the variation. Here, I use the Corpus of Spontaneous Japanese (CSJ; Maekawa et al. 2000), which contains a large amount of Japanese speech data that was produced in naturalistic settings. In this chapter, I describe the methods that are used to construct a dataset for an analysis of high vowel devoicing, and the phonological tendencies found from the dataset.

2.1. Corpus

The entire body of the Corpus of Spontaneous Japanese contains about 660 hours of speech that amount to 7.5 million words spoken by 137 native speakers of so-called “Standard” or “Common” Japanese, which is the standardized variety of Tokyo Japanese. The content of the corpus consists of academic presentations, simulated public speech, dialogues, and read speech, all of which are orthographically transcribed and morphologically tagged.

The dataset for the current study comes from the “Core” subset of the corpus. In addition to orthographic transcription and morphological tagging, the “Core” subset contains phonetic information such as segmental alignment, phonetic transcription, prosodic information marked by X-JToBI labels (Maekawa et al. 2002), and most importantly for the current study, information on vowel devoicing. Devoiced vowels are transcribed as such by human labelers using the information from "the wide-band spectrogram, speech waveform, extracted speech fundamental frequency, peak value of the auto-correlation function, in addition to audio playback" (Maekawa and Kikuchi 2005). The Core subset amounts to about 500,000 words and 45 hours of speech.

From the annotation files of the Core, I extracted words that contain the potential target sequence of high vowel devoicing; that is, phonemically short high vowels flanked by voiceless consonants. In those words, I analyzed sequences of /$C_1V_1C_2V_2$/, where V_1 is a high vowel and C_1 and C_2 are voiceless consonants (See (2) for the phonemic inventory and some relevant allophonic rules of Japanese). C_2 can be a geminate consonant, which is contrastive with singleton consonants in Japanese (e.g., /sikaku/ 'qualification' vs. /sikkaku/ 'disqualification'). The identity of the following vowel V_2 is also included in the dataset to see the effect of V_2 height, which is reported to affect devoicing (Yoshida 2002, Byun 2012).

(2) Phonemic inventory of Japanese (Kubozono 2015, Shibatani 1990)

(a) Consonants

	labial	dental-alveolar	palatal	velar	glottal
plosive	p b	t d		k g	
fricative		s z			h
nasal	m	n			
liquid		r			
glide	w		j		

(b) Vowels (long vowels are marked with [ː])

	front	back
high	i iː	u uː
mid	e eː	o oː
low		a aː

(c) Allophonic rules

/t/ → [tʃ] / __ i

→ [ts] / __ u

/s/ → [ʃ] / __ i

/h/ → [ç] / __ i

→ [ɸ] / __ u

Several criteria were used in order to construct the final dataset. First, if V_2 is a short high vowel and the consonant following V_2 is voiceless, V_2 is also a target for devoicing, such as the second /u/ in /tukusu/ 'to exhaust', a case often referred to as a "consecutive devoicing environment" in the literature (Nielsen 2015, Varden 1998, Tsuchida 1997, 2001). In this case, the two devoicing environments may interact with each other, which would complicate the basic patterns of vowel devoicing. Therefore, I separated the tokens that are in a consecutive devoicing environment from those that are in a single devoicing environment. The patterns of consecutive devoicing will be treated in a separate section.

Second, tokens that contain a prosodic phrase boundary after a target vowel were excluded from the dataset. This is because if there is a prosodic boundary, another vowel devoicing process in Japanese, namely phrase-final vowel devoicing, overlaps with the focus of this study. Phrase-final vowel devoicing occurs when a high vowel is preceded by a voiceless consonant and followed by a phrase boundary. Two devoicing processes overlap if the environment for the phrase-final devoicing is followed by a voiceless consonant (e.g., *…desu* 'politeness copula' # *kore* 'this'…, where # denotes a phrase boundary). Using the CSJ data, Kilbourn-Ceron and Sonderegger (2018) claim that two devoicing processes are qualitatively different. For example, a longer pause duration between a target high vowel and the following phrase-initial consonant increases the likelihood of phrase-final devoicing, but when there is no prosodic boundary, a longer pause duration between a target vowel and its following consonant decreases the likelihood of interconsonantal devoicing. Given this, in order to avoid the effects of phrase-final devoicing, only phrase-internal tokens were analyzed.

Third, phonotactically foreign sequences were excluded due to their idiosyncratic behavior. There are certain consonants or CV combinations that almost only occur in loanwords, and thus have lower lexical frequency. For example, /p/ does not occur in the native or Sino-Japanese vocabulary unless it is a part of a geminate or preceded by a nasal coda. Also, sequences like [ti] or [tu] only occur in loanwords since /t/ is affricated before a high vowel in the native and Sino-Japanese strata. As shown by Nogita (2016) and Hirayama and Vance (2018), sequences like these tend to have low rates of devoicing, which might be attributed to their low lexical frequency or a different lexical stratum. Although previous studies (Kilbourn-Ceron and Sonderegger 2018, Maekawa and Kikuchi 2005) find little effect of lexical frequency on devoicing or even a negative correlation between lexical frequency and devoicing rates, Hirayama and Vance (2018) find that frequency of particular moras[2] has a weak positive correlation with devoicing rates in the CSJ data. However, they do not show any relationship between lexical frequency and lexical strata. Since the reason why these sequences result in low devoicing rates is yet inconclusive, I excluded tokens that contain foreign sequences, such as [p], or [t] followed by a high vowel.

Fourth, tokens containing /h/ in C_2 position are excluded due to variable /h/-voicing in intervocalic position (i.e., /h/ →[ɦ] / V_V; Tsuchida 1997, Yoshioka 1981), which might affect vowel devoicing. In terms of rule ordering, if the /h/-voicing rule applies before high vowel devoicing, the target vowel will be no longer in the context of devoicing and thus not

---

[2] Moras are a prosodic unit that is reported to be active in many phonological processes such as accent assignment or timing patterns in Japanese, as in other languages (Kubozono 2015). A CV syllable is monomoraic (light), and a bimoraic (heavy) syllable is constituted by a closed syllable, or an open syllable containing a long vowel or a diphthong in the nucleus.

devoiced. Regarding this issue, many studies (Maekawa and Kikuchi 2005, Fujimoto 2004, Tsuchida 1997, among others) report that /h/ in C_2 position greatly suppresses devoicing, although it does not result in categorical blocking. Unfortunately, there is little information in the literature on the interaction of vowel devoicing and /h/-voicing, so we cannot determine whether low devoicing rates before /h/ should be attributed to the result of /h/-voicing or to the independently motivated effect of /h/ in C_2 position. Further, the information on /h/-voicing is not present in the annotation of the corpus, so it is difficult to collect tokens where /h/ in C_2 is realized as voiceless. Therefore, I excluded tokens with /h/ in C_2 position, leaving the relationship between /h/-voicing and vowel devoicing for future research.

Finally, tokens with speech disfluencies, which are annotated in the corpus, were excluded. The final dataset consists of 31,572 tokens in a single devoicing environment, which include 2,309 words. Additionally, 1,827 tokens are in a consecutive devoicing environment. The consonants that occur in C_1 position are /s, h, t, k/ and those occurring in C_2 position are /s, t, k/ (singleton or geminate). In terms of the manner of articulation of these consonants, there are three classes: fricatives, affricates, and stops. Note that affricates are derived as allophones of /t/ before high vocoids.

There are several other factors that were additionally coded in the dataset. First, presence of pitch accent on target vowels was coded. Second, if C_2 is a geminate, this information was also included. Finally, a word boundary (inside a prosodic phrase) might be present after a target vowel (e.g., *iku* + *si* ‘to go + and’, where + denotes a word boundary), or after the first part of a geminate C_2 (e.g., *sit* + *te* ‘to know + connective suffix’), and this information was also coded in the dataset. It should be noted that bound morphemes such as case markers or inflectional suffixes are counted as a “word” in the corpus, so many of the word boundaries that are transcribed as such in the corpus might actually be morpheme boundaries.

2.2. Phonological tendencies

Having constructed the dataset for the analysis, I report phonological tendencies that contribute to the variation of high vowel devoicing. It will be shown that high vowel devoicing is a gradient phenomenon whereby its probability is determined by various phonological factors. Crucially, I explore the following factors that could affect the probability of devoicing: C_1 and C_2 manner, V_2 height, geminacy of C_2, pitch accent, and the presence of a word/morpheme boundary.

2.2.1. C_1 and C_2 manner

The average rate of high vowel devoicing in the dataset is almost at ceiling (96.38%; 30,428 tokens devoiced out of 31,572). However, as shown in (3), devoicing rates noticeably drop when C_1 and C_2 are both fricatives (Fric_Fric), or when C_1 is an affricate and C_2 is a fricative (Affr_Fric). This is in accordance with the previous findings (Fujimoto 2004, Yoshida 2002) that devoicing is almost obligatory except when the target vowel is surrounded by fricatives.

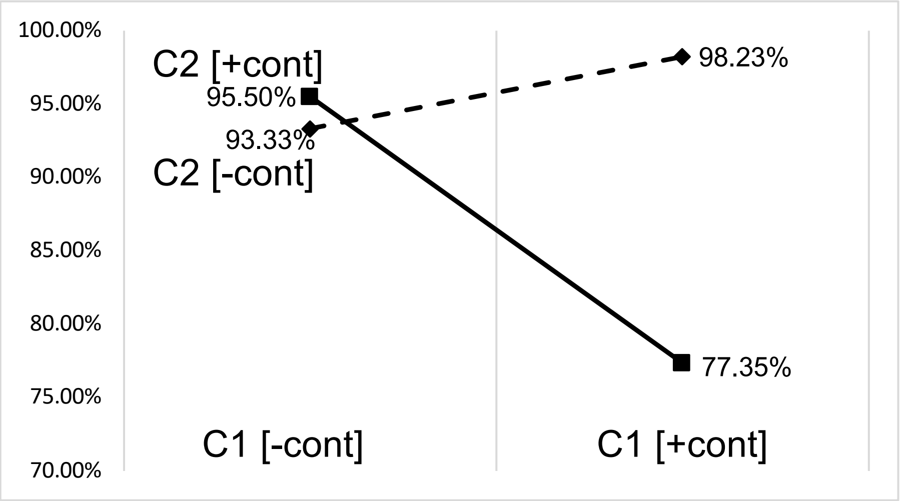
(3) Devoicing rates by C_1 and C_2 manner (only singleton C_2)

C_1	C_2	Devoicing rates (%)	Number of devoiced tokens / total
Fric	Stop	99.11	19821/19999
Fric	Affr	98.50	590/599
Affr	Stop	96.32	2434/2527
Stop	Fric	95.50	2336/2446
Stop	Affr	93.92	479/510
Affr	Affr	93.89	123/131
Stop	Stop	93.24	3597/3858
Affr	**Fric**	**82.53**	**203/246**
Fric	**Fric**	**75.22**	**422/562**

While most of the environments show high devoicing rates above 90%, Affr_Fric and Fric_Fric environments show lower devoicing rates of 82.53% and 75.22%, respectively. This is not attributable only to the effect of a C_2 fricative environment, however, since Stop_Fric environment shows a high devoicing rate of 95.50%. The effect of surrounding fricatives on blocking devoicing has been observed by many researchers (Maekawa and Kikuchi 2005, Fujimoto 2004, Fujimoto and Kiritani 2003, Yoshida 2002, Tsuchida 1997, among others), while the effect of Affr_Fric environment has been pointed out by relatively few (Byun 2012, Maekawa and Kikuchi 2005).

While affricates in C_1 position pattern together with fricatives as in Affr_Fric and Fric_Fric environments, the environments where an affricate is in C_2 position (Affr_Affr (93.89%) or Fric_Affr (98.50%)) do not show a similar suppressing effect as Affr_Fric or Fric_Fric environments. This seems to stem from the fact that when an affricate is in C_1 position, the target vowel is adjacent to the fricative portion of the affricate, while when an affricate is in C_2 position, the vowel is adjacent to the stop portion of the affricate. In other words, affricates exhibit *edge effects* (Sagey 1986, Lombardi 1990, Lin 2005), such that they behave like fricatives prevocalically, while they behave like stops postvocalically. Therefore, I group affricates in C_1 and fricatives as [+continuant] whereas affricates in C_2 and stops are [−continuant] from now on, and I treat Fric_Fric and Affr_Fric environments as a single category where C_1 and C_2 agree in [+continuant]. Given this, I claim that devoicing is suppressed if C_1 and C_2 have matching [+continuant], and I will refer to this environment as *the matching manner condition*. The effect of the matching manner condition is more clearly shown with reference to [continuant] in (4). Note that the matching manner condition only includes the environment where both C_1 and C_2 are [+continuant], since matching [−continuant] does not result in a noticeable blocking effect. In Chapter 3, I propose that the potential matching effect of [−continuant] is overridden by the preference for devoicing when C_2 is [−continuant].

(4) Devoicing rates by manner



Further similarity between consonants in the matching manner condition seems to lower devoicing rates. Below are the devoicing rates of Affr_/s/, /h/_/s/, and /s/_/s/ environments. Devoicing rates are lower when C_1 and C_2 are both /s/ than when they are /h/_/s/ or Affr_/s/, but due to the relatively small number of tokens, it is inconclusive if the differences are significant. As such, I will not further divide these environments into different categories, such as sibilant fricatives and non-sibilant fricatives.

(5) Devoicing rates of Affr_/s/, /h/_/s/, and /s/_/s/ environments

(only singleton C_2)

C_1	C_2	Devoicing rates (%)	Number of devoiced tokens / total
Affr	/s/	82.52	203/246
/h/	/s/	82.05	64/78
/s/	/s/	74.12	358/483

2.2.2. V_2 height

In addition to the matching manner condition, the effect of the vowel in the syllable following the target syllable is observed in the data. That is, given the matching manner condition where both C_1 and C_2 are [+continuant], devoicing is further suppressed if V_2 is high, which I refer to as *the matching height condition*. In other words, devoicing rates decrease if C_1 and C_2 agree in [+continuant] *and* V_1 (target vowel) and V_2 agree in [+high] (since the target vowel is always high). This is shown in (6). As mentioned earlier, in C_1 position, fricatives and affricates are classified as [+cont] while stops are [−cont]. In C_2 position, on the other hand, only fricatives are classified as [+cont] and stops and affricates are [−cont].

(6) Devoicing rates according to C_1/C_2 manner and V_2 height

(only singleton C_2)

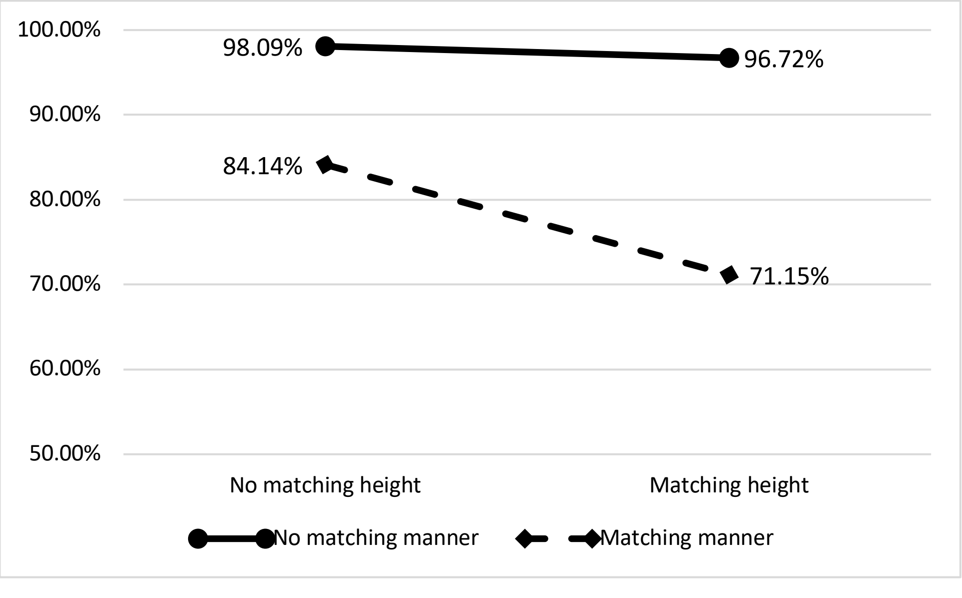
C_1	C_2	V_2	Devoicing rates (%)	Number of devoiced tokens / total
[+cont]	[−cont]	non-high	98.81	20387/20633
		high	98.40	2581/2623
[−cont]	[+cont]	non-high	96.81	1365/1410
		high	93.73	971/1036
[−cont]	[−cont]	non-high	93.34	3701/3965
		high	93.05	375/403
[+cont]	**[+cont]**	**non-high**	**84.14**	**329/391**
		high	**71.15**	**296/416**

Among the tokens in the matching manner condition ([+cont]_[+cont]), the devoicing rate drops from 84.14% to 71.15% when V_2

is high. On the other hand, there is little difference in devoicing rates depending on V_2 height in other combinations of manner.

Although Yoshida (2002) states that the V_2 height effect holds across the manner combinations, no study, to my knowledge, has observed that devoicing is further degraded if both manner and height match. This double-matching-based degradation means that the degree of similarity in two phonological dimensions (consonant manner and vowel height) is negatively correlated with the likelihood of devoicing, such that the more similarities there are between consonants and between vowels, the less likely devoicing is. I refer to this double matching effect of C_1/C_2 manner and V_2 height as *the additive similarity effect* on blocking devoicing. This is expressed in (7). In (7), matching manner only includes matching [+cont], excluding matching [−cont].

(7) Devoicing rates depending on the degree of similarity



In (8), I summarize the phonological tendencies of devoicing by making reference to matching profiles of adjacent syllables. C_1 and C_2 are the onset of adjacent syllables since /CVCV/ sequences are parsed into [CV.CV] in Japanese. Note that the matching height condition is subordinate to the matching manner condition, since in other manner conditions V_2 height does not seem to have a significant effect.

(8) Devoicing likelihood by the degree of similarity

(i) Devoicing is the most likely when C_1 and C_2 (onset of adjacent syllables) does not have matching [+cont]. (e.g., /kusa/ ‘grass’; 97.67%),

(ii) Devoicing is less likely when C_1 and C_2 have matching [+cont] but V_1 and V_2 (nuclei of adjacent syllables) do not have matching [+high]. (e.g., /suso/ ‘hem’; 84.14%),

(iii) Devoicing is the least likely when C_1 and C_2 have matching [+cont] and V_1 and V_2 have matching [+high]. (e.g., /susi/ ‘sushi’; 71.15%).

Among the tokens in the matching height (and manner) condition, total V_1-V_2 identity (i.e., /i/-/i/ or /u/-/u/) does not show greater blocking effects than partial identity (i.e., /i/-/u/ or /u/-/i/), as shown in (9).

(9) Partial identity vs. total identity of V_1/V_2 in the matching manner condition

	V_1/V_2 partial identity (/i/-/u/, /u/-/i/)	V_1/V_2 total identity (/i/-/i/, /u/-/u/)
Devoicing rate	69.70% (138/198)	72.48% (158/218)

In summary, the current survey has identified two effects of similarity where devoicing is probabilistically blocked: (i) if both the preceding and the following consonants are [+continuant] (the matching manner condition), and (ii) if both the target vowel and the following vowel are [+high] (the matching height condition). The matching height condition has an effect only when the matching manner condition is met, exhibiting an additive similarity effect. This indicates that these two dimensions of similarity interact in a certain phonological domain to block vowel devoicing. Although the matching manner and matching height effects have been reported separately in the literature, this study first presents a novel finding about the interaction of the two effects with quantitative data.

2.2.3. Pitch accent

Pitch accent is contrastively used in Japanese, as in /ame/ ‘taffy’ vs. /áme/ ‘rain’. While some studies report that pitch accent on the target vowel categorically blocks devoicing (Shibatani 1990), others claim that pitch accent does not affect devoicing at all (Kondo 1997, Nagano-Madsen 1994), or that its blocking effect is gradient (Imai 2004, Fujimoto 2004).

Here, the results from the corpus data show that the blocking effect of pitch accent is indeed gradient, as illustrated in (10). Although accent lowers devoicing probability, its blocking effect is by no means categorical.

(10) Devoicing rates by presence/absence of pitch accent

Unaccented	Accented
96.90% (29359/30297)	83.84% (1069/1275)

2.2.4. Geminacy of C_2

Length is contrastive in Japanese for consonants as well as vowels. For example, /kita/ ‘north’ and /kitta/ ‘to cut-past’ are primarily differentiated by the constriction duration of /t/. The distribution of geminate consonants differs across lexical strata. Geminacy is licensed to voiceless obstruents and nasals in the native and Sino-Japanese vocabulary, while its range is extended to voiced obstruents in loanwords (Itô and Mester 1995).

What is relevant to the current focus is the case where the consonant following the target vowel is a geminate (as in /kitta/). As shown in (11), geminate C_2 inhibits vowel devoicing.

(11) Difference in devoicing rates when C_2 is a singleton or geminate

Singleton C_2	Geminate C_2
97.18% (30005/30877)	60.86% (423/695)

Geminate C_2 lowers devoicing rates in this dataset, which is in accordance with the results of previous studies (Kondo 2001, Han 1994, Shrosbee 2013). A possible cause for the geminate effect, suggested by Kawahara (2015), is that longer duration of a vowel before a geminate (See Kawahara 2015 and references therein) inhibits application of vowel devoicing. This is based on the idea that the phonetic motivation for high vowel devoicing is the intrinsic short duration of high vowels (Solé and Ohala 2010, Ohala 1983, Lehiste 1970). In light of this, lengthened duration of a high vowel caused by a geminate in C_2 might fail to satisfy the durational requirements for vowel devoicing. Considering this, a formal analysis of the effect of geminacy will be offered in Section 3.3., based on the perceptual consequences of pre-geminate vowel lengthening for vowel devoicing.

2.2.5. Morpheme boundary

Presence of a word or morpheme boundary has been reported to affect vowel devoicing. For example, Imai (2004) investigated different types of morphological conditions (morpheme-internal, morpheme boundary followed by a bound morpheme, compound boundary, and word boundary), and shows that word and compound boundaries significantly suppress devoicing. Kilbourn-Ceron and Sonderegger (2018) also show from the CSJ dataset that devoicing rates are lower if a word boundary is present in a devoicing context. In the current data, the devoicing rate in tokens containing a word boundary is not much different to the rate of word-internal tokens, as shown in (12).

(12) Effect of a word boundary on devoicing

No boundary	Boundary
96.94% (11901/12277)	96.02% (18527/19295)

However, as has been noted, many of the word boundaries that are annotated as such might not actually be a word boundary. The corpus counts bound morphemes such as case markers or inflectional suffixes as a separate word, so the types of morpheme boundaries (word boundary, compound boundary, or morpheme boundary before a bound morpheme) are not differentiated in the dataset. Given this, it is difficult to offer a reliable analysis on the effects of word/morpheme boundaries based on the current data, since it is possible that different types of morpheme boundaries, which are pooled in the corpus, have varying effects on devoicing, as suggested by Imai (2004). Therefore, a proper analysis would require further classification of the boundary types. Given the current lack of such a further classification in the corpus, I will leave an analysis on the effects of morpheme boundaries for future research.

2.2.6. Consecutive devoicing environment

Before ending the description of the phonological factors on devoicing, I present the devoicing patterns of multiple adjacent vowels that are in the devoicing context. For the sake of simplicity, I included tokens that only have two vowels that are in the devoicing context and in adjacent syllables (i.e., /$C_{[\text{-vce}]}V_{[\text{+hi}]}C_{[\text{-vce}]}V_{[\text{+hi}]}C_{[\text{-vce}]}$/). These amount to 1,827 tokens in total. Here, I only show which of the vowels devoice in the consecutive devoicing environments, which I have pooled across manner combinations, since the distribution by manner was not sufficiently counterbalanced to see any effects that might be present (See Nielsen (2015) for the manner effects in consecutive devoicing environments). As shown in (13), the most common pattern in the consecutive devoicing environment is that only one of the vowels devoices; tokens where both vowels devoice or no vowel devoices are less likely. More specifically, the majority of the tokens show only single devoicing (66.11%, 1208/1827). Consecutive devoicing is less likely, constituting 32.51% of the cases (594/1827). This is comparable to the results by Nielsen (2015), where she found a 27.1% rate of consecutive devoicing in her experiment. Given that the average rate of single devoicing is over 90%, the rate of consecutive devoicing is much lower than expected.. Finally, non-devoicing of both vowels is very rare (1.37%, 25/1827).

(13) Pattern of vowel devoicing in the consecutive devoicing environment

Both vowels devoice	**Only one vowel devoices** **66.11%** (1208/1827)		**Neither vowel devoices**
	First vowel devoices	Second vowel devoices	
32.51% (594/1827)	26.16% (478/1827)	39.96% (730/1827)	**1.37%** (25/1827)

2.2.7. Statistical analysis

Below are the phonological factors that probabilistically block devoicing (in a single devoicing environment) observed from the corpus data.

(14) Summary of findings

(i) The matching manner condition:	Devoicing is less likely if the target vowel is surrounded by [+cont] consonants.
(ii) The matching height condition:	In the matching manner condition, devoicing is even less likely if the syllable following the target vowel contains another high vowel.
(iii) Pitch accent:	Devoicing is less likely if the target vowel is accented.

(iv) C_2 geminacy: Devoicing is less likely if the consonant following the target vowel is a geminate.

In order to statistically assess these effects, I constructed a mixed-effects logistic regression model, using the *glmer()* function of the *lmerTest* package (Kuznetsova et al. 2017) in R (R Core Team 2020). The dependent variable is whether high vowels in the devoicing context (flanked by voiceless consonants) are devoiced or not (binary). Independent variables are as follows. First, in order to test the similarity effects, *the matching profile* was included as a predictor. In this predictor, the similarity conditions were numerically coded; 0 for the mismatching manner, 1 for the matching manner and mismatching height, and 2 for the matching manner and height. Then, this predictor was Helmert-coded, to see if there is any significant difference between the matching and mismatching manner conditions, and between the matching and mismatching height conditions among the tokens in the matching manner condition. Second, the distinction between singleton vs. geminate C_2 was included, with the singleton condition as the baseline. Third, presence of pitch accent on the target vowel was included, with the unaccented condition as the baseline. Finally, presence of a morpheme boundary was also included in the model as a control predictor. The baseline for this predictor was no boundary condition. The predictors for geminacy, pitch accent, and morpheme boundaries were binary and sum-coded, in order to see the main effects of each predictor, excluding the influence of the baseline of other predictors.

By-word and by-subject random intercepts, as well as by-word and by-subject random slope terms for the morpheme boundary predictor were included in the model.[3] The number of words and speakers was 2,309 and 137, respectively. Below are the results from the model. A negative value of a coefficient estimate means that likelihood of devoicing decreases by the predictor.

(15) Results of the mixed-effect model

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	0.116	0.204	0.571	0.568
Matching profile				
Matching vs. mismatching manner	-2.832	0.229	12.389	< 0.001 ***
Matching vs. mismatching height	-1.379	0.37	3.729	< 0.001 ***
Accent	-0.675	0.074	-9.112	< 0.001 ***
Geminate	-1.918	0.102	-18.788	< 0.001 ***
Boundary	-0.418	0.12	-3.499	< 0.001 ***

As shown in (15), the devoicing rate in the matching manner condition was significantly lower than the mismatching manner condition ($\hat{\beta} = -2.832, p < 0.001$). Moreover, among the tokens in the matching manner condition, the matching height significantly lowers the devoicing rates ($\hat{\beta} = -1.379, p < 0.001$). Therefore, we can see that the devoicing rates decrease as

---

[3] More complex models failed to converge.

the similarity between adjacent syllables increases.

Geminate consonants ($\hat{\beta} = -1.918$, $p < 0.001$) and pitch accent ($\hat{\beta} = -0.675$, $p < 0.001$) also significantly lower devoicing rates. Finally, devoicing rates are lower if there is a morpheme boundary ($\hat{\beta} = -0.418$, $p < 0.001$). Although the difference of the average rates between boundary vs. no boundary condition was small, the results of the regression analysis indicate that certain kinds of morpheme boundaries might have inhibitory effects on devoicing. In summary, among the findings in (14), the following phonological tendencies are found to be significant by the statistical analysis:

(16) Tendencies confirmed by the statistical analysis

(i) The matching manner condition:	Devoicing is less likely if the target vowel is surrounded by [+cont] consonants.
(ii) The matching height condition:	In the matching manner condition, devoicing is even less likely if the syllable following the target vowel contains another high vowel.
(iii) Pitch accent:	Devoicing is less likely if the target vowel is accented.
(iv) C_2 geminacy:	Devoicing is less likely if the consonant following the target vowel is a geminate.

3. Analysis

In this chapter, I analyze the phonological tendencies observed in the previous chapter within the framework of Optimality Theory (Prince and Smolensky 1993/2004). First, I argue the basic pattern of Japanese high vowel devoicing is derived from a typical constraint ranking for allophonic variation: Context-sensitive Markedness >> Context-free Markedness >> Faithfulness. A potentially pathological case of consecutive devoicing is mended by an OCP constraint on voiceless vowels in adjacent syllables.

I further argue that the matching manner and height effects observed in high vowel devoicing are the results of similarity preservation via correspondence between word-internal substrings. Adopting Zuraw's (2002) Aggressive Reduplication, I propose that blocking of devoicing arises in order to prevent disruption of self-similarity between adjacent syllables. By positing a reduplication-like structure, established by the constraint REDUP, identity between substrings is checked by a family of correspondence constraints. An important consequence of the current proposal is that it captures the additive similarity effect; the more similar the two syllables are, the more likely blocking of devoicing is, since fewer correspondence constraints are violated.

Furthermore, I propose positional faithfulness constraints (Beckman 1998) to account for the effects of pitch accent and C_2 geminacy. Finally, variation observed in the data is explained by adopting partial ordering of constraints (Anttila 1997). I show that the probabilities of each observed ranking can be calculated from the devoicing rates observed in the data.

3.1. Basic mechanism of high vowel devoicing

In this section, I argue that devoicing is driven by a contextual markedness constraint DEVOICE, which prohibits a voiced short high vowel between voiceless consonants. Note that all voiceless consonants are obstruents in Japanese.

(17) DEVOICE

*[−syll, −vce][+syll, +hi, −long, +vce][−syll, −vce]

(No voiced short high vowel between voiceless consonants)

This markedness constraint may have functional roots in articulatory effort. Since high vowels are intrinsically shorter than non-high vowels cross-linguistically (Solé and Ohala 2010), it can be difficult to achieve a glottal adduction gesture of a short high vowel between two glottal abduction gestures of surrounding voiceless consonants. Therefore, the undershoot of voicing might result in devoicing. As mentioned in Chapter 1, this mechanism of devoicing is put forward by a more general theory of *gestural overlap*, as proposed by Jun et al. (1998) and Beckman (1996).

However, simply taking gestural overlap as the sole explanation for devoicing has limitations for following reasons. First, it cannot account for the blocking effects of the matching manner and height, as shown in the current data. It is not clear why a high vowel surrounded by [+continuant] consonants, which have a glottal opening gesture on their own, resists undershoot, while devoicing is relatively frequent in other contexts. Furthermore, the effect of the following vowel is more difficult to explain under the gestural overlap account, since the glottal gestures of the two vowels are not adjacent. Second, the size of the glottal opening in the vowel devoicing context is often greater than that of a single voiceless consonant (Funatsu and Fujimoto 2011). This is not expected by the gestural overlap account, where devoicing is assumed to result due to concatenation of surrounding glottal gestures. Finally, physiological studies show that the activation of the glottal opening muscle occurs only once during the devoicing context (Hirose 1971). This suggests that devoicing is driven by motor control, rather than by a passive assimilation to the surrounding segment. Taken together, treating devoicing as a phonetic by-product is not consistent to the patterns observed in the corpus data and the results from experimental studies. Rather, it would be better to treat high vowel devoicing as a phonological phenomenon that is motivated by phonetics, as many phonological processes are.

Since voicing of the vowel is sacrificed by devoicing, the faithfulness constraint that is in conflict with Devoice is IO-Id(vce)[4]. Devoicing occurs if Devoice is ranked above IO-Id(vce), and the tableau for the basic devoicing pattern is given in (18).

(18) Basic devoicing pattern

/kusa/ ‘grass’	Devoice	IO-Id(vce)
a. kusa	*!	
☞ b. ku̥sa		*

An additional context-free markedness constraint, SonVoi (Itô, Mester and Padgett 1995), which bans a voiceless sonorant, should be taken into consideration. This is based on the observation that voiceless sonorants, including vowels, are marked cross-linguistically (Maddieson 1984), and that the distinction between voiced vs. voiceless sonorants is not contrastive in Japanese, as in many languages. This constraint, defined in (19), should be ranked above IO-Id(vce) in order to derive the correct pattern.

(19) SonVoi (Itô, Mester and Padgett 1995)

*[+son, −vce] (Sonorants must be voiced)

Under the ranking of SonVoi >> IO-Id(vce), even if a hypothetical

---

[4] Alternatively, Tsuchida (1997, 2001) uses [spread glottis] for representation of devoiced vowels. Her argument is based on cross-linguistic markedness of devoiced vowels and the phonetic fact that devoiced vowels are produced with wide glottis. However, if we use [spread glottis] for the analysis, it is unclear why the repair strategy for Devoice should be marking [spread glottis] for the vowel. For this reason, I consider that using [voice] is more intuitive for the analysis.

input contains a voiceless vowel under the assumption of Richness of the Base (Prince and Smolensky 1993/2004), it must surface as voiced, as shown in (20). Nevertheless, since voiceless vowels are allowed only in a certain context, a context-sensitive markedness DEVOICE must outrank SONVOI. This constitutes a classic example of allophonic variation, with the ranking of Context-Sensitive Markedness >> Context-Free Markedness >> IO-Faithfulness.

(20) Hypothetical input voiceless vowel

/bḁ/	DEVOICE	SONVOI	IO-ID(vce)
a. bḁ		*!	
☞ b. ba			*

Although the ranking of SONVOI is required for completeness of the argument, it does not affect the results of the analysis with regard to the inhibitory factors of vowel devoicing. For this reason, I will omit the ranking of SONVOI in the tableaux presented below, for the sake of simplicity.

Before moving on to inhibitory factors of vowel devoicing, I would like to point out a case which seems to contradict the current analysis. This is the case of *consecutive vowel devoicing*, where consecutive syllables are in the devoicing context. For example, the first two vowels in /tukusu/ ‘to exhaust’ are surrounded by voiceless consonants, and hence subject to devoicing. If DEVOICE is ranked above IO-ID(vce), the optimal output candidate for /tukusu/ should be [tsu̥ku̥su], where both vowels devoice.

However, this is not the case; as shown in Section 2.2.6., an important tendency observed in words in consecutive devoicing environments is that devoicing of two consecutive vowels does not occur very often (also shown by Nielsen 2015, Maekawa and Kikuchi 2005). In the current data, while devoicing in only one vowel comprises the majority of the consecutive devoicing pattern (66.11% of the tokens with two adjacent vowels that are subject to devoicing), devoicing of two vowels occurs less frequently (32.51%).

Formulating a complete analysis for consecutive devoicing environments is beyond the scope of the current study, but it seems that there is a general tendency to avoid consecutive devoiced vowels. A similar tendency to avoid consecutive devoicing has also been observed in Comanche (Cho 1993). It is not the two violations of IO-ID[vce] or SONVOI that creates the observed pattern, since having multiple devoiced vowels in non-adjacent syllables is common, as confirmed by several studies (Kawai et al. 1995, Han 1962). For example, a word with three vowels in the devoicing context like /kikitukeru/ 'to overhear' is often pronounced as [ki̥kitsu̥keru], where devoicing occurs every other syllable, leaving the second vowel voiced. Therefore, to capture the observed pattern, I propose OCP-V̥, which is a markedness constraint that penalizes an occurrence of two voiceless vowels in adjacent syllables.

(21) OCP-V̥:

Two voiceless vowels in adjacent syllables are banned.

As shown in (22), the ranking of OCP-V̥ >> DEVOICE >> IO-ID(vce) produces only single devoicing, banning consecutive devoicing. Nevertheless, the ranking between OCP-V̥ and DEVOICE should not be treated as categorical, because consecutive devoicing does happen, but only rarely. Therefore, I assume that the ranking between OCP-V̥ and DEVOICE is variable, but the ranking of OCP-V̥ over DEVOICE is more likely, deriving the observed distribution.

(22) Blocking consecutive devoicing

/tukusu/ 'to exhaust'	OCP-V̥	DEVOICE	IO-ID(vce)
a. tsukusu		**	
☞ b. tsu̥kusu[5]		*	*
☞ c. tsuku̥su		*	*
☞ d. tsu̥ku̥su	*		**

3.2. Aggressive Reduplication

In this section, I deal with the similarity-driven blocking effects in Japanese high vowel devoicing. As shown in Chapter 2, the more the syllable containing the target for devoicing is similar to the following syllable, the

---

[5] The more likely output is [tsu̥kusu], probably due to the inhibitory effect of pitch accent on the second syllable.

more devoicing is suppressed. More specifically, if onsets of the two syllables are both [+continuant], devoicing is suppressed (e.g., *suso*). In addition to the matching onset manner, if the nuclei of the two syllables matches in [+high], devoicing is further suppressed (e.g., *susi*).

I propose that these effects are the results of similarity preservation between two adjacent syllables. Note that in a /$C_1V_1C_2V_2$/ sequence in Japanese where V_1 is targeted for devoicing, the syllable boundary is placed between the target vowel and the following consonant, i.e., [$C_1V_1.C_2V_2$]. Given this, I claim that similarity preservation happens in the domain of syllables, since devoicing of a vowel disrupts the similarity with the adjacent syllable which contains a voiced vowel. To formalize this claim, I adopt Zuraw's (2002) Aggressive Reduplication account. Based on the observation that Tagalog vowel raising variably fails to apply if the preceding syllable is similar to the target syllable, Zuraw argues that there is a phonological drive to make word-internal substrings similar, even without any morphological relation between the two substrings. Her analysis consists of two parts, the first of which is the constraint REDUP, which demands a correspondence relation between word-internal substrings, which she dubs "coupling". Zuraw defines REDUP is as follows: "A word must contain some substrings that are coupled". Here, I adopt a slightly modified definition of REDUP, which is used in Stanton (2020). The difference from the original definition is that REDUP requires correspondent units to be adjacent syllables, rather than just any two word-internal substrings. As mentioned in Zuraw (2002), the conditions for locality and the size of a correspondent unit can be attributed to different constraints, as is true for morphological reduplication (e.g., ALIGN, RED=σ). Instead of introducing this sort of constraints, I will use a restricted definition of REDUP for the sake of simplicity. The definition of REDUP used here is presented below.

(23) REDUP (Zuraw 2002, Stanton 2020)

A word must contain adjacent syllables[6] that are coupled.

From now on, I use square brackets and subscripted *κ* to indicate syllables that are coupled. I assume that output candidates with or without coupling are generated by GEN in OT. For example, an output candidate like [su]$_{\kappa}$[ʃi]$_{\kappa}$ for an input /susi/ contains two syllables that are coupled, and thus satisfies REDUP. On the other hand, output candidates like s[u]$_{\kappa}$[ʃi]$_{\kappa}$ or [s]u$_{\kappa}$[ʃi]$_{\kappa}$ does not satisfy REDUP since the two correspondents are not isomorphic or adjacent.[7]

The second part of the Aggressive Reduplication account consists of a family of *κκ*-CORR constraints that demands identity between syllables in correspondence. These correspondence constraints are parallel to the standard faithfulness constraints that operate between input-output pairs (IO-CORR), or

---

[6] Since the unit of correspondence here consists of CV sequences, the unit can also be defined as mora. However, since there was not enough empirical evidence from the data to confirm whether syllable weight contributes to the similarity effects, I use syllable without any theoretical commitment.

[7] I assume that the correct syllable parsing, i.e., [CV.CV], is derived by relevant constraints.

between a reduplicant and a base in case of morphological reduplication (BR-CORR). What differentiates *κκ*-CORR constraints from other dimensions of correspondence is that they are solely driven by the surface form, working across prosodic units in output realizations without any morphological exponence.

For example, again for the input /susi/, an output candidate [su]$_{\kappa}$[ʃi]$_{\kappa}$ satisfies *κκ*-ID(vce) since two correspondent vowels match in terms of voicing, as well as IO-ID(vce), since the input voicing values have not changed in the output. In contrary, [su̥]$_{\kappa}$[ʃi]$_{\kappa}$ violates both *κκ*-ID[vce] and IO-ID(vce), since devoicing yields two syllables that do not have a matching [voice] value. Finally, [su̥]$_{\kappa}$[ʃi̥]$_{\kappa}$ satisfies *κκ*-ID[vce] since both vowels are voiceless, but it violates IO-ID(vce) twice. Note that a word without coupling vacuously satisfies *κκ*-CORR constraints.

With the interaction of REDUP, *κκ*-CORR, IO-CORR and markedness constraints, a phonological process can be blocked or extended in order to maintain similarity between correspondent units. In this aspect, the Aggressive Reduplication account is parallel to McCarthy and Prince's (1995) correspondence-based account of reduplication. McCarthy and Prince show that reduplicative identity can have effects on a phonological rule in such a way that the rule either underapplies or overapplies in both a base and a reduplicant in order to maintain identity between the two. For example, in Southern Paiute (McCarthy and Prince 1995, Sapir 1930), the labial glide *w* undergoes nasalization (represented as *ŋ*^{w}) between vowels (24i). However, this rule fails to apply in reduplication (24ii).

(24) Southern Paiute glide nasalization (McCarthy and Prince 1995, Sapir 1930)

(i) Stem	Derived
a. **w**a’aŋi- ‘to shout’	tïˑ′-**ŋʷ**a’aŋi- ‘to give a good shout’
b. **w**aix̭a- ‘to have a council’	nıaˑ′vi-**ŋʷ**aix̭a-pˑɪ ‘council of (chiefs)’

(ii) Base	Reduplicated
a. **w**ïɣï- ‘vulva’	wï-**w**ï′xïᴀ ‘vulvas (obj.)’
b. **w**aɣi-‘several enter’	wa-**w**a′xˑɪpïɣᴀ‘ ‘all entered’

The above case is what McCarthy and Prince (1995) and Wilbur (1973) refer to as *underapplication*, where a phonological process is expected to apply but fails to do so due to the reduplicative identity with a correspondent unit that is not in the environment of the rule. The opposite pattern is referred to as *overapplication*, where application of a phonological rule is extended to a correspondent unit in order to maintain identity, despite the fact that the correspondent unit is not in the environment of the rule. In (24), even though the intervocalic glides in the reduplicated forms in (24ii) are subject to nasalization, they resist it in order to maintain similarity with the reduplicant, which is not in the environment of the rule.

According to McCarthy and Prince (1995), both underapplication and overapplication result due to the effects of BR-Corr that demands identity between a base and a reduplicant. Specifically, they claim that underapplication is derived when BR-Corr constraints and a markedness constraint are ranked above IO-Corr, and an additional constraint is ranked higher than BR-Corr in order to rule out overapplication. As an example, the tableau proposed by McCarthy and Prince (1995) for the Southern Paiute example is given in (25).

(25) Tableau for underapplication in Southern Paiute reduplication

/RED + **w**ïɣï-/	*[ŋ^{w}	BR-Id(nasal)	*VwV	IO-Id(nas)
☞ a. wï-**w**ï′xïA			*	
b. wï-**ŋ**^{w}ï′xïA		*!		*
c. **ŋ**^{w}ï-**ŋ**^{w}ï′xïA	*!			**

The basic pattern of glide nasalization is derived by the ranking of *VwV >> IO-Id(nas). In addition to these constraints, BR-Id(nas) is introduced which favors matching [nasal] values between the base and the reduplicant. Due to the higher ranking of BR-Id(nas) over the markedness constraint, the output candidate where nasalization applies only in the expected environment (25b) is ruled out. Finally, output (25c) with overapplied nasalization is ruled out because a word-initial nasal glide is banned for markedness reasons. As such, a markedness constraint, *[ŋ^{w}, is ranked over BR-Id(nas) to block overapplication. In summary, derivations of underapplication in reduplication can be generalized into the schematic ranking of ℂ (any constraint) >> BR-Corr >> Markedness >> IO-Corr, as proposed by McCarthy and Prince (1995).

For the current case, despite the absence of any reduplicant morpheme, vowel devoicing underapplies due to the similarity effects between adjacent syllables. In order to account for this, adjacent syllables are understood to stand in a correspondence relation just like Base-Reduplicant, in order to enable similarity effects to block devoicing. In the Aggressive Reduplication account, the role of a reduplicant morpheme that requires a correspondence structure is substituted by the constraint Redup.

Let us now see how a correspondence-based analysis works in devoicing. Recall that the basic devoicing pattern is derived by the ranking of Devoice >> IO-Id(vce). Furthermore, OCP-V̥ variably outranks Devoice to inhibit consecutive devoicing. If Redup and *κκ*-Id(vce) are ranked above Devoice, coupling will block any disruption of self-similarity in voicing. Thus, the ranking of OCP-V̥ >> *κκ*-Id(vce), Redup >> Devoice >> IO-Id(vce) blocks devoicing. Note that the structure of this ranking is exactly the same with the one proposed to generate underapplication in reduplication, except for the presence of Redup: a phonological rule would have applied in normal circumstances due to the ranking of markedness over IO-Corr, but given a correspondence relation, the higher-ranked *κκ*-Corr derives overapplication or underapplication for identity between correspondents. Finally, an additional constraint, OCP-V̥, rules out overapplication, so only underapplication results.[8] In this way, the Aggressive Reduplication account enables an extension of the correspondence-based account for identity effects to cases without a reduplicant morpheme.

(26) Blocking of devoicing via coupling

/susi/	OCP-V̥	κκ-ID(vce)	REDUP	DEVOICE	IO-ID(vce)
☞ a. [su]$_{\kappa}$[ʃi]$_{\kappa}$				*	
b. [su̥]$_{\kappa}$[ʃi]$_{\kappa}$		*!			*
c. [su̥]$_{\kappa}$[ʃi̥]$_{\kappa}$	*!				**
d. suʃi			*!	*	
e. su̥ʃi			*!		*

Tableau (26) shows that the optimal output is (a), where coupling is imposed and devoicing underapplies. This is because the high ranking of REDUP favors output candidates with a correspondence relation, and κκ-ID(vce) favors voicing identity between correspondent syllables. The ranking between REDUP and κκ-ID(vce) does not affect the optimal choice here. Note that candidate (c), which undergoes consecutive devoicing, is banned for a reason external to Aggressive Reduplication. Neither REDUP nor κκ-ID(vce) penalizes the consecutively devoiced output (c) as much as the non-devoiced output (a), since both the outputs have the same [voice] value. Rather, were it not for OCP-V̥, the ranking would select the overapplied output (c) since it

---

[8] Overapplication (i.e., consecutive devoicing) rarely occurs, as in [su̥su̥mu] 'to advance', at the expense of violating OCP-V̥. Note that the second vowel in *susumu* is not followed by a voiceless consonant and hence not expected to undergo devoicing. Although the number of relevant tokens is relatively small, the rate of overapplication is 15.38% (94/611) when both manner and height match. This is higher than the overall rate of overapplication (where the first high vowel is surrounded by voiceless consonants but the following high vowel is followed by a voiced consonant or a vowel) observed in the dataset, which is 7.24% (444/6130).

satisfies DEVOICE. Thus, it is the high ranking of OCP-V̥, together with REDUP and *κκ*-ID(vce), that results in underapplication of devoicing.

Although I have shown the parallelism between morphological reduplication and Aggressive Reduplication, there is a point that differentiates the two: namely, coupling should only be possible if two substrings are similar enough, whereas in morphological reduplication correspondence is always presumed. Such a divergence arises due to the nature of reduplicant morphemes and REDUP. Since a reduplicant morpheme itself does not have any phonological content underlyingly, identity with the base is always presupposed. On the other hand, by its definition, REDUP does not require prior identity, and similarity between correspondent units is a matter of constraint interaction. That is, because of the interaction of REDUP with *κκ*-CORR, coupling will only select self-similar syllables. For instance, if certain *κκ*-CORR constraints are ranked above REDUP, dissimilar substrings with coupling will satisfy REDUP but violate higher-ranked *κκ*-CORR constraints. In such cases, the ranking will choose an output candidate without coupling at the expense of violating REDUP, since it vacuously satisfies *κκ*-CORR. However, for an input word that has self-similar substrings, an output with coupling does not suffer from violations of *κκ*-CORR constraints nor REDUP, and as a result, an output with coupling will be chosen. In this way, the Aggressive Reduplication account can capture the similarity-driven blocking effects; only self-similar strings are allowed for coupling, and thus they resist the rule application.

At the same time, the ranking of IO-CORR constraints will determine whether similarity between substrings is preserved or enhanced. If they outrank κκ-CORR, input specifications will not change to enhance self-similarity, whereas if κκ-CORR outranks IO-CORR, the difference between substrings will be repaired.

For the current devoicing pattern, coupling should only be possible if (i) two syllables have matching onset manner ([+cont]) or (ii) they have matching onset manner *plus* vowel height. To account for this, I propose κκ-ID(cont) and κκ-ID(hi), as defined in (27). κκ-ID(cont) requires matching [continuant] values between two correspondent syllables, so given the ranking of κκ-ID(cont) >> REDUP, coupling will be only possible if adjacent syllables contain an identical [cont] value. Similarly, κκ-ID(hi) require matching height between two correspondent syllables, so given the ranking of κκ-ID(hi) >> REDUP, coupling will be only possible if a target vowel and the vowel in the following syllable are both high.[9] Other possible κκ-CORR constraints, such as κκ-ID(back) or κκ-ID(place) are apparently low-ranked since the effects of other matching features do not seem to manifest in the dataset analyzed.

---

[9] One might wonder why correspondence is only possible between the target syllable and the following syllable, because there is no requirement in REDUP about directionality. However, note that if there is a syllable before the target syllable (e.g., *utsusu* 'to move'), the preceding syllable is already different from the target syllable. Since the preceding syllable is not in the devoicing context, it would not contain a voiceless onset, while the following syllable always contain a voiceless onset due to the context of devoicing (C_2 in the devoicing context is voiceless, and it is the onset of the following syllable). Thus, correspondence with the preceding syllable would be dispreferred since it would incur more violations of κκ-ID constraints than correspondence with the following syllable.

(27) *κκ*-CORR constraints to account for the similarity effects

(a) *κκ*-ID(cont): Let a word contain two substrings S_1 and S_2 that are coupled, and a segment in S_1 correspond to a segment in S_2. If a segment in S_1 is [α continuant], the correspondent segment in S_2 must contain [α continuant].[10]

(b) *κκ*-ID(hi): Let a word contain two substrings S_1 and S_2 that are coupled, and a segment in S_1 correspond to a segment in S_2. If a segment in S_1 is [α high], the correspondent segment in S_2 must contain [α high].

Now, in addition to the ranking of OCP-V̥ >> REDUP >> *κκ*-ID(vce) >> DEVOICE >> IO-ID(vce), if *κκ*-ID(cont) or *κκ*-ID(hi) outranks REDUP, coupling is only chosen if the correspondent syllables have the matching feature, and then devoicing will be only blocked in such a case to preserve self-similarity. IO-CORR constraints like IO-ID(cont) and IO-ID(hi) are undominated, since changing the input specifications in order to enhance self-similarity is banned (e.g., /suso/ → *[susu] or /kusa/ → *[husa]).

It should be noted that the definition of *κκ*-ID(cont) also allows coupling if two onsets are stops (e.g., *kuki* ‘stem’), and thus blocks devoicing. This is problematic since the data shows that the devoicing rates in Stop_Stop

---

[10] This kind of “existential” definition of ID constraints (à la Struijke 2000) is required to allow partial identity between a complex/contour segment and a simple segment (e.g., [ts] vs. [t] or [s]) as discussed later. Presumably, MAX[F] or DEP[F] can account for the difference between partial and total identity.

environment are quite high (i.e., over 90%). Furthermore, it is unclear at this point how we should evaluate correspondence of [cont] for affricates given their position-specific behaviors. These issues are accounted for by adopting a few additional hypotheses motivated by the nature of affricates and stops.

As mentioned in Chapter 2, I assume the contour representation for affricates for the edge effects (Sagey 1986). That is, affricates have linearly ordered representation of [−cont] and [+cont], which correspond to the stop and fricative portion, respectively, as illustrated in (28). In this fashion, the edge effects of affricates are accounted for; affricates pattern with stops in postvocalic position (i.e., left edge of an affricate) while they pattern with fricatives in prevocalic position (i.e., right edge of an affricate).

(28) Contour representation of [cont] in affricates (Sagey 1986)

```
            [ts]
           /    \
    [−cont]      [+cont]
```

By specifying two [cont] values for the segment, affricates satisfy *κκ*-Id(cont) either with stops or fricatives.[11] By its definition in (27), *κκ*-Id(cont)

---

[11] An alternative way of explaining the matching manner effect of affricates is to assume that affricates are strident stops (Jakobson et al. 1952, Rubach 1994), so that affricates and /s/ share [+strident] but not [+continuant]. Then, the matching effect of Affr_/s/ can be attributed to *κκ*-Id(strident) rather than *κκ*-Id(cont), without resorting to the contour segment analysis. A consequence of introducing *κκ*-Id(str) is that /s/_/s/ environment is expected to show a stronger blocking effect than /h/_/s/ or Affr_/s/, since it satisfies both *κκ*-Id(cont) and *κκ*-Id(str). This might be true, since /s/_/s/ environment does evince lower devoicing rates than /h/_/s/ or Affr_/s/, as shown in Chapter 2. However, due to the small number of the tokens, I could not verify the significance, so the evidence

is only violated if correspondent segments do not contain an identical value for [cont]. For example, *κκ*-ID(cont) is satisfied in words like *ʃitʃi* ‘seven’, *tsuʃima* ‘Tsushima (name of a region)’, or *tsutʃi* ‘soil’. However, non-devoicing via correspondence of [−cont] (e.g., Stop_Stop, Affr_Affr, Stop_Affr, Affr_Stop environments) is ruled out due to an additional constraint: namely, devoicing of a vowel followed by [−cont] is highly preferred. Formulating this constraint is supported by cross-linguistic tendencies and phonetics. Both in Korean (Jun et al. 1998) and Turkish (Jannedy 1995), high vowel devoicing is more likely before a stop and affricate than a fricative. Both studies attribute this tendency to the articulatory characteristics of fricatives and stops. The ballistic oral closure of stops and affricates impedes voicing, while before voiceless fricatives, voicing of vowels ceases rather gradually (Sawashima and Hirose 1980).

Based on this, I propose an additional context-sensitive markedness constraint, namely, DEVOICE/_[−cont]. This constraint demands vowel devoicing if a short high vowel is preceded by a voiceless consonant and followed by a stop or affricate, and it should be ranked above *κκ*-ID(cont), so that the matching effect of [−cont] is overridden by the preference for devoicing when the target vowel is followed by [−cont]. The contour representation of [cont] allows affricates to pattern with stops post-vocalically.

---

for *κκ*-ID(str) is not entirely conclusive. Also, even if *κκ*-ID(str) is introduced, the positional behaviors of affricates still require additional explanation.

(29) DEVOICE/_[−cont]

*[−syll, −vce][+syll, −long, +high, +vce][−syll, −vce, −cont]

(No voiced short high vowel between a voiceless consonant and a voiceless stop or affricate)

Let us now illustrate how matching [+cont] onsets result in blocking devoicing. In (30), I present the ranking of DEVOICE/_[−cont] over *κκ*-ID(cont), and *κκ*-ID(cont) over REDUP. Undominated IO-ID(cont) and IO-ID(hi) are omitted from the tableaux, due to space limitations. (30i) shows an example where adjacent syllables have mismatching [±cont] onsets. Outputs with coupling (a, b, c) lose due to a fatal violation of *κκ*-ID(cont), and the devoiced output without coupling (e) is chosen as optimal based on the ranking of DEVOICE >> IO-ID(vce). In (30ii), adjacent syllables have matching [+cont] onsets, so non-devoicing (a) results via coupling. Finally, (30iii) shows an example with matching [−cont]. Although the non-devoiced output with coupling (a) satisfy *κκ*-ID(cont), it fatally violates a higher-ranked DEVOICE/_[−cont], and hence the devoiced output without coupling (e) is chosen.

(30) Blocking of devoicing if onsets match in [+cont]

(i) Mismatching [±cont] → Devoicing applies

/kusa/ ‘grass’	OCP-V̥	DEVOICE/ _[−cont]	κκ-ID(cont)	κκ-ID(vce)	REDUP	DEVOICE	IO-ID(vce)
a. [ku]$_{\kappa}$[sa]$_{\kappa}$			*!			*	
b. [ku̥]$_{\kappa}$[sa]$_{\kappa}$			*!	*			*
c. [ku̥]$_{\kappa}$[sḁ]$_{\kappa}$	*!		*				**
d. kusa					*	*!	
☞ e. ku̥sa					*		*

(ii) Matching [+cont] onsets → Non-devoicing

/suso/ ‘hem’	OCP-V̥	DEVOICE/ _[−cont]	κκ-ID(cont)	κκ-ID(vce)	REDUP	DEVOICE	IO-ID(vce)
☞ a. [su]$_{\kappa}$[so]$_{\kappa}$						*	
b. [su̥]$_{\kappa}$[so]$_{\kappa}$				*!			*
c. [su̥]$_{\kappa}$[so̥]$_{\kappa}$	*!						**
d. suso					*!	*	
e. su̥so					*!		*

(iii) Matching [−cont] onsets → Devoicing applies

/tuti/ ‘soil’	OCP-V̥	DEVOICE/ _[−cont]	κκ-ID(cont)	κκ-ID(vce)	REDUP	DEVOICE	IO-ID(vce)
a. [tsu]$_{\kappa}$[tʃi]$_{\kappa}$		*!				*	
b. [tsu̥]$_{\kappa}$[tʃi]$_{\kappa}$				*!			*
c. [tsu̥]$_{\kappa}$[tʃi̥]$_{\kappa}$	*!						**
d. tsutʃi		*!			*	*!	
☞ e. tsu̥tʃi					*		*

In addition to the ranking in (30), another ranking is required for the case where devoicing is blocked if both onset manner and vowel height match. This is possible if both κκ-ID(cont) and κκ-ID(hi) are ranked above REDUP, which is shown in (31). In a word like *suso* ‘hem’ (31i), outputs with coupling (a, b, c) lose due to the violation of the high-ranked κκ-ID(hi). On the other hand, if both manner and height match, as in *susi* ‘sushi’ (31ii), non-devoicing results via coupling.

(31) Blocking of devoicing if onsets match in [+cont] and vowels match in height

(i) Matching [+cont] onsets but mismatching vowel height

→ Devoicing applies

/suso/ ‘hem’	OCP-V̥	Devoice/ _[−cont]	κκ-Id(cont)	κκ-Id(hi)	κκ-Id(vce)	Redup	Devoice	IO-Id(vce)
a. [su]$_\kappa$[so]$_\kappa$				*!			*	
b. [su̥]$_\kappa$[so]$_\kappa$				*!	*			*
c. [su̥]$_\kappa$[so̥]$_\kappa$	*!			*				**
d. suso						*	*!	
☞ e. su̥so̥						*		*

(ii) Matching [+cont] onsets and vowel height → Non-devoicing

/susi/ ‘sushi’	OCP-V̥	Devoice/ _[−cont]	κκ-Id(cont)	κκ-Id(hi)	κκ-Id(vce)	Redup	Devoice	IO-Id(vce)
☞ a. [su]$_\kappa$[ʃi]$_\kappa$							*	
b. [su̥]$_\kappa$[ʃi]$_\kappa$					*!			*
c. [su̥]$_\kappa$[ʃi̥]$_\kappa$	*!							**
d. suʃi						*!	*	
e. su̥ʃi						*!		*

Based on the analysis so far, I summarized the rankings established to account for the observed distribution of Japanese high vowel devoicing in (32). Since devoicing occurs even in self-similar syllables, Redup and *κκ*-Corr can be ranked below and do not have any effect, as in ranking (32a). The second ranking (32b) derives non-devoicing if onset manner matches in [+cont], regardless of vowel height. Finally, ranking (32c) derives non-devoicing if both onset manner and vowel height match.

Notice that the current analysis captures the additive similarity effect. If a word has mismatching manner (e.g., *kusa*), all three rankings derive devoicing. If a word has matching [+cont] onsets but mismatching vowel height (e.g. *suso*), two out of three rankings derive devoicing. Finally, if a word has matching [+cont] onsets and vowel height (e.g., *susi*), only one ranking derives devoicing. In other words, if the adjacent syllables have more matching features, more rankings will allow blocking. While each ranking in (32) produces (non-)devoicing categorically, the variation observed in the data will be accounted for in Section 3.4., by employing partial ordering of constraints (Anttila 1997) and probability assignment to each ranking.

(32) Observed rankings to account for the similarity effects

(a) Devoicing across the board

OCP-V̥, IO-ID(cont), IO-ID(hi), DEVOICE/_[−cont]

>> DEVOICE >> SONVOI >> IO-ID(vce)

(REDUP and *κκ*-CORR constraints are ranked below)

(b) Devoicing blocked if onset manner matches

OCP-V̥, IO-ID(cont), IO-ID(hi), DEVOICE/_[−cont]

>> *κκ*-ID(cont) >> *κκ*-ID(vce), REDUP

>> DEVOICE >> SONVOI >> IO-ID(vce)

(*κκ*-ID(hi) is ranked below)

(c) Devoicing blocked if both onset manner and vowel height match

OCP-V̥, IO-ID(cont), IO-ID(hi), DEVOICE/_[−cont]
>> *κκ*-ID(cont), *κκ*-ID(hi) >> *κκ*-ID(vce), REDUP
>> DEVOICE >> IO-ID(vce)

In summary, I have shown that the similarity-driven blocking effects in Japanese high vowel devoicing is successfully explained by introducing a reduplication-like structure to substrings in a word. As in the case of reduplicative identity in morphological reduplication, correspondence constraints have effects in keeping the two substrings similar. The only difference with reduplication is that the current analysis requires an additional constraint that imposes a correspondence relation, and this relation is only posited if certain correspondence constraints are satisfied. In this manner, the additive similarity effects are explained; the more similar two substrings are, the fewer violations of correspondence constraints, and more rankings will block devoicing.

3.3. Accent and geminacy

In this section, I present additional constraints for the blocking effects of pitch accent, C_2 geminacy, which achieved significance in the statistical analysis performed in Chapter 2.

First, I ascribe the blocking effect of pitch accent to positional faithfulness (Beckman 1998) of accented syllables. As the positional privilege of stressed syllables is widely attested across languages, as shown in Beckman (1998), it is plausible to assume that accented syllables hold such privilege because pitch accent gives prominence to a syllable. Furthermore, given that the primary cue for pitch accent is fundamental frequency, losing it by devoicing would greatly endanger its perceptibility.[12] Therefore, I propose IO-ID(vce)/σ́[13] that penalizes change of [voice] in an accented syllable. If this faithfulness constraint is ranked above DEVOICE, devoicing will be blocked. Since devoicing of an accented vowel does happen, the ranking between DEVOICE and IO-ID(vce)/σ́ should be variable.

(33) Blocking of devoicing by pitch accent

/títi/ ‘father’	IO-ID(vce)/σ́	DEVOICE	IO-ID(vce)
☞ a. tʃitʃi		*	
☞ b. tʃi̥tʃi	*!		*

Next, the reason why C_2 geminacy blocks devoicing can be ascertained by considering the effect of geminates on duration of the preceding vowel. In Japanese, the duration of pre-geminate vowels is longer than pre-singleton vowels (Kawahara 2015 and references therein). For example, Kawahara (2006) found that the duration of vowels before a

---

[12] Nevertheless, Japanese speakers might be able to detect the pitch accent of a word since pitch accent is largely lexically determined, and there are secondary cues for pitch accent such as f0 contour of the preceding/following syllables (Matsui 1993, Sugito 1982).

[13] Similar constraints have been employed in Cho (1993) and Tsuchida (2001).

singleton voiceless consonant is on average 36.9ms, whereas the duration before a voiceless geminate is 53.4ms. From this, I assume that devoicing before a geminate is impeded since perceptual change of the [voice] value is quantitatively greater for pre-geminate high vowels than pre-singleton high vowels. In other words, voicing change of a short vowel would be less distinguishable than that of a lengthened vowel. This idea is based on the P-map hypothesis by Steriade (2001, 2009) who claims that greater perceptual change of an input yields a greater violation of a relevant constraint. Under this hypothesis, I assume that the perceptual difference between a lengthened voiced high vowel and its devoiced counterpart would be greater than that between a short voiced high vowel and its devoiced counterpart, i.e., Δ(iˑ-i̥ˑ)/C̥_C̥ > Δ(i-i̥)/C̥_C̥ ([ˑ] denotes lengthening). This is reflected in the grammar in a way that IO-ID(vce) constraints are relativized in terms of their target or context, and the ranking of these faithfulness constraints is fixed so that a greater perceptual change is penalized by a higher-ranked constraint.

To account for the current pattern, I propose another positional faithfulness constraint, IO-ID(vce)/_GEM, that bans a voicing change of a vowel before a geminate consonant. To achieve the blocking effect, the ranking between two faithfulness constraint should be IO-ID(vce)/_GEM >> DEVOICE, although this ranking is variable as the blocking effect is not categorical. The result of this analysis is shown in (34). Based on the ranking between IO-ID(vce)/_GEM and DEVOICE, devoicing before a geminate consonant is determined.

(34) Blocking by C_2 geminacy

/kitte/ ‘stamp’	IO-Id(vce)/_Gem	Devoice	IO-Id(vce)
☞ a. kitte		*	
☞ b. ki̥tte	*		*

3.4. Calculating ranking probabilities

So far, the constraints that block devoicing would do so categorically if they consistently outrank the relevant other constraints. However, given the variable nature of vowel devoicing in Japanese, a mechanism by which phonological grammar can produce variation is necessary. There are many proposals in order to model variation in phonological theory (see Coetzee and Pater 2011 for a review of many such proposals), but in this study, I demonstrate a preliminary implementation of the Partially Ordered Constraints approach proposed by Anttila (1997), whereby some parts of a ranking are not fixed by grammar and a different ranking is generated at each evaluation in OT. In this theory, the probability of a certain output is determined by the number of possible rankings which produce it. However, as pointed out by Boersma and Hayes (2001), this can generate only limited gradience of probabilities. To resolve this issue, I diverge from Anttila’s original proposal in proposing that grammar can directly assign probabilities for each ranking (See Jarosz 2013, 2015 for a similar approach) based on the formula of total probability. The current proposal also departs from Stochastic OT (Boersma 1997, Boersma and Hayes 2001), under which ranking distributions follow a normal distribution. As such, Stochastic OT is more restricted than the current proposal, which model a wider range of potential output probabilities by allowing rankings to be, in principle, assigned any probability. In this approach, learners' job is to find the relevant constraints that account for the output distributions, and the ranking probabilities that best fit the observed output probabilities. For example, say a ranking of A >> B produces application of a rule while the opposite ranking B >> A blocks it. Suppose that the observed data shows that the rate of rule application is 70%, and there is no additional constraint that can affect the output realization. Then, the probability of the observed output $P(\mathrm{O})$ is equal to the sum of the products of the probability of a possible ranking $P(\mathrm{R_i})$ and the conditional probability of the observed output given a specific ranking $P(\mathrm{O}|\mathrm{R_i})$. This is expressed in the formula of total probability (35). Intuitively, the formula indicates that the probability of an output is determined by the rankings that produce the rule application and how likely those rankings are.

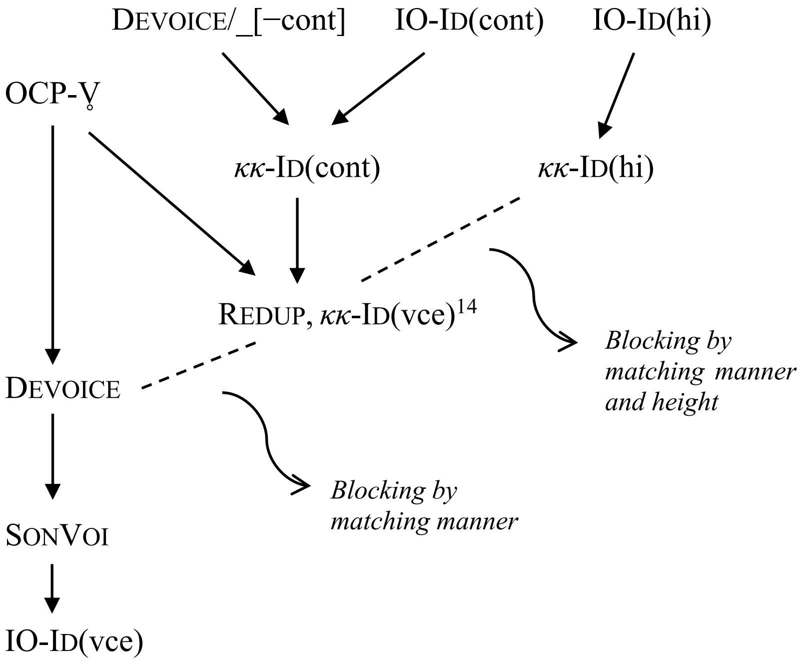
(35) Formula of total probability for calculating ranking probabilities

$$P(O) = \textstyle\sum_i P(O \cap R_i), \text{ or } P(O) = \textstyle\sum_i P(O|R_i)\, P(\mathrm{R}_i)$$

In the above example, we know that the probability of the observed output $P(\mathrm{O})$ is 0.7, $P(\mathrm{O}|\mathrm{A}>>\mathrm{B}) = 1$ and $P(\mathrm{O}|\mathrm{B}>>\mathrm{A}) = 0$. All the conditional probabilities are 0 or 1 because each ranking produces or blocks rule application categorically. Then, assuming that only the two rankings are possible, $P(\mathrm{O}) = P(\mathrm{O}|\mathrm{A}{>>}\mathrm{B})*P(\mathrm{A}{>>}\mathrm{B}) + P(\mathrm{O}|\mathrm{B}{>>}\mathrm{A})*\mathrm{P}(\mathrm{B}{>>}\mathrm{A})$. Since $P(\mathrm{O}|\mathrm{B}{>>}\mathrm{A}) = 0$, we can deduce that $0.7 = 1*P(\mathrm{A}{>>}\mathrm{B}) + 0* P(\mathrm{B}{>>}\mathrm{A})$, so the probability of ranking A >> B is 0.7, and that of ranking B >> A is 0.3.

To instantiate this idea to the current data, let us consider again the rankings for the similarity-driven blocking effects. In (36), I present the Hasse diagram for the relevant rankings. The solid lines indicate that the ranking between two constraints is fixed, and the upper constraint dominates the lower constraint. The broken lines indicate that the ranking is variable.

(36) Hasse diagram for the variable ranking



---

[14] The ranking between REDUP and κκ-ID(vce) is irrelevant to the outcome.

In (36), all the constraint rankings are fixed except for that between Devoice and Redup, and between Redup and *κκ*-Id(hi).[15] The ranking between Devoice and Redup determines whether devoicing is blocked by the matching manner, given that the ranking of *κκ*-Id(cont) above Redup is fixed. In addition, the ranking Redup and *κκ*-Id(hi) determines whether the matching height will block devoicing in addition to the matching manner. By setting the ranking between Redup and *κκ*-Id(cont) as fixed and the ranking between Redup and *κκ*-Id(hi) as variable, it is possible to model the subordinate position of the matching height condition where *κκ*-Id(hi) outranks Redup only when *κκ*-Id(cont) also outranks Redup.[16]

I assume that the ranking of Devoice over IO-Id(vce) is learned as fixed by learners of Tokyo Japanese, given the generality of devoicing. That is, devoicing usually occurs unless there are specific environments that block it, supporting the ranking of Devoice over IO-Id(vce). Note that if the ranking between Devoice and IO-Id(vce) is variable, it causes a learning problem since a non-devoiced output can be ambiguously attributed to the effect of either IO-Id(vce) or coupling, assuming that there is no audible difference between the output realizations with or without coupling. I will leave the

---

[15] Although the ranking between and is in fact variable (See Section 3.1), I assume that it is fixed to focus on the derivation of underapplication of devoicing.

[16] It seems difficult to model this type of constraint interaction in probabilistic variants of Harmonic Grammar (HG; Legendre et al. 1990), such as Noisy HG (Boersma and Pater 2008) or MaxEnt HG (Goldwater and Johnson 2003). This is because any amount of weight assigned to *κκ*-Id(hi) by grammar to account for the additive similarity effect will also result in blocking devoicing to some degree when only *κκ*-Id(hi) is violated.

question of how a learner can determine a ranking in the presence of hidden structure such as coupling for future research (see Jarosz 2013, 2015 for learning hidden structure given ambiguity), and assume that the ranking between Devoice and IO-Id(vce) is fixed in order to circumvent this learning problem.

The data in (8) showed the devoicing rates arranged by degree of similarity. The result is roughly that, if there is no matching feature between adjacent syllables, the devoicing rate is almost at ceiling (≈ 100%). If the two syllables have matching manner but do not have matching height, the devoicing rate is about 80%. Finally, if the two syllables have matching manner and height, the rate is about 70%.

With this data in hand, we can calculate the probabilities of each variable ranking shown in (36) from the conditional probabilities of the observed output given a specific ranking. In (37), I present devoicing probabilities in different similarity contexts and possible rankings. Each cell shows the conditional probability of devoicing in a given context with a specific ranking. There are three rankings under consideration. The first (R_1) is where Devoice outranks Redup, so that devoicing occurs across the board. The second (R_2) is where Redup outranks Devoice and *κκ*-Id(hi) is ranked below Redup, so that non-devoicing results when the onset manner matches. Finally, the third ranking (R_2) is where Redup outranks Devoice and *κκ*-Id(hi) outranks Redup, so that non-devoicing results when the onset manner and vowel height match. Note that when Devoice outranks Redup, the ranking between *κκ*-ID(hi) and REDUP is irrelevant since coupling will not have any effect on the output. In addition to these rankings, the three contexts with differing degrees of similarity are referred to as C$_1$, C$_2$, and C$_3$ in (37).

(37) Devoicing rates by context and ranking

Probabilities	R$_1$ Devoicing across the board	R$_2$ Non-devoicing if manner matches	R$_3$ Non-devoicing if manner and height match
C$_1$ no matching features	100%	100%	100%
C$_2$ only manner matches	100%	0%	100%
C$_3$ both manner and height match	100%	0%	0%

As mentioned earlier, based on the formula (35), the devoicing rate in each similarity condition can be calculated by summing up the products of the probability of each ranking $P(\mathrm{R})$ and the conditional probability of devoicing in a context given a specific ranking $P(\mathrm{C}|\mathrm{R})$ (= the devoicing rate in each cell in (37)). All the conditional probabilities are either 0 or 1 (i.e., each ranking produces a categorical result), and the rankings that block devoicing will not be relevant for calculating the results, since their conditional probabilities are zero. In other words, the devoicing probability in a certain environment is derived by the rankings that lead to devoicing and the probabilities of those rankings. The job of the analyst, much as for the learner, is to deduce the probability of a ranking $P(R_i)$ from the devoicing probability $P(C_j)$ and the conditional probability of devoicing given a certain ranking $P(C_j|R_i)$. From the production frequencies observed from the data, we know that $P(C_1) = 1$, $P(C_2) = 0.8$, and $P(C_3) = 0.7$. Thus, because we are assuming that the only possible rankings are R_1, R_2, and R_3, $P(R_1) + P(R_2) + P(R_3) = 1$. As such, based on the formula (36):

$$P(C_j) = P(C_j|R_1) * P(R_1) + P(C_j|R_2) * P(R_2) + P(C_j|R_3) * P(R_3).$$

For C_3 (both manner and height match), which has a devoicing rate of 0.7, devoicing is only produced by R_1, under which devoicing occurs across the board. That is, $P(C_3) = 0.7$, and the relevant conditional probabilities are as follows: $P(C_3|R_1) = 1$, $P(C_3|R_2) = 0$, $P(C_3|R_3) = 0$. Then, since $P(C_1) = P(C_3|R_1) * P(R_1) + P(C_3|R_2) * P(R_2) + P(C_3|R_3) * P(R_3)$, $0.7 = 1 * P(R_1) + 0 * P(R_2) + 0 * P(R_3)$. Therefore, $P(R_1) = 0.7$

For C_2 (only manner matches) with a devoicing rate of 0.8, devoicing is produced by R_1 (devoicing across the board) or R_3 (devoicing if either manner or height does not match). That is, $P(C_2) = 0.8$, and $P(C_2|R_1) = 1$, $P(C_2|R_2) = 0$, $P(C_2|R_3) = 1$. From the previous calculation, we know that $P(R_1) = 0.7$. Then, since $P(C_2) = P(C_2|R_1) * P(R_1) + P(C_2|R_2) * P(R_2) + P(C_2|R_3) * P(R_3)$, $0.8 = 1 * 0.7 + 0 * P(R_2) + 1 * P(R_3)$. Therefore, $P(R_3) = 0.1$. Finally, since $P(R_1) + P(R_2) + P(R_3) = 1$, $P(R_2)$ is 0.2.

This method of calculating the probabilities of rankings is possible for other variable rankings that are proposed to account for other blocking effects on devoicing such as pitch accent or geminacy. The number of possible rankings will increase, but with the knowledge of the output probabilities and the observed rankings that can produce the rule application, the ranking probabilities can be estimated. Although exploring a learning algorithm of this sort is a subject for future research, this section has served to demonstrate a preliminary method by which grammar can assign probabilities of rankings when the learner is equipped with the knowledge of observed rankings and probabilities of output realizations.

4. Potential alternatives

In this chapter, I discuss some potential alternative accounts for the similarity-driven blocking effects on devoicing. I show that these alternative accounts are not as successful as the Aggressive Reduplication account in capturing the similarity effects between adjacent syllables, since they rely only on segmental similarity and cannot be extended to similarity between prosodic units such as syllables.

4.1. OCP

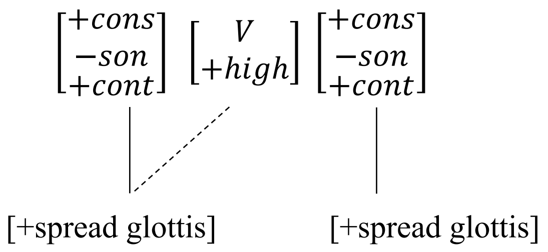
The Obligatory Contour Principle, first suggested by Leben (1978) for tonal phonology and extended to segmental phonology by McCarthy (1986), is a constraint scheme to capture the prevalent tendency of avoiding similar segments in adjacency. Different criteria for "adjacency" have been suggested in the literature (See Suzuki 1998 among others), but it is usually the case that the closer two similar segments are, the stronger the avoidance effect is. For example, adjacency can be hierarchized into immediate adjacency (i.e., two similar segments right next to each other are prohibited), transvocalic adjacency (i.e., two similar segments with a vowel in between are prohibited), or unbounded adjacency (i.e., any occurrence of two similar segments in a word is prohibited). Regarding this, Suzuki (1998) proposes that this hierarchy is encoded in grammar so that more adjacency is penalized by a higher-ranked proximity constraint.

For the current case, blocking of devoicing between [+cont] consonants can be ascribed to an OCP constraint that bans adjacent [+cont] consonants. However, in order to pursue this analysis, we should ask how vowel devoicing makes two consonants more adjacent. One might assume that devoicing is actually deletion so that it yields two surrounding [+cont] segments in immediate adjacency. Transvocalic adjacency would not work since it would penalize non-devoiced outputs too. If devoicing is actually deletion, this is parallel to the cross-linguistically attested pattern of antigemination, where deletion of a vowel is blocked between similar or identical consonants (McCarthy 1986).

However, acoustic (Whang 2018) and articulatory (Shaw and Kawahara 2018, 2021) studies show that traces of oral gestures often remain even though the target high vowel is devoiced. Total deletion of vowel gestures is found in some devoiced tokens, but it seems like total deletion is not conditioned by similarity. For example, Shaw and Kawahara (2021) show that for some speakers, vowels are deleted more often between fricatives than between a preceding fricative and a following stop. If OCP were active in devoicing/deletion, we would expect that fricatives surrounding a vowel would resist deletion more often than other consonantal environments, but this is not confirmed by previous experimental results.

On the other hand, Tsuchida (1997, 2001) pursues a different formulation of OCP to account for the matching manner effect in Japanese high vowel devoicing. She proposes that OCP is violated due to the adjacency between a devoiced vowel and the following fricative, rather than between two fricatives. She proposes that voiceless vowels and fricatives are specified for [+spread glottis] since they are both articulated with wide glottal opening. Furthermore, she claims that if the preceding consonant is a fricative, the vowel becomes [+spread glottis] via spreading from the preceding fricative. Then, if the following consonant is also fricative, two [+spread glottis] values are immediately adjacent, so OCP(+spread glottis) penalizes such a structure. This is illustrated in (38).

(38) Violation of OCP(+spread glottis) (Tsuchida 1997, 2001)



Tsuchida argues that if the preceding consonant is a stop, which does not have the [+spread glottis] specification, [+spread glottis] is not spread from the stop to the vowel but instead GEN generates outputs where the vowel is specified for [+spread glottis], which prevents them from violating a markedness constraint *VOICECONTOUR, which bans a [−vce][+vce][−vce] sequence.[17] If so, her argument incorrectly predicts that devoicing will also

---

[17] Tsuchida implicitly assumes that specification of [+spread glottis] for vowels implies [−voice] specification.

be blocked if the preceding consonant is a stop and the following consonant is a fricative, since a devoiced vowel is specified for [+spread glottis] and is adjacent to the following fricative, which is also marked for [+spread glottis]. Therefore, the formulation of OCP(+spread glottis) proposed by Tsuchida does not single out the cases where the target vowel is surrounded by fricatives.

As shown above, both immediate and transvocalic adjacency condition for OCP cannot explain why devoicing is avoided if the surrounding consonants have the same manner. Regarding this, one might propose a phonetically fine-grained or functional version of OCP (à la Boersma 2000) based on the fact that devoicing between fricatives renders a succession of aperiodic noise, which might be a difficult condition in which to perceive the segments. In a similar vein, Maekawa and Kikuchi (2005) suggest that the succession of noise resulted from devoicing between frication noise would make perception of a syllable boundary less clear. Considering this, one might propose a stipulation to the proximity condition of OCP such that similar consonants across a devoiced vowel must be penalized, or put differently, similar consonants must be separated by a voiced vowel.

However, even if such a stipulation worked successfully, all the formulations of OCP miss the generalization that the matching manner condition interacts with the matching height condition. Additional constraints are required to account for the matching height condition, and they should be able to interact with the OCP constraint such that satisfying two dimensions of similarity would render devoicing less likely. In Aggressive Reduplication, this is possible since the unit of correspondence can be larger than a segment; similarity across syllables, moras or feet is checked by *κκ*-CORR constraints without having to separate different dimensions of similarity such as consonants or vowels.

4.2. ABC

Based on the cross-linguistic observation that similar segments in a word increase their similarity in long-distance harmony, Rose and Walker (2004) propose Agreement by Correspondence (henceforth ABC), which is a correspondence-based theory of similarity. This theory consists of constraints that require correspondence between similar segments, such as sibilants, liquids, or vowels, and a set of correspondence constraints that requires matching features among those segments which are in such a correspondence relation. In this way, the basic architecture of ABC looks similar to Aggressive Reduplication in that similarity preservation or enhancement stems from correspondence. However, there are some significant divergences between the two, as pointed out by Zuraw (2002). First, ABC requires prior similarity for correspondence. For example, constraints such as VV-CORR or SS-CORR, which demand that only vowels or sibilants, respectively, participate in correspondence, and identity is checked only for those segments. On the other hand, Zuraw's REDUP says nothing about similarity. It only requires substrings to be coupled, and it is the interaction with κκ-CORR constraints that results in selecting similar substrings. Second, participants in correspondence are limited to individual segments in ABC, while such restrictions are not required by Aggressive Reduplication. Although a more recent version of the ABC theory, namely ABC+Q (Inkelas and Shih 2013 and their subsequent works), proposes that the size of a correspondent unit can be up to three subsegmental representations (e.g., /ⁿʧ/ can be subdivided into three subsegments [n], [t], and [ʃ]), these subsegments are assumed to belong to a single segment, so that it is still impossible to posit a correspondence relation beyond individual segments.

As such, for the current case, ABC can only account for the matching height condition on blocking devoicing. That is, if two vowels are high, they participate in correspondence, and hence change of [voice] is banned among those. For example, suppose a constraint, IU-CORR, which requires that all high vowels (i.e., /i/ or /u/) in a word to stand in correspondence. Then, a correspondence constraint, IU-ID(voice), penalizes difference in voicing among high vowels. If these constraints are ranked above DEVOICE, devoicing will be blocked if the following vowel is also high.[18]

However, the ABC account cannot account for the matching manner condition, since the target of a phonological process (i.e., high vowels) is

---

[18] Since all high vowels in a word must be in correspondence, the current analysis in ABC predicts that not only the following high vowel but also any high vowel in a word can interact with the target vowel, which is not supported by the data. However, a constraint such as PROXIMITY (Rose and Walker 2004) is able to force the correspondent vowel to be adjacent.

different from what conditions its applications (i.e., surrounding consonants). Even if surrounding consonants are in correspondence by CC-CORR, they can only interact with each other (as in consonant harmony) but never with the target vowel. This is because correspondent units are restricted to be individual segments in ABC.

One might wonder if an analysis is possible with both the ABC and OCP constraints. For the matching manner condition, OCP blocks devoicing if [+cont] segments are separated by a devoiced vowel (as stipulated in the previous section) and for the matching height condition, ABC constraints block devoicing due to correspondence between high vowels. In (39), I present tableaux for such an analysis. ABC constraints and OCP(+cont) are both ranked above DEVOICE, so that devoicing is blocked in similarity conditions. Three possible cases are presented: (i) a word with matching height, (ii) a word with matching manner but mismatching height, and (iii) a word with matching height and manner. As seen in the tableaux, in all the cases devoicing is blocked. This is because OCP and ABC constraints work independently with each other; OCP blocks devoicing without any reference to correspondence, and ABC constraints are ignorant of the identity of surrounding consonants. In (40i), blocking is attributed to the violation of IU-ID(vce). In (40ii), ABC is irrelevant since correspondence between a high vowel and a non-high vowel is not required, and IU-CORR is vacuously satisfied. Blocking is attributed to the violation of OCP. Finally, in (40iii), the optimal output satisfies both ABC and OCP constraints.

However, the results in (39) are different from the observed distribution; what is necessary is the rankings that differentiate the degrees of similarity between syllables. However, the ABC+OCP alternative cannot produce a ranking where devoicing is blocked only when both manner and height match, since the manner-based OCP effect and height-based ABC effect are totally independent of one another.

(39) OCP+ABC analysis of the blocking effects

(i) Only matching height → Non-devoicing

/kusi/ ‘skewer’	OCP-V̥	IU-Id(vce)	IU-Corr	OCP (+cont)	Devoice	IO-Id(vce)
☞ a. k[u]$_x$ʃ[i]$_x$					*	
b. k[u̥]$_x$ʃ[i]$_x$		*!				*
c. k[u̥]$_x$ʃ[i̥]$_x$	*!					**
d. kuʃi			*!		*	
e. ku̥ʃi			*!			*

(ii) Only matching manner → Non-devoicing

/suso/ ‘hem’	OCP-V̥	IU-Id(vce)	IU-Corr	OCP (+cont)	Devoice	IO-Id(vce)
☞ a. s[u]$_x$s[o]$_x$					*	
b. s[u̥]$_x$s[o]$_x$				*!		*
c. s[u̥]$_x$s[o̥]$_x$	*!			*		**
☞ d. suso					*	
e. su̥so				*!		*

(iii) Both matching height and manner → Non-devoicing

/susi/ ‘sushi	OCP-V̥	IU-ID(vce)	IU-CORR	OCP (+cont)	DEVOICE	IO-ID(vce)
☞ a. s[u]$_x$ʃ[i]$_x$					*	
b. s[u̥]$_x$ʃ[i]$_x$		*!		*		*
c. s[u̥]$_x$ʃ[i̥]$_x$	*!			*		**
d. suʃi			*!		*	
e. su̥ʃi			*!	*		*

In summary, in order to account for the attested pattern of additive similarity effects, similarity relations beyond segments are necessary. While the Aggressive Reduplication is able to capture this intuition, alternative accounts such as OCP or ABC are incapable of doing. High vowel devoicing in Japanese constitutes a case where similarities between two dimensions (between consonants and between vowels) interact in order to block a rule, and, of the accounts considered, only Aggressive Reduplication can successfully capture it.

5. Conclusion

The goals of this study are to identify the phonological tendencies in the variable pattern of Japanese high vowel devoicing and to provide an analysis of the observed tendencies. Using a large-scale dataset from a speech corpus, I showed that devoicing is suppressed when the syllable containing the target vowel for devoicing is similar to the following syllable. Specifically, devoicing is suppressed if the consonants surrounding a target vowel, which are the onset of the adjacent syllables, is both [+continuant], which I referred to as the matching manner condition. In addition to the matching manner condition, the devoicing rates further drop when the vowel in the syllable following the target vowel is also [+high], which I referred to as the matching height condition. Furthermore, pitch accent on the target vowel and geminacy of the consonant following the target vowel also suppress devoicing. Finally, I found that devoicing of two vowels in adjacent syllables tends to be avoided in consecutive devoicing environments.

I further provided a formal analysis for Japanese high vowel devoicing within the framework of Optimality Theory (Prince and Smolensky 1993/2004). First, I argued that devoicing is derived by a typical constraint ranking for allophonic variation, which consists of Context-sensitive Markedness >> Context-free Markedness >> Faithfulness. Next, I proposed that the similarity-driven blocking effects observed in high vowel devoicing stems from an effort to preserve self-similarity between adjacent syllables.

Adopting Zuraw's (2002) Aggressive Reduplication, I argued that correspondence between adjacent self-similar syllables is responsible for blocking devoicing. Finally, I introduced additional constraints to account for the effects of pitch accent and geminacy. I claimed that the inhibitory effect of pitch accent is due to the positional faithfulness (Beckman 1998) of accented syllables. Based on the P-map hypothesis (Steriade 2001, 2009), I ascribed the effect of geminates to the positional faithfulness that reflects the perceptual difference of vowel voicing before a singleton consonant and before a geminate.

Given the variable nature of vowel devoicing, I claimed that some parts of the observed ranking are variably ordered, so that different outputs may be produced at each evaluation. In order to produce the devoicing rates observed in the data, I proposed that each observed ranking is assigned a probability by grammar, and learners can calculate it if they are equipped with the knowledge of observed rankings and output probabilities.

I also discussed some alternative similarity-based accounts proposed in the literature. I showed that they do not produce the correct results for the current case, since they all rely on segmental similarity, and different dimensions of similarity such as consonants and vowels cannot interact together to exhibit the additive similarity effect. Therefore, the current study supports Zuraw's (2002) claim that word-internal correspondence exists beyond the segmental level, so that correspondence is possible across prosodic units such as syllables or moras.

One of the remaining questions in this study is how speakers learn correspondence relations. Regarding this, it is worth mentioning that the previous cases that have been claimed to be analyzable by Aggressive Reduplication are only neutralizing processes. For example, Tagalog vowel raising (i.e., e → i, o → u) in penultimate syllables variably fails to apply if the preceding syllable is similar to the target syllable (Zuraw 2002). In Sundanese, long-distance liquid dissimilation fails to apply in local position but rather assimilation happens (i.e., l…l or r…r) (Stanton 2020). Both cases involve neutralization, and these authors suggest that the lexicon of these languages is replete with pseudo-reduplicated words (i.e., words that are not morphologically reduplicated but contain highly self-similar substrings), and thus may provide evidence for the Aggressive Reduplication structure.

However, Japanese high vowel devoicing is non-neutralizing and non-structure-preserving, in that voiceless vowels are not contrastive with the voiced counterparts and thus they are not part of the underlying forms. This study shows that the Aggressive Reduplication account can be extended to non-neutralizing processes, even when the distinction in vowel voicing is not present in lexicon. I speculate that there are several possible ways that REDUP and *κκ*-CORR constraints might be acquired. One is that they are innately provided by Universal Grammar, so that speakers activate those constraints if they encounter words with self-similar substrings. On the other hand, it is possible that similarity relations in lexicon can provide at least partial evidence for the Aggressive Reduplication structure. In this regard, Kawahara et al. (2006) report that many verbal stems in Japanese contain identical consonants in the first and second syllable (e.g., *tataku* ‘to hit’, *sasuru* ‘to rub’), but it remains an open question whether this identity relations among contrastive features can be extended to vowel voicing, which is not contrastive in the language.

Related to this, supposing that correspondence structure is acquired by speakers via Universal Grammar or the lexicon, we may speculate that speakers make adjustments to phonetic realizations of outputs with or without coupling, and listeners would be able to hear them. Therefore, correspondence structure might not actually be “hidden” at all, and learners might infer it from output realizations. For example, the difference in durations of vowels in adjacent syllables might be smaller if the two onsets have the same manner (e.g., [s**a**s**a**] vs. [s**a**n**a**]). Future research is necessary to test these speculations on how correspondence structure is learned.

In conclusion, this study contributes to finding and understanding phonological tendencies observed in Japanese high vowel devoicing by performing a data-rich quantitative study. Although Japanese high vowel devoicing has often been claimed to be a phonetic process, this study argues that phonological abstract structure such as correspondence plays an important role even in non-contrastive processes.

References

Anttila, Arto (1997). Deriving variation from grammar. In Frans Hinskens, Roeland van Hout and W. Leo Wetzels (eds.) *Variation, change and phonological theory*. Amsterdam: Benjamins. 35-68.

Beckman, Jill N. (1998). *Positional Faithfulness*. PhD dissertation, University of Massachusetts Amherst.

Beckman, Mary E. (1996). When is a syllable not a syllable? In Takashi Otake and Anne Cutler (eds.), *Phonological structure and language processing: Cross-linguistic studies*. Berlin: Mouton de Gruyter. 95-123.

Boersma, Paul (1997). How we learn variation, optionality, and probability. *IFA Proceedings* 21. University of Amsterdam. 43-58.

Boersma, Paul (2000). The OCP in the perception grammar. *Rutgers Optimality Archive* 435.

Boersma, Paul and Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32. 45-86.

Boersma, Paul and Joe Pater. 2008. *Convergence Properties of a Gradual Learner in Harmonic Grammar*. Manuscript. Amsterdam and Amherst, MA: Meertens Institute and University of Massachusetts Amherst. [Available on Rutgers Optimality Archive, ROA-970.]

Browman, Catherine P., and Louis Goldstein. (1992). Articulatory phonology: An overview. *Phonetica* 49, 155-180.

Byun, Hi-Gyung (2012). *High Vowel Devoicing in Japanese: as an Indicator of Standardization of Dialect* [in Japanese]. PhD dissertation, University of Tokyo.

Cho, Young-mee Yu (1993). The phonology and phonetics of 'voiceless' vowels. *Proceedings of the 9th Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Semantic Typology and Semantic Universals*. 64-75.

Coetzee, Andries W. and Joe Pater (2011). The place of variation in phonological theory. In John Goldsmith, Jason Riggle and Alan Yu (eds.) *The Handbook of Phonological Theory*. 2nd edn. Maiden, Mass. & Oxford: Wiley-Blackwell. 401-434.

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

Flemming, Edward S. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18. 7-44.

Fujimoto, Masako (2004). Effects of consonant type and syllable position within a word on vowel devoicing in Japanese. *Proceedings of the International Conference on Speech Prosody*, Nara, Japan, 625-628.

Fujimoto, Masako (2012). Effects of consonantal environment and speech rate on vowel devoicing: an analysis of glottal opening pattern. *Journal of the Phonetic Society of Japan* 16(3). 1-13.

Fujimoto, Masako (2015). Vowel devoicing. In Haruo Kubozono (ed.) *Handbook of Japanese Phonetics and Phonology*. Berlin, München, Boston: De Gruyter Mouton. 167-214.

Fujimoto, Masako and Shigeru Kiritani (2003). Comparison of vowel devoicing for speakers of Tokyo and Kinki dialects. *Journal of the Phonetic Society of Japan* 7(1). 58-69.

Funatsu, Seiya and Masako Fujimoto (2011). Physiological realization of Japanese vowel devoicing. *Proceedings of Forum Acousticum 2011*, Aalborg, Denmark, 2709-2714.

Goldwater, Sharon and Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson and Östen Dahl. (eds.) *Proceedings of the Workshop on Variation within Optimality Theory*. Stockholm: Stockholm University. 111-120.

Han, Mieko Shimizu (1962). Unvoicing of vowels in Japanese. *The Study of Sounds* 10 (The Phonetic Society of Japan). 81-100.

Han, Mieko Shimizu (1994). Acoustic manifestations of mora timing in Japanese. *Journal of the Acoustical Society of America* 96(1). 73-82.

Hayes, Bruce, Robert Kirchner, and Donca Steriade (eds) (2004). *Phonetically Based Phonology*. Cambridge: Cambridge University Press.

Hirayama, Manami and Timothy J. Vance (2018). Onset *Cy* and high vowel devoicing in Japanese. *Journal of Japanese Linguistics* 34(1). 103-126.

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

Imai, Terumi (2004). *Vowel devoicing in Tokyo Japanese: a variationist approach*. PhD dissertation, Michigan State University.

Inkelas, Sharon and Stephanie S. Shih (2013). Contour segments and tones in (sub)segmental Agreement by Correspondence. Paper presented at the 21st Manchester Phonology Meeting, University of Manchester, UK.

Itô, Junko and Armin Mester (1995). Japanese phonology. In John A. Goldsmith (ed.) *The Handbook of Phonological Theory*. Cambridge: Blackwell. 817-838.

Itô, Junko, Armin Mester and Jaye Padgett. Licensing and underspecification in Optimality Theory. *Linguistic Inquiry* 26(4). 571-613.

Jaeger, Jeri J. (1978). Speech aerodynamics and phonological universals. *Proceedings of the 4th Annual Meeting of the Berkeley Linguistics Society*. 312-329.

Jakobson, Roman, Gunnar Fant, and Morris Halle (1952). *Preliminaries to speech analysis*. Cambridge, Mass.: MIT Press.

Jarosz, Gaja (2013). Learning with Hidden Structure in Optimality Theory and Harmonic Grammar: Beyond Robust Interpretive Parsing. *Phonology* 30(1). 27–71.

Jarosz, Gaja (2015). Expectation Driven Learning of Phonology. University of Massachusetts, Amherst, ms.

Jun, Sun-Ah, Mary E. Beckman, Seiji Niimi and Mark Tiede (1998). Electromyographic evidence for a gestural-overlap analysis of vowel devoicing in Korean. *UCLA Working Papers in Phonetics* 96. 1-42.

Kawahara, Shigeto (2006). A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language* 82(3). 536–574.

Kawahara, Shigeto (2015). The phonetics of *sokuon*, or geminate obstruents. In Haruo Kubozono (ed.) *The Handbook of Japanese Phonetics and Phonology*. Berlin, München, Boston: De Gruyter Mouton. 43-78.

Kawahara, Shigeto, Hajime Ono and Kiyoshi Sudo (2006). Consonant Cooccurrence Restrictions in Yamato Japanese. *Japanese/Korean Linguistics* 14. Stanford: CSLI. 27-38.

Kawai, Hisashi, Norio Higuchi, Tohru Shimizu and Sei'ichi Yamamoto (1995). Devoicing rules for text-to-speech synthesis of Japanese. *Journal of the Acoustical Society of Japan* 51(9). 698-705.

Kilbourn-Ceron, Oriana and Morgan Sonderegger (2018). Boundary phenomena and variability in Japanese high vowel devoicing. *Natural Language & Linguistic Theory* 36. 175-217.

Kondo, Mariko (1997). *Mechanisms of vowel devoicing in Japanese*. PhD dissertation, University of Edinburgh.

Kondo, Mariko (2001). Vowel devoicing and syllable structure in Japanese. In Mineharu Nakayama and Charles J. Quinn, Jr. (eds.)

Japanese/Korean linguistics 9. Stanford: CSLI. 125-138.

Kubozono, Haruo (2015). Introduction to Japanese phonetics and phonology. In Haruo Kubozono (ed.) *The Handbook of Japanese Phonetics and Phonology*. Berlin, München, Boston: De Gruyter Mouton. 1-40.

Kuznetsova, Alexandra, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen (2017). lmertest: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26.

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

Leben, William (1978). *Suprasegmental Phonology*. PhD dissertation, MIT, Cambridge, Mass.

Lehiste, Ilse (1970). Suprasegmentals. Cambridge, MA: The MIT Press.

Lin, Yen-Hwei (2005). Piro affricates: Phonological edge effects and phonetic anti-edge effects. In Jeroen van de Weijer and Marc van Oostendorp (eds.) *The internal organization of phonological segments*. Berlin & New York: Mouton de Gruyter. 121–151.

Maddieson, Ian (1984). *Patterns of Sounds*. New York: Cambridge University Press.

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

Maekawa, Kikuo, Hanae Koiso, Sadaoki Furui and Hitoshi Isahara (2002). X-JToBI: an extended J_ToBi for spontaneous speech. *Proceedings of the 7th International Conference on Spoken Language Processing* 3: 1545-1548. Denver.

Maekawa, Kikuchi, Hanae Koiso, Sadaoki Furui and Hitoshi Isahara (2000). Spontaneous speech corpus of Japanese. *Proceedings of the 2nd Conference of Language Resources and Evaluation*, Athens. 947-952.

Matsui, Michinao (1993). Auditory scene analysis of fundamental frequency declination [in Japanese]. *Studies in Language and Culture* 2 (Osaka University). 117-124.

McCawley, James (1968). *The phonological component of a grammar of Japanese*. The Hague: Mouton.

McCarthy, John J. (1986). OCP Effects: gemination and antigemination. *Linguistic Inquiry* 17. 207-264.

McCarthy, John and Alan Prince (1995). Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey and Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*. Amherst: GLSA. 249-384.

Nagano-Madsen, Yasuko (1994). Vowel devoicing rates in Japanese from a sentence corpus. *Lund Working Papers in Linguistics* 42. 117-127.

Nielsen, Kuniko Y. (2015). Continuous versus categorical aspects of Japanese consecutive devoicing. *Journal of Phonetics* 52. 70-88.

Nogita, Akitsugu 2016 Arguments that Japanese [Cj]s are complex onsets: durations of Japanese [Cj]s and Russian [C^j]s and blocking of Japanese vowel devoicing. *Working papers of the Linguistics Circle of the University of Victoria* 26. 73-99.

Ohala, John J. (1983). The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.) *The Production of Speech*. New York: Springer-Verlag. 189-216.

Prince, Alan and Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Ms, Rutgers University & University of Colorado, Boulder. Revised version published by Blackwell (2004).

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rose, Sharon and Rachel Walker (2004) A Typology of consonant agreement as correspondence. *Language* 80. 475–531.

Rubach, Jerzy (1994). Affricates as strident stops in Polish. *Linguistic Inquiry* 25(1). 119-143.

Sapir, Edward (1930). Southern Paiute, a Shoshonean language. *Proceedings of the American Academy of Arts and Sciences* 65, nos. 1–3.

Sawashima, Masayuki and Hajime Hirose (1980). Laryngeal gestures in speech production. *Annual Bulletin Research Institute of Logopedics and Phoniatrics* 14. 29-51.

Shaw, Jason and Shigeto Kawahara (2018). The lingual gesture of devoiced [u] in Japanese. *Journal of Phonetics* 66. 100–119.

Shaw, Jason and Shigeto Kawahara (2021). More on the articulation of devoiced [u] in Tokyo Japanese: effects of surrounding consonants. *Phonetica* 78(5/6). 467-513.

Shibatani, Masayoshi (1990). *The languages of Japan.* Cambridge: Cambridge University Press.

Shrosbee, Miki (2013). A study of a blocking factor of Japanese devoicing: comparison of pre-geminate vs. pre-singleton consonant environments. *Phonological Studies* 16. 61-68.

Solé, Maria-Josep and John J. Ohala (2010). What is and what is not under the control of the speaker: intrinsic vowel duration. In Cécile Fougeron, Barbara Kuehnert, Mariapaola D'Imperio, and Nathalie Vallee, (eds.). *Laboratory Phonology 10*, Berlin, New York: De Gruyter.

Stanton, Juliet (2020). Aggressive reduplication and dissimilation in Sundanese. *Phonological Data and Analysis* 2(5). 1-35.

Steriade, Donca (2001). Directional asymmetries in place assimilation: a perceptual account. In Elizabeth Hume and Keith Johnson (eds.) *The role of speech perception in phonology*. San Diego: Academic Press. 219–250.

Steriade, Donca (2009). The phonology of perceptibility effects: the P-map and its consequences for constraint organization. In Kristin Hanson and Sharon Inkelas (eds.) *The nature of the word: studies in honor of Paul Kiparsky*. Cambridge, Mass.: MIT Press. 151–179.

Struijke, Caro (2000). *Existential Faithfulness: A study of reduplicative TETU, featural movement, and dissimilation*. PhD dissertation, University of Maryland at College Park.

Sugito, Miyoko (1982). Study on Japanese accent [in Japanese]. Tokyo: Sanseido.

Suzuki, Keiichiro (1998). *A Typological Investigation of Dissimilation*. PhD dissertation, University of Arizona.

Tsuchida, Ayako (1997). *The Phonetics and Phonology of Japanese Vowel Devoicing*. PhD dissertation, Cornell University.

Tsuchida, Ayako (2001). Japanese vowel devoicing: cases of consecutive devoicing environments. *Journal of East Asian Linguistics* 10(3). 225-245.

Varden, J. Kevin (1998). *On High Vowel Devoicing in Standard Modern Japanese: Implications for Current Phonological Theory*. PhD dissertation, University of Washington.

Whang, James (2018). Recoverability-driven coarticulation: Acoustic evidence from Japanese high vowel devoicing. *Journal of the Acoustical Society of America*, 143(2), 1159-1172.

Wilbur, Ronnie (1973). *The Phonology of Reduplication*. PhD dissertation, University of Illinois, Urbana-Champaign.

Yoshida, Natsuya (2002). The effects of phonetic environment on vowel devoicing in Japanese [in Japanese]. *Kokugogaku* 53(3). 34-47.

Yoshioka, Hirohide (1981). Laryngeal adjustment in the production of the fricative consonants and devoiced vowels in Japanese. *Phonetica* 38. 236-351.

Yoshioka, Hirohide, Anders Löfqvist and Hajime Hirose (1982). Laryngeal adjustment in Japanese voiceless sound production. *Journal of Phonetics* 10. 1–10.

Zuraw, Kie (2002). Aggressive reduplication. *Phonology* 19. 195-439.

국문 초록

일본어 고모음 무성음화에서 나타나는 적극 중첩

본고는 일본어 고모음 무성음화의 변이 양상에 기여하는 음운론적인 요소를 탐구하고자 한다. 고모음 무성음화란, 무성 자음 사이에서 고모음이 유성성을 잃는 현상으로, 이는 종종 음성학적인 현상으로 간주되어 왔다. 예컨대 Jun et al. (1998)과 같은 기존 연구에 따르면, 고모음 무성음화는 고모음을 둘러싼 무성 자음의 성문 개방 동작이 겹쳐짐에 따라 고모음의 유성성이 실현되지 못하는 것으로 이해되었다. 이러한 음성학적 설명은 고모음 무성음화의 실현이 오로지 주변 자음의 조음적 특성 및 대상 모음과 주변 자음 사이의 시간적 배열에 따라 결정될 것임을 가정한다. 따라서, 이러한 설명은 일본어 고모음 무성음화가 일본어의 음운론적 구조와 무관할 것임을 예측한다.

이러한 예측과는 반대로, 본고는 인접 음절 간의 유사성에 대한 구조적 지식이 고모음 무성음화의 실현에 중요한 역할을 한다는 것을 보인다. 일본어 발화 코퍼스(Corpus of Spontaneous Japanese; Maekawa et al. 2000)에서 추출한 대규모의 자료를 기반으로, 본고는 고모음 무성음화의 빈도에 영향을 주는 음운론적 경향성을 다음과 같이 밝혀내었다.

(i) 무성음화는 마찰음이나 파찰음이 대상 모음에 선행하고, 마찰음이 후행할 때 저지된다.(조음 양식의 일치 조건)

(ii) 조음 양식의 일치 조건 하에서, 무성음화는 대상 모음에 후행하는 음절이 또다른 고모음을 포함할 때 더욱 더 저지된다.(모음 높이의 일치 조건)

(iii) 무성음화는 대상 모음이 액센트를 지닐 때 덜 일어난다.

(iv) 무성음화는 중자음이 대상 모음에 후행할 때 덜 일어난다.

(v) 무성음화 환경이 연속적으로 일어나는 환경에서, 인접 음절의 두 모음이 모두 무성음화되는 것은 회피되는 경향이 있다.

본고는 위의 경향성들을 최적성이론(Optimality Theory; Prince and Smolensky 1993/2004)에 입각하여 형식적으로 분석하고자 한다. 먼저, 본고는 무성음화가 변이음적 교체를 산출해내는 일반적인 제약의 위계(ranking)에 의해 도출된다고 주장한다. 즉, 이는 환경의존적 유표성(context-sensitive markedness) 제약 (**DEVOICE**: "무성 자음 사이에서 짧은 유성 고모음을 금함") >> 환경독립적 유표성(context-free markedness) 제약(**SONVOI**: "무성 공명음을 금함") >> 충실성(faithfulness) 제약(**IO-ID(vce)**: "입력형의 [유성성]([voice]) 값을 출력형에서 바꾸는 것을 금함")으로 나타난다. 추가적으로, 연속 무성음화가 회피되는 경향성을 설명하기 위해 **OCP-V̥** ("인접 음절 내의 두 무성 모음을 금함")을 제안한다.

위의 경향성 중, 조음 양식과 모음 높이의 일치 조건은 음절 구조를 참조하여 분석될 수 있다. V_1이 무성음화의 대상이 되는 고모음이고 C_1과 C_2가 무성 자음이라고 할 때, /$C_1V_1C_2V_2$/ 연속은 일본어에서 $[C_1V_1]_{\sigma 1}[C_2V_2]_{\sigma 2}$ 의 두 음절로 나뉜다. 이를 바탕으로, 무성음화는 인접 음절의 두음(onset), 즉 C_1과 C_2가 [+지속성]([+continuant])로 일치할 때 저지된다고 할 수 있다. ((예) *suso* '옷자락') 파찰음은 위치에 따라 다른 행동을 보이는데, 모음에 선행하는 환경(즉, C_1)에서는 마찰음과 함께 [+지속성]로 행동하고 모음에 후행하는 환경(즉, C_2)에서는 파열음과 함께 [-지속성]로 행동한다. 조음 양식의 일치 조건에 더하여, 무성음화는 인접 음절의 음절핵(nuclei), 즉 V_1과 V_2의 [고설성]([high]) 자질이 일치할 때 더욱 더 저지된다고 할 수 있다. 조음 양식과 모음 높이 간의 부가적인 효과는 인접 음절 간의 유사도가 높아질수록 무성음화 빈도가

낮아짐을 시사한다.

본고는 위와 같이 유사성에 의한 저지 효과가 인접하고 서로 유사한 음절 내에 있는 모음 간의 유성성 일치를 보존하기 위해 일어난다고 주장한다. 이러한 주장을 형식화하기 위해, 본고는 Zuraw(2002)의 적극 중첩(Aggressive Reduplication)을 차용한 분석을 제시한다. McCarthy and Prince(1995)에서 어기와 중첩어 사이의 동일성이 대응(correspondence)에 의해 발생한다고 주장되는 것처럼, 적극 중첩 이론은 단어 내 하위 연쇄(substring)간의 대응이 REDUP 이라는 제약에 의해 부과되고, REDUP 에 의해 촉발된 대응 제약들(κκ-CORR)이 대응 연쇄 간의 상호유사성이 떨어지는 것을 막는다고 주장한다. 이 주장에 따라, 본고는 유사성에 의한 저지 효과가 인접 음절 간에서 작용하는 대응 제약들이 무성음화가 일어났을 때 발생하는 대응 모음 간의 유성성 불일치를 비선호하기 때문에 일어난다고 설명한다. 나아가서, κκ-CORR 제약이 REDUP 제약보다 상위에 놓이게 되면, 대응 구조가 상호유사한 음절에만 놓이게 되고, 따라서 무성음화가 이러한 환경에서만 저지되는 결과를 도출해 낸다. 본고는 유사성과 관련한 대안적 이론인 의무 굴곡 원리(Obligatory Contour Principle; McCarthy 1986)와 대응을 통한 일치 이론(Agreement by Correspondence; Rose and Walker 2004)을 살펴보고, 오로지 적극 중첩 이론만이 분절음보다 더 큰 단위의 대응을 허용하기 때문에 일본어 고모음화에서 나타나는 유사성에 의한 저지 효과를 성공적으로 설명할 수 있다고 주장한다.

본고는 고저 액센트와 중자음의 효과를 설명하기 위한 추가적인 제약들 또한 제시한다. 액센트를 지닌 음절의 심리언어학적, 청취적 현저성을 바탕으로, 액센트를 지닌 음절이 더 큰 위치적 충실성(positional faithfulness; Beckman 1998)을 요구하기에 무성음화를 저지한다고 보고, 이러한 위치적 충실성을 반영하여 (IO-ID(vce)/σ́(“액센트를 지닌 음절 내의 입력형 [유성성] 값을 출력형에서 바꾸는 것을 금함”)을 제시한다. 중자음이 무성음화를 저지하는 효과는 일본어에서 중자음에 선행하는 모음이 길이가 길어진다는 사실(Kawahara 2015)로부터 기인한다고 주장

한다. 청취적 사상 가설(P-map hypothesis; Steriade 2001, 2009)에 입각하여, 유성성의 청취적 차이가 짧은 모음과 그것의 무성음 짝 사이보다 장음화된 모음과 그것의 무성음 짝 사이에서 더 클 것이라고 가정한다. 이에 따라 또다른 위치적 충실성 제약인 IO-ID(vce)/_GEM("중자음에 선행하는 모음에 대해, 입력형의 [voice] 값을 출력형에서 바꾸는 것을 금함")을 제시한다.

마지막으로, 고모음 무성음화에서 나타나는 변이를 설명하기 위해, 본고는 Anttila(1997)의 부분 위계 이론(Partially Ordered Constraints)을 차용한다. 이에 입각하여, 제약의 일부 위계는 고정되어 있지만, 다른 부분은 매 산출 시마다 바뀔 수 있다고 가정한다.

본고는 음절 간의 유사성이라는 구조적 지식이 일본어 고모음 무성음화에 영향을 준다는 것을 발견하고, 이는 기존에 이 현상이 "음성적"이거나 "어휘부 외적(post-lexical)"이라는 주장과 상충됨을 보인다. 본고의 발견은 일본어 화자들에게 모음의 유성성에 대한 정보가 순전히 변이음적일지라도, 화자들이 모음 간의 유성성이 서로 같거나 다른지에 대한 지식을 가지고 있음을 시사한다.

주요어 : 일본어 고모음 무성음화, 적극 중첩, 유사성, 대응, 변이

학번: 2018-29432